TheoretiCS

Upper and Lower Bounds on the Smoothed Complexity of the Simplex Method

Received May 16, 2024
Revised Apr 30, 2025
Accepted Sep 1, 2025
Published Oct 15, 2025

Key words and phrases
simplex method, linear programming, smoothed analysis, beyond worst case analysis, extended formulation

Sophie Huiberts [a,b]
Yin Tat Lee [c,d]
Xinzhi Zhang [c]

a CNRS / LIMOS, Clermont-Ferrand, France

b Université Clermont Auvergne, Clermont-Ferrand, France

c University of Washington, Seattle, USA

d Microsoft Research, Redmond, USA

ABSTRACT. The simplex method for linear programming is known to be highly efficient in practice, and understanding its performance from a theoretical perspective is an active research topic. The framework of smoothed analysis, first introduced by Spielman and Teng (JACM '04) for this purpose, defines the smoothed complexity of solving a linear program with d variables and n constraints as the expected running time when Gaussian noise of variance σ^2 is added to the LP data. We prove that the smoothed complexity of the simplex method is $O(\sigma^{-3/2} d^{13/4} \log^{5/4} n)$, improving the dependence on $1/\sigma$ compared to the previous bound of $O(\sigma^{-2} d^2 \sqrt{\log n})$. We accomplish this through a new analysis of the *shadow bound*, key to earlier analyses as well. Illustrating the power of our new approach, we moreover prove a nearly tight upper bound on the smoothed complexity of two-dimensional polygons.

We also establish the first non-trivial lower bound on the smoothed complexity of the simplex method, proving that the *shadow vertex simplex method* requires, with a given auxiliary objective, at least $\Omega\Big(\min\big(\sigma^{-1/2} d^{-1/2} \log^{-1/4} d, 2^d\big)\Big)$ pivot steps with high probability. A key part of our analysis is a new variation on the extended formulation for the regular 2^k-gon. We end with a numerical experiment that suggests our lower bound could be further improved.

A preliminary version of this article appeared in STOC 2023 [35].
Sophie Huiberts' work was supported by grants from the Simons Foundation (966674) and ANR (ANR-24-CE48-2762). Xinzhi Zhang's research supported by NSF grant CCF-1813135, NSF CAREER Award and Packard Fellowships.

Cite as Sophie Huiberts, Yin Tat Lee, Xinzhi Zhang. Upper and Lower Bounds on the Smoothed Complexity of the Simplex Method. TheoretiCS, Volume 4 (2025), Article 23, 1-56.

https://theoretics.episciences.org
DOI 10.46298/theoretics.25.23

1. Introduction

Introduced by Dantzig [20], the simplex method is one of the primary methods for solving linear programs (LP's) in practice and is an essential component in many software packages for combinatorial optimization. It is a family of local search algorithms which begin by finding a vertex of the set of feasible solutions and iteratively move to a better neighboring vertex along the edges of the feasible polyhedron until an optimal solution is reached. These moves are known as *pivot steps*. Variants of the simplex method can be differentiated by the choice of *pivot rule*, which determines which neighboring vertex is chosen in each iteration, as well as by the method for obtaining the initial vertex. Some well-known pivot rules are the most negative reduced cost rule, the steepest edge rule, and an approximate steepest edge rule known as the Devex rule. In theoretical work, the parametric objective rule, also known as the shadow vertex rule, plays an important role.

Empirical evidence suggests that the simplex algorithm typically takes $O(d+n)$ pivot steps, see [47, 5, 29]. However, obtaining a rigorous explanation for this excellent performance has proven challenging. In contrast to the practical success of the simplex method, all studied variants are known to have super-polynomial or even exponential worst-case running times. For deterministic variants, many published bad inputs are based on *deformed cubes*, see [40, 36, 6, 32, 43, 30] and a unified construction in [4]. For randomized and history-dependent variants, bad inputs have been constructed based on *Markov Decision Processes* [38, 41, 25, 24, 34, 23]. The fastest provable (randomized) simplex algorithm takes $O(2^{\sqrt{d \log n}})$ pivot steps in expectation [38, 41, 34].

Average-case analyses of the simplex method have been performed for a variety of random distributions over linear programs [14, 13, 12, 48, 33, 42, 1, 50, 2]. While insightful, the results from average-case analyses might not be fully realistic due to the fact that "random" linear programs tend to have certain properties that "typical" linear programs do not.

To better explain why the simplex algorithm performs well in practice, while avoiding some of the pitfalls of average-case analysis, Spielman and Teng [49] introduced the smoothed complexity framework. For any base LP data $\bar{A} \in \mathbb{R}^{n \times d}, \bar{b} \in \mathbb{R}^n, c \in \mathbb{R}^d \setminus \{0\}$ where the rows of $(\bar{A}, \bar{b})$ are normalized to have ℓ_2 norm at most 1, they consider the *smoothed LP* obtained by adding independent Gaussian perturbations to the constraints:

$$\max_{x \in \mathbb{R}^d} c^\top x \quad \text{subject to} \quad (\bar{A} + \hat{A})x \leq (\bar{b} + \hat{b}).$$

The entries of $\hat{A}$ and $\hat{b}$ are i.i.d. Gaussian random variables with mean 0 and variance σ^2. The *smoothed complexity* of a simplex algorithm $\mathcal{A}$ is defined to be the maximum (over $\bar{A}, \bar{b}, c$)

expected number of pivot steps the algorithm takes to solve the smoothed LP, i.e.,

$$\mathcal{SC}_{\mathcal{A},n,d,\sigma} := \max_{\substack{\bar{A}\in\mathbb{R}^{n\times d},\bar{b}\in\mathbb{R}^n,c\in\mathbb{R}^d \\ \|[\bar{A},\bar{b}]\|_{1,2}\le 1}} \left(\mathbb{E}_{\hat{A},\hat{b}}\left[T_{\mathcal{A}}(\bar{A}+\hat{A},\bar{b}+\hat{b},c)\right]\right).$$

Here $T_{\mathcal{A}}(A,b,c)$ is the number of pivot steps that the algorithm $\mathcal{A}$ takes to solve the linear program $\max_{x\in\mathbb{R}^d}\{c^\top x : Ax \le b\}$. We may note that if $\sigma\to\infty$ then $\mathcal{SC}_{\mathcal{A},n,d,\sigma}$ approaches the average-case complexity of $\mathcal{A}$ on independent Gaussian distributed input data. In contrast, if $\sigma\to 0$ then $\mathcal{SC}_{\mathcal{A},n,d,\sigma}$ will approach the worst-case complexity of $\mathcal{A}$. As a result, most interest has been directed at understanding the dependence on σ in the regime where $\sigma \ge 2^{-\Omega(d)}$ but $\sigma \le 1/\mathrm{poly}(d)$.

The motivation for smoothed analysis lies in the observation that the above-mentioned worst-case instances are very "brittle" to perturbations, and computer implementations require great care in handling numerical inaccuracies to obtain the theorized running times even on problems with a small number of variables. When implemented with a larger number of variables, the limited accuracy of floating-point numbers make it impossible to reach the theorized running times.

An algorithm is said to have polynomial smoothed complexity if under the perturbation of constraints, it has expected running time $\mathrm{poly}(n,d,\sigma^{-1})$, and [49] proved that the smoothed complexity of a specific simplex method based on the shadow vertex simplex method (which we will describe next) is at most $O(d^{55}n^{86}\sigma^{-30}+d^{70}n^{86})$. The best bound available in the literature is $O(\sigma^{-2}d^2\sqrt{\log n})$ pivot steps due to [17], assuming $\sigma \le 1/\sqrt{d\log n}$. We note that assuming an upper bound on σ can be done without loss of generality; its influence can be captured as an additive term in the upper bound that does not depend on σ.

This work improves the dependence on σ of the smoothed complexity, obtaining an upper bound of $O(\sigma^{-3/2}d^{13/4}\log^{5/4} n)$ for $\sigma \le 1/d\sqrt{\log n}$. As a second contribution, we prove the first non-trivial lower bound on the smoothed complexity of a simplex method, finding that the shadow vertex simplex method requires $\Omega(\min(\frac{1}{\sqrt{\sigma d\sqrt{\log n}}}, 2^d))$ pivot steps.

Shadow Vertex Simplex Algorithm One of the most extensively studied simplex algorithms in theory is the shadow vertex simplex algorithm [27, 14]. Given an LP

$$\max_{x\in\mathbb{R}^d} c^\top x, \quad Ax \le b,$$

for $A\in\mathbb{R}^{n\times d}, b\in\mathbb{R}^n, c\in\mathbb{R}^d$, let $P=\{x\in\mathbb{R}^d : Ax\le b\}$ denote the feasible polyhedron of the linear program. The algorithm starts from an initial vertex $x_0\in P$ that optimizes an initial objective c_0[1]. During the execution, it maintains an intermediate objective $c_\lambda = \lambda c + (1-\lambda)c_0$ and a vertex that optimizes c_λ. Thus by slowly increasing λ from 0 to 1 during different pivot

1 There are many standard methods of finding such initialization with at most multiplicative $O(d)$ overhead in running time, so we can assume that both x_0 and c_0 are already given. See the discussion in [17].

steps, the temporary objective gradually changes from c_0 to c, revealing the desired solution at the end. Since each pivot step requires $\mathrm{poly}(d, n)$ computational work, theoretical analysis has focused on analysing the number of pivot steps.

The algorithm is called the shadow vertex simplex method because, when performing orthogonal projection of the feasible set onto the two-dimensional linear subspace $W = \mathrm{span}(c_0, c)$, the vertices visited by the algorithm project onto the boundary of the projection ("shadow") $\pi_W(P)$. Assuming certain non-degeneracy conditions, which will hold with probability 1 for the distributions we consider, this projection gives an injective map from iterations of the method to vertices of the shadow, meaning that we can upper bound the number of pivot steps in the algorithm by the number of vertices of the shadow polygon. This characterization makes the shadow vertex simplex method ideally suited for probabilistic analysis.

To analyse the "shadow size", the number of vertices of the shadow polygon, we follow earlier work of [52] and reduce to the case that $b = \mathbf{1}_n$. In this case, well-established principles of polyhedral duality show that we can bound the number of vertices of a convex polygon by the number of edges of its dual polygon:

$$\mathrm{vertices}(\pi_W(P)) \le \mathrm{edges}(W \cap \mathrm{conv}(0, a_1, \dots, a_n)) \le \mathrm{edges}(W \cap \mathrm{conv}(a_1, \dots, a_n)) + 1.$$

Here, $a_1, \dots, a_n$ denote the rows of the matrix A used to define $P = \{x \in \mathbb{R}^d : Ax \le \mathbf{1}_n\}$. Note that the second inequality holds because $0 \in W$.

The smoothed complexity of shadow vertex simplex algorithm can thus be reduced to the smoothed complexity of a two-dimensional slice of a convex hull. For this reason, let us define the maximum smoothed shadow size as

$$\mathcal{S}(n, d, \sigma) = \max_{\substack{\bar{a}_1,\dots,\bar{a}_n \in \mathbb{R}^d \\ \max_{i\in[n]} \|\bar{a}_i\|_2 \le 1 \\ W \subseteq \mathbb{R}^d, \dim(W)=2}} \mathbb{E}_{\tilde{a}_1,\dots,\tilde{a}_n \sim \mathcal{N}(0,\sigma^2)} \left[\mathrm{edges}\left(\mathrm{conv}(\bar{a}_1 + \tilde{a}_1, \dots, \bar{a}_n + \tilde{a}_n) \cap W\right)\right] \tag{1}$$

The following upper bound we take from [17], which states that the analysis of [52] can be strengthened to obtain the claimed bound. This upper bound should be understood as stating that there exists a shadow vertex rule based simplex algorithm which satisfies that smoothed complexity bound. The lower bound is due to [13] and shows that the shadow vertex simplex rule, with two adversarially specified objectives c_0, c, can be made to follow paths of this length.

THEOREM 1.1 (Smoothed Complexity of Shadow Vertex Simplex Algorithm). *Given any $n \ge d \ge 2, \sigma > 0$, the smoothed complexity of the shadow vertex simplex algorithm satisfies*

$$\mathcal{S}(n, d, \sigma)/4 \le \mathcal{SC}_{\textit{ShadowSimplex},n,d,\sigma} \le 2 \cdot \mathcal{S}\left(n + d, d, \min(\sigma, \frac{1}{\sqrt{d}\log d}, \frac{1}{\sqrt{d \log n}})\right) + 4.$$

With this reduction, analysing the smoothed complexity of the simplex method comes down to bounding the smoothed shadow size $\mathcal{S}(n, d, \sigma)$. As such, that will be the focus of the remainder of this paper.

1.1 Our Results

The previous best shadow bound is due to [17], who prove that $\mathcal{S}(n, d, \sigma) \leq O(d^2\sqrt{\log n}\sigma^{-2})$. Our first main result strengthens this result for small values of σ.

THEOREM 1.2. *For $n \geq d \geq 3$ and $\sigma \leq \frac{1}{8d\sqrt{\log n}}$, the smoothed shadow size satisfies*

$$\mathcal{S}(n, d, \sigma) = O\left(\sigma^{-3/2} d^{13/4} \log^{5/4} n\right).$$

A full overview of bounds on the smoothed shadow size, including previous results in the literature, can be found in Table 1.

Second, we prove the first non-trivial lower bound on the smoothed shadow size, establishing that $\mathcal{S}(4d - 13, d, \sigma) \geq \Omega(\min(\frac{1}{\sqrt{\sigma d\sqrt{\log d}}}, 2^d))$ for $d > 5$. This lower bound is proven by constructing a polyhedron $P = \{x \in \mathbb{R}^d : Ax \leq \mathbf{1}_n\}$ and a two-dimensional subspace W such that for any small perturbation of A, the new polyhedron P projected onto W will have many vertices. The construction is based on an extended formulation similar to those first constructed by [8, 28].

THEOREM 1.3. *For any $d > 5$ and $\sigma \leq \frac{1}{360d\sqrt{\log(4d)}}$, the smoothed shadow size satisfies*

$$\mathcal{S}(4d - 13, d, \sigma) = \Omega\left(\min\left(\frac{1}{\sqrt{d\sigma\sqrt{\log d}}}, 2^d\right)\right).$$

It is possible that the exponent of σ in our bound can be further optimized. In Section 6.6, we describe numerical experiments which suggest that the actual shadow size for random perturbations of our constructed polytope might be as high as $\Omega(\min(\sigma^{-3/4}, 2^d))$. We leave open the question whether having $n/d > 4$ can lead to stronger lower bounds in the regime of $\sigma < 2^{-d}$.

Two-dimensional polygons To better understand the smoothed complexity of the intersection polygon $\mathrm{conv}(a_1, \dots, a_n) \cap W$, we also analyse its two-dimensional analogue introduced by [19]. Taking $\bar{a}_1, \dots, \bar{a}_n \in \mathbb{R}^2$, each satisfying $\|\bar{a}_i\|_2 \leq 1$, we are interested in the number of edges of the smoothed polygon $\mathrm{conv}(\bar{a}_1 + \hat{a}_1, \dots, \bar{a}_n + \hat{a}_n)$, where $\hat{a}_1, \dots, \hat{a}_n \sim N(0, \sigma^2)$ are independent. The previous best upper bound on the smoothed complexity of this polygon is $O(\sigma^{-1} + \sqrt{\log n})$, due to [18]. Their analysis is based on an adaptation of the shadow bound by [17]. In Section 4 we improve this upper bound, obtaining the following theorem.

THEOREM 1.4 (Two-Dimensional Upper Bound). *Let $\bar{a}_1, \dots, \bar{a}_n \in \mathbb{R}^2$ be $n > 2$ vectors with norm at most 1. For each $i \in [n]$, let a_i be independently distributed as $\mathcal{N}(\bar{a}_i, \sigma^2 I_2)$. Then*

$$\mathbb{E}\left[\mathrm{edges}\left(\mathrm{conv}(a_1, \dots, a_n)\right)\right] \leq O\left(\sqrt{\log n} + \frac{\sqrt[4]{\log n}}{\sqrt{\sigma}}\right).$$

Reference	Shadow size	Model
[13]	$\Theta(d^{3/2}\sqrt{\log n})$	Average-case, Gaussian
[49]	$O(\sigma^{-6}d^3 n + d^6 n \log^3 n)$	Smooth
[21]	$O(\sigma^{-2}dn^2 \log n + d^2n^2 \log^2 n)$	Smooth
[52]	$O(\sigma^{-4}d^3 + d^5 \log^2 n)$	Smooth
[17]	$O(\sigma^{-2}d^2\sqrt{\log n} + d^3 \log^{1.5} n)$	Smooth
This paper	$O(\sigma^{-3/2}d^{13/4} \log^{5/4} n + d^{19/4} \log^2 n)$	Smooth
This paper	$\Omega(\min(\frac{1}{\sqrt{\sigma d\sqrt{\log d}}}, 2^d))$	Smooth
[53, 43, 31]	2^d	Worst

Table 1. Bounds of expected number of pivots in previous literature, assuming $d \geq 3$. Logarithmic factors are simplified. The lower bound of [13] holds in the smoothed model as well.

To confirm that the above upper bound is stronger than that of [18], one may use the AM-GM inequality to verify that $2\frac{\sqrt[4]{\log n}}{\sqrt{\sigma}} \leq \sqrt{\log n} + \sigma^{-1}$. Combined with the trivial upper bound of n vertices, our bound nearly matches the lower bound of $\Omega(\min(\sqrt{\log n} + \frac{\sqrt[4]{\log(n\sqrt{\sigma})}}{\sqrt{\sigma}}, n))$ proven by [22]. A full overview of previous results on the smoothed complexity of the two-dimensional convex hull can be found in Table 2.

1.2 Related work

Shadow Vertex Simplex Method The shadow vertex simplex algorithm has played a key role in many analyses of simplex and simplex-like algorithms. On well-conditioned polytopes, such as those of the form $\{x \in \mathbb{R}^d : Ax \leq b\}$ where A is integral with subdeterminants bounded by Δ, the shadow vertex method has been studied by in [15, 16]. The shadow vertex method on polytopes all whose vertices are integral was studied in [9, 10]

On random polytopes of the form $\{x \in \mathbb{R}^d : Ax \leq \mathbf{1}_n\}$, assuming the constraint vectors are independently drawn from any rotationally symmetric distribution, the expected iteration complexity of the shadow vertex simplex method was studied by [14, 13, 12]. In the case when the rows of A arise from a Poisson distribution on the unit sphere, concentration results for the

Reference	Smoothed polygon complexity
[19]	$O(\log(n)^2 + \sigma^{-2}\log n)$
[45]	$O(\log n + \sigma^{-2})$
[22]	$O(\sqrt{\log n} + \sigma^{-1}\sqrt{\log n})$
[18]	$O(\sqrt{\log n} + \sigma^{-1})$
This paper	$O(\sqrt{\log n} + \frac{\sqrt[4]{\log(n)}}{\sqrt{\sigma}})$
[22]	$\Omega(\min(\sqrt{\log n} + \frac{\sqrt[4]{\log(n\sqrt{\sigma})}}{\sqrt{\sigma}}, n))$

Table 2. Bounds on the smoothed complexity of a two-dimensional polygon.

shadow size and diameter bounds were proven in [11]. The diameter of smoothed polyhedra was studied by [44], who used the shadow bound of [17] to show that most vertices, according to some measure, are connected by short paths.

A randomized algorithm for solving linear programs in weakly polynomial time, using the shadow vertex simplex method as a subroutine, was proposed in [39]. The shadow vertex algorithm was recently used as part of the analysis of an interior point method by [3].

Extended Formulations For a polyhedron $P \subseteq \mathbb{R}^d$, an *extended formulation* is any polyhedron $Q \subseteq \mathbb{R}^{d'}$, $d' \geq d$, such that P can be obtained as the orthogonal projection of Q to some d-dimensional subspace. Importantly, Q can have much fewer facets than P. While there is a wider literature on extended formulations, here we describe only what is most relevant to the construction in Section 6.

The construction in our lower bound is based on an adaptation of the extended formulation by [8] of the regular 2^k-gon. They used this construction to obtain a polyhedral approximation of the second order cone $\{x \in \mathbb{R}^{n+1} : \sum_{i=1}^n x_i^2 \leq x_{n+1}^2\}$. A variant on their construction using fewer variables and inequalities was given by [28]. A more general construction based on *reflection relations* is used to construct extended formulation for the regular 2^k-gon, as well as other polyhedra, in [37]. Extended formulations for regular n-gons, when n is not a power of 2, can be found in [51].

Approximations of the second order cone based on the work of [8, 28] have been used to solve second order conic programs, see, e.g., [7]. These approximations were included in the open-source MIP solver SCIP until version 7.0 [26].

1.3 Proof Overview

1.3.1 Smoothed Complexity Upper Bound

We write the random polytope from (1) as $Q = \mathrm{conv}(a_1, \dots, a_n)$ where each a_i is sampled independently from $\mathcal{N}_d(\bar{a}_i, \sigma^2 I)$ such that $\|\bar{a}_i\| \le 1$. Our goal is to upper bound the expected number of edges of the polygon $Q \cap W$ for fixed two-dimensional plane $W \subseteq \mathbb{R}^d$ and $\bar{a}_1, \dots, \bar{a}_n$. This will immediately give us an upper bound of $\mathcal{S}(n, d, \sigma)$.

A fact used since the early days of smoothed analysis [49] states that the intersection polygon $Q \cap W$ is *non-degenerate* with probability measure 1: every edge on $Q \cap W$ is uniquely given by the intersection between W and a facet of Q, and every facet of Q is spanned by exactly d vertices. For any index set $I \in \binom{[n]}{d}$, write E_I as the event that $\mathrm{conv}(a_i : i \in I) \cap W$ is an edge of $Q \cap W$. Non-degeneracy implies that every edge of $Q \cap W$ uniquely corresponds to an index set $I \in \binom{[n]}{d}$ such that E_I holds.

Before sketching our proof, we first review the approach of [17], then discuss the main technical challenges in achieving an upper bound with better dependence on σ.

As a first step in [17], the authors replace the Gaussian distribution with a distribution they call the Laplace-Gaussian distribution[2]. The latter distribution approximates the probability density of the former, in particular having nearly equivalent smoothed shadow size, while being $O(\sigma^{-1}\sqrt{d \log n})$-log-Lipschitz for any point on its domain. A probability distribution with probability density function μ is L-log-Lipschitz for some $L > 0$ if, for any $x, y \in \mathbb{R}^d$, the condition $|\log(\mu(x)) - \log(\mu(y))| \le L\|x - y\|$ holds.

Next, define ℓ_I as the length of the edge on $Q \cap W$ that corresponds to I, i.e., when $\mathrm{conv}(a_i : i \in I) \cap W$ is an edge of $Q \cap W$ then ℓ_I gives the length of this line segment, and otherwise $\ell_I = 0$. [17] showed that, for any family $S \subseteq \binom{[n]}{d}$, the expected number of edges of $Q \cap W$ coming from S is at most

$$\mathbb{E}\left[\sum_{I \in S} 1[E_I]\right] \le \frac{\mathbb{E}[\mathrm{perimeter}(Q \cap W)]}{\min_{I \in S} \mathbb{E}[\ell_I \mid E_I]}. \tag{2}$$

Therefore, by taking $S = \{I \in \binom{[n]}{d} : \Pr[E_I] \ge \binom{n}{d}^{-1}\}$, the expected number of edges of $Q \cap W$ is at most

$$\mathbb{E}[\mathrm{edges}(Q \cap W)] = \mathbb{E}\left[\sum_{I \in \binom{[n]}{d}} 1[E_I]\right] \le 1 + \mathbb{E}\left[\sum_{I \in S} 1[E_I]\right] \le 1 + \frac{\mathbb{E}[\mathrm{perimeter}(Q \cap W)]}{\min_{I \in S} \mathbb{E}[\ell_I \mid E_I]}. \tag{3}$$

To upper bound the numerator of (3), notice that $Q \cap W$ is a convex polygon contained in the two-dimensional disk centered at $\mathbf{0}$ with radius $\max_{i \in [n]} \|\pi_W(a_i)\|$. It follows that the

2 The Laplace-Gaussian distribution follows the Gaussian probability density function near its mean but exhibits exponentially decaying tails.

perimeter of $Q \cap W$ is at most the perimeter of this disk, namely,

$$\mathbb{E}[\text{perimeter}(Q \cap W)] \leq 2\pi \cdot \mathbb{E}[\max_{i \in [n]} \|\pi_W(a_i)\|] \leq 2\pi \cdot (1 + 4\sigma\sqrt{\log n}), \tag{4}$$

where the last step comes from a Gaussian tail bound. For the denominator of (3), [17] showed that for any $I \in \binom{[n]}{d}$ with $\Pr[E_I] \geq \binom{n}{d}^{-1}$ that, conditional on E_I, the expected edge length is at least

$$\mathbb{E}[\ell_I \mid E_I] \geq \Omega(\frac{\sigma^2}{d^2\sqrt{\log n}} \cdot \frac{1}{1 + \sigma\sqrt{d \log n}}) \tag{5}$$

By combining (3), (4) and (5), they proved an upper bound of $O(\sigma^{-2} d^2 \sqrt{\log n} + d^3 \log^{1.5} n)$.

New Strategy for Counting Edges While [17] made the best possible analysis based of their edge-counting strategy as outlined in (3), the strategy itself is sub-optimal. The main drawback is that using the minimum expected length of edge $\min_{I \in \binom{[n]}{d}} \mathbb{E}[\ell_I \mid E_I]$ at the denominator of (3) is too pessimistic when the edges of $Q \cap W$ are long. For instance, consider the case where an edge on $Q \cap W$ has length $\Omega(1)$ without perturbation. After the perturbation, is very likely that the length of this edge is still $\Omega(1)$, but [17] uses a lower-bound of $\Omega(\frac{\sigma^2}{d^2\sqrt{\log n}})$.

To improve this, we developed a new edge-counting strategy that can handle the long and short edges separately with two different ways of counting the edges. Take any index set $I \in \binom{[n]}{d}$; conditional on E_I, we write e_I for its edge $\text{conv}(a_i : i \in I) \cap W$. We call the next edge in clockwise direction as e_{I^+} and say it has length ℓ_{I^+}. We say e_{I^+} is *likely to be long* if $\Pr[\ell_{I^+} \geq t \mid E_I] \geq 0.05$ for some parameter $t > 0$ to be specified later. Here 0.05 can be replaced by some arbitrary constant in $(0, 0.5)$. Let $S_0 \subseteq \binom{[n]}{d}$ denote the set of $I \in \binom{[n]}{d}$ such that e_{I^+} is likely to be long. Following (2), the expected number of edges in $Q \cap W$ that are likely to be long is at most

$$\mathbb{E}[\sum_{I \in S_0} 1[E_I]] \leq \frac{\mathbb{E}[\text{perimeter}(Q \cap W)]}{0.05t} \leq \frac{(2\pi + O(\sigma\sqrt{\log n}))}{0.05t}. \tag{6}$$

Here the second step uses the exact same upper bound of $\mathbb{E}[\text{perimeter}(Q \cap W)]$ as in (4).

In the other case, e_{I^+} is *unlikely to be long*, i.e., $\Pr[\ell_{I^+} \geq t \mid E_I] < 0.05$. Now we will upper bound the number of such edges by claiming that their exterior angles each are large in expectation. Let θ_I be the exterior angle at the endpoint of $\text{conv}(a_i : i \in I) \cap W$ that comes last in clockwise order, i.e., the vertex where e_I and e_{I^+} meet. Our key observation is that $\sin(\theta_I) \cdot \ell_{I^+}$ equals the distance from the affine hull of e_I to the second vertex of e_{I^+} in clockwise order. Therefore, by establishing a universal lower bound for such line-to-vertex distances, we can derive a lower bound for $E[\theta_I \mid E_I]$ for any edge e_{I^+} that is unlikely to be long. See Figure 1 for an illustration.

More formally, let p_{I^+} denote the next vertex of $Q \cap W$ after e_I in clockwise order.

Let $S_1 \subseteq \binom{[n]}{d}$ be the collection of index sets $I \in \binom{[n]}{d}$ for which $\Pr[E_I] \geq 10\binom{n}{d}^{-1}$. For a specific value of $\gamma > 0$, suppose that for each $I \in S_1$ we have

$$\Pr[\mathrm{dist}(p_{I^+}, \mathrm{affhull}(e_I)) \geq \gamma \mid E_I] \geq 0.1. \tag{7}$$

Then, for $I \in S_1 \setminus S_0$, conditional on E_I, we have $\theta_I \geq \sin(\theta_I) \geq \frac{\gamma}{t}$ with probability at least 0.05, and the expectation of the exterior angle at the shared endpoint of e_I and e_{I^+} is at least $\frac{\gamma}{20t}$.

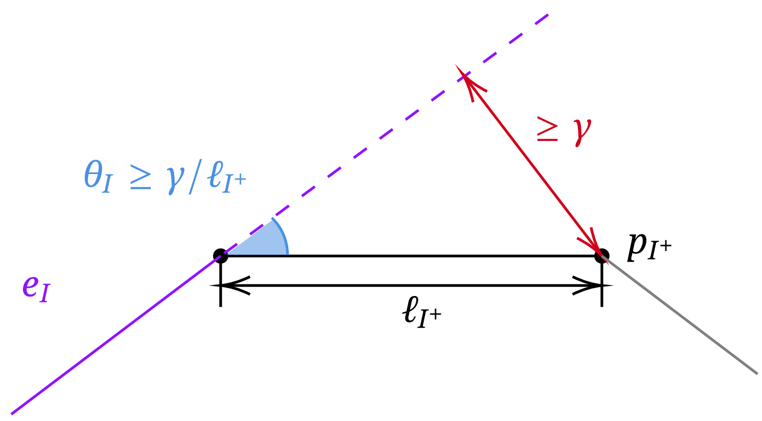

Figure 1. Illustration of the case when e_{I^+} is short. In purple is the edge e_I with its extension line dashed. The next edge in clockwise direction, e_I^+, has length ℓ_{I^+} and is drawn in black. In red is the line-to-vertex distance $\mathrm{dist}(p_{I^+}, \mathrm{affhull}(e_I))$, and in blue is the angle θ_I. If $\mathrm{dist}(p_{I^+}, \mathrm{affhull}(e_I)) \geq \gamma$ then $\theta_I \geq \sin(\theta_I) \geq \gamma/\ell_{I^+}$.

On the other hand, the sum of exterior angles of a polygon equals to 2π. Therefore we can upper bound the expected number of edges that are not likely to be long of at most

$$\mathbb{E}[\sum_{I \in S_1 \setminus S_0} 1[E_I]] \leq \frac{2\pi}{\min_{I \in S_1 \setminus S_0} \mathbb{E}[\theta_I \mid E_I]} \leq \frac{2\pi \cdot 20t}{\gamma}. \tag{8}$$

We will select $\gamma > 0$ as large as possible subject to the fact that every $I \in \binom{[n]}{d}$ with $\Pr[E_I] \geq 10\binom{n}{d}^{-1}$ satisfies $I \in S_1$.

Summing up the number of edges induced by sets in $S_0, S_1 \setminus S_0$ and $\binom{[n]}{d} \setminus S_1$, we get an upper bound on the expected edge-count of $Q \cap W$ of at most

$$\mathbb{E}[\mathrm{edges}(Q \cap W)] \leq \frac{2\pi + O(\sigma\sqrt{\log n})}{t} + \frac{40\pi t}{\gamma} + 10 = O\left(\sqrt{\frac{1 + \sigma\sqrt{\log n}}{\gamma}}\right) \tag{9}$$

where the final step follows from optimizing $t > 0$ to get the strongest possible bound. We summarize our result in Theorem 3.4. For details of the edge-counting strategy, see Section 3.

Two-dimensional Upper Bound In the second part of our proof, we need to show a lower bound of the expected distance from the affine hull of an edge of $Q \cap W$ to the next vertex in clockwise order, which is the quantity γ mentioned in (7).

As a warm-up, we first discuss the two-dimensional case $d = 2$, which will be explained in detail in Section 4. In this case, W will become the entire two-dimensional space and will disappear from consideration. Therefore, we can focus on lower-bounding the distance from any edge e of the polygon $Q = \mathrm{conv}(a_1, \dots, a_n)$ to any other vertices, i.e., it suffices to find $\gamma > 0$ such that for any $I \in \binom{[n]}{2}$,

$$\Pr[\mathrm{dist}(\{a_j : j \notin I\}, \mathrm{affhull}(e_I)) \geq \gamma \mid E_I] \geq 0.1.$$

We can obtain a lower bound on this quantity for any L-log-Lipschitz distribution. Through an appropriate coordinate transformation we prove that, irrespective of the values of a_j, $j \notin I$, the distance $\mathrm{dist}(\mathrm{conv}(a_j, j \notin I), \mathrm{affhull}(e_I))$, conditional on being non-zero, follows from a $2L$-log-Lipschitz distribution. We will calculate that we may choose $\gamma = \Omega(1/L)$. This result can be applied with small changes to our Gaussian random variables $a_1, \dots, a_n$ by substituting for the Laplace-Gaussian distribution following [17]. Plugging into (9), we get that in the two-dimensional case, $\mathbb{E}[\mathrm{edges}(Q)] \leq O(\sqrt{(1 + \sigma\sqrt{\log n})/(\sigma/\sqrt{\log n})}) = O(\sqrt[4]{\log n}/\sqrt{\sigma} + \sqrt{\log n})$ as stated in Theorem 1.4.

Multi-Dimensional Upper Bound As in the two-dimensional case, we must lower-bound of the line-to-vertex distance $\mathrm{dist}(p_{I^+}, \mathrm{affhull}(e_I))$ (see (7)) of $Q \cap W$. Analysing this, however, becomes more challenging. In two-dimensional case, each edge is the convex hull of two vertices among $a_1, \dots, a_n$ and is independent of the other potential vertices a_i. In contrast, if $d \geq 3$ then each edge on $Q \cap W$ will be the intersection between W and a $(d-1)$-dimensional facet of Q (which is the convex hull of d vertices), and each vertex will be the intersection between W and a $(d-2)$-dimensional ridge of Q (which is the convex hull of $d-1$ vertices). So the distributions of e_I and p_{I^+} are correlated.

To overcome these difficulties, we first factor $\mathrm{dist}(p_{I^+}, \mathrm{affhull}(e_I))$ into the product of separate parts which are easier to analyse, and then use log-Lipschitzness of $a_1, \dots, a_n$ to lower-bound each part with good probability. Fix without loss of generality $e = e_{[d]} = \mathrm{conv}(a_1, \dots, a_d) \cap W$, as the potential edge of interest. Consider the second endpoint p on e in clockwise direction and let $J \in \binom{[d]}{d-1}$ be the index set such that $\{p\} = \mathrm{conv}(a_j : j \in J) \cap W$. Let $p' = \mathrm{conv}(a_i : i \in J') \cap W$ (with $J' \in \binom{[n]}{d-1}$) be the vertex next to the edge e in clockwise direction. From the non-degeneracy conditions, we know that J' only differs from J with two vertices almost surely, so we can assume without loss of generality that $J = \{2, \dots, d\}$ and $J' = \{3, \dots, d\} \cup \{k\}$ for some $k \in \{d+1, \dots, n\}$.

The main idea of our analysis is the observation that if the Euclidean diameter of Q is bounded above by $O(1)$ (which happens with overwhelming probability due to Gaussian tail bound), then we can lower bound the two-dimensional line-to-vertex distance $\mathrm{dist}(p', \mathrm{affhull}(e))$ by the product of two distances $\Omega(\delta \cdot r)$, where

— δ is the d-dimensional distance from the facet-defining hyperplane $\mathrm{affhull}(a_1, \ldots, a_d)$ containing e, to the vertices of Q that are not in the facet, i.e.,

$$\delta = \mathrm{dist}(\mathrm{conv}(a_{d+1}, \ldots, a_n), \mathrm{affhull}(a_1, \ldots, a_d));$$

— r is the distance from the boundary of the ridge $\partial\,\mathrm{conv}(a_2, \ldots, a_d)$ to the one-dimensional line $\mathrm{affhull}(e)$, i.e., $r = \mathrm{dist}(\mathrm{affhull}(e), \partial\,\mathrm{conv}(a_2, \ldots, a_d))$.

We will give the formal statement of the distance splitting lemma in Lemma 5.5.

It then remains to show that r and δ are both unlikely to be too small. Similar to the two-dimensional case, we will also use log-Lipschitzness of $a_1, \ldots, a_d$ as our main tool.

— After specifying $\mathrm{affhull}(a_1, \ldots, a_d)$, the lower bound on δ is derived from the remaining randomness in $a_{d+1}, \ldots, a_n$. Here we use both the L-log-Lipschitzness of the distributions of $a_{d+1}, \ldots, a_n$, as well as the knowledge that we only need to consider hyperplanes $\mathrm{affhull}(a_1, \ldots, a_d)$ which are likely to have all points $a_{d+1}, \ldots, a_n$ on its one side. This is made precise in Section 5.3.

— The lower bound of r resembles a more technical version of the proof of the "distance lemma" of [49].

Write $\pi : \mathrm{affhull}(a_1, \ldots, a_d) \to \mathrm{affhull}(e)^\perp$ for the orthogonal projection sending e to a single point $p = \pi(e)$. With this notation we have $r = \mathrm{dist}(p, \partial\,\mathrm{conv}(\pi(a_2), \ldots, \pi(a_d)))$.

First we show that each vertex of the projected ridge $\mathrm{conv}(\pi(a_2), \ldots, \pi(a_d))$ is a distance $\Omega(1/(d^2L))$ away from the hyperplane spanned by its other vertices. That means that the projected $(d-2)$-dimensional ridge $\mathrm{conv}(\pi(a_2), \ldots, \pi(a_d))$ is likely to be wide in every direction.

In the second step, we show that W intersects $\mathrm{conv}(a_2, \ldots, a_d)$ "through the center". Specifically, we show that if we write $p = \sum_{i\in[d]} \lambda_i \pi(a_i)$ as the convex combination of $a_2, \cdots, a_d$, then with constant probability $\min_{i\in[d]} \lambda_i \geq \Omega(1/(d^2L))$.

The product of the lower bounds $\Omega(1/(d^2L))$ and $\Omega(1/(d^2L))$ for the individual quantities will yield a lower bound for r with good probability. This is included in Section 5.4.

We conclude our main result of the line-to-vertex distance lower bound in Lemma 5.2. Readers are referred to Section 5 for detailed discussions.

1.3.2 Smoothed Complexity Lower Bound

Our smoothed complexity lower bound (Theorem 6.1) is based on two geometric observations using the inner and outer radius of the perturbed polytope. For a polytope P and a unit norm ball $\mathbb{B}$, its outer radius with center x is the smallest R such that there exists $P \subseteq R \cdot \mathbb{B} + x$. Its inner radius with center x is the largest r such that $r \cdot \mathbb{B} + x \subseteq P$.

The first observation is that, if a two-dimensional polygon T has inner ℓ_2-radius of r and outer ℓ_2-radius of $(1+\varepsilon)\cdot r$ with respect to the same center, then T has at least $\Omega(\varepsilon^{-1/2})$ edges

(Lemma 6.10). This comes from the fact that every edge of T has length at most $O(r\sqrt{\varepsilon})$, whereas the perimeter of T is at least $2\pi r$.

Second, if two polytopes $Q, \tilde{Q} \subseteq \mathbb{R}^d$, each with inner radius t, have Hausdorff distance $\varepsilon < t/2$ to each other, then Q will approximate $\tilde{Q}$ in the way that (Lemma 6.9)

$$(1 - 2\varepsilon/t) \cdot Q \subseteq \tilde{Q} \subseteq (1 + \varepsilon/t) \cdot Q.$$

In particular, for any two-dimensional linear subspace W we have

$$(1 - 2\varepsilon/t) \cdot Q \cap W \subseteq \tilde{Q} \cap W \subseteq (1 + \varepsilon/t) \cdot Q \cap W. \tag{10}$$

To prove our lower bound, we construct a polytope $Q = \mathrm{conv}(\bar{a}_1, \dots, \bar{a}_n) \subseteq \mathbb{R}^d$ and a two-dimensional linear subspace W such that $\Omega(1) \cdot \mathbb{B}_1^d \subseteq Q \subseteq \mathbb{B}_1^d$, and $Q \cap W$ has outer ℓ_2-radius $r > 0$ and inner ℓ_2-radius $\frac{r}{(1+4^{-d})}$. Perturbing the vertices of Q, we obtain $\tilde{Q} = \mathrm{conv}(a_1, \dots, a_n)$, where $a_i \sim N(\bar{a}_i, \sigma^2 I_{d\times d})$ for each $i \in [n]$. Note that $Q \subseteq \mathbb{B}_1^d$ implies that $\bar{a}_1, \dots, \bar{a}_n$ satisfy the normalization requirement in (1). With high probability the Hausdorff distance in ℓ_1 between Q and $\tilde{Q}$ is bounded by $\max_{i\in[n]} \|a_i - \tilde{a}_i\|_1 \le O(\sigma d\sqrt{\log n})$. Using (10), we bound the inner and outer radius of $\tilde{Q} \cap W$. A lower bound on the number of edges of $\tilde{Q} \cap W$ then follows from Lemma 6.10 as described above.

We remark that the polytopes $Q = \mathrm{conv}(a_1, \dots, a_n) \subseteq \mathbb{R}^d$ with $n = O(d)$ and two-dimensional subspaces W such that $Q \cap W$ has inner ℓ_2-radius $\frac{r}{1+4^{-d}}$ and outer ℓ_2-radius $r > 0$ were first obtained by [8] as an *extended formulation* for a regular 2^k-gon with $O(k)$ variables and $O(k)$ inequalities. Their polytope, however, has an outer and inner radius that differ by a factor $2^{\Omega(k)}$, meaning that we cannot apply Lemma 6.9 for $\sigma > 2^{-k}$. We construct an alternative extended formulation where the ratio between inner and outer ℓ_1-radius is only $O(1)$. With an appropriate scaling to get $Q \subseteq B_1^d$, we find that the perturbed polytope $\tilde{Q}$ will have intersection $\tilde{Q} \cap W$ with inner radius $\frac{r}{1+4^{-d}}(1 - 2\varepsilon/t)$ and outer radius $(1 + \varepsilon/t)r$, where $\varepsilon = O(\sigma d\sqrt{\log n})$, and thus has $\Omega(\min(\frac{1}{\sqrt{\varepsilon}}, 2^d))$ edges, with high probability.

2. Preliminaries

We write $\mathbf{1}_n$ for the all-ones vector in $\mathbb{R}^n$, $\mathbf{0}_n$ for the all-zeroes vector in $\mathbb{R}^n$, and $I_{n\times n}$ for the n by n identity matrix. The standard basis vectors are denoted by $e_1, \dots, e_n \in \mathbb{R}^n$. For a linear subspace $W \subseteq \mathbb{R}^n$ we denote the orthogonal projection onto W by π_W. The subspace of vectors orthogonal to a given vector $\omega \in \mathbb{R}^n$ is denoted $\omega^\perp$.

For a vector $x \in \mathbb{R}^n$, the ℓ_1 norm is $\|x\|_1 = \sum_{i\in[n]} |x_i|$, the ℓ_2-norm is $\|x\|_2 = \sqrt{\sum_{i\in[n]} x_i^2}$ and the ℓ_∞-norm is $\|x\|_\infty = \max_{i\in[n]} |x_i|$. A norm without a subscript is always the ℓ_2-norm. Given $p \ge 1, d \in \mathbb{Z}_+$, define $\mathbb{B}_p^d = \{x \in \mathbb{R}^d : \|x\|_p \le 1\}$ as the d-dimensional unit ball of ℓ_p norm.

We write $[n] := \{1, \dots, n\}$. By $\mathrm{conv}(a_1, \dots, a_n) = \mathrm{conv}(a_i : i \in [n])$ we denote the convex hull of vectors $a_1, \dots, a_n$, and similarly we write for the affine hull as $\mathrm{affhull}(a_i : i \in [n])$. For sets $A, B \subseteq \mathbb{R}^d$, the distance between the two is $\mathrm{dist}(A, B) = \inf_{a\in A, b\in B} \|a - b\|$. For a point $x \in \mathbb{R}^d$ we write $\mathrm{dist}(x, A) = \mathrm{dist}(A, x) = \mathrm{dist}(A, \{x\})$.

We say a random event happens *almost surely* if it occurs with probability 1.

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

2.1 Polytopes

DEFINITION 2.1 (Polytope). A convex set $P \subseteq \mathbb{R}^d$ is a polyhedron if it can be expressed as $P = \{x \in \mathbb{R}^d : Ax \leq b\}$ for some $A \in \mathbb{R}^{n\times d}$, $b \in \mathbb{R}^n$ where $n \in \mathbb{Z}_+$. A bounded polyhedron is also called a polytope.

DEFINITION 2.2 (Valid Condition and Facet). Given a polytope $P \subseteq \mathbb{R}^d$, vector $c \in \mathbb{R}^d$ and $d \in \mathbb{R}$, we say the linear condition $x^\top c \leq d$ is valid for P if the condition holds for all $x \in P$.

A subset $F \subseteq P$ is called a face of P if $F = P \cap \{x \in \mathbb{R}^d : x^\top c = d\} \neq \emptyset$ for some valid condition $x^\top c \leq d$. A facet is a $d-1$-dimensional face, a ridge is a $d-2$-dimensional face, an edge is a 1-dimensional face and a vertex is a 0-dimensional face.

DEFINITION 2.3 (Polar dual of a convex body). Let $P \subseteq \mathbb{R}^d$ be a convex set. Define the polar dual of P as

$$P^\circ = \{y \in \mathbb{R}^d : y^\top x \leq 1, \forall x \in P\}.$$

We state some basic facts from duality theory:

FACT 2.4 (Polar dual of polytope). *Let $P \subseteq \mathbb{R}^d$ be a polytope given by the linear system $P = \{x \in \mathbb{R}^d, Ax \leq \mathbf{1}_n\} \subseteq \mathbb{R}^d$ for some $A \in \mathbb{R}^{n\times d}$. Then the polar dual of P equals to*

$$P^\circ := \mathrm{conv}(\mathbf{0}_d, a_1, a_2, \ldots, a_n).$$

where $a_1, \ldots, a_n \in \mathbb{R}^d$ are the row vectors of A. Moreover, P is bounded if and only if $\mathbf{0}_d \in \mathrm{int}(\mathrm{conv}(a_1, \ldots, a_n))$.

FACT 2.5. *Let $P, Q \subseteq \mathbb{R}^d$ be two convex sets such that $P \subseteq Q$. Then $Q^\circ \subseteq P^\circ$.*

FACT 2.6. *Let $P \subseteq \mathbb{R}^d$ be a polytope, and let $W \subseteq \mathbb{R}^d$ be any $k \leq d$-dimensional linear subspace. Then the polar dual of $\pi_W(P)$, considered as a subset of the linear space W, is equal to $P^\circ \cap W$.*

2.2 Probability Distributions

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

DEFINITION 2.7 (Gaussian distribution). The d-dimensional Gaussian distribution $\mathcal{N}_d(\bar{a}, \sigma^2 I)$ with support $\mathbb{R}^d$, mean $\bar{a} \in \mathbb{R}^d$, and standard deviation σ is defined by the probability density function

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

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

A basic property of Gaussian distribution is the following strong tail bound:

LEMMA 2.8 (Gaussian tail bound). *Let $x \in \mathbb{R}^d$ be a random vector sampled from $\mathcal{N}_d(\mathbf{0}, \sigma^2 I)$. For any $t \geq 1$ and any $\theta \in \mathbb{S}^{d-1}$ where $\mathbb{S}^{d-1}$ is the unit sphere in the d-dimensional space, we have*

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

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

COROLLARY 2.9 (Global diameter of Gaussian random variables). *For any $n \geq 2$, let $x_1, \ldots, x_n \in \mathbb{R}^d$ be random variables where each $x_i \sim \mathcal{N}_d(\mathbf{0}_d, \sigma^2 I)$. Then with probability at least $1 - \binom{n}{d}^{-1}$, $\max_{i \in [n]} \|x_i\| \leq 4\sigma\sqrt{d \log n}$.*

PROOF. From Lemma 2.8, we have for each $i \in [n]$ that

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

Then the statement follows from the union bound over all choices of $i \in [n]$. ∎

A helpful technical substitute for the Gaussian distribution was introduced by [17]:

DEFINITION 2.10 ((σ, r)-Laplace-Gaussian distribution). For any $\sigma, r > 0$, $\bar{a} \in \mathbb{R}^d$, define the d-dimensional (σ, r)-Laplace-Gaussian distribution with mean $\bar{a}$, or $LG_d(\bar{a}, \sigma, r)$, if its density function is proportional to

$$f(x) = \begin{cases} \exp\left(-\frac{\|x-\bar{a}\|^2}{2\sigma^2}\right) & \text{, if } \|x - \bar{a}\| \leq r\sigma \\ \exp\left(-\frac{\|x-\bar{a}\| r}{\sigma} + \frac{r^2}{2}\right) & \text{, if } \|x - \bar{a}\| > r\sigma. \end{cases} \tag{11}$$

The Laplace-Gaussian random variables satisfies many desirable properties. Like the Gaussian distribution, the distance to its mean is bounded above with high probability. Moreover, its probability density is log-Lipschitz throughout its domain (as a contrast, the probability density of Gaussian distribution is only log-Lipschitz close to the expectation). The definition of L-log-Lipschitz is as follows:

DEFINITION 2.11 (L-log-Lipschitz random variable). Given $L > 0$, we say a random variable $x \in \mathbb{R}^d$ with probability density μ is L-log-Lipschitz (or μ is L-log-Lipschitz), if for all $x, y \in \mathbb{R}^d$, we have

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

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

LEMMA 2.12 (Properties of Laplace-Gaussian random variables, Lemmas 3.7 and 3.33 of [17]). *Given any $n \geq d$, $\sigma > 0$. Let $a_1, \ldots, a_n \in \mathbb{R}^d$ be independent random variables each sampled form $LG_d(\bar{a}, \sigma, 4\sigma\sqrt{d \log n})$ (see Definition 2.10). Then $a_1, \ldots, a_n$ satisfy the follows:*

1. *(Log-Lipschitzness) For each $i \in [n]$, the probability density of a_i is $(4\sigma^{-1}\sqrt{d \log n})$-log-Lipschitz.*
2. *(Bounded maximum norm) With probability at least $1 - \binom{n}{d}^{-1}$, $\max_{i \in n} \|a_i - \bar{a}_i\| \leq 4\sigma\sqrt{d \log n}$.*
3. *(Bounded expected radius of projection) For any $k \leq d$, any fixed k-dimensional linear subspace $H \subseteq \mathbb{R}^d$, we have $\mathbb{E}[\max_{i \in [n]} \|\pi_H(a_i - \bar{a}_i)\|] \leq 8\sigma\sqrt{k \log n}$.*

Most importantly, Laplace-Gaussian perturbations lead to nearly the same shadow size as Gaussian perturbations.

LEMMA 2.13 (Lemma 3.34 of [17]). *Given any $n \geq d \geq 2, \sigma > 0$, any two-dimensional linear subspace $W \subseteq \mathbb{R}^d$, and any $\bar{a}_1, \ldots, \bar{a}_n \in \mathbb{R}^d$. For every $i \in [n]$, let $a_i \sim \mathcal{N}_d(\bar{a}_i, \sigma)$ and $\widehat{a}_i \sim LG_d(\bar{a}_i, \sigma, 4\sigma\sqrt{d \log n})$ be independently sampled. Then the following holds*

$$\mathbb{E}\left[\mathrm{edges}(\mathrm{conv}(a_1, \ldots, a_n) \cap W)\right] \leq 1 + \mathbb{E}\left[\mathrm{edges}(\mathrm{conv}(\widehat{a}_1, \ldots, \widehat{a}_n) \cap W)\right].$$

Although [17] state the above lemma only for $d \geq 3$, their proof applies without change to the case $d = 2$.

2.3 Change of Variables

We will make use of a specific change of variables, which is a standard tool in stochastic and integral geometry. It will allow us to investigate, for points $a_1, \ldots, a_d \in \mathbb{R}^d$, the distribution of these points when conditioning on a specific affine hyperplane $\mathrm{affhull}(a_1, \ldots, a_d)$. This will enable us to make conclusions about the shapes of the faces of the convex hull $\mathrm{conv}(a_1, \ldots, a_n)$ of n vectors.

DEFINITION 2.14 (Change of variables). Let $\theta \in \mathbb{S}^{d-1}, t \in \mathbb{R}$ be such that $\forall i \in [d], \theta^\top a_i = t$ and suppose $\theta^\top e_1 > 1$ without loss of generality where $e_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^d$ is the unit vector that has nonzero element on the first coordinate.

Fix h as any isometric embedding from $\mathbb{R}^{d-1} \to e_1^\perp$. Let $\tilde{R}_\theta : \mathbb{R}^d \to \mathbb{R}^d$ denote the rotation that rotates e_1 to θ in the two-dimensional subspace $\mathrm{span}(e_1, \theta)$, and is the identity transformation on $\mathrm{span}(e_1, \theta)^\perp$. Define $R_\theta = \tilde{R}_\theta \circ h$ to be the resulting isometric embedding from $\mathbb{R}^{d-1}$, identified with $e_1^\perp$, to $\theta^\perp$. Now define the transformation ϕ from $\theta \in \mathbb{S}^{d-1}, t \in \mathbb{R}, b_1, \dots, b_d \in \mathbb{R}^{d-1}$ to $a_1, \dots, a_d \in \mathbb{R}^d$ as follows:

$$\phi(\theta, t, b_1, \dots, b_d) = (R_\theta(b_1) + t\theta, \dots, R_\theta(b_d) + t\theta). \tag{12}$$

The choice of the embeddings $\tilde{R}_\theta$ is largely arbitrary. What matters is that the transformation ϕ and its inverse are continuous almost everywhere.

LEMMA 2.15 (Jacobian of the transformation, see Theorem 7.2.7 in [46]). *Let $\phi : \mathbb{S}^{d-1} \times \mathbb{R} \times \mathbb{R}^{(d-1)\times d} \to \mathbb{R}^{d\times d}$ be the transformation defined in Definition 2.14. The transformation ϕ is defined almost everywhere and has Jacobian determinant that equals to*

$$\left|\det\left(\frac{\partial \phi}{\partial(\theta, t, b_1, \dots, b_d)}\right)\right| = C_d \cdot (d-1)! \cdot \mathrm{vol}_{d-1}(\mathrm{conv}(b_1, \dots, b_d))$$

for some constant C_d depending only on the dimension. As a consequence, if $a_1, \dots, a_d \in \mathbb{R}^d$ are points with probability density $\mu(a_1, \dots, a_d)$ and if $\theta \in \mathbb{S}^{d-1}, t \in \mathbb{R}, b_1, \dots, b_d \in \mathbb{R}^{d-1}$ have probability density proportional to

$$\mathrm{vol}_{d-1}(\mathrm{conv}(b_1, \dots, b_d)) \cdot \mu(t\theta + R_\theta(b_1), \dots, t\theta + R_\theta(b_d))$$

then $\mathbb{E}[f(a_1, \dots, a_d)] = \mathbb{E}[f(\phi(\theta, t, b_1, \dots, b_d))]$ for any measurable function f.

The interested reader might observe that if $(a_1, \dots, a_d) = \phi(\theta, t, b_1, \dots, b_d)$ then the Jacobian

$$(d-1)! \cdot \mathrm{vol}_{d-1}(\mathrm{conv}(a_1, \dots, a_d)) = (d-1)! \cdot \mathrm{vol}_{d-1}(\mathrm{conv}(b_1, \dots, b_d))$$

is equal to the determinant of the $d \times d$ matrix $(\theta^\top, (a_1 - a_d)^\top, (a_2 - a_d)^\top, \dots, (a_{d-1} - a_d)^\top)$, which is equal to the determinant of the $(d-1) \times (d-1)$ matrix with rows or columns given by $b_1 - b_d, b_2 - b_d, \dots, b_{d-1} - b_d$.

In particular, we will use this transformation to condition on the value of θ and consider events in the variables $t, b_1, \dots, b_d$. For this purpose, we have the following fact.

LEMMA 2.16 (Log-Lipschitzness of the Position of Affine Hull). *Let $a_1, \dots, a_d \in \mathbb{R}^d$ be d independent L-log-Lipschitz random variables, and let $(\theta, t, b_1, \dots, b_d) = \phi^{-1}(a_1, \dots, a_d)$, where $\phi : \mathbb{S}^{d-1} \times \mathbb{R} \times \mathbb{R}^{(d-1)\times d} \to \mathbb{R}^{d\times d}$ is defined in Definition 2.14. Then conditional on the values of $\theta, b_1, \dots, b_d$, the random variable t is (dL)-log-Lipchitz.*

PROOF. By Lemma 2.15, the joint probability density of $(a_1, \dots, a_d)$ is proportional to

$$\mathrm{vol}_{d-1}(\mathrm{conv}(b_1, \dots, b_d)) \cdot \prod_{i=1}^{d} \mu_i(R_\theta(b_i) + t\theta)$$

where μ_i is the probability density of a_i. Conditioning on $b_1, \dots, b_d \in \mathbb{R}^{d-1}$ and $\theta \in \mathbb{S}^{d-1}$, the volume $\mathrm{vol}_{d-1}(\mathrm{conv}(b_1, \dots, b_d))$ is fixed. The statement then follows from the fact that for each $i \in [d]$, $\mu_i(R_\theta(b_i) + t\theta)$ is L-log-Lipschitz in t for any b_i. ∎

Conditional on θ and t, the points $b_1, \dots, b_d$ are not L-log-Lipschitz due to the volume term. When it becomes relevant, we will show that this factor does not affect the argument in any negative way.

2.4 Non-Degenerate Conditions

DEFINITION 2.17 (Non-degenerate polytope). A polytope $Q = \mathrm{conv}(a_1, \dots, a_n) \subseteq \mathbb{R}^d$ is called non-degenerate, if it is simplicial (every facet is a simplex) and if, for $i \in [n]$, $a_i \in \partial Q$ implies that a_i is a vertex of Q.

DEFINITION 2.18 (Non-degenerate intersection with a 2D-plane). Let $Q \subseteq \mathbb{R}^d$ be a non-degenerate polytope and let $W \subseteq \mathbb{R}^d$ be a two-dimensional linear subspace. We say Q has non-degenerate intersection with W, if

1. the edges of the two-dimensional polygon $Q \cap W$ have one-to-one correspondence to the facets of Q that have non-empty intersection with W; and
2. the vertices of $Q \cap W$ have one-to-one correspondence to the $(d-2)$-dimensional faces (ridges) of Q that have non-empty intersection with W

FACT 2.19 (Non-degenerate conditions of random polytope). *Given any $n \geq d \geq 2$ and any fixed two-dimensional plane $W \subseteq \mathbb{R}^d$. For $a_1, \dots, a_n \in \mathbb{R}^d$, the polytope $Q = \mathrm{conv}(a_1, \dots, a_n)$ satisfies the following properties everywhere except for a set of measure 0:*

1. *Q is non-degenerate;*
2. *Q has non-degenerate intersection with W;*
3. *For every normal vector v to any facet of Q, $e_1^\top v \neq 0$.*

Assume the polytope $Q = \mathrm{conv}(a_1, \dots, a_n)$ and the two-dimensional linear subspace $W \subseteq \mathbb{R}^d$ satisfy the non-degenerate conditions in Fact 2.19. Each edge of the two-dimensional polygon formed by the intersection $W \cap Q$ can be described by a set of d vertices, where the edge is equivalent to the intersection of W with the convex hull of these d vertices. Furthermore, each vertex of $Q \cap W$ corresponds to a set of $(d-1)$ vertices. The following lemma characterizes the relation of these sets for adjacent vertices and edges:

FACT 2.20 (Properties of neighboring vertices on non-degenerate intersection polygon). *Let $W \subseteq \mathbb{R}^d$ be a two-dimensional linear subspace, $Q = \mathrm{conv}(a_1, \dots, a_n) \subseteq \mathbb{R}^d$ is simplicial and has non-degeneracy intersection with W. Given $J_1, J_2 \in \binom{[n]}{d-1}$, $I \in \binom{[n]}{d}$, suppose (1) $V_{J_1} = \mathrm{conv}(a_j : j \in J_1) \cap W$ and $V_{J_2} = \mathrm{conv}(a_j : j \in J_2) \cap W$ are two adjacent vertices of $Q \cap W$, and (2) $\mathrm{conv}(a_i : i \in I) \cap W$ is an edge of $Q \cap W$ that contains V_{J_1} but not contains V_{J_2}. Then we have $|J_1 \setminus J_2| = |J_2 \setminus J_1| = 1$ and $|I \setminus J_2| = 2$.*

PROOF. Let $I' = J_1 \cup J_2$. Then $\mathrm{conv}(V_{J_1}, V_{J_2}) = \mathrm{conv}(a_i : i \in I) \cap W$ is an edge of the polygon $Q \cap W$. Since Q has non-degenerate intersection with W, we have that $|I'| = d$. Combining with $|J_1| = |J_2| = d - 1$ gives us that $|J_1 \backslash J_2| = |J_2 \backslash J_1| = 1$.

Next we consider $|I \backslash J_2|$. Since $J_1 \subseteq I$ and $|J_1 \backslash J_2| = 1$, it could only be the case that $|I \backslash J_2| \in \{1, 2\}$. If $|I \backslash J_2| = 1$, then since $|I| = |J_2| + 1$ we must have $J_2 \subseteq I$, but this contradicts to the fact that $\mathrm{conv}(J_2) \not\subset \mathrm{conv}(I)$. Therefore we could only have $|I \backslash J_2| = 2$. ■

3. Smoothed Complexity Upper Bound

In this section, we establish our key theorem for upper bounding the number of edges of a random polygon $\mathrm{conv}(a_1, \ldots, a_n) \cap W$ for W a fixed 2-dimensional linear subspace and $a_1, \ldots, a_n \in \mathbb{R}^d$. We demonstrate that if for any edge on the shadow polygon $\mathrm{conv}(a_1, \ldots, a_n) \cap W$, the expected distance between the affine hull of the edge and the next vertex on the shadow is sufficiently large in expectation, then the expected number of edges of $\mathrm{conv}(a_1, \ldots, a_n) \cap W$ can be upper-bounded.

DEFINITION 3.1 (Facet and edge event). For $I \subseteq [n]$, we write $F_I = \mathrm{conv}(a_i : i \in I)$. Define E_I to be the event that both F_I is a facet of $\mathrm{conv}(a_1, \ldots, a_n)$ and $F_I \cap W \neq \emptyset$.

REMARK 3.2. Any edge e of $\mathrm{conv}(a_1, \ldots, a_n) \cap W$ can be written as $e = F_I \cap W$ for some $I \subseteq [n]$ for which E_I holds. Assuming non-degeneracy, this relation between edges and index sets is a one-to-one correspondence, and moreover every $I \subseteq [n]$ for which E_I holds satisfies $|I| = d$.

To state the key theorem's assumption, we need a concept of 'clockwise' to characterize the order of edges and vertices on the shadow polygon.

DEFINITION 3.3 (Clockwise order of edges and vertices). For any given two-dimensional linear subspace $W \subseteq \mathbb{R}^d$ we denote an arbitrary but fixed rotation as "clockwise". For the polygon $\mathrm{conv}(a_1, \ldots, a_n) \cap W$ of our interest, let $p_1, \ldots, p_k$ denote its vertices in clockwise order and write $p_{k+1} = p_1, p_{k+2} = p_2$. Then for any edge $e = [p_{i-1}, p_i]$, we call p_i its second vertex in clockwise order and we call p_{i+1} the next vertex after e in clockwise order. The edge $[p_i, p_{i+1}]$ is the next edge after e in clockwise order.

Note that the above terms are well-defined in the sense that they depend only on the polygon and the orientation of the subspace, not on the vertex labels. With this definition in place, we can now state the theorem itself:

THEOREM 3.4 (Smoothed complexity upper bound for continuous perturbations). *Fix any $n, d \geq 2$, $\sigma \geq 0$, and any two-dimensional linear subspace $W \subseteq \mathbb{R}^d$. Let $a_1, \ldots, a_n \in \mathbb{R}^d$ be independently distributed each according to a continuous probability distribution.*

For any $I \in \binom{[n]}{d}$, conditional on E_I, define $y_I \in W$ as the outer unit normal of the edge $F_I \cap W$. For $\gamma > 0$, suppose that for each $I \in \binom{[n]}{d}$ such that $\Pr[E_I] \geq 10\binom{n}{d}^{-1}$ we have

$$\Pr[y_I^\top p_2 - y_I^\top p_3 \geq \gamma \mid E_I] \geq 0.1,$$

where we write $[p_1, p_2] = F_I \cap W$ and $p_3 \in \mathrm{conv}(a_1, \ldots, a_n) \cap W$ as the next vertex after $F_I \cap W$ in clockwise order. Then we have

$$\begin{aligned}\mathbb{E}\left[\mathrm{edges}\left(\mathrm{conv}(a_1, \ldots, a_n)\right)\right] &\leq 10 + 80\pi\sqrt{\frac{\mathbb{E}[\max_{i\in[n]} \|\pi_W(a_i)\|]}{\gamma}} \\ &= O\left(\sqrt{\frac{\mathbb{E}[\max_{i\in[n]} \|\pi_W(a_i)\|]}{\gamma}}\right).\end{aligned}$$

Note that, assuming non-degeneracy, y_I is well-defined if and only if E_I happens. In this case, we are guaranteed that $y_I^\top p_2 - y_I^\top p_3 > 0$.

To prove the above theorem, we show that any $I \in \binom{[n]}{d}$ with $\Pr[E_I] \geq \binom{n}{d}^{-1}$ can be charged to either a portion of the perimeter of the polygon $\mathrm{conv}(a_1, \ldots, a_n) \cap W$ or to a portion of its sum 2π of exterior angles at its vertices.

DEFINITION 3.5 (Exterior angle and length of the next edge). Given any $I \in \binom{[n]}{d}$, we define two random variables $\theta_I, \ell_{I^+} \geq 0$. If E_I happens, write $p_1, p_2 \in F_I \cap W$ for the first and the second endpoint of $F_I \cap W$ in clockwise order. Let θ_I to be the (two-dimensional) exterior angle of $\mathrm{conv}(a_1, \ldots, a_n) \cap W$ at p_2; If E_I doesn't happen then let $\theta_I = 0$.

Let ℓ_{I^+} denote the following random variable: If E_I happens, then ℓ_{I^+} equals to the length of the next edge after $F_I \cap W$ in clockwise order, i.e., the other edge of $\mathrm{conv}(a_1, \ldots, a_n) \cap W$ containing p_2. If E_I doesn't happen then let $\ell_{I^+} = 0$.

PROOF OF THEOREM 3.4. Since we have non-degeneracy with probability 1, by Fact 2.19 and linearity of expectation we find

$$\mathbb{E}\left[\mathrm{edges}\left(\mathrm{conv}(a_1, \ldots, a_n) \cap W\right)\right] = \sum_{I \in \binom{[n]}{d}} \Pr[E_I].$$

We can give an upper bound on the expected number of edges of $\mathrm{conv}(a_1, \ldots, a_n) \cap W$ by upper-bounding each $\Pr[E_I]$. Fix any $I \in \binom{[n]}{d}$ and let $t > 0$ be a parameter to be determined later. We consider three different possible upper bounds on $\Pr[E_I]$, at least one of which will always hold:

Case 1: $\Pr[E_I] \leq 10\binom{n}{d}^{-1}$.

Since $\sum_{I \in \binom{[n]}{d}} 10\binom{n}{d}^{-1} = 10$, one can immediately see that the total contribution of edges counted in this case is at most 10.

Case 2: $\Pr[E_I] > 10\binom{n}{d}^{-1}$ **and** $\Pr[\ell_{I^+} \geq t \mid E_I] \geq \frac{1}{20}$.

In this case, by Markov's inequality $\mathbb{E}[\ell_{I^+} \mid E_I] \geq \frac{t}{20}$, therefore we obtain from $\mathbb{E}[\ell_{I^+}] = \mathbb{E}[\ell_{I^+} \mid E_I]\Pr[E_I]$ and $E[\ell_{I^+}|E_I^C] = 0$ that

$$\Pr[E_I] = \frac{\mathbb{E}[\ell_{I^+}]}{\mathbb{E}[\ell_{I^+} \mid E_I]} \leq \frac{20}{t} \cdot \mathbb{E}[\ell_{I^+}].$$

Case 3: $\Pr[E_I] > 10\binom{n}{d}^{-1}$ **and** $\Pr[\ell_{I^+} \leq t \mid E_I] \geq \frac{19}{20}$. Readers are referred to Figure 1 for more illustration of the proof.

Conditional on E_I, without loss of generality we write $[p_1, p_2] = F_I \cap W$ and let p_3 denote the next vertex after $F_I \cap W$ in clockwise direction. From the theorem's assumption we have $\Pr[\mathrm{dist}(\mathrm{affhull}(p_1, p_2), p_3) \geq \gamma \mid E_I] \geq \frac{1}{10}$. Then from the union bound,

$$\begin{aligned}&\Pr[(\ell_{I^+} \leq t) \wedge (\mathrm{dist}(\mathrm{affhull}(p_1, p_2), p_3) \geq \gamma) \mid E_I] \\ &\geq 1 - \Pr[\ell_{I^+} > t \mid E_I] - \Pr[\mathrm{dist}(\mathrm{affhull}(p_1, p_2), p_3) < \gamma \mid E_I] \geq \frac{1}{20}.\end{aligned}$$

Referring back to Figure 1, we know that $\theta_I \geq 0$ and thus

$$\theta_I \geq \sin(\theta_I) = \frac{\mathrm{dist}(\mathrm{affhull}(p_1, p_2), p_3)}{\ell_{I^+}}$$

we have $\mathbb{E}[\theta_I \mid E_I] \geq \frac{1}{20} \cdot \frac{\gamma}{t}$. Combining this with the fact that $\theta_I = 0$ if E_I does not hold, we know that $\mathbb{E}[\theta_I] = \mathbb{E}[\theta_I \mid E_I]\Pr[E_I]$ and we can upper bound

$$\Pr[E_I] = \frac{\mathbb{E}[\theta_I]}{\mathbb{E}[\theta_I \mid E_I]} \leq \frac{20t}{\gamma} \cdot \mathbb{E}[\theta_I].$$

Combining the upper bounds for each $\Pr[E_I]$ for the above three cases, we get that

$$\begin{aligned}\mathbb{E}\left[\mathrm{edges}\left(\mathrm{conv}(a_1, \ldots, a_n) \cap W\right)\right] &= \sum_{I \in \binom{[n]}{d}} \Pr[E_I] \\ &\leq \sum_{I \in \binom{[n]}{d}} \left(10\binom{n}{d}^{-1} + \frac{20}{t} \cdot \mathbb{E}[\ell_{I^+}] + \frac{20t}{\gamma} \cdot \mathbb{E}[\theta_I]\right) \\ &= 10 + \frac{20}{t} \cdot \mathbb{E}[\sum_{I \in \binom{[n]}{d}} \ell_{I^+}] + \frac{20t}{\gamma} \cdot \mathbb{E}[\sum_{I \in \binom{[n]}{d}} \theta_I] \qquad (13)\end{aligned}$$

To upper bound the second term of (13), we notice that $\sum_{I \in \binom{[n]}{d}} \ell_{I^+}$ exactly equals the perimeter of $\mathrm{conv}(a_1, \ldots, a_n) \cap W$. Since the shadow polygon $\mathrm{conv}(a_1, \ldots, a_n) \cap W$ is contained in the two-dimensional disk of radius $\max_{i \in [n]} \|\pi_W(a_i)\|$, by the monotonicity of surface area for convex sets we have

$$\mathbb{E}[\sum_{I \in \binom{[n]}{d}} \ell_{I^+}] \leq 2\pi \cdot \mathbb{E}[\max_{i \in [n]} \|\pi_W(a_i)\|].$$

To upper bound the third term of (13), we notice that the sum of exterior angles for any polygon always equals 2π. Thus

$$\mathbb{E}[\sum_{I \in \binom{[n]}{d}} \theta_I] = 2\pi \tag{14}$$

Finally, we combine (13 - 14) and minimize over all $t > 0$:

$$\begin{aligned}\mathbb{E}\left[\text{edges}\left(\text{conv}(a_1, \ldots, a_n) \cap W\right)\right] &\leq \min_{t>0}\left(10 + \frac{40\pi\mathbb{E}[\max_{i\in[n]} \|\pi_W(a_i)\|]}{t} + \frac{40\pi t}{\gamma}\right) \\ &= 10 + 80\pi\sqrt{\frac{\mathbb{E}[\max_{i\in[n]} \|\pi_W(a_i)\|]}{\gamma}}.\end{aligned}$$

where in the final step, we set $t = \sqrt{\gamma\mathbb{E}[\max_{i\in[n]} \|\pi_W(a_i)\|]}$. ■

In the subsequent sections, we will show a lower bound for the edge-to-vertex distance γ assuming the independently distributed vectors $a_1, \cdots, a_n$ follow Laplace-Gaussian distributions. This allows us to directly apply Theorem 3.4 to derive an upper bound on the expected number of edges of $\text{conv}(a_1, \cdots, a_n) \cap W$. Furthermore, by using Lemma 2.13, we can further reduce our upper bound to the case when $a_1, \ldots, a_n$ are Gaussian distributed vectors.

4. Upper Bound in Two Dimension

In this section, we establish the smoothed complexity upper bound for $d = 2$. For this scenario, the shadow plane W encompasses the entire two-dimensional Euclidean space, and $P \cap W$ is identical to $P = \text{conv}(a_1, \cdots, a_n)$. From Theorem 3.4 and Lemma 2.13, it remains to lower bound the distance from the affine hull of an edge to its neighboring vertex in clockwise order (denoted by γ in Theorem 3.4), where the vertices of the polygon $a_1, \cdots, a_n$ is sampled from a Laplace-Gaussian distribution with the center of $\bar{a}_1, \cdots, \bar{a}_n$. We will demonstrate a slightly stronger result: a lower bound for the distance between the affine hull of an edge to all of the remaining $(n-2)$ vertices.

LEMMA 4.1 (Edge-to-vertex distance in Two Dimension). *Let $a_1, \ldots, a_n \in \mathbb{R}^2$ be n independent L-log-Lipschitz random variables. Then, for any $I \in \binom{[n]}{2}$, the outer unit normal $y \in W$ of the edge* $\text{conv}(a_i : i \in I)$ *satisfies*

$$\Pr[y^\top a_i - \max_{j \notin I} y^\top a_j \geq \frac{1}{L} \mid E_I] \geq 0.1,$$

for any $i \in I$.

By applying Theorem 3.4, Lemma 2.13, and the Laplace-Gaussian tail bound of Lemma 2.12, we find the following upper bound for two-dimensional polygons under Gaussian perturbation:

THEOREM 4.2 (Two-Dimensional Upper Bound). *Let $\bar{a}_1, \dots, \bar{a}_n \in \mathbb{R}^2$ be $n > 2$ vectors with norm at most 1. For each $i \in [n]$, let a_i be independently distributed as $\mathcal{N}_2(\bar{a}_i, \sigma^2 I)$. Then*

$$\mathbb{E}\left[\text{edges}\left(\text{conv}(a_1, \dots, a_n)\right)\right] \le O\left(\frac{\sqrt[4]{\log n}}{\sqrt{\sigma}} + \sqrt{\log n}\right).$$

PROOF. For each $i \in [n]$, let $\hat{a}_i$ be independently sampled form the 2-dimensional Laplace-Gaussian distribution $LG_2(\bar{a}_i, \sigma, 4\sigma\sqrt{2\log n})$. It follows from Lemma 2.12 that $\hat{a}_i$ is $(4\sigma^{-1}\sqrt{2\log n})$-log-Lipschitz and $\mathbb{E}[\max_{i\in[n]} \|\hat{a}_i\|] \le 1 + 4\sigma\sqrt{2\log n}$. We use Lemma 4.1, setting $L = 4\sigma^{-1}\sqrt{2\log n}$, and Theorem 3.4, setting $\gamma = \frac{1}{L} = \frac{\sigma}{4\sqrt{2\log n}}$, to find

$$\mathbb{E}\left[\text{edges}\left(\text{conv}(\hat{a}_1, \dots, \hat{a}_n)\right)\right] \le O\left(\sqrt{\frac{\sqrt{\log n}}{\sigma} + \log n}\right) \le O\left(\frac{\sqrt[4]{\log n}}{\sqrt{\sigma}} + \sqrt{\log n}\right).$$

Finally, from Lemma 2.13, we conclude that $\mathbb{E}\left[\text{edges}\left(\text{conv}(a_1, \dots, a_n)\right)\right] \le 1 + O(\frac{\sqrt[4]{\log n}}{\sqrt{\sigma}} + \sqrt{\log n})$. ∎

PROOF OF LEMMA 4.1. Fix any set $I = \{i, i'\} \subseteq [n]$. Define $z \in \mathbb{S}^1$ and t to satisfy $z^\top a_i = z^\top a_{i'} = t$ and $z^\top e_1 > 0$. Both are well-defined with probability 1.

Note that E_I is now equivalent to either having $z^\top a_j < t$ for all $j \notin I$ or having $z^\top a_j > t$ for all $j \notin I$. Write E_I^+ for the former case and E_I^- for the latter. The vector z is always defined, assuming non-degeneracy, and is equal to the outer normal unit vector y conditional on E_I^+ and equal to $-y$ conditional on E_I^-.

We want to apply the principle of deferred decisions to fix the values of a_j for each $j \notin I$ and z, and now only allow a_i and $a_{i'}$ to vary while fixing the slope of the line between a_i and $a_{i'}$. Let $\mu : \mathbb{R} \to \mathbb{R}_{\ge 0}$ denote the induced probability density function of $t = y^\top a_i = y^\top a_{i'}$. Lemma 2.16 tells us that μ is $(2L)$-log-Lipschitz.

In the first case, for E_I^+, we have, still only considering the randomness over t,

$$\begin{aligned}
&\Pr[(t - \max_{j\notin I} z^\top a_j \ge \frac{1}{L}) \wedge E_I^+] \\
=& \int_{\max_{j\notin I} z^\top a_j + 1/L}^{\infty} \mu(t)\mathrm{d}t \\
=& \int_{\max_{j\notin I} z^\top a_j}^{\infty} \mu(s + 1/L)\mathrm{d}s \\
\ge& \int_{\max_{j\notin I} z^\top a_j}^{\infty} e^{-2}\mu(s)\mathrm{d}s && \text{(By $(2L)$-log-Lipschitzness of μ)} \\
=& e^{-2}\Pr[E_I^+].
\end{aligned}$$

Similarly for the other case, E_I^-, we find

$$\Pr[(\min_{j\notin I} z^\top a_j - t \geq \frac{1}{L}) \wedge E_I^-] \geq e^{-2}\Pr[E_I^-].$$

Now observe that, for any $i \in I$, we have

$$\begin{aligned}&\Pr[y^\top a_i - \max_{j\notin I} y^\top a_j \geq \frac{1}{L} \wedge E_I]\\ =&\Pr[(t - \max_{j\notin I} z^\top a_j \geq \frac{1}{L}) \wedge E_I^+] + \Pr[(\min_{j\notin I} z^\top a_j - t \geq \frac{1}{L}) \wedge E_I^-]\\ \geq& e^{-2}\Pr[E_I^+] + e^{-2}\Pr[E_I^-] = e^{-2}\Pr[E_I].\end{aligned}$$

This finishes the proof since

$$\Pr[y^\top a_i - \max_{j\notin I} y^\top a_j \geq \frac{1}{L} \mid E_I] = \Pr[y^\top a_i - \max_{j\notin I} y^\top a_j \geq \frac{1}{L} \wedge E_I]/\Pr[E_I] \geq e^{-2} \geq 0.1.$$

■

5. Multi-Dimensional Upper Bound

In this section, we establish the upper bound for the higher-dimensional case (i.e., $d \geq 3$):

THEOREM 5.1 (Multi-dimensional Upper Bound). *Let $d > 2, n \geq d$, and $\sigma \leq \frac{1}{16d\sqrt{\log n}}$. Let $\bar{a}_1, \ldots, \bar{a}_n$ be n vectors with $\max_{i\in[n]} \|\bar{a}_i\| \leq 1$. For each $i \in [n]$, let a_i be independently distributed as $\mathcal{N}_d(\bar{a}_i, \sigma^2 I)$. Then*

$$\mathbb{E}[\mathrm{edges}(\mathrm{conv}(a_1, \ldots, a_n) \cap W)] = O\left(\sigma^{-3/2} d^{13/4} \log^{5/4} n\right). \tag{15}$$

Similar to the two-dimensional case (see Section 4), the main technical ingredient of Theorem 5.1 is a lower-bound of the edge-to-vertex distance (the quantity γ in Theorem 3.4) on the shadow polygon:

LEMMA 5.2 (Edge-to-vertex distance of shadow polygon in multi-dimension). *For any $d \geq 3$, let $a_1, \ldots, a_n \in \mathbb{R}^d$ be independent L-log-Lipschitz random variables. For any $I \in \binom{[n]}{d}$ that satisfies $\Pr[E_I] \geq 10\binom{n}{d}^{-1}$, (where E_I is defined in Definition 3.1), we have*

$$\Pr[y^\top p - y^\top p' \geq \Omega(\frac{1}{L^3 d^5 \log n}) \mid E_I] \geq 0.1,$$

where p is any point in $F_I \cap W$, and $p' \in \mathrm{conv}(a_1, \ldots, a_n) \cap W$ is the next vertex after $F_I \cap W$ in clockwise direction. Here $y \in W$ is the outer unit normal to the edge $F_I \cap W$ on $\mathrm{conv}(a_1, \ldots, a_n) \cap W$.

Theorem 5.1 then immediately follows from Lemma 5.2, Theorem 3.4, and Lemma 2.13:

PROOF OF THEOREM 5.1. For each $i \in [n]$, let $\hat{a}_i$ be independently sampled from the Laplace-Gaussian distribution $LG_d(\bar{a}_i, \sigma, 4\sigma\sqrt{d \log n})$. From Lemma 2.12, we know that

1. Each $\hat{a}_i$ is $L = (4\sigma^{-1}\sqrt{d \log n})$-log-Lipschitz;
2. $\mathbb{E}[\max_{i\in[n]} \|\pi_W(\hat{a}_i)\|] \le 1 + 8\sigma\sqrt{2\log n} \le 1.5$.

From Lemma 5.2, we get that for any $p \in F_I \cap W$, if p' is the next vertex after the edge $F_I \cap W$ in clockwise order, then

$$\Pr[y_I^\top p \ge y_I^\top p' + \Omega(\frac{1}{L^3 d^5 \log n}) \mid E_I] \ge 0.1.$$

Here $y_I \in W$ is the outer unit normal vector of the polygon $\mathrm{conv}(\hat{a}_1, \ldots, \hat{a}_n) \cap W$ on the edge $F_I \cap W$. Then we can use Theorem 3.4 by setting $L = 4\sigma^{-1}\sqrt{d\log n}$, $\gamma = \Omega(\frac{1}{L^3 d^5 \log n})$ and $\mathbb{E}[\max_{i\in[n]} \|\pi_W(a_i)\|] = 1.5$, to find

$$\begin{aligned}
\mathbb{E}\left[|\,\mathrm{edges}\,(\mathrm{conv}(\hat{a}_1, \ldots, \hat{a}_n))\,|\right] &\le 10 + O(\sqrt{\frac{E[\max_{i\in[n]} \|\pi_W(a_i)\|]}{\gamma}}) \\
&\le 10 + O(\sqrt{\frac{1.5}{\frac{1}{L^3 d^5 \log n}}}) \\
&\le 10 + O(\sqrt{L^3 d^5 \log n}) \\
&\le 10 + O(\sqrt{\sigma^{-3} d^{13/2} \log^{5/2} n}).
\end{aligned}$$

Finally, from Lemma 2.13, we conclude that

$$\begin{aligned}
\mathbb{E}\left[|\,\mathrm{edges}\,(\mathrm{conv}(a_1, \ldots, a_n) \cap W)\,|\right] &\le 11 + O(\sqrt{\sigma^{-3} d^{13/2} \log^{5/2} n}) \\
&= O\left(\sigma^{-3/2} d^{13/4} \log^{5/4} n\right).
\end{aligned}$$

■

The rest of this section is dedicated to the proof of Lemma 5.2 and will be structured as follows. In Section 5.1 we define some basic notation that will be used in the proof. In Section 5.2 we establish two sufficient criteria for the conclusion of Lemma 5.2 to hold. In Section 5.3 and Section 5.4, we prove that these conditions hold with good probability conditional on E_I. Section 5.5 to Section 5.7 include the proof of the auxiliary lemmas. Finally, we finish the proof of Lemma 5.2 in Section 5.8.

5.1 Notations

Since we assume that the constraint matrix rows $a_1, \ldots, a_n$ each have a continuous probability density function, $\mathrm{conv}(a_1, \ldots, a_n)$ and W satisfy the non-degeneracy conditions (see Fact 2.19) almost surely. In this case, each edge of the polygon $\mathrm{conv}(a_1, \ldots, a_n) \cap W$ is given by $F_I \cap W = \mathrm{conv}(a_i : i \in I)$ for which $I \in \binom{[n]}{d}$ and E_I holds (where F_I and E_I are defined in Definition 3.1). In addition, each vertex of the polygon $\mathrm{conv}(a_1, \ldots, a_n) \cap W$ is given by the intersection between W and $(d-2)$-dimensional ridges of $\mathrm{conv}(a_1, \ldots, a_n)$, which are convex hulls of $(d-1)$ vertices of $\mathrm{conv}(a_1, \ldots, a_n)$. We define the following notations for a ridge and its corresponding vertex:

DEFINITION 5.3 (Ridge and vertex event). For any $J \subseteq [n]$, write $R_J = \mathrm{conv}(a_j : j \in J)$. Define A_J to be the event that R_J is a ridge of $\mathrm{conv}(a_1, \ldots, a_n)$ and $R_J \cap W \neq \emptyset$.

REMARK 5.4. Any vertex v of $\mathrm{conv}(a_1, \ldots, a_n)$ can be written as $v = R_J \cap W$ for some $J \subseteq [n]$ for which A_J holds. Assuming non-degeneracy, each J for which A_J holds satisfies $|J| = d-1$ and the relation between vertices and index sets $J \in \binom{[n]}{d-1}$ with A_J is a one-to-one correspondence.

5.2 Deterministic Conditions for a Good Edge-to-Vertex Separator

In this subsection, we present a series of sufficient conditions such that an edge on the polygon $\mathrm{conv}(a_1, \ldots, a_n) \cap W$ maintains a significant separation from its next vertex in clockwise order. When the vertex set $\{a_i\}_{i=1}^n$ is fixed, this edge-to-vertex distance can be decomposed into two geometric components:

- the distance from all other vertices of Q to the supporting hyperplane of the facet containing the edge, and
- the depth at which the intersection point $p = R \cap W$ lies in the interior of the ridge R.

Lemma 5.5 shows that if these two quantities are bounded below by δ and r respectively, then their product, $r\delta/3$, guarantees a significant edge-to-vertex gap in the projected polygon $Q \cap W$.

LEMMA 5.5. *Let $W \subseteq \mathbb{R}^d$ be a two-dimensional linear subspace, $Q = \mathrm{conv}(a_1, \ldots, a_n) \subseteq \mathbb{R}^d$ be a non-degenerate polytope with a non-degenerate intersection with W such that $\max_{i,j \in [n]} \|a_i - a_j\| \leq 3$ and $W \cap Q \neq \emptyset$. Fix any facet F of Q such that $F \cap W \neq \emptyset$ and any ridge $R \subseteq F$ of F such that $W \cap R$ is a singleton set $\{p\}$. Let $\delta, r \geq 0$ be such that*

1. *(distance between F and other vertices) $\forall a_k \notin F, \mathrm{dist}(\mathrm{affhull}(F), a_k) \geq \delta$;*
2. *(Inner radius of R) $\mathrm{dist}(\mathrm{affhull}(F \cap W), \partial R) \geq r$.*

Then for any $p \in F \cap W$, the outer unit normal vector $\bar{\theta} \in W$ to the edge $F \cap W$ satisfies

$$\bar{\theta}^\top p - \bar{\theta}^\top p' \geq \delta r/3,$$

where $p' \in Q \cap W$ is the next vertex after $F \cap W$ in clockwise order.

The reader who desires a more intuitive illustration of the geometry involved may take a look at Figure 2 from the next lemma, where relevant concepts are depicted as they happen for $d = 4$. The simplex is the "next" facet after F, and its bottom face B is the ridge of Q that is shared with F. The distance from b_1 to B is large due to the first assumption of Lemma 5.5, and the distance from q to the boundary of B is large due to the second assumption. To map the depicted facet to a three-dimensional space for the purpose of the illustration, it has been projected orthogonally to the subspace perpendicular to $F \cap W$.

We remark that Lemma 5.5 gives a sufficient condition assuming that $\{a_i : i \in [n]\}$ is fixed. In later subsections, our goal is to show this sufficient condition actually occurs with high probability even when $\{a_i : i \in [n]\}$ are Laplace-Gaussian distributed random variables.

To start, we show a lemma about the distance from a point in the simplex to the affine hull of its neighboring facet.

LEMMA 5.6. *Given $b_1, \ldots, b_d \in \mathbb{R}^{d-1}$ such that* $\mathrm{conv}(b_1, \ldots, b_d)$ *is non-degenerate. For any* $D > 0$, *suppose*

1. $\forall i, j \in [d]$, $\|b_i - b_j\| \le D$;
2. $\mathrm{dist}(b_1, \mathrm{affhull}(b_2, \ldots, b_d)) \ge \delta$;
3. *There exists* $q \in \mathrm{conv}(b_2, \ldots, b_d)$ *such that* $\mathrm{dist}(q, \partial(\mathrm{conv}(b_2, \ldots, b_d))) \ge r$.

Then we have $\mathrm{dist}(q, \mathrm{affhull}(b_1, \ldots, b_{d-1})) \ge r\delta/D$.

PROOF. For simplicity, write $B = \mathrm{conv}(b_2, \ldots, b_d)$ and $B' = \mathrm{conv}(b_1, \ldots, b_{d-1})$. Let $q' = \pi_{B'}(q)$ be the point closest to q on $\mathrm{affhull}(B')$, i.e., $\|q - q'\| = \mathrm{dist}(q, \mathrm{affhull}(b_1, \ldots, b_{d-1}))$.

Let $x = (B \cap B') \cap \mathrm{affhull}(b_1, q, q')$ be its intersection between the two-dimensional plane $\mathrm{affhull}(b_1, q, q')$ and the $(d-3)$-dimensional ridge $B \cap B'$ (which gives a unique point). (See Figure 2 for an illustration). Consider the triangle $\mathrm{conv}(b_1, q, x)$ and calculate its area in two

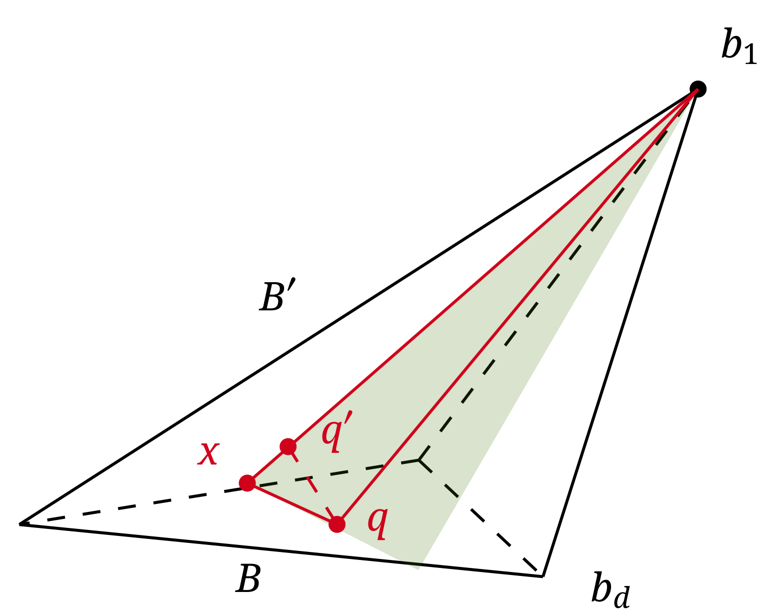

Figure 2. Illustration of Lemma 5.6 when $d - 1 = 3$. In light green is the intersection between the two-dimensional plane $\mathrm{affhull}(b_1, q, q')$ and $\mathrm{conv}(b_1, \ldots, b_d)$. The red triangle is $\mathrm{conv}(b_1, x, q)$. The bottom face is B and the left-facing back face is B'.

different ways. On one hand, it has base $\mathrm{conv}(b_1, x)$ of length $\|b_1 - x\| \le D$ with height $\mathrm{dist}(q, \mathrm{affhull}(b_1, x)) = \|q - q'\|$, which gives that the area of the triangle is at most $D\|q - q'\|/2$. On the other hand, this triangle has base $\mathrm{conv}(x, q)$ of length $\|x - q\| \ge \mathrm{dist}(q, \partial(B)) \ge r$ with height $\mathrm{dist}(b_1, \mathrm{affhull}(x, q)) \ge \mathrm{dist}(b_1, \mathrm{affhull}(B)) \ge \delta$, which gives that the area of the triangle is at least $r\delta/2$.

Combining the above two ways of determining the area of triangle $\mathrm{conv}(b_1, q, x)$, we have $D\|q - q'\|/2 \ge r\delta/2$. Therefore we have $\mathrm{dist}(q, \mathrm{affhull}(B')) = \|q - q'\| \ge \frac{r\delta}{R}$ as desired. ∎

What Lemma 5.6 tells us is that, if we have sufficiently strong information about the geometry in base of a simplex (conditions 1 and 2) then we can relate the height of that simplex to the length of a given chord. In our setting, we let a ridge of the polar polytope play the role of the base, and take the apex to be one of the vertices a_i not contained in the chosen facet.

To prove our main deterministic geometric result (Lemma 5.5), we *squash* the configuration by projecting it orthogonally onto the $(d-1)$-dimensional subspace $s^\perp$ orthogonal to the edge direction. Under this projection, the two neighboring ridges R and R' become adjacent facets of a simplex, with the apex corresponding to the first vertex outside the facet F. The bounded diameter, height, and inner radius conditions in Lemma 5.5 then align precisely with the assumptions of Lemma 5.6, which allow us to apply it and lift the resulting bound back to the original two-dimensional slice W.

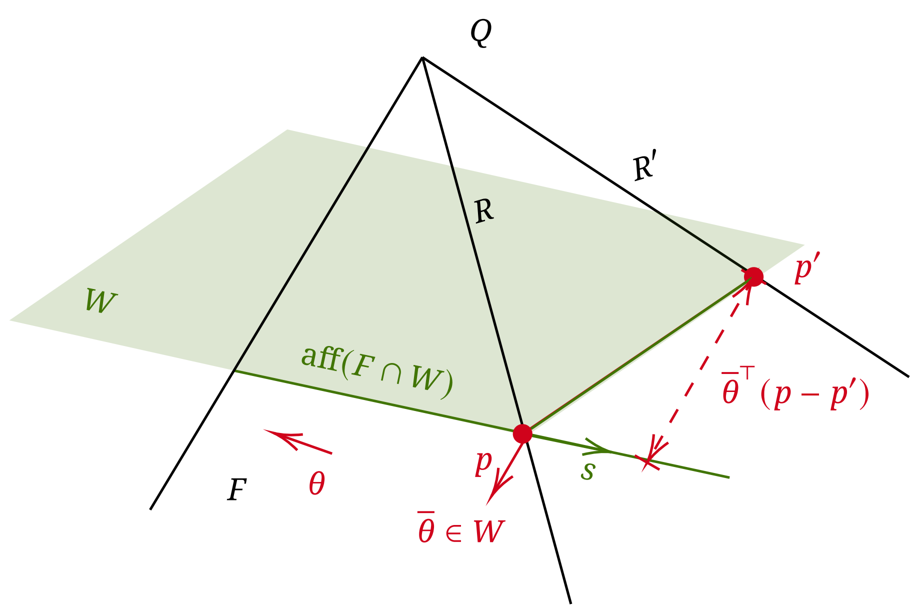

Figure 3. Illustration of the variables in Lemma 5.5 for $d = 3$. The light-green parallelogram is the two-dimensional plane W. The red arrows are θ (outward unit normal of F) and $\bar{\theta} = \pi_W(\theta)/\|\pi_W(\theta)\|$. The red points $p = W \cap R$ and $p' = W \cap R'$ are two consecutive points on the polygon $Q \cap W$. The red dashed line demonstrates the edge-to-vertex distance $\bar{\theta}^\top(p - p')$.

PROOF OF LEMMA 5.5. Write R' for the ridge of Q such that $\{p'\} = R' \cap W$. Since $p' \in Q \cap W$ is adjacent to vertex p and the edge $F \cap W$, by Fact 2.20 we may relabel the a_i such that $R' = \mathrm{conv}(a_1, \dots, a_{d-1})$, $R = \mathrm{conv}(a_2, \dots, a_d)$, and $F = \mathrm{conv}(a_2, \dots, a_{d+1})$ without loss of generality. Let $\theta \in \mathbb{S}^{d-1}$ denote the outward unit normal to F. This normal vector satisfies

$$\delta \le \min_{\substack{i \in [n] \\ a_i \notin F}} \theta^T (p - a_i) \le \theta^\top (p - a_1).$$

Let $s \in \mathbb{S}^{d-1}$ be a unit vector indicating the direction of the (one-dimensional) line $F \cap W$. This vector is unique up to sign. Also, let $\bar{\theta} = \pi_W(\theta)/\|\pi_W(\theta)\|$ be the outward unit normal to $F \cap W$ in the two-dimensional plane W. See Figure 3 for an illustration of the variables for $d = 3$. Notice that $\bar{\theta}$ and s form an orthonormal basis of W. Therefore we get

$$\bar{\theta}^\top(p-p') = \bar{\theta}^\top \pi_{s^\perp}(p-p') = \|\pi_{s^\perp}(p-p')\| \tag{16}$$

Here the last equality comes from $(p-p') \in W = \mathrm{span}(\bar{\theta}, s)$, so $\pi_{s^\perp}(p-p') = \pi_{\bar{\theta}}(p-p')$.

Now we focus on the $(d-1)$-dimensional space $s^\perp$, and consider the orthogonal projections $\pi_{s^\perp}(a_1), \dots, \pi_{s^\perp}(a_d)$. Since the diameter of $\mathrm{conv}(a_1, \dots, a_d)$ is at most 3, we have $\max_{i,j\in[d]} \|\pi_{s^\perp}(a_i) - \pi_{s^\perp}(a_j)\| \leq 3$. By definition, θ is a unit normal of R, so since $s \in R$, we have $\theta \in s^\perp$. It follows that θ is also a unit normal of $\pi_{s^\perp}(R) = \pi_{s^\perp}(\mathrm{conv}(a_2, \dots, a_d))$. This gives

$$\mathrm{dist}(\pi_{s^\perp}(a_1), \mathrm{affhull}(\pi_{s^\perp}(R))) = \theta^\top(p - a_1) \geq \delta.$$

Also, since $\mathrm{dist}(F \cap W, \partial R) \geq r$ and $F \cap W$ is one-dimensional, after the projection to $s^\perp$ we have

$$\mathrm{dist}(\pi_{s^\perp}(p), \partial \pi_{s^\perp}(R)) = \mathrm{dist}(\mathrm{affhull}(F \cap W), \partial R) \geq r.$$

Therefore we can use Lemma 5.6 to get

$$\|\pi_{s^\perp}(p) - \pi_{s^\perp}(p')\| \geq \mathrm{dist}(\pi_{s^\perp}(p), \mathrm{affhull}(\pi_{s^\perp}(R'))) \geq r\delta/3,$$

where the first step comes from $\pi_{s^\perp}(p') \in \mathrm{affhull}(\pi_{s^\perp}(R'))$. The lemma then follows from (16). ∎

5.3 Randomized Lower-Bound for δ: Distance between vertices and facets

In this section, we show that the affine hull of a given facet F of the polytope $\mathrm{conv}(a_1, \dots, a_n)$ is $\Omega(\frac{1}{Ld\log n})$-far away to other vertices with good probability, or in other words, the distance δ in Lemma 5.5 is at least $\Omega(\frac{1}{Ld\log n})$ with good probability. Our main result of this section is as follows:

LEMMA 5.7 (Randomized lower-bound for δ). *Let $a_1, \dots, a_n \in \mathbb{R}^d$ be independent L-log-Lipschitz random vectors. For any $I \in \binom{[n]}{d}$ such that $\Pr[E_I] \geq 10\binom{n}{d}^{-1}$, we have*

$$\Pr[\min_{k\in[n]\setminus I} \mathrm{dist}(\mathrm{affhull}(F_I), a_k) \geq \frac{1}{10e^3 dL\log n}) \mid E_I] \geq 0.72.$$

Intuition The proof of this lemma will span this entire subsection. Let us start with some words on the intuition behind it. Assume $\mathrm{affhull}(F_I)$ is fixed arbitrarily. Then the quantities $\mathrm{dist}(\mathrm{affhull}(F_I), a_k)$ are determined solely by the points a_k, $k \in [n] \setminus I$. The points are L-log-Lipschitz, which makes each signed distance $|\theta^\top a_k| = \mathrm{dist}(\mathrm{affhull}(F_I), a_k) \in \mathbb{R}$ into an L-log-Lipschitz random variable as well. Any L-log-Lipschitz random variable has its probability density function pointwise upper bounded by L, and hence the probability that for a given $k \in [n] \setminus I$ we have $\Pr[\theta^\top a_k \in [-\varepsilon, \varepsilon]] \leq 2\varepsilon L$. A union bound would then give

$$\Pr[\min_{k\in[n]\setminus I} \mathrm{dist}(\mathrm{affhull}(F_I), a_k) \leq \varepsilon] \leq 2(n-d)\varepsilon L.$$

Taking $\varepsilon = \frac{0.28}{2(n-d)L}$ would give us a bound on the probability.

It is clear that this weaker version of Lemma 5.7 would be relatively easy to prove. However, it has a linear dependence on n and thus it would add a factor of $\sqrt{n}$ to our shadow bound. This is undesirable. To obtain the stronger conclusion, we consider the expected number of points at close distance to affhull(F_I).

$$\Pr\left[\min_{k\in[n]\setminus I} \operatorname{dist}(\operatorname{affhull}(F_I), a_k) \le \varepsilon \,\middle|\, E_I\right] \le \mathbb{E}\left[|\{i \in [k] \setminus I : \operatorname{dist}(\operatorname{affhull}(F_I), a_k) \le \varepsilon\}| \,\middle|\, E_I\right]. \quad (17)$$

If $\varepsilon \le 1/L$ then we can use L-log-Lipschitzness to derive a lower bound on the number of points a_k lying above (or below) the affine subspace affhull(F_I). What we prove is that if (17) is large then (without conditioning on E_I) both the expected number of points above affhull(F_I) and the expected number of points below affhull(F_I) are at least $\Omega(\frac{1}{\varepsilon L})$ times as large as (17).

However, recall that the event E_I requires all points to lie on the same side of affhull(F_I). If there is simultaneously a point above affhull(F_I) and a point below affhull(F_I) then E_I does not hold. Using the Chernoff bound we can show, conditional on some affhull$(F_I) = H$, that if we have $\Pr[E_I \mid \operatorname{affhull}(F_I) = H] \ge n^{-d}$ then at least one of the expected number of points above affhull(F_I) or the expected number of points below affhull(F_I) must be at most $2d \log n$. If one of these is bounded from above, then (17) must be bounded from above. Taking proper care to observe that those affine subspaces H for which $\Pr[E_I \mid \operatorname{affhull}(F_I) = H] \ge n^{-d}$ together account for most of the probability mass, this will yield the desired result.

Proving the lemma To show Lemma 5.7, we fix any $I \in \binom{[n]}{d}$ of consideration. Without loss of generality, assume $I = [d]$ and write $E = E_{[d]}$. We define the following event B_ε indicating that the distance from affhull$(F_{[d]})$ to other vertices is at least ε.

DEFINITION 5.8 (Separation by the margin of a facet). Let $\theta \in \mathbb{S}^{d-1}, t \in \mathbb{R}$ be as in Definition 2.14. For any $\varepsilon > 0$, let B_ε^+ denote the event that $\theta^\top a_i < t - \varepsilon$ for all $i \in [n] \setminus [d]$ and B_ε^- denote the event that $\theta^\top a_i > t + \varepsilon$ for all $i \in [n] \setminus [d]$. We write $B_\varepsilon = B_\varepsilon^+ \vee B_\varepsilon^-$.

In the following lemma, we show that for sufficiently small ε, $\Pr[E \wedge B_\varepsilon]$ is still a constant fraction of $\Pr[E]$.

LEMMA 5.9. *For any $0 < \varepsilon \le \frac{1}{10e^3 Ld \log n}$ it holds that*

$$\Pr[E] \le \binom{n}{d}^{-1} + \frac{5}{4} \cdot \Pr[E \wedge B_\varepsilon].$$

PROOF. Writing random variables as subscripts to denote which expectation is over which variables, we start by using Fubini's theorem to write

$$\Pr_{a_1,\dots,a_n}[E] = \mathbb{E}_{a_1,\dots,a_d}[\Pr_{a_{d+1},\dots,a_n}[E]].$$

Fix any $a_1, \ldots, a_d \in \mathbb{R}^n$ subject to $\mathrm{conv}(a_1, \ldots, a_d) \cap W \neq \emptyset$ and the non-degeneracy assumptions in Fact 2.19. Define $\theta \in \mathbb{S}^{d-1}, t > 0$ as described in Definition 2.14, i.e., $\theta^\top a_i = t$ for each $i \in [d]$. Write $s_i = \theta^\top a_i$ for each $i \in [n] \setminus [d]$. We note that s_i is an L-log-Lipschitz random variable for all $i \in [n] \setminus [d]$. Moreover, over the remaining randomness in $a_{d+1}, \ldots, a_n$, we have $\Pr[E] = \Pr[B_0^+] + \Pr[B_0^-]$ and $\Pr[B_\varepsilon] = \Pr[B_\varepsilon^+] + \Pr[B_\varepsilon^-]$. We will show that

$$\Pr[B_0^+] \leq \frac{1}{2\binom{n}{d}} + \frac{5}{4}\Pr[B_\varepsilon^+] \tag{18}$$

and the appropriate statement will follow for B_ε^- analogously. Putting together this will prove the lemma. If $\Pr[B_0^+] \leq \frac{1}{2}\binom{n}{d}^{-1}$ then the desired inequality holds directly.

In order to prove (18) we require the following claim:

CLAIM 5.10. *Conditional on θ, t, if $\Pr[B_0^+ \mid \theta, t] \geq n^{-d}$ then $\mathbb{E}[\#\{i \in [n] \setminus [d] : s_i \geq t\}] \leq 2d \log n$. If $\Pr[B_0^- \mid \theta, t] \geq n^{-d}$ then $\mathbb{E}[\#\{i \in [n] \setminus [d] : s_i \leq t\}] \leq 2d \log n$.*

PROOF. We prove the first implication, and the second follows analogously. For each $i \in [n] \setminus [d]$, let $X_i \in \{0, 1\}$ have value 1 if and only if $s_i \geq t$. Since θ, t are fixed and depend only on $a_1, \ldots, a_d$, the random variables $X_{d+1}, \ldots, X_n$ are independent. Write $X = \sum_{i=d+1}^n X_i$. The Chernoff bound gives

$$\Pr[X = 0] \leq \exp\left(-\frac{\mathbb{E}[X]}{2}\right).$$

As such, $\mathbb{E}[X] > 2d \log n$ would imply $\Pr[X = 0] < n^{-d}$, contradicting the original assumption that $\Pr[X = 0] \geq n^{-d}$. It follows that $\mathbb{E}[X] \leq 2d \log n$. ■

Thus, in what remains, we may suppose that $Pr[B_0^+] > 1/2\binom{n}{d}^{-1}$. Fix any $i \in [n] \setminus [d]$ and let μ_i denote the induced probability density function of s_i. We then have a sequence of inequalities as found below. The first two inequalities above follow from L-log-Lipschitzness of μ_i. For the first inequality in particular, note that for any $s \in [-\frac{1}{L}, 0]$, we have $|s - \varepsilon L s| \leq 1/L$ from $\varepsilon \leq 1/L$. This then gives $\frac{\mu_i(t+\varepsilon L s)}{\mu_i(t+s)} \leq \exp(L \cdot (1 - \varepsilon L)s) \leq 1$.

$$\begin{aligned}
\Pr[s_i \geq t - \varepsilon \mid s_i \leq t] &= \frac{\int_{-\varepsilon}^0 \mu_i(t+s)\mathrm{d}s}{\int_{-\infty}^0 \mu_i(t+s)\mathrm{d}s} \\
&= \frac{\varepsilon L \int_{-1/L}^0 \mu_i(t+\varepsilon L s)\mathrm{d}s}{\int_{-\infty}^0 \mu_i(t+s)\mathrm{d}s} \\
&\leq e\frac{\varepsilon L \int_{-1/L}^0 \mu_i(t+s)\mathrm{d}s}{\int_{-\infty}^0 \mu_i(t+s)\mathrm{d}s} \\
&= e\frac{\varepsilon L \int_0^{1/L} \mu_i(t+s-1/L)\mathrm{d}s}{\int_{-\infty}^{1/L} \mu_i(t+s-1/L)\mathrm{d}s}
\end{aligned}$$

$$\begin{aligned}
&\le e^3 \frac{\varepsilon L \int_0^{1/L} \mu_i(t+s)\mathrm{d}s}{\int_{-\infty}^{1/L} \mu_i(t+s)\mathrm{d}s} \\
&= e^3 \varepsilon L \Pr[s_i \ge t \mid s_i \le t + 1/L] \\
&\le e^3 \varepsilon L \Pr[s_i \ge t]. \qquad (19)
\end{aligned}$$

The third inequality, on the final line, follows from the fact that $s_i \ge t + 1/L$ implies $s_i \ge t$, and hence $\Pr[s_i \ge t \mid s_i > t + 1/L] = 1$. As such we can, for fixed t, θ, upper-bound the probability over $s_1, \ldots, s_d$ that, conditional on B_0^+, there exists a vertex being ε-close to $\mathrm{affhull}(F_I)$:

$$\begin{aligned}
\Pr[\neg B_\varepsilon^+ \mid B_0^+] &= \Pr[\exists i \in [n] \setminus [d] : s_i \ge t - \varepsilon \mid B_0^+] && \text{(By union bound)} \\
&\le \sum_{i \in [n] \setminus [d]} \Pr[s_i \ge t - \varepsilon \mid B_0^+] \\
&\le \sum_{i \in [n] \setminus [d]} e^3 \varepsilon L \Pr[s_i \ge t] && \text{(By (19))} \\
&= e^3 \varepsilon L \, \mathbb{E}[\#\{i \in [n] \setminus [d] : s_i \ge t\}]. && (20)
\end{aligned}$$

To interpret the last equality above, we observe that $\#\{i \in [n] \setminus [d] : s_i \ge t\} = 0$ if and only if B_0^+ happens. Then by applying (20) to Claim 5.10 (note that we are using the assumption $Pr[B_0^+] > 1/2\binom{n}{d}^{-1}$) with our choice of ε we conclude that

$$\Pr[B_0^+] \le \frac{5}{4} \Pr[B_\varepsilon^+] \le \frac{1}{2} \binom{n}{d}^{-1} + \frac{5}{4} \Pr[B_\varepsilon^+]. \qquad \blacksquare$$

Now we can prove Lemma 5.7 using Lemma 5.9.

PROOF OF LEMMA 5.7. Fix any $I \in \binom{[n]}{d}$. By Lemma 5.9, we have that $\Pr[E_I] \le \binom{n}{d}^{-1} + \frac{5}{4} \cdot \Pr[E_I \wedge (\delta \ge \varepsilon)]$ for $\varepsilon = \frac{1}{10e^3 L d \log n}$. This gives that

$$\frac{\Pr[E_I \wedge (\delta \ge \varepsilon)]}{\Pr[E_I]} \ge \frac{4}{5} - \binom{n}{d}^{-1} \cdot \frac{4}{5 \Pr[E_I]}$$

Moreover, since $\Pr[E_I] \ge 10\binom{n}{d}^{-1}$, we have

$$\begin{aligned}
\Pr[(\delta \ge \varepsilon) \mid E_I] &= \frac{\Pr[E_I \wedge (\delta \ge \varepsilon)]}{\Pr[E_I]} \\
&\ge \frac{4}{5} - \binom{n}{d}^{-1} \cdot \frac{4}{5 \Pr[E_I]} \ge 0.72,
\end{aligned}$$

as desired. ■

5.4 Randomized Lower-Bound for r: Inner Radius of a Ridge Projected onto $(d-1)$-Dimensional Subspace

In the next sections, we demonstrate that for any ridge R of the polytope P, wherein $R \cap W$ is a vertex of $P \cap W$, its inner radius—after projection onto the subspace orthogonal to the adjacent edge of $P \cap W$—is at least $\Omega(d^{-4}L^{-2})$ with high probability. Essentially, this establishes that the parameter r, as referred to in Lemma 5.5, is at least $\Omega(d^{-4}L^{-2})$ with good probability. We remark that Lemma 5.11 does not have an analogue when $d = 2$. Moreover, it will require substantially more technical effort. Its proof is similar to Lemma 4.1.1 (Distance bound) in [49], their main technical result. In an effort to help the ease of understanding the larger structure, some lemmas will be stated while their proofs will be given in later sections.

LEMMA 5.11 (Randomized Lower-bound for r). *Let $a_1, \dots, a_n \in \mathbb{R}^d$ be independent L-log-Lipschitz random vectors. Let D denote the event that $\forall i, j \in [n], \|a_i - a_j\| \leq 3$. Fix any $I \in \binom{[n]}{d}$ and any $J \in \binom{I}{d-1}$, we have*

$$\Pr[\operatorname{dist}(W \cap \operatorname{affhull}(a_i : i \in I), \partial \operatorname{conv}(a_j : j \in J)) \leq \frac{1}{19200 d^4 L^2} \mid E_I \wedge A_J] \leq 0.1 + \Pr[\neg D \mid E_I \wedge A_J].$$

Then Lemma 5.11 will quickly from its lower-dimensional equivalent:

LEMMA 5.12 (Randomized lower bound for r after change of variables). *Let $b_1, \dots, b_d \in \mathbb{R}^{d-1}$ be random vectors with joint probability density proportional to*

$$\operatorname{vol}_{d-1}(\operatorname{conv}(b_1, \dots, b_d)) \cdot \prod_{i=1}^{d} \bar{\mu}_i(b_i)$$

where $\bar{\mu}_i$ is L-log-Lipschitz for each $i \in [d]$. Let D' denote the event that the set $\{b_1, \dots, b_d\}$ has Euclidean diameter of at most 3. Fox any fixed one-dimensional line $\ell \subseteq \mathbb{R}^{d-1}$, we have that

$$\Pr\left[\left(\operatorname{dist}\left(\ell, \partial \operatorname{conv}(b_1, \dots, b_{d-1})\right) < \frac{1}{19200 d^4 L^2}\right) \mid \ell \cap \operatorname{conv}(b_1, \dots, b_{d-1}) \neq \emptyset\right]$$
$$\leq 0.1 + \Pr[\neg D' \mid \ell \cap \operatorname{conv}(b_1, \dots, b_{d-1}) \neq \emptyset].$$

The proof of Lemma 5.12 will be presented in Section 5.5.

PROOF OF LEMMA 5.11. We may assume without loss of generality that $I = [d]$ and $J = [d-1]$. Apply the change of variables ϕ as in Definition 2.14 to $\{a_i : i \in [d]\}$ and obtain

$$\phi(\theta, t, b_1, \dots, b_d) = (a_1, \dots, a_d).$$

where $\theta \in \mathbb{S}^{d-1}, t \in \mathbb{R}, b_1, \dots, b_d \in \mathbb{R}^{d-1}$. For any $i \in [n]$, let μ_i denote the probability density function of a_i. Writing the conditioning to $(E_{[d]} \wedge A_{[d-1]})$ as part of the pdf, we find that the joint probability density of $t, \theta, b_1, \dots, b_d, a_{d+1}, \dots, a_n$ is proportional to

$$\operatorname{vol}_{d-1}(\operatorname{conv}(b_1, \dots, b_d)) \cdot \prod_{i=1}^{d} \bar{\mu}_i(t, \theta, b_i) \cdot \prod_{i=d+1}^{n} \mu_i(a_i) \cdot 1[E_{[d]} \wedge A_{[d-1]}], \tag{21}$$

where $\operatorname{vol}_{d-1}(\cdot)$ is the volume function of $(d-1)$-dimensional simplex in its spanning hyperplane, $\bar{\mu}_i(t, \theta, b_i) = \mu_i(t\theta + R_\theta(b_i))$ is the induced probability density of b_i, which is L-log-Lipschitz, and

$1[\cdot]$ denotes the indicator function. Write S for the event that

$$\operatorname{dist}(W \cap \operatorname{affhull}(a_i : i \in I), \partial \operatorname{conv}(a_j : j \in J)) \leq \frac{1}{19200 d^4 L^2}.$$

In this language, our goal is to prove that $\Pr[S] \leq 0.1 + \Pr[\neg D]$.

Let D' denote the event that $\|b_i - b_j\| \leq 3$ for all $i, j \in [d]$. Each of the events E_I, A_J, S, D', D are functions of the random variables $\theta, t, b_1, \dots, b_d, a_{d+1}, \dots, a_n$. We then use Fubini's theorem to write

$$\Pr_{\theta,t,b_1,\dots,b_d,a_{d+1},\dots,a_n}[S] = \mathbb{E}_{\theta,t,a_{d+1},\dots,a_n}[\Pr_{b_1,\dots,b_d}[S]]$$

With probability 1 over the choice of $\theta, t, a_{d+1}, \dots, a_n$, the inner term satisfies all the conditions of Lemma 5.12. Specifically, since the value of $1[E_{[d]}]$ is already fixed, the intersection $(t\theta + \theta^\perp) \cap W$ is a line. Let $\ell \subseteq \mathbb{R}^{d-1}$ be the image of such line under the inverse change of variables ϕ^{-1}, i.e., $(t\theta + \theta^\perp) \cap W = t\theta + R_\theta(\ell)$. Then the event $A_{[d-1]}$ is equivalent to $\ell \cap \operatorname{conv}(b_1, \dots, b_{d-1}) \neq \emptyset$. From Lemma 2.15, the joint probability distribution of $b_1, \dots, b_d$ is thus proportional to

$$\operatorname{vol}_{d-1}(\operatorname{conv}(b_1, \dots, b_d)) \cdot \prod_{i=1}^{d} \bar{\mu}_i(b_i) \cdot 1[\ell \cap \operatorname{conv}(b_1, \dots, b_{d-1} \neq \emptyset]$$

Applying Lemma 5.12 to the term $\Pr_{b_1,\dots,b_d}[S]$ we find

$$\begin{aligned}\mathbb{E}_{\theta,t,a_{d+1},\dots,a_n}\Big[\Pr_{b_1,\dots,b_d}[S]\Big] &\leq \mathbb{E}_{\theta,t,a_{d+1},\dots,a_n}\Big[0.1 + \Pr_{b_1,\dots,b_d}[\neg D']\Big] \\ &= 0.1 + \Pr[\neg D'] \leq 0.1 + \Pr[\neg D],\end{aligned}$$

using Fubini's theorem for the equality and the fact that $\neg D'$ implies $\neg D$ for the final inequality. ∎

5.5 Proof of Lemma 5.12: Randomized lower bound for *r* after change of variables

In this section, we deliver the proof of Lemma 5.12. We use the following two technical lemmas. As in from the proof of Lemma 5.12 we define $\lambda \in \mathbb{R}^{d-1}_{\geq 0}$ to be the unique solution to $\sum_{i=1}^{d-1} \lambda_i b_i = \ell \cap \operatorname{conv}(b_1, \dots, b_{d-1})$ and $\sum_{i=1}^{d-1} \lambda_i = 1$. First, we describe Lemma 5.13 which we will use to show that every convex parameter λ_i is at least $\Omega(1/d^2 L)$ with constant probability.

LEMMA 5.13 (Lower-bound for Convex Parameters of Vertices on the Ridge). *Let $b_1, \dots, b_d \in \mathbb{R}^{d-1}$ be random vectors with joint probability density proportional to*

$$\operatorname{vol}_{d-1}(\operatorname{conv}(b_1, \dots, b_d)) \cdot \prod_{i=1}^{d} \bar{\mu}_i(b_i)$$

where each $\bar{\mu}_i : \mathbb{R}^{d-1} \to \mathbb{R}_+$ is L-log-Lipschitz. Let one-dimensional line $\ell \subseteq \mathbb{R}^{d-1}$ and conditional on $\ell \cap \operatorname{conv}(b_i : i \in [d-1]) \neq \emptyset$. Let $\lambda \in \mathbb{R}^{d-1}_+$ be the unique solution to $\sum_{i=1}^{d-1} \lambda_i b_i \in \ell \cap \operatorname{conv}(b_i :$

$i \in [d-1]$). Let D' denote the event that $\forall i, j \in [d]$, $\|b_i - b_j\| \leq 3$. Then we have

$$\Pr\left[\forall i \in [d-1] : \lambda_i \geq \frac{1}{120d^2L} \middle| D' \wedge \ell \cap \operatorname{conv}(b_i : i \in [d-1]) \neq \emptyset\right] \geq 0.95.$$

We will prove Lemma 5.13 in Section 5.6.

Secondly, we present a lemma to lower bound the distance between each vertex b_j (where $j \in [d-1]$) and the $(d-3)$-dimensional hyperplane spanned by the other vertices affhull$(b_j : j \in [d-1], j \neq i)$. Specifically, we show the following lemma:

LEMMA 5.14. *Let $b_1, \ldots, b_d \in \mathbb{R}^{d-1}$ be random vectors with joint probability density proportional to*

$$\operatorname{vol}_{d-1}(\operatorname{conv}(b_1, \ldots, b_d)) \cdot \prod_{i=1}^{d} \bar{\mu}_i(b_i)$$

where each $\bar{\mu}_i : \mathbb{R}^{d-1} \to [0,1]$ is L-log-Lipschitz. Given a one-dimensional line $\ell \subseteq \mathbb{R}^{d-1}$, let $w \in \mathbb{S}^{d-2}$ be any unit direction of ℓ. For any $i \in [d-1]$ we have

$$\Pr\left[\operatorname{dist}\left(\pi_{w^\perp}(b_i), \operatorname{affhull}(\pi_{w^\perp}(b_j) : j \in [d-1], j \neq i)\right) \geq \frac{1}{160d^2L} \middle| \ell \cap \operatorname{conv}(b_i : i \in [d-1]) \neq \emptyset\right] \geq 1 - \frac{1}{20d}.$$

We defer the proof of Lemma 5.14 to Section 5.7. In addition to the above two lemmas, we need the following two basic linear algebraic statements.

FACT 5.15. *Let $\ell \subseteq \mathbb{R}^k$ be an affine line and $x \in \mathbb{R}^k$ be a point. Suppose $w \in \mathbb{S}^{k-1} \setminus \{0\}$ points in the direction of ℓ, i.e., that $\ell + w = \ell$. Then $\operatorname{dist}(x, \ell) = \operatorname{dist}(\pi_{w^\perp}(x), \pi_{w^\perp}(\ell))$.*

PROOF. Let $\{z\} = \ell \cap (x + w^\perp)$. This intersection is non-empty because w points in the direction along ℓ and so must intersect any affine subspace orthogonal to w. The intersection is a singleton because if z, z' were two distinct points in this set then $z - z' \in w\mathbb{R} \cap w^\perp = \{0\}$.

Since $x - z \in w^\perp$, we have $\operatorname{dist}(x, z) = \operatorname{dist}(\pi_{w^\perp}(x), \pi_{w^\perp}(\ell))$. Also notice that for any $z' \in \ell, z' \neq z$,

$$\|x - z'\|^2 = \|\pi_w(x - z')\|^2 + \|\pi_{w^\perp}(x - z')\|^2 \geq \|\pi_{w^\perp}(x - z')\|^2 = \|x - z\|^2.$$

Therefore, $\operatorname{dist}(x, \ell) = \|x - z\| = \operatorname{dist}(\pi_{w^\perp}(x), \pi_{w^\perp}(\ell))$. ■

FACT 5.16. *Let $H \subseteq \mathbb{R}^k$ be a hyperplane, and let $p_1, \ldots, p_k \in H$ and $p_{k+1} \in \mathbb{R}^k$, and assume $\lambda_{k+1} \geq 0$.*

Then

$$\operatorname{dist}\left(\sum_{i=1}^{k+1} \lambda_i p_i, H\right) = \lambda_{k+1} \operatorname{dist}(p_{k+1}, H).$$

PROOF. Let $y \in \mathbb{R}^k, t \in \mathbb{R}$ be such that $H = \{x \in \mathbb{R}^k : y^\top x = t\}$ and $\|y\| = 1$.

Now we have

$$\begin{aligned}\operatorname{dist}(\sum_{i=1}^{k+1}\lambda_i p_i, H) &= |t - y^\top(\sum_{i=1}^{k+1}\lambda_i p_i)| \\ &= |\sum_{i=1}^{k+1}\lambda_i(t - y^\top p_i)| \\ &= |\lambda_{k+1}(t - y^\top p_{k+1})| \\ &= \lambda_{k+1}\operatorname{dist}(p_{k+1}, H),\end{aligned}$$

using that $y^\top p_i = t$ for all $i = 1, \dots, k$. ■

Assuming Lemma 5.13 and Lemma 5.14, we can prove Lemma 5.12.

PROOF OF LEMMA 5.12. We can bound the distance from ℓ to $\partial \operatorname{conv}(b_1, \dots, b_{d-1})$. Note that generically w is not parallel to any direction in $\operatorname{affhull}(b_1, \dots, b_{d-1})$ which makes it so that $\pi_{w^\perp}(\operatorname{conv}(b_1, \dots, b_{d-1}))$ is a simplex of the same dimension as $\operatorname{conv}(b_1, \dots, b_{d-1})$.

$$\begin{aligned}&\operatorname{dist}\left(\ell, \partial \operatorname{conv}(b_1, \dots, b_{d-1})\right) \\ =&\operatorname{dist}\left(\pi_{w^\perp}(\ell), \pi_{w^\perp}(\partial \operatorname{conv}(b_1, \dots, b_{d-1}))\right) && \text{(By Fact 5.15)} \\ =&\operatorname{dist}\left(\pi_{w^\perp}(\ell), \partial \operatorname{conv}(\pi_{w^\perp}(b_1), \dots, \pi_{w^\perp}(b_{d-1}))\right) \\ =&\min_{i\in[d-1]}\operatorname{dist}\left(\pi_{w^\perp}(\ell), \operatorname{conv}(\pi_{w^\perp}(b_1), \dots, \pi_{w^\perp}(b_{i-1}), \pi_{w^\perp}(b_{i+1}), \dots, \pi_{w^\perp}(b_{d-1}))\right) \\ \geq&\min_{i\in[d-1]}\operatorname{dist}\left(\pi_{w^\perp}(\ell), \operatorname{affhull}(\pi_{w^\perp}(b_1), \dots, \pi_{w^\perp}(b_{i-1}), \pi_{w^\perp}(b_{i+1}), \dots, \pi_{w^\perp}(b_{d-1}))\right) \\ =&\min_{i\in[d-1]}\lambda_i \cdot \operatorname{dist}(\pi_{w^\perp}(b_i), \operatorname{affhull}(\pi_{w^\perp}(b_j) : j \in [d-1], j \neq i)) && \text{(By Fact 5.16)} \\ \geq&\min_{i\in[d-1]}\lambda_i \cdot \min_{k\in[d-1]}\operatorname{dist}(\pi_{w^\perp}(b_k), \operatorname{affhull}(\pi_{w^\perp}(b_j) : j \in [d-1], j \neq k))\end{aligned}$$

Where in the second step, we use the fact that $\pi_{w^\perp}$ is an affine isomorphism restricting to $b_1, \dots, b_{d-1}$, thus taking the boundary of $\operatorname{conv}(\pi_{w^\perp}(b_1), \dots, \pi_{w^\perp}(b_{d-1}))$ commutes with the projection $\pi_w^\perp$. In the fifth step, $\lambda \in \mathbb{R}^{d-1}_{\geq 0}$ is the unique solution to $\sum_{i=1}^{d-1}\lambda_i b_i = \ell \cap \operatorname{conv}(b_1, \dots, b_{d-1})$ and $\sum_{i=1}^{d-1}\lambda_i = 1$. Additionally, assume $\ell = w\mathbb{R}$ for a non-zero $w \in \mathbb{S}^{d-2}$. Abbreviate, for $k \in [d-1]$,

$$r_k = \operatorname{dist}(\pi_{w^\perp}(b_k), \operatorname{affhull}(\pi_{w^\perp}(b_j) : j \in [d-1], j \neq k)).$$

Let T denote the event that $\ell \cap \operatorname{conv}(b_1, \dots, b_{d-1}) \neq \emptyset$. We now find using a union bound, for any $\alpha, \beta > 0$,

$$\begin{aligned}&\Pr[\operatorname{dist}(\ell, \partial \operatorname{conv}(b_1, \dots, b_{d-1})) < \alpha\beta \mid T] \\ \leq&\Pr[\min_{i\in[d-1]}\lambda_i < \alpha \mid T] + \Pr[\min_{k\in[d-1]} r_k < \beta \mid T] \\ \leq&\Pr[\min_{i\in[d-1]}\lambda_i < \alpha \mid D' \wedge T] + \Pr[\min_{k\in[d-1]} r_k < \beta \mid T] + \Pr[\neg D' \mid T]\end{aligned}$$

By Lemma 5.13 we know that $\Pr[\min_{i\in[d-1]} \lambda_i < \alpha \mid D' \wedge T] \leq 0.05$ for $\alpha = \frac{1}{120d^2L}$. By Lemma 5.14, we know that $\Pr[\min_{k\in[d-1]} r_k < \beta \mid T] \leq \sum_{k\in[d-1]} \Pr[r_k < \beta \mid T] \leq 0.05$ for $\beta = \frac{1}{160d^2L}$. This proves the lemma. ■

5.6 Proof of Lemma 5.13

Now we show that, with good probability, the convex multipliers λ are not too small.

PROOF OF LEMMA 5.13. By using a union bound, it suffices to prove for each $i \in [d-1]$ that

$$\Pr[\lambda_i < \frac{1}{120d^2L} \mid D' \wedge \ell \cap \mathrm{conv}(b_j : j \in [d-1]) \neq \emptyset] \leq \frac{1}{20(d-1)}.$$

Fix any $i \in [d-1]$, without loss of generality $i = 1$. Recall that $w \in \mathbb{S}^{d-1}$ is such that $\ell = \ell + w$, and hence that $\pi_{w^\perp}(\ell)$ is a singleton point. For ease of exposition we assume that the plane $w^\perp$ is coordinatized such that $\pi_{w^\perp}(\ell) = 0$ and hence $\ell = w\mathbb{R}$. Thus λ is defined to satisfy $\sum_{j=1}^{d-1} \lambda_j \pi_{w^\perp}(b_j) = 0$.

Using the principle of deferred decision, we fix the values of $b_1 - b_j$, $j \in [d]$. This determines the shape of the simplex $\mathrm{conv}(b_j : j \in [d])$, including its volume. The remainder of this proof will use the randomness in the position of the simplex in the subspace orthogonal to the line, which we represent using $\pi_{w^\perp}(b_1)$. For the remainder of this proof, we can consider all b_j, $j \in [d]$ to be functions of b_1. The position $\pi_{w^\perp}(b_1)$ has probability density $\mu'(\pi_{w^\perp}(b_1)) \propto \prod_{j=1}^{d} \mu_j(b_j)$, which is dL-log-Lipschitz in $\pi_{w^\perp}(b_1)$ with respect to the $(d-2)$-dimensional Lebesgue measure on $w^\perp$.

Define $M = \mathrm{conv}(\pi_{w^\perp}(b_1 - b_j) : j \in [d-1]) \subseteq w^\perp$ and note that, due to our fixing the values of $b_1 - b_j$ in the previous paragraph, the shape M is fixed and we can see that that $\pi_{w^\perp}(b_1) \in M$ if and only if $\lambda \geq 0$. It remains to show that

$$\Pr[\lambda_1 < \frac{1}{120d^2L} \mid D' \wedge \pi_{w^\perp}(b_1) \in M] < \frac{1}{20d}. \tag{22}$$

For any $j \in [d-1]$, let $l_j : M \to [0,1]$ be the function sending any point to its j'th convex coefficient, i.e., the functions satisfy $\sum_{j=1}^{d-1} l_j(x) = 1$ and $\sum_{j=1}^{d-1} l_j(x) \cdot \pi_{w^\perp}(b_1 - b_j) = x$ for every $x \in M$. For any $1 \geq \alpha \geq 0$, observe that l_1 takes values in the interval $[\alpha, 1]$ on the set $(1-\alpha)M$. Hence we get

$$\begin{aligned}
\Pr[\lambda_1 \geq \alpha \mid \pi_{w^\perp}(b_1) \in M] &= \frac{\int_M \mu'(x)\mathbf{1}[l_1(x) \geq \alpha]\mathrm{d}x}{\int_M \mu'(x)\mathrm{d}x} \\
&\geq \frac{\int_{(1-\alpha)M} \mu'(x)\mathrm{d}x}{\int_M \mu'(x)\mathrm{d}x} && (\forall x \in (1-\alpha)M, l_1(x) \geq \alpha) \\
&= \frac{(1-\alpha)^{d-2}\int_M \mu'((1-\alpha)x)\mathrm{d}x}{\int_M \mu'(x)\mathrm{d}x}
\end{aligned}$$

$$\geq (1-\alpha)^{d-2} \max_{x \in M} e^{-dL\|\alpha x\|},$$

where in the last inequality, we use dL-log-Lipschitzness of μ' to see that for any $s \in M$ we have $\frac{\mu'((1-\alpha)s)}{\mu'(s)} \leq \max_{x \in M} e^{-dL\|\alpha x\|}$.

By definition of D', we know that M has Euclidean diameter at most 3. Thus we can bound $\|\alpha x\| \leq 3\alpha$ for any $x \in M$. Now take $\alpha = \frac{1}{120d^2L}$, we find

$$\begin{aligned}\Pr[(\lambda_i < \frac{1}{120d^2L}) \mid D' \wedge \pi_{w^\perp}(b_i) \in M] &\leq 1 - \Pr[\lambda_i \geq \frac{1}{120d^2L} \mid D' \wedge \pi_{w^\perp}(b_i) \in M] \\ &\leq 1 - (1 - \frac{1}{120d^2L})^{d-2} e^{-1/40d} \leq \frac{1}{20(d-1)},\end{aligned}$$

where the last inequality comes from $d \geq 3$ and $L \geq 1$. Thus (22) holds as desired. ∎

5.7 Proof of Lemma 5.14

To show Lemma 5.14 on the width of the facet, we need the following upper bound about the mass around zero for a random variable whose density is formed by a log-Lipschitz function multiplied by a convex function. Its function is to deal with the volume term that we receive from the Jacobian in Lemma 2.15.

LEMMA 5.17. *Assume that $h : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is a K-log-Lipschitz function and $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is a convex function such that $\int_{-\infty}^{\infty} g(x) \cdot h(x) \mathrm{d}x = 1$. Suppose that $X \in \mathbb{R}$ is distributed with probability density $g(X) \cdot h(X)$. For any $\varepsilon > 0$ we have* $\Pr[X \in [-\varepsilon, \varepsilon]] \leq 8\varepsilon K$.

PROOF. We can assume that $\varepsilon < 1/(8K)$, for otherwise the bound is trivial. First, we use the rudimentary upper bound

$$\Pr\left[X \in [-\varepsilon, \varepsilon]\right] \leq \Pr\left[X \in [-\varepsilon, \varepsilon] \mid X \in [-1/K, 1/K]\right] = \frac{\int_{-\varepsilon}^{\varepsilon} g(x) \cdot h(x) \mathrm{d}x}{\int_{-1/K}^{1/K} g(x) \cdot h(x) \mathrm{d}x}.$$

Log-Lipschitzness implies that for any $\gamma > 0$ we have

$$e^{-\gamma K} h(0) \int_{-\gamma}^{\gamma} g(x) \mathrm{d}x \leq \int_{-\gamma}^{\gamma} g(x) \cdot h(x) \mathrm{d}x \leq e^{\gamma K} h(0) \int_{-\gamma}^{\gamma} g(x) \mathrm{d}x,$$

and hence we get

$$\Pr[X \in [-\varepsilon, \varepsilon]] \leq e^{(1/K+\varepsilon)K} \frac{\int_{-\varepsilon}^{\varepsilon} g(x) \mathrm{d}x}{\int_{-1/K}^{1/K} g(x) \mathrm{d}x} \leq e^{1+\varepsilon K} \frac{2\varepsilon \cdot \max_{x \in [-\varepsilon,\varepsilon]} g(x)}{\int_{-1/K}^{1/K} g(x) \mathrm{d}x}$$

Since $g(x)$ is convex, at least one of

$$\max_{x \in [-\varepsilon,\varepsilon]} g(x) \leq \min_{x \in [-1/K, -\varepsilon]} g(x) \quad \text{or} \quad \max_{x \in [-\varepsilon,\varepsilon]} g(x) \leq \min_{x \in [\varepsilon, 1/K]} g(x)$$

holds. Without loss of generality, assume the second case holds. Then we bound

$$\frac{\max_{x\in[-\varepsilon,\varepsilon]} g(x)}{\int_{-1/K}^{1/K} g(x)\mathrm{d}x} \leq \frac{\max_{x\in[-\varepsilon,\varepsilon]} g(x)}{\int_{\varepsilon}^{1/K} g(x)\mathrm{d}x} \leq \frac{1}{1/K-\varepsilon}.$$

To summarize, we find $\Pr[X \in [-\varepsilon,\varepsilon]] \leq e^{1+\varepsilon K} \cdot \frac{2\varepsilon}{1/K-\varepsilon}$. Since $\varepsilon < 1/(8K)$ this implies

$$\Pr[X \in [-\varepsilon,\varepsilon]] \leq 2e^{9/8} \cdot \frac{8}{7} \cdot \varepsilon K \leq 8\varepsilon K.$$ ∎

Now we show Lemma 5.14.

PROOF OF LEMMA 5.14. In the following arguments, we condition on $\ell \cap \mathrm{conv}(b_i : i \in [d-1]) \neq \emptyset$. Without loss of generality, set $i = d-1$ and assume that ℓ is a linear subspace, i.e., $\pi_{w^\perp}(\ell) = 0$.

We start with a coordinate transformation. Let $\phi \in w^\perp \cap \mathbb{S}^{d-2}$ denote the unit vector satisfying $\phi^\top b_1 = \phi^\top b_j > 0$ for all $j = 1,\dots,d-1$. Note that ϕ is uniquely defined almost surely: $w^\perp$ is a $(d-2)$-dimensional linear space and we impose $(d-3)$ linear constraints $\{\phi^\top b_1 = \phi^\top b_j, \forall j \in [d-2]\}$. Almost surely, these give a one-dimensional linear subspace which, after adding the unit norm and $b_1^\top\phi > 0$ constraint, leaves a unique choice of ϕ.

Now define $h \in \mathbb{R}$ by $h = \phi^\top b_1$ and define $\alpha \in \mathbb{R}$ by $\alpha h = -\phi^\top b_{d-1}$. Since $0 \in \mathrm{conv}(\pi_{w^\perp}(b_i) : i \in [d-1])$ but $\phi^\top b_i > 0$ for all $i \in [d-2]$, we must have $\alpha \geq 0$ for otherwise ϕ would separate $\mathrm{conv}(\pi_{w^\perp}(b_i) : i \in [d-1])$ from 0. Again from almost-sure non-degeneracy we get $\alpha > 0$ and $h \neq 0$. We define the following coordinate transformation:

$$\begin{aligned} b_j &= h\phi + c_j, \quad \forall j \in [d-2] \\ b_{d-1} &= -\alpha h\phi + c_{d-1} \end{aligned}$$

where for each $j \in [d-1]$, $c_j \in \phi^\perp \cap \mathrm{affhull}(b_1,\dots,b_{d-1})$ has $(d-3)$ degrees of freedom. From here on out, we consider the vertices $(b_1,\dots,b_{d-1})$ to be a function of $(h,\alpha,\phi,c_1,\dots,c_{d-1})$. Again by Lemma 2.15, the induced joint probability density on $(h,\alpha,\phi,c_1,\dots,c_{d-1},b_d)$, is proportional to

$$\mathrm{vol}_{d-1}(\mathrm{conv}(b_1,\dots,b_d)) \cdot \mathrm{vol}_{d-3}(\mathrm{conv}(c_1,\dots,c_{d-2})) \cdot \prod_{j=1}^{d} \bar{\mu}_j(b_j)$$

Using the principle of deferred decision, fix the exact values of $(\alpha,\phi,c_1,\dots,c_{d-1},b_d)$. When this is the case we find that

$$\begin{aligned} &\mathrm{vol}_{d-1}(\mathrm{conv}(b_1,\dots,b_d)) \cdot \mathrm{vol}_{d-3}(\mathrm{conv}(c_1,\dots,c_{d-2})) \cdot \prod_{j=1}^{d} \bar{\mu}_j(b_j) \\ \propto\ &\mathrm{vol}_{d-1}(\mathrm{conv}(b_1,\dots,b_d)) \cdot \mathrm{vol}_{d-3}(\mathrm{conv}(\pi_{w^\perp}(c_1),\dots,\pi_{w^\perp}(c_{d-2}))) \cdot \prod_{j=1}^{d} \bar{\mu}_j(b_j). \end{aligned}$$

To see why $\mathrm{vol}_{d-3}(\mathrm{conv}(c_1,\ldots,c_{d-2}))$ and $\mathrm{vol}_{d-3}(\mathrm{conv}(\pi_{w^\perp}(c_1),\ldots,\pi_{w^\perp}(c_{d-2}))$ are proportional, consider the following. Note that $\pi_{w^\perp}$ is a fixed projection. Generically we know that $w \notin \mathrm{span}(b_2-b_1,\ldots,b_{d-1}-b_1)$, this is a non-degeneracy condition that holds true with probability 1. This implies that $\pi_{w^\perp}$ is a bijection between the affine hyperplane $\mathrm{affhull}(b_1,\ldots,b_{d-1})$ and its image $\pi_{w^\perp}(\mathrm{affhull}(b_1,\ldots,b_{d-1}))$, and also a bijection between the subsets

$$\phi^\perp \cap \mathrm{affhull}(b_1,\ldots,b_{d-1}) \text{ and } \pi_{w^\perp}(\phi^\perp \cap \mathrm{affhull}(b_1,\ldots,b_{d-1})).$$

This is the space where $c_1,\ldots,c_{d-2}$ live. It follows that $\mathrm{conv}(c_1,\ldots,c_{d-2})$ has the same dimension as its projection $\mathrm{conv}(\pi_{w^\perp}(c_1),\ldots,\pi_{w^\perp}(c_{d-2}))$. The ratio between their volumes depends only on w and $\phi^\perp \cap \mathrm{span}(b_2-b_1,\ldots,b_{d-1}-b_1)$. Note that $h=\phi^\perp b_1$ and so, after fixing $(\alpha,\phi,c_1,\ldots,c_{d-1},b_d)$ the ratio between the volumes is constant.

Note that the event $0\in\mathrm{conv}(\pi_{w^\perp}(b_i) : i\in[d-1])$ depends only on these variables and not on h, and the same is true for $\mathrm{vol}_{d-3}(\mathrm{conv}(\pi_{w^\perp}(c_1),\ldots,\pi_{w^\perp}(c_{d-2})))$. We are looking only at the randomness in h, and so we can ignore any constant factors in the probability density function. The induced probability density on h is now proportional to

$$\mathrm{vol}_{d-1}(\mathrm{conv}(b_1,\ldots,b_d)) \cdot \prod_{j=1}^{d-1} \tilde{\mu}_j(h),$$

where $\tilde{\mu}_j(h) := \bar{\mu}_j(h\phi + c_j)$, $j\in[d-2]$ and $\tilde{\mu}_{d-1}(h) := \bar{\mu}_{d-1}(\alpha h\phi + c_{d-1})$. Since each $\bar{\mu}_j$ is L-log-Lipschitz, it follows that the product $\prod_{j=1}^{d-1}\tilde{\mu}_j(h)$ is $(d-2+\alpha)L \le d(1+\alpha)L$-log-Lipschitz in h.

Next, consider the volume term. We can write $\mathrm{vol}_{d-1}(\mathrm{conv}(b_1,\ldots,b_d))$ as a constant depending on d times the absolute value of the determinant of the following $(d-1)\times(d-1)$ matrix

$$\begin{bmatrix} (b_1-b_d)^\top \\ \vdots \\ (b_{d-1}-b_d)^\top \end{bmatrix} = \begin{bmatrix} (h\phi+c_1-b_d)^\top \\ \vdots \\ (h\phi+c_{d-2}-b_d)^\top \\ (-\alpha h\phi + c_{d-1}-b_d)^\top \end{bmatrix} = \begin{bmatrix} (c_1-b_d)^\top \\ \vdots \\ (c_{d-2}-b_d)^\top \\ (c_{d-1}-b_d)^\top \end{bmatrix} + h\cdot \begin{bmatrix} 1 \\ \vdots \\ 1 \\ -\alpha \end{bmatrix} \phi^\top,$$

Define

$$B := \begin{bmatrix} (c_1-b_d)^\top \\ \vdots \\ (c_{d-2}-b_d)^\top \\ (c_{d-1}-b_d)^\top \end{bmatrix}, v := \begin{bmatrix} 1 \\ \vdots \\ 1 \\ -\alpha \end{bmatrix}.$$

Recall that both B and v are fixed and we are only interested in the distribution of h. Then by the matrix determinant lemma, we can write the volume as the absolute value of an affine function of h (which is a convex function):

$$k(h) := \mathrm{vol}_{d-1}(\mathrm{conv}(b_1,\ldots,b_d)) \propto \left|\det(B + hv\phi^\top)\right|$$

$$= \left|\det(B)(1 + j\phi^\top B^{-1}v)\right|$$

Hence, we have found a convex function $k : \mathbb{R} \to \mathbb{R}_{\geq 0}$ and a $d(1+\alpha)L$-log-Lipschitz function $v : \mathbb{R} \to \mathbb{R}_{\geq 0}$ such that h has probability density proportional to $k(h) \cdot v(h)$.

To finalize the argument, we write $\mathrm{dist}(\pi_{W^\perp}(b_i), \mathrm{affhull}(\pi_{W^\perp}(b_j) : j \in [d-1], j \neq i)) = |(1+\alpha)h|$. It follows that the signed distance $(1+\alpha)h$ has a probability density function proportional to the product of a dL-log-Lipschitz function and a convex function. The result follows from Lemma 5.17 by plugging in the signed distance $(1+\alpha)h$, $K = dL$, and $\varepsilon = \frac{1}{160d^2L}$. ■

5.8 Combining Together and Proof of Lemma 5.2

In this section, we combine the deterministic argument in Lemma 5.5 with the probabilistic arguments in Lemma 5.7 and Lemma 5.11. We can finally show the main technical lemma (Lemma 5.2).

PROOF OF LEMMA 5.2. Without loss of generality, let $I = [d]$ and write $E = E_I$. Suppose $p' = A_J = \mathrm{conv}(a_j : j \in J) \cap W$ is the next vertex after the edge $F_I \cap W$. Here $J \in \binom{[n]}{d-1}$ and R_J is the $(d-2)$-dimensional ridge. With probability 1, the polytope $\mathrm{conv}(a_1, \ldots, a_n)$ is non-degenerate and $W \cap R'$ is a single point for any ridge R' of $\mathrm{conv}(a_1, \ldots, a_n)$ that intersects with W. We will show that conditional on E, each of the following conditions in the deterministic argument (Lemma 5.5) is satisfied with good probability:

1. (Bounded diameter) $\forall i, j \in [n]$, $\|a_i - a_j\| \leq 3$;
2. (Lower bound of δ) $\min_{k \in [n] \setminus I} \mathrm{dist}(\mathrm{affhull}(F_I), a_k) \geq \Omega(\frac{1}{Ld\log n})$;
3. (Lower bound of r) $\forall J \in \binom{I}{d-1}$ for which the ridge $R_J = \mathrm{conv}(a_j : j \in J)$ has nonempty intersection with W, we have $\mathrm{dist}(F_I \cap W, \partial R_J) \geq \Omega(\frac{1}{d^4L^2})$.

Note for the last point that Lemma 5.5 only requires this for the set J which indexes the second vertex of $F_I \cap W$ in clockwise direction, but we prove it for both of the sets J for which $R_J \cap W \neq \emptyset$.

First, we write D as the event that $\forall i, j \in [n]$ for which $\|a_i - a_j\| \leq 3$. From Lemma 2.12, for any $\sigma \leq \frac{1}{8\sqrt{d\log n}}$, with probability at least $1 - \binom{n}{d}^{-1}$, we have $\max_{i \in [n]} \|\hat{a}_i\| \leq 1 + 4\sigma\sqrt{d\log n} \leq \frac{3}{2}$, i.e., $\Pr[D] \geq 1 - \binom{n}{d}^{-1}$. Using the assumption that $\Pr[E_I] \geq 10\binom{n}{d}^{-1}$, we have

$$\Pr[\neg D \mid E] = \Pr[\neg D \wedge E]/\Pr[E] \leq \frac{\Pr[\neg D]}{\Pr[E]} \leq 0.1,$$

This immediately implies $\Pr[D \mid E] \geq 0.9$.

Next, we consider $\delta := \mathrm{dist}(\mathrm{affhull}(a_1, \ldots, a_d), \{a_{d+1}, \ldots, a_n\})$. Using Lemma 5.7, we have $\Pr[\delta \geq \frac{1}{10e^3Ld\log n} \mid E] \geq 0.72$.

Finally, we consider $r := \max_J \mathrm{dist}(\mathrm{affhull}(a_1, \ldots, a_d) \cap W, \partial R_J)$ subject to all $J \in \binom{I}{d-1}$ such that A_J happens (in other words, $R_J = \mathrm{conv}(a_j : j \in J)$ is a ridge of F_I such that $R_J \cap W \neq \emptyset$).

By a union bound,

$$\begin{aligned}
&\Pr\left[\exists J \in \binom{I}{d-1}, A_J \wedge \operatorname{dist}(\operatorname{affhull}(F \cap W), \partial R_J \geq \frac{1}{19200 d^4 L^2} \mid E\right] \\
\geq& 1 - \sum_{J \in \binom{I}{d-1}} \Pr[A_J \wedge \operatorname{dist}(\operatorname{affhull}(F \cap W), \partial R_J < \frac{1}{19200 d^4 L^2} \mid E] \\
=& 1 - \sum_{J \in \binom{I}{d-1}} \Pr[\operatorname{dist}(\operatorname{affhull}(F \cap W), \partial R_J < \frac{1}{19200 d^4 L^2} \mid E \wedge A_J] \Pr[A_J \mid E]. \qquad (23)
\end{aligned}$$

From Lemma 5.11, for each $J \in \binom{I}{d-1}$, we know that

$$\Pr[\operatorname{dist}(\operatorname{affhull}(F \cap W), \partial R_J < \frac{1}{19200 d^4 L^2} \mid E \wedge A_J] \leq 0.1 + \Pr[\neg D \mid E \wedge A_J],$$

Notice that when E happens, there are exactly two distinct ridges $R_J, R_{J'}$ that has nonempty intersection with W (or A_J happens), thus $\sum_{J \in \binom{I}{d-1}} \Pr[A_J \mid E] = 2$. Therefore

$$\begin{aligned}
&\sum_{J \in \binom{I}{d-1}} \Pr[\operatorname{dist}(\operatorname{affhull}(F \cap W), \partial R_J < \frac{1}{19200 d^4 L^2} \mid E \wedge A_J] \Pr[A_J \mid E] \\
\leq& \sum_{J \in \binom{I}{d-1}} (0.1 + \Pr[\neg D \mid E \wedge A_J]) \Pr[A_J \mid E] \\
\leq& 0.1 \cdot 2 - 2 \cdot \Pr[\neg D \mid E],
\end{aligned}$$

and (23) becomes

$$\begin{aligned}
&\Pr\left[\exists J \in \binom{I}{d-1}, A_J \wedge \operatorname{dist}(\operatorname{affhull}(F \cap W), \partial R_J \geq \frac{1}{19200 d^4 L^2} \mid E\right] \\
\geq& 1 - 0.1 \cdot 2 - 2 \cdot \Pr[\neg D \mid E] \geq 0.6.
\end{aligned}$$

Therefore, by a union bound, the three conditions hold with probability at least $1 - (1 - 0.9) - (1 - 0.72) - (1 - 0.6) \geq 0.1$, and the lemma directly follows from Lemma 5.5. ■

6. Smoothed Complexity Lower Bound

In this section, we present the lower bound of the smoothed complexity of the shadow vertex simplex method by studying the intersection of the smoothed polar polytope $\operatorname{conv}(a_1, \ldots, a_n) \subseteq \mathbb{R}^d$ (where each a_i is a Gaussian perturbation from a $\bar{a}_i$), and the two-dimensional shadow plane $W \subseteq \mathbb{R}^d$. Our main result is as follows:

THEOREM 6.1. *For any $d > 5, n = 4d - 13$, there exists a two-dimensional linear subspace $W \subseteq \mathbb{R}^d$ and vectors $\bar{a}_1, \ldots, \bar{a}_n \in \mathbb{R}^d$, $\max_{i \in n} \|\bar{a}_i\|_1 \leq 1$ such that the following holds. Let $a_1, \ldots, a_n$ be independent Gaussian random variables where each $a_i \sim \mathcal{N}_d(\bar{a}_i, \sigma^2 I), \sigma \leq \frac{1}{7200 d \sqrt{\log n}}$. Then* *with probability at least* $1 - \binom{n}{d}^{-1}$ *we have*

$$\mathrm{edges}(\mathrm{conv}(a_1, \dots, a_n) \cap W) \geq \Omega\left(\min\left(\frac{1}{\sqrt{d\sigma\sqrt{\log n}}}, 2^d\right)\right).$$

Theorem 6.1 is a direct consequence of the next theorem, which is a lower bound for adversarial perturbations of bounded magnitude:

THEOREM 6.2. *For any* $d > 5, n = 4d - 13$, *there exists a two-dimensional linear subspace* $W \subseteq \mathbb{R}^d$ *and vectors* $\bar{a}_1, \dots, \bar{a}_n \in \mathbb{R}^d$, $\max_{i \in n} \|\bar{a}_i\|_1 \leq 1$ *such that the following holds. For any* $\varepsilon < \frac{1}{180}$, *if* $a_1, \dots, a_n \in \mathbb{R}^d$ *satisfy* $\|a_i - \bar{a}_i\|_1 \leq \varepsilon$ *for all* $i \in [n]$ *then we have*

$$\mathrm{edges}(\mathrm{conv}(a_1, \dots, a_n) \cap W) \geq \Omega\left(\min\left(\frac{1}{\sqrt{\varepsilon}}, 2^d\right)\right).$$

The rest of this section is organized as follows. In Section 6.1, we construct an auxiliary polytope $P \subseteq \mathbb{R}^d$ and a two-dimensional shadow plane W. In Section 6.2, we show that the projection $\pi_W(P)$ approximates the unit disk $\mathbb{B}_2^2$. In Section 6.3, we analyse the largest ℓ_∞-ball contained in P and the smallest ℓ_∞-ball containing P. Section 6.4 investigates the polar polytope $Q = (P - x)^\circ$ of a shift of P, such that we may choose $\bar{a}_1, \dots, \bar{a}_n$ to satisfy $Q = \mathrm{conv}(\bar{a}_1, \dots, \bar{a}_n)$. The largest contained ℓ_1-ball in Q is derived from the smallest ℓ_∞-ball containing P, and the smallest ℓ_1-ball containing Q is derived from the largest ℓ_∞-ball contained in P. Similarly, the section $Q \cap W$ approximates a circular disk because the projection $\pi_W(P)$ approximates a circular disk. Finally, Section 6.5 shows that the bounded ratio between the inner and outer radii implies that any sufficiently small perturbation $\tilde{Q}$ still has $\tilde{Q} \cap W$ approximate the unit disk $\mathbb{B}_2^2$ well and uses this to prove Theorem 6.1.

6.1 Construction of the Auxiliary Polytope

In this subsection, we construct the auxiliary polytope P and the two-dimensional plane W. For $k \in \mathbb{N}$, we construct a $(k + 5)$-dimensional polytope. We will use the following vectors in the definition:

— Define $e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \in \mathbb{R}^2$ and $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

— For every $i \in \{0, 1, \dots, k\}$, define the pair of orthogonal unit vectors $w_i = \begin{bmatrix} \cos(\pi/2^{i+2}) \\ \sin(\pi/2^{i+2}) \end{bmatrix} \in \mathbb{R}^2$ and $v_i = \begin{bmatrix} \sin(\pi/2^{i+2}) \\ -\cos(\pi/2^{i+2}) \end{bmatrix} \in \mathbb{R}^2$.

With these definitions in mind, construct an auxiliary $P' \subseteq \mathbb{R}^{3k+5}$ as the set of points $(x, y, p_0, \dots, p_k, t, s)$ where $x, y \in \mathbb{R}$, $p_0, p_1, \dots, p_k \in \mathbb{R}^2$, $t \in \mathbb{R}^k$ and $s \in \mathbb{R}$ satisfy the following system of linear inequalities:

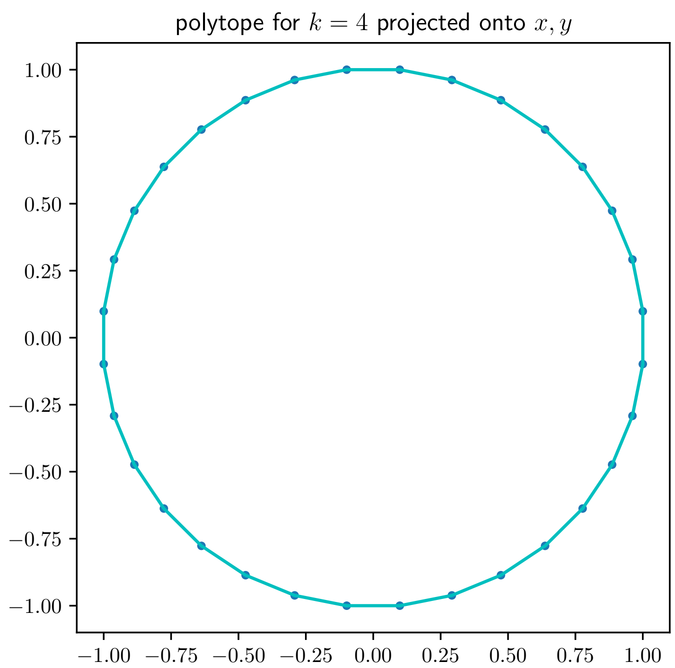

Figure 4. Vertices of the projected auxiliary polytope $\pi_W(P)$ (see (30)) without perturbation for $k = 4$.

$$e_1^\top p_0 \geq |x|, e_2^\top p_0 \geq |y| \tag{24}$$

$$w_i^\top p_i = w_i^\top p_{i-1}, \ \forall i \in [k] \tag{25}$$

$$t_i + is = v_i^\top p_i \geq |v_i^\top p_{i-1}|, \ \forall i \in [k] \tag{26}$$

$$e_1^\top p_k \leq 1 \tag{27}$$

$$\mathbf{0}_k \leq t \leq \mathbf{1}_k \tag{28}$$

$$0 \leq s \leq 1. \tag{29}$$

We remark that p_0, t, s uniquely define the values of $p_1, p_2, \dots, p_k$ via (25) and (26). As such, define the polytope $P \subseteq \mathbb{R}^{k+5}$ as the projection of P' onto the subspace spanned by the variables (x, y, p_0, t, s):

$$P = \{(x, y, p_0, t, s) : \exists\, p_1, \dots, p_k, \text{ s.t. } (x, y, p_0, \dots, p_k, t, s) \in P'\}. \tag{30}$$

We choose the plane W to be the one that is spanned by the unit vectors in the x and y directions.

An illustration of the vertices of the projected polytope $\pi_W(P)$ can be found in Figure 4 for $k = 4$. Note that the figure appears to depict a regular polygon with 2^{k+1} vertices. Our construction of P' is similar to those of [8, 28]. The primary difference lies in the addition of the variables t and s in (26). This change is made to ensure that the projected polytope P has its largest contained and smallest containing ℓ_∞-ball be of similar sizes.

6.2 Projected Auxiliary Polytope Approximates Two-Dimensional Unit Disk

In this subsection, we will show that the polytope P we constructed in (30) has a projection $\pi_W(P)$ which approximates the two-dimensional unit disk $\mathbb{B}_2^2 = \{x, y \in \mathbb{R} : x^2 + y^2 \leq 1\}$ within exponentially small error:

LEMMA 6.3 (Projected auxiliary polytope approximates the two-dimensional disk). *For any $k \in \mathbb{N}$, let $P \subseteq \mathbb{R}^{k+5}$ be the polytope defined by the linear system* (30) *with variables $x, y, s \in \mathbb{R}$, $p_0 \in \mathbb{R}^2$, $t \in \mathbb{R}^k$. Let W be the two-dimensional subspace spanned by the directions of x and y. Then we have*

$$\mathbb{B}_2^2 \subseteq \pi_W(P) \subseteq \cos(\pi/2^{k+2})^{-1}\mathbb{B}_2^2.$$

Lemma 6.3 directly follows from the next two lemmas. First, we show that the two-dimensional unit disk is contained in $\pi_W(P)$:

LEMMA 6.4 (Inner ℓ_2-radius of the projected auxiliary polytope). *For every $x, y \in \mathbb{R}$ with $x^2 + y^2 \leq 1$ there exist $p_0 \in \mathbb{R}^2$, $t \in \mathbb{R}^k$ and $s \in \mathbb{R}$ such that $(x, y, p_0, t, s) \in P$.*

PROOF. Suppose $x, y \in \mathbb{R}$ with $x^2 + y^2 \leq 1$. We want to exhibit $p_0 \in \mathbb{R}^2$, $t \in \mathbb{R}^k$, and $s \in \mathbb{R}$ such that $(x, y, p_0, t, s) \in P$. We proceed as follows:

First, set $s = 0$ and define $p_0 = (x, y)$. Then define p_i inductively by the recurrence

$$p_i = \begin{bmatrix} v_i^\top \\ w_i^\top \end{bmatrix}^{-1} \begin{bmatrix} |v_{i-1}^\top p_{i-1}| \\ w_{i-1}^\top p_{i-1} \end{bmatrix},$$

with the base case being $p_0 = (|x|, |y|)$. Finally, for each $i \in [k]$ define $t_i = v_i^\top p_i = |v_i^\top p_{i-1}|$.

Then (24), (25), (26) and (29) follow by definition. It remains to argue that (27) and (28) are satisfied. Since v_i and w_i are orthogonal and defined to have norm 1, the matrix $\begin{bmatrix} v_i^\top \\ w_i^\top \end{bmatrix}$ is an isometry. Thus, it preserves norms, meaning that each of the p_i have ℓ_2-norm $\sqrt{x^2 + y^2}$. Since $t_i = |v_i^\top p_{i-1}| \leq \|v_i\| \cdot \|p_{i-1}\| \leq 1$, we know that (28) is satisfied. Furthermore, we have $e_1^\top p_k \leq \|e_1\| \cdot \|p_k\| = \|e_1\| \cdot \sqrt{x^2 + y^2} \leq 1$, which ensures that (27) is satisfied. ∎

In the next lemma, we show that the projection $\pi_W(P)$ is contained in the two-dimensional disk $\cos(\pi/2^{k+2})^{-1}\mathbb{B}_2^2$:

LEMMA 6.5 (Outer ℓ_2-radius of the projected auxiliary polytope). *For every $x, y \in \mathbb{R}$ such that $\sqrt{x^2 + y^2} \geq \cos(\pi/2^{k+2})^{-1}$, there exist no $p_0 \in \mathbb{R}^2$, $t \in \mathbb{R}^k$ and $s \in \mathbb{R}$ such that $(x, y, p_0, t, s) \in P$.*

PROOF. Fix any $(x, y) \in \mathbb{R}^2$ and $p_0 \in \mathbb{R}^2$, such that $\sqrt{x^2 + y^2} > \cos(\pi/2^{k+2})^{-1}$ and $p_0 \geq \begin{bmatrix} |x|, |y| \end{bmatrix}^\top$. Also fix any $p_1, \dots, p_k \in \mathbb{R}^2$ satisfying (25) and (26). We will show that such $p_1, \dots, p_k$ would violate (27), i.e., $e_1^\top p_k > 1$. To simplify our notation, for all $i \in \{0, 1, \dots, k\}$, let $(p_i)_v = v_i^\top p_i \in \mathbb{R}$ and $(p_i)_w = w_i^\top p_i \in \mathbb{R}$. Then $p_i = (p_i)_v v_i + (p_i)_w w_i$.

Notice that for all $i \in [k]$, the increment of the first coordinate from p_{i-1} to p_i is

$$\begin{aligned}
e_1^\top p_i - e_1^\top p_{i-1} &= e_1^\top \left((p_i)_w w_i + (p_i)_v v_i - (w_i^\top p_{i-1})w_i - (v_i^\top p_{i-1})v_i\right) \\
&= e_1^\top \left((p_i)_v v_i - (v_i^\top p_{i-1})v_i\right) && \text{(By (25))} \\
&\geq e_1^\top v_i \left(|v_i^\top p_{i-1}| - v_i^\top p_{i-1}\right) && \text{(By } e_1^\top v_i > 0 \text{ and (26))} \quad (31) \\
&\geq 0
\end{aligned}$$

where the inequality in (31) is tight when $v_i^\top p_i = |v_i^\top p_{i-1}|$. Let $p_0^*, p_1^*, \ldots, p_k^* \in \mathbb{R}^2$ be the (unique) sequence defined by

$$\begin{aligned}
p_0^* &= p_0 \\
w_i^\top p_i^* &= w_i^\top p_{i-1}^*, && \forall i \in [k] && \text{(Tight for (25))} \\
v_i^\top p_i^* &= |v_i^\top p_{i-1}^*|, && \forall i \in [k] && \text{(Tight for (26))}
\end{aligned}$$

Then $e_1^\top p_i - e_1^\top p_{i-1} \geq e_1^\top p_i^* - e_1^\top p_{i-1}^*$ for each $i \in [k]$. Also, notice that $e_1^\top p_0 = e_1^\top p_0^*$, therefore for each $i \in [k]$, $e_1^\top p_i \geq e_1^\top p_i^*$.

It remains to show that $e_1^\top p_k^* > 1$. For all $i \in \{0, 1, \ldots, k\}$, let $\theta_i \in [-\pi, \pi]$ denote the angle between $p_i^* \in \mathbb{R}^2$ and e_1. Then since $e_1^\top p_0 \geq 0$ and $e_2^\top p_0 \geq 0$, we have $0 \leq \theta_0 \leq \frac{\pi}{2}$. For any $i \in [k]$, notice that p_i^* equals to p_{i-1}^* (if $\theta_{i-1} \leq \frac{\pi}{2^{i+2}}$), or equals to the mirror of p_{i-1}^* with respect to the line spanned by w_i (if $\theta_{i-1} \geq \frac{\pi}{2^{i+2}}$). By induction, this gives

$$\|p_i^*\|_2 = \|p_{i-1}^*\|_2 = \ldots = \|p_0^*\|_2 = \|p_0\|_2,$$

and

$$\theta_i = \frac{\pi}{2^{i+2}} - |\theta_{i-1} - \frac{\pi}{2^{i+2}}| \leq \frac{\pi}{2^{i+2}}.$$

Therefore, we get

$$e_1^\top p_k^* = \|p_k^*\| \cdot \cos(\theta_k) \geq \|p_0\| \cdot \cos\left(\frac{\pi}{2^{k+2}}\right) > 1.$$

Thus we have shown $x^2 + y^2 > \cos(\pi/2^{k+2})^{-1}$ implies that $e_1^\top p_k \geq e_1^\top p_k^* > 1$ as desired. ∎

6.3 Inner and Outer ℓ_∞-Radius of the Auxiliary Polytope

In this subsection, we will show that the auxiliary polytope P has large inner ℓ_∞-radius and small outer ℓ_∞-radius.

LEMMA 6.6 (Inner and Outer ℓ_∞-Radius of the Auxiliary Polytope). *For $k \in \mathbb{N}$, let $P \subseteq \mathbb{R}^{k+5}$ be the polytope defined by the linear system* (30). *Then for or $\bar{x} = \bar{y} = 0, \bar{p}_0 = (1/6, 1/6)^\top, \bar{t} = \mathbf{1}_k/30, \bar{s} = \frac{1}{3}$, it holds that*

$$\frac{1}{30} \cdot \mathbb{B}_\infty^{k+5} \subseteq P - (\bar{x}, \bar{y}, \bar{p}_0, \bar{t}, \bar{s}) \subseteq \frac{3}{2}\mathbb{B}_\infty^{k+5}.$$

The following lemma is the key to show Lemma 6.6, where we construct a point in B such that the ℓ_∞ ball with radius $\frac{1}{30}$ centered at that point is contained in P. Obtaining a large inner ball is relatively easy using linear programming. In this exposition we demonstrate the large inner ball by hand, which will require a somewhat tedious calculation to 3 significant digits.

LEMMA 6.7 (Inner ℓ_∞-Radius of the Polytope). *For $\bar{x} = \bar{y} = 0$, $\bar{p}_0 = (1/6, 1/6)^\top$, $\bar{t} = \mathbf{1}_k/30$, $\bar{s} = \frac{1}{3}$, we have $(\bar{x}, \bar{y}, \bar{p}_0, \bar{t}, \bar{s}) + r \cdot \mathbb{B}_\infty^{k+5} \subseteq P$ for $r = \frac{1}{30}$.*

PROOF. Fix any $(x, y, p_0, t, s) \in \mathbb{R}^{k+5}$ such that $\|(x - \bar{x}, y - \bar{y}, p_0 - \bar{p}_0, t - \bar{t}, s - \bar{s})\|_\infty \le r$. We set $p_1, \ldots, p_k \in \mathbb{R}^2$ inductively by (25) and the equations in (26), i.e., $w_i^\top p_i = w_i^\top p_{i-1}$ and $t_i + is = v_i^\top p_i$ start from the base case of p_0, s and t.

We will show that $(x, y, p_0, t, s) \in P$ by verifying the conditions in (24 - 29). To simplify our notation, we define $(p_i)_v = v_i^\top p_i \in \mathbb{R}$ and $(p_i)_w = w_i^\top p_i \in \mathbb{R}$ for all $i \in \{0, 1, \ldots, k\}$. Note that v_i and w_i are orthogonal, $p_i = (p_i)_v v_i + (p_i)_w w_i$.

First, observe that $e_1^\top p_0 \ge \frac{1}{6} - r \ge r \ge |x|$ and $e_2^\top p_0 \ge \frac{1}{6} - r \ge r \ge |y|$, confirming that (24) holds. Also, notice that $t_i \in [\bar{t}_i - r, \bar{t}_i + r] \subseteq [0, 1]$ and $s \in [\bar{s} - r, \bar{s} + r] \subseteq [0, 1]$. Thus (28) and (29) hold. The equality constraint (25) holds directly by definition of $p_1, \ldots, p_k$. It remains to show (26) and (27), i.e., $(p_i)_v \ge |v_i^\top p_{i-1}|$ for all $i \in [k]$ and $e_1^\top p_k \le 1$.

To aid in the remaining steps of the proof, we show the claim that $(p_i)_w \ge 0$ for all $i \in \{0, 1, \ldots, k\}$ by induction. Observe that $w_0^\top p_0 \ge w_0^\top \bar{p}_0 - \|w_0\| \cdot \|p_0 - \bar{p}_0\| \ge \frac{\sqrt{2}}{6} - \sqrt{2} r \ge 0$. Also, for all $i \in [k]$,

$$
\begin{aligned}
(p_i)_w &= w_i^\top p_{i-1} && \text{(By (25))} \\
&= w_i^\top \left((p_{i-1})_w w_{i-1} + (p_{i-1})_v v_{i-1}\right) \\
&= (p_{i-1})_w w_i^\top w_{i-1} + (p_{i-1})_v w_i^\top v_{i-1} \\
&= (p_{i-1})_w \cdot \cos\left(\frac{\pi}{2^{i+2}}\right) + (p_{i-1})_v \cdot \sin\left(\frac{\pi}{2^{i+2}}\right) && (32) \\
&\ge (p_{i-1})_w \cdot \cos\left(\frac{\pi}{2^{i+2}}\right), && \text{(By } (p_{i-1})_v = t_{i-1} + (i-1)s \ge 0\text{)}
\end{aligned}
$$

where (32) follows from $w_i^\top w_{i-1} = \|w_i\| \|w_{i-1}\| \cos(\theta) = \cos(\theta)$ where $\theta = \pi/2^{i+2}$ is the angle between w_i and w_{i-1}; similarly, we have $w_i^\top v_{i-1} = \|w_i\| \|v_{i-1}\| \cos(\pi/2 - \theta) = \sin(\theta) = \sin(\pi/2^{i+2})$. It then follows by induction that $(p_i)_w \ge 0$ for all $i \in \{0, 1, \ldots, k\}$.

Verify (26): To verify the inequality listed in (26), i.e., $(p_i)_v \ge |v_i^\top p_{i-1}|$ for all $i \in [k]$, we upper-bound the right-hand-side by expanding it into the inner product with v_{i-1} and the inner product with w_{i-1}. Notice that for all $i \in [k]$,

$$
\begin{aligned}
|v_i^\top p_{i-1}| &= \left|v_i^\top \left((p_{i-1})_v v_{i-1} + (p_{i-1})_w w_{i-1}\right)\right| \\
&\le |(p_{i-1})_v| v_i^\top v_{i-1} + |(p_{i-1})_w| \cdot |v_i^\top w_{i-1}| && \text{(Triangle inequality)} \\
&= (t_{i-1} + (i-1)s) \cdot \cos\left(\frac{\pi}{2^{i+2}}\right) + (p_{i-1})_w \cdot \sin\left(\frac{\pi}{2^{i+2}}\right) && \text{(By (26) and } (p_{i-1})_w \ge 0\text{)} \quad (33)
\end{aligned}
$$

Next, we require an upper bound on $(p_{i-1})_w$. For all $i \in [k]$, from (32)

$$\begin{aligned}(p_i)_w &= (p_{i-1})_w \cdot \cos(\frac{\pi}{2^{i+2}}) + (p_{i-1})_v \cdot \sin(\frac{\pi}{2^{i+2}}) \\ &\le (p_{i-1})_w + (t_{i-1} + (i-1)s) \cdot \sin(\frac{\pi}{2^{i+2}}) && \text{(By } (p_{i-1})_w \ge 0 \text{ and (26))}\end{aligned}$$

Let $t_0 = v_0^\top p_0$ and $\bar{t}_0 = v_0^\top \bar{p}_0 = 0$. By applying the above inequality to $1, 2, \cdots, i-1$, we have

$$\begin{aligned}(p_i)_w &\le w_0^\top p_0 + \sum_{j=0}^{i-1} (t_j + js) \cdot \sin(\frac{\pi}{2^{j+3}}) \\ &\le w_0^\top p_0 + \sum_{j=0}^{i-1} (t_j + js) \cdot \sin(\frac{\pi}{8}) \cdot 1.9^{-j} \\ &\le w_0^\top p_0 + \sum_{j=0}^{i-1} (\bar{t}_j + r + js) \cdot \sin(\frac{\pi}{8}) \cdot 1.9^{-j} \\ &\le (\frac{\sqrt{2}}{6} + \sqrt{2}r) + \sin(\frac{\pi}{8}) \sum_{j=0}^{i-1} (\frac{1}{30} + r + js) \cdot 1.9^{-j} && \text{(By } w_0^\top \bar{p}_0 = \tfrac{\sqrt{2}}{6}, \bar{t}_j = \tfrac{1}{30}) \\ &\le \frac{\sqrt{2}}{6} + \sin(\frac{\pi}{8}) \cdot \frac{1}{30} \cdot \sum_{j=0}^{\infty} 1.9^{-j} + r\left(\sqrt{2} + \sin(\frac{\pi}{8}) \sum_{j=0}^{\infty} 1.9^{-j}\right) + s\left(\sin(\frac{\pi}{8}) \sum_{j=0}^{\infty} j(1.9)^{-j}\right) \\ &\le 0.263 + 2.226r + 0.898s. && (34)\end{aligned}$$

Here the second inequality uses the fact that for every $x \in [0, \frac{\pi}{8}]$ one has $\sin(x)/1.9 \ge \sin(x/2)$. This is because by the half angle formula, $\sin(x/2) = \pm\sqrt{\frac{1-\cos(x)}{2}}$, thus

$$\frac{\sin(x)^2}{\sin(x/2)^2} = \frac{1-\cos(x)^2}{(1-\cos(x))/2} = 2 + 2\cos(x) \ge 2 + 2\cos(\pi/8) \ge 1.9^2.$$

Plugging (34) back into (33), we have for all $i \in [k]$,

$$\begin{aligned}|v_i^\top p_{i-1}| &\le (t_{i-1} + (i-1)s) \cdot \cos(\frac{\pi}{2^{i+2}}) + (0.263 + 2.226r + 0.898s) \cdot \sin(\frac{\pi}{2^{i+2}}) \\ &\le (\frac{1}{30} + r + (i-1)s) + (0.263 + 2.226r + 0.898s) \cdot \sin(\frac{\pi}{8}) && \text{(By } \bar{t}_{i-1} = \tfrac{1}{30} \text{ and } i \ge 1) \\ &\le 0.134 + 1.852r + (i - 0.656)s \\ &\le 0.134 + 1.852r + is - 0.656(\bar{s} - r) \\ &\le is && \text{(By } \bar{s} = \tfrac{1}{3}, r \le \tfrac{1}{30}) \\ &\le (p_i)_v.\end{aligned}$$

Therefore, $v_i^\top p_i \ge |v_i^\top p_{i-1}|$ for all $i \in [k]$ and (26) holds.

Verify (27): To verify (27), notice that increment of the first coordinate from p_{i-1} to p_i is

$$e_1^\top p_i - e_1^\top p_{i-1} = e_1^\top \left((p_i)_v v_i - (v_i^\top p_{i-1}) v_i\right) \qquad \text{(By } w_i^\top p_i = w_i^\top p_{i-1})$$

$$
\begin{aligned}
&= \sin(\frac{\pi}{2^{i+2}}) \cdot (t_i + is - v_i^\top p_{i-1}) && (\text{By } e_1^\top v_i = \sin(\tfrac{\pi}{2^{i+2}}) \text{ and (26)})\\
&= \sin(\frac{\pi}{2^{i+2}}) \cdot (t_i + is - v_i^\top((p_{i-1})_w w_{i-1} + (p_{i-1})_v v_{i-1}))\\
&= \sin(\frac{\pi}{2^{i+2}}) \cdot \Big(t_i + is + (p_{i-1})_w \cdot \sin(\frac{\pi}{2^{i+2}}) - (t_{i-1} + (i-1)s) \cdot \cos(\frac{\pi}{2^{i+2}})\Big) && (35)
\end{aligned}
$$

where the last step comes from $v_i^\top v_{i-1} = \cos(\frac{\pi}{2^{i+2}})$ and $v_i^\top w_{i-1} = -\sin(\frac{\pi}{2^{i+2}})$. For all $i \geq 2$, we can show that in (35), the third term in the brackets is at most the fourth term, thus the right-hand-side of (35) is at most $\sin(\pi/2^{i+2}) \cdot (t_i + is)$:

$$
\begin{aligned}
(p_{i-1})_w \cdot \sin(\frac{\pi}{2^{i+2}}) &\leq (0.263 + 2.226r + 0.898s) \cdot \sin(\frac{\pi}{2^{i+2}}) && (\text{By (34)})\\
&\leq (0.263 + 2.226r + 0.898s) \cdot \tan(\frac{\pi}{2^{i+2}}) \cdot \cos(\frac{\pi}{2^{i+2}})\\
&\leq (0.263 + 2.226r + 0.898s) \cdot \tan(\frac{\pi}{8}) \cdot \cos(\frac{\pi}{2^{i+2}}) && (\text{By } \tfrac{\pi}{2^{i+2}} \leq \tfrac{\pi}{8})\\
&\leq 0.277 \cdot \cos(\frac{\pi}{2^{i+2}}) && (\text{By } r \leq \tfrac{1}{30} \text{ and } s \leq \bar{s} + r = \tfrac{11}{30})\\
&\leq (\bar{t}_{i-1} - r + (i-1)s) \cdot \cos(\frac{\pi}{2^{i+2}})\\
& && (\text{By } \bar{t}_{i-1} - r = 0 \text{ and } (i-1)s \geq s \geq \bar{s} - r = 0.3)\\
&\leq (t_{i-1} + (i-1)s) \cdot \cos(\frac{\pi}{2^{i+2}}).
\end{aligned}
$$

Plugging back into (35), we have for all $i \geq 2$ that

$$
\begin{aligned}
e_1^\top p_i - e_1^\top p_{i-1} &\leq \sin(\frac{\pi}{2^{i+2}}) \cdot (t_i + is)\\
&\leq \sin(\frac{\pi}{2^{i+2}}) \cdot (\bar{t}_i + i\bar{s} + (i+1)r)\\
&\leq \sin(\frac{\pi}{8}) \cdot 1.9^{-(i-1)} \cdot (\frac{1}{30} + \frac{i}{3} + (i+1)r) && (\text{By } \sin(\tfrac{\pi}{2^{i+2}}) \leq \sin(\tfrac{\pi}{8}) \cdot 1.9^{-(i-1)})
\end{aligned}
$$

For $i = 1$ we recall from above that $(p_{i-1})_w \cdot \sin(\frac{\pi}{2^{i+2}}) \leq 0.277 \cdot \cos(\frac{\pi}{2^{i+2}})$. Now take (35) and observe $e_1^\top p_1 - e_1^\top p_0 = \sin(\pi/8)(t_1 + s + (p_0)_w \cdot \sin(\pi/8) - t_0 \cdot \cos(\pi/8)) \leq \sin(\pi/8)(11/30 + 2r + 0.277 \cdot \cos(\pi/8))$. Hence we find that $e_1^\top p_1 - e_1^\top p_0 \leq 0.239 + 0.766r$ and therefore,

$$
\begin{aligned}
e_1^\top p_k &\leq e_1^\top p_0 + \sum_{i=1}^{k}(e_1^\top p_i - e_1^\top p_{i-1})\\
&\leq (\frac{1}{6} + r) + 0.239 + 0.766r + \sin(\frac{\pi}{8}) \cdot \sum_{i=2}^{k} 1.9^{-(i-1)} \cdot (\frac{1}{30} + \frac{i}{3} + (i+1)r)\\
&\leq (\frac{1}{6} + r) + 0.239 + 0.766r + 0.456 + 1.749r\\
&\leq 0.862 + 3.515r \leq 1.
\end{aligned}
$$

where the last inequality holds for any $r \leq \frac{1}{30}$. Therefore (27) holds and $(x, y, p_0, t, s) \in P$. ■

PROOF OF LEMMA 6.6. In Lemma 6.7 we construct a point $(\bar{x}, \bar{y}, \bar{p}_0, \bar{t}, \bar{s})$ such that $\frac{1}{30}\mathbb{B}_\infty^{k+5} \subseteq P - (\bar{x}, \bar{y}, \bar{p}_0, \bar{t}, \bar{s})$.

For the second inclusion, we show that $P \subseteq (\cos(\pi/2^{k+2})^{-1} \cdot \mathbb{B}_\infty^{k+5}$. Suppose $(x, y, p_0, t, s) \in P$ is arbitrary. From Lemma 6.5 we know that $\|(x,y)\|_\infty \leq \sqrt{x^2+y^2} \leq \|p_0\|_2 \leq \cos(\pi/2^{k+2})^{-1}$. Since $\mathbf{0}_k \leq t \leq \mathbf{1}_k$ we get $\|t\|_\infty \leq 1$, and lastly we have $0 \leq s \leq 1$. Put together, we find that $\|(x,y,p_0,t,s)\|_\infty \leq \cos(\pi/2^{k+2})^{-1}$. Since $(x,y,p_0,t,s) \in P$ was arbitrary, we find that $P \subseteq \cos(\pi/2^{k+2})^{-1} \cdot \mathbb{B}_\infty^{k+5}$. By the triangle inequality we find that $P - (\bar{x}, \bar{y}, \bar{p}_0, \bar{t}, \bar{s}) \subseteq (\cos(\pi/2^{k+2})^{-1} + \|(\bar{x}, \bar{y}, \bar{p}_0, \bar{t}, \bar{s})\|_\infty) \cdot \mathbb{B}_\infty^{k+5}$ and we see in Lemma 6.7 that $\|(\bar{x}, \bar{y}, \bar{p}_0, \bar{t}, \bar{s})\|_\infty = 1/3$. Finally, note that $\cos(\pi/2^{k+2})^{-1} \leq \cos(\pi/8)^{-1} \leq 1.1$, we have

$$\cos(\frac{\pi}{2^{k+2}})^{-1} \leq \cos(\frac{\pi}{8})^{-1} + \frac{1}{3} \leq \frac{3}{2}. \qquad \blacksquare$$

6.4 Properties of the Polar Polytope

In this section, we will analyse the scaled polar polytope $Q = \frac{1}{30}(P - (\bar{x}, \bar{y}, \bar{p}_0, \bar{t}, \bar{s}))^\circ$. From well-known duality properties, we will find that Q satisfies the following desirable properties:

1. $Q \cap W$ approximates a two-dimensional disk;
2. The inner ℓ_1-radius of Q is at least $\frac{1}{45}$ when centered at $\mathbf{0}$;
3. The outer ℓ_1-radius of Q is at most 1 when centered at $\mathbf{0}$.

LEMMA 6.8. *For any $k \in \mathbb{N}$, there exists a two-dimensional linear subspace $W \subseteq \mathbb{R}^{k+5}$ and $n = 4k+7$ points $\bar{a}_1, \ldots, \bar{a}_n \in \mathbb{B}_1^{k+5}$ such that $Q := \mathrm{conv}(\bar{a}_1, \ldots, \bar{a}_n)$ satisfies*

$$\frac{\cos(\pi/2^{k+2})}{30} \cdot \mathbb{B}_2^{k+5} \cap W \subseteq Q \cap W \subseteq \frac{1}{30} \cdot \mathbb{B}_2^{k+5} \cap W$$

and

$$\frac{1}{45} \cdot \mathbb{B}_1^{k+5} \subseteq Q \subseteq \mathbb{B}_1^{k+5}$$

PROOF. Let $P \subseteq \mathbb{R}^{k+5}$ be the polytope defined by the linear system in (30), and let

$$\tilde{P} = P - (\bar{x}, \bar{y}, \bar{p}_0, \bar{t}, \bar{s})$$

denote the polytope obtained from shifting its center $(\bar{x}, \bar{y}, \bar{p}_0, \bar{t}, \bar{s})$ to $\mathbf{0}_d$. Here, as in Lemma 6.7,

$$\bar{x} = \bar{y} = 0, \bar{p}_0 = (1/6, 1/6)^\top, \bar{t} = \mathbf{1}_k/30, \bar{s} = \frac{1}{3}.$$

By applying row rescalings we transform the constraints (24 - 29) into a matrix $A \in \mathbb{R}^{(4k+7)\times(k+5)}$ such that

$$\tilde{P} = \{z \in \mathbb{R}^{k+5} : Az \leq 1\}.$$

Let $\tilde{Q} = (\tilde{P})^\circ \subseteq \mathbb{R}^d$ denote the polar body of $\tilde{P}$. Since $\tilde{P}$ is bounded, $\tilde{Q}$ is the convex hull of the rows of the matrix A, i.e.

$$\tilde{Q} = \{A^\top \lambda : \lambda \in [0,1]^{4k+7} \text{ s.t. } \sum_{i=1}^{4k+7} \lambda_i = 1\}.$$

Then by Lemma 6.6 and Fact 2.5, the inner and outer ball of $\tilde{Q}$ satisfy

$$\frac{2}{3} \cdot \mathbb{B}_1^{k+5} \subseteq \tilde{Q} \subseteq 30 \cdot \mathbb{B}_1^{k+5}.$$

Also, by Lemma 6.3, Fact 2.6 and Fact 2.5, the inner and outer ball of $\tilde{Q} \cap W$ satisfy

$$\cos(\pi/2^{k+2}) \cdot \mathbb{B}_2^{k+5} \cap W \subseteq \tilde{Q} \cap W \subseteq \mathbb{B}_2^{k+5} \cap W.$$

The lemma then follows from taking $Q = \frac{1}{30}\tilde{Q}$. ■

6.5 Perturbation Analysis and Proof of the Lower Bound

In this subsection, we study the number of edges of the intersection polygon $Q \cap W$ after perturbation and prove our main lower bound theorem (Theorem 6.1). To show that our construction has many edges even after perturbation, we require the following two statements:

LEMMA 6.9. *Let $\bar{a}_1, \dots, \bar{a}_n \in \mathbb{R}^d$ be points with $r\mathbb{B}_1^d \subseteq \operatorname{conv}(\bar{a}_1, \dots, \bar{a}_n)$ for some $r > 0$. If $\varepsilon \leq r/2$ and points $a_1, \dots, a_n \in \mathbb{R}^d$ satisfy $\|a_i - \bar{a}_i\|_1 \leq \varepsilon$ for all $i \in [n]$ then it follows that*

$$(1 - \frac{2\varepsilon}{r}) \operatorname{conv}(\bar{a}_1, \dots, \bar{a}_n) \subseteq \operatorname{conv}(a_1, \dots, a_n) \subseteq (1 + \frac{\varepsilon}{r}) \operatorname{conv}(\bar{a}_1, \dots, \bar{a}_n).$$

PROOF. Write $Q = \operatorname{conv}(a_1, \dots, a_n)$ and $\bar{Q} = \operatorname{conv}(\bar{a}_1, \dots, \bar{a}_n)$. The second inclusion follows by $Q \subseteq \bar{Q} + \varepsilon\mathbb{B}_1^d \subseteq \bar{Q} + \frac{\varepsilon}{r}\bar{Q}$. For the first inclusion, we observe that $r\mathbb{B}_1^d \subseteq \bar{Q} \subseteq Q + \varepsilon\mathbb{B}_1^d \subseteq Q + \frac{r}{2}\mathbb{B}_1^d$. This implies that $\frac{r}{2}\mathbb{B}_1^d \subseteq Q$, for if there were to exist $x \in \frac{r}{2}\mathbb{B}_1^d$ such that $x \notin Q$ then, since Q is closed and convex, then we could find $y \in \mathbb{R}^d$ such that $y^\top x > y^\top z$ for all $z \in Q$. Writing $f(S) = \max_{z \in S} y^\top z$ for $S \subseteq \mathbb{R}^d$, this would give

$$f(r\mathbb{B}_1^d) \geq f(x + \frac{r}{2}\mathbb{B}_1^d) = y^\top x + f(\frac{r}{2}\mathbb{B}_1^d) > f(Q) + f(\frac{r}{2}\mathbb{B}_1^d) \geq f(Q + \frac{r}{2}\mathbb{B}_1^d) \geq f(r\mathbb{B}_1^d).$$

By contradiction it follows that $\frac{r}{2}\mathbb{B}_1^d \subseteq Q$.

Note that $1 - x^2 \leq 1$, so $1 - x \leq 1/(1+x)$ for any $1 + x > 0$. Combining with $\bar{Q} \subseteq Q + \varepsilon\mathbb{B}_1^d \subseteq Q + \frac{\varepsilon}{r/2}Q$ we conclude the desired result. ■

LEMMA 6.10. *If a polygon $T \subseteq \mathbb{R}^2$ satisfies $\alpha \cdot \mathbb{B}_2^2 \subseteq T \subseteq \beta \cdot \mathbb{B}_2^2$ for some $\alpha, \beta > 0$ then T has at least $\sqrt{\alpha/(\beta - \alpha)}$ edges.*

PROOF. If $\beta > 2\alpha$ then the bound is trivially true, so assume that $\beta \leq 2\alpha$.

Without loss of generality, re-scale T so that $\mathbb{B}_2^2 \subseteq T \subseteq (1+\varepsilon) \cdot \mathbb{B}_2^2$, where $\varepsilon = \beta/\alpha - 1 > 0$. Note that since $\beta \leq 2\alpha$ we have $\varepsilon \leq 1$.

Consider any edge $[q_1, q_2] \subseteq T$ and let $p \in [q_1, q_2]$ denote the minimum-norm point in this edge. Then we have $\|q_1 - p\|^2 = \|q_1\|^2 + \|p\|^2 - 2\langle q_1, p\rangle$. Since p is the minimum-norm point, we have $\langle q_1, p\rangle \geq \|p\|^2$, and hence $\|q_1 - p\|^2 \leq \|q_1\|^2 - \|p\|^2 \leq (1+\varepsilon)^2 - \|p\|^2$. Since p lies on the boundary of T we have $\|p\| \geq 1$, which implies that $\|q_1 - p\|^2 \leq (1+\varepsilon)^2 - 1 = 2\varepsilon + \varepsilon^2$. The analogous argument for $\|q_2 - p\|$ and the triangle inequality tell us that $\|q_1 - q_2\| \leq 2\sqrt{2\varepsilon + \varepsilon^2} \leq 4\sqrt{\varepsilon}$, where we use $\varepsilon \leq 1$ at the second step. The choice of the edge $[q_1, q_2]$ was arbitrary, hence every edge of T has length as most $4\sqrt{\varepsilon}$.

But T has perimeter at least 2π. Since the perimeter of a polygon is equal to the sum of the lengths of its edges, this implies that T has at least $\frac{2\pi}{4\sqrt{\varepsilon}} > \frac{1}{\sqrt{\varepsilon}}$ edges. ■

Now, we can prove the generic lower bound claimed in Theorem 6.2 on the shadow size under adversarial ℓ_1-perturbations.

PROOF OF THEOREM 6.2. Fix any $d > 5$, let $k = d - 5$ and observe that $n = 4k + 7$. Let $\bar{a}_1, \dots, \bar{a}_n$ be as constructed in Lemma 6.8. Then we have

$$\frac{\cos(\pi/2^{k+2})}{30} \cdot \mathbb{B}_2^{k+5} \cap W \subseteq \operatorname{conv}(\bar{a}_1, \dots, \bar{a}_n) \cap W \subseteq \frac{1}{30} \cdot \mathbb{B}_2^{k+5} \cap W,$$

and

$$\frac{1}{45} \cdot \mathbb{B}_1^{k+5} \subseteq \operatorname{conv}(\bar{a}_1, \dots, \bar{a}_n) \subseteq \mathbb{B}_1^{k+5}.$$

For any set of points $a_1, \dots, a_n$ such that $\|a_i - \bar{a}_i\|_1 \leq \varepsilon$ for each $i \in [n]$. By Lemma 6.9 and setting $r = 1/45$, we have

$$\operatorname{conv}(\bar{a}_1, \dots, \bar{a}_n) \subseteq \frac{1}{1 - 2\varepsilon/r} \operatorname{conv}(a_1, \dots, a_n) = \frac{1}{1 - 90\varepsilon} \operatorname{conv}(a_1, \dots, a_n).$$

Therefore,

$$\frac{\cos(\pi/2^{k+2})}{30} \cdot \mathbb{B}_2^{k+5} \cap W \subseteq \operatorname{conv}(\bar{a}_1, \dots, \bar{a}_n) \cap W \subseteq (1 - 90\varepsilon)^{-1} \operatorname{conv}(a_1, \dots, a_n) \cap W,$$

and

$$\operatorname{conv}(a_1, \dots, a_n) \cap W \subseteq (1 + 45\varepsilon) \operatorname{conv}(\bar{a}_1, \dots, \bar{a}_n) \cap W \subseteq \frac{1 + 45\varepsilon}{30} \cdot \mathbb{B}_2^{k+5} \cap W.$$

Therefore, we can bound the inner and outer ℓ_1-radius of $\operatorname{conv}(a_1, \dots, a_n) \cap W$ by

$$\frac{(1 - 90\varepsilon) \cdot \cos(\pi/2^{k+2})}{30} \cdot \mathbb{B}_2^{k+5} \cap W \subseteq \operatorname{conv}(a_1, \dots, a_n) \cap W \subseteq \cdot \frac{1 + 45\varepsilon}{30} \cdot \mathbb{B}_2^{k+5} \cap W.$$

It then follows from Lemma 6.10 that the polygon $\operatorname{conv}(a_1, \dots, a_n) \cap W$ has at least

$$\begin{aligned}
&\left(\frac{(1 - 90\varepsilon) \cdot \cos(\pi/2^{k+2})}{(1 + 45\varepsilon) - (1 - 90\varepsilon) \cdot \cos(\pi/2^{k+2})}\right)^{1/2} \\
\geq\ &\left(\frac{(1 - 90\varepsilon) \cdot (1 - \frac{\pi^2}{4^{k+2}})}{(1 + 45\varepsilon) - (1 - 90\varepsilon) \cdot (1 - \frac{\pi^2}{4^{k+2}})}\right)^{1/2} && \text{(By } \cos(x) \geq 1 - x^2 \text{ for } x \leq \tfrac{\pi}{2}\text{)} \\
\geq\ &\left(\frac{1 - 90\varepsilon - \frac{\pi^2}{4^{k+2}}}{(1 + 45\varepsilon) - (1 - 90\varepsilon - \frac{\pi^2}{4^{k+2}})}\right)^{1/2}
\end{aligned}$$

$$\geq \ \Omega\left(\frac{1}{\sqrt{\varepsilon + 4^{-k}}}\right)$$

edges. ∎

Finally, we can prove our main result using Gaussian tail bound:

PROOF OF THEOREM 6.1. Using concentration of Gaussian distribution in Corollary 2.9, we find that if $\sigma \leq 1/(360d\sqrt{\log n})$, then with probability at least $1 - \binom{n}{d}^{-1}$, we have $\max_{i\in[n]} \|a_i - \bar{a}_i\|_2 \leq 4\sigma\sqrt{d \log n} \leq \frac{1}{90\sqrt{d}}$. The result follows from Theorem 6.2 and the fact that $\|x\|_1 \leq \sqrt{d}\|x\|_2$ for every $x \in \mathbb{R}^d$. ∎

6.6 Experimental Results

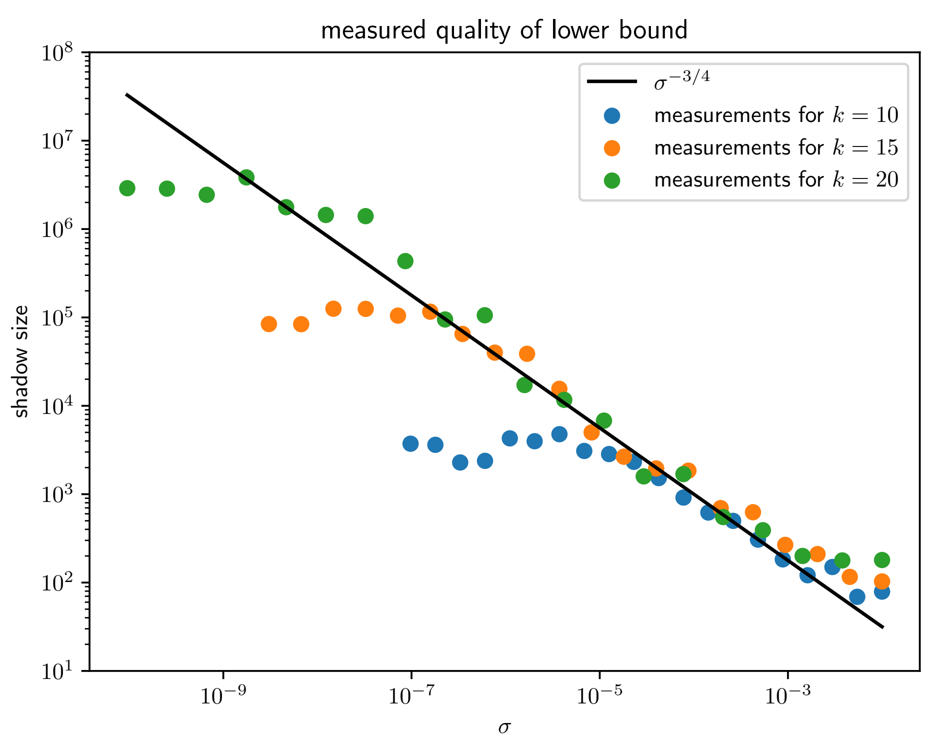

Figure 5. Measured shadow sizes for sampled perturbations of our construction, for different values of k and σ.

To measure whether the analysis in Theorem 6.1 is tight or not, we ran numerical experiments. Using `Python` and `Gurobi 10.0.3`, we constructed a matrix A such that

$$P - (\bar{x}, \bar{y}, \bar{p}_0, \bar{t}, \bar{s}) = \{z \in \mathbb{R}^{k+5} : Az \leq \mathbf{1}_{4k+7}\},$$

as described earlier in this section. Writing R as the maximum Euclidean norm among the row vectors of A, we sampled $\hat{A}$ with independent Gaussian distributed entries with standard deviation σR and $\mathbb{E}[\hat{A}] = A$. To approximate the shadow size, we optimized the objective vectors

$\cos(\frac{(i+0.3)\pi}{2^{k+4}})x + \sin(\frac{(i+0.3)\pi}{2^{k+4}})y$, with $i = 0, \dots, 2^{k+5} - 1$, over the polyhedron $\{z \in \mathbb{R}^{k+5} : \hat{A}z \leq 1\}$ and counted the number of distinct values (x, y) found among the solutions. Two consecutive solutions were counted as distinct if their x, y coordinates differed in ℓ_1 norm by at least 10^{-11}.[3] Note that the number of vertices of this projection of the perturbed auxiliary polytope is equal to the number of vertices of the section of the perturbed polar polytope, which then describes the shadow size.

When $\sigma = 0$, our code found 2^{k+1} such distinct points. For $\sigma > 0$, Theorem 6.1 shows that we expect to find at least $\Omega\Big(\min\Big(\frac{1}{\sqrt{d\sigma\sqrt{\log d}}}, 2^k\Big)\Big)$ distinct pairs (x, y).

For $k = 10, 15, 20$, we measured the shadow size for 20 different values of σ ranging from 0.01 to $0.0001/2^k$. The resulting data is depicted in Figure 5 along with a graph of the function $\sigma \mapsto \sigma^{-3/4}$. We observe that for each k, the measured shadow size appears to follow the graphed function up to a point, plateauing slightly above 2^{k+1} when σ is small. The fact that some measurements come out higher than 2^{k+1}, the shadow size for $\sigma = 0$, is not unexpected: the polytope P is highly degenerate, whereas the perturbed polytope is simple and can thus have many more vertices.

The measured shadow sizes appear to grow much faster than $1/\sqrt{\sigma}$ as σ gets small, closer to the $\sigma^{-3/4}$ line that we plotted. These results suggest that the behaviour of the shadow size is substantially different in $d = 2$, where we have an upper bound of $O\left(\frac{\sqrt[4]{\log(n)}}{\sqrt{\sigma}} + \sqrt{\log n}\right)$, and $d > 3$, where one might expect a lower bound with a higher dependence on σ.

References

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

[2] **Ilan Adler and Nimrod Megiddo**. A simplex algorithm whose average number of steps is bounded between two quadratic functions of the smaller dimension. *Journal of the ACM (JACM)*, 32(4):871–895, 1985. DOI (2)

[3] **Xavier Allamigeon, Daniel Dadush, Georg Loho, Bento Natura, and László A. Végh**. Interior point methods are not worse than simplex. *SIAM Journal on Computing*:FOCS22-178–264, 2025. DOI (7)

[4] **Nina Amenta and Günter M. Ziegler**. Deformed products and maximal shadows. *Contemporary Mathematics*, 223:57–90, 1998. DOI (2)

[5] **Neculai Andrei**. On the complexity of MINOS package for linear programming. *Studies in Informatics and Control*, 13(1):35–46, 2004. (2)

[6] **David Avis and Vasek Chvátal**. Notes on Bland's pivoting rule, *Polyhedral Combinatorics*, pages 24–34. Springer, 1978. DOI (2)

[7] **Andreas Bärmann, Andreas Heidt, Alexander Martin, Sebastian Pokutta, and Christoph Thurner**. Polyhedral approximation of ellipsoidal uncertainty sets via extended formulations: a computational case study. *Computational Management Science*, 13(2):151–193, October 2015. DOI (7)

[8] **Aharon Ben-Tal and Arkadi Nemirovski**. On polyhedral approximations of the second-order cone. *Mathematics of Operations Research*, 26(2):193–205, 2001. DOI (5, 7, 13, 44)

[9] **Alexander E Black, Jesús A De Loera, Sean Kafer, and Laura Sanità**. On the simplex method for 0/1-polytopes. *Mathematics of Operations Research*, 2024. DOI (6)

3 The code used for this experiment is available at `https://github.com/sophiehuiberts/shadow-size/tree/main/HLZ23`

[10] **Alexander E. Black**. Small shadows of lattice polytopes. *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1669–1679. Society for Industrial and Applied Mathematics, 2023. DOI (6)

[11] **Gilles Bonnet**, **Daniel Dadush**, **Uri Grupel**, **Sophie Huiberts, and Galyna Livshyts**. Asymptotic bounds on the combinatorial diameter of random polytopes. *38th International Symposium on Computational Geometry (SoCG)*, volume 224 of *LIPIcs*, 18:1–18:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. DOI (7)

[12] **Karl Heinz Borgwardt**. A sharp upper bound for the expected number of shadow vertices in lp-polyhedra under orthogonal projection on two-dimensional planes. *Mathematics of Operations Research*, 24(3):544–603, 1999. DOI (2, 6)

[13] **Karl Heinz Borgwardt**. The Simplex Method: A Probabilistic Analysis, volume 1. Springer-Verlag, Berlin, 1987. DOI (2, 4, 6)

[14] **Karl-Heinz Borgwardt**. The average number of pivot steps required by the simplex-method is polynomial. *Zeitschrift für Operations Research*, 26(1):157–177, 1982. DOI (2, 3, 6)

[15] **Tobias Brunsch and Heiko Röglin**. Finding short paths on polytopes by the shadow vertex algorithm. *Automata, Languages, and Programming*, ICALP '13, pages 279–290, Berlin, Heidelberg. Springer Berlin Heidelberg, 2013. DOI (6)

[16] **Daniel Dadush and Nicolai Hähnle**. On the shadow simplex method for curved polyhedra. *Discrete & Computational Geometry*, 56(4):882–909, 2016. DOI (6)

[17] **Daniel Dadush and Sophie Huiberts**. A friendly smoothed analysis of the simplex method. *SIAM Journal on Computing*, 49(5):STOC18-449–499, January 2020. DOI (3–9, 11, 15, 16)

[18] **Daniel Dadush and Sophie Huiberts**. Smoothed analysis of the simplex method. *Beyond the Worst-Case Analysis of Algorithms*. **Tim Roughgarden**, editor. Cambridge University Press, 2021., pages 309–333. DOI (5–7)

[19] **Valentina Damerow and Christian Sohler**. Extreme points under random noise. *European Symposium on Algorithms (ESA)*, pages 264–274. Springer, 2004. DOI (5, 7)

[20] **George B Dantzig**. Maximization of a linear function of variables subject to linear inequalities. *Cowles Commission Monograph 13: Activity analysis of production and allocation*:339–347, 1951. (2)

[21] **Amit Deshpande and Daniel A Spielman**. Improved smoothed analysis of the shadow vertex simplex method. *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 349–356, 2005. DOI (6)

[22] **Olivier Devillers**, **Marc Glisse**, **Xavier Goaoc**, and **Rémy Thomasse**. Smoothed complexity of convex hulls by witnesses and collectors. *Journal of Computational Geometry*, 7(2):101–144, 2016. DOI (6, 7)

[23] **Yann Disser**, **Oliver Friedmann**, and **Alexander V. Hopp**. An exponential lower bound for Zadeh's pivot rule. *Mathematical Programming*, July 2022. DOI (2)

[24] **Oliver Friedmann**. A subexponential lower bound for Zadeh's pivoting rule for solving linear programs and games. *Proceedings of 15th International Conference on Integer Programming and Combinatorial Optimization (IPCO)*, pages 192–206. Springer, 2011. DOI (2)

[25] **Oliver Friedmann**, **Thomas Dueholm Hansen**, and **Uri Zwick**. Subexponential lower bounds for randomized pivoting rules for the simplex algorithm. *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing (STOC)*, pages 283–292, 2011. DOI (2)

[26] **Gerald Gamrath**, **Daniel Anderson**, **Ksenia Bestuzheva**, **Wei-Kun Chen**, **Leon Eifler**, **Maxime Gasse**, **Patrick Gemander**, **Ambros Gleixner**, **Leona Gottwald**, **Katrin Halbig**, **Gregor Hendel**, **Christopher Hojny**, **Thorsten Koch**, **Pierre Le Bodic**, **Stephen J. Maher**, **Frederic Matter**, **Matthias Miltenberger**, **Erik Mühmer**, **Benjamin Müller**, **Marc E. Pfetsch**, **Franziska Schlösser**, **Felipe Serrano**, **Yuji Shinano**, **Christine Tawfik**, **Stefan Vigerske**, **Fabian Wegscheider**, **Dieter Weninger**, and **Jakob Witzig**. The SCIP Optimization Suite 7.0. Technical Report, Optimization Online, March 2020. URL (7)

[27] **Saul Gass and Thomas Saaty**. The computational algorithm for the parametric objective function. *Naval research logistics quarterly*, 2(1-2):39–45, 1955. (3)

[28] **François Glineur**. Computational Experiments With a Linear Approximation of Second-Order Cone Optimization. Technical report, 2000. URL (5, 7, 44)

[29] **Donald Goldfarb**. On the complexity of the simplex method. *Advances in Optimization and Numerical Analysis*. Springer, 1994., pages 25–38. DOI (2)

[30] **Donald Goldfarb**. Using the steepest-edge simplex algorithm to solve sparse linear programs, *Sparse matrix computations*, pages 227–240. Elsevier, 1976. DOI (2)

[31] **Donald Goldfarb**. Worst case complexity of the shadow vertex simplex algorithm. Technical report, Columbia University, New York, 1983. (6)

[32] **Donald Goldfarb and William Y Sit**. Worst case behavior of the steepest edge simplex method. *Discrete Applied Mathematics*, 1(4):277–285, 1979. DOI (2)

[33] **M. Haimovich**. The simplex method is very good! On the expected number of pivot steps and related properties of random linear programs. Technical report, Columbia University, 1983. (2)

[34] **Thomas Dueholm Hansen and Uri Zwick**. An improved version of the random-facet pivoting rule for the simplex algorithm. *Proceedings of the 47th annual ACM Symposium on Theory of Computing (STOC)*, pages 209–218, 2015. DOI (2)

[35] **Sophie Huiberts**, **Yin Tat Lee**, and **Xinzhi Zhang**. Upper and lower bounds on the smoothed complexity of the simplex method. *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1904–1917. ACM, 2023. DOI (1)

[36] **Robert G Jeroslow**. The simplex algorithm with the pivot rule of maximizing criterion improvement. *Discrete Mathematics*, 4(4):367–377, 1973. DOI (2)

[37] **Volker Kaibel and Kanstantsin Pashkovich**. Constructing extended formulations from reflection relations. *Facets of Combinatorial Optimization: Festschrift for Martin Grötschel*. **Michael Jünger and Gerhard Reinelt**, editors. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013., pages 77–100. DOI (7)

[38] **Gil Kalai**. A subexponential randomized simplex algorithm. *Proceedings of the 24th annual ACM Symposium on Theory of Computing (STOC)*, pages 475–482, 1992. DOI (2)

[39] **Jonathan A. Kelner and Daniel A. Spielman**. A randomized polynomial-time simplex algorithm for linear programming. *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, pages 51–60, 2006. DOI (7)

[40] **Victor Klee and George J Minty**. How good is the simplex algorithm. *Inequalities : III : proceedings of the 3rd Symposium on inequalities*:159–175, 1972. (2)

[41] **Jiří Matoušek**, **Micha Sharir**, and **Emo Welzl**. A subexponential bound for linear programming. *Algorithmica*, 16(4-5):498–516, 1996. DOI (2)

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

[43] **Katta G Murty**. Computational complexity of parametric linear programming. *Mathematical Programming*, 19(1):213–219, 1980. DOI (2, 6)

[44] **Hariharan Narayanan**, **Rikhav Shah**, and **Nikhil Srivastava**. A spectral approach to polytope diameter. *Discrete & Computational Geometry*, 72(4):1647–1674, March 2024. DOI (7)

[45] **Emanuel Schnalzger**. Lineare Optimierung mit dem Schatteneckenalgorithmus im Kontext probabilistischer Analysen. PhD thesis, Universität Augsburg, 2014. English translation by K.H. Borgwardt. (7)

[46] **Rolf Schneider and Wolfgang Weil**. Stochastic and Integral Geometry. Springer Berlin Heidelberg, 2008. DOI (17)

[47] **Ron Shamir**. The efficiency of the simplex method: a survey. *Management science*, 33(3):301–334, 1987. DOI (2)

[48] **Steve Smale**. On the average number of steps of the simplex method of linear programming. *Mathematical Programming*, 27(3):241–262, 1983. DOI (2)

[49] **Daniel A Spielman and Shang-Hua Teng**. Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, 51(3):385–463, 2004. DOI (2, 3, 6, 8, 12, 33)

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

[51] **Arnaud Vandaele**, **Nicolas Gillis**, and **François Glineur**. On the linear extension complexity of regular n-gons. *Linear Algebra and its Applications*, 521:217–239, 2017. DOI (7)

[52] **Roman Vershynin**. Beyond hirsch conjecture: walks on random polytopes and smoothed complexity of the simplex method. *SIAM Journal on Computing*, 39(2):646–678, 2009. DOI (4, 6)

[53] **Norman Zadeh**. A bad network problem for the simplex method and other minimum cost flow algorithms. *Mathematical Programming*, 5:255–266, 1973. DOI (6)