

# A Friendly Smoothed Analysis of the Simplex Method

Daniel Dadush\*

Centrum Wiskunde & Informatica  
The Netherlands  
dadush@cwi.nl

Sophie Huiberts

Centrum Wiskunde & Informatica  
The Netherlands  
sophie.huiberts@gmail.com

## ABSTRACT

Explaining the excellent practical performance of the simplex method for linear programming has been a major topic of research for over 50 years. One of the most successful frameworks for understanding the simplex method was given by Spielman and Teng (JACM '04), who developed the notion of smoothed analysis. Starting from an arbitrary linear program with  $d$  variables and  $n$  constraints, Spielman and Teng analyzed the expected runtime over random perturbations of the LP (smoothed LP), where variance  $\sigma^2$  Gaussian noise is added to the LP data. In particular, they gave a two-stage shadow vertex simplex algorithm which uses an expected  $\tilde{O}(d^{55}n^{86}(1+\sigma^{-30}))$  number of simplex pivots to solve the smoothed LP. Their analysis and runtime was substantially improved by Deshpande and Spielman (FOCS '05) and later Vershynin (SICOMP '09). The fastest current algorithm, due to Vershynin, solves the smoothed LP using an expected  $O(d^3 \log^3 n \sigma^{-4} + d^9 \log^7 n)$  number of pivots, improving the dependence on  $n$  from polynomial to logarithmic.

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

$$O(d^2 \sqrt{\log n} \sigma^{-2} + d^5 \log^{3/2} n)$$

number of simplex pivots. We obtain our results via an improved *shadow bound*, key to earlier analyses as well, combined with algorithmic techniques of Borgwardt (ZOR '82) and Vershynin. As an added bonus, our analysis is completely *modular*, allowing us to obtain non-trivial bounds for perturbations beyond Gaussians, such as Laplace perturbations.

## CCS CONCEPTS

• **Mathematics of computing** → *Probabilistic algorithms*; • **Theory of computation** → *Randomness, geometry and discrete structures*; Linear programming;

\*Supported by NWO Veni grant 639.071.510.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

STOC'18, June 25–29, 2018, Los Angeles, CA, USA

© 2018 Copyright held by the owner/author(s). Publication rights licensed to the Association for Computing Machinery.

ACM ISBN 978-1-4503-5559-9/18/06...\$15.00

<https://doi.org/10.1145/3188745.3188826>

## KEYWORDS

Linear Programming, Shadow Vertex Simplex Method, Smoothed Analysis

### ACM Reference Format:

Daniel Dadush and Sophie Huiberts. 2018. A Friendly Smoothed Analysis of the Simplex Method. In *Proceedings of 50th Annual ACM SIGACT Symposium on the Theory of Computing (STOC'18)*. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3188745.3188826>

## 1 INTRODUCTION

The simplex method for linear programming (LP) is one of the most important algorithms of the 20th century. Invented by Dantzig in 1947 [23, 24], it remains to this day one of the fastest methods for solving LPs in practice. The simplex method is not one algorithm however, but a class of LP algorithms, each differing in the choice of *pivot rule*. At a high level, the simplex method moves from vertex to vertex along edges of the feasible polyhedron, where the pivot rule decides which edges to cross, until an optimal vertex or unbounded ray is found. Important examples include Dantzig's most negative reduced cost [24], the Gass and Saaty parametric objective [35] and Goldfarb's steepest edge [38] method. We note that for solving LPs in the context of branch & bound and cutting plane methods for integer programming, where the successive LPs are "close together", the dual steepest edge method [32] is the *dominant* algorithm in practice [10, 11], due its observed ability to quickly re-optimize.

The continued success of the simplex method in practice is remarkable for two reasons. Firstly, there is no known polynomial time simplex method for LP. Indeed, there are exponential examples for almost every major pivot rule starting with constructions based on *deformed products* [5, 7, 39, 43, 48, 54, 63], such as the Klee-Minty cube [54], which defeat most classical pivot rules, and more recently based on Markov decision processes (MDP) [33, 34], which notably defeat randomized and history dependent pivot rules. Furthermore, for an LP with  $d$  variables and  $n$  constraints, the fastest (randomized) simplex method requires  $2^{O(\sqrt{d \ln(2+(n-d)/d)})}$  pivots [45, 49, 58], while the observed practical behavior is linear  $O(d+n)$  [72]. Secondly, it remains the most popular way to solve LPs despite the tremendous progress for polynomial time methods [53], mostly notably, interior point methods [51, 56, 62, 68]. How can we explain the simplex method's excellent practical performance?

This question has fascinated researchers for decades. An immediate question is how does one model instances in "practice", or at least instances where simplex should perform well? The research on this subject has broadly speaking followed three different lines: the analysis of average case LP models, where natural distributions of LPs are studied, the smoothed analysis of arbitrary LPs, where small random perturbations are added to the LP data, and work on structured LPs, such as totally unimodular systems and MDPs. We

review the major results for the first two lines in the next section, as they are the most relevant to the present work, and defer additional discussion to the related work section. To formalize the model, we consider LPs in  $d$  variables and  $n$  constraints of the following form:

$$\begin{aligned} \max \mathbf{c}^\top \mathbf{x} & \quad (\text{Main LP}) \\ \mathbf{Ax} \leq \mathbf{b} \end{aligned}$$

where the feasible polyhedron  $\mathbf{Ax} \leq \mathbf{b}$  is denoted by  $P$ . We now introduce relevant details for the simplex methods of interest to this work.

*Parametric Simplex Algorithms.* While a variety of pivot rules have been studied, the most successfully analyzed in theory are the so-called parametric simplex methods, due to the useful geometric characterization of the paths they follow. The first such method, and the main one used in the context of smoothed analysis, is the parametric objective method of Gass and Saaty [35], dubbed the *shadow (vertex) simplex* method by Borgwardt [14]. Starting at a *known* vertex  $\mathbf{v}$  of  $P$  maximizing an objective  $\mathbf{c}'$ , the parametric objective method computes the path corresponding to the sequence of maximizers for the objectives obtained by interpolating  $\mathbf{c}' \rightarrow \mathbf{c}$ <sup>1</sup>. The name shadow vertex method is derived from the fact that the visited vertices are in correspondence with those on the projection of  $P$  onto  $W := \text{span}(\mathbf{c}, \mathbf{c}')$ , the 2D convex polygon known as the shadow of  $P$  on  $W$ . In particular, the number of edges traversed by the method is bounded by the number of edges on the shadow, known as the *size of the shadow*.

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

A more general method, popular in the context of average case analysis, is the self-dual parametric simplex method of Dantzig [25]. In this method, one *simultaneously* interpolates both the objectives  $\mathbf{c}' \rightarrow \mathbf{c}$  and right hand sides  $\mathbf{b}' \rightarrow \mathbf{b}$  which has the effect of combining Phase I and II. Here  $\mathbf{c}'$  and  $\mathbf{b}'$  are chosen to induce a *known* initial maximizer. While the polyhedron is no longer fixed, the breakpoints in the path of maximizers (now a piecewise linear curve) can be computed via certain primal and dual pivots. This procedure was in fact generalized by Lemke [57] to solve linear complementarity problems. We note that the self dual method can roughly speaking be simulated in a higher dimensional space by adding an interpolation variable  $\lambda$ , i.e.  $\mathbf{Ax} \leq \lambda \mathbf{b} + (1-\lambda)\mathbf{b}'$ ,  $0 \leq \lambda \leq 1$ , which has been the principal approach in smoothed analysis.

## 1.1 Prior Work

Here we present the main works in both average case and smoothed analysis which inform our main results, presented in the next section. A common theme in these works, which all study parametric simplex methods, is to first obtain a bound on the expected parametric path length, with respect to some distribution on interpolations and LPs, and then find a way to use the bounds algorithmically. This second step can be non-obvious, as it is often the case that one

cannot directly find a starting vertex on the paths in question. We now present the main random LP models that have been studied, presenting path bounds and algorithms. Lastly, as our results are in the smoothed analysis setting, we explain the high level strategies used to prove smoothed (shadow) path bounds.

*Average case Models.* The first model, introduced in the seminal work of Borgwardt [14–17], examined distributions over LPs of the form  $\max \mathbf{c}^\top \mathbf{x}, \mathbf{Ax} \leq \mathbf{1}$ , possibly with  $\mathbf{x} \geq \mathbf{0}$  constraints (note that this model is always feasible at  $\mathbf{0}$ ), where the rows of  $\mathbf{A}$  are drawn i.i.d. from a rotationally symmetric distribution (RSM). Borgwardt proved tight bounds on the expected shadow size of the feasible polyhedron when projected onto any fixed plane. For general RSM, he proved a sharp  $\Theta(d^2 n^{1/(d-1)})$  [16, 17] bound, tight for rows drawn uniformly from the sphere, and for Gaussians a sharp  $\Theta(d^{1.5} \sqrt{\log n})$  bound [16], though this last bound is only known to hold asymptotically as  $n \rightarrow \infty$  (i.e. very large compared to  $d$ ). On the algorithmic side, Borgwardt [15] gave a *dimension by dimension* (DD) algorithm which optimizes over such polytopes by traversing  $d - 2$  different shadow vertex paths. The DD algorithm proceeds by iteratively solving the restrictions  $\max \mathbf{c}^\top \mathbf{x}, \mathbf{Ax} \leq \mathbf{1}, x_i = 0, i \in \{k + 1, \dots, d\}$ , for  $k \geq 2$ , which are all of RSM type. The key observation is that the optimal solution at phase  $k \in \{2, \dots, d - 1\}$  is (generically) on an edge of the shadow at stage  $k + 1$  for the plane generated by  $(c_1, \dots, c_k, 0, \dots, 0)$  and  $\mathbf{e}_{k+1}$  (the standard basis vector), and hence the shadow bound can be used to bound the algorithms complexity.

For the next class, Smale [73] analyzed the standard self dual method for LPs where  $\mathbf{A}$  and  $(\mathbf{c}, \mathbf{b})$  are chosen from independent RSM distributions, where Megiddo [61] gave the best known bound of  $f(\min\{d, n\})$  iterations, for some exponentially large function  $f$ . Adler [2] and Haimovich [44] examined a much weaker model where the data is fixed, but where the signs of all the inequalities, including non-negativity constraints, are flipped uniformly at random. Using the combinatorics of hyperplane arrangements, they achieved a remarkable bound of  $O(\min\{d, n\})$  for the expected length of parametric paths. These results were made algorithmic shortly thereafter [3, 4, 76], where it was shown that a lexicographic version of the parametric self dual simplex method<sup>2</sup> requires  $\Theta(\min\{d, n\}^2)$  iterations, where tightness was established in [4]. While these results are impressive, a notable criticism of the symmetry model is that it results in infeasible LPs almost surely once  $n$  is a bit larger than  $d$ .

*Smoothed LP Models.* The *smoothed* analysis framework, introduced in the breakthrough work of Spielman and Teng [75], helps explain the performance of algorithms whose worst-case examples are in essence *pathological*, i.e. which arise from very brittle structures in instance data. To get rid of these structures, the idea is to add a small amount of noise to the data, quantified by a parameter  $\sigma$ , where the general goal is then to prove an expected runtime bound over any *smoothed instance* that scales inverse polynomially with  $\sigma$ . Beyond the simplex method, smoothed analysis has been successfully applied to many other algorithms such as interior

<sup>1</sup>This path is well-defined under mild non-degeneracy assumptions

<sup>2</sup>These works use seemingly different algorithms, though they were shown to be equivalent to a lexicographic self-dual simplex method by Megiddo [60].

point methods [74], Gaussian elimination [69], Lloyd’s  $k$ -means algorithm [6], the 2-OPT heuristic for the TSP [31], and much more.

The smoothed LP model introduced by [75], starts with any  $\max \mathbf{c}^\top \mathbf{x}, \mathbf{A}\mathbf{x} \leq \bar{\mathbf{b}}$  – the average LP – normalized so that the rows of  $(\bar{\mathbf{A}}, \bar{\mathbf{b}})$  have  $\ell_2$  norm at most 1, and adds i.i.d. variance  $\sigma^2$  Gaussian noise to the entries of  $(\bar{\mathbf{A}}, \bar{\mathbf{b}})$  yielding  $(\mathbf{A}, \mathbf{b})$  – the smoothed LP data. Note that in this model  $\mathbf{c}$  is not perturbed. For smoothed LPs, Spielman and Teng provided a two phase shadow vertex method which uses an expected  $\tilde{O}(d^{55} n^{86} \sigma^{-30})$  number of pivots. This bound was substantially improved by Deshpande and Spielman [27] and Vershynin [78], where Vershynin gave the fastest such method requiring an expected  $O(d^3 \log^3 n \sigma^{-4} + d^9 \log^7 n)$  number of pivots.

In all these works, the complexity of the algorithms is reduced in a black box manner to a shadow bound for *smoothed unit LPs*. In particular, a smoothed unit LP has an expected system  $\tilde{\mathbf{A}}\mathbf{x} \leq 1$ , where  $\tilde{\mathbf{A}}$  has row norms at most 1, and smoothing is performed only to  $\tilde{\mathbf{A}}$ . Here the goal is to obtain a bound on the expected shadow size with respect to any fixed plane. Note that if  $\tilde{\mathbf{A}}$  is the zero matrix, then this is exactly Borgwardt’s Gaussian model, where he achieved the asymptotically tight bound of  $\Theta(d^{1.5} \sqrt{\ln n})$ . For smoothed unit LPs, Spielman and Teng [75] gave the first bound of  $O(d^3 n \sigma^{-6} + d^6 n \ln^3 n)$ . Deshpande and Spielman [27] derived a bound of  $O(dn^2 \ln n \sigma^{-2} + d^2 n^2 \ln^2 n)$ , substantially improving the dependence on  $\sigma$  while doubling the dependence on  $n$ . Lastly, Vershynin achieved a bound of  $O(d^3 \sigma^{-4} + d^5 \ln^2 n)$ , dramatically improving the dependence on  $n$  to logarithmic, though still with a larger dependence on  $\sigma$  than [27].

Before discussing the high level ideas for how these bounds are proved, we overview how they are used algorithmically. In this context, [75] and [78] provide two different reductions to the unit LP analysis, each via an interpolation method. Spielman and Teng first solve the smoothed LP with respect to an artificial “somewhat uniform” right hand side  $\mathbf{b}'$ , constructed to force a randomly chosen basis of  $\mathbf{A}$  to yield a vertex of the artificial system. From here they use shadow vertex to compute a maximizer for right hand side  $\mathbf{b}'$ , and continue via interpolation to derive an optimal solution for  $\mathbf{b}$ . Here the analysis is quite challenging, since in both steps the LPs are not quite smoothed unit LPs and the used shadow planes correlate with the perturbations. To circumvent these issues, Vershynin uses a *random vertex* (RV) algorithm, which starts with  $\mathbf{b}' = 1$  and adds a random additional set of  $d$  inequalities to the system to induce an “uncorrelated known vertex”. From this random vertex, he proceeds similarly to Spielman and Teng, but now at every step the LP is of smoothed unit type and the used shadow planes are (almost) independent of the perturbations.

We note that beyond the above model, smoothed analysis techniques have been used to analyze the simplex method in other interesting settings. In [20], the successive shortest path algorithm for min-cost flow, which is a shadow vertex algorithm, was shown to be efficient when only the objective (i.e. edge costs) is perturbed. In [52], Kelner and Spielman used smoothed analysis techniques to give a “simplex like” algorithm which solves arbitrary LPs in polynomial time. Here they developed a technique to analyze the expected shadow size when only the right hand side of an LP is perturbed.

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

$$\text{conv}(Q, \mathbf{0}) \cap W. \quad (1)$$

As  $\mathbf{0}$  is already in  $W$ , removing it from the convex hull can at worst decrease the number of edges by 1, and hence it is sufficient to bound the edges formed by  $D := Q \cap W$ .

We overview the different approaches used in [27, 75, 78] to bound the number of edges of  $D$ . Let  $\mathbf{u}_\theta$ ,  $\theta \in [0, 2\pi]$ , denote an angular parametrization of the unit circle in  $W$ , and let  $\mathbf{r}_\theta = \mathbf{u}_\theta \cdot \mathbb{R}_{\geq 0}$  denote the corresponding ray. Spielman and Teng [75] bounded the probability that any two nearby rays  $\mathbf{r}_\theta$  and  $\mathbf{r}_{\theta+\varepsilon}$  intersect different edges of  $D$  by a linear function of  $\varepsilon$ . Summing this probability over any fine enough discretization of the circle upper bounds the expected number of edges of  $D$ <sup>3</sup>. Their probability bound proceeds in two steps, first they estimate the probability that the Euclidean distance between the intersection of  $\mathbf{r}_\theta$  with its corresponding edge and the boundary of that edge is small (the *distance lemma*), and second they estimate the probability that angular distance is small compared to Euclidean distance (the *angle of incidence bound*). Vershynin [78] avoided the use of the angle of incidence bound by measuring the intersection probabilities with respect to the “best” of three different viewpoints, i.e. where the rays emanate from a well-chosen set of three equally spaced viewpoints as opposed to just the origin. This gave a much more efficient reduction to the distance lemma, and in particular allowed Vershynin to reduce the dependence on  $n$  from linear to logarithmic. Deshpande and Spielman [27] bounded different probabilities to get their shadow bound. Namely, they bounded the probability that nearby objectives  $\mathbf{u}_\theta$  and  $\mathbf{u}_{\theta+\varepsilon}$  are maximized at different vertices of  $D$ . The corresponding discretized sum over the circle directly bounds the number of vertices of  $D$ , which is the same as the number of edges.

## 1.2 Results

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

Recalling the models, the results in the following table bound the expected number of edges in the projection of a random polytope  $\mathbf{A}\mathbf{x} \leq 1$ ,  $\mathbf{A} \in \mathbb{R}^{n \times d}$ , onto any fixed 2-dimensional plane. The models differ in the class of distributions examined for  $\mathbf{A}$ . In the RSM model, the rows of  $\mathbf{A}$  are distributed i.i.d. according to an *arbitrary* rotationally symmetric distribution. In the Gaussian model, the rows of  $\mathbf{A}$  are i.i.d. mean zero standard Gaussian vectors. Note that this is a special case of the RSM model. The  $n \rightarrow \infty$  in the table

<sup>3</sup>One must a bit more careful when  $D$  does not contain the origin, but the details are similar.

**Table 1: Shadow Bounds. Logarithmic factors are simplified.**

Works	Expected Number of Edges	Model
[17]	$\Theta(d^2 n^{1/(d-1)})$	RSM
[16]	$\Theta(d^{3/2} \sqrt{\ln n})$	Gaussian, $n \rightarrow \infty$
[75]	$O(d^3 n \sigma^{-6} + d^6 n \ln^3 n)$	Smooth
[27]	$O(d n^2 \ln n \sigma^{-2} + d^2 n^2 \ln^2 n)$	Smooth
[78]	$O(d^3 \sigma^{-4} + d^5 \ln^2 n)$	Smooth
This paper	$O(d^2 \sqrt{\ln n} \sigma^{-2} + d^{2.5} \ln^{3/2} n (1 + \sigma^{-1}))$	Smooth

indicates that bound only holds for  $n$  large enough (compared to  $d$ ). In the smoothed model, the rows of  $\mathbf{A}$  are  $d$ -dimensional Gaussian random vectors with standard deviation  $\sigma$  centered at vectors of norm at most 1, i.e. the expected matrix  $\mathbb{E}[\mathbf{A}]$  has rows of  $\ell_2$  norm at most 1.

As can be seen, our new shadow bound yields a substantial improvement over earlier smoothed bounds in all regimes of  $\sigma$  and is also competitive in the Gaussian model. For small  $\sigma$ , our bound improves the dependence on  $d$  from  $d^3$  to  $d^2$ , achieves the same  $\sigma^{-2}$  dependence as [27], and improves the dependence on  $n$  to  $\ln^{3/2} n$ . For  $\sigma \geq 1$ , our bound becomes  $O(d^{2.5} \ln^{3/2} n)$ , which in comparison to Borgwardt's optimal (asymptotic) Gaussian bound is only off by a  $d \ln n$  factor. Furthermore, our proof is substantially simpler than Borgwardt's. In terms of the optimal bounds, given Borgwardt's result one may conjecture that the correct dependence on  $n$  and  $d$  should be  $d^{3/2} \sqrt{\ln n}$  for the smoothed Gaussian case as well, though it is unclear what the correct dependence on  $\sigma$  should be. We leave these questions as open problems.

An interesting point of our analysis is that it is *completely modular*, and it can give bounds for perturbations beyond Gaussians, in particular, we also get good bounds for Laplace perturbations (see section 4 of the full paper for details). The range of analyzable perturbations still remains limited however, our analysis doesn't extend to bounded perturbations such as uniform  $[-1/\sigma, 1/\sigma]$  for example, which we leave as an important open problem.

From the algorithmic perspective, our shadow bound naturally leads to improved shadow vertex running times via a two phase interpolation approach, using for Phase I either Vershynin's random vertex (RV) or Borgwardt's dimension by dimension algorithm (DD) depending on the value of  $\sigma$ .

Borgwardt's DD is faster for  $1/\sigma \leq d \sqrt{\log n}$  while Vershynin's RV is faster for all smaller  $\sigma$ . The tradeoff between the two is explained by the fact that DD works for all  $\sigma$  but requires following  $d-2$  shadow vertex paths, whereas RV requires  $1/\sigma \geq \sqrt{\log nd}^{1.5}$ <sup>4</sup> (always achievable by scaling down  $\mathbf{A}$ ) but follows only an expected  $O(1)$  number of shadow vertex paths. We note that [46] performed an amortized analysis of the DD algorithm in the RSM model yielding a  $\sqrt{d}$  factor improvement, using the fact that the interpolated objectives in later stages get closer and closer together, however

<sup>4</sup>In fact  $1/\sigma \geq \max\{\sqrt{d \log n}, d^{3/2} \sqrt{\log d}\}$  is sufficient. We rely on a worse bound for simplicity.

it is unknown whether such an improvement carries over in the smoothed setting.

Interestingly, the combination of interpolation and DD, while perhaps less efficient for small  $\sigma$ , completely removes all dependencies between the choice of shadow planes to follow and the instance data, a major issue in [75] and the main motivation for the RV algorithm, and its analysis (given a smoothed shadow bound) is essentially elementary. This combination was recently and explicitly suggested in [36] as a way to turn DD into a full LP algorithm, in the context of analyzing a generalized version of the RSM model. We note that Meggido [60] showed that the lexicographic self dual simplex used in the average case analyses, which combines Phase I and II, can simulate the DD algorithm and its dual the *constraint by constraint* algorithm, and thus one can view the combination of interpolation and DD as essentially simulating a lexicographic self dual simplex method. After the completion of this work, we learned that the idea of applying DD in the context of smoothed analysis was also recently presented by Schnalzer [71] in his thesis<sup>5</sup>.

We note that the above runtimes essentially follow by plugging in our shadow bound into the extant analyses of Vershynin and Borgwardt. We are however able to simplify and improve the analysis of a slight modification of Vershynin's RV algorithm, where we remove additional polylogarithmic runtime factors incurred by the original analysis. We defer further discussion of this to the full version of the paper.

### 1.3 Techniques: Improved Shadow Bound

We now give a detailed sketch of the proof of our improved shadow bound. Proofs of all claims can be found in section 2. The outline of the presentation is as follows. To begin, we explain our general edge counting strategy, where we depart from the previously discussed analyses. In particular, we adapt the approach of Kelner and Spielman (KS) [52], who analyzed a smoothing model where only the right-hand side is perturbed, to the present setting. Following this, we present a parametrized shadow bound, which applies to any class of perturbations for which the relevant parameters are bounded. Lastly, we give the high-level idea of how we estimate the relevant quantities in the KS approach within the parametrized model.

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

In [52], Kelner and Spielman developed a very elegant and useful alternative strategy to bound the expected number of edges, which can be applied to many distributions over 2D convex polygons. Whereas they analyzed the geometry of the primal shadow polygon, the projection of  $P$  onto  $W$ , we will instead work with the geometry of the polar polygon  $Q \cap W$ . The analysis begins with the following elementary identity:

$$\mathbb{E}[\text{perimeter}(Q \cap W)] = \mathbb{E}\left[\sum_{e \in \text{edges}(Q \cap W)} \text{length}(e)\right]. \quad (2)$$

<sup>5</sup>The thesis was originally published in German. An English translation by K.H. Borgwardt has recently been made available via the following [link](#).

**Table 2: Runtime bounds. Logarithmic factors are simplified.**

Works	Expected Number of Pivots	Model	Algorithm
[16, 17, 46]	$O(d^{2.5} n^{1/(d-1)})$	RSM	DD
[78]	$O(d^3 \log^3 n \sigma^{-4} + d^9 \log^7 n)$	Smooth	Int. + RV Phase I
This paper	$O(d^3 \sqrt{\log n} \sigma^{-2} + d^{3.5} \log^{3/2} n (1 + \sigma^{-1}))$	Smooth	Int. + DD Phase I
This paper	$O(d^2 \sqrt{\log n} \sigma^{-2} + d^3 \log^{3/2} n)$	Smooth	Int. + RV Phase I

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

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

$$\mathbb{E}[\text{perimeter}(Q \cap W)] \leq \mathbb{E}[2\pi \max_{\mathbf{x} \in Q \cap W} \|\mathbf{x}\|] \leq \mathbb{E}[2\pi \max_{i \in [n]} \|\pi_W(\mathbf{a}_i)\|]. \quad (3)$$

To extract the expected number of edges from the right hand side of (2), we first note that every edge of  $Q \cap W$  is derived from a facet of  $Q$  intersected with  $W$ . Assuming non-degeneracy, the possible facets of  $Q$  are  $F_I := \text{conv}(\mathbf{a}_i)_{i \in I}$ , where  $I \subseteq [n]$  is any subset of size  $d$ . Let  $E_I$  denote the event that  $F_I$  induces an edge of  $Q \cap W$ , more precisely, that  $F_I$  is a facet of  $Q$  and that  $F_I \cap W \neq \emptyset$ . From here, we get that

$$\begin{aligned} & \mathbb{E}\left[\sum_{\mathbf{e} \in \text{edges}(Q \cap W)} \text{length}(\mathbf{e})\right] \\ &= \sum_{|I|=d} \mathbb{E}[\text{length}(F_I \cap W) \mid E_I] \Pr[E_I] \\ &\geq \min_{|I|=d} \mathbb{E}[\text{length}(F_I \cap W) \mid E_I] \cdot \sum_{|I|=d} \Pr[E_I] \\ &= \min_{|I|=d} \mathbb{E}[\text{length}(F_I \cap W) \mid E_I] \cdot \mathbb{E}[|\text{edges}(Q \cap W)|]. \end{aligned} \quad (4)$$

Combining (2), (3), (4), we derive our main fundamental bound:

$$\mathbb{E}[|\text{edges}(Q \cap W)|] \leq \frac{\mathbb{E}[2\pi \max_{i \in [n]} \|\pi_W(\mathbf{a}_i)\|]}{\min_{|I|=d} \mathbb{E}[\text{length}(F_I \cap W) \mid E_I]}. \quad (5)$$

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

We note that applying the KS approach already improves the situation with respect to the maximum perturbation size compared to earlier analyses, as [27, 75, 78] all require a bound to hold almost surely as opposed to on expectation. For this purpose, they enforced

the condition  $1/\sigma \geq \sqrt{d \ln n}$  (for Gaussian perturbations), which we do not require here.

*Bound for Parametrized Distributions.* We now present the parameter bounds we require of the perturbations to obtain our parametrized shadow bound. We also discuss how these parameters behave for the Gaussian distribution.

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

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

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

We formalize the definitions of these parameters exactly in subsection 2.1.1.

The first two parameters are smoothness related while the last two relate to tail bounds. Assuming the above parameter bounds for  $\hat{\mathbf{a}}_1, \dots, \hat{\mathbf{a}}_n$ , our main “plug and play” bound on the expected shadow size is as follows (see Theorem 2.9):

$$\mathbb{E}[|\text{edges}(\text{conv}(\mathbf{a}_1, \dots, \mathbf{a}_n) \cap W)|] = O\left(\frac{d^{1.5} L}{\tau} (1 + R_{n,d})(1 + r_n)\right). \quad (6)$$

For the variance  $\sigma^2$  Gaussian distribution in  $\mathbb{R}^d$ , it is direct to verify that  $\tau = \sigma$  for any line (since every line restriction results in a 1D variance  $\sigma^2$  Gaussian), and from standard tail bounds that  $R_{n,d} = O(\sigma\sqrt{d \log n})$  and  $r_n = O(\sigma\sqrt{\log n})$ . The only parameter that cannot be bounded directly is the log-Lipschitz parameter  $L$ , since  $\|\mathbf{x}/\sigma\|^2/2$ , the log of the Gaussian density, is quadratic. Nevertheless, as noted in previous analyses, the Gaussian is locally smooth inside any fixed radius. Indeed the main radius of interest will be  $R_{n,d}$ , since events that happen with probability  $\ll \binom{n}{d}^{-1}$  have little effect on the shadow bound. Within radius  $R_{n,d}$  of the mean the density is  $O(\sqrt{d \log n}/\sigma)$ -log-Lipschitz.

As opposed to conditioning the perturbations to land in this ball as in prior analyses, which leads to complications, we instead replace the Gaussian with an essentially equivalent distribution (i.e., having the same properties and shadow bound), that is everywhere  $O(\sqrt{d \log n}/\sigma)$ -log-Lipschitz, which we call the Laplace-Gaussian distribution (see the full version of the paper for details). This helps simplify the analysis and also establishes the utility of the above parametrized model.

*Bounding the Perimeter and Edge Length.* We now briefly describe how the perimeter and minimum edge length are bounded in our parametrized perturbation model. As this is the most technical part of the analysis, we refer the reader to the proofs in section 2 and give only a very rough discussion here. As above, we will assume that the perturbations satisfy the bounds given by  $L, \tau, R_{n,d}, r_n$ .

For the perimeter bound, we immediately derive the bound

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

by the triangle inequality. From here, we must bound the minimum expected edge length, which requires the majority of the work. For this task, we provide a clean analysis, which shares high-level similarities with the Spielman and Teng distance lemma, though our task is simpler. Firstly, we only need to show that an edge is large on average, whereas the distance lemma has the more difficult task of proving that an edge is unlikely to be small. Second, our conditioning is much milder. Namely, the distance lemma conditions a facet  $F_I$  on intersecting a specified ray  $r_\theta$ , whereas we only condition  $F_I$  on intersecting  $W$ . This conditioning gives the edge much more “wobble room”, and is the main leverage we use to get the factor  $d$  improvement.

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

- (1) Show that if  $F$  induces an edge, then under this conditioning  $F$  has *small diameter* with good probability, namely its vertices are all at distance at most  $O(1 + R_{n,d})$  from each other (Lemma 2.17). This uses the tailbound defining  $R_{n,d}$  and the fact that  $E$  occurs with non-trivial probability.
- (2) Condition on  $F$  being a facet by fixing its containing hyperplane  $H$  (Lemma 2.20). This is standard and done via a change of variables analyzed by Blaschke.
- (3) Let  $l := H \cap W$  denote the line which intersects  $F$  to form an edge of  $Q \cap W$ . Show that on average the longest chord of  $F$  parallel to  $l$  is long. We achieve the bound  $\Omega(\tau/\sqrt{d})$  (Lemma 2.28) using that the vertices of  $F$  restricted to lines parallel to  $l$  have variance at least  $\tau^2$ .
- (4) Show that on average  $F$  is pierced by  $l$  through a chord that is not too much shorter than the longest one. Here we derive the final bound on the expected edge length of

$$\Omega((\tau/\sqrt{d}) \cdot 1/(dL(1 + R_{n,d}))) \text{ (Lemma 2.27),}$$

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

This concludes the high-level discussion of the proof.

## 1.4 Related Work

*Structured Polytopes.* An important line of work has been to study LPs with good geometric or combinatorial properties. Much work has been done to analyze primal and dual network simplex algorithms for fundamental combinatorial problems on flow polyhedra such as bipartite matching [47], shortest path [28, 42], maximum flow [37, 40] and minimum cost flow [41, 65, 66]. Generalizing on the purely combinatorial setting, LPs where the constraint matrix  $A \in \mathbb{Z}^{n \times d}$  is totally unimodular (TU), i.e. the determinant of any square submatrix of  $A$  is in  $\{0, \pm 1\}$ , were analyzed by Dyer and Frieze [29], who gave a random walk based simplex algorithm which requires  $\text{poly}(d, n)$  pivots. Recently, an improved random walk approach was given by Eisenbrand and Vempala [30], which works in the more general setting where the subdeterminants are bounded in absolute value by  $\Delta$ , who gave an  $O(\text{poly}(d, \Delta))$  bound on the number of pivots (note that there is no dependence on  $n$ ). Furthermore, randomized variants of the shadow vertex algorithm were analyzed in this setting by [21, 22], where in particular [22] gave an expected  $O(d^5 \Delta^2 \ln(d\Delta))$  bound on the number of pivots. Another interesting class of structured polytopes comes from the LPs associated with Markov Decision Processes (MDP), where simplex rules such as Dantzig’s most negative reduced cost correspond to variants of policy iteration. Ye [79] gave polynomial bounds for Dantzig’s rule and Howard’s policy iteration for MDPs with a fixed discount rate, and Ye and Post [67] showed that Dantzig’s rule converges in strongly polynomial time for deterministic MDPs with variable discount rates.

*Diameter Bounds.* Another important line of research has been to establish diameter bounds for polyhedra, namely to give upper bounds on the shortest path length between any two vertices of a polyhedron as a function of the dimension  $d$  and the number of inequalities  $n$ . For any simplex method pivoting on the vertices of a fixed polytope, the diameter is clearly a lower bound on the worst-case number of pivots. The famous Hirsch conjecture from 1957, posited that for polytopes (bounded polyhedra) the correct bound should be  $n - d$ . This precise bound was recently disproven by Santos [70], who gave a 43 dimensional counter-example, improved to 20 in [59], where the Hirsch bound is violated by about 5% (these counter-examples can also be extended to infinite families). However, the possibility of a polynomial (or even linear) bound is still left open, and is known as the polynomial Hirsch conjecture. From this standpoint, the best general results are the  $O(2^d n)$  bound by Barnette [9] and Larman [55], and the quasi-polynomial  $n^{\log_2 d + 2}$  bound of Kalai and Kleitman [50], recently refined to  $(n - d)^{\log_2 d}$  by Todd [77]. As above, such bounds have been studied for structured classes of polytopes. In particular, the diameter of polytopes with bounded subdeterminants was studied by various authors [13, 22, 29], where the best known bound of  $O(d^3 \Delta^2 \ln(d\Delta))$  was given in [22]. The diameters of other classes such as 0/1 polytopes [64], transportation polytopes [8, 18, 19, 26] and flag polytopes [1] have also been studied.

## 1.5 Conclusions and Open Problems

We have given a substantially simplified and improved shadow bound and used it to derive faster simplex methods. We are hopeful

that our modular approach to the shadow bound will help spur the development of a more robust smoothed analysis of the simplex method, in particular, one that can deal with a much wider class of perturbations such as those coming from bounded distributions. A natural open problem is to improve the dependence on the parameters, both for the shadow bound and its algorithmic applications. On the lower bound side, there is currently no lower bound on the expected shadow size in the smoothed Gaussian model apart from that of Borgwardt, which does not depend on  $\sigma$ . As a final open problem: is it possible to show polynomial diameter bounds for any interesting class of random polyhedra (smoothed or otherwise)? As far as the authors are aware, even for polyhedra of the form  $\mathbf{A}\mathbf{x} \leq 1$ , where  $\mathbf{A}$  has i.i.d. Gaussian entries, the known diameter bounds are the same as those for worst-case polyhedra.

## 2 SHADOW BOUNDS

In this section, we first state our new and improved shadow bounds for Gaussian and Laplace distributed perturbations.

**THEOREM 2.1.** *Let  $W \subset \mathbb{R}^d$  be a fixed two-dimensional subspace, and let  $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^d$ , where  $n \geq d \geq 3$ , be independent Gaussian random vectors with variance  $\sigma^2$  and centers of norm at most 1. Then the expected number of edges is bounded by*

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

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

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

**THEOREM 2.2.** *Let  $W \subset \mathbb{R}^d$  be a fixed two-dimensional subspace, and let  $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^d$ , where  $n \geq d \geq 3$ , be independent Laplace distributed random vectors with parameter  $\sigma$  and centers of norm at most 1. Then the expected number of edges is bounded by*

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

We achieve these results through a shadow bound for parametrized distributions, which is then applied to the Gaussian and Laplace setting in an essentially blackbox manner. We defer the proofs of these reductions to the full version of the paper. For the rest of this extended abstract, we focus of stating and proving our shadow bound for parametrized distributions.

### 2.1 Shadow Bound for Parametrized Distributions

In this subsection, we bound the expected number of edges in the polygon

$$\text{conv}(\mathbf{a}_1, \dots, \mathbf{a}_n) \cap W,$$

where  $W \subset \mathbb{R}^d$  is a fixed two-dimensional plane and the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_n$  are independently distributed according to distributions with probability density functions  $\mu_1, \dots, \mu_n$  with centers of norm at most 1 and have bounds on the parameters specified below. When we talk about the *center* of a distribution, we indicate the mean vector. We write  $\mathbb{S}^{d-1} \subset \mathbb{R}^d$  for the  $d-1$ -dimensional sphere of unit vectors. The parameters we will use are defined below.

#### 2.1.1 Distribution Parameters.

**Definition 2.3.** A distribution with density  $\mu$  on  $\mathbb{R}^d$  is *L-log-Lipschitz* if for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$  we have  $|\ln(\mu(\mathbf{x})) - \ln(\mu(\mathbf{y}))| \leq L\|\mathbf{x} - \mathbf{y}\|$ . Equivalently,  $\mu$  is *L-log-Lipschitz* if  $\mu(\mathbf{x})/\mu(\mathbf{y}) \leq \exp(L\|\mathbf{x} - \mathbf{y}\|)$  for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ .

**Definition 2.4.** Given a distribution with probability density  $\mu$  on  $\mathbb{R}^d$ , we define the line variance  $\tau^2$  as the infimum of the variances when restricted to any fixed line  $l \subset \mathbb{R}^d$ :

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

**Definition 2.5.** Given a distribution with probability density  $\mu$  on  $\mathbb{R}^d$  with expectation  $\mathbb{E}_{\mathbf{x} \sim \mu}[\mathbf{x}] = \mathbf{y}$  we define the  $n$ -th deviation  $r_n$  to be the smallest number such that for any unit vector  $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$ ,

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

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

The  $n$ -th deviation allows us to bound the expected maximum norm  $\mathbb{E}[\max_{i \leq n} |\mathbf{x}_i^\top \boldsymbol{\theta}|]$  of  $n$  separate perturbations in a given direction  $\boldsymbol{\theta}$ . We formalize this in Lemma 2.7.

**Definition 2.6.** Given a distribution with probability density  $\mu$  on  $\mathbb{R}^d$  with expectation  $\mathbb{E}_{\mathbf{x} \sim \mu}[\mathbf{x}] = \mathbf{y}$ , we define the cutoff distance  $R(p)$  as the smallest number satisfying

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

The cutoff radius of interest is  $R_{n,d} := R(\frac{1}{d \binom{n}{d}})$ . The cutoff norm tells us how concentrated the probability mass of the random variable is, while the log-Lipschitzness tells us how spread out the probability mass is. Not both of these quantities can be arbitrarily good at the same time. We formalize this notion in Lemma 2.8.

**LEMMA 2.7.** *If  $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^d$  are mean  $\mathbf{0}$  random variables with  $n$ -th deviations bounded by  $r_n \geq 0$ , then for any  $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$ ,*

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

**PROOF.** We rewrite the expectation as

$$\mathbb{E}[\max_{i \in [n]} |\boldsymbol{\theta}^\top \mathbf{x}_i|] = \int_0^{\infty} \Pr[\max_{i \in [n]} |\boldsymbol{\theta}^\top \mathbf{x}_i| > t] dt.$$

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

$$\int_0^{r_n} \Pr[\max_{i \in [n]} |\boldsymbol{\theta}^\top \mathbf{x}_i| > t] dt \leq r_n,$$

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

$$\begin{aligned} \int_{r_n}^{\infty} \Pr[\max_{i \in [n]} |\boldsymbol{\theta}^\top \mathbf{x}_i| > t] dt &\leq \sum_{i \in [n]} \int_{r_n}^{\infty} \Pr[|\boldsymbol{\theta}^\top \mathbf{x}_i| > t] \\ &\leq r_n. \end{aligned}$$

Together this makes

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

LEMMA 2.8. *For a  $d$ -dimensional distribution with probability density  $\mu$ , where  $d \geq 3$ , with parameters  $L, R$  as described above, we have the inequality  $LR(1/2) \geq d/3$ .*

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

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

Taking logarithms, we find

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

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

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

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

### 2.1.2 Proof of Shadow Bound for Parametrized Distributions.

The main result of this subsection is the following parametrized shadow bound.

THEOREM 2.9 (PARAMETRIZED SHADOW BOUND). *Let  $n \geq d \geq 3$  and let  $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^d$  be independently distributed according to  $L$ -log-Lipschitz distributions with centers of norm at most 1, line variances at least  $\tau^2$ , cutoff radii at most  $R_{n,d}$  and  $n$ -th deviations at most  $r_n$ . For any fixed two-dimensional linear subspace  $W \subset \mathbb{R}^d$ , the expected number of edges satisfies*

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

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

In the rest of the subsection,  $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^d$ , where  $n \geq d \geq 3$ , will be as in Theorem 2.9. We use  $Q := \text{conv}(\mathbf{a}_1, \dots, \mathbf{a}_n)$  to denote the convex hull of the constraint vectors and  $W$  to denote the two-dimensional shadow plane.

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

- (1) Every  $d + 1$  vectors from  $\mathbf{a}_1, \dots, \mathbf{a}_n$  are affinely independent. Thus, every facet of  $Q$  is the convex hull of exactly  $d$  vectors from  $\mathbf{a}_1, \dots, \mathbf{a}_n$ .
- (2) For every edge  $e$  of  $Q \cap W$  there is a unique facet  $F$  of  $Q$  such that  $e = F \cap W$ .
- (3) Any  $d$  vectors  $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_d}, i_1, \dots, i_d \in [n]$ , have a unique hyperplane through them. This hyperplane intersects  $W$  in an one-dimensional line, does not contain the origin  $\mathbf{0}$ , and its unit normal vector pointing away from the origin is not  $-\mathbf{e}_1$ .

In what follows we will always assume the above conditions hold. The last condition will allow us to uniquely define a coordinate transformation  $(\mathbf{a}_1, \dots, \mathbf{a}_d) \mapsto (\boldsymbol{\theta}, t, \mathbf{b}_1, \dots, \mathbf{b}_d)$ , where  $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t > 0$  define the hyperplane containing  $\mathbf{a}_1, \dots, \mathbf{a}_d$ , i.e.,  $\boldsymbol{\theta}^\top \mathbf{a}_i = t, \forall i \leq d$ , and  $\mathbf{b}_1, \dots, \mathbf{b}_d \in \mathbb{R}^{d-1}$  isometrically parametrize the position of  $\mathbf{a}_1, \dots, \mathbf{a}_d$  within their containing hyperplane. This coordinate transformation, first analyzed by Blaschke and commonly used to study convex hulls, is given together with its Jacobian in the following theorem.

THEOREM 2.10. *Let  $\boldsymbol{\theta} \in \mathbb{S}^{d-1} \setminus \{-\mathbf{e}_1\}$  denote a unit vector,  $t > 0$  a scalar and  $\mathbf{b}_1, \dots, \mathbf{b}_d \in \mathbb{R}^{d-1}$ . Let  $h : \mathbb{R}^{d-1} \rightarrow \mathbb{e}_1^\perp$  be an isometric embedding and let  $R'_\theta : \mathbb{R}^d \rightarrow \mathbb{R}^d$  be the rotation sending  $\mathbf{e}_1$  to  $\boldsymbol{\theta}$  and fixes  $\text{span}(\mathbf{e}_1, \boldsymbol{\theta})^\perp$ . Now define  $R_\theta : \mathbb{R}^{d-1} \rightarrow \mathbb{R}^d$  by  $R_\theta = R'_\theta \circ h$ . Consider the map*

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

The Jacobian of this map equals

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

where  $\text{vol}_{d-1}$  denotes the volume of a set.

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

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

Definition 2.11. For each set  $I \in \binom{[n]}{d}$ , let  $E_I$  denote the event that  $\text{conv}(\mathbf{a}_i)_{i \in I} \cap W$  forms an edge of  $Q \cap W$ .

Definition 2.12. We define the set  $B \subseteq \binom{[n]}{d}$  to be the set of those  $I \subseteq [n]$  satisfying  $|I| = d$  and  $\Pr[E_I] \geq 2\binom{n}{d}^{-1}$ .

The next lemma is inspired by Theorem 3.2 of [52].

LEMMA 2.13. *Let  $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^d, Q = \text{conv}(\mathbf{a}_1, \dots, \mathbf{a}_n)$  and let  $W$  be a fixed two-dimensional plane. The expected number of edges in  $Q \cap W$  satisfies*

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

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



and we continue from that:

$$\begin{aligned}
& \mathbb{E}[\text{perimeter}(Q \cap W)] \\
&= \sum_{I \in \binom{[n]}{d}} \mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_i)_{i \in I} \cap W) \mid E_I] \Pr[E_I] \\
&\geq \sum_{I \in B} \mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_i)_{i \in I} \cap W) \mid E_I] \Pr[E_I] \\
&\geq \min_{I \in B} \mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_i)_{i \in I} \cap W) \mid E_I] \sum_{J \in B} \Pr[E_J] \\
&\geq \min_{I \in B} \mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_i)_{i \in I} \cap W) \mid E_I] \left( \sum_{J \in \binom{[n]}{d}} \Pr[E_J] - 2 \right) \\
&= \min_{I \in B} \mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_i)_{i \in I} \cap W) \mid E_I] (\mathbb{E}[\text{edges}(Q \cap W)] - 2).
\end{aligned}$$

The last line holds because whenever  $E_J$  holds,  $F_J \cap W$  is an edge of  $Q \cap W$ , and every edge of  $Q \cap W$  is formed by exactly one face  $F_J$ , by the non-degeneracy conditions we have assumed. By linearity of expectation, the expected number of edges is the sum of probabilities of bases to form an edge. By dividing on both sides of the inequality, we can now conclude

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

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

The perimeter is bounded using a standard convexity argument. A convex region in  $\mathbb{R}^2$  has perimeter no larger than that of any circle containing it. We exploit the fact that all centers have norm at most 1 and the expected perturbation sizes are not too big along any axis.

**LEMMA 2.14.** *Let  $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^d$  be distributed with  $n$ -th deviation at most  $r_n$  and let  $W$  be a fixed two-dimensional plane. For  $Q = \text{conv}(\mathbf{a}_1, \dots, \mathbf{a}_n)$ , the expected perimeter of  $Q \cap W$  is bounded by*

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

**PROOF.** By convexity, the perimeter is bounded from above by the perimeter of any containing circle. Let  $\hat{\mathbf{a}}_i = \mathbf{a} + \mathbf{v}_i$  denote the perturbation of  $\mathbf{a}_i$  from the center of its distribution. We can now derive the bound

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

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

size satisfies

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

The expectations satisfy, by Lemma 2.7,  $\mathbb{E}[\max_{i \leq n} |\mathbf{v}_j^\top \hat{\mathbf{a}}_i|] \leq 2r_n$ , thereby concluding the proof.  $\square$

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

To lower bound the length  $\mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_1, \dots, \mathbf{a}_d) \cap W) \mid E]$  we will need the pairwise distances between the different  $\mathbf{a}_i$ 's for  $i \in \{1, \dots, d\}$  to be small along  $\omega^\perp$ . This way we can know that the ‘‘wobble room’’ in the location of the facet is not much smaller than the size of the facet.

**Definition 2.15 (Containing Hyperplane).** Define the hyperplane  $H = \text{aff}(\mathbf{a}_1, \dots, \mathbf{a}_d) = t\boldsymbol{\theta} + \boldsymbol{\theta}^\perp \subset \mathbb{R}^d$ , where  $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$ ,  $t > 0$ , to be the hyperplane containing  $\mathbf{a}_1, \dots, \mathbf{a}_d$ . Define  $l = H \cap W$ . From our non-degeneracy conditions we know that  $l$  is a line. Express  $l = \mathbf{p} + \omega \cdot \mathbb{R}$ , where  $\omega \in \mathbb{S}^{d-1}$  and  $\mathbf{p} \in \omega^\perp$ .

**Definition 2.16 (Bounded Diameter Event).** We define the event  $D$  to hold exactly when  $\|\pi_{\omega^\perp}(\mathbf{a}_i) - \pi_{\omega^\perp}(\mathbf{a}_j)\| \leq 2 + 2R_{n,d}$  for all  $i, j \in [d]$ .

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

**LEMMA 2.17.** *Let  $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^d$  be distributed as above and let  $W \subset \mathbb{R}^d$  be a fixed two-dimensional plane. If  $\Pr[E] \geq 2\binom{n}{d}^{-1}$ , then the expected edge length satisfies*

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

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

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

By the triangle inequality and the bounded centers of distributions we continue

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

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

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

$\square$

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

*Definition 2.18 (Change of Variables).* Recall the change of variables we saw in Theorem 2.10,  $(\mathbf{a}_1, \dots, \mathbf{a}_d) \mapsto (\boldsymbol{\theta}, t, \mathbf{b}_1, \dots, \mathbf{b}_d)$ , where  $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$ ,  $t > 0$ ,  $\mathbf{b}_1, \dots, \mathbf{b}_d \in \mathbb{R}^{d-1}$ . We abbreviate

$$\bar{\mu}_i(\boldsymbol{\theta}, t, \mathbf{b}_i) := \mu_i(R_{\boldsymbol{\theta}}(\mathbf{b}_i) + t\boldsymbol{\theta})$$

and we write  $\bar{\mu}_i(\mathbf{b}_i)$  when the values of  $\boldsymbol{\theta}, t$  are clear.

By Theorem 2.10 of Blaschke [12] we know that for any fixed values of  $\boldsymbol{\theta}, t$  the vectors  $\mathbf{b}_1, \dots, \mathbf{b}_d$  have joint probability density proportional to

$$\text{vol}_{d-1}(\text{conv}(\mathbf{b}_1, \dots, \mathbf{b}_d)) \prod_{i=1}^d \bar{\mu}_i(\mathbf{b}_i). \quad (7)$$

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

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

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

We will condition on the shape of the projected simplex.

*Definition 2.19 (Projected shape).* Define the projected shift variable  $\mathbf{x} := \mathbf{x}_{\bar{\boldsymbol{\omega}}}(\mathbf{b}_1) = \pi_{\bar{\boldsymbol{\omega}}^\perp}(\mathbf{b}_1)$  and shape variable  $S := S_{\bar{\boldsymbol{\omega}}}(\mathbf{b}_1, \dots, \mathbf{b}_d)$  by

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

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

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

Observe that whenever  $S \in \mathcal{S}$  the event  $D$  holds.

We already know that  $\mathbf{b}_1, \dots, \mathbf{b}_d$  are affinely independent by our non-degeneracy conditions, i.e., they do not all lie in a  $d - 2$ -dimensional affine subspace. This means that  $\mathbf{s}_1, \dots, \mathbf{s}_d$  do not all lie in a  $d - 3$ -dimensional affine subspace, from which it follows that  $\text{rank}(\mathbf{s}_2, \dots, \mathbf{s}_d) = d - 2$  (recalling that  $\mathbf{s}_1 = \mathbf{0}$ ).

**LEMMA 2.20.** *Let  $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$ ,  $t > 0$ ,  $\mathbf{b}_1, \dots, \mathbf{b}_d \in \mathbb{R}^{d-1}$  denote the change of variables of  $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^d$  as in (2.18). Then, the expected*

*length satisfies*

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

A proof of Lemma 2.20 can be found in the full version of the paper.

*Definition 2.21 (Kernel Combination).* For  $S \in \mathcal{S}$ , define the combination  $\mathbf{z} := \mathbf{z}(S)$  to be the unique (up to sign)  $\mathbf{z} = (z_1, \dots, z_d) \in \mathbb{R}^d$  satisfying

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

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

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

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

*Definition 2.22 (Chord combinations).* We define the set of convex combinations of the shape  $S = (\mathbf{s}_1, \dots, \mathbf{s}_d) \in \mathcal{S}$  that equal  $\mathbf{q} \in \bar{\boldsymbol{\omega}}^\perp$  by

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

When  $S$  is clear we drop the subscript.

Observe that  $C(\mathbf{q})$  is a line segment of the form  $C(\mathbf{q}) = \lambda_{\mathbf{q}} + \mathbf{z} \cdot [0, d_{\mathbf{q}}]$ . We write  $\|C(\mathbf{q})\|_1$  for the  $\ell_1$ -diameter of  $C(\mathbf{q})$ . Since  $C(\mathbf{q})$  is a line segment,  $\|C(\mathbf{q})\|_1 = d_{\mathbf{q}}$ . We prove two basic properties of  $\|C(\mathbf{q})\|_1$  as a function of  $\mathbf{q}$ .

**LEMMA 2.23 (PROPERTIES OF CHORD COMBINATIONS).** *Let  $S \in \mathcal{S}$ , and  $\mathbf{y} := \mathbf{y}(S) = \sum_{i=1}^d |z_i| \mathbf{s}_i$ , with  $\mathbf{z} := \mathbf{z}(S)$  as in Definition 2.22. Then the following holds:*

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

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

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

$$\sum_{i=1}^d (\gamma \alpha_i + (1 - \gamma) \beta_i) \mathbf{s}_i = \gamma \mathbf{x} + (1 - \gamma) \mathbf{y},$$

from which we derive that

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

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

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

$$\sum_{i \in I} y'_i > \sum_{i \in I} z_i = 1/2, \quad \sum_{j \in J} y'_j > \sum_{j \in J} -z_j = 1/2,$$

so  $\|\mathbf{y}'\| > 1$ . This contradicts our earlier assumption, so we conclude that  $\max_{\mathbf{q} \in \text{conv}(S)} \|C(\mathbf{q})\|_1 = 2$ .  $\square$

The  $\ell_1$ -diameter  $\|C(\mathbf{q})\|_1$  specified by  $\mathbf{q} \in \text{conv}(S(\mathbf{b}_1, \dots, \mathbf{b}_d))$  relates to the length of the chord  $(\mathbf{q} + \mathbf{x} + \bar{\omega} \cdot \mathbb{R}) \cap \text{conv}(\mathbf{b}_1, \dots, \mathbf{b}_d)$ , which projects to  $\mathbf{q}$  under  $\pi_{\bar{\omega}^\perp}$ . Specifically,  $\|C(\mathbf{q})\|_1$  measures how long the chord is compared to the longest chord through the simplex. The exact relation is given below.

LEMMA 2.24. For fixed values of  $(h_1, \dots, h_d) = (\bar{\omega}^\top \mathbf{b}_1, \dots, \bar{\omega}^\top \mathbf{b}_d)$ ,  $(\mathbf{s}_1, \dots, \mathbf{s}_d) = S(\mathbf{b}_1, \dots, \mathbf{b}_d)$  and  $\mathbf{x} = \pi_{\bar{\omega}^\perp}(\mathbf{b}_1)$ , we have for any  $\mathbf{q} \in \text{conv}(S)$  the following equality:

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

PROOF. By definition of  $\|C(\mathbf{q})\|_1$ , there exists a convex combination  $\lambda_1, \dots, \lambda_d \geq 0, \sum_{i=1}^d \lambda_i = 1$  satisfying  $\sum_{i=1}^d \lambda_i \mathbf{s}_i = \mathbf{q}$  such that

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

From this we deduce

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

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

We can view the terms in the above product as follows: the length of the longest chord of  $\text{conv}(\mathbf{b}_1, \dots, \mathbf{b}_d)$  parallel to  $\bar{l}$  is  $2|\sum_{i=1}^d z_i h_i|$ , and the ratio of the length of the chord  $\text{conv}(\mathbf{b}_1, \dots, \mathbf{b}_d) \cap \bar{l}$  to the length of the longest chord parallel to  $\bar{l}$  equals  $\|C(\mathbf{q})\|_1/2$ . This follows from Lemma 2.23 since  $\|C(\mathbf{q})\|_1$  achieves a maximum value of 2 at  $\mathbf{q} = \mathbf{y}$ . As discussed in the high-level description, we will bound the expected values of these two quantities separately.

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

LEMMA 2.25. For fixed  $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t > 0, S \in \mathcal{S}$ , define  $\mathbf{x} \in \bar{\omega}^\perp, h_1, \dots, h_d \in \mathbb{R}$  conditioned on  $\boldsymbol{\theta}, t, S$  to have joint probability density function proportional to

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

where  $\mathbf{z} := \mathbf{z}(S)$  is as in Definition 2.21. Then for  $\mathbf{b}_1, \dots, \mathbf{b}_d \in \mathbb{R}^{d-1}$  distributed as in Lemma 2.20, conditioned on the hyperplane as specified by  $\boldsymbol{\theta}, t$  and the shape  $S = (\mathbf{s}_1, \dots, \mathbf{s}_d)$ , where  $\mathbf{s}_1 = \mathbf{0}$ , we have equivalence of the distributions

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

The proof of Lemma 2.25 can be found in the full version of the paper.

Recall that  $\bar{l} = \bar{\mathbf{p}} + \bar{\omega} \cdot \mathbb{R}$ . The event  $\bar{E}$  occurs if and only if  $\bar{\mathbf{p}} \in \mathbf{x} + \text{conv}(S)$ , hence if and only if  $\bar{\mathbf{p}} - \mathbf{x} \in \text{conv}(S)$ .

LEMMA 2.26. Let  $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t > 0, S \in \mathcal{S}$  be fixed. Let the random variables  $\mathbf{b}_1, \dots, \mathbf{b}_d \in \mathbb{R}^{d-1}, h_1, \dots, h_d \in \mathbb{R}, \mathbf{x} \in \bar{\omega}^\perp$  be distributed as in Lemma 2.25. Define  $\mathbf{q} := \bar{\mathbf{p}} - \mathbf{x}$ . Then, the expected edge length satisfies

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

PROOF. We start with the assertion of Lemma 2.24:

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

We take expectation on both sides to derive the equality

$$\begin{aligned} & \mathbb{E}[\text{length}(\text{conv}(\mathbf{b}_1, \dots, \mathbf{b}_d) \cap \bar{l}) \mid \boldsymbol{\theta}, t, S, \bar{E}] \\ & = \mathbb{E}[\|C(\mathbf{q})\|_1 \cdot \left| \sum_{i=1}^d z_i h_i \right| \mid \boldsymbol{\theta}, t, S, \bar{E}]. \end{aligned}$$

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

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

$\square$

We will first bound the expected  $\ell_1$ -diameter of  $C(\mathbf{q})$ , where  $\mathbf{q} = \bar{\mathbf{p}} - \mathbf{x}$ , which depends on where  $\bar{\mathbf{p}} - \mathbf{x}$  intersects the projected simplex  $\text{conv}(S)$ : where this quantity tends to get smaller as we approach to boundary of  $\text{conv}(S)$ . We recall that  $\bar{E}$  occurs if and only if  $\mathbf{q} \in \text{conv}(S)$ .

LEMMA 2.27 (CHORD COMBINATION BOUND). *Let  $\theta \in \mathbb{S}^{d-1}$ ,  $t > 0$  and  $S \in \mathcal{S}$  be fixed. Let  $\mathbf{q} = \bar{\mathbf{p}} - \mathbf{x}$  be distributed as in Lemma 2.26. Then, the expected  $\ell_1$ -diameter of  $C(\mathbf{q})$  satisfies*

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

PROOF. To get a lower bound on the expected value of  $\|C(\mathbf{q})\|_1$ , we will use concavity of  $\|C(\mathbf{q})\|_1$  over  $\text{conv}(S) = \text{conv}(s_1, \dots, s_d)$  and the fact that  $\max_{\mathbf{q} \in \text{conv}(S)} \|C(\mathbf{q})\|_1 \geq 2$ . These facts are proven in Lemma 2.23. We show that shifting the projected simplex does not change the probability density too much (using log-Lipschitzness), and use the properties of  $\|C(\mathbf{q})\|_1$  mentioned above.

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

$$\int \cdots \int \prod_{i=1}^d \bar{\mu}_i(\bar{\mathbf{p}} - \mathbf{q} + s_i + h_i \bar{\omega}) dh_1 \cdots dh_d.$$

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

$$\begin{aligned} & \int \cdots \int \prod_{i=1}^d \bar{\mu}_i(\bar{\mathbf{p}} - \mathbf{a} + s_i + h_i \bar{\omega}) dh_1 \cdots dh_d \\ & \leq \int \cdots \int \prod_{i=1}^d e^{L\|\mathbf{b}-\mathbf{a}\|} \bar{\mu}_i(\bar{\mathbf{p}} - \mathbf{b} + s_i + h_i \bar{\omega}) dh_1 \cdots dh_d \\ & = e^{dL\|\mathbf{b}-\mathbf{a}\|} \int \cdots \int \prod_{i=1}^d \bar{\mu}_i(\bar{\mathbf{p}} - \mathbf{b} + s_i + h_i \bar{\omega}) dh_1 \cdots dh_d, \end{aligned}$$

as needed.

Let  $\alpha \in (0, 1)$  be a scaling factor to be chosen later, and let  $\mathbf{y} = \mathbf{y}(S)$  be as in Lemma 2.23. Now we can write

$$\mathbb{E}[\|C(\mathbf{q})\| \mid \theta, t, S, \bar{E}] \quad (8)$$

$$\begin{aligned} & = \int_{\text{conv}(S)} \|C(\mathbf{q})\|_1 \hat{\mu}(\mathbf{q}) d\mathbf{q} \\ & \geq \int_{\alpha \text{conv}(S) + (1-\alpha)\mathbf{y}} \|C(\mathbf{q})\|_1 \hat{\mu}(\bar{\mathbf{q}}) d\mathbf{q}, \quad (9) \end{aligned}$$

because the integrand is non-negative. By concavity of  $\|C(\mathbf{q})\|_1$  we have the lower bound  $\|C(\alpha\mathbf{q} + (1-\alpha)\mathbf{y})\| \geq 2(1-\alpha)$  for all  $\mathbf{q} \in \text{conv}(S)$ . Therefore, (9) is lower bounded by

$$\begin{aligned} & \geq \int_{\alpha \text{conv}(S) + (1-\alpha)\mathbf{y}} 2(1-\alpha) \hat{\mu}(\mathbf{q}) d\mathbf{q} \\ & = 2\alpha^d (1-\alpha) \int_{\text{conv}(S)} \hat{\mu}(\alpha\mathbf{q} + (1-\alpha)\mathbf{y}) d\mathbf{q} \\ & \geq 2\alpha^d (1-\alpha) e^{-\max_{\mathbf{q} \in \text{conv}(S)} (1-\alpha)\|\mathbf{q}-\mathbf{y}\| \cdot dL} \int_{\text{conv}(S)} \hat{\mu}(\mathbf{q}) d\mathbf{q}, \\ & = 2\alpha^d (1-\alpha) e^{-\max_{i \in [d]} (1-\alpha)\|s_i - \mathbf{y}\| \cdot dL}, \quad (10) \end{aligned}$$

where we used a change of variables in the first equality, the  $dL$ -log-Lipschitzness of  $\hat{\mu}$  in the second inequality, and the convexity of the  $\ell_2$  norm in the last equality. Using the diameter bound of  $2 + 2R_{n,d}$  for  $\text{conv}(S)$ , (10) is lower bounded by

$$\geq 2\alpha^d (1-\alpha) e^{-(1-\alpha)dL(2+2R_{n,d})}. \quad (11)$$

Setting  $\alpha = 1 - \frac{1}{dL(2+2R_{n,d})} \geq 1 - 1/d$  (by Lemma 2.8) gives a lower bound for (11) of

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

□

Recall that we have now fixed the position  $\mathbf{x}$  and shape  $S$  of the projected simplex. The randomness we have left is in the positions  $h_1, \dots, h_d$  of  $\mathbf{b}_1, \dots, \mathbf{b}_d$  along lines parallel to the vector  $\bar{\omega}$ . As  $\theta$  and  $t$  are also fixed, restricting  $\mathbf{b}_i$  to lie on a line is the same as restricting  $\mathbf{a}_i$  to lie on a line. Thus, were it not for the correlation between  $h_1, \dots, h_d$ , i.e., the factor  $|\sum_{i=1}^d z_i h_i|$  in the joint probability density function, each  $h_i$  would be independent and have variance  $\tau^2$  by assumption, and thus one would expect  $\mathbb{E}[|\sum_{i=1}^d z_i h_i|] = \Omega(\|z\|\tau)$ . The following lemma establish this, and shows that in fact, the correlation term only helps.

LEMMA 2.28 (HEIGHT OF SIMPLEX BOUND). *Let  $\theta \in \mathbb{S}^{d-1}$ ,  $t \geq 0$ ,  $S \in \mathcal{S}$ ,  $\mathbf{x} \in \bar{\omega}^\perp$  be fixed and let  $\mathbf{z} := \mathbf{z}(S)$  be as in Definition 2.21. Then for  $h_1, \dots, h_d \in \mathbb{R}$  distributed as in Lemma 2.26, the expected inner product satisfies*

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

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

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

Note that, by assumption, the variables  $g_1, \dots, g_d$  each have variance at least  $\tau^2$ . We recall from Lemma 2.25 that the joint probability density of  $h_1, \dots, h_d$  is proportional to  $|\sum_{i=1}^d z_i h_i| \prod_{i=1}^d \tilde{\mu}_i(h_i)$ . Thus, may rewrite the above expectation as

$$\begin{aligned} & \mathbb{E}[|\sum_{i=1}^d z_i h_i| \mid \theta, t, S, \mathbf{x}] \\ & = \frac{\int \cdots \int_{\mathbb{R}} |\sum_{i=1}^d z_i h_i|^2 \prod_{i=1}^d \tilde{\mu}_i(h_i) dh_1 \cdots dh_d}{\int \cdots \int_{\mathbb{R}} |\sum_{i=1}^d z_i h_i| \prod_{i=1}^d \tilde{\mu}_i(h_i) dh_1 \cdots dh_d} \\ & = \frac{\mathbb{E}[|\sum_{i=1}^d z_i g_i|^2]}{\mathbb{E}[|\sum_{i=1}^d z_i g_i|]}. \end{aligned}$$

By the additivity of variance for independent random variables, we see that

$$\text{Var}\left(\sum_{i=1}^d z_i g_i\right) = \sum_{i=1}^d z_i^2 \text{Var}(g_i) \geq \tau^2 \|z\|^2 \geq \tau^2 \|z\|^2 / d = \tau^2 / d.$$

We reach the desired conclusion by applying Lemma 2.29:

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

□

LEMMA 2.29. *Let  $X$  be a random variable with  $\mathbb{E}[X] = \mu$  and  $\text{Var}(X) = \sigma^2$ . Then  $X$  satisfies*

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

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

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

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

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

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

PROOF OF THEOREM 2.9 (PARAMETRIZED SHADOW BOUND). By Lemma 2.13, we derive the shadow bound by combining an upper bound on  $\mathbb{E}[\text{perimeter}(Q \cap W)]$  and a lower bound on the expected edge length  $\mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_i)_{i \in I} \cap W) \mid E_I]$  for all  $I \in B$ . For the perimeter upper bound, by Lemma 2.14 we have that

$$\mathbb{E}[\text{perimeter}(Q \cap W)] \leq 2\pi(1 + 4r_n). \tag{12}$$

For the edge length bound, we assume w.l.o.g. as above that  $I = [d]$ . Combining lemmas 2.17, 2.20, 2.26, 2.27 and 2.28 we have that

$$\begin{aligned} & \mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_1, \dots, \mathbf{a}_d) \cap W) \mid E] \\ & \geq \frac{1}{2} \cdot \mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_1, \dots, \mathbf{a}_d) \cap W) \mid D, E] \\ & \geq \frac{1}{2} \cdot \inf_{\substack{\boldsymbol{\theta} \in \mathbb{S}^{d-1}, \\ t > 0}} \mathbb{E}[\text{length}(\text{conv}(\mathbf{b}_1, \dots, \mathbf{b}_d) \cap \bar{I}) \mid \boldsymbol{\theta}, t, S \in \mathcal{S}, \bar{E}] \\ & \geq \frac{1}{2} \cdot \inf_{\substack{\boldsymbol{\theta} \in \mathbb{S}^{d-1}, \\ t > 0, S \in \mathcal{S}}} \left( \mathbb{E}[\|C(\bar{\mathbf{p}} - \mathbf{x})\|_1 \mid \boldsymbol{\theta}, t, S, \bar{E}] \right. \\ & \quad \cdot \left. \inf_{\mathbf{x} \in \bar{\omega}^\perp} \mathbb{E}\left[\sum_{i=1}^d z_i h_i \mid \boldsymbol{\theta}, t, S, \mathbf{x}\right] \right) \\ & \geq \frac{1}{2} \cdot \frac{e^{-2}}{dL(1 + R_{n,d})} \cdot \frac{\tau}{2\sqrt{d}}. \end{aligned} \tag{13}$$

The theorem now follows by taking the ratio of (12) and (13).  $\square$

## REFERENCES

- [1] Karim A. Adiprasito and Bruno Benedetti. 2014. The Hirsch Conjecture Holds for Normal Flag Complexes. *Mathematics of Operations Research* 39, 4 (2014), 1340–1348. <https://doi.org/10.1287/moor.2014.0661>
- [2] Ilan Adler. 1983. *The expected number of pivots needed to solve parametric linear programs and the efficiency of the self-dual simplex method*. Technical Report. University of California, Dept. of IE and OR.
- [3] Ilan Adler, Richard M. Karp, and Ron Shamir. 1987. A simplex variant solving an  $m \times d$  linear program in  $O(\min(m^2, d^2))$  expected number of pivot steps. *J. Complexity* 3, 4 (1987), 372–387. [https://doi.org/10.1016/0885-064X\(87\)90007-0](https://doi.org/10.1016/0885-064X(87)90007-0)
- [4] Ilan Adler and Nimrod Megiddo. 1985. 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 (1985), 871–895.
- [5] Nina Amenta and Günter M. Ziegler. 1998. Deformed Products and Maximal Shadows. *Contemporary Math.* 223 (1998), 57–90.
- [6] David Arthur, Bodo Manthey, and Heiko Röglin. 2011. Smoothed analysis of the  $k$ -means method. *J. ACM* 58, 5 (2011), Art. 19, 31. <https://doi.org/10.1145/2027216.2027217> Preliminary version in FOCS '09.
- [7] David Avis and Vasek Chvátal. 1978. Notes on Bland's pivoting rule. In *Polyhedral Combinatorics*. Springer, 24–34.
- [8] Michel L. Balinski. 1984. The Hirsch conjecture for dual transportation polyhedra. *Math. Oper. Res.* 9, 4 (1984), 629–633. <https://doi.org/10.1287/moor.9.4.629>
- [9] David Barnette. 1974. An upper bound for the diameter of a polytope. *Discrete Math.* 10 (1974), 9–13.
- [10] Robert E. Bixby. 2012. A brief history of linear and mixed-integer programming computation. *Doc. Math.* (2012), 107–121.
- [11] Robert E. Bixby, Mary Fenelon, Zonghao Gu, Ed Rothberg, and Roland Wunderling. 2000. MIP: theory and practice—closing the gap. In *System modelling and optimization (Cambridge, 1999)*. Kluwer Acad. Publ., Boston, MA, 19–49.
- [12] Wilhelm Blaschke. 1935. INTEGRALGEOMETRIE 2. *Bulletin mathématique de la Société Roumaine des Sciences* 37, 1 (1935), 3–11. <http://www.jstor.org/stable/43769768>
- [13] Nicolas Bonifas, Marco Di Summa, Friedrich Eisenbrand, Nicolai Hähnle, and Martin Niemeier. 2014. On sub-determinants and the diameter of polyhedra. *Discrete Comput. Geom.* 52, 1 (2014), 102–115. <https://doi.org/10.1007/s00454-014-9601-x> Preliminary version in SOCG '12.
- [14] Karl-Heinz Borgwardt. 1977. *Untersuchungen zur Asymptotik der mittleren Schrittzahl von Simplexverfahren in der linearen Optimierung*. Ph.D. Dissertation. Universität Kaiserslautern.
- [15] Karl-Heinz Borgwardt. 1982. The average number of pivot steps required by the simplex-method is polynomial. *Z. Oper. Res. Ser. A-B* 26, 5 (1982), A157–A177.
- [16] Karl-Heinz Borgwardt. 1987. *The simplex method: A probabilistic analysis*. Algorithms and Combinatorics: Study and Research Texts, Vol. 1. Springer-Verlag, Berlin. xii+268 pages.
- [17] Karl-Heinz Borgwardt. 1999. A sharp upper bound for the expected number of shadow vertices in LP-polyhedra under orthogonal projection on two-dimensional planes. *Math. Oper. Res.* 24, 3 (1999), 544–603. <https://doi.org/10.1287/moor.24.3.544> A corrected version of Figure 1 appears in Volume 24(4), pp. 925–984.
- [18] Steffen Borgwardt, Jesús A. De Loera, and Elisabeth Finhold. 2017. The diameters of network-flow polytopes satisfy the Hirsch conjecture. *Mathematical Programming* (11 Jul 2017). <https://doi.org/10.1007/s10107-017-1176-x>
- [19] Graham Brightwell, Jan van den Heuvel, and Leen Stougie. 2006. A linear bound on the diameter of the transportation polytope. *Combinatorica* 26, 2 (2006), 133–139. <https://doi.org/10.1007/s00493-006-0010-5>
- [20] Tobias Brunsch, Kamiel Cornelissen, Bodo Manthey, Heiko Röglin, and Clemens Rönsner. 2015. Smoothed analysis of the successive shortest path algorithm. *SIAM J. Comput.* 44, 6 (2015), 1798–1819. <https://doi.org/10.1137/140989893> Preliminary version in SODA '13.
- [21] Tobias Brunsch, Anna Großwendt, and Heiko Röglin. 2015. Solving totally unimodular LPs with the shadow vertex algorithm. In *Proceedings of the 32nd Symposium on Theoretical Aspects of Computer Science (STACS '15)*.
- [22] Daniel Dadush and Nicolai Hähnle. 2016. On the shadow simplex method for curved polyhedra. *Discrete Comput. Geom.* 56, 4 (2016), 882–909. <https://doi.org/10.1007/s00454-016-9793-3> Preliminary version in SOCG '15.
- [23] George B. Dantzig. 1948. *Programming in a linear structure*. Technical Report. U.S. Air Force Comptroller, USAF, Washington, D.C.
- [24] George B. Dantzig. 1951. Maximization of a linear function of variables subject to linear inequalities. In *Activity Analysis of Production and Allocation*. John Wiley & Sons, Inc., New York, N. Y.; Chapman & Hall, Ltd., London, 339–347.
- [25] George B. Dantzig. 1959. *Linear programming and extensions*. Princeton university press.
- [26] Jesús A. De Loera, Edward D. Kim, Shmuel Onn, and Francisco Santos. 2009. Graphs of transportation polytopes. *J. Combin. Theory Ser. A* 116, 8 (2009), 1306–1325. <https://doi.org/10.1016/j.jcta.2009.03.010>
- [27] Amit Deshpande and Daniel A. Spielman. 2005. Improved Smoothed Analysis of the Shadow Vertex Simplex Method. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS '05)*. 349–356. <https://doi.org/10.1109/SFCS.2005.44>
- [28] Robert Dial, Fred Glover, David Karney, and Darwin Klingman. 1979. A computational analysis of alternative algorithms and labeling techniques for finding shortest path trees. *Networks* 9, 3 (1979), 215–248.
- [29] Martin Dyer and Alan Frieze. 1994. Random walks, totally unimodular matrices, and a randomised dual simplex algorithm. *Mathematical Programming* 64, 1 (1994), 1–16.
- [30] Friedrich Eisenbrand and Santosh Vempala. 2017. Geometric random edge. *Mathematical Programming* (2017), 1–15.
- [31] Matthias Englert, Heiko Röglin, and Berthold Vöcking. 2014. Worst case and probabilistic analysis of the 2-opt algorithm for the TSP. *Algorithmica* 68, 1 (2014), 190–264. <https://doi.org/10.1007/s00453-013-9801-4> Preliminary version in SODA '07.
- [32] John J. Forrest and Donald Goldfarb. 1992. Steepest-edge simplex algorithms for linear programming. *Math. Programming* 57, 3, Ser. A (1992), 341–374. <https://doi.org/10.1007/BF01581089>
- [33] Oliver Friedmann. 2011. A Subexponential Lower Bound for Zadeh's Pivoting Rule for Solving Linear Programs and Games. In *Proceedings of 15th International Conference on Integer Programming and Combinatorial Optimization (IPCO '11)*, Oktay Günlük and Gerhard J. Woeginger (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 192–206. [https://doi.org/10.1007/978-3-642-20807-2\\_16](https://doi.org/10.1007/978-3-642-20807-2_16)

- [34] Oliver Friedmann, Thomas Dueholm Hansen, and Uri Zwick. 2011. Subexponential Lower Bounds for Randomized Pivoting Rules for the Simplex Algorithm. In *Proceedings of the Forty-third Annual ACM Symposium on Theory of Computing (STOC '11)*. ACM, New York, NY, USA, 283–292. <https://doi.org/10.1145/1993636.1993675>
- [35] Saul Gass and Thomas Saaty. 1955. The computational algorithm for the parametric objective function. *Naval Res. Logist. Quart.* 2 (1955), 39–45. <https://doi.org/10.1002/nav.3800020106>
- [36] Markus Göhl and Karl-Heinz Borgwardt. 2014. The average number of pivot steps of the simplex-algorithm based on a generalized rotation-symmetry-model. *Math. Methods Oper. Res.* 80, 3 (2014), 329–366. <https://doi.org/10.1007/s00186-014-0483-8>
- [37] Andrew V. Goldberg, Michael D. Grigoriadis, and Robert E. Tarjan. 1991. Use of dynamic trees in a network simplex algorithm for the maximum flow problem. *Math. Programming* 50, 3, (Ser. A) (1991), 277–290. <https://doi.org/10.1007/BF01594940>
- [38] Donald Goldfarb. 1976. Using the steepest-edge simplex algorithm to solve sparse linear programs. *Sparse matrix computations* (1976), 227–240.
- [39] Donald Goldfarb. 1983. *Worst case complexity of the shadow vertex simplex algorithm*. Technical Report. Columbia University, New York.
- [40] Donald Goldfarb and Jianxiu Hao. 1990. A primal simplex algorithm that solves the maximum flow problem in at most  $nm$  pivots and  $O(n^2m)$  time. *Mathematical Programming* 47, 1 (1990), 353–365.
- [41] Donald Goldfarb and Jianxiu Hao. 1992. Polynomial-time primal simplex algorithms for the minimum cost network flow problem. *Algorithmica* 8, 2 (1992), 145–160. <https://doi.org/10.1007/BF01758840>
- [42] Donald Goldfarb, Jianxiu Hao, and Sheng-Roan Kai. 1990. Efficient shortest path simplex algorithms. *Oper. Res.* 38, 4 (1990), 624–628. <https://doi.org/10.1287/opre.38.4.624>
- [43] Donald Goldfarb and William Y. Sit. 1979. Worst case behavior of the steepest edge simplex method. *Discrete Appl. Math.* 1, 4 (1979), 277–285. [https://doi.org/10.1016/0166-218X\(79\)90004-0](https://doi.org/10.1016/0166-218X(79)90004-0)
- [44] M. Haimovich. 1983. *The simplex method is very good! On the expected number of pivot steps and related properties of random linear programs*. Technical Report. Columbia University.
- [45] Thomas Dueholm Hansen and Uri Zwick. 2015. An Improved Version of the Random-Facet Pivoting Rule for the Simplex Algorithm. In *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing (STOC '15)*. ACM, New York, NY, USA, 209–218. <https://doi.org/10.1145/2746539.2746557>
- [46] Gabriele Höfner. 1995. *Lineare Optimierung mit dem Schatteneckenalgorithmus: Untersuchungen zum mittleren Rechenaufwand und Entartungsverhalten*. Ph.D. Dissertation. Universität Augsburg.
- [47] Ming S. Hung. 1983. A polynomial simplex method for the assignment problem. *Oper. Res.* 31, 3 (1983), 595–600. <https://doi.org/10.1287/opre.31.3.595>
- [48] Robert G. Jeroslow. 1973. The simplex algorithm with the pivot rule of maximizing criterion improvement. *Discrete Math.* 4 (1973), 367–377. [https://doi.org/10.1016/0012-365X\(73\)90171-4](https://doi.org/10.1016/0012-365X(73)90171-4)
- [49] Gil Kalai. 1992. A Subexponential Randomized Simplex Algorithm (Extended Abstract). In *Proceedings of the Twenty-fourth Annual ACM Symposium on Theory of Computing (STOC '92)*. ACM, New York, NY, USA, 475–482. <https://doi.org/10.1145/129712.129759>
- [50] Gil Kalai and Daniel J. Kleitman. 1992. A quasi-polynomial bound for the diameter of graphs of polyhedra. *Bull. Amer. Math. Soc. (N.S.)* 26, 2 (1992), 315–316. <https://doi.org/10.1090/S0273-0979-1992-00285-9>
- [51] Narendra Karmarkar. 1984. A new polynomial-time algorithm for linear programming. *Combinatorica* 4, 4 (1984), 373–395. <https://doi.org/10.1007/BF02579150>
- [52] Jonathan A. Kelner and Daniel A. Spielman. 2006. A randomized polynomial-time simplex algorithm for linear programming. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC '06)*. ACM, New York, 51–60. <https://doi.org/10.1145/1132516.1132524>
- [53] Leonid G. Khachiyan. 1979. A polynomial algorithm in linear programming. *Dokl. Akad. Nauk SSSR* 244, 5 (1979), 1093–1096.
- [54] Victor Klee and George J. Minty. 1970. *How good is the simplex algorithm*. Technical Report. University of Washington.
- [55] David G. Larman. 1970. Paths of polytopes. *Proc. London Math. Soc. (3)* 20 (1970), 161–178.
- [56] Yin Tat Lee and Aaron Sidford. 2014. Path Finding Methods for Linear Programming: Solving Linear Programs in  $\sqrt{\text{rank}}$  Iterations and Faster Algorithms for Maximum Flow. In *Proceedings of the IEEE 55th Annual Symposium on Foundations of Computer Science (FOCS '14)*. 424–433. <https://doi.org/10.1109/FOCS.2014.52>
- [57] Carlton E. Lemke. 1965. Bimatrix equilibrium points and mathematical programming. *Management science* 11, 7 (1965), 681–689.
- [58] Jiri Matoušek, Micha Sharir, and Emo Welzl. 1996. A subexponential bound for linear programming. *Algorithmica* 16, 4-5 (1996), 498–516. <https://doi.org/10.1007/s004539900062>
- [59] Benjamin Matschke, Francisco Santos, and Christophe Weibel. 2015. The width of five-dimensional prisms. *Proc. Lond. Math. Soc. (3)* 110, 3 (2015), 647–672. <https://doi.org/10.1112/plms/pdu064>
- [60] Nimrod Megiddo. 1985. A note on the generality of the self-dual algorithm with various starting points. *Methods of operations research* 49 (1985), 271–275.
- [61] Nimrod Megiddo. 1986. Improved asymptotic analysis of the average number of steps performed by the self-dual simplex algorithm. *Math. Programming* 35, 2 (1986), 140–172. <https://doi.org/10.1007/BF01580645>
- [62] Sanjay Mehrotra. 1992. On the implementation of a primal-dual interior point method. *SIAM J. Optim.* 2, 4 (1992), 575–601. <https://doi.org/10.1137/0802028>
- [63] Katta G. Murty. 1980. Computational complexity of parametric linear programming. *Math. Programming* 19, 2 (1980), 213–219. <https://doi.org/10.1007/BF01581642>
- [64] Denis Naddef. 1989. The Hirsch conjecture is true for  $(0, 1)$ -polytopes. *Math. Programming* 45, 1, Ser. B (1989), 109–110. <https://doi.org/10.1007/BF01580999>
- [65] James B. Orlin. 1984. *Genuinely polynomial simplex and non-simplex algorithms for the minimum cost flow problem*. Technical Report. Massachusetts Institute of Technology.
- [66] James B. Orlin, Serge A. Plotkin, and Éva Tardos. 1993. Polynomial dual network simplex algorithms. *Math. Programming* 60, 3, Ser. A (1993), 255–276. <https://doi.org/10.1007/BF01580615>
- [67] Ian Post and Yinyu Ye. 2015. The simplex method is strongly polynomial for deterministic markov decision processes. *Mathematics of Operations Research* 40, 4 (2015), 859–868. Preliminary version in SODA '14.
- [68] James Renegar. 1988. A polynomial-time algorithm, based on Newton's method, for linear programming. *Math. Programming* 40, 1, (Ser. A) (1988), 59–93. <https://doi.org/10.1007/BF01580724>
- [69] Arvind Sankar, Daniel A. Spielman, and Shang-Hua Teng. 2006. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM J. Matrix Anal. Appl.* 28, 2 (2006), 446–476. <https://doi.org/10.1137/S0895479803436202>
- [70] Francisco Santos. 2012. A counterexample to the Hirsch conjecture. *Ann. of Math. (2)* 176, 1 (2012), 383–412. <https://doi.org/10.4007/annals.2012.176.1.7>
- [71] Emanuel Schnalzer. 2014. *Lineare Optimierung mit dem Schatteneckenalgorithmus im Kontext probabilistischer Analysen*. Ph.D. Dissertation. Universität Augsburg. English translation available at this link. [https://www.math.uni-augsburg.de/prof/opt/mitarbeiter/Ehemalige/borgwardt/Downloads/Abschlussarbeiten/Doc\\_Habil.pdf](https://www.math.uni-augsburg.de/prof/opt/mitarbeiter/Ehemalige/borgwardt/Downloads/Abschlussarbeiten/Doc_Habil.pdf)
- [72] Ron Shamir. 1987. The efficiency of the simplex method: a survey. *Management Sci.* 33, 3 (1987), 301–334. <https://doi.org/10.1287/mnsc.33.3.301>
- [73] Steve Smale. 1983. On the average number of steps of the simplex method of linear programming. *Mathematical programming* 27, 3 (1983), 241–262.
- [74] Daniel A. Spielman and Shang-Hua Teng. 2003. Smoothed analysis of termination of linear programming algorithms. *Math. Program.* 97, 1-2, Ser. B (2003), 375–404. <https://doi.org/10.1007/s10107-003-0448-9> ISMP, 2003 (Copenhagen).
- [75] Daniel A. Spielman and Shang-Hua Teng. 2004. Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time. *J. ACM* 51, 3 (2004), 385–463 (electronic). <https://doi.org/10.1145/990308.990310>
- [76] Michael J. Todd. 1986. Polynomial expected behavior of a pivoting algorithm for linear complementarity and linear programming problems. *Mathematical Programming* 35, 2 (1986), 173–192.
- [77] Michael J. Todd. 2014. An Improved Kalai–Kleitman Bound for the Diameter of a Polyhedron. *SIAM Journal on Discrete Mathematics* 28, 4 (2014), 1944–1947.
- [78] Roman Vershynin. 2009. Beyond Hirsch conjecture: walks on random polytopes and smoothed complexity of the simplex method. *SIAM J. Comput.* 39, 2 (2009), 646–678. <https://doi.org/10.1137/070683386> Preliminary version in FOCS '06.
- [79] Yinyu Ye. 2011. The simplex and policy-iteration methods are strongly polynomial for the Markov decision problem with a fixed discount rate. *Mathematics of Operations Research* 36, 4 (2011), 593–603.