A Friendly Smoothed Analysis of the Simplex Method

Daniel Dadush [*1] and Sophie Huiberts [†1]

[1]Centrum Wiskunde & Informatica, The Netherlands

November 24, 2017

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 the 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 σ^2 Gaussian noise is added to the LP data. In particular, they gave a two-stage shadow vertex simplex algorithm which uses an expected $\widetilde{O}(d^{55}n^{86}\sigma^{-30})$ 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 $O(d^3 \log^7 n\sigma^{-4} + d^9 \log^7 n)$ 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.

Keywords. Linear Programming, Shadow Vertex Simplex Method, Smoothed Analysis.

*Email: dadush@cwi.nl. Supported by NWO Veni grant 639.071.510.
†Email: sophie.huiberts@gmail.com.

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 [Dan48, Dan51], 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 [Dan51], the Gass and Saaty parametric objective [GS55] and Goldfarb's steepest edge [Gol76] 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 [FG92] is the *dominant* algorithm in practice [BFG$^+$00, Bix12], 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* [KM70, Jer73, AC78, GS79, Mur80, Gol83, AZ98], such as the Klee-Minty cube [KM70], which defeat most classical pivot rules, and more recently based on Markov decision processes (MDP) [FHZ11, Fri11], which notably defeat randomized and history dependent pivot rules. Furthermore, for an LP with d variables and n constraints, the fastest provable (randomized) simplex method requires $2^{O(\sqrt{d \ln(1+(n-d)/d)})}$ pivots [Kal92, MSW96, HZ15], while the observed practical behavior is linear $O(d + n)$ [Sha87]. Secondly, it remains the most popular way to solve LPs despite the tremendous progress for polynomial time methods [Kha79], mostly notably, interior point methods [Kar84, Ren88, Meh92, LS14]. 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 over 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:

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

where the feasible polyhedron $\mathbf{A}\mathbf{x} \leq \mathbf{b}$ 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 [GS55], dubbed the *shadow (vertex) simplex* method by Borgwardt [Bor77]. Starting at a *known* vertex $\mathbf{v}$ of P maximizing an objective $\mathbf{c}'$, the parametric objective method computes the path corresponding to the sequence of maximizers for the objectives obtained by interpolating $\mathbf{c}' \to \mathbf{c}$ [1]. The name shadow simplex method is derived from the fact that the visited vertices are in correspondence with those on the projection of P onto $W := \text{span}(\mathbf{c}, \mathbf{c}')$, the 2D convex polygon known as the shadow of P on W (see figure 3 for an illustration). 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 [Dan59]. In this method, one *simultaneously* interpolates the objectives $\mathbf{c}' \to \mathbf{c}$ and right hand sides $\mathbf{b}' \to \mathbf{b}$ which has the effect of combining Phase I and II. Here $\mathbf{c}'$ and $\mathbf{b}'$ are chosen to induce a *known* initial maximizer. While the polyhedron is no longer fixed, the breakpoints in the path of maximizers (now a piecewise linear curve) can be computed via certain primal and dual pivots. This procedure was in fact generalized by Lemke [Lem65] to solve linear complementarity problems. We note that the self dual method can roughly speaking be simulated in a higher dimensional space by adding an interpolation variable λ, 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.

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

Average case Models. The first model, introduced in the seminal work of Borgwardt [Bor77, Bor82, Bor87, Bor99], examined LPs of the form $\max \mathbf{c}^\mathsf{T}\mathbf{x}, \mathbf{A}\mathbf{x} \le \mathbf{1}$, possibly with $\mathbf{x} \ge \mathbf{0}$ 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 $\Theta(d^2 n^{1/(d-1)})$ [Bor87, Bor99] bound, tight for rows drawn uniformly from the sphere, and for Gaussians a sharp $\Theta(d^{1.5}\sqrt{\log n})$ bound [Bor87], though this last bound is only known to hold asymptotically as $n \to \infty$ (i.e. very large compared to d). On the algorithmic side, Borgwardt [Bor82] gave a *dimension by dimension* (DD) algorithm which optimizes over such polytopes by traversing $d-2$ different shadow simplex paths. The DD algorithm proceeds by iteratively solving the restrictions $\max \mathbf{c}^\mathsf{T}\mathbf{x}, \mathbf{A}\mathbf{x} \le 1, x_i = 0, i \in \{k+1, \dots, d\}$, for $k \ge 2$, 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 $(c_1, \cdots, c_k, 0, \dots, 0)$ 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 [Sma83] analyzed the standard self dual method for LPs where $\mathbf{A}$ and $(\mathbf{c}, \mathbf{b})$ are chosen from independent RSM distributions, where Meggido [Meg86] gave the best known bound of $f(\min\{d, n\})$ iterations, for some exponentially large function f. Adler [Adl83] and Haimovich [Hai83] 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 [Tod86, AM85, AKS87], where it was shown that a lexicographic version of the parametric self dual simplex method[2] requires $\Theta(\min\{d, n\}^2)$ iterations, where tightness was established in [AM85]. 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 [ST04], helps explain the performance of algorithms whose worst-case examples are in essence *pathological*, i.e. which arise from very brittle structures in instance data. To get rid of these structures, the idea is to add a small amount of noise to the data, quantified by a parameter σ, where the general goal is then to prove an expected runtime bound over any *smoothed instance* that scales inverse polynomially with σ. Beyond the simplex method, smoothed analysis has been successfully applied to many other algorithms such as interior point methods [ST03], Gaussian elimination [SST06], Loyd's k-means algorithm [AMR11], the 2-OPT heuristic for the TSP [ERV14], and much more.

The smoothed LP model introduced by [ST04], starts with any $\max \mathbf{c}^\mathsf{T}\mathbf{x}, \bar{\mathbf{A}}\mathbf{x} \le \bar{\mathbf{b}}$ – the

[2] These works use seemingly different algorithms, though they were shown to be equivalent to a lexicographic self-dual simplex method by Meggiddo [Meg85].

average LP – normalized so that the rows of $(\bar{\mathbf{A}}, \bar{\mathbf{b}})$ have ℓ_2 norm at most 1, and adds i.i.d. variance σ^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 simplex method which uses an expected $\widetilde{O}(d^{55}n^{86}\sigma^{-30})$ number of pivots. This bound was substantially improved by Deshpande and Spielman [DS05] and Vershynin [Ver09], where Vershynin gave the fastest such method requiring an expected $O(d^3 \log^7 n\sigma^{-4} + d^9 \log^7 n)$ 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 $\bar{\mathbf{A}}\mathbf{x} \leq \mathbf{1}$, where $\bar{\mathbf{A}}$ has row norms at most 1, and smoothing is performed only to $\bar{\mathbf{A}}$. Here the goal is to obtain a bound on the expected shadow size with respect to any fixed plane. Note that if $\bar{\mathbf{A}}$ is the zero matrix, then this is exactly Borgwardt's Gaussian model, where he achieved the asymptotically tight bound of $\Theta(d^{1.5}\sqrt{\ln n})$. For smoothed unit LPs, Spielman and Teng [ST04] gave the first bound of $O(d^3 n\sigma^{-6} + d^6 n \ln^3 n)$. Deshpande and Spielman [DS05] derived a bound of $O(dn^2 \ln n\sigma^{-2} + d^2 n^2 \ln^2 n)$, substantially improving the dependence on σ 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 σ than [DS05].

Before discussing the high level ideas for how these bounds are proved, we overview how they are used algorithmically. In this context, [ST04] and [Ver09] 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 simplex to compute a maximizer for right hand side $\mathbf{b}'$, and continue via interpolation to derive an optimal solution for $\mathbf{b}$. Here the analysis is quite challenging, since in both steps the LPs are not quite smoothed unit LPs and the used shadow planes correlate with the perturbations. To circumvent these issues, Vershynin uses a *random vertex* (RV) algorithm, which starts with $\mathbf{b}' = \mathbf{1}$ 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 [BCM$^+$15], the successive shortest path algorithm for min-cost flow, which is a shadow simplex algorithm, was shown to be efficient when only the objective (i.e. edge costs) is perturbed. In [KS06], 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 $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d, i \in [n]$, denote the rows of the constraint matrix of the smoothed unit LP $\mathbf{A}\mathbf{x} \leq \mathbf{1}$. 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 *polar polygon* (see figure 2.4 for an illustration). Letting $Q := \mathrm{conv}(\mathbf{a}_1, \dots, \mathbf{a}_n)$, the convex hull of the rows, the polar polygon can be expressed as

$$\mathrm{conv}(Q, \mathbf{0}) \cap W \,. \tag{1}$$

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 [ST04, DS05, Ver09] 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 [ST04] bounded the probability that any two nearby rays $\mathbf{r}_\theta$ and $\mathbf{r}_{\theta+\varepsilon}$ intersect different edges of D by a linear function of ε. Summing this probability over any fine enough discretization of the circle upper bounds the expected number of edges of D [3]. Their probability bound proceeds in two steps, first they estimate the probability that the Euclidean distance between the intersection of $\mathbf{r}_\theta$ with its corresponding edge and the boundary of that edge is small (the *distance lemma*), and second they estimate the probability that angular distance is small compared to Euclidean distance (the *angle of incidence bound*). Vershynin [Ver09] 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 [DS05] 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 $\mathbf{A}\mathbf{x} \leq \mathbf{1}$, $\mathbf{A} \in \mathbb{R}^{n \times d}$, onto any fixed 2-dimensional plane. The models differ in the class of distributions examined for $\mathbf{A}$. In the 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

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

Works	Expected Number of Edges	Model
[Bor99]	$\Theta(d^2 n^{1/(d-1)})$	RSM
[Bor87]	$\Theta(d^{3/2}\sqrt{\ln n})$	Gaussian: $n \to \infty$
[ST04]	$O(d^3 n\sigma^{-6} + d^6 n \ln^3 n)$	Smooth
[DS05]	$O(dn^2 \ln n\sigma^{-2} + d^2 n^2 \ln^2 n)$	Smooth
[Ver09]	$O(d^3\sigma^{-4} + d^5 \ln^2 n)$	Smooth
This paper	$O(d^2\sqrt{\ln n}\sigma^{-2} + d^{2.5} \ln^{3/2} n(1+\sigma^{-1}))$	Smooth

Figure 1: Shadow Bounds. Logarithmic factors are simplified.

Gaussian vectors. Note that this is a special case of the RSM model. The $n \to \infty$ in the table 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 σ centered at vectors of norm at most 1, i.e. the expected matrix $\mathbb{E}[\mathbf{A}]$ has rows of ℓ_2 norm at most 1.

As can be seen, our new shadow bound yields a substantial improvement over earlier smoothed bounds in all regimes of σ and is also competitive in the Gaussian model. For small σ, our bound improves the dependence on d from d^3 to d^2, achieves the same σ^{-2} dependence as [DS05], 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 σ 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 3 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.

Works	Expected Number of Pivots	Model	Algorithm
[Bor87, Höf95, Bor99]	$O(d^{2.5} n^{1/(d-1)})$	RSM	DD
[Ver09]	$O(d^3 \log^7 n\sigma^{-4} + d^9 \log^7 n)$	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	$O(d^2\sqrt{\log n}\sigma^{-2} + d^5 \log^{3/2} n)$	Smooth	Int. + RV Phase I

Figure 2: Runtime bounds. Logarithmic factors are simplified.

From the algorithmic perspective, our shadow bound naturally leads to improved shadow simplex 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 σ.

Borgwardt's DD is faster for $1/\sigma \leq d\sqrt{\log n}$ while Vershynin's RV is faster for all smaller σ. The tradeoff between the two is explained by the fact that DD works for all σ but requires following $d-2$ shadow simplex paths, whereas RV requires $1/\sigma \geq \sqrt{\log n} d^{1.5}$ [4] (always achievable by scaling down $\mathbf{A}$) but follows only an expected $O(1)$ number of shadow simplex paths. We note that [Höf95] 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 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 σ, completely removes all dependencies between the choice of shadow planes to follow and the instance data, a major issue in [ST04] 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 [GB14] 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 [Meg85] 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 Schnalzger [Sch14] in his thesis [5].

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 section 4 of the paper.

1.3 Techniques: Improved Shadow Bound

We now give a detailed sketch of the proof of our improved shadow bound. Proofs of all claims can be found in section 3. The outline of the presentation is as follows. To begin, we explain our general edge counting strategy, where we depart from the previously discussed analyses. In particular, we adapt the approach of Kelner and Spielman (KS) [KS06], which they analyzed in a smoothing model where only the right hand side is perturbed, to the present setting. Following this, we present a parametrized shadow

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

[5] The thesis was originally published in German. An English translation by K.H. Borgwardt has recently been made available via the following link.

bound, which applies to any class of perturbations for which the 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 compute a bound on the expected number of edges of the polygon $Q \cap W$, where W is the two-dimensional shadow plane, $Q := \mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ and $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d$ are the smoothed constraints of a unit LP.

In [KS06], 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:

$$\mathbb{E}[\mathrm{perimeter}(Q \cap W)] = \mathbb{E}[\sum_{\mathbf{e} \in \mathrm{edges}(Q \cap W)} \mathrm{length}(\mathbf{e})] \,. \tag{2}$$

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

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

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

To extract the expected number of edges from the right hand side of 2, we first note that every of edge $Q \cap W$ is derived from a facet of Q intersected with W (see figure 2.4 for an illustration). The possible facets are $F_I := \mathrm{conv}(\mathbf{a}_i)_{i \in I}$, where $I \subseteq [n]$ is any subset of size d. Let E_I denote the event that F_I induces an edge of $Q \cap W$, more precisely, that F_I is a facet of Q and that $F_I \cap W \neq \emptyset$. From here, we get that

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

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

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

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 this 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 [ST04, DS05, Ver09] 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 parameters 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 $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^n$ are independently distributed, where as before we assume that the centers $\bar{\mathbf{a}}_i := \mathbb{E}[\mathbf{a}_i]$, $i \in [n]$, have norm at most 1. We denote the perturbations by $\hat{\mathbf{a}}_i := \mathbf{a}_i - \bar{\mathbf{a}}_i$, $i \in [n]$. We will assume for simplicity of the presentation, that all the perturbations $\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_n$ are i.i.d. distributed according to a distribution μ (in general, they could each have a distinct distribution).

At a high level, the main property we require of μ 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. μ has an L-log-Lipschitz probability density function $f : \mathbb{R}^n \to \mathbb{R}_+$, that is $|\log f(\mathbf{x}) - \log f(\mathbf{y})| \leq L\|\mathbf{x} - \mathbf{y}\|$, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$.
2. The variance of $\mathbf{X}$ when restricted to any line $l \subset \mathbb{R}^d$ is at least τ^2.
3. $\Pr[\|\mathbf{X}\| \geq R_{n,d}] \leq \frac{1}{d\binom{n}{d}}$.
4. For all $\boldsymbol{\theta}$, $\|\boldsymbol{\theta}\| = 1$, $\mathbb{E}[\max_{i \in [n]} |\langle \mathbf{X}_i, \boldsymbol{\theta} \rangle|] \leq r_n$, when $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are i.i.d. μ-distributed.

The first two parameters are smoothness related while the last two relate to tail bounds. Assuming the above parameter bounds for $\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_n$, our main "plug-n-play" bound on the expected shadow size is as follows (see Theorem 14):

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

For the variance σ^2 Gaussian distribution in $\mathbb{R}^d$, it is direct to verify that $\tau = \sigma$ for any line (since every line restriction results in a 1D variance σ^2 Gaussian), and from standard tail bounds that $R_{n,d} = O(\sigma\sqrt{d \ln n})$ and $r_n = O(\sigma\sqrt{\ln n})$. 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 $R_{n,d}$, inside which the density is $O(\sqrt{d \ln n}/\sigma)$-log-Lipschitz, since events that happen with probability $\ll \binom{n}{d}^{-1}$ have little effect on the shadow bound. 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 $O(\sqrt{d \ln n}/\sigma)$-log-Lipschitz, which we call the Laplace-Gaussian distribution (see section 3.3 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 3, and give only a very rough discussion here. As above, we will assume that the perturbations satisfy the bounds given by $L, \tau, R_{n,d}, r_n$.

For the perimeter bound, we immediately derive the bound

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

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

Let us fix $F := F_{[d]} = \mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_d)$ as the potential facet of interest, under the assumption that the event $E := E_{[d]}$, 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 21). 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 24). This is standard and done via a change of variables analyzed by Blashke.

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 31) using that the vertices of F restricted to lines parallel to l have variance at least τ^2.

4. Show that on average F is pierced by l through a chord that is not too much shorter than the longest one. Here we derive the final bound on the expected edge length of

$$\Omega((\tau/\sqrt{d}) \cdot 1/(dL(1+R_{n,d}))) \text{ (Lemma 30),}$$

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 [Hun83], shortest path [DGKK79, GHK90], maximum flow [GH90, GGT91] and minimum cost flow [Orl84, GH92, OPT93]. Generalizing on the purely combinatorial setting, LPs where the constraint matrix $\mathbf{A} \in \mathbb{Z}^{n \times d}$ is totally unimodular (TU), i.e. the determinant of any square submatrix of $\mathbf{A}$ is in $\{0, \pm 1\}$, were analyzed by Dyer and Frieze [DF94], 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 [EV17], which works in the more general setting where the subdeterminants are bounded in absolute value by Δ, who gave an $O(\text{poly}(d, \Delta))$ bound on the number of pivots (note that there is no dependence on n). Furthermore, randomized variants of the shadow simplex algorithm were analyzed in this setting by [BGR15, DH16], where in particular [DH16] 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 [Ye11] gave polynomial bounds for Dantzig's rule and Howard's policy iteration for MDPs with a fixed discount rate, and Ye and Post [PY15] 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 [San12], who gave a 43 dimensional counter-example, improved to 20 in [MSW15], 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 [Bar74] and Larman [Lar70], and the quasi-polynomial $n^{O(\log d)}$ bound of Kalai and Kleitman [KK92], recently refined to $(n-d)^{\log d}$ by Todd [Tod14]. 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 [DF94, BDSE+14, DH16], where the best known bound of $O(d^3 \Delta^2 \ln(d\Delta))$ was given in [DH16]. The diameters of other classes such as 0/1 polytopes [Nad89], transportation polytopes [Bal84, BvdHS06, DLKOS09, BDLF17] and flag polytopes [AB14] 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. 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 σ, and so we leave this as open problem. A final natural open problem is to improve the dependence on the parameters, both for the shadow bound and its algorithmic applications.

1.6 Organization

Section 2 contains basic definitions and background material. The proofs of our shadow bounds are given in section 3. In particular, the proof of our shadow bound for parametrized distributions is given in subsection 3.1, and its applications to Laplace and Gaussian perturbations are given in subsections 3.2 and 3.3 respectively. The details regarding the two phase shadow simplex algorithms we use, which rely in a black box way on the shadow bound, are presented in section 4.

2 Preliminaries

2.1 Notation

1. Vectors are printed in bold to contrast with scalars: $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$. The space $\mathbb{R}^d$ comes with a standard basis $\mathbf{e}_1, \dots, \mathbf{e}_d$.

2. The inner product of $\mathbf{x}$ and $\mathbf{y}$ is written with two notations $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\mathsf{T}\mathbf{y} = \sum_{i=1}^d x_i y_i$. We use the ℓ_2-norm $\|\mathbf{x}\|_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ and the ℓ_1-norm $\|\mathbf{x}\|_1 = \sum_{i=1}^d |x_i|$. Every norm without subscript is the ℓ_2-norm. We use the unit sphere $\mathbb{S}^{d-1} = \left\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\right\}$ and the unit ball $\mathcal{B}_2^d = \left\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| \leq 1\right\}$.

3. For a linear subspace $V \subseteq \mathbb{R}^d$ we denote the orthogonal complement by writing $V^\perp = \left\{\mathbf{x} \in \mathbb{R}^d : \langle \mathbf{v}, \mathbf{x} \rangle = 0, \forall\, \mathbf{v} \in V\right\}$. For $\mathbf{v} \in \mathbb{R}^d$ we abbreviate $\mathbf{v}^\perp := \mathrm{span}(\mathbf{v})^\perp$.

4. For sets $A, B \subset \mathbb{R}^d$ we denote the Minkowski sum $A + B = \{\mathbf{a} + \mathbf{b} : \mathbf{a} \in A, \mathbf{b} \in B\}$. For a vector $\mathbf{v} \in \mathbb{R}^d$ we write $A + \mathbf{v} = A + \{\mathbf{v}\}$. For a set of scalars $S \subset \mathbb{R}$ we write $\mathbf{v} \cdot S = \{s\mathbf{v} : s \in S\}$.

5. A set $V + \mathbf{p}$ is an affine subspace if $V \subset \mathbb{R}^d$ is a linear subspace. If $S \subset \mathbb{R}^d$ then the affine hull $\mathrm{aff}(S)$ is the smallest affine subspace containing S. We say $\dim(S) = k$ if $\dim(\mathrm{aff}(S)) = k$ and write $\mathrm{vol}_k(S)$ for the k-dimensional volume of S. The 1-dimensional volume of a line segment l will also be written as $\mathrm{length}(l)$.

6. We abbreviate $[n] := \{1, \ldots, n\}$ and $\binom{[n]}{d} = \{I \subset [n] \mid |I| = d\}$. For $a, b \in \mathbb{R}$ we denote the intervals $[a, b] = \{r \in \mathbb{R} : a \leq r \leq b\}$ and $(a, b) = \{r \in \mathbb{R} : a < r < b\}$. For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ the line segment between $\mathbf{x}$ and $\mathbf{y}$ is $[\mathbf{x}, \mathbf{y}] = \{\lambda\mathbf{x} + (1-\lambda)\mathbf{y} : \lambda \in [0, 1]\}$.

7. A set $S \subset \mathbb{R}^d$ is convex if for all $\mathbf{x}, \mathbf{y} \in S, \lambda \in [0, 1]$ we have $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in S$. We write $\mathrm{conv}(S)$ to denote the convex hull of S, which is the intersection of all convex sets $T \supset S$. In a d-dimensional vector space, the convex hull equals
$$\mathrm{conv}(S) = \left\{ \sum_{i=1}^{d+1} \lambda_i \mathbf{s}_i : \lambda_1, \ldots, \lambda_{d+1} \geq 0, \sum_{i=1}^{d+1} \lambda_i = 1, \mathbf{s}_1, \ldots, \mathbf{s}_d \in S \right\}.$$

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

9. For any linear or affine subspace $V \subset \mathbb{R}^d$ the orthogonal projection onto V is denoted by π_V.

10. For $\mathbf{A} \in \mathbb{R}^{n \times d}$ a matrix and $B \subset [n]$ we write $\mathbf{A}_B \in \mathbb{R}^{|B| \times d}$ for the submatrix of $\mathbf{A}$ consisting of the rows indexed in B, and for $\mathbf{b} \in \mathbb{R}^n$ we write $\mathbf{b}_B$ for the restriction of $\mathbf{b}$ to the coordinates indexed in B.

11. For a set $A \subseteq \mathbb{R}^d$, we use the notation $\mathbb{1}[\mathbf{x} \in A]$ to denote the indicator function of A, i.e. $\mathbb{1}[\mathbf{x} \in A] = 1$ if $\mathbf{x} \in A$ and 0 otherwise.

2.2 Random variables

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

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

In a slight abuse of notation, we interchangeably use distributions and density functions: if $\mathbf{X} \in \mathbb{R}^d$ is distributed according to μ then $\Pr[\mathbf{X} \in D] = \int_D \mu(\mathbf{x})\, \mathrm{d}\mathbf{x}$.

For the distribution of perturbations of vectors, we will specifically look at the Gaussian distribution $N_d(\bar{\mathbf{a}}, \sigma) := N(\bar{\mathbf{a}}, \sigma^2 I_d)$ (probability density $(2\pi)^{-d/2} e^{-\|\mathbf{x}-\bar{\mathbf{a}}\|^2/(2\sigma^2)}$) and the d-dimensional Laplace distribution $L_d(\bar{\mathbf{a}}, \sigma)$ which has probability function $\frac{\sqrt{d}^d}{d!\sigma^d \mathrm{vol}_d(\mathcal{B}_2^{d-1})} e^{-\|\mathbf{x}-\bar{\mathbf{a}}\|\sqrt{d}/\sigma}$. The d-dimensional Laplace distribution is normalized to have expected norm $\sqrt{d}\sigma$. We abbreviate $N_d(\sigma) = N_d(\mathbf{0}, \sigma)$ and $L_d(\sigma) = L_d(\mathbf{0}, \sigma)$.

When we talk about the *center* of a distribution we indicate the mean vector, and when we say that distributions are centered at points of norm at most 1 it means that the mean vectors of these distributions have norms bounded by 1.

We recall that the Gamma distribution $\Gamma(\alpha, \beta)$ on the non-negative real numbers has probability density $\frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}$ and moment generating function $\mathbb{E}_{x \sim \Gamma(\alpha,\beta)}[e^{\lambda x}] = (1 - \lambda/\beta)^{-\alpha}$, for $\lambda < \beta$.

A useful fact is that one can generate a d-dimensional Laplace distribution $L_d(\sigma)$ as the product of independent random variables $\boldsymbol{\theta} \cdot s$, where $\boldsymbol{\theta}$ is sampled uniformly from the sphere $\mathbb{S}^{d-1}$ and $s \sim \Gamma(d, \sqrt{d}/\sigma)$.

Lemma 1. *Let X be a random variable with $\mathbb{E}[X] = \mu$ and $Var(X) = \sigma^2$. Then X satisfies*

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

Proof. By definition one has $\mathbb{E}\left[X^2\right] = \mu^2 + \sigma^2$. We will show that $\mathbb{E}\left[|X|\right] \leq |\mu| + \sigma$, 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 $\mathbb{E}\left[X^2\right] / \mathbb{E}\left[|X|\right] \geq (|\mu| + \sigma)/2$.

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

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

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

2.2.1 Tail bounds for Gaussian and Laplace distribution

We state some basic tail bounds for Gaussian and Laplace distributions. We include proofs for completeness.

Lemma 2 (Gaussian tail bounds). *For $\mathbf{X} \in \mathbb{R}^d$ distributed as $N_d(\mathbf{0}, \sigma)$, $t \geq 1$,*

$$\Pr[\|\mathbf{X}\| \geq t\sigma\sqrt{d}] \leq e^{-(d/2)(t-1)^2} . \tag{7}$$

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

$$\Pr[|\langle \mathbf{X}, \boldsymbol{\theta} \rangle| \geq t\sigma] \leq 2e^{-t^2/2} . \tag{8}$$

Proof. By homogeneity, we may w.l.o.g. assume that $\sigma = 1$.

Proof of (7)**.**

$$\begin{aligned}
\Pr[\|\mathbf{X}\| \geq \sqrt{d}t] &\leq \min_{\lambda \in [0,1/2]} \mathbb{E}[e^{\lambda \|\mathbf{X}\|^2}] e^{-\lambda t^2 d} \\
&= \min_{\lambda \in [0,1/2]} (\frac{1}{1-2\lambda})^{d/2} e^{-\lambda t^2 d} \\
&\leq e^{-(d/2)(t^2 - 2\ln t - 1)} \text{ , setting } \lambda = \frac{1}{2}(1 - 1/t^2) \\
&\leq e^{-(d/2)(t-1)^2}. \quad (\text{ since } \ln t \leq t - 1 \text{ for } t \geq 1 \text{ })
\end{aligned}$$

Proof of (8)**.**

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

□

Lemma 3 (Laplace tail bounds). *For $\mathbf{X} \in \mathbb{R}^d$ distributed as $(\mathbf{0}, \sigma)$-Laplace, $t \geq 1$,*

$$\Pr[\|\mathbf{X}\| \geq t\sigma\sqrt{d}] \leq e^{-d(t - \ln t - 1)} . \tag{9}$$

In particular, for $t \geq 2$,

$$\Pr[\|\mathbf{X}\| \geq t\sigma\sqrt{d}] \leq e^{-dt/7} . \tag{10}$$

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

$$\Pr[|\langle \mathbf{X}, \boldsymbol{\theta} \rangle| \geq t\sigma] \leq \begin{cases} 2e^{-t^2/16} : 0 \leq t \leq 2\sqrt{d} \\ e^{-\sqrt{d}t/7} : t \geq 2\sqrt{d} \end{cases} . \tag{11}$$

Proof. By homogeneity, we may w.l.o.g. assume that $\sigma = 1$.

Proof of (9)**.**

$$\begin{aligned}
\Pr[\|\mathbf{X}\| \geq \sqrt{d}t\sigma] &\leq \min_{\lambda \in [0,\sqrt{d}]} \mathbb{E}[e^{\lambda\|\mathbf{X}\|}]e^{-\lambda t} \\
&\leq \min_{\lambda \in [0,\sqrt{d}]} (1-\lambda/\sqrt{d})^{-d} e^{-\lambda\sqrt{d}t} \\
&\leq e^{-d(t-\ln t-1)} \text{, setting } \lambda = \sqrt{d}(1-1/t).
\end{aligned}$$

In particular, it follows from the inequality $t - \ln t - 1 \geq t/7$ for $t \geq 2$, noting that $(t - \ln t - 1)/t$ is an increasing function on $t \geq 1$.

Proof of (11)**.** For $t \geq 2\sqrt{d}$ we directly apply equation (10):

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

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

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

For the first term we follow [Bal97], where the third line follows from an inclusion as sets.

$$\begin{aligned}
\Pr[|\langle \boldsymbol{\omega}, \boldsymbol{\theta}\rangle| \geq t/(2\sqrt{d})] &= \Pr_{\mathbf{x} \in \mathcal{B}_2^d}[|\langle \mathbf{x}/\|\mathbf{x}\|, \boldsymbol{\theta}\rangle| \geq t/(2\sqrt{d})] \\
&= \frac{\operatorname{vol}_d(\left\{\mathbf{x} \in \mathcal{B}_2^d : |\langle \mathbf{x}/\|\mathbf{x}\|, \boldsymbol{\theta}\rangle| \geq t/(2\sqrt{d})\right\})}{\operatorname{vol}_d(\mathcal{B}_2^d)} \\
&\leq \frac{\operatorname{vol}_d(\left\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x} - \boldsymbol{\theta}t/(2\sqrt{d})\| \leq \sqrt{1 - t^2/(4d)}\right\})}{\operatorname{vol}_d(\mathcal{B}_2^d)} \\
&\leq (1 - \frac{t^2}{4d})^{d/2} \leq e^{-t^2/8}.
\end{aligned}$$

The desired conclusion follows since $e^{-t^2/8} + e^{-d/4} \leq 2e^{-t^2/16}$ for $0 \leq t \leq 2\sqrt{d}$.

□

2.3 Coordinate transformation

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

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

where $\left|\det\left(\frac{\partial\phi(\mathbf{x})}{\partial\mathbf{x}}\right)\right|$ is the Jacobian of ϕ. We describe a particular change of variables which has often been used for studying convex hulls and by Spielman and Teng's [ST04] for deriving shadow bounds.

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

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

where the unit vector $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$ and scalar $t \geq 0$ satisfy $\langle \boldsymbol{\theta}, \mathbf{a}_i \rangle = t$ for every $i \in \{1, \dots, d\}$, and the vectors $\mathbf{b}_1, \dots, \mathbf{b}_d \in \mathbb{R}^{d-1}$ parametrize the positions of $\mathbf{a}_1, \dots, \mathbf{a}_d$ within the hyperplane $\{\mathbf{x} \in \mathbb{R}^d \mid \langle \boldsymbol{\theta}, \mathbf{x} \rangle = t\}$. We achieve a consistent coordinatization of the hyperplanes as follows:

Fix a reference unit vector $\mathbf{v} \in \mathbb{S}^{d-1}$, and pick an isometric embedding $h : \mathbb{R}^{d-1} \to \mathbf{v}^\perp$. For any unit vector $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$, define the map $R'_{\boldsymbol{\theta}} : \mathbb{R}^d \to \mathbb{R}^d$ as the unique map that rotates $\mathbf{v}$ to $\boldsymbol{\theta}$ along $\mathrm{span}(\mathbf{v}, \boldsymbol{\theta})$ and fixes the orthogonal subspace $\mathrm{span}(\mathbf{v}, \boldsymbol{\theta})^\perp$. We define $R_{\boldsymbol{\theta}} = R'_{\boldsymbol{\theta}} \circ h$. The change of variables takes the form

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

The change of variables as specified above is not uniquely determined when $\mathbf{a}_1, \dots, \mathbf{a}_d$ are affinely dependent or when $\boldsymbol{\theta} = -\mathbf{v}$. Since these events happen with probability 0, we ignore this issue.

The Jacobian of this change of variables has been determined by Blaschke[Bla35].

Theorem 4. *Let $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$ be a unit vector, $t \geq 0$ be a scalar and $\mathbf{b}_1, \dots, \mathbf{b}_d \in \mathbb{R}^{d-1}$. Consider the map*

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

The Jacobian of this map equals

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

2.4 Shadow simplex algorithm

Let $P = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$ be a polyhedron, where $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^d$ correspond to the rows of $\mathbf{A}$. We call a set $B \subset [n]$ a basis of $\mathbf{A}\mathbf{x} \leq \mathbf{b}$ if $\mathbf{A}_B$ is invertible, and we call B a feasible basis if $\mathbf{x}_B = \mathbf{A}_B^{-1}\mathbf{b}_B$ satisfies $\mathbf{A}\mathbf{x}_B \leq \mathbf{b}$. Note that a feasible basis induces a vertex of P. We say a feasible basis B is optimal for an objective $\mathbf{c} \in \mathbb{R}^d$ if $\mathbf{c}^\mathsf{T}\mathbf{A}_B^{-1} \geq \mathbf{0}$. Note that when this holds, $\max_{\mathbf{x} \in P} \langle \mathbf{c}, \mathbf{x} \rangle = \langle \mathbf{c}, \mathbf{x}_B \rangle$. We say the polyhedron is non-degenerate (or simple) if every vertex has exactly d tight inequalities.

The Shadow Simplex algorithm is a pivot rule for the simplex method. Given a feasible basis $B \subset [n]$ that is optimal for an objective $\mathbf{d} \in \mathbb{R}^d$ and an objective function $\mathbf{c} \in \mathbb{R}^d$ to optimize, the Shadow Simplex algorithm specifies which pivot steps to take to reach an optimal basis for $\mathbf{c}$. We parametrize $\mathbf{c}_\lambda := (1-\lambda)\mathbf{d} + \lambda\mathbf{c}$ and start at $\lambda = 0$. The shadow

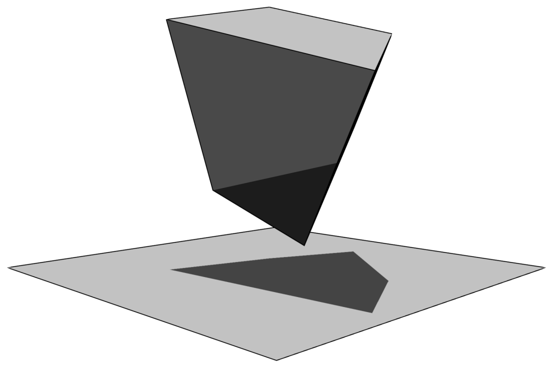

Figure 3: The shadow of a polyhedron.

pivot rule increases λ until there are $j \neq k \in [n]$ such that a new basis $B \cup \{j\} - \{k\}$ is optimal for $\mathbf{c}_\lambda$, and increases λ again. The index $k \in B$ is such that the coordinate for k in $\mathbf{c}_\lambda^\top \mathbf{A}_B^{-1}$ first lowers to 0, and $j \notin B$ is such that $B \cup \{j\} - \{k\}$ is a feasible basis: we follow the edge $\mathbf{A}_B^{-1}\mathbf{b}_B - \mathbf{A}_B^{-1}\mathbf{e}_k\mathbb{R}_+$ until we hit the first constraint $\mathbf{a}_j^\top \mathbf{x} \leq b_j$, and then replace k by j to get the new basis $B \cup \{j\} - \{k\}$.

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

We say the shadow path is non-degenerate if for every λ no more than two vertices of

Algorithm 1: Shadow Simplex method for non-degenerate LPs and shadow paths.

Input: $P = \left\{\mathbf{x} \in \mathbb{R}^d : A\mathbf{x} \leq \mathbf{b}\right\}$, $\mathbf{c}, \mathbf{d} \in \mathbb{R}^d$, feasible basis $B \subset [n]$ optimal for $\mathbf{d}$.
Output: optimal basis $B \subset [n]$ for $\mathbf{c}$.
$\lambda_0 \leftarrow 0$.
$i \leftarrow 0$.
loop
  $i \leftarrow i + 1$.
  $\lambda_i :=$ maximum λ such that $\mathbf{c}_\lambda^\top \mathbf{A}_B^{-1} \geq 0$, i.e. where B is optimal for $\lambda \in [\lambda_{i-1}, \lambda_i]$.
  if $\lambda_i \geq 1$ **then return** B.
  Set k as the index with $(c_{\lambda_i}^\top \mathbf{A}_B^{-1})_k = 0$.
  $j \leftarrow \arg\min \left\{ \frac{\mathbf{a}_j^\top \mathbf{A}_B^{-1}\mathbf{b}_B - b_j}{\mathbf{a}_j^\top \mathbf{A}_B^{-1}\mathbf{e}_k} : j \in [n] \setminus B, \frac{\mathbf{a}_j^\top \mathbf{A}_B^{-1}\mathbf{b}_B - b_j}{\mathbf{a}_j^\top \mathbf{A}_B^{-1}\mathbf{e}_k} > 0 \right\}$.
  if no such j exists **then return** *unbounded*.
  $B \leftarrow B \cup \{j\} - \{k\}$.

the polyhedron are optimal for $\mathbf{c}_\lambda$. If both the polyhedron and the shadow path are non-degenerate, a pivot step can be performed in $O(nd)$ time. From this point on we always assume this is the case, since any $d+1$ vectors in our model are affinely independent with probability 1 and any d vectors are linearly independent with probability 1.

For our purpose, we will mainly work with polyhedra of the form $\mathbf{Ax} \leq \mathbf{1}$, in which case $\mathbf{0}$ is always contained in the polyhedron. It is instructive to examine the geometry of shadow paths on such polyhedra from a polar perspective. For any polyhedron $P = \left\{\mathbf{x} \in \mathbb{R}^d : \mathbf{Ax} \leq \mathbf{1}\right\}$, the polar polytope is defined as the convex hull $P^\circ := \mathrm{conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n)$ of the origin and the constraint vectors. For any index-set $I \subset [n], |I| = d$, the set $\mathrm{conv}(\mathbf{a}_i)_{i \in I}$ forms a facet of the polytope $\mathrm{conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n)$ if and only if the (unique) solution $\mathbf{x}_I$ to the equations

$$\langle \mathbf{a}_i, \mathbf{x} \rangle = 1 \qquad \forall i \in I$$

is a vertex of the original polyhedron $\left\{\mathbf{x} \in \mathbb{R}^d : \mathbf{Ax} \leq 1\right\}$. For $I \subset [n], |I| = d-1$, the set $\mathrm{conv}(\mathbf{0}, (\mathbf{a}_i)_{i \in I})$ is a facet of $\mathrm{conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n)$ if and only if

$$\left\{\mathbf{x} \in \mathbb{R}^d : \langle \mathbf{a}_i, \mathbf{x} \rangle = 1, \forall i \in I, \left\langle \mathbf{a}_j, \mathbf{x} \right\rangle \leq \mathbf{1}, \forall i \in [n] \setminus I\right\}$$

is an unbounded ray.

In the polar perspective a pivot moves from one face of $\mathrm{conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n)$ to a neighboring face. The shadow simplex method moves the objective $\mathbf{c}_\lambda$ along the line segment $[\mathbf{d}, \mathbf{c}]$ and keeps track of which face of the polar is intersected by the ray $\mathbf{c}_\lambda \mathbb{R}^+$.

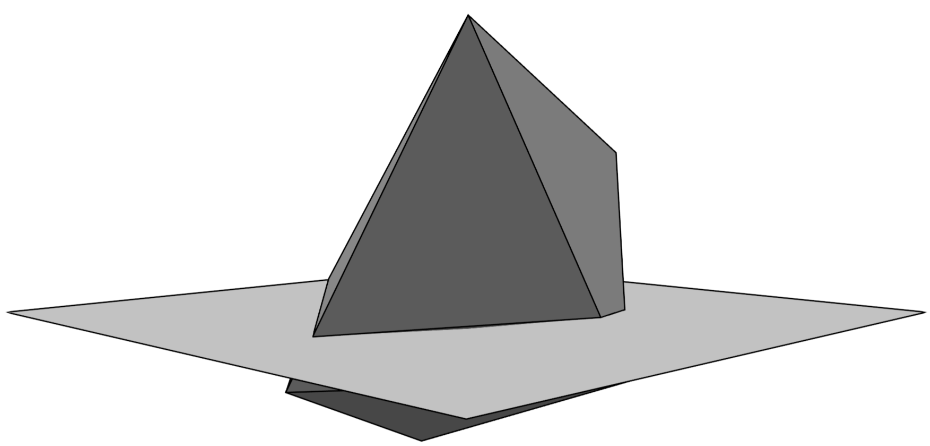

Figure 4: The polar of the above polytope intersected with the corresponding plane.

The number of pivot steps taken in a Shadow Simplex phase is bounded from above by the number of edges of the intersection $\mathrm{conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n) \cap \mathrm{span}(\mathbf{d}, \mathbf{c})$. Hence it suffices if we prove an upper bound on this geometric quantity. The following theorem gives the properties we will use of the shadow simplex algorithm, which we state without proof.

Theorem 5. *Let $P = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$ denote a non-degenerate polyhedron, where $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d$ are the rows of $\mathbf{A}$ and $\mathbf{b} \in \mathbb{R}^n$. Let $\mathbf{c}, \mathbf{d} \in \mathbb{R}^d$ denote two objectives inducing a non-degenerate shadow path on P, and let $W = \mathrm{span}(\mathbf{d}, \mathbf{c})$. Then given feasible basis $I \in \binom{[n]}{d}$ for $\mathbf{A}\mathbf{x} \leq \mathbf{b}$ which is optimal for $\mathbf{d}$, Algorithm 1 (Shadow Simplex) finds a basis $J \in \binom{n}{d}$ optimal for $\mathbf{c}$ in a number of pivot steps bounded by $|\mathrm{edges}(\pi_W(P))|$, where π_W is the orthogonal projection onto W. In particular, when $\mathbf{b} = \mathbf{1}$, we have that*

$$|\mathrm{edges}(\pi_W(P))| = |\mathrm{edges}(\mathrm{conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n) \cap W)|.$$

3 Shadow bounds

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

Theorem 6. *Let $W \subset \mathbb{R}^d$ be a fixed two-dimensional subspace, and let $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d, n \geq d \geq 3$ independent Laplace distributed random vectors with parameter σ and centers of norm at most 1. Then the expected number of edges is bounded by*

$$\mathbb{E}[|\mathrm{edges}(\mathrm{conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n) \cap W)|] = O(d^{2.5}\sigma^{-2} + d^3 \ln n\, \sigma^{-1} + d^3 \ln^2 n).$$

Theorem 7. *Let $W \subset \mathbb{R}^d$ be a fixed two-dimensional subspace, and let $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d, n \geq d \geq 3$ be independent Gaussian random vectors with variance σ^2 and centers of norm at most 1. Then the expected number of edges is bounded by*

$$\mathbb{E}[|\mathrm{edges}(\mathrm{conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n) \cap W)|] \leq \mathcal{D}_g(n, d, \sigma),$$

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

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

The proofs of Theorems 6 and 7 are given in subsections 3.2 and 3.3 respectively.

3.1 Shadow bound for parametrized distributions

In this subsection, we bound the expected number of edges of the two-dimensional polygon

$$\mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n) \cap W,$$

where $W \subset \mathbb{R}^d$ is a fixed two-dimensional plane and $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are independent vectors distributed according to parametrized distributions $\mu_1, \ldots, \mu_n$. The parameters we will use are defined below.

3.1.1 Distribution parameters

Definition 8. A probability distribution μ on $\mathbb{R}^d$ with density function $f : \mathbb{R}^d \to \mathbb{R}^+$ is *L-log-Lipschitz* if for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $|\ln(f(\mathbf{x})) - \ln(f(\mathbf{y}))| \leq L\|\mathbf{x} - \mathbf{y}\|$. Equivalently, μ is L-log-Lipschitz if $f(\mathbf{x})/f(\mathbf{y}) \leq \exp(L\|\mathbf{x} - \mathbf{y}\|)$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$.

Definition 9. Given a distribution μ on $\mathbb{R}^d$ we define the line variance τ^2 as the infimum of the variances when restricted to any fixed line $l \subset \mathbb{R}^d$:

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

Definition 10. Given a distribution μ on $\mathbb{R}^d$ with expectation $\mathbb{E}_{\mathbf{x} \sim \mu}[\mathbf{x}] = \mathbf{y}$ we define the n-th deviation r_n to be the smallest number such that for any unit vector $\boldsymbol{\theta} \in \mathbb{R}^d$,

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

Note that as r_n decreases, the left hand side of the above inequality increases while the right hand side decreases, so r_n is well-defined.

Definition 11. Given a distribution μ 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 $R_{n,d} := R(\frac{1}{d\binom{n}{d}})$.

We will use two relations between the parameters, which we prove in the lemmas below.

Lemma 12. *If a distribution μ is L-log-Lipschitz then its line variance satisfies $\tau \geq 1/(\sqrt{e}L)$.*

Proof. Let f be the probability density function of μ, and assume for simplicity of notation that μ is a measure on the real line $\mathbb{R}$ with expectation 0. With probability at least $1/e$ the variable has distance at least $1/L$ from its expectation:

$$\int_0^{\infty} f(t)\, \mathrm{d}t = \int_{1/L}^{\infty} f(t - 1/L)\, \mathrm{d}t \leq e \int_{1/L}^{\infty} f(t)\, \mathrm{d}t.$$

Similarly, $\int_{-\infty}^{0} f(t)\, \mathrm{d}t \leq e \int_{-\infty}^{-1/L} f(t)\, \mathrm{d}t$. Hence, the variance τ^2 is at least $1/(eL^2)$. □

Lemma 13. *For a d-dimensional distribution μ, $d \geq 3$, with parameters L, R as described above we have the inequality $LR(1/2) \geq d/3$.*

Proof. Let $\bar{R} := R(1/2)$. Suppose $L\bar{R} < d$. For $\alpha > 1$ to be chosen later we know

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

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:

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

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

3.1.2 Proof of shadow bound for parametrized distributions

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

Theorem 14 (Parametrized Shadow Bound)**.** *Let $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d$, $n \geq d \geq 3$, be independently distributed according to L-log-Lipschitz distributions $\mu_1, \ldots, \mu_n$ with centers of norm at most 1, line variances at least τ^2, cutoff radii at most $R_{n,d}$ and n-th deviations at most r_n. For any fixed two-dimensional linear subspace $W \subset \mathbb{R}^d$, the expected number of edges satisfies*

$$\mathbb{E}[|\text{edges}(\text{conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n) \cap W)|] \leq O(\frac{d^{1.5}L}{\tau}(1 + R_{n,d})(1 + r_n)).$$

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

In the rest of the section, $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d$, $n \geq d \geq 3$, will be as in Theorem 14. We use $Q := \text{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ to denote the convex hull of the constraint vectors and W to denote the two dimensional shadow plane.

As in the introduction, we will restrict ourselves to bounding $|\text{edges}(Q \cap W)|$, as $\mathbf{0}$ adds at most 1 edge to the convex hull.

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 2 to the expected number of edges, as there are only $\binom{n}{d}$ possible facets with non-zero probability of forming. We treat those separately.

Definition 15. For each set $I \in \binom{[n]}{d}$, let E_I denote the event that $\mathrm{conv}(\mathbf{a}_i)_{i\in I} \cap W$ forms an edge of $Q \cap W$.

Definition 16. 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 [KS06].

Lemma 17. *The expected number of edges in $Q \cap W$ satisfies*

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

Proof. We give a lower bound on the perimeter of the intersection $Q \cap W$ in terms of the number of edges.

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

By dividing on both sides of the inequality we can now conclude

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

□

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 section.

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

Lemma 18. *The expected perimeter of $Q \cap W$ is bounded by*

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

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

Proof. By convexity, the perimeter is bounded from above by 2π times the norm of the maximum norm point. Let $\hat{\mathbf{a}}_i$ denote the perturbation of $\mathbf{a}_i$ from the center of its distribution. We can now derive the bound

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

where the last inequality follows 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

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

Each of the two expectations satisfy, by the definition of the n-th deviation,

$$\begin{aligned}
\mathbb{E}[\max_{i \le n} |\langle \mathbf{v}_j, \hat{\mathbf{a}}_i \rangle|] &\le \int_0^\infty \Pr[\max_{i \le n} |\langle \mathbf{v}_j, \hat{\mathbf{a}}_i \rangle| > t] \, \mathrm{d}t \\
&\le \int_0^{r_n} \Pr[\max_{i \le n} |\langle \mathbf{v}_j, \hat{\mathbf{a}}_i \rangle| > t] \, \mathrm{d}t + \int_{r_n}^\infty \sum_{i \le n} \Pr[|\langle \mathbf{v}_j, \hat{\mathbf{a}}_i \rangle| > t] \, \mathrm{d}t \\
&\le 2r_n.
\end{aligned}$$

□

The rest of this section 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 $E := E_{[d]}$.

We assume $\mathbf{a}_1, \dots, \mathbf{a}_n$ satisfy the following non-degeneracy conditions:

1. For any $I \subseteq [n]$ with $|I| = d + 1$, the vectors $\mathbf{a}_i$ with $i \in I$ are affinely independent.
2. For any $I \subseteq [n]$ with $|I| = d$, there is a unique $(d-1)$-dimensional hyperplane through all $\mathbf{a}_i$ with $i \in I$.

These conditions are satisfied with probability 1, so from now on we always assume this to be the case. Any other degenerate situations of probability 0 are also ignored.

To lower bound the length $\mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_1, \dots, \mathbf{a}_d) \cap W) \mid E]$ we will need the pairwise distances between the different $\mathbf{a}_i$'s for $i \in \{1, \dots, d\}$ to be small along $\boldsymbol{\omega}^\perp$.

Definition 19 (Containing hyperplane)**.** Define $H = \text{aff}(\mathbf{a}_1, \dots, \mathbf{a}_d) = t\boldsymbol{\theta} + \boldsymbol{\theta}^\perp$, $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$, $t \geq 0$ to be the hyperplane containing $\mathbf{a}_1, \dots, \mathbf{a}_d$. Define $l = H \cap W$. Express $l = \mathbf{p} + \boldsymbol{\omega} \cdot \mathbb{R}$, where $\boldsymbol{\omega} \in \mathbb{S}^{d-1}$ and $\mathbf{p} \in \boldsymbol{\omega}^\perp$. Note that this representation is justified since l is a line in $\mathbb{R}^d$ with probability 1.

Definition 20 (Bounded diameter event)**.** We define the event D to hold exactly when $\|\pi_{\boldsymbol{\omega}^\perp}(\mathbf{a}_i) - \pi_{\boldsymbol{\omega}^\perp}(\mathbf{a}_j)\| \leq 2 + 2R_{n,d}$ 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 $I \in B$.

Lemma 21. *The expected edge length satisfies*

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

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

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

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

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

□

Definition 22 (Change of variables)**.** Recall the map $(\mathbf{a}_1, \dots, \mathbf{a}_d) \mapsto (\boldsymbol{\theta}, t, \mathbf{b}_1, \dots, \mathbf{b}_d)$, where $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t \geq 0, \mathbf{b}_1, \dots, \mathbf{b}_d \in \mathbb{R}^{d-1}$. For any $\boldsymbol{\theta}, t, \mathbf{b}_i$ we write $\bar{\mu}_i(\boldsymbol{\theta}, t, \mathbf{b}_i) = \mu_i(R_\theta(\mathbf{b}_i) + t\boldsymbol{\theta})$ and we write $\bar{\mu}_i(\mathbf{b}_i)$ when the values of $\boldsymbol{\theta}, t$ are clear.

By Theorem 4 of Blaschke [Bla35] we know that for any fixed values of $\boldsymbol{\theta}, t$ the vectors $\mathbf{b}_1, \ldots, \mathbf{b}_d$ have joint probability distribution proportional to

$$\mathrm{vol}_{d-1}(\mathrm{conv}(\mathbf{b}_1, \ldots, \mathbf{b}_d)) \prod_{i=1}^{d} \bar{\mu}_i(\mathbf{b}_i) \; . \tag{12}$$

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

We will condition on the shape of the projected simplex.

Definition 23 (Projected shape)**.** Define the projected shift variable $\mathbf{x} := \mathbf{x}_{\boldsymbol{\omega}}(\mathbf{b}_1) = \pi_{\bar{\boldsymbol{\omega}}^\perp}(\mathbf{b}_1)$ and shape variable $S := S_\omega(\mathbf{b}_1, \ldots, \mathbf{b}_d)$ by

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

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

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

Observe that $S \in \mathcal{S}$ exactly when D holds.

Lemma 24. *Let $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, \mathbf{b}_1, \ldots, \mathbf{b}_d \in \mathbb{R}^{d-1}$ denote the change of variables of $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^d$ as in* (22)*. Then, the expected length satisfies*

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

Proof. To derive the desired inequality, we first understand the effect of conditioning on E. Let E_0 denote the event that $F := \mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_d)$ induces a facet of Q. Note that E is equivalent to $E_0 \cap \bar{E}$, where $\bar{E}$ is as above. We now perform the change of variables from $\mathbf{a}_1, \ldots, \mathbf{a}_d \in \mathbb{R}^d$ to $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t \in \mathbb{R}_+$, $\mathbf{b}_1, \ldots, \mathbf{b}_d \in \mathbb{R}^{d-1}$ as in Definition 22. Recall that F induces a facet of Q iff $\langle \boldsymbol{\theta}, \mathbf{a}_{d+i} \rangle \leq t, \forall\, i \in [n-d]$. Given this, we see that

$$\begin{aligned}
&\mathbb{E}[\mathrm{length}(\mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_d) \cap W) \mid D, E] \\
&= \mathbb{E}[\mathrm{length}(\mathrm{conv}(\mathbf{b}_1, \ldots, \mathbf{b}_d) \cap \bar{l}) \mid D, E_0, \bar{E}] \\
&= \frac{\mathbb{E}[\, \mathbb{1}[\forall i \in [n-d], \langle \boldsymbol{\theta}, \mathbf{a}_{d+i} \rangle \leq t] \cdot \mathrm{length}(\mathrm{conv}(\mathbf{b}_1, \ldots, \mathbf{b}_d) \cap \bar{l}) \mid D, \bar{E}]}{\Pr[\forall i \in [n-d], \langle \boldsymbol{\theta}, \mathbf{a}_{d+i} \rangle \leq t \mid D, \bar{E}]} \\
&= \frac{\mathbb{E}_{\boldsymbol{\theta},t}[\, \mathbb{E}[\mathbb{1}[\forall i \in [n-d], \langle \boldsymbol{\theta}, \mathbf{a}_{d+i} \rangle \leq t] \cdot \mathrm{length}(\mathrm{conv}(\mathbf{b}_1, \ldots, \mathbf{b}_d) \cap \bar{l}) \mid \boldsymbol{\theta}, t, D, \bar{E}] \,]}{\mathbb{E}_{\boldsymbol{\theta},t}[\, \Pr[\forall i \in [n-d], \langle \boldsymbol{\theta}, \mathbf{a}_{d+i} \rangle \leq t \mid \boldsymbol{\theta}, t, D, \bar{E}] \,]}
\end{aligned} \tag{13}$$

Since $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are independent, conditioned on $\boldsymbol{\theta}, t$, the random vectors $\mathbf{b}_1, \ldots, \mathbf{b}_d$ are independent of $\langle \boldsymbol{\theta}, \mathbf{a}_{d+1} \rangle, \ldots, \langle \boldsymbol{\theta}, \mathbf{a}_n \rangle$. Since the events D and $\bar{E}$ only depend on $\mathbf{b}_1, \ldots, \mathbf{b}_d$, continuing from (13), we get that

$$\begin{aligned}
&\frac{\mathbb{E}_{\boldsymbol{\theta},t}[\ \mathbb{E}[\mathbb{1}[\forall i \in [n-d], \langle \boldsymbol{\theta}, \mathbf{a}_{d+i} \rangle \le t] \cdot \mathrm{length}(\mathrm{conv}(\mathbf{b}_1, \ldots, \mathbf{b}_d) \cap \bar{l}) \mid \boldsymbol{\theta}, t, D, \bar{E}]\ ]}{\mathbb{E}_{\boldsymbol{\theta},t}[\ \Pr[\forall i \in [n-d], \langle \boldsymbol{\theta}, \mathbf{a}_{d+i} \rangle \le t \mid \boldsymbol{\theta}, t, D, \bar{E}]\ ]} \\
&= \frac{\mathbb{E}_{\boldsymbol{\theta},t}[\ \Pr[\forall i \in [n-d], \langle \boldsymbol{\theta}, \mathbf{a}_{d+i} \rangle \le t \mid \boldsymbol{\theta}, t] \cdot \mathbb{E}[\mathrm{length}(\mathrm{conv}(\mathbf{b}_1, \ldots, \mathbf{b}_d) \cap \bar{l}) \mid \boldsymbol{\theta}, t, D, \bar{E}]\ ]}{\mathbb{E}_{\boldsymbol{\theta},t}[\ \Pr[\forall i \in [n-d], \langle \boldsymbol{\theta}, \mathbf{a}_{d+i} \rangle \le t \mid \boldsymbol{\theta}, t]\ ]} \\
&\ge \inf_{\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t \ge 0} \mathbb{E}[\mathrm{length}(\mathrm{conv}(\mathbf{b}_1, \ldots, \mathbf{b}_d) \cap \bar{l}) \mid \boldsymbol{\theta}, t, D, \bar{E}]\,.
\end{aligned}$$

Lastly, note that the event D is equivalent to $S := S_\omega(\mathbf{b}_1, \ldots, \mathbf{b}_d) \in \mathcal{S}$ as in (23). Therefore, for any fixed $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t \ge 0$,

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

□

Definition 25 (Kernel combination)**.** For $S \in \mathcal{S}$, define the combination $\mathbf{z} := \mathbf{z}(S)$ to be the unique (up to sign) $\mathbf{z} = (z_1, \ldots, z_d) \in \mathbb{R}^d$ satisfying

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

We note that $\mathbf{z}$ is well-defined as $\mathbf{z}$ is specified by d variables and set of $d-1$ *generic* linear equations. Note that for $S := S_\omega(\mathbf{b}_1, \ldots, \mathbf{b}_d)$, $\mathbf{z}$ satisfies $\pi_{\bar{\omega}^\perp}(\sum_{i=1}^d z_i \mathbf{b}_i) = \mathbf{0}$.

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

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

When S is clear we drop the subscript.

We write $\|C(\mathbf{q})\|_1$ for the ℓ_1-diameter of $C(\mathbf{q})$. Observe that $C(\mathbf{q})$ is a one-dimensional line segment of the form $C(\mathbf{q}) = \boldsymbol{\lambda}_\mathbf{q} + \mathbf{z} \cdot [0, d_\mathbf{q}]$, hence $\|C(\mathbf{q})\|_1 = d_\mathbf{q}$. The ℓ_1 diameter $\|C(\mathbf{q})\|_1$ specified by $\mathbf{q} \in \mathrm{conv}(S(\mathbf{b}_1, \ldots, \mathbf{b}_d))$ directly relates to the length of the chord $(\mathbf{q} + \mathbf{x} + \bar{\omega} \cdot \mathbb{R}) \cap \mathrm{conv}(\mathbf{b}_1, \ldots, \mathbf{b}_d)$, which projects to $\mathbf{q}$ under $\pi_{\bar{\omega}^\perp}$. The exact relation is given below.

Lemma 27. *Write* $(h_1, \ldots, h_d) = (\langle \bar{\omega}, \mathbf{b}_1 \rangle, \ldots, \langle \bar{\omega}, \mathbf{b}_d \rangle)$, $(\mathbf{s}_1, \ldots, \mathbf{s}_d) = S(\mathbf{b}_1, \ldots, \mathbf{b}_d)$ *and* $\mathbf{x} = \pi_{\bar{\omega}^\perp}(\mathbf{b}_1)$. *For any* $\mathbf{q} \in \mathrm{conv}(S)$ *the following equality holds:*

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

Proof. By construction there is a convex combination $\lambda_1, \ldots, \lambda_d \geq 0, \sum_{i=1}^d \lambda_i = 1$ satisfying $\sum_{i=1}^d \lambda_i \mathbf{s}_i = \mathbf{q}$ such that

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

From this we deduce

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

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

We can view the terms in the above product as follows: $\|C(\mathbf{y})\|_1$ measures if the line intersects the simplex in a wide or a narrow part, while $|\sum_{i=1}^d z_i h_i|$ measures the size of the simplex along the direction of $\bar{\omega}$. The last term can also be used to simplify the volume term in the probability density of $\mathbf{b}_1, \ldots, \mathbf{b}_d$ after we condition on the shape S. We prove this in the next lemma.

Lemma 28. *For fixed* $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t \geq 0, S \in \mathcal{S}$, *define* $\mathbf{x} \in \bar{\omega}^\perp, h_1, \ldots, h_d \in \mathbb{R}$ *conditioned on* $\boldsymbol{\theta}, t, S$ *to have joint probability density function proportional to*

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

where $\mathbf{z} := \mathbf{z}(S)$ *is as in Definition 25. Then for* $\mathbf{b}_1, \ldots, \mathbf{b}_d \in \mathbb{R}^{d-1}$ *distributed as in Lemma 24, conditioned on* $\boldsymbol{\theta}, t$ *and the shape* $S = (\mathbf{s}_1, \ldots, \mathbf{s}_d)$, *where* $\mathbf{s}_1 = \mathbf{0}$, *we have equivalence of the distributions*

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

Proof. By (12), the variables $\mathbf{b}_1, \dots, \mathbf{b}_d$ conditioned on $\boldsymbol{\theta}, t$ have density proportional to $\mathrm{vol}_{d-1}(\mathrm{conv}(\mathbf{b}_1, \dots, \mathbf{b}_d)) \prod_{i=1}^d \bar{\mu}_i(\mathbf{b}_i)$. We make a change of variables from $\mathbf{b}_1, \dots, \mathbf{b}_d$ to $\mathbf{x}, \mathbf{s}_2, \dots, \mathbf{s}_d \in \bar{\boldsymbol{\omega}}^\perp, h_1, \dots, h_d \in \mathbb{R}$, defined by

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

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

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

By Lemma 27 we find

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

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

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

□

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

Lemma 29. *Let $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t \geq 0, S \in \mathcal{S}$ be fixed. Let $\mathbf{b}_1, \dots, \mathbf{b}_d \in \mathbb{R}^{d-1}$, $\mathbf{x} \in \omega^\perp$, $h_1, \dots, h_d \in \mathbb{R}$ be distributed as in Lemma 28. Define $\mathbf{q} := \bar{\mathbf{p}} - \mathbf{x}$. Then, the expected edge length satisfies*

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

Proof. We start with the assertion of Lemma 27:

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

We take expectation on both sides to derive the equality

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

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

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

□

We will first bound the expected ℓ_1 diameter of $C(\mathbf{q})$, where $\mathbf{q} = \bar{\mathbf{p}} - \mathbf{x}$, which depends on where $\bar{\mathbf{p}} - \mathbf{x}$ intersects the projected simplex $\text{conv}(S)$: 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 30. *Let $\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t \geq 0$ and $S \in \mathcal{S}$ be fixed. Let $\mathbf{q} = \bar{\mathbf{p}} - \mathbf{x}$ be distributed as in Lemma 29. Then, the expected ℓ_1-diameter of $C(\mathbf{q})$ satisfies*

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

Proof. Let $\hat{\mu}$ denote the probability density of $\mathbf{q}$ conditioned on $\boldsymbol{\theta}, t, S, \bar{E}$. Note that $\hat{\mu}$ is supported on $\text{conv}(S)$ and has density proportional to $\int \cdots \int \prod_{i=1}^d \bar{\mu}_i(\bar{\mathbf{p}} - \mathbf{q} + \mathbf{s}_i + h_i \bar{\boldsymbol{\omega}}) \, \mathrm{d}h_1 \cdots \, \mathrm{d}h_d$. We claim that $\hat{\mu}$ is dL-log-Lipschitz. To see this, note that since $\bar{\mu}_1, \ldots, \bar{\mu}_d$ are L-log-Lipschitz, for $\mathbf{a}, \mathbf{b} \in \text{conv}(S)$ we have that

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

To get a lower bound on $\|C(\mathbf{q})\|_1$, we will use the fact that $\|C(\mathbf{q})\|_1$ is concave over $\text{conv}(S) = \text{conv}(\mathbf{s}_1, \ldots, \mathbf{s}_d)$, which follows by construction, and that $\max_{\mathbf{q} \in \text{conv}(S)} \|C(\mathbf{q})\|_1 \geq 2$. For the latter claim, we use the combination $\mathbf{y} := \sum_{i=1}^n |z_i| \mathbf{s}_i \in \text{conv}(S)$. Note that for $\gamma \in [-1, 1]$, $\sum_{i=1}^d (|z_i| + \gamma z_i) \mathbf{s}_i = \mathbf{y}$, $\sum_{i=1}^d |z_i| + \gamma z_i = \|\mathbf{z}\|_1 = 1$ and $|z_i| + \gamma z_i \geq 0, \forall i \in [d]$. Hence, $\|C(\mathbf{y})\|_1 \geq 2$ as desired.

Let $\alpha \in (0,1)$ be a scale factor to be chosen later. Now we can write

$$\begin{aligned}
\mathbb{E}[\|C(\mathbf{q})\| \mid \boldsymbol{\theta}, t, S, \bar{E}] &= \int_{\mathrm{conv}(S)} \|C(\mathbf{q})\|_1 \hat{\mu}(\mathbf{q}) \, \mathrm{d}\mathbf{q} \\
&\geq \int_{\alpha \mathrm{conv}(S)+(1-\alpha)\mathbf{y}} \|C(\mathbf{q})\|_1 \hat{\mu}(\bar{\mathbf{q}}) \, \mathrm{d}\mathbf{q}, \qquad (14)
\end{aligned}$$

because the integrand is non-negative. By concavity, we have the lower bound $\|C(\alpha\mathbf{q} + (1-\alpha)\mathbf{y})\| \geq 2(1-\alpha)$ for all $\mathbf{q} \in \mathrm{conv}(S)$. Therefore, (14) is lower bounded by

$$\begin{aligned}
&\geq \int_{\alpha \mathrm{conv}(S)+(1-\alpha)\mathbf{y}} 2(1-\alpha)\hat{\mu}(\mathbf{q}) \, \mathrm{d}\mathbf{q} \\
&= 2\alpha^d(1-\alpha) \int_{\mathrm{conv}(S)} \hat{\mu}(\alpha\mathbf{q} + (1-\alpha)\mathbf{y}) \, \mathrm{d}\mathbf{q} \\
&\geq 2\alpha^d(1-\alpha) e^{-\max_{\mathbf{q}\in\mathrm{conv}(S)}(1-\alpha)\|\mathbf{q}-\mathbf{y}\|\cdot dL} \int_{\mathrm{conv}(S)} \hat{\mu}(\mathbf{q}) \, \mathrm{d}\mathbf{q}, \\
&= 2\alpha^d(1-\alpha) e^{-\max_{i\in[d]}(1-\alpha)\|\mathbf{s}_i-\mathbf{y}\|\cdot dL}, \qquad (15)
\end{aligned}$$

where we used a change of variables in the first equality, the dL-log-Lipschitzness of $\hat{\mu}$ in the second inequality, and the convexity of the ℓ_2 norm in the last equality. Using the diameter bound of $2 + 2R_{n,d}$ for $\mathrm{conv}(S)$, (15) is lower bounded by

$$\geq 2\alpha^d(1-\alpha)e^{-(1-\alpha)dL(2+2R_{n,d})}. \qquad (16)$$

Setting $\alpha = 1 - \frac{1}{dL(2+2R_{n,d})} \geq 1 - 1/d$ (by Lemma 13) gives a lower bound for (16) 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 $\boldsymbol{\theta}$ and t are also fixed, restricting $\mathbf{b}_i$ to lie on a line is the same as restricting $\mathbf{a}_i$ to lie on a line. Thus, were it not for the correlation between $h_1, \dots, h_d$, i.e. the factor $|\sum_{i=1}^d z_i h_i|$ in the joint pdf, each h_i would be independent and have variance τ^2 by assumption, and thus one would expect $\mathbb{E}[|\sum_{i=1}^d z_i h_i|] = \Omega(\|\mathbf{z}\|\sigma)$. The following lemma establishes this, and shows that in fact, the correlation term only helps.

Lemma 31. *Let $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$, $t \geq 0$, $S \in \mathcal{S}$, $\mathbf{x} \in \bar{\omega}^\perp$ be fixed and let $\mathbf{z} := \mathbf{z}(S)$ be as in Definition 25. Then for $h_1, \dots, h_d \in \mathbb{R}$ distributed as in Lemma 29, the expected inner product satisfies*

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

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

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

Note that, by assumption, the variables $g_1, \ldots, g_d$ each have variance at least τ^2. We recall from Lemma 28 that the joint probability density of $h_1, \ldots, h_d$ is proportional to $|\sum_{i=1}^d z_i h_i| \prod_{i=1}^d \tilde{\mu}_i(h_i)$. Thus, may rewrite the above expectation as

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

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

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

We reach the desired conclusion by applying Lemma 1:

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

□

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

Proof of Theorem 14 (Parametrized Shadow Bound). We first observe that, since $\mathbf{0} \in W$, the number of edges can decrease by at most 1 after removing $\mathbf{0}$:

$$|\mathrm{edges}(\mathrm{conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n) \cap W)| \leq |\mathrm{edges}(\mathrm{conv}(\mathbf{a}_1, \ldots, \mathbf{a}_n) \cap W)| + 1.$$

We will bound the right hand side in expectation. By Lemma 17, we may derive the shadow bound by combining an upper bound on $\mathbb{E}[\mathrm{perimeter}(Q \cap W)]$ and a uniform lower bound on $\mathbb{E}[\mathrm{length}(\mathrm{conv}(\mathbf{a}_i)_{i \in I} \cap W) \mid E_I]$ for all $I \in B$. For the perimeter upper bound, by Lemma 18 we have that

$$\mathbb{E}[\mathrm{perimeter}(Q \cap W)] \leq 2\pi(1 + 4r_n). \tag{17}$$

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

$$
\begin{aligned}
&\mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_1, \dots, \mathbf{a}_d) \cap W) \mid E] \\
&\geq \frac{1}{2} \cdot \mathbb{E}[\text{length}(\text{conv}(\mathbf{a}_1, \dots, \mathbf{a}_d) \cap W) \mid D, E] \quad (\text{ Lemma 21 }) \\
&\geq \frac{1}{2} \cdot \inf_{\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t \geq 0} \mathbb{E}[\text{length}(\text{conv}(\mathbf{b}_1, \dots, \mathbf{b}_d) \cap \bar{l}) \mid \boldsymbol{\theta}, t, S \in \mathcal{S}, \bar{E}] \quad (\text{ Lemma 24 }) \\
&\geq \frac{1}{2} \cdot \inf_{\boldsymbol{\theta} \in \mathbb{S}^{d-1}, t \geq 0, S \in \mathcal{S}} \left(\mathbb{E}[\|C(\bar{\mathbf{p}} - \mathbf{x})\|_1 \mid \boldsymbol{\theta}, t, S, \bar{E}] \cdot \inf_{\mathbf{x} \in \bar{\omega}^{\perp}} \mathbb{E}[|\sum_{i=1}^{d} z_i h_i| \mid \boldsymbol{\theta}, t, S, \mathbf{x}] \right) \quad (\text{ Lemma 29 }) \\
&\geq \frac{1}{2} \cdot \frac{e^{-2}}{dL(1 + R_{n,d})} \cdot \frac{\tau}{2\sqrt{d}} \quad (\text{ Lemmas 30 and 31 }) .
\end{aligned} \tag{18}
$$

The theorem now follows by taking the ratio of (17) and (18). □

3.2 Shadow bound for Laplace distributed perturbations

In this section, we use Theorem 14 to prove a shadow bound for Laplace perturbations. To achieve this, we bound the needed parameters of the Laplace distribution below.

Lemma 32. *For $n \geq d \geq 3$, the Laplace distribution $L_d(\bar{\mathbf{a}}, \sigma)$, satisfies the following properties:*

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

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

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

2. The second property follows from Lemma 3:

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

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

$$\int_{7\sigma\ln n}^{\infty} \Pr[|\langle \mathbf{X}, \boldsymbol{\theta}\rangle| \geq t]\,\mathrm{d}t \leq \int_{7\sigma\ln n}^{\infty} e^{-\sqrt{d}t/(7\sigma)}\,\mathrm{d}t$$
$$= \frac{7\sigma}{\sqrt{d}} n^{-\sqrt{d}} \leq \frac{7\sigma\ln n}{n}.$$

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

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

4. This follows from the $\sqrt{d}/\sigma$-log-Lipschitzness and Lemma 12. □

Proof of Theorem 6. By Theorem 14, we get the desired bound on the expected shadow size by plugging in the bounds from 32 for L, $R_{n,d}$, r_n and τ into the upper bound $O((d^{1.5}L/\tau)(1 + R_{n,d})(1 + r_n))$. □

3.3 Shadow bound for Gaussian distributed perturbations

The Gaussian distribution is not log-Lipschitz, so we can not directly apply Theorem 14. We will define a *smoothed out* version of the Gaussian distribution to remedy this problem, which we call the Laplace-Gaussian distribution. The Laplace-Gaussian distribution, defined below, matches the Gaussian distribution with respect to every meaningful parameter, while also being log-Lipschitz. We will first bound the shadow size with respect to Laplace-Gaussian perturbations, and then show that the expected number of edges for Gaussian perturbations is not much higher.

Definition 33. We define a random variable $\mathbf{X} \in \mathbb{R}^d$ to be (σ, r)*-Laplace-Gaussian distributed with mean* $\bar{\mathbf{a}}$, or $\mathbf{X} \sim LG_d(\bar{\mathbf{a}}, \sigma, r)$, if its density is proportional to $f_{(\bar{\mathbf{a}},\sigma,r)} : \mathbb{R}^d \to \mathbb{R}_+$ given by

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

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

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

Lemma 34 (Laplace-Gaussian tail bounds)**.** *Let* $\mathbf{X} \in \mathbb{R}^d$ *be* (σ, r)*-Laplace-Gaussian distributed with mean* $\mathbf{0}$*, where* $r := c\sqrt{d \ln n}$*,* $c \geq 4$*. Then for* $t \geq r$*,*

$$\Pr[\|\mathbf{X}\| \geq \sigma t] \leq e^{-(1/4)(c\sqrt{d \ln n})t}. \tag{19}$$

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

$$\Pr[|\langle \mathbf{X}, \boldsymbol{\theta} \rangle| \geq \sigma t] \leq \begin{cases} e^{-(1/4)(c\sqrt{d \ln n})t} & : t \geq r \\ 3e^{-t^2/4} & : 0 \leq t \leq r. \end{cases} \tag{20}$$

Lemma 35. *For* $n \geq d \geq 3$*, the* $(\sigma, 4\sqrt{d \ln n})$*-Laplace-Gaussian distribution in* $\mathbb{R}^d$ *with mean* $\bar{\mathbf{a}}$ *satisfies the following properties:*

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

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

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

2. Applying the tail bound from Lemma 34, we get that

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

3. Again using Lemma 34,

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

4. For the line variance, by rotational symmetry, we may w.l.o.g. assume that $l := (\mathbf{y}, 0) + \mathbb{R}\mathbf{e}_d$, where $\mathbf{y} \in \mathbb{R}^{d-1}$, and so $(\mathbf{y}, 0)$ is the point on l closest to the origin. Since $f_{(\sigma,r)}(\mathbf{y}, \lambda) = f_{(\sigma,r)}(\mathbf{y}, -\lambda)$ for every $\lambda \in \mathbb{R}$, the expectation $\mathbb{E}[\mathbf{X} \mid \mathbf{X} \in l] = (\mathbf{y}, 0)$. Thus, $\mathrm{Var}(\mathbf{X} \mid \mathbf{X} \in l) = \mathbb{E}[X_d^2 \mid \mathbf{X} \in l]$.

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

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

$$1 \geq \frac{f_{(\sigma,r)}(\mathbf{y}, \lambda)}{f_{(\sigma,r)}(\mathbf{y}, 0)} \geq \frac{e^{-\|(\mathbf{y},\lambda)\|^2/(2\sigma^2)}}{e^{-\|(\mathbf{y},0)\|^2/(2\sigma^2)}} = e^{-\lambda^2/(2\sigma^2)} \geq e^{-1/2} \,. \tag{21}$$

Given the above, we have that

$$\begin{aligned} \mathbb{E}[X_d^2 \mid \mathbf{X} \in \bar{l}] &\geq (\sigma^2/4) \Pr[|X_d| \geq \sigma/2 \mid \mathbf{X} \in \bar{l}] \\ &= (\sigma^2/4) \frac{\int_{\sigma/2}^{\sigma} f_{(\sigma,r)}(\mathbf{y}, t)\, \mathrm{d}t}{\int_0^{\sigma} f_{(\sigma,r)}(\mathbf{y}, t)\, \mathrm{d}t} \\ &\geq (\sigma^2/4) \frac{\int_{\sigma/2}^{\sigma} f_{(\sigma,r)}(\mathbf{y}, 0) e^{-1/2}\, \mathrm{d}t}{\int_0^{\sigma} f_{(\sigma,r)}(\mathbf{y}, 0)\, \mathrm{d}t} \quad (\text{ by (21) }) \\ &= (\sigma^2/4)(e^{-1/2}/2) \geq \sigma^2/16 \;\text{, as needed .} \end{aligned} \tag{22}$$

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

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

In particular, we get that

$$1 \geq \frac{f_{(\sigma,r)}(\mathbf{y}, \lambda)}{f_{(\sigma,r)}(\mathbf{y}, 0)} = \frac{e^{-\|(\mathbf{y},\lambda)\|(r/\sigma)}}{e^{-\|(\mathbf{y},0)\|(r/\sigma)}} \geq e^{-\lambda^2/(2\sigma^2)} \geq e^{-1/2} \,. \tag{23}$$

The desired lower bound now follows by combining (22), (23). □

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

Lemma 36. *For $d \geq 3$, the number of edges of* $\mathrm{conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n) \cap W$ *satisfies*

$$\mathbb{E}_{N_d(\sigma)}[|\mathrm{edges}(\mathrm{conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n))|] \leq 1 + \mathbb{E}_{LG_d(\sigma, 4\sqrt{d \ln n})}[|\mathrm{edges}(\mathrm{conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n))|].$$

Proof. Let us abbreviate edges $:=$ edges(conv($\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n$)) and let $r := 4\sqrt{d \ln n}$. We make use of the fact that $N_d(\sigma)$ and $LG_d(\sigma, r)$ are equal when restricted to distance at most σr from their centers.

$$\begin{aligned}\mathbb{E}_{N(\sigma)}[|\mathrm{edges}|] = &\Pr_{N_d(\sigma)}[\exists i \in [n]\ \|\hat{\mathbf{a}}_i\| > \sigma r]\mathbb{E}_{N_d(\sigma)}[|\mathrm{edges}| \mid \exists i \in [n]\ \|\hat{\mathbf{a}}_i\| > \sigma r] \\ &+ \Pr_{N_d(\sigma)}[\forall i \in [n]\ \|\hat{\mathbf{a}}_i\| \leq \sigma r]\mathbb{E}_{N_d(\sigma)}[|\mathrm{edges}| \mid \forall i \in [n]\ \|\hat{\mathbf{a}}_i\| \leq \sigma r]. \qquad (24)\end{aligned}$$

By Lemma 34, the first probability is at most $n^{-4d} \leq n^{-d}/4$, so we upper bound the first number of edges by $\binom{n}{d}$ making a total contribution of less than $1/4$. Now we use the fact that within radius $4\sigma\sqrt{d \ln n}$ we have equality of densities between $N_d(\sigma)$ and $LG_d(\sigma, r)$. Continuing from (24),

$$\begin{aligned}&\leq 1/4 + \mathbb{E}_{N_d(\sigma)}[|\mathrm{edges}| \mid \forall i \in [n]\ \|\hat{\mathbf{a}}_i\| \leq \sigma r] \\ &= 1/4 + \mathbb{E}_{LG_d(\sigma,r)}[|\mathrm{edges}| \mid \forall i \in [n]\ \|\hat{\mathbf{a}}_i\| \leq \sigma r] \\ &\leq 1/4 + \mathbb{E}_{LG_d(\sigma,r)}[|\mathrm{edges}|] / \Pr_{LG_d(\sigma,r)}[\forall i \in [n]\ \|\hat{\mathbf{a}}_i\| \leq \sigma r]. \qquad (25)\end{aligned}$$

The inequality above is true by non-negativity of the number of edges. Next we lower bound the denominator and continue (25),

$$\begin{aligned}&\leq 1/4 + \mathbb{E}_{LG_d(\sigma,r)}[|\mathrm{edges}|]/(1 - n^{-d}/4) \\ &\leq 1/4 + (1 + n^{-d}/2)\mathbb{E}_{LG_d(\sigma,r)}[|\mathrm{edges}|]. \qquad (26)\end{aligned}$$

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

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

□

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

Proof of Theorem 7. By Lemma 36 we know that

$$\mathbb{E}_{N_d(\sigma)}[|\mathrm{edges}(\mathrm{conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n))|] \leq 1 + \mathbb{E}_{LG_d(\sigma, 4\sqrt{d \ln n})}[|\mathrm{edges}(\mathrm{conv}(\mathbf{0}, \mathbf{a}_1, \ldots, \mathbf{a}_n))|].$$

We now derive the shadow bound for Laplace-Gaussian perturbations by combining the parameter bounds in Lemma 35 with the parameterized shadow bound in Theorem 14.

□

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

Proof of Lemma 34. By homogeneity, we may w.l.o.g. assume that $\sigma = 1$. Define auxilliary distributions $\mathbf{Y} \in \mathbb{R}^d$ distributed as $(\mathbf{0}, 1/(c\sqrt{\ln n}))$-Laplace and $\mathbf{Z} \in \mathbb{R}^d$ be distributed as $N_d(\mathbf{0}, 1)$.

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

$$\begin{aligned} \mathbf{Z} \mid \|\mathbf{Z}\| \leq r \equiv \mathbf{X} \mid \|\mathbf{X}\| \leq r \\ \mathbf{Y} \mid \|\mathbf{Y}\| \geq r \equiv \mathbf{X} \mid \|\mathbf{X}\| \geq r \end{aligned} \tag{27}$$

Proof of (19)**.** By the above, for any $t \geq r$, we have that

$$\Pr[\|\mathbf{X}\| \geq t] = \Pr[\|\mathbf{Y}\| \geq t] \cdot \frac{\Pr[\|\mathbf{X}\| \geq r]}{\Pr[\|\mathbf{Y}\| \geq r]} \,. \tag{28}$$

For the first term, by the Laplace tail bound (9), we get that

$$\Pr[\|\mathbf{Y}\| \geq t] \leq e^{-c\sqrt{d \ln n}t - d\ln(\frac{c\sqrt{\ln n}t}{\sqrt{d}}) - d} \,. \tag{29}$$

For the second term,

$$\begin{aligned} \frac{\Pr[\|\mathbf{X}\| \geq r]}{\Pr[\|\mathbf{Y}\| \geq r]} &= e^{r^2/2} \frac{\int_{\mathbb{R}^n} e^{-r\|\mathbf{x}\|} \, \mathrm{d}\mathbf{x}}{\int_{\mathbb{R}^n} f_{(\sigma,r)}(\mathbf{x}) \, \mathrm{d}\mathbf{x}} \leq e^{r^2/2} \frac{\int_{\mathbb{R}^n} e^{-r\|\mathbf{x}\|} \, \mathrm{d}\mathbf{x}}{\int_{\mathbb{R}^n} e^{-\|\mathbf{x}\|^2/2} \, \mathrm{d}\mathbf{x}} \\ &\leq e^{r^2/2} \frac{r^{-d} d! \mathrm{vol}_d(\mathcal{B}_2^d)}{\sqrt{2\pi}^d} \leq e^{(dc^2 \ln n)/2} (\frac{\sqrt{e}}{c\sqrt{\ln n}})^d \\ &\leq e^{(dc^2 \ln n)/2} \,, \end{aligned} \tag{30}$$

where we have used the upper bound $\mathrm{vol}_d(\mathcal{B}_2^d) \leq (2\pi e/d)^{d/2}$, $r = c\sqrt{d} \ln n$ and $c \geq \sqrt{e}$. Combining (29), (30) and that $t \geq r$, $c \geq 4$, we get

$$\begin{aligned} \Pr[\|\mathbf{X}\| \geq t] &\leq e^{-c\sqrt{d \ln n}t - d\ln(\frac{c\sqrt{\ln n}t}{\sqrt{d}}) - d} \cdot e^{(dc^2 \ln n)/2} \\ &\leq e^{-c\sqrt{d \ln n}t/2 - d\ln(\frac{c\sqrt{\ln n}t}{\sqrt{d}}) - d} = e^{-d(\frac{rt}{2d} - \ln(\frac{rt}{d}) - 1)} \\ &\leq e^{-d(\frac{rt}{4d})} = e^{-c\sqrt{d \ln n}t/4} , \end{aligned} \tag{31}$$

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

Proof of (20). For $t \geq r$, using the bound (19), we get

$$\Pr[|\langle \mathbf{X}, \boldsymbol{\theta} \rangle| \geq t] \leq \Pr[\|\mathbf{X}\| \geq t] \leq e^{-c\sqrt{d \ln n} t/4} . \tag{32}$$

For $t \leq r$, we see that

$$\begin{aligned} \Pr[|\langle \mathbf{X}, \boldsymbol{\theta} \rangle| \geq t] &\leq \Pr[|\langle \mathbf{X}, \boldsymbol{\theta} \rangle| \geq t, \|\mathbf{X}\| \leq r] + \Pr[\|\mathbf{X}\| \geq r] \\ &\leq \Pr[|\langle \mathbf{X}, \boldsymbol{\theta} \rangle| \geq t, \|\mathbf{X}\| \leq r] + n^{-(c^2 d)/4} . \end{aligned} \tag{33}$$

By the identity (27), for the first term, using the Gaussian tailbound (8), we have that

$$\begin{aligned} \Pr[|\langle \mathbf{X}, \boldsymbol{\theta} \rangle| \geq t, \|\mathbf{X}\| \leq r] &= \Pr[|\langle \mathbf{Z}, \boldsymbol{\theta} \rangle| \geq t, \|\mathbf{Z}\| \leq r] \cdot \frac{\Pr[\|\mathbf{X}\| \leq r]}{\Pr[\|\mathbf{Z}\| \leq r]} \\ &= \Pr[|\langle \mathbf{Z}, \boldsymbol{\theta} \rangle| \geq t, \|\mathbf{Z}\| \leq r] \cdot \frac{\int_{\mathbb{R}^n} e^{-\|\mathbf{x}\|^2/2} \, \mathrm{d}\mathbf{x}}{\int_{\mathbb{R}^n} f_{(1,r)}(\mathbf{x}) \, \mathrm{d}\mathbf{x}} \\ &\leq \Pr[|\langle \mathbf{Z}, \boldsymbol{\theta} \rangle| \geq t] \leq 2e^{-t^2/2} . \end{aligned} \tag{34}$$

The desired inequality (20) now follows directly by combining (32), (33), (34), noting that $2e^{-t^2/2} + n^{-(c^2 d)/4} \leq 3e^{-t^2/4}$ for $0 \leq t \leq r$. □

4 Simplex algorithms

In this section, we give details on how to use the shadow bound in a black box manner to bound the complexity of shadow simplex algorithms. We restrict our attention here to Gaussian perturbations, as the details for Laplace perturbations are similar. We will follow the two stage interpolation strategy given by Vershynin in [Ver09]. For large σ however, we will replace Vershynin's random vertex (RV) phase I by Borgwardt's dimension by dimension (DD) algorithm [Bor82]. We will also use a slightly simplified version of the RV algorithm, which improves the polylogarithmic dependence on n. As most of the results here follow rather directly from prior work, we keep the exposition short and refer the reader to [Ver09, Bor87] for additional details.
To recall, our goal is to solve the smoothed LP

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

where $\mathbf{A} \in \mathbb{R}^{n \times d}$, $\mathbf{b} \in \mathbb{R}^n$. Here each row $(\mathbf{a}_i, b_i)$, $i \in [n]$, of $(\mathbf{A}, \mathbf{b})$ is a variance σ^2 Gaussian random vector with mean $(\bar{\mathbf{a}}_i, \bar{b}_i) := \mathbb{E}[(\mathbf{a}_i, b_i)]$ of ℓ_2 norm at most 1. We will say that (Smooth LP) is unbounded (bounded) if the system $\mathbf{c}^\mathsf{T}\mathbf{x} > 0, \mathbf{A}\mathbf{x} \leq 0$ is feasible (infeasible). Note that (Smooth LP) can be unbounded and infeasible under this definition. Furthermore, if it is bounded and feasible, then it has an optimal solution.

The runtimes we achieve are given below.

Theorem 37. *(Smooth LP) can be solved via a two phase shadow simplex method, with either RV or DD for Phase I, using an expected number of pivots given below:*

1. *DD:* $O(d^3\sqrt{\ln n}\sigma^{-2} + d^{3.5}\ln n\sigma^{-1} + d^{3.5}\ln^{3/2} n)$
2. *RV:* $O(d^2\sqrt{\ln n}\sigma^{-2} + d^5\ln^{3/2} n)$

Proof. For the DD based algorithm, combining Lemma 38 and Corollary 40, the expected number of simplex pivots is bounded by

$$2 + \mathcal{D}_g(d,n,\sigma/2) + \sum_{k=3}^{d} \mathcal{D}_g(k,n,\sigma) \, .$$

By the smoothed Gaussian shadow bound 7, the above is bounded by

$$O(d\mathcal{D}_g(d,n,\sigma)) = O(d^3\sqrt{\ln n}\sigma^{-2} + d^{3.5}\ln n\sigma^{-1} + d^{3.5}\ln^{3/2} n) \, ,$$

as needed. For the RV based algorithm, combining Lemma 38 and lemma 42, the expected number of simplex pivots is bounded by

$$3 + \mathcal{D}_g(d,n,\sigma/2) + 4\mathcal{D}_g(d,n+d,\min\{\sigma,\sigma_0\}/4) \, .$$

where σ_0 is as defined in 35. Noting that $1/\sigma_0 = O(d^{3/2}\sqrt{\log n})$ since $n \geq d$, by the smoothed Gaussian shadow bound 7, the above is bounded by

$$O(\mathcal{D}_g(d,n,\sigma) + \mathcal{D}_g(d,n,(d^{3/2}\sqrt{\log n})^{-1})) = O(d^2\sqrt{\ln n}\sigma^{-2} + d^5\ln^{3/2} n) \, ,$$

as needed. □

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

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

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

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

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

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

If (Smooth LP) is unbounded (i.e. the system $\mathbf{c}^\mathsf{T}\mathbf{x} > 0, \mathbf{A}\mathbf{x} \leq \mathbf{0}$ is feasible), this will be detected during Phase I as (Unit LP) is also unbounded. If (Smooth LP) is infeasible but bounded, then the shadow simplex run will terminate at a vertex having $\lambda < 1$. Thus, all cases can be detected by the two phase procedure (see [Ver09, Proposition 4.1] for a formal proof).

From here, we notice that (Int. LP), apart from the constraints $0 \leq \lambda \leq 1$, is of smoothed unit type. Namely, the rows of $(\mathbf{A}, \mathbf{1} - \mathbf{b})$ are variance σ^2 Gaussians centered at means of norm at most 2 (by the triangle inequality). Furthermore, note the rows of (Unit LP) are also variance σ^2 Gaussians centered at means of norm at most 1 (since $\mathbf{b}$ is deleted). Thus, with the constraints $0 \leq \lambda \leq 1$ deleted, the expected size of the (Int. LP) shadow is bounded by $\mathcal{D}_g(d, n, \sigma/2)$ ($\sigma/2$ since the centers have norm at most 2). Furthermore, since the shadow plane contains $(\mathbf{0}, 1)$, the constraints $0 \leq \lambda \leq 1$ can only increase the number of edges in the shadow by 2. By combining these observations together we directly derive the following lemma of Vershynin [Ver09].

Lemma 38. *If (Unit LP) is unbounded, then (Smooth LP) is unbounded. If (Unit LP) is bounded, then given an optimal solution to (Unit LP) one can solve (Smooth LP) using at most an expected $\mathcal{D}_g(d, n, \sigma/2) + 2$ shadow simplex pivots over (Int. LP).*

Given the above, our main task is now to solve (Unit LP), i.e. either to find an optimal solution or to determine unboundedness. To solve this Phase I, we provide the guarantees of Borgwardt's DD algorithm and continue with our simplification of Vershynin's RV.

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

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

where $k \in \{2, \dots, d\}$ and $\mathbf{c}_k := (c_1, \dots, c_k, 0, \dots, 0)$. The main idea here is that the solution of (Unit LP$_k$), $k \in \{2, \dots, d-1\}$, is generically on an edge of the shadow of (Unit LP$_{k+1}$) on the span of $\mathbf{c}_k$ and $\mathbf{e}_{k+1}$, which is sufficient to initialize the shadow simplex path in the next step. We note that Borgwardt's algorithm can be applied to any LP of (Unit LP) form (not necessarily smoothed) as long as appropriate non-degeneracy conditions hold (which occur w.p.1 for smoothed LPs). To avoid degeneracy, we will assume that $\mathbf{c}_k \neq \mathbf{0}$ for all $k \in \{2, \dots, d\}$, which can always be achieved by permuting the coordinates.

Theorem 39 ([Bor82]). *Let W_k, $k \in \{2, \dots, d\}$, denote the shadow of (Unit LP$_k$) on the span of $\mathbf{c}_{k-1}$ and $\mathbf{e}_k$. Then if each (Unit LP$_k$) and shadow W_k is non-degenerate, for $k \in \{2, \dots, d\}$, there exists a shadow simplex method which solves (Unit LP) using at most $\sum_{k=3}^{d} |\mathrm{edges}(W_k)|$ number of pivots, as well as $O(n)$ arithmetic operations to solve (Unit LP$_2$).*

We note that the solution of (Unit LP$_2$) is distinguished here since our shadow bounds only hold for $d \geq 3$. We ignore this contribution in the sequel as a single simplex pivot already requires $O(dn)$ time in the full space. From the above, we derive the following immediate corollary.

Corollary 40. *The smoothed (Unit LP) can be solved using an expected $\sum_{k=3}^{d} \mathcal{D}_g(k, n, \sigma)$ number of shadow simplex pivots.*

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

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

$$1/\sigma \geq 1/\sigma_0 := \max c_0 \left\{ \sqrt{d \log n}, d^{3/2} \sqrt{\log d} \right\} \tag{35}$$

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

We now present a slightly simplified version of the RV algorithm (algorithm 2). For this purpose, define $\mathbf{s}_0 \in \mathbb{R}^d$ be a unit vector, and let $\mathbf{s}_1, \dots, \mathbf{s}_d$ denote the vertices of $d-1$ dimensional regular simplex centered at $\mathbf{s}_0$, such that $\langle \mathbf{s}_i - \mathbf{s}_0, \mathbf{s}_0 \rangle = 0$ and $\|\mathbf{s}_i - \mathbf{s}_0\| = 1/(c_1 \sqrt{\log d})$, for all $i \in [d]$.

Compared to Vershynin's implementation, our algorithm 2 does not increase the length of the added constraints as a function of the maximum row norm of $\mathbf{A}$. Instead we use a single cutoff radius at distance 1.5, which simplifies the analysis and avoids additional polylogarithmic factors in n.

We begin with some preliminary remarks for algorithm 2. First, the goal of lines 2-4 is to create a new artificial LP, (Unit LP') $\max \mathbf{c}^\mathsf{T}\mathbf{x}$, $\mathbf{A}\mathbf{x} \leq \mathbf{1}$, $\mathbf{R}\mathbf{x} \leq \mathbf{1}$, such that $\mathbf{v}$ is a vertex of

Algorithm 2: Random Vertex algorithm

Input: $\mathbf{c} \in \mathbb{R}^d$, $\mathbf{A} \in \mathbb{R}^{n \times d}$, $\mathbf{A}$ is variance $\sigma^2 \leq \sigma_0^2$ Gaussian with rows having centers of norm at most 1.

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

1 If the rows of $\mathbf{A}$ have norm at most 1.5, continue. Otherwise, solve (Unit LP) via the shadow simplex method starting at an arbitrary feasible vertex (obtained via ray shooting from the origin), and return either "unbounded" or the optimal solution.

loop

2 Let $\mathbf{U} \in \mathbb{R}^{d \times d}$ be a random rotation, and let $\mathbf{R} \in \mathbb{R}^{d \times d}$ be the random matrix satisfying $\mathbb{E}[\mathbf{R}^\mathsf{T}|\mathbf{U}] := 3\mathbf{U}(\mathbf{s}_1, \dots, \mathbf{s}_d)$ and $\mathbf{R}|\mathbf{U}$ is variance σ^2 Gaussian.

3 Compute $\mathbf{v} := \mathbf{R}^{-1}\mathbf{1}$ and $\mathbf{d} := \mathbf{U}\mathbf{s}_0$.

4 If not $\mathbf{A}\mathbf{v} \leq \mathbf{1}$ and $\mathbf{d}^\mathsf{T}\mathbf{R}^{-1} \geq \mathbf{0}$, restart the loop.

5 Solve the (Unit LP') $\max \mathbf{c}^\mathsf{T}\mathbf{x}, \mathbf{A}\mathbf{x} \leq \mathbf{1}, \mathbf{R}\mathbf{x} \leq \mathbf{1}$ by following the shadow simplex path induced by $\mathbf{d}$ and $\mathbf{c}$ initialized at $\mathbf{v}$.

6 If (Unit LP') is unbounded return "unbounded". If (Unit LP') is bounded and the optimal vertex $\mathbf{v}^*$ does not hit the constraints $\mathbf{R}\mathbf{x} \leq \mathbf{1}$, return $\mathbf{v}^*$ as the optimal solution to (Unit LP). Otherwise, restart the loop.

the corresponding system which maximizes $\mathbf{d}$ (checked on line 4). Having passed these checks, (Unit LP') is solved on line 5 via shadow simplex initialized at vertex $\mathbf{v}$ with objective $\mathbf{d}$. Lastly, on line 6, it is checked whether the solution to (Unit LP') is a solution to (Unit LP). Correctness of the algorithm is thus straightforward.

For the runtime, the crux of the analysis is the following theorem of Vershynin, which establishes that conditions (i) and (ii), as stated earlier, hold with reasonable probability. We note that the original statement requires $1/\sigma_0 \geq c_0 d^{3/2} \log d$, however an inspection of the proof reveals that $1/\sigma_0 \geq c_0 d^{3/2} \sqrt{\log d}$ (as defined in 35) is sufficient.

Theorem 41 (Theorem 5.4 [Ver09])**.** *Given $\sigma \leq \sigma_0$, then for $c_0, c_1 \geq 1$ large enough constants, conditioned on entering the main loop in algorithm 2, the probability that a main loop iteration succeeds in solving (Unit LP) is at least $1/4$.*

Using the above theorem, we easily derive a bound on the expected number of pivots to solve (Unit LP).

Lemma 42. *Given $\sigma \leq \sigma_0$, then for $c_0, c_1 \geq 1$ large enough absolute constants, algorithm 2 solves (Unit LP) in at most an expected $1 + 4\mathcal{D}_g(d, n+d, \sigma/4)$ number of shadow simplex pivots.*

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

Pivots from line 1. Let E denote the event that the rows of $\mathbf{a_1}, \dots, \mathbf{a_n}$ all have norm at most 1.5. Noting that each $\mathbf{a}_i$, $i \in [n]$, is a variance σ^2 Gaussian and $1/\sigma \geq c_0\sqrt{d \ln n}$, by Lemma 2 (Gaussian concentration), we have that

$$\begin{aligned}\Pr[E^c] &= \Pr[\exists i \in [n] : \|\mathbf{a}_i\| \geq 1.5] \leq n \Pr[\|\mathbf{a}_1 - \bar{\mathbf{a}}_1\| \geq 1/2] \\ &\leq n \Pr[\|\mathbf{a}_1 - \bar{\mathbf{a}}_1\| \geq (c_0/2)\sqrt{d \ln n}\sigma] \leq e^{-(d/2)((c_0/2)\sqrt{\ln n}-1)^2} \leq n^{-d} ,\end{aligned}$$

for $c_0 \geq 6$. Therefore, the shadow simplex run on line 1 is executed with probability at most n^{-d} incurring at most $n^{-d}\binom{n}{d} \leq 1$ pivots on expectation for $c_0 \geq 6$.

Pivots from the main loop. To begin, recall that we enter the main loop exactly when event E occurs, thus all expectations are conditioned on E below. We first bound the expected number pivots in any iteration of the loop.

Fix any instantiation of $\mathbf{U}$ in line 2, and let $\mathbf{R} := \mathbf{R}|\mathbf{U}$ denote the corresponding Gaussian matrix. Index the rows of $\mathbf{R}$ by $\mathbf{a}_{n+1}, \dots, \mathbf{a}_{n+d}$. Note that these are variance σ^2 Gaussian random vectors and have centers of norm bounded by

$$\|\mathbb{E}[\mathbf{a}_{n+i}]\| = \|3\mathbf{U}\mathbf{s}_i\| \leq 3(\|\mathbf{s}_0\| + \frac{1}{c_1\sqrt{\ln d}}) \leq 4,\ \forall i \in [d],$$

for $c_1 \geq 3$. Let $F := F \mid E, \mathbf{U}$ denote the event that the checks on line 4 succeed. Note that shadow plane $W := \operatorname{span}(\mathbf{c}, \mathbf{U}\mathbf{s}_0)$ used by the shadow simplex algorithm on line 5 is fixed when conditioning on $\mathbf{U}$. Thus, conditioned on $E, \mathbf{U}$, the expected number of pivots in one iteration is equal to

$$\begin{aligned}&\Pr[F \mid E, \mathbf{U}] \cdot \mathbb{E}[|\text{edges}(\text{conv}(\mathbf{0}, \mathbf{a}_1, \dots, \mathbf{a}_{n+d}) \cap W)| \mid E, F, \mathbf{U}] \\ &\leq \Pr[F \mid E, \mathbf{U}] \cdot \frac{\mathbb{E}[|\text{edges}(\text{conv}(\mathbf{0}, \mathbf{a}_1, \dots, \mathbf{a}_{n+d}) \cap W)| \mid \mathbf{U}]}{\Pr[F, E \mid \mathbf{U}]} \\ &= \frac{\mathbb{E}[|\text{edges}(\text{conv}(\mathbf{0}, \mathbf{a}_1, \dots, \mathbf{a}_{n+d}) \cap W)| \mid \mathbf{U}]}{\Pr[E]} \leq \frac{\mathcal{D}_g(d, n+d, \sigma/4)}{\Pr[E]} .\end{aligned}$$

Since this bound holds for any $\mathbf{U}$, it clearly upper bounds the expected number of pivots in any iteration. By Theorem 41, conditioned on E, each main loop iteration succeeds with probability at least $1/4$, and hence we execute in expectation at most 4 iterations. Thus, the total expected number of simplex pivots coming from the main loop is bounded by

$$\Pr[E] \cdot (4\mathcal{D}_g(d, n+d, \sigma/4)/\Pr[E]) = 4\mathcal{D}_g(d, n+d, \sigma/4) .$$

Final Bound. Combining the results from the above paragraphs, we get that the total expected number of simplex pivots in algorithm 2 is bounded by:

$$\Pr[E^c]\binom{n}{d} + \Pr[E](4\mathcal{D}_g(d, n+d, \sigma/4)/\Pr[E]) \leq 1 + 4\mathcal{D}_g(d, n+d, \sigma/4) ,$$

for $c_0, c_1 \geq 1$ large enough, as needed. □

References

[AB14] Karim A. Adiprasito and Bruno Benedetti. The hirsch conjecture holds for normal flag complexes. *Mathematics of Operations Research*, 39(4):1340–1348, 2014.

[AC78] David Avis and Vasek Chvátal. Notes on blands pivoting rule. In *Polyhedral Combinatorics*, pages 24–34. Springer, 1978.

[Adl83] Ilan Adler. The expected number of pivots needed to solve parametric linear programs and the efficiency of the self-dual simplex method. Technical report, University of California, Dept. of IE and OR, 1983.

[AKS87] Ilan Adler, Richard M. Karp, and Ron Shamir. A simplex variant solving an $m \times d$ linear program in $O(\min(m^2, d^2))$ expected number of pivot steps. *J. Complexity*, 3(4):372–387, 1987.

[AM85] Ilan Adler and Nimrod Megiddo. A simplex algorithm whose average number of steps is bounded between two quadratic functions of the smaller dimension. *Journal of the ACM (JACM)*, 32(4):871–895, 1985.

[AMR11] David Arthur, Bodo Manthey, and Heiko Röglin. Smoothed analysis of the k-means method. *J. ACM*, 58(5):Art. 19, 31, 2011. Preliminary version in FOCS '09.

[AZ98] N. Amenta and G.M. Ziegler. Deformed products and maximal shadows. *Contemporary Math.*, 223:57–90, 1998.

[Bal84] M. L. Balinski. The Hirsch conjecture for dual transportation polyhedra. *Math. Oper. Res.*, 9(4):629–633, 1984.

[Bal97] Keith Ball. An elementary introduction to modern convex geometry. *Flavors of geometry*, 31:1–58, 1997.

[Bar74] David Barnette. An upper bound for the diameter of a polytope. *Discrete Math.*, 10:9–13, 1974.

[BCM+15] Tobias Brunsch, Kamiel Cornelissen, Bodo Manthey, Heiko Röglin, and Clemens Rösner. Smoothed analysis of the successive shortest path algorithm. *SIAM J. Comput.*, 44(6):1798–1819, 2015. Preliminary version in SODA '13.

[BDLF17] S. Borgwardt, J. A. De Loera, and E. Finhold. The diameters of network-flow polytopes satisfy the hirsch conjecture. *Mathematical Programming*, Jul 2017.

[BDSE+14] Nicolas Bonifas, Marco Di Summa, Friedrich Eisenbrand, Nicolai Hähnle, and Martin Niemeier. On sub-determinants and the diameter of polyhedra. *Discrete Comput. Geom.*, 52(1):102–115, 2014. Preliminary version in SOCG '12.

[BFG+00] Robert E. Bixby, Mary Fenelon, Zonghao Gu, Ed Rothberg, and Roland Wunderling. MIP: theory and practice—closing the gap. In *System modelling and optimization (Cambridge, 1999)*, pages 19–49. Kluwer Acad. Publ., Boston, MA, 2000.

[BGR15] Tobias Brunsch, Anna Großwendt, and Heiko Röglin. Solving totally unimodular LPs with the shadow vertex algorithm. In *Proceedings of the 32nd Symposium on Theoretical Aspects of Computer Science*, STACS '15, 2015.

[Bix12] Robert E. Bixby. A brief history of linear and mixed-integer programming computation. *Doc. Math.*, pages 107–121, 2012.

[Bla35] Wilhelm Blaschke. INTEGRALGEOMETRIE 2. *Bulletin mathématique de la Société Roumaine des Sciences*, 37(1):3–11, 1935.

[Bor77] K.H. Borgwardt. *Untersuchungen zur Asymptotik der mittleren Schrittzahl von Simplexverfahren in der linearen Optimierung*. PhD thesis, Universität Kaiserslautern, 1977.

[Bor82] K.-H. Borgwardt. The average number of pivot steps required by the simplex-method is polynomial. *Z. Oper. Res. Ser. A-B*, 26(5):A157–A177, 1982.

[Bor87] Karl-Heinz Borgwardt. *The simplex method: A probabilistic analysis*, volume 1 of *Algorithms and Combinatorics: Study and Research Texts*. Springer-Verlag, Berlin, 1987.

[Bor99] Karl Heinz Borgwardt. A sharp upper bound for the expected number of shadow vertices in LP-polyhedra under orthogonal projection on two-dimensional planes. *Math. Oper. Res.*, 24(3):544–603, 1999. A corrected version of Figure 1 appears in Volume 24(4), pp. 925984.

[BvdHS06] Graham Brightwell, Jan van den Heuvel, and Leen Stougie. A linear bound on the diameter of the transportation polytope. *Combinatorica*, 26(2):133–139, 2006.

[Dan48] G. Dantzig. Programming in a linear structure. Technical report, U.S. Air Force Comptroller, USAF, Washington, D.C., 1948.

[Dan51] George B. Dantzig. Maximization of a linear function of variables subject to linear inequalities. In *Activity Analysis of Production and Allocation*, Cowles Commission Monograph No. 13, pages 339–347. John Wiley & Sons, Inc., New York, N. Y.; Chapman & Hall, Ltd., London, 1951.

[Dan59] George Dantzig. *Linear programming and extensions*. Princeton university press, 1959.

[DF94] Martin Dyer and Alan Frieze. Random walks, totally unimodular matrices, and a randomised dual simplex algorithm. *Mathematical Programming*, 64(1):1–16, 1994.

[DGKK79] Robert Dial, Fred Glover, David Karney, and Darwin Klingman. A computational analysis of alternative algorithms and labeling techniques for finding shortest path trees. *Networks*, 9(3):215–248, 1979.

[DH16] Daniel Dadush and Nicolai Hähnle. On the shadow simplex method for curved polyhedra. *Discrete Comput. Geom.*, 56(4):882–909, 2016. Preliminary version in SOCG '15.

[DLKOS09] Jesús A. De Loera, Edward D. Kim, Shmuel Onn, and Francisco Santos. Graphs of transportation polytopes. *J. Combin. Theory Ser. A*, 116(8):1306–1325, 2009.

[DS05] Amit Deshpande and Daniel A. Spielman. Improved smoothed analysis of the shadow vertex simplex method. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '05, pages 349–356, 2005.

[ERV14] Matthias Englert, Heiko Röglin, and Berthold Vöcking. Worst case and probabilistic analysis of the 2-opt algorithm for the TSP. *Algorithmica*, 68(1):190–264, 2014. Preliminary version in SODA '07.

[EV17] Friedrich Eisenbrand and Santosh Vempala. Geometric random edge. *Mathematical Programming*, pages 1–15, 2017.

[FG92] John J. Forrest and Donald Goldfarb. Steepest-edge simplex algorithms for linear programming. *Math. Programming*, 57(3, Ser. A):341–374, 1992.

[FHZ11] Oliver Friedmann, Thomas Dueholm Hansen, and Uri Zwick. Subexponential lower bounds for randomized pivoting rules for the simplex algorithm. In *Proceedings of the Forty-third Annual ACM Symposium on Theory of Computing*, STOC '11, pages 283–292, New York, NY, USA, 2011. ACM.

[Fri11] Oliver Friedmann. A subexponential lower bound for zadeh's pivoting rule for solving linear programs and games. In Oktay Günlük and Gerhard J. Woeginger, editors, *Proceedings of 15th International Conference on Integer Programming and Combinatorial Optimization*, IPCO '11, pages 192–206, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[GB14] Markus Göhl and Karl Heinz Borgwardt. The average number of pivot steps of the simplex-algorithm based on a generalized rotation-symmetry-model. *Math. Methods Oper. Res.*, 80(3):329–366, 2014.

[GGT91] Andrew V. Goldberg, Michael D. Grigoriadis, and Robert E. Tarjan. Use of dynamic trees in a network simplex algorithm for the maximum flow problem. *Math. Programming*, 50(3, (Ser. A)):277–290, 1991.

[GH90] Donald Goldfarb and Jianxiu Hao. A primal simplex algorithm that solves the maximum flow problem in at most nm pivots and $O(n^2m)$ time. *Mathematical Programming*, 47(1):353–365, 1990.

[GH92] Donald Goldfarb and Jianxiu Hao. Polynomial-time primal simplex algorithms for the minimum cost network flow problem. *Algorithmica*, 8(2):145–160, 1992.

[GHK90] Donald Goldfarb, Jianxiu Hao, and Sheng-Roan Kai. Efficient shortest path simplex algorithms. *Oper. Res.*, 38(4):624–628, 1990.

[Gol76] Donald Goldfarb. Using the steepest-edge simplex algorithm to solve sparse linear programs. *Sparse matrix computations*, pages 227–240, 1976.

[Gol83] Donald Goldfarb. Worst case complexity of the shadow vertex simplex algorithm. Technical report, Columbia University, New York, 1983.

[GS55] Saul Gass and Thomas Saaty. The computational algorithm for the parametric objective function. *Naval Res. Logist. Quart.*, 2:39–45, 1955.

[GS79] Donald Goldfarb and William Y. Sit. Worst case behavior of the steepest edge simplex method. *Discrete Appl. Math.*, 1(4):277–285, 1979.

[Hai83] M. Haimovich. The simplex method is very good! on the expected number of pivot steps and related properties of random linear programs. Technical report, Columbia University, 1983.

[Höf95] G. Höfner. *Lineare Optimierung mit dem Schatteneckenalgorithmus: Untersuchungen zum mittleren Rechenaufwand und Entartungsverhalten*. PhD thesis, Universität Augsburg, 1995.

[Hun83] Ming S. Hung. A polynomial simplex method for the assignment problem. *Oper. Res.*, 31(3):595–600, 1983.

[HZ15] Thomas Dueholm Hansen and Uri Zwick. An improved version of the random-facet pivoting rule for the simplex algorithm. In *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pages 209–218, New York, NY, USA, 2015. ACM.

[Jer73] R. G. Jeroslow. The simplex algorithm with the pivot rule of maximizing criterion improvement. *Discrete Math.*, 4:367–377, 1973.

[Kal92] Gil Kalai. A subexponential randomized simplex algorithm (extended abstract). In *Proceedings of the Twenty-fourth Annual ACM Symposium on Theory of Computing*, STOC '92, pages 475–482, New York, NY, USA, 1992. ACM.

[Kar84] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984.

[Kha79] L. G. Khachiyan. A polynomial algorithm in linear programming. *Dokl. Akad. Nauk SSSR*, 244(5):1093–1096, 1979.

[KK92] Gil Kalai and Daniel J. Kleitman. A quasi-polynomial bound for the diameter of graphs of polyhedra. *Bull. Amer. Math. Soc. (N.S.)*, 26(2):315–316, 1992.

[KM70] Victor Klee and George J Minty. How good is the simplex algorithm. Technical report, WASHINGTON UNIV SEATTLE DEPT OF MATHEMATICS, 1970.

[KS06] Jonathan A. Kelner and Daniel A. Spielman. A randomized polynomial-time simplex algorithm for linear programming. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, STOC '06, pages 51–60. ACM, New York, 2006.

[Lar70] D. G. Larman. Paths of polytopes. *Proc. London Math. Soc. (3)*, 20:161–178, 1970.

[Lem65] Carlton E Lemke. Bimatrix equilibrium points and mathematical programming. *Management science*, 11(7):681–689, 1965.

[LS14] Y. T. Lee and A. Sidford. Path finding methods for linear programming: Solving linear programs in $\sqrt{\text{rank}}$ iterations and faster algorithms for maximum flow. In *Proceedings of the IEEE 55th Annual Symposium on Foundations of Computer Science*, FOCS '14, pages 424–433, Oct 2014.

[Meg85] N Megiddo. A note on the generality of the self-dual algorithm with various starting points. *Methods of operations research*, 49:271–275, 1985.

[Meg86] Nimrod Megiddo. Improved asymptotic analysis of the average number of steps performed by the self-dual simplex algorithm. *Math. Programming*, 35(2):140–172, 1986.

[Meh92] Sanjay Mehrotra. On the implementation of a primal-dual interior point method. *SIAM J. Optim.*, 2(4):575–601, 1992.

[MSW96] J. Matoušek, M. Sharir, and E. Welzl. A subexponential bound for linear programming. *Algorithmica*, 16(4-5):498–516, 1996.

[MSW15] Benjamin Matschke, Francisco Santos, and Christophe Weibel. The width of five-dimensional prismatoids. *Proc. Lond. Math. Soc. (3)*, 110(3):647–672, 2015.

[Mur80] Katta G. Murty. Computational complexity of parametric linear programming. *Math. Programming*, 19(2):213–219, 1980.

[Nad89] Denis Naddef. The Hirsch conjecture is true for $(0,1)$-polytopes. *Math. Programming*, 45(1, Ser. B):109–110, 1989.

[OPT93] James B. Orlin, Serge A. Plotkin, and Éva Tardos. Polynomial dual network simplex algorithms. *Math. Programming*, 60(3, Ser. A):255–276, 1993.

[Orl84] James B Orlin. Genuinely polynomial simplex and non-simplex algorithms for the minimum cost flow problem. Technical report, Massachusetts Institute of Technology, 1984.

[PY15] Ian Post and Yinyu Ye. The simplex method is strongly polynomial for deterministic markov decision processes. *Mathematics of Operations Research*, 40(4):859–868, 2015. Preliminary version in SODA '14.

[Ren88] James Renegar. A polynomial-time algorithm, based on Newton's method, for linear programming. *Math. Programming*, 40(1, (Ser. A)):59–93, 1988.

[San12] Francisco Santos. A counterexample to the Hirsch conjecture. *Ann. of Math. (2)*, 176(1):383–412, 2012.

[Sch14] E. Schnalzger. *Lineare Optimierung mit dem Schatteneckenalgorithmus im Kontext probabilistischer Analysen*. PhD thesis, Universität Augsburg, 2014. English translation by K.H. Borgwardt available via the following link.

[Sha87] Ron Shamir. The efficiency of the simplex method: a survey. *Management Sci.*, 33(3):301–334, 1987.

[Sma83] Steve Smale. On the average number of steps of the simplex method of linear programming. *Mathematical programming*, 27(3):241–262, 1983.

[SST06] Arvind Sankar, Daniel A. Spielman, and Shang-Hua Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM J. Matrix Anal. Appl.*, 28(2):446–476, 2006.

[ST03] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of termination of linear programming algorithms. *Math. Program.*, 97(1-2, Ser. B):375–404, 2003. ISMP, 2003 (Copenhagen).

[ST04] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time. *J. ACM*, 51(3):385–463 (electronic), 2004.

[Tod86] Michael J Todd. Polynomial expected behavior of a pivoting algorithm for linear complementarity and linear programming problems. *Mathematical Programming*, 35(2):173–192, 1986.

[Tod14] Michael J Todd. An improved kalai–kleitman bound for the diameter of a polyhedron. *SIAM Journal on Discrete Mathematics*, 28(4):1944–1947, 2014.

[Ver09] Roman Vershynin. Beyond Hirsch conjecture: walks on random polytopes and smoothed complexity of the simplex method. *SIAM J. Comput.*, 39(2):646–678, 2009. Preliminary version in FOCS '06.

[Ye11] Yinyu Ye. The simplex and policy-iteration methods are strongly polynomial for the Markov decision problem with a fixed discount rate. *Mathematics of Operations Research*, 36(4):593–603, 2011.