Smoothed Analysis of Algorithms: Why the Simplex Algorithm Usually Takes Polynomial Time *

Daniel A. Spielman †
Department of Mathematics
Massachusetts Institute of Technology

Shang-Hua Teng ‡
Department of Computer Science
Boston University, and
Akamai Technologies Inc.

July 8, 2004

Abstract

We introduce the *smoothed analysis of algorithms*, which continuously interpolates between the worst-case and average-case analyses of algorithms. In smoothed analysis, we measure the maximum over inputs of the expected performance of an algorithm under small random perturbations of that input. We measure this performance in terms of both the input size and the magnitude of the perturbations. We show that the simplex algorithm has *smoothed complexity* polynomial in the input size and the standard deviation of Gaussian perturbations.

*An extended abstract of this paper appeared in the *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*, pp. 296-305, 2001.

†Partially supported by an Alfred P. Sloan Foundation Fellowship, NSF CAREER award CCR-9701304, NSF grant CCR-0112487, and a Junior Faculty Research Leave sponsored by the M.I.T. School of Science

‡Partially suppoted by an Alfred P. Sloan Foundation Fellowship, and NSF grant CCR: 99-72532. Part of this work was done while at UIUC and visiting the department of mathematics at M.I.T.

Contents

List of Theorems, Lemmas, Corollaries and Propositions

1 Introduction

The Analysis of Algorithms community has been challenged by the existence of remarkable algorithms that are known by scientists and engineers to work well in practice, but whose theoretical analyses are negative or inconclusive. The root of this problem is that algorithms are usually analyzed in one of two ways: by worst-case or average-case analysis. Worst-case analysis can improperly suggest that an algorithm will perform poorly by examining its performance under the most contrived circumstances. Average-case analysis was introduced to provide a less pessimistic measure of the performance of algorithms, and many practical algorithms perform well on the random inputs considered in average-case analysis. However, average-case analysis may be unconvincing as the inputs encountered in many application domains may bear little resemblance to the random inputs that dominate the analysis.

We propose an analysis that we call smoothed analysis which can help explain the success of algorithms that have poor worst-case complexity and whose inputs look sufficiently different from random that average-case analysis cannot be convincingly applied. In smoothed analysis, we measure the performance of an algorithm under slight random perturbations of arbitrary inputs. In particular, we consider Gaussian perturbations of inputs to algorithms that take real inputs, and we measure the running times of algorithms in terms of their input size and the standard deviation of the Gaussian perturbations.

We show that the simplex method has polynomial smoothed complexity. The simplex method is the classic example of an algorithm that is known to perform well in practice but which takes exponential time in the worst case [KM72, Mur80, GS79, Gol83, AC78, Jer73, AZ99]. In the late 1970's and early 1980's the simplex method was shown to converge in expected polynomial time on various distributions of random inputs by researchers including Borgwardt, Smale, Haimovich, Adler, Karp, Shamir, Megiddo, and Todd [Bor80, Bor77, Sma83, Hai83, AKS87, AM85, Tod86]. These works introduced novel probabilistic tools to the analysis of algorithms, and provided some intuition as to why the simplex method runs so quickly. However, these analyses are dominated by "random looking" inputs: even if one were to prove very strong bounds on the higher moments of the distributions of running times on random inputs, one could not prove that an algorithm performs well in any particular small neighborhood of inputs.

To bound expected running times on small neighborhoods of inputs, we consider linear programming problems in the form

$$\begin{aligned} &\text{maximize} && \boldsymbol{z}^T \boldsymbol{x} \\ &\text{subject to} && \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{y}, \end{aligned} \tag{1}$$

and prove that for every vector $\boldsymbol{z}$ and every matrix $\bar{\boldsymbol{A}}$ and vector $\bar{\boldsymbol{y}}$, the expectation over standard deviation $\sigma\left(\max_i \|(\bar{y}_i, \bar{\boldsymbol{a}}_i)\|\right)$ Gaussian perturbations $\boldsymbol{A}$ and $\boldsymbol{y}$ of $\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{y}}$ of the time taken by a two-phase shadow-vertex simplex method to solve such a linear program is polynomial in $1/\sigma$ and the dimensions of $\boldsymbol{A}$.

1.1 Linear Programming and the Simplex Method

It is difficult to overstate the importance of linear programming to optimization. Linear programming problems arise in innumerable industrial contexts. Moreover, linear program-

ming is often used as a fundamental step in other optimization algorithms. In a linear programming problem, one is asked to maximize or minimize a linear function over a polyhedral region.

Perhaps one reason we see so many linear programs is that we can solve them efficiently. In 1947, Dantzig [Dan51] introduced the simplex method, which was the first practical approach to solving linear programs and which remains widely used today. To state it roughly, the simplex method proceeds by walking from one vertex to another of the polyhedron defined by the inequalities in (1). At each step, it walks to a vertex that is better with respect to the objective function. The algorithm will either determine that the constraints are unsatisfiable, determine that the objective function is unbounded, or reach a vertex from which it cannot make progress, which necessarily optimizes the objective function.

Because of its great importance, other algorithms for linear programming have been invented. In 1979, Khachiyan [Kha79] applied the ellipsoid algorithm to linear programming and proved that it always converged in time polynomial in d, n, and L—the number of bits needed to represent the linear program. However, the ellipsoid algorithm has not been competitive with the simplex method in practice. In contrast, the interior-point method introduced in 1984 by Karmarkar [Kar84], which also runs in time polynomial in d, n, and L, has performed very well: variations of the interior point method are competitive with and occasionally superior to the simplex method in practice.

In spite of half a century of attempts to unseat it, the simplex method remains the most popular method for solving linear programs. However, there has been no satisfactory theoretical explanation of its excellent performance. A fascinating approach to understanding the performance of the simplex method has been the attempt to prove that there always exists a short walk from each vertex to the optimal vertex. The Hirsch conjecture states that there should always be a walk of length at most $n - d$. Significant progress on this conjecture was made by Kalai and Kleitman [KK92], who proved that there always exists a walk of length at most $n^{\log_2 d+2}$. However, the existence of such a short walk does not imply that the simplex method will find it.

A simplex method is not completely defined until one specifies its *pivot rule*—the method by which it decides which vertex to walk to when it has many to choose from. There is no deterministic pivot rule under which the simplex method is known to take a subexponential number of steps. In fact, for almost every deterministic pivot rule there is a family of polytopes on which it is known to take an exponential number of steps [KM72, Mur80, GS79, Gol83, AC78, Jer73]. (See [AZ99] for a survey and a unified construction of these polytopes). The best present analysis of randomized pivot rules shows that they take expected time $n^{O(\sqrt{d})}$[Kal92, MSW96], which is quite far from the polynomial complexity observed in practice. This inconsistency between the exponential worst-case behavior of the simplex method and its everyday practicality leave us wanting a more reasonable theoretical analysis.

Various average-case analyses of the simplex method have been performed. Most relevant to this paper is the analysis of Borgwardt [Bor77, Bor80], who proved that the simplex method with the shadow vertex pivot rule runs in expected polynomial time for polytopes whose constraints are drawn independently from spherically symmetric distributions (*e.g.* Gaussian distributions centered at the origin). Independently, Smale [Sma83, Sma82] proved bounds on the expected running time of Lemke's self-dual parametric simplex algorithm on

linear programming problems chosen from a spherically-symmetric distribution. Smale's analysis was substantially improved by Megiddo [Meg86].

While these average-case analyses are significant accomplishments, it is not clear whether they actually provide intuition for what happens on typical inputs. Edelman [Ede92] writes on this point:

> What is a mistake is to psychologically link a random matrix with the intuitive notion of a "typical" matrix or the vague concept of "any old matrix."

Another model of random linear programs was studied in a line of research initiated independently by Haimovich [Hai83] and Adler [Adl83]. Their works considered the maximum over matrices, $\boldsymbol{A}$, of the expected time taken by parametric simplex methods to solve linear programs over these matrices in which the directions of the inequalities are chosen at random. As this framework considers the maximum of an average, it may be viewed as a precursor to smoothed analysis—the distinction being that the random choice of inequalities cannot be viewed as a perturbation, as different choices yield radically different linear programs. Haimovich and Adler both proved that parametric simplex methods would take an expected linear number of steps to go from the vertex minimizing the objective function to the vertex maximizing the objective function, even conditioned on the program being feasible. While their theorems confirmed the intuitions of many practitioners, they were geometric rather than algorithmic[1] as it was not clear how an algorithm would locate either vertex. Building on these analyses, Todd [Tod86], Adler and Megiddo [AM85], and Adler, Karp and Shamir [AKS87] analyzed parametric algorithms for linear programming under this model and proved quadratic bounds on their expected running time. While the random inputs considered in these analyses are not as special as the random inputs obtained from spherically symmetric distributions, the model of randomly flipped inequalities provokes some similar objections.

1.2 Smoothed Analysis of Algorithms and Related Work

We introduce the *smoothed analysis of algorithms* in the hope that it will help explain the good practical performance of many algorithms that worst-case does not and for which average-case analysis is unconvincing. Our first application of the smoothed analysis of algorithms will be to the simplex method. We will consider the maximum over $\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{y}}$ of the expected running time of the simplex method on inputs of the form

$$\begin{aligned} &\text{maximize} && \boldsymbol{z}^T\boldsymbol{x} \\ &\text{subject to} && (\bar{\boldsymbol{A}} + \boldsymbol{G})\boldsymbol{x} \leq (\bar{\boldsymbol{y}} + \boldsymbol{h}), \end{aligned} \tag{2}$$

where we let $\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{y}}$ be arbitrary and $\boldsymbol{G}$ and $\boldsymbol{h}$ be a matrix and a vector of independently chosen Gaussian random variables of mean 0 and standard deviation $\sigma\left(\max_i \|(\bar{y}_i, \bar{\boldsymbol{a}}_i)\|\right)$. If we let σ go to 0, then we obtain the worst-case complexity of the simplex method; whereas, if we let σ be so large that $\boldsymbol{G}$ swamps out $\boldsymbol{A}$, we obtain the average-case analyzed by Borgwardt. By choosing polynomially small σ, this analysis combines advantages of worst-case and average-case analysis, and roughly corresponds to the notion of imprecision in low-order digits.

[1] Our results in Section 4 are analogous to these results.

In a smoothed analysis of an algorithm, we assume that the inputs to the algorithm are subject to slight random perturbations, and we measure the complexity of the algorithm in terms of the input size and the standard deviation of the perturbations. If an algorithm has low smoothed complexity, then one should expect it to work well in practice since most real-world problems are generated from data that is inherently noisy. Another way of thinking about smoothed complexity is to observe that if an algorithm has low smoothed complexity, then one must be unlucky to choose an input instance on which it performs poorly.

We now provide some definitions for the smoothed analysis of algorithms that take real or complex inputs. For an algorithm A and input $\boldsymbol{x}$, let

$$\mathcal{C}_A(\boldsymbol{x})$$

be a complexity measure of A on input $\boldsymbol{x}$. Let X be the domain of inputs to A, and let X_n be the set of inputs of size n. The size of an input can be measured in various ways. Standard measures are the number of real variables contained in the input and the sums of the bit-lengths of the variables. Using this notation, one can say that A has worst-case $\mathcal{C}$-complexity $f(n)$ if

$$\max_{\boldsymbol{x} \in X_n} (\mathcal{C}_A(\boldsymbol{x})) = f(n).$$

Given a family of distributions μ_n on X_n, we say that A has average-case $\mathcal{C}$-complexity $f(n)$ under μ if

$$\mathop{\mathbf{E}}_{\boldsymbol{x} \stackrel{\mu_n}{\leftarrow} X_n} [\mathcal{C}_A(\boldsymbol{x})] = f(n).$$

Similarly, we say that A has *smoothed $\mathcal{C}$-complexity* $f(n, \sigma)$ if

$$\max_{\boldsymbol{x} \in X_n} \mathop{\mathbf{E}}_{\boldsymbol{g}} [\mathcal{C}_A(\boldsymbol{x} + (\sigma \left\|\boldsymbol{x}\right\|_?) \boldsymbol{g})] = f(n, \sigma), \tag{3}$$

where $(\sigma \left\|\boldsymbol{x}\right\|_?) \boldsymbol{g}$ is a vector of Gaussian random variables of mean 0 and standard deviation $\sigma \left\|\boldsymbol{x}\right\|_?$ and $\left\|\boldsymbol{x}\right\|_?$ is a measure of the magnitude of $\boldsymbol{x}$, such as the largest element or the norm. We say that an algorithm has *polynomial smoothed complexity* if its smoothed complexity is polynomial in n and $1/\sigma$. In Section 6, we present some generalizations of the definition of smoothed complexity that might prove useful. To further contrast smoothed analysis with average-case analysis, we note that the probability mass in (3) is concentrated in a region of radius $O(\sigma\sqrt{n})$ and volume at most $O(\sigma\sqrt{n})^n$, and so, when σ is small, this region contains an exponentially small fraction of the probability mass in an average-case analysis. Thus, even an extension of average-case analysis to higher moments will not imply meaningful bounds on smoothed complexity.

A discrete analog of smoothed analysis has been studied in a collection of works inspired by Santha and Vazirani's *semi-random source* model [SV86]. In this model, an adversary generates an input, and each bit of this input has some probability of being flipped. Blum and Spencer [BS95] design a polynomial-time algorithm that k-colors k-colorable graphs generated by this model. Feige and Krauthgamer [FK] analyze a model in which the adversary is more powerful, and use it to show that Turner's algorithm [Tur86] for approximating the bandwidth performs well on semi-random inputs. They also improve Turner's analysis. Feige and Kilian [FK98] present polynomial-time algorithms that recover large independent sets, k-colorings, and optimal bisections in semi-random graphs. They also demonstrate that significantly better results would lead to surprising collapses of complexity classes.

1.3 Our Results

We consider the maximum over $\boldsymbol{z}$, $\bar{\boldsymbol{y}}$, and $\bar{\boldsymbol{a}}_1, \ldots, \bar{\boldsymbol{a}}_n$ of the expected time taken by a two-phase shadow vertex simplex method to solve linear programming problems of the form

$$\begin{aligned} &\text{maximize} && \boldsymbol{z}^T \boldsymbol{x} \\ &\text{subject to} && \langle \boldsymbol{a}_i | \boldsymbol{x} \rangle \leq y_i, \text{ for } 1 \leq i \leq n, \end{aligned} \tag{4}$$

where each $\boldsymbol{a}_i$ is a Gaussian random vector of standard deviation $\sigma \max_i \|(\bar{y}_i, \bar{\boldsymbol{a}}_i)\|$ centered at $\bar{\boldsymbol{a}}_i$, and each y_i is a Gaussian random variable of standard deviation $\sigma \max_i \|(\bar{y}_i, \bar{\boldsymbol{a}}_i)\|$ centered at $\bar{y}_i$.

We begin by considering the case in which $\boldsymbol{y} = \mathbf{1}$, $\|\bar{\boldsymbol{a}}_i\| \leq 1$, and $\sigma < 1/3\sqrt{d \ln n}$. In this case, our first result, Theorem 4.0.1, says that for every vector $\boldsymbol{t}$ the expected size of the *shadow* of the polytope—the projection of the polytope defined by the equations (4) onto the plane spanned by $\boldsymbol{t}$ and $\boldsymbol{z}$—is polynomial in n, the dimension, and $1/\sigma$. This result is the geometric foundation of our work, but it does not directly bound the running time of an algorithm, as the shadow relevant to the analysis of an algorithm depends on the perturbed program and cannot be specified beforehand as the vector $\boldsymbol{t}$ must be. In Section 3.3, we describe a two-phase shadow-vertex simplex algorithm, and in Section 5 we use Theorem 4.0.1 as a black box to show that it takes expected time polynomial in n, d, and $1/\sigma$ in the case described above.

Efforts have been made to analyze how much the solution of a linear program can change as its data is perturbed. For an introduction to such analyses, and an analysis of the complexity of interior point methods in terms of the resulting condition number, we refer the reader to the work of Renegar [Ren95b, Ren95a, Ren94].

1.4 Intuition Through Condition Numbers

For those already familiar with the simplex method and condition numbers, we include this section to provide some intuition for why our results should be true.

Our analysis will exploit geometric properties of the condition number of a matrix, rather than of a linear program. We start with the observation that if a corner of a polytope is specified by the equation $A_I \boldsymbol{x} = \boldsymbol{y}_I$, where I is a d-set, then the condition number of the matrix A_I provides a good measure of how far the corner is from being flat. Moreover, it is relatively easy to show that if A is subject to perturbation, then it is unlikely that A_I has poor condition number. So, it seems intuitive that if A is perturbed, then most corners of the polytope should have angles bounded away from being flat. This already provides some intuition as to why the simplex method should run quickly: one should make reasonable progress as one rounds a corner if it is not too flat.

There are two difficulties in making the above intuition rigorous: the first is that even if A_I is well-conditioned for most sets I, it is not clear that A_I will be well-conditioned for most sets I that are bases of corners of the polytope. The second difficulty is that even if most corners of the polytope have reasonable condition number, it is not clear that a simplex method will actually encounter many of these corners. By analyzing the shadow vertex pivot rule, it is possible to resolve both of these difficulties.

The first advantage of studying the shadow vertex pivot rule is that its analysis comes down to studying the expected sizes of shadows of the polytope. From the specification of

the plane onto which the polytope will be projected, one obtains a characterization of all the corners that will be in the shadow, thereby avoiding the complication of an iterative characterization. The second advantage is that these corners are specified by the property that they optimize a particular objective function, and using this property one can actually bound the probability that they are ill-conditioned. While the results of Section 4 are not stated in these terms, this is the intuition behind them.

Condition numbers also play a fundamental role in our analysis of the shadow-vertex algorithm. The analysis of the algorithm differs from the mere analysis of the sizes of shadows in that, in the study of an algorithm, the plane onto which the polytope is projected depends upon the polytope itself. This correlation of the plane with the polytope complicates the analysis, but is also resolved through the help of condition numbers. In our analysis, we view the perturbation as the composition of two perturbations, where the second is small relative to the first. We show that our choice of the plane onto which we project the shadow is well-conditioned with high probability after the first perturbation. That is, we show that the second perturbation is unlikely to substantially change the plane onto which we project, and therefore unlikely to substantially change the shadow. Thus, it suffices to measure the expected size of the shadow obtained after the second perturbation onto the plane that would have been chosen after just the first perturbation.

The technical lemma that enables this analysis, Lemma 5.1.1, is a concentration result that proves that it is highly unlikely that almost all of the minors of a random matrix have poor condition number. This analysis also enables us to show that it is highly unlikely that we will need a large "big-M" in phase I of our algorithm.

We note that the condition numbers of the A_Is have been studied before in the complexity of linear programming algorithms. The condition number $\bar{\chi}_A$ of Vavasis and Ye [VY96] measures the condition number of the worst sub-matrix A_I, and their algorithm runs in time proportional to $\ln(\bar{\chi}_A)$. Todd, Tunçel, and Ye [TTY01] have shown that for a Gaussian random matrix the expectation of $\ln(\bar{\chi}_A)$ is $O(\min(d \ln n, n))$. That is, they show that it is unlikely that any A_I is exponentially ill-conditioned. It is relatively simple to apply the techniques of Section 5.1 to obtain a similar result in the smoothed case. We wonder whether our concentration result that it is exponentially unlikely that many A_I are even polynomially ill-conditioned could be used to obtain a better smoothed analysis of the Vavasis-Ye algorithm.

1.5 Discussion

One can debate whether the definition of *polynomial smoothed complexity* should be that an algorithm have complexity polynomial in $1/\sigma$ or $\log(1/\sigma)$. We believe that the choice of being polynomial in $1/\sigma$ will prove more useful as the other definition is too strong and quite similar to the notion of being polynomial in the worst case. In particular, one can convert any algorithm for linear programming whose smoothed complexity is polynomial in d, n and $\log(1/\sigma)$ into an algorithm whose worst-case complexity is polynomial in d, n, and L. That said, one should certainly prefer complexity bounds that are lower as a function of $1/\sigma$, d and n.

We also remark that a simple examination of the constructions that provide exponential lower bounds for various pivot rules [KM72, Mur80, GS79, Gol83, AC78, Jer73] reveals that

none of these pivot rules have smoothed complexity polynomial in n and sub-polynomial in $1/\sigma$. That is, these constructions are unaffected by exponentially small perturbations.

2 Notation and Mathematical Preliminaries

In this section, we define the notation that will be used in the paper. We will also review some background from mathematics and derive a few simple statements that we will need. The reader should probably skim this section now, and save a more detailed examination for when the relevant material is referenced.

- $[n]$ denotes the set of integers between 1 and n, and $\binom{[n]}{k}$ denotes the subsets of $[n]$ of size k.
- Subsets of $[n]$ are denoted by the capital Roman letters I, J, L, K. $\mathcal{M}$ will denote a subset of integers, and $\mathcal{K}$ will denote a set of subsets of $[n]$.
- Subsets of $\mathbb{R}^?$ are denoted by the capital Roman letters $A, B, P, Q, R, S, T, U, V$.
- Vectors in $\mathbb{R}^?$ are denoted by bold lower-case Roman letters, such as $\boldsymbol{a}_i, \bar{\boldsymbol{a}}_i, \tilde{\boldsymbol{a}}_i$, $\boldsymbol{b}_i, \boldsymbol{c}_i$, $\boldsymbol{d}_i, \boldsymbol{h}$, $\boldsymbol{t}, \boldsymbol{q}, \boldsymbol{z}, \boldsymbol{y}$.
- Whenever a vector, say $\boldsymbol{a} \in \mathbb{R}^d$ is present, its components will be denoted by lower-case Roman letters with subscripts, such as $a_1, \ldots, a_d$.
- Whenever a collection of vectors, such as $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, are present, the similar bold upper-case letter, such as $\boldsymbol{A}$, will denote the matrix of these vectors. For $I \in \binom{[n]}{k}$, $\boldsymbol{A}_I$ will denote the matrix of those $\boldsymbol{a}_i$ for which $i \in I$.
- Matrices are denoted by bold upper-case Roman letters, such as $\boldsymbol{A}, \bar{\boldsymbol{A}}, \tilde{\boldsymbol{A}}, \boldsymbol{B}, \boldsymbol{M}$ and $\boldsymbol{R}_{\boldsymbol{\omega}}$.
- S^{d-1} denotes the unit sphere in $\mathbb{R}^d$.
- Vectors in $S^?$ will be denoted by bold Greek letters, such as $\boldsymbol{\omega}, \boldsymbol{\psi}, \boldsymbol{\tau}$.
- Generally speaking, univariate quantities with scale, such as lengths or heights, will be represented by lower case Roman letters such as c, h, l, r, s, and t. The principal exceptions are that κ and M will also denote such quantities.
- Quantities without scale, such as the ratios of quantities with scale or affine coordinates, will be represented by lower case Greek letters such as $\alpha, \beta, \lambda, \xi, \zeta$. $\boldsymbol{\alpha}$ will denote a vector of such quantities such as $(\alpha_1, \ldots, \alpha_d)$.
- Density functions are denoted by lower case Greek letters such as μ and ν.
- The standard deviations of Gaussian random variables are denoted by lower-case Greek letters such as σ, τ and ρ.
- Indicator random variables are denoted by upper case Roman letters, such as A, B, E, F, V, W, X, Y, and Z
- Functions into the reals or integers will be denoted by calligraphic upper-case letters, such as $\mathcal{F}, \mathcal{G}, \mathcal{S}^+, \mathcal{S}', \mathcal{T}$.

- Functions into $\mathbb{R}^?$ are denoted by upper-case Greek letters, such as $\Phi_\epsilon, \Upsilon, \Psi$.
- $\langle \boldsymbol{x} | \boldsymbol{y} \rangle$ denotes the inner product of vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.
- For vectors $\boldsymbol{\omega}$ and $\boldsymbol{z}$, we let $\mathbf{angle}(\boldsymbol{\omega}, \boldsymbol{z})$ denote the angle between these vectors at the origin.
- The logarithm base 2 is written lg and the natural logarithm is written ln.
- The probability of an event A is written $\mathbf{Pr}[A]$, and the expectation of a variable X is written $\mathbf{E}[X]$.
- The indicator random variable for an event A is written $[A]$.

2.1 Geometric Definitions

For the following definitions, we let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ denote a set of vectors in $\mathbb{R}^d$.

- $\mathbf{Span}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)$ denotes the subspace spanned by $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$.
- $\mathbf{Aff}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)$ denotes the hyperplane that is the affine span of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$: the set of points $\sum_i \alpha_i \boldsymbol{a}_i$, where $\sum_i \alpha_i = 1$, for all i.
- $\mathbf{ConvHull}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)$ denotes the convex hull of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$.
- $\mathbf{Cone}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)$ denotes the positive cone through $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$: the set of points $\sum_i \alpha_i \boldsymbol{a}_i$, for $\alpha_i \geq 0$.
- $\triangle(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d)$ denotes the simplex $\mathbf{ConvHull}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d)$.

For a linear program specified by $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, $\boldsymbol{y}$ and $\boldsymbol{z}$, we will say that the linear program is in *general position* if

- The points $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ are in general position with respect to $\boldsymbol{y}$, which means that for all $I \subset \binom{[n]}{d}$ and $\boldsymbol{x} = \boldsymbol{A}_I^{-1} \boldsymbol{y}_I$, and all $j \notin I$, $\langle \boldsymbol{a}_j | \boldsymbol{x} \rangle \neq y_j$.
- For all $I \subset \binom{[n]}{d-1}$, $\boldsymbol{z} \notin \mathbf{Cone}(\boldsymbol{A}_I)$.

Furthermore, we will say that the linear program is in *general position with respect to a vector* $\boldsymbol{t}$ if the set of λ for which there exists an $I \in \binom{[n]}{d-1}$ such that

$$(1 - \lambda)\boldsymbol{t} + \lambda \boldsymbol{z} \in \mathbf{Cone}(\boldsymbol{A}_I)$$

is finite and does not contain 0.

2.2 Vector and Matrix Norms

The material of this section is principally used in Sections 3.3 and 5.1. The following definitions and propositions are standard, and may be found in standard texts on Numerical Linear Algebra.

Definition 2.2.1 (Vector Norms) *For a vector $\boldsymbol{x}$, we define*

- $\|\boldsymbol{x}\| = \sqrt{\sum_i x_i^2}$.
- $\|\boldsymbol{x}\|_1 = \sum_i |x_i|$.
- $\|\boldsymbol{x}\|_\infty = \max_i |x_i|$.

Proposition 2.2.2 (Vectors norms) *For a vector $\boldsymbol{x} \in \mathbb{R}^d$,*

$$\|\boldsymbol{x}\| \leq \|\boldsymbol{x}\|_1 \leq \sqrt{d}\,\|\boldsymbol{x}\|\,.$$

Definition 2.2.3 (Matrix norm) *For a matrix $\boldsymbol{A}$, we define*

$$\|\boldsymbol{A}\| \stackrel{\text{def}}{=} \max_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x}\| \,/\, \|x\|\,.$$

Proposition 2.2.4 (Properties of matrix norm) *For d-by-d matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, and a d-vector $\boldsymbol{x}$,*

(a) $\|\boldsymbol{A}\boldsymbol{x}\| \leq \|\boldsymbol{A}\|\,\|\boldsymbol{x}\|$.

(b) $\|\boldsymbol{A}\boldsymbol{B}\| \leq \|\boldsymbol{A}\|\,\|\boldsymbol{B}\|$.

(c) $\|\boldsymbol{A}\| = \left\|\boldsymbol{A}^T\right\|$.

(d) $\|\boldsymbol{A}\| \leq \sqrt{d}\max_i \|\boldsymbol{a}_i\|$, *where* $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d)$.

(e) $\mathbf{det}\,(\boldsymbol{A}) \leq \|\boldsymbol{A}\|^d$.

Definition 2.2.5 ($\mathbf{s_{min}}\,()$) *For a matrix $\boldsymbol{A}$, we define*

$$\mathbf{s}_{\boldsymbol{min}}\,(\boldsymbol{A}) \stackrel{\text{def}}{=} \left\|\boldsymbol{A}^{-1}\right\|^{-1}.$$

We recall that $\mathbf{s_{min}}\,(\boldsymbol{A})$ is the smallest singular value of the matrix $\boldsymbol{A}$, and that it is not a norm.

Proposition 2.2.6 (Properties of $\mathbf{s_{min}}\,()$) *For d-by-d matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,*

(a) $\mathbf{s}_{\boldsymbol{min}}\,(\boldsymbol{A}) = \min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x}\| \,/\, \|\boldsymbol{x}\|$.

(b) $\mathbf{s}_{\boldsymbol{min}}\,(\boldsymbol{B}) \geq \mathbf{s}_{\boldsymbol{min}}\,(\boldsymbol{A}) - \|\boldsymbol{A} - \boldsymbol{B}\|$.

2.3 Probability

For an event, A, we let $[A]$ denote the indicator random variable for the event. We generally describe random variables by their density functions. If $\boldsymbol{x}$ has density μ, then

$$\mathbf{Pr}\left[A(\boldsymbol{x})\right] \stackrel{\text{def}}{=} \int \left[A(\boldsymbol{x})\right] \mu(\boldsymbol{x})\, d\boldsymbol{x}\,.$$

If B is another event, then

$$\Pr_{B}\left[A(\boldsymbol{x})\right] \stackrel{\text{def}}{=} \mathbf{Pr}\left[A(\boldsymbol{x})\middle|B(\boldsymbol{x})\right] \stackrel{\text{def}}{=} \frac{\int \left[B(\boldsymbol{x})\right]\left[A(\boldsymbol{x})\right]\mu(\boldsymbol{x})\, d\boldsymbol{x}}{\int \left[B(\boldsymbol{x})\right]\mu(\boldsymbol{x})\, d\boldsymbol{x}}.$$

In a context where multiple densities are present, we will use use the notation $\mathbf{Pr}_{\mu}\left[A(\boldsymbol{x})\right]$ to indicate the probability of A when $\boldsymbol{x}$ is distributed according to μ.

In many situations, we will not know the density μ of a random variable $\boldsymbol{x}$, but rather a function ν such that $\nu(\boldsymbol{x}) = c\mu(\boldsymbol{x})$ for some constant c. In this case, we will say that $\boldsymbol{x}$ *has density proportional to* ν.

The following Propositions and Lemmas will play a prominent role in the proofs in this paper. The only one of these which might not be intuitively obvious is Lemma 2.3.5.

Proposition 2.3.1 (Average $\leq$ maximum) *Let $\mu(x,y)$ be a density function, and let x and y be distributed according to $\mu(x,y)$. If $A(x,y)$ is an event and $X(x,y)$ is random variable, then*

$$\begin{aligned}\underset{x,y}{\boldsymbol{Pr}}\left[A(x,y)\right] &\leq \max_{x} \underset{y}{\boldsymbol{Pr}}\left[A(x,y)\right], \text{ and}\\ \underset{x,y}{\boldsymbol{E}}\left[X(x,y)\right] &\leq \max_{x} \underset{y}{\boldsymbol{E}}\left[X(x,y)\right],\end{aligned}$$

where in the right-hand terms, y is distributed according to the induced distribution $\mu(x,y)$.

Proposition 2.3.2 (Expectation on sub-domain) *Let $\boldsymbol{x}$ be a random variable and $A(\boldsymbol{x})$ an event. Let P be a measurable subset of the domain of $\boldsymbol{x}$. Then,*

$$\underset{\boldsymbol{x}\in P}{\boldsymbol{Pr}}\left[A(\boldsymbol{x})\right] \leq \boldsymbol{Pr}\left[A(\boldsymbol{x})\right] / \boldsymbol{Pr}\left[\boldsymbol{x}\in P\right].$$

Proof By the definition of conditional probability,

$$\begin{aligned}\Pr_{\boldsymbol{x}\in P}\left[A(\boldsymbol{x})\right] &= \mathbf{Pr}\left[A(\boldsymbol{x})|\boldsymbol{x}\in P\right]\\ &= \mathbf{Pr}\left[A(\boldsymbol{x}) \text{ and } \boldsymbol{x}\in P\right] / \mathbf{Pr}\left[\boldsymbol{x}\in P\right], && \text{by Bayes' rule,}\\ &\leq \mathbf{Pr}\left[A(\boldsymbol{x})\right] / \mathbf{Pr}\left[\boldsymbol{x}\in P\right].\end{aligned}$$

■

Lemma 2.3.3 (Comparing expectations) *Let X and Y be non-negative random variables and A an event satisfying (1) $X \leq k$, (2) $\boldsymbol{Pr}\left[A\right] \geq 1-\epsilon$, and (3) there exists a constant c such that $\boldsymbol{E}\left[X|A\right] \leq c\,\boldsymbol{E}\left[Y|A\right]$. Then,*

$$\boldsymbol{E}\left[X\right] \leq c\,\boldsymbol{E}\left[Y\right] + \epsilon k.$$

Proof

$$\begin{aligned}\mathbf{E}[X] &= \mathbf{E}[X|A]\,\mathbf{Pr}[A] + \mathbf{E}[X|\mathbf{not}(A)]\,\mathbf{Pr}[\mathbf{not}(A)] \\ &\leq c\,\mathbf{E}[Y|A]\,\mathbf{Pr}[A] + \epsilon k \\ &\leq c\,\mathbf{E}[Y] + \epsilon k,\end{aligned}$$

by Proposition 2.3.2. ■

Lemma 2.3.4 (Similar distributions) *Let X be a non-negative random variable such that $X \leq k$. Let ν and μ be density functions for which there exists a set S such that (1) $\boldsymbol{Pr}_{\nu}[S] > 1 - \epsilon$ and (2) there exists a constant $c \geq 1$ such that for all $a \in S$, $\nu(a) \leq c\mu(a)$. Then,*

$$\underset{\nu}{\boldsymbol{E}}[X(a)] \leq c\,\underset{\mu}{\boldsymbol{E}}[X(a)] + k\epsilon.$$

Proof We write

$$\begin{aligned}\underset{\nu}{\mathbf{E}}[X] &= \int_{a\in S} X(a)\nu(a)\,da + \int_{a\notin S} X(a)\nu(a)\,da \\ &\leq c\int_{a\in S} X(a)\mu(a)\,da + k\epsilon \\ &\leq c\int_{a} X(a)\mu(a)\,da + k\epsilon \\ &= c\,\underset{\mu}{\mathbf{E}}[X] + k\epsilon.\end{aligned}$$

■

Lemma 2.3.5 (Combination lemma) *Let x and y be random variables distributed according to $\mu(x, y)$. Let $\mathcal{F}(x)$ and $\mathcal{G}(x, y)$ be non-negative functions and α and β be constants such that*

- $\forall \epsilon \geq 0$, $\boldsymbol{Pr}_{x,y}[\mathcal{F}(x) \leq \epsilon] \leq \alpha\epsilon$, *and*
- $\forall \epsilon \geq 0$, $\max_x \boldsymbol{Pr}_y[\mathcal{G}(x, y) \leq \epsilon] \leq (\beta\epsilon)^2$,

where in the second line y is distributed according to the induced density $\mu(x, y)$. Then

$$\underset{x,y}{\boldsymbol{Pr}}[\mathcal{F}(x)\mathcal{G}(x, y) \leq \epsilon] \leq 4\alpha\beta\epsilon.$$

Proof Consider any x and y for which $\mathcal{F}(x)\mathcal{G}(x, y) \leq \epsilon$. If i is the integer for which

$$2^i\beta\epsilon < \mathcal{F}(x) \leq 2^{i+1}\beta\epsilon,$$

then $\mathcal{G}(x, y) \leq 2^{-i}/\beta$. Thus, $\mathcal{F}(x)\mathcal{G}(x, y) \leq \epsilon$, implies that either $\mathcal{F}(x) \leq 2\beta\epsilon$, or there exists an integer $i \geq 1$ for which

$$\mathcal{F}(x) \leq 2^{i+1}\beta\epsilon \qquad \text{and} \qquad \mathcal{G}(x, y) \leq 2^{-i}/\beta.$$

So, we obtain the bound

$$\begin{aligned}
\mathop{\mathbf{Pr}}_{x,y}\left[\mathcal{F}(x)\mathcal{G}(x,y) \le \epsilon\right] &\le \mathop{\mathbf{Pr}}_{x,y}\left[\mathcal{F}(x) \le 2\beta\epsilon\right] + \sum_{i\ge 1} \mathop{\mathbf{Pr}}_{x,y}\left[\mathcal{F}(x) \le 2^{i+1}\beta\epsilon \text{ and } \mathcal{G}(x,y) \le 2^{-i}/\beta\right] \\
&\le 2\alpha\beta\epsilon + \sum_{i\ge 1} \mathop{\mathbf{Pr}}_{x,y}\left[\mathcal{F}(x) \le 2^{i+1}\beta\epsilon\right] \mathop{\mathbf{Pr}}_{x,y}\left[\mathcal{G}(x,y) \le 2^{-i}/\beta \middle| \mathcal{F}(x) \le 2^{i+1}\beta\epsilon\right] \\
&\le 2\alpha\beta\epsilon + \sum_{i\ge 1} \mathop{\mathbf{Pr}}_{x,y}\left[\mathcal{F}(x) \le 2^{i+1}\beta\epsilon\right] \max_x \mathop{\mathbf{Pr}}_{y}\left[\mathcal{G}(x,y) \le 2^{-i}/\beta\right] \\
&\le 2\alpha\beta\epsilon + \sum_{i\ge 1} \left(2^{i+1}\alpha\beta\epsilon\right)\left(2^{-i}\right)^2, \text{ by Proposition 2.3.1,} \\
&= 2\alpha\beta\epsilon + \alpha\beta\epsilon \sum_{i\ge 1} 2^{1-i} \\
&= 4\alpha\beta\epsilon.
\end{aligned}$$

■

As we have found this lemma very useful in our work, and we suspect others may as well, we state a more broadly applicable generalization. It's proof is similar.

Lemma 2.3.6 (Generalized combination lemma) *Let x and y be random variables distributed according to $\mu(x,y)$. There exists a function $c(a,b)$ such that if $\mathcal{F}(x)$ and $\mathcal{G}(x,y)$ are non-negative functions and α, β, a and b are constants such that*

- $\boldsymbol{Pr}_{x,y}\left[\mathcal{F}(x) \le \epsilon\right] \le (\alpha\epsilon)^a$, *and*
- $\max_x \boldsymbol{Pr}_y\left[\mathcal{G}(x,y) \le \epsilon\right] \le (\beta\epsilon)^b$,

where in the second line y is distributed according to the induced density $\mu(x,y)$, then

$$\mathop{\boldsymbol{Pr}}_{x,y}\left[\mathcal{F}(x)\mathcal{G}(x,y) \le \epsilon\right] \le c(a,b)\alpha\beta\epsilon^{\min(a,b)} \lg(1/\epsilon)^{[a=b]},$$

where $[a=b]$ is 1 if $a=b$ and 0 otherwise.

Lemma 2.3.7 (Almost polynomial densities) *Let $k > 0$ and let t be a non-negative random variable with density proportional to $\mu(t)t^k$ such that, for some $t_0 > 0$,*

$$\frac{\max_{0\le t\le t_0} \mu(t)}{\min_{0\le t\le t_0} \mu(t)} \le c.$$

Then,

$$\boldsymbol{Pr}[t < \epsilon] < c(\epsilon/t_0)^{k+1}.$$

Proof For $\epsilon \geq t_0$, the lemma is vacuously true. Assuming $\epsilon < t_0$,

$$\begin{aligned}
\mathbf{Pr}\left[t < \epsilon\right] &\leq \frac{\mathbf{Pr}\left[t < \epsilon\right]}{\mathbf{Pr}\left[t < t_0\right]} \\
&= \frac{\int_{t=0}^{\epsilon} \mu(t) t^k \, dt}{\int_{t=0}^{t_0} \mu(t) t^k \, dt} \\
&\leq \frac{\max_{0 \leq t \leq t_0} \mu(t) \int_{t=0}^{\epsilon} t^k \, dt}{\min_{0 \leq t \leq t_0} \mu(t) \int_{t=0}^{t_0} t^k \, dt} \\
&\leq c \frac{\epsilon^{k+1}/(k+1)}{t_0^{k+1}/(k+1)} \\
&= c(\epsilon/t_0)^{k+1}.
\end{aligned}$$

■

2.4 Gaussian Random Vectors

For the convenience of the reader, we recall some standard facts about Gaussian random variables and vectors. These may be found in [Fel68, VII.1] and [Fel71, III.6]. We then draw some corollaries of these facts and derive some lemmas that we will need later in the paper.

We first recall that a univariate Gaussian distribution with mean 0 and standard deviation σ has density

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-a^2/2\sigma^2},$$

and that a Gaussian random vector in $\mathbb{R}^d$ centered at a point $\bar{\boldsymbol{a}}$ with covariance matrix $\boldsymbol{M}$ has density

$$\frac{1}{\left(\sqrt{2\pi}\right)^d \det(\boldsymbol{M})} e^{-(\boldsymbol{a}-\bar{\boldsymbol{a}})^T \boldsymbol{M}^{-1} (\boldsymbol{a}-\bar{\boldsymbol{a}})/2}.$$

For positive-definite $\boldsymbol{M}$, there exists a basis in which the density can be written

$$\prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-a_i^2/2\sigma_i^2},$$

where $\sigma_1^2 \leq \cdots \leq \sigma_d^2$ are the eigenvalues of $\boldsymbol{M}$. When all the eigenvalues of $\boldsymbol{M}$ are the same and equal to σ, then we will refer to the density as a *Gaussian distribution of standard deviation* σ.

Proposition 2.4.1 (Additivity of Gaussians) *If $\boldsymbol{a}_1$ is a Gaussian random vector with covariance matrix $\boldsymbol{M}_1$ centered at a point $\bar{\boldsymbol{a}}_1$ and $\boldsymbol{a}_2$ is a Gaussian random vector with covariance matrix $\boldsymbol{M}_2$ centered at a point $\bar{\boldsymbol{a}}_2$, then $\boldsymbol{a}_1 + \boldsymbol{a}_2$ is the Gaussian random vector with covariance matrix $\boldsymbol{M}_1 + \boldsymbol{M}_2$ centered at $\bar{\boldsymbol{a}}_1 + \bar{\boldsymbol{a}}_2$.*

Lemma 2.4.2 (Smoothness of Gaussians) *Let $\mu(\boldsymbol{x})$ be a Gaussian distribution of standard deviation σ centered at a point $\bar{\boldsymbol{a}}$. Let $k \geq 1$, let $\mathbf{dist}(\boldsymbol{x}, \bar{\boldsymbol{a}}) \leq k$ and let $\mathbf{dist}(\boldsymbol{x}, \boldsymbol{y}) < \epsilon \leq k$. Then,*

$$\frac{\mu(\boldsymbol{y})}{\mu(\boldsymbol{x})} \geq e^{-3k\epsilon/2\sigma^2}.$$

Proof By translating $\bar{\boldsymbol{a}}$, $\boldsymbol{x}$ and $\boldsymbol{y}$, we may assume $\bar{\boldsymbol{a}} = 0$ and $\|\boldsymbol{x}\| \leq k$. We then have

$$\begin{aligned}
\frac{\mu(\boldsymbol{y})}{\mu(\boldsymbol{x})} &= e^{-(\|\boldsymbol{y}\|^2 - \|\boldsymbol{x}\|^2)/2\sigma^2} \\
&\geq e^{-(2\epsilon\|\boldsymbol{x}\| + \epsilon^2)/2\sigma^2}, && \text{as } \|\boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \epsilon \\
&\geq e^{-(2\epsilon k + \epsilon^2)/2\sigma^2}, && \text{as } \|\boldsymbol{x}\| \leq k \\
&\geq e^{-3\epsilon k/2\sigma^2} && \text{as } \epsilon \leq k.
\end{aligned}$$

■

Proposition 2.4.3 (Restrictions of Gaussians) *Let μ be a Gaussian distribution of standard deviation σ centered at a point $\bar{\boldsymbol{a}}$. Let $\boldsymbol{v}$ be any vector and r any real. Then, the induced distribution*

$$\mu(\boldsymbol{x} | \boldsymbol{v}^T \boldsymbol{x} = r)$$

is a Gaussian distribution of standard deviation σ centered at the projection of $\bar{\boldsymbol{a}}$ onto the plane $\{\boldsymbol{x} : \boldsymbol{v}^T \boldsymbol{x} = r\}$.

Proposition 2.4.4 (Gaussian measure of halfspaces) *Let $\boldsymbol{\omega}$ be any unit vector in $\mathbb{R}^d$ and r any real. Then,*

$$\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^d \int_{\boldsymbol{g}} [\langle \boldsymbol{\omega} | \boldsymbol{g} \rangle \leq r] \, e^{-\|\boldsymbol{g}\|^2/2\sigma^2} \, d\boldsymbol{g} = \frac{1}{\sqrt{2\pi}\sigma} \int_{t=-\infty}^{t=r} e^{-t^2/2\sigma^2} \, dt$$

Proof Immediate if one expresses the Gaussian density in a basis containing $\boldsymbol{\omega}$. ■

The distribution of the square of the norm of a Gaussian random vector is the Chi-Square distribution. We use the following weak bound on the Chi-Square distribution, which follows from Equality (26.4.8) of [AS70].

Proposition 2.4.5 (Chi-Square bound) *Let $\boldsymbol{x}$ be a Gaussian random vector in $\mathbb{R}^d$ of standard deviation σ centered at the origin. Then,*

$$\boldsymbol{Pr}[\|\boldsymbol{x}\| \geq k\sigma] \leq \frac{\left(k^2\right)^{d/2-1} e^{-k^2/2}}{2^{d/2-1}\Gamma(\frac{d}{2})}. \tag{5}$$

From this, we derive

Corollary 2.4.6 (A chi-square bound) *Let $\boldsymbol{x}$ be a Gaussian random vector in $\mathbb{R}^d$ of standard deviation σ centered at the origin. Then, for $n \geq 3$*

$$\boldsymbol{Pr}\left[\|\boldsymbol{x}\| \geq 3\sqrt{d \ln n}\sigma\right] \leq n^{-2.9d}.$$

Moreover, if $n > d \geq 3$, and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are such vectors, then

$$\boldsymbol{Pr}\left[\max_i \|\boldsymbol{x}_i\| \geq 3\sqrt{d \ln n}\sigma\right] \leq n^{-2.9d+1} \leq 0.0015\binom{n}{d}^{-1}.$$

Proof For $\alpha = 3\sqrt{\ln n}\sigma$ we can apply Stirling's formula [AS70] to (5) to find

$$\begin{aligned}
\mathbf{Pr}\left[\|\boldsymbol{x}\| \geq \alpha\sqrt{d}\right] &\leq \frac{(\alpha^2 d)^{d/2-1} e^{-\alpha^2 d/2} e^{d/2}\sqrt{d/2}}{2^{d/2-1}(d/2)^{d/2}\sqrt{2\pi}} \\
&= \left(\alpha^2\right)^{d/2-1} e^{-(\alpha^2-1)d/2} \frac{d^{d/2-1}\sqrt{d}}{2^{d/2-1}(d/2)^{d/2} 2\sqrt{\pi}} \\
&= \left(\alpha^2\right)^{d/2-1} e^{-(\alpha^2-1)d/2} \frac{1}{\sqrt{d\pi}} \\
&\leq \left(\alpha^2\right)^{d/2} e^{-(\alpha^2-1)d/2} \\
&= e^{-(\alpha^2 - \ln(\alpha^2) - 1)d/2} \\
&\leq e^{-2.9d \ln n} \\
&= n^{-2.9d},
\end{aligned}$$

as

$$(\alpha^2 - \ln(\alpha^2) - 1) = 9\ln(n) - \ln(9\ln n) - 1 \geq \ln(n)(9 - \ln 9 - 1) \geq 5.8\ln(n).$$

■

We also prove it is unlikely that a Gaussian random variable has small norm.

Proposition 2.4.7 (Gaussian near point or plane) *Let $\boldsymbol{x}$ be a d-dimensional Gaussian random vector of standard deviation σ centered anywhere. Then,*

(a) For any point $\boldsymbol{p}$, $\boldsymbol{Pr}\left[\mathbf{dist}\left(\boldsymbol{x}, \boldsymbol{p}\right) \leq \epsilon\right] \leq \left(\min\left(1, \sqrt{e/d}\right)(\epsilon/\sigma)\right)^d$, and

(b) For a plane H of dimension h, $\boldsymbol{Pr}\left[\mathbf{dist}\left(\boldsymbol{x}, H\right) \leq \epsilon\right] \leq (\epsilon/\sigma)^{d-h}$.

Proof Let $\bar{\boldsymbol{x}}$ be the center of the Gaussian distribution, and let $B_\epsilon(\boldsymbol{p})$ denote the ball of radius ϵ around $\boldsymbol{p}$. Recall that the volume of $B_\epsilon(\boldsymbol{p})$ is

$$\frac{2\pi^{d/2}\epsilon^d}{d\Gamma(d/2)}.$$

To prove part (a), we bound the probability that $\mathbf{dist}\left(\boldsymbol{x}, \boldsymbol{p}\right) \leq \epsilon$ by

$$\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^d \int_{\boldsymbol{x} \in B_\epsilon(\boldsymbol{p})} e^{-\|(\boldsymbol{x}-\bar{\boldsymbol{x}})\|^2/2\sigma^2}\, d\boldsymbol{x} \leq \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^d \left(\frac{2\pi^{d/2}\epsilon^d}{d\Gamma(d/2)}\right) = \left(\frac{\epsilon}{\sigma}\right)^d \frac{2}{d2^{d/2}\Gamma(d/2)}.$$

By Proposition 2.4.8, we have for $d \geq 3$

$$\frac{2}{d2^{d/2}\Gamma(d/2)} \leq (e/d)^{d/2}.$$

Combining with the fact that $2/(d2^{d/2}\Gamma(d/2)) \leq 1$ for all $d \geq 1$, we establish (a).

To prove part (b), we consider a basis in which $d-h$ vectors are perpendicular to H, and apply part (a) to the components of $\boldsymbol{x}$ in the span of those basis vectors. ■

Proposition 2.4.8 (Gamma Inequality) *For $d \geq 3$*

$$\frac{2}{d2^{d/2}\Gamma(d/2)} \quad \leq \quad (e/d)^{d/2}$$

Proof For $d \geq 3$, we apply the inequality $\Gamma(x+1) \geq \sqrt{2\pi}\sqrt{x}(x/e)^x$ to show

$$\begin{aligned}
\frac{2}{d2^{d/2}\Gamma(d/2)} &\leq \frac{2}{d2^{d/2}\sqrt{2\pi}\sqrt{(d-2)/2}}\left(\frac{2e}{d-2}\right)^{(d-2)/2} \\
&= \left(\frac{e^{(d-2)/2}}{d^{d/2}\sqrt{2\pi}\sqrt{(d-2)/2}}\right)\left(\frac{d}{d-2}\right)^{(d-2)/2} \\
&\leq (e/d)^{d/2},
\end{aligned}$$

where the last inequality used the inequalities $1+2/(d-2) \leq e^{2/(d-2)}$ and $\sqrt{2\pi}\sqrt{(d-1)/2} > 1$ when $d \geq 3$. ■

Proposition 2.4.9 (Non-central Gaussian near the origin) *For any integer $d \geq 3$, let $\boldsymbol{x}$ be a d-dimensional Gaussian random vector of standard deviation σ centered at $\bar{\boldsymbol{x}}$. Then, for $\epsilon \leq 1/(\sqrt{2}e)$*

$$\boldsymbol{Pr}\left[\|\boldsymbol{x}\| \leq \left(\sqrt{\|\bar{\boldsymbol{x}}\|^2 + d\sigma^2}\right)\epsilon\right] \leq \left(\sqrt{2e}\epsilon\right)^d$$

.

Proof Let $\lambda = \|\bar{\boldsymbol{x}}\|$. We divide the analysis into two cases: (1) $\lambda \leq \sqrt{d}\sigma$, and (2) $\lambda \geq \sqrt{d}\sigma$.

For $\lambda \leq \sqrt{d}\sigma$,

$$\mathbf{Pr}\left[\|\boldsymbol{x}\| \leq (\sqrt{\lambda^2 + d\sigma^2})\epsilon\right] \quad \leq \quad \mathbf{Pr}\left[\|\boldsymbol{x}\| \leq (\sqrt{2d}\sigma)\epsilon\right] \leq (\sqrt{2e}\epsilon)^d,$$

by Part (a) of Lemma 2.4.7.

For $\lambda > \sqrt{d}\sigma$, let B_r be the ball of radius r around the origin. Applying the assumption $\epsilon \leq 1/(\sqrt{2}e)$ and letting $\lambda = c\sqrt{d}\sigma$ for $c \geq 1$, we have

$$\begin{aligned}
\mathbf{Pr}\left[\|\boldsymbol{x}\| \leq (\sqrt{\lambda^2 + d\sigma^2})\epsilon\right] &\leq \mathbf{Pr}\left[\|\boldsymbol{x}\| \leq (\sqrt{2}\lambda)\epsilon\right] \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^d \int_{\boldsymbol{x} \in B_{\sqrt{2}\epsilon\lambda}} e^{-\|(\boldsymbol{x}-\bar{\boldsymbol{x}})\|^2/2\sigma^2}\, d\boldsymbol{x} \\
&\leq \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^d \left(\frac{2\pi^{d/2}}{d\Gamma(d/2)}\right) (\sqrt{2}\epsilon\lambda)^d e^{-(1-1/e)^2\lambda^2/2\sigma^2} \\
&\leq (\sqrt{2e}\epsilon)^d \frac{\lambda^d}{d^{d/2}\sigma^d} e^{-(1-1/e)^2\lambda^2/2\sigma^2} \\
&= (\sqrt{2e}\epsilon)^d e^{d(\ln c - c^2(1-1/e)^2/2)} \\
&\leq (\sqrt{2e}\epsilon)^d,
\end{aligned}$$

where the second inequality holds because $\epsilon \leq 1/(\sqrt{2}e)$ and for any point $\boldsymbol{x} \in B_{\sqrt{2}\epsilon\lambda}$,

$$e^{-\|(\boldsymbol{x}-\bar{\boldsymbol{x}})\|^2/2\sigma^2} \leq e^{-(1-\sqrt{2}\epsilon)^2\lambda^2/2\sigma^2} \leq e^{-(1-1/e)^2\lambda^2/2\sigma^2},$$

the third inequality follows from Proposition 2.4.8, and the last inequality holds because one can prove for any $c \geq 1$, $\ln c - c^2(1-1/e)^2/2 < 0$. ■

Bounds such as the following on the tails of Gaussian distributions are standard (see, for example [Fel68, Section VII.1])

Proposition 2.4.10 (Gaussian tail bound)

$$\left(\frac{\sigma}{x}\right)\frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi}} \geq \frac{1}{\sqrt{2\pi}\sigma}\int_{t=x}^{\infty} e^{-t^2/2\sigma^2}\, dt \geq \left(\frac{\sigma}{x} - \frac{\sigma^3}{x^3}\right)\frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi}}.$$

Using this, we prove:

Lemma 2.4.11 (Comparing Gaussian tails) *Let $\sigma \leq 1$ and let*

$$\mu(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2/2\sigma^2}.$$

Then, for $x \leq 2$ and $|x - y| \leq \epsilon$,

$$\frac{\int_{t=y}^{\infty} \mu(t)\, dt}{\int_{t=x}^{\infty} \mu(t)\, dt} \geq 1 - \frac{8\epsilon}{3\sigma^2}. \tag{6}$$

Proof If $y < x$, the ratio is greater than 1 and the lemma is trivially true. Assuming $y \geq x$, the ratio is minimized when $y = x + \epsilon$. In this case, the lemma will follow from

$$\frac{\int_{t=x}^{x+\epsilon} \mu(t)\, dt}{\int_{t=x}^{\infty} \mu(t)\, dt} \leq \frac{8\epsilon}{3\sigma^2}. \tag{7}$$

It follows from part (b) of Proposition 2.4.12 that the left-hand ratio in (7) is monotonically increasing in x, and therefore is maximized when x is maximized at 2. For $x = 2$, we apply Proposition 2.4.10 to show

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{t=x}^{\infty}\mu(t)\,dt \;\geq\; \left(\frac{\sigma}{2}-\frac{\sigma^3}{8}\right)\frac{e^{-2/\sigma^2}}{\sqrt{2\pi}} \geq \frac{3\sigma e^{-2/\sigma^2}}{8\sqrt{2\pi}}.$$

We then combine this bound with

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{t=x}^{x+\epsilon}\mu(t)\,dt \;\leq\; \frac{\epsilon e^{-2/\sigma^2}}{\sqrt{2\pi}\sigma},$$

to obtain

$$\frac{\int_{t=x}^{x+\epsilon}\mu(t)\,dt}{\int_{t=x}^{\infty}\mu(t)\,dt} \leq \left(\frac{\epsilon e^{-2/\sigma^2}}{\sqrt{2\pi}\sigma}\right)\left(\frac{8\sqrt{2\pi}}{3\sigma e^{-2/\sigma^2}}\right) = \frac{8\epsilon}{3\sigma^2}.$$

■

Proposition 2.4.12 (Monotonicity of Gaussian density) *Let*

$$\mu(t) = \frac{1}{\sqrt{2\pi}\sigma}e^{-t^2/2\sigma^2}.$$

(a) For all $a > 0$, $\mu(x)/\mu(x+a)$ is monotonically increasing in x;

(b) The following ratio is monotonically increasing in x

$$\frac{\mu(x)}{\int_{t=x}^{\infty}\mu(t)\,dt}$$

Proof Part (a) follows from

$$\frac{\mu(x)}{\mu(x+a)} = e^{(2ax+a^2)/2\sigma^2},$$

and that e^{2ax} is monotonically increasing in x.

To prove part (b) note that for all $a > 0$

$$\frac{\int_{t=x}^{\infty}\mu(t)\,dt}{\mu(x)} = \frac{\int_{t=0}^{\infty}\mu(x+t)\,dt}{\mu(x)} \geq \frac{\int_{t=0}^{\infty}\mu(x+a+t)\,dt}{\mu(x+a)} = \frac{\int_{t=x+a}^{\infty}\mu(t)\,dt}{\mu(x+a)},$$

where the inequality follows from part (a). ■

2.5 Changes of Variables

The main proof technique used in Section 4 is change of variables. For the reader's convenience, we recall how a change of variables affects probability distributions.

Proposition 2.5.1 (Change of variables) *Let $\boldsymbol{y}$ be a random variable distributed according to density μ. If $\boldsymbol{y} = \Phi(\boldsymbol{x})$, then $\boldsymbol{x}$ has density*

$$\mu(\Phi(\boldsymbol{x})) \left| \mathbf{det} \left(\frac{\partial \Phi(\boldsymbol{x})}{\partial \boldsymbol{x}} \right) \right| .$$

Recall that $\left| \mathbf{det} \left(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} \right) \right|$ is the Jacobian of the change of variables.

We now introduce the fundamental change of variables used in this paper. Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d$ be linearly independent points in $\mathbb{R}^d$. We will represent these points by specifying the plane passing through them and their positions on that plane. Many studies of the convex hulls of random point sets have used this change of variables (for example, see [RS63, RS64, Efr65, Mil71]). We specify the plane containing $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d$ by $\boldsymbol{\omega}$ and r, where $\|\boldsymbol{\omega}\| = 1$, $r \geq 0$ and $\langle \boldsymbol{\omega} | \boldsymbol{a}_i \rangle = r$ for all i. We will not concern ourselves with the issue that $\boldsymbol{\omega}$ is ill-defined if the $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d$ are affinely dependent, as this is an event of probability zero. To specify the positions of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d$ on the plane specified by $(\boldsymbol{\omega}, r)$, we must choose a coordinate system for that plane. To choose a canonical set of coordinates for each $(d-1)$-dimensional hyperplane specified by $(\boldsymbol{\omega}, r)$, we first fix a reference unit vector in $\mathbb{R}^d$, say $\boldsymbol{q}$, and an arbitrary coordinatization of the subspace orthogonal to $\boldsymbol{q}$. For any $\boldsymbol{\omega} \neq -\boldsymbol{q}$, we let

$$\boldsymbol{R}_{\boldsymbol{\omega}}$$

denote the linear transformation that rotates $\boldsymbol{q}$ to $\boldsymbol{\omega}$ in the two-dimensional subspace through $\boldsymbol{q}$ and $\boldsymbol{\omega}$ and that is the identity in the orthogonal subspace. Using $\boldsymbol{R}_{\boldsymbol{\omega}}$, we can map points specified in the $d-1$ dimensional hyperplane specified by r and $\boldsymbol{\omega}$ to $\mathbb{R}^d$ by

$$\boldsymbol{a}_i = \boldsymbol{R}_{\boldsymbol{\omega}} \boldsymbol{b}_i + r\boldsymbol{\omega},$$

where $\boldsymbol{b}_i$ is viewed both as a vector in $\mathbb{R}^{d-1}$ and as an element of the subspace orthogonal to $\boldsymbol{q}$. We will not concern ourselves with the fact that this map is not well defined if $\boldsymbol{q} = -\boldsymbol{\omega}$, as the set of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d$ that result in this coincidence has measure zero.

The Jacobian of this change of variables is computed by a famous theorem of integral geometry due to Blaschke [Bla35] (for more modern treatments, see [Mil71] or [San76, 12.24]), and actually depends only marginally on the coordinatizations of the hyperplanes.

Theorem 2.5.2 (Blaschke) *For variables $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d$ taking values in $\mathbb{R}^{d-1}$, $\boldsymbol{\omega} \in S^{d-1}$ and $r \in \mathbb{R}$, let*

$$(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d) = (\boldsymbol{R}_{\boldsymbol{\omega}} \boldsymbol{b}_1 + r\boldsymbol{\omega}, \ldots, \boldsymbol{R}_{\boldsymbol{\omega}} \boldsymbol{b}_d + r\boldsymbol{\omega})$$

The Jacobian of this map is

$$\left| \mathbf{det} \left(\frac{\partial (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d)}{\partial (\boldsymbol{\omega}, r, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)} \right) \right| = (d-1)! \mathbf{Vol} \left(\triangle (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d) \right) .$$

That is,

$$d\boldsymbol{a}_1 \ldots d\boldsymbol{a}_d = (d-1)! \mathbf{Vol} \left(\triangle (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d) \right) d\boldsymbol{\omega}\, dr\, d\boldsymbol{b}_1 \ldots d\boldsymbol{b}_d$$

We will also find it useful to specify the plane by $\boldsymbol{\omega}$ and s, where $\langle s\boldsymbol{q}|\boldsymbol{\omega}\rangle = r$, so that $s\boldsymbol{q}$ lies on the plane specified by $\boldsymbol{\omega}$ and r. We will also arrange our coordinate system so that the origin on this plane lies at $s\boldsymbol{q}$.

Corollary 2.5.3 (Blaschke with s) *For variables $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d$ taking values in $\mathbb{R}^{d-1}$, $\boldsymbol{\omega} \in S^{d-1}$ and $s \in \mathbb{R}$, let*

$$(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d) = (\boldsymbol{R}_{\boldsymbol{\omega}}\boldsymbol{b}_1 + s\boldsymbol{q}, \ldots, \boldsymbol{R}_{\boldsymbol{\omega}}\boldsymbol{b}_d + s\boldsymbol{q})$$

The Jacobian of this map is

$$\left|\mathbf{det}\left(\frac{\partial(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d)}{\partial(\boldsymbol{\omega}, s, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)}\right)\right| = (d-1)!\,\langle\boldsymbol{\omega}|\boldsymbol{q}\rangle\,\mathbf{Vol}\left(\triangle\left(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d\right)\right).$$

Proof So that we can apply Theorem 2.5.2, we will decompose the map into three simpler maps:

$$\begin{aligned}
(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d, s, \boldsymbol{\omega}) &\mapsto \left(\boldsymbol{b}_1 + \boldsymbol{R}_{\boldsymbol{\omega}}^{-1}(s\boldsymbol{q} - r\boldsymbol{\omega}), \ldots, \boldsymbol{b}_d + \boldsymbol{R}_{\boldsymbol{\omega}}^{-1}(s\boldsymbol{q} - r\boldsymbol{\omega}), s, \boldsymbol{\omega}\right) \\
&\mapsto \left(\boldsymbol{b}_1 + \boldsymbol{R}_{\boldsymbol{\omega}}^{-1}(s\boldsymbol{q} - r\boldsymbol{\omega}), \ldots, \boldsymbol{b}_d + \boldsymbol{R}_{\boldsymbol{\omega}}^{-1}(s\boldsymbol{q} - r\boldsymbol{\omega}), r, \boldsymbol{\omega}\right) \\
&\mapsto \left(R_{\boldsymbol{\omega}}\left(\boldsymbol{b}_1 + \boldsymbol{R}_{\boldsymbol{\omega}}^{-1}(s\boldsymbol{q} - r\boldsymbol{\omega})\right) + r\boldsymbol{\omega}, \ldots, R_{\boldsymbol{\omega}}\left(\boldsymbol{b}_d + \boldsymbol{R}_{\boldsymbol{\omega}}^{-1}(s\boldsymbol{q} - r\boldsymbol{\omega})\right) + r\boldsymbol{\omega}\right) \\
&= (R_{\boldsymbol{\omega}}\boldsymbol{b}_1 + s\boldsymbol{q}, \ldots, R_{\boldsymbol{\omega}}\boldsymbol{b}_d + s\boldsymbol{q})
\end{aligned}$$

As $s\boldsymbol{q} - r\boldsymbol{\omega}$ is orthogonal to $\boldsymbol{\omega}$, $\boldsymbol{R}_{\boldsymbol{\omega}}^{-1}(s\boldsymbol{q} - r\boldsymbol{\omega})$ can be interpreted as a vector in the $d-1$ dimensional space in which $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d$ lie. So, the first map is just a translation, and its Jacobian is 1. The Jacobian of the second map is

$$\left|\frac{\partial r}{\partial s}\right| = \langle\boldsymbol{q}|\boldsymbol{\omega}\rangle.$$

Finally, we note

$$\mathbf{Vol}\left(\boldsymbol{b}_1 + \boldsymbol{R}_{\boldsymbol{\omega}}^{-1}(s\boldsymbol{q} - r\boldsymbol{\omega}), \ldots, \boldsymbol{b}_d + \boldsymbol{R}_{\boldsymbol{\omega}}^{-1}(s\boldsymbol{q} - r\boldsymbol{\omega})\right) = \mathbf{Vol}\left(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d\right),$$

and that the third map is one described in Theorem 2.5.2. ■

In Section 4.2, we will need to represent $\boldsymbol{\omega}$ by $c = \langle\boldsymbol{\omega}|\boldsymbol{q}\rangle$ and $\boldsymbol{\psi} \in S^{d-2}$, where $\boldsymbol{\psi}$ gives the location of $\boldsymbol{\omega}$ in the cross-section of S^{d-1} for which $\langle\boldsymbol{\omega}|\boldsymbol{q}\rangle = c$. Formally, the map can be defined in a coordinate system with first coordinate $\boldsymbol{q}$ by

$$\boldsymbol{\omega} = (c, \boldsymbol{\psi}\sqrt{1-c^2}).$$

For this change of variables, we have:

Proposition 2.5.4 (Latitude and longitude) *The Jacobian of the change of variables from $\boldsymbol{\omega}$ to $(c, \boldsymbol{\psi})$ is*

$$\left|\mathbf{det}\left(\frac{\partial(\boldsymbol{\omega})}{\partial(c, \boldsymbol{\psi})}\right)\right| = (1-c^2)^{(d-3)/2}.$$

Proof We begin by changing $\boldsymbol{\omega}$ to $(\theta, \boldsymbol{\psi})$, where θ is the angle between $\boldsymbol{\omega}$ and $\boldsymbol{q}$, and $\boldsymbol{\psi}$ represents the position of $\boldsymbol{\omega}$ in the $d-2$ dimensional sphere of radius $\sin(\theta)$ of points at angle θ to $\boldsymbol{q}$. To compute the Jacobian of this change of variables, we choose a local coordinate system on S^{d-1} at $\boldsymbol{\omega}$ by taking the great circle through $\boldsymbol{\omega}$ and $\boldsymbol{q}$, and then an arbitrary coordinatization of the great $d-2$ dimensional sphere through $\boldsymbol{\omega}$ orthogonal to the great circle. In this coordinate system, θ is the position of $\boldsymbol{\omega}$ along the first great circle. As the $d-2$ dimensional sphere of points at angle θ to $\boldsymbol{q}$ is orthogonal to the great circle at $\boldsymbol{\omega}$, the coordinates in $\boldsymbol{\psi}$ can be mapped orthogonally into the coordinates of the great $d-2$ dimensional sphere–the only difference being the radii of the sub-spheres. Thus,

$$\left|\mathbf{det}\left(\frac{\partial(\boldsymbol{\omega})}{\partial(\theta, \boldsymbol{\psi})}\right)\right| = \sin(\theta)^{d-2}$$

If we now let $c = \cos(\theta)$, then we find

$$\left|\mathbf{det}\left(\frac{\partial(\boldsymbol{\omega})}{\partial(c, \boldsymbol{\psi})}\right)\right| = \left|\mathbf{det}\left(\frac{\partial(\boldsymbol{\omega})}{\partial(\theta, \boldsymbol{\psi})}\right)\right| \left|\mathbf{det}\left(\frac{\partial(\theta)}{\partial(c)}\right)\right| = \left(\sqrt{1-c^2}\right)^{d-2} \frac{1}{\sqrt{1-c^2}} = \left(\sqrt{1-c^2}\right)^{d-3}.$$

■

3 The Shadow Vertex Method

In this section, we will review the shadow vertex method and formally state the two-phase method analyzed in this paper. We will begin by motivating the method. In Section 3.1, we will explain how the method works assuming a feasible vertex is known. In Section 3.2, we present a polar perspective on the method, from which our analysis is most natural. We then present a complete two-phase method in Section 3.3. For a more complete exposition of the Shadow Vertex Method, we refer the reader to [Bor80, Chapter 1].

The shadow-vertex simplex method is motivated by the observation that the simplex method is very simple in two-dimensions: the set of feasible points form a (possibly open) polygon, and the simplex method merely walks along the exterior of the polygon. The shadow-vertex method lifts the simplicity of the simplex method in two dimensions to higher dimensions. Let $\boldsymbol{z}$ be the objective function of a linear program and let $\boldsymbol{t}$ be an objective function optimized by $\boldsymbol{x}$, a vertex of the polytope of feasible points for the linear program. The shadow-vertex method considers the *shadow* of the polytope—the projection of the polytope onto the plane spanned by $\boldsymbol{z}$ and $\boldsymbol{t}$. One can verify that

(1) this shadow is a (possibly open) polygon,

(2) each vertex of the polygon is the image of a vertex of the polytope,

(3) each edge of the polygon is the image of an edge between two adjacent vertices of the polytope,

(4) the projection of $\boldsymbol{x}$ onto the plane is a vertex of the polygon, and

(5) the projection of the vertex optimizing $\boldsymbol{z}$ onto the plane is a vertex of the polygon.

Thus, if one walks along the vertices of the polygon starting from the image of $\boldsymbol{x}$, and keeps track of the vertices' pre-images on the polytope, then one will eventually encounter the vertex of the polytope optimizing $\boldsymbol{z}$. Given one vertex of the polytope that maps to a vertex of the polygon, it is easy to find the vertex of the polytope that maps to the next vertex of the polygon: fact (3) implies that it must be a neighbor of the vertex on the polytope; moreover, for a linear program that is in general position with respect to $\boldsymbol{t}$, there will be d such vertices. Thus, the method will be efficient provided that the shadow polygon does not have too many vertices. This is the motivation for the shadow vertex method.

3.1 Formal Description

Our description of the shadow vertex simplex method will be facilitated by the following definition:

Definition 3.1.1 (optVert) *Given vectors $\boldsymbol{z}$, $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ in $\mathbb{R}^d$ and $\boldsymbol{y} \in \mathbb{R}^n$, we define* $\mathbf{optVert}_{\boldsymbol{z}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n; \boldsymbol{y})$ *to be the set of $\boldsymbol{x}$ solving*

$$\begin{array}{ll} \textit{maximize} & \boldsymbol{z}^T \boldsymbol{x} \\ \textit{subject to} & \langle \boldsymbol{a}_i | \boldsymbol{x} \rangle \leq y_i, \ \textit{for } 1 \leq i \leq n. \end{array}$$

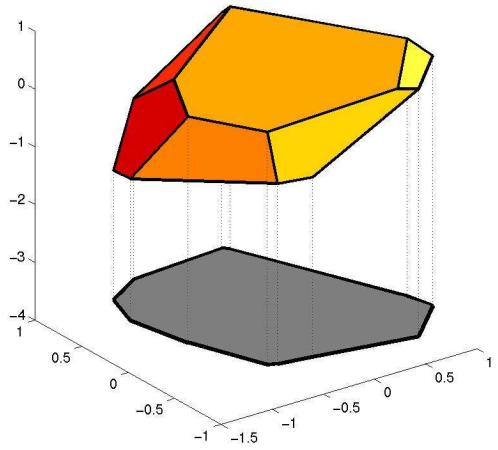

Figure 1: A shadow of a polytope

If there are no such $\boldsymbol{x}$*, either because the program is unbounded or infeasible, we let* $\mathbf{optVert}_{\boldsymbol{z}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n; \boldsymbol{y})$ *be* $\emptyset$*. When* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ *and* $\boldsymbol{y}$ *are understood, we will use the notation* $\mathbf{optVert}_{\boldsymbol{z}}$*.*

We note that, for linear programs in general position, $\mathbf{optVert}_{\boldsymbol{z}}$ will either be empty or contain one vertex.

Using this definition, we will give a description of the shadow vertex method assuming that a vertex $\boldsymbol{x}_0$ and a vector $\boldsymbol{t}$ are known for which $\mathbf{optVert}_{\boldsymbol{t}} = \boldsymbol{x}_0$. An algorithm that works without this assumption will be described in Section 3.3. Given $\boldsymbol{t}$ and $\boldsymbol{z}$, we define objective functions interpolating between the two by

$$\boldsymbol{q}_\lambda = (1 - \lambda)\boldsymbol{t} + \lambda \boldsymbol{z}.$$

The shadow-vertex method will proceed by varying λ from 0 to 1, and tracking $\mathbf{optVert}_{\boldsymbol{q}_\lambda}$. We will denote the vertices encountered by $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$, and we will set λ_i so that $\boldsymbol{x}_i \in \mathbf{optVert}_{\boldsymbol{q}_\lambda}$ for $\lambda \in [\lambda_i, \lambda_{i+1}]$.

As our main motivation for presenting the primal algorithm is to develop intuition in the reader, we will not dwell on issues of degeneracy in its description. We will present a polar version of this algorithm with a proof of correctness in the next section.

primal shadow-vertex method
Input: $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, $\boldsymbol{y}$, $\boldsymbol{z}$, and $\boldsymbol{x}_0$ and $\boldsymbol{t}$ satisfying $\{\boldsymbol{x}_0\} = \mathbf{optVert}_{\boldsymbol{t}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n; \boldsymbol{y})$.

(1) Set $\lambda_0 = 0$, and $i = 0$.

(2) Set λ_1 to be maximal such that $\{\boldsymbol{x}_0\} = \mathbf{optVert}_{\boldsymbol{q}_\lambda}$ for $\lambda \in [\lambda_0, \lambda_1]$.

(3) while $\lambda_{i+1} < 1$,

  (a) Set $i = i + 1$.

  (b) Find an $\boldsymbol{x}_i$ for which there exists a $\lambda_{i+1} > \lambda_i$ such that $\boldsymbol{x}_i \in \mathbf{optVert}_{\boldsymbol{q}_\lambda}$ for $\lambda \in [\lambda_i, \lambda_{i+1}]$. If no such $\boldsymbol{x}_i$ exists, return *unbounded*.

  (c) Let λ_{i+1} be maximal such that $\boldsymbol{x}_i \in \mathbf{optVert}_{\boldsymbol{q}_\lambda}$ for $\lambda \in [\lambda_i, \lambda_{i+1}]$.

(4) return $\boldsymbol{x}_i$.

Step (b) of this algorithm deserves further explanation. Assuming that the linear program is in general position with respect to $\boldsymbol{t}$, each vertex $\boldsymbol{x}_i$ will have exactly d neighbors, and the vertex $\boldsymbol{x}_{i+1}$ will be one of these [Bor80, Lemma 1.3]. Thus, the algorithm can be described as a simplex method. While one could implement the method by examining these d vertices in turn, more efficient implementations are possible. For an efficient implementation of this algorithm in tableau form, we point the reader to the exposition in [Bor80, Section 1.3].

3.2 Polar Description

Following Borgwardt [Bor80], we will analyze the shadow vertex method from a polar perspective. This polar perspective is natural provided that all $y_i > 0$. In this section, we will describe a polar variant of the shadow-vertex method that works under this assumption. In the next section, we will describe a two-phase shadow vertex method that uses this polar variant to solve linear programs with arbitrary y_is.

While it is not strictly necessary for the results in this paper, we remind the reader that for a polytope $P = \{\boldsymbol{x} : \langle \boldsymbol{x} | \boldsymbol{a}_i \rangle \leq 1, \forall i\}$, the polar of P is $\{\boldsymbol{y} : \langle \boldsymbol{x} | \boldsymbol{y} \rangle \leq 1, \forall \boldsymbol{x} \in P\}$. An equivalent definition of the polar is $\mathbf{ConvHull}(\mathbf{0}, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$. We remark that P is bounded if and only if $\mathbf{0}$ is in the interior of $\mathbf{ConvHull}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$. The polar motivates:

Definition 3.2.1 (optSimp) *For $\boldsymbol{z}$ and $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ in $\mathbb{R}^d$ and $\boldsymbol{y} \in \mathbb{R}^n$, $y_i > 0$, we let $\mathbf{optSimp}_{\boldsymbol{z}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n; \boldsymbol{y})$ denote the set of $I \in \binom{[n]}{d}$ such that $\boldsymbol{A}_I$ has full rank, $\triangle((\boldsymbol{a}_i/y_i)_{i \in I})$ is a facet of $\boldsymbol{ConvHull}(\mathbf{0}, \boldsymbol{a}_1/y_1, \ldots, \boldsymbol{a}_n/y_n)$ and $\boldsymbol{z} \in \mathbf{Cone}((\boldsymbol{a}_i)_{i \in I})$. When $\boldsymbol{y}$ is understood to be $\mathbf{1}$, we will use the notation $\mathbf{optSimp}_{\boldsymbol{z}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ When $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ and $\boldsymbol{y}$ are understood, we will use the notation $\mathbf{optSimp}_{\boldsymbol{z}}$.*

We remark that for $\boldsymbol{y}$, $\boldsymbol{z}$ and $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ in general position, $\mathbf{optSimp}_{\boldsymbol{z}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n; \boldsymbol{y})$ will be the empty set or contain just one set of indices I.

The following proposition follows from the duality theory of linear programming:

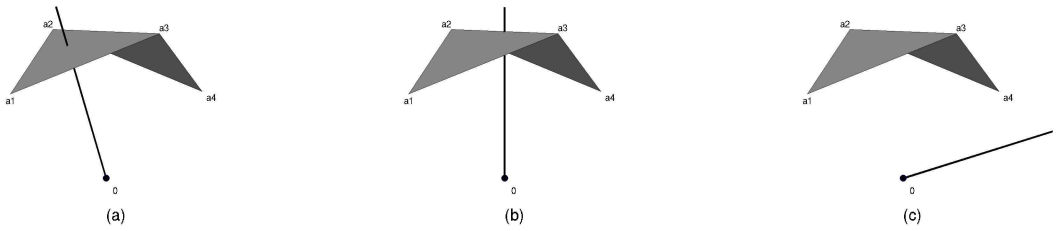

Figure 2: In example (a), optSimp $= \{\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}\}$. In example (b), optSimp $= \{\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}, \{\boldsymbol{a}_2, \boldsymbol{a}_3, \boldsymbol{a}_4\}\}$. In example (c), optSimp $= \emptyset$,

Proposition 3.2.2 (Duality) *For $y_1, \ldots, y_n > 0$, $I \in \mathbf{optSimp}_{\boldsymbol{z}}(\boldsymbol{a}_1/y_1, \ldots, \boldsymbol{a}_n/y_n)$ if and only if there exists an $\boldsymbol{x}$ such that $\boldsymbol{x} \in \mathbf{optVert}_{\boldsymbol{z}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n; \boldsymbol{y})$ and $\langle \boldsymbol{x} | \boldsymbol{a}_i \rangle = y_i$, for $i \in I$.*

We now state the polar shadow vertex method.

polar shadow-vertex method
Input:

- $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, $\boldsymbol{z}$, and $y_1, \ldots, y_n > 0$,
- $I \in \binom{[n]}{d}$ and $\boldsymbol{t}$ satisfying $I \in \mathbf{optSimp}_{\boldsymbol{t}}(\boldsymbol{a}_1/y_1, \ldots, \boldsymbol{a}_n/y_n)$.

(1) Set $\lambda_0 = 0$ and $i = 0$.

(2) Set λ_1 to be maximal such that for $\lambda \in [\lambda_0, \lambda_1]$,

$$I \in \mathbf{optSimp}_{\boldsymbol{q}_\lambda}(\boldsymbol{a}_1/y_1, \ldots, \boldsymbol{a}_n/y_n).$$

(3) while $\lambda_{i+1} < 1$,

  (a) Set $i = i + 1$.

  (b) Find a j and k for which there exists a $\lambda_{i+1} > \lambda_i$ such that

$$I \cup \{j\} - \{k\} \in \mathbf{optSimp}_{\boldsymbol{q}_\lambda}(\boldsymbol{a}_1/y_1, \ldots, \boldsymbol{a}_n/y_n)$$

  for $\lambda \in [\lambda_i, \lambda_{i+1}]$. If no such j and k exist, return *unbounded*.

  (c) Set $I = I \cup \{j\} - \{k\}$.

  (d) Let λ_{i+1} be maximal such that $I \in \mathbf{optSimp}_{\boldsymbol{t}}(\boldsymbol{a}_1/y_1, \ldots, \boldsymbol{a}_n/y_n)$ for $\lambda \in [\lambda_i, \lambda_{i+1}]$.

(4) return I.

The $\boldsymbol{x}$ optimizing the linear program, namely $\mathbf{optVert}_{\boldsymbol{z}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n; \boldsymbol{y})$, is given by the

equations $\langle \boldsymbol{x} | \boldsymbol{a}_i \rangle = y_i$, for $i \in I$.

Borgwardt [Bor80, Lemma 1.9] establishes that such j and k can be found in step (b) if there exists an ϵ for which $\mathbf{optSimp}_{\boldsymbol{q}_{\lambda_i+\epsilon}}(\boldsymbol{a}_1/y_1, \ldots, \boldsymbol{a}_n/y_n) \neq \emptyset$. That the algorithm may conclude that the program is unbounded if a j and k cannot be found in step (b) follows from:

Proposition 3.2.3 (Detecting unbounded programs) *If there is an i and an $\epsilon > 0$ such that $\lambda_i + \epsilon < 1$ and $\mathbf{optSimp}_{\boldsymbol{q}_{\lambda_i+\epsilon}}(\boldsymbol{a}_1/y_1, \ldots, \boldsymbol{a}_n/y_n) = \emptyset$, then $\mathbf{optSimp}_{\boldsymbol{z}}(\boldsymbol{a}_1/y_1, \ldots, \boldsymbol{a}_n/y_n) = \emptyset$.*

Proof $\mathbf{optSimp}_{\boldsymbol{q}_{\lambda_i+\epsilon}}(\boldsymbol{a}_1/y_1, \ldots, \boldsymbol{a}_n/y_n) = \emptyset$ if and only if $\boldsymbol{q}_{\lambda_i+\epsilon} \notin \mathbf{Cone}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$. The proof now follows from the facts that $\mathbf{Cone}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ is a convex set and $\boldsymbol{q}_{\lambda_i+\epsilon}$ is a positive multiple of a convex combination of $\boldsymbol{t}$ and $\boldsymbol{z}$. ■

The running time of the shadow-vertex method is bounded by the number of vertices in shadow of the polytope defined by the constraints of the linear program. Formally, this is

Definition 3.2.4 (Shadow) *For independent vectors $\boldsymbol{t}$ and $\boldsymbol{z}$, $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ in $\mathbb{R}^d$ and $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{y} > 0$,*

$$\mathbf{Shadow}_{\boldsymbol{t},\boldsymbol{z}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n; \boldsymbol{y}) \stackrel{\text{def}}{=} \bigcup_{\boldsymbol{q} \in \boldsymbol{Span}(\boldsymbol{t},\boldsymbol{z})} \left\{ \mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1/y_1, \ldots, \boldsymbol{a}_n/y_n) \right\}.$$

If $\boldsymbol{y}$ is understood to be $\mathbf{1}$, we will just write $\mathbf{Shadow}_{\boldsymbol{t},\boldsymbol{z}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$.

3.3 Two-Phase Method

We now describe a two-phase shadow vertex method that solves linear programs of form

$$\begin{array}{ll} \text{maximize} & \langle \boldsymbol{z} | \boldsymbol{x} \rangle \\ \text{subject to} & \langle \boldsymbol{a}_i | \boldsymbol{x} \rangle \leq y_i, \text{ for } 1 \leq i \leq n. \end{array} \qquad (LP)$$

There are three issues that we must resolve before we can apply the polar shadow vertex method as described in Section 3.2 to the solution of such programs:

(1) the method must know a feasible vertex of the linear program,

(2) the linear program might not even be feasible, and

(3) some y_i might be non-positive.

The first two issues are standard motivations for two-phase methods, while the third is motivated by the polar perspective from which we prefer to analyze the shadow vertex method. We resolve these issues in two stages. We first relax the constraints of LP to construct a linear program LP' such that

(a) the right-hand vector of the linear program is positive, and

(b) we know a feasible vertex of the linear program.

After solving LP', we construct another linear program, LP^+, in one higher dimension that interpolates between LP and LP'. LP^+ has properties (a) and (b), and we can use the shadow vertex method on LP^+ to transform the solution to LP' into a solution of LP.

Our two-phase method first chooses a d-set I to define the known feasible vertex of LP'. The linear program LP' is determined by $\boldsymbol{A}$, $\boldsymbol{z}$ and the choice of I. However, the magnitude of the right-hand entries in LP' depends upon $\mathbf{s_{min}}(A_I)$. To reduce the chance that these entries will need to be large, we examine several randomly chosen d-sets, and use the one maximizing $\mathbf{s_{min}}$.

The algorithm then sets

$$\begin{aligned} M &= 2^{\lceil \lg(\max_i \|y_i, \boldsymbol{a}_i\|)\rceil + 2}, \\ \kappa &= 2^{\lfloor \lg(\mathbf{s_{min}}(\boldsymbol{A}_I))\rfloor}, \text{and} \\ y'_i &= \begin{cases} M & \text{for } i \in I \\ \sqrt{d}M^2/4\kappa & \text{otherwise.} \end{cases} \end{aligned}$$

These define the program LP':

$$\begin{array}{ll} \text{maximize} & \langle \boldsymbol{z} | \boldsymbol{x} \rangle \\ \text{subject to} & \langle \boldsymbol{a}_i | \boldsymbol{x} \rangle \leq y'_i, \text{ for } 1 \leq i \leq n. \end{array} \qquad (LP')$$

By Proposition 3.3.1, $\boldsymbol{A}_I$ is a feasible basis for LP', and optimizes any objective function of the form $\boldsymbol{A}_I\boldsymbol{\alpha}$, for $\boldsymbol{\alpha} > 0$. Our two-phase algorithm will solve LP' by starting the polar shadow-vertex algorithm at the basis I and the objective function $\boldsymbol{A}_I\boldsymbol{\alpha}$ for a randomly chosen $\boldsymbol{\alpha}$ satisfying $\sum \alpha_i = 1$ and $\alpha_i \geq 1/d^2$, for all i.

Proposition 3.3.1 (Initial simplex of LP') *For any $\boldsymbol{\alpha} > 0$, $I = \mathbf{optSimp}_{\boldsymbol{A}_I\boldsymbol{\alpha}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n; \boldsymbol{y}')$.*

Proof Let $\boldsymbol{x}'$ be the solution of the linear system

$$\left\langle \boldsymbol{a}_i | \boldsymbol{x}' \right\rangle = y'_i, \qquad \text{for} \qquad i \in I.$$

By Definition 2.2.3 and Proposition 2.2.4 (a),

$$\left\|\boldsymbol{x}'\right\| \leq \left\|\boldsymbol{y}'_I\right\| \left\|\boldsymbol{A}_I^{-1}\right\| \leq M\sqrt{d} \left\|\boldsymbol{A}_I^{-1}\right\| = M\sqrt{d}/\mathbf{s_{min}}(\boldsymbol{A}_I).$$

So, for all $i \notin I$,

$$\left\langle \boldsymbol{a}_i | \boldsymbol{x}' \right\rangle \leq (\max_i \|\boldsymbol{a}_i\|) M\sqrt{d}/\mathbf{s_{min}}(\boldsymbol{A}_I) < M^2\sqrt{d}/4\kappa.$$

Thus, for all $i \notin I$,

$$\left\langle \boldsymbol{a}_i | \boldsymbol{x}' \right\rangle < y'_i,$$

and, by Definition 3.2.1, $I = \mathbf{optSimp}_{\boldsymbol{A}_I\boldsymbol{\alpha}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n; \boldsymbol{y}')$. ■

We will now define a linear program LP^+ that interpolates between LP' and LP. This linear program will contain an extra variable x_0 and constraints of the form

$$\langle \boldsymbol{a}_i | \boldsymbol{x} \rangle \leq \left(\frac{1 + x_0}{2}\right) y_i + \left(\frac{1 - x_0}{2}\right) y'_i,$$

and $-1 \leq x_0 \leq 1$. So, for $x_0 = 1$ we see the original program LP while for $x_0 = -1$ we get LP'. Formally, we let

$$\boldsymbol{a}_i^+ = \begin{cases} ((y_i' - y_i)/2, \boldsymbol{a}_i) & \text{for } 1 \leq i \leq n \\ (1, 0, \ldots, 0) & \text{for } i = 0 \\ (-1, 0, \ldots, 0) & \text{for } i = -1 \end{cases}$$

$$y_i^+ = \begin{cases} (y_i' + y_i)/2 & \text{for } 1 \leq i \leq n \\ 1 & \text{for } i = 0 \\ 1 & \text{for } i = -1 \end{cases}$$

$$\boldsymbol{z}^+ = (1, 0, \ldots, 0),$$

and we define LP^+ by

$$\begin{array}{ll} \text{maximize} & \left\langle \boldsymbol{z}^+ | (x_0, \boldsymbol{x}) \right\rangle \\ \text{subject to} & \left\langle \boldsymbol{a}_i^+ | (x_0, \boldsymbol{x}) \right\rangle \leq y_i^+, \text{ for } -1 \leq i \leq n, \end{array} \qquad (LP^+)$$

and we set

$$\boldsymbol{y}^+ \stackrel{\text{def}}{=} (y_{-1}^+, \ldots, y_n^+).$$

By Proposition 3.3.2, $\sqrt{d}M/4\kappa \geq 1$, so $y_i' \geq M$ and $y_i^+ > 0$, for all i. If LP is infeasible, then the solution to LP^+ will have $x_0 < 1$. If LP is feasible, then the solution to LP^+ will have form $(1, \boldsymbol{x})$ where $\boldsymbol{x}$ is a feasible point for LP. If we use the shadow-vertex method to solve LP^+ starting from the appropriate initial vector, then $\boldsymbol{x}$ will be an optimal solution to LP.

Proposition 3.3.2 (relation of M and κ) *For M and κ as set by the algorithm, $\sqrt{d}M/4\kappa \geq 1$.*

Proof By definition, $\kappa \leq \mathbf{s_{min}}(\boldsymbol{A}_I)$. On the other hand, $\mathbf{s_{min}}(\boldsymbol{A}_I) \leq \|\boldsymbol{A}_I\| \leq \sqrt{d} \max_i \|\boldsymbol{a}_i\|$, by Proposition 2.2.4 (d). Finally, $M \geq 4 \max_i \|\boldsymbol{a}_i\|$. ■

We now state and prove the correctness of the two-phase shadow vertex method.

two-phase shadow-vertex method
Input: $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$, $\boldsymbol{y}$, $\boldsymbol{z}$.

(1) Let $\mathcal{I} = \{I_1, \ldots, I_{3nd \ln n}\}$ be a collection of randomly chosen sets in $\binom{[n]}{d}$, and let $I \in \mathcal{I}$ be the set maximizing $\mathbf{s_{min}}(\boldsymbol{A}_I)$.

(2) Set $M = 2^{\lceil \lg(\max_i \|y_i, \boldsymbol{a}_i\|) \rceil + 2}$ and $\kappa = 2^{\lfloor \lg(\mathbf{s_{min}}(\boldsymbol{A}_I)) \rfloor}$.

(3) Set $y'_i = \begin{cases} M & \text{for } i \in I \\ \sqrt{d}M^2/4\kappa & \text{otherwise.} \end{cases}$.

(4) Choose $\boldsymbol{\alpha}$ uniformly at random from $\{\boldsymbol{\alpha} : \sum \alpha_i = 1 \text{ and } \alpha_i \geq 1/d^2\}$. Set $\boldsymbol{t}' = \boldsymbol{A}_I \boldsymbol{\alpha}$.

(5) Let J be the output of the polar shadow vertex algorithm on LP' on input I and $\boldsymbol{t}'$. If LP' is *unbounded*, then return *unbounded*.

(6) Let $\zeta > 0$ be such that

$$\{-1\} \cup J \in \mathbf{optSimp}_{(-\zeta, \boldsymbol{z})}\big(\boldsymbol{a}^+_{-1}/y^+_{-1}, \ldots, \boldsymbol{a}^+_n/y^+_n\big).$$

(7) Let K be the output of the polar shadow vertex algorithm on LP^+ on input $\{-1\} \cup J$, $(-\zeta, \boldsymbol{z})$.

(8) Compute $(x_0, \boldsymbol{x})$ satisfying $\big\langle (x_0, \boldsymbol{x}) | \boldsymbol{a}^+_i \big\rangle = y_i$ for $i \in K$.

(9) If $x_0 < 1$, return *infeasible*. Otherwise, return $\boldsymbol{x}$.

The following propositions prove the correctness of the algorithm.

Proposition 3.3.3 (Unbounded programs) *The following are equivalent*

(a) LP is unbounded;

(b) LP' is unbounded;

(c) there exists a $1 > \lambda > 0$ such that $\mathbf{optSimp}_{\lambda(1,\mathbf{0})+(1-\lambda)(-\zeta,\boldsymbol{z})}\big(\boldsymbol{a}^+_{-1}, \ldots, \boldsymbol{a}^+_n; \boldsymbol{y}^+\big) = \emptyset$;

(d) for all $1 > \lambda > 0$, $\mathbf{optSimp}_{\lambda(1,\mathbf{0})+(1-\lambda)(-\zeta,\boldsymbol{z})}\big(\boldsymbol{a}^+_{-1}, \ldots, \boldsymbol{a}^+_n; \boldsymbol{y}^+\big) = \emptyset$.

Proposition 3.3.4 (Bounded programs) *If LP' is bounded and has solution J, then*

(a) there exists ζ_0 such that for all $\zeta > \zeta_0$, $\{-1\} \cup J \in \mathbf{optSimp}_{(-\zeta,\boldsymbol{z})}\big(\boldsymbol{a}^+_{-1}, \ldots, \boldsymbol{a}^+_n; \boldsymbol{y}^+\big)$,

(b) If LP is feasible, then for $K' \in \mathbf{optSimp}_{\boldsymbol{z}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n; \boldsymbol{y})$, there exists ξ_0 such that for all $\xi > \xi_0$, $\{0\} \cup K' \in \mathbf{optSimp}_{(\xi,\boldsymbol{z})}\big(\boldsymbol{a}^+_{-1}, \ldots, \boldsymbol{a}^+_n; \boldsymbol{y}^+\big)$, and

(c) if we use the shadow vertex method to solve LP^+ starting from $\{-1, J\}$ and objective function $(-\zeta, \boldsymbol{z})$, then the output of the algorithm will have form $\{0\} \cup K'$, where K' is a solution to LP.

Proof of Proposition 3.3.3 LP is unbounded if and only if there exists a vector $\boldsymbol{v}$ such that $\langle \boldsymbol{z}|\boldsymbol{v}\rangle > 0$ and $\langle \boldsymbol{a}_i|\boldsymbol{v}\rangle \leq 0$ for all i. The same holds for LP', and establishes the equivalence of (a) and (b). To show that (a) or (b) implies (d), observe

$$\begin{aligned}
\langle \lambda(1,\mathbf{0}) + (1-\lambda)(-\zeta,\boldsymbol{z})|(0,\boldsymbol{v})\rangle &= (1-\lambda)\langle \boldsymbol{z}|\boldsymbol{v}\rangle > 0, && (8)\\
\left\langle \boldsymbol{a}_i^+|(0,\boldsymbol{v})\right\rangle &= \langle \boldsymbol{a}_i|\boldsymbol{v}\rangle, \text{ for } i = 1,\ldots,n, && (9)\\
\left\langle \boldsymbol{a}_0^+|(0,\boldsymbol{v})\right\rangle &= 0, \text{ and}\\
\left\langle \boldsymbol{a}_{-1}^+|(0,\boldsymbol{v})\right\rangle &= 0.
\end{aligned}$$

To show that (c) implies (a) and (b), note that $\boldsymbol{a}_0^+$ and $\boldsymbol{a}_{-1}^+$ are arranged so that if for some v_0 we have

$$\left\langle \boldsymbol{a}_i^+|(v_0,\boldsymbol{v})\right\rangle \leq 0, \text{ for } -1 \leq i \leq n,$$

then $v_0 = 0$. This identity allows us to apply (8) and (9) to show (c) implies (a) and (b). ■

Proof of Proposition 3.3.4 Let J be the solution to LP' and let $\boldsymbol{x}' = \boldsymbol{A}_J^{-1}\boldsymbol{y}'_J$ be the corresponding vertex. We then have

$$\begin{aligned}
\left\langle \boldsymbol{x}'|\boldsymbol{a}_i\right\rangle &= y_i', && \text{for } i \in J, \text{ and}\\
\left\langle \boldsymbol{x}'|\boldsymbol{a}_i\right\rangle &\leq y_i', && \text{for } i \notin J.
\end{aligned}$$

Therefore, it is clear that

$$\begin{aligned}
\left\langle (-1,\boldsymbol{x}')|\boldsymbol{a}_i^+\right\rangle &= y_i^+, && \text{for } i \in \{-1\}\cup J, \text{ and}\\
\left\langle (-1,\boldsymbol{x}')|\boldsymbol{a}_i^+\right\rangle &\leq y_i^+, && \text{for } i \notin \{-1\}\cup J.
\end{aligned}$$

Thus, $\triangle\left(\boldsymbol{a}_{-1}^+,(\boldsymbol{a}_i^+)_{i\in J}\right)$ is a facet of LP^+. To see that there exists a ζ_0 such that it optimizes $(-\zeta,\boldsymbol{z})$ for all $\zeta > \zeta_0$, first observe that there exist $\alpha_i > 0$, for $i \in J$, such that $\sum_{i\in J}\alpha_i\boldsymbol{a}_i = \boldsymbol{z}$. Now, let $(-\zeta_0,\boldsymbol{z}) = \sum_{i\in J}\alpha_i\boldsymbol{a}_i^+$. For $\zeta > \zeta_0$, we have

$$(-\zeta,\boldsymbol{z}) = (\zeta-\zeta_0)\boldsymbol{a}_{-1}^+ + \sum_{i\in J}\alpha_i\boldsymbol{a}_i^+,$$

which proves $(-\zeta,\boldsymbol{z}) \in \mathbf{Cone}\left(\boldsymbol{a}_{-1}^+,(\boldsymbol{a}_i^+)_{i\in J}\right)$ and completes the proof of (a).

The proof of (b) is similar.

To prove part (c), let K be as in step (7). Then, there exists a λ_k such that for all $\lambda \in (\lambda_k,1)$,

$$K = \mathbf{optSimp}_{(1-\lambda)(-\zeta,\boldsymbol{z})+\lambda\boldsymbol{z}^+}\left(\boldsymbol{a}_{-1}^+,\ldots,\boldsymbol{a}_n^+;\boldsymbol{y}^+\right).$$

Let $(x_0,\boldsymbol{x})$ satisfy $\left\langle (x_0,\boldsymbol{x})|\boldsymbol{a}_i^+\right\rangle = y_i^+$, for $i \in K$. Then, by Proposition 3.2.2,

$$(x_0,\boldsymbol{x}) = \mathbf{optVert}_{(1-\lambda)(-\zeta,\boldsymbol{z})+\lambda\boldsymbol{z}^+}\left(\boldsymbol{a}_{-1}^+,\ldots,\boldsymbol{a}_n^+;\boldsymbol{y}^+\right).$$

If $x_0 < 1$, then LP was infeasible. Otherwise, let $\boldsymbol{x}^* = \mathbf{optVert}_{\boldsymbol{z}}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n;\boldsymbol{y})$. By part (b), there exists ξ_0 such that for all $\xi > \xi_0$,

$$(1,\boldsymbol{x}^*) = \mathbf{optVert}_{(\xi,\boldsymbol{z})}\left(\boldsymbol{a}_{-1}^+,\ldots,\boldsymbol{a}_n^+;\boldsymbol{y}^+\right).$$

For $\xi = -\zeta + \lambda/(1-\lambda)$, we have

$$(\xi, \boldsymbol{z}) = \frac{1}{1-\lambda}\left((1-\lambda)(-\zeta, \boldsymbol{z}) + \lambda \boldsymbol{z}^+\right).$$

So, as λ approaches 1, $\xi = -\zeta + \lambda/(1-\lambda)$ goes to infinity and we have

$$\mathbf{optVert}_{(1-\lambda)(-\zeta,\boldsymbol{z})+\lambda \boldsymbol{z}^+}\left(\boldsymbol{a}_{-1}^+, \ldots, \boldsymbol{a}_n^+; \boldsymbol{y}^+\right) = \mathbf{optVert}_{(\xi,\boldsymbol{z})}\left(\boldsymbol{a}_{-1}^+, \ldots, \boldsymbol{a}_n^+; \boldsymbol{y}^+\right),$$

which implies $(x_0, \boldsymbol{x}) = (1, \boldsymbol{x}^*)$. ■

Finally, we bound the number of steps taken in step (7) by the shadow size of a related polytope:

Lemma 3.3.5 (Shadow path of LP^+) *Let $\boldsymbol{a}_{-1}^+, \ldots, \boldsymbol{a}_n^+$ and $y_{-1}^+, \ldots, y_n^+$ be as defined in LP^+. Let $\zeta > 0$ be such that $\{-1\} \cup J = \mathbf{optSimp}_{(-\zeta,\boldsymbol{z})}\left(\boldsymbol{a}_{-1}^+/y_{-1}^+, \ldots, \boldsymbol{a}_n^+/y_n^+\right)$. Then the number of simplex steps made by the polar shadow vertex algorithm while solving LP^+ from initial basis $\{-1\} \cup J$ and vector $(-\zeta, \boldsymbol{z})$ is at most*

$$2 + \left|\mathbf{Shadow}_{(0,\boldsymbol{z}),\boldsymbol{z}^+}\left(\boldsymbol{a}_1^+/y_1^+, \ldots, \boldsymbol{a}_n^+/y_n^+\right)\right|.$$

Proof We will establish that $\{-1\} \in I$ for the first step only. One can similarly prove that $\{0\} \in I$ is only true at termination.

Let $I \in \mathbf{optSimp}_{\boldsymbol{q}_\lambda}\left(\boldsymbol{a}_{-1}^+/y_{-1}^+, \ldots, \boldsymbol{a}_n^+/y_n^+\right)$ have form $\{-1\} \cup L$. As $\boldsymbol{q}_0 = \boldsymbol{a}_{-1}^+ \in \mathbf{Cone}\left(\boldsymbol{A}_{\{-1\}\cup L}\right)$, and $\mathbf{Cone}\left(\boldsymbol{A}_{\{-1\}\cup L}\right)$ is a convex set, we have $\boldsymbol{q}_{\lambda'} \in \mathbf{Cone}\left(\boldsymbol{A}_{\{-1\}\cup L}\right)$ for all $0 \leq \lambda' \leq \lambda$. As $[\lambda_i, \lambda_{i+1}]$ is exactly the set of λ optimized by $\triangle\left(\boldsymbol{A}_I\right)$ in the ith step of the polar shadow vertex method, I must be the initial set.

■

3.4 Discussion

We also note that our analysis of the two-phase algorithm actually takes advantage of the fact that κ and M have been set to powers of two. In particular, this fact is used to show that there are not too many likely choices for κ and M. For the reader who would like to drop this condition, we briefly explain how the argument of Section 5 could be modified to compensate: first, we could consider setting κ and M to powers of $1 + 1/poly(n, d, 1/\sigma)$. This would still result in a polynomially bounded number of choices for κ and M. One could then drop this assumption by observing that allowing κ and M to vary in a small range would not introduce too much dependency between the variables.

4 Shadow Size

In this section, we bound the expected size of the shadow of the perturbation of a polytope onto a fixed plane. This is the main geometric result of the paper. The algorithmic results of this paper will rely on extensions of this theorem derived in Section 4.3.

Theorem 4.0.1 (Shadow Size) *Let $d \geq 3$ and $n > d$. Let $\boldsymbol{z}$ and $\boldsymbol{t}$ be independent vectors in $\mathbb{R}^d$, and let $\mu_1, \ldots, \mu_n$ be Gaussian distributions in $\mathbb{R}^d$ of standard deviation σ centered at points each of norm at most 1. Then,*

$$\underset{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}{\boldsymbol{E}}\left[|\mathbf{Shadow}_{\boldsymbol{t},\boldsymbol{z}}\left(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n\right)|\right] \leq \mathcal{D}(n,d,\sigma), \tag{10}$$

where

$$\mathcal{D}(n,d,\sigma) = \frac{58,888,678\ nd^3}{\min\left(\sigma, 1/3\sqrt{d\ln n}\right)^6},$$

and $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ have density $\prod_{i=1}^n \mu_i(\boldsymbol{a}_i)$.

The proof of Theorem 4.0.1, will use the following definitions.

Definition 4.0.2 (ang) *For a vector $\boldsymbol{q}$ and a set S, we define*

$$\mathbf{ang}\left(\boldsymbol{q}, S\right) = \min_{\boldsymbol{x}\in S} \mathbf{angle}\left(\boldsymbol{q}, \boldsymbol{x}\right),$$

If S is empty, we set $\mathbf{ang}\left(\boldsymbol{q}, \emptyset\right) = \infty$.

Definition 4.0.3 ($\mathbf{ang}_{\boldsymbol{q}}$) *For a vector $\boldsymbol{q}$ and points $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ in $\mathbb{R}^d$, we define*

$$\mathbf{ang}_{\boldsymbol{q}}\left(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n\right) = \mathbf{ang}\left(\boldsymbol{q}, \partial \triangle \left(\mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n)\right)\right),$$

where $\partial\triangle\left(\mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n)\right)$ denotes the boundary of the simplex $\triangle\left(\mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n)\right)$.

These definitions are arranged so that if the ray through $\boldsymbol{q}$ does not pierce the convex hull of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, then $\mathbf{ang}_{\boldsymbol{q}}\left(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n\right) = \infty$.

In our proofs, we will make frequent use of the fact that it is very unlikely that a Gaussian random variable is far from its mean. To capture this fact, we define:

Definition 4.0.4 (P) *P is the set of $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ for which $\|\boldsymbol{a}_i\| \leq 2$, for all i.*

Applying a union bound to Corollary 2.4.6, we obtain

Proposition 4.0.5 (Measure of P)

$$\boldsymbol{Pr}\left[(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n) \in P\right] \geq 1 - n(n^{-2.9d}) = 1 - n^{-2.9d+1}.$$

Proof of Theorem 4.0.1 We first observe that we can assume $\sigma \leq 1/3\sqrt{d \ln n}$: if $\sigma > 1/3\sqrt{d \ln n}$, then we can scale down all the data until $\sigma = 1/3\sqrt{d \ln n}$. As this could only decrease the norms of the centers of the distributions, the theorem statement would be unaffected.

Assume without loss of generality that $\boldsymbol{z}$ and $\boldsymbol{t}$ are orthogonal. Let

$$\boldsymbol{q}_\theta = \boldsymbol{z} \sin(\theta) + \boldsymbol{t} \cos(\theta). \tag{11}$$

We discretize the problem by using the intuitively obvious fact, which we prove as Lemma 4.0.6, that the left-hand of (10) equals

$$\lim_{m \to \infty} \mathop{\mathbf{E}}_{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n} \left[\left| \bigcup_{\theta \in \left\{ \frac{2\pi}{m}, \frac{2 \cdot 2\pi}{m}, \ldots, \frac{m \cdot 2\pi}{m} \right\}} \left\{ \mathbf{optSimp}_{\boldsymbol{q}_\theta}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) \right\} \right| \right].$$

Let E_i denote the event

$$\left[\mathbf{optSimp}_{\boldsymbol{q}_{2\pi i/m}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) \neq \mathbf{optSimp}_{\boldsymbol{q}_{2\pi((i+1) \mod m)/m}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) \right].$$

Then, for any $m \geq 2$ and for all $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$,

$$\left| \bigcup_{\theta \in \left\{ \frac{2\pi}{m}, \frac{2 \cdot 2\pi}{m}, \ldots, \frac{m \cdot 2\pi}{m} \right\}} \left\{ \mathbf{optSimp}_{\boldsymbol{q}_\theta}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) \right\} \right| = \sum_{i=1}^{m} E_i(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n).$$

We bound this sum by

$$\begin{aligned}
\mathbf{E}\left[\sum_{i=1}^{m} E_i\right] &= \mathop{\mathbf{E}}_{P}\left[\sum_i E_i\right] \mathbf{Pr}\left[P\right] + \mathop{\mathbf{E}}_{\bar{P}}\left[\sum_i E_i\right] \mathbf{Pr}\left[\bar{P}\right] \\
&\leq \mathop{\mathbf{E}}_{P}\left[\sum_i E_i\right] + \binom{n}{d} n^{-2.9d+1} \\
&\leq \mathop{\mathbf{E}}_{P}\left[\sum_i E_i\right] + 1
\end{aligned}$$

Thus, we will focus on bounding $\mathbf{E}_P\left[\sum_i E_i\right]$.

Observing that E_i implies $\left[\mathbf{ang}_{\boldsymbol{q}_{2\pi i/m}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) \leq 2\pi/m\right]$, and applying linearity of expectation, we obtain

$$\begin{aligned}
\mathop{\mathbf{E}}_{P}\left[\sum_i E_i\right] &= \sum_{i=1}^{m} \mathop{\mathbf{Pr}}_{P}\left[E_i\right] \\
&\leq \sum_{i=1}^{m} \mathop{\mathbf{Pr}}_{P}\left[\mathbf{ang}_{\boldsymbol{q}_{2\pi i/m}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) < \frac{2\pi}{m}\right] \\
&\leq 2\pi \frac{9,372,424\ nd^3}{\sigma^6} && \text{by Lemma 4.0.7,} \\
&\leq \frac{58,888,677\ nd^3}{\sigma^6}.
\end{aligned}$$

■

Lemma 4.0.6 (Discretization in limit) *Let $\boldsymbol{z}$ and $\boldsymbol{t}$ be orthogonal vectors in $\mathbb{R}^d$, and let $\mu_1, \ldots, \mu_n$ be non-degenerate Gaussian distributions. Then,*

$$\mathop{\boldsymbol{E}}_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\left|\bigcup_{\boldsymbol{q}\in \boldsymbol{Span}(\boldsymbol{z},\boldsymbol{t})}\left\{\mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n)\right\}\right|\right] = \lim_{m\to\infty}\mathop{\boldsymbol{E}}_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\left|\bigcup_{\theta\in\left\{\frac{2\pi}{m},\frac{2\cdot 2\pi}{m},\ldots,\frac{m\cdot 2\pi}{m}\right\}}\left\{\mathbf{optSimp}_{\boldsymbol{q}_\theta}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n)\right\}\right|\right], \quad (12)$$

where $\boldsymbol{q}_\theta$ is as defined in (11).

Proof For a $I \in \binom{[n]}{d}$, let

$$F_I(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n) = \int_\theta \left[\mathbf{optSimp}_{\boldsymbol{q}_\theta}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n) = I\right]\, d\theta\,.$$

The left and right hand sides of (12) can differ only if there exists a $\delta > 0$ such that for all $\epsilon > 0$,

$$\Pr_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\exists I \middle| \begin{array}{l} I = \mathbf{optSimp}_{\boldsymbol{q}_\theta}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n) \text{ for some } \theta, \text{ and} \\ F_I(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n) < \epsilon \end{array}\right] \geq \delta.$$

As there are only finitely many choices for I, this would imply the existence of a δ' and a particular I such that for all $\epsilon > 0$,

$$\Pr_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\begin{array}{l} I = \mathbf{optSimp}_{\boldsymbol{q}_\theta}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n) \text{ for some } \theta, \text{ and} \\ F_I(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n) < \epsilon \end{array}\right] \geq \delta'.$$

As $F_I(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n) = F_I(\boldsymbol{A}_I)$ given that $I = \mathbf{optSimp}_{\boldsymbol{q}_\theta}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n)$ for some θ, this implies that for all $\epsilon > 0$,

$$\Pr_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\begin{array}{l} I = \mathbf{optSimp}_{\boldsymbol{q}_\theta}(\boldsymbol{A}_I) \text{ for some } \theta, \text{ and} \\ F_I(\boldsymbol{A}_I) < \epsilon \end{array}\right] \geq \delta'. \quad (13)$$

Note that $I = \mathbf{optSimp}_{\boldsymbol{q}_\theta}(\boldsymbol{A}_I)$ if and only if $\boldsymbol{q}_\theta \in \mathbf{Cone}\,(\boldsymbol{A}_I)$. Now, let

$$G(\boldsymbol{A}_I) = \int_\theta \left[\boldsymbol{q}_\theta \in \mathbf{Cone}\,(\boldsymbol{A}_I)\right] \left(\mathbf{ang}\,(\boldsymbol{q}_\theta, \partial\triangle\,(\boldsymbol{A}_I))\,/\pi\right)\, d\theta\,.$$

As $G(\boldsymbol{A}_I) \leq F_I(\boldsymbol{A}_I)$, (13) implies that for all $\epsilon > 0$

$$\Pr_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\begin{array}{l} I = \mathbf{optSimp}_{\boldsymbol{q}_\theta}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n) \text{ for some } \theta, \text{ and} \\ G(\boldsymbol{A}_I) < \epsilon \end{array}\right] \geq \delta'.$$

However, G is a continuous function, and therefore measurable, so this would imply

$$\Pr_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\begin{array}{l} I = \mathbf{optSimp}_{\boldsymbol{q}_\theta}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n) \text{ for some } \theta, \text{ and} \\ G(\boldsymbol{A}_I) = 0 \end{array}\right] \geq \delta',$$

which is clearly false as the set of $\boldsymbol{A}_I$ satisfying

- $G(\boldsymbol{A}_I) = 0$, and
- $\exists \theta : \mathbf{optSimp}_{\boldsymbol{q}_\theta}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = \{\boldsymbol{A}_I\}$

has co-dimension 1, and so has measure zero under the product distribution of non-degenerate Gaussians. ■

Lemma 4.0.7 (Angle bound) *Let $d \geq 3$ and $n > d$. Let $\boldsymbol{q}$ be any unit vector and let $\mu_1, \ldots, \mu_n$ be Gaussian measures in $\mathbb{R}^d$ of standard deviation $\sigma \leq 1/3\sqrt{d \ln n}$ centered at points of norm at most 1. Then,*

$$\Pr_P\left[\boldsymbol{ang}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) < \epsilon\right] \leq \frac{9,372,424\ nd^3}{\sigma^6}\epsilon$$

where $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ have density

$$\prod_{i=1}^{n} \mu_i(\boldsymbol{a}_i).$$

The proof will make use of the following definition:

Definition 4.0.8 ($\boldsymbol{P_I^j}$) *For a $I \in \binom{[n]}{d}$ and $j \in I$, we define P_I^j to be the set of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d$ satisfying*

(1) For all $\boldsymbol{q}$, if $\mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) \neq \emptyset$, then $s \leq 2$, where s is the real number for which $s\boldsymbol{q} \in \triangle\left(\mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)\right)$,

(2) $\mathbf{dist}\left(\boldsymbol{a}_i, \boldsymbol{a}_k\right) \leq 4$, for $i, k \in I - \{j\}$,

(3) $\mathbf{dist}\left(\boldsymbol{a}_j, \mathbf{Aff}\left(\boldsymbol{A}_{I-\{j\}}\right)\right) \leq 4$, and

(4) $\mathbf{dist}\left(\boldsymbol{a}_j^{\perp}, \boldsymbol{a}_i\right) \leq 4$, for all $i \in I - \{j\}$, where $\boldsymbol{a}_j^{\perp}$ is the orthogonal projection of $\boldsymbol{a}_j$ onto $\mathbf{Aff}\left(\boldsymbol{A}_{I-\{j\}}\right)$.

Proposition 4.0.9 ($\boldsymbol{P \subset P_I^j}$) *For all j, I, $P \subset P_I^j$.*

Proof Parts (2), (3), and (4) follow immediately from the restrictions $\|\boldsymbol{a}_i\| \leq 2$. To see why part (1) is true, note that $s\boldsymbol{q}$ lies in the convex hull of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, and so its norm, s, can be at most $\max_i \|\boldsymbol{a}_i\| \leq 2$, for $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) \in P$. ■

Proof of Lemma 4.0.7 Applying a union bound twice, we write

$$\begin{aligned}
&\Pr_P\left[\mathbf{ang}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) < \epsilon\right] \\
&\leq \sum_I \Pr_P\left[\begin{array}{l}\mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = I \text{ and} \\ \mathbf{ang}(\boldsymbol{q}, \partial \triangle (\boldsymbol{A}_I)) < \epsilon\end{array}\right] \\
&\leq \sum_I \sum_{j=1}^{d} \Pr_P\left[\begin{array}{l}\mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = I \text{ and} \\ \mathbf{ang}(\boldsymbol{q}, \triangle\left(\boldsymbol{A}_{I-\{j\}}\right)) < \epsilon\end{array}\right] \\
&\leq \sum_I \sum_{j=1}^{d} \Pr_{P_I^j}\left[\begin{array}{l}\mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = I \text{ and} \\ \mathbf{ang}(\boldsymbol{q}, \triangle\left(\boldsymbol{A}_{I-\{j\}}\right)) < \epsilon\end{array}\right] \Big/ \Pr_{P_I^j}[P]
\end{aligned}$$

(by Proposition 2.3.2)

$$\leq \sum_{I} \sum_{j=1}^{d} \mathop{\mathbf{Pr}}_{P_I^j} \left[\begin{array}{l} \mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = I \text{ and} \\ \mathbf{ang}(\boldsymbol{q}, \triangle\left(\boldsymbol{A}_{I-\{j\}}\right)) < \epsilon \end{array} \right] \Big/ \mathbf{Pr}\left[P\right]$$

(by $P \subset P_I^j$)

$$\leq \frac{1}{1 - n^{-2.9d+1}} \sum_{I} \sum_{j=1}^{d} \mathop{\mathbf{Pr}}_{P_I^j} \left[\begin{array}{l} \mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = I \text{ and} \\ \mathbf{ang}(\boldsymbol{q}, \triangle\left(\boldsymbol{A}_{I-\{j\}}\right)) < \epsilon \end{array} \right]$$

(by Proposition 4.0.5)

$$\leq \frac{1}{1 - n^{-2.9d+1}} \sum_{j=1}^{d} \sum_{I} \mathop{\mathbf{Pr}}_{P_I^j} \left[\begin{array}{l} \mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = I \text{ and} \\ \mathbf{ang}(\boldsymbol{q}, \triangle\left(\boldsymbol{A}_{I-\{j\}}\right)) < \epsilon, \end{array} \right],$$

by changing the order of summation.

We now expand the inner summation using Bayes' rule to get

$$\begin{aligned} &\sum_{I} \mathop{\mathbf{Pr}}_{P_I^j} \left[\begin{array}{l} \mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = I \text{ and} \\ \mathbf{ang}(\boldsymbol{q}, \triangle\left(\boldsymbol{A}_{I-\{j\}}\right)) < \epsilon \end{array} \right] \\ &\qquad = \sum_{I} \mathop{\mathbf{Pr}}_{P_I^j} \left[\mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = I\right] \cdot \qquad (14) \\ &\qquad\qquad \mathop{\mathbf{Pr}}_{P_I^j} \left[\begin{array}{l} \mathbf{ang}(\boldsymbol{q}, \triangle\left(\boldsymbol{A}_{I-\{j\}}\right)) < \epsilon \big| \\ \qquad \mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = I \end{array} \right] \end{aligned}$$

As $\mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ is a set of size zero or one with probability 1,

$$\sum_{I} \mathbf{Pr}\left[\mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = I\right] \leq 1;$$

from which we derive

$$\begin{aligned} &\sum_{I} \mathop{\mathbf{Pr}}_{P_I^j} \left[\mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = I\right] \\ &\leq \sum_{I} \mathbf{Pr}\left[\mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = I\right] / \mathbf{Pr}\left[P_I^j\right] \end{aligned}$$

(by Proposition 2.3.2)

$$\leq \frac{1}{1 - n^{-2.9d+1}} \sum_{I} \mathbf{Pr}\left[\mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = I\right]$$

(by $P \subset P_I^j$ and Proposition 4.0.5)

$$\leq \frac{1}{1 - n^{-2.9d+1}}.$$

So,

$$(14) \leq \frac{1}{1-n^{-2.9d+1}} . \max_I \mathop{\mathbf{Pr}}_{P_I^j} \left[\begin{array}{l} \mathbf{ang}(\boldsymbol{q}, \triangle\left(\boldsymbol{A}_{I-\{j\}}\right)) < \epsilon \Big| \\ \qquad \mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = I \end{array} \right].$$

Plugging this bound in to the first inequality derived in the proof, we obtain the bound of

$$\begin{aligned} &\mathop{\mathbf{Pr}}_{P}\left[\mathbf{ang}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) < \epsilon\right] \\ &\leq \frac{d}{(1-n^{-2.9d+1})^2} \max_{j,I} \mathop{\mathbf{Pr}}_{P_I^j} \left[\begin{array}{l} \mathbf{ang}(\boldsymbol{q}, \triangle\left(\boldsymbol{A}_{I-\{j\}}\right)) < \epsilon \Big| \\ \qquad \mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = I \end{array} \right] \\ &\leq d\frac{9,372,424\ nd^3}{\sigma^6}\epsilon, \text{ by Lemma 4.0.11, } d \geq 3 \text{ and } n \geq d+1, \\ &= \frac{9,372,424\ nd^3}{\sigma^6}\epsilon. \end{aligned}$$

■

Definition 4.0.10 (Q) *We define Q to be the set of $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d) \in \mathbb{R}^{d-1}$ satisfying*

(1) $\mathbf{dist}\left(\boldsymbol{b}_1, \mathbf{Aff}\left(\boldsymbol{b}_2, \ldots, \boldsymbol{b}_d\right)\right) \leq 4$,

(2) $\mathbf{dist}\left(\boldsymbol{b}_i, \boldsymbol{b}_j\right) \leq 4$ *for all* $i, j \geq 2$,

(3) $\mathbf{dist}\left(\boldsymbol{b}_1^{\perp}, \boldsymbol{b}_i\right) \leq 4$ *for all* $i \geq 2$, *where* $\boldsymbol{b}_1^{\perp}$ *is the orthogonal projection of* $\boldsymbol{b}_1$ *onto* $\mathbf{Aff}\left(\boldsymbol{b}_2, \ldots, \boldsymbol{b}_d\right)$, *and*

(4) $\mathbf{0} \in \triangle\left(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d\right)$.

Lemma 4.0.11 (Angle bound given optSimp) *Let $\mu_1, \ldots, \mu_n$ be Gaussian measures in $\mathbb{R}^d$ of standard deviation $\sigma \leq 1/3\sqrt{d \ln n}$ centered at points of norm at most 1. Then*

$$\mathop{\boldsymbol{Pr}}_{P^1_{1,\ldots,d}} \left[\begin{array}{l} \boldsymbol{ang}(\boldsymbol{q}, \triangle\left(\boldsymbol{a}_2, \ldots, \boldsymbol{a}_d\right)) < \epsilon \Big| \\ \qquad \mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = \{1, \ldots, d\} \end{array} \right] \leq \frac{9,371,990\ nd^2\epsilon}{\sigma^6} \tag{15}$$

where $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ have density

$$\prod_{i=1}^{n} \mu_i(\boldsymbol{a}_i).$$

Proof We begin by making the change of variables from $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d$ to $\boldsymbol{\omega}, s, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_d$ described in Corollary 2.5.3, and we recall that the Jacobian of this change of variables is

$$(d-1)!\left\langle \boldsymbol{\omega} | \boldsymbol{q} \right\rangle \mathbf{Vol}\left(\triangle\left(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d\right)\right).$$

As this change of variables is arranged so that $s\boldsymbol{q} \in \triangle\left(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d\right)$ if and only if $\mathbf{0} \in \triangle\left(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d\right)$, the condition that $\mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = \{1, \ldots, d\}$ can be expressed as

$$\left[\mathbf{0} \in \triangle\left(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d\right)\right] \prod_{j>d} \left[\left\langle \boldsymbol{\omega} | \boldsymbol{a}_j \right\rangle \leq \left\langle \boldsymbol{\omega} | s\boldsymbol{q} \right\rangle\right].$$

Let $\boldsymbol{x}$ be any point on $\triangle(\boldsymbol{a}_2, \ldots, \boldsymbol{a}_d)$. Given that $s\boldsymbol{q} \in \triangle(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d)$, conditions (3) and (4) for membership in $P^1_{1,\ldots,d}$ imply that

$$\mathbf{dist}(s\boldsymbol{q}, \boldsymbol{x}) \leq \mathbf{dist}(\boldsymbol{a}_1, \boldsymbol{x}) \leq \sqrt{\mathbf{dist}(\boldsymbol{a}_1, \mathbf{Aff}(\boldsymbol{a}_2, \ldots, \boldsymbol{a}_d))^2 + \mathbf{dist}\left(\boldsymbol{a}_1^{\perp}, \boldsymbol{x}\right)^2} \leq 4\sqrt{2},$$

where $\boldsymbol{a}_1^{\perp}$ is the orthogonal projection of $\boldsymbol{a}_1$ onto $\mathbf{Aff}(\boldsymbol{a}_2, \ldots, \boldsymbol{a}_d)$. So, Lemma 4.0.12 implies

$$\mathbf{ang}(\boldsymbol{q}, \triangle(\boldsymbol{a}_2, \ldots, \boldsymbol{a}_d)) \geq \frac{\mathbf{dist}(s\boldsymbol{q}, \mathbf{Aff}(\boldsymbol{a}_2, \ldots, \boldsymbol{a}_d))\, \langle \boldsymbol{\omega} | \boldsymbol{q} \rangle}{2 + 4\sqrt{2}} = \frac{\mathbf{dist}(\boldsymbol{0}, \mathbf{Aff}(\boldsymbol{b}_2, \ldots, \boldsymbol{b}_d))\, \langle \boldsymbol{\omega} | \boldsymbol{q} \rangle}{2 + 4\sqrt{2}}.$$

Finally, observe that $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d) \in P^1_{1,\ldots,d}$ is equivalent to the conditions $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d) \in Q$ and $s \leq 2$, given that $\mathbf{optSimp}_{\boldsymbol{q}}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d) = \{1, \ldots, d\}$. Now, the left-hand side of (15) can be bounded by

$$\Pr_{\substack{\boldsymbol{\omega}, s \leq 2 \\ (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d) \in Q}} \left[\frac{\mathbf{dist}(\boldsymbol{0}, \mathbf{Aff}(\boldsymbol{b}_2, \ldots, \boldsymbol{b}_d))\, \langle \boldsymbol{\omega} | \boldsymbol{q} \rangle}{2 + 4\sqrt{2}} < \epsilon \right], \tag{16}$$

where the variables have density proportional to

$$\langle \boldsymbol{\omega} | \boldsymbol{q} \rangle\, \mathbf{Vol}(\triangle(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)) \left(\prod_{j>d} \int_{\boldsymbol{a}_j} [\langle \boldsymbol{\omega} | \boldsymbol{a}_j \rangle \leq s \langle \boldsymbol{\omega} | \boldsymbol{q} \rangle]\, \mu_j(\boldsymbol{a}_j)\, d\boldsymbol{a}_j \right) \prod_{i=1}^{d} \mu_i(\boldsymbol{R}_{\boldsymbol{\omega}} \boldsymbol{b}_i + s\boldsymbol{q}).$$

As Lemma 4.1.1 implies

$$\Pr_{\substack{\boldsymbol{\omega}, s \leq 2 \\ (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d) \in Q}} [\mathbf{dist}(\boldsymbol{0}, \mathbf{Aff}(\boldsymbol{b}_2, \ldots, \boldsymbol{b}_d)) < \epsilon] \leq \frac{900 e^{2/3} d^2 \epsilon}{\sigma^4},$$

and Lemma 4.2.1 implies

$$\max_{s \leq 2, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_d \in Q} \Pr_{\boldsymbol{\omega}} [\langle \boldsymbol{\omega} | \boldsymbol{q} \rangle < \epsilon] < \left(\frac{340 n \epsilon}{\sigma^2} \right)^2,$$

we can apply Lemma 2.3.5 to prove

$$(16) \leq 4 \cdot (2 + 4\sqrt{2}) \cdot \left(\frac{900 e^{2/3} d^2}{\sigma^4} \right) \left(\frac{340 n}{\sigma^2} \right) \epsilon \leq \frac{9,371,990\ n d^2 \epsilon}{\sigma^6}.$$

■

Lemma 4.0.12 (Division into distance and angle) *Let $\boldsymbol{x}$ be a vector, let $0 < s \leq 2$, and let $\boldsymbol{q}$ and $\boldsymbol{\omega}$ be unit vectors satisfying*

(a) $\langle \boldsymbol{\omega} | \boldsymbol{x} - s\boldsymbol{q} \rangle = 0$, *and*

(b) $\mathbf{dist}(\boldsymbol{x}, s\boldsymbol{q}) \leq 4\sqrt{2}$.

Then,

$$\mathbf{angle}\,(\boldsymbol{q},\boldsymbol{x}) \geq \frac{\mathbf{dist}\,(\boldsymbol{x},s\boldsymbol{q})\,\langle\boldsymbol{\omega}|\boldsymbol{q}\rangle}{2+4\sqrt{2}}.$$

Proof Let $\boldsymbol{r} = \boldsymbol{x} - s\boldsymbol{q}$. Then, (a) implies

$$\langle\boldsymbol{\omega}|\boldsymbol{q}\rangle^2 + \left\langle \frac{\boldsymbol{r}}{\|\boldsymbol{r}\|}\middle|\boldsymbol{q}\right\rangle^2 \leq \|\boldsymbol{q}\| = 1;$$

so,

$$\langle\boldsymbol{r}|\boldsymbol{q}\rangle \leq \sqrt{1-\langle\boldsymbol{\omega}|\boldsymbol{q}\rangle^2}\,\|\boldsymbol{r}\|\,.$$

Let h be the distance from $\boldsymbol{x}$ to the ray through $\boldsymbol{q}$. Then,

$$h^2 + \langle\boldsymbol{r}|\boldsymbol{q}\rangle^2 = \|\boldsymbol{r}\|^2\,;$$

so,

$$h \geq \langle\boldsymbol{\omega}|\boldsymbol{q}\rangle\,\|\boldsymbol{r}\| = \langle\boldsymbol{\omega}|\boldsymbol{q}\rangle\,\mathbf{dist}\,(\boldsymbol{x},s\boldsymbol{q})$$

Now,

$$\mathbf{angle}\,(\boldsymbol{q},\boldsymbol{x}) \geq \sin(\mathbf{angle}\,(\boldsymbol{q},\boldsymbol{x})) = \frac{h}{\|\boldsymbol{x}\|} \geq \frac{h}{s+\mathbf{dist}\,(\boldsymbol{x},s\boldsymbol{q})} \geq \frac{h}{2+4\sqrt{2}} \geq \frac{\langle\boldsymbol{\omega}|\boldsymbol{q}\rangle\,\mathbf{dist}\,(\boldsymbol{x},s\boldsymbol{q})}{2+4\sqrt{2}}.$$

■

4.1 Distance

The goal of this section is to prove it is unlikely that $\mathbf{0}$ is near $\partial\triangle(\boldsymbol{b}_1,\ldots,\boldsymbol{b}_d)$.

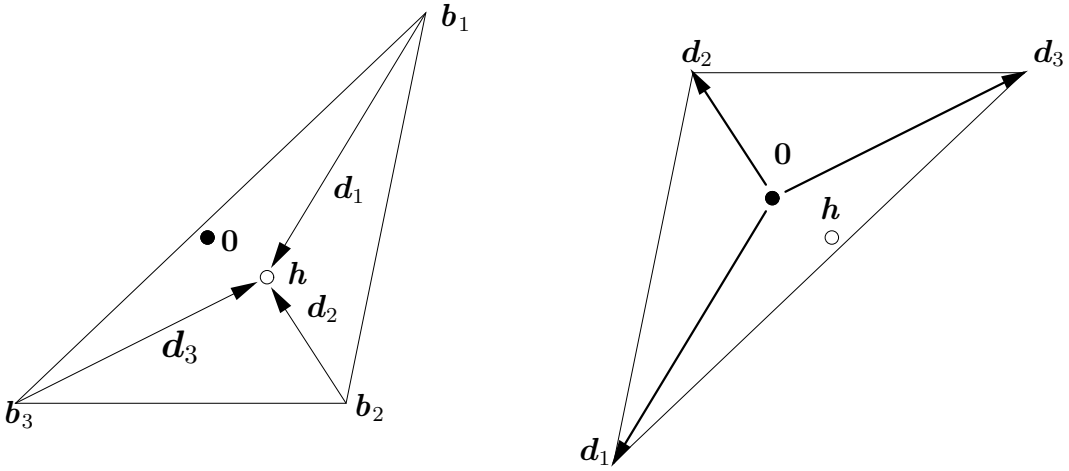

Figure 3: The change of variables in Lemma 4.1.2.

Lemma 4.1.1 (Distance bound) *Let $\boldsymbol{q}$ be a unit vector and let $\mu_1, \ldots, \mu_n$ be Gaussian measures in $\mathbb{R}^d$ of standard deviation $\sigma \leq 1/3\sqrt{d \ln n}$ centered at points of norm at most 1. Then,*

$$\underset{\substack{\boldsymbol{\omega}, s \leq 2 \\ (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d) \in Q}}{\boldsymbol{Pr}} [\mathbf{dist}\,(\mathbf{0}, \mathbf{Aff}\,(\boldsymbol{b}_2, \ldots, \boldsymbol{b}_d)) < \epsilon] \leq \frac{900 e^{2/3} d^2 \epsilon}{\sigma^4}, \tag{17}$$

where the variables have density proportional to

$$\langle \boldsymbol{\omega} | \boldsymbol{q} \rangle \, \mathbf{Vol}\,(\triangle\,(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)) \left(\prod_{j>d} \int_{\boldsymbol{a}_j} [\langle \boldsymbol{\omega} | \boldsymbol{a}_j \rangle \leq s \, \langle \boldsymbol{\omega} | \boldsymbol{q} \rangle] \, \mu_j(\boldsymbol{a}_j) \, d\boldsymbol{a}_j \right) \prod_{i=1}^{d} \mu_i(\boldsymbol{R}_{\boldsymbol{\omega}} \boldsymbol{b}_i + s\boldsymbol{q}).$$

Proof Note that if we fix $\boldsymbol{\omega}$ and s, then the first and third terms in the density become constant. For any fixed plane specified by $(\boldsymbol{\omega}, s)$, Proposition 2.4.3 tells us that the induced density on $\boldsymbol{b}_i$ remains a Gaussian of standard deviation σ and is centered at the projection of the center of μ_i onto the plane. As the origin of this plane is the point $s\boldsymbol{q}$, and $s \leq 2$, these induced Gaussians have centers of norm at most 3. Thus, we can use Lemma 4.1.2 to bound the left-hand side of (17) by

$$\max_{\boldsymbol{\omega}, s \leq 2} \underset{(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d) \in Q}{\mathbf{Pr}} [\mathbf{dist}\,(\mathbf{0}, \mathbf{Aff}\,(\boldsymbol{b}_2, \ldots, \boldsymbol{b}_d)) < \epsilon] \leq \frac{900 e^{2/3} d^2 \epsilon}{\sigma^4}.$$

■

Lemma 4.1.2 (Distance bound in plane) *Let $\mu_1, \ldots, \mu_d$ be Gaussian measures in $\mathbb{R}^{d-1}$. of standard deviation $\sigma \leq 1/3\sqrt{d \ln n}$ centered at points of norm at most 3. Then*

$$\underset{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d \in Q}{\boldsymbol{Pr}} [\mathbf{dist}\,(\mathbf{0}, \mathbf{Aff}\,(\boldsymbol{b}_2, \ldots, \boldsymbol{b}_d)) < \epsilon] \leq \frac{900 e^{2/3} d^2 \epsilon}{\sigma^4}, \tag{18}$$

where $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d$ have density proportional to

$$\mathbf{Vol}\,(\triangle\,(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)) \prod_{i=1}^{d} \mu_i(\boldsymbol{b}_i).$$

Proof In Lemma 4.1.3, we will prove it is unlikely that $\boldsymbol{b}_1$ is close to $\mathbf{Aff}\,(\boldsymbol{b}_2, \ldots, \boldsymbol{b}_d)$. We will exploit this fact by proving that it is unlikely that $\mathbf{0}$ is much closer than $\boldsymbol{b}_1$ to $\mathbf{Aff}\,(\boldsymbol{b}_2, \ldots, \boldsymbol{b}_d)$. We do this by fixing the shape of $\triangle\,(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)$, and then considering slight translations of this simplex. That is, we make a change of variables to

$$\begin{aligned} \boldsymbol{h} &= \frac{1}{d} \sum_{i=1}^{d} \boldsymbol{b}_i \\ \boldsymbol{d}_i &= \boldsymbol{h} - \boldsymbol{b}_i, \text{ for } i \geq 2. \end{aligned}$$

The vectors $\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d$ specify the shape of the simplex, and $\boldsymbol{h}$ specifies its location. As this change of variables is a linear transformation, its Jacobian is constant. For convenience, we also define $\boldsymbol{d}_1 = \boldsymbol{h} - \boldsymbol{b}_1 = -\sum_{i \geq 2} \boldsymbol{d}_i$.

It is easy to verify that

$$\begin{aligned}
\mathbf{0} \in \triangle(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d) &\Leftrightarrow \boldsymbol{h} \in \triangle(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d), \\
\mathbf{dist}(\mathbf{0}, \mathbf{Aff}(\boldsymbol{b}_2, \ldots, \boldsymbol{b}_d)) &= \mathbf{dist}(\boldsymbol{h}, \mathbf{Aff}(\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d)), \\
\mathbf{dist}(\boldsymbol{b}_1, \mathbf{Aff}(\boldsymbol{b}_2, \ldots, \boldsymbol{b}_d)) &= \mathbf{dist}(\boldsymbol{d}_1, \mathbf{Aff}(\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d)), \text{ and} \\
\mathbf{Vol}(\triangle(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)) &= \mathbf{Vol}(\triangle(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d)).
\end{aligned}$$

Note that the relation between $\boldsymbol{d}_1$ and $\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d$ guarantees $\mathbf{0} \in \triangle(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d)$ for all $\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d$. So, $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d) \in Q$ if and only if $(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d) \in Q$ and $h \in \triangle(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d)$. As $\boldsymbol{d}_1$ is a function of $\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d$, we let Q' be the set of $\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d$ for which $(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d) \in Q$.

So, the left-hand side of (18) equals

$$\mathop{\mathbf{Pr}}_{\substack{(\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d) \in Q' \\ \boldsymbol{h} \in \triangle(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d)}} [\mathbf{dist}(\boldsymbol{h}, \mathbf{Aff}(\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d)) < \epsilon]$$

where $\boldsymbol{h}, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_d$ have density proportional to

$$\mathbf{Vol}(\triangle(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d)) \prod_{i=1}^{d} \mu_i(\boldsymbol{h} - \boldsymbol{d}_i). \tag{19}$$

Similarly, Lemma 4.1.3 can be seen to imply

$$\mathop{\mathbf{Pr}}_{\substack{(\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d) \in Q' \\ \boldsymbol{h} \in \triangle(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d)}} [\mathbf{dist}(\boldsymbol{d}_1, \mathbf{Aff}(\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d)) < \epsilon] \leq \left(\frac{\epsilon 3 e^{2/3} d}{\sigma^2} \right)^3 \leq \left(\frac{\epsilon 3 e^{2/3} d}{\sigma^2} \right)^2 \tag{20}$$

under density proportional to (19). We take advantage of (20) by proving

$$\max_{\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d \in Q'} \mathop{\mathbf{Pr}}_{\boldsymbol{h} \in \triangle(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d)} \left[\frac{\mathbf{dist}(\boldsymbol{h}, \mathbf{Aff}(\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d))}{\mathbf{dist}(\boldsymbol{d}_1, \mathbf{Aff}(\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d))} < \epsilon \right] < \frac{75 d \epsilon}{\sigma^2}, \tag{21}$$

where $\boldsymbol{h}$ has density proportional to

$$\prod_{i=1}^{d} \mu_i(\boldsymbol{h} - \boldsymbol{d}_i).$$

Before proving (21), we point out that using Lemma 2.3.5 to combine (20) and (21), we obtain

$$\mathop{\mathbf{Pr}}_{\substack{(\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d) \in Q' \\ \boldsymbol{h} \in \triangle(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d)}} [\mathbf{dist}(\boldsymbol{h}, \mathbf{Aff}(\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d) < \epsilon)] \leq \frac{900 e^{2/3} d^2 \epsilon}{\sigma^4},$$

from which the lemma follows.

To prove (21), we let

$$U_\epsilon = \left\{ \boldsymbol{h} \in \triangle(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d) : \frac{\mathbf{dist}(\boldsymbol{h}, \mathbf{Aff}(\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d))}{\mathbf{dist}(\boldsymbol{d}_1, \mathbf{Aff}(\boldsymbol{d}_2, \ldots, \boldsymbol{d}_d))} \geq \epsilon \right\},$$

and we set $\nu(\boldsymbol{h}) = \prod_{i=1}^{d} \mu_i(\boldsymbol{h} - \boldsymbol{d}_i)$. Under this notation, the probability in (21) is equal to

$$(\nu(U_0) - \nu(U_\epsilon))/\nu(U_0).$$

To bound this ratio, we construct an isomorphism from U_0 to U_ϵ. The natural isomorphism, which we denote Φ_ϵ, is the map that contracts the simplex by a factor of $(1-\epsilon)$ at $\boldsymbol{d}_1$. To use this isomorphism to compare the measures of the sets, we use the facts that for $\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d \in Q$ and $\boldsymbol{h} \in \triangle(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d)$,

(a) $\|\boldsymbol{h} - \boldsymbol{d}_i\| \leq \max_{i,j} \|\boldsymbol{d}_i - \boldsymbol{d}_j\| \leq 4\sqrt{2}$, so the distance from $\boldsymbol{h} - \boldsymbol{d}_i$ to the center of its distribution is at most $\|\boldsymbol{h} - \boldsymbol{d}_i\| + 3 \leq 4\sqrt{2} + 3$;

(b) $\mathbf{dist}(\boldsymbol{h}, \Phi_\epsilon(\boldsymbol{h})) \leq \epsilon \max_i \mathbf{dist}(\boldsymbol{d}_1, \boldsymbol{d}_i) \leq 4\sqrt{2}\epsilon$

to apply Lemma 2.4.2 to show that for all $\boldsymbol{h} \in \triangle(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d)$,

$$\frac{\mu_i(\Phi_\epsilon(\boldsymbol{h}) - \boldsymbol{d}_i)}{\mu_i(\boldsymbol{h} - \boldsymbol{d}_i)} \geq e^{-\frac{3 \cdot 4\sqrt{2}(4\sqrt{2}+3)\epsilon}{2\sigma^2}} = e^{-\frac{(48+18\sqrt{2})\epsilon}{\sigma^2}}.$$

So,

$$\min_{\boldsymbol{h} \in \triangle(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d)} \frac{\nu(\Phi_\epsilon(\boldsymbol{h}))}{\nu(\boldsymbol{h})} = \min_{\boldsymbol{h} \in \triangle(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d)} \prod_{i=1}^{d} \frac{\mu_i(\Phi_\epsilon(\boldsymbol{h}) - \boldsymbol{d}_i)}{\mu_i(\boldsymbol{h} - \boldsymbol{d}_i)} \geq e^{-\frac{(48+18\sqrt{2})d\epsilon}{\sigma^2}} \geq 1 - \frac{(48+18\sqrt{2})d\epsilon}{\sigma^2}. \tag{22}$$

As the Jacobian

$$\left| \frac{\partial \Phi_\epsilon(\boldsymbol{h})}{\partial \boldsymbol{h}} \right| = (1-\epsilon)^d \geq 1 - d\epsilon,$$

using the change of variables $\boldsymbol{x} = \Phi_\epsilon(\boldsymbol{h})$ we can compute

$$\nu(U_\epsilon) = \int_{\boldsymbol{x} \in U_\epsilon} \nu(\boldsymbol{x})\, d\boldsymbol{x} = \int_{\boldsymbol{h} \in U_0} \nu(\Phi_\epsilon(\boldsymbol{h})) \left| \frac{\partial \Phi_\epsilon(\boldsymbol{h})}{\partial \boldsymbol{h}} \right| d\boldsymbol{h} \geq (1 - d\epsilon) \int_{\boldsymbol{h} \in U_0} \nu(\Phi_\epsilon(\boldsymbol{h}))\, d\boldsymbol{h}\,. \tag{23}$$

So,

$$\begin{aligned}
\frac{\nu(U_\epsilon)}{\nu(U_0)} &\geq \frac{(1-d\epsilon) \int_{\boldsymbol{h} \in U_0} \nu(\Phi_\epsilon(\boldsymbol{h}))\, d\boldsymbol{h}}{\int_{\boldsymbol{h} \in U_0} \nu(\boldsymbol{h})\, d\boldsymbol{h}} && \text{by (23)} \\
&\geq (1-d\epsilon) \left(\min_{\boldsymbol{h} \in \triangle(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_d)} \frac{\nu(\Phi_\epsilon(\boldsymbol{h}))}{\nu(\boldsymbol{h})} \right) \frac{\int_{\boldsymbol{h} \in U_0} d\boldsymbol{h}}{\int_{\boldsymbol{h} \in U_0} d\boldsymbol{h}} \\
&\geq (1-d\epsilon) \left(1 - \frac{(48+18\sqrt{2})d\epsilon}{\sigma^2} \right) && \text{by (22)} \\
&\geq 1 - \frac{75 d\epsilon}{\sigma^2}, && \text{as } \sigma \leq 1.
\end{aligned}$$

(21) now follows from $(\nu(U_0) - \nu(U_\epsilon))/\nu(U_0) < \frac{75d\epsilon}{\sigma^2}$. ■

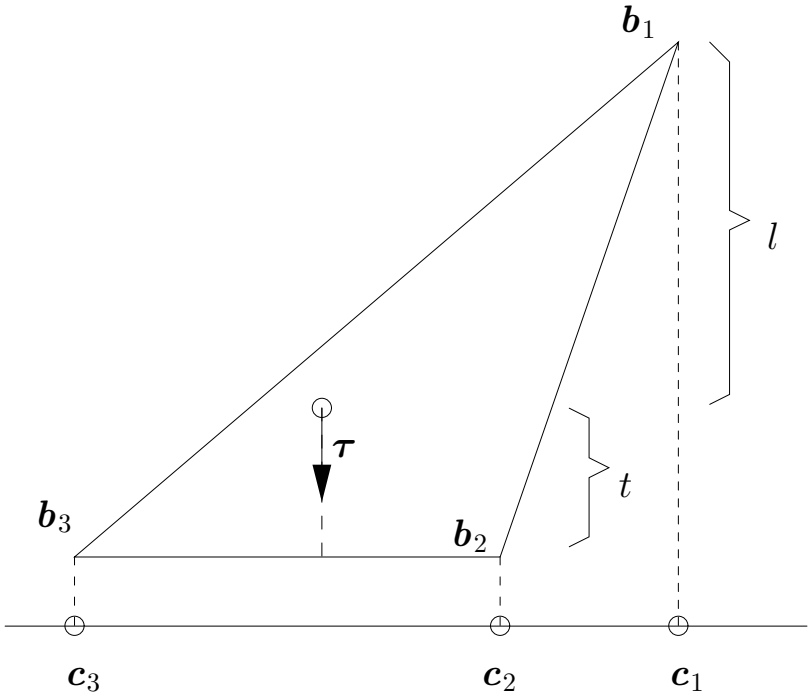

Figure 4: The change of variables in Lemma 4.1.3.

Lemma 4.1.3 (Height of simplex) *Let $\mu_1, \ldots, \mu_d$ be Gaussian measures in $\mathbb{R}^{d-1}$ of standard deviation $\sigma \leq 1/3\sqrt{d \ln n}$ centered at points of norm at most 3. Then*

$$\underset{\boldsymbol{b}_1,\ldots,\boldsymbol{b}_d \in Q}{\boldsymbol{Pr}} \left[\mathbf{dist}\left(\boldsymbol{b}_1, \mathbf{Aff}\left(\boldsymbol{b}_2, \ldots, \boldsymbol{b}_d\right) < \epsilon\right)\right] \leq \left(\frac{3\epsilon e^{2/3} d}{\sigma^2}\right)^3$$

where $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d$ have density proportional to

$$\mathbf{Vol}\left(\triangle\left(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d\right)\right) \prod_{i=1}^{d} \mu_i(\boldsymbol{b}_i).$$

Proof We begin with a simplifying change of variables. As in Theorem 2.5.2, we let

$$\left(\boldsymbol{b}_2, \ldots, \boldsymbol{b}_d\right) = \left(\boldsymbol{R}_{\boldsymbol{\tau}} \boldsymbol{c}_2 + t\boldsymbol{\tau}, \ldots, \boldsymbol{R}_{\boldsymbol{\tau}} \boldsymbol{c}_d + t\boldsymbol{\tau}\right),$$

where $\boldsymbol{\tau} \in S^{d-2}$ and $t \geq 0$ specify the plane through $\boldsymbol{b}_2, \ldots, \boldsymbol{b}_d$, and $\boldsymbol{c}_2, \ldots, \boldsymbol{c}_d \in \mathbb{R}^{d-2}$ denote the local coordinates of these points on that plane. Recall that the Jacobian of this change of variables is $\mathbf{Vol}\left(\triangle\left(\boldsymbol{c}_2, \ldots, \boldsymbol{c}_d\right)\right)$. Let $l = -\left\langle \boldsymbol{\tau} | \boldsymbol{b}_1 \right\rangle$, and let $\boldsymbol{c}_1$ denote the coordinates in $\mathbb{R}^{d-2}$ of the projection of $\boldsymbol{b}_1$ onto the plane specified by $\boldsymbol{\tau}$ and t. Note that $l \geq 0$. In this notation, we have

$$\mathbf{dist}\left(\boldsymbol{b}_1, \mathbf{Aff}\left(\boldsymbol{b}_2, \ldots, \boldsymbol{b}_d\right)\right) = l + t.$$

The Jacobian of the change from $\boldsymbol{b}_1$ to $(l, \boldsymbol{c}_1)$ is 1 as the transformation is just an orthogonal change of coordinates. The conditions for $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d) \in Q$ translate into the conditions

(a) $\mathbf{dist}(\boldsymbol{c}_i, \boldsymbol{c}_j) \leq 4$ for all $i \neq j$;

(b) $(l+t) \leq 4$; and

(c) $\mathbf{0} \in \triangle(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)$.

Let R denote the set of $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_d$ satisfying the first condition. As the lemma is vacuously true for $\epsilon \geq 4$, we will drop the second condition and note that doing so cannot decrease the probability that $(t+l) < \epsilon$. Thus, our goal is to bound

$$\mathop{\mathbf{Pr}}_{\boldsymbol{\tau}, t, l, (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_d) \in R} [(l+t) < \epsilon], \tag{24}$$

where the variables have density proportional to[2]

$$[\mathbf{0} \in \triangle(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)] \, \mathbf{Vol}(\triangle(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)) \, \mathbf{Vol}(\triangle(\boldsymbol{c}_2, \ldots, \boldsymbol{c}_d)) \prod_{i=1}^{d} \mu_i(\boldsymbol{b}_i).$$

As $\mathbf{Vol}(\triangle(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)) = (l+t)\mathbf{Vol}(\triangle(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_d))/d$, this is the same as having density proportional to

$$(l+t)\,[\mathbf{0} \in \triangle(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)] \, \mathbf{Vol}(\triangle(\boldsymbol{c}_2, \ldots, \boldsymbol{c}_d))^2 \prod_{i=1}^{d} \mu_i(\boldsymbol{b}_i).$$

Under a suitable system of coordinates, we can express $\boldsymbol{b}_1 = (-l, \boldsymbol{c}_1)$ and $\boldsymbol{b}_i = (t, \boldsymbol{c}_i)$ for $i \geq 2$. The key idea of this proof is that multiplying the first coordinates of these points by a constant does not change whether or not $\mathbf{0} \in \triangle(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)$; so, we can determine whether $\mathbf{0} \in \triangle(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)$ from the data $(l/t, \boldsymbol{c}_1, \ldots, \boldsymbol{c}_d)$. Thus, we will introduce a new variable α, set $l = \alpha t$, and let S denote the set of $(\alpha, \boldsymbol{c}_1, \ldots, \boldsymbol{c}_d)$ for which $\mathbf{0} \in \triangle(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)$ and $(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_d) \in R$. This change of variables from l to α incurs a Jacobian of $\frac{\partial l}{\partial \alpha} = t$, so (24) equals

$$\mathop{\mathbf{Pr}}_{\boldsymbol{\tau}, t, (\alpha, \boldsymbol{c}_1, \ldots, \boldsymbol{c}_d) \in S} [(1+\alpha)t < \epsilon],$$

where the variables have density proportional to

$$t^2(1+\alpha)\mathbf{Vol}(\triangle(\boldsymbol{c}_2, \ldots, \boldsymbol{c}_d))^2 \, \mu_1(-\alpha t, \boldsymbol{c}_1) \prod_{i=2}^{d} \mu_i(t, \boldsymbol{c}_i).$$

We upper bound this probability by

$$\max_{\boldsymbol{\tau}, (\alpha, \boldsymbol{c}_1, \ldots, \boldsymbol{c}_d) \in S} \mathop{\mathbf{Pr}}_{t} [(1+\alpha)t < \epsilon] \quad \leq \quad \max_{\boldsymbol{\tau}, (\alpha, \boldsymbol{c}_1, \ldots, \boldsymbol{c}_d) \in S} \mathop{\mathbf{Pr}}_{t} [\max(1, \alpha)t < \epsilon],$$

where t has density proportional to

$$t^2 \mu_1(-\alpha t, \boldsymbol{c}_1) \prod_{i=2}^{d} \mu_i(t, \boldsymbol{c}_i).$$

[2]While we keep terms such as $\boldsymbol{b}_1$ in the expression of the density, they should be interpreted as functions of $\boldsymbol{\tau}, t, l, \boldsymbol{c}_1, \ldots, \boldsymbol{c}_d$.

For $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_d$ fixed, the points $(-\alpha t, \boldsymbol{c}_1), (t, \boldsymbol{c}_2), \ldots, (t, \boldsymbol{c}_d)$ become univariate Gaussians of standard deviation σ and mean of absolute value at most 3. Let $t_0 = \sigma^2/(3\max(1,\alpha)d)$. Then, for t in the range $[0, t_0]$, $-\alpha t$ is at most $3 + \alpha t_0$ from the mean of the first distribution and t is at most $3 + t_0$ from the means of the other distributions. We will now observe that if t is restricted to a sufficiently small domain, then the densities of these Gaussians will have bounded variation. In particular, Lemma 2.4.2 implies that

$$\begin{aligned}
\frac{\max_{t\in[0,t_0]} \mu_1(-\alpha t, \boldsymbol{c}_1) \prod_{i=2}^d \mu_i(t, \boldsymbol{c}_i)}{\min_{t\in[0,t_0]} \mu_1(-\alpha t, \boldsymbol{c}_1) \prod_{i=2}^d \mu_i(t, \boldsymbol{c}_i)} &\leq e^{3(3+\alpha t_0)\alpha t_0/2\sigma^2} \prod_{i=2}^d e^{3(3+t_0)t_0/2\sigma^2} \\
&\leq e^{9\alpha t_0/2\sigma^2} \left(\prod_{i=2}^d e^{9t_0/2\sigma^2}\right) \cdot e^{3(\alpha t_0)^2/2\sigma^2} \left(\prod_{i=2}^d e^{3t_0^2/2\sigma^2}\right) \\
&\leq e^{3/2d} \left(\prod_{i=2}^d e^{3/2d}\right) \cdot e^{\sigma^2/6d^2} \left(\prod_{i=2}^d e^{\sigma^2/6d^2}\right) \\
&\leq e^{3/2} \cdot e^{1/6d} \\
&\leq e^2.
\end{aligned}$$

Thus, we can now apply Lemma 2.3.7 to show that

$$\Pr_t [t < \epsilon] < e^2 \left(\frac{3\epsilon(\max(1,\alpha)d}{\sigma^2}\right)^3,$$

from which we conclude

$$\Pr_t [\max(1,\alpha)t < \epsilon] < \left(\frac{3\epsilon e^{2/3} d}{\sigma^2}\right)^3.$$

■

4.2 Angle of $\boldsymbol{q}$ to $\boldsymbol{\omega}$

Lemma 4.2.1 (Angle of incidence) *Let $d \geq 3$ and $n > d$. Let $\mu_1, \ldots, \mu_n$ be Gaussian densities in $\mathbb{R}^d$ of standard deviation σ centered at points of norm at most 1 in $\mathbb{R}^d$. Let $s \leq 2$ and let $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d) \in Q$. Then,*

$$\Pr_{\boldsymbol{\omega}} [\langle \boldsymbol{\omega} | \boldsymbol{q} \rangle < \epsilon] < \left(\frac{340\epsilon n}{\sigma^2}\right)^2, \tag{25}$$

where $\boldsymbol{\omega}$ has density proportional to

$$\langle \boldsymbol{\omega} | \boldsymbol{q} \rangle \left(\prod_{j>d} \int_{\boldsymbol{a}_j} [\langle \boldsymbol{\omega} | \boldsymbol{a}_j \rangle \leq s \langle \boldsymbol{\omega} | \boldsymbol{q} \rangle] \, \mu_j(\boldsymbol{a}_j) \, d\boldsymbol{a}_j\right) \prod_{i=1}^d \mu_i(\boldsymbol{R}_{\boldsymbol{\omega}} \boldsymbol{b}_i + s\boldsymbol{q}).$$

Proof First note that the conditions for $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)$ to be in Q imply that for $1 \leq i \leq d$, $\boldsymbol{b}_i$ has norm at most $\sqrt{(4)^2 + (4)^2} = 4\sqrt{2}$ by properties (1), (3) and (4) of Q.

As in Proposition 2.5.4, we change $\boldsymbol{\omega}$ to $(c, \boldsymbol{\psi})$, where $c = \langle \boldsymbol{\omega} | \boldsymbol{q} \rangle$ and $\boldsymbol{\psi} \in S^{d-2}$.The Jacobian of this change of variables is

$$(1 - c^2)^{(d-3)/2}.$$

In these variables, the bound follows from Lemma 4.2.2. ■

Lemma 4.2.2 (Angle of incidence, II) *Let $d \geq 3$ and $n > d$. Let $\mu_{d+1}, \ldots, \mu_n$ be Gaussian densities in $\mathbb{R}^d$ of standard deviation σ centered at points of norm at most 1 in $\mathbb{R}^d$. Let $s \leq 2$, and let $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d$ each have norm at most $4\sqrt{2}$. Let $\boldsymbol{\psi} \in S^{d-2}$. Then*

$$\boldsymbol{Pr}[c < \epsilon] < \left(\frac{340\epsilon n}{\sigma^2} \right)^2,$$

where c has density proportional to

$$(1 - c^2)^{(d-3)/2} \cdot c \cdot \left(\prod_{j>d} \int_{\boldsymbol{a}_j} [\langle \boldsymbol{\omega}_{\boldsymbol{\psi},c} | \boldsymbol{a}_j \rangle \leq s \langle \boldsymbol{\omega}_{\boldsymbol{\psi},c} | \boldsymbol{q} \rangle] \, \mu_j(\boldsymbol{a}_j) \, d\boldsymbol{a}_j \right) \prod_{i=1}^{d} \mu_i(\boldsymbol{R}_{\boldsymbol{\omega}_{\boldsymbol{\psi},c}} \boldsymbol{b}_i + s\boldsymbol{q}) \quad (26)$$

Proof Let

$$\begin{aligned}
\nu_1(c) &= (1 - c^2)^{(d-3)/2}, \\
\nu_2(c) &= \prod_{j>d} \int_{\boldsymbol{a}_j} [\langle \boldsymbol{\omega}_{\boldsymbol{\psi},c} | \boldsymbol{a}_j \rangle \leq s \langle \boldsymbol{\omega}_{\boldsymbol{\psi},c} | \boldsymbol{q} \rangle] \, \mu_j(\boldsymbol{a}_j) \, d\boldsymbol{a}_j \,, \text{ and} \\
\nu_3(c) &= \prod_{i=1}^{d} \mu_i(\boldsymbol{R}_{\boldsymbol{\omega}_{\boldsymbol{\psi},c}} \boldsymbol{b}_i + s\boldsymbol{q}).
\end{aligned}$$

Then, the density of c is proportional to

$$(26) = c \cdot \nu_1(c)\nu_2(c)\nu_3(c).$$

Let

$$c_0 = \frac{\sigma^2}{240n}. \quad (27)$$

We will show that, for c between 0 and c_0, the density will vary by a factor no greater than 2. We begin by letting $\theta_0 = \pi/2 - \arccos(c_0)$, and noticing that a simple plot of the arccos function reveals $c_0 < 1/26$ implies

$$\theta_0 \leq 1.001 c_0. \quad (28)$$

So, as c varies in the range $[0, c_0]$, $\boldsymbol{\omega}_{\boldsymbol{\psi},c}$ travels in an arc of angle at most θ_0 and therefore travels a distance at most θ_0. As $c = \langle \boldsymbol{q} | \boldsymbol{\omega}_{\boldsymbol{\psi},c} \rangle$, we can apply Lemma 4.2.3 to show

$$\frac{\min_{0 \leq c \leq c_0} \nu_2(c)}{\max_{0 \leq c \leq c_0} \nu_2(c)} \geq 1 - \frac{8n(1+s)\theta_0}{3\sigma^2} \geq 1 - \frac{24n\theta_0}{3\sigma^2} \geq 1 - \frac{1.001}{30}, \quad (29)$$

by (27) and (28).

We similarly note that as c varies between 0 and c_0, the point $\boldsymbol{R}_{\boldsymbol{\omega}_{\psi,c}}\boldsymbol{b}_i + s\boldsymbol{q}$ moves a distance of at most

$$\theta_0 \left\|\boldsymbol{b}_i\right\| \leq 4\sqrt{2}\theta_0.$$

As this point is at distance at most

$$1 + s + \left\|\boldsymbol{b}_i\right\| \leq 4\sqrt{2} + 3$$

from the center of μ_i, Lemma 2.4.2 implies

$$\frac{\min_{0\leq c\leq c_0} \mu_i(\boldsymbol{R}_{\boldsymbol{\omega}_{\psi,c}}\boldsymbol{b}_i + s\boldsymbol{q})}{\max_{0\leq c\leq c_0} \mu_i(\boldsymbol{R}_{\boldsymbol{\omega}_{\psi,c}}\boldsymbol{b}_i + s\boldsymbol{q})} \geq e^{-\left(3(4\sqrt{2}+3)4\sqrt{2}\theta_0\right)/2\sigma^2} \geq e^{-147\theta_0/\sigma^2}.$$

So,

$$\frac{\min_{0\leq c\leq c_0} \nu_3(c)}{\max_{0\leq c\leq c_0} \nu_3(c)} \geq e^{-147d\theta_0/\sigma^2} \geq e^{-148/240}, \tag{30}$$

by (27) and (28) and $d \leq n$.

Finally, we note that

$$1 \geq \nu_1(c) = (1-c^2)^{(d-3)/2} \geq (1 - 1/26d)^{(d-3)/2} \geq \left(1 - \frac{1}{52}\right). \tag{31}$$

So, combining equations (29), (30), and (31), we obtain

$$\frac{\min_{0\leq c\leq c_0} \nu_1(c)\nu_2(c)\nu_3(c)}{\max_{0\leq c\leq c_0} \nu_1(c)\nu_2(c)\nu_3(c)} \geq \left(1 - \frac{1}{52}\right) e^{-\frac{148}{240}} \left(1 - \frac{1.001}{30}\right) \geq 1/2.$$

We conclude by using Lemma 2.3.7 to show

$$\Pr_c\left[c < \epsilon\right] \leq 2(\epsilon/c_0)^2 = 2\left(\frac{240\epsilon n}{\sigma^2}\right)^2 \leq \left(\frac{340\epsilon n}{\sigma^2}\right)^2.$$

■

Lemma 4.2.3 (Points under plane) *For $n > d$, let $\mu_{d+1}, \ldots, \mu_n$ be Gaussian distributions in $\mathbb{R}^d$ of standard deviation σ centered at points of norm at most 1. Let $s \geq 0$ and let $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ be unit vectors such that $\langle \boldsymbol{\omega}_1 | \boldsymbol{q}\rangle$ and $\langle \boldsymbol{\omega}_2 | \boldsymbol{q}\rangle$ are non-negative. Then,*

$$\frac{\prod_{j>d}\int_{\boldsymbol{a}_j}\left[\langle \boldsymbol{\omega}_2|\boldsymbol{a}_j\rangle \leq s\,\langle \boldsymbol{\omega}_2|\boldsymbol{q}\rangle\right]\mu_j(\boldsymbol{a}_j)\,d\boldsymbol{a}_j}{\prod_{j>d}\int_{\boldsymbol{a}_j}\left[\langle \boldsymbol{\omega}_1|\boldsymbol{a}_j\rangle \leq s\,\langle \boldsymbol{\omega}_1|\boldsymbol{q}\rangle\right]\mu_j(\boldsymbol{a}_j)\,d\boldsymbol{a}_j} \geq 1 - \frac{8n(1+s)\left\|\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2\right\|}{3\sigma^2}.$$

Proof As the integrals in the statement of the lemma are just the integrals of Gaussian measures over half-spaces, they can be reduced to univariate integrals. If μ_j is centered at $\bar{\boldsymbol{a}}_j$, then

$$\begin{aligned}\int_{\boldsymbol{a}_j}\left[\langle \boldsymbol{\omega}_1|\boldsymbol{a}_j\rangle \leq s\,\langle \boldsymbol{\omega}_1|\boldsymbol{q}\rangle\right]\mu_j(\boldsymbol{a}_j)\,d\boldsymbol{a}_j &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^d \int_{\boldsymbol{a}_j}\left[\langle \boldsymbol{\omega}_1|\boldsymbol{a}_j\rangle \leq s\,\langle \boldsymbol{\omega}_1|\boldsymbol{q}\rangle\right] e^{-\left\|\boldsymbol{a}_j - \bar{\boldsymbol{a}}_j\right\|^2/2\sigma^2}\,d\boldsymbol{a}_j \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^d \int_{\boldsymbol{g}_j}\left[\left\langle \boldsymbol{\omega}_1|\boldsymbol{g}_j + \bar{\boldsymbol{a}}_j\right\rangle \leq s\,\langle \boldsymbol{\omega}_1|\boldsymbol{q}\rangle\right] e^{-\left\|\boldsymbol{g}_j\right\|^2/2\sigma^2}\,d\boldsymbol{g}_j\,,\end{aligned}$$

(setting $\boldsymbol{g}_j = \boldsymbol{a}_j - \bar{\boldsymbol{a}}_j$)

$$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^d \int_{\boldsymbol{g}_j} \left[\langle \boldsymbol{\omega}_1 | \boldsymbol{g}_j \rangle \leq \langle \boldsymbol{\omega}_1 | s\boldsymbol{q} - \bar{\boldsymbol{a}}_j \rangle\right] e^{-\|\boldsymbol{g}_j\|^2/2\sigma^2}\, d\boldsymbol{g}_j$$
$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{t=-\infty}^{t=\langle \boldsymbol{\omega}_1 | s\boldsymbol{q} - \bar{\boldsymbol{a}}_j \rangle} e^{-t^2/2\sigma^2}\, dt$$

(by Proposition 2.4.4)

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{t=-\langle \boldsymbol{\omega}_1 | s\boldsymbol{q} - \bar{\boldsymbol{a}}_j \rangle}^{t=\infty} e^{-t^2/2\sigma^2}\, dt\,.$$

As $\|\bar{\boldsymbol{a}}_j\| \leq 1$, we know

$$-\langle \boldsymbol{\omega}_1 | s\boldsymbol{q} - \bar{\boldsymbol{a}}_j \rangle \;=\; -\langle \boldsymbol{\omega}_1 | s\boldsymbol{q} \rangle + \langle \boldsymbol{\omega}_1 | \bar{\boldsymbol{a}}_j \rangle \;\leq\; \langle \boldsymbol{\omega}_1 | \bar{\boldsymbol{a}}_j \rangle \;\leq\; 1 \tag{32}$$

Similarly,

$$\begin{aligned} \left| -\langle \boldsymbol{\omega}_1 | s\boldsymbol{q} - \bar{\boldsymbol{a}}_j \rangle + \langle \boldsymbol{\omega}_2 | s\boldsymbol{q} - \bar{\boldsymbol{a}}_j \rangle \right| &= \left| -\langle \boldsymbol{\omega}_1 - \boldsymbol{\omega}_2 | s\boldsymbol{q} - \bar{\boldsymbol{a}}_j \rangle \right| \\ &\leq \|\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2\|\, \|s\boldsymbol{q} - \bar{\boldsymbol{a}}_j\| \qquad (33) \\ &\leq \|\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2\|\, (s+1). \qquad (34) \end{aligned}$$

Thus, by applying Lemma 2.4.11 to (32) and (34), we obtain

$$\begin{aligned} \frac{\int_{\boldsymbol{a}_j} \left[\langle \boldsymbol{\omega}_2 | \boldsymbol{a}_j \rangle \leq s \langle \boldsymbol{\omega}_2 | \boldsymbol{q} \rangle\right] \mu_j(\boldsymbol{a}_j)\, d\boldsymbol{a}_j}{\int_{\boldsymbol{a}_j} \left[\langle \boldsymbol{\omega}_1 | \boldsymbol{a}_j \rangle \leq s \langle \boldsymbol{\omega}_1 | \boldsymbol{q} \rangle\right] \mu_j(\boldsymbol{a}_j)\, d\boldsymbol{a}_j} &= \frac{\int_{t=-\langle \boldsymbol{\omega}_2 | s\boldsymbol{q} - \bar{\boldsymbol{a}}_j \rangle}^{t=\infty} e^{-t^2/2\sigma^2}\, dt\,.}{\int_{t=-\langle \boldsymbol{\omega}_1 | s\boldsymbol{q} - \bar{\boldsymbol{a}}_j \rangle}^{t=\infty} e^{-t^2/2\sigma^2}\, dt\,.} \\ &\geq \left(1 - \frac{8(1+s)\,\|\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2\|}{3\sigma^2}\right). \end{aligned}$$

Thus,

$$\begin{aligned} \frac{\prod_{j>d} \int_{\boldsymbol{a}_j} \left[\langle \boldsymbol{\omega}_2 | \boldsymbol{a}_j \rangle \leq s \langle \boldsymbol{\omega}_2 | \boldsymbol{q} \rangle\right] \mu_j(\boldsymbol{a}_j)\, d\boldsymbol{a}_j}{\prod_{j>d} \int_{\boldsymbol{a}_j} \left[\langle \boldsymbol{\omega}_1 | \boldsymbol{a}_j \rangle \leq s \langle \boldsymbol{\omega}_1 | \boldsymbol{q} \rangle\right] \mu_j(\boldsymbol{a}_j)\, d\boldsymbol{a}_j} &\geq \left(1 - \frac{8(1+s)\,\|\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2\|}{3\sigma^2}\right)^{n-d} \\ &\geq \left(1 - \frac{8n(1+s)\,\|\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2\|}{3\sigma^2}\right). \end{aligned}$$

■

4.3 Extending the shadow bound

In this section, we relax the restrictions made in the statement of Theorem 4.0.1. The extensions of Theorem 4.0.1 are needed in the proof of Theorem 5.0.1.

We begin by removing the restrictions on where the distributions are centered in the shadow bound.

Corollary 4.3.1 ($\|\boldsymbol{a}_i\|$ free) *Let $\boldsymbol{z}$ and $\boldsymbol{t}$ be unit vectors and let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be Gaussian random vectors in $\mathbb{R}^d$ of standard deviation $\sigma \leq 1/3\sqrt{d \ln n}$ centered at points $\bar{\boldsymbol{a}}_1, \ldots, \bar{\boldsymbol{a}}_n$. Then,*

$$\boldsymbol{E}\left[\mathbf{Shadow}_{\boldsymbol{z},\boldsymbol{t}}\left(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\right)\right] \leq \mathcal{D}\left(d, n, \frac{\sigma}{\max\left(1, \max_i \|\bar{\boldsymbol{a}}\|\right)}\right)$$

where $\mathcal{D}(d, n, \sigma)$ is as given in Theorem 4.0.1.

Proof Let $k = \max_i \|\bar{\boldsymbol{a}}_i\|$. Assume without loss of generality that $k \geq 1$, and let $\boldsymbol{b}_i = \boldsymbol{a}_i/k$ for all i. Then, $\boldsymbol{b}_i$ is a Gaussian random variable of standard deviation (σ/k) centered at a point of norm at most 1. So, Theorem 4.0.1 implies

$$\mathbf{E}\left[\mathbf{Shadow}_{\boldsymbol{z},\boldsymbol{t}}\left(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\right)\right] \leq \mathcal{D}\left(d, n, \frac{\sigma}{k}\right).$$

On the other hand, the shadow of the polytope defined by the $\boldsymbol{b}_i$s can be seen to be a dilation of the polytope defined by the $\boldsymbol{a}_i$s: the division of the $\boldsymbol{b}_i$s by a factor of k is equivalent to the multiplication of $\boldsymbol{x}$ by k. So, we may conclude that for all $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$,

$$\left|\mathbf{Shadow}_{\boldsymbol{z},\boldsymbol{t}}\left(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\right)\right| = \left|\mathbf{Shadow}_{\boldsymbol{z},\boldsymbol{t}}\left(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\right)\right|.$$

■

Corollary 4.3.2 (Gaussians free) *Let $\boldsymbol{z}$ and $\boldsymbol{t}$ be unit vectors and let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be Gaussian random vectors in $\mathbb{R}^d$ with covariance matrices $\boldsymbol{M}_1, \ldots, \boldsymbol{M}_n$ centered at points $\bar{\boldsymbol{a}}_1, \ldots, \bar{\boldsymbol{a}}_n$, respectively. If the eigenvalues of each $\boldsymbol{M}_i$ lie between σ^2 and $1/9d \ln n$, then*

$$\boldsymbol{E}\left[\mathbf{Shadow}_{\boldsymbol{z},\boldsymbol{t}}\left(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\right)\right] \leq \mathcal{D}\left(d, n, \frac{\sigma}{1 + \max_i \|\bar{\boldsymbol{a}}\|}\right) + 1$$

where $\mathcal{D}(d, n, \sigma)$ is as given in Theorem 4.0.1.

Proof By Proposition 2.4.1, each $\boldsymbol{a}_i$ can be expressed as

$$\boldsymbol{a}_i = \bar{\boldsymbol{a}}_i + \boldsymbol{g}_i + \tilde{\boldsymbol{g}}_i,$$

where $\tilde{\boldsymbol{g}}_i$ is a Gaussian random vector of standard deviation σ centered at the origin and $\boldsymbol{g}_i$ is a Gaussian random vector centered at the origin with covariance matrix $\boldsymbol{M}_i^0 = \boldsymbol{M}_i - \sigma^2 I$, each of whose eigenvalues is at most $1/9d \ln n$. Let $\tilde{\boldsymbol{a}}_i = \bar{\boldsymbol{a}}_i + \boldsymbol{g}_i$. If $\|\tilde{\boldsymbol{a}}_i\| \leq 1 + \|\bar{\boldsymbol{a}}_i\|$, for all i, then we can apply Corollary 4.3.1 to show

$$\mathop{\mathbf{E}}_{\tilde{\boldsymbol{g}}_1, \ldots, \tilde{\boldsymbol{g}}_n}\left[\mathbf{Shadow}_{\boldsymbol{z},\boldsymbol{t}}\left(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\right)\right] \leq \mathcal{D}\left(d, n, \frac{\sigma}{\max\left(1, \max_i \|\tilde{\boldsymbol{a}}\|\right)}\right) \leq \mathcal{D}\left(d, n, \frac{\sigma}{1 + \max_i \|\bar{\boldsymbol{a}}\|}\right).$$

On the other hand, Corollary 2.4.6 implies

$$\mathop{\mathbf{Pr}}_{\boldsymbol{g}_1, \ldots, \boldsymbol{g}_n}\left[\exists i : \|\tilde{\boldsymbol{a}}_i\| > 1 + \|\bar{\boldsymbol{a}}_i\|\right] \leq 0.0015 \binom{n}{d}^{-1}$$

So, using Lemma 2.3.3 and $\mathbf{Shadow}_{\boldsymbol{z},\boldsymbol{t}}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n) \leq \binom{n}{d}$, we can show

$$\mathop{\mathbf{E}}_{\tilde{\boldsymbol{g}}_1,\ldots,\tilde{\boldsymbol{g}}_n}\left[\mathop{\mathbf{E}}_{\boldsymbol{g}_1,\ldots,\boldsymbol{g}_n}\left[\mathbf{Shadow}_{\boldsymbol{z},\boldsymbol{t}}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n)\right]\right] \leq \mathcal{D}\left(d,n,\frac{\sigma}{1+\max_i \|\bar{\boldsymbol{a}}\|}\right)+1.$$

from which the Corollary follows. ■

Corollary 4.3.3 ($\boldsymbol{y_i}$ free) *Let $\boldsymbol{y} \in \mathbb{R}^n$ be a positive vector. Let $\boldsymbol{z}$ and $\boldsymbol{t}$ be unit vectors and let $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n$ be Gaussian random vectors in $\mathbb{R}^d$ with covariance matrices $\boldsymbol{M}_1,\ldots,\boldsymbol{M}_n$ centered at points $\bar{\boldsymbol{a}}_1,\ldots,\bar{\boldsymbol{a}}_n$, respectively. If the eigenvalues of each $\boldsymbol{M}_i$ lie between σ^2 and $1/9d\ln n$, then*

$$\boldsymbol{E}[\mathbf{Shadow}_{\boldsymbol{z},\boldsymbol{t}}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n)\,;\boldsymbol{y}] \leq \mathcal{D}\left(d,n,\frac{\sigma}{(1+\max_i \|\bar{\boldsymbol{a}}_i\|)(\max_i y_i)/(\min_i y_i)}\right)+1$$

where $\mathcal{D}(d,n,\sigma)$ is as given in Theorem 4.0.1.

Proof Nothing in the statement is changed if we rescale the y_is. So, assume without loss of generality that $\min_i y_i = 1$.

Let $\boldsymbol{b}_i = \boldsymbol{a}_i/y_i$. Then $\boldsymbol{b}_i$ is a Gaussian random vector with covariance matrix $\boldsymbol{M}_i/y_i^2$ centered at a point of norm at most $\|\boldsymbol{a}_i\|/y_i \leq \|\boldsymbol{a}_i\|$. Then, the eigenvalues of each $\boldsymbol{M}_i$ lie between σ^2/y_i^2 and $1/(9d\ln n y_i^2) \leq 1/9d\ln n$, so we may complete the proof by applying Corollary 4.3.2. ■

5 Smoothed Analysis of a Two-Phase Simplex Algorithm

In this section, we will analyze the smoothed complexity of the two-phase shadow-vertex simplex method introduced in Section 3.3. The analysis of the algorithm will use as a black-box the bound on the expected sizes of shadows proved in the previous section. However, the analysis is not immediate from this bound.

The most obvious difficulty in applying the shadow bound to the analysis of an algorithm is that, in the statement of the shadow bound, the plane onto which the polytope was projected to form the shadow was fixed, and unrelated to the data defining the polytope. However, in the analysis of the shadow-vertex algorithm, the plane onto which the polytope is projected will necessarily depend upon data defining the linear program. This is the dominant complication in the analysis of the number of steps taken to solve LP'.

Another obstacle will stem from the fact that, in the analysis of LP^+, we need to consider the expected sizes of shadows of the convex hulls of points of the form $\boldsymbol{a}_i^+/y_i^+$, which do not have a Gaussian distribution. In our analysis of LP^+, we essentially handle this complication by demonstrating that in almost every small region the distribution can be approximated by some Gaussian distribution.

The last issue we need to address is that if $\mathbf{s_{min}}(\boldsymbol{A}_I)$ is too small, then the resulting values for y_i' and y_i^+ can be too large. In Section 5.1 we resolve this problem by proving that one of $3nd \ln n$ randomly chosen I will have reasonable $\mathbf{s_{min}}(\boldsymbol{A}_I)$ with very high probability. Having a reasonable $\mathbf{s_{min}}(\boldsymbol{A}_I)$ is also essential for the analysis of LP'.

As our two-phase shadow-vertex simplex algorithm is randomized, we will measure its expected complexity on each input. For an input linear program specified by $\boldsymbol{A}$, $\boldsymbol{y}$ and $\boldsymbol{z}$, we let

$$\mathcal{C}(\boldsymbol{A}, \boldsymbol{y}, \boldsymbol{z})$$

denote the expected number of simplex steps taken by the algorithm on input $(\boldsymbol{A}, \boldsymbol{y}, \boldsymbol{z})$. As this expectation is taken over the choices for $\mathcal{I}$ and $\boldsymbol{\alpha}$, and can be divided into the number of steps taken to solve LP^+ and LP', we introduce the functions

$$\mathcal{S}'_{\boldsymbol{z}}(\boldsymbol{A}, \boldsymbol{y}, \mathcal{I}, \boldsymbol{\alpha}),$$

to denote the number of simplex steps taken by the algorithm in step (5) to solve LP' for a given $\boldsymbol{A}$, $\boldsymbol{y}$, $\mathcal{I}$ and $\boldsymbol{\alpha}$, and

$$\mathcal{S}^+_{\boldsymbol{z}}(\boldsymbol{A}, \boldsymbol{y}, \mathcal{I}) + 2$$

to denote the number of simplex steps[3] taken by the algorithm in step (7) to solve LP^+ for a given $\boldsymbol{A}$, $\boldsymbol{y}$ and $\mathcal{I}$. We note that the complexity of the second phase does not depend upon $\boldsymbol{\alpha}$, however it does depend upon $\mathcal{I}$ as $\mathcal{I}$ affects the choice of κ and M. We have

$$\mathcal{C}(\boldsymbol{A}, \boldsymbol{y}, \boldsymbol{z}) \leq \mathop{\mathbf{E}}_{\mathcal{I}, \boldsymbol{\alpha}} \left[\mathcal{S}'_{\boldsymbol{z}}(\boldsymbol{A}, \boldsymbol{y}, \mathcal{I}, \boldsymbol{\alpha})\right] + \mathop{\mathbf{E}}_{\mathcal{I}, \boldsymbol{\alpha}} \left[\mathcal{S}^+_{\boldsymbol{z}}(\boldsymbol{A}, \boldsymbol{y}, \mathcal{I}, \boldsymbol{\alpha})\right] + 2.$$

Theorem 5.0.1 (Main) *There exists a polynomial $\mathcal{P}$ and a constant σ_0 such that for every $n > d \geq 3$, $\bar{\boldsymbol{A}} = [\bar{\boldsymbol{a}}_1, \ldots, \bar{\boldsymbol{a}}_n] \in \mathbb{R}^{n \times d}$, $\bar{\boldsymbol{y}} \in \mathbb{R}^n$ and $\boldsymbol{z} \in \mathbb{R}^d$, and $\sigma > 0$,*

$$\mathop{\boldsymbol{E}}_{\boldsymbol{A}, \boldsymbol{y}} \left[\mathcal{C}(\boldsymbol{A}, \boldsymbol{y}, \boldsymbol{z})\right] \leq \min\left(\mathcal{P}(d, n, 1/\min(\sigma, \sigma_0)), \binom{n}{d} + \binom{n}{d+1} + 2\right),$$

[3]The seemingly odd appearance of $+2$ in this definition is explained by 3.3.5.

where $\boldsymbol{A}$ is a Gaussian random matrix centered at $\bar{\boldsymbol{A}}$ of standard deviation $\sigma \max_i \|(\bar{y}_i, \bar{\boldsymbol{a}}_i)\|$, and $\boldsymbol{y}$ is a Gaussian random vector centered at $\bar{\boldsymbol{y}}$ of standard deviation $\sigma \max_i \|(\bar{y}_i, \bar{\boldsymbol{a}}_i)\|$.

Proof We first observe that the behavior of the algorithm is unchanged if one multiplies $\boldsymbol{A}$ and $\boldsymbol{y}$ by a power of two. That is,

$$\mathcal{C}(\boldsymbol{A}, \boldsymbol{y}, \boldsymbol{z}) = \mathcal{C}(2^k \boldsymbol{A}, 2^k \boldsymbol{y}, \boldsymbol{z}),$$

for any integer k. When $\boldsymbol{A}$ and $\boldsymbol{y}$ are Gaussian random variables centered at $\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{y}}$ of standard deviation $\sigma \max_i \|(\bar{y}_i, \bar{\boldsymbol{a}}_i)\|$, $2^k \boldsymbol{A}$ and $2^k \boldsymbol{y}$ are Gaussian random variables centered at $2^k \bar{\boldsymbol{A}}$ and $2^k \bar{\boldsymbol{y}}$ of standard deviation $\sigma \max_i \left\|(2^k \bar{y}_i, 2^k \bar{\boldsymbol{a}}_i)\right\|$. Accordingly, we may assume without loss of generality in our analysis that $\max_i \|(\bar{y}_i, \bar{\boldsymbol{a}}_i)\| \in (1/2, 1]$.

The Theorem now follows from Proposition 5.0.2 and Lemmas 5.2.1 and 5.3.1.

■

Before proceeding with the proof of Theorem 5.0.1, we state a trivial upper bound on $\mathcal{S}'$ and $\mathcal{S}^+$:

Proposition 5.0.2 (trivial shadow bounds) *For all $\boldsymbol{A}$, $\boldsymbol{y}$, $\boldsymbol{z}$, $\mathcal{I}$ and $\boldsymbol{\alpha}$:*

$$\mathcal{S}'_{\boldsymbol{z}}(\boldsymbol{A}, \boldsymbol{y}, \mathcal{I}, \boldsymbol{\alpha}) \leq \binom{n}{d} \qquad \text{and} \qquad \mathcal{S}^+_{\boldsymbol{z}}(\boldsymbol{A}, \boldsymbol{y}, \mathcal{I}, \boldsymbol{\alpha}) \leq \binom{n}{d+1}.$$

Proof The bound on $\mathcal{S}'$ follows from the fact that there are $\binom{n}{d}$ d-subsets of $[n]$. The bound on $\mathcal{S}^+$ follows from the observation in Lemma 3.3.5 that the number of steps taken by the second phase is at most 2 plus the number of $(d+1)$-subsets of $[n]$. ■

5.1 Many Good Choices

For a Gaussian random d-by-d matrix $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d)$, it is possible to show that the probability that the smallest singular value of $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d)$ is less than ϵ is at most $O(d^{1/2}\epsilon)$. In this section, we consider the probability that almost all of the d-by-d minors of a d-by-n matrix $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ have small singular value. If the events for different minors were independent, then the proof would be straightforward. However, distinct minors may have significant overlap. While we believe stronger concentration results should be obtainable, we have only been able to prove:

Lemma 5.1.1 (Many good choices) *For $n > d \geq 3$, let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be Gaussian random variables in $\mathbb{R}^d$ of standard deviation σ centered at points of norm at most 1. Let $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$. Then, we have*

$$\Pr_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\sum_{I\in\binom{[n]}{d}} [\mathbf{s}_{\min}(\boldsymbol{A}_I) \leq \kappa_0] \geq \left(1-\frac{1}{n}\right)\binom{n}{d}\right] \leq n^{-d} + n^{-n+d-1} + n^{-2.9d+1},$$

where

$$\kappa_0 \stackrel{\text{def}}{=} \frac{\sigma \min(1,\sigma)}{12d^2n^7\sqrt{\ln n}}. \tag{35}$$

In the analyses of LP' and LP^+, we use the following consequence of Lemma 5.1.1, whose statement is facilitated by the following notation for a set of d-sets, $\mathcal{I}$

$$\mathcal{I}(\boldsymbol{A}) \stackrel{\text{def}}{=} \operatorname{argmax}_{I\in\mathcal{I}} (\mathbf{s_{min}}(\boldsymbol{A}_I)).$$

Corollary 5.1.2 (probability of small $\mathbf{s_{min}}(\boldsymbol{A}_{\mathcal{I}(\boldsymbol{A})})$) *For $n > d \geq 3$, let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be Gaussian random variables in $\mathbb{R}^d$ of standard deviation σ centered at points of norm at most 1, and let $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$. For $\mathcal{I}$ a set of $3nd\ln n$ randomly chosen d-subsets of $[n]$,*

$$\Pr_{\boldsymbol{A},\mathcal{I}}\left[\mathbf{s}_{\min}\left(\boldsymbol{A}_{\mathcal{I}(\boldsymbol{A})}\right) \leq \kappa_0\right] \leq 0.417\binom{n}{d}^{-1}.$$

Proof

$$
\begin{aligned}
&\Pr_{\boldsymbol{A},\mathcal{I}}\left[\mathbf{s_{min}}\left(\boldsymbol{A}_{\mathcal{I}(\boldsymbol{A})}\right) \le \kappa_0\right] \\
&= \Pr_{\boldsymbol{A},\mathcal{I}}\left[\forall I \in \mathcal{I} : \mathbf{s_{min}}\left(\boldsymbol{A}_I\right) \le \kappa_0\right] \\
&\le \Pr_{\boldsymbol{A}}\left[\sum_{I\in\binom{[n]}{d}} \left[\mathbf{s_{min}}\left(\boldsymbol{A}_I\right) \le \kappa_0\right] < \left(1-\frac{1}{n}\right)\binom{n}{d}\right] \\
&\quad + \Pr_{\mathcal{I},\boldsymbol{A}}\left[\forall I \in \mathcal{I} : \mathbf{s_{min}}\left(\boldsymbol{A}_I\right) \le \kappa_0 \,\middle|\, \sum_{I\in\binom{[n]}{d}} \left[\mathbf{s_{min}}\left(\boldsymbol{A}_I\right) \le \kappa_0\right] \ge \left(1-\frac{1}{n}\right)\binom{n}{d}\right] \\
&\le n^{-d} + n^{-n+d-1} + n^{-2.9d+1} + \left(1-\frac{1}{n}\right)^{|\mathcal{I}|}, \text{ by Lemma 5.1.1} \\
&\le n^{-d} + n^{-n+d-1} + n^{-2.9d+1} + n^{-3d}, \text{ as } |\mathcal{I}| = 3nd\ln n, \\
&\le 0.417\binom{n}{d}^{-1},
\end{aligned}
$$

for $n > d \ge 3$. ■

We also use the following corollary, which states that it is highly unlikely that κ falls outside the set $\mathcal{K}$, which we now define:

$$\mathcal{K} = \left\{2^{\lfloor \lg(x)\rfloor} : \kappa_0 \le x \le \sqrt{d} + 3d\sqrt{\ln n}\sigma\right\}. \tag{36}$$

Corollary 5.1.3 (probability of κ in $\mathcal{K}$) *For $n > d \ge 3$, let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be Gaussian random variables in $\mathbb{R}^d$ of standard deviation σ centered at points of norm at most 1, and let $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$. For $\mathcal{I}$ a set of $3nd\ln n$ randomly chosen d-subsets of $[n]$,*

$$\Pr_{\boldsymbol{A},\mathcal{I}}\left[2^{\lfloor \lg(\mathbf{s}_{min}(\boldsymbol{A}_{\mathcal{I}(\boldsymbol{A})}))\rfloor} \notin \mathcal{K}\right] \le 0.42\binom{n}{d}^{-1}.$$

Proof

It follows from Corollary 5.1.2 that

$$\Pr_{\boldsymbol{A},\mathcal{I}}\left[\mathbf{s_{min}}\left(\boldsymbol{A}_{\mathcal{I}(\boldsymbol{A})}\right) \le \kappa_0\right] \le 0.417\binom{n}{d}^{-1}.$$

On the other hand, as

$$\mathbf{s_{min}}\left(\boldsymbol{A}_I\right) \le \|\boldsymbol{A}_I\| \le \sqrt{d}\max_i \|\boldsymbol{a}_i\|,$$

$$\Pr_{\boldsymbol{A},\mathcal{I}}\left[\mathbf{s_{min}}\left(\boldsymbol{A}_{\mathcal{I}(\boldsymbol{A})}\right) \ge \sqrt{d}\left(1+3\sqrt{d\ln n}\sigma\right)\right] \le \Pr_{\boldsymbol{A}}\left[\max_i \|\boldsymbol{a}_i\| \ge 1+3\sqrt{d\ln n}\sigma\right] \le 0.0015\binom{n}{d}^{-1},$$

by Corollary 2.4.6. ■

Proposition 5.1.4 (size of $\mathcal{K}$)

$$|\mathcal{K}| \leq 9 \lg(nd/\min(\sigma, 1)).$$

The rest of this section is devoted to the proof of Lemma 5.1.1. The key to the proof is an examination of the relation between the events which we now define.

Definition 5.1.5 *For $I \in \binom{[n]}{d}$, $K \in \binom{[n]}{d-1}$, and $j \notin K$, we define the indicator random variables*

$$X_I = [\mathbf{s_{min}}(\boldsymbol{A}_I) \leq \kappa_0], \text{ and}$$
$$Y_K^j = [\mathbf{dist}(\boldsymbol{a}_j, \boldsymbol{Span}(A_K)) \leq h_0],$$

where

$$h_0 \stackrel{\text{def}}{=} \frac{\sigma}{4n^4}.$$

In Lemma 5.1.8, we obtain a concentration result on the Y_K^js using the fact that the Y_K^j are independent for fixed K and different j. To relate this concentration result to the X_Is, we show in Lemma 5.1.9 that when X_I is true, it is probably the case that $Y_{I-\{j\}}^j$ is true for most j.

Proof of Lemma 5.1.1 The proof has two parts. The first, and easier, part is Lemma 5.1.8 which implies

$$\mathop{\mathbf{Pr}}_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\sum_{K\in\binom{[n]}{d-1}}\sum_{j\notin K} Y_K^j \leq \left\lceil\frac{n-d-1}{2}\right\rceil\binom{n}{d-1}\right] > 1 - n^{-n+d-1}.$$

To apply this fact, we use Lemma 5.1.9, which implies

$$\mathop{\mathbf{Pr}}_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\sum_{K\in\binom{[n]}{d-1}}\sum_{j\notin K} Y_K^j > \frac{d}{2}\sum_I X_I\right] > 1 - n^{-d} - n^{-2.9d+1}.$$

Combining these two Lemmas, we obtain

$$\mathop{\mathbf{Pr}}_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\frac{d}{2}\sum_I X_I < \left\lceil\frac{n-d-1}{2}\right\rceil\binom{n}{d-1}\right] \geq 1 - n^{-d} - n^{-n+d-1} - n^{-2.9d+1}.$$

Observing,

$$\begin{aligned}\frac{d}{2}\sum_I X_I < \left\lceil\frac{n-d-1}{2}\right\rceil\binom{n}{d-1} \quad \Longrightarrow \quad \sum_I X_I &< \frac{n-d}{d}\binom{n}{d-1} \\ &= \frac{n-d}{n-d+1}\binom{n}{d} \\ &= \left(1 - \frac{1}{n-d+1}\right)\binom{n}{d} \\ &\leq \left(1 - \frac{1}{n}\right)\binom{n}{d},\end{aligned}$$

we obtain

$$\mathop{\mathbf{Pr}}_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\sum_I X_I \geq \left(1-\frac{1}{n}\right)\binom{n}{d}\right] \leq n^{-d}+n^{-n+d-1}+n^{-2.9d+1}.$$

■

Lemma 5.1.6 (Probability of Y_K^j) *Under the conditions of Lemma 5.1.1, for all $K \in \binom{[n]}{d-1}$ and $j \notin K$,*

$$\mathop{\boldsymbol{Pr}}_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[Y_K^j\right] \leq \frac{h_0}{\sigma}.$$

Proof Follows from Proposition 2.4.7. ■

Lemma 5.1.7 (Sum over j of Y_K^j) *Under the conditions of Lemma 5.1.1, for all $K \in \binom{[n]}{d-1}$,*

$$\mathop{\boldsymbol{Pr}}_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\sum_{j\notin K} Y_K^j \geq \lceil (n-d+1)/2 \rceil\right] \leq \left(\frac{4h_0}{\sigma}\right)^{\lceil (n-d+1)/2 \rceil}$$

Proof Using the fact that for fixed K, the events Y_K^j are independent, we compute

$$\begin{aligned}
\mathop{\mathbf{Pr}}_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\sum_{j\notin K} Y_K^j \geq \lceil (n-d+1)/2 \rceil\right] &\leq \sum_{J\in\binom{[n]-K}{\lceil (n-d+1)/2 \rceil}} \mathop{\mathbf{Pr}}_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\forall j \in J, Y_K^j\right] \\
&= \sum_{J\in\binom{[n]-K}{\lceil (n-d+1)/2 \rceil}} \prod_{j\in J} \mathop{\mathbf{Pr}}_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[Y_K^j\right] \\
&\leq \sum_{J\in\binom{[n]-K}{\lceil (n-d+1)/2 \rceil}} \left(\frac{h_0}{\sigma}\right)^{\lceil (n-d+1)/2 \rceil}, \qquad \text{by Lemma 5.1.6,} \\
&\leq \left(\frac{4h_0}{\sigma}\right)^{\lceil (n-d+1)/2 \rceil},
\end{aligned}$$

as $\left|\binom{[n]-K}{\lceil (n-d+1)/2 \rceil}\right| \leq 2^{|[n]-K|} = 2^{n-d+1}$. ■

Lemma 5.1.8 (Sum over K and j of Y_K^j) *Under the conditions of Lemma 5.1.1,*

$$\mathop{\boldsymbol{Pr}}_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\sum_{K\in\binom{[n]}{d-1}} \sum_{j\notin K} Y_K^j > \left\lceil \frac{n-d-1}{2} \right\rceil \binom{n}{d-1}\right] \leq n^{-n+d-1}.$$

Proof If $\sum_{K\in\binom{[n]}{d-1}}\sum_{j\notin K} Y_K^j > \left\lceil\frac{n-d-1}{2}\right\rceil\binom{n}{d-1}$, then there must exist a K for which $\sum_{j\notin K} Y_K^j > \left\lceil\frac{n-d-1}{2}\right\rceil$, which implies for that K

$$\sum_{j\notin K} Y_K^j \geq \left\lceil\frac{n-d-1}{2}\right\rceil + 1 = \left\lceil\frac{n-d+1}{2}\right\rceil.$$

Using this trick, we compute

$$\begin{aligned}
\mathop{\mathbf{Pr}}_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\sum_{K\in\binom{[n]}{d-1}}\sum_{j\notin K} Y_K^j \geq \left\lceil\frac{n-d-1}{2}\right\rceil\binom{n}{d-1}\right] &\leq \mathop{\mathbf{Pr}}_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\exists K\in\binom{[n]}{d-1} : \sum_{j\notin K} Y_K^j \geq \left\lceil\frac{n-d+1}{2}\right\rceil\right] \\
&\leq \binom{n}{d-1}\mathop{\mathbf{Pr}}_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\sum_{j\notin K} Y_K^j \geq \left\lceil\frac{n-d+1}{2}\right\rceil\right] \\
&\leq \binom{n}{d-1}\left(\frac{4h_0}{\sigma}\right)^{\lceil(n-d+1)/2\rceil}
\end{aligned}$$

(by Lemma 5.1.7)

$$\begin{aligned}
&= \binom{n}{n-d+1}\left(\frac{4h_0}{\sigma}\right)^{\lceil(n-d+1)/2\rceil} \\
&\leq n^{n-d+1}\left(\frac{1}{n^4}\right)^{\lceil(n-d+1)/2\rceil} \\
&\leq n^{-n+d-1}.
\end{aligned}$$

■

The other statement needed for the proof of Lemma 5.1.1 is:

Lemma 5.1.9 (Relating Xs to Ys) *Under the conditions of Lemma 5.1.1,*

$$\mathop{\boldsymbol{Pr}}_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}\left[\sum_{K\in\binom{[n]}{d-1}}\sum_{j\notin K} Y_K^j \leq \frac{d}{2}\sum_I X_I\right] \leq n^{-d} + n^{-2.9d+1}$$

Proof Follows immediately from Lemmas 5.1.10 and 5.1.12. ■

Lemma 5.1.10 (Geometric condition for bad I) *If there exists a d-set I such that*

$$X_I \quad \textit{and} \quad \sum_{j\in I} Y_{I-\{j\}}^j \leq d/2,$$

then there exists a set $L\subset I$, $|L| = \lfloor d/2-1\rfloor$ and a $j_0\in I-L$ such that

$$\mathbf{dist}\left(\boldsymbol{a}_{j_0}, \boldsymbol{Span}\left(\boldsymbol{A}_L\right)\right) \leq \sqrt{d}\kappa_0\left(1+\left\lceil\frac{d}{2}\right\rceil\frac{\max_i\|\boldsymbol{a}_i\|}{h_0}\right).$$

Proof Let $I = \{i_1, \ldots, i_d\}$. By Proposition 2.2.6 (a), X_I implies the existence of $u_{i_1}, \ldots, u_{i_d}$, $\|(u_{i_1}, \ldots, u_{i_d})\| = 1$, such that

$$\left\| \sum_{i \in I} u_i \boldsymbol{a}_i \right\| \leq \kappa_0.$$

On the other hand, $\sum_{j \in I} Y^j_{I-\{j\}} \leq d/2$ implies the existence of a $J \subset I$, $|J| = \lceil d/2 \rceil$, such that $Y^j_{I-\{j\}} = 0$ for all $j \in J$. By Lemma 5.1.11, this implies $|u_j| < \kappa_0/h_0$ for all $j \in J$. As $\|(u_{i_1}, \ldots, u_{i_d})\| = 1$ and $\kappa_0/h_0 \leq 1/\sqrt{d}$, there exists some $j_0 \in I - J$ such that $|u_{j_0}| \geq 1/\sqrt{d}$. Setting $L = I - J - \{j_0\}$, we compute

$$\begin{aligned}
\left\| \sum_{j \in I} u_j \boldsymbol{a}_j \right\| \leq \kappa_0 \quad &\Longrightarrow \quad \left\| u_{j_0} \boldsymbol{a}_{j_0} + \sum_{j \in L} u_j \boldsymbol{a}_j + \sum_{j \in J} u_j \boldsymbol{a}_j \right\| \leq \kappa_0 \\
&\Longrightarrow \quad \left\| u_{j_0} \boldsymbol{a}_{j_0} + \sum_{j \in L} u_j \boldsymbol{a}_j \right\| \leq \kappa_0 + \left\| \sum_{j \in J} u_j \boldsymbol{a}_j \right\| \\
&\Longrightarrow \quad \left\| \boldsymbol{a}_{j_0} + \sum_{j \in L} (u_j/u_{j_0}) \boldsymbol{a}_j \right\| \leq (1/|u_{j_0}|) \left(\kappa_0 + \left\| \sum_{j \in J} u_j \boldsymbol{a}_j \right\| \right) \\
&\Longrightarrow \quad \left\| \boldsymbol{a}_{j_0} + \sum_{j \in L} (u_j/u_{j_0}) \boldsymbol{a}_j \right\| \leq \sqrt{d} \left(\kappa_0 + \sum_{j \in J} |u_j| \, \|\boldsymbol{a}_j\| \right) \\
&\Longrightarrow \quad \mathbf{dist}\left(\boldsymbol{a}_{j_0}, \mathbf{Span}\left(A_L\right)\right) \leq \sqrt{d} \left(\kappa_0 + \left\lceil \frac{d}{2} \right\rceil \frac{\kappa_0 \max_i \|\boldsymbol{a}_i\|}{h_0} \right).
\end{aligned}$$

■

Lemma 5.1.11 (Big height, small coefficient) *Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d$ be vectors and $\boldsymbol{u}$ be a unit vector such that*

$$\left\| \sum_{i=1}^{d} u_i \boldsymbol{a}_i \right\| \leq \kappa_0.$$

If $\mathbf{dist}\left(\boldsymbol{a}_j, \boldsymbol{Span}\left(\{\boldsymbol{a}_i\}_{i \neq j}\right)\right) > h_0$, *then* $|u_j| < \kappa_0/h_0$.

Proof We have

$$\begin{aligned}
\left\| \sum_{i=1}^{d} u_i \boldsymbol{a}_i \right\| \leq \kappa_0 \quad &\Longrightarrow \quad \left\| u_j \boldsymbol{a}_j + \sum_{i \neq j} u_i \boldsymbol{a}_i \right\| \leq \kappa_0 \\
&\Longrightarrow \quad \left\| \boldsymbol{a}_j + \sum_{i \neq j} (u_i/u_j) \boldsymbol{a}_i \right\| \leq \kappa_0/|u_j| \\
&\Longrightarrow \quad \mathbf{dist}\left(\boldsymbol{a}_j, \mathbf{Span}\left(\{\boldsymbol{a}_i\}_{i \neq j}\right)\right) \leq \kappa_0/|u_j|,
\end{aligned}$$

from which the lemma follows. ■

Lemma 5.1.12 (Probability of bad geometry) *Under the conditions of Lemma 5.1.1,*

$$\underset{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}{\boldsymbol{Pr}}\left[\begin{array}{l}\exists L \in \binom{[n]}{\lfloor d/2-1\rfloor}, j_0 \notin L \text{ such that} \\ \quad \mathbf{dist}\left(\boldsymbol{a}_{j_0}, \boldsymbol{Span}\left(A_L\right)\right) \leq \sqrt{d}\kappa_0\left(1+\left\lceil\frac{d}{2}\right\rceil\frac{\max_i\|\boldsymbol{a}_i\|}{h_0}\right)\end{array}\right] \leq n^{-d}+n^{-2.9d+1}.$$

Proof We first note that

$$\begin{aligned}&\underset{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}{\mathbf{Pr}}\left[\begin{array}{l}\exists L \in \binom{[n]}{\lfloor d/2-1\rfloor}, j_0 \notin L \text{ such that} \\ \quad \mathbf{dist}\left(\boldsymbol{a}_{j_0}, \mathbf{Span}\left(A_L\right)\right) \leq \sqrt{d}\kappa_0\left(1+\left\lceil\frac{d}{2}\right\rceil\frac{\max_i\|\boldsymbol{a}_i\|}{h_0}\right)\end{array}\right] \\ &\quad\leq \underset{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}{\mathbf{Pr}}\left[\begin{array}{l}\exists L \in \binom{[n]}{\lfloor d/2-1\rfloor}, j_0 \notin L \text{ such that} \\ \quad \mathbf{dist}\left(\boldsymbol{a}_{j_0}, \mathbf{Span}\left(A_L\right)\right) \leq \sqrt{d}\kappa_0\left(1+\left\lceil\frac{d}{2}\right\rceil\frac{1+3\sqrt{d\ln n}\sigma}{h_0}\right)\end{array}\right] \qquad (37) \\ &\qquad+\underset{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}{\mathbf{Pr}}\left[\max_i\|\boldsymbol{a}_i\| > 1+3\sqrt{d\ln n}\sigma\right]. \qquad (38)\end{aligned}$$

We now apply Proposition 2.4.7 to bound (37) by

$$\begin{aligned}&\sum_{L\in\binom{[n]}{\lfloor d/2-1\rfloor}}\sum_{j_0\notin L}\underset{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n}{\mathbf{Pr}}\left[\mathbf{dist}\left(\boldsymbol{a}_{j_0}, \mathbf{Span}\left(A_L\right)\right) \leq \sqrt{d}\kappa_0\left(1+\left\lceil\frac{d}{2}\right\rceil\frac{1+3\sqrt{d\ln n}\sigma}{h_0}\right)\right] \\ &\quad\leq \binom{n}{\lfloor d/2-1\rfloor}(n-d/2+1)\left(\frac{\sqrt{d}\kappa_0}{\sigma}\left(1+\left\lceil\frac{d}{2}\right\rceil\frac{1+3\sqrt{d\ln n}\sigma}{h_0}\right)\right)^{d-|L|} \qquad (39)\end{aligned}$$

To simplify this expression, we note that $\left\lceil\frac{d}{2}\right\rceil \leq \frac{2d}{3}$, for $d \geq 3$. We then recall

$$\frac{\kappa_0}{h_0} = \frac{\min(\sigma,1)}{3d^2n^3\sqrt{\ln n}},$$

and apply $d \geq 3$ to show

$$\begin{aligned}\frac{\sqrt{d}\kappa_0}{\sigma}\left(1+\left\lceil\frac{d}{2}\right\rceil\frac{1+3\sqrt{d\ln n}\sigma}{h_0}\right) &\leq \frac{\sqrt{d}\kappa_0}{\sigma}+\frac{\kappa_0}{h_0}\left(\frac{2d^{3/2}}{3\sigma}+2d^2\sqrt{\ln n}\right) \\ &\leq \frac{1}{n^3}.\end{aligned}$$

So, we have

$$\begin{aligned}(39) &\leq \binom{n}{\lfloor d/2-1\rfloor}(n-d/2+1)\left(\frac{1}{n^3}\right)^{\lceil d/2\rceil} &(40)\\ &\leq n^{\lfloor d/2-1\rfloor+1}n^{-3d/2} &(41)\\ &\leq n^{-d}. &(42)\end{aligned}$$

On the other hand, we can use Corollary 2.4.6, to bound (38) by $n^{-2.9d+1}$. ■

5.1.1 Discussion

It is natural to ask whether one could avoid the complication of this section by setting $I = \{1, \ldots, d\}$, or even choosing I to be the best d-set in $\{1, \ldots, d+k\}$ for some constant k. It is possible to show that the probability that all d-by-d minors of a perturbed d-by-$(d+k)$ matrix have condition number at most ϵ grows like $(\sqrt{d}\epsilon/\sigma)^k$. Thus, the best of these sets would have reasonable condition number with polynomially high probability. This bound would be sufficient to handle our concerns about the magnitude of y_i'. The analysis in Lemma 5.2.4 might still be possible in this situation; however, it would require considering multiple possible splittings of the perturbation (for multiple values of τ_1), and it is not clear whether such an analysis can be made rigorous. Finally, it seems difficult in this situation to apply the trick in the proofs of Lemma 5.3.1 and 5.2.1 of summing over all likely values for κ. If the algorithm is given σ as input, then it is possible to avoid the need for this trick (and an such an analysis appeared in an earlier draft of this paper). However, we believe that it is preferable for the algorithm to make sense without taking σ as an input.

While choosing I in such a simple fashion could possibly simplify this section, albeit at the cost of complicating others, we feel that once Lemma 5.1.1 has been improved and the correct concentration bound has been obtained, this technique will provide the best bounds.

One of the anonymous referees pointed out that it should be possible to use the rank revealing QR factorization to find an I with almost maximal $\mathbf{s_{min}}(\boldsymbol{A}_I)$ (see [CH92]). While doing so seems to be the best choice algorithmically, it is not clear to us how we could analyze the smoothed complexity of the resulting two-phase algorithm. The difficulty is that the assumption that a particular I was output by the rank revealing QR factorization would impose conditions on $\boldsymbol{A}$ that we are currently not able to analyze.

5.2 Bounding the shadow of LP'

Before beginning our analysis of the shadow of LP', we define the set from which $\boldsymbol{\alpha}$ is chosen to be A_{1/d^2}, where we define

$$A = \{\boldsymbol{\alpha} : \langle \boldsymbol{\alpha} | \mathbf{1} \rangle = 1\}, \text{ and}$$
$$A_\delta = \{\boldsymbol{\alpha} : \langle \boldsymbol{\alpha} | \mathbf{1} \rangle = 1 \text{ and } \alpha_i \geq \delta, \forall i\}.$$

The principal obstacle to proving the bound for LP' is that Theorem 4.0.1 requires one to specify the plane on which the shadow of the perturbed polytope will be measured before the perturbation is known, whereas the shadow relevant to the analysis of LP' depends on the perturbation—it is the shadow onto $\mathbf{Span}(\boldsymbol{A}\boldsymbol{\alpha}, \boldsymbol{z})$. To overcome this obstacle, we prove in Lemma 5.2.4 that if $\mathbf{s_{min}}\left(\bar{\boldsymbol{A}}_{\mathcal{I}(A)}\right) \geq \kappa_0/2$, then the expected size of the shadow onto $\mathbf{Span}(\boldsymbol{A}\boldsymbol{\alpha}, \boldsymbol{z})$ is close to the expected size of the shadow onto $\mathbf{Span}\left(\bar{\boldsymbol{A}}\bar{\boldsymbol{\alpha}}, \boldsymbol{z}\right)$, where $\bar{\boldsymbol{\alpha}}$ is chosen from A_0. As this plane is independent of the perturbation, we can apply Theorem 4.0.1 to bound the size of the shadow on this plane. Unfortunately, $\bar{\boldsymbol{A}}$ is arbitrary, so we cannot make any assumptions about $\mathbf{s_{min}}\left(\bar{\boldsymbol{A}}_{\mathcal{I}(A)}\right)$. Instead, we decompose the perturbation into two parts, as in Corollary 4.3.2, and can then use Corollary 5.1.2 to show that with high probability $\mathbf{s_{min}}\left(\tilde{\boldsymbol{A}}_{\mathcal{I}(A)}\right) \geq \kappa_0/2$. We begin the proof with this decomposition, and build to the point at which we can apply Lemma 5.2.4.

A secondary obstacle in the analysis is that κ and M are correlated with $\boldsymbol{A}$ and $\boldsymbol{y}$. We overcome this obstacle by considering the sum of the expected sizes of the shadows when κ and M are fixed to each of their likely values. This analysis is facilitated by the notation

$$\mathcal{T}'_{\boldsymbol{z}}(\boldsymbol{A}, I, \boldsymbol{\alpha}, \kappa, M) \stackrel{\text{def}}{=} \left|\mathbf{Shadow}_{A_I \boldsymbol{\alpha}, \boldsymbol{z}}\left(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n; \boldsymbol{y}'\right)\right|, \quad \text{where } y'_i = \begin{cases} M & \text{if } i \in I \\ \sqrt{d}M^2/4\kappa & \text{otherwise.} \end{cases}$$

We note that

$$\mathcal{S}'_{\boldsymbol{z}}(\boldsymbol{A}, \boldsymbol{y}, \mathcal{I}, \boldsymbol{\alpha}) = \mathcal{T}'_{\boldsymbol{z}}\left(\boldsymbol{A}, \mathcal{I}(\boldsymbol{A}), \boldsymbol{\alpha}, 2^{\lfloor \lg \mathbf{s_{min}}\left(\boldsymbol{A}_{\mathcal{I}(\boldsymbol{A})}\right) \rfloor}, 2^{\lceil \lg(\max_i \|(y_i, \boldsymbol{a}_i)\|) \rceil + 2}\right).$$

Lemma 5.2.1 (LP') *Let $d \geq 3$ and $n \geq d+1$. Let $\bar{\boldsymbol{A}} = [\bar{\boldsymbol{a}}_1, \ldots, \bar{\boldsymbol{a}}_n] \in \mathbb{R}^{n \times d}$, $\bar{\boldsymbol{y}} \in \mathbb{R}^n$ and $\boldsymbol{z} \in \mathbb{R}^d$ satisfy $\max_i \|(\bar{y}_i, \bar{\boldsymbol{a}}_i)\| \in (1/2, 1]$. For any $\sigma > 0$, let $\boldsymbol{A}$ be a Gaussian random matrix centered at $\bar{\boldsymbol{A}}$ of standard deviation σ, and let $\boldsymbol{y}$ by a Gaussian random vector centered at $\bar{\boldsymbol{y}}$ of standard deviation σ. Let $\boldsymbol{\alpha}$ be chosen uniformly at random from A_{1/d^2} and let $\mathcal{I}$ be a collection of $3nd \ln n$ randomly chosen d-subsets of $[n]$. Then,*

$$\mathop{\boldsymbol{E}}_{\boldsymbol{A}, \boldsymbol{y}, \mathcal{I}, \boldsymbol{\alpha}}\left[\mathcal{S}'_{\boldsymbol{z}}(\boldsymbol{A}, \boldsymbol{y}, \mathcal{I}, \boldsymbol{\alpha})\right] = 326nd(\ln n) \lg(dn/\min(1, \sigma))\ \mathcal{D}\left(d, n, \frac{\min(1, \sigma^4)}{12,960 d^{8.5} n^{14} \ln^{2.5} n}\right),$$

where $\mathcal{D}(d, n, \sigma)$ is as given in Theorem 4.0.1.

Proof Instead of treating $\boldsymbol{A}$ as a perturbation of standard deviation σ of $\bar{\boldsymbol{A}}$, we will view $\boldsymbol{A}$ as the result of applying a perturbation of standard deviation τ_0 followed by a

perturbation of standard deviation τ_1, where $\tau_0^2 + \tau_1^2 = \sigma^2$. Formally, we will let $\boldsymbol{G}$ be a Gaussian random matrix of standard deviation τ_0 centered at the origin, $\tilde{\boldsymbol{A}} = \bar{\boldsymbol{A}} + \boldsymbol{G}$, $\tilde{\boldsymbol{G}}$ be a Gaussian random matrix of standard deviation τ_1 centered at the origin, and $\boldsymbol{A} = \tilde{\boldsymbol{A}} + \tilde{\boldsymbol{G}}$, where

$$\tau_1 \stackrel{\text{def}}{=} \frac{\kappa_0}{6d^3\sqrt{\ln n}}.$$

and $\tau_0^2 = \sigma^2 - \tau_1^2$. We similarly decompose the perturbation to $\boldsymbol{y}$ into a perturbation of standard deviation τ_0 from which we obtain $\tilde{\boldsymbol{y}}$, and a perturbation of standard deviation τ_1 from which we obtain $\boldsymbol{y}$. We will let $\tilde{\boldsymbol{h}} = \boldsymbol{y} - \tilde{\boldsymbol{y}}$.

We can then apply Lemma 5.2.2 to show

$$\Pr_{\mathcal{I},\tilde{\boldsymbol{A}},\tilde{\boldsymbol{G}}}\left[\mathbf{s_{min}}\left(\tilde{\boldsymbol{A}}_{\mathcal{I}(\boldsymbol{A})}\right) < \kappa_0/2\right] < 0.42\binom{n}{d}^{-1}. \tag{43}$$

One difficulty in bounding the expectation of $\mathcal{T}'$ is that its input parameters are correlated. To resolve this difficulty, we will bound the expectation of $\mathcal{T}'$ by the sum over the expectations obtained by substituting each of the likely choices for κ and M.

In particular, we set

$$\mathcal{M} = \left\{2^{\lceil \lg x\rceil+2} : \left(\max_i \|(\tilde{y}_i, \tilde{\boldsymbol{a}}_i)\|\right) - 3\sqrt{d\ln n}\tau_1 \le x \le \left(\max_i \|(\tilde{y}_i, \tilde{\boldsymbol{a}}_i)\|\right) + 3\sqrt{d\ln n}\tau_1\right\}.$$

We now define indicator random variables V, W, X, Y, and Z by

$$\begin{aligned}
V &= [|\mathcal{M}| \le 2]\,,\\
W &= \left[\max_i \|(\tilde{y}_i, \tilde{\boldsymbol{a}}_i)\| \le 1 + 3\sqrt{(d+1)\ln n}\sigma\right],\\
X &= \left[\mathbf{s_{min}}\left(\tilde{\boldsymbol{A}}_{\mathcal{I}(\boldsymbol{A})}\right) \ge \kappa_0/2\right],\\
Y &= \left[2^{\lfloor \lg \mathbf{s_{min}}(\boldsymbol{A}_{\mathcal{I}(\boldsymbol{A})})\rfloor} \in \mathcal{K}\right], \text{ and}\\
Z &= \left[2^{\lceil \lg \max_i \|(y_i,\boldsymbol{a}_i)\|\rceil+2} \in \mathcal{M}\right],
\end{aligned}$$

and then expand

$$\begin{aligned}
&\mathop{\mathbf{E}}_{\mathcal{I},A,\boldsymbol{y},\boldsymbol{\alpha}}\left[\mathcal{S}'(\boldsymbol{A},\boldsymbol{y},\mathcal{I},\boldsymbol{\alpha})\right]\\
&\quad= \mathop{\mathbf{E}}_{\mathcal{I},A,\boldsymbol{y},\boldsymbol{\alpha}}\left[\mathcal{S}'(\boldsymbol{A},\boldsymbol{y},\mathcal{I},\boldsymbol{\alpha})VWXYZ\right] + \mathop{\mathbf{E}}_{\mathcal{I},A,\boldsymbol{y},\boldsymbol{\alpha}}\left[\mathcal{S}'(\boldsymbol{A},\boldsymbol{y},\mathcal{I},\boldsymbol{\alpha})(1-VWXYZ)\right].
\end{aligned} \tag{44}$$

From Corollary 5.1.3, we know

$$\Pr_{\boldsymbol{A},\mathcal{I}}[\text{not}(Y)] = \Pr_{\boldsymbol{A},\mathcal{I}}\left[2^{\lfloor \lg \mathbf{s_{min}}(\boldsymbol{A}_{\mathcal{I}(\boldsymbol{A})})\rfloor} \notin \mathcal{K}\right] \le 0.42\binom{n}{d}^{-1}. \tag{45}$$

Similarly, Corollary 2.4.6 implies for any $\tilde{\boldsymbol{A}}$ and $\tilde{\boldsymbol{y}}$ and $n > d \ge 3$.

$$\Pr_{\tilde{\boldsymbol{G}},\tilde{\boldsymbol{h}},\mathcal{I}}[\text{not}(Z)] = \Pr_{\tilde{\boldsymbol{G}},\tilde{\boldsymbol{h}},\mathcal{I}}\left[2^{\lceil \lg \max_i \|(y_i,\boldsymbol{a}_i)\|\rceil+2} \notin \mathcal{M}\right] \le 0.0015\binom{n}{d}^{-1}. \tag{46}$$

From Corollary 2.4.6 we have

$$\mathop{\mathbf{Pr}}_{\tilde{\boldsymbol{A}},\tilde{\boldsymbol{y}}}[\mathrm{not}(W)] \leq n^{-2.9(d+1)+1} \leq 0.0015\binom{n}{d}^{-1}.$$

For i_0 an index for which $\|(y_{i_0}, \boldsymbol{a}_{i_0})\| \geq 1/2$, Proposition 2.4.9 implies

$$\mathop{\mathbf{Pr}}_{\tilde{\boldsymbol{A}}}[\mathrm{not}(V)] \leq \mathop{\mathbf{Pr}}_{\tilde{\boldsymbol{a}}_{i_0},\tilde{y}_{i_0}}\left[\|(\tilde{y}_{i_0}, \tilde{\boldsymbol{a}}_{i_0})\| < 9\sqrt{(d+1)\ln n}\tau_1\right] \leq 0.01\binom{n}{d}^{-1}.$$

By also applying inequality (43) to bound the probability of $\mathrm{not}(X)$, we find

$$\mathop{\mathbf{Pr}}_{\boldsymbol{A},\boldsymbol{y},\mathcal{I}}[(1 - VWXYZ) = 1] \leq 0.86\binom{n}{d}^{-1}.$$

As

$$\mathcal{S}'(\boldsymbol{A}, \boldsymbol{y}, \mathcal{I}, \boldsymbol{\alpha}) \leq \binom{n}{d}, \text{ (by Proposition 5.0.2)}$$

the second term of (44) can be bounded by 1.

To bound the first term of (44), we note

$$\begin{aligned}
&\mathop{\mathbf{E}}_{\mathcal{I},A,\boldsymbol{y},\boldsymbol{\alpha}}\left[\mathcal{S}'(\boldsymbol{A}, \boldsymbol{y}, \mathcal{I}, \boldsymbol{\alpha})VWXYZ\right] \\
&\qquad \leq \mathop{\mathbf{E}}_{\mathcal{I},\tilde{\boldsymbol{A}},\tilde{\boldsymbol{y}}}\left[VW \sum_{\kappa\in\mathcal{K},M\in\mathcal{M}} \mathop{\mathbf{E}}_{\tilde{\boldsymbol{G}},\tilde{\boldsymbol{h}},\boldsymbol{\alpha}}\left[\mathcal{T}'\left(\boldsymbol{A}, \mathcal{I}(\boldsymbol{A}), \boldsymbol{\alpha}, \kappa, M\right) XW\right]\right] \qquad (47)
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\mathop{\mathbf{E}}_{\tilde{\boldsymbol{G}},\tilde{\boldsymbol{h}},\boldsymbol{\alpha}}\left[\mathcal{T}'\left(\boldsymbol{A}, \mathcal{I}(\boldsymbol{A}), \boldsymbol{\alpha}, \kappa, M\right) XW\right] \\
&= \mathop{\mathbf{E}}_{\tilde{\boldsymbol{G}},\tilde{\boldsymbol{h}},\boldsymbol{\alpha}}\left[\sum_{I\in\mathcal{I}} \mathcal{T}'\left(\boldsymbol{A}, I, \boldsymbol{\alpha}, \kappa, M\right) W\left[\mathbf{s_{min}}\left(\tilde{\boldsymbol{A}}_I\right) \geq \kappa_0/2\right][\mathcal{I}(\boldsymbol{A}) = I]\right] \\
&\leq \mathop{\mathbf{E}}_{\tilde{\boldsymbol{G}},\tilde{\boldsymbol{h}},\boldsymbol{\alpha}}\left[\sum_{I\in\mathcal{I}} \mathcal{T}'\left(\boldsymbol{A}, I, \boldsymbol{\alpha}, \kappa, M\right) W\left[\mathbf{s_{min}}\left(\tilde{\boldsymbol{A}}_I\right) \geq \kappa_0/2\right]\right] \\
&\leq \mathop{\mathbf{E}}_{\tilde{\boldsymbol{G}},\tilde{\boldsymbol{h}},\boldsymbol{\alpha}}\left[\sum_{I\in\mathcal{I}} \mathcal{T}'\left(\boldsymbol{A}, I, \boldsymbol{\alpha}, \kappa, M\right) \middle| W \text{ and } \mathbf{s_{min}}\left(\tilde{\boldsymbol{A}}_I\right) \geq \kappa_0/2\right] \\
&= \sum_{I\in\mathcal{I}} \mathop{\mathbf{E}}_{\tilde{\boldsymbol{G}},\tilde{\boldsymbol{h}},\boldsymbol{\alpha}}\left[\mathcal{T}'\left(\boldsymbol{A}, I, \boldsymbol{\alpha}, \kappa, M\right) \middle| W \text{ and } \mathbf{s_{min}}\left(\tilde{\boldsymbol{A}}_I\right) \geq \kappa_0/2\right] \\
&\leq \sum_{I\in\mathcal{I}} (6 + 10^{-4})\mathcal{D}\left(d, n, \frac{\tau_1}{(2 + 3\sqrt{d\ln n}\sigma)(\sqrt{d}M^2/4\kappa M)}\right), \qquad \text{by Lemma 5.2.3,} \\
&\leq 3(6 + 10^{-4})nd(\ln n)\left(\mathcal{D}\left(d, n, \frac{4\tau_1\kappa}{(2 + 3\sqrt{d\ln n}\sigma)(\sqrt{d}M)}\right)\right).
\end{aligned}$$

Thus,

$$
\begin{aligned}
(47) &\leq \mathop{\mathbf{E}}_{\mathcal{I},\tilde{\boldsymbol{A}},\tilde{\boldsymbol{y}}}\left[VW\sum_{\kappa\in\mathcal{K},M\in\mathcal{M}} 3(6+10^{-4})nd(\ln n)\mathcal{D}\left(d,n,\frac{4\tau_1\kappa}{(2+3\sqrt{d\ln n}\sigma)(\sqrt{d}M)}\right)\right]\\
&\leq \mathop{\mathbf{E}}_{\mathcal{I},\tilde{\boldsymbol{A}},\tilde{\boldsymbol{y}}}\left[3(6+10^{-4})nd(\ln n)(V\left|\mathcal{M}\right|)\left|\mathcal{K}\right|W\mathcal{D}\left(d,n,\frac{4\tau_1\min(\mathcal{K})}{(2+3\sqrt{d\ln n}\sigma)(\sqrt{d}\max(\mathcal{M}))}\right)\right]\\
&\leq \mathop{\mathbf{E}}_{\mathcal{I},\tilde{\boldsymbol{A}},\tilde{\boldsymbol{y}}}\left[6(6+10^{-4})nd(\ln n)\left|\mathcal{K}\right|W\mathcal{D}\left(d,n,\frac{4\tau_1\min(\mathcal{K})}{(2+3\sqrt{d\ln n}\sigma)(\sqrt{d}\max(\mathcal{M}))}\right)\right]\\
&\leq 6(6+10^{-4})nd(\ln n)\left|\mathcal{K}\right|\mathcal{D}\left(d,n,\frac{2\tau_1\kappa_0}{\sqrt{d}(2+3\sqrt{d\ln n}\sigma)(1+6\sqrt{(d+1)\ln n}\sigma)}\right),
\end{aligned}
$$

where the last inequality follows from $\max(\mathcal{M}) \leq 1+6\sqrt{(d+1)\ln n}\sigma$ when W is true.

To simplify, we first bound the third argument of the function $\mathcal{D}$ by:

$$
\begin{aligned}
&\frac{2\tau_1\kappa_0}{\sqrt{d}(2+3\sqrt{d\ln n}\sigma)(1+6\sqrt{(d+1)\ln n}\sigma)}\\
&= \frac{1}{3d^3\sqrt{\ln n}}\frac{\kappa_0^2}{\sqrt{d}(2+3\sqrt{d\ln n}\sigma)(1+6\sqrt{(d+1)\ln n}\sigma)}\\
&= \frac{1}{3d^{3.5}\sqrt{\ln n}}\left(\frac{1}{12d^2n^7\sqrt{\ln n}}\right)^2\frac{\sigma^2(\min(1,\sigma))^2}{(2+3\sqrt{d\ln n}\sigma)(1+6\sqrt{(d+1)\ln n}\sigma)}\\
&\geq \frac{1}{432d^{7.5}n^{14}(\ln n)^{1.5}}\frac{\min(1,\sigma^4)}{(2+3\sqrt{d\ln n})(1+6\sqrt{(d+1)\ln n})}\\
&\geq \frac{1}{432d^{7.5}n^{14}(\ln n)^{1.5}}\frac{\min(1,\sigma^4)}{30d\ln n}\\
&= \frac{\min(1,\sigma^4)}{12,960d^{8.5}n^{14}\ln^{2.5}n}
\end{aligned}
$$

where the last inequality follows from the assumption that $n > d \geq 3$.

Applying Proposition 5.1.4 to show $|\mathcal{K}| \leq 9\lg(dn/\min(1,\sigma))$, we now obtain

$$
\begin{aligned}
(47) &\leq 6(6+10^{-4})\left|\mathcal{K}\right|nd(\ln n)\mathcal{D}\left(d,n,\frac{\min(1,\sigma^4)}{12,960d^{8.5}n^{14}\ln^{2.5}n}\right)\\
&\leq 325nd(\ln n)\lg(dn/\min(1,\sigma))\mathcal{D}\left(d,n,\frac{\min(1,\sigma^4)}{12,960d^{8.5}n^{14}\ln^{2.5}n}\right).
\end{aligned}
$$

■

Lemma 5.2.2 (probability of small $\mathbf{s_{min}}\left(\tilde{\boldsymbol{A}}_{\boldsymbol{\mathcal{I}(A)}}\right)$) *For $\boldsymbol{A}$, $\tilde{\boldsymbol{A}}$, and $\mathcal{I}$ as defined in the proof of Lemma 5.2.1,*

$$
\mathop{\boldsymbol{Pr}}_{\mathcal{I},\tilde{\boldsymbol{A}},\tilde{\boldsymbol{G}}}\left[\mathbf{s}_{\boldsymbol{min}}\left(\tilde{\boldsymbol{A}}_{\mathcal{I}(\boldsymbol{A})}\right) < \kappa_0/2\right] < 0.42\binom{n}{d}^{-1}.
$$

Proof Let $I = \mathcal{I}(\boldsymbol{A})$, we have

$$\mathbf{Pr}\left[\mathbf{s_{min}}\left(\tilde{\boldsymbol{A}}_I\right) < \kappa_0/2\right] \leq \frac{\mathbf{Pr}\left[\mathbf{s_{min}}\left(\boldsymbol{A}_I\right) < \kappa_0\right]}{\mathbf{Pr}\left[\mathbf{s_{min}}\left(\boldsymbol{A}_I\right) < \kappa_0 \middle| \mathbf{s_{min}}\left(\tilde{\boldsymbol{A}}_I\right) < \kappa_0/2\right]}.$$

From Corollary 5.1.2, we have

$$\mathbf{Pr}\left[\mathbf{s_{min}}\left(\boldsymbol{A}_I\right) < \kappa_0\right] \leq 0.417\binom{n}{d}^{-1},$$

On the other hand, we have

$$\begin{aligned}
&\mathbf{Pr}\left[\mathbf{s_{min}}\left(\boldsymbol{A}_I\right) \geq \kappa_0 \middle| \mathbf{s_{min}}\left(\tilde{\boldsymbol{A}}_I\right) < \kappa_0/2\right] \\
&\leq \mathbf{Pr}\left[\mathbf{s_{min}}\left(\boldsymbol{A}_I\right) \geq \kappa_0 \text{ and } \mathbf{s_{min}}\left(\tilde{\boldsymbol{A}}_I\right) < \kappa_0/2\right] \\
&\leq \mathbf{Pr}\left[\left\|\boldsymbol{A}_I - \tilde{\boldsymbol{A}}_I\right\| \geq \kappa_0/2\right], && \text{by Proposition 2.2.6 (b),} \\
&\leq \mathop{\mathbf{Pr}}_{\boldsymbol{A}}\left[\max_i \|\boldsymbol{a}_i - \tilde{\boldsymbol{a}}_i\| \geq \kappa_0/2\sqrt{d}\right], && \text{by Proposition 2.2.4 (d),} \\
&= \mathop{\mathbf{Pr}}_{\boldsymbol{A}}\left[\max_i \|\boldsymbol{a}_i - \tilde{\boldsymbol{a}}_i\| \geq 3d^{5/2}\sqrt{\ln n}\tau_1\right] \\
&\leq n^{-2.9d+1},
\end{aligned}$$

by Corollary 2.4.6. Thus,

$$(5.2) \leq \frac{0.417\binom{n}{d}^{-1}}{1 - n^{-2.9d+1}} \leq 0.42\binom{n}{d}^{-1},$$

for $n > d \geq 3$. ■

Lemma 5.2.3 (From $\tilde{\boldsymbol{a}}$) *Let I be a set in $\binom{[n]}{d}$ and let $\tilde{\boldsymbol{a}}_1, \ldots, \tilde{\boldsymbol{a}}_n$ be points each of norm at most $1 + 3\sqrt{(d+1)\ln n}\sigma$ such that*

$$\mathbf{s}_{\boldsymbol{min}}\left(\tilde{\boldsymbol{A}}_I\right) \geq \kappa_0/2.$$

Then,

$$\mathop{\boldsymbol{E}}_{\boldsymbol{A},\boldsymbol{\alpha}\in A_{1/d^2}}\left[\mathbf{Shadow}_{A_I\boldsymbol{\alpha},\boldsymbol{z}}\left(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n; \boldsymbol{y}'\right)\right] \leq (6+10^{-4})\mathcal{D}\left(d, n, \frac{\tau_1}{(2 + 3\sqrt{d\ln n}\sigma)(\max_i y_i'/\min_i y_i')}\right). \tag{48}$$

Proof We apply Lemma 5.2.4 to show

$$\begin{aligned}
\mathop{\mathbf{E}}_{\boldsymbol{A},\boldsymbol{\alpha}\in A_{1/d^2}}\left[\left|\mathbf{Shadow}_{\boldsymbol{A}_I\boldsymbol{\alpha},\boldsymbol{z}}\left(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n;\boldsymbol{y}'\right)\right|\right] &\leq 6\mathop{\mathbf{E}}_{\boldsymbol{A},\tilde{\boldsymbol{\alpha}}\in A_0}\left[\left|\mathbf{Shadow}_{\tilde{\boldsymbol{A}}_I\tilde{\boldsymbol{\alpha}},\boldsymbol{z}}\left(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n;\boldsymbol{y}'\right)\right|\right]+1\\
&\leq 6\max_{\tilde{\boldsymbol{\alpha}}\in A_0}\mathop{\mathbf{E}}_{\boldsymbol{A}}\left[\left|\mathbf{Shadow}_{\tilde{\boldsymbol{A}}_I\tilde{\boldsymbol{\alpha}},\boldsymbol{z}}\left(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n;\boldsymbol{y}'\right)\right|\right]+1\\
&\leq 6\mathcal{D}\left(d,n,\frac{\tau_1}{(2+3\sqrt{d\ln n}\sigma)(\max_i y_i'/\min_i y_i')}\right)+7\\
&\leq (6+10^{-4})\mathcal{D}\left(d,n,\frac{\tau_1}{(2+3\sqrt{d\ln n}\sigma)(\max_i y_i'/\min_i y_i')}\right),
\end{aligned}$$

by Corollary 4.3.3 and fact that $\mathcal{D}(n,d,\sigma)\geq 58,888,678$ for any positive n,d,σ. ■

Lemma 5.2.4 (Changing $\boldsymbol{\alpha}$ to $\tilde{\boldsymbol{\alpha}}$) *Let $I\in\binom{[n]}{d}$. Let $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n$ be Gaussian random vectors in $\mathbb{R}^d$ of standard deviation τ_1, centered at points $\tilde{\boldsymbol{a}}_1,\ldots,\tilde{\boldsymbol{a}}_n$. If $\mathbf{s}_{\boldsymbol{min}}\left(\tilde{\boldsymbol{A}}_I\right)\geq\kappa_0/2$, then*

$$\mathop{\boldsymbol{E}}_{\boldsymbol{A},\boldsymbol{\alpha}\in A_{1/d^2}}\left[\left|\mathbf{Shadow}_{\boldsymbol{A}_I\boldsymbol{\alpha},\boldsymbol{z}}\left(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n;\boldsymbol{y}'\right)\right|\right]\leq 6\mathop{\boldsymbol{E}}_{\boldsymbol{A},\tilde{\boldsymbol{\alpha}}\in A_0}\left[\left|\mathbf{Shadow}_{\tilde{\boldsymbol{A}}_I\tilde{\boldsymbol{\alpha}},\boldsymbol{z}}\left(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n;\boldsymbol{y}'\right)\right|\right]+1.$$

Proof The key to our proof is Lemma 5.2.5. To ready ourselves for the application of this lemma, we let

$$\mathcal{F}_{\boldsymbol{A}}(\boldsymbol{t})=\left|\mathbf{Shadow}_{\boldsymbol{t},\boldsymbol{z}}\left(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n;\boldsymbol{y}'\right)\right|,$$

and note that $\mathcal{F}_{\boldsymbol{A}}(\boldsymbol{t})=\mathcal{F}_{\boldsymbol{A}}(\boldsymbol{t}/\|\boldsymbol{t}\|)$. If $\left\|\tilde{\boldsymbol{A}}-\boldsymbol{A}\right\|\leq 3d\sqrt{\ln n}\tau_1$, then

$$\left\|I-\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}\right\| \leq \left\|\tilde{\boldsymbol{A}}^{-1}\right\|\left\|\tilde{\boldsymbol{A}}-\boldsymbol{A}\right\| \leq \left(\frac{2}{\kappa_0}\right)3d\sqrt{\ln n}\tau_1 \leq \left(\frac{2}{\kappa_0}\right)\frac{3d\sqrt{\ln n}\kappa_0}{12d^3\sqrt{\ln n}} \leq \frac{1}{2d^2}.$$

By Proposition 2.2.6 (b),

$$\begin{aligned}
\mathbf{s_{min}}\left(\boldsymbol{A}_I\right) &\geq \mathbf{s_{min}}\left(\tilde{\boldsymbol{A}}_I\right)-\left\|\tilde{\boldsymbol{A}}-\boldsymbol{A}\right\|\\
&\geq \kappa_0/2-3d\sqrt{\ln n}\tau_1\\
&\geq \frac{\kappa_0}{2}\left(1-\frac{1}{2d^2}\right)\\
&\geq \frac{\kappa_0}{2}\left(\frac{17}{18}\right),
\end{aligned}$$

for $d\geq 3$. So, we can similarly bound

$$\left\|I-\boldsymbol{A}^{-1}\tilde{\boldsymbol{A}}\right\|\leq\frac{9}{17d^2}.$$

We can then apply Lemma 5.2.5 to show

$$\mathop{\mathbf{E}}_{\boldsymbol{\alpha}\in A_{1/d^2}}\left[\left|\mathbf{Shadow}_{\boldsymbol{A}_I\boldsymbol{\alpha},\boldsymbol{z}}\left(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n;\boldsymbol{y}'\right)\right|\right]\leq 6\mathop{\mathbf{E}}_{\tilde{\boldsymbol{\alpha}}\in A}\left[\left|\mathbf{Shadow}_{\tilde{\boldsymbol{A}}_I\tilde{\boldsymbol{\alpha}},\boldsymbol{z}}\left(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n;\boldsymbol{y}'\right)\right|\right].$$

From Corollary 2.4.6 and Proposition 2.2.4 (d), we know that the probability that $\left\|\tilde{\boldsymbol{A}} - \boldsymbol{A}\right\| > 3d\sqrt{\ln n}\tau_1$ is at most $n^{-2.9d+1}$. As $\mathbf{Shadow}_{\tilde{\boldsymbol{A}}_I\tilde{\boldsymbol{\alpha}},\boldsymbol{z}}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n;\boldsymbol{y}') \leq \binom{n}{d}$, we can apply Lemma 2.3.3 to show

$$\mathop{\mathbf{E}}_{\boldsymbol{A}}\left[\mathop{\mathbf{E}}_{\boldsymbol{\alpha}\in A_{1/d^2}}\left[\left|\mathbf{Shadow}_{\boldsymbol{A}_I\boldsymbol{\alpha},\boldsymbol{z}}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n;\boldsymbol{y}')\right|\right]\right] \leq 6\mathop{\mathbf{E}}_{\boldsymbol{A}}\left[\mathop{\mathbf{E}}_{\tilde{\boldsymbol{\alpha}}\in A}\left[\left|\mathbf{Shadow}_{\tilde{\boldsymbol{A}}_I\tilde{\boldsymbol{\alpha}},\boldsymbol{z}}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n;\boldsymbol{y}')\right|\right]\right]+1.$$

■

To compare the expected sizes of the shadows, we will show that the distribution $\mathbf{Span}(\boldsymbol{A}\boldsymbol{\alpha},\boldsymbol{z})$ is close to the distribution $\mathbf{Span}\left(\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}},\boldsymbol{z}\right)$. To this end, we note that for a given $\tilde{\boldsymbol{\alpha}} \in A_0$ the $\boldsymbol{\alpha} \in A$ for which $\boldsymbol{A}\boldsymbol{\alpha}$ is a positive multiple of $\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}}$ is given by

$$\boldsymbol{\alpha} = \Psi(\tilde{\boldsymbol{\alpha}}) \stackrel{\text{def}}{=} \frac{\boldsymbol{A}^{-1}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}}}{\left\langle \boldsymbol{A}^{-1}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}}|\boldsymbol{1}\right\rangle}. \tag{49}$$

To derive this equation, note that $\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}}$ is the point in $\triangle(\tilde{\boldsymbol{a}}_1,\ldots,\tilde{\boldsymbol{a}}_d)$ specified by $\tilde{\boldsymbol{\alpha}}$. $\boldsymbol{A}^{-1}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}}$ provides the coordinates of this point in the basis $\boldsymbol{A}$. Dividing by $\left\langle \boldsymbol{A}^{-1}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}}|\boldsymbol{1}\right\rangle$ provides the $\boldsymbol{\alpha} \in A$ specifying the parallel point in $\mathbf{Aff}(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_d)$. We can similarly derive

$$\Psi^{-1}(\boldsymbol{\alpha}) = \frac{\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}\boldsymbol{\alpha}}{\left\langle \tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}\boldsymbol{\alpha}|\boldsymbol{1}\right\rangle}.$$

Our analysis will follow from a bound on the Jacobian of Ψ.

Lemma 5.2.5 (Approximation of $\boldsymbol{\alpha}$ by $\tilde{\boldsymbol{\alpha}}$) *Let $\mathcal{F}(\boldsymbol{x})$ be a non-negative function depending only on $\boldsymbol{x}/\|\boldsymbol{x}\|$. If $\delta = 1/d^2$, $\left\|I - \tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}\right\| \leq \epsilon$, and $\left\|I - \boldsymbol{A}^{-1}\tilde{\boldsymbol{A}}\right\| \leq \epsilon$, where $\epsilon \leq 9/17d^2$, then*

$$\mathop{\boldsymbol{E}}_{\boldsymbol{\alpha}\in A_\delta}[\mathcal{F}(\boldsymbol{A}\boldsymbol{\alpha})] \leq 6\mathop{\boldsymbol{E}}_{\tilde{\boldsymbol{\alpha}}\in A_0}\left[\mathcal{F}(\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}})\right]$$

Proof Expressing the expectations as integrals, the lemma is equivalent to

$$\frac{1}{\mathbf{Vol}(A_\delta)}\int_{\boldsymbol{\alpha}\in A_\delta}\mathcal{F}(\boldsymbol{A}\boldsymbol{\alpha})\,d\boldsymbol{\alpha} \leq \frac{6}{\mathbf{Vol}(A_0)}\int_{\tilde{\boldsymbol{\alpha}}\in A_0}\mathcal{F}(\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}})\,d\tilde{\boldsymbol{\alpha}}\,.$$

Applying Lemma 5.2.7 and setting $\boldsymbol{\alpha} = \Psi(\tilde{\boldsymbol{\alpha}})$, we bound

$$\begin{aligned}
\frac{1}{\mathbf{Vol}(A_\delta)}\int_{\boldsymbol{\alpha}\in A_\delta}\mathcal{F}(\boldsymbol{A}\boldsymbol{\alpha})\,d\boldsymbol{\alpha} &\leq \frac{1}{\mathbf{Vol}(A_\delta)}\int_{\boldsymbol{\alpha}\in\Psi(A_0)}\mathcal{F}(\boldsymbol{A}\boldsymbol{\alpha})\,d\boldsymbol{\alpha} \\
&= \frac{1}{\mathbf{Vol}(A_\delta)}\int_{\tilde{\boldsymbol{\alpha}}\in A_0}\mathcal{F}(\boldsymbol{A}\Psi(\tilde{\boldsymbol{\alpha}}))\left|\frac{\partial\Psi(\tilde{\boldsymbol{\alpha}})}{\partial\tilde{\boldsymbol{\alpha}}}\right|\,d\tilde{\boldsymbol{\alpha}} \\
&= \frac{1}{\mathbf{Vol}(A_\delta)}\int_{\tilde{\boldsymbol{\alpha}}\in A_0}\mathcal{F}(\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}})\left|\frac{\partial\Psi(\tilde{\boldsymbol{\alpha}})}{\partial\tilde{\boldsymbol{\alpha}}}\right|\,d\tilde{\boldsymbol{\alpha}}
\end{aligned}$$

(as $\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}}$ is a positive multiple of $\boldsymbol{A}\Psi(\tilde{\boldsymbol{\alpha}})$ and $\mathcal{F}(\boldsymbol{x})$ only depends on $\boldsymbol{x}/\left\|\boldsymbol{x}\right\|$)

$$\begin{aligned}
&\leq \max_{\tilde{\boldsymbol{\alpha}}\in A_0}\left(\left|\frac{\partial\Psi(\tilde{\boldsymbol{\alpha}})}{\partial\tilde{\boldsymbol{\alpha}}}\right|\right)\frac{1}{\mathbf{Vol}\,(A_\delta)}\int_{\tilde{\boldsymbol{\alpha}}\in A_0}\mathcal{F}(\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}})\,d\tilde{\boldsymbol{\alpha}}\\
&= \max_{\tilde{\boldsymbol{\alpha}}\in A_0}\left(\left|\frac{\partial\Psi(\tilde{\boldsymbol{\alpha}})}{\partial\tilde{\boldsymbol{\alpha}}}\right|\right)\left(\frac{\mathbf{Vol}\,(A_0)}{\mathbf{Vol}\,(A_\delta)}\right)\frac{1}{\mathbf{Vol}\,(A_0)}\int_{\tilde{\boldsymbol{\alpha}}\in A_0}\mathcal{F}(\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}})\,d\tilde{\boldsymbol{\alpha}}\\
&\leq \frac{(1+\epsilon)^d}{(1-\epsilon\sqrt{d})^d(1-\epsilon)}\left(\frac{1}{1-d\delta}\right)^d\frac{1}{\mathbf{Vol}\,(A_0)}\int_{\tilde{\boldsymbol{\alpha}}\in A_0}\mathcal{F}(\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}})\,d\tilde{\boldsymbol{\alpha}}
\end{aligned}$$

(by Proposition 5.2.6 and Lemma 5.2.10)

$$\leq 6\frac{1}{\mathbf{Vol}\,(A_0)}\int_{\tilde{\boldsymbol{\alpha}}\in A_0}\mathcal{F}(\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}})\,d\tilde{\boldsymbol{\alpha}}\,,$$

for $\epsilon \leq 9/17d^2$, $\delta = 1/d^2$ and $d \geq 3$. ■

Proposition 5.2.6 (Volume dilation)

$$\frac{\mathbf{Vol}\,(A_0)}{\mathbf{Vol}\,(A_\delta)} = \left(\frac{1}{1-d\delta}\right)^d.$$

Proof The set A_δ may be obtained by contracting the set A_0 at the point $(1/d, 1/d, \ldots, 1/d)$ by the factor $(1-d\delta)$. ■

Lemma 5.2.7 (Proper subset) *Under the conditions of Lemma 5.2.5,*

$$A_\delta \subset \Psi(A_0).$$

Proof We will prove

$$\Psi^{-1}(A_\delta) \subset A_0.$$

Let $\boldsymbol{\alpha} \in A_\delta$, $\boldsymbol{\alpha}' = \tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}\boldsymbol{\alpha}$ and $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}'/\left\langle\boldsymbol{\alpha}'|\mathbf{1}\right\rangle$. Using Proposition 2.2.2 to show $\left\|\boldsymbol{\alpha}\right\| \leq \left\|\boldsymbol{\alpha}\right\|_1 = 1$ and Proposition 2.2.4 (a), we bound

$$\alpha_i' \quad\geq\quad \alpha_i - \left|\alpha_i - \alpha_i'\right| \quad\geq\quad \delta - \left\|\boldsymbol{\alpha}-\boldsymbol{\alpha}'\right\| \quad\geq\quad \delta - \left\|I - \tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}\right\|\left\|\boldsymbol{\alpha}\right\| \quad\geq\quad \delta-\epsilon \quad>\quad 0.$$

So, all components of $\boldsymbol{\alpha}'$ are positive and therefore all components of $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}'/\left\langle\boldsymbol{\alpha}'|\mathbf{1}\right\rangle$ are positive. ■

We will now begin a study of the Jacobian of Ψ. This study will be simplified by decomposing Ψ into the composition of two maps. The second of these maps is given by:

Definition 5.2.8 ($\boldsymbol{\Gamma_{u,v}}$) *Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be vectors in $\mathbb{R}^d$ and let $\Gamma_{\boldsymbol{u},\boldsymbol{v}}(\boldsymbol{x})$ be the map from $\{\boldsymbol{x} : \langle\boldsymbol{x}|\boldsymbol{u}\rangle = 1\}$ to $\{\boldsymbol{x} : \langle\boldsymbol{x}|\boldsymbol{v}\rangle = 1\}$ by*

$$\Gamma_{\boldsymbol{u},\boldsymbol{v}}(\boldsymbol{x}) = \frac{\boldsymbol{x}}{\langle\boldsymbol{x}|\boldsymbol{v}\rangle}.$$

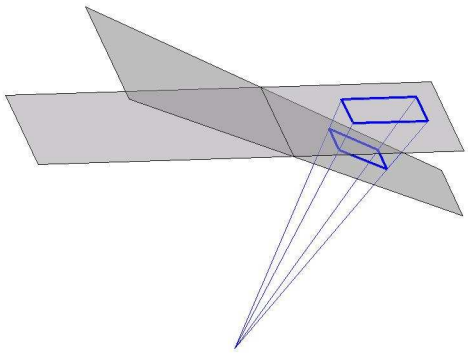

Figure 5: $\Gamma_{\boldsymbol{u},\boldsymbol{v}}$ can be understood as the projection through the origin from one plane onto the other.

Lemma 5.2.9 (Jacobian of Ψ)

$$\left|\frac{\partial\Psi(\tilde{\boldsymbol{\alpha}})}{\partial\tilde{\boldsymbol{\alpha}}}\right| = \det\left(\boldsymbol{A}^{-1}\tilde{\boldsymbol{A}}\right)\frac{\|\mathbf{1}\|}{\left\langle \boldsymbol{A}^{-1}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}}|\mathbf{1}\right\rangle^{d}\left\|\left(\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}\right)^{T}\mathbf{1}\right\|}.$$

Proof Let $\boldsymbol{\alpha} = \Psi(\tilde{\boldsymbol{\alpha}})$ and let $\boldsymbol{\alpha}' = \boldsymbol{A}^{-1}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}}$. As $\langle\tilde{\boldsymbol{\alpha}}|\mathbf{1}\rangle = 1$, we have

$$\left\langle \boldsymbol{\alpha}'|\left(\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}\right)^{T}\mathbf{1}\right\rangle = 1.$$

So, $\boldsymbol{\alpha} = \Gamma_{\boldsymbol{u},\boldsymbol{v}}(\boldsymbol{\alpha}')$, where $\boldsymbol{u} = \left(\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}\right)^{T}\mathbf{1}$ and $\boldsymbol{v} = \mathbf{1}$. By Lemma 5.2.11,

$$\begin{aligned}\left|\frac{\partial\boldsymbol{\alpha}}{\partial\tilde{\boldsymbol{\alpha}}}\right| = \left|\frac{\partial\boldsymbol{\alpha}}{\partial\boldsymbol{\alpha}'}\right|\left|\frac{\partial\boldsymbol{\alpha}'}{\partial\tilde{\boldsymbol{\alpha}}}\right| &= \det\left(\frac{\partial\Gamma_{\boldsymbol{u},\boldsymbol{v}}(\boldsymbol{\alpha}')}{\partial\boldsymbol{\alpha}'}\right)\det\left(\boldsymbol{A}^{-1}\tilde{\boldsymbol{A}}\right)\\ &= \det\left(\boldsymbol{A}^{-1}\tilde{\boldsymbol{A}}\right)\frac{\|\mathbf{1}\|}{\left\langle \boldsymbol{A}^{-1}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}}|\mathbf{1}\right\rangle^{d}\left\|\left(\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}\right)^{T}\mathbf{1}\right\|}.\end{aligned}$$

■

Lemma 5.2.10 (Bound on Jacobian of Ψ) *Under the conditions of Lemma 5.2.5,*

$$\left|\frac{\partial\Psi(\tilde{\boldsymbol{\alpha}})}{\partial\tilde{\boldsymbol{\alpha}}}\right| \leq \frac{(1+\epsilon)^{d}}{(1-\epsilon\sqrt{d})^{d}(1-\epsilon)}.$$

for all $\tilde{\boldsymbol{\alpha}} \in A_0$.

Proof The condition $\left\| I - \boldsymbol{A}^{-1}\tilde{\boldsymbol{A}} \right\| \leq \epsilon$ implies $\left\| \boldsymbol{A}^{-1}\tilde{\boldsymbol{A}} \right\| \leq 1+\epsilon$, so Proposition 2.2.4 (e) implies

$$\mathbf{det}\left(\boldsymbol{A}^{-1}\tilde{\boldsymbol{A}}\right) \leq (1+\epsilon)^d.$$

Observing that $\|\mathbf{1}\| = \sqrt{d}$, and $\left\| I - (\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A})^T \right\| = \left\| I - (\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}) \right\|$, we compute

$$\left\| (\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A})^T\mathbf{1} \right\| \geq \|\mathbf{1}\| - \left\| \mathbf{1} - (\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A})^T\mathbf{1} \right\| \geq \sqrt{d} - \left\| I - (\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A})^T \right\| \|\mathbf{1}\| \geq \sqrt{d} - \epsilon\sqrt{d}.$$

So,

$$\frac{\|\mathbf{1}\|}{\left\| \left(\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}\right)^T \mathbf{1} \right\|} \leq \frac{1}{1-\epsilon}.$$

Finally, as $\langle\tilde{\boldsymbol{\alpha}}|\mathbf{1}\rangle = 1$ and $\|\tilde{\boldsymbol{\alpha}}\| \leq 1$, we have

$$\begin{aligned}
\left\langle \boldsymbol{A}^{-1}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}}|\mathbf{1} \right\rangle &= \langle\tilde{\boldsymbol{\alpha}}|\mathbf{1}\rangle + \left\langle \boldsymbol{A}^{-1}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}|\mathbf{1} \right\rangle \\
&= 1 + \left\langle (\boldsymbol{A}^{-1}\tilde{\boldsymbol{A}} - I)\tilde{\boldsymbol{\alpha}}|\mathbf{1} \right\rangle \\
&\geq 1 - \left\| \boldsymbol{A}^{-1}\tilde{\boldsymbol{A}} - I \right\| \|\tilde{\boldsymbol{\alpha}}\| \|\mathbf{1}\| \\
&\geq 1 - \epsilon\sqrt{d}.
\end{aligned}$$

Applying Lemma 5.2.9, we have

$$\left| \frac{\partial\Psi(\tilde{\boldsymbol{\alpha}})}{\partial\tilde{\boldsymbol{\alpha}}} \right| = \mathbf{det}\left(\boldsymbol{A}^{-1}\tilde{\boldsymbol{A}}\right) \frac{\|\mathbf{1}\|}{\left\langle \boldsymbol{A}^{-1}\tilde{\boldsymbol{A}}\boldsymbol{\alpha}|\mathbf{1} \right\rangle^d \left\| \left(\tilde{\boldsymbol{A}}^{-1}\boldsymbol{A}\right)^T \mathbf{1} \right\|} \leq \frac{(1+\epsilon)^d}{(1-\epsilon\sqrt{d})^d(1-\epsilon)}.$$

■

Lemma 5.2.11 (Jacobian of $\boldsymbol{\Gamma}_{\boldsymbol{u},\boldsymbol{v}}$)

$$\left| \mathbf{det}\left(\frac{\partial\Gamma_{\boldsymbol{u},\boldsymbol{v}}(\boldsymbol{x})}{\partial\boldsymbol{x}} \right) \right| = \frac{\|\boldsymbol{v}\|}{\langle\boldsymbol{x}|\boldsymbol{v}\rangle^d \|\boldsymbol{u}\|}.$$

Proof Consider dividing $\mathbb{R}^d$ into $\mathbf{Span}(\boldsymbol{u},\boldsymbol{v})$ and the space orthogonal to $\mathbf{Span}(\boldsymbol{u},\boldsymbol{v})$. In the $(d-2)$-dimensional orthogonal space, $\Gamma_{\boldsymbol{u},\boldsymbol{v}}$ acts as a multiplication by $1/\langle\boldsymbol{x}|\boldsymbol{v}\rangle$. On the other hand, the Jacobian of the restriction of $\Gamma_{\boldsymbol{u},\boldsymbol{v}}$ to $\mathbf{Span}(\boldsymbol{u},\boldsymbol{v})$ is computed by Lemma 5.2.12 to be

$$\frac{\|\boldsymbol{v}\|}{\langle\boldsymbol{x}|\boldsymbol{v}\rangle^2 \|\boldsymbol{u}\|}.$$

So,

$$\left| \mathbf{det}\left(\frac{\partial\Gamma_{\boldsymbol{u},\boldsymbol{v}}(\boldsymbol{x})}{\partial\boldsymbol{x}} \right) \right| = \left(\frac{1}{\langle\boldsymbol{x}|\boldsymbol{v}\rangle} \right)^{d-2} \frac{\|\boldsymbol{v}\|}{\langle\boldsymbol{x}|\boldsymbol{v}\rangle^2 \|\boldsymbol{u}\|} = \frac{\|\boldsymbol{v}\|}{\langle\boldsymbol{x}|\boldsymbol{v}\rangle^d \|\boldsymbol{u}\|}.$$

■

Lemma 5.2.12 (Jacobian of $\Gamma_{u,v}$ in 2D) *Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be vectors in $\mathbb{R}^2$ and let $\Gamma_{\boldsymbol{u},\boldsymbol{v}}(\boldsymbol{x})$ be the map from $\{\boldsymbol{x} : \langle \boldsymbol{x}|\boldsymbol{u}\rangle = 1\}$ to $\{\boldsymbol{x} : \langle \boldsymbol{x}|\boldsymbol{v}\rangle = 1\}$ by*

$$\Gamma_{\boldsymbol{u},\boldsymbol{v}}(\boldsymbol{x}) = \frac{\boldsymbol{x}}{\langle \boldsymbol{x}|\boldsymbol{v}\rangle}.$$

Then,

$$\left|\det\left(\frac{\partial \Gamma_{\boldsymbol{u},\boldsymbol{v}}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right)\right| = \frac{\|\boldsymbol{v}\|}{\langle \boldsymbol{x}|\boldsymbol{v}\rangle^2 \|\boldsymbol{u}\|}.$$

Proof Let $\boldsymbol{R} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, the 90^o rotation counter-clockwise. Let

$$\boldsymbol{u}^\perp = \boldsymbol{R}\boldsymbol{u}/\|\boldsymbol{u}\| \qquad \text{and} \qquad \boldsymbol{v}^\perp = \boldsymbol{R}\boldsymbol{v}/\|\boldsymbol{v}\|.$$

Express the $\boldsymbol{x}$ such that $\langle \boldsymbol{x}|\boldsymbol{u}\rangle = 1$, as $\boldsymbol{x} = \boldsymbol{u}/\|\boldsymbol{u}\|^2 + x\boldsymbol{u}^\perp$. Similarly, parameterize the line $\{\boldsymbol{x} : \langle \boldsymbol{x}|\boldsymbol{v}\rangle = 1\}$ by $\boldsymbol{v}/\|\boldsymbol{v}\|^2 + y\boldsymbol{v}^\perp$. Then, we have

$$\Gamma_{\boldsymbol{u},\boldsymbol{v}}\left(u/\|\boldsymbol{u}\|^2 + x\boldsymbol{u}^\perp\right) = \boldsymbol{v}/\|\boldsymbol{v}\|^2 + y\boldsymbol{v}^\perp,$$

where

$$y = \frac{\left\langle u/\|\boldsymbol{u}\|^2 + x\boldsymbol{u}^\perp | \boldsymbol{v}^\perp \right\rangle}{\left\langle u/\|\boldsymbol{u}\|^2 + x\boldsymbol{u}^\perp | \boldsymbol{v} \right\rangle} = \frac{\left\langle u/\|\boldsymbol{u}\|^2 + x\boldsymbol{u}^\perp | \boldsymbol{v}^\perp \right\rangle}{\langle \boldsymbol{x}|\boldsymbol{v}\rangle}.$$

So,

$$\begin{aligned}
\left|\det\left(\frac{\partial \Gamma_{\boldsymbol{u},\boldsymbol{v}}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right)\right| &= \left|\det\left(\frac{\partial y}{\partial x}\right)\right| \\
&= \left|\frac{\left\langle \boldsymbol{u}^\perp \middle| \boldsymbol{v}^\perp \right\rangle \left\langle \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|^2} + x\boldsymbol{u}^\perp \middle| \boldsymbol{v} \right\rangle - \left\langle \boldsymbol{u}^\perp \middle| \boldsymbol{v} \right\rangle \left\langle \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|^2} + x\boldsymbol{u}^\perp \middle| \boldsymbol{v}^\perp \right\rangle}{\langle \boldsymbol{x}|\boldsymbol{v}\rangle^2}\right| \\
&= \left|\frac{\left\langle \boldsymbol{u}^\perp \middle| \boldsymbol{v}^\perp \right\rangle \left\langle \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|^2} \middle| \boldsymbol{v} \right\rangle - \left\langle \boldsymbol{u}^\perp \middle| \boldsymbol{v} \right\rangle \left\langle \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|^2} \middle| \boldsymbol{v}^\perp \right\rangle}{\langle \boldsymbol{x}|\boldsymbol{v}\rangle^2}\right| \\
&= \left|\frac{\|\boldsymbol{v}\| \left(\left\langle \boldsymbol{u}^\perp \middle| \boldsymbol{v}^\perp \right\rangle \left\langle \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|} \middle| \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} \right\rangle - \left\langle \boldsymbol{u}^\perp \middle| \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} \right\rangle \left\langle \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|} \middle| \boldsymbol{v}^\perp \right\rangle \right)}{\|\boldsymbol{u}\| \langle \boldsymbol{x}|\boldsymbol{v}\rangle^2}\right| \\
&= \left|\frac{\|\boldsymbol{v}\| \left(\left\langle \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|} \middle| \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} \right\rangle^2 + \left\langle \boldsymbol{u}^\perp \middle| \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} \right\rangle^2 \right)}{\|\boldsymbol{u}\| \langle \boldsymbol{x}|\boldsymbol{v}\rangle^2}\right|, \text{ as } \boldsymbol{R} \text{ is orthogonal and } \boldsymbol{R}^2 = -1, \\
&= \frac{\|\boldsymbol{v}\|}{\|\boldsymbol{u}\| \langle \boldsymbol{x}|\boldsymbol{v}\rangle^2}, \text{ as } \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|}, \boldsymbol{u}^\perp \text{ is a basis.}
\end{aligned}$$

■

5.3 Bounding the shadow of LP^+

The main obstacle to proving a bound on the size of the shadow of LP^+ is that the vectors $\boldsymbol{a}_i^+/y_i^+$ are not Gaussian random vectors. To resolve this problem, we will show that, in almost every sufficiently small region, we can construct a family of Gaussian random vectors with distributions similar to the vectors $\boldsymbol{a}_i^+/y_i^+$. We will then bound the expected size of the shadow of the vectors $\boldsymbol{a}_i^+/y_i^+$ by a small multiple of the expected size of the shadow of these Gaussian vectors. These regions are defined by splitting the original perturbation into two, and letting the first perturbation define the region.

As in the analysis of LP', a secondary obstacle is the correlation of κ and M with $\boldsymbol{A}$ and $\boldsymbol{y}$. We again overcome this obstacle by considering the sum of the expected sizes of the shadows when κ and M are fixed to each of their likely values, and use the notation

$$\mathcal{T}_{\boldsymbol{z}}^+(\boldsymbol{A},\boldsymbol{y},\kappa,M) \stackrel{\text{def}}{=} \begin{cases} \left|\mathbf{Shadow}_{(0,z),z^+}\left(\boldsymbol{a}_1^+/y_1^+,\ldots,\boldsymbol{a}_n^+/y_n^+\right)\right|, & \text{if } \sqrt{d}M/4\kappa \geq 1 \\ 0 & \text{otherwise,} \end{cases}$$

where

$$\begin{aligned} \boldsymbol{a}_i^+ &= \left((y_i' - y_i)/2, \boldsymbol{a}_i\right), \\ y_i^+ &= (y_i' + y_i)/2, \text{ and} \\ y_i' &= \begin{cases} M & \text{if } i \in I \\ \sqrt{d}M^2/4\kappa & \text{otherwise.} \end{cases} \end{aligned}$$

By Lemma 3.3.5 and Proposition 3.3.2, we then have

$$\mathcal{S}_{\boldsymbol{z}}^+(\boldsymbol{A},\boldsymbol{y},\mathcal{I}) = \mathcal{T}_{\boldsymbol{z}}^+\left(\boldsymbol{A},\boldsymbol{y},2^{\lfloor \lg \mathbf{S_{min}}(\boldsymbol{A}_{\mathcal{I}(A)})\rfloor}, 2^{\lceil \lg(\max_i \|(y_i,\boldsymbol{a}_i)\|)\rceil+2}\right).$$

Lemma 5.3.1 (LP+) *Let $d \geq 3$ and $n \geq d+1$. Let $\bar{\boldsymbol{A}} = [\bar{\boldsymbol{a}}_1,\ldots,\bar{\boldsymbol{a}}_n] \in \mathbb{R}^{n\times d}$, $\bar{\boldsymbol{y}} \in \mathbb{R}^n$ and $\boldsymbol{z} \in \mathbb{R}^d$, satisfy $\max_i \|(\bar{y}_i, \bar{\boldsymbol{a}}_i)\| \in (1/2, 1]$. For any $\sigma > 0$, let $\boldsymbol{A}$ be a Gaussian random matrix centered at $\bar{\boldsymbol{A}}$ of standard deviation σ, and let $\boldsymbol{y}$ by a Gaussian random vector centered at $\bar{\boldsymbol{y}}$ of standard deviation σ. Let $\mathcal{I}$ be a set of $3nd\ln n$ randomly chosen d-subsets of $[n]$. Then,*

$$\mathop{\boldsymbol{E}}_{\boldsymbol{A},\boldsymbol{y},\mathcal{I}}\left[\mathcal{S}^+(\boldsymbol{A},\boldsymbol{y},\mathcal{I})\right] \leq 49\lg(nd/\min(\sigma,1))\mathcal{D}\left(d,n,\frac{\min(1,\sigma^5)}{2^{23}(d+1)^{11/2}n^{14}(\ln n)^{5/2}}\right) + n.$$

where $\mathcal{D}(d,n,\sigma)$ is as given in Theorem 4.0.1.

Proof For ρ_0 and ρ_1 defined below, we let $\boldsymbol{G}$ and $\tilde{\boldsymbol{G}}$ be Gaussian random matrices centered at the origin of standard deviations ρ_0 and ρ_1, respectively. We then let $\tilde{\boldsymbol{A}} = \bar{\boldsymbol{A}} + \boldsymbol{G}$ and $\boldsymbol{A} = \tilde{\boldsymbol{A}} + \tilde{\boldsymbol{G}}$. We similarly let $\boldsymbol{h}$ and $\tilde{\boldsymbol{h}}$ be Gaussian random vectors centered at the origin of standard deviations ρ_0 and ρ_1, respectively, and let $\tilde{\boldsymbol{y}} = \boldsymbol{y}' + \boldsymbol{h}$ and $\boldsymbol{y} = \tilde{\boldsymbol{y}} + \tilde{\boldsymbol{h}}$. If

$$\sigma \leq \frac{3\sqrt{1/4}}{\sqrt{2e}n(60n(d+1)^{3/2}(\ln n)^{3/2})},$$

we set $\rho_1 = \sigma$. Otherwise, we set ρ_1 so that

$$\rho_1 = \frac{3\sqrt{1/4 + d(\sigma^2 - \rho_1^2)}}{\sqrt{2e}n(60n(d+1)^{3/2}(\ln n)^{3/2})},$$

and set $\rho_0^2 = \sigma^2 - \rho_1^2$. We note that

$$\rho_1 = \min\left(\sigma, \frac{3\sqrt{1/4 + d\rho_0^2}}{\sqrt{2e}n(60n(d+1)^{3/2}(\ln n)^{3/2})}\right).$$

As in the proof of Lemma 5.2.1, we define the set of likely values for M:

$$\mathcal{M} = \left\{ 2^{\lceil \lg x \rceil + 2} : \left(\max_i \|(\tilde{y}_i, \tilde{\boldsymbol{a}}_i)\|\right)\left(1 - \frac{9\sqrt{(d+1)\ln n}}{(60n(d+1)^{3/2}(\ln n)^{3/2})}\right) \leq x \right.$$
$$\left. \leq \left(\max_i \|(\tilde{y}_i, \tilde{\boldsymbol{a}}_i)\|\right)\left(1 + \frac{9\sqrt{(d+1)\ln n}}{(60n(d+1)^{3/2}(\ln n)^{3/2})}\right) \right\}.$$

Observed that $|\mathcal{M}| \leq 2$.

As in the proof of Lemma 5.2.1, we define random variables:

$$W = \left[\max_i \|(\tilde{y}_i, \tilde{\boldsymbol{a}}_i)\| \leq 1 + 3\sqrt{(d+1)\ln n}\rho_0\right],$$
$$X = \left[\max_i \|(\tilde{y}_i, \tilde{\boldsymbol{a}}_i)\| \geq \frac{\sqrt{1/4 + d\rho_0^2}}{\sqrt{2e}n}\right],$$
$$Y = \left[2^{\lfloor \lg \mathbf{S_{min}}(\boldsymbol{A}_{\mathcal{I}(A)}) \rfloor} \in \mathcal{K}\right], \text{ and}$$
$$Z = \left[2^{\lceil \lg \max_i \|(y_i, \boldsymbol{a}_i)\| \rceil + 2} \in \mathcal{M}\right].$$

In order to apply the shadow bound proved below in Lemma 5.3.2, we need

$$M \geq 3 \max_i \|(\tilde{y}_i, \tilde{\boldsymbol{a}}_i)\|,$$

and

$$M \geq (60n(d+1)^{3/2}(\ln n)^{3/2})\rho_1.$$

From the definition of $\mathcal{M}$ and the inequality $1 - 9\sqrt{(d+1)\ln n}/(60n(d+1)^{3/2}(\ln n)^{3/2}) \geq 3/4$, the first of these inequalities holds if Z is true. Given that Z is true, the second inequality holds if X is also true.

From Corollary 5.1.3, we know

$$\mathop{\mathbf{Pr}}_{\boldsymbol{A},\mathcal{I}}[\text{not}(Y)] \leq \mathop{\mathbf{Pr}}_{\boldsymbol{A},\mathcal{I}}\left[2^{\lfloor \lg \mathbf{S_{min}}(\boldsymbol{A}_{\mathcal{I}(A)}) \rfloor} \notin \mathcal{K}\right] \leq 0.42\binom{n}{d}^{-1} \leq 0.42n\binom{n}{d+1}^{-1}. \tag{50}$$

From Corollary 2.4.6 we have

$$\mathop{\mathbf{Pr}}_{\tilde{\boldsymbol{A}},\tilde{\boldsymbol{y}}}[\text{not}(W)] \leq n^{-2.9(d+1)+1} \leq 0.0015\binom{n}{d+1}^{-1}. \tag{51}$$

From Proposition 2.4.9, we know

$$\mathbf{Pr}\left[\text{not}(X)\right] = \mathbf{Pr}\left[\max_i \|(\tilde{y}_i, \tilde{\boldsymbol{a}}_i)\| < \frac{\sqrt{1/4 + d\rho_0^2}}{\sqrt{2e}n}\right] < n^{-(d+1)} \leq \frac{1}{24}\binom{n}{d+1}^{-1}. \tag{52}$$

To bound the probability that Z fails, we note that

$$\max_i \|(\tilde{y}_i, \tilde{\boldsymbol{a}}_i)\| \geq \frac{\sqrt{1/4 + d\rho_0^2}}{\sqrt{2e}n}$$

and

$$\max_i \|(y_i - \tilde{y}_i, \boldsymbol{a}_i - \tilde{\boldsymbol{a}}_i)\| \leq \rho_1 3\sqrt{(d+1)\ln n},$$

imply Z is true. Hence, by Corollary 2.4.6 and (52),

$$\mathbf{Pr}\left[\text{not(Z)}\right] \leq n^{-2.9(d+1)+1} + n^{-(d+1)} \leq .044\binom{n}{d+1}^{-1}. \tag{53}$$

As in the proof of Lemma 5.2.1, we now expand

$$\mathop{\mathbf{E}}_{\mathcal{I},A,\boldsymbol{y}}\left[\mathcal{S}^+(\boldsymbol{A},\boldsymbol{y},\mathcal{I})\right] = \mathop{\mathbf{E}}_{\mathcal{I},A,\boldsymbol{y}}\left[\mathcal{S}^+(\boldsymbol{A},\boldsymbol{y},\mathcal{I})WXYZ\right] + \mathop{\mathbf{E}}_{\mathcal{I},A,\boldsymbol{y}}\left[\mathcal{S}^+(\boldsymbol{A},\boldsymbol{y},\mathcal{I})(1-WXYZ)\right]. \tag{54}$$

To bound the second term by n, we apply (51) , (52) , (50) and (53) to show

$$\mathop{\mathbf{Pr}}_{\boldsymbol{A},\mathcal{I}}\left[\text{not}(W) \text{ or not}(X) \text{ or not}(Y) \text{ or not}(Z)\right] \leq n\binom{n}{d+1}^{-1},$$

and then combine this inequality with Proposition 5.0.2.

To bound the first term of (54), we note

$$\begin{aligned}
&\mathop{\mathbf{E}}_{\mathcal{I},A,\boldsymbol{y}}\left[\mathcal{S}^+(\boldsymbol{A},\boldsymbol{y},\mathcal{I})WXYZ\right] \\
&\leq \mathop{\mathbf{E}}_{\mathcal{I},\tilde{\boldsymbol{A}},\tilde{\boldsymbol{y}}}\left[WX\sum_{\kappa\in\mathcal{K},M\in\mathcal{M}}\mathop{\mathbf{E}}_{\tilde{\boldsymbol{G}},\tilde{\boldsymbol{h}}}\left[\mathcal{T}^+\left(\boldsymbol{A},\boldsymbol{y},\kappa,M\right)XZ\right]\right] \\
&\leq \mathop{\mathbf{E}}_{\mathcal{I},\tilde{\boldsymbol{A}},\tilde{\boldsymbol{y}}}\left[WX\sum_{\kappa\in\mathcal{K},M\in\mathcal{M}}\mathop{\mathbf{E}}_{\tilde{\boldsymbol{G}},\tilde{\boldsymbol{h}}}\left[\mathcal{T}^+\left(\boldsymbol{A},\boldsymbol{y},\kappa,M\right)\Big|XZ\right]\right] \\
&\leq \mathop{\mathbf{E}}_{\mathcal{I},\tilde{\boldsymbol{A}},\tilde{\boldsymbol{y}}}\left[WX\sum_{\kappa\in\mathcal{K},M\in\mathcal{M}}e\mathcal{D}\left(d,n,\frac{\rho_1\min_i y_i'}{3(\max_i y_i')^2}\right)+1,\right] \text{ by Lemma 5.3.2} \qquad (55) \\
&\leq \mathop{\mathbf{E}}_{\mathcal{I},\tilde{\boldsymbol{A}},\tilde{\boldsymbol{y}}}\left[WX\sum_{\kappa\in\mathcal{K},M\in\mathcal{M}}e\mathcal{D}\left(d,n,\frac{\sigma M}{3(M^2/4\kappa)^2}\right)+1\right] \\
&\leq \mathop{\mathbf{E}}_{\mathcal{I},\tilde{\boldsymbol{A}},\tilde{\boldsymbol{y}}}\left[WX\left|\mathcal{K}\right|\left|M\right|e\mathcal{D}\left(d,n,\frac{16\sigma\min(\mathcal{K})^2}{3\max(\mathcal{M})^3}\right)+1\right]. \qquad (56)
\end{aligned}$$

As $\min(\mathcal{K}) \geq \kappa_0/2$ and W implies $\max(\mathcal{M}) \leq 9\left(1+3\sqrt{(d+1)\ln n}\sigma\right)$,

$$\begin{aligned}\frac{16\sigma\min(\mathcal{K})^2}{3\max(\mathcal{M})^3} &\geq \frac{16\sigma^3\min(1,\sigma)^2}{3\cdot 4\left(9(1+3\sqrt{(d+1)\ln n}\sigma\right)^3\left(12d^2n^7\sqrt{\ln n}\right)^2}\\ &\geq \frac{16\min(1,\sigma^5)}{3\cdot 4\left(9(1+3\sqrt{(d+1)\ln n}\right)^3\left(12d^2n^7\sqrt{\ln n}\right)^2}\\ &\geq \frac{\min(1,\sigma^5)}{2^{23}(d+1)^{11/2}n^{14}(\ln n)^{5/2}}.\end{aligned}$$

Applying this inequality, Proposition 5.1.4, and the fact that X implies $|\mathcal{M}| \leq 2$, we obtain

$$(56) \leq 49\lg(nd/\min(\sigma,1))\mathcal{D}\left(d,n,\frac{\min(1,\sigma^5)}{2^{23}(d+1)^{11/2}n^{14}(\ln n)^{5/2}}\right).$$

■

Lemma 5.3.2 (LP^+ Shadow, part 2) *Let $d \geq 3$ and $n \geq d+1$. Let $\boldsymbol{y}$ be a Gaussian random vector of standard deviation ρ_1 centered at a point $\tilde{\boldsymbol{y}}$, and let $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n$ be Gaussian random vectors in $\mathbb{R}^d$ of standard deviation ρ_1 centered at $\tilde{\boldsymbol{a}}_1,\ldots,\tilde{\boldsymbol{a}}_n$ respectively. Under the conditions*

$$y_i' > 3(\|\tilde{y}_i,\tilde{\boldsymbol{a}}_i\|),\forall i, \text{ and} \tag{57}$$

$$y_i' > 60n(d+1)^{3/2}(\ln n)^{3/2}\rho_1,\forall i. \tag{58}$$

Let

$$\begin{aligned}\boldsymbol{a}_i^+ &= \left((y_i'-y_i)/2,\boldsymbol{a}_i\right), \text{and}\\ y_i^+ &= (y_i'+y_i)/2.\end{aligned}$$

Then,

$$\underset{(y_1,\boldsymbol{a}_1),\ldots,(y_n,\boldsymbol{a}_n)}{\boldsymbol{E}}\left[\left|\mathbf{Shadow}_{(0,z),z^+}\left(\boldsymbol{a}_1^+/y_1^+,\ldots,\boldsymbol{a}_n^+/y_n^+\right)\right|\right] \leq e\mathcal{D}\left(d,n,\frac{\rho_1\min_i y_i'}{3(\max_i y_i')^2}\right)+1.$$

Proof We use the notation

$$(p_{i,0}(\tilde{h}_i),\boldsymbol{p}_i(\tilde{h}_i,\tilde{\boldsymbol{g}}_i)) = \boldsymbol{a}_i^+/y_i^+ = \left(\frac{y_i'-\tilde{y}_i-\tilde{h}_i}{y_i'+\tilde{y}_i+\tilde{h}_i},\frac{2(\tilde{\boldsymbol{a}}_i+\tilde{\boldsymbol{g}}_i)}{y_i'+\tilde{y}_i+\tilde{h}_i}\right),$$

where $\tilde{\boldsymbol{g}}_1,\ldots,\tilde{\boldsymbol{g}}_n$ are the columns of $\tilde{\boldsymbol{G}}$ and $(\tilde{h}_1,\ldots,\tilde{h}_n) = \tilde{\boldsymbol{h}}$ as defined in the proof of Lemma 5.3.1.

The Gaussian random vectors that we will use to approximate these will come from their first-order approximations:

$$(\hat{p}_{i,0}(\tilde{h}_i),\hat{\boldsymbol{p}}(\tilde{h}_i,\tilde{\boldsymbol{g}}_i)) = \left(\frac{y_i'-\tilde{y}_i-\tilde{h}_i(2y_i'/(y_i'+\tilde{y}_i))}{y_i'+\tilde{y}_i},\frac{2\tilde{\boldsymbol{a}}_i+2\tilde{\boldsymbol{g}}_i-\tilde{h}_i(2\tilde{\boldsymbol{a}}_i/(y_i'+\tilde{y}_i))}{y_i'+\tilde{y}_i}\right)$$

Let $\hat{\nu}_i(\hat{p}_{i,0}, \hat{\boldsymbol{p}}_i)$ be the induced density on $(\hat{p}_{i,0}, \hat{\boldsymbol{p}}_i)$. In Lemma 5.3.4, we prove that there exists a set B of $((p_{1,0}, \boldsymbol{p}_1), \ldots, (p_{n,0}, \boldsymbol{p}_n))$ such that

$$\mathop{\mathbf{Pr}}_{\prod_{i=1}^n \nu_i(p_{i,0}, \boldsymbol{p}_i)}[((p_{1,0}, \boldsymbol{p}_1), \ldots, (p_{n,0}, \boldsymbol{p}_n)) \in B] \geq 1 - 0.0015\binom{n}{d+1}^{-1},$$

and for $((p_{1,0}, \boldsymbol{p}_1), \ldots, (p_{n,0}, \boldsymbol{p}_n)) \in B$,

$$\prod_{i=1}^n \nu_i(p_{i,0}, \boldsymbol{p}_i) \leq e \prod_{i=1}^n \mu_i'(p_{i,0}, \boldsymbol{p}_i).$$

Consequently, Lemma 2.3.4 allows us to prove

$$\begin{aligned}\mathop{\mathbf{E}}_{\prod_{i=1}^n \nu_i(p_{i,0}, \boldsymbol{p}_i)} &\left[\left|\mathbf{Shadow}_{(0,\boldsymbol{z}),\boldsymbol{z}^+}((p_{1,0}, \boldsymbol{p}_1), \ldots, (p_{n,0}, \boldsymbol{p}_n))\right|\right] \\ &\leq e \mathop{\mathbf{E}}_{\prod_{i=1}^n \hat{\nu}_i(p_{i,0}, \boldsymbol{p}_i)} \left[\left|\mathbf{Shadow}_{(0,\boldsymbol{z}),\boldsymbol{z}^+}((p_{1,0}, \boldsymbol{p}_1), \ldots, (p_{n,0}, \boldsymbol{p}_n))\right|\right] + 1.\end{aligned}$$

By Lemma 5.3.3, the densities $\hat{\nu}_i$ represent Gaussian distributions centered at points of norm at most

$$\left\| \left(\frac{y_i' - \tilde{y}_i}{y_i' + \tilde{y}_i}, \frac{2\tilde{\boldsymbol{a}}_i}{y_i' + \tilde{y}_i} \right) \right\| \leq \sqrt{5}, \quad \text{(by condition (57))}$$

whose covariance matrices have eigenvalues at most

$$\left(9\rho_1/2y_i'\right)^2 \leq \left(9/2(60n(d+1)^{3/2}(\ln n)^{3/2})\right)^2 \leq 1/9d\ln n, \quad \text{(by condition (58))}$$

and at least

$$\left(9\rho_1/8y_i'\right)^2.$$

Thus, we can apply Corollary 4.3.3 to bound

$$\begin{aligned}&\mathop{\mathbf{E}}_{\prod_{i=1}^n \hat{\nu}_i(p_{i,0}, \boldsymbol{p}_i)} \left[\left|\mathbf{Shadow}_{(0,\boldsymbol{z}),\boldsymbol{z}^+}((p_{1,0}, \boldsymbol{p}_1), \ldots, (p_{n,0}, \boldsymbol{p}_n))\right|\right] \\ &\leq e\mathcal{D}\left(d, n, \frac{9\rho_1/8\max_i y_i'}{(1+\sqrt{5})(\max_i y_i'/\min_i y_i')}\right) + 1, \\ &\leq e\mathcal{D}\left(d, n, \frac{\rho_1 \min_i y_i'}{3(\max_i y_i')^2}\right) + 1,\end{aligned}$$

thereby proving the Lemma. ■

Lemma 5.3.3 ($\hat{\boldsymbol{\nu}}$) *Under the conditions of Lemma 5.3.2, the vector $(\hat{p}_{i,0}(\tilde{h}_i), \hat{\boldsymbol{p}}(\tilde{h}_i, \tilde{\boldsymbol{g}}_i))$ is a Gaussian random vector centered at*

$$\left(\frac{y_i' - \tilde{y}_i}{y_i' + \tilde{y}_i}, \frac{2\tilde{\boldsymbol{a}}_i}{y_i' + \tilde{y}_i} \right),$$

and has a covariance matrix with eigenvalues between $\left(9\rho_1/8y_i'\right)^2$ and $\left(9\rho_1/2y_i'\right)^2$.

Proof Because $(\hat{p}_{i,0}(\tilde{h}_i), \hat{\boldsymbol{p}}(\tilde{h}_i, \tilde{\boldsymbol{g}}_i))$ is linear in $(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)$ and $(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)$ is a Gaussian random vector, $(\hat{p}_{i,0}(\tilde{h}_i), \hat{\boldsymbol{p}}(\tilde{h}_i, \tilde{\boldsymbol{g}}_i))$ is a Gaussian vector. The statement about the center of the distributions follows immediately from the fact that $(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)$ is centered at the origin. To construct the covariance matrix, we note that the matrix corresponding to the transformation from $(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)$ to $(\hat{p}_{i,0}(\tilde{h}_i), \hat{\boldsymbol{p}}(\tilde{h}_i, \tilde{\boldsymbol{g}}_i))$ is

$$C_i \stackrel{\text{def}}{=} \left(\begin{array}{c|c} \frac{-2y_i'}{(y_i'+\tilde{y}_i)^2}, & 0, \ldots, 0 \\ \hline \frac{-2\tilde{a}_{i,1}}{(y_i'+\tilde{y}_i)^2} & \\ \frac{-2\tilde{a}_{i,2}}{(y_i'+\tilde{y}_i)^2} & \frac{2}{y_i'+\tilde{y}_i} I \\ \vdots & \\ \frac{-2\tilde{a}_{i,d}}{(y_i'+\tilde{y}_i)^2} & \end{array}\right)$$

Thus, the covariance matrix of $(\hat{p}_{i,0}(\tilde{h}_i), \hat{\boldsymbol{p}}(\tilde{h}_i, \tilde{\boldsymbol{g}}_i))$ is given by $\rho_1^2 C_i^T C_i$.

We now note that

$$\frac{y_i'+\tilde{y}_i}{2} C_i - \left(\begin{array}{c|c} -1 & 0, \ldots, 0 \\ \hline 0 & \\ 0 & I \\ \vdots & \\ 0 & \end{array}\right) = \left(\begin{array}{c|c} \frac{\tilde{y}_i}{y_i'+\tilde{y}_i}, & 0, \ldots, 0 \\ \hline -\frac{\tilde{a}_{i,1}}{y_i'+\tilde{y}_i} & \\ -\frac{\tilde{a}_{i,2}}{y_i'+\tilde{y}_i} & 0 \\ \vdots & \\ -\frac{\tilde{a}_{i,d}}{y_i'+\tilde{y}_i} & \end{array}\right)$$

As all the singular values of the middle matrix are 1, and the norm of the right-hand matrix is $\|(\tilde{y}_i, \tilde{\boldsymbol{a}}_i)\| / (y_i' + \tilde{y}_i)$, all the singular values of C_i lie between

$$\frac{2}{y_i'+\tilde{y}_i}\left(1 - \frac{\|(\tilde{y}_i, \tilde{\boldsymbol{a}}_i)\|}{y_i'+\tilde{y}_i}\right) \text{ and } \frac{2}{y_i'+\tilde{y}_i}\left(1 + \frac{\|(\tilde{y}_i, \tilde{\boldsymbol{a}}_i)\|}{y_i'+\tilde{y}_i}\right)$$

The stated bounds now follow from inequality (57). ■

Lemma 5.3.4 (Almost Gaussian) *Under the conditions of Lemma 5.3.2, let $\nu_i(p_{i,0}, \boldsymbol{p}_i)$ be the induced density on $(p_{i,0}, \boldsymbol{p}_i)$, and let $\hat{\nu}_i(\hat{p}_{i,0}, \hat{\boldsymbol{p}}_i)$ be the induced density on $(\hat{p}_{i,0}, \hat{\boldsymbol{p}}_i)$. Then, there exists a set B of $((p_{1,0}, \boldsymbol{p}_1), \ldots, (p_{n,0}, \boldsymbol{p}_n))$ such that*

(a) $\boldsymbol{Pr}[((p_{1,0}, \boldsymbol{p}_1), \ldots, (p_{n,0}, \boldsymbol{p}_n)) \in B] \geq 1 - 0.0015\binom{n}{d+1}^{-1}$; *and*

(b) *for all $((p_{1,0}, \boldsymbol{p}_1), \ldots, (p_{n,0}, \boldsymbol{p}_n)) \in B$,*

$$\prod_{i=1}^{n} \nu_i(p_{i,0}, \boldsymbol{p}_i) \leq e \prod_{i=1}^{n} \hat{\nu}_i(p_{i,0}, \boldsymbol{p}_i).$$

Proof Let

$$B = \left\{ \begin{array}{c} ((p_{1,0}(\tilde{h}_1), \boldsymbol{p}_1(\tilde{h}_1, \tilde{\boldsymbol{g}}_1)), \ldots, (p_{n,0}(\tilde{h}_n), \boldsymbol{p}_n(\tilde{h}_n, \tilde{\boldsymbol{g}}_n))) \\ \text{such that } \left\| (\tilde{h}_i, \tilde{\boldsymbol{g}}_i) \right\| \leq 3\sqrt{(d+1)\ln n}\rho_1, \text{ for } 1 \leq i \leq n \end{array} \right\}.$$

From inequalities (57) and (58), and the assumption $\left|\tilde{h}_i\right| \leq 3\sqrt{(d+1)\ln n}\rho_1$, we can show $y_i' + \tilde{y}_i + \tilde{h}_i > 0$, and so the map from $(\tilde{h}_1, \tilde{\boldsymbol{g}}_1), \ldots, (\tilde{h}_n, \tilde{\boldsymbol{g}}_n)$ to $(p_{1,0}, \boldsymbol{p}_1), \ldots, (p_{n,0}, \boldsymbol{p}_n)$ is invertible for $(p_{1,0}, \boldsymbol{p}_1), \ldots, (p_{n,0}, \boldsymbol{p}_n) \in B$. Thus, we may apply Corollary 2.4.6 to establish part (a).

Part (b) of follows directly Lemma 5.3.5. ■

Lemma 5.3.5 (Almost Gaussian, single variable) *Under the conditions of Lemma 5.3.2, for all $\tilde{h}_i$ and $\tilde{\boldsymbol{g}}_i$ such that $\left\| (\tilde{h}_i, \tilde{\boldsymbol{g}}_i) \right\| \leq 3\sqrt{(d+1)\ln n}\rho_1$,*

$$\nu_i(p_{i,0}(\tilde{h}_i), \boldsymbol{p}_i(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)) \leq e^{1/n}\hat{\nu}_i(p_{i,0}(\tilde{h}_i), \boldsymbol{p}_i(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)).$$

Proof Let $\mu(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)$ be the density on $(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)$. As observed in the proof of Lemma 5.3.4, the map from $(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)$ to $(p_{i,0}(\tilde{h}_i), \boldsymbol{p}_i(\tilde{h}_i, \tilde{\boldsymbol{g}}_i))$ is injective for $\left\| (\tilde{h}_i, \tilde{\boldsymbol{g}}_i) \right\| \leq 3\sqrt{(d+1)\ln n}\rho_1$; so, by Proposition 2.5.1, the induced density on ν_i is

$$\nu_i(p_{i,0}, \boldsymbol{p}_i) = \frac{1}{\left|\mathbf{det}\left(\frac{\partial(p_{i,0}, \boldsymbol{p}_i)}{\partial(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)}\right)\right|}\mu(\tilde{h}_i, \tilde{\boldsymbol{g}}_i), \text{where } (p_{i,0}, \boldsymbol{p}_i) = (p_{i,0}(\tilde{h}_i), \boldsymbol{p}_i(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)).$$

Similarly,

$$\hat{\nu}_i(\hat{p}_{i,0}, \hat{\boldsymbol{p}}_i) = \frac{1}{\left|\mathbf{det}\left(\frac{\partial(\hat{p}_{i,0}, \hat{\boldsymbol{p}}_i)}{\partial(\hat{h}_i, \hat{\boldsymbol{g}}_i)}\right)\right|}\mu(\hat{h}_i, \hat{\boldsymbol{g}}_i), \text{where } (\hat{p}_{i,0}, \hat{\boldsymbol{p}}_i) = (\hat{p}_{i,0}(\hat{h}_i), \hat{\boldsymbol{p}}_i(\hat{h}_i, \hat{\boldsymbol{g}}_i)).$$

The proof now follows from Lemma 5.3.6, which tells us that

$$\frac{\mu(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)}{\mu(\hat{h}_i, \hat{\boldsymbol{g}}_i)} \leq e^{0.81/n},$$

and Lemma 5.3.7, which tells us that

$$\frac{\left|\mathbf{det}\left(\frac{\partial(\hat{p}_{i,0}, \hat{\boldsymbol{p}}_i)}{\partial(\hat{h}_i, \hat{\boldsymbol{g}}_i)}\right)\right|}{\left|\mathbf{det}\left(\frac{\partial(p_{i,0}, \boldsymbol{p}_i)}{\partial(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)}\right)\right|} \leq e^{1/10n}.$$

■

Lemma 5.3.6 (Almost Gaussian, pointwise) *Under the conditions of Lemma 5.3.5, If $p_{i,0}(\tilde{h}_i) = \hat{p}_0(\hat{h}_i)$, $\boldsymbol{p}_i(\tilde{h}_i, \tilde{\boldsymbol{g}}_i) = \hat{\boldsymbol{p}}_i(\hat{h}_i, \hat{\boldsymbol{g}}_i)$, and $\left\| \tilde{h}_i, \tilde{\boldsymbol{g}}_i \right\| \leq 3\sqrt{(d+1)\ln n}\rho_1$, then*

$$\frac{\mu(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)}{\mu(\hat{h}_i, \hat{\boldsymbol{g}}_i)} \leq e^{0.81/n}.$$

Proof We first observe that the conditions of the lemma imply

$$\hat{h}_i = \frac{\tilde{h}_i(y'_i + \tilde{y}_i)}{y'_i + \tilde{y}_i + \tilde{h}_i}, \qquad \text{and} \qquad \hat{\boldsymbol{g}}_i = \frac{\tilde{\boldsymbol{g}}_i(y'_i + \tilde{y}_i)}{y'_i + \tilde{y}_i + \tilde{h}_i}.$$

We then compute

$$\frac{\mu(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)}{\mu(\hat{h}_i, \hat{\boldsymbol{g}}_i)} = \exp\left(\frac{-1}{2\rho_1^2}\left\|(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)\right\|^2 \left(\frac{2\tilde{h}_i(y'_i + \tilde{y}_i) + \tilde{h}_i^2}{(y'_i + \tilde{y}_i + \tilde{h}_i)^2}\right)\right). \tag{59}$$

Assuming $\left\|(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)\right\| \leq 3\sqrt{(d+1)\ln n}\rho_1$, the absolute value of the exponent in (59) is at most

$$\frac{9(d+1)\ln n}{2}\left(\frac{2\tilde{h}_i(y'_i + \tilde{y}_i) + \tilde{h}_i^2}{(y'_i + \tilde{y}_i + \tilde{h}_i)^2}\right).$$

From inequalities (57) and (58), we find

$$\frac{y'_i + \tilde{y}_i}{(y'_i + \tilde{y}_i + \tilde{h}_i)^2} \leq \frac{40}{(37)^2 n(d+1)^{3/2}(\ln n)^{3/2}\rho_1}.$$

Observing that $\tilde{h}_i \leq (1/40)(y'_i + \tilde{y}_i)$, we can now lower bound the exponent in (59) by

$$\frac{9(d+1)\ln n}{2}\left(\frac{2\tilde{h}_i(81/80)40}{(37)^2 n(d+1)^{3/2}(\ln n)^{3/2}\rho_i}\right) \leq 0.81/n.$$

■

Lemma 5.3.7 (Almost Gaussian, Jacobians) *Under the conditions of Lemma 5.3.5,*

$$\frac{\left|\mathbf{det}\left(\frac{\partial(\hat{p}_0, \hat{\boldsymbol{p}}_i)}{\partial(\hat{h}_i, \hat{\boldsymbol{g}}_i)}\right)\right|}{\left|\mathbf{det}\left(\frac{\partial(p_{i,0}, \boldsymbol{p}_i)}{\partial(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)}\right)\right|} \leq e^{.0094/n}$$

Proof We first note that

$$\left|\mathbf{det}\left(\frac{\partial(\hat{p}_0, \hat{\boldsymbol{p}}_i)}{\partial(\hat{h}_i, \hat{\boldsymbol{g}}_i)}\right)\right| = |\mathbf{det}\,(C_i)| = \frac{2^{d+1}y'_i}{(y'_i + \tilde{y}_i)^{d+2}}.$$

To compute $\left|\mathbf{det}\left(\frac{\partial(p_{i,0}, \boldsymbol{p}_i)}{\partial(\tilde{h}_i, \tilde{\boldsymbol{g}}_i)}\right)\right|$, we note that

$$\left|\frac{\partial p_{i,0}}{\partial \tilde{h}_i}\right| = \frac{-2y'_i}{(y'_i + \tilde{y}_i + \tilde{h}_i)^2}, \text{and}$$

$$\left|\frac{\partial \boldsymbol{p}_{i,j}(\tilde{h}_i, g_{i,k})}{\partial g_{i,k}}\right| = \begin{cases} 0 & \text{if } j \neq k \\ \frac{2}{y'_i + \tilde{y}_i + \tilde{h}_i} & \text{otherwise.} \end{cases}$$

Thus, the matrix of partial derivatives is lower-triangular, and its determinant has absolute value

$$\left|\mathbf{det}\left(\frac{\partial(p_{i,0},\boldsymbol{p}_i)}{\partial(\tilde{h}_i,\tilde{\boldsymbol{g}}_i)}\right)\right| = \frac{2^{d+1}y_i'}{(y_i'+\tilde{y}_i+\tilde{h}_i)^{d+2}}.$$

Thus,

$$\begin{aligned}
\frac{\left|\mathbf{det}\left(\frac{\partial(\hat{p}_0,\hat{\boldsymbol{p}}_i)}{\partial(\tilde{h}_i,\hat{\boldsymbol{g}}_i)}\right)\right|}{\left|\mathbf{det}\left(\frac{\partial(p_{i,0},\boldsymbol{p}_i)}{\partial(\tilde{h}_i,\tilde{\boldsymbol{g}}_i)}\right)\right|} &= \left(\frac{y_i'+\tilde{y}_i+\tilde{h}_i}{y_i'+\tilde{y}_i}\right)^{d+2} \\
&= \left(1+\frac{\tilde{h}_i}{y_i'+\tilde{y}_i}\right)^{d+2} \\
&\leq \left(1+\frac{3\tilde{h}_i}{2y_i'}\right)^{d+2}, && \text{by (57)} \\
&\leq e^{\frac{3(d+2)\tilde{h}_i}{2y_i'}} \\
&\leq e^{0.094/n}, && \text{by } d\geq 3 \text{ and (58)}.
\end{aligned}$$

■

6 Discussion and Open Questions

The results proved in this paper support the assertion that the shadow-vertex simplex algorithm usually runs in polynomial time. However, our understanding of the performance of the simplex algorithm is far from complete. In this section, we discuss problems in the analysis of the simplex algorithm and in the smoothed analysis of algorithms that deserve further study.

6.1 Practicality of the analysis

While we have demonstrated that the smoothed complexity of the shadow-vertex algorithm is polynomial, the polynomial we obtain is quite large. Yet, we believe that the present analysis provides some intuition for why the shadow-vertex simplex algorithm should run quickly. It is clear that the proofs in this paper are very loose and make many worst-case assumptions that are unlikely to be simultaneously valid. We did not make any attempt to optimize the coefficients or exponents of the polynomial we obtained. We have not attempted such optimization for two reasons: they would increase the length of the paper and probably make it more difficult to read; and, we believe that it should be possible to improve the bounds in this paper by *simplifying* the analysis rather than making it more complicated. Finally, we point out that most of our intuition comes from the shadow size bound, which is not so bad as the bound for the two-phase algorithm.

6.2 Further analysis of the simplex algorithm

- While we have analyzed the shadow-vertex pivot rule, there are many other pivot rules that are more commonly used in practice. Knowing that one pivot rule usually takes polynomial time makes it seem reasonable that others should as well. We consider the maximum-increase and steepest-increase rules, as well as randomized pivot rules, to be good candidates for smoothed analysis. However, the reader should note that there is a reason that the shadow-vertex pivot rule was the first to be analyzed: there is a simple geometric description of the vertices encountered by the algorithm. For other pivot rules, the only obvious characterization of the vertices encountered is by iterative application of the pivot rule. This iterative characterization introduces dependencies that make probabilistic analysis difficult.

- Even if we cannot perform a smoothed analysis of other pivot rules, we might be able to measure the diameter of a polytope under smoothed analysis. We conjecture that it is expected polynomial in m, d, and $1/\sigma$.

- Given that the shadow-vertex simplex algorithm can solve the perturbations of linear programs efficiently, it seems natural to ask if we can follow the solutions as we *unperturb* the linear programs. For example, having solved an instance of type (4), it makes sense to follow the solution as we let σ approach zero. Such an approach is often called a *homotopy* or *path-following* method. So far, we know of no reason that there should exist an $\boldsymbol{A}$ for which one cannot follow these solutions in expected polynomial time, where the expectation is taken over the choice of $\boldsymbol{G}$. Of course, if

one could follow these solutions in expected polynomial time for every $\boldsymbol{A}$, then one would have a randomized strongly-polynomial time algorithm for linear programming!

6.3 Degeneracy

One criticism of our model is that it does not allow for degenerate linear programs. It is an interesting problem to find a model of local perturbations that will preserve meaningful degeneracies. It seems that one might be able to expand upon the ideas of Todd [Tod91] to construct such a model. Until such a model presents itself and is analyzed, we make the following observations about types of degeneracies.

- In primal degeneracy, a single feasible vertex may correspond to multiple bases, I. In the polar formulation, this corresponds to an unexpectedly large number of the $\boldsymbol{a}_i$s lying in a $(d-1)$-dimensional affine subspace. In this case, a simplex method may cycle—spending many steps switching among bases for this vertex, failing to make progress toward the objective function. Unlike many simplex methods, the shadow-vertex method may still be seen to be making progress in this situation: each successive basis corresponds to a simplex that maps to an edge further along the shadow. It just happens that these edges are co-linear.

  A more severe version of this phenomenon occurs when the set of feasible points of a linear program lies in an affine subspace of fewer then d dimensions. By considering perturbations to the constraints under the condition that they do not alter the affine span of the set of feasible points, the results on the sizes of shadows obtained in Section 4 carry over unchanged. However, how such a restriction would affect the results in Section 5 is presently unclear.

- In dual degeneracy, the optimal solution of the linear program is a face of the polyhedron rather than a vertex. This does not appear to be a very strong condition, and we expect that one could extend our analysis to a model that preserves such degeneracies.

6.4 Smoothed Analysis

We believe that many algorithms will be better understood through smoothed analysis. Scientists and engineers routinely use algorithms with poor worst-case performance. Often, they solve problems that appear intractable from the worst-case perspective. While we do not expect smoothed analysis to explain every such instance, we hope that it can explain away a significant fragment of the discrepancy between the algorithmic intuitions of engineers and theorists. To make it easier to apply smoothed analyses, we briefly discuss some alternative definitions of smoothed analysis.

Zero-preserving perturbations: One criticism of *smoothed complexity* as defined in Section 1.2 is that the additive Gaussian perturbations destroy any zero-structure that the problem has, as it will replace the zeros with small values. One can refine the model to fix this problem by studying *zero-preserving perturbations*. In this model, one applies Gaussian perturbations only to non-zero entries. Zero entries remain zero.

Relative perturbations: A further refinement is the model of *relative perturbations*. Under a relative perturbation, an input is mapped to a constant multiple of itself. For

example, a reasonable definition would be to map each variable by

$$x \mapsto x(1 + \sigma g),$$

where g is a Gaussian random variable of mean zero and variance 1. Thus, each number is usually mapped to one of similar magnitude, and zero is always mapped to zero. When we measure smoothed complexity under relative perturbations, we call it *relative smoothed complexity*. Smooth complexity as defined in Section 1.2 above can be called *absolute smoothed complexity* if clarification is necessary. It would be very interesting to know if the simplex method has polynomial relative smoothed complexity.

ϵ-smoothed-complexity: Even if we cannot bound the expectation of the running time of an algorithm under perturbations, we can still obtain computationally meaningful results for an algorithm by proving that it has ϵ-smoothed-complexity $f(n, \sigma, \epsilon)$, by which we mean that the probability that it takes time more than $f(n, \sigma, \epsilon)$ is at most ϵ:n

$$\forall_{x \in X_n} \mathop{\mathbf{Pr}}_{g} [C(A, x + \sigma \max(x) g) \leq f(n, \sigma)] \geq 1 - \epsilon.$$

7 Acknowledgments

We thank Chloe for support in writing this paper, and Donna, Julia and Diana for their patience while we wrote, wrote, and re-wrote. We thank Marcos Kiwi for his comments on the earliest draft of this paper, and Mohammad Madhian for his detailed comments on the first draft. We thank Alan Edelman, Charles Leiserson, and Michael Sipser for helpful conversations. We thank Alan Edelman for proposing the name "smoothed analysis". Finally, we thank the referees for extrodinary efforts and many helpful suggestions.

References

[AC78] David Avis and Vasek Chvátal. Notes on Bland's pivoting rule. In *Polyhedral Combinatorics*, volume 8 of *Math. Programming Study*, pages 24–34. 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 at Berkeley, May 1983.

[AKS87] I. Adler, R. M. Karp, and R. Shamir. A simplex variant solving an m x d linear program in $O(min(m^2, d^2))$ expected number of pivot steps. *J. Complexity*, 3:372–387, 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*, 32(4):871–895, October 1985.

[AS70] Milton Abramowitz and Irene A. Stegun, editors. *Handbook of mathematical functions*, volume 55 of *Applied Mathematics Series*. National Bureau of Standards, 9 edition, 1970.

[AZ99] Nina Amenta and Gunter Ziegler. Deformed products and maximal shadows of polytopes. In B. Chazelle, J.E. Goodman, and R. Pollack, editors, *Advances in Discrete and Computational Geometry*, number 223 in Contemporary Mathematics, pages 57–90. Amer. Math. Soc., 1999.

[Bla35] W. Blaschke. Integralgeometrie 2: Zu ergebnissen von m. w. crofton. *Bull. Math. Soc. Roumaine Sci.*, 37:3–11, 1935.

[Bor77] Karl Heinz Borgwardt. *Untersuchungen zur Asymptotik der mittleren Schrittzahl von Simplexverfahren in der linearen Optimierung.* PhD thesis, Universitat Kaiserslautern, 1977.

[Bor80] Karl Heinz Borgwardt. *The Simplex Method: a probabilistic analysis.* Number 1 in Algorithms and Combinatorics. Springer-Verlag, 1980.

[BS95] Avrim Blum and Joel Spencer. Coloring random and semi-random k-colorable graphs. *J. Algorithms*, 19(2):204–234, 1995.

[CH92] T. F. Chan and P. C. Hansen. Some applications of the rank revealing qr factorization. *SIAM J. Sci. Stat. Comput.*, 13(3):727–741, May 1992.

[Dan51] G. B. Dantzig. Maximization of linear function of variables subject to linear inequalities. In T. C. Koopmans, editor, *Activity Analysis of Production and Allocation*, pages 339–347. 1951.

[Ede92] Alan Edelman. Eigenvalue roulette and random test matrices. In Marc S. Moonen, Gene H. Golub, and Bart L. R. De Moor, editors, *Linear Algebra for Large Scale and Real-Time Applications*, NATO ASI Series, pages 365–368. 1992.

[Efr65] Bradley Efron. The convex hull of a random set of points. *Biometrika*, 52(3/4):331–343, 1965.

[Fel68] William Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. John Wiley & Sons, 1968.

[Fel71] William Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. John Wiley & Sons, 1971.

[FK] Uri Feige and Robert Krauthgamer. Improved performance guarantees for bandwidth minimization heuristics. November 1998.

[FK98] U. Feige and J. Kilian. Heuristics for finding large independent sets, with applications to coloring semi-random graphs. In IEEE, editor, *39th Annual Symposium on Foundations of Computer Science: proceedings: November 8–11, 1998, Palo Alto, California*, pages 674–683, 1109 Spring Street, Suite 300, Silver Spring, MD 20910, USA, 1998. IEEE Computer Society Press.

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

[GS79] Donald Goldfarb and William T. Sit. Worst case behaviour of the steepest edge simplex method. *Discrete Applied Math*, 1:277–285, 1979.

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

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

[Kal92] G. Kalai. A subexponential randomized simplex algorithm. In *Proc. 24th Ann. ACM Symp. on Theory of Computing*, pages 475–482, Victoria, B.C., Canada, May 1992.

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

[Kha79] L. G. Khachiyan. A polynomial algorithm in linear programming. *Doklady Akademia Nauk SSSR*, pages 1093–1096, 1979.

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

[KM72] V. Klee and G. J. Minty. How good is the simplex algorithm ? In Shisha, O., editor, *Inequalities – III*, pages 159–175. Academic Press, 1972.

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

[Mil71] R. E. Miles. Isotropic random simplices. *Adv. Appl. Prob.*, 3:353–382, 1971.

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

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

[Ren94] J. Renegar. Some perturbation theory for linear programming. *Math. Programming*, 65(1, Ser. A):73–91, 1994.

[Ren95a] J. Renegar. Incorporating condition measures into the complexity theory of linear programming. *SIAM J. Optim.*, 5(3):506–524, 1995.

[Ren95b] J. Renegar. Linear programming, complexity theory and elementary functional analysis. *Math. Programming*, 70(3, Ser. A):279–351, 1995.

[RS63] A. Renyi and R. Sulanke. Uber die convexe hulle von is zufallig gewahlten punkten, i. *Z. Whar.*, 2:75–84, 1963.

[RS64] A. Renyi and R. Sulanke. Uber die convexe hulle von is zufallig gewahlten punkten, ii. *Z. Whar.*, 3:138–148, 1964.

[San76] Luis A. Santalo. *Integral Geometry and Geometric Probability*. Encyclopedia of Mathematics and its Applications. Addison-Wesley, 1976.

[Sma82] S. Smale. The problem of the average speed of the simplex method. In *Proceedings of the 11th International Symposium on Mathematical Programming*, pages 530–539, August 1982.

[Sma83] S. Smale. On the average number of steps in the simplex method of linear programming. *Mathematical Programming*, 27:241–262, 1983.

[SV86] Miklos Santha and Umesh Vazirani. Generating quasi-random sequences from semi-random sources. *JCSS*, 33:75–87, 1986.

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

[Tod91] M. J. Todd. Probabilistic models for linear programming. *Mathematics of Operations Research*, 16(4):671–693, 1991.

[TTY01] M. J. Todd, L. Tunçel, and Y. Ye. Characterizations, bounds, and probabilistic analysis of two complexity measures for linear programming problems. *Math. Program., Ser. A*, 90:59–69, 2001.

[Tur86] Johathan S. Turner. On the probable performance of a heuristic for bandwidth minimization. *SIAM Journal on Computing*, 15(2):561–580, 1986.

[VY96] S. A. Vavasis and Y. Ye. A primal-dual interior-point method whose running time depends only on the constraint matrix. *Math. Program.*, 74:79–120, 1996.

Index