Smoothed Analysis of Linear Programming

Martin Licht

December 19, 2014

Und die Bewegung in den Rosen, sieh:
Gebärden von so kleinem Ausschlagswinkel,
daß sie unsichtbar blieben, liefen ihre
Strahlen nicht auseinander in das Weltall.

Rainer Maria Rilke

Contents

1 Introduction

Many scientific texts on the complexity of simplex methods begin with the statement that simplex methods generally feature acceptable running times on practical problems while for each variant of simplex methods we known a series of instances which realizes worst-case exponential running time. While a deterministic provably polynomial-time simplex method remained elusive, theorists grew to formalize the observation that these wort-case instances are special and 'rare' among the possible input data. The pursuit of new measures in complexity theory is justified by the practical relevance of linear programming problems and the gap between theoretical upper bounds and the observed performance of simplex methods. Furthermore, fundamental research might lay the ground for applications beyond the initial area of research and provide useful tools in other contexts.

The principle endeavour of this thesis is to elaborate on the smoothed analysis of the shadow vertex simplex method by Vershynin [15], which itself builds on top and improves the smoothed analysis in the seminal publication of Spielman and Teng [13]. The underlying idea is to estimate the expected number of pivot steps in the simplex method when the input data are Gaussian variables. We expect such a presentation to be beneficial for the scientific community.

To ensure self-containedness of the thesis and to impose a common background, we also include a steep introduction into the theory of linear programming problems. Throughout all of this thesis, we lay stress on an sound and rigorous mathematical elaboration of the topic.

Smoothed Analysis and its Historical Context The relevance of linear programming for industrial applications, e.g. as a computational framework for resource allocation, has been elucidated in the first place by the US-American mathematician George Dantzig and the Russian mathematician Leonid Kantorovich. The invention of the simplex method, which seems to have been the most important practical method in computional linear programming, is generally credited to Dantzig. The underlying idea is that an optimal solution of a linear programming problem is located at a vertex of the underlying polyhedron; then, having found an initial vertex, the simplex traverses between neighbouring vertices until an optimal solution has been found. Each such step is called a pivot step and can be performed with complexity polynomial in the input size.

Notwithstanding the favourable performance of the simplex method in practical applications, in comparison to provably polynomial time methods such as ellipsoid-type methods, for any variant of the simplex method we are aware of a series of instances, parametrized over the dimension of the search space, which forces the simplex method to perform a number of pivot steps asympotically exponential in the input size. It has soon been realized that these instances are pathological only, and that understanding this theorical gap (or closing it), is considered one of the most important issues in computational mathematics.

The intuition that the expected number of pivot steps is merely polynomial in the input size, provided that the input has a random component, can be found at least as early as in [5]. A notable precursion to the contempary smoothed analysis is the work of Borgwardt [2], who studies the expected running time of the shadow vertex simplex method [4] on unit linear programming problems where the constraint vectors are distributed independently, identically and rotationally symmetric. These and similar approaches obtained complexity bounds polynomial in the input size.

It is a natural idea to view input variables as Gaussian variables, because Gaussian probability distributions model the effects measurement noise in many applications. The complexity

estimate depends on the dimension of variables d, the number of constraints n, and the standard deviation σ of the Gaussian input variables. In [13] (as cited in [15]), the authors obtain the bound

$$\mathcal{O}^*(n^{86}d^{55}\sigma^{-30}),$$

neglecting polylogarithmic factors. This upper bound has been improved in [15, Theorem 1.2] to

$$\mathcal{O}((d^9 + d^3\sigma^{-4})\log^7 n)$$

for $n > d \geq 3$ and σ small enough. Notably, this bound is merely polylogarithmic in n. Intuitively, the deterministic limit is approached as $\sigma \to 0$. The blow-up coincides with the lack of a polynomial classical complexity bound. Conversely, as $\sigma \to \infty$, the input approaches the setting of [2], where a bound polynomial in d and sublinear in n is expected.

Thesis Structure and Results The sections of this thesis follow a loose grouping. We first assemble an outline of linear programming theory. Second, we collect pertinent elements of stochastic analysis. In the description and the smoothed analysis of the shadow vertex simplex method we closely follow the order of [15]. The algorithm and its subroutines are described in a bottom approach, until only the shadow sizes remain to be estimated. Due to the inherent difficulty, this last step is accomplished in the two final sections.

We begin with an essentially self-contained introduction into the theory of linear programmig problems in Sections 2 – 5. We consider only the structural theory of convex sets, polyhedrons and the variational theory of functionals over them, and completely disregard algorithmic aspects of linear programming. We only assume familiarity with basic concepts from first-year courses in mathematics, in particular linear algebra, multivariate analysis and basic topology. The motivation for this elaboration is two-fold. On the one hand, the author's knowledge in the theory of linear programming was only superficial prior to this thesis, thus the writing of this section served to familiarize himself with the theory. On the other hand, considering the vast amount of literature on the topic and the diversity with respect to the proficiency of possible readers, it is ought to provide a point of reference in the presentation of the actual topic of this thesis.

In Section 2 we outline basic results on the optimization of linear functionals over convex sets. This lays the ground for the specialized polyhedral theory, which is not yet treated explicitly here. The main results are the variational theory of faces, and the Caratheodory-Steinitz-type results. We do not refrain from modest usage of analytical tools. While this is unusual in theoretical computer science, it is not at odds with scientific history: Indeed, important historical contributors to convex optimization, like, e.g., Krein, Milman and Choquet, have conducted seminal work in functional analysis, too. (cmp. [18, p.452])

The general convex theory is specialized to the polyhedral theory in Section 3, which introduces basic definitions and structural results. Again, the notion of faces and the natural partial order on them is the most important notion.

Section 4 puts the spotlight on the fundamental theorem of linear inequalities, whose main application is a short proof of Farkas' lemma. We subsequently prove classical results on the decomposition of polyhedra.

Eventually, Section 5 formally introduces the notion of linear programming problem, the dual linear programming problem and the notion of polar polyhedra.

Having gathered the basal aspects of linear programming theory, we turn our attention towards the stochastic analysis pertaining to the smoothed analysis in Section 6. We give a self-contained review of selected results on Gaussian random vectors, on the uniform distribution on unit spheres, the Haar measure on the orthogonal group, and on random variables over sequences. We necessarily take the perspective of a non-probabilist, since the author only has an eclectic background by himself. Notably, we neither follow the track of the seminal paper by Spielman and Teng [13], nor the selection of results in [15], but combine the pertinent sources to a different presentation.

Next, the pivotal algorithm `Shadow Vertex Simplex Method` is introduced in Section 7. We lean our presentation to the book of Borgwardt [2], but our formalism is notably different. We introduce basic notions of the solution theory of unit linear programming problems, including concepts specifically relevant to the algorithmic implementation. We study basic properties of projections of polyhedra onto planes, and eventually outline algorithm `Shadow Vertex Simplex Method` with attention to the algorithmic details.

We approach the algorithmic implementation of the complete solution algorithm with a bottom-up approach. Step by step we introduce the algorithmic details of Phase I and Phase II. However, the proofs for the bounds of the pivot steps are postponed to later sections.

In Section 8 we introduce the randomized (Monte-Carlo) algorithm `Adding Constraints`. Given constraint vectors $a_1, \dots, a_n$, that algorithm produces a randomized set of additional constraint vectors $a_{n+1}, \dots, a_{n+d}$ and a functional direction z_0, such that the index set $\{n+1, \dots, n+d\}$ describes an optimal solution of the unit linear programming problem with functional z and constraint vectors $a_1, \dots, a_{n+d}$. We follow the presentation of [15, Appendix B], but our results for the success probability are stated in very general parametrized form, which allows for the improvement of some absolute constants in the smoothed analysis and, more importantly, an exposition of the scaling invariance of the probability estimates.

The full Phase I algorithm is briefly outlined in Section 9. Phase I is a randomized algorithm of Las-Vegas-type. We iteratively build a sequence of linear programming problems with addtional constraints until suitable initial data for an application of the shadow vertex method have been found. The eventual bound for the expected total number of pivot steps is given in abstract form, to be instantiated later.

In Phase II of the solution algorithm, not the original linear programming is solved, but an "interpolation" linear programming problem, which introduces an additional variable τ and whose goal functional is parametrized over $\mathbb{R}$. We describe the essential properties of the interpolation program in Section 10. The Phase II algorithm formally corresponds to solving a sequence of instatiations of the interpolation problem where the parameter of the goal functional formally moves from $-\infty$ to ∞.

The top level algorithm is eventually described in Section 11. This incorporates the ideas in Sections 8 and 9, and the uses the interpolation problem of the previuos section in the Phase II part. On the one hand, we instantiate the abstract results of the previous sections; on the other hand, we anticipate the bonds of the shadow size in the followings sections.

The remainder of the thesis elaborates the estimates of the shadow sizes in Phase I and Phase II.

In Section 12 we describe the pivotal result on the number of edges of the intersection of a random polytope with a plane, and some of its variations. Notably we slightly generalize Lemmas 7.2 and 7.3 of [15], and show a distinctive generalization of Lemma 4.0.6 of [13]. We thus are able to provide a clear presentation of Section 7 of [15]. A minor concession, due to the desired size of this thesis, is that Lemma 7.5 of [15] is only cited, but not proved. Eventually, the section provides the desired Phase II estimate.

The estimate for Phase I builds on top of the previous bound. The analysis boils down to two steps. First, the maximum norm of the input constraints is located within a sutiable band, where it serves as an input for `Adding Constraints`. Second, we use a variation of Theorem 12.3 to estimate the shadow size of the auxiliary linear programming problems in Phase I.

We close the work with Section 14, where some open questions and possible extensions are outlined.

This thesis has achieved its goal to provide a (mostly) self-contained presentation of the smoothed analysis in [15]. The original paper is very dense and concise, and references results of [13]. Instead, we include all necessary preleminary results and avoid conceptual leaps, at the expense (and with the gain) of a rather technical and formal presentation. A particular endeavour has been to elucidate the behaviour of the smoothed analysis under scaling of the variables. This behaviour has been completely understood and elaborated with the completion of this thesis.

Acknowledgements I would like to thank Prof. Heiko Röglin for acquainting me with this very interesting topic, assigning this thesis to me and examing it, and his continual willingness for constructive advice. Furthermore, I would like to thank Prof. Rolf Klein for coexamining the present work. The progression has greatly benefited from the stimulating and pleasant working atmosphere at the Institute of Computer Science I of the University of Bonn, for which I would like to express my appreciation to the complete and exceptionlessly enjoyable staff, and I wish the very best for their futures. Last but not least, a considerable amount of financial support has been due to my parents during the past year.

Conventions

We use the following notational guidelines throughout this thesis, which are never to be understood as too strict. We use Latin uppercase letters (A,B,C,...) for matrices; we use Latin lowercase letters (a,b,c,...) for vectors and for integer indices; we use Greek lowercase letters (α,β,γ,...) for scalar entities in $\mathbb{R}$; we use calligraphic Latin uppercase letters ($\mathcal{A}$,$\mathcal{B}$,$\mathcal{C}$,...) for subsets of $\mathbb{R}^n$; we use fractured Latin uppercase letters ($\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$, $\cdots$) for events in the manner of probability theory.

We freely use notation like $a+\mathcal{U} := \{a+u \mid u \in \mathcal{U}\}$ or $\mathcal{U}+\mathcal{V} = \{u+v \mid u \in \mathcal{U}, v \in \mathcal{V}\}$, and similarly for other operations. We write $B_r(x)$ for the r-ball around some point x. For any set $\mathcal{A} \subset \mathbb{R}^d$, we let $\operatorname{vol}^d \mathcal{A}$ the d-dimensional Hausdorff volume of $\mathcal{A}$.

When $n \in \mathbb{N}$ we let $[n] := \{1, \ldots, n\}$.

We assume that all vector spaces in this work are real vector spaces. We denote by $\vec{1}$ and $\vec{0}$ the all-one and the all-zero vectors, respectively, of finite-dimensional real vector spaces. For a sequence of vectors $x_0, x_1, \ldots$ in $\mathbb{R}^d$ indexed by natural numbers, we write $x_n \to x$ if that sequence converges to $x \in \mathbb{R}^d$.

2 Convex Geometry

The aim of this section is to introduce some basic results from linear algebra and convex geometry. We recall basic notions from linear algebra and topology in Subsection 2.1. Then, in Subsection 2.2, we derive the separation theorems of convex sets. In Subsection 2.3 we introduce the concept of faces of convex sets, with a particular focus of the variational theory of linear functionals. This leads to a proof of the Krein-Milman theorem. Finally, in Subsection 2.4, we give elementary proofs of two theorems of Caratheodory-Steinitz-type.

2.1 Basic Concepts

We begin with basic definitions. Throughout this subsection, let $\mathcal{A} \subseteq \mathbb{R}^d$. We say that

$$\begin{aligned}
\mathcal{A} \text{ is convex} &\iff \forall a_1, a_2 \in \mathcal{A}, \lambda \in [0,1] : \lambda a_1 + (1-\lambda) a_2 \in \mathcal{A} \\
&\iff \forall a_1, a_2 \in \mathcal{A}, \lambda_1, \lambda_2 \in \mathbb{R}_0^+, \lambda_1 + \lambda_2 = 1 : \lambda_1 a_1 + (1-\lambda_2) a_2 \in \mathcal{A}, \\
\mathcal{A} \text{ is affine} &\iff \forall a_1, a_2 \in \mathcal{A}, \lambda_1, \lambda_2 \in \mathbb{R}, \lambda_1 + \lambda_2 = 1 : \lambda_1 a_1 + \lambda_2 a_2 \in \mathcal{A}, \\
\mathcal{A} \text{ is conical} &\iff \forall a_1, a_2 \in \mathcal{A}, \lambda_1, \lambda_2 \in \mathbb{R}_0^+ : \lambda_1 a_1 + \lambda_2 a_2 \in \mathcal{A}, \\
\mathcal{A} \text{ is linear} &\iff \forall a_1, a_2 \in \mathcal{A}, \lambda_1, \lambda_2 \in \mathbb{R} : \lambda_1 a_1 + \lambda_2 a_2 \in \mathcal{A}.
\end{aligned}$$

Affine and conical spaces are convex. The intersection of convex, affine or conical sets is again a convex, affine or conical set, respectively. We define $\operatorname{convex} \mathcal{A}$, $\operatorname{aff} \mathcal{A}$ and $\operatorname{cone} \mathcal{A}$ to be the smallest convex, affine or conical set, respectively, containing $\mathcal{A}$. Furthermore, we let $\operatorname{lin} \mathcal{A}$ be the smallest linear space containing $\mathcal{A}$. Their existence can be proven by Zorn's lemma; direct constructions are given by

$$\begin{aligned}
\operatorname{convex} \mathcal{A} &:= \left\{ \sum_{i=1}^k \lambda_i a_i \,\middle|\, k \geq 1, a_i \in \mathcal{A}, \lambda_i \in \mathbb{R}_0^+, \sum_{i=0}^k \lambda_i = 1 \right\}, \\
\operatorname{aff} \mathcal{A} &:= \left\{ \sum_{i=1}^k \lambda_i a_i \,\middle|\, k \geq 1, a_i \in \mathcal{A}, \lambda_i \in \mathbb{R}, \sum_{i=0}^k \lambda_i = 1 \right\}, \\
\operatorname{cone} \mathcal{A} &:= \left\{ \sum_{i=1}^k \lambda_i a_i \,\middle|\, k \geq 1, a_i \in \mathcal{A}, \lambda_i \in \mathbb{R}_0^+ \right\}, \\
\operatorname{lin} \mathcal{A} &:= \left\{ \sum_{i=1}^k \lambda_i a_i \,\middle|\, k \geq 1, a_i \in \mathcal{A}, \lambda_i \in \mathbb{R} \right\}.
\end{aligned}$$

A non-empty set $\mathcal{A}$ is affine if and only if there does exist a linear subspace $\mathcal{U} \subseteq \mathbb{R}^d$ such that for some $a \in \mathcal{A}$ we have $\mathcal{A} = a + \mathcal{U}$. Note that $\mathcal{U}$ is independent of a and that $\mathcal{A} = a' + \mathcal{U}$ for any $a' \in \mathcal{A}$. An affine space $\mathcal{A}$ is a linear subspace if and only if $0 \in \mathcal{A}$. We observe

$$\operatorname{convex} \mathcal{A} \subseteq \operatorname{aff} \mathcal{A} \cap \operatorname{cone} \mathcal{A}, \quad \operatorname{aff} \mathcal{A} \subseteq \operatorname{lin} \mathcal{A}, \quad \operatorname{cone} \mathcal{A} \subseteq \operatorname{lin} \mathcal{A}.$$

We can reasonably talk about the dimensions of convex sets. If $\mathcal{A} = a + \mathcal{U}$ is affine, then $\dim \mathcal{A} = \dim \mathcal{U}$ is called the dimension of $\mathcal{A}$. If instead $\mathcal{A}$ is merely convex, then we define the dimension as $\dim \mathcal{A} = \dim \operatorname{aff} \mathcal{A}$. A different characterization, which is inspired by Definition 3.2 in [7], can be given in terms of matrices.

Theorem 2.1.
Let $\mathcal{A}$ be a non-empty convex set. Then

$$\dim \mathcal{A} = \min \left\{ \dim \ker M \mid M \in \mathbb{R}^{d \times d}, \forall x, y \in \mathcal{A} : Mx = My \right\}$$

$$= \min \left\{ \dim \ker M \mid M \in \mathbb{R}^{d\times d}, \forall x, y \in \operatorname{aff} \mathcal{A} : Mx = My \right\}.$$

Proof. Let $x \in \mathcal{A}$. Then $\operatorname{aff} \mathcal{A} - x$ is a linear subspace. Let M be the orthogonal projection onto $(\operatorname{aff} \mathcal{A} - x)^{\perp}$, so $M\mathcal{A} = \{Mx\}$. Then $\dim \ker M = \dim \operatorname{aff} \mathcal{A}$. It remains to show that there is no M that is constant on $\mathcal{A}$ and whose kernel has a smaller dimension. Indeed, if $Mx = My$ for $x, y \in \mathcal{A}$, then for $\alpha, \beta \in \mathbb{R}$ with $\alpha + \beta = 1$ we see

$$M(\alpha x + \beta y) = \alpha Mx + \beta My = (\alpha + \beta)Mx = Mx.$$

Hence M is constant on $\operatorname{aff} \mathcal{A}$. But then it vanishes on $\operatorname{aff} \mathcal{A} - x$, so $\dim \ker M \geq \dim \operatorname{aff} \mathcal{A}$. □

We call the vectors $a_0, \ldots, a_k$ affinely independent if the vectors $a_1 - a_0, \ldots, a_k - a_0$ are linearly independent. Note that this does not depend on the order of the vectors.

Next we introduce some basic topological concepts that we will use frequently. We refer to [6] or [17] for basic notions in topology of $\mathbb{R}^d$.

Lemma 2.2.
If $\mathcal{A}$ is convex, then its closure $\overline{\mathcal{A}}$ is convex.

Proof. Let $\mathcal{A} \neq \emptyset$ and x_k and y_k be convergent sequences in $\mathcal{A}$. Let $x, y \in \overline{\mathcal{A}}$ such that $x_k \to x$ and $y_k \to y$. For $\lambda \in [0, 1]$ let $z = \lambda x + (1 - \lambda)y$ and $z_k = \lambda x_k + (1 - \lambda)y_k$. Then $z_k \to z$ and $z_k \in \mathcal{A}$. This implies that $z \in \overline{\mathcal{A}}$, so $\overline{\mathcal{A}}$ is convex. □

We note that $\operatorname{lin} \mathcal{A}$ and $\operatorname{aff} \mathcal{A}$ are always closed, while $\operatorname{convex} \mathcal{A}$ and $\operatorname{cone} \mathcal{A}$ need not be closed, even if $\mathcal{A}$ is a closed set.

Lemma 2.3.
A convex set $\mathcal{A} \subseteq \mathbb{R}^d$ has an interior point if and only if $\mathcal{A}$ is d-dimensional.

Proof. If $\mathcal{A}$ has an interior point x, then it contains a closed d-dimensional ball $B_\rho(x)$ for some $\rho > 0$. Then its affine closure must have dimension d, so $\dim \mathcal{A} = d$. Conversely, suppose that $\mathcal{A}$ is d-dimensional. We may assume $\vec{0} \in \mathcal{A}$ after a translation, so $\operatorname{aff} \mathcal{A} = \operatorname{lin} \mathcal{A}$. Then $\mathcal{A}$ must contain a set of d linearly independent points $a_1, \ldots, a_d$. In conclusion, the d-simplex $\operatorname{convex}\{\vec{0}, a_1, \ldots, a_d\}$ is contained in $\mathcal{A}$, so $\mathcal{A}$ contains an interior point. □

We call a point $a \in \mathcal{A}$ a relative boundary point if it is a boundary point of $\mathcal{A}$ within $\operatorname{aff} \mathcal{A}$ with that set equipped with the relative topology. We call a point a relative interior point, if it is an interior point of $\mathcal{A}$ within $\operatorname{aff} \mathcal{A}$ with that set equipped with the relative topology. Note that a singleton has no boundary point and its single member is relative interior. It is easy to see that any non-empty convex set $\mathcal{A}$ has a non-empty relative interior.

Corollary 2.4.
Any non-empty convex set $\mathcal{A}$ has a relative interior point. □

Lemma 2.5.
Let $x \in \mathcal{A}$ be a relative interior point. Then for any $a \in \mathcal{A}$ there exist $b \in \mathcal{A}$ and $\lambda \in (0, 1)$ such that $x = \lambda a + (1 - \lambda)b$.

Proof. Let $\mathcal{L}$ be the infinite line through a and x. Then $\mathcal{L} \setminus \{x\}$ consists of two non-empty segments, because x is a relative interior point. Then we may choose $b \in \mathcal{A}$ in the segment of $\mathcal{L}$ that does not contain a. □

2.2 Separation Theorems and Linear Functionals

Recall that we are given the canonical Euclidean scalar product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^d$, and that $\|x\| = \sqrt{\langle x, x \rangle}$ for any $x \in \mathbb{R}^d$. For any vector $c \in \mathbb{R}^d$ we have a linear functional $\langle c, \cdot \rangle$ over $\mathbb{R}^d$. For any $\alpha \in \mathbb{R}$, the α-level set of that functional (or, shorter, of c) is the set of vectors x such that $\langle c, x \rangle = \alpha$. Note that α-level sets are hyperplanes, provided that $c \neq \vec{0}$. Indeed, if $\langle c, x_\alpha \rangle = \alpha$, then the α level set is given by $x_\alpha + c^\perp$.

Let $\mathcal{A} \subseteq \mathbb{R}^d$. Then we write

$$\operatorname{dist}_{\mathcal{A}}(x) := \inf_{a \in \mathcal{A}} \|x - a\|$$

for the distance of a point x from $\mathcal{A}$. Note that $\operatorname{dist}_{\mathcal{A}}(x) = 0$ if and only if $x \in \overline{\mathcal{A}}$. The infimum is in fact a minimum provided that $\mathcal{A}$ is closed:

Lemma 2.6.
Let $\mathcal{A} \neq \emptyset$ be closed. For $x \in \mathbb{R}^d$ there exists $a \in \mathcal{A}$ such that $\operatorname{dist}_{\mathcal{A}}(x) = \|x - a\|$.

Proof. Let $\rho = \operatorname{dist}_{\mathcal{A}}(x)$ and set $\mathcal{B} = \mathcal{A} \cap B_{\rho+1}(x)$. Then $\mathcal{B}$ is closed and bounded, i.e., it is compact. The continuous function $\|x - \cdot\|$ thus has a minimizer a over $\mathcal{B}$. Because $\mathcal{A} \setminus \mathcal{B}$ has a distance from x larger than ρ, the element a is the minimizer of the distance over $\mathcal{A}$. □

Our next goal is deriving the standard separation theorems. The presentation is inspired by the usual treatment of the separation theorems in functional analysis. Notably, we do not utilize Minkowski functionals for the proofs.

Lemma 2.7 ([18, Lemma V.3.3], [8, Appendix 1, Theorem 3])**.**
Suppose that $\mathcal{A}$ is non-empty, closed and convex, $a \in \mathcal{A}$ and $x \in \mathbb{R}^d$. Then

$$\|a - x\| = \operatorname{dist}_{\mathcal{A}}(x) \quad \Longleftrightarrow \quad \forall y \in \mathcal{A} : \langle x - a, y - a \rangle \leq 0.$$

The last inequality is strict for $x \notin \mathcal{A}$. Furthermore, such a vector $a \in \mathcal{A}$ is unique.

Proof. Suppose the right-hand statement holds. Then

$$\|x - y\|^2 = \|x - a + a - y\|^2 = \|x - a\|^2 + \|a - y\|^2 - 2\langle x - a, y - a \rangle \geq \|x - a\|^2.$$

Conversely, suppose the left-hand statement holds. For any $\tau \in [0, 1]$ we see

$$\begin{aligned} \|x - a\|^2 \leq \|x - (1 - \tau)a + \tau y\|^2 &= \langle x - (1 - \tau)a + \tau y, x - (1 - \tau)a + \tau y \rangle \\ &= \|x - a\|^2 + \tau^2 \|a - y\|^2 + 2\tau \langle x - a, a - y \rangle, \end{aligned}$$

which implies $\langle x - a, y - a \rangle \leq \frac{\tau}{2} \|a - y\|^2$ for $0 < \tau \leq 1$. For $\tau \to 0$ the inequality of the right-hand statement follows. This proves the equivalence of both statements. In case that $x \notin \mathcal{A}$, strictness of the inequality follows from $a \neq x$ and

$$0 \leq \langle a - x, y - a \rangle = \langle a - x, y - x \rangle - \langle a - x, a - x \rangle = \langle a - x, y - x \rangle - \|a - x\|^2,$$

because $\|a - x\| > 0$.

To prove the uniqueness, suppose that $a, a' \in \mathcal{A}$ satisfy $\|a' - x\| = \|a - x\| = \operatorname{dist}_{\mathcal{A}}(x)$, and without loss of generality we assume $x = 0$. Then by convexity we have $\frac{1}{2}a + \frac{1}{2}a' \in \mathcal{A}$, and a and a' are not colinear by the minimization property. Then the uniqueness follows from Minkowski's inequality. □

We call $\mathcal{H}$ a supporting hyperplane of $\mathcal{A}$ at $a \in \mathbb{R}^d$ if $a \in \mathcal{H}$ and $\mathcal{A}$ lies within one of the two closed halfspaces induced by $\mathcal{H}$.

Theorem 2.8 ([8, Appendix 1, Theorem 4])**.**
If $a \in \partial\mathcal{A}$, then there exists a supporting hyperplane of $\mathcal{A}$ at a.

Proof. If the claim holds for $\overline{\mathcal{A}}$, then it holds for $\mathcal{A}$ as well. Therefore we may assume that $\mathcal{A}$ is closed. Let $b_k \to a$ with $b_k \notin \mathcal{A}$. Let $q_k \in \mathcal{A}$ be the unique minimizer of the distance to b_k over $\mathcal{A}$. Then we know that $\langle q_k - b_k, \cdot - b_k\rangle > 0$ over $\mathcal{A}$, and conclude that $\|q_k - b_k\|^{-1}\langle q_k - b_k, \cdot - b_k\rangle > 0$ over $\mathcal{A}$. The sequence $\|q_k - b_k\|^{-1}(q_k - b_k)$ has a accumulation point $q \in S_1$. By limit arguments we infer that $\langle q, \cdot - a\rangle \geq 0$ over $\mathcal{A}$. The desired hyperplane is $a + \operatorname{lin}\{q\}^\perp$. □

Theorem 2.9 (separation theorem, first version)**.**
Suppose that $\mathcal{A}$ is convex and closed and $\vec{0} \notin \mathcal{A}$. Then there exists $c \in \mathbb{R}^d$ such that

$$\forall x \in \mathcal{A} : \langle c, x\rangle > 0.$$

Proof. Let $\mathcal{A} \neq \emptyset$ and $a \in \mathcal{A}$ minimize $\|\cdot\|$ over $\mathcal{A}$. Then $a \neq 0$, i.e., $\|a\| \neq 0$, and Lemma 2.7 implies that $\langle a, y\rangle \geq \|a\|^2 > 0$ for $y \in \mathcal{A}$. □

Theorem 2.10 (separation theorem, second version)**.**
Suppose that $\mathcal{A}$ is convex and open and $\vec{0} \notin \mathcal{A}$. Then there exists $c \in \mathbb{R}^d$ such that

$$\forall x \in \mathcal{A} : \langle c, x\rangle > 0.$$

Proof. If $\vec{0} \notin \overline{\mathcal{A}}$, then we may apply the previous theorem to $\overline{\mathcal{A}}$. Otherwise $\vec{0} \in \partial\mathcal{A}$. Let $\mathcal{H}$ be a supporting hyperplane of $\mathcal{A}$ through $\vec{0}$, and let $c \perp \mathcal{H}$ point towards $\mathcal{A}$. Because $\mathcal{A}$ is open we know that $\mathcal{H} \cap \mathcal{A} = \emptyset$, and the claim follows. □

Theorem 2.11 (separation theorem, third version)**.**
If $\mathcal{A}$ and $\mathcal{B}$ are convex and disjoint, with $\mathcal{A}$ open, then there exists $c \in \mathbb{R}^d$ such that

$$\forall a \in \mathcal{A}, b \in \mathcal{B} : \langle c, a\rangle < \langle c, b\rangle.$$

Proof. Let $\mathcal{C} := \mathcal{A} - \mathcal{B}$. Since $\mathcal{A}$ and $\mathcal{B}$ are disjoint we find $\vec{0} \notin \mathcal{C}$. Since $\mathcal{A}$ is open, so is $\mathcal{C}$. We derive the existence of c as desired by the Theorem 2.10. □

2.3 Faces of Convex Sets

Our goal in this section is to discuss faces of convex sets, and to derive the Krein-Milman theorem.

Let $\mathcal{A}$ be convex. We call $\mathcal{F} \subseteq \mathcal{A}$ a face of $\mathcal{A}$ if $\mathcal{F}$ is convex and

$$\forall x, y \in \mathcal{A}, \lambda \in [0,1] : (\lambda x + (1-\lambda)y \in \mathcal{F} \implies x, y \in \mathcal{F}).$$

We call $x \in \mathcal{A}$ an extremal point, if $\{x\}$ is a face, i.e., if x is not the convex combination of other points in $\mathcal{A}$. We let $\operatorname{ex}\mathcal{A}$ denote the set of extremal points of $\mathcal{A}$.

Lemma 2.12.
Suppose that $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{A}$ are convex subsets of $\mathbb{R}^d$ such that $\mathcal{F} \subseteq \mathcal{G}$ is a face of $\mathcal{G}$ and $\mathcal{G} \subseteq \mathcal{A}$ is a face of $\mathcal{A}$. Then $\mathcal{F}$ is a face of $\mathcal{A}$.

Proof. Suppose that $\mathcal{F} \neq \emptyset$ and $z \in \mathcal{F}$, $\lambda \in [0,1]$ and $x, y \in \mathcal{C} \setminus \mathcal{F}$ such that $\lambda x + (1-\lambda)y = z$. Then $x \in \mathcal{A} \setminus \mathcal{G}$ contradicts that $\mathcal{G}$ is a face of $\mathcal{A}$, while $x \in \mathcal{G} \setminus \mathcal{F}$ contradicts $\mathcal{F}$ being a face of $\mathcal{G}$. Thus $x \in \mathcal{F}$, and therefore $\mathcal{F}$ is a face of $\mathcal{A}$. □

Lemma 2.13.
Suppose that $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{A}$ are convex subsets of $\mathbb{R}^d$ such that $\mathcal{F} \subseteq \mathcal{G}$ is a face of $\mathcal{A}$ and $\mathcal{G} \subseteq \mathcal{A}$ is a face of $\mathcal{A}$. Then $\mathcal{F}$ is a face of $\mathcal{G}$. □

Proof. If $\mathcal{F}$ is not a face of $\mathcal{G}$, then there exist $x, y \in \mathcal{G}$ and $\lambda \in (0,1)$ with x or y in $\mathcal{G} \setminus \mathcal{F}$, such that $\lambda x + (1-\lambda)y \in \mathcal{F}$. But then $\mathcal{F}$ is not a face of $\mathcal{A}$, which is a contradiction. So $\mathcal{F}$ is a face of $\mathcal{G}$. □

Corollary 2.14.
If $\mathcal{F}$ is a face of $\mathcal{A}$, then $\operatorname{ex}\mathcal{F} = \operatorname{ex}\mathcal{A} \cap \mathcal{F}$. □

Lemma 2.15.
A convex set $\mathcal{F} \subseteq \mathcal{A}$ is a face of $\mathcal{A}$ if and only if $\mathcal{A} \setminus \mathcal{F}$ is convex.

Proof. If $\mathcal{A} \setminus \mathcal{F}$ is convex then

$$\forall x, y \in \mathcal{A} : \lambda \in [0,1] : \lambda x + (1-\lambda)y \notin \mathcal{F}.$$

Then $\mathcal{F}$ satisfies the definition of a face. If $\mathcal{F}$ is a face, then

$$\forall x, y \in \mathcal{A} \setminus \mathcal{F}, \lambda \in [0,1] : \lambda x + (1-\lambda)y \in \mathcal{A} \setminus \mathcal{F},$$

which had to be shown. □

Corollary 2.16.
Relative interior points of a convex set $\mathcal{A}$ are not contained in any proper face. Any convex set has a point not contained in any proper face.

Lemma 2.17.
If $\mathcal{F} \subset \mathcal{A}$ is a proper face of $\mathcal{A}$, then $\dim \mathcal{F} < \dim \mathcal{A}$.

Proof. We have $\dim F \leq \dim A$ since $\operatorname{aff}\mathcal{F} \subseteq \operatorname{aff}\mathcal{A}$. Suppose that $\dim \mathcal{F} = \dim \mathcal{A}$. Then $\dim \mathcal{F}$ contains an interior point of $\mathcal{A}$, which is an interior of $\mathcal{A}$, too. But this contradicts the previous corollary. □

Lemma 2.18.
If $\mathcal{F}$ is a non-empty proper face of $\mathcal{A}$, then there exists $\vec{0} \neq c \in \mathbb{R}^d$ such that $\mathcal{F}$ is contained within the maximum level set of c over $\mathcal{A}$, and such that c is not constant over $\mathcal{A}$.

Proof. After a coordinate transform, we may assume that $\mathcal{A}$ is full-dimensional. Since $\mathcal{F}$ is a face, it does not contain a relative interior point of $\mathcal{A}$. So $\mathcal{F}$ is a subset of the relative boundary. Let $x \in \mathcal{F}$ be a relative interior point of $\mathcal{F}$. There exists a supporting hyperplane of $\mathcal{A}$ at x, therefore there exists $c \in \mathbb{R}^d$ that takes on a maximum at x. But then it attains this maximum everywhere on $\mathcal{F}$ since otherwise we had a contradiction. □

Lemma 2.19.
If $\mathcal{A}$ is closed and convex, then its faces are closed.

Proof. Suppose that $\mathcal{F}$ is a face of $\mathcal{A}$. Then there exists a hyperplane such that $\mathcal{G} = \mathcal{A} \cap \mathcal{H}$ is a proper closed face of $\mathcal{A}$ and $\mathcal{F} \subseteq \mathcal{G}$, using Lemma 2.18. If $\mathcal{F} = \mathcal{G}$, then there is nothing more

to show. Otherwise, $\dim \mathcal{G} < \dim \mathcal{A}$, and we may repeat the argument until $\dim \mathcal{F} = \dim \mathcal{G}$, in which case $\mathcal{F} = \emptyset$ or $\mathcal{F} = \mathcal{G}$. □

Theorem 2.20 (Krein-Milman, [18, Theorem VIII.4.4])**.**
Let $\mathcal{A} \neq \emptyset$ be compact and convex. Then $\operatorname{ex} \mathcal{A} \neq \emptyset$ and $\mathcal{A} = \operatorname{convex} \operatorname{ex} \mathcal{A}$.

Proof. Since $\mathcal{A}$ is closed, the faces of $\mathcal{A}$ are closed. Let all non-empty faces be partially ordered by inclusion. By Lemma 2.17, or by Zorn's lemma, there does exists a minimal non-empty face $\mathcal{F}$, not containing a proper subface.

Suppose that $\mathcal{F}$ is not a singleton. For $x, y \in \mathcal{F}$ distinct, there exists $c \in \mathbb{R}^d$ such that $\langle c, x\rangle < \langle c, y\rangle$. Let $\gamma \in \mathbb{R}$ be the maximum of c over $\mathcal{F}$, which exists because $\mathcal{F}$ is compact. Let $\mathcal{F}^+$ be the intersection of $\mathcal{F}$ and the γ level set of c. Then $\mathcal{F}^+$ is a proper subface of $\mathcal{F}$, which is a contradiction. So a minimal face $\mathcal{F}$ must be a singleton.

Let $\mathcal{B} = \operatorname{convex} \operatorname{ex} \mathcal{A}$. Assume there exists $z \in \mathcal{A} \setminus \mathcal{B}$. By the separation theorem there exists $c \in \mathbb{R}^d$ such that $\langle c, z\rangle > \langle c, \mathcal{B}\rangle$. Let $\mathcal{F}$ be the maximizing face of c in $\mathcal{A}$. Then $\operatorname{ex} \mathcal{F} \neq \emptyset$ and $\operatorname{ex} \mathcal{F} \subseteq \operatorname{ex} \mathcal{A}$, but $\operatorname{ex} \mathcal{F} \cap \mathcal{B} = \emptyset$, which is a contradiction. Thus we have $\mathcal{A} = \mathcal{B}$. □

2.4 Caratheodory-Steinitz-type Results

We provide proofs of two Caratheodory-Steinitz-type results. These results state that, if a vector $x \in \mathbb{R}^d$ lies in the convex or conal hull of a set $\mathcal{A}$ of vectors, then it can be written as the convex or conal combination, respectively, of only a few elements of $\mathcal{A}$. The first one, Lemma 2.21 states this for convex hulls, and its proof is common in literature. The second one, Lemma 2.22, is a variation for conal hulls, and the author is not aware of a direct proof in literature. There are only minor differences in the respective proofs.

Lemma 2.21 (Convex hull version)**.**
Let $\mathcal{A}$ be a subset of $\mathbb{R}^d$. Then any $x \in \operatorname{convex} \mathcal{A}$ is the convex combination of at most $d+1$ points of $\mathcal{A}$.

Proof. Assume $\mathcal{A} \neq \emptyset$ and let $x \in \mathcal{A}$ and assume that

$$x = \sum_{i=0}^{k} \lambda_i a_i, \quad \lambda_i > 0, \quad \sum_{i=0}^{k} \lambda_i = 1, \quad a_i \in \mathcal{A}.$$

If $k > d$, then the vectors $a_1 - a_0, \ldots, a_k - a_0$ are linearly dependent. Then there exist $\mu_1, \ldots, \mu_k \in \mathbb{R}$, not all 0, such that, when writing $\mu_0 = -\mu_1 - \cdots - \mu_k$, we have

$$\sum_{i=1}^{k} \mu_i (a_i - a_0) = \vec{0}, \quad \sum_{j=0}^{k} \mu_j = 0, \quad \sum_{j=0}^{k} \mu_j a_j = \vec{0}.$$

Because not all μ_j are 0, there exist $\mu_j > 0$. Choose i indices in $\{0, \ldots, k\}$ with $\mu_i > 0$ such that $\alpha = \lambda_i / \mu_i$ is minimal among those. Without loss of generality $i = 0$. Then we see for every $j \in [k]$ that $\lambda_j - \alpha \mu_j \geq 0$, and, in particular, $\lambda_0 - \alpha \mu_0 = 0$. So we see that

$$x = \sum_{i=0}^{k} \lambda_i a_i = \sum_{i=0}^{k} \lambda_i a_i - \alpha \sum_{j=0}^{k} \mu_j a_j = \sum_{j=0}^{k} (\lambda_j - \alpha \mu_j) a_j = \sum_{j=1}^{k} (\lambda_j - \alpha \mu_j) a_j.$$

The coefficients $(\lambda_j - \alpha \mu_j)$ are non-negative. Furthermore, their sum is 1. This process can be repeated, until we have constructed the desired affinely independent points. There are at most $d+1$ of them. □

Lemma 2.22 (Conical hull version)**.**
Let $\mathcal{A}$ be a subset of $\mathbb{R}^d$. Then any $x \in \operatorname{cone} \mathcal{A}$ is the conical combination of at most d linearly independent vectors of $\mathcal{A}$.

Proof. Assume $\mathcal{A} \neq \emptyset$ and let $x \in \mathcal{A}$ and assume that

$$x = \sum_{i=1}^{k} \lambda_i a_i, \quad \lambda_i > 0, \quad a_i \in \mathcal{A}.$$

Assume that the vectors $a_1, \ldots, a_k$ are linearly dependent. Then there exist $\mu_1, \ldots, \mu_k \in \mathbb{R}$, not all 0, such that

$$\sum_{i=1}^{k} \mu_i a_i = \vec{0},$$

Because not all μ_i are 0, we may assume w.l.o.g. that some of these coefficients are positive. Choose $1 \leq l \leq k$ with $\mu_l > 0$ such that $\alpha = \lambda_i/\mu_i$ is minimal. W.l.o.g., $k = l$. Then we see for every $i \in [k]$ that $\lambda_i - \alpha\mu_i \geq 0$, and, in particular, $\lambda_k - \alpha\mu_k = 0$. So we see that

$$x = \sum_{i=1}^{k} \lambda_i a_i = \sum_{i=1}^{k} \lambda_i a_i - \alpha \sum_{i=1}^{k} \mu_i a_i = \sum_{i=1}^{k} (\lambda_i - \alpha\mu_i) a_i = \sum_{j=1}^{k-1} (\lambda_j - \alpha\mu_j) a_j.$$

The coefficients $(\lambda_j - \alpha\mu_j)$ are non-negative. This process can be repeated, until we have constructed the desired linearly independent vectors. There are at most d of them. □

3 Basics of Polyhedrons

We introduce polyhedrons and explore their combinatorial and geometric structure. The usage of the variational theory of linear functionals forshades the investigation of linear programming, which is the point of this work. In the following subsections, we elaborate on the notion of faces of a polyhedron, inspect the classes of polyhedrons constituted by polyhedral cones and polytopes, and finally introduce minimal and maximal faces.

Let $m, d \in \mathbb{N}_0$. Let $A \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$. We call

$$[A, b] := \left\{ x \in \mathbb{R}^d \mid Ax \leq b \right\}$$

the polyhedron generated by A and b. Note that we also allow $m = 0$, i.e., the case of no constraints, in which case we define $[A, b] = \mathbb{R}^d$. Because $[A, b]$ is the intersection of convex and topologically closed sets, we immediately observe:

Lemma 3.1.
$[A, b]$ is convex and topologically closed.

Given a class of objects, it is always a good idea to study structure-preserving transformations between these objects. We do this only tentatively, in order to provide some technical tools. Let us consider linear mappings of polyhedrons.

Lemma 3.2.
Let $\mathcal{P} = [A, b] \subseteq \mathbb{R}^d$ be a polyhedron, and let $T : \mathbb{R}^d \to \mathbb{R}^l$ be a linear transform. Then $T(\mathcal{P}) \subseteq \mathbb{R}^l$ is a polyhedron.

Proof. Suppose that T is the orthogonal projection from $\mathbb{R}^d$ to $\mathbb{R}^{d-1}$ with $T(x_1, \ldots, x_d) = (x_2, \ldots, x_d)$. Without loss of generality, we assume that the first column of A has entries in $\{-1, 0, 1\}$. Let I_{-1}, I_0 and I_1 be the corresponding index sets. If $(x_1, x_2, \ldots, x_d) \in [A, b]$, then

$$\begin{aligned} a_{i2}x_2 + \cdots + a_{id}x_d \leq b_i \text{ for } i \in I_0, \\ (a_{j2} - a_{k2})x_2 + \cdots + (a_{jd} - a_{kd})x_d \leq b_j - b_k \text{ for } j \in I_1, k \in I_{-1}. \end{aligned}$$

Conversely, if $(x_2, \ldots, x_d)$ satisfies the above equations, then $\beta^+ \leq \beta^-$, where

$$\begin{aligned} \beta^+ &:= \sup_{j \in I_1} a_{j2}x_2 + \cdots + a_{jd}x_d - b_j \in [-\infty, \infty), \\ \beta^- &:= \inf_{k \in I_{-1}} b_k - a_{k2}x_2 - \cdots - a_{kd}x_d \in (-\infty, \infty]. \end{aligned}$$

We can choose $x_1 \in [\beta^+, \beta^-]$ to obtain $(x_1, \ldots, x_d) \in [A, b]$. After coordinate permutation and iteration, we conclude the statement holds for arbitrary coordinate projections $T : \mathbb{R}^d \to \mathbb{R}^l$, $l < d$. In the case of general T, we verify the following equalities:

$$\begin{aligned} T(\mathcal{P}) &= \left\{ y \in \mathbb{R}^l \mid \exists x \in \mathbb{R}^d : Tx = y,\ Ax \leq b \right\} \\ &= T_{\mathbb{R}^l}\left(\left\{ (x, y) \in \mathbb{R}^d \times \mathbb{R}^l \mid Tx = y,\ Ax \leq b \right\} \right) \\ &= T_{\mathbb{R}^l}\left(\left\{ (x, y) \in \mathbb{R}^d \times \mathbb{R}^l \mid -Tx + y \leq \vec{0},\ Tx - y \leq \vec{0},\ Ax \leq b \right\} \right), \end{aligned}$$

where $T_{\mathbb{R}^l} : \mathbb{R}^d \times \mathbb{R}^l \to \mathbb{R}^l$ is the coordinate projection. Therefore $T(\mathcal{P})$ is the orthogonal projection of a polytope. This proves the result. □

Note that the proof implicitly involves the Fourier-Motzkin-elimination, which turns out to be a statement on the orthogonal projection of polyhedra.

3.1 Faces

We specialize the general theory of faces as in Section 2.

Lemma 3.3.
Let $[A, b]$ be a polyhedron. Then for each relative boundary point $x \in \partial[A, b]$ there exists a unique non-empty maximal index set $I \subseteq [m]$ such that $A_I x = b_I$.

Proof. The set $[A, b]$ is the intersection of halfspaces

$$\mathcal{H}_i = \{ x \in \mathbb{R}^d \mid \langle a_i, x \rangle \leq b_i \}$$

for $1 \leq i \leq m$. Assume that $x \in \partial[A, b]$ is a relative boundary of $[A, b]$, but there exists $\epsilon > 0$ such that $B_\epsilon(x) \subset \mathcal{H}_i$ for $1 \leq i \leq m$. Then x is contained in a finite intersection of open sets, and thus x must be an interior point of $[A, b]$, which is a contradiction. This proves the lemma. □

Theorem 3.4.
For each non-empty proper face $\mathcal{F}$ of $[A, b]$ there does exist a unique non-empty maximal index set $I \subseteq [m]$ such that $\mathcal{F} = \{x \in [A, b] \mid A_I x = b_I\}$.

Proof. After a coordinate transform, we may assume no inequality in $Ax \leq b$ is redundant, and that no inequality is satisfied with equality. Let $\mathcal{F} \subseteq \partial[A, b]$ be a non-empty proper face.

Let $I \subseteq [m]$ be a maximal index set such that $A_I z = b_I$ for all $z \in \mathcal{F}$. Suppose that $I = \emptyset$. We let $z_1, \ldots, z_m \in \mathcal{F}$ be vectors such that $a_i^t z_i < b_i$ for $1 \leq i \leq m$, and set $z := m^{-1}(z_1 + \cdots + z_m)$. Then $z \in \mathcal{F}$ does not satisfy any equality constraint, so it must be an relative interior point of $[A, b]$. But this is a contradiction, and we conclude that $I \neq \emptyset$.

Let $\mathcal{G} := \{x \in [A, b] \mid A_I x = b_I\}$. Note that $\mathcal{G}$ is a face of $[A, b]$, and that $\mathcal{F}$ is a face of the polyhedron $\mathcal{G}$. Let $J = [m] - I$. Because I has been chosen maximal, no inequality with index J is satisfied with equality everywhere over $\mathcal{F}$. But then after another coordinate transform, we see that $\mathcal{F}$ is the face of polyhedron that does not satisfy any inequalities of the polyhedron with equality. But then $\mathcal{F}$ is not a proper face of $\mathcal{G}$, so $\mathcal{F} = \mathcal{G}$.

This proves the theorem. □

Theorem 3.5.
Let $[A, b] \subseteq \mathbb{R}^d$ be a polyhedron. For any face $\mathcal{F}$ of $[A, b]$ there exists $c \in \mathbb{R}^d$ such that c attains its maximum over $[A, b]$ precisely over $\mathcal{F}$.

Proof. Let I be the maximal index set such that $\mathcal{F} = \{x \in [A, b] \mid A_I x = b_I\}$. Then $c := \sum_{i \in I} a_i$ is the desired vector. □

Corollary 3.6.
Each polyhedron has only a finite number of faces. In particular, it has only a finite number of extremal points. □

We introduce a technical definition. Let $\langle a_i, x\rangle \leq b_i$ be an inequality of $Ax \leq b$, for $A \in \mathbb{R}^{m \times d}$, $b \in \mathbb{R}^m$ and $1 \leq i \leq m$. We call $\langle a_i, x\rangle \leq b_i$ an implicit equality, if already $\langle a_i, x\rangle = b_i$ for $x \in [A, b]$. We partition the rows of A and b into the set of implicit equalities, $A^= x \leq b$, and the other ones, $A^+ x \leq b$. We let $I^=$ and I^+ denote the respective sets of indices. Note that there exists an element $x \in [A, b]$ such that $A^= x = b$ and $A^+ x < b$. We see

Lemma 3.7 ([12, Chapter 8, Equation (8)])**.**
For a polyhedron $\mathcal{P} = [A, b]$ we have

$$\operatorname{aff} \mathcal{P} = \left\{x \in \mathbb{R}^d \mid A^= x = b\right\}.$$

Proof. The left-hand side is included in the right-hand side by definition. On the other hand, suppose that $x_0 \in \mathbb{R}^d$ with $A^= x_0 = b^=$, and let $x_1 \in \mathcal{P}$ with $A^= x_1 = b^=$ and $A^+ x_1 < b^+$. Now, either $x_0 \in \mathcal{P} \subseteq \operatorname{aff} \mathcal{P}$, or $\operatorname{aff}\{x_0, x_1\}$ contains a point $z \in \mathcal{P}$ with $z \neq x_1$. But then $x_0 \in \operatorname{aff} \mathcal{P}$. □

This allows us to characterize $\dim \mathcal{P}$ as follows:

$$\dim \mathcal{P} = \dim \operatorname{aff} \mathcal{P} = \dim \left\{x \in \mathbb{R}^d \mid A^= x = b\right\} = \dim \ker A = n - \operatorname{rng} A^=.$$

Lemma 3.8.
A polyhedron has no proper subfaces if and only if it is an affine subspace.

Proof. Affine subspaces do not have non-empty proper subfaces. Conversely, if a polyhedron $[A, b]$ has no proper subfaces, then no $x \in [A, b]$ satisfies an inequality of $A^+ x \leq b^+$ with equality. Let $j \in I^+$ and $\tilde{A}x \leq \tilde{b}$ be the subsystem of $Ax \leq b$ after removing the constraint j. Let $\tilde{x} \in [\tilde{A}, \tilde{b}] \setminus [A, b]$, so $\langle a_j, \tilde{x}\rangle > b_j$, and the boundary of the half-space $\mathcal{H}_j := \{\langle a_j, \tilde{x}\rangle \leq b_j\}$ separates $\tilde{x}$ and $[A, b]$. Let $x \in [A, b]$, then the line segment $[x, \tilde{x}]$ is contained in $[\tilde{A}, \tilde{b}]$. Then $[x, \tilde{x}] \cap \mathcal{H}_j$ is contained in $[A, b]$ and contains a point in $\partial \mathcal{H}_j$. But this is a contradiction, so $[\tilde{A}, \tilde{b}] \setminus [A, b] = \emptyset$. Hence j is a redundant constraint. A repetition of this yields that I^+ is redundant in $Ax \leq b$. But then $[A, b] = [A^=, b^=]$. □

We now turn our attention to the behaviour of faces under linear transformations.

Theorem 3.9.
Let $\mathcal{P} \subseteq \mathbb{R}^D$ and $\mathcal{S} \subseteq \mathbb{R}^d$ be polyhedra. Let $T : \mathbb{R}^D \to \mathbb{R}^d$ be a linear mapping, such that $T\mathcal{P} = \mathcal{S}$. Then for any face $\mathcal{F}$ of $\mathcal{S}$ the preimage of $\mathcal{F}$ under T in $\mathcal{P}$ is a face $\mathcal{G}$ of $\mathcal{P}$.

Proof. Let $\mathcal{F}$ be a face of $\mathcal{S}$. Then there exists a maximal subset $\mathcal{G}$ of $\mathcal{P}$ such that $T\mathcal{G} = \mathcal{F}$. Suppose that $\mathcal{G}$ is not a face of $\mathcal{P}$. Then there exist $z \in \mathcal{G}$, $x, y \in \mathcal{P}$ and $0 < \lambda < 1$ such that $\lambda x + (1-\lambda)y = z$, and w.l.o.g. we have $y \notin \mathcal{G}$. But then

$$\lambda Tx + (1-\lambda)Ty = Tz \in \mathcal{F}$$

whereas $y \notin \mathcal{F}$, since G has been chosen maximal. But this contradicts $\mathcal{F}$ being a face. □

3.2 Polyhedral Cones and Polytopes

Polyhedral cones are sets that are both a polyhedron and a cone. They have a very simple characterization.

Theorem 3.10.
Let $[A, b]$ be a non-empty polyhedral cone. Then $[A, b] = [A, \vec{0}]$.

Proof. Because $[A, b]$ is a cone, we have $\vec{0} \in [A, b]$. Therefore, all entries of b are non-negative. Suppose that for $i \in [m]$ we have $b_i > 0$. Then either $\langle a_i, x\rangle \leq 0$ for all $x \in [A, b]$, so that we may replace b_i by 0, or this exist $x \in [A, b]$ such that $\langle a_i, x\rangle = \gamma$ with $0 < \gamma \leq b_i$. But because $[A, b]$ is a cone, there exists $\tau > 0$ such that $\tau x \in [A, b]$ and $\langle a_i, \tau x\rangle = \tau\gamma > b_i$. This is a contradiction, so we may replace b_i by 0. This proves the claim. □

In fact, the conal combination of finitely many vectors is a polyhedral cone.

Lemma 3.11.
Let $\mathcal{A} = \{a_1, \ldots, a_m\} \subset \mathbb{R}^d$. Then $\operatorname{cone} \mathcal{A}$ is a polyhedral cone.

Proof. We consider the mapping

$$T : \mathbb{R}^m \to \mathbb{R}^d, \quad x \mapsto \sum_{i=1}^{m} x_i a_i.$$

and let $\mathcal{C} = \{x \in \mathbb{R}^m \mid x \geq \vec{0}\}$. Then obviously $T\mathcal{C} = \operatorname{cone} \mathcal{A}$. Since $\mathcal{C}$ is a polyhedron and linear mapping preserve polyhedrons, we conclude the desired statement. □

Another subclass of polyhedrons are polytopes. We define polytopes as compact polyhedra. Recall that polyhedra are topologically closed; by the Heine-Borel theorem [6, Chapter I.4], we already infer that a polyhedron $\mathcal{P}$ is a polytope if and only if it is a bounded set. By the Krein-Milman theorem, we find that

Corollary 3.12.
A polytope is the convex closure of finitely many points. □

The converse result is easy to prove as well.

Lemma 3.13.
The convex closure of finitely many points is a polytope.

Proof. Let $z_0, \ldots, z_t \in \mathbb{R}^d$. We consider

$$T : \mathbb{R}^t \to \mathbb{R}^d, \quad x \mapsto \sum_{i=1}^{t} x_i z_i + (1 - \sum_{i=1}^{t} x_i) z_0.$$

Let $\Delta^t = \text{convex}\{0, e_1, \ldots, e_t\}$. Then $T(\Delta^t)$ is the linear transform of a polytope. Because T is continuous and Δ^t is compact, $T(\Delta^t)$ is compact. Because Δ^t is a polyhedron and T is linear, $T(\Delta^t)$ is a polyhedron. So $T(\Delta^t)$ is a polytope by definition. □

3.3 Extremal Faces

The purpose of this subsection is to study faces of $\mathcal{P}$ that are extremal with respect to the inclusion order. For a polyhedron $\mathcal{P}$ a face $\mathcal{F}$ is called minimal if its only proper subface is the empty set, and a proper face $\mathcal{F}$ is called a facet of $\mathcal{P}$ if it is not a proper face of any proper face of $\mathcal{P}$. A minimal face that is a singleton is called a vertex of $\mathcal{P}$. We study these concepts in more detail.

We see that a face of $\mathcal{P}$ is minimal if and only if it is an affine subspace, which implies that the minimal faces of polytopes are vertices.

Lemma 3.14.
Suppose that no inequality in $A^+x \leq b^+$ is implied by the other inequalities in $Ax \leq b$. Then there is a one-to-one correspondence between the facets $\mathcal{P}$ and I^+, given by

$$\mathcal{F} = \{x \in \mathcal{P} \mid \langle a_i, x \rangle = b_i\}.$$

Proof. Let $\mathcal{F} = \{x \in \mathcal{P} \mid A_I x = b_I\}$ be a facet of $\mathcal{P}$, where $I \subseteq I^+$, and let $i \in I$. But then $\mathcal{F}' = \{x \in \mathcal{P} \mid \langle a_i, x \rangle = b_i\}$ satisfies

$$\mathcal{F}' \neq \mathcal{P}, \quad \mathcal{F} \subseteq \mathcal{F}' \subseteq \mathcal{P},$$

This implies that $I = \{i\}$. Therefore, for each facet there exists an index as desired. We furthermore show that the choice of index is unique. Let $J = I^+ - \{i\}$. Let $x_1 \in [A, b]$ with $A^+x_1 < b^+$. Because we assume irredudancy of $A^+x \leq b^+$ in $Ax \leq b$, there exists x_2 with $A^=x_2 = b^=$, $\langle a_i, x_2 \rangle > b_i$ and $A_J x_0 \leq b_J$. Then a suitable convex combination x_0 of x_1 and x_2 satisfies $A^=x_0 = b^=$, $\langle a_i, x_0 \rangle = b_i$ and $A_J x_0 < b_J$. This implies that the choice of the index is unique. □

Lemma 3.15.
Let $\mathcal{F} \subseteq \mathcal{P}$. Then $\mathcal{F}$ is a minimal face of $\mathcal{P}$ if and only if $\emptyset \neq \mathcal{F} \subseteq \mathcal{P}$ and there exists $I \subseteq [m]$ such that

$$\mathcal{F} = \left\{x \in \mathbb{R}^d \mid A_I x = b_I\right\}.$$

Proof. A face of the form $\left\{x \in \mathbb{R}^d \mid A_I x = b_I\right\}$ is an affine subspace, and thus has no facets. Suppose that $\mathcal{F}$ is a minimal face. Let $I, J \subseteq [m]$ be such that

$$\mathcal{F} = \left\{x \in \mathbb{R}^d \mid A_I x = b_I,\ A_J x \leq b_J\right\},$$

and suppose that J has minimum cardinality. Then no inequality in $A_J x \leq b_J$ is implied by $A_I x = b_I$ and the other inequalities. By Lemma 3.14, $J = \emptyset$, so that $\mathcal{F}$ has no facets. But then $\mathcal{F}$ must be a minimal face. □

This has a nice consequence for the dimensions of the faces of a polyhedron:

Lemma 3.16.
Suppose that $\mathcal{P} = [A, b]$ has no implicit inequalities. Then $\mathcal{P}$ is full-dimensional.

Proof. Suppose that $Ax \leq b$ has no implicit inequalities. Then there exists $x_0 \in \mathcal{P}$ such that $Ax_0 < b$. Suppose now that $\mathcal{P}$ is not full-dimensional. W.l.o.g., we assume $0 \in \mathcal{P}$. Let then $\mathcal{H}$ be a hyperplane with normal $\vec{n}$ such that $\mathcal{P} \subseteq \mathcal{H}$. Then there exists $\epsilon > 0$ such that $A(x_0 + \epsilon\vec{n}) < b$. This is a contradiction, so $\mathcal{P}$ is full-dimensional. □

Lemma 3.17.
Let $\mathcal{F}$ be a facet of $\mathcal{P}$. Then $\dim \mathcal{F} = \dim \mathcal{P} - 1$.

Proof. Let $\mathcal{P} = [A, b]$ and let $\mathcal{F}$ be a facet of $\mathcal{P}$. After translation, rotation, and skipping of trivial coordinate axes, we may assume that $A^{=}x \leq b^{=}$ is the empty system, so $A^{+} = A$, $b^{+} = b$ and $\mathcal{P}$ is full-dimensional. Let i be associated index to $\mathcal{F}$. Then there exists $x_0 \in \mathcal{F}$ which satisfies the i-th inquality with equality, but no other inequalities with equality. There exists $\gamma > 0$ such that $B_\gamma(x_0)$ satisfies all inequalities except the i-th, but no inequality with equality. Then there exists $\alpha > 0$ small such that $x_0 + \alpha a_i$ satisfies no inequality, so $x_0 + \alpha a_i$ is an interior point. Then there exists $\beta > 0$ such that $B_\beta(x_0 + \alpha a_i) \subset \mathcal{P}$, because $\mathcal{P}$ is full-dimensional. We conclude that for α and β small enough, $\mathcal{F}$ contains the translate of an $(d-1)$-dimensional ball, so $\dim \mathcal{F} = d - 1$. □

Corollary 3.18.
Let $\mathcal{F}$ be a face of the polyhedron $\mathcal{P}$. If $\dim \mathcal{P} - \dim \mathcal{F} > 1$, then there exists a face $\mathcal{G}$ of $\mathcal{P}$ such that $\mathcal{F} \subsetneq \mathcal{G} \subsetneq \mathcal{P}$ and $\dim \mathcal{F} < \dim \mathcal{G} < \dim \mathcal{P}$. □

In order to characterize minimal faces, we define the lineality space of $\mathcal{P}$ as

$$\operatorname{lineal} \mathcal{P} := \left\{ y \in \mathbb{R}^d \mid Ay = \vec{0} \right\} \equiv \ker A.$$

We call $\mathcal{P}$ pointed if $\operatorname{lineal} \mathcal{P} = \vec{0}$.

Lemma 3.19.
All minimal faces are translates of the lineality space of $\mathcal{P}$.

Proof. Let $\mathcal{F}$ be a minimal face of $\mathcal{P}$. Let $I \subseteq [m]$ such that $\mathcal{F} = \left\{ x \in \mathbb{R}^d \mid A_I x = b_I \right\}$. We then observe $\operatorname{lineal} \mathcal{P} = \ker A \subseteq \ker A_I$, so $\mathcal{F}$ contains a suitable translate of $\operatorname{lineal} \mathcal{P}$. On the other hand, suppose there exists $z \in \ker A_I$ and $j \in [m] - I$ such that $\langle a_j, z \rangle \neq 0$. Let $x \in \mathcal{F}$, then $x + \alpha z \in \mathcal{F}$ for $\alpha \in \mathbb{R}$. But since $\mathcal{F} \subseteq \mathcal{P}$, there exists $\alpha \in \mathbb{R}$ such that $x + \alpha z \in \mathcal{F}$ while $\langle a_j, x + \alpha z \rangle > b_j$. We conclude that $\ker A_I = \ker A$, from which the claim follows. □

Corollary 3.20.
For $\mathcal{F}$ a minimal face of $\mathcal{P}$ we have $\dim \mathcal{F} = \dim \operatorname{lineal} \mathcal{P}$. □

Corollary 3.21.
For each vertex of $\mathcal{P}$ there exists $I \subseteq [m]$ with $|I| = d$ such that $A_I x = b_I$. □

We conclude this section with further results on the structure of polyhedral cones.

Theorem 3.22.
Let $\mathcal{P}$ be a polyhedral cone spanned by the vectors $a_1, \ldots, a_m$. Then each face of $\mathcal{P}$ is spanned by a subset of these vectors.

Proof. Again, we consider the mapping

$$T : \mathbb{R}^m \to \mathbb{R}^d, \quad x \mapsto \sum_{i=1}^{m} x_i a_i.$$

and let $\mathcal{C} = \{x \in \mathbb{R}^m \mid x \geq \vec{0}\}$, so $T\mathcal{C} = \mathcal{P}$. Let $\mathcal{F}$ be a face of $\mathcal{P}$. Then we apply Theorem 3.9 to find that there exists a face $\mathcal{G}$ of $\mathcal{C}$ such that $T\mathcal{G} = \mathcal{F}$. This implies the desired result. □

4 The Fundamental Theorem of Linear Inequalities and its Consequences

The fundamental theorem of linear inequalities is a central result in the theory of linear programming. It facilitates a proof of Farkas' lemma and a decomposition result on polyhedra. A proof, which describes an application of the simplex method, can be found in [12, Theorem 7.1], but the theorem follows can be proven with our previous results.

Theorem 4.1 (Fundamental theorem of linear inequalities)**.**
Let $a_1, \ldots, a_m, b \in \mathbb{R}^d$. Then the following alternative holds: Either b is a non-negative linear combination of linearly independent vectors from $a_1, \ldots, a_m$, or there does exist a vector c whose 0-level set contains $(t-1)$ linearly independent vectors from $a_1, \ldots, a_m$ such that $\langle c, b\rangle < 0$ and $\langle c, a_1\rangle, \ldots, \langle c, a_m\rangle \geq 0$, where $t = \dim \operatorname{lin}\{a_1, \ldots, a_m, b\}$. □

Proof. Let $\mathcal{P} = \operatorname{cone}\{a_1, \ldots, a_m\}$.

By definition, $b \in \mathcal{P}$ if and only if b is the convex cone generated by the vectors $a_1, \ldots, a_m$. But by the Caratheodory-Steinitz lemma for conical hulls, Lemma 2.22, this is equivalent to b being the non-negative linear combination of at most d vectors from this set.

Now, suppose that $b \notin \mathcal{P}$, and let $\mathcal{H} = \operatorname{aff}\mathcal{P}$. If $b \notin \mathcal{H}$, then clearly the second alternative holds. If instead $b \in \mathcal{H}$ holds, then we assume without loss of generality that $\mathcal{P}$ is full-dimensional. From Lemma 3.11 and Theorem 3.10 we find that $\mathcal{P}$ is a polyhedron, say, $\mathcal{P} = [A, \vec{0}]$ with $A \in \mathbb{R}^{p\times d}$ and $p \in \mathbb{N}$. We can assume that $[A^=, \vec{0}^=]$ is an empty system, and that no inequality of $[A, \vec{0}] = [A^+, \vec{0}^+]$ is implied by the others. Since we assume that $b \notin \mathcal{P}$, there exists an inequality of $[A, \vec{0}]$, associated to some index i, such that $a_i^t b > 0$. Then we let $\mathcal{F}$ be the unique facet of $\mathcal{P}$ associated to the index i. We know that $\dim \mathcal{F} = \dim \mathcal{P} - 1$ by Corollary 3.17, and using Theorem 3.22, we see that there exists a subset $g_1, \ldots, g_k$ of vectors from $a_1, \ldots, a_m$ that conically generates $\mathcal{F}$. But we have $k \geq \dim \mathcal{F}$, and therefore we may choose $\dim \mathcal{P} - 1$ linearly independent vectors of $g_1, \ldots, g_k$. Choosing an appropriate normal of the linear hull of $\mathcal{F}$, which is a hyperplane, the statement follows.

□

An immediate consequence is a structural observation on polyhedral cones.

Lemma 4.2.
The cone $\mathcal{C} = \operatorname{cone}\{a_1, \ldots, a_p\} \subseteq \mathbb{R}^d$ is the intersection of the halfspaces $\mathcal{H} \subset \operatorname{lin}\{a_1, \ldots, a_p\}$ that are generated by $\dim \operatorname{lin}\{a_1, \ldots, a_p\} - 1$ linearly independent vectors of $a_1, \ldots, a_p$ and that do not separate $\mathcal{C}$.

Proof. After a coordinate transform we may assume w.l.o.g. that $\mathbb{R}^d = \operatorname{lin}\{a_1, \ldots, a_p\}$. For any $b \in \mathbb{R}^d$ we have that $b \notin \mathcal{C}$ if and only if there does exist a hyperplane $\mathcal{H}$ that is spanned by $d-1$ linearly independent vectors of $a_1, \ldots, a_p$ and that separates b and $\mathcal{C}$. This implies that $z \in \mathcal{C}$ if and only if for all hyperplanes $\mathcal{H}$ that are spanned by $d-1$ linearly independent

vectors of $a_1, \ldots, a_p$ that do not separate $\mathcal{C}$, we have that z lies on the same side of $\mathcal{H}$ as $\mathcal{C}$. This proves the claim. □

Farkas' lemma follows immediately from the fundamental theorem of linear inequalities. It can also be proven by the Hahn-Banach separation theorem.

Lemma 4.3 (Farkas' lemma)**.**
Let $A \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$. Then there does exist $x \in \mathbb{R}^d$ with $x \geq \vec{0}$ and $Ax = b$ if and only if for all $y \in \mathbb{R}^m$ with $A^t y \geq \vec{0}$ we have $\langle y, b \rangle \geq 0$.

Proof. It is obvious that the first condition implies the second condition. We prove the converse direction. Let $a_1, \ldots, a_d$ be the columns of A. Suppose that there does not exist $x \in \mathbb{R}^d$ with $x \geq \vec{0}$ and $Ax = b$. Then $b \notin \text{cone}\{a_1, \ldots, a_d\}$. By the fundamental theorem of linear inequalities (or the Hahn-Banach separation theorem), there does exist $y \in \mathbb{R}^m$ with $y^t A \geq \vec{0}$ such that $\langle y, b \rangle < 0$. □

Corollary 4.4 (Farkas' lemma, variant)**.**
Let $A \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$. Then there does exist $x \in \mathbb{R}^d$ with $x \geq \vec{0}$ and $Ax \leq b$ if and only if for all $y \in \mathbb{R}^m$ with $y \geq \vec{0}$ and $A^t y \geq \vec{0}$ we have $\langle y, b \rangle \geq 0$.

Proof. There exists $x \geq \vec{0}$ with $Ax \leq b$ if and only if there does exist $(x_1, x_2) \in \mathbb{R}^{m+d}$ with $x_1 \geq \vec{0}$, $x_2 \geq \vec{0}$ and $x_1 + Ax_2 = b$. By Farkas' lemma, (x_1, x_2) do exist if and only if for all $y \in \mathbb{R}^m$ with $y \geq \vec{0}$ and $A^t y \geq \vec{0}$ we have $\langle y, b \rangle \geq \vec{0}$. □

4.1 Decomposition of Polyhedra

Another important application of the fundamental theorem of linear inequalities is the characterization of conical and general polyhedral sets.

Lemma 4.5.
Finitely generated cones are polyhedral.

Proof. Let $a_1, \ldots, a_p \in \mathbb{R}^d$. Let $\mathcal{C} = \text{cone}\{a_1, \ldots, a_p\}$. By Lemma 4.2 we conclude $\mathcal{C}$ is the intersection of finitely many halfspaces whose boundaries are spanned by subsets of $\{a_1, \ldots, a_p\}$. □

Theorem 4.6.
Let $a_1, \ldots, a_p$ and $b_1, \ldots, b_q$ be column vectors of matrices $A \in \mathbb{R}^{m \times p}$ and $B \in \mathbb{R}^{m \times q}$, respectively. Then we have

$$\text{cone}\{a_1, \ldots, a_p\} = [B^t, \vec{0}] \quad \Longleftrightarrow \quad \text{cone}\{b_1, \ldots, b_q\} = [A^t, \vec{0}].$$

Proof. Suppose that the left-hand statement holds. We have to show the right-hand statement. For $i \in [p]$ and $j \in [q]$ we have $\langle a_i, b_j \rangle \leq 0$, so $b_1, \ldots, b_q \in [A^t, \vec{0}]$. Suppose that $y \in [A^t, \vec{0}]$ but $y \notin \text{cone}\{b_1, \ldots, b_q\}$. We know that the latter cone is polyhedral by Lemma 4.5, thus there exists $w \in [B^t, \vec{0}]$ with $\langle w, y \rangle > 0$. Then $w \in \text{cone}\{a_1, \ldots, a_p\}$, but also $\langle w, c \rangle \leq 0$ for all $c \in [B^t, \vec{0}]$, such that in particular $\langle w, a_i \rangle \leq 0$ for $i \in [p]$. But then $w \in [A^t, \vec{0}]$ and $\langle w, y \rangle > 0$, which is a contradiction. By symmetry we obtain the desired result. □

Corollary 4.7.
A cone is polyhedral if and only if it is finitely generated.

Proof. We already know that finitely generated cones are polyhedral. Let $[A, \vec{0}]$ with $A \in \mathbb{R}^{m\times p}$ be a polyhedral cone. Then $\operatorname{cone} A$ is finitely generated, and therefore it is polyhedral, say $\operatorname{cone} A = [B, \vec{0}]$ with $B \in \mathbb{R}^{m\times q}$. But then $\operatorname{cone} B = [A, \vec{0}]$, which proves the corollary. □

Theorem 4.8.
A set $\mathcal{P} \subseteq \mathbb{R}^d$ is a polyhedron if and only if there exist a polytope $\mathcal{Q} \subseteq \mathbb{R}^d$ and a polyhedral cone $\mathcal{C} \subseteq \mathbb{R}^d$ such that $\mathcal{P} = \mathcal{Q} + \mathcal{C}$.

Proof. Suppose that $\mathcal{Q} = \operatorname{convex}\{x_1, \ldots, x_p\}$ is a polytope and $\mathcal{C} = \operatorname{cone}\{y_1, \ldots, y_q\}$ is a cone. Then $x_0 \in \mathcal{Q} + \mathcal{C}$ if and only if

$$\begin{pmatrix} x_0 \\ 1 \end{pmatrix} \in \operatorname{cone}\left\{ \begin{pmatrix} x_1 \\ 1 \end{pmatrix}, \ldots, \begin{pmatrix} x_p \\ 1 \end{pmatrix}, \begin{pmatrix} y_1 \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} y_q \\ 0 \end{pmatrix}, \right\}.$$

Since this cone is finitely generated, there exist $m \geq 0$, $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$ such that this cone equals $[(A, b), \vec{0}]$. But $(x_0, 1) \in [(A, b), \vec{0}]$ is equivalent to $Ax \leq -b$.

Conversely, we have

$$x \in [A, b] \quad \Longleftrightarrow \quad \begin{pmatrix} x \\ 1 \end{pmatrix} \in \left\{ \begin{pmatrix} x \\ \lambda \end{pmatrix} \middle| \begin{array}{c} x \in \mathbb{R}^d, \\ \lambda \in \mathbb{R}_0^+, \\ Ax - \lambda b \leq \vec{0} \end{array} \right\}.$$

The latter set is a polyhedral cone, and therefore it is finitely generated. We conclude that there exist $x_1, \ldots, x_t \in \mathbb{R}^d$ and $\lambda_1, \ldots, \lambda_t \in \mathbb{R}$ such that

$$\left\{ \begin{pmatrix} x \\ \lambda \end{pmatrix} \middle| \begin{array}{c} x \in \mathbb{R}^d, \\ \lambda \in \mathbb{R}_0^+, \\ Ax - \lambda b \leq \vec{0} \end{array} \right\} = \operatorname{cone}\left\{ \begin{pmatrix} x_1 \\ \lambda_1 \end{pmatrix}, \ldots, \begin{pmatrix} x_t \\ \lambda_t \end{pmatrix} \right\}.$$

After rescaling, we may assume that the λ_i are either 0 or 1. It follows that we can choose $\mathcal{Q}$ to be the polytope generated by those x_i with $\lambda_i = 1$, and $\mathcal{C}$ to be the cone with indices generated by those x_i with $\lambda_i = 0$. □

It is a startling fact that this decomposition is almost unique: the polyhedral cone in the decomposition is the same in each decomposition of a fixed polyhedron $\mathcal{P}$. We introduce the recession cone

$$\text{rec. cone}\, \mathcal{P} := \left\{ y \in \mathbb{R}^d \mid \forall x \in \mathcal{P} : x + y \in \mathcal{P} \right\}.$$

Lemma 4.9.
If $\mathcal{P} = [A, b]$, then

$$\text{rec. cone}\, \mathcal{P} = \left\{ y \in \mathbb{R}^d \mid Ay \leq \vec{0} \right\}.$$

Proof. Clearly, if $y \in \mathbb{R}^d$ with $Ay \leq 0$, then y is a member of the left-hand side. Conversely, if y satisfies $x + y \in \mathcal{P}$ for all $x \in \mathcal{P}$ but has $\langle a_i, y\rangle = \gamma$ for some $i \in [n]$ and $\gamma \in \mathbb{R}$, $\gamma > 0$, then we may choose $k \in \mathbb{N}$ and $x \in \mathcal{P}$ such that $x + ky \in \mathcal{P}$ while $\langle a_i, x + ky\rangle > b_i$. This is a contradiction, and so the desired claim follows. □

Corollary 4.10.
If $\mathcal{C}$ is a polyhedral cone and $\mathcal{Q}$ is a polytope such that $\mathcal{P} = \mathcal{Q} + \mathcal{C}$, then $\mathcal{C} = \text{rec. cone}\, A$.

Proof. Clearly, $\mathcal{C} \subseteq \text{rec. cone}\,\mathcal{P}$. Conversely, suppose that $z \in (\text{rec. cone}\,\mathcal{P}) \setminus \mathcal{C}$. There exists $\alpha \in \mathbb{R}_0^+$ such that $\alpha z + \mathcal{Q}$ has positive distance from $\mathcal{C} + \mathcal{Q} = \mathcal{P}$. But $\alpha z + \mathcal{Q} \subset \mathcal{Q}$, since $\alpha z \in \text{rec. cone}\,\mathcal{P}$. We conclude that $\mathcal{C} = \text{rec. cone}\,\mathcal{P}$. □

The former corollary is useful for proving Theorem 5.2 below. Note furthermore that $\text{lineal}\,\mathcal{P} = (\text{rec. cone}\,\mathcal{P}) \cap (-\text{rec. cone}\,\mathcal{P})$.

4.2 Polarity theory

We introduce the polar polyhedron and some related results. Properties of a polyhedron directly correspond to properties its polar polyhedron and vice versa, so switching from the primal to the polar polyhedron can be useful.

Let $\mathcal{P} \subseteq \mathbb{R}^d$ be a polyhedron. The polar set of $\mathcal{P}$ is

$$\mathcal{P}^* := \left\{ w \in \mathbb{R}^d \mid \forall x \in \mathcal{P} : \langle w, x \rangle \leq 1 \right\}.$$

Suppose that we have the decomposition

$$\mathcal{P} = \text{convex}\left\{ \vec{0}, x_1, \ldots, x_p \right\} + \text{cone}\left\{ y_1, \ldots, y_q \right\}. \tag{1}$$

Then it is easy to see that

$$\begin{aligned} \mathcal{P}^* &= \left\{ w \in \mathbb{R}^d \mid \forall 1 \leq j \leq q : \langle w, y_j \rangle \leq 0 \text{ and } \forall x \in \text{convex}\left\{ \vec{0}, x_1, \ldots, w_p \right\} : \langle w, x \rangle \leq 1 \right\} \\ &= \left\{ w \in \mathbb{R}^d \mid \forall 1 \leq j \leq q : \langle w, y_j \rangle \leq 0 \text{ and } \forall 1 \leq i \leq p : \langle w, x_i \rangle \leq 1 \right\}. \end{aligned}$$

We conclude that $\mathcal{P}^*$ is polyhedron again. In fact, polarization is self-inverse under mild conditions on the polyhedron $\mathcal{P}$.

Theorem 4.11.
Assume that $0 \in \mathcal{P}$. Then $\mathcal{P}^{**} = \mathcal{P}$.

Proof. We first verify that $P \subseteq \mathcal{P}^{**}$ from definitions. Conversely, if $x \in \mathcal{P}^{**}$, suppose that $x \notin \mathcal{P}$. Let $\langle a_i, x \rangle \leq b_i$ be an inequality of $\mathcal{P}$ not satisfied by x. We have $\vec{0} \in \mathcal{P}$, so $b_i \geq 0$. If $b_i = 0$, then $\mathbb{R}_0^+ a_i \subseteq \mathcal{P}^*$; but then $\langle a_i, x \rangle > 0$, and so for some $\lambda > 0$ we have $\lambda \langle a_i, x \rangle > 1$. This is a contradiction. If instead $b_i > 0$, then $b_i^{-1} a_i \in \mathcal{P}^*$; thus $x \notin \mathcal{P}^{**}$, which is a contradiction, too. This proves that $x \in \mathcal{P}$. □

Theorem 4.12.
The decomposition

$$\mathcal{P} = \text{convex}\left\{ \vec{0}, x_1, \ldots, w_p \right\} + \text{cone}\left\{ y_1, \ldots, y_q \right\}.$$

holds if and only if

$$\mathcal{P}^* = \left\{ w \in \mathbb{R}^d \mid \forall 1 \leq j \leq q : \langle w, y_j \rangle \leq 0 \text{ and } \forall 1 \leq i \leq p : \langle w, x_i \rangle \leq 1 \right\}.$$

Proof. The first implication has been outlined above. For the proof of the other direction, set

$$\mathcal{Q} := \text{convex}\left\{ \vec{0}, x_1, \ldots, w_p \right\} + \text{cone}\left\{ y_1, \ldots, y_q \right\}.$$

Then $\mathcal{Q}^* = \mathcal{P}^*$ and $\vec{0} \in \mathcal{Q}$, so $\mathcal{Q} = \mathcal{Q}^{**} = \mathcal{P}^{**} = \mathcal{P}$. □

Having clarified the relation between the polyhedron and the polar polyhedron, it is now possible to infer properties of one of this polyhedrons from properties of the other.

Lemma 4.13.
We have $\dim \mathcal{P} = k$ if and only if $\operatorname{lineal} \mathcal{P}^* = d - k$, provided that $\vec{0} \in \mathcal{P}$.

Proof. We can write $\mathcal{P}$ as in (1). Let $Y \in \mathbb{R}^{(p+q)\times d}$ be the matrix with rows $x_1^t, \ldots, x_p^t, y_1^t, \ldots, y_q^t$. Then

$$d - \operatorname{lineal} \mathcal{P}^* = d - \dim \ker Y = \operatorname{rng} Y.$$

But $\dim \mathcal{P} = \dim \operatorname{aff} \mathcal{P} = \dim \operatorname{lin} \mathcal{P}$, since $\vec{0} \in \mathcal{P}$. So $\operatorname{rng} Y = \dim \mathcal{P}$. □

Lemma 4.14.
Assume that $\vec{0} \in \mathcal{P}$. Then $\vec{0}$ is an internal point of $\mathcal{P}$ if and only if $\mathcal{P}^*$ is bounded.

Proof. We immediately see that $\vec{0}$ is an internal point of $\mathcal{P}$ if and only if there exist $a_1, \ldots, a_p$, $p \in \mathbb{N}_0$, such that $\mathcal{P} = \left\{x \in \mathbb{R}^d \mid \langle a_i, x\rangle \leq 1\right\}$. But this is the case if and only if $\mathcal{P}^* = \operatorname{convex}\left\{\vec{0}, a_1, \ldots, a_p\right\}$, which is equivalent to $\mathcal{P}^*$ being bounded, due to Theorem 4.8. □

5 Linear Programming

The theory of extremal values of linear functionals on polyhedrons appears in two flavors in this thesis. On the one hand, as a tool for polyhedral theory, as serves as a tool in Sections 3 and 4. On the other hand, the maximizations of linear functionals over polyhedrons is central in optimization theory. We therefore elaborate this perspective in the present section.

Let $A \in \mathbb{R}^{n\times d}$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}^d$. A linear programming problem (A, b, c) is the problem of solving

$$\text{Maximize} \quad \langle c, x\rangle \quad \text{subject to} \quad Ax \leq b. \tag{2}$$

We call the problem feasible, if $[A, b]$ is non-empty, and infeasible otherwise. If $[A, b] \neq \emptyset$, then we call any $z \in [A, b]$ a (feasible) solution. We call the problem unbounded, if $\langle c, \cdot\rangle$ does not have a supremum over $[A, b]$, otherwise we call it bounded. If it is bounded, then call x an optimal solution if it maximizes $\langle c, \cdot\rangle$ over $[A, b]$.

It is not immediately clear that such an optimal solution exists, either because $[A, b]$ is empty, or because $\langle c, \cdot\rangle$ has no supremum over $[A, b]$, or because an existing supremum is not realized over $[A, b]$. The last possibility can be ruled out.

Lemma 5.1.
If (A, b, c) is bounded, then the supremum is attained, i.e., $\langle c, \cdot\rangle$ has a maximum over $[A, b]$.

Proof. Let $\mathcal{C}$ be a polyhedral cone, generated by vectors $c_1, \ldots, c_m \in \mathbb{R}^d$, and let $\mathcal{D}$ be a polytope, such that $[A, b] = \mathcal{C} + \mathcal{D}$. Because $\langle c, \cdot\rangle$ is continuous and $\mathcal{D}$ is compact, the functional attains a maximum over $\mathcal{D}$, say, at point $x_{\mathcal{D}} \in \mathcal{D}$. If $\mathcal{C} = \emptyset$, then there is nothing to show. So we assume that $\mathcal{C}$ is non-empty. Because c is bounded from above over $[A, b]$, it must be bounded from above over $\mathcal{C}$. This necessitates that $\langle c, c_i\rangle \leq 0$ for $1 \leq i \leq m$. But then the supremum of c over $\mathcal{C}$ is attained at 0. On the other hand, $x_{\mathcal{D}} \in [A, b]$, and for $y \in \mathcal{C}$, $z \in \mathcal{D}$ we have

$$\langle c, x_{\mathcal{D}}\rangle \geq \langle c, z\rangle \geq \langle c, z + y\rangle$$

This proves the claim. □

Therefore, the three most constitutive questions in linear programming theory are whether the problem is feasible, whether the problem is bounded, and whether we can find an optimal solution. For usage in the sequel, we adopt the terminology to call two linear programming problems (A_1, b_1, z_1) and (A_2, b_2, z_2) equivalent if either both are infeasible, both are unbounded, or both are bounded with exactly the same set of optimal solutions. Surprisingly, the boundedness of a linear programming problem essentially depends on the system matrix A and the direction z, but not on the right-hand side b.

Theorem 5.2.
Let $A \in \mathbb{R}^{n \times d}$, $z \in \mathbb{R}^d$ and $b_1, b_2 \in \mathbb{R}^n$. If $[A, b_1]$ and $[A, b_2]$ are both non-empty, then (A, b_1, z) is bounded if and only if (A, b_2, z) is bounded.

Proof. Let $[A, b_1] = \mathcal{Q}_1 + \mathcal{C}_1$ and $[A, b_2] = \mathcal{Q}_2 + \mathcal{C}_2$, where $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are polytopes and $\mathcal{C}_1$ and $\mathcal{C}_2$ are polyhedral cones, by Theorem 4.8. But by Corollary 4.10, we have $\mathcal{C}_1 = \mathcal{C}_2 = \text{rec. cone } A$. Thus, z is bounded on $[A, b_1]$ or $[A, b_2]$ if and only if it is bounded on rec. cone A. This had to be shown. □

The notion of dual linear programming problem is vital for the overall understanding of linear programming theory. Let (A, b, c) be a linear programming problem. Then the dual linear programming problem $(A, b, c)^*$ is given by

$$\text{Minimize} \quad \langle b, y \rangle \quad \text{subject to} \quad A^t y = c,\ y \geq \vec{0}. \tag{3}$$

This problem can be formulated as a linear programming problem again. To see this, we note that

$$\begin{array}{c} y^t A = c^t, \\ y \geq \vec{0} \end{array} \quad \Longleftrightarrow \quad \begin{array}{c} A^t y = c, \\ y \geq \vec{0} \end{array} \quad \Longleftrightarrow \quad \begin{pmatrix} A^t \\ -A^t \\ -I \end{pmatrix} y \leq \begin{pmatrix} c \\ -c \\ \vec{0} \end{pmatrix}.$$

Hence (3) is equivalent to

$$\text{Maximize} \quad -\langle b, y \rangle \quad \text{subject to} \quad \begin{pmatrix} A^t \\ -A^t \\ -I \end{pmatrix} y \leq \begin{pmatrix} c \\ -c \\ \vec{0} \end{pmatrix}.$$

Dualization is "self-inverse", i.e., the primal is the dual of the dual of the primal. To see this, we note that $(A, b, c)^{**}$ equals

$$\text{Find } z \in \operatorname{argmin} \left\{ (-c, c, \vec{0})^t (z_1, z_2, z_3) \mid (A, -A, -I)(z_1, z_2, z_3)^t = (-b),\ (z_1, z_2, z_3)^t \geq \vec{0} \right\}.$$

Then (z_1, z_2, z_3) optimally solves this problem if and only if $(z, w) = (z_1 - z_2, z_3)$ optimally solves

$$\text{Find } (z, w) \in \operatorname{argmin} \left\{ (-c, \vec{0})^t (z, w) \mid (A, -I)(z, w)^t = (-b),\ w \geq \vec{0} \right\},$$

which in turn is equivalent to z being an optimal solution of

$$\text{Find } z \in \operatorname{argmin} \left\{ -\langle c, z \rangle \mid z \in \mathbb{R}^d,\ Az \leq b \right\}.$$

But this problem is equivalent to (2).

Lemma 5.3.
Let x be a feasible solution of (A, b, c), and let y be a feasible solution of $(A, b, c)^*$. Then we have $\langle c, x\rangle \leq \langle y, b\rangle$.

Proof. We verify $\langle c, x\rangle = \langle A^t y, x\rangle = \langle y, Ax\rangle \geq \langle y, b\rangle$. □

Corollary 5.4.
If a linear programming problem is unbounded, then its dual is infeasible. If the dual of a linear programming problem is feasible, then it is bounded.

Theorem 5.5.
Let (A, b, c) and $(A, b, c)^*$ be feasible. Then both are bounded, and their respective optimal solutions x and y satisfy $\langle c, x\rangle = \langle y, b\rangle$.

Proof. This is Theorem 2.7 of [11]. From Lemma 5.3 we know that both linear programming problems are bounded and that their respective optimal solutions x' and y' satisfy $\langle c, x'\rangle \leq \langle b, y'\rangle$. It suffices to prove that $\langle c, x\rangle \geq \langle b, y\rangle$ for some feasible solutions x of (A, b, c) and y of $(A, b, c)^*$, respectively. But this means that (x, y) satisfies the linear inequalities

$$Ax \leq b, \quad \langle -c, x\rangle + \langle b, y\rangle \leq 0, \quad A^t y \leq c, \quad -A^t y \leq -c, \quad -y \leq \vec{0}.$$

We know by Farkas' Lemma that such a (x, y) exists if and only if

$$\forall \begin{pmatrix} u \\ v \\ w \\ z \end{pmatrix} \geq \begin{pmatrix} \vec{0} \\ \vec{0} \\ \vec{0} \\ \vec{0} \end{pmatrix}, \beta \in \mathbb{R}_0^+ : \left(\begin{array}{cc} Au - \beta c = \vec{0} & \wedge \\ \beta b + Av - Aw - z = \vec{0} & \end{array} \implies \langle u, b\rangle + \langle v - w, c\rangle \geq 0 \right).$$

We eliminate the variable z and find the equivalent statement

$$\forall \begin{pmatrix} u \\ v \\ w \end{pmatrix} \geq \begin{pmatrix} \vec{0} \\ \vec{0} \\ \vec{0} \end{pmatrix}, \beta \in \mathbb{R}_0^+ : \left(\begin{array}{cc} Au - \beta c = \vec{0} & \wedge \\ \beta b + Av - Aw \geq \vec{0} & \end{array} \implies \langle u, b\rangle + \langle v - w, c\rangle \geq 0 \right).$$

Let $u \geq 0, v \geq 0, w \geq 0$. In the case that $\beta > 0$ we find

$$\langle u, b\rangle = \beta^{-1}\beta\langle b, u\rangle \geq \beta^{-1}\langle A(w - v), u\rangle = \beta^{-1}\beta\langle w - v, c\rangle = \langle w - v, c\rangle.$$

If instead $\beta = 0$, then for any pair (x, y) of feasible solutions of (A, b, c) and $(A, b, c)^*$ we have

$$\langle u, b\rangle \geq \langle u, Ax\rangle = 0 \geq \langle A(w - v), y\rangle = \langle w - v, c\rangle.$$

This means that Farkas' lemma can be applied, and so the desired result follows. □

Lemma 5.6.
Let (A, b, c) be a linear program. Let x and y be solutions to (A, b, c) and $(A, b, c)^*$. Then

$$x \text{ and } y \text{ are optimal solutions} \quad \Leftrightarrow \quad \langle c, x\rangle = \langle y, b\rangle \quad \Leftrightarrow \quad \langle y, b - Ax\rangle = 0.$$

Proof. The first equivalence follows from the strong duality theorem (Theorem 5.5). For the second equivalence we verify

$$\langle y, b - Ax\rangle = \langle y, b\rangle - \langle y, Ax\rangle = \langle y, b\rangle - \langle c, x\rangle = 0$$

from the definition of the dual linear programming problem. □

6 Elements of Stochastic Analysis and Randomized Algorithms

Within this section we assemble a number of results from probability theory for later reference. The focus is on Gaussian vectors in $\mathbb{R}^d$, but the uniform distribution on unit spheres and the normalized Haar measure on the orthogonal group are briefly examined, too. For this section, and similar for the whole of the present work, we freely employ basic notions of probability and its canonical foundation in measure theory, but at no point will an immersion into the foundational and technical details be necessary. The interest reader is recommended [3, Kapitel I-V] for further background.

Let $\mathcal{X}$ denote the set of Lebesgue-measurable subsets of $\mathbb{R}^d$. A probability measure or probability distribution is a measure defined over $\mathcal{X}$ under which $\mathbb{R}^d$ evaluates to 1. A random variable x in $\mathbb{R}^d$ is just a probability measure on $\mathbb{R}^d$ which we think of as a "random object in $\mathbb{R}^d$" rather than measure. The measurable function $\mu_x : \mathbb{R}^d \to \mathbb{R}$ is called a density function of x if

$$x(\mathcal{A}) = \int_{\mathcal{A}} \mu_x(a)\mathrm{d}a, \quad \mathcal{A} \in \mathcal{X}.$$

Let $F : \mathbb{R}^d \to \mathbb{R}^d$ be a diffeomorphism. Then we define

$$(Fx)(\mathcal{A}) := x(F^{-1}\mathcal{A}), \quad \mathcal{A} \in \mathcal{X}.$$

In particular, the law of substitution [3, Kapital V.4] implies for a density function that

$$x(F^{-1}\mathcal{A}) = \int_{F^{-1}\mathcal{A}} \mu_x(a)\mathrm{d}a = \int_{\mathcal{A}} \mu_x(F^{-1}(a)) \cdot |\det DF^{-1}(a)|\mathrm{d}a, \quad \mathcal{A} \in \mathcal{X}.$$

Let x and y be random variables in $\mathbb{R}^d$ with continuous density functions μ_x and μ_y. Then $x + y$ is a random variable with density function

$$\mu_{x+y}(\circ) = \int_{\mathbb{R}^d} \mu_x(t)\mu_y(\circ - t)\mathrm{d}t.$$

The most important type of random variables in this thesis are Gaussian vectors in $\mathbb{R}^d$. These random variables are defined only by two parameters and their analytical properties are comparatively easy to assess. An $(\bar{x}, \sigma)$-Gaussian vector x is a random variable in $\mathbb{R}^d$ that has a density function

$$\mu_x(\circ) = \left(\sqrt{2\pi}\sigma\right)^{-d} \exp\left(-\frac{1}{2}\sigma^{-2}\| \circ - \bar{x}\|^2\right). \tag{4}$$

Note that a Gaussian vector can be seen as a vector whose entries are Gaussian variables on $\mathbb{R}$. Due to the exponential decay of μ_x and due to symmetry arguments, we can directly infer that $\mathbb{E}_x(x)$ exists and equals $\bar{x}$. We call $\bar{x}$ the mean value of x, and call σ the standard deviation of x. For an $y \in \mathbb{R}^d$, the random variable $x + y \equiv y + x$ has density

$$\mu_{x+y}(\circ) = \left(\sqrt{2\pi}\sigma\right)^{-d} \exp\left(-\frac{1}{2}\sigma^{-2}\| \circ - (\bar{x} + y)\|^2\right),$$

so $x + y$ is a $(\bar{x} + y, \sigma)$-Gaussian. Let $\alpha \in \mathbb{R} \setminus \{0\}$, then αx has density

$$\mu_{\alpha x}(\circ) = \left(\sqrt{2\pi}\alpha\sigma\right)^{-d} \exp\left(-\frac{1}{2}\sigma^{-2}\|\alpha^{-1} \circ - \bar{x}\|^2\right) \alpha^{-d}$$

$$= \left(\sqrt{2\pi}\alpha\sigma\right)^{-d} \exp\left(-\frac{1}{2}(\alpha\sigma)^{-2}\| \circ -\alpha\bar{x}\|^2\right),$$

so αx is a $(\alpha\bar{x}, \alpha\sigma)$-Gaussian vector. Let x be a $(\bar{x}, \sigma_x)$-Gaussian and let y be a $(\bar{y}, \sigma_y)$-Gaussian. For the density function $\mu_{x+y}(\circ)$ of the sum $x + y$, we see

$$\begin{aligned}
&\mu_{x+y}(\circ) = \int_{\mathbb{R}^d} \mu_x(t)\mu_y(\circ - t)\mathrm{d}t \\
&= \int_{\mathbb{R}^d} (2\pi\sigma_x\sigma_y)^{-d} \exp\left(-\frac{1}{2\sigma_x^2}\|t - \bar{x}\|^2 - \frac{1}{2\sigma_y^2}\| \circ -t - \bar{y}\|^2\right) \mathrm{d}t \\
&= (2\pi)^{-d/2}\left(\sigma_x^2 + \sigma_y^2\right)^{-d/2} \exp\left(-\frac{1}{2}\left(\sigma_x^2 + \sigma_y^2\right)^{-1}\| \circ -\bar{x} - \bar{y}\|^2\right) \\
&\quad \cdot \int_{\mathbb{R}^d} (2\pi)^{-d/2}\left(\frac{\sigma_x^2\sigma_y^2}{\sigma_x^2 + \sigma_y^2}\right)^{-d/2} \exp\left(-\frac{\|t - \bar{x}\|^2}{2\sigma_x^2} - \frac{\| \circ -t - \bar{y}\|^2}{2\sigma_y^2} + \frac{\| \circ -\bar{x} - \bar{y}\|^2}{2(\sigma_x^2 + \sigma_y^2)}\right) \mathrm{d}t.
\end{aligned}$$

After substituting t by $t + \bar{x}$ we can derive by simple calculations that

$$\begin{aligned}
&\frac{\|t - \bar{x}\|^2}{2\sigma_x^2}\frac{\| \circ -t - \bar{y}\|^2}{2\sigma_y^2} - \frac{\| \circ -\bar{x} - \bar{y}\|^2}{2(\sigma_x^2 + \sigma_y^2)} \\
&= \frac{1}{2}\left(\frac{\sigma_x^2\sigma_y^2}{\sigma_x^2 + \sigma_y^2}\right)^{-1} \left\|t - \sigma_x^2(\sigma_x^2 + \sigma_y^2)^{-1}(\circ - \bar{x} - \bar{y})\right\|^2.
\end{aligned}$$

So the integrand of the last integral is the density of a Gaussian distribution, whose integral is 1. We conclude that $x + y$ is a $\left(\bar{x} + \bar{y}, \sqrt{\sigma_x^2 + \sigma_y^2}\right)$-Gaussian. Furthermore, the orthogonal projections of Gaussian vectors are Gaussian vectors within the target space, too.

Lemma 6.1.
Let $\mathcal{H} \subset \mathbb{R}^d$ be an affine space, and let $P_\mathcal{H} : \mathbb{R}^d \to \mathcal{H}$ be the orthogonal projection onto $\mathcal{H}$. Let x be a $(\bar{x}, \sigma)$-Gaussian vector. Then $P_\mathcal{H}x$ is a $(P_\mathcal{H}\bar{x}, \sigma)$-Gaussian vector.

Proof. Without loss of generality, we assume that $0 \in \mathcal{H}$ and that $\mathcal{H}$ is a coordinate plane, $\mathcal{H} = \mathrm{lin}\{e_1, \ldots, e_k\}$. Set $\mathcal{Z} = \mathrm{lin}\{e_{k+1}, \ldots, e_d\}$ and let $\mathcal{A} \subseteq \mathcal{H}$ be measurable in $\mathcal{H}$. Then a generalization of the law of substitution [3, Kapitel III.1, Satz 1.7] implies

$$\begin{aligned}
\mu_{P_\mathcal{H}x}(\mathcal{A}) = x(P_\mathcal{H}^{-1}\mathcal{A}) &= \int_{\mathcal{A}+\mathcal{Z}} \mu_x(\mathcal{A} + z)\mathrm{d}z \\
&= \int_{\mathcal{A}+\mathcal{Z}} \left(\sqrt{2\pi}\sigma\right)^{-d} \exp\left(-\frac{1}{2}\sigma^{-2}\|\mathcal{A} + z - \bar{x}\|^2\right) \mathrm{d}z \\
&= \left(\sqrt{2\pi}\sigma\right)^{-k} e^{-\frac{1}{2\sigma^2}\|(\mathcal{A}-\bar{x})_{1,\ldots,k}\|^2} \int_{\mathcal{A}+\mathcal{Z}} \left(\sqrt{2\pi}\sigma\right)^{-d+k} e^{-\frac{1}{2\sigma^2}\|(\mathcal{A}+z-\bar{x})_{k+1,\ldots,d}\|^2} \mathrm{d}z.
\end{aligned}$$

The last integral evaluates to 1, which completes the proof. □

Remark 6.2.
Spielman and Teng [13, Subsection 2.4] take into account an even broader class of random variables, where, speaking in loose terms, the randomness is anisotropic. However, their subsequent analysis considers only Gaussian random variables with isotropic density functions, as above (4), so this more general case remains for future study.

We derive a series of estimates on the probability that Gaussian vectors happen to be contained in certain sets. The most pivotal tool is the following lemma, which we state without proof.

Lemma 6.3 ([16, Proposition 5.34])**.**
Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz function with Lipschitz constant K. Let x be the $(0, 1)$-Gaussian in $\mathbb{R}^d$. Then $\mathbb{E}_x f(x)$ exists and for $t \geq 0$ we have

$$\mathrm{Prob}_x \left\{ f(x) - \mathbb{E}_x f(x) \geq t \right\} \leq \exp\left(-\frac{t^2}{2K^2} \right).$$

□

We can obtain a simple probability estimate of a Gaussian vector deviating too far from its center. A similar estimate holds for Gaussian matrices as well.

Corollary 6.4.
Let x be a $(\bar{x}, \sigma)$-Gaussian vector in $\mathbb{R}^d$. Let κ and θ be in $\mathbb{R}_0^+$. Then

$$\mathrm{Prob}_X\{\|x - \bar{x}\| \geq \kappa\sigma\} \leq \exp\left(-\frac{1}{2}\kappa^2 \right),$$
$$\mathrm{Prob}_X\{\|x - \bar{x}\| \geq \theta\} \leq \exp\left(-\frac{1}{2}\theta^2\sigma^{-2} \right).$$

Proof. The random vector $\sigma^{-1}(x - \bar{x})$ is the $(0, 1)$-Gaussian in $\mathbb{R}^d$. The inverse triangle inequality implies that the norm $\|\cdot\|$ has Lipschitz constant 1. Furthemore, $\mathbb{E}_X \sigma^{-1}\|x - \bar{x}\| = 0$ exists. Since $\|x - \bar{x}\| > \kappa\sigma$ is equivalent to $\|\sigma^{-1}(x - \bar{x})\| > \kappa$, we may apply Lemma 6.3 to find the first result. The second result follows immediately from $\theta = \theta\sigma^{-1}\sigma$, choosing $\kappa = \theta\sigma^{-1}$, and the first result. □

Corollary 6.5.
Let G be a random matrix with $(0, \sigma)$-Gaussian entries. Let κ and θ be in $\mathbb{R}_0^+$. Then

$$\mathrm{Prob}_G\{\|G\|_2 \geq \kappa\sigma\} \leq \exp\left(-\frac{1}{2}\kappa^2 \right),$$
$$\mathrm{Prob}_G\{\|G\|_2 \geq \theta\} \leq \exp\left(-\frac{1}{2}\theta^2\sigma^{-2} \right).$$

Proof. It suffices to bound the probability of $\|G\|_F > t\sigma$, because $\|G\|_F \geq \|G\|_2$. But the center of G is the 0-matrix, and so Corollary 6.4 applies. The second result follows immediately from $\theta = \theta\sigma^{-1}\sigma$, choosing $\kappa = \theta\sigma^{-1}$ and the first result. □

In the one-dimensional case, a tail estimate, widely regarded as canonical, can help to find a sharper bound if $\tau > \sigma/2$ holds:

Lemma 6.6.
Let x be a $(0, \sigma)$-Gaussian vector in $\mathbb{R}$. Then for $\tau > 0$ we have

$$\mathrm{Prob}\{x \geq \tau\} \leq \frac{\sigma}{2\tau} \exp\left(-\frac{\tau^2}{2\rho^2} \right).$$

Proof. We verify

$$\begin{aligned}\mathrm{Prob}\{x > \tau\} = \mathrm{Prob}\{\sigma^{-1}x > \sigma^{-1}\tau\} &= \int_{\sigma^{-1}\tau}^{\infty} e^{-\theta^2}\mathrm{d}\theta \leq \int_{\sigma^{-1}\tau}^{\infty} \frac{\sigma\theta}{\tau} e^{-\theta^2}\mathrm{d}\theta \\ &\leq \left[-\frac{\sigma}{2\tau}e^{-\theta^2}\right]_{\sigma^{-1}\tau}^{\infty} = \frac{\sigma}{2\tau}e^{-\tau^2}.\end{aligned}$$

□

The following result proves useful if the tail bound is parametric.

Lemma 6.7 ([15, Lemma 7.4, 1.])**.**
Let x be a $(\bar{x}, \sigma)$-Gaussian vector, $\beta \geq 0$ and $n \geq d$. Then

$$\mathrm{Prob}\left\{\|x - \bar{x}\| \geq \beta\sqrt{d\log n}\sigma\right\} \leq n^{-\frac{\beta^2 d}{2}}.$$

Proof. With a simple application of Corollary 6.4, we derive

$$\mathrm{Prob}\left\{\|x - \bar{x}\| \geq \beta\sqrt{d\log n}\sigma\right\} \leq \exp\left(-\frac{\beta^2}{2}d\log n\right) = \exp\log\left(n^{-\frac{\beta^2}{2}d}\right) = n^{-\frac{\beta^2 d}{2}}.$$

□

Surprisingly, there exist non-trivial upper bounds for the probability that a Gaussian vector is close to an arbitrary point. The proof is independent of the previous corollaries. Let us recall the volume formula for the d-dimensional unit ball:

$$|B_1^d(0)| = \frac{\pi^{d/2}}{\Gamma\left(1 + \frac{d}{2}\right)} = \frac{2\pi^{d/2}}{d\Gamma\left(\frac{d}{2}\right)}.$$

Lemma 6.8 ([13, Proposition 2.4.7])**.**
Let x be a $(\bar{x}, \sigma)$-Gaussian in $\mathbb{R}^d$. Let $p \in \mathbb{R}^d$. Then

$$\mathrm{Prob}\{\|x - p\| \leq \gamma\} \leq \left(\min\left(1, \sqrt{\frac{e}{d}}\right)\frac{\gamma}{\sigma}\right)^d.$$

Proof. We derive

$$\begin{aligned}\int_{B_\gamma(p)} \mu_x \mathrm{d}t &= \left(\sqrt{2\pi}\sigma\right)^{-d}\int_{B_\gamma(p)} \exp\left(-\frac{1}{2}\sigma^{-2}\|t - \bar{x}\|^2\right)\mathrm{d}t = \left(\sqrt{2\pi}\sigma\right)^{-d}\int_{B_\gamma(p)} 1 dt \\ &= \left(\sqrt{2\pi}\sigma\right)^{-d}|B_\gamma(p)| = \left(\sqrt{2\pi}\sigma\right)^{-d}\frac{2\pi^{d/2}\gamma^d}{d\Gamma(d/2)} = \left(\frac{\gamma}{\sigma}\right)^d\frac{2}{d2^{d/2}\Gamma(d/2)}.\end{aligned}$$

Note that $2(d2^{d/2}\Gamma(d/2))^{-1} \leq 1$ in general. Furthermore, the Gamma inequality [13, Proposition 2.4.8] implies for $d \geq 3$ that $2(d2^{d/2}\Gamma(d/2))^{-1} \leq (e/d)^{d/2}$. This completes the proof. □

Corollary 6.9 ([15, Lemma 7.4, 2.])**.**
Let $d \geq 3$ and x be a $(\bar{x}, \sigma)$-Gaussian vector. Then

$$\mathrm{Prob}\left\{\|x\| \leq e^{-3/2}\sigma\sqrt{d}\right\} \leq e^{-d}.$$

Proof. We apply Lemma 6.8 with $p = 0$ and immediately derive

$$\mathrm{Prob}\left\{\|x\| \leq e^{-3/2}\sigma\sqrt{d}\right\} \leq \left(\min\left(1, \sqrt{\frac{e}{d}}\right)\frac{e^{-3/2}\sigma\sqrt{d}}{\sigma}\right)^d \leq e^{-d}.$$

□

We close the treatment of Gaussian variables with the algorithic remark that it is possible to simulate a Gaussian distribution provided that a random variable on $[0,1]$ and, e.g., trigonometric functions are available. We assume this for the computational model that we use in this thesis.

We turn our attention to the uniform probability distribution on the unit spheres. — Measure spaces with bounded total measure can be interpreted in probabilistic terms after normalization of the measure. An important example is the hypersphere, where the measurable sets are the intersection of measurable subsets of $\mathbb{R}^{d+1}$ with $\mathcal{S}^d$. Let $\mathcal{S}^d \subset \mathbb{R}^{d+1}$ denote the d-dimensional unit sphere. Then $\mathcal{S}^d$ has surface measure

$$|\mathcal{S}^d| = 2\pi^{\frac{d+1}{2}} \Gamma\left(\frac{d+2}{2}\right)^{-1}.$$

Considering the curved part of the surface of the spherical cap

$$\mathcal{S}^d(h) := \left\{u \in \mathcal{S}^d \mid \langle u, e_1\rangle > h\right\},$$

a formula by Li [9] states that

$$|\mathcal{S}^d(h)| = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \int_0^{\pi/2-\sin^{-1}(h)} \sin^{d-1}(\theta)\mathrm{d}\theta.$$

Let $\nu \in \mathcal{S}^d$, $h \in [0,1]$ and u be the uniformly distributed random variable on $\mathcal{S}^d$. Then

$$\begin{aligned}
&\mathrm{Prob}_u\left\{\langle u,\nu\rangle \geq h\right\} = \frac{|\mathcal{S}^d(h)|}{|\mathcal{S}^d|} = |\mathcal{S}^d|^{-1} \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \int_0^{\pi/2-\sin^{-1}(h)} \sin^{d-1}(\theta)\mathrm{d}\theta \\
&= \frac{1}{2}\left(\int_0^{\pi/2} \sin^{d-1}(\theta)\mathrm{d}\theta\right)^{-1} \int_0^{\pi/2-\sin^{-1}(h)} \sin^{d-1}(\theta)\mathrm{d}\theta \\
&= \frac{1}{2}\left(\int_0^{\pi/2} \sin^{d-1}(\theta)\mathrm{d}\theta\right)^{-1} \left(\int_0^{\pi/2} \sin^{d-1}(\theta)\mathrm{d}\theta - \int_{\pi/2-\sin^{-1}(h)}^{\pi/2} \sin^{d-1}(\theta)\mathrm{d}\theta\right) \\
&= \frac{1}{2} - \frac{1}{2}\left(\int_0^{\pi/2} \sin^{d-1}(\theta)\mathrm{d}\theta\right)^{-1} \int_{\pi/2-\sin^{-1}(h)}^{\pi/2} \sin^{d-1}(\theta)\mathrm{d}\theta.
\end{aligned}$$

A lower estimate, useful for small h, thus is

$$\mathrm{Prob}_u\left\{\langle u,\nu\rangle \geq h\right\} \geq 1 - \frac{1}{\pi}\sin^{-1}(h). \tag{5}$$

An upper estimate for small h is also known as concentration of measure on the sphere. Let $B^{d+1}(h) := \mathrm{convex}\{0, S^d(h)\}$ be the "cone" element suspended from the spherical cap $S^d(h)$ to the origin. We observe that $B^{d+1}(h)$ is contained within a ball of radius $\sqrt{1-h^2}$ provided that $h \leq \frac{1}{\sqrt{2}}$. More precisely, the ball can be chosen to be centered at $h\nu$ for ν being the symmetry axis of $B^{d+1}(h)$. Under this condition we have $|B^{d+1}(h)| \leq \sqrt{1-h^2}^{d+1}|B^{d+1}|$ and thus

$$\mathrm{Prob}_u\left\{\langle u,\nu\rangle \geq h\right\} \leq \frac{|\mathcal{S}^d(h)|}{|S^d|} = \frac{|B^{d+1}(h)|}{|B^{d+1}|} \leq (1-h^2)^{\frac{d+1}{2}} \leq e^{-\frac{1}{2}(d+1)h^2}. \tag{6}$$

This proof is taken from [1, Lecture 2].

Finally, the Haar measure on the orthogonal group $\mathcal{O}(d)$ is of minor relevance for this thesis. We only state a simplified statement on its existence, combined from [3, Kapitel VIII]. [i]

Theorem 6.10.
There does exist a unique measure $\mathfrak{h}$ on $\mathcal{O}(d)$, called normalized Haar measure, defined on the σ-algebra $\mathfrak{O}$ of intersections of open subsets of $\mathbb{R}^d$ with $\mathcal{O}(d)$, such that

$$\mathfrak{h}\left(\mathcal{O}(d)\right) = 1, \quad \forall Q \in \mathcal{O}(d), \mathcal{U} \in \mathfrak{O} : \mathfrak{h}(Q\mathcal{U}) = \mathfrak{h}(\mathcal{U})$$

□

In other words, the Haar measure is a probability measure that is invariant under the group operation from the left. The normalized Haar measure on $\mathcal{O}(d)$ is the conceptual sibling of the uniform measure on the unit sphere and the natural choice for a probability measure. It can implemented algorithmicly, provided that Gaussian variables in $\mathbb{R}$ are available in the computational model; see also Remark 8.1.

6.1 The change of variables formula

Blaschke's formula and its consequences are used frequently in the work of Spielman, Teng and Vershynin. We present it in a manner more akin to the background in integration theory.

Let $q \in S^1(\mathbb{R}^d)$ be arbitrary but fixed, and assume that $\omega \in S^1(\mathbb{R}^d)$ with $\omega \neq -q$. Then there exists a unique $R^q_\omega \in O(d)$ which maps q to ω and acts as the identity on $\operatorname{lin}\{\omega, q\}^\perp$. Furthermore, we let I^q be any isometric linear mapping from $\mathbb{R}^{d-1}$ onto $q^\perp$. Note that R^q_ω maps $q^\perp$ to $\omega^\perp$. We consider the coordinate transform

$$\begin{aligned} \Phi^q : \quad & \mathbb{R}^{(d-1)\times d} \times \mathbb{R}^+_0 \times S^1(\mathbb{R}^d) \to \mathbb{R}^{d\times d}, \\ & (y_1, \ldots, y_d, r, \omega) \mapsto (R^q_\omega I^q y_1 + r\omega, \ldots, R^q_\omega I^q y_d + r\omega). \end{aligned} \tag{7}$$

With abuse of notation, we ignore here that Φ^q is not defined for $\omega = -q$. A famous theorem by Blaschke provides a convenient formula for the coordinate transform induced by Φ^q.

Theorem 6.11.
Let $F : \mathbb{R}^{d\times d} \to \mathbb{R}$ be a measurable function, and let

$$\mathcal{X} \subseteq \mathbb{R}^{d\times d}, \quad \mathcal{Y} \subseteq \mathbb{R}^{(d-1)\times d} \times \mathbb{R}^+_0 \times S^1(\mathbb{R}^d)$$

be measurable sets such that $\mathcal{X} = \Phi^q(\mathcal{Y})$. Then

$$\int_{\mathcal{X}} F(x_1, \ldots, x_d) \mathrm{d}x_1 \ldots \mathrm{d}x_d \tag{8}$$

$$= \int_{\mathcal{Y}} (F \circ \Phi^q)(y_1, \ldots, y_d, r, \omega) \cdot (d-1)! \operatorname{vol}^{d-1} \operatorname{convex}\{y_1, \ldots, y_d\} \mathrm{d}\omega \mathrm{d}r \mathrm{d}y_1 \ldots \mathrm{d}y_d \tag{9}$$

Note that this is basically a statement on the term $|\det D\Phi^q|$ in the change of variables of integration theory. In our applications, F generally is the joint distribution of Gaussian random variables $a_1, \ldots, a_d$ in $\mathbb{R}^d$. — We first note a basic observation:

[i] While not stated explicitly there, our version follows from Definition 1.1, Theorem 3.12, Theorem 3.15 b) and Theorem 1.16 of the referenced chapter, together with elementary topology.

Lemma 6.12.
Suppose that $x = \Phi^q(y, r, \omega)$ for $(y, r, \omega) \in \mathbb{R}^{(d-1)\times d} \times \mathbb{R}_0^+ \times S^1(\mathbb{R}^d)$. Then $\|x\| \geq \|y\|$.

Proof. With the triangle inequality we verify that

$$\|x\| = \|R_\omega^q I^q y + r\omega\| \geq \|R_\omega^q I^q y\| = \|y\|,$$

where we have used $\omega \perp \operatorname{Im} R_\omega^q$. □

Now assume that $a_1, \ldots, a_d$ are Gaussian variables in $\mathbb{R}^d$ with center of norm $\bar{a}_1, \ldots, \bar{a}_d$ and common standard deviation σ, and let F denote their joint density function. Then the function

$$G(y_1, \ldots, y_d, r, \omega) = (F \circ \Phi^q)(y_1, \ldots, y_d, r, \omega) \cdot (d-1)! \operatorname{vol}^{d-1} \operatorname{convex}\{y_1, \ldots, y_d\}$$

is the density function of a probability distribution on $\mathbb{R}^{(d-1)\times d} \times \mathbb{R}_0^+ \times S^1(\mathbb{R}^d)$. For fixed r and ω, it takes the form of a joint density of Gaussian vectors itself.

Theorem 6.13.
Let $a_1, \ldots, a_d$, F and G be as above, and let r and ω be fixed. Then G has the form

$$G(y_1, \ldots, y_d, r, \omega) = (d-1)! \operatorname{vol}^{d-1} \operatorname{convex}\{y_1, \ldots, y_d\} \cdot \prod_{i=1}^{d} \phi_i(y_1, \ldots, y_d)$$

where the ϕ_i are probability density functions of $(\bar{b}_i, \sigma)$-Gaussians on $\mathbb{R}^{d-1}$ with $\|\bar{b}_i\| \leq \|\bar{a}_i\|$.

Proof. A direct application of Blaschke's Lemma yields that

$$G(y_1, \ldots, y_d, r, \omega) = (d-1)! \operatorname{vol}^{d-1} \operatorname{convex}\{y_1, \ldots, y_d\} \cdot \prod_{i=1}^{d} \mu_i(\Phi^q(y_1, \ldots, y_d, r, \omega))$$

where the $\mu_1, \ldots, \mu_d$ are the density functions of the random variables $a_1, \ldots, a_d$. Let $1 \leq i \leq d$ be fixed. We use the results at the beginning of the section. First, note that $\mu_i(\cdot + r\omega)$ is a $(\bar{a}_i - r\omega, \sigma)$-Gaussian. Then $\mu_i(R_\omega^q \cdot + r\omega)$ must be a $((R_\omega^q)^{-1}\bar{a}_i - r(R_\omega^q)^{-1}\omega, \sigma)$ Gaussian. We see that the new center of norm is in fact $(R_\omega^q)^{-1}\bar{a}_i - rq$. Similarly as in the proof of Lemma 6.1, we can use [3, Kapitel III.1, Satz 1.7] to infer the desired statement along the lines of the aforementioned proof. □

The latter two results can be found in Appendix C of [15].

7 The Shadow Vertex Method

We provide a concise description of the shadow vertex method, which is a variant of the simplex method. We draw from the book of Borgwardt [2] and the seminal publication of Spielman and Teng [13]. These authors discuss the algorithm in a 'primal' and a 'polar' formulation. The difference between the two presentations is mostly formal, and instead we choose a presentation of the algorithm that already demonstrates a (possible) implementation. We precede the algorithmic discussion with an outline of some duality relations for pointed unit linear programming problems and properties of polyhedra projected onto hyperplanes.

For the rest of this section, let $A \in \mathbb{R}^{n\times d}$ and $z \in \mathbb{R}^d$.

7.1 Properties of Unit Linear Programming Problems

Suppose a unit linear programming problem $(A, \vec{1}, z)$ is to be solved.

$$\text{Maximize} \quad \langle z, x \rangle \quad \text{subject to} \quad Ax \leq \vec{1}.$$

We write $\mathcal{P} := [A, \vec{1}]$ for the corresponding polyhedron in this section, and let $a_1, \ldots, a_n \in \mathbb{R}^d$ denote the rows of A. The polar $\mathcal{P}^*$ of $\mathcal{P}$ has a useful characterization.

Lemma 7.1 ([2, Lemma 1.5])**.**
We have

$$\mathcal{P}^* \equiv \left\{ y \in \mathbb{R}^d \mid \forall x \in \mathcal{P} : \langle x, y \rangle \leq 1 \right\} = \text{convex} \left\{ \vec{0}, a_1, \ldots, a_n \right\}. \tag{10}$$

Proof. Suppose that y is contained in the right-hand side, so there exist $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ with

$$\lambda_1, \ldots, \lambda_n \geq 0, \quad \sum_{i=1}^{n} \lambda_i \leq 1, \quad y = \sum_{i=1}^{n} \lambda_i a_i.$$

For $x \in \mathcal{P}$ we observe $\langle x, y \rangle = \lambda_1 \langle a_1, x \rangle + \cdots + \lambda_n \langle a_n, x \rangle \leq 1$, so $y \in \mathcal{P}^*$.

Conversely, suppose that $y \in \mathbb{R}^d$ with $\langle y, \mathcal{P} \rangle \leq 1$. Assume that y is not a member of the right-hand side of (10). By the Hahn-Banach separation theorem (Theorem 2.7) there exists $z \in \mathbb{R}^d$ such that

$$\langle z, y \rangle > 1, \quad \forall w \in \text{convex}\{\vec{0}, a_1, \ldots, a_n\} : \langle w, z \rangle \leq 1.$$

But then $z \in \mathcal{P}$ and so $y \notin \mathcal{P}^*$, which is a contradiction. This proves the lemma. □

We introduce a general position condition on the matrix A, which holds almost surely if the entries are Gaussian random variables:

$$\forall I \subset \{1, \ldots, n\}, |I| = d + 1 : \text{vol}^d \, \text{convex}\{a_i \mid i \in I\} > 0. \tag{GP1}$$

We assume (GP1) for the rest of this section. This condition implies that all submatrices of A with d rows are invertible, and therefore that $\mathcal{P}$ is pointed, and that an optimal solution of the unit linear program is realized at a vertex of $\mathcal{P}$. The general position condition implies furthermore that the vertices can almost surely be identified with an index d-set $I \subseteq [n]$.

In other words, (GP1) implies

$$A_I := (a_i)_{i \in I} \in GL(d), \quad \dim \text{convex}\{a_i \mid i \in I\} = d - 1.$$

For notational simplicity we write

$$x_I := A_I^{-1} \vec{1} \in \mathbb{R}^d, \quad \mathcal{F}_I := \text{convex}\{a_i \mid i \in I\},$$
$$\mathcal{H}_I := \left\{ y \in \mathbb{R}^d \mid y^t x_I = 1 \right\}, \quad \mathcal{C}_I := \text{cone}\{a_i \mid i \in I\}.$$

It is essential to understand that the solution properties of a pointed unit linear programming problem can be viewed through the following definitions and the subsequent equivalence theorem.

Definition 7.2.
We say that I describes a vertex of $\mathcal{P}$ if x_I is a vertex of $\mathcal{P}$. We say that I describes a facet of $\mathcal{P}$ if $\mathcal{F}_I$ is a facet of $\mathcal{P}^*$.

Definition 7.3.
Let $I \subseteq [n]$ be a d-set of indices. We say that I satisfies the primal optimal vertex condition if x_I is a vertex of $\mathcal{P}$ that is an optimal solution of the linear programming problem. We say that I satisfies the polar optimal simplex condition if

$$z \in \mathcal{C}_I, \quad \mathcal{F}_I \text{ is a facet of } \operatorname{convex}\{\vec{0}, a_1, \ldots, a_n\}.$$

Theorem 7.4 ([2, Lemma 1.6, Lemma 1.7, Lemma 1.8])**.**
Let $I \subseteq [n]$ be a d-set of indices. Then I describes a vertex of $\mathcal{P}$ if and only if I describes a facet of $\mathcal{P}^*$, and I satisfies the primal optimal vertex condition if and only if I satisfies the polar optimal facet condition.

If I satisfies the primal optimal vertex condition or the polar optimal facet condition, then we say that I is optimal with respect to z.

Proof. For an auxiliary result, let $I \subseteq [n]$ describe a vertex x_I of $\mathcal{P}$. Then

$$\mathcal{F}_I = \mathcal{P}^* \cap \mathcal{H}_I \equiv \left\{ y \in \mathbb{R}^d \mid \langle y, x_I \rangle = 1, \forall x \in \mathcal{P} : \langle y, x \rangle \leq 1 \right\}. \tag{11}$$

To see this, assume that $y \in \mathcal{H}_I \cap \mathcal{P}^*$. Then there exist $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ such that

$$\lambda_i \geq 0, \quad \sum_{i=1}^{n} \lambda_i \leq 1, \quad y = \sum_{i=1}^{n} \lambda_i a_i.$$

We can infer $\lambda_1 + \cdots + \lambda_n = 1$ by

$$1 = \langle y, x_I \rangle = \sum_{i=1}^{n} \lambda_i \langle a_i, x_I \rangle \leq \sum_{i=1}^{n} \lambda_i \leq 1.$$

Hence, $y \in \mathcal{F}_I$, and the opposite inclusion in (11) holds by definition. The general position condition (GP1) implies that I is unique.

Now, we show that x_I is a vertex if and only if $\mathcal{F}_I$ is a facet of $\mathcal{P}^*$. Suppose that x_I is a vertex. Then $\{a_i \mid i \in I\}$ is a subset of $\mathcal{H}_I$. Note that $\mathcal{F}_I \subset \mathcal{H}_I$, while $\langle a_j, x_I \rangle \leq 1$ for $j \notin I$. Thus $\mathcal{H}_I$ is a supporting hyperplane of $\mathcal{P}^*$, and $\mathcal{F}_I$ is a facet of $\mathcal{P}^*$. Conversely, assume that $\mathcal{F}_I$ is a facet of $\mathcal{P}^*$. Then $\mathcal{H}_I$ is a supporting hyperplane of $\mathcal{P}^*$, and so x_I satisfies the definition of being a vertex of $\mathcal{P}$.

We have shown that x_I is a vertex of $\mathcal{P}$ if and only if $\mathcal{F}_I$ is a facet of $\mathcal{P}^*$. If a vertex x_I optimally solves the primal problem, then the dual problem has an optimal solution $y \in \mathbb{R}^n$ that satisfies

$$y \geq 0, \quad y^t(\vec{1} - Ax_I) = 0, \quad y^t A = z^t.$$

From this we infer that $z \in \mathcal{C}_I$. Conversely, if $z \in \mathcal{C}_I$ and $\mathcal{F}_I$ is a facet of $\mathcal{P}^*$, then $\langle a_i, x_I \rangle = 1$ for $i \in I$ and $\langle a_i, x \rangle \leq 1$ for $x \in \mathcal{P}$. Hence

$$\forall x \in \mathcal{P} : \langle z, x_I \rangle \geq \langle z, x \rangle.$$

We conclude that x_I is a vertex of $\mathcal{P}$ that maximizes z over $\mathcal{P}$.

This completes the proof. □

7.2 Projected Polyhedra

We continue to consider the unit linear programming problem with the general position condition (GP1). Let $u \in \mathbb{R}^d$ be non-colinear to z, so $\mathcal{E} = \operatorname{lin}\{u, z\}$ is two-dimensional. Let $\mathcal{P}^{\mathcal{E}}$ be the orthogonal projection of $\mathcal{P}$ to $\mathcal{E}$, and write $x^{\mathcal{E}}$ for the orthogonal projection to $\mathcal{E}$ of $x \in \mathcal{P}$. To understand the shadow vertex simplex method, we have to relate concepts on $\mathcal{P}^{\mathcal{E}}$ with corresponding concepts on $\mathcal{P}$.

We have for $x \neq 0$ that every vector $v \in \mathcal{E}^{\perp}$ is already contained in $z^{\perp}$, and therefore $\langle x, z\rangle = \langle x^{\mathcal{E}}, z\rangle$ for $x \in \mathcal{P}$. In particular, if x_I maximizes z over $\mathcal{P}$, then $x_I^{\mathcal{E}}$ maximizes z over $\mathcal{P}^{\mathcal{E}}$, and the two maxima are the same. We introduce a second general position condition:

$$\forall V \subset \{a_1, \ldots, a_n, z, u\}, |V| = d : \dim \operatorname{lin} V = d. \tag{GP2}$$

If A is a Gaussian matrix, then (GP2) holds almost surely, and therefore we assume (GP2) in the following.

Definition 7.5.
We call x_I a shadow vertex of $\mathcal{P}$ with respect to $\mathcal{E}$ if $x_I^{\mathcal{E}}$ is a vertex of $\mathcal{P}^{\mathcal{E}}$.

Lemma 7.6 ([2, Lemma 1.2])**.**
Let x_I be a vertex of $\mathcal{P}$. Then x_I is a shadow vertex if and only if there exists $w \in \mathcal{E}, w \neq 0$ such that x_I maximizes w over $\mathcal{P}^{\mathcal{E}}$.

Proof. If x_I is a shadow vertex, then $x_I^{\mathcal{E}}$ is a vertex of $\mathcal{P}^{\mathcal{E}}$, and we can find a supporting hyperplane at $x_I^{\mathcal{E}}$ of $\mathcal{P}^{\mathcal{E}}$ in $\mathcal{E}$.

Conversely, let x_I be a vertex that maximizes $w = \alpha u + \beta v \in \mathcal{E} \setminus \{0\}$, $\alpha, \beta \in \mathbb{R}$ over $\mathcal{P}$. Then $x_I^{\mathcal{E}}$ must lie on an edge of $\mathcal{E}$, to which x_I must be orthogonal. There exists $v \in \mathcal{E}$ such that $x_I^{\mathcal{E}} + [-\epsilon, \epsilon]v$ lies within that edge. Note that $\{a_i \mid i \in I\}$ is a basis of $\mathbb{R}^d$. Let us parameterize $w + \mathbb{R}v$ and write

$$w + \gamma v = \sum_{i \in I} \lambda_i(\gamma) a_i,$$

where the $\lambda_i(\cdot)$ are affine. Note that w is the only point on the edge that intersects with $\mathcal{C}_I$, so there do exist at least two indices, say, $k, l \in I$, such that $\lambda_k(\cdot)$ and $\lambda_l(\cdot)$ vanish at $\gamma = 0$. But then

$$\vec{0} \neq w = \alpha z + \beta u = \sum_{\substack{i \in I \\ i \neq k, i \neq l}} \lambda_i(\gamma) a_i,$$

which contradicts non-degeneracy. □

Lemma 7.7 ([2, Lemma 1.3])**.**
Let x_I and x_J be shadow vertices of $\mathcal{P}$. If $x_I^{\mathcal{E}}$ and $x_I^{\mathcal{E}}$ are adjacent, then so are x_I and x_J.

Proof. The segment $[x_I^{\mathcal{E}}, x_J^{\mathcal{E}}]$ is an edge of $\mathcal{P}^{\mathcal{E}}$. Therefore there exists $w = \alpha u + \beta v \in \mathcal{E} \setminus \{\vec{0}\}$, $\alpha, \beta \in \mathbb{R}$, $w \neq \vec{0}$ such that

$$1 = \langle x_I, w\rangle = \langle x_J, w\rangle = \langle x_I^{\mathcal{E}}, w\rangle = \langle x_J^{\mathcal{E}}, w\rangle$$

is maximized over $\mathcal{P}$. It can be seen by Farkas' lemma that w is contained in the cone of the restriction vectors that are satisfied exactly over $[x_I^{\mathcal{E}}, x_J^{\mathcal{E}}]$. Suppose that $[x_I, x_J]$ is not an edge

of $\mathcal{P}$. Then it is not in the intersection of $(d-1)$ supporting hyperplanes of $\mathcal{P}$. But then w is in the convex cone of at most $(d-2)$ vectors of $\{a_i \mid i \in I\}$,

$$\alpha u + \beta v = \sum_{j \in J} \gamma_j a_i, \quad J \subset [n], |J| < d-1.$$

But this contradicts the general position condition (GP2) so the lemma follows. □

Lemma 7.8.
Let y be a vertex of $\mathcal{P}^{\mathcal{E}}$. Then it is the projection of a vertex of $\mathcal{P}$.

Proof. If y is a vertex of $\mathcal{P}^{\mathcal{E}}$, then there exists $w \in \mathcal{E}$ such that y is the only maximizer of w over $\mathcal{P}^{\mathcal{E}}$. But then $y + \{g \in \mathcal{E} \mid w \perp g\} + \mathcal{E}^{\perp}$ is a supporting hyperplane of $\mathcal{P}$, which intersects with $\mathcal{P}$ only in $y + \mathcal{E}^{\perp}$, where it contains a vertex of $\mathcal{P}$. □

Remark 7.9.
For a matrix with Gaussian entries, the projections of two vertices are almost surely distinct.

7.3 Algorithmic Details

This leads to a method to solve linear programming problems: If we choose a vector $u \in \mathbb{R}^d$ as above, called initial direction, we find a solution to the problem if we maximize z over $\mathcal{E} = \text{lin}\{u, z\}$. This is the idea of the shadow vertex method.

Let $q_\lambda := \lambda z + (1-\lambda)u$, and let $x_0 \in \mathcal{P}^{\mathcal{E}}$ be a vertex that maximizes $u \equiv q_0$ over $\mathcal{P}^{\mathcal{E}}$. Then x_0 is the shadow of a vertex x of $\mathcal{P}$, called initial solution. We try to construct a sequence $\lambda_0 \equiv 0, \lambda_1, \lambda_2, \ldots$ of real numbers, and vertices $x_0^{\mathcal{E}}, x_1^{\mathcal{E}}, \ldots$ of $\mathcal{P}^{\mathcal{E}}$, such that $x_i^{\mathcal{E}}$ maximizes q_{λ_i} over $\mathcal{P}^{\mathcal{E}}$. Furthermore, we keep track of the original shadow vertices $x_0, x_1, \ldots$ of the $x_0^{\mathcal{E}}, x_1^{\mathcal{E}}, \ldots$. The problem has a solution if and only if we can find a vertex of $\mathcal{P}^{\mathcal{E}}$ that maximizes $q_1 \equiv z$ over $\mathcal{P}^{\mathcal{E}}$.

A pseudo-code for the shadow vertex method is shown in Algorithm 1. We estimate the time and space complexity of this algorithm, and we inspect how the respective steps can be implemented in detail.

An elementary subtask is to check whether an index d-set $I \subseteq [n]$ describes a vertex. On the one hand, the point x_I must be well-defined, i.e., A_I must be invertible. This either can be checked by matrix inversion algorithms, or holds by the general position conditions. On the other hand, the point x_I must satisfy the constraint set I of the linear programming problem; this can be checked by a series of scalar products. Therefore, the costs for this step are a matrix inversion and $\mathcal{O}(n)$ scalar products. Alternatively, we may check whether x_I describes a facet of $\mathcal{P}^*$. For this, we compute a normal to $\mathcal{F}_I$ and check whether all points are below the affine hull of $\mathcal{F}_I$, using $\mathcal{O}(n)$ operations. Interestingly, if we use the (stabilized) Gram-Schmidt process, then such a normal vector can be computed almost surely, in time $\mathcal{O}(d^3)$, using the generalized position assumptions.

Another elementary subtask is to determine the interval $[\mu_{\min}, \mu_{\max}] \subseteq [-\infty, \infty]$ such that I describes an optimal solution for $q_\mu = (1-\mu)u + \mu z$ with $\mu \in [\mu_{\min}, \mu_{\max}]$. We define

$$p^u := A_I^{-1} u, \quad p^z := A_I^{-1} z,$$

so that $A_I^{-1} q_\mu = (1-\mu)p^u + \mu p^z$. We know that I describes an optimal solution if and only if $q_\mu \in \text{cone}\{a_i \mid i \in I\}$. But this is equivalent to $A_I^{-1} q_\mu$ having non-negative coordinates. Now

Input: $A \in \mathbb{R}^{n\times d}, z \in \mathbb{R}^d, u \in \mathbb{R}^d, I_0 \subset [n]$ such that x_{I_0} is a vertex of $\mathcal{P}$ that maximizes u over $\mathcal{P}$.

Output: Either 'failure' or $I \subset [n]$ such that x_I is a vertex of $\mathcal{P}$ that maximizes z over $\mathcal{P}$.

Algorithm:

(I) Let $i = 0$ and $\lambda_0 = 0$. Check that I_0 describes a vertex of $\mathcal{P}$.

(II) Let $[\lambda_{\min}, \lambda_{\max}] \subseteq [-\infty, \infty]$ be the maximal interval such that I_0 is optimal for q_λ with $\lambda \in [\lambda_{\min}, \lambda_{\max}]$. Check that $0 \in [\lambda_{\min}, \lambda_{\max}]$.

(III) If $1 \in [\lambda_{\min}, \lambda_{\max}]$, then return I_i.

(IV) Find $j \in I_i$ and $k \notin I_i$ such that

(i) $I_i \cup \{k\} - \{j\}$ describes a vertex of $\mathcal{P}$.

(ii) When $[\mu_{\min}, \mu_{\max}]$ is the maximal interval such that $I_i \cup \{k\} - \{j\}$ is optimal for q_μ with $\mu \in [\mu_{\min}, \mu_{\max}]$, then $\mu_{\max} > \lambda_{\max}$.

If this is not possible, then return 'failure'. Otherwise, let $I_{i+1} = I_i \cup \{k\} - \{j\}$, $\lambda_{\min} = \mu_{\min}$, $\lambda_{\max} = \mu_{\max}$ and $i = i + 1$.

(V) Go to (III)

Algorithm 1: Shadow vertex simplex algorithm

it is easy to determine the desired maximal interval $[\mu_{\min}, \mu_{\max}]$ from p^u and p^z. This last step requires $O(d)$ operations, and two matrix inversions have been employed before.

These two subtasks are solved at each iteration of the main loop of the shadow vertex algorithm. Note that at each iteration we cycle through all indices $j \in I_i$ and $k \notin I_i$ until either the range is exhausted, or we have found an index set I_{i+1} which improves upon the current solution. A crude upper bound for the number of those iterations is $\mathcal{O}(nd)$.

We recall that scalar products use $\mathcal{O}(d)$ operations and matrix inversions use $\mathcal{O}(d^3)$ operations. Consolidating these thoughts, we obtain the following bounds on the time and space complexity of the method.

Theorem 7.10 (Time complexity of the simplex method, naive estimate)**.**
Let t be the number of pivot steps in the shadow vertex method. Then the number of operations used in the shadow vertex method is

$$\mathcal{O}(tn^2d^4)$$

□

Remark 7.11.
In each of the subtasks, multiplications and divisions dominate the overall complexity. Thus, this estimate also applies if we only count multiplications and divisions and disregard additions and subtractions.

Remark 7.12.
The estimate can be improved at several points, and further aspects of the implementation should be mentioned briefly.

First, it is not necessary to actually invert any matrix throughout the algorithm; instead, we only apply a matrix inversion algorithm to several right-hand sides. This does not affect the asymptotic complexity of the method, but the numerical stability of most operations.

An important observation is that there are only three right-hand sides $z, u, \vec{1} \in \mathbb{R}^d$ for which we actually compute preimages under A_I, and the matrix A_I between pivot steps changes only by one row. It is therefore possible to recycle information from previous iterations, which may help to reduce the asymptotic complexity of the method in variable d.

Lastly, the number of indices checked at each pivot step may be reduced. For example, under the general position condition each vertex has only d neighbors, so we may stop the search after having encountered at least d neighbours. A clever ordering of the rows may also help to reduce the complexity of each pivot step.

Remark 7.13.
The general position conditions are not only relevant for a stochastic analysis of the simplex method, but also are important for the understanding of the method itself. Let $k \geq 2$ and $1 \leq i \leq k-1$ and set

$$\begin{aligned} e &= (0,0,1), \\ a_0 &= (1,0,0), \qquad\qquad a_k = (-1,0,0), \\ a_i^+ &= \left(\cos(\tfrac{i}{2\pi}), \sin(\tfrac{i}{2\pi}), 0\right), \quad a_i^- = \left(\cos(\tfrac{i}{2\pi}), -\sin(\tfrac{i}{2\pi}), 0\right). \end{aligned}$$

Let $\mathcal{P}$ be the convex closure of these points. Let $\mathcal{Q}$ its projection onto the x-z-plane. Then $\mathcal{Q}$ has an edge between a_0 and a_1 that is not the projection of an edge of $\mathcal{P}$. The shadow vertex method then follows the vertices in the upper or lower 'arc' of $\mathcal{P}$, but these vertices are not shadow simplices. If the algorithm followed directly the vertices in $\mathcal{Q}$, then the path would be arbitrarily shorter for k arbitrarily large, but this would only constitute a generalized simplex method. This does not affect the algorithmic implementation of the shadow vertex method, but the relation between the number of vertices of $\mathcal{Q}$ and the number of pivot steps depends on the general position condition.

8 Adding constraints

Phase I of Vershynin's simplex method requires solving a sequence of unit linear programming problems, each of which is derived from the original problem by imposure of additional constraints. These new constraints are assembled by the algorithm `Adding Constraints` which this section describes and analysis; see Algorithm 2 for the pseudo-code. The underlying idea is that `Adding Constraints` creates random constraint vectors $a_{n+1}, \ldots, a_{n+d}$ and a functional vector z_0 such that, with high probability, the maximum of z_0 over the polyhedron defined by the constraint vectors $a_1, \ldots, a_{n+d}$ with unit right-hand side is realized at the vertex x_I defined by the index set $I = \{n+1, \ldots, n+d\}$.

For notational purposes we introduce the vectors

$$\begin{aligned} w = d^{-\frac{1}{2}}\vec{1}, \quad \bar{v}_i &= ld(d^2-d)^{-\frac{1}{2}}e_i + d^{-\frac{1}{2}}\vec{1} - l(d^2-d)^{-\frac{1}{2}}\vec{1}. \\ &= l(d^2-d)^{-\frac{1}{2}}(de_i - \vec{1}) + d^{-\frac{1}{2}}\vec{1}. \end{aligned}$$

Here $l > 0$. The simplex $\Delta = \text{convex}\{\bar{v}_1, \ldots, \bar{v}_d\}$ has center and normal w, which is a unit vector, and the distance between center and each vertex is l. For usage in the sequel, let T be the invertible matrix whose columns are precisely the vectors $\bar{v}_i$. One can see that T

acts on $\text{lin}\{\vec{1}\}^\perp$ as scaling by $ld(d^2-d)^{-\frac{1}{2}}$ and on $\text{lin}\{\vec{1}\}$ as scaling by $d^{\frac{1}{2}}$. This describes the eigenspace decomposition of T. The maximal and minimal singular values of T and T^{-1}, respectively, satisfy

$$s_{\max}(T) = \max\{d^{\frac{1}{2}}, ld(d^2-d)^{-\frac{1}{2}}\}, \quad s_{\max}(T^{-1}) = \max\{d^{-\frac{1}{2}}, l^{-1}d^{-1}(d^2-d)^{\frac{1}{2}}\},$$
$$s_{\min}(T) = \min\{d^{\frac{1}{2}}, ld(d^2-d)^{-\frac{1}{2}}\}, \quad s_{\min}(T^{-1}) = \min\{d^{-\frac{1}{2}}, l^{-1}d^{-1}(d^2-d)^{\frac{1}{2}}\}.$$

Intuitively, T scales $\text{convex}\{e_1, \dots, e_d\}$ into a regular simplex of diamter l, centered at and normal to w, as can also be seen by direct calculation.

Input:	$\zeta > 0$, $l > 0$, $\rho > 0$
Output:	Either 'failure' or $a_{n+1}, \dots, a_{n+d}, z_0 \in \mathbb{R}^d$.
Algorithm:	(I) Let $v_1, \dots, v_d$ be random variables in $\mathbb{R}^d$ with standard deviation ρ and centers of norm $\bar{v}_{n+1}, \dots, \bar{v}_{n+d}$.
	(II) Let Q be a random variable in $\mathcal{O}(d)$ drawn by Haar measure. Let $z_0 = 2\zeta Q w$ and $a_{n+i} = 2\zeta Q v_i$ for $1 \le i \le d$.
	(III) Check that $z_0 \in \text{cone}\{a_{n+1}, \dots, a_{n+d}\}$ and $\text{aff}\{a_{n+1}, \dots, a_{n+d}\}$ has distance from the origin at least ζ. If not, return 'failure', otherwise, return $a_{n+1}, \dots, a_{n+d}, z_0$.

Algorithm 2: `Adding Constraints` algorithm

The analysis of `Adding Constraints` is subject to the next subsection. We note that the random variables $a_{n+i} = 2\zeta Q v_i$ for $1 \le i \le d$ have standard deviation $2\zeta\rho$ and respective mean values $\bar{a}_{n+i} = 2\zeta Q \bar{v}_i$.

Remark 8.1.
Algorithm 2 requires drawing a random element of $Q \in \mathcal{O}(d)$ according to the Haar measure. This can be implemented by drawing d^2 Gaussian variables on the real line (with arbitrary but common standard deviations) to obtain a random matrix M, which is invertible almost surely, and an application of the LQ-decomposition to M. The matrix Q then is a orthogonal matrix drawn from $\mathcal{O}(d)$ by the Haar measure. It is easy to see that Algorithm 2 utilizes $\mathcal{O}(d^3)$ numerical operations and draws $\mathcal{O}(d^2)$ Gaussian variables in $\mathbb{R}$. We refer to [14] and [10] for further information.

It is easy to verify that `Adding Constraints` produces a new linear programming problem already together with its optimal solution. Let $\widehat{A}$ be a matrix that is derived from A by appending the additional rows $a_{n+1}, \dots, a_{n+d}$.

Theorem 8.2.
Let $\zeta \ge \max_{1 \le j \le n} \|a_j\|$. If Algorithm `Adding Constraints` does not return failure, then the linear programming problem $(\widehat{A}, \vec{1}, z_0)$, i.e., (Aug Unit LP), is solved by the index set $I = \{n+1, \dots, n+d\}$.

Proof. Suppose that `Adding Constraints` does not return failure. We need to show that I as above describes a facet of the polar polyhedron $\left(\widehat{\mathcal{P}}\right)^*$ of $\widehat{\mathcal{P}}$, and that $\mathbb{R}_0^+ \cdot z_0$ intersects with this facet. First, if I describes such a facet, then $z_0 \in \mathcal{C}_I$ is asserted in the final step of the algorithm. So it remains to show that I describes a facet. By the second assertion checked at

the end of `Adding Constraints`,

$$\operatorname{dist}\left(\operatorname{aff}\{a_{n+1},\ldots,a_{n+d}\},0\right) \geq \zeta + \epsilon \geq \zeta \geq \max_{1\leq j\leq n} \|a_j\|$$

for some $\epsilon > 0$, almost surely. But then, none of the vectors a_i, $n+1 \leq i \leq n+d$, lies in

$$\operatorname{convex}\{a_1,\ldots,a_{i-1},a_{i+1},\ldots,a_{n+d}\}.$$

This implies the desired property. □

Having shown correctness, The remainder of the section aims at estimating the probability that `Adding Constraints` returns a set of vectors, and furthermore that these vectors lie within some given halfspace $\mathcal{H}$. In later applications, that halfspace will be specified to be a fixed halfspace in the "numb set" of the linear programming problem considered; see Section 9. Throughout the following proofs, we let G be the random matrix in $\mathbb{R}^{d\times d}$ that satisfies

$$v_i = (T+G)e_i.$$

In other words, the entries of G are $(0,\rho)$-Gaussians. The following probabilistic results are drawn from [15], but we state the results in a more general manner, in order to detach the analysis from any particular instantiation of parameters.

Lemma 8.3 (Estimate for (B.1) in [15])**.**
Let $a_{n+1},\ldots,a_{n+d}$ and z_0 be the output of `Adding Constraints`, then

$$\operatorname{Prob}_{G,Q}\left\{z_0 \in \operatorname{cone}\{a_{n+1},\ldots,a_{n+d}\}\right\} \geq 1 - \exp\left(-\frac{1}{2}\rho^{-2}\kappa^{-2}s_{\max}^{-2}\left(T^{-1}\right)\right),$$

where $\kappa \in \mathbb{R}^+$ satisfies

$$\kappa^{-1} \leq \left(1 + 2d\cdot s_{\max}\left(T^{-1}\right)\right)^{-1}.$$

Proof. The statement whose probability we estimate is invariant under scaling and orthogonal transformations. Therefore it is sufficient to derive a lower bound for

$$\operatorname{Prob}_G\left\{w \in \operatorname{cone}\{v_1,\ldots,v_d\}\right\}.$$

Invertibility of T is known, and $T+G$ is invertible almost surely. We condition on T and $T+G$ being invertible. This implies that there exist real numbers $c_1,\ldots,c_d$ and $\bar{c}_1,\ldots,\bar{c}_d$ such that

$$w = c_1v_1 + \cdots + c_dv_d, \quad w = \bar{c}_1\bar{v}_1 + \cdots + \bar{c}_d\bar{v}_d.$$

Since $\bar{v}_i = Te_i$ and $v_1 = (T+G)e_i$ for $1 \leq i \leq d$, we see

$$(T+G)^{-1}w = c_1e_1 + \cdots + c_de_d, \quad T^{-1}w = \bar{c}_1e_1 + \cdots + \bar{c}_de_d.$$

In particular,

$$c_i = \langle (T+G)^{-1}w, e_i\rangle, \quad \bar{c}_i = \langle T^{-1}w, e_i\rangle = 1/d.$$

So it suffices to bound $\|T^{-1} - (T+G)^{-1}\|$ from above. Since T and $T+G$ are invertible,

$$\begin{aligned} T^{-1} - (T+G)^{-1} &= (T+G)^{-1}(T+G)(T^{-1} - (T+G)^{-1}) \\ &= (T+G)^{-1}(1 + GT^{-1} - 1) \end{aligned}$$

$$= (T+G)^{-1}GT^{-1}$$
$$= (1+T^{-1}G)^{-1}T^{-1}GT^{-1}$$

holds, and so

$$\|T^{-1}-(T+G)^{-1}\| = \|(1+T^{-1}G)^{-1}T^{-1}GT^{-1}\| \leq \|(1+T^{-1}G)^{-1}\| \cdot \|T^{-1}G\| \cdot \|T^{-1}\|.$$

Next, we recall that

$$\begin{aligned}\mathrm{Prob}_G\left\{\|T^{-1}G\| < \kappa^{-1}\right\} &\geq \mathrm{Prob}_G\left\{\|G\| < \kappa^{-1}s_{\max}^{-1}(T^{-1})\right\}\\ &\geq \mathrm{Prob}_G\left\{\|\rho^{-1}G\| < \kappa^{-1}\rho^{-1}s_{\max}^{-1}(T^{-1})\right\}\\ &\geq 1-\exp\left(-\frac{1}{2}\rho^{-2}\kappa^{-2}s_{\max}^{-2}(T^{-1})\right).\end{aligned}$$

via Corollary 6.5. Provided that $\kappa^{-1} < 1$, we have $\|T^{-1}G\| < 1$ with high probability, so the Neumann series identity ([18, Satz II.1.11]) is applicable:

$$(1-\|T^{-1}G\|)^{-1} \geq \sum_{p=0}^{\infty}\|T^{-1}G\|^p \geq \left\|\sum_{p=0}^{\infty}(-T^{-1}G)^p\right\| = \left\|(1+T^{-1}G)^{-1}\right\|.$$

We can then derive

$$\|(1+T^{-1}G)^{-1}\| \cdot \|T^{-1}G\| \cdot \|T^{-1}\| \leq \frac{\|T^{-1}G\|\|T^{-1}\|}{1-\|T^{-1}G\|} \leq \frac{s_{\max}(T^{-1})\kappa^{-1}}{1-\kappa^{-1}} = \frac{s_{\max}(T^{-1})}{\kappa-1}.$$

The last term is bounded from above by $(2d)^{-1}$, provided that T satisfies

$$\kappa^{-1} \leq \left(1+2d\cdot s_{\max}(T^{-1})\right)^{-1} < 1.$$

This implies the desired result. □

Lemma 8.4 (Estimate for (B.2) in [15])**.**
Let $a_{n+1},\ldots,a_{n+d}$ and z_0 be the output of `Adding Constraints`, then

$$\mathrm{Prob}_{G,Q}\left\{\mathrm{dist}(\vec{0},\mathrm{aff}\{a_{n+1},\ldots,a_{n+d}\}) \geq \zeta\right\} \geq 1-\exp\left(-\frac{1}{2}\rho^{-2}\kappa^{-2}s_{\max}^{-2}(T^{-1})\right),$$

where $\kappa \in \mathbb{R}^+$ satisfies

$$\kappa^{-1} \leq \left(1+2d\cdot s_{\max}(T^{-1})\right)^{-1}.$$

Proof. The result whose probability we want to estimate is invariant under scaling und orthogonal transformations. Therefore if suffices to estimate

$$\mathrm{Prob}_G\left\{\mathrm{dist}(0,\mathrm{aff}\{v_1,\ldots,v_d\}) \geq \frac{1}{2}\right\}.$$

Recall that $v_1,\ldots,v_d$ have standard deviation ρ and centers of norm $\bar{v}_{n+1},\ldots,\bar{v}_{n+d}$. Let $h \in \mathbb{R}^d$ be the defining normal to $\mathcal{H} := \mathrm{aff}\{v_1,\ldots,v_d\}$ and let $\bar{h} \in \mathbb{R}^d$ be the defining normal to $\overline{\mathcal{H}} := \mathrm{aff}\{\bar{v}_1,\ldots,\bar{v}_d\}$; this means that $x \in \mathcal{H}$ if and only if $\langle x,h\rangle = 1$ and $x \in \overline{\mathcal{H}}$ if and only if $\langle x,\bar{h}\rangle = 1$. We then know by linear algebra that

$$\bar{h} = (T^*)^{-1}\vec{1}, \quad h = ((T+G)^*)^{-1}\vec{1},$$

as follows from

$$1 = \langle \vec{1}, e_i \rangle = \langle \vec{1}, T^{-1}\bar{v}_i \rangle = \langle (T^*)^{-1}\vec{1}, \bar{v}_i \rangle,$$
$$1 = \langle \vec{1}, e_i \rangle = \langle \vec{1}, (T+G)^{-1}v_i \rangle = \langle (T^*+G^*)^{-1}\vec{1}, v_i \rangle.$$

We can then estimate

$$\|h - \bar{h}\| \leq \|((T+G)^*)^{-1} - (T^*)^{-1}\| \cdot \|\vec{1}\| = \|(T+G)^{-1} - T^{-1}\| \cdot \|\vec{1}\|.$$

Using exactly the same argument as in the previous proof, we obtain that with high probability we have $\|(T+G)^{-1} - T^{-1}\| \leq \frac{1}{2}d^{-1}$, so with high probability

$$\|h - \bar{h}\| \cdot \|\vec{1}\| \leq \frac{1}{2}d^{-\frac{1}{2}}.$$

Since $\bar{h} = w$ has norm 1, with high probability h has norm near 1. This completes the proof. □

Remark 8.5.
The parameter κ depends on ρ, d and the spectrum of T. The latter depends on the parameter l, which in turn will be chosen depending on ρ, n and d. In Section 11, a possible choice of κ is examined.

The previous two theorems allow us to estimate the probability that `Adding Constraints` does not return failure. It is relevant for applications of `Adding Constraints` that the output vectors $a_{n+1}, \ldots, a_{n+d}$ are located within some (fixed) halfspace.

Lemma 8.6 (Estimate for (B.3) in [15])**.**
Let $\mathcal{H}$ be a halfspace. Then

$$\begin{aligned}&\mathrm{Prob}_{a_{n+1},\ldots,a_{n+d}}\{a_{n+1}, \ldots, a_{n+d} \in \mathcal{H}\}\\ &\geq \frac{|\mathcal{S}^{d-1}(h)|}{|\mathcal{S}^{d-1}|} - \frac{d\rho}{2\tau}\exp\left(-\frac{\tau^2}{2\rho^2}\right) - d\frac{|\mathcal{S}^{d-2}(\delta/l)|}{|\mathcal{S}^{d-2}|},\end{aligned}$$

provided that $h \in [0,1]$, $\tau \in (0,\infty)$, $\delta \in [0,l]$, and $h - \tau - \delta \geq 0$.

Remark 8.7.
The parameters h, τ and δ must be chosen with dependence on d and ρ. Upon appropiate choice of ρ with dependence on d and n only, the parameters can be chosen with dependence only on d in the context of our application.

Proof. Let $\nu \in S^{d-1}$ such that $\mathcal{H} = \{x \in \mathbb{R}^d \mid \langle \nu, x \rangle \geq 0\}$. We want to estimate the probability

$$\mathrm{Prob}_{a_{n+1},\ldots,a_{n+d}}\left\{\langle a_{n+1}, \nu\rangle, \ldots, \langle a_{n+d}, \nu\rangle \in \mathbb{R}_0^+\right\}.$$

With abuse of notation, we may neglect the successive scaling of z_0, $\bar{a}_{n+i}$ and a_{n+i} by the factor 2ζ as in `Adding Constraints`. We split the products

$$\langle \nu, a_{n+i} \rangle = \langle \nu, z_0 \rangle + \langle \nu, a_{n+i} - \bar{a}_{n+i} \rangle + \langle \nu, \bar{a}_{n+i} - z_0 \rangle, \tag{12}$$

and show that these terms are contained within suitable ranges with high probability.

First, since z_0 is the uniformly distributed random variable in $\mathcal{S}^{d-1}$, the musings of Section 6 state that

$$\mathrm{Prob}_Q\{\langle z_0, \nu \rangle \geq h\} = \frac{|\mathcal{S}^{d-1}(h)|}{|\mathcal{S}^{d-1}|},$$

where $h \in [0,1]$ and $\mathcal{S}^{d-1}(h) \subset S^{d-1}$ is the spherical cap suspended from e_1 with height h above the plane orthogonal to e_1.

Second, we observe that the vectors $a_{n+i} - \bar{a}_{n+i}$ are Gaussian variables with centers at the origin and standard deviation ρ. According to Lemma 6.1, the product $\langle a_{n+i} - \bar{a}_{n+i}, \nu\rangle$ is a Gaussian variable on $\mathbb{R}$ centered at 0 with standard deviation ρ. Since

$$\mathrm{Prob}_{A,Q}\left\{-\langle a_{n+i} - \bar{a}_{n+i}, \nu\rangle > \tau\right\} \leq \frac{\rho}{2\tau}\exp\left(-\frac{\tau^2}{2\rho^2}\right)$$

according to Lemma 6.6, we obtain the estimate

$$\begin{aligned}
&\mathrm{Prob}_{A,Q}\left\{\min_{1\leq i\leq d}\langle a_{n+i} - \bar{a}_{n+i}, \nu\rangle > -\tau\right\}\\
&\geq \mathrm{Prob}_{A,Q}\left\{-\min_{1\leq i\leq d}\langle a_{n+i} - \bar{a}_{n+i}, \nu\rangle < \tau\right\}\\
&\geq \mathrm{Prob}_{A,Q}\left\{\max_{1\leq i\leq d}\langle a_{n+i} - \bar{a}_{n+i}, -\nu\rangle < \tau\right\}\\
&= 1 - \sum_{i=1}^{d}\mathrm{Prob}_{A,Q}\left\{-\langle a_{n+i} - \bar{a}_{n+i}, \nu\rangle > \tau\right\} \geq 1 - \frac{d\rho}{2\tau}\exp\left(-\frac{\tau^2}{2\rho^2}\right).
\end{aligned}$$

The third term of Equation (12) is the most complicated to estimate. We decompose $Q = VW$, where W is a random variable in $O(d)$ and V is random variable in the subgroup of $O(d)$ whose elements leave the space $\mathrm{lin}(Ww)$ invariant. Let $l_i = W(\bar{v}_i - w)$, let P be the orthogonal projection onto $(Ww)^\perp$, and let ν' be the normalization of $P\nu$. We verify

$$\langle l_i, Ww\rangle = \langle W(\bar{v}_i - w), Ww\rangle = \langle \bar{v}_i - w, w\rangle = 0,$$

by the definition of w and $\bar{v}_i$. We conclude that $Vl_i \perp Ww$, and that

$$|\langle \nu, \bar{a}_{n+i} - z_0\rangle| = |\langle \nu, Vl_i\rangle| = |\langle P\nu, Vl_i\rangle| = |\langle V^*P\nu, l_i\rangle| \leq |\langle V^*\nu', l_i\rangle|.$$

We have $\|l_i\| = l$, and without loss of generality we may apply a coordinate transformation such that the l_i are vectors in $\mathbb{R}^{d-1} \subset \mathbb{R}^d$ of norm l, and $u := V^*\nu'$ is a uniformly distributed random vector on the 1-sphere of $\mathbb{R}^{d-1}$. We estimate

$$\begin{aligned}
&\mathrm{Prob}_u\left\{\max_{1\leq i\leq d}|\langle \bar{a}_{n+i} - z_0, \nu\rangle| \leq \delta\right\} \geq \mathrm{Prob}_u\left\{\max_{1\leq i\leq d}|\langle u, l_i\rangle| \leq \delta\right\}\\
&= 1 - \sum_{i=1}^{d}\mathrm{Prob}_u\left\{|\langle u, l_i\rangle| \geq \delta\right\} = 1 - d\,\mathrm{Prob}_u\left\{|\langle u, e_1\rangle| \geq \frac{\delta}{l}\right\} = 1 - d\frac{|S^{d-2}(\delta/l)|}{|S^{d-2}|},
\end{aligned}$$

provided that $\delta \in [0, l]$. With these estimates, we can finally derive

$$\begin{aligned}
&\mathrm{Prob}_{a_{n+1},\ldots,a_{n+d}}\left\{\langle a_{n+1}, \nu\rangle, \ldots, \langle a_{n+d}, \nu\rangle \in \mathbb{R}_0^+\right\}\\
&\geq \mathrm{Prob}_{a_{n+1},\ldots,a_{n+d}}\left\{\begin{array}{c}\langle z_0, \nu\rangle \geq h,\\ \forall 1 \leq i \leq d : \langle a_{n+i} - \bar{a}_{n+i}, \nu\rangle > -\tau,\\ \forall 1 \leq i \leq d : |\langle \bar{a}_{n+i} - z_0, \nu\rangle| \leq \delta\end{array}\right\}\\
&\geq \frac{|\mathcal{S}^{d-1}(h)|}{|\mathcal{S}^{d-1}|} - \frac{d\rho}{2\tau}\exp\left(-\frac{\tau^2}{2\rho^2}\right) - d\frac{|\mathcal{S}^{d-2}(\delta/l)|}{|\mathcal{S}^{d-2}|}.
\end{aligned}$$

This completes the proof. □

Remark 8.8.
As one might reasonably expect, not only are the events mentioned in Theorem 8.3 and Theorem 8.4 invariant under scaling, but so are the lower bounds, too. This can be verified by tracking the effects of scaling through the respective proofs. The scaling invariance of the event in Theorem 8.6 is obvious, which is not as easy for the lower bound. However, scaling the random vectors $a_{n+1}, \ldots, a_{n+d}$ by a factor α leads to the emergence of rescaled parameters $\frac{h}{\alpha}$, $\frac{\tau}{\alpha}$ and $\frac{\delta}{\alpha}$. The parametrized set of lower bounds remains the same.

The previous three results serve as building blocks to eventually estimate the probability of success for `Adding Constraints`.

Theorem 8.9.
Let $\mathcal{H}$ be a halfspace. Then `Adding Constraints` returns z_0, $a_{n+1}, \ldots, a_{n+d}$ such that $a_{n+1}, \ldots, a_{n+d} \in \mathcal{H}$ with probability at least

$$\frac{|\mathcal{S}^{d-1}(h)|}{|\mathcal{S}^{d-1}|} - \frac{d\rho}{2\tau} \exp\left(-\frac{\tau^2}{2\rho^2}\right) - d\frac{|\mathcal{S}^{d-2}(\delta/l)|}{|\mathcal{S}^{d-2}|} - 2\exp\left(-\frac{1}{2}\rho^{-2}\kappa^{-2}s_{\max}^{-2}(T^{-1})\right)$$

where $h \in [0,1]$, $\tau \in (0,\infty)$, $\delta \in [0,l]$, $h - \tau - \delta \geq 0$, and $\kappa \in \mathbb{R}$ such that

$$\kappa^{-1} \leq \left(1 + 2d \cdot s_{\max}(T^{-1})\right)^{-1}.$$

9 Phase I Method

Each variant of the simplex method requires an initial vertex of the input problem and possible additional initial information, which entails the need to construct a so-called Phase I method. The Phase I algorithm of the shadow vertex algorithm according to Vershynin's work constructs a randomized sequence of linear programming problems by employing the algorithm `Adding Constraints`. The output of each call of `Adding Constraints` is an augmented system matrix $\widehat{A}$ and an initial direction z_0, such that the indices of the additional constraint describe a vertex that maximizes z_0 over $\left[\widehat{A}, \vec{1}\right]$. With sufficient probability, as remains to be shown in Section 11, the shadow vertex method applied to the linear programming problem provides an optimal solution of $\left(A, \vec{1}, z\right)$. The Phase I method, algorithm `Unit Solver`, is therefore a Las-Vegas-method for solving unit linear programming problems. In this section, we describe `Unit Solver`, inspect its correctness and provide probabilistic estimates for its running time. Again, the run-time estimate is parametrized, to be instantiated in Section 11.

Let us consider the following unit linear programming problem

$$\text{Maximize } \langle z, x\rangle \quad \text{subject to} \quad Ax \leq \vec{1}. \tag{Unit LP}$$

Let the augmented matrix $\widehat{A} \in \mathbb{R}^{(n+d)\times d}$ be the extension of the matrix $A \in \mathbb{R}^{d\times d}$ by the additional constraint row vectors. This induces an augmented linear programming problem

$$\text{Maximize } \langle z, x\rangle \quad \text{subject to} \quad \widehat{A}x \leq \vec{1}, \tag{Aug Unit LP}$$

which serves as an auxiliary problem. Let us denote its underlying polyhedron by $\widehat{\mathcal{P}}$ and its dual polyhedron by $\left(\widehat{\mathcal{P}}\right)^*$. Our first concern is which rows can be appended to A such that (Unit LP) and (Aug Unit LP) are equivalent. Our second concern is whether their equivalence can be checked algorithmically, if any set of additional rows is given.

We call $\mathcal{N} \subset \mathbb{R}^d$ a numb set of the linear programming problem (Unit LP) if imposing any finite number of additional constraint vectors to (Unit LP) gives a new linear programming problem equivalent to the (Unit LP).

Lemma 9.1.
The numb set of a unit linear programming problem $(A, \vec{1}, z)$ always contains a closed halfspace.

Proof. If $(A, \vec{1}, z)$ is infeasible, then the statement is trivial. Suppose that $(A, \vec{1}, z)$ is feasible and bounded, and that $I \subseteq [n]$ describes a solution. Then $z \in \mathcal{C}_I$. Let $\mathcal{G}$ be the unique halfspace that contains $\mathcal{P}^*$ and has $\mathcal{F}_I$ in its boundary. Then $\mathcal{G}$ is a numb set. Suppose that $(A, \vec{1}, z)$ is feasible but unbounded, which means that there exists a hyperplane through the origin that separates $\mathbb{R}^+ \cdot z$ from $\mathcal{P}^*$. The closed halfspace containing $\mathcal{P}^*$ is a numb set. □

The reader might have noted that in the case of a bounded feasible program, constraints from the numb halfspace added may lead to vertices in non-general position. This event happens almost never in our algorithmic construction. Anyways, the solution theory and the correctness of the algorithms are not critically affected by this. This is important for the following result.

Lemma 9.2.
Assume that (Aug Unit LP) is in general position. The linear programming problem (Aug Unit LP) is equivalent to (Unit LP) if and only if either (Aug Unit LP) is unbounded or if (Aug Unit LP) has a solution that satisfies d different constraint vectors of A with equality.

Proof. We note that both linear programming problems are trivially feasible, since 0 is a feasible point in both cases. If (Aug Unit LP) is unbounded, then (Unit LP) is unbounded, too. Otherwise, suppose that (Aug Unit LP) is bounded with optimal solution set I. If $I \subseteq [n]$, then there is nothing to show. Otherwise, the optimal value of (Aug Unit LP) is almost surely strictly smaller than the optimal value of (Unit LP), since almost surely no two vertices lie in the same level set of z, which means that the solution space is different. □

We now provide the pseudo-code for algorithm `Unit Solver`, see Algorithm 3, and inspect its correctness and smoothed complexity.

Input: $A \in \mathbb{R}^{n \times d}$, $z \in \mathbb{R}^d$, $\rho \in \mathbb{R}^+$, $l \in \mathbb{R}^+$.

Output: Either 'unbounded' or the index set I of an optimal solution of $(A, \vec{1}, z)$.

Algorithm:

(I) Apply `Adding Constraints` with parameters μ_0, ρ and l, where

$$\mu_0 = \exp \operatorname{ceil} \log \max_{1 \leq i \leq n} \|a_i\|,$$

If 'failure' is returned, then repeat. Otherwise, keep the output vector z_0 and the augmented matrix $\widehat{A}$.

(II) Apply the polar shadow vertex method with initial direction z_0, initial solution $\{n+1, \ldots, n+d\}$ and objective direction z to $(\widehat{A}, \vec{1}, z)$.

(III) If 'unbounded' is returned, then return 'unbounded'. If a solution is returned that contains an index among $\{n+1, \ldots, n+d\}$, then go to Step (I). Otherwise, return the solution I.

Algorithm 3: Randomized solver for unit linear programming problems

Theorem 9.3.
`Unit Solver` is correct.

Proof. There is a fixed non-zero probability that z_0 and $\widehat{A}$ are returned in Step (I), so almost surely, after a finite number of iterations of Step (I), it finds z_0 and $\widehat{A}$. The application of the shadow vertex method in Step (II) is almost surely well-defined, because almost surely z and z_0 are linearly independent. There is a fixed non-zero probabibility that z_0 and the additional constraint vectors lie within a numb halfspace of the original problem. So almost surely, after a finite number of iterations of Steps (I) – (III), the output of Step (II) is not an index set containing indices from $[n+d]-[n]$. If the output is an index set $I \subseteq [n]$, then a solution has been found. If 'unbounded' is found, then the original is unbounded. □

On the one hand, we want to estimate the number of iterations of the main loop, and on the other hand we need to estimate the number of pivot steps in each iteration. Let $F(k)$ be the binary random variable in $\{0,1\}$ that is 1 if and only if either `Adding Constraints` returns 'failure' in the k-th iteration or (Aug Unit LP) in the k-th iteration is not equivalent to (Unit LP). Let $T(k)$ be the random variable in $\mathbb{N}_0$ that describes the number of pivot steps in the call of the shadow vertex method in the k-th iteration. Then the random variable V, defined by

$$T_I := \sum_{k=1}^{\infty} T(k) \prod_{j=1}^{k-1} F(j), \tag{13}$$

is the number of pivot steps in the call of Algorithm 3 until the first success is made. Our aim is to derive its expected value.

The random variables about which we take the expected value are the input matrix A and the random choices $\mathfrak{C}_i$ in each i-th iteration, thus

$$\mathbb{E}_{A,\mathfrak{C}_1,\mathfrak{C}_2,\ldots} T_I = \mathbb{E}_A \sum_{k=1}^{\infty} \mathbb{E}_{\mathfrak{C}_k} T(k) \prod_{j=1}^{k-1} \mathbb{E}_{\mathfrak{C}_j} F(j).$$

The expected values $\mathbb{E}_{A,\mathfrak{C}_k} T(k)$ are all independent of k, so we may write $\mathbb{E}_{A,\mathfrak{C}_k} T(k) = \mathbb{E}_{A,\mathfrak{C}_k} T(1)$. Let $p \in [0,1)$ be an upper bound independent from A for the probability that the algorithm `Adding Constraints` returns 'failure' or that (Aug Unit LP) is not equivalent to (Unit LP). With the elementary fact on the geometric series we find

$$\begin{aligned}\mathbb{E}_{A,\mathfrak{C}_1,\mathfrak{C}_2,\ldots} T_I = \mathbb{E}_A \sum_{k=1}^{\infty} \mathbb{E}_{\mathfrak{C}_k} T(k) \prod_{j=1}^{k-1} p &\leq \mathbb{E}_{\widehat{A}_1} T(1) \cdot \sum_{k=1}^{\infty} p^{k-1} \\ &= \mathbb{E}_{\widehat{A}_1} T(1) \cdot \sum_{k=0}^{\infty} p^k = \frac{\mathbb{E}_{\widehat{A}_1} T(1)}{1-p}.\end{aligned}$$

An explicit instantion for p and a bound for $\mathbb{E}_{\widehat{A}_1} T(1)$ is provided in Section 11

10 Interpolation of Linear Programming Problems

The shadow vertex simplex method, Algorithm 1, applies to unit linear programming problems in general position. Employing the algorithm indirectly for general linear programming

problems not only requires the construction of auxiliary problems with unit right-hand side, but also should the construction preserve stochastic properties of the input. We describe such a construction in this section. Recall that our initial task is to find the optimal solution of the linear programming problem

$$\text{Maximize} \quad \langle z, x\rangle \quad \text{subject to} \quad Ax \leq b. \tag{LP}$$

We consider the corresponding unit linear linear programming problem

$$\text{Maximize} \quad \langle z, x\rangle \quad \text{subject to} \quad Ax \leq \vec{1}. \tag{LP Unit}$$

for auxiliary purposes because its solution properties are easier to comprehend. We relate (LP) and (LP Unit) through a third problem. The interpolation linear programming problem depends on a parameter λ and is given by

$$\text{Maximize} \quad \langle z, x\rangle + \lambda \cdot \tau \quad \text{subject to} \quad Ax \leq \tau b + (1-\tau)\vec{1}, \quad 0 \leq \tau \leq 1. \tag{LP Int}$$

We write $\mathcal{P}_{[0,1]}$ for the underlying polyhedron, called interpolation polyhedron. Occassionally, we might fix the variable τ with some value t. This yields the linear programming problem

$$\text{Maximize} \quad \langle z, x\rangle + \lambda \cdot \tau \quad \text{subject to} \quad Ax \leq \tau b + (1-\tau)\vec{1}, \quad \tau = t, \tag{LP Int t}$$

and we write $\mathcal{P}_t$ for its polyhedron.

The following results rigorously relate and describe (LP), (LP Unit) and (LP Int). Theorems 10.1, 10.2 and 10.4, and Corollary 10.3 can be found in Appendix A of [15]. We only give some minor alterations. For notational brevity, we write

$$\beta := \max_{\tau\in[0,1]} \|\tau\vec{1} - (1-\tau)b\|_\infty \in [1,\infty), \quad \mu(\tau) := \sup_{x\in\mathcal{P}_\tau} \langle z, x\rangle \in [-\infty, \infty].$$

Theorem 10.1 ([15, Proposition 4.1, (i)])**.**
The following statements are equivalent

- (LP) is unbounded.
- (LP Unit) is unbounded.
- (LP Int) is unbounded for all λ.
- (LP Int) is unbounded for some λ.

Proof. We know by Theorem 5.2 that (LP) is bounded if and only if (LP Unit) is bounded.

Assume that (LP Unit) is unbounded. Since $\mathcal{P}_0$ contains the feasible set of (LP Unit), this set is contained in the feasible set of (LP Int), hence (LP Int) is unbounded for every λ, too.

Now assume that (LP Int) is unbounded for some λ, i.e., there exists a sequence (x_i, τ_i) of feasible points such that $\langle (x_i, \tau_i), (z, \lambda)\rangle \to \infty$. Because $0 \leq \tau_i \leq 1$, we have

$$Ax_i \leq \beta\vec{1} \quad \Longleftrightarrow \quad \beta^{-1}Ax_i \leq \vec{1}.$$

This shows that $\beta^{-1}x_i$ is a sequence of feasible vectors for (LP Unit). But we also observe that $\langle z, \beta^{-1}x_i\rangle \to \infty$ because $0 \leq \tau_i \leq 1$.

This completes the proof. □

Theorem 10.2 ([15, Proposition 4.1, (iii)])**.**
Assume that (LP) is not unbounded. Then (LP) is feasible if and only if for λ sufficiently large (LP Int) has an optimal solution with $\tau = 1$. Furthermore, $x \in \mathbb{R}^d$ is an optimal solution of (LP) if and only if for λ sufficiently large (LP Int) has an optimal solution $(x, 1)$.

Proof. Suppose that for λ sufficiently large (LP Int) has an optimal solution $(x, 1)$, then $(x, 1) \in \mathcal{P}_1$ and x is an optimal solution of (LP). In particular, (LP) is feasible.

Suppose that (LP) is feasible. We see that for $x \in \mathcal{P}_1$ we have $\tau x \in \mathcal{P}_\tau$, so (LP_τ) is feasible for $\tau \in [0, 1]$, so their optimal values are larger than $-\infty$. By assumption and Theorem 10.1, we know that (LP Unit) is not unbounded. Because $\vec{0} \in \mathcal{P}_0$, we infer that (LP Unit) is feasible and bounded. Note that $\mu(0) \geq 0$. Since we have $(\beta^{-1}x) \in \mathcal{P}_0$ for $x \in \mathcal{P}_\tau$, we derive

$$\forall \tau \in [0, 1] : |\mu(\tau)| = \left| \sup_{x \in P_\tau} \langle z, x \rangle \right| = \beta \left| \sup_{x \in P_\tau} \langle z, \beta^{-1}x \rangle \right| \leq \beta \left| \sup_{x \in P_0} \langle z, x \rangle \right| = \beta\mu(0).$$

Thus, (LP Int) is bounded. Let $(x_\lambda, \tau_\lambda)$ optimally solve (LP Int) with some parameter $\lambda > 0$. This implies

$$\tau_\lambda \in \operatorname{argmax}_{\tau \in [0,1]} \max_{x \in \mathcal{P}_\tau} \langle (z, \lambda), (x, \tau_\lambda) \rangle \quad \Longleftrightarrow \quad \tau_\lambda \in \operatorname{argmax}_{\tau \in [0,1]} \lambda^{-1}\mu(\tau) + \tau.$$

We also know that

$$\left| (\lambda^{-1}\mu(\tau) + \tau) - \tau \right| = \lambda^{-1}|\mu(\tau)| \leq \beta\lambda^{-1}\mu(0).$$

Because τ_λ maximizes we find

$$\begin{aligned} \tau_\lambda &\geq \left(\lambda^{-1}\mu(\tau_\lambda) + \tau_\lambda\right) - \beta\lambda^{-1}\mu(0) \\ &\geq \left(\lambda^{-1}\mu(1) + 1\right) - \beta\lambda^{-1}\mu(0) \\ &\geq 1 - 2\beta\lambda^{-1}\mu(0). \end{aligned}$$

We have shown that an optimal solution $(x_\lambda, \tau_\lambda)$ of (LP Int) with parameter λ satisfies $\tau_\lambda \to 1$ as $\lambda \to \infty$. But we may assume without loss of generality that there are only finitely many different points in the family $(x_\lambda, \tau_\lambda)$. It follows that $\tau_\lambda = 1$ for all sufficiently large λ. This completes the proof. □

Corollary 10.3 ([15, Proposition 4.1, (iv)])**.**
Assume that (LP) is feasible and bounded. Then $x \in \mathbb{R}^d$ is an optimal solution of (LP) if and only if $(x, 1)$ is an optimal solution of (LP Int) for all sufficiently large λ.

Theorem 10.4 ([15, Proposition 4.1, (ii)])**.**
Assume that (LP)) is not unbounded. Then $x \in \mathbb{R}^d$ is an optimal solution of (LP Unit) if and only if for λ sufficiently small (LP Int) has an optimal solution $(x, 0)$.

Proof. Suppose that for λ sufficiently small (LP Int) has an optimal solution $(x, 0)$, then $(x, 0) \in \mathcal{P}_0$ and x is an optimal solution of (LP Unit).

By assumption, (LP Unit) is feasible and bounded. As $P_\tau = \emptyset$ is possible for general $\tau \in [0, 1]$, we introduce $\mathcal{T} \subset [0, 1]$, where $\tau \in \mathcal{T}$ if and only if $P_\tau \neq \emptyset$. Since we have $\beta^{-1}x \in \mathcal{P}_0$ for $x \in \mathcal{P}_\tau$, we derive

$$\forall \tau \in \mathcal{T} : |\mu(\tau)| = \left| \sup_{x \in P_\tau} \langle z, x \rangle \right| = \beta \left| \sup_{x \in P_\tau} \langle z, \beta^{-1}x \rangle \right| \leq \beta \left| \sup_{x \in P_0} \langle z, x \rangle \right| = \beta\mu(0).$$

This implies that (LP Int) is bounded.

Suppose that (LP Int) has an optimal solution. Let $(x_\lambda, \tau_\lambda)$ optimally solve (LP Int) with some parameter $\lambda < 0$. This implies

$$\tau_\lambda \in \operatorname{argmax}_{\tau\in\mathcal{T}} \max_{x\in\mathcal{P}_\tau} \langle (z,\lambda),(x,\tau_\lambda)\rangle \quad\Longleftrightarrow\quad \tau_\lambda \in \operatorname{argmax}_{\tau\in\mathcal{T}} |\lambda|^{-1}\mu(\tau) - \tau.$$

We also know that

$$\big|-\big(|\lambda|^{-1}\mu(\tau)-\tau\big)-\tau\big| = \big|\big(|\lambda|^{-1}\mu(\tau)-\tau\big)+\tau\big| = |\lambda|^{-1}|\mu(\tau)| \le \beta|\lambda|^{-1}\mu(0).$$

Since τ_λ is a maximizer,

$$\begin{aligned}\tau_\lambda &\le -\big(|\lambda|^{-1}\mu(\tau_\lambda)-\tau_\lambda\big)+\beta|\lambda|^{-1}\mu(0)\\ &\le -\big(|\lambda|^{-1}\mu(0)-0\big)+\beta|\lambda|^{-1}\mu(0)\\ &\le 0+2\beta|\lambda|^{-1}\mu(0).\end{aligned}$$

We have shown that an optimal solution (x_λ,τ_λ) of (LP Int) with parameter λ satisfies $\tau_\lambda \to 0$ as $\lambda\to-\infty$. But we may assume without loss of generality that there are only finitely many different points in the family (x_λ,τ_λ). It follows that $\tau_\lambda = 0$ for all sufficiently small λ. This completes the proof. □

Remark 10.5.
The optimal solutions of (LP Int) for $\lambda\to-\infty$ and of (LP Int) for $\lambda\to+\infty$ are optimal solutions for the directions $(0,-1)$ and $(0,1)$, respectively. But the converse implication does not hold, because any vertex of $\mathcal{P}_0$ and $\mathcal{P}_1$ is an optimal solution for $(0,-1)$ and $(0,1)$, respectively. This is related to the subspace $(0,\mathbb{R})$ being orthogonal to $\mathcal{P}_0$ and $\mathcal{P}_1$. We conclude that the optimal solution of the limit problems depends on z. This has implications for the choice of the plane of rotation in Phase II of the full solution algorithm.

Remark 10.6.
Suppose that the linear programming problem (LP) has Gaussian entries in A and b. The previous reduction of (LP) to a unit problem is ostensibly complicated – indeed, one might object that an instance has no zero entries in b with probability 1, so a trivial renormalization would suffice to derive a unit problem. We appeal to the reader to recall that the system matrix of that unit problem then has rows distributed by a ratio distribution. It seems difficult to estimate $\mathcal{S}(\mathcal{Q},\mathcal{E})$ if $\mathcal{Q}$ is the convex closure of random variables having such a distribution.

11 The Full Solution Algorithm

In this section we eventually describe a randomized variant of the simplex algorithm and estimate its smoothed complexity. We first provide a reduction from general linear programming problems to the interpolation linear programming problem (LP Int). We prove correctness of the method, which employs the previously discussed subroutines `Shadow Vertex Simplex Method`, `Unit Solver`, and, indirectly, `Adding Constraints`. The smoothed complexity estimate reduces to estimating the number of pivot steps taken within all calls of Subroutine 1. The latter estimates depend on two results on the sizes of two-dimensional shadows of polyhedrons, which are proven in the remaining sections of this thesis.

11.1 Proof of Correctness

We recall the definition of the interpolated linear programming problem:

$$\text{Maximize} \quad \langle z, x\rangle + \lambda \cdot \tau \quad \text{subject to} \quad \begin{array}{c} Ax - \tau(\vec{1} - b) \leq \vec{1}, \\ 0 \leq \tau \leq 1. \end{array} \tag{LP Int}$$

and the linear programming problems (LP) and (LP Unit). Algorithm `Full Solution Algorithm` formally uses (LP Int), but we need to transform it into a unit linear programming problem. We introduce a parameter $\gamma > 0$ and a perturbed interpolated problem:

$$\text{Maximize} \quad \langle z, x\rangle + \lambda \cdot \tau \quad \text{subject to} \quad \begin{array}{c} Ax - \tau(\vec{1} - b) \leq \vec{1}, \\ -\gamma \leq \tau \leq 1. \end{array} \tag{LP Int'}$$

We write $\mathcal{P}_{[-\gamma,1]}$ for the underlying polyhedron. The vertices of the interpolated polyhedron are either located on the planes $\{t = 0\}$ or $\{t = 1\}$, or have positive distance from these planes. Similarly, we infer that there exists $\gamma > 0$ such that the limit solutions of (LP Int') and (LP Int) are the same. The transformation into a unit linear programming problem is now obvious:

$$\text{Maximize} \quad \langle z, x\rangle + \lambda \cdot \tau \quad \text{subject to} \quad \begin{array}{c} Ax - \tau(\vec{1} - b) \leq \vec{1}, \\ -\frac{1}{\gamma}\tau \leq 1, \quad \tau \leq 1. \end{array} \tag{LP Int"}$$

Suppose we are given an arbitrary linear programming problem (A, b, z). The pseudo-code for `Full Solution Algorithm` is given by Algorithm 4. Step (I) describes Phase I of the algorithm, while Step (II) and Step (III) formalize Phase II. Notably, Step (II) describes the rotation of the search direction q from $(0, -1)$ to $(0, 1)$ in the plane $\mathcal{E} = \text{lin}\{(0, 1), (z, 0)\}$. This process of rotation is formally split up into two parts, because $(0, 1)$ and $(0, -1)$ are colinear, but we refer to it as a single rotation in the sequel. Furthermore, we ignore the difference between $\mathcal{P}_{[-\gamma,1]}$ and $\mathcal{P}_{[0,1]}$ formally. Alternatively to computing a suitable γ, one may modify the beginning of the shadow vertex method in Phase II to ignore the single 0 in the right-hand side.

We first prove that this algorithm is well-posed and returns the correct result.

If `Unit Solver` returns 'unbounded', then Theorem (10.1) implies unboundedness of (LP). Otherwise `Unit Solver` produces an index set $I_0 \subseteq [n]$ which describes an optimal solution of (LP Unit). We then infer from Theorem (10.4) that $(x_{I_0}, 0)$ is an optimal solution of (LP Int) for $\lambda \to -\infty$. Note that the constraint $\tau \geq 0$ is satisfied exactly, so I_0 can be reused without modification. This implies that $(x_{I_0}, 0)$ maximizes $(0, -1)$ over the polyhedron of the interpolated problem. A fortiori, we may utilize the initial direction $(0, -1)$ and the index set I_0 of initial solution as input of Phase II.

Since (LP Unit) is bounded, we infer that (LP Int) is bounded for all λ, and we know that (LP Int) is feasible. Therefore, when we use $(0, 1)$ as the objective direction and $(z, 0)$ as the direction of rotation, the two successive applications of algorithm `Shadow Vertex Simplex Method` provide a vertex (x_{I_1}, τ). Obviously, (x_{I_1}, τ) is an optimal solution of (LP Int) for $\lambda \to \infty$. Then we conclude via Theorem 10.2 that $\tau \leq 1$ if and only if (LP) is feasible with optimal solution x_{I_1}. Note that the constraint $\tau \leq 1$ is satisfied exactly by $(x_{I_1}, 1)$ then, so I_1 can be extracted without modification.

We conclude that algorithm `Full Solution Algorithm` is correct.

Input:	$A \in \mathbb{R}^{n\times d}$, $b \in \mathbb{R}^n$, $z \in \mathbb{R}^d$
Output:	Either 'unbounded', 'infeasible' or the index set I_1 of an optimal solution.
Algorithm:	(I) Apply the algorithm `Unit Solver` to the unit problem $(A, \vec{1}, z)$. If 'unbounded' is returned, then return 'unbounded'. If I_0 is returned, then proceed.
	(II) Apply the shadow vertex method to the interpolated polyhedron with initial direction $(0, -1)$, initial solution $(x_{I_0}, 0)$ and objective direction $(z, 0)$, and obtain a solution $(x, \tau)_{I_{0.5}}$. Apply the shadow vertex method to the interpolated polyhedron with initial direction $(z, 0)$, initial solution $(x, \tau)_{I_{0.5}}$ and objective direction $(1, 0)$, and obtain a solution (x_{I_1}, τ_1).
	(III) If $\tau \neq 1$, then return 'infeasible'. Otherwise, $(x_{I_1}, \tau_1) = (x_{I_1}, 1)$ for some $I_1 \subseteq [n]$. Return I_1.

Algorithm 4: Full solution algorithm

11.2 Smoothed Complexity Estimates

We separately estimate the expected number of pivot steps undertaken by the calls of `Shadow Vertex Simplex Method` in Phase I and Phase II of `Full Solution Algorithm`.

We make the following technical assumptions:

$$\forall i \in [n] : \left\|(\bar{a}_i, \bar{b}_i)\right\|_2 \leq 1, \quad \sigma \leq \frac{1}{6\sqrt{d \log n}}. \tag{14}$$

These conditions are satisfied if the variables are scaled by a suitable factor. This scaling does not have to be implemented and is only of technical relevance; in fact, we show below that it is not necessary.

The linear programming problem (LP Int") is defined by n Gaussian constraint vectors $(a_i, 1-b_i)$ and 2 additional deterministic constraint vectors, say $(0, 1)$ and $(0, -\frac{1}{\gamma})$ in $\mathcal{E}$. TODO: As the `Shadow Vertex Simplex Method` produces a sequence of intermediate solutions, the corresponding index set ranges over facets of the polar polytope that intersect with a vector from $\mathcal{E}$. Hence, the number of pivot steps in Phase II is bounded by the number of facets of the polytope

$$Q := \text{convex}\{0, (0,1), (0, -\frac{1}{\gamma}), (a_1, 1-b_1), \ldots, (a_n, 1-b_n)\}$$

that intersect with the plane $\mathcal{E}$, in other words, the number of edges $\mathcal{S}(\mathcal{Q}, \mathcal{E})$ of the random polytope $\mathcal{E} \cap \mathcal{Q}$. Since $0, (0,1), (0, -\frac{1}{\gamma}) \in \mathcal{E}$, we conclude

$$\mathcal{S}(\mathcal{Q}, \mathcal{E}) \leq 3 + \mathcal{S}\left(\text{convex}\{(a_1, 1-b_1), \ldots, (a_n, 1-b_n)\}, \mathcal{E}\right)$$

Thus it remains to estimate the number of facets of the random polytope

$$Q^- := \text{convex}\{(a_1, 1-b_1), \ldots, (a_n, 1-b_n)\} \subset \mathbb{R}^{d+1}$$

that intersect with $\mathcal{E}$. To simplify matters, we use the scaling invariance, Corollary 12.2; without changing the expected value $\mathcal{S}(\mathcal{Q}, \mathcal{E})$, we may assume

$$\forall i \in [n] : \left\|(\bar{a}_i, \bar{b}_i)\right\|_2 \leq 1, \quad \sigma \leq \frac{1}{6\sqrt{d \log n}}.$$

We thus can utilize the following upper bound.

Theorem 11.1 ([15, Theorem 6.2])**.**
Suppose that $a_1, \ldots, a_n$ are independent Gaussian vectors in $\mathbb{R}^d$, whose mean values have norm at most 1 and whose standard deviation is bounded from above by $\sigma \leq (6\sqrt{d \log n})^{-1}$. Let $\mathcal{E} \subset \mathbb{R}^d$ be a plane, and let $\mathcal{Q} = \text{convex}\{a_1, \ldots, a_n\}$. Then the shadow size $\mathcal{S}(\mathcal{Q}, \mathcal{E})$ is a random variable whose mean value satisfies

$$\mathbb{E}_A \mathcal{S}(\mathcal{P}, \mathcal{E}) \leq C_{II} \mathcal{D}(d, \sigma),$$

where $\mathcal{D}(d, \sigma) = d^3 \sigma^{-4}$ and C_{II} is a universal constant. □

When we denote the expected value of pivot steps in Phase II by T_{II}, then

$$T_{II} \leq \mathcal{D}(d+1, \sigma) + 3 = C_{II}(d+1)^3 \sigma^{-4} + 3.$$

This establishes the upper bound of the number of pivot steps in Phase II.

It is more complicated to estimate the number of pivot steps in Phase I. On the one hand, we need to estimate the number of iterations in `Unit Solver`, on the other hand we need to estimate the number of pivot steps in each call `Shadow Vertex Simplex Method`, where the additional constraint vectors correlate with the input vectors; so a different shadow size estimate proves necessary.

A first observation considers the behaviour of Phase I under scaling. Our estimation of the number of calls to Algorithm `Shadow Vertex Simplex Method`, which is no a tight bound, is invariant under rescaling of the constraint vectors $a_1, \ldots a_n$. Furthermore, the upper bound on the number of facets in the random polytopes is invariant under rescaling of the $a_1, \ldots, a_n$, having chosen ρ as below. In conclusion, we may again assume

$$\forall i \in [n] : \left\|(\bar{a}_i, \bar{b}_i)\right\|_2 \leq 1, \quad \sigma \leq \frac{1}{6\sqrt{d \log n}}.$$

The algorithms in Phase I depend on several parameters. We set

$$l := \left(300 \log^{\frac{1}{2}} d\right)^{-1}, \quad \rho := \min\left\{\frac{1}{6\sqrt{d \log n}}, \frac{1}{9000000 d^{3/2} \log d}\right\}. \tag{15}$$

This implies for the largest singular value of the matrix T^{-1} in Section 8 that

$$\begin{aligned} s_{\max}(T^{-1}) &= \max\left\{d^{-\frac{1}{2}}, l^{-1} d^{-1} (d^2 - d)^{\frac{1}{2}}\right\} \\ &= \max\left\{d^{-\frac{1}{2}}, 300 \log^{\frac{1}{2}} d \cdot (1 - 1/d)^{\frac{1}{2}}\right\} = 300 \log^{\frac{1}{2}} d \cdot (1 - 1/d)^{\frac{1}{2}}. \end{aligned}$$

because $d \geq 3$. Having chosen these parameters, we compute for the terms in Lemma 8.3 and Lemma 8.4 that

$$\left(\rho \kappa s_{\max}(T^{-1})\right)^{-1}$$

$$
\begin{aligned}
&= \frac{\max\{6\sqrt{d\log n}, 9000000 d^{3/2}\log d\}}{300\log^{\frac{1}{2}} d\cdot(1-1/d)^{\frac{1}{2}}\left(1+2d\cdot 300\log^{\frac{1}{2}} d\cdot(1-1/d)^{\frac{1}{2}}\right)}\\
&\geq \frac{9000000 d^{3/2}\log d}{300\log^{\frac{1}{2}} d\cdot(1-1/d)^{\frac{1}{2}}\left(1+2d\cdot 300\log^{\frac{1}{2}} d\cdot(1-1/d)^{\frac{1}{2}}\right)}\\
&\geq \frac{9000000 d^{3/2}\log d}{100\log^{\frac{1}{2}} d\left(1+d\cdot 300\log^{\frac{1}{2}} d\right)}\\
&\geq \frac{90000 d^{3/2}\log^{\frac{1}{2}} d}{\left(1+d\cdot 300\log^{\frac{1}{2}} d\right)} \geq \frac{90000 d^{3/2}\log^{\frac{1}{2}} d}{d\cdot 600\log^{\frac{1}{2}} d} \geq 150 d^{1/2},
\end{aligned}
$$

hence

$$
\exp\left(-\frac{1}{2}\rho^{-2}\kappa^{-2}s_{\max}^{-2}(T^{-1})\right) \leq \exp\left(-11250d\right) < 10^{-10000}.
$$

Next, we choose the parameters of Lemma 8.6 as

$$
h = \frac{1}{60\sqrt{d}}, \quad \tau = \frac{1}{120\sqrt{d}}, \quad \delta = \frac{1}{120\sqrt{d}}.
$$

We instantiate the three terms of the probability estimate in Lemma 8.6 with these values. For the first term, the lower estimate (5) provides

$$
\frac{|\mathcal{S}^{d-1}(h)|}{|S^{d-1}|} \geq 1 - \frac{1}{\pi}\sin^{-1}(h) = 1 - \frac{1}{\pi}\sin^{-1}\left(\frac{1}{60\sqrt{d}}\right) \geq 0.496937.
$$

The upper bound for the second term is

$$
\begin{aligned}
&\frac{d\rho}{2\tau}\exp\left(\frac{-\tau^2}{2\rho^2}\right)\\
&= \frac{60 d^{3/2}\exp\left(\frac{-\tau^2}{2\rho^2}\right)}{\max\left\{6\sqrt{d\log n}, 9\cdot 10^6 d^{3/2}\log d\right\}} \leq \frac{60 d^{3/2}\exp\left(\frac{-\tau^2}{2\rho^2}\right)}{9\cdot 10^6 d^{3/2}\log d} \leq \frac{\exp\left(\frac{-\tau^2}{2\rho^2}\right)}{10^5\log d}\\
&\leq 10^{-5}\log^{-1} d\cdot\exp\left(\frac{-\max\left\{6\sqrt{d\log n}, 9\cdot 10^6 d^{3/2}\log d\right\}^2}{28800 d}\right)\\
&\leq 10^{-5}\log^{-1} d\cdot\exp\left(-\frac{81}{28800}\cdot 10^{12} d^2\log^2 d\right)\\
&\leq 10^{-5}\log^{-1} d\cdot d^{-(10^{10})} \leq 0.0001.
\end{aligned}
$$

For the third term, we apply the upper estimate (6) to find

$$
\begin{aligned}
d\frac{|S^{d-2}(\delta/l)|}{|S^{d-2}|} &\leq d\exp\left(-\frac{(d-3)^2}{4\cdot 14400 d^2 l^2}\right) = d\exp\left(-\frac{(d-3)^2 90000}{4\cdot 14400 d^2}\log d\right)\\
&= d^{-\frac{90000(d-3)^2}{57600 d^2}+1} = d^{-\frac{32400 d^2 - 180000 d + 810000}{57600 d^2}} < 0.1
\end{aligned}
$$

for, say, $d > 200$. This last term contributes most to the failure probability of `Adding Constraints`, and decreases only faster than $d^{-\frac{1}{2}}$. We conclude that the success probability of Algorithm

`Adding Constraints` is very close to $\frac{1}{2}$. In fact, the expected number of iterations at most $N_I \leq 2.04$, and the value rapidly approaches 2 for d large.

The estimate for the number of pivot steps in the iterations of `Unit Solver` is different than the estimate in Phase II, because the additional constraints are coupled with the original constraints. A shadow bound can be developed separately. We have

$$T_I \leq C_I \ln \ln n \cdot \left(D\left(d, \frac{\min(\sigma, \rho)}{\sqrt{\ln n}}\right) + 1 \right) + 1,$$

Eventually, we may summarize for the total number of steps:

$$\begin{aligned}
&N_I \cdot T_I + T_{II} \\
&\leq N_I C_I \ln \ln n \cdot \left(D\left(d, \frac{\min(\sigma, \rho)}{\sqrt{\ln n}}\right) + 1 \right) + 1 + C_{II} D(d+1, \sigma) + 3 \\
&\leq N_I C' \ln^2 n \ln \ln n \cdot D\left(d, \min(\sigma, \rho)\right) + C' D(d, \sigma) + 4. \\
&\leq N_I C'' \ln^2 n \ln \ln n \cdot D\left(d, \min(\sigma, \rho)\right) + 4.
\end{aligned}$$

Here, C' and C'' are universal constants. We see

$$\min(\sigma, \rho) = \min\left(\sigma, \frac{1}{6\sqrt{d \log n}}, \frac{1}{9000000 d^{3/2} \log d}\right),$$

so, with C''' being another universal constant,

$$\begin{aligned}
&N_I C'' \ln^2 n \ln \ln n \cdot \left(D(d, \sigma) + D\left(d, \frac{1}{6\sqrt{d \log n}}\right) + D\left(d, \frac{1}{9000000 d^{3/2} \log d}\right) \right) + 4 \\
&\leq N_I C''' \ln^2 n \ln \ln n \cdot \left(d^3 \sigma^{-4} + d^3 d^2 \log^2 n + d^3 d^6 \log^4 n \right) + 4
\end{aligned}$$

This reproduces Theorem 6.1 of [15].

12 Shadows of Random Polytopes

The major part of the smoothed analysis is the upper estimate of the expected number of edges of a random polytope with a plane, which, in itself, is a question of stochastic geometry and of independent intrinsic interest.

Let $\mathcal{E} \subset \mathbb{R}^d$ be a plane and let $\mathcal{P}$ be a polytope. Then $\mathcal{P} \cap \mathcal{E}$ is a polyhedron of dimension at most two. We denote the number of its edges, called shadow size, by $\mathcal{S}(\mathcal{P}, \mathcal{E}) \in \mathbb{N}$. Note that $\mathcal{S}(\mathcal{P}, \mathcal{E}) \leq \binom{n}{d}$. Let $a_1, \ldots, a_n$ be Gaussian vectors in $\mathbb{R}^d$. More precisely, we assume that a_i is a $(\bar{a}_i, \sigma_i)$-Gaussian for $\bar{a}_i \in \mathbb{R}^d$ and $\sigma_i > 0$. We assume that $\mathcal{P} = \text{convex}\{a_1, \ldots, a_n\}$, thus $\mathcal{P}$ is a random polytope. Then $\mathcal{S}(\mathcal{P}, \mathcal{E})$ is a random variable in $\mathbb{N}$, actually in $\{0, \ldots, \binom{n}{d}\}$. Our goal is to bound its expected value $\mathbb{E}_{a_1, \ldots, a_n} \mathcal{S}(\mathcal{P}, \mathcal{E})$.

In general, if an event is invariant under positive scaling, then its probability under a Gaussian vector is invariant under positive scaling of that random variable. We formalize this this in manner that is of direct relevance in Section 11 for the understanding of the smoothed complexity of the shadow vertex simplex method.

Theorem 12.1.
Let $n, d \in \mathbb{N}$ and let $\mathcal{A} \subset \mathbb{R}^{n \times d}$ be invariant under scaling by $\alpha \in \mathbb{R}^+$. Let X be a random

matrix in $\mathbb{R}^{n\times d}$ with rows $x_1,\dots,x_n$, each x_i being a $(\bar{x}_i,\sigma_i)$-Gaussian. Then for $\alpha\in\mathbb{R}^+$ we have

$$X(\mathcal{A}) = (\alpha X)(\mathcal{A}), \quad \mathcal{A}\subset\mathbb{R}^{d\times d} \text{ measurable.}$$

Proof. Consider the diffeomorphism $F(x)=\alpha x$. We know that X has a density function

$$\mu_X(a_1,\dots,a_n) = \mu_{x_1}(a_1)\cdots\mu_{x_n}(a_n).$$

An elementary computation now shows:

$$\begin{aligned}\int_{\mathcal{A}}\mu_X(a)\mathrm{d}a &= \int_{F^{-1}\mathcal{A}}\mu_X(a)\mathrm{d}a = \int_{\mathcal{A}}\mu_X\left(F^{-1}(a)\right)\alpha^{-dn}\mathrm{d}a = \int_{\mathcal{A}}\mu_{x_1}\left(\frac{a_1}{\alpha}\right)\cdots\mu_{x_n}\left(\frac{a_n}{\alpha}\right)\alpha^{-dn}\mathrm{d}a\\ &= \int_{\mathcal{A}}\mu_{F(x_1)}(a_1)\cdots\mu_{F(x_n)}(a_n)\mathrm{d}a = \int_{\mathcal{A}}\mu_{F(X)}(a)\mathrm{d}a = \int_{\mathcal{A}}\mu_{\alpha X}(a)\mathrm{d}a\end{aligned}$$

which had to be shown. □

Corollary 12.2.
Let $\alpha>0$. Then $\alpha\mathcal{P} = \mathrm{convex}\{\alpha a_1,\dots,\alpha a_n\}$, and

$$\mathbb{E}_{a_1,\dots,a_n}\mathcal{S}(\mathcal{P},\mathcal{E}) = \mathbb{E}_{\alpha a_1,\dots,\alpha a_n}\mathcal{S}(\alpha\mathcal{P},\mathcal{E}).$$

□

We may therefore apply any scaling to the random variables $a_1,\dots,a_n$ at the outset of our proofs. Thus, without loss of generality, we assume a scaling condition within this section:

$$\forall 1\le i\le n: \left(\|\bar{a}_i\|\le 1, \text{ and } \sigma_i \le \frac{1}{6\sqrt{d\log n}}\right). \tag{16}$$

It can be achieved by a suitable scaling of the variables. Our aim is proving the following result.

Theorem 12.3 ([15, Theorem 6.2])**.**
Suppose that $n>d\ge 3$ and the scaling condition (16) holds. Suppose furthermore that $\sigma_i=\sigma$ for $1\le i\le n$. There exists a universal constant $C_{II}>0$ such that

$$\mathbb{E}_{a_1,\dots,a_n}\mathcal{S}(\mathcal{P},\mathcal{E}) \le C_{II} D\left(d,\sigma\right).$$

Here, $D\left(d,\sigma\right) = d^3\sigma^{-4}$.

We refer to the universal constant C_{II} in the sequel. — This bound detoriates for d growing and σ decreasing. As has been observed before, $\mathbb{E}_{a_1,\dots,a_n}\mathcal{S}(\mathcal{P},\mathcal{E})$ is invariant under scaling of the constraint vectors, so one could assume without loss of generality that $\sigma^{-1}=6\sqrt{d\log n}$; however, this can only be assumed if the centers of the rescaled a_i are still contained in the unit ball then, i.e., $(6\sigma\sqrt{d\log n})^{-1}\|\bar{a}_i\|\le 1$. We conclude and interpret this as follows: If the centers of norm are small enough, then the lower bound for σ^{-1} can be assumed after rescaling. This corresponds to the intuition that for vectors with relatively large standard deviation σ, the random polytope $\mathcal{P}$ will be "smeared out". The lower bound, quadratic in d and $\log n$, is then still good enough. If the centers of norm are too large, then the scaling condition (16) might only hold with σ being rescaled too a comparatively small value, an possibly degenerate instances can not be "healed" by the stochastic pertubations then. The blow-up of σ^{-1} in the upper bound of Theorem 12.3 has impact within that asymptoticly deterministic setting.

These insights can be formalized partially as follows:

Corollary 12.4 ([13, Corollary 4.3.1])**.**
Let $n > d \geq 3$ and let $\mathcal{E} \subset \mathbb{R}^d$ be a plane. Let $a_1, \ldots, a_n$ be random vectors in $\mathbb{R}^d$ centered at points $\bar{a}_1, \ldots, \bar{a}_n$ with common standard deviation $\sigma \leq \frac{1}{6}\sqrt{d \ln n}$. Let $\mathcal{P}$ denote the convex closure of $a_1, \ldots, a_n$. Then

$$\mathbb{E}_{a_1,\ldots,a_n} \mathcal{S}(\mathcal{P}, \mathcal{E}) \leq C_{II} D\left(d, \frac{\sigma}{\max(1, \|\bar{a}_1\|, \ldots, \|\bar{a}_n\|)}\right).$$

Proof. Let $\bar{\mu} := \max_i \|\bar{a}_i\|$. If $\bar{\mu} \leq 1$, then Theorem 12.3 holds, and there is nothing to show. Otherwise, we use Corollary 12.2 and Theorem 12.3, again, to infer

$$\mathbb{E}_{a_1,\ldots,a_n} \mathcal{S}(\mathcal{P}, \mathcal{E}) \leq \mathbb{E}_{a_1,\ldots,a_n} \mathcal{S}(\bar{\mu}^{-1}\mathcal{P}, \mathcal{E}) \leq D\left(d, \bar{\mu}^{-1}\sigma\right).$$

This completes the proof. □

An important auxiliary result implies that the Gaussian vectors are contained with the 2-ball, and will be referenced several times throughout this thesis.

Lemma 12.5.
Assume that $n > d \geq 3$ and that (16) holds, and that $\sigma_i \leq 3\sqrt{d \log n}$ for $1 \leq i \leq n$. Then

$$\operatorname*{Prob}_{a_1,\ldots,a_n}\left\{\max_{1\leq i\leq n} \|a_i\| \leq 2\right\} \geq \operatorname*{Prob}_{a_1,\ldots,a_n}\left\{\forall 1 \leq i \leq n : \|a_i\| \leq 3\sigma_i\sqrt{d \log n}\right\} \geq 1 - \omega_0 \binom{n}{d}^{-1},$$

where $\omega_0 \leq 0.00003$ is a universal constant.

Proof. Using the scaling condition (16) and Lemma 6.7, we see

$$\begin{aligned}
&\operatorname{Prob}_A\left\{\max_{1\leq i\leq n} \|a_i\| \geq 2\right\} \leq \operatorname{Prob}_A\left\{\max_{1\leq i\leq n} (\|a_i\| - 1) \geq 1\right\} \\
&\leq \operatorname{Prob}_A\left\{\max_{1\leq i\leq n} (\|a_i\| - \|\bar{a}_i\|) \geq 1\right\} \leq \operatorname{Prob}_A\left\{\max_{1\leq i\leq n} |\|a_i\| - \|\bar{a}_i\|| \geq 1\right\} \\
&\leq \operatorname{Prob}_A\left\{\max_{1\leq i\leq n} \|a_i - \bar{a}_i\| \geq 1\right\} \leq \operatorname{Prob}_A\left\{\max_{1\leq i\leq n} \|a_i - \bar{a}_i\| - 3\sigma_i\sqrt{d \log n} \geq 0\right\} \\
&\leq n^{-\frac{9}{2}d+1} \leq \frac{(d)\cdots(1)}{(n)\cdots(n-d+1)} n^{-\frac{7}{2}d+1} \leq \omega_0 \binom{n}{d}^{-1},
\end{aligned}$$

where $\omega_0 \leq 3^{-9.5} \leq 0.00003$. This proves the desired result. □

Hereafter, we let ω_0 denote the constant from the previous result. As a first application, we show how the assumption of uniform standard deviations of the $a_1, \ldots, a_n$ can be relieved. The remaining effort of this section is directed towards a proof of Theorem 12.3.

Lemma 12.6 ([13, Corollary 4.3.2])**.**
Let $n > d \geq 3$ and let $\mathcal{E} \subset \mathbb{R}^d$ be a plane. For $1 \leq i \leq n$, let a_i be a random vector in $\mathbb{R}^d$ centered at $\bar{a}_i$ with standard deviation σ_i. Let σ_0 be such that $\sigma_0 \leq \sigma_i \leq \frac{1}{3}\sqrt{d \ln n}$. Let $\mathcal{P}$ denote the convex closure of $a_1, \ldots, a_n$. Then

$$\mathbb{E}_{a_1,\ldots,a_n} \mathcal{S}(\mathcal{P}, \mathcal{E}) \leq C_{II} D\left(d, \frac{\sigma_0}{1 + \max_i \|\bar{a}_i\|}\right) + \omega_0.$$

Proof. We write $a_i = \bar{a}_i + u_i + v_i$, where u_i is a $(0, \sigma_0)$-Gaussian and v_i is a $(0, \sqrt{\sigma_i^2 - \sigma_0^2})$-Gaussian. Consider the event $\mathfrak{V}$ that $\|\bar{a}_i + v_i\| \leq \|\bar{a}_i\| + 1$ for $1 \leq i \leq n$. We know from Lemma 12.5 that

$$\begin{aligned}\mathrm{Prob}_{v_1,\ldots,v_n} \mathfrak{V}^c &= \mathrm{Prob}_{v_1,\ldots,v_n} \{\exists 1 \leq i \leq n : \|\bar{a}_i + v_i\| \geq \|\bar{a}_i\| + 1\} \\ &\leq \mathrm{Prob}_{v_1,\ldots,v_n} \{\exists 1 \leq i \leq n : |\|\bar{a}_i + v_i\| - \|\bar{a}_i\|| \geq 1\} \\ &\leq \mathrm{Prob}_{v_1,\ldots,v_n} \{\exists 1 \leq i \leq n : \|v_i\| \geq 1\} \leq \omega_0 \binom{n}{d}^{-1}.\end{aligned}$$

We continue with

$$\mathbb{E}_{v_1,\ldots,v_n}\mathbb{E}_{u_1,\ldots,u_n}\mathcal{S}(\mathcal{P},\mathcal{E}) \leq \mathbb{E}_{v_1,\ldots,v_n|\mathfrak{V}}\mathbb{E}_{u_1,\ldots,u_n}\mathcal{S}(\mathcal{P},\mathcal{E}) + \omega_0.$$

Using Lemma 12.4, we estimate from above by

$$\begin{aligned}\mathbb{E}_{v_1,\ldots,v_n|\mathfrak{V}}\mathbb{E}_{u_1,\ldots,u_n}\mathcal{S}(\mathcal{P},\mathcal{E}) &\leq \mathbb{E}_{v_1,\ldots,v_n|\mathfrak{V}} C_{II} D\left(d, \frac{\sigma_0}{\max(1, \|\bar{a}_1 + v_1\|, \ldots, \|\bar{a}_n + v_n\|)}\right). \\ &\leq C_{II} D\left(d, \frac{\sigma_0}{1 + \max(\|\bar{a}_1\|, \ldots, \|\bar{a}_n\|)}\right).\end{aligned}$$

which completes the proof. □

12.1 Preparatory Geometric Results

Given a random polyhedron $\mathcal{P}$ and a hyperplane $\mathcal{E}$, we estimate the number of edges of the random polyhedron $\mathcal{P} \cap \mathcal{E}$. A series of geometric auxiliary lemmas is proven before we investigate the shadow size.

We begin with some auxiliary lemmas.

Lemma 12.7.
Let $a_1, \ldots, a_n \in \mathbb{R}^2$ be convexly independent, and $v_1, v_2, v_3 \in \mathbb{R}^2$ be the vertices of an equilateral triangle centered at 0. Assume that $r, R > 0$ with $\|a_i\| \leq r$ for $1 \leq i \leq n$ and $\|v_1\| = \|v_2\| = \|v_3\| = R$, and such that $r < R \cdot \cos(\pi/6)$. Write $\mathcal{A} = \mathrm{convex}\{a_1, \ldots, a_n\}$. Then for any edge (a_s, a_t) of $\mathcal{A}$, where $1 \leq s, t \leq n$, there exists $1 \leq j \leq 3$ such that (a_s, a_t) is an edge of $\mathrm{convex}\{\mathcal{A}, v_j\}$ and

$$\mathrm{dist}(v_i, \mathrm{aff}\{a_s, a_t\}) \geq R - r.$$

Proof. First, we observe that for any line L through 0 there exist indices $1 \leq i, j \leq 3$ such that v_i and v_j are separated by L and their distance to L is at least γR, where $\gamma = \cos(\pi/6)$. To see this, we consider the case where L passes through a vertex v_t and rotate this geometric setting.

Second, let (a_s, a_t) be an edge of $\mathcal{A}$, and let L be the line through 0 parallel to this edge. So there exist v_i and v_j whose distance to L is at least γR. Moreover, $\|a_s\|, \|a_t\| \leq r$ implies that $\mathrm{dist}(L, \mathrm{aff}\{a_s, a_t\}) < r$. Since $r < R \cdot \cos(\pi/6)$ by assumption, we conclude that $\mathrm{aff}\{a_s, a_t\}$ separates v_i and v_j, too, and that their distance to the affine line is at least $\cos(\pi/6)R - r > 0$. Finally, we see that a vertex, say, v_j lies on the same side of $\mathrm{aff}\{a_s, a_t\}$ as $\mathcal{A}$. Therefore (a_s, a_t) has the desired property. □

Lemma 12.8.
Let $0 < r \leq R$. If L is an affine line with distance r to the origin, and $x, y \in L$ have norm at

most R, then

$$\left(r + \frac{R^2}{r}\right)^{-1} \|x - y\| \leq \sphericalangle(x, y) \leq \frac{1}{r}\|x - y\|.$$

Proof. Note first that $\left(r + R^2/r\right)^{-1} = \left(1 + R^2/r^2\right)^{-1} r^{-1}$, We see that $\tilde{L} := L/r$ has distance 1 to the origin, and that $\tilde{x} := x/r$, $\tilde{y} := y/r$ on L' have norms smaller than $\tilde{R} := R/r$. The claim is equivalent to

$$\left(1 + \tilde{R}^2\right)^{-1} \|\tilde{x} - \tilde{y}\| \leq \sphericalangle\left(\tilde{x}, \tilde{y}\right) \leq \|\tilde{x} - \tilde{y}\|.$$

Let $u, v \in S^1$ with $u \perp v$ and write $\tilde{L} = u + \mathbb{R}v$. Furthermore, let $\tilde{x} = u + sv$ and $\tilde{y} = u + tv$. Without loss of generality, $s \leq t$. We observe

$$\sphericalangle(x, y) = \arctan t - \arctan s = \int_t^s \frac{d}{d\tau}(\arctan \tau) d\tau = \int_t^s (1 + \tau^2)^{-1} d\tau.$$

We estimate for $s \leq \tau \leq t$ that $1 \leq 1 + \tau^2 \leq 1 + t^2 \leq 1 + \tilde{R}^2$, which implies

$$\frac{t - s}{1 + \tilde{R}^2} \leq \sphericalangle(x, y) \leq t - s.$$

The identity $\|\tilde{y} - \tilde{x}\| = t - s$ proves the claim. □

In the next theorem, and in the sequel, we write $\mathbb{T}^{\mathcal{E}}$ for the 1-sphere in $\mathcal{E}$. Furthermore, after choice of an arbitrary but fixed isometry $f : S^1(0) \subset \mathbb{R}^2 \to \mathbb{T}^{\mathcal{E}}$, we write

$$\mathbb{T}^{\mathcal{E}}_m = \left\{ f\left(\frac{2k\pi}{m}\right) \in \mathbb{T}^{\mathcal{E}} \;\middle|\; 0 \leq k \leq m - 1 \right\}.$$

Then we can show

Theorem 12.9.
Let $a_1, \ldots, a_n$ denote Gaussian random variables in $\mathbb{R}^d$ and let $\mathcal{P} = \text{convex}\{0, a_1, \ldots, a_n\}$. Let $\mathcal{E}$ be a plane in $\mathbb{R}^d$ and $\gamma > 0$.

$$\begin{aligned} &\mathbb{E}_{a_1,\ldots,a_n}\left|\left\{I \subset [n],\ |I| = d \;\middle|\; \begin{array}{c}\exists q \in \mathbb{T}^E : \mathcal{F}_I \text{ facet of } \mathcal{P} \text{ intersecting } \mathbb{R}^+_0 q,\\ \text{dist}(0, \text{aff}\,\mathcal{F}_I) \geq \gamma\end{array}\right\}\right| \\ = &\lim_{m\to\infty} \mathbb{E}_{a_1,\ldots,a_n}\left|\left\{I \subset [n],\ |I| = d \;\middle|\; \begin{array}{c}\exists q \in \mathbb{T}^E_m : \mathcal{F}_I \text{ facet of } \mathcal{P} \text{ intersecting } \mathbb{R}^+_0 q,\\ \sphericalangle(\mathcal{F}_I \cap \mathcal{E}) > \frac{2\pi}{m},\ \text{dist}(0, \text{aff}\,\mathcal{F}_I) \geq \gamma\end{array}\right\}\right|. \end{aligned} \tag{17}$$

Proof. Note that $\geq$ holds trivially in (17). We introduce binary random variables

$$X_I = \chi\left\{\exists q \in \mathbb{T}^E : \mathcal{F}_I \text{ facet of } \mathcal{P} \text{ intersecting } \mathbb{R}^+_0 q,\ \text{dist}(0, \text{aff}\,\mathcal{F}_I) \geq \gamma\right\},$$
$$X^m_I = \chi\left\{\exists q \in \mathbb{T}^E_m : \mathcal{F}_I \text{ facet of } \mathcal{P} \text{ intersecting } \mathbb{R}^+_0 q,\ \sphericalangle(\mathcal{F}_I \cap \mathcal{E}) > \frac{2\pi}{m},\ \text{dist}(0, \text{aff}\,\mathcal{F}_I) \geq \gamma\right\}.$$

indexed over $I \subset [n]$ with $|I| = d$. The first expected term is the sum $\mathbb{E}_{a_1,\ldots,a_n} X_I$, whereas the second term is $\lim_{m\to\infty} \mathbb{E}_{a_1,\ldots,a_n} X^m_I$. We derive

$$\mathbb{E}_{a_1,\ldots,a_n} X_I - \lim_{m\to\infty} \mathbb{E}_{a_1,\ldots,a_n} X^m_I = \sum_I \left(\int \mu_I \mathrm{d}a - \lim_{m\to\infty} \int \mu^m_I \mathrm{d}a\right)$$

where μ_I and μ_I^m are the characteristic functions of the respective events, and the integrals are taken over $\mathbb{R}^{n\times d}$ with the probability measure induced by $a_1, \ldots, a_n$. Note that $\mu_I \geq \mu_I^m$ for all m. By a variation of Fatou's lemma [3, Kapitel IV, §5, Aufgabe 5.5], or direct consideration, the limit and the integral can be rearranged. The last term equals

$$\sum_I \left(\int \mu_I \mathrm{d}a - \int \limsup_{m\to\infty} \mu_I^m \mathrm{d}a \right) = \sum_I \int \mu_I - \max_{m\in\mathbb{N}} \mu_I^m \mathrm{d}a.$$

We have $>$ in (17) if and only if there exists I such that

$$\mathrm{Prob}_{a_1,\ldots,a_n} \left\{ \begin{array}{c} \exists q \in S^1(\mathcal{E}) : \mathcal{F}_I \text{ facet of } \mathcal{P} \text{ intersecting } \mathbb{R}_0^+ q,\ \mathrm{dist}(0, \mathrm{aff}\, \mathcal{F}_I) \geq \gamma, \\ \forall m \in \mathbb{N} : \neg \left(\begin{array}{c} \exists q \in \mathbb{T}_m^E : \mathcal{F}_I \text{ facet of } \mathcal{P} \text{ intersecting } \mathbb{R}_0^+ q, \\ \sphericalangle(\mathcal{F}_I \cap \mathcal{E}) > \frac{2\pi}{m},\ \mathrm{dist}(0, \mathrm{aff}\, \mathcal{F}_I) \geq \gamma \end{array} \right) \end{array} \right\} > 0.$$

It is easy to see that this can be simplified to

$$\mathrm{Prob}_{a_1,\ldots,a_n} \left\{ \begin{array}{c} \exists q \in S^1(\mathcal{E}) : \mathcal{F}_I \text{ facet of } \mathcal{P} \text{ intersecting } \mathbb{R}_0^+ q,\ \mathrm{dist}(0, \mathrm{aff}\, \mathcal{F}_I) \geq \gamma, \\ \forall m \in \mathbb{N} : \sphericalangle(\mathcal{F}_I \cap \mathcal{E}) > \frac{2\pi}{m} \end{array} \right\} > 0.$$

Obviously, this probability is dominated by the probability of $\sphericalangle(\mathcal{F}_I \cap \mathcal{E}) = 0$. But this corresponds to a set of Lebesgue-measure 0. Thus, $>$ cannot hold in (17), and the theorem follows. □

Remark 12.10.
The above theorem can be vastly generalized. The statement $\mathrm{dist}(0, \mathrm{aff}\, \mathcal{F}_I) \geq \gamma$ has not been used explicitly and can be replaced by any event that does depend on A and I (but not on m).

12.2 Main Estimate

We say that the event $\mathfrak{E}$ holds if $\max_{1\leq i\leq n} \|a_i\| \leq 2$. The random variables are located within a small ball with high probability with high probability. Using the scaling condition (14) and Lemma 12.5, we see

$$\mathrm{Prob}_A\, \mathfrak{E}^c \leq \omega_0 \binom{n}{d}^{-1}. \tag{18}$$

Therefore we may condition the expected value to $\mathfrak{E}$ with only a minor penalty:

$$\mathbb{E}_A \mathcal{S}(\mathcal{P}, \mathcal{E}) = \mathbb{E}_{A|\mathfrak{E}} \mathcal{S}(\mathcal{P}, \mathcal{E}) \cdot \mathrm{Prob}\, \mathfrak{E} + \mathbb{E}_{A|\mathfrak{E}^c} \mathcal{S}(\mathcal{P}, \mathcal{E}) \cdot \mathrm{Prob}\, \mathfrak{E}^c \leq \mathbb{E}_{A|\mathfrak{E}} \mathcal{S}(\mathcal{P}, \mathcal{E}) + \omega_0.$$

We have the simple equality

$$\mathbb{E}_{A|\mathfrak{E}} \mathcal{S}(\mathcal{P}, \mathcal{E}) = \mathbb{E}_{A|\mathfrak{E}} \left| \left\{ \mathcal{F} \mid \mathcal{F} \text{ facet of } \mathcal{P}^{\mathcal{E}} \text{ intersecting } \mathbb{R}q,\ q \in \mathbb{T}^{\mathcal{E}} \right\} \right|.$$

The event $\mathfrak{E}$ implies that $\mathcal{P}^{\mathcal{E}} \subset \mathcal{S}^2(\mathcal{E})$. Let $z_1, z_2, z_3 \in \mathcal{S}^8(\mathcal{E})$ be points of an equilateral triangle in $\mathcal{E}$ centered at the origin with diameter 8. We define

$$\mathcal{P}_i = \mathrm{convex}\{\mathcal{P} - z_i, 0\} = \mathrm{convex}\{\mathcal{P}, z_i\} - z_i.$$

We can apply Lemma 12.7 with $r = 2$ and $R = 8$, and infer that for each edge of $\mathcal{P}^{\mathcal{E}}$ there exists at least one $1 \leq i \leq 3$ such that this edge is still an edge of $\mathrm{convex}\{\mathcal{P}^{\mathcal{E}}, z_i\} = \mathrm{convex}\{\mathcal{P}, z_i\}^{\mathcal{E}}$,

and furthermore that its affine span has distance $R - r = 6$ from the origin. We have a correspondence between edges of convex$\{\mathcal{P}, z_i\}^{\mathcal{E}}$ that do not include z_i, and edges of $\mathcal{P}_i^{\mathcal{E}}$ that do not include 0. These observations give

$$\begin{aligned}
&\mathbb{E}_{A|\mathfrak{E}} \left|\left\{\mathcal{F} \mid \mathcal{F} \text{ facet of } \mathcal{P}^{\mathcal{E}} \text{ intersecting } \mathbb{R}_0^+ q,\ q \in \mathbb{T}^{\mathcal{E}}\right\}\right| \\
&\leq \sum_{i=1}^{3} \mathbb{E}_{A|\mathfrak{E}} \left|\left\{\mathcal{F} \mid \mathcal{F} \text{ facet of } \mathcal{P}_i^{\mathcal{E}} \text{ intersecting } \mathbb{R}_0^+ q,\ q \in \mathbb{T}^{\mathcal{E}},\ \operatorname{dist}(\operatorname{aff} \mathcal{F}, 0) \geq 6\right\}\right|.
\end{aligned}$$

Under the general position assumption, it is possible to remove the intersection with $\mathcal{E}$, i.e.,

$$\mathbb{E}_{A|\mathfrak{E}} \mathcal{S}(\mathcal{P}, \mathcal{E}) \leq \sum_{i=1}^{3} \mathbb{E}_{A|\mathfrak{E}} \left|\left\{\mathcal{F} \mid \mathcal{F} \text{ facet of } \mathcal{P}_i \text{ intersecting } \mathbb{R}_0^+ q,\ q \in \mathbb{T}^{\mathcal{E}},\ \operatorname{dist}(\mathcal{E} \cap \operatorname{aff} \mathcal{F}, 0) \geq 6\right\}\right|.$$

In order to apply the torus lemma, we first remove the conditioning on $\mathfrak{E}$, i.e.,

$$\begin{aligned}
&\sum_{i=1}^{3} \mathbb{E}_{A|\mathfrak{E}} \left|\left\{\mathcal{F} \mid \mathcal{F} \text{ facet of } \mathcal{P}_i \text{ intersecting } \mathbb{R}_0^+ q,\ q \in \mathbb{T}^{\mathcal{E}},\ \operatorname{dist}(\mathcal{E} \cap \operatorname{aff} \mathcal{F}, 0) \geq 6\right\}\right| \\
&\leq \frac{1}{\operatorname{Prob} \mathfrak{E}} \cdot \sum_{i=1}^{3} \mathbb{E}_{A} \left|\left\{\mathcal{F} \mid \mathcal{F} \text{ facet of } \mathcal{P}_i \text{ intersecting } \mathbb{R}_0^+ q,\ q \in \mathbb{T}^{\mathcal{E}},\ \operatorname{dist}(\mathcal{E} \cap \operatorname{aff} \mathcal{F}, 0) \geq 6\right\}\right|.
\end{aligned}$$

Note that the additional factor is bounded from above by, say, $\omega_1 \leq 1.0001$. Now Lemma 12.9 shows

$$\begin{aligned}
&\sum_{i=1}^{3} \mathbb{E}_{A} \left|\left\{\mathcal{F} \;\middle|\; \begin{array}{c} \mathcal{F} \text{ facet of } \mathcal{P}_i \text{ intersecting } \mathbb{R}_0^+ q,\ q \in \mathbb{T}^{\mathcal{E}}, \\ \operatorname{dist}(\mathcal{E} \cap \operatorname{aff} \mathcal{F}, 0) \geq 6 \end{array}\right\}\right| \\
&\leq \sum_{i=1}^{3} \lim_{m \to \infty} \mathbb{E}_{A} \left|\left\{\mathcal{F} \;\middle|\; \begin{array}{c} \mathcal{F} \text{ facet of } \mathcal{P}_i \text{ intersecting } \mathbb{R}_0^+ q,\ q \in \mathbb{T}_m^{\mathcal{E}}, \\ \operatorname{dist}(\mathcal{E} \cap \operatorname{aff} \mathcal{F}, 0) \geq 6,\ \sphericalangle(I) > \frac{2\pi}{m} \end{array}\right\}\right|.
\end{aligned}$$

Now we turn back towards the setting where $\mathfrak{E}$ holds, and obtain

$$\begin{aligned}
&\sum_{i=1}^{3} \lim_{m \to \infty} \mathbb{E}_{A} \left|\left\{\mathcal{F} \;\middle|\; \begin{array}{c} \mathcal{F} \text{ facet of } \mathcal{P}_i \text{ intersecting } \mathbb{R}_0^+ q,\ q \in \mathbb{T}_m^{\mathcal{E}}, \\ \sphericalangle(\mathcal{F}) > \frac{2\pi}{m},\ \operatorname{dist}(\mathcal{E} \cap \operatorname{aff} \mathcal{F}, 0) \geq 6 \end{array}\right\}\right| \\
&= \sum_{i=1}^{3} \lim_{m \to \infty} \mathbb{E}_{A|\mathfrak{E}} \left|\left\{\mathcal{F} \;\middle|\; \begin{array}{c} \mathcal{F} \text{ facet of } \mathcal{P}_i \text{ intersecting } \mathbb{R}_0^+ q,\ q \in \mathbb{T}_m^{\mathcal{E}}, \\ \sphericalangle(\mathcal{F}) > \frac{2\pi}{m},\ \operatorname{dist}(\mathcal{E} \cap \operatorname{aff} \mathcal{F}, 0) \geq 6 \end{array}\right\}\right| \cdot \operatorname{Prob}_A \mathfrak{E} \\
&\quad + \sum_{i=1}^{3} \lim_{m \to \infty} \mathbb{E}_{A|\mathfrak{E}^c} \left|\left\{\mathcal{F} \;\middle|\; \begin{array}{c} \mathcal{F} \text{ facet of } \mathcal{P}_i \text{ intersecting } \mathbb{R}_0^+ q,\ q \in \mathbb{T}_m^{\mathcal{E}}, \\ \sphericalangle(\mathcal{F}) > \frac{2\pi}{m},\ \operatorname{dist}(\mathcal{E} \cap \operatorname{aff} \mathcal{F}, 0) \geq 6 \end{array}\right\}\right| \cdot \operatorname{Prob}_A \mathfrak{E}^c \\
&\leq \sum_{i=1}^{3} \lim_{m \to \infty} \mathbb{E}_{A|\mathfrak{E}} \left|\left\{\mathcal{F} \;\middle|\; \begin{array}{c} \mathcal{F} \text{ facet of } \mathcal{P}_i \text{ intersecting } \mathbb{R}_0^+ q,\ q \in \mathbb{T}_m^{\mathcal{E}}, \\ \sphericalangle(\mathcal{F}) > \frac{2\pi}{m},\ \operatorname{dist}(\mathcal{E} \cap \operatorname{aff} \mathcal{F}, 0) \geq 6 \end{array}\right\}\right| \cdot \operatorname{Prob}_A \mathfrak{E} + 3\omega_0.
\end{aligned}$$

Now Lemma 12.8 applies. For any facet $\mathcal{F}$ included in one of the sums above we have $\operatorname{dist}(0, \operatorname{aff}(\mathcal{F})) \geq 1$ and $\mathcal{F} \subset B^{\vartheta}(0)$, where $\vartheta = 2 + R = 8$, so that we also have

$$\left(1 + \frac{\vartheta^2}{6}\right)^{-1} \operatorname{diam}(\mathcal{F}) \leq \sphericalangle(\mathcal{F}) \leq \operatorname{diam}(\mathcal{F}).$$

We write $q_{\mathcal{F}}$ for the intersection point of $\mathbb{R}_0^+ q$ and $\mathcal{F}$. The comparability of angle and diameter of $\mathcal{F}$ implies that we may write

$$\begin{aligned}
&\sum_{i=1}^{3} \lim_{m\to\infty} \mathbb{E}_{A|\mathfrak{E}} \left| \left\{ \mathcal{F} \;\middle|\; \begin{array}{c} \mathcal{F} \text{ facet of } \mathcal{P}_i \text{ intersecting } \mathbb{R}_0^+ q,\ q \in \mathbb{T}_m^{\mathcal{E}}, \\ \sphericalangle(\mathcal{F}) > \frac{2\pi}{m},\ \operatorname{dist}(0, \operatorname{aff}(\mathcal{F})) \geq 6 \end{array} \right\} \right| \\
&\leq \sum_{i=1}^{3} \lim_{m\to\infty} \mathbb{E}_{A|\mathfrak{E}} \left| \left\{ \mathcal{F} \;\middle|\; \begin{array}{c} \mathcal{F} \text{ facet of } \mathcal{P}_i \text{ intersecting } \mathbb{R}_0^+ q \text{ at } q_{\mathcal{F}},\ q \in \mathbb{T}_m^{\mathcal{E}}, \\ \sphericalangle(\mathcal{F}) > \frac{2\pi}{m},\ \operatorname{dist}(0, \operatorname{aff}(\mathcal{F})) \geq 6,\ \sphericalangle(\mathcal{F}, q_{\mathcal{F}}) < \frac{2\pi}{m} \end{array} \right\} \right| \\
&\leq \sum_{i=1}^{3} \lim_{m\to\infty} \mathbb{E}_{A|\mathfrak{E}} \left| \left\{ \mathcal{F} \;\middle|\; \begin{array}{c} \mathcal{F} \text{ facet of } \mathcal{P}_i \text{ intersecting } \mathbb{R}_0^+ q \text{ at } q_{\mathcal{F}},\ q \in \mathbb{T}_m^{\mathcal{E}}, \\ \sphericalangle(\mathcal{F}) > \frac{2\pi}{m},\ \operatorname{dist}(0, \operatorname{aff}(\mathcal{F})) \geq 6,\ \operatorname{dist}(\partial\mathcal{F}, q_{\mathcal{F}}) < (1 + \frac{\vartheta^2}{6})\frac{2\pi}{m} \end{array} \right\} \right| \\
&\leq \sum_{i=1}^{3} \lim_{m\to\infty} \mathbb{E}_{A|\mathfrak{E}} \left| \left\{ \mathcal{F} \;\middle|\; \begin{array}{c} \mathcal{F} \text{ facet of } \mathcal{P}_i \text{ intersecting } \mathbb{R}_0^+ q \text{ at } q_{\mathcal{F}},\ q \in \mathbb{T}_m^{\mathcal{E}}, \\ \operatorname{dist}(\partial\mathcal{F}, q_{\mathcal{F}}) < (1 + \frac{\vartheta^2}{6})\frac{2\pi}{m} \end{array} \right\} \right|.
\end{aligned}$$

Summarizing the present estimate, we may conclude

$$\begin{aligned}
&\mathbb{E}_A \mathcal{S}(\mathcal{P}, \mathcal{E}) \\
&\leq \omega_1 \cdot \sum_{i=1}^{3} \lim_{m\to\infty} \mathbb{E}_{A|\mathfrak{E}} \left| \left\{ \mathcal{F} \;\middle|\; \begin{array}{c} \mathcal{F} \text{ facet of } \mathcal{P}_i \text{ intersecting } \mathbb{R}_0^+ q \text{ at } q_{\mathcal{F}}, \\ q \in \mathbb{T}_m^{\mathcal{E}},\ \operatorname{dist}(\partial\mathcal{F}, q_{\mathcal{F}}) < (1 + \frac{\vartheta^2}{6})\frac{2\pi}{m} \end{array} \right\} \right| + 4\omega_0 \\
&\leq 3\omega_1 \cdot \max_{1\leq i\leq 3} \lim_{m\to\infty} \mathbb{E}_{A|\mathfrak{E}} \left| \left\{ \mathcal{F} \;\middle|\; \begin{array}{c} \mathcal{F} \text{ facet of } \mathcal{P}_i \text{ intersecting } \mathbb{R}_0^+ q \text{ at } q_{\mathcal{F}}, \\ q \in \mathbb{T}_m^{\mathcal{E}},\ \operatorname{dist}(\partial\mathcal{F}, q_{\mathcal{F}}) < (1 + \frac{\vartheta^2}{6})\frac{2\pi}{m} \end{array} \right\} \right| + 4\omega_0.
\end{aligned}$$

For the remaining estimates we forget about most of the special structure of the $\mathcal{P}_i$. We only use that the vertices $a_1 - z_i, \ldots, a_n - z_i$ have mean values within the 9-ball, and introduce the set of tupels of random variables

$$\mathfrak{Z} := \left\{ \begin{array}{c} a_1, \ldots, a_n \text{ are random variables with} \\ \text{centers in } B^9(0) \text{ and standard deviation } \sigma \end{array} \right\}.$$

We generalize our to estimate to

$$\begin{aligned}
&\mathbb{E}_A \mathcal{S}(\mathcal{P}, \mathcal{E}) \\
&\leq 3\omega_1 \cdot \max_{Z\in\mathfrak{Z}} \lim_{m\to\infty} \mathbb{E}_{Z|\mathfrak{E}} \left| \left\{ \mathcal{F} \;\middle|\; \begin{array}{c} \mathcal{F} \text{ facet of } \mathcal{P}_Z \text{ intersecting } \mathbb{R}_0^+ q \text{ at } q_{\mathcal{F}}, \\ q \in \mathbb{T}_m^{\mathcal{E}},\ \operatorname{dist}(\partial\mathcal{F}, q_{\mathcal{F}}) < (1 + \frac{\vartheta^2}{6})\frac{2\pi}{m} \end{array} \right\} \right| + 4\omega_0.
\end{aligned}$$

We may employ Lemma 7.5 of [15] to find:

$$\mathbb{E}_A \mathcal{S}(\mathcal{P}, \mathcal{E}) \leq 3\omega_1 \cdot C_0 \cdot \max_{Z\in\mathfrak{Z}} d^3 \sigma^{-4} + 4\omega_0,$$

where C_0 is another universal constant. This completes the estimate.

13 Shadows of Random Polytopes with an Added Facet

In Phase I we solve unit linear programming problems whose constraint vectors are correlated Gaussian random variables. More precisely, the additional Gaussian constraint vectors correlate with the maximal norm μ of the input constraint vectors. We may in fact exclude

the origin from the definition of $\mathcal{P}$ at the cost of neglecting 1 additional edge of $\mathcal{P} \cap \mathcal{E}$. The eventual goal in this section is therefore to find an upper bound for

$$\mathbb{E}_{a_1,\dots,a_{n+d}} \mathcal{S}(\mathcal{P}, \mathcal{E})$$

where $\mathcal{P} := \operatorname{convex}\{a_1, \dots, a_{n+d}\}$ is a random polytope of the type encountered in Phase I.

In our attempt to find an upper bound, we take a look at the following random variables:

$$\bar{\mu} := \max_{1 \le i \le n} \|\bar{a}_i\|, \quad \mu := \max_{1 \le i \le n} \|a_i\|, \quad \mu_0 := e^{\operatorname{ceil}(\ln \mu)}.$$

Note μ_0 is a random variable over the set $e^{\mathbb{N}}$. We first inspect the random variable μ. Let $C_t = e^{-3/2}$ be the constant from the tail estimate, Lemma 6.9.

Lemma 13.1.
With the definitions above, we have

$$\operatorname{Prob}_{a_1,\dots,a_n} \left\{ \frac{C_t}{9} (\ln n)^{-\frac{1}{2}} (\bar{\mu} + \sigma\sqrt{d \ln n}) \le \mu \le \bar{\mu} + 3\sigma\sqrt{d \ln n} \right\} \ge 1 - \binom{n}{d}^{-1}.$$

Proof. We separately estimate the probability that μ transgresses the upper or lower bounds. As for transgressions of the upper bound, we observe

$$\begin{aligned}
&\operatorname{Prob}_{a_1,\dots,a_n} \left\{ \mu \ge \bar{\mu} + 3\sigma\sqrt{d \ln n} \right\} \\
&= \operatorname{Prob}_{a_1,\dots,a_n} \left\{ \mu - \bar{\mu} \ge 3\sigma\sqrt{d \ln n} \right\} \\
&\le \operatorname{Prob}_{a_1,\dots,a_n} \left\{ \max_{1 \le i \le n} \|a_i\| \ge 3\sigma\sqrt{d \ln n} + \bar{\mu} \right\} \\
&\le \operatorname{Prob}_{a_1,\dots,a_n} \left\{ \max_{1 \le i \le n} \|a_i\| \ge 3\sigma\sqrt{d \ln n} \right\} \le \omega_0 \binom{n}{d}^{-1}
\end{aligned}$$

where we have used Lemma 12.5. As for transgressions of the lower bound, we distinguish two cases. On the one hand, if $\bar{\mu} \ge 8\sigma\sqrt{d \log n}$, then we use $n \ge 3$ and $\frac{C_t}{9} < \frac{1}{2}$ to see

$$\begin{aligned}
&\operatorname{Prob}_{a_1,\dots,a_n} \left\{ \mu \le \frac{C_t}{9} (\ln n)^{-\frac{1}{2}} (\bar{\mu} + \sigma\sqrt{d \ln n}) \right\} \\
&\le \operatorname{Prob}_{a_1,\dots,a_n} \left\{ \mu \le \frac{1}{2} (\bar{\mu} + \sigma\sqrt{d \ln n}) \right\} \\
&\le \operatorname{Prob}_{a_1,\dots,a_n} \left\{ \mu \le \bar{\mu} - 3\sigma\sqrt{d \ln n} \right\} \\
&= \operatorname{Prob}_{a_1,\dots,a_n} \left\{ \mu - \bar{\mu} \le -3\sigma\sqrt{d \ln n} \right\} \\
&= \operatorname{Prob}_{a_1,\dots,a_n} \left\{ \bar{\mu} - \mu \ge 3\sigma\sqrt{d \ln n} \right\} \le \omega_0 \binom{n}{d}^{-1}
\end{aligned}$$

as we have already seen in the previous estimate. On the other hand, if $\bar{\mu} \le 8\sigma\sqrt{d \log n}$, then

$$\begin{aligned}
&\operatorname{Prob}_{a_1,\dots,a_n} \left\{ \mu \le \frac{C_t}{9} (\ln n)^{-\frac{1}{2}} (\bar{\mu} + \sigma\sqrt{d \ln n}) \right\} \\
&\le \operatorname{Prob}_{a_1,\dots,a_n} \left\{ \mu \le C_t \sigma \sqrt{d} \right\}
\end{aligned}$$

$$\leq \operatorname{Prob}_{a_1,\dots,a_n}\left\{\max\|a_i\| \leq C_t\sigma\sqrt{d}\right\}$$
$$\leq e^{-dn} \leq e^{-d\log n} = n^{-d} \leq 0.5\binom{n}{d}^{-1}$$

via Lemma 6.9 and direct computation. These estimates prove the claim. □

This essentially means that if μ and $\bar{\mu}$ are represented in a real number system with basis e, then with high probability μ needs about as many digits as $\bar{\mu}$ does. More formally, with probability larger than $1 - \binom{n}{d}^{-1}$, the random variable μ takes values within a set

$$\mathcal{M} := \left\{\mu \in \mathbb{R} \middle| \frac{C_t}{9}(\ln n)^{-\frac{1}{2}}(\bar{\mu} + \sigma\sqrt{d\ln n}) \leq \mu \leq \bar{\mu} + 3\sigma\sqrt{d\ln n}\right\}.$$

We want to bound the cardinality of the set

$$\mathcal{M}_0 := e^{\operatorname{ceil}\ln\mathcal{M}}.$$

We verify that

$$\begin{aligned}
|\mathcal{M}_0| &= |\ln\mathcal{M}_0| = |\operatorname{ceil}\ln\mathcal{M}| \\
&\leq \ln\left(\bar{\mu} + 3\sigma\sqrt{d\ln n}\right) - \ln\left(\frac{C_t}{9}(\ln n)^{-\frac{1}{2}}\left(\bar{\mu} + \sigma\sqrt{d\ln n}\right)\right) + 1 \\
&\leq \ln\left(\bar{\mu} + \sigma\sqrt{d\ln n}\right) + \ln 3 - \ln\left(\frac{C_t}{9}(\ln n)^{-\frac{1}{2}}\left(\bar{\mu} + \sigma\sqrt{d\ln n}\right)\right) + 1 \\
&\leq \ln\left(\bar{\mu} + \sigma\sqrt{d\ln n}\right) + \ln 3 - \ln C_t + \ln 9 + \frac{1}{2}\ln\ln n - \ln\left(\bar{\mu} + \sigma\sqrt{d\ln n}\right) + 1 \\
&\leq \frac{1}{2}\ln\ln n + 3\ln 3 + \frac{5}{2} \leq 63\ln\ln n.
\end{aligned}$$

Remark 13.2.
Notably, the factor 63 can be replaced by 5 for $d \geq 100$.

Remark 13.3.
While $\mathcal{M}$ is an interval whose length is not invariant under scaling of the random variables a_i, the cardinality of $\mathcal{M}_0$ is indeed invariant (up to a unit due to rounding errors), so this cardinality estimate is scaling invariant.

The net result is that $\mathcal{M}_0$ is a set of cardinality $O(\log\log n)$. Recall that the additional constraints are coupled with input constraints only through the maximum norm of these. We reduce this to the uncoupled case. Let $\mathfrak{E}$ denote the event that $\mu \in \mathcal{M}$. Using basic properties of the expected value, we derive

$$\begin{aligned}
&\mathbb{E}_{a_1,\dots,a_{n+d}}\mathcal{S}(\mathcal{P},\mathcal{E}) \\
&= \mathbb{E}_{a_1,\dots,a_{n+d}|\mathfrak{E}^c}\mathcal{S}(\mathcal{P},\mathcal{E}) \cdot \operatorname*{Prob}_{a_1,\dots,a_n}\{\mu_0 \notin \mathcal{M}_0\} + \mathbb{E}_{a_1,\dots,a_{n+d}|\mathfrak{E}}\mathcal{S}(\mathcal{P},\mathcal{E}) \cdot \operatorname*{Prob}_{a_1,\dots,a_n}\{\mu_0 \in \mathcal{M}_0\} \\
&\leq 1 + \mathbb{E}_{a_1,\dots,a_{n+d}|\mathfrak{E}}\mathcal{S}(\mathcal{P},\mathcal{E}).
\end{aligned}$$

The joint probability measure of the variables $a_1,\dots,a_{n+d}$ has a density of the form

$$\Psi(a_1,\dots,a_{n+d}) = \Psi_1(a_1,\dots,a_n)\cdot\Psi_2(a_{n+1},\dots,a_{n+d};Q,\mu_0)$$

where Ψ_1 is the product of the densities of the independent Gaussian variables $a_1, \ldots, a_n$, and Ψ_2 is the product of the densities of the independent Gaussian variables $a_{n+1}, \ldots, a_{n+d}$, and $\mu_0 = e^{\mathrm{ceil}(\ln \max_{1\le i\le n} \|a_i\|)}$. The parameters Q and μ_0 capture the dependency of the additional random vectors on $\mathcal{O}(d)$ and $\max_{1\le i\le n} \|a_i\|$. This gives

$$\begin{aligned}
&\sum_t t\,\mathrm{Prob}\{\mathcal{S}(\mathcal{P},\mathcal{E}) = t \text{ and } \mathfrak{E}\} \\
&= \sum_t t \int_{\substack{\mathcal{S}(\mathcal{P},\mathcal{E})=t \\ \mathfrak{E}}} \Psi(a_1,\ldots,a_{n+d})\mathrm{d}a_1\cdots\mathrm{d}a_{n+d} \\
&= \sum_t t \int_{\substack{\mathcal{S}(\mathcal{P},\mathcal{E})=t \\ \mathfrak{E}}} \Psi_1(a_1,\ldots,a_n)\cdot\Psi_2\left(a_{n+1},\ldots,a_{n+d};Q,\mu_0\right)\mathrm{d}a_1\cdots\mathrm{d}a_{n+d} \qquad (19) \\
&= \sum_{\zeta_0\in\mathcal{M}_0}\sum_t t \int_{\substack{\mathcal{S}(\mathcal{P},\mathcal{E})=t \\ \mu_0=\zeta}} \Psi_1(a_1,\ldots,a_n)\cdot\Psi_2\left(a_{n+1},\ldots,a_{n+d};Q,\zeta_0\right)\mathrm{d}a_1\cdots\mathrm{d}a_{n+d} \\
&= |\mathcal{M}_0|\cdot\max_{\zeta_0\in\mathcal{M}_0}\sum_t t \int_{\mathcal{S}(\mathcal{P},\mathcal{E})=t} \Psi_1(a_1,\ldots,a_n)\cdot\Psi_2\left(a_{n+1},\ldots,a_{n+d};Q,\zeta_0\right)\mathrm{d}a_1\cdots\mathrm{d}a_{n+d}
\end{aligned}$$

Furthermore, for d large enough

$$\operatorname*{Prob}_{a_1,\ldots,a_n}\{\mu_0\in\mathcal{M}_0\} \le \operatorname*{Prob}_{a_1,\ldots,a_n}\{\mu\in\mathcal{M}\} \ge 0.99.$$

So the condition on $\mathfrak{E}$ can be removed again and we obtain the upper estimate that

$$1 + \mathbb{E}_{a_1,\ldots,a_{n+d}|\mathfrak{E}}\mathcal{S}(\mathcal{P},\mathcal{E}) \le 1 + \omega_2|\mathcal{M}_0|\cdot\max_{\zeta_0\in\mathcal{M}_0}\max_{\mathfrak{Z}}\mathbb{E}_{z_1,\ldots,z_{n+d}}\mathcal{S}(\mathcal{P},\mathcal{E}).$$

Here, say, $\omega_2 \le 1.02$, and $\mathfrak{Z}$ is a set of random variables as follows. Each z_i is a $(\bar{z}_i,\sigma_i)$-Gaussian such that $\sigma_i = \rho$ for $n+1 \le i \le n+d$, such that $\sigma_i = \sigma$ for $1 \le i \le n$, and where

$$\max_{1\le i\le n}\|\bar{z}_i\| \le \bar{\mu}, \qquad \max_{n+1\le i\le n+d}\|\bar{z}_i\| \le 2\zeta_0\cdot\|\bar{v}_1\|$$

holds. It remains to find an upper bound for the expression

$$\mathbb{E}_{z_1,\ldots,z_{n+d}}\mathcal{S}(\mathcal{P},\mathcal{E}).$$

Because of the scaling invariance of the shadow size, Corollary 12.2, we may assume furthermore without loss of generality

$$\max_{1\le i\le n}\|\bar{z}_i\| \le 1, \quad \sigma \le (6\sqrt{d\log n})^{-1}, \quad \rho \le (6\sqrt{d\log n})^{-1}.$$

In a further utilization of the rescaling invariance, we introduce the rescaled variables

$$b_i := \left(27\zeta_0\ln^{\frac{1}{2}} n\right)^{-1} C_t z_i.$$

We write that the variables have centers $\bar{b}_i$ and standard deviations σ_i^b. We know that

$$\mathbb{E}_{z_1,\ldots,z_{n+d}}\mathcal{S}(\mathcal{P},\mathcal{E}) = \mathbb{E}_{b_1,\ldots,b_{n+d}}\mathcal{S}(\mathcal{P},\mathcal{E})$$

In order to apply Theorem 12.6 for this shadow size, we estimate the norm of the centers and standard deviations of these vectors. Write $\zeta := \exp(\mathrm{floor}\ln\zeta_0) \in \mathcal{M}$.

We estimate the norms of their centers from above. Using $\zeta \leq \zeta_0$ and $\zeta \in \mathcal{M}$, we derive for $1 \leq i \leq n$ and n sufficiently large that

$$\begin{aligned}\|\bar{b}_i\| = C_t \left(27\zeta_0 \log^{\frac{1}{2}} n\right)^{-1} \|\bar{z}_i\| &\leq C_t \left(27\zeta \log^{\frac{1}{2}} n\right)^{-1} \bar{\mu} \\ &\leq C_t \left(27 \left(C_t \frac{\bar{\mu}}{9} \sqrt{\ln n}\right) \log^{\frac{1}{2}} n\right)^{-1} \bar{\mu} \leq \frac{1}{3}.\end{aligned}$$

For the indices $n+1 \leq i \leq n+d$, we first recall the norms of their centers have norms bounded by $2\mu_0$ from above, by construction in `Adding Constraints`. Thus we derive for n sufficiently large that

$$\|\bar{b}_i\| = C_t \left(27\zeta_0 \log^{\frac{1}{2}} n\right)^{-1} \|\bar{z}_i\| \leq C_t \left(13.5 \log^{\frac{1}{2}} n\right)^{-1} \|\bar{v}_1\| \leq 1$$

due to the choice of l in Section 11. In particular, the centers of norm $\bar{b}_1, \ldots, \bar{b}_{n+d}$ are contained within the unit sphere.

Our next aim is to estimate the standard deviations of the dilated variables from above and below. For the indices $1 \leq i \leq n$, we derive the upper bound

$$\begin{aligned}\sigma_i^b = C_t \left(27\zeta_0 \log^{\frac{1}{2}} n\right)^{-1} \sigma &\leq C_t \left(27\zeta \log^{\frac{1}{2}} n\right)^{-1} \sigma \\ &\leq C_t \left(27 \left(C_t \frac{\sigma}{9} \sqrt{d}\right) \log^{\frac{1}{2}} n\right)^{-1} \sigma \leq \left(3\sqrt{d \log n}\right)^{-1},\end{aligned}$$

again by $\zeta \leq \zeta_0$, $\zeta \in \mathcal{M}$ and largeness of n. The lower bound is shown by

$$\begin{aligned}\sigma_i^b = C_t \left(27\zeta_0 \log^{\frac{1}{2}} n\right)^{-1} \sigma &\geq C_t \left(27e\zeta \log^{\frac{1}{2}} n\right)^{-1} \sigma \\ \geq C_t \left(27e \log^{\frac{1}{2}} n\right)^{-1} \left(\bar{\mu} + 3\sigma\sqrt{d \log n}\right)^{-1} \sigma &\geq C_t \left(324\sqrt{\log n}\right)^{-1} \sigma,\end{aligned}$$

where we use $\zeta_0 \leq e\zeta$, $\zeta \in \mathcal{M}$ and the scaling condition (14). For the variables $n+1 \leq i \leq n+d$, we need no additional lower bound. We derive an upper bound using the definition of ρ in (15) and largeness of n:

$$\begin{aligned}\sigma_i^b &= C_t \left(27\zeta_0 \log^{\frac{1}{2}} n\right)^{-1} \cdot 2\zeta_0\rho \\ &= 2C_t \left(27 \log^{\frac{1}{2}} n\right)^{-1} \rho \\ &\leq 2C_t \left(27 \log^{\frac{1}{2}} n\right)^{-1} \left(6\sqrt{d \log n}\right)^{-1} \\ &\leq C_t \left(27 \log^{\frac{1}{2}} n\right)^{-1} \left(3\sqrt{d \log n}\right)^{-1} \\ &\leq C_t \, (27)^{-1} \left(3\sqrt{d \log n}\right)^{-1} \leq \left(3\sqrt{d \log n}\right)^{-1}.\end{aligned}$$

Summarizing, we have seen that

$$\max_i \|\bar{b}_i\| \leq 1, \quad \max_{1 \leq i \leq n+d} \sigma_i^b \leq \left(3\sqrt{d \ln n}\right)^{-1}, \quad \min_{1 \leq i \leq n+d} \sigma_i^b \geq \frac{C_t}{324} \ln^{-\frac{1}{2}} n \cdot \min(\sigma, \rho).$$

We can leverage these bounds for an application of Lemma 12.6, and derive

$$\mathbb{E}_{b_1, \ldots, b_n} \mathcal{S}(\mathcal{P}, \mathcal{E}) \leq C_{II} D \left(d, \frac{C_t}{648} \ln^{-\frac{1}{2}} n \cdot \min(\sigma, \rho)\right) + 1.$$

In combination, we find that

$$\begin{aligned}\mathbb{E}_{a_1,\dots,a_n}\mathcal{S}(\mathcal{P},\mathcal{E}) &\leq 1+\omega_2 64\ln\ln n\cdot\left(C_{II}D\left(d,\frac{C_t}{648}\ln^{-\frac{1}{2}}n\cdot\min(\sigma,\rho)\right)+1\right)\\ &\leq C_I\ln\ln n\cdot\left(D\left(d,\frac{\min(\sigma,\rho)}{\sqrt{\ln n}}\right)+1\right)+1,\end{aligned}$$

for some universal constant C_I. This completes the purpose of this section.

14 Conclusion and Outlook

There is a number possibilities to extend and improve the content of this thesis.

We have used a small pertubation of the interpolation linear programming problem to make the shadow vertex simplex method applicable. Either one circumvents this by a suitable modification of the algorithm, e.g., one checks all neigbours of the initial vertex at $t = 0$, or one computationally implements a suitable upper estimate of the parameter γ. We have not explored the computational efficiency of such an estimate.

We have left out the proof of [15, Lemma 7.5] as the last part of the original paper that is described here. It appears natural to include a complete description of Blaschke's theorem [13, Theorem 2.5.2] then, and, going further, a complete proof of Lemma 7.6 in the same publication. Eventually, large and relatively technical parts of Section 4 of [13] are ought to be elaborated upon if such an extension is made.

As already mentioned in [15, p.14], a factor of asymptotically $\mathcal{O}(\log^2 n\log\log n)$ can be removed from the complexity estimate. Albeit not expanded within this thesis, the general path to this improvement has become visible during the production of this thesis. On the one hand, a term of order $\log\log n$ can be removed by a change of Derivation 19; it is not obligatory to replace $\exp(\operatorname{ceil}\ln\mu)$ by μ then. On the other hand, the factor $\log^2 n$ can be removed using the scaling invariance of the shadow size.

The author of [15] also mentioned the possibility to derandomize Phase I. First, we note that letting $\rho\to 0$ does not detoriate the success probability of algorith `Adding Constraints`. Instead if the randomization is only exerted via the random variable Q, the algorithm has success probability slightly below 0.5. Second, the randomization can be skipped altogether, because one of the $2d$ directions $\pm e_i$, $1\leq i\leq d$, can serve as a choice of z_0. However, as $\rho\to 0$, the estimate on $\mathcal{S}(\mathcal{E},\mathcal{P})$ suffers a blow-up. It is not clear how this can be fixed.

These are extensive modifications and expansions to the presentation at hand. Subsequently, it is an interesting question with which random pertubations other than Gaussians we can implement a smoothed analysis. This has also been an initial motiviation for this thesis. Note that Gaussian vectors are generated from the $(0,1)$-Gaussian by rescaling and translation; we conjecture that scalings and translations of other rotationally symmetric random "urvariables" can be accessed in ways very similar. Assuming that the above modifications have been employed, the major difficulty is then to generalize Lemma 7.5 of [15], which likely entails considerable generalizations of Section 4 of [13]. Theorem 12.9 seems to hold for rather large class of random variables. We expect similar generality for Lemma 6.9 and Lemma 12.5. An immediate extension is likely possible for Gaussian variables with anisotropic covariance matrices, in a manner similar to Lemma 12.6. Finally, in order to model zero-preserving random pertubations of A, the random pertubations can be restricted to subspaces of $\mathbb{R}^d$. But how this can be attained is completely open.

In the wake of Michelangelo's aphorism on the sculpture that already resides within the untreated marmor, we conclude that the theory is already there; ours is just to disclose it.

Register

References

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

[2] K. H. Borgwardt, *The Simplex Method - A Probabilistic Analysis*, no. 1 in Algorithms and Combinatorics, Springer, 1987.

[3] J. Elstrodt, *Maß-und Integrationstheorie*, Springer DE, 2007.

[4] S. Gass and T. Saaty, *The computational algorithm for the parametric objective function*, Naval research logistics quarterly, 2 (1955), pp. 39–45.

[5] M. Haimovich, *The simplex algorithm is very good*, On the expected number of pivot steps and related properties of random linear programs. Report, Columbia University, New York, (1983).

[6] K. Königsberger, *Analysis 2*, vol. 2, Springer, 2004.

[7] B. Korte and J. Vygen, *Combinatorial Optimization: Theory and Algorithms*, Springer Publishing Company, Incorporated, 4th ed., 2007.

[8] S. Lang, *Linear algebra*, Addison-Wesley, 1966.

[9] S. Li, *Concise formulas for the area and volume of a hyperspherical cap*, Asian Journal of Mathematics and Statistics, 4 (2011), pp. 66–70.

[10] F. Mezzadri, *How to generate random matrices from the classical compact groups*, arXiv preprint math-ph/0609050, (2006).

[11] T. Nieberg, *Lineare und ganzzahlige programmierung.* University Lecture, 2012.

[12] A. Schrijver, *Theory of linear and integer programming*, Wiley, 1998.

[13] D. A. Spielman and S.-H. Teng, *Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time*, Journal of the ACM (JACM), 51 (2004), pp. 385–463.

[14] G. Stewart, *The efficient generation of random orthogonal matrices with an application to condition estimators*, SIAM Journal on Numerical Analysis, 17 (1980), pp. 403–409.

[15] R. Vershynin, *Beyond hirsch conjecture: Walks on random polytopes and smoothed complexity of the simplex method*, SIAM Journal on Computing, 39 (2009), pp. 646–678.

[16] ——, *Introduction to the non-asymptotic analysis of random matrices*, arXiv preprint arXiv:1011.3027, (2010).

[17] B. Von Querenburg and G. Bengel, *Mengentheoretische topologie*, vol. 3, Springer, 1973.

[18] D. Werner, *Funktionalanalysis*, Springer-Lehrbuch, Springer London, Limited, 2007.

Ausgelöst worden ist alles dadurch, dass ich mir nach und nach sagte, ich hätte vielleicht eine Chance.

Elementarteilchen
Michel Houellebecq