Finding Short Paths on Polytopes by the Shadow Vertex Algorithm[*]

Tobias Brunsch   Heiko Röglin

Department of Computer Science
University of Bonn, Germany
{brunsch,roeglin}@cs.uni-bonn.de

Abstract

We show that the shadow vertex algorithm can be used to compute a short path between a given pair of vertices of a polytope $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ along the edges of P, where $A \in \mathbb{R}^{m \times n}$. Both, the length of the path and the running time of the algorithm, are polynomial in m, n, and a parameter $1/\delta$ that is a measure for the flatness of the vertices of P. For integer matrices $A \in \mathbb{Z}^{m \times n}$ we show a connection between δ and the largest absolute value Δ of any sub-determinant of A, yielding a bound of $O(\Delta^4 m n^4)$ for the length of the computed path. This bound is expressed in the same parameter Δ as the recent non-constructive bound of $O(\Delta^2 n^4 \log(n\Delta))$ by Bonifas et al. [1].

For the special case of totally unimodular matrices, the length of the computed path simplifies to $O(mn^4)$, which significantly improves the previously best known constructive bound of $O(m^{16} n^3 \log^3(mn))$ by Dyer and Frieze [7].

1   Introduction

We consider the following problem: Given a matrix $A = [a_1, \ldots, a_m]^{\mathrm{T}} \in \mathbb{R}^{m \times n}$, a vector $b \in \mathbb{R}^m$, and two vertices x_1 and x_2 of the polytope $P = \{x \in \mathbb{R}^n : Ax \leq b\}$, find a short path from x_1 to x_2 along the edges of P efficiently. In this context *efficient* means that the running time of the algorithm is polynomially bounded in m, n, and the length of the path it computes. Note, that the polytope P does not have to be bounded.

The *diameter* $d(P)$ of the polytope P is the smallest integer d that bounds the length of the shortest path between any two vertices of P from above. The *polynomial Hirsch conjecture* states that the diameter of P is polynomially bounded in m and n for any matrix A and any vector b. As long as this conjecture remains unresolved, it is unclear whether there always exists a path of polynomial length between the given vertices x_1 and x_2. Moreover, even if such a path exists, it is open whether there is an efficient algorithm to find it.

Related work   The diameter of polytopes has been studied extensively in the last decades. In 1957 Hirsch conjectured that the diameter of P is bounded by $m - n$ for any matrix A and any vector b (see Dantzig's seminal book about linear programming [6]). This conjecture has been disproven by Klee and Walkup [9] who gave an unbounded counterexample. However, it remained open for quite a long time whether the conjecture holds for bounded polytopes. More than forty years later Santos [12] gave the first counterexample to this refined conjecture showing that there are bounded polytopes P for which $d(P) \geq (1 + \varepsilon) \cdot m$ for some $\varepsilon > 0$. This is the best known

---

lower bound today. On the other hand, the best known upper bound of $O(m^{1+\log n})$ due to Kalai and Kleitman [8] is only quasi-polynomial. It is still an open question whether $d(P)$ is always polynomially bounded in m and n. This has only been shown for special classes of polytopes like 0/1 polytopes, flow-polytopes, and the transportation polytope. For these classes of polytopes bounds of $m - n$ (Naddef [10]), $O(mn \log n)$ (Orlin [11]), and $O(m)$ (Brightwell et al. [3]) have been shown, respectively. On the other hand, there are bounds on the diameter of far more general classes of polytopes that depend polynomially on m, n, and on additional parameters. Recently, Bonifas et al. [1] showed that the diameter of polytopes P defined by integer matrices A is bounded by a polynomial in n and a parameter that depends on the matrix A. They showed that $d(P) = O(\Delta^2 n^4 \log(n\Delta))$, where Δ is the largest absolute value of any sub-determinant of A. Although the parameter Δ can be very large in general, this approach allows to obtain bounds for classes of polytopes for which Δ is known to be small. For example, if the matrix A is totally unimodular, i.e., if all sub-determinants of A are from $\{-1, 0, 1\}$, then their bound simplifies to $O(n^4 \log n)$, improving the previously best known bound of $O(m^{16} n^3 \log^3(mn))$ by Dyer and Frieze [7].

We are not only interested in the existence of a short path between two vertices of a polytope but we want to compute such a path efficiently. It is clear that lower bounds for the diameter of polytopes have direct (negative) consequences for this algorithmic problem. However, upper bounds for the diameter do not necessarily have algorithmic consequences as they might be non-constructive. The aforementioned bounds of Orlin, Brightwell et al., and Dyer and Frieze are constructive, whereas the bound of Bonifas et al. is not.

Our contribution We give a constructive upper bound for the diameter of the polytope $P = \{x \in \mathbb{R}^n : Ax \le b\}$ for arbitrary matrices $A \in \mathbb{R}^{m \times n}$ and arbitrary vectors $b \in \mathbb{R}^m$.[1] This bound is polynomial in m, n, and a parameter $1/\delta$, which depends only on the matrix A and is a measure for the angle between edges of the polytope P and their neighboring facets. We say that a facet F of the polytope P is neighboring an edge e if exactly one of the endpoints of e belongs to F. The parameter δ denotes the smallest sine of any angle between an edge and a neighboring facet in P. If, for example, every edge is orthogonal to its neighboring facets, then $\delta = 1$. On the other hand, if there exists an edge that is almost parallel to a neighboring facet, then $\delta \approx 0$. The formal definition of δ is deferred to Section 5.

A well-known pivot rule for the simplex algorithm is the shadow vertex rule, which gained attention in recent years because it has been shown to have polynomial running time in the framework of smoothed analysis [13]. We will present a randomized variant of this pivot rule that computes a path between two given vertices of the polytope P. We will introduce this variant in Section 2 and we call it *shadow vertex algorithm* in the following.

Theorem 1. *Given vertices x_1 and x_2 of P, the shadow vertex algorithm efficiently computes a path from x_1 to x_2 on the polytope P with expected length $O\left(\frac{mn^2}{\delta^2}\right)$.*

Let us emphasize that the algorithm is very simple and its running time depends only polynomially on m, n and the length of the path it computes.

Theorem 1 does not resolve the polynomial Hirsch conjecture as the value δ can be exponentially small. Furthermore, it does not imply a good running time of the shadow vertex method for optimizing linear programs because for the variant considered in this paper both vertices have to be known. Contrary to this, in the optimization problem the objective is to determine the optimal vertex. To compare our results with the result by Bonifas et al. [1], we show that, if A is an integer matrix, then $\frac{1}{\delta} \le n \cdot \Delta^2$, which yields the following corollary.

Corollary 2. *Let $A \in \mathbb{Z}^{m \times n}$ be an integer matrix and let $b \in \mathbb{R}^m$ be a real-valued vector. Given vertices x_1 and x_2 of P, the shadow vertex algorithm efficiently computes a path from x_1 to x_2 on the polytope P with expected length $O(\Delta^4 mn^4)$.*

---

[1]Note that we do not require the polytope to be bounded.

This bound is worse than the bound of Bonifas et al., but it is constructive. Furthermore, if A is a totally unimodular matrix, then $\Delta = 1$. Hence, we obtain the following corollary.

Corollary 3. *Let $A \in \mathbb{Z}^{m \times n}$ be a totally unimodular matrix and let $b \in \mathbb{R}^m$ be a vector. Given vertices x_1 and x_2 of P, the shadow vertex algorithm efficiently computes a path from x_1 to x_2 on the polytope P with expected length $O(mn^4)$.*

This is a significant improvement upon the previously best known constructive bound of $O(m^{16} n^3 \log^3(mn))$ due to Dyer and Frieze because we can assume $m \geq n$. Otherwise, P does not have vertices and the problem is ill-posed.

Organization of the paper In Section 2 we describe the shadow vertex algorithm. In Section 4 we give an outline of our analysis and present the main ideas. After that, in Section 5, we introduce the parameter δ and discuss some of its properties. Section 6 is devoted to the proof of Theorem 1. The probabilistic foundations of our analysis are provided in Section 7.

2 The Shadow Vertex Algorithm

Let us first introduce some notation. For an integer $n \in \mathbb{N}$ we denote by $[n]$ the set $\{1, \ldots, n\}$. Let $A \in \mathbb{R}^{m \times n}$ be an $m \times n$-matrix and let $i \in [m]$ and $j \in [n]$ be indices. With $A_{i,j}$ we refer to the $(m-1) \times (n-1)$-submatrix obtained from A by removing the ith row and the jth column. We call the determinant of any $k \times k$-submatrix of A a *sub-determinant of A of size k*. By $\mathbb{I}_n$ we denote the $n \times n$-identity matrix $\text{diag}(1, \ldots, 1)$ and by $\mathbb{O}_{m \times n}$ the $m \times n$-zero matrix. If $n \in \mathbb{N}$ is clear from the context, then we define vector e_i to be the ith column of $\mathbb{I}_n$. For a vector $x \in \mathbb{R}^n$ we denote by $\|x\| = \|x\|_2$ the Euclidean norm of x and by $\mathcal{N}(x) = \frac{1}{\|x\|} \cdot x$ for $x \neq 0$ the normalization of vector x.

2.1 Shadow Vertex Pivot Rule

Our algorithm is inspired by the shadow vertex pivot rule for the simplex algorithm. Before describing our algorithm, we will briefly explain the geometric intuition behind this pivot rule. For a complete and more formal description, we refer the reader to [2] or [13]. Let us consider the linear program $\min c^{\text{T}} x$ subject to $x \in P$ for some vector $c \in \mathbb{R}^n$ and assume that an initial vertex x_1 of the polytope P is known. For the sake of simplicity, we assume that there is a unique optimal vertex $x^\star$ of P that minimizes the objective function $c^{\text{T}} x$. The shadow vertex pivot rule first computes a vector $w \in \mathbb{R}^n$ such that the vertex x_1 minimizes the objective function $w^{\text{T}} x$ subject to $x \in P$. Again for the sake of simplicity, let us assume that the vectors c and w are linearly independent.

In the second step, the polytope P is projected onto the plane spanned by the vectors c and w. The resulting projection is a polygon P' and one can show that the projections of both the initial vertex x_1 and the optimal vertex $x^\star$ are vertices of this polygon. Additionally every edge between two vertices x and y of P' corresponds to an edge of P between two vertices that are projected onto x and y, respectively. Due to these properties a path from the projection of x_1 to the projection of $x^\star$ along the edges of P' corresponds to a path from x_1 to $x^\star$ along the edges of P.

This way, the problem of finding a path from x_1 to $x^\star$ on the polytope P is reduced to finding a path between two vertices of a polygon. There are at most two such paths and the shadow vertex pivot rule chooses the one along which the objective $c^{\text{T}} x$ improves.

2.2 Our Algorithm

As described in the introduction we consider the following problem: We are given a matrix $A = [a_1, \ldots, a_m]^{\text{T}} \in \mathbb{R}^{m \times n}$, a vector $b \in \mathbb{R}^m$, and two vertices x_1, x_2 of the polytope $P =$

$\{x \in \mathbb{R}^n : Ax \le b\}$. Our objective is to find a short path from x_1 to x_2 along the edges of P.

We propose the following variant of the shadow vertex pivot rule to solve this problem: First choose two vectors $w_1, w_2 \in \mathbb{R}^n$ such that x_1 uniquely minimizes $w_1^{\mathrm{T}} x$ subject to $x \in P$ and x_2 uniquely maximizes $w_2^{\mathrm{T}} x$ subject to $x \in P$. Then project the polytope onto the plane spanned by w_1 and w_2 in order to obtain a polygon P'. Let us call the projection π. By the same arguments as for the shadow vertex pivot rule, it follows that $\pi(x_1)$ and $\pi(x_2)$ are vertices of P' and that a path from $\pi(x_1)$ to $\pi(x_2)$ along the edges of P' can be translated into a path from x_1 to x_2 along the edges of P. Hence, it suffices to compute such a path to solve the problem. Again computing such a path is easy because P' is a two-dimensional polygon.

The vectors w_1 and w_2 are not uniquely determined, but they can be chosen from cones that are determined by the vertices x_1 and x_2 and the polytope P. We choose w_1 and w_2 randomly from these cones. A more precise description of this algorithm is given as Algorithm 1.

Algorithm 1 Shadow Vertex Algorithm

1: Determine n linearly independent rows u_k^{T} of A for which $u_k^{\mathrm{T}} x_1 = b_k$.
2: Determine n linearly independent rows v_k^{T} of A for which $v_k^{\mathrm{T}} x_2 = b_k$.
3: Draw vectors $\lambda, \mu \in (0,1]^n$ independently and uniformly at random.
4: Set $w_1 = -[\mathcal{N}(u_1), \ldots, \mathcal{N}(u_n)] \cdot \lambda$ and $w_2 = [\mathcal{N}(v_1), \ldots, \mathcal{N}(v_n)] \cdot \mu$.
5: Use the function $\pi : x \mapsto (w_1^{\mathrm{T}} x, w_2^{\mathrm{T}} x)$ to project P onto the Euclidean plane and obtain the shadow vertex polygon $P' = \pi(P)$.
6: Walk from $\pi(x_1)$ along the edges of P' in increasing direction of the second coordinate until $\pi(x_2)$ is found.
7: Output the corresponding path of P.

Let us give some remarks about the algorithm above. The vectors $u_1, \ldots, u_n$ in Line 1 and the vectors $v_1, \ldots, v_n$ in Line 2 must exist because x_1 and x_2 are vertices of P. The only point where our algorithm makes use of randomness is in Line 3. By the choice of w_1 and w_2 in Line 4, x_1 is the unique optimum of the linear program $\min w_1^{\mathrm{T}} x$ s.t. $x \in P$ and x_2 is the unique optimum of the linear program $\max w_2^{\mathrm{T}} x$ s.t. $x \in P$. The former follows because for any $y \in P$ with $y \ne x_1$ there must be an index $k \in [n]$ with $u_k^{\mathrm{T}} x_1 < b_k$. The latter follows analogously. Note, that $\|w_1\| \le \sum_{k=1}^n \lambda_k \cdot \|\mathcal{N}(u_k)\| \le \sum_{k=1}^n \lambda_k \le n$ and, similarly, $\|w_2\| \le n$. The shadow vertex polygon P' in Line 5 has several important properties: The projections of x_1 and x_2 are vertices of P' and all edges of P' correspond to projected edges of P. Hence, any path on the edges of P' is the projection of a path on the edges of P. Though we call P' a polygon, it does not have to be bounded. This is the case if P is unbounded in the directions w_1 or $-w_2$. Nevertheless, there is always a path from x_1 to x_2 which will be found in Line 6. For more details about the shadow vertex pivot rule and formal proofs of these properties, we refer to the book of Borgwardt [2].

To give a bit intuition why these statements hold true, consider the projection depicted in Figure 1. We denote the first coordinate of the Euclidean plane by ξ and the second coordinate by η. Since w_1 and w_2 are chosen such that x_1 and x_2 are, among the points of P, optimal for the function $x \mapsto w_1^{\mathrm{T}} x$ and $x \mapsto w_2^{\mathrm{T}} x$, respectively, the projections $\pi(x_1)$ and $\pi(x_2)$ of x_1 and x_2 must be the leftmost vertex and the topmost vertex of $P' = \pi(P)$, respectively. As P' is a (not necessarily bounded) polygon, this implies that if we start in vertex $\pi(x_1)$ and follow the edges of P' in direction of increasing values of η, then we will end up in $\pi(x_2)$ after a finite number of steps. This is not only true if P' is bounded (as depicted by the dotted line and the dark gray area) but also if P is unbounded (as depicted by the dashed lines and the dark gray plus the light gray area). Moreover, note that the slopes of the edges of the path from $\pi(x_1)$ to $\pi(x_2)$ are positive and monotonically decreasing.

Figure 1: Shadow polygon P'

3 Degeneracy

Any degenerate polytope P can be made non-degenerate by perturbing the vector b by a tiny amount of random noise. This way, another polytope $\tilde{P}$ is obtained that is non-degenerate with probability one. Any degenerate vertex of P at which $\ell > n$ constraints are tight generates at most $\binom{\ell}{n}$ vertices of $\tilde{P}$ that are all very close to each other if the perturbation of b is small. We say that two vertices of $\tilde{P}$ that correspond to the same vertex of P are in the same equivalence class.

If the perturbation of the vector b is small enough, then any edge between two vertices of $\tilde{P}$ in different equivalence classes corresponds to an edge in P between the vertices that generated these equivalence classes. We apply the shadow vertex algorithm to the polytope $\tilde{P}$ to find a path R between two arbitrary vertices from the equivalence classes generated by x_1 and x_2, respectively. Then we translate this path into a walk from x_1 to x_2 on the polytope P by mapping each vertex on the path R to the vertex that generated its equivalence class. This way, we obtain a walk from x_1 to x_2 on the polytope P that may visit vertices multiple times and may also stay in the same vertex for some steps. In the latter type of steps only the algebraic representation of the current vertex is changed. As this walk on P has the same length as the path that the shadow vertex algorithm computes on $\tilde{P}$, the upper bound we derive for the length of R also applies to the degenerate polytope P.

Of course the perturbation of the vector b might change the shape of the polytope P. In this context it is important to point out that the parameter δ, which we define in the following, only depends on the matrix A and, thus, is independent of the right-hand side b. Consequently, the parameter δ of the original polytope P can also be used to describe the behavior of the shadow vertex simplex algorithm on the polytope $\tilde{P}$.

4 Outline of the Analysis

In the remainder of this paper we assume that the polytope P is non-degenerate, i.e., for each vertex x of P there are exactly n indices i for which $a_i^{\mathrm{T}} x = b_i$. This implies that for any edge between two vertices x and y of P there are exactly $n-1$ indices i for which $a_i^{\mathrm{T}} x = a_i^{\mathrm{T}} y = b_i$. According to Section 3 this assumption is justified.

From the description of the shadow vertex algorithm it is clear that the main step in proving Theorem 1 is to bound the expected number of edges on the path from $\pi(x_1)$ to $\pi(x_2)$ on the polygon P'. In order to do this, we look at the slopes of the edges on this path. As we discussed above, the sequence of slopes is monotonically decreasing. We will show that due to the randomness in the objective functions w_1 and w_2, it is even strictly decreasing with probability one. Furthermore

all slopes on this path are bounded from below by 0.

Instead of counting the edges on the path from $\pi(x_1)$ to $\pi(x_2)$ directly, we will count the number of different slopes in the interval $[0, 1]$ and we observe that the expected number of slopes from the interval $[0, \infty)$ is twice the expected number of slopes from the interval $[0, 1]$. In order to count the number of slopes in $[0, 1]$, we partition the interval $[0, 1]$ into several small subintervals and we bound for each of these subintervals I the expected number of slopes in I. Then we use linearity of expectation to obtain an upper bound on the expected number of different slopes in $[0, 1]$, which directly translates into an upper bound on the expected number of edges on the path from $\pi(x_1)$ to $\pi(x_2)$.

We choose the subintervals so small that, with high probability, none of them contains more than one slope. Then, the expected number of slopes in a subinterval $I = (t, t+\varepsilon]$ is approximately equal to the probability that there is a slope in the interval I. In order to bound this probability, we use a technique reminiscent of the principle of deferred decisions that we have already used in [5]. The main idea is to split the random draw of the vectors w_1 and w_2 in the shadow vertex algorithm into two steps. The first step reveals enough information about the realizations of these vectors to determine the last edge $e = (\hat{p}, p^\star)$ on the path from $\pi(x_1)$ to $\pi(x_2)$ whose slope is bigger than t (see Figure 2). Even though e is determined in the first step, its slope is not. We argue that there is still enough randomness left in the second step to bound the probability that the slope of e lies in the interval $(t, t + \varepsilon]$ from above, yielding Theorem 1.

We will now give some more details on how the random draw of the vectors w_1 and w_2 is partitioned. Let $\hat{x}$ and $x^\star$ be the vertices of the polytope P that are projected onto $\hat{p}$ and $p^\star$, respectively. Due to the non-degeneracy of the polytope P, there are exactly $n - 1$ constraints that are tight for both $\hat{x}$ and $x^\star$ and there is a unique constraint $a_i^{\mathrm{T}} x \leq b_i$ that is tight for $x^\star$ but not for $\hat{x}$. In the first step the vector w_1 is completely revealed while instead of w_2 only an element $\tilde{w}_2$ from the ray $\{w_2 + \gamma \cdot a_i : \gamma \geq 0\}$ is revealed. We then argue that knowing w_1 and $\tilde{w}_2$ suffices to identify the edge e. The only randomness left in the second step is the exact position of the vector w_2 on the ray $\{\tilde{w}_2 - \gamma \cdot a_i : \gamma \geq 0\}$, which suffices to bound the probability that the slope of e lies in the interval $(t, t + \varepsilon]$.

Let us remark that the proof of Theorem 1 is inspired by the recent smoothed analysis of the successive shortest path algorithm for the minimum-cost flow problem [4]. Even though the general structure bears some similarity, the details of our analysis are much more involved.

5 The Parameter δ

In this section we define the parameter δ that describes the flatness of the vertices of the polytope and state some relevant properties.

Definition 4.

1. *Let $z_1, \ldots, z_n \in \mathbb{R}^n$ be linearly independent vectors and let $\varphi \in (0, \frac{\pi}{2}]$ be the angle between z_n and the hyperplane $\mathrm{span}\{z_1, \ldots, z_{n-1}\}$. By $\hat{\delta}(\{z_1, \ldots, z_{n-1}\}, z_n) = \sin \varphi$ we denote the sine of angle φ. Moreover, we set $\delta(z_1, \ldots, z_n) = \min_{k \in [n]} \hat{\delta}(\{z_i : i \in [n] \setminus \{k\}\}, z_k)$.*

2. *Given a matrix $A = [a_1, \ldots, a_m]^{\mathrm{T}} \in \mathbb{R}^{m \times n}$, we set*

   $$\delta(A) = \min\left\{\delta(a_{i_1}, \ldots, a_{i_n}) : a_{i_1}, \ldots, a_{i_n} \text{ linearly independent}\right\}.$$

The value $\hat{\delta}(\{z_1, \ldots, z_{n-1}\}, z_n)$ describes how orthogonal z_n is to the span of $z_1, \ldots, z_{n-1}$. If $\varphi \approx 0$, i.e., z_n is close to the span of $z_1, \ldots, z_{n-1}$, then $\hat{\delta}(\{z_1, \ldots, z_{n-1}\}, z_n) \approx 0$. On the other hand, if z_n is orthogonal to $z_1, \ldots, z_{n-1}$, then $\varphi = \frac{\pi}{2}$ and, hence, $\hat{\delta}(\{z_1, \ldots, z_{n-1}\}, z_n) = 1$. The value $\hat{\delta}(\{z_1, \ldots, z_{n-1}\}, z_n)$ equals the distance between both faces of the parallelotope Q, given by $Q = \{\sum_{i=1}^n \alpha_i \cdot \mathcal{N}(z_i) : \alpha_i \in [0, 1]\}$, that are parallel to $\mathrm{span}\{z_1, \ldots, z_{n-1}\}$ and is scale invariant.

The value $\delta(z_1, \ldots, z_n)$ equals twice the inner radius r_n of the parallelotope Q and, thus, is a measure of the flatness of Q: A value $\delta(z_1, \ldots, z_n) \approx 0$ implies that Q is nearly $(n-1)$-dimensional. On the other hand, if $\delta(z_1, \ldots, z_n) = 1$, then the vectors $z_1, \ldots, z_n$ are pairwise orthogonal, that is, Q is an n-dimensional unit cube.

The next lemma lists some useful statements concerning the parameter $\delta := \delta(A)$ including a connection to the parameters Δ_1, Δ_{n-1}, and Δ introduced in the paper of Bonifas et al. [1].

Lemma 5. *Let $z_1, \ldots, z_n \in \mathbb{R}^n$ be linearly independent vectors, let $A \in \mathbb{R}^{m \times n}$ be a matrix, let $b \in \mathbb{R}^m$ be a vector, and let $\delta = \delta(A)$. Then, the following claims hold true:*

1. *If M is the inverse of $[\mathcal{N}(z_1), \ldots, \mathcal{N}(z_n)]^{\mathrm{T}}$, then*

$$\delta(z_1, \ldots, z_n) = \frac{1}{\max_{k \in [n]} \|m_k\|} \leq \frac{\sqrt{n}}{\max_{k \in [n]} \|M_k\|},$$

   where $[m_1, \ldots, m_n] = M$ and $[M_1, \ldots, M_n] = M^{\mathrm{T}}$.

2. *If $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, then $\delta(Qz_1, \ldots, Qz_n) = \delta(z_1, \ldots, z_n)$.*

3. *Let y_1 and y_2 be two neighboring vertices of $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ and let a_i^{T} be a row of A. If $a_i^{\mathrm{T}} \cdot (y_2 - y_1) \neq 0$, then $|a_i^{\mathrm{T}} \cdot (y_2 - y_1)| \geq \delta \cdot \|y_2 - y_1\|$.*

4. *If A is an integral matrix, then $\frac{1}{\delta} \leq n\Delta_1\Delta_{n-1} \leq n\Delta^2$, where Δ, Δ_1, and Δ_{n-1} are the largest absolute values of any sub-determinant of A of arbitrary size, of size 1, and of size $n-1$, respectively.*

Proof. First of all we derive a simple formula for $\hat{\delta}(\{z_1, \ldots, z_{n-1}\}, z_n)$. For this, assume that the vectors $z_1, \ldots, z_n$ are normalized. Now consider a normal vector $x \neq 0$ of $\mathrm{span}\{z_1, \ldots, z_{n-1}\}$ that lies in the same halfspace as z_n. Let $\varphi \in (0, \frac{\pi}{2}]$ be the angle between z_n and $\mathrm{span}\{z_1, \ldots, z_{n-1}\}$ and let $\psi \in [0, \frac{\pi}{2})$ be the angle between z_n and x. Clearly, $\varphi + \psi = \frac{\pi}{2}$. Consequently,

$$\hat{\delta}(\{z_1, \ldots, z_{n-1}\}, z_n) = \sin \varphi = \sin\left(\frac{\pi}{2} - \psi\right) = \cos \psi = \frac{z_n^{\mathrm{T}} x}{\|x\|}.$$

The last fraction is invariant under scaling of x. Since x and z_n lie in the same halfspace, w.l.o.g. we can assume that $z_n^{\mathrm{T}} x = 1$. Hence, $\hat{\delta}(\{z_1, \ldots, z_{n-1}\}, z_n) = \frac{1}{\|x\|}$, where x is the unique solution of the equation $[z_1, \ldots, z_{n-1}, z_n]^{\mathrm{T}} \cdot x = (0, \ldots, 0, 1)^{\mathrm{T}} = e_n$. If the vectors $z_1, \ldots, z_n$ are not normalized, then we obtain

$$\hat{\delta}(\{z_1, \ldots, z_{n-1}\}, z_n) = \hat{\delta}(\{\mathcal{N}(z_1), \ldots, \mathcal{N}(z_{n-1})\}, \mathcal{N}(z_n)) = \frac{1}{\|x\|},$$

where $x = [\mathcal{N}(z_1), \ldots, \mathcal{N}(z_{n-1}), \mathcal{N}(z_n)]^{-\mathrm{T}} \cdot e_n$. Since for the previous line of reasoning we can relabel the vectors $z_1, \ldots, z_n$ arbitrarily, this implies

$$\delta(z_1, \ldots, z_n) = \min_{k \in [n]} \frac{1}{\|[\mathcal{N}(z_1), \ldots, \mathcal{N}(z_n)]^{-\mathrm{T}} \cdot e_k\|}$$

$$= \frac{1}{\max\left\{\|x\| : x \text{ is column of } [\mathcal{N}(z_1), \ldots, \mathcal{N}(z_n)]^{-\mathrm{T}}\right\}}.$$

This yields the equation in Claim 1. Due to

$$\left(\max_{k \in [n]} \|M_k\|\right)^2 \leq \sum_{k \in [n]} \|M_k\|^2 = \sum_{k \in [n]} \|m_k\|^2 \leq n \cdot \left(\max_{k \in [n]} \|m_k\|\right)^2$$

we obtain the inequality $\frac{1}{\max_{k \in [n]} \|m_k\|} \leq \frac{\sqrt{n}}{\max_{k \in [n]} \|M_k\|}$ stated in Claim 1.

For Claim 2 observe that

$$[\mathcal{N}(Qz_1),\ldots,\mathcal{N}(Qz_n)]^{-\mathrm{T}} = [Q\mathcal{N}(z_1),\ldots,Q\mathcal{N}(z_n)]^{-\mathrm{T}}$$
$$= (Q\cdot[\mathcal{N}(z_1),\ldots,\mathcal{N}(z_n)])^{-\mathrm{T}}$$
$$= ([\mathcal{N}(z_1),\ldots,\mathcal{N}(z_n)]^{-1}\cdot Q^{\mathrm{T}})^{\mathrm{T}}$$
$$= Q\cdot[\mathcal{N}(z_1),\ldots,\mathcal{N}(z_n)]^{-\mathrm{T}}$$

for any orthogonal matrix Q. Therefore, we get

$$\frac{1}{\delta(Qz_1,\ldots,Qz_n)} = \max\big\{\|x\| : x \text{ is column of } [\mathcal{N}(Qz_1),\ldots,\mathcal{N}(Qz_n)]^{-\mathrm{T}}\big\}$$
$$= \max\big\{\|Qy\| : y \text{ is column of } [\mathcal{N}(z_1),\ldots,\mathcal{N}(z_n)]^{-\mathrm{T}}\big\}$$
$$= \max\big\{\|y\| : y \text{ is column of } [\mathcal{N}(z_1),\ldots,\mathcal{N}(z_n)]^{-\mathrm{T}}\big\}$$
$$= \frac{1}{\delta(z_1,\ldots,z_n)}\,.$$

For Claim 3 let y_1 and y_2 be two neighboring vertices of P. Then, there are exactly $n-1$ indices j for which $a_j^{\mathrm{T}}\cdot(y_2-y_1)=0$. We denote them by $j_1,\ldots,j_{n-1}$. If there is an index i for which $a_i^{\mathrm{T}}\cdot(y_2-y_1)\neq 0$, then $a_{j_1},\ldots,a_{j_{n-1}},a_i$ are linearly independent. Consequently, $\delta(a_{j_1},\ldots,a_{j_{n-1}},a_i)\geq\delta$. Let us assume that $a_i^{\mathrm{T}}\cdot(y_2-y_1)\geq 0$. (Otherwise, consider $a_i^{\mathrm{T}}\cdot(y_1-y_2)$ instead.) Since y_2-y_1 is a normal vector of $\mathrm{span}\{a_{j_1},\ldots,a_{j_{n-1}}\}$ that lies in the same halfspace as a_i, we obtain

$$\frac{a_i^{\mathrm{T}}\cdot(y_2-y_1)}{\|y_2-y_1\|} = \hat{\delta}(\{a_{j_1},\ldots,a_{j_{n-1}}\},a_i) \geq \delta(a_{j_1},\ldots,a_{j_{n-1}},a_i) \geq \delta$$

and, thus, $a_i^{\mathrm{T}}\cdot(y_2-y_1)\geq\delta\cdot\|y_2-y_1\|$.

For proving Claim 4 we can focus on showing the first inequality. The second one follows from $\Delta\geq\max\{\Delta_1,\Delta_{n-1}\}$. For this, it suffices to show that for n arbitrary linearly independent rows $a_{i_1}^{\mathrm{T}},\ldots,a_{i_n}^{\mathrm{T}}$ of A the inequality

$$\frac{1}{\hat{\delta}(\{a_{i_1},\ldots,a_{i_{n-1}}\},a_{i_n})} \leq n\Delta_1\Delta_{n-1}$$

holds. By previous observations we know that

$$\frac{1}{\hat{\delta}(\{a_{i_1},\ldots,a_{i_{n-1}}\},a_{i_n})} = \|x\|$$

where x is the unique solution of $\hat{A}x=e_n$ for $\hat{A}=[\mathcal{N}(a_{i_1}),\ldots,\mathcal{N}(a_{i_n})]^{\mathrm{T}}$. Let $\tilde{A}=[a_{i_1},\ldots,a_{i_n}]^{\mathrm{T}}$. Then,

$$\|x\|^2 = \sum_{k=1}^{n}x_k^2 = \sum_{k=1}^{n}\left(\frac{\det(\hat{A}_{n,k})}{\det(\hat{A})}\right)^2 = \sum_{k=1}^{n}\left(\frac{\det(\tilde{A}_{n,k})\cdot\prod_{j=1}^{n-1}\frac{1}{\|a_{i_j}\|}}{\det(\tilde{A})\cdot\prod_{j=1}^{n}\frac{1}{\|a_{i_j}\|}}\right)^2$$
$$= \sum_{k=1}^{n}\left(\frac{\det(\tilde{A}_{n,k})\cdot\|a_{i_n}\|}{\det(\tilde{A})}\right)^2 \leq \sum_{k=1}^{n}\left(\frac{\Delta_{n-1}\cdot\sqrt{n}\Delta_1}{1}\right)^2 = n^2\Delta_1^2\Delta_{n-1}^2\,.$$

Some of the equations need further explanation: Due to Cramer's rule, we have $x_k=\frac{\det(\bar{A})}{\det(\hat{A})}$, where $\bar{A}$ is obtained from $\hat{A}$ by replacing the kth column by the right-hand side e_n of the equation $\hat{A}x=e_n$. Laplace's formula yields $|\det(\bar{A})|=|\det(\hat{A}_{n,k})|$. Hence, the second equation is true. For

the third equation note that the kth row of matrix $\hat{A}$ is the same as the kth row of matrix $\tilde{A}$ up to a factor of $\frac{1}{\|a_{i_k}\|}$. The inequality follows from $|\det(\tilde{A}_{n,k})| \le \Delta_{n-1}$ since this is a sub-determinant of A of size $n-1$, from $\|a_{i_n}\| \le \sqrt{n} \cdot \|a_{i_n}\|_\infty \le \sqrt{n}\Delta_1$, since $\|a_{i_n}\|_\infty$ is a sub-determinant of A of size 1, and from $|\det(\tilde{A})| \ge 1$ since $\tilde{A}$ is invertible and integral by assumption. Hence,

$$\frac{1}{\hat{\delta}(\{a_{i_1}, \ldots, a_{i_{n-1}}\}, a_{i_n})} = \|x\| \le n\Delta_1\Delta_{n-1}. \qquad \square$$

6 Analysis

For the proof of Theorem 1 we assume that $\|a_i\| = 1$ for all $i \in [m]$. This entails no loss of generality since normalizing the rows of matrix A (and scaling the right-hand side b appropriately) does neither change the behavior of our algorithm nor does it change the parameter $\delta = \delta(A)$.

For given linear functions L_1 and L_2, we denote by $\pi = \pi_{L_1,L_2}$ the function $\pi \colon \mathbb{R}^n \to \mathbb{R}^2$, given by $\pi(x) = (L_1(x), L_2(x))$. Note, that n-dimensional vectors can be treated as linear functions. By $P' = P'_{L_1,L_2}$ we denote the projection $\pi(P)$ of polytope P onto the Euclidean plane, and by $R = R_{L_1,L_2}$ we denote the path from $\pi(x_1)$ to $\pi(x_2)$ along the edges of polygon P'.

Our goal is to bound the expected number of edges of the path $R = R_{w_1,w_2}$ which is random since w_1 and w_2 depend on the realizations of the random vectors λ and μ. Each edge of R corresponds to a slope in $(0, \infty)$. These slopes are pairwise distinct with probability one (see Lemma 8). Hence, the number of edges of R equals the number of distinct slopes of R. In order to bound the expected number of distinct slopes we first restrict our attention to slopes in the interval $(0, 1]$.

Definition 6. *For a real $\varepsilon > 0$ let $\mathbb{F}_\varepsilon$ denote the event that there are three pairwise distinct vertices z_1, z_2, z_3 of P such that z_1 and z_3 are neighbors of z_2 and such that*

$$\left| \frac{w_2^{\text{T}} \cdot (z_2 - z_1)}{w_1^{\text{T}} \cdot (z_2 - z_1)} - \frac{w_2^{\text{T}} \cdot (z_3 - z_2)}{w_1^{\text{T}} \cdot (z_3 - z_2)} \right| \le \varepsilon.$$

Note that if event $\mathbb{F}_\varepsilon$ does not occur, then all slopes of R differ by more than ε. Particularly, all slopes are pairwise distinct. First of all we show that event $\mathbb{F}_\varepsilon$ is very unlikely to occur if ε is chosen sufficiently small.

Lemma 7. *The probability that there are two neighboring vertices z_1, z_2 of P such that $|w_1^{\text{T}} \cdot (z_2 - z_1)| \le \varepsilon \cdot \|z_2 - z_1\|$ is bounded from above by $\frac{2m^n\varepsilon}{\delta}$.*

Proof. Let z_1 and z_2 be two neighbors of P. Let $\Delta_z = z_2 - z_1$. Because the claim we want to show is invariant under scaling, we can assume without loss of generality that $\|\Delta_z\| = 1$. There are $n-1$ indices $i_1, \ldots, i_{n-1} \in [m]$ such that $a_{i_k}^{\text{T}} z_1 = b_{i_k} = a_{i_k}^{\text{T}} z_2$. Recall that $w_1 = -[u_1, \ldots, u_n] \cdot \lambda$, where $\lambda = (\lambda_1, \ldots, \lambda_n)$ is drawn uniformly at random from $(0, 1]^n$. There must be an index i such that $a_{i_1}, \ldots, a_{i_{n-1}}, u_i$ are linearly independent. Hence, $\kappa := u_i^{\text{T}} \Delta_z \ne 0$ and, thus, $|\kappa| \ge \delta$ due to Lemma 5, Claim 3.

We apply the principle of deferred decisions and assume that all λ_j for $j \ne i$ are already drawn. Then

$$w_1^{\text{T}} \Delta_z = -\sum_{j=1}^n \lambda_j \cdot u_j^{\text{T}} \Delta_z = \underbrace{-\sum_{j \ne i} \lambda_j \cdot u_j^{\text{T}} \Delta_z}_{=: \gamma} - \lambda_i \cdot \kappa.$$

Thus,

$$|w_1^{\text{T}} \Delta_z| \le \varepsilon \iff w_1^{\text{T}} \Delta_z \in [-\varepsilon, \varepsilon] \iff \lambda_i \cdot \kappa \in [\gamma - \varepsilon, \gamma + \varepsilon]$$
$$\iff \lambda_i \in \left[\frac{\gamma}{\kappa} - \frac{\varepsilon}{|\kappa|}, \frac{\gamma}{\kappa} + \frac{\varepsilon}{|\kappa|} \right].$$

The probability for the latter event is bounded by the length of the interval, i.e., by $\frac{2\varepsilon}{|\kappa|} \leq \frac{2\varepsilon}{\delta}$. Since we have to consider at most $\binom{m}{n-1} \leq m^n$ pairs of neighbors (z_1, z_2), applying a union bound yields the additional factor of m^n. $\qquad\square$

Lemma 8. *The probability of event $\mathbb{F}_\varepsilon$ tends to 0 for $\varepsilon \to 0$.*

Proof. Let z_1, z_2, z_3 be pairwise distinct vertices of P such that z_1 and z_3 are neighbors of z_2 and let $\Delta_z := z_2 - z_1$ and $\Delta_z' := z_3 - z_2$. We assume that $\|\Delta_z\| = \|\Delta_z'\| = 1$. This entails no loss of generality as the fractions in Definition 6 are invariant under scaling. Let $i_1, \ldots, i_{n-1} \in [m]$ be the $n-1$ indices for which $a_{i_k}^T z_1 = b_{i_k} = a_{i_k}^T z_2$. The rows $a_{i_1}, \ldots, a_{i_{n-1}}$ are linearly independent because P is non-degenerate. Since z_1, z_2, z_3 are distinct vertices of P and since z_1 and z_3 are neighbors of z_2, there is exactly one index i_ℓ for which $a_{i_\ell}^T z_3 < b_{i_\ell}$, i.e., $a_{i_\ell}^T \Delta_z' \neq 0$. Otherwise, z_1, z_2, z_3 would be collinear which would contradict the fact that they are distinct vertices of P. Without loss of generality assume that $\ell = n - 1$. Since $a_{i_k}^T \Delta_z = 0$ for each $k \in [n-1]$, the vectors $a_{i_1}, \ldots, a_{i_{n-1}}, \Delta_z$ are linearly independent.

We apply the principle of deferred decisions and assume that w_1 is already fixed. Thus, $w_1^T \Delta_z$ and $w_1^T \Delta_z'$ are fixed as well. Moreover, we assume that $w_1^T \Delta_z \neq 0$ and $w_1^T \Delta_z' \neq 0$ since this happens almost surely due to Lemma 7. Now consider the matrix $M = [a_{i_1}, \ldots, a_{i_{n-2}}, \Delta_z, a_{i_{n-1}}]$ and the random vector $(Y_1, \ldots, Y_{n-1}, Z)^T = M^{-1} \cdot w_2 = M^{-1} \cdot [v_1, \ldots, v_n] \cdot \mu$. For fixed values $y_1, \ldots, y_{n-1}$ let us consider all realizations of μ for which $(Y_1, \ldots, Y_{n-1}) = (y_1, \ldots, y_{n-1})$. Then

$$
\begin{aligned}
w_2^T \Delta_z &= \left(M \cdot (y_1, \ldots, y_{n-1}, Z)^T \right)^T \Delta_z \\
&= \sum_{k=1}^{n-2} y_k \cdot a_{i_k}^T \Delta_z + y_{n-1} \cdot \Delta_z^T \Delta_z + Z \cdot a_{i_{n-1}}^T \Delta_z \\
&= y_{n-1},
\end{aligned}
$$

i.e., the value of $w_2^T \Delta_z$ does not depend on the outcome of Z since Δ_z is orthogonal to all a_{i_k}. For Δ_z' we obtain

$$
\begin{aligned}
w_2^T \Delta_z' &= \left(M \cdot (y_1, \ldots, y_{n-1}, Z)^T \right)^T \Delta_z' \\
&= \sum_{k=1}^{n-2} y_k \cdot a_{i_k}^T \Delta_z' + y_{n-1} \cdot \Delta_z^T \Delta_z' + Z \cdot a_{i_{n-1}}^T \Delta_z' \\
&= \underbrace{y_{n-1} \cdot \Delta_z^T \Delta_z'}_{=:\,\kappa} + Z \cdot a_{i_{n-1}}^T \Delta_z'
\end{aligned}
$$

as Δ_z' is orthogonal to all a_{i_k} except for $k = \ell = n - 1$. The chain of equivalences

$$
\begin{aligned}
&\left| \frac{w_2^T \Delta_z}{w_1^T \Delta_z} - \frac{w_2^T \Delta_z'}{w_1^T \Delta_z'} \right| \leq \varepsilon \iff \frac{w_2^T \Delta_z'}{w_1^T \Delta_z'} \in \left[\frac{w_2^T \Delta_z}{w_1^T \Delta_z} - \varepsilon, \frac{w_2^T \Delta_z}{w_1^T \Delta_z} + \varepsilon \right] \\
&\iff w_2^T \Delta_z' \in \left[\frac{w_2^T \Delta_z}{w_1^T \Delta_z} \cdot w_1^T \Delta_z' - \varepsilon \cdot |w_1^T \Delta_z'|, \frac{w_2^T \Delta_z}{w_1^T \Delta_z} \cdot w_1^T \Delta_z' + \varepsilon \cdot |w_1^T \Delta_z'| \right] \\
&\iff Z \cdot a_{i_{n-1}}^T \Delta_z' \in \left[\frac{w_2^T \Delta_z}{w_1^T \Delta_z} \cdot w_1^T \Delta_z' - \kappa - \varepsilon \cdot |w_1^T \Delta_z'|, \frac{w_2^T \Delta_z}{w_1^T \Delta_z} \cdot w_1^T \Delta_z' - \kappa + \varepsilon \cdot |w_1^T \Delta_z'| \right]
\end{aligned}
$$

implies, that for event $\mathbb{F}_\varepsilon$ to occur Z must fall into an interval $I = I(y_1, \ldots, y_{n-1})$ of length $2\varepsilon \cdot \frac{|w_1^T \Delta_z'|}{|a_{i_{n-1}}^T \Delta_z'|}$. The probability of this is bounded from above by

$$
\frac{2n \cdot 2\varepsilon \cdot \frac{|w_1^T \Delta_z'|}{|a_{i_{n-1}}^T \Delta_z'|}}{\delta(r_1, \ldots, r_n) \cdot \min_{k \in [n]} \|r_k\|} = \underbrace{\frac{4n \cdot |w_1^T \Delta_z'|}{\delta(r_1, \ldots, r_n) \cdot \min_{k \in [n]} \|r_k\| \cdot |a_{i_{n-1}}^T \Delta_z'|}}_{=:\,\gamma} \cdot \varepsilon,
$$

Figure 2: Slopes of the vertices of R

where $[r_1, \ldots, r_n] = M^{-1} \cdot [v_1, \ldots, v_n]$. This is due to $(Y_1, \ldots, Y_{n-1}, Z)^{\mathrm{T}} = [r_1, \ldots, r_n] \cdot \mu$ and Theorem 15. Since the vectors $r_1, \ldots, r_n$ are linearly independent, we have $\delta(r_1, \ldots, r_n) > 0$ and $\min_{k \in [n]} \|r_k\| > 0$. Furthermore, $|a_{i_{n-1}}^{\mathrm{T}} \Delta_z'| > 0$ since i_{n-1} is the constraint which is not tight for z_3, but for z_2. Hence, $\gamma < \infty$, and thus $\mathbf{Pr}\left[\left|\frac{w_2^{\mathrm{T}} \Delta_z}{w_1^{\mathrm{T}} \Delta_z} - \frac{w_2^{\mathrm{T}} \Delta_z'}{w_1^{\mathrm{T}} \Delta_z'}\right| \leq \varepsilon\right] \to 0$ for $\varepsilon \to 0$.

As there are at most m^{3n} triples (z_1, z_2, z_3) we have to consider, the claim follows by applying a union bound. $\qquad\square$

Let $p \neq \pi(x_2)$ be a vertex of R. We call the slope s of the edge incident to p to the right of p *the slope of* p. As a convention, we set the slope of $\pi(x_2)$ to 0 which is smaller than the slope of any other vertex p of R.

Let $t \geq 0$ be an arbitrary real, let $\hat{p}$ be the right-most vertex of R whose slope is larger than t, and let $p^\star$ be the right neighbor of $\hat{p}$ (see Figure 2). Let $\hat{x}$ and $x^\star$ be the neighboring vertices of P with $\pi(\hat{x}) = \hat{p}$ and $\pi(x^\star) = p^\star$. Now let $i = i(x^\star, \hat{x}) \in [m]$ be the index for which $a_i^{\mathrm{T}} x^\star = b_i$ and for which $\hat{x}$ is the (unique) neighbor x of $x^\star$ for which $a_i^{\mathrm{T}} x < b_i$. This index is unique due to the non-degeneracy of the polytope P. For an arbitrary real $\gamma \geq 0$ we consider the vector $\tilde{w}_2 = w_2 + \gamma \cdot a_i$.

Lemma 9. *Let $\tilde{\pi} = \pi_{w_1, \tilde{w}_2}$ and let $\tilde{R} = R_{w_1, \tilde{w}_2}$ be the path from $\tilde{\pi}(x_1)$ to $\tilde{\pi}(x_2)$ in the projection $\tilde{P}' = P'_{w_1, \tilde{w}_2}$ of polytope P. Furthermore, let $\tilde{p}^\star$ be the left-most vertex of $\tilde{R}$ whose slope does not exceed t. Then, $\tilde{p}^\star = \tilde{\pi}(x^\star)$.*

Let us reformulate the statement of Lemma 9 as follows: The vertex $\tilde{p}^\star$ is defined for the path $\tilde{R}$ of polygon $\tilde{P}'$ with the same rules as used to define the vertex $p^\star$ of the original path R of polygon P'. Even though R and $\tilde{R}$ can be very different in shape, both vertices, $p^\star$ and $\tilde{p}^\star$, correspond to the same solution $x^\star$ in the polytope P, that is, $p^\star = \pi(x^\star)$ and $\tilde{p}^\star = \tilde{\pi}(x^\star)$. Let us remark that Lemma 9 is a significant generalization of Lemma 4.3 of [4].

Proof. We consider a linear auxiliary function $\bar{w}_2 \colon \mathbb{R}^n \to \mathbb{R}$, given by $\bar{w}_2(x) = \tilde{w}_2^{\mathrm{T}} x - \gamma \cdot b_i$. The paths $\bar{R} = R_{w_1, \bar{w}_2}$ and $\tilde{R}$ are identical except for a shift by $-\gamma \cdot b_i$ in the second coordinate because for $\bar{\pi} = \pi_{w_1, \bar{w}_2}$ we obtain

$$\bar{\pi}(x) = (w_1^{\mathrm{T}} x, \tilde{w}_2^{\mathrm{T}} x - \gamma \cdot b_i) = (w_1^{\mathrm{T}} x, \tilde{w}_2^{\mathrm{T}} x) - (0, \gamma \cdot b_i) = \tilde{\pi}(x) - (0, \gamma \cdot b_i)$$

for all $x \in \mathbb{R}^n$. Consequently, the slopes of $\bar{R}$ and $\tilde{R}$ are exactly the same (see Figure 3a).

Let $x \in P$ be an arbitrary point from the polytope P. Then, $\tilde{w}_2^{\mathrm{T}} x = w_2^{\mathrm{T}} x + \gamma \cdot a_i^{\mathrm{T}} x \leq w_2^{\mathrm{T}} x + \gamma \cdot b_i$. The inequality is due to $\gamma \geq 0$ and $a_i^{\mathrm{T}} x \leq b_i$ for all $x \in P$. Equality holds, among others, for $x = x^\star$ due to the choice of a_i. Hence, for all points $x \in P$ the two-dimensional points $\pi(x)$ and $\bar{\pi}(x)$ agree in the first coordinate while the second coordinate of $\pi(x)$ is at least the second coordinate of $\bar{\pi}(x)$ as $\bar{w}_2(x) = \tilde{w}_2^{\mathrm{T}} x - \gamma \cdot b_i \leq w_2^{\mathrm{T}} x$. Additionally, we have $\pi(x^\star) = \bar{\pi}(x^\star)$. Thus, path $\bar{R}$ is

11

(a) Relation between $\bar{R}$ and $\tilde{R}$ (b) Relation between $\bar{R}$ an R

Figure 3: Relations between R, $\tilde{R}$, and $\bar{R}$

below path R but they meet at point $p^\star = \pi(x^\star)$. Hence, the slope of $\bar{R}$ to the left (right) of $p^\star$ is at least (at most) the slope of R to the left (right) of $p^\star$ which is greater than (at most) t (see Figure 3b). Consequently, $p^\star$ is the left-most vertex of $\bar{R}$ whose slope does not exceed t. Since $\bar{R}$ and $\tilde{R}$ are identical up to a shift of $-(0, \gamma \cdot b_i)$, $\tilde{\pi}(x^\star)$ is the left-most vertex of $\tilde{R}$ whose slope does not exceed t, i.e., $\tilde{\pi}(x^\star) = \tilde{p}^\star$. $\qquad\square$

Lemma 9 holds for any vector $\tilde{w}_2$ on the ray $\vec{r} = \{w_2 + \gamma \cdot a_i : \gamma \geq 0\}$. As $\|w_2\| \leq n$ (see Section 2.2), we have $w_2 \in [-n, n]^n$. Hence, ray $\vec{r}$ intersects the boundary of $[-n, n]^n$ in a unique point z. We choose $\tilde{w}_2 = \tilde{w}_2(w_2, i) := z$ and obtain the following result.

Corollary 10. *Let $\tilde{\pi} = \pi_{w_1, \tilde{w}_2(w_2, i)}$ and let $\tilde{p}^\star$ be the left-most vertex of path $\tilde{R} = R_{w_1, \tilde{w}_2(w_2, i)}$ whose slope does not exceed t. Then, $\tilde{p}^\star = \tilde{\pi}(x^\star)$.*

Note, that Corollary 10 only holds for the right choice of index $i = i(x^\star, \hat{x})$. The vector $\tilde{w}_2(w_2, i)$ is defined for any vector $w_2 \in [-n, n]^n$ and any index $i \in [m]$. In the remainder, index i is an arbitrary index from $[m]$.

We can now define the following event that is parameterized in i, t, and a real $\varepsilon > 0$ and that depends on w_1 and w_2.

Definition 11. *For an index $i \in [m]$ and a real $t \geq 0$ let $\tilde{p}^\star$ be the left-most vertex of $\tilde{R} = R_{w_1, \tilde{w}_2(w_2, i)}$ whose slope does not exceed t and let $y^\star$ be the corresponding vertex of P. For a real $\varepsilon > 0$ we denote by $\mathbb{E}_{i, t, \varepsilon}$ the event that the conditions*

- *$a_i^{\mathrm{T}} y^\star = b_i$ and*

- *$\frac{w_2^{\mathrm{T}}(\hat{y} - y^\star)}{w_1^{\mathrm{T}}(\hat{y} - y^\star)} \in (t, t + \varepsilon]$, where $\hat{y}$ is the neighbor y of $y^\star$ for which $a_i^{\mathrm{T}} y < b_i$,*

are met. Note, that the vertex $\hat{y}$ always exists and that it is unique since the polytope P is non-degenerate.

Let us remark that the vertices $y^\star$ and $\hat{y}$, which depend on the index i, equal $x^\star$ and $\hat{x}$ if we choose $i = i(x^\star, \hat{x})$. For other choices of i, this is, in general, not the case.

Observe that all possible realizations of w_2 from the line $L := \{w_2 + x \cdot a_i : x \in \mathbb{R}\}$ are mapped to the same vector $\tilde{w}_2(w_2, i)$. Consequently, if w_1 is fixed and if we only consider realizations of μ for which $w_2 \in L$, then vertex $\tilde{p}^\star$ and, hence, vertex $y^\star$ from Definition 11 are already determined. However, since w_2 is not completely specified, we have some randomness left for event $\mathbb{E}_{i, t, \varepsilon}$ to occur. This allows us to bound the probability of event $\mathbb{E}_{i, t, \varepsilon}$ from above (see proof of Lemma 13). The next lemma shows why this probability matters.

Lemma 12. *For reals $t \geq 0$ and $\varepsilon > 0$ let $\mathbb{A}_{t,\varepsilon}$ denote the event that the path $R = R_{w_1,w_2}$ has a slope in $(t, t+\varepsilon]$. Then, $\mathbb{A}_{t,\varepsilon} \subseteq \bigcup_{i=1}^{m} \mathbb{E}_{i,t,\varepsilon}$.*

Proof. Assume that event $\mathbb{A}_{t,\varepsilon}$ occurs. Let $\hat{p}$ be the right-most vertex of R whose slope exceeds t, let $p^\star$ be the right neighbor of $\hat{p}$, and let $\hat{x}$ and $x^\star$ be the neighboring vertices of P for which $\pi(\hat{x}) = \hat{p}$ and $\pi(x^\star) = p^\star$, where $\pi = \pi_{w_1,w_2}$. Moreover, let $i = i(x^\star, \hat{x})$ be the index for which $a_i^T x^\star = b_i$ but $a_i^T \hat{x} < b_i$. We show that event $\mathbb{E}_{i,t,\varepsilon}$ occurs.

Consider the left-most vertex $\tilde{p}^\star$ of $\tilde{R} = R_{w_1, \tilde{w}_2(w_2, i)}$ whose slope does not exceed t and let $y^\star$ be the corresponding vertex of P. In accordance with Corollary 10 we obtain $y^\star = x^\star$. Hence, $a_i^T y^\star = b_i$, i.e., the first condition of event $\mathbb{E}_{i,t,\varepsilon}$ holds. Now let $\hat{y}$ be the unique neighbor y of $y^\star$ for which $a_i^T y < b_i$. Since $y^\star = x^\star$, we obtain $\hat{y} = \hat{x}$. Consequently,

$$\frac{w_2^T(\hat{y} - y^\star)}{w_1^T(\hat{y} - y^\star)} = \frac{w_2^T(\hat{x} - x^\star)}{w_1^T(\hat{x} - x^\star)} \in (t, t+\varepsilon],$$

since this is the smallest slope of R that exceeds t and since there is a slope in $(t, t+\varepsilon]$ by assumption. Hence, event $\mathbb{E}_{i,t,\varepsilon}$ occurs since the second condition for event $\mathbb{E}_{i,t,\varepsilon}$ to happen holds as well. $\square$

With Lemma 12 we can now bound the probability of event $\mathbb{A}_{t,\varepsilon}$.

Lemma 13. *For reals $t \geq 0$ and $\varepsilon > 0$ the probability of event $\mathbb{A}_{t,\varepsilon}$ is bounded by $\mathbf{Pr}[\mathbb{A}_{t,\varepsilon}] \leq \frac{4mn^2\varepsilon}{\delta^2}$.*

Proof. Due to Lemma 12 it suffices to show that $\mathbf{Pr}[\mathbb{E}_{i,t,\varepsilon}] \leq \frac{1}{m} \cdot \frac{4mn^2\varepsilon}{\delta^2} = \frac{4n^2\varepsilon}{\delta^2}$ for any index $i \in [m]$.

We apply the principle of deferred decisions and assume that vector $\lambda \in (0,1]^n$ is not random anymore, but arbitrarily fixed. Thus, vector w_1 is already fixed. Now we extend the normalized vector a_i to an orthonormal basis $\{q_1, \ldots, q_{n-1}, a_i\}$ of $\mathbb{R}^n$ and consider the random vector $(Y_1, \ldots, Y_{n-1}, Z)^T = Q^T w_2$ given by the matrix vector product of the transpose of the orthogonal matrix $Q = [q_1, \ldots, q_{n-1}, a_i]$ and the vector $w_2 = [v_1, \ldots, v_n] \cdot \mu$. For fixed values $y_1, \ldots, y_{n-1}$ let us consider all realizations of μ such that $(Y_1, \ldots, Y_{n-1}) = (y_1, \ldots, y_{n-1})$. Then, w_2 is fixed up to the ray

$$w_2(Z) = Q \cdot (y_1, \ldots, y_{n-1}, Z)^T = \sum_{j=1}^{n-1} y_j \cdot q_j + Z \cdot a_i = w + Z \cdot a_i$$

for $w = \sum_{j=1}^{n-1} y_j \cdot q_j$. All realizations of $w_2(Z)$ that are under consideration are mapped to the same value $\tilde{w}_2$ by the function $w_2 \mapsto \tilde{w}_2(w_2, i)$, i.e., $\tilde{w}_2(w_2(Z), i) = \tilde{w}_2$ for any possible realization of Z. In other words, if $w_2 = w_2(Z)$ is specified up to this ray, then the path $R_{w_1, \tilde{w}_2(w_2, i)}$ and, hence, the vectors $y^\star$ and $\hat{y}$ used for the definition of event $\mathbb{E}_{i,t,\varepsilon}$, are already determined.

Let us only consider the case that the first condition of event $\mathbb{E}_{i,t,\varepsilon}$ is fulfilled. Otherwise, event $\mathbb{E}_{i,t,\varepsilon}$ cannot occur. Thus, event $\mathbb{E}_{i,t,\varepsilon}$ occurs iff

$$(t, t+\varepsilon] \ni \frac{w_2^T \cdot (\hat{y} - y^\star)}{w_1^T \cdot (\hat{y} - y^\star)} = \underbrace{\frac{w^T \cdot (\hat{y} - y^\star)}{w_1^T \cdot (\hat{y} - y^\star)}}_{=:\,\alpha} + Z \cdot \underbrace{\frac{a_i^T \cdot (\hat{y} - y^\star)}{w_1^T \cdot (\hat{y} - y^\star)}}_{=:\,\beta} \,.$$

The next step in this proof will be to show that the inequality $|\beta| \geq \frac{\delta}{n}$ is necessary for event $\mathbb{E}_{i,t,\varepsilon}$ to happen. For the sake of simplicity let us assume that $\|\hat{y} - y^\star\| = 1$ since β is invariant under scaling. If event $\mathbb{E}_{i,t,\varepsilon}$ occurs, then $a_i^T y^\star = b_i$, $\hat{y}$ is a neighbor of $y^\star$, and $a_i^T \hat{y} \neq b_i$. That is, by Lemma 5, Claim 3 we obtain $|a_i^T \cdot (\hat{y} - y^\star)| \geq \delta \cdot \|\hat{y} - y^\star\| = \delta$ and, hence,

$$|\beta| = \left| \frac{a_i^T \cdot (\hat{y} - y^\star)}{w_1^T \cdot (\hat{y} - y^\star)} \right| \geq \frac{\delta}{|w_1^T \cdot (\hat{y} - y^\star)|} \geq \frac{\delta}{\|w_1\| \cdot \|\hat{y} - y^\star\|} \geq \frac{\delta}{n \cdot 1} \,.$$

Summarizing the previous observations we can state that if event $\mathbb{E}_{i,t,\varepsilon}$ occurs, then $|\beta| \geq \frac{\delta}{n}$ and $\alpha + Z \cdot \beta \in (t, t+\varepsilon] \subseteq [t-\varepsilon, t+\varepsilon]$. Hence,

$$Z \in \left[\frac{t-\alpha}{\beta} - \frac{\varepsilon}{|\beta|}, \frac{t-\alpha}{\beta} + \frac{\varepsilon}{|\beta|}\right] \subseteq \left[\frac{t-\alpha}{\beta} - \frac{\varepsilon}{\frac{\delta}{n}}, \frac{t-\alpha}{\beta} + \frac{\varepsilon}{\frac{\delta}{n}}\right] =: I(y_1, \dots, y_{n-1}).$$

Let $\mathbb{B}_{i,t,\varepsilon}$ denote the event that Z falls into the interval $I(Y_1, \dots, Y_{n-1})$ of length $\frac{2n\varepsilon}{\delta}$. We showed that $\mathbb{E}_{i,t,\varepsilon} \subseteq \mathbb{B}_{i,t,\varepsilon}$. Consequently,

$$\mathbf{Pr}\left[\mathbb{E}_{i,t,\varepsilon}\right] \leq \mathbf{Pr}\left[\mathbb{B}_{i,t,\varepsilon}\right] \leq \frac{2n \cdot \frac{2n\varepsilon}{\delta}}{\delta(Q^{\mathrm{T}}v_1, \dots, Q^{\mathrm{T}}v_n)} \leq \frac{4n^2\varepsilon}{\delta^2},$$

where the second inequality is due to first claim of Theorem 15: By definition, we have

$$(Y_1, \dots, Y_{n-1}, Z)^{\mathrm{T}} = Q^{\mathrm{T}}w_2 = Q^{\mathrm{T}} \cdot [v_1, \dots, v_n] \cdot \mu = [Q^{\mathrm{T}}v_1, \dots, Q^{\mathrm{T}}v_n] \cdot \mu.$$

The third inequality stems from the fact that $\delta(Q^{\mathrm{T}}v_1, \dots, Q^{\mathrm{T}}v_n) = \delta(v_1, \dots, v_n) \geq \delta$, where the equality is due to the orthogonality of Q (Claim 2 of Lemma 5). □

Lemma 14. *Let Y be the number of slopes of $R = R_{w_1,w_2}$ that lie in the interval $(0,1]$. Then, $\mathbf{E}[Y] \leq \frac{4mn^2}{\delta^2}$.*

Proof. For a real $\varepsilon > 0$ let $\mathbb{F}_\varepsilon$ denote the event from Definition 6. Recall that all slopes of R differ by more than ε if $\mathbb{F}_\varepsilon$ does not occur. Let $Z_{t,\varepsilon}$ be the random variable that indicates whether R has a slope in the interval $(t, t+\varepsilon]$ or not, i.e., $Z_{t,\varepsilon} = 1$ if there is such a slope and $Z_{t,\varepsilon} = 0$ otherwise. Then, for any integer $k \geq 1$

$$Y \leq \begin{cases} \sum_{i=0}^{k-1} Z_{\frac{i}{k}, \frac{1}{k}} & \text{if } \mathbb{F}_{\frac{1}{k}} \text{ does not occur}, \\ m^n & \text{otherwise}. \end{cases}$$

This is true since $\binom{m}{n-1} \leq m^n$ is a worst-case bound on the number of edges of P and, hence, of the number of slopes of R. Consequently,

$$\mathbf{E}[Y] \leq \sum_{i=0}^{k-1} \mathbf{E}\left[Z_{\frac{i}{k}, \frac{1}{k}}\right] + \mathbf{Pr}\left[\mathbb{F}_{\frac{1}{k}}\right] \cdot m^n = \sum_{i=0}^{k-1} \mathbf{Pr}\left[\mathbb{A}_{\frac{i}{k}, \frac{1}{k}}\right] + \mathbf{Pr}\left[\mathbb{F}_{\frac{1}{k}}\right] \cdot m^n$$

$$\leq \sum_{i=0}^{k-1} \frac{4mn^2 \cdot \frac{1}{k}}{\delta^2} + \mathbf{Pr}\left[\mathbb{F}_{\frac{1}{k}}\right] \cdot m^n = \frac{4mn^2}{\delta^2} + \mathbf{Pr}\left[\mathbb{F}_{\frac{1}{k}}\right] \cdot m^n,$$

where the second inequality stems from Lemma 13. The claim follows since the bound on $\mathbf{E}[Y]$ holds for any integer $k \geq 1$ and since $\mathbf{Pr}[\mathbb{F}_\varepsilon] \to 0$ for $\varepsilon \to 0$ in accordance with Lemma 8. □

Proof of Theorem 1. Lemma 14 bounds only the expected number of edges on the path R that have a slope in the interval $(0,1]$. However, the lemma can also be used to bound the expected number of edges whose slope is larger than 1. For this, one only needs to exchange the order of the objective functions $w_1^{\mathrm{T}}x$ and $w_2^{\mathrm{T}}x$ in the projection π. Then any edge with a slope of $s > 0$ becomes an edge with slope $\frac{1}{s}$. Due to the symmetry in the choice of w_1 and w_2, Lemma 14 can also be applied to bound the expected number of edges whose slope lies in $(0,1]$ for this modified projection, which are exactly the edges whose original slope lies in $[1, \infty)$.

Formally we can argue as follows. Consider the vertices $x_1' = x_2$ and $x_2' = x_1$, the directions $w_1' = -w_2$ and $w_2' = -w_1$, and the projection $\pi' = \pi_{w_1', w_2'}$, yielding a path R' from $\pi'(x_1')$ to $\pi'(x_2')$. Let X be the number of slopes of R and let Y and Y' be the number of slopes of R and of R', respectively, that lie in the interval $(0,1]$. The paths R and R' are identical except for the linear

transformation $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$. Consequently, s is a slope of R if and only if $\frac{1}{s}$ is a slope of R' and, hence, $X \le Y + Y'$. One might expect equality here but in the unlikely case that R contains an edge with slope equal to 1 we have $X = Y + Y' - 1$. The expectation of Y is given by Lemma 14. Since this result holds for any two vertices x_1 and x_2 it also holds for x_1' and x_2'. Note, that w_1' and w_2' have exactly the same distribution as the directions the shadow vertex algorithm computes for x_1' and x_2'. Therefore, Lemma 14 can also be applied to bound $\mathbf{E}[Y']$ and we obtain $\mathbf{E}[X] \le \mathbf{E}[Y] + \mathbf{E}[Y'] = \frac{8mn^2}{\delta^2}$. $\qquad\square$

The proof of Corollary 2 follows immediately from Theorem 1 and Claim 4 of Lemma 5.

7 Some Probability Theory

The following theorem is a variant of Theorem 35 from [5]. The two differences are as follows: In [5] arbitrary densities are considered. We only consider uniform distributions. On the other hand, instead of considering matrices with entries from $\{-1, 0, 1\}$ we consider real-valued square matrices. This is why the results from [5] cannot be applied directly.

Theorem 15. *Let $X_1, \ldots, X_n$ be independent random variables uniformly distributed on $(0, 1]$, let $A = [a_1, \ldots, a_n] \in \mathbb{R}^{n \times n}$ be an invertible matrix, let $(Y_1, \ldots, Y_{n-1}, Z)^{\mathrm{T}} = A \cdot (X_1, \ldots, X_n)^{\mathrm{T}}$ be the linear combinations of $X_1, \ldots, X_n$ given by A, and let $I \colon \mathbb{R}^{n-1} \to \{[x, x + \varepsilon] : x \in \mathbb{R}\}$ be a function mapping a tuple $(y_1, \ldots, y_{n-1})$ to an interval $I(y_1, \ldots, y_{n-1})$ of length ε. Then the probability that Z lies in the interval $I(Y_1, \ldots, Y_{n-1})$ can be bounded by*

$$\mathbf{Pr}[Z \in I(Y_1, \ldots, Y_{n-1})] \le \frac{2n\varepsilon}{\delta(a_1, \ldots, a_n) \cdot \min_{k \in [n]} \|a_k\|} \,.$$

Proof. Let us consider the proof of Theorem 35 of [5] for $m = n$ and $k = 1$. We obtain

$$\mathbf{Pr}[Z \in I(Y_1, \ldots, Y_{n-1})] \le \varepsilon \cdot |\det(A^{-1})| \cdot \int_{y \in \mathbb{R}^{n-1}} \max_{z \in \mathbb{R}} f_X(A^{-1} \cdot (y, z)^{\mathrm{T}}) \mathrm{d}y \,,$$

where f_X denotes the common density of the variables $X_1, \ldots, X_n$. In our case, f_X is 1 on $(0, 1]^n$ and 0 otherwise. Note, that in the proof of Theorem 35 matrix A was an integer matrix and so $|\det(A^{-1})| \le 1$. In this proof considering this factor is crucial.

It remains to bound $\int_{y \in \mathbb{R}^{n-1}} \max_{z \in \mathbb{R}} f_X(A^{-1} \cdot (y, z)^{\mathrm{T}}) \mathrm{d}y$. For this we only have to consider the proof of Lemma 36 of [5] since all densities are rectangular functions. Here, we have $\chi = 1$ and $\ell_i = 1$ and $\phi_i = 1$ for any $i \in [n]$. The only point where the structure of matrix A is exploited is where $|\det(PAT)|$ for $P = [\mathbb{I}_{n-1}, \mathbb{O}_{n-1 \times 1}]$ and $T = [e_1, \ldots, e_{i-1}, e_{i+1}, \ldots, e_n]$ for an arbitrary index $i \in [n]$ is bounded. Since $PAT = A_{n,i}$, we obtain

$$\int_{y \in \mathbb{R}^{n-1}} \max_{z \in \mathbb{R}} f_X(A^{-1} \cdot (y, z)^{\mathrm{T}}) \mathrm{d}y \le \chi \cdot \sum_{i \in [n]} \sum_{j=0}^{1} |\det(A_{n,i})| \cdot \prod_{i' \ne i} \ell_{i'}$$

$$= 2 \cdot \sum_{i \in [n]} |\det(A_{n,i})| \,.$$

Summarizing both bounds, we obtain

$$\mathbf{Pr}[Z \in I(Y_1, \ldots, Y_{n-1})] \le 2\varepsilon \cdot |\det(A^{-1})| \cdot \sum_{i \in [n]} |\det(A_{n,i})| = 2\varepsilon \cdot \sum_{i \in [n]} \frac{|\det(A_{n,i})|}{|\det(A)|} \,.$$

We are now going to bound the fraction $\frac{|\det(A_{n,i})|}{|\det(A)|}$. To do this, consider the equation $Ax = e_n$. We obtain

$$|x_i| = \frac{|\det([a_1, \ldots, a_{i-1}, e_n, a_{i+1}, \ldots, a_n])|}{|\det(A)|} = \frac{|\det(A_{n,i})|}{|\det(A)|},$$

where the first equality is due to Cramer's rule and the second equality is due to Laplace's formula. Hence,

$$\mathbf{Pr}\left[Z \in I(Y_1, \ldots, Y_{n-1})\right] \leq 2\varepsilon \cdot \sum_{i \in [n]} \frac{|\det(A_{n,i})|}{|\det(A)|} = 2\varepsilon \cdot \|x\|_1 \leq 2\sqrt{n}\varepsilon \cdot \|x\|_2$$

Now consider the equation $\hat{A}\hat{x} = e_n$ for $\hat{A} = [\mathcal{N}(a_1), \ldots, \mathcal{N}(a_n)]$. Vector $\hat{x} = \hat{A}^{-1}e_n$ is the nth column of the matrix $\hat{A}^{-1}$. Thus, we obtain

$$\|\hat{x}\| \leq \max_{\substack{r \text{ column} \\ \text{of } \hat{A}^{-1}}} \|r\| \leq \frac{\sqrt{n}}{\delta(a_1, \ldots, a_n)},$$

where second inequality is due to Claim 1 of Lemma 5. Due to $A = \hat{A} \cdot \mathrm{diag}(\|a_1\|, \ldots, \|a_n\|)$, we have

$$x = A^{-1}e_n = \mathrm{diag}\left(\frac{1}{\|a_1\|}, \ldots, \frac{1}{\|a_n\|}\right) \cdot \hat{A}^{-1}e_n = \mathrm{diag}\left(\frac{1}{\|a_1\|}, \ldots, \frac{1}{\|a_n\|}\right) \cdot \hat{x}.$$

Consequently, $\|x\| \leq \frac{\|\hat{x}\|}{\min_{k \in [n]} \|a_k\|}$ and, thus,

$$\mathbf{Pr}\left[Z \in I(Y_1, \ldots, Y_{n-1})\right] \leq 2\sqrt{n}\varepsilon \cdot \frac{\|\hat{x}\|}{\min_{k \in [n]} \|a_k\|}$$

$$\leq \frac{2n\varepsilon}{\delta(a_1, \ldots, a_n) \cdot \min_{k \in [n]} \|a_k\|}. \qquad \square$$

References

[1] Nicolas Bonifas, Marco Di Summa, Friedrich Eisenbrand, Nicolai Hähnle, and Martin Niemeier. On sub-determinants and the diameter of polyhedra. In *Proceedings of the 28th ACM Symposium on Computational Geometry (SoCG)*, pages 357–362, 2012.

[2] Karl Heinz Borgwardt. *A probabilistic analysis of the simplex method*. Springer-Verlag New York, Inc., New York, NY, USA, 1986.

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

[4] Tobias Brunsch, Kamiel Cornelissen, Bodo Manthey, and Heiko Röglin. Smoothed analysis of the successive shortest path algorithm. In *Proceedings of the 24th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1180–1189, 2013.

[5] Tobias Brunsch and Heiko Röglin. Improved smoothed analysis of multiobjective optimization. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC)*, pages 407–426, 2012.

[6] George B. Dantzig. *Linear programming and extensions*. Rand Corporation Research Study. Princeton University Press, 1963.

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

[8] Gil Kalai and Daniel J. Kleitman. A quasi-polynomial bound for the diameter of graphs of polyhedra. *Bulletin of the AMS*, 26(2):315–316, 1992.

[9] Victor Klee and David W. Walkup. The d-step conjecture for polyhedra of dimension $d < 6$. *Acta Mathematica*, 117:53–78, 1967.

[10] Denis Naddef. The hirsch conjecture is true for $(0, 1)$-polytopes. *Mathematical Programming*, 45:109–110, 1989.

[11] James B. Orlin. A polynomial time primal network simplex algorithm for minimum cost flows. *Mathematical Programming*, 78(2):109–129, 1997.

[12] Francisco Santos. A counterexample to the hirsch conjecture. *CoRR*, abs/1006.2814, 2010.

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