Short Paths for the Simplex Algorithm

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von M.Sc. Kirill Kukharenko
geb. am 15. November 1997 in Minsk

genehmigt durch die Fakultät für Mathematik
der Otto–von–Guericke–Universität Magdeburg

Gutachter: Prof. Dr. Volker Kaibel
Otto–von–Guericke–Universität Magdeburg

Prof. Dr. Laura Sanità
Università Bocconi, Mailand, Italien

eingereicht am: 30. April 2025
Verteidigung am: 18. August 2025

Contents

List of Figures

List of Tables

List of Algorithms

Zusammenfassung

Der Simplex-Algorithmus ist einer der bekanntesten und am häufigsten verwendeten Algorithmen zur Lösung von linearen Optimierungsproblemen. Obwohl er in der Praxis sehr effizient ist, wurde keine Version des Simplex-Algorithmus jemals theoretisch als ein Polynomialzeitalgorithmus nachgewiesen, im Gegensatz zu seinem Hauptkonkurrenten, der Interior-Point-Methode. Die beiden Rivalen unterscheiden sich in ihrer Ideologie: Während die Interior-Point-Methode, wie der Name schon verrät, durch das Innere des Polyeders der zulässigen Lösungen läuft, bewegt sich der Simplex-Algorithmus entlang der Kanten dieses Polyeders.

Es gibt viele vorangegangene Arbeiten, die zeigen, dass jede bekannte Version des Simplex-Algorithmus auf Instanzen stößt, bei denen der von ihm gefolgte Pfad im zulässigen Polyeder eine exponentielle Anzahl von Kanten enthält. Noch problematischer ist, dass trotz Jahrzehnten der Forschung immer noch nicht bekannt ist, ob der Durchmesser jedes Polytops durch ein Polynom in seiner Dimension und der Anzahl der Facetten begrenzt werden kann (eine Annahme, die als polynomiale Hirsch-Vermutung bekannt ist), was eine notwendige Bedingung für die Existenz einer Polynomialzeitversion des Simplex-Verfahrens darstellt.

In Kapitel 2 dieser Arbeit beschreiben wir eine Konstruktion, die eine Umgehungslösung für dieses Problem bietet. Wir präsentieren eine erweiterte Formulierung, die eine bestimmte entspannte Version der Hirsch-Vermutung etabliert. Wir verbessern unsere Konstruktion für nieder-dimensionale Polytope, passen sie an, um Monotonie zu berücksichtigen, und stellen sicher, dass alle beschriebenen Erweiterungen in stark polynomialer Zeit berechnet werden können. Darüber hinaus reduzieren wir das allgemeine lineare Optimierungsproblem auf Optimierung über den vorgestellten Erweiterungen. Damit beweisen wir, dass, wenn es eine Pivotregel für den Simplex-Algorithmus gibt, bei der man die Anzahl der Schritte durch ein Polynom im Durchmesser des Polyeders der zulässigen Lösungen begrenzen kann, das allgemeine lineare Optimierungsproblem in stark polynomialer Zeit gelöst werden kann.

Zusätzlich zu dem oben erwähnten Problem der exponentiell langen Pfade wird der Simplex-Algorithmus durch Degeneration behindert. In der Tat kann er an einer degenerierten Ecke des zulässigen Polyeders für exponentiell viele aufeinanderfolgenden Iterationen „festhängen“. Während es viele vorangegangene Arbeiten gibt, die zeigen, wie dieses Phänomen für verschiedene spezielle Klassen von linearen Optimierungsproblemen vermieden werden kann, wurde noch kein einheitlicher Ansatz vorgeschlagen.

In Kapitel 3 beweisen wir, dass es immer möglich ist, die Anzahl der aufeinanderfolgenden degenerierten Pivots, die der Simplex-Algorithmus ausführt, auf $n - m - 1$ zu begrenzen, wobei n die Anzahl der Variablen und m die Anzahl der Gleichungen eines gegebenen linearen Programms im Gleichungsformat ist. Wir erhalten auch eine Schranke für die Gesamtanzahl der Simplex-Pivots und zeigen, dass diese in der Tat stark polynomial für bestimmte Klassen von kombinatorischen LPs ist.

Abstract

The simplex algorithm is one of the most popular and widely used algorithms for solving linear programming problems. Although being very efficient in practice, no version of the simplex was ever proven to be a polynomial time algorithm from the theoretical prospective, in contrast to its main competitor, the interior point method. The two rivals differ in ideology; while the interior point method, as the name suggests, runs through the interior of the polyhedron of feasible solutions, the simplex proceeds along the edges of it.

There has been a lot of prior work showing that every popular simplex version runs into instances where the path that it follows on the feasible polyhedron contains an exponential number of edges. Ever worse than that, despite decades of research, is is still not known whether the diameter of every polytope can be bounded by a polynomial in its dimension and the number of facets (an assumption known as the polynomial Hirsch conjecture), which is a necessary condition for the existence of a polynomial time simplex method.

In Chapter 2 of this work, we describe a construction that offers a workaround for the latter issue. We present an extended formulation that establishes a certain relaxed version of the Hirsch conjecture. We improve our construction for low-dimensional polytopes, modify it to account for monotonicity, and ensure that all described extensions can be computed in strongly polynomial time. Moreover, we reduce the general linear programming problem to optimization over the presented extensions. With that we prove that if there is a pivot rule for the simplex algorithm for which one can bound the number of steps by a polynomial in the diameter of the polyhedron of feasible solutions, then the general linear programming problem can be solved in strongly polynomial time.

In addition to the issue of exponentially long paths mentioned above, the simplex algorithm is obstructed by degeneracy. In fact, it can "get stuck" at a degenerate vertex of the feasible polyhedron for exponentially many consecutive iterations. While there have been many prior works showing how to avoid this phenomenon for a number of special classes of linear programs, no unified approach has ever been suggested.

In Chapter 3 we prove that it is always possible to limit the number of consecutive degenerate pivots that the simplex algorithm performs to $n - m - 1$, where n is the number of variables and m is the number of equality constraints of a given linear program in standard equality form. We also obtain a bound on the total number of simplex pivots and show that it is, in fact, strongly polynomial for certain classes of combinatorial LPs.

1 Introduction and motivation

Linear programming (LP) is the problem of finding a point within a given polyhedral region that optimizes a specified linear function. Mathematically we write

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax \le b \end{aligned} \tag{1.1}$$

where $x \in \mathbb{R}^d$ is a variable vector, $c \in \mathbb{R}^d$ is the objective vector, and $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$ are the system matrix and the right hand side, respectively, of the system of linear inequalities $Ax \le b$. The *feasible region* $\{x \in \mathbb{R}^d \mid Ax \le b\}$ of a linear programming problem is a *polyhedron* (a finite intersection of closed half-spaces induced by the row inequalities of the system). The polyhedral structure of the feasible region, together with the linearity of the objective function, calls for the invention of optimization techniques specifically tailored for linear programming. Historically the first, and still one of the most widely used methods for solving linear programming problems is the simplex algorithm introduced by Dantzig (1951).

Geometrically, the simplex algorithm proceeds from vertex to vertex of the feasible polyhedron along the edges of it, improving the objective function value with each move, until it reaches an optimal vertex. See Figure 1.1 for an illustration.

Behind the scenes, the simplex algorithm relies on the concept of *basis*, where a basis corresponds to an inclusion-wise minimal set of tight inequalities which defines a vertex of the underlying feasible polyhedron. In each step, it performs a basis exchange by replacing one constraint currently in the basis with a different one. Such an exchange is called a *pivot*. From a geometric perspective, a basis exchange identifies a direction to move from the current vertex. The simplex algorithm only considers pivots that yield an *improving* direction, with respect to the objective function to be optimized. As the result of pivoting, it is possible for the algorithm to either stay at the same vertex, or to move to an adjacent vertex of the underlying polyhedron with a better objective function value. We refer to a pivot of the former type as *degenerate*, in contrast to a pivot of the latter type, which will be called *non-degenerate*. Refer to Figure 1.2 for an illustration of the both types of pivots. The decision on how to perform these basis

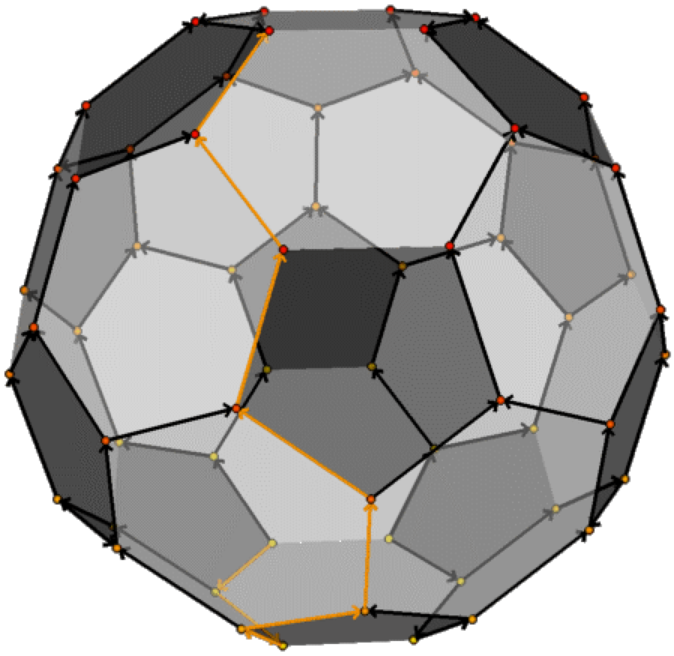

Fig. 1.1: A simplex path in orange on a polytope. Image: Prof. Dr. Marc Pfetsch, TU Darmstadt.

exchanges (i.e., which inequality leaves and which inequality enters the basis) is made by the *pivot rule* which constitutes the core of the simplex algorithm.

Even though the simplex algorithm performs as a linear time algorithm for many real world linear programming models (see, e.g., the survey of Shamir, 1987), none of the numerous pivot rules that have been introduced over decades yields a polynomial time version of it. It should be noted, that a variant of the simplex algorithm, the shadow simplex, has been shown to have polynomial smoothed complexity (see, e.g., Bach and Huiberts (2025); Borgwardt (1987, 1999); Dadush and Huiberts (2020); Megiddo (1986a); Spielman and Teng (2004); Vershynin (2009)).

The search for a polynomial time pivot rule is also considered highly relevant in light of the question of whether there is a *strongly polynomial* time algorithm for linear programming (i.e., an algorithm for which not only the number of bit-operations can be bounded by a polynomial in the entire input length, but also the number of its arithmetic operations can be bounded by a polynomial in the number of inequalities and in the number of variables), which is most prominent in Smale's list of 18 open problems for the 21st century (Smale, 1998).

What does actually prevent the simplex from being a polynomial time algorithm? The two main obstructions are

(1) the length (i.e., the number of edges) of a path that the simplex takes on the feasible polyhedron might be exponential,
(2) degeneracy, causing so-called cycling and stalling.

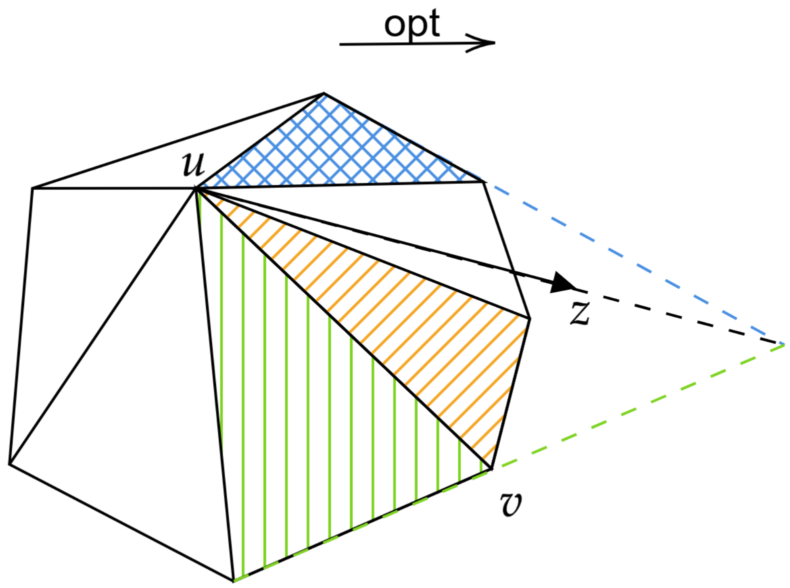

Fig. 1.2: The simplex algorithm optimizing the depicted optimization direction opt is currently at the vertex u of the pyramid with the basis consisting of the three inequalities corresponding to the colored facets. Here will we refer to an inequality by the color of the corresponding facet. If the blue inequality leaves the basis, the simplex "slides down" the edge formed by the green and the orange facets until it "bumps" into the lower facet of the pyramid at the vertex v. The new basis obtained by this *non-degenerate* pivot is comprised of the inequalities defining the green, the orange, and the lower facets. If, on the other hand, the orange inequality leaves the basis, the simplex tries to move along the direction z formed by the green and the blue inequalities. However this direction is outside of the feasible polyhedron, and the movement along it is obstructed by the white facet located in between the orange and the blue ones. Therefore the simplex stays at the vertex u. The inequality defining the said white facet, together with the green and the blue inequalities constitute the new basis after this *degenerate* pivot.

The first issue is known to affect all popular pivot rules. They are cleverly tricked into visiting an exponential number of vertices of some concrete "twisted" polytopes (see, e.g., Avis and Friedmann (2017); Black (2024); Disser and Hopp (2019); Goldfarb (1994); Goldfarb and Sit (1979); Hansen and Zwick (2015); Klee and Minty (1972); Terlaky and Zhang (1993); Zadeh (2009) and the references therein) before reaching an optimal one. Even worse than that, it is not known whether a path of polynomial (in the dimension and the number of facets) length between the starting vertex and an optimal one always exists in polytopes.

The second point refers to situations when the linear programming problem is *degenerate*, meaning the feasible polyhedron has at least one *degenerate vertex*, i.e., a vertex, such that the number of facets that contain it is larger than the dimension of the polyhedron (the apex of the pyramid in Figure 1.2 is an example for that). Whenever the simplex algorithm encounters such a vertex, it can potentially fall into an exponentially long sequence of degenerate pivots, termed *stalling* or, even worse, loop indefinitely, a phenomenon called *cycling*.

Our work, titled Short Paths for the Simplex Algorithm, is dedicated to studying the two mentioned issues and obtaining results that at least somewhat lower the hurdles preventing the simplex from becoming a (strongly) polynomial time algorithm. This work is structured as follows. We work on problem (1) in Chapter 2 by introducing a construction that allows us to guarantee the existence of a path of linear (in the number of inequalities) length between the starting vertex and an optimal vertex of a feasible polyhedron, without resorting to degeneracy or blowing up the problem size too much. We modify our construction to account for monotonicity and improve it for two- and three-dimensional polytopes. Moreover, we prove that if there is a pivot rule for the simplex algorithm for which one can bound the number of steps by a polynomial in the diameter of the polyhedron of feasible solutions, then the general linear programming problem can be solved in strongly polynomial time. After that, we switch to the problem (2) in Chapter 3. We show that it is always possible to bound the number of consecutive degenerate pivots that the simplex algorithm performs by $n-m-1$, where n is the number of variables and m is the number of equality constraints of a given linear program in standard equality form (which will be formally defined later). As a result, we ensure the existence of a short (monotone) path leading out of degeneracy in the *basis exchange graph* of a polytope. Moreover, we provide a bound on the total number of simplex pivots and test our pivot rule on a benchmark LP dataset.

Chapter 2 is based on the work of Kaibel and Kukharenko (2024) while most of the results in Chapter 3 are presented in Kukharenko and Sanità (2024).

In the remainder of this chapter, the reader will find notations and definitions used throughout this work.

1.1 Preliminaries

In this section we introduce our basic notations without going into much detail. If appropriate, we point the reader to the relevant literature. Concepts that are only relevant for certain parts of the work may be defined locally.

Throughout this work, we assume familiarity with basic facts about linear algebra, graph theory, optimization, and polyhedra. For detailed background information we refer to the books of Schrijver (1986) and Ziegler (1994). The basic concepts of algorithm complexity follow the standard textbook of Garey and Johnson (1990).

We use $\mathbb{O}_d$ and $\mathbb{1}_d$ for the all-zeros and all-ones vectors in $\mathbb{R}^d$, respectively. When the subscript is clear from the context, it will be dropped. For a row-vector $\alpha \in \mathbb{R}^{1\times d} \setminus \{\mathbb{O}^T\}$ and a number $\beta \in \mathbb{R}$ we call the sets $H^{\leq}(\alpha,\beta) := \{x \in \mathbb{R}^d \mid \alpha x \leq \beta\}$ and $H^{=}(\alpha,\beta) := \{x \in \mathbb{R}^d \mid \alpha x = \beta\}$ a *halfspace* and a *hyperplane*, respectively. Moreover, we naturally extend the above notation by $H^{\sigma}(\alpha,\beta)$ to denote the set $\{x \in R^d \mid \alpha x \, \sigma \, \beta\}$ where $\sigma \in \{<,>\}$.

For $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$ we use $P^{\leq}(A,b)$ to denote the *polyhedron* $\{x \in \mathbb{R}^d \mid Ax \leq b\}$. For a matrix $A \in \mathbb{R}^{m\times d}$, a row subset $I \subseteq [m]$, and a column subset $J \in [d]$, we use $A_{I,J}$ to denote the submatrix of A indexed by the corresponding rows and columns. Instead of $A_{I,[d]}$ and $A_{[m],J}$ we also write $A_{I,\star}$ and $A_{\star,J}$, respectively, and for $I = \{i\}, J = \{j\}$ where $i \in [m]$ and $j \in [d]$ we use $A_{i,\star}$, $A_{j,\star}$, and A_{ij}. We use Δ_A for the largest absolute value of a sub-determinant of A and $\Delta_{A,k}$ with $1 \leq k \leq \min\{m,d\}$ when restricted to $k\times k$ sub-determinants. Similarly, for a vector $b \in \mathbb{R}^m$ and $I \subseteq [m]$, the notation b_I is used to denote a vector consisting of the entries of b indexed by I. We use $||b||_1$ and $||b||_2$ to denote the 1-norm and the Euclidean norm of b, respectively. The Euclidean (straight line) distance between two points $u, v \in \mathbb{R}^d$ is given by $||u-v||_2$.
We use the concept of encoding size in accordance with Schrijver (1986)[p.15].

A polyhedron that has vertices is called *pointed*. A bounded polyhedron is called a *polytope*. Note that every polytope is naturally a pointed polyhedron. We call a d-dimensional polytope a *d-polytope*. A d-dimensional polytope $P \subset \mathbb{R}^d$ is called *full-dimensional*. A d-polytope is called *simple* if each of its vertices is contained in exactly d facets. We denote the *convex hull* of a set $S \subseteq \mathbb{R}^d$ by conv$\{S\}$.

The *diameter* of a polytope P is the smallest number κ such that in the *graph* of P formed by its vertices and one-dimensional faces (*edges*) of P every pair of vertices is connected by a path with at most κ edges.

The next concept play a central role in this work. *Extended formulation* of a polytope $P \subseteq \mathbb{R}^d$ is a pair (Q,π), where Q is a higher-dimensional polyhedron called an *extension* of P and π is an affine map called *projection*, such that $\pi(Q) = P$. For an illustration see Figure 1.3. Extended formulations play an important role in "reducing complexity" of linear programming problems in various ways, since the lifted problem might be easier, in some sense, than the original one.

In this work we write linear programming problems in *inequality form* (1.1) in Chapter 2 and in the so-called *standard equality form*

$$\begin{aligned} \min \quad & c^T x \\ s.t. \quad & Ax = b \\ & x \geq 0 \end{aligned} \tag{1.2}$$

in Chapter 3. Although many authors define LP in standard equality form as a maximization problem, we still use this wording for (1.2), since simply multiplying our

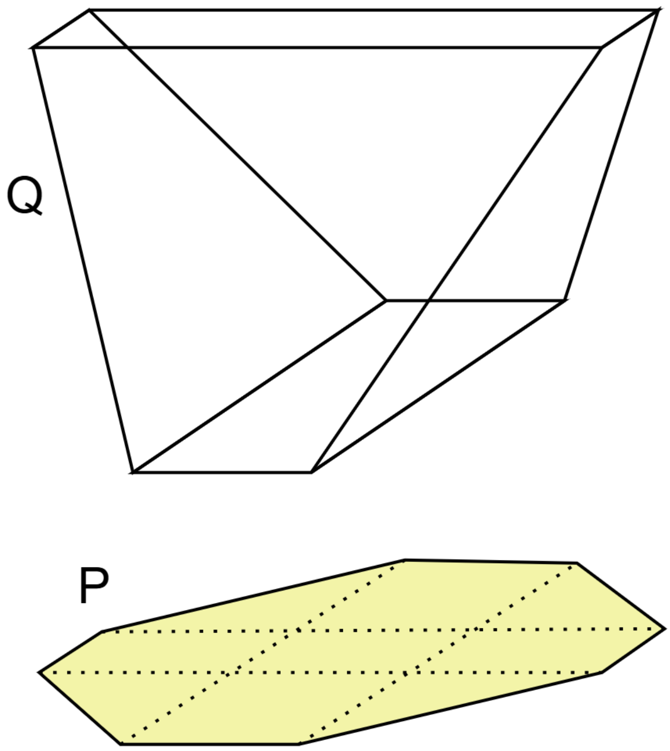

Fig. 1.3: A three-dimensional simple extension Q of the two-dimensional P. Note that the diameter of Q is by one smaller than the diameter of the original polytope P.

objective c by -1 brings our LP to this standard representation. Note that (1.2) is a special case of (1.1) (after writing $Ax = b$ as $Ax \le b$ and $-Ax \le -b$, and the non-negativity constraint as $-x \le 0$). Conversely, a problem in the inequality form can be transformed into the standard equality form by splitting the unknowns $x = x^+ - x^-$, where $x^+, x^- \ge 0$ and introducing a slack variable for each inequality.

When dealing with both forms of linear programs, (1.1) and (1.2), we naturally work with two types of systems of linear inequalities and equalities:

$$Ax \le b \tag{1.3a}$$

$$Ax = b, x \ge 0 \tag{1.3b}$$

Both of these systems have their own concepts of *basis* and *basic feasible* or *infeasible solution.* We introduce these notions next. A *basis,* also called a *row basis,* of a system (1.3a)/of an LP (1.1) with an $m \times d$-matrix A and $\operatorname{rank}(A) = d$ is a subset $B \subseteq [m]$ with $|B| = d$ such that the submatrix $A_{B,\star}$ of A formed by the rows of A indexed by B is non-singular. Such a basis defines the *basic solution* $x^B = A_{B,\star}^{-1} b_I$ of the system (1.3a).

A *basis,* also called a *column basis,* of a system (1.3b)/of an LP (1.2) with an $m \times n$-matrix A and $\operatorname{rank}(A) = m$ is a subset $B \subseteq [n]$ with $|B| = m$ and $A_{\star,B}$ being non-singular. The point $x \in \mathbb{R}^n$ with $x_B = A^{-1}_{\star,B} b$, $x_N = \mathbb{O}$ where $N := [n] \setminus B$ is a *basic solution* of (1.3b) with basis B.

If a basis B with a basic solution x of a system (1.3a) or (1.3b) satisfies all inequalities of the said system, both the basic solution x and the basis B are called *feasible*, and *infeasible* otherwise. Note that the feasible basic solutions are exactly the vertices of the polyhedra defined by (1.3a) and (1.3b), respectively.

Since a basis of a system (1.3a)/(1.3b) depends only on the matrix A, sometimes we will refer to it as simply a basis of A, if the associated system is clear from the context. Note that bases of systems (1.3a) and (1.3b) differ in concept. While a basis of (1.3a) only contains inequalities, that are tight at the defined vertex, a basis of (1.3b) can be associated with a subset of variables having their non-negativity constraints not (necessarily) tight. To avoid confusion, from now on we will refer to a basis of $Ax \le b$ as a *row basis*. The word *basis* will be reserved for column bases of systems (1.3b).

A row/an inequality of (1.3a) is called *redundant*, if its deletion does not affect the feasible set. A system $Ax \le b$ without redundant rows is called *irredundant*. Observe that a d-dimensional polytope in $\mathbb{R}^d$ defined by an irredundant system $Ax \le b$ is simple if and only if each vertex of it is defined by exactly one row basis.

Let x^B be a basic feasible solution of (1.3a) with basis B, then $Eq_{Ax \le b}(x^B) := \{i \in [m] \mid A_{i,\star} x^B = b_i\}$. The *feasible* and the *basic cones* at x^B are the sets $C\big(Eq_{Ax\le b}(x^B)\big) := \{x \in \mathbb{R}^d \mid A_{Eq_{Ax\le b}(x^B),\star} x \le \mathbb{O}\}$ and $C(B) := \{x \in \mathbb{R}^d \mid A_{B,\star} x \le \mathbb{O}\}$, respectively.

Finally, we briefly turn to the concept of LP duality. For more information see, e.g., Dantzig (1963). For a pair of primal and dual problems

$$\text{(P)} \qquad \begin{aligned} \max \quad & c^T x \\ s.t. \quad & Ax \le b \end{aligned} \tag{1.4}$$

$$\text{(D)} \qquad \begin{aligned} \min \quad & y^T b \\ s.t. \quad & A^T y = c \\ & y \ge 0 \end{aligned} \tag{1.5}$$

we use the concept of *weak duality*: $c^T x = y^T A x \le y^T b$ holds for any pair (x, y) of feasible solutions to (P) and (D), respectively. If the latter inequality is tight, then x and y are optimal solutions to (P) and (D), respectively.

2

Polytope extensions with linear diameters

2.1 Introduction

As we've already mentioned in the previous chapter, the simplex algorithm proceeds along (monotone) paths in the graph of a polytope. And therefore, for each polytope P the worst-case running time of the simplex algorithm (over all linear objective functions) is bounded from below by the diameter of P. Hence, the diameter of polytopes is necessarily bounded by a polynomial in the number of its facets if a polynomial time pivot rule for the simplex algorithm for linear programming exists.

Warren M. Hirsch conjectured in 1957 (see, e.g., Ziegler, 1994) that the diameter of each d-dimensional polytope with n facets is bounded from above by $n-d$. Disproving this bound took substantial effort and was achieved only 53 years later by Santos (2012) using a polytope in dimension 43 with 86 facets and diameter 44. Today, it is known that no upper bound better than $\frac{21}{20}(n-d)$ is valid in general (Matschke et al., 2015). The belief, that the diameter of polytopes is bounded by a polynomial in d and n is called *polynomial Hirsch conjecture*. The best-known upper bounds in terms of n and d are are derived from a result by Kalai and Kleitman (1992), who presented an upper bound of $n^{\log_2 d+2}$. Todd (2014) improved the latter bound to $(n-d)^{\log_2 d}$, which was further refined by Sukegawa (2019) to $(n-d)^{\log_2 O(d/\log_2 d)}$. Another line of research, which was carried out by Bonifas, di Summa, Eisenbrand, Hähnle, and Niemeier (2014), lead to the upper bound $\mathrm{O}\left(\Delta_A^2 n^{3.5} \log_2(n\Delta_A)\right)$, where A is the integral coefficient matrix of some inequality description of a rational polytope P. The latter result was improved to $\mathrm{O}\left(n^3\Delta_A^2 \ln(n\Delta_A)\right)$ by Dadush and Hähnle (2016).

While not presenting a new bound on the diameters of polytopes, the first main contribution (see Theorem 2.7, and in particular Corollary 2.8) we make is to prove the following: for each d-dimensional polytope P in $\mathbb{R}^d$ with n facets that satisfies a certain non-degeneracy assumption there is a simple $(d+1)$-dimensional extension Q with $n+1$ facets and diameter at most $2(n-d)$. We further show in Theorem 2.14 that such an extension Q is even computable in strongly polynomial time if a vertex of P is specified within the input.

We remark that without requiring the number of facets and the dimension of Q to be polynomially bounded in n and d, the polytope Q can trivially be chosen as a high-dimensional simplex (which even has diameter one). However, the number of facets of that simplex equals the number of vertices of P (which might easily be exponential in n and d). Similarly, without the requirement of Q being simple such a construction can trivially be obtained by forming a pyramid over P (which has diameter at most two). On the other hand, the results in Kaibel and Walter (2015) show that the combination of those two requirements on Q implies some (non-degeneracy) condition on P.

Our second main contribution is to use the extensions of small diameters that we described in the first part in order to show the following: in order to devise a strongly polynomial time algorithm for the general linear programming problem it suffices to find a polynomial time pivot rule for the simplex algorithm just for the class of linear programs whose feasible region is a simple polytope whose diameter is bounded linearly in the number of inequalities (see Theorems 2.21 and 2.24). Thus, even if it turns out that the polynomial Hirsch conjecture fails, it still might be possible to come up with a strongly polynomial time algorithm for general linear programming by devising a polynomial time pivot rule for only that special class of problems.

This chapter is organized as follows. Section 2.2 introduces a special type of extended formulations that we call *rock extensions* which will allow us to realize the claimed diameter bounds. Special properties of rock extensions for two- and three-dimensional polytopes are discussed in Section 2.3. In Section 2.4 we ensure that the procedure we devise in Section 2.2 for obtaining a rock extension with certain additional properties (that we need to maintain in our inductive construction) can be adjusted to produce a rational extension having its encoding size polynomially bounded in the encoding size of the input. We consider a reduction of the general linear programming problem to its special case for rock extensions in Section 2.5 and upgrade our extensions to allow for short monotone paths in Section 2.6.

2.2 Rock extensions

Let $Ax \le b$ be a system of linear inequalities with $A \in \mathbb{R}^{m \times d}, b \in \mathbb{R}^m$. Then we call the family of hyperplanes $H^=(A_{1,\star}, b_1), \ldots, H^=(A_{m,\star}, b_m)$ the *hyperplane arrangement associated with $Ax \le b$* and denote it by $\mathcal{H}(A,b)$. *Vertices* and *lines* of a hyperplane arrangement are 0-dimensional and 1-dimensional intersections of its hyperplanes, respectively. The polyhedron $P^{\le}(A,b)$ is called a *chamber* of $\mathcal{H}(A,b)$.

2.2.1 Foundations

We start by introducing two types of systems of linear inequalities which will be crucial throughout this chapter.

Definition 2.1. *A feasible system of linear inequalities $Ax \le b$ with $A \in \mathbb{R}^{m\times d}$, $b \in \mathbb{R}^m$ is said to be* ***non-degenerate*** *if each vertex of $\mathcal{H}(A,b)$ is contained in exactly d of the m hyperplanes. The system is called* ***totally non-degenerate*** *if, for any collection of k hyperplanes of $\mathcal{H}(A,b)$, their intersection is a $(d-k)$-dimensional affine subspace for $1 \le k \le d$ and the empty set for $k > d$.*

Note that total non-degeneracy implies non-degeneracy. Additionally, observe that non-degeneracy can be achieved using perturbation arguments. We elaborate on that in Section 2.5 in more detail. We introduce corresponding notions for polytopes in the following way.

Definition 2.2. *A polytope is called* ***strongly non-degenerate****, respectively,* ***totally non-degenerate*** *if there is a* ***non-degenerate****, respectively,* ***totally non-degenerate*** *system of linear inequalities defining it.*

We observe that each strongly non-degenerate polytope is full-dimensional and simple.

Definition 2.3. *A non-degenerate system $Ax \le b$ with $A \in \mathbb{R}^{m\times d}, b \in \mathbb{R}^m$ is said to be* ***simplex-containing*** *if there exists a subset $I \subseteq [m]$ of with $|I| = d+1$ such that $P^{\le}(A_{I,\star}, b_I)$ is a d-simplex.*

Note that each strongly non-degenerate polytope P can be described by a simplex-containing non-degenerate system $Ax \le b$. This is due to the fact that one can add $d+1$ redundant inequalities defining a simplex $S \supseteq P$ to any non-degenerate description of P without violating non-degeneracy (in fact, later we establish that a single auxiliary inequality is enough to ensure the simplex-containing property). In addition, it turns out that any *totally* non-degenerate system defining a polytope is simplex-containing. We proceed with a proof of this fact.

Proposition 2.4. *Let $P \subseteq \mathbb{R}^d$ be a d-polytope given by a totally non-degenerate system $Ax \le b$ of m linear inequalities. There exists a subset $I \subseteq [m]$ with $|I| = d+1$ such that the polyhedron $P^{\le}(A_{I,\star}, b_I)$ is bounded.*

Proof. Consider an inequality $A_{i,\star}x \le b_i$ that defines a facet F_i of P. First, we show that the vertex $v \in P$ that minimizes $A_{i,\star}x$ over P is unique. For the sake of contradiction, assume that a k-dimensional face F with $k \ge 1$ minimizes $A_{i,\star}$. Note that $F \not\subset F_i$ holds due to full-dimensionality of P, and hence the intersection of $d-k$ hyperplanes $H^{=}(A_{j,\star}, b_j)$, $j \in J \subseteq [m], |J| = d-k, k \ge 1$, $i \notin J$ containing F and the hyperplane

$H^{=}(A_{i,\star}, b_i)$ is empty, which contradicts total non-degeneracy. Then, the d inequalities in $Ax \le b$ satisfied at equality by v and the inequality $A_{i,\star}x \le b_i$ define a simplex around P, since each edge containing v is $A_{i,\star}$-increasing and therefore each extreme ray of the feasible cone emanating from v intersects $H^{=}(A_{i,\star}, b_i)$. □

Next, we introduce a special type of extensions we will be working with.

Definition 2.5. *Let P be the polytope defined by a system $Ax \le b$ with $A \in \mathbb{R}^{m\times d}$, $b \in \mathbb{R}^m$. Any polytope $Q := \{(x,z) \in \mathbb{R}^{d+1} \mid Ax + az \le b, z \ge 0\}$ with $a \in \mathbb{R}^m_{>0}$ will be called a **rock extension** of P.*

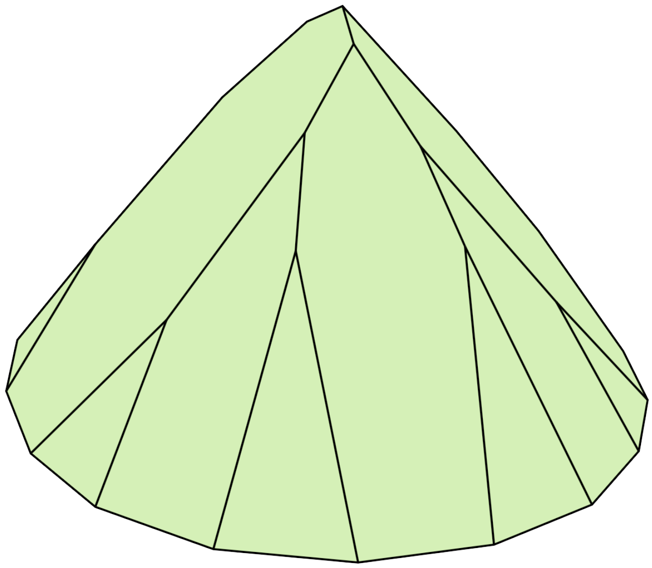

Fig. 2.1: A rock extension with diameter eight of the regular 20-gon.

Note that a rock extension Q together with the orthogonal projection onto the first d coordinates indeed provides an extended formulation of P. We henceforth assume that P is a full-dimensional d-polytope. Then Q is a $(d+1)$-dimensional polytope that has at most $m+1$ facets, including the polytope P itself (identified with $P \times \{0\}$) as the facet defined by the inequality $z \ge 0$. In case $Ax \le b$ is an irredundant description of P, a rock extension Q has exactly $m+1$ facets defined by $z \ge 0$ and $A_{i,\star}x + a_i z \le b_i$ for $i \in [m]$, where the latter m inequalities are in one-to-one correspondence with the facets of P. See Figure 2.1 for an illustration.

We call the facet P of Q the *base* and partition the vertices and edges of Q into *base* and *non-base* vertices and edges accordingly. A vertex of Q with maximal z-coordinate is called a *top vertex*. A path in the graph of a rock extension will be called *z-increasing* if the sequence of z-coordinates of vertices along the path is strictly increasing. To shorten our notation, we denote a hyperplane $\{(x,z) \in \mathbb{R}^{d+1} \mid z = h\}$ and a halfspace $\{(x,z) \in \mathbb{R}^{d+1} \mid z \le h\}$ by $\{z = h\}$ and $\{z \le h\}$, respectively. We also use the notation $B^d_\epsilon(q)$ for the d-dimensional open Euclidean ball of radius ϵ centered in $q \in \mathbb{R}^d$.

Definition 2.6. *Let $\epsilon > 0$ be a positive number. We say that a rock extension Q of P is* ***ϵ-concentrated*** *around $(o, h) \in \mathbb{R}^d \times \mathbb{R}_{>0}$ if*

1. *(o, h) is the unique top vertex of Q,*
2. *$B^d_\epsilon(o) \subseteq P$, and*
3. *all non-base vertices of Q are contained in the open ball $B^{d+1}_\epsilon\big((o, h)\big)$.*

It turns out that maintaining the ϵ-concentrated property opens the door for inductive constructions of "well-behaved" rock extensions.

2.2.2 The main theorem of rock extensions

We proceed with the result that provides the foundation for the further development of the theory of rock extensions.

Theorem 2.7. *For every d-polytope P given by a simplex-containing non-degenerate system $Ax \le b$ of m linear inequalities, every $\epsilon > 0$, and every point o with $B^d_\epsilon(o) \subseteq P$, there exists a simple rock extension Q that is ϵ-concentrated around $(o, 1)$ so that for each vertex of Q there exists a z-increasing path of length at most $m - d$ to the top vertex $(o, 1)$.*

For *totally* non-degenerate polytopes the latter result immediately implies the following bound that is only twice as large as the bound originally conjectured by Hirsch.

Corollary 2.8. *Each totally non-degenerate d-polytope P with n facets admits a simple $(d + 1)$-dimensional extension Q with $n + 1$ facets and diameter at most $2(n - d)$.*

For a more general result for all strongly non-degenerate polytopes along with algorithmic considerations see Section 2.5. Now we turn to the proof of Theorem 2.7.

Proof of Theorem 2.7. We proceed by induction on the number m of linear inequalities in $Ax \le b$.

Suppose first that we have $m = d + 1$. Then the polytope P is a d-simplex and hence the $(d + 1)$-dimensional pyramid Q over P with $(o, 1)$ as the top vertex has the required properties.

So let us consider the case $m \ge d + 2$. Since $Ax \le b$ is simplex-containing, there exists an inequality $A_{i,\star} x \le b_i$ ($i \in [m] \setminus I$ can be chosen arbitrarily for some I as in Definition 2.3), whose deletion from $Ax \le b$ yields a system of linear inequalities defining a polyhedron $\widetilde{P}$. Note that $\widetilde{P}$ is bounded due to the simplex-containing property of $Ax \le b$. By the induction hypothesis and due to $B^d_\epsilon(o) \subseteq P \subseteq \widetilde{P}$, for every $0 < \mu \le \epsilon$ the polytope $\widetilde{P}$ defined by the simplex-containing non-degenerate system $A_{J,\star}x \le b_J$ with $J := [m] \setminus \{i\}$ admits a simple rock extension $\widetilde{Q}$ that is μ-concentrated around $(o,1)$ with each vertex having a z-increasing path of length at most $m-d-1$ to the top vertex $(o,1)$ of $\widetilde{Q}$.

To complete the proof we pick a certain $0 < \mu < \epsilon$ and add the inequality $A_{i,\star}x + a_i z \le b_i$ with an appropriate choice of a_i to the inequality description of the μ-concentrated extension $\widetilde{Q}$ of $\widetilde{P}$ in order to obtain a simple rock extension Q of P that is ϵ-concentrated around $(o,1)$. We then show that the vertices of Q admit similar paths to the top vertex as the vertices of $\widetilde{Q}$ do.

Here we choose the coefficient $a_i > 0$ that determines the "tilt angle" of the corresponding hyperplane in such a way that $H^{=}\big((A_{i,\star}, a_i), b_i\big)$ is tangential to $B_\mu^{d+1}\big((o,1)\big)$ with $B_\mu^{d+1}\big((o,1)\big) \subseteq H^{\le}\big((A_{i,\star}, a_i), b_i\big)$, which is possible since $B_\mu^d(o) \subsetneq B_\epsilon^d(o) \subseteq P$ due to $\mu < \epsilon$. Then the inequality $A_{i,\star}x + a_i z \le b_i$ will not cut off any non-base vertices from $\widetilde{Q}$ (as they are all contained in $B_\mu^{d+1}\big((o,1)\big)$), and hence $(o,1)$ is the unique top vertex of Q as well. Note that each "new" non-base vertex of Q is the intersection of $H^{=}\big((A_{i,\star}, a_i), b_i\big)$ with the relative interior of some non-base edge of $\widetilde{Q}$ connecting a base vertex of $\widetilde{Q}$ cut off by $H^{\le}\big((A_{i,\star}, a_i), b_i\big)$ to a non-base vertex contained in $B_\mu^{d+1}\big((o,1)\big)$. We use the following statement, which will be proven separately.

Claim 2.9. *There exists a number $D \ge 7$ such that for every $0 < \mu \le \frac{1}{2}$ with $\mu < \epsilon$ the Euclidean distance from any "new" non-base vertex of Q to $(o,1)$ is less than μD.*

Hence by choosing any $0 < \mu \le \min\left\{\frac{1}{2}, \frac{\epsilon}{D}\right\}$ (in particular, $\mu < \epsilon$), we guarantee that all non-base vertices of Q (including the "new" ones) are contained in $B_\epsilon^{d+1}\big((o,1)\big)$.

As $\widetilde{Q}$ is simple, every base vertex of Q has exactly one edge not lying in the base, which will be called the *increasing* edge (since the z-coordinate of its non-base endpoint is greater than 0, the z-coordinate of its base endpoint). Note that a z-increasing path connecting a base vertex u to the top vertex necessarily contains the increasing edge incident to u.

Now suppose v is a (base or non-base) vertex of Q, that is a vertex of $\widetilde{Q}$ as well, then $v \in H^{<}\big((A_{i,\star}, a_i), b_i\big)$ holds, which for the non-base vertices follows from their membership in $B_\mu^{d+1}\big((o,1)\big)$ and the choice of a_i, and for the base vertices this is due to $Ax \le b$ being non-degenerate. In particular, v is still contained in exactly d facets of Q. Hence v has the same z-increasing path of length at most $m-d-1$ to the top vertex in Q as in $\widetilde{Q}$, since v itself and all non-base vertices of $\widetilde{Q}$ are contained in $H^{<}\big((A_{i,\star}, a_i), b_i\big)$.

Finally consider a "new" base vertex v of Q, which is the intersection of the hyperplane $H^{=}\big((A_{i,\star}, a_i), b_i\big)$ with the relative interior of some base edge e of $\widetilde{Q}$ (again due to the non-degeneracy of $Ax \le b$). Denote the endpoint of e contained in $H^{>}\big((A_{i,\star}, a_i), b_i\big)$ by u. Since u is a base vertex of $\widetilde{Q}$, it has a unique increasing edge which we denote by

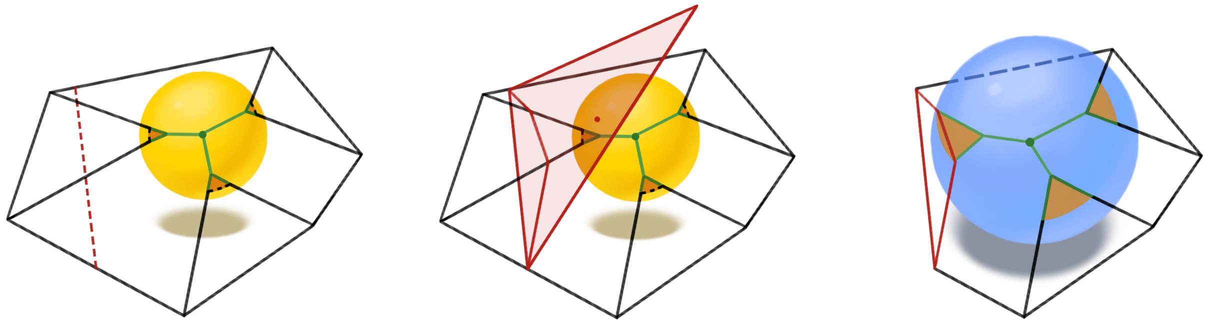

Fig. 2.2: Visualization of the proof of Theorem 2.7 for two-dimensional polytopes. Illustrated by Maryia Kukharenka.

g. Lets denote the other endpoint of g by w. Then, since $w \in B_\mu\big((o,1)\big)$, the hyperplane $H^=\big((A_{i,\star}, a_i), b_i\big)$ intersects g in a relative interior point that we denote by y. As $\widetilde{Q}$ is simple, both v and y are contained in exactly d facets of Q and there exists a 2-face F of $\widetilde{Q}$ containing both edges e and g incident to u. Since the hyperplane $H^=\big((A_{i,\star}, a_i), b_i\big)$ intersects both edges e and g in points v and y, respectively, it intersects F in the edge $\{v, y\}$ of the rock extension Q. Since there exists a z-increasing path of length at most $m-d-1$ connecting u and the top vertex $(o,1)$ in $\widetilde{Q}$, the same path with only the edge $\{w,u\}$ replaced by the two edges $\{w,y\},\{y,v\}$ (which are both z-increasing since u is a base vertex and y is contained in the relative interior of the increasing edge $\{w,u\}$) connects the base vertex v to $(o,1)$ in Q and has length at most $m-d$. Note that every "new" non-base vertex of Q arises like the vertex y described above, thus admitting a z-increasing path to the top vertex $(o,1)$ of length at most $m-d$ (in fact at most $m-d-1$). Therefore, Q is indeed a simple rock extension that is ϵ-concentrated around $(o,1)$ with each vertex of Q admitting a z-increasing path to the top vertex of length at most $m-d$. See Figure 2.2 for an illustration. $\square$

We still have to prove Claim 2.9. For that, recall the notions of a row basis and a basic solution of a system $Ax \le b$.

Definition 2.10. *Let δ_1 denote the maximum Euclidean distance from any (feasible or infeasible) basic solution of the system $Ax \le b$ to the point o. And let δ_2 be the minimum nonzero Euclidean distance from any (again feasible or infeasible) basic solution u of $Ax \le b$ to any hyperplane induced by a row of $Ax \le b$.*

Proof of Claim 2.9. Let u be a base vertex of $\widetilde{Q}$ cut off by $H^\le\big((A_{i,\star}, a_i), b_i\big)$. We denote the other vertex of the increasing edge of u by w. Note that the following argumentation only relies on $w \in B_\mu\big((o,1)\big)$ and the fact that w doesn't lie above $\{z=1\}$, which will be useful for considerations in Section 2.4. Let y be the intersection point of $H^=\big((A_{i,\star}, a_i), b_i\big)$ with the edge $\{u,v\}$. We aim to bound the distance from y to $(o,1)$.

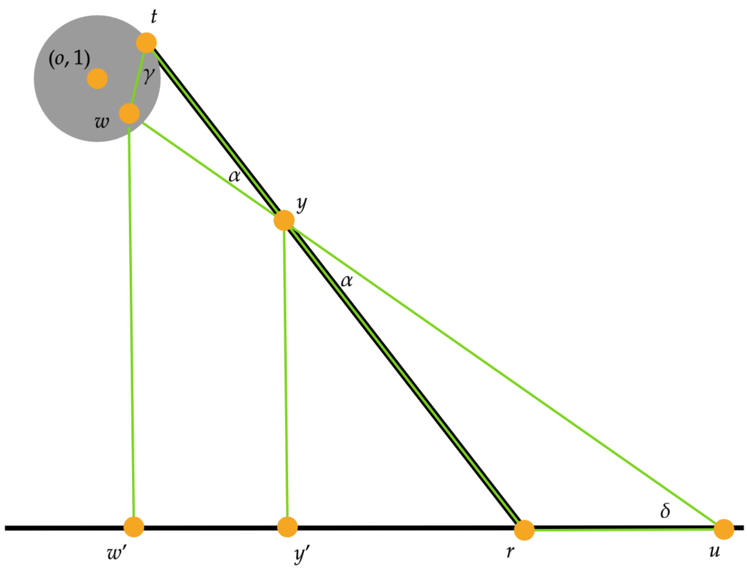

Fig. 2.3: Objects of dimensionality $d+1$, d, 1, and 0 are depicted in gray, **black**, **light green**, and **orange** respectively. The gray ball has radius μ. The points w, w', y, y', and u are contained in a two-dimensional plane, which, however, in general does not contain r and t.

Note that y lies below $\{z=1\}$ because $w \in \{z \leq 1\}$. Furthermore, due to the choice of a_i, the hyperplane $H^{=}\big((A_{i,\star}, a_i), b_i\big)$ is tangential to $B_{\mu}^{d+1}\big((o,1)\big)$ at a point we denote by t. Note that t lies above $\{z=1\}$ since we have $B_{\mu}^{d}(o) \subsetneq B_{\epsilon}^{d}(o) \subseteq P$. Thus the line through t and y intersects $\{z=0\}$ in a point r. Since both t and y are contained in $H^{=}\big((A_{i,\star}, a_i), b_i\big)$, so is that line. We denote the angles $\angle ryu = \angle tyw, \angle wty$, and $\angle yur$ by α, γ, and δ, respectively. See Figure 2.3 for an illustration.

For the sake of readability, we further use $|ab|$ for the length of the line segment $[a,b]$ (the Euclidean distance) between any two points a and b. Applying the law of sines to $\triangle ryu$ we obtain $\frac{\sin\alpha}{|ur|} = \frac{\sin\delta}{|yr|}$. On the other hand, for $\triangle tyw$ the same implies $\frac{\sin\alpha}{|tw|} = \frac{\sin\gamma}{|wy|}$. Solving both equations for $\sin\alpha$ we get $\frac{|ur|}{|yr|}\sin\delta = \frac{|tw|}{|wy|}\sin\gamma$. Then, solving the last equality for $|wy|$ we obtain

$$|wy| = \frac{|tw|\cdot|yr|}{|ur|}\frac{\sin\gamma}{\sin\delta} \leq \frac{2\mu(|yu|+|ur|)|yu|}{|ur|\cdot h_{y,[u,r]}}, \tag{2.1}$$

where the last inequality holds since $|tw| \leq \operatorname{dist}\big(t,(o,1)\big) + \operatorname{dist}\big(w,(o,1)\big) \leq 2\mu$, $\sin\gamma \leq 1$, $|yr| \leq |yu| + |ur|$, and $\sin\delta = \frac{h_{y,[u,r]}}{|yu|}$, where $h_{y,[u,r]}$ is the height of vertex y in $\triangle ryu$.

We denote the orthogonal projections of y and w to the hyperplane $\{z=0\}$ by y' and w', respectively. Since $|yy'|$ is the distance between y and the hyperplane $\{z=0\}$ containing both u and r, we conclude $h_{y,[u,r]} \geq |yy'|$. Moreover, the triangles $\triangle yuy'$ and $\triangle wuw'$ are similar and therefore $h_{y,[u,r]} \geq |yy'| = \frac{|yu|}{|yu|+|wy|}|ww'| \geq \frac{|yu|}{|yu|+|wy|}(1-\mu)$, where the last inequality follows from the fact that $W \in B_{\mu}\big((o,1)\big)$. Plugging that estimate into (2.1) gives

$$|wy| \le \frac{2\mu(|yu|+|ur|)|yu|(|yu|+|wy|)}{|ur|(1-\mu)|yu|} = \frac{2\mu(|yu|+|wy|)}{1-\mu}\Big(1+\frac{|yu|}{|ur|}\Big). \tag{2.2}$$

Finally we bound the length of all the remaining line segments appearing in the right-hand side of (2.2) to obtain an upper bound on $|wy|$. First, we observe $|yu| \le |yu|+|wy| \le \operatorname{dist}\big(u,(o,1)\big)+\mu \le \sqrt{\delta_1^2+1}+\mu$. Second, $|ur| \ge \operatorname{dist}(u, H^{=}\big(A_{i,\star}, b_i)\big) \ge \delta_2$. Plugging those inequalities into (2.2) we obtain

$$\begin{aligned} |wy| &\le \frac{2\mu(\sqrt{\delta_1^2+1}+\mu)}{1-\mu}\Big(1+\frac{\sqrt{\delta_1^2+1}+\mu}{\delta_2}\Big) \\ &\le 4\mu(\delta_1+1.5)\Big(1+\frac{\delta_1+1.5}{\delta_2}\Big), \end{aligned} \tag{2.3}$$

where for the last inequality we used $\mu \le 0.5$ and $\sqrt{\delta_1^2+1} \le \delta_1+1$. It follows that $\operatorname{dist}\big((o,1),y\big) < \mu+|wy| \le \mu D$, with $D := 4\big(\delta_1+1.5\big)\Big(1+\frac{\delta_1+1.5}{\delta_2}\Big)+1 \ge 7$. □

2.2.3 Paths in hyperplane arrangements

In this subsection, we briefly elaborate on some properties of paths on rock extensions and establish a connection to paths on hyperplane arrangements.

By the definition of a rock extension Q of $P = P^{\le}(A,b)$, the orthogonal projection of any path on Q that does not only use the base edges onto the original space runs through the interior of P for at least some part. Therefore, it is hard to interpret the projected path in terms of the original system $Ax \le b$.

However, if we restrict the scope to the simple rock extensions Q and totally non-degenerate systems $Ax \le b$, a different relation to the original space can be observed. In fact, every path on Q can be associated with a path in the hyperplane arrangement $\mathcal{H}(A,b)$ where it is allowed to take "long steps" along the lines of the arrangement. Let us elaborate on how the said correspondence works. Firstly, all the base edges of Q are naturally contained in the lines of $\mathcal{H}(A,b)$. Secondly, since each non-base edge of Q is contained in d facets of it, the intersection of the corresponding d hyperplanes in $\mathbb{R}^d$ is a basic (feasible or infeasible) solution of $Ax \le b$ due to total non-degeneracy. Moreover, for any two consecutive non-base edges e_1 and e_2 on a path T in Q let us denote the sets of d facets of Q containing e_1 and e_2 by U_1 and U_2, respectively. Let $U := U_1 \cap U_2$, then $|U| = d-1$ since Q is simple. Again, the hyperplanes of $\mathcal{H}(A,b)$, corresponding to facets of U intersect along the line l of the arrangement, that contains the two (feasible or infeasible) basic solutions u_1 and u_2 of $Ax \le b$ associated with e_1 ans e_2, respectively. Therefore, moving from e_1 to e_2 along the path T in Q corresponds

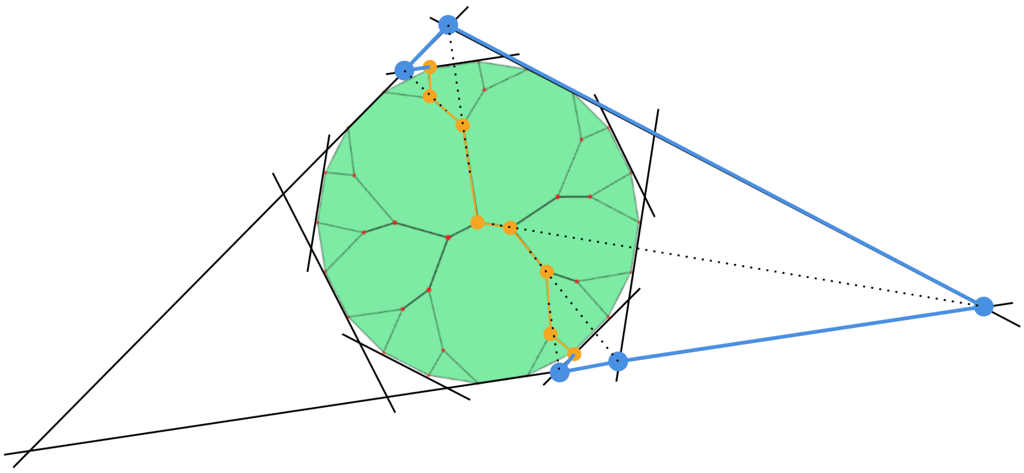

Fig. 2.4: An orange path on a rock extension (top view) and the associated blue path on the hyperplane arrangement in the original space. The correspondence of edges of the orange path to vertices of the blue path is visualized by dotted lines.

to taking a "long step" from u_1 to u_2 along the line l possibly "jumping" over vertices of $\mathcal{H}(A,b)$. See Figure 2.4 for an illustration.

Note that by allowing for such "long steps" along the lines of a hyperplane arrangement $\mathcal{H}(A,b)$ in $\mathbb{R}^d$ associated with a totally non-degenerate system of linear inequalities $Ax \leq b$, one can get between any two vertices u and v of $\mathcal{H}(A,b)$ in at most d steps in fact. This is due to the fact that given the sets U and V of hyperplanes containing u and v, respectively (recall $|U| = |V| = d$ due to total degeneracy), we can, starting at u, follow any line of the arrangement until it intersects a hyperplane from $|V \setminus U|$ (which is bound to happen due to total non-degeneracy). We denote the found vertex by w with W being the set of hyperplanes that contain it and repeat the latter for w and v. Note that $|V \setminus W| = |V \setminus U| - 1$. The rest follows by induction.

2.3 Low-dimensional polytopes

This section improves the diameter bound from the last section for rock extensions of two- and three-dimensional polytopes.

Let us consider again the setting of the proof of Theorem 2.7. The main source of improvement for $d \in \{2,3\}$ originates from applying the induction hypothesis to a polytope obtained by deleting a batch of inequalities defining *pairwise disjoint facets* of the original polytope. It turns out that subsequently constructing a rock extension by adding all of the batch inequalities back one after another (with coefficients a as in the proof of Theorem 2.7) has the effect of increasing the combinatorial distances to the top vertex by at most one overall. Next we elaborate on the latter fact.

Let $Ax \leq b$ be a simplex-containing non-degenerate system of $m \geq d+3$ inequalities defining a polytope $P = P^{\leq}(A,b)$ with an interior point o, and let ϵ be a positive number such that $B^d_\epsilon(o) \subseteq P$. Furthermore, let the inequalities $A_{i,\star}x \leq b_i$ and $A_{j,\star}x \leq b_j$ with $i,j \in [m] \setminus I$ (where, again, I is as in Definition 2.3) and $i \neq j$ define *disjoint* facets F_i and F_j of P, respectively. Note that each vertex of F_j is contained in $H^{<}(A_{i,\star}, b_i)$ and vice versa. Consider the polytopes $P_J := P^{\leq}(A_{J,\star}, b_J)$ with $J := [m] \setminus \{i\}$ and $P_K = P^{\leq}(A_K, b_K)$ with $K := [m] \setminus \{i,j\}$. Theorem 2.7 establishes that for a positive number $\nu := \min\{\frac{1}{2D}, \frac{\epsilon}{D^2}\} < \epsilon$ with D as in Claim 2.9, the polytope P_K admits a simple rock extension Q_K that is ν-concentrated around $(o,1)$ such that for every vertex of Q_K there exists a z-increasing path of length at most $m-d-2$ to the top vertex $(o,1)$. Now we argue that adding the inequality $A_{j,\star}x + a_j z \leq b_j$ to a system describing Q_K (with a_j chosen as discussed in the proof of Theorem 2.7, where we use $\mu := \min\{\frac{1}{2}, \frac{\epsilon}{D}\}$ for ϵ in said theorem), and then further adding $A_{i,\star}x + a_i z \leq b_i$ (with a_i as in the proof of Theorem 2.7 again) yields a simple rock extension Q of P that is ϵ-concentrated around $(o,1)$ and has diameter at most $2(m-d-1)$. In other words, despite subsequently adding two cutting halfspaces, the length of all paths to the top has increased by at most one.

Let v be a "new" base vertex of Q_J, which is the intersection of $H^{=}\big((A_{j,\star}, a_j), b_j\big)$ with the relative interior of some base edge e of Q_K, admitting a z-increasing path to the top vertex of Q_J of length at most $m-d-1$ as in the proof of Theorem 2.7. Since v is identified with a vertex of facet F_j of P and since F_i and F_j are disjoint, $v \in H^{<}\big((A_{i,\star}, a_i), b_i\big)$ holds and hence v is a vertex of Q as well. Moreover, recall that all non-base vertices of Q_J are vertices of Q since they are contained in $B^{d+1}_\mu\big((o,1)\big) \subseteq H^{<}\big((A_{i,\star}, a_i), b_i\big)$ and hence they admit an increasing path of length at most $m-d-2$ to the top of Q. Therefore, v admits the very same z-increasing path of length at most $m-d-1$ to the top vertex of Q as in Q_J. On the other hand any "old" base vertex u of Q_J (which is a base vertex of Q_K too) admits a path to the top vertex of Q_J of length at most $m-d-2$. Since the vertices of the latter kind are the only ones that could be cut off by $A_{i,\star}x + a_i z \leq b_i$ when constructing Q, all the "new" base and non-base vertices of Q admit an increasing path of length at most $m-d-1$, respectively, $m-d-2$ to the top vertex of Q.

Note that the above argumentation naturally extends to any number of inequalities, defining pairwise disjoint facets of P where the sequence $\mu = \min\{\frac{1}{2}, \frac{\epsilon}{D}\}$, $\nu = \frac{\mu}{D}$ is extended to $\mu, \frac{\mu}{D}, \frac{\mu}{D^2}, \frac{\mu}{D^3}, \ldots$.

We now exploit the latter consideration to improve the diameter bounds for rock extensions of two- and three-dimensional polytopes. Let us also note upfront that any non-degenerate system of m inequalities $Ax \leq b$ defining a d-polytope P can be augmented to a non-degenerate simplex-containing system describing P by adding a single redundant inequality to $Ax \leq b$ as follows.Let v be a vertex of P. Then the redundant inequality $\alpha x \leq \beta$ can be chosen in such a way that together with d inequalities defining v it defines a simplex containing P and so that the system $Ax \leq b, \alpha x \leq \beta$ is non-degenerate. We will elaborate on how to choose α and β in Section 2.5 in more detail.

The following statement holds for polygons .

Theorem 2.11. *Each n-gon admits a simple three-dimensional extension with at most $n+2$ facets and diameter at most* $2\log_2(n-2)+4$.

Proof. We start with the observation that any irredundant system of inequalities describing an n-gon P is non-degenerate, since no three distinct edge-containing lines intersect in a point. Hence, as discussed above, P can be described by a non-degenerate system $Ax \leq b$ of $m = n+1$ inequalities, consisting of n edge-defining inequalities for P and an artificially added inequality. Two inequalities defining edges incident to a vertex of P and the auxiliary inequality, such that the three of them form a simplex containing P, are indexed by $I \subseteq [m]$. As in Theorem 2.7, we prove by induction that for any interior point o of P and every $\epsilon > 0$ with $B^d_\epsilon(o) \subseteq P$ there exists a simple rock extension Q of P that is ϵ-concentrated around $(o,1)$ such that for each vertex of Q there exists a z-increasing path of length at most $\log_2(m-3)+2$ to the top vertex. Clearly Q then has diameter at most $2\log_2(n-2)+4$.

It is easy to see that the claim holds for $m = 4,5$. Note that $\lceil \frac{n-2}{2} \rceil = \lceil \frac{m-3}{2} \rceil$ of the facets defined by inequalities from $[m] \setminus I$ are pairwise disjoint. For that we just pick every second edge while traversing the graph of the (not necessarily bounded) polygon $P^{\leq}(A_{[m]\setminus I,\star}, b_{[m]\setminus I})$ since the corresponding edges are pairwise disjoint in P as well. Deleting the inequalities corresponding to all those facets at once yields a polygon $\widetilde{P}$ described by a system of $\widetilde{m} := \lfloor \frac{m+3}{2} \rfloor \leq \frac{m+3}{2}$ inequalities. By the induction hypothesis for $\mu := D^{-\lceil \frac{n-2}{2} \rceil} \min\{\frac{D}{2}, \epsilon\}$ with D as in Claim 2.9 there exists a simple rock extension $\widetilde{Q}$ of $\widetilde{P}$ that is μ-concentrated around $(o,1)$ so that for each vertex of $\widetilde{Q}$ there exists a z-increasing path of length at most $\log_2(\widetilde{m}-3)+2$ to the top vertex. Subsequently adding all $\lceil \frac{n-2}{2} \rceil$ deleted inequalities back one by one with appropriate a-coefficients yields a sequence of $\lceil \frac{n-2}{2} \rceil$ rock extensions λ_k-concentrated around $(o,1)$ with $\lambda_0 = \mu, \lambda_{k+1} = D\lambda_k, k = 1,..,\lceil \frac{n-2}{2} \rceil - 1$. Note that $\lambda_0 < \lambda_1 < \cdots < \lambda_{\lceil \frac{n-2}{2} \rceil} \leq \epsilon$. According to the arguments discussed above, the last rock extension Q of the latter sequence is a simple extension of P that is ϵ-concentrated around $(o,1)$ such that each vertex of Q admits a z-increasing path to the top vertex of length at most $\log_2(\widetilde{m}-3)+2+1 \leq \log_2(\frac{m+3}{2}-3)+2+1 = \log_2(m-3)+2 = \log_2(n-2)+2$. □

Figure 2.5 illustrates how a rock extension from the latter theorem drastically reduces the diameter of a 400-gon.

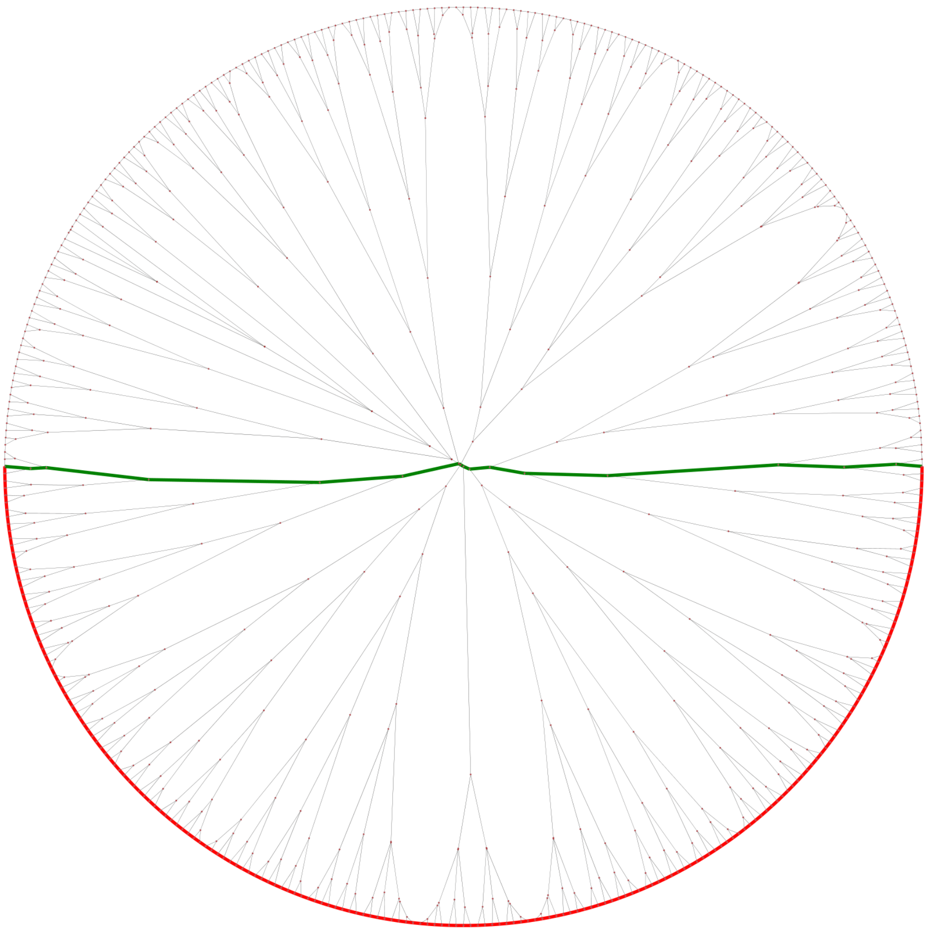

Fig. 2.5: The graph of a rock extension Q of a 400-gon P. The paths of length 200 and 24 determining the diameters of P and Q are highlighted in red and green, respectively.

Similarly, we prove the following bound for three-dimensional polytopes (recall that each strongly non-degenerate polytope with n facets can be described by a non-degenerate simplex-containing system of at most $m = n+1$ inequalities).

Theorem 2.12. *Each three-dimensional polytope P described by a non-degenerate simplex-containing system with m inequalities admits a simple four-dimensional extension with at most $m+1$ facets and diameter at most* $2\log_{\frac{4}{3}}(m-4)+4$

Proof. Once more, the set of indices of four inequalities defining the simplex containing P is referred to as I. To estimate the number of pairwise disjoint facets of P, consider the graph G_F whose vertices are the facets of P where two vertices are adjacent if and only if the corresponding facets are non-disjoint. Since P is simple, two facets are non-disjoint if and only if they share an edge. Therefore G_F is the graph of the polar polytope P°. Since P° is three-dimensional, $G(P^\circ)$ is planar, and so is the graph $G'_F := G(P^\circ) \setminus V(I)$, where $V(I)$ contains vertices of $G(P^\circ)$ corresponding to the facets of P defined by the inequalities indexed by I. It is a consequence of the four-color theorem (Appel and Haken, 1977; Appel et al., 1977; Robertson et al., 1997) that any planar graph G admits a stable set of cardinality at least $\frac{|V(G)|}{4}$. Let $S \subseteq V(G'_F)$ be a stable set in G'_F of cardinally at least $\frac{|V(G'_F)|}{4} = \frac{m-4}{4}$. By deleting the inequalities that correspond to the vertices in S from $Ax \le b$, applying the induction hypothesis as in Theorem 2.11, and subsequently adding these deleted inequalities back with appropriate a-coefficients we again obtain a simple rock extension with diameter at most $2\big(\log_{\frac{4}{3}}(\frac{3m+4}{4} - 4) + 2 + 1\big) = 2\big(\log_{\frac{4}{3}} \frac{3(m-4)}{4} + 2 + 1\big) = 2\log_{\frac{4}{3}}(m-4) + 4$. □

To conclude this section, we note that similar argumentation does not work for some (strongly non-degenerate) polytopes in dimension four and higher. An example of that is the polar dual of the cyclic polytope, since any pair of its facets intersect (as any two vertices of the cyclic polytope of dimension at least four are adjacent). Furthermore, the dual of the cyclic polytope is simple, since the cyclic polytope is simplicial. A mild perturbation would even make it strongly non-degenerate (and again by adding redundant inequalities one can achieve simplex-containment). Therefore, since the (perturbed) polar of the cyclic polytope of dimension at least four does not even have two disjoint facets, the same line of reasoning yielding logarithmic bounds in two- and three-dimensional cases cannot be applied here.

2.4 Rational polytopes and encoding sizes

In this section we consider rational polytopes. We revisit Theorem 2.7 and adjust its proof to ensure that for a rational polytope P defined by $A \in \mathbb{Q}^{m\times d}, b \in \mathbb{Q}^m$ the constructed rock extension Q is rational as well. We also show that the encoding size of Q (with respect to the inequality description) is polynomially bounded in the encoding size of P, denoted by $\langle A, b\rangle$.

We can assume that A and b are integral, since one can multiply the system $Ax \le b$ by the least common denominator of entries of A and b (which has encoding size polynomially bounded in $\langle A, b\rangle$). We now adjust the proof of Theorem 2.7 so that the extension Q being constructed meets the additional requirements.

Theorem 2.13. *For each polynomial $q_1(\cdot)$ there exists a polynomial $q_2(\cdot)$ such that for every simplex-containing non-degenerate system defining a d-polytope $P = P^{\le}(A, b)$ with $A \in \mathbb{Z}^{m\times d}, b \in \mathbb{Z}^m$, every rational $\epsilon > 0$, and every rational point o such that $B^d_\epsilon(o) \subseteq P$ with $\langle\epsilon\rangle, \langle o\rangle \le q_1(\langle A, b\rangle)$, there exists a simple rational rock extension Q that is ϵ-concentrated around $(o, 1)$ such that for each vertex of Q there exists a z-increasing path of length at most $m - d$ to the unique top vertex, and such that $\langle a\rangle \le q_2(\langle A, b\rangle)$ holds for the coefficient vector a corresponding to the additional variable in the description of Q.*

Proof. Without loss of generality we assume that $o = \mathbb{O} \in \operatorname{int}(P)$, which implies $b > 0$.

Again, for $m = d+1$ the statement trivially holds true for $Q = \operatorname{conv}\{P \cup \{(\mathbb{O},1)\}\} = \{x \in \mathbb{R}^d \mid Ax + bz \le b, z \ge 0\}$. Now we consider the induction step.

First, let us obtain explicit bounds on δ_1 and δ_2 (from Definition 2.10). Due to Cramer's rule and the integrality of A and b each coordinate of any basic solution of $Ax \le b$ is at most $\Delta_{(A,b),d}$ in absolute value. Moreover, since $\langle \det(M)\rangle \le 2\langle M\rangle$ holds for any rational square matrix M (Schrijver, 1986, Theorem 3.2), we have $\Delta_{(A,b),d} \le 2^{2\langle A,b\rangle}$, and therefore

$$\delta_1 \le \Delta_{(A,b),d}\sqrt{d} \le 2^{2\langle A,b\rangle} d\,. \tag{2.4}$$

Now assume that a basic solution u and a hyperplane $H^=(A_{i,\star}, b_i)$ corresponding to a row of $Ax \le b$ with $u \notin H^=(A_{i,\star}, b_i)$ have the Euclidean distance $\frac{|A_{i,\star}u - b_i|}{||A_{i,\star}||_2} = \delta_2$. Since the least common denominator of all entries of u is at most $\Delta_{(A,b),d}$ (due to Cramer's rule again), and since $|A_{i,\star}u - b_i| \neq 0$ and due to integrality of $A_{i,\star}, b_i$, we obtain $|A_{i,\star}u - b_i| \ge \frac{1}{\Delta_{(A,b),d}}$. Therefore,

$$\delta_2 \ge \frac{1}{\Delta_{(A,b),d}||A_{i,\star}||_2} \ge (2^{2\langle A,b\rangle} d\Delta_{(A,b),1})^{-1}\,, \tag{2.5}$$

where the last inequality follows from the aforementioned bound on $\Delta_{(A,b),d}$ and from $||A_{i,\star}||_2 \le \Delta_{(A,b),1}\sqrt{d} \le d\Delta_{(A,b),1}$.

Now we can adjust the choice of the constant D from Claim 2.9. We plug (2.4) and (2.5) into the upper bound on the Euclidean distance between the top vertex $(\mathbb{O},1)$ and a "new" non-basic vertex y of Q (see the proof of Theorem 2.7 and Claim 2.9 for more details):

$$\begin{aligned}
\operatorname{dist}((\mathbb{O},1), y) &< \mu + |wy| \\
&\overset{(2.3)}{\le} \mu + 4\mu(\delta_1 + 1.5)\left(1 + \tfrac{\delta_1 + 1.5}{\delta_2}\right) \\
&\overset{(2.4)(2.5)}{\le} \mu + 4\mu(2^{2\langle A,b\rangle}d + 1.5)\big(1 + (2^{2\langle A,b\rangle}d + 1.5)(2^{2\langle A,b\rangle}d\Delta_{(A,b),1})\big) \\
&\le \mu \cdot \underbrace{25d^3\Delta_{(A,b),1}2^{6\langle A,b\rangle}}_{=:\widehat{D}}
\end{aligned}$$

Additionally, later in the proof we will use the fact that μ does not exceed either $\frac{1}{4d}$ or $\frac{4db_i}{||A_{i,\star}||_1 + b_i}$ for any $i \in [m]$. Therefore, we choose

$$\mu := \min\left\{\left\{\frac{4db_i}{||A_{i,\star}||_1 + b_i}\right\}_{i\in[m]}, \frac{1}{4d}, \frac{\epsilon}{\widehat{D}}\right\}.$$

Note that this choice guarantees $\mu \le \frac{1}{2}$ and $\mu < \epsilon$ as well as that μ is rational with $\langle \mu \rangle = O(\langle A,b\rangle + \langle \epsilon \rangle)$.

In the proof of Theorem 2.7 we chose a_i such that $H^=\big((A_{i,\star}, a_i), b_i\big)$ is tangential to $B_\mu^{d+1}\big((\mathbb{O}, 1)\big)$. We now want to quantify this value. Once again, we denote the tangential point of the ball by $t \in \mathbb{R}^{d+1}$. We have $(A_{i,\star}, a_i)t = b_i$ since $t \in H^=\big((A_{i,\star}, a_i), b_i\big)$ and $t = (\mathbb{O}, 1) + \frac{(A_{i,\star},a_i)^T}{||(A_{i,\star},a_i)||_2}\mu$ since t lies on the boundary of $B_\mu^{d+1}\big((\mathbb{O}, 1)\big) \subseteq H^\leq\big((A_{i,\star}, a_i), b_i\big)$. Plugging the second equality into the first one, we obtain

$$(A_{i,\star}, a_i)(\mathbb{O}, 1) + \frac{||(A_{i,\star}, a_i)||_2^2}{||(A_{i,\star}, a_i)||_2}\mu = a_i + \mu\sqrt{||A_{i,\star}||_2^2 + a_i^2} = b_i\,.$$

Note that $b_i \geq b_i - a_i > 0$ holds. By taking a_i to the right in the last equation and squaring both sides we get

$$\mu^2(||A_{i,\star}||_2^2 + a_i^2) = b_i^2 + a_i^2 - 2a_i b_i\,.$$

After rearranging the terms we obtain a quadratic equation

$$a_i^2(1-\mu^2) - 2a_i b_i + b_i^2 - \mu^2||A_{i,\star}||_2^2 = 0$$

with roots

$$\begin{aligned} a_i^{+,-} &= \frac{b_i \pm \sqrt{b_i^2 - (1-\mu^2)(b_i^2 - \mu^2||A_{i,\star}||_2^2)}}{1-\mu^2} \\ &= \frac{b_i \pm \mu\sqrt{(1-\mu^2)||A_{i,\star}||_2^2 + b_i^2}}{1-\mu^2}\,. \end{aligned}$$

We deduce that $a_i = a_i(\mu) := a_i^-$, since $a_i^+ \geq \frac{b_i}{1-\mu^2} \geq b_i$. Unfortunately, $a_i(\mu)$ is not necessarily rational. However, we will show that one can use the rational number

$$\widehat{a}_i(\mu) := \frac{b_i - \frac{\mu}{2d}\big(||A_{i,\star}||_1 + b_i\big)}{1-\mu^2},$$

whose encoding size is polynomially bounded in $\langle A, b\rangle + \langle\mu\rangle$ instead of $a_i(\mu)$ when constructing the rock extension. This is due to a crucial fact that $\widehat{a}_i(\mu)$ satisfies

$$a_i(\mu') \geq \widehat{a}_i(\mu) \geq a_i(\mu)\,, \tag{2.6}$$

with $\mu' := \frac{\mu}{4d}$, which will be shown at the end of this section. Note that due to $\langle\mu'\rangle = \langle\mu\rangle + \mathrm{O}(\langle d\rangle)$ (and the above estimate $\langle\mu\rangle = \mathrm{O}(\langle A, b\rangle + \langle\epsilon\rangle)$), throughout all less than m recursive steps the encoding length of μ' will be bounded by $\mathrm{O}(m\langle A, b\rangle + \langle\epsilon\rangle) = \mathrm{O}(\langle A, b\rangle + \langle\epsilon\rangle)$ with the "original" ϵ.

Then, in order to construct a rational rock extension Q of P, we use a recursively constructed rational rock extension $\widetilde{Q}$ of $\widetilde{P}$ that is in fact μ'-concentrated around $(\mathbb{O}, 1)$

and add the inequality $A_{i,\star}x+\widehat{a}_i(\mu)\le b$. Due to $B^{d+1}_{\mu'}\big((\mathbb{O},1)\big)\subseteq H^{\le}\big((A_{i,\star},a_i(\mu'),b_i\big)$ and $a_i(\mu')\ge\widehat{a}_i(\mu)$, we have $B^{d+1}_{\mu'}\big((\mathbb{O},1)\big)\subseteq H^{\le}\big((A_{i,\star},\widehat{a}_i(\mu)),b_i\big)$. Therefore, the argument for the existence of z-increasing paths to the top vertex of length at most $m-d$ in Q is the same as in the proof of Theorem 2.7. On the other hand, since $\widehat{a}_i(\mu)\ge a_i(\mu)$, all "new" non-base vertices of Q are contained in $B^{d+1}_{\epsilon}\big((\mathbb{O},1)\big)$. Let us prove the latter. Consider Figure 2.3 once again. The point w is now contained in a smaller ball $B^{d+1}_{\mu'}\big((\mathbb{O},1)\big)\subseteq B_{\mu}\big((\mathbb{O},1)\big)$ and lies in $\{z\le 1\}$. Since $\widehat{a}_i(\mu)\ge a_i(\mu)$, the hyperplane $H^{\le}\big((A_{i,\star},\widehat{a}_i(\mu),b_i\big)$ intersects the edge $\{u,w\}$ in a point $\widehat{y}$ that lies on the line segment $[w,y]$. Therefore $|w\widehat{y}|\le|wy|$ and hence $\widehat{y}\in B^{d+1}_{\epsilon}\big((\mathbb{O},1)\big)$ as well. It remains to prove (2.6).

We start with a sequence of estimations:

$$\begin{aligned}
\tfrac{\mu}{4d}\sqrt{\big(1-(\tfrac{\mu}{4d})^2\big)||A_{i,\star}||_2^2+b_i^2} &\overset{\mu\ge 0}{\le} \tfrac{\mu}{4d}\sqrt{||A_{i,\star}||_2^2+b_i^2}\\
&\overset{2||A_{i,\star}||_2 b_i\ge 0}{\le} \tfrac{\mu}{4d}\big(||A_{i,\star}||_2+b_i\big)\\
&\overset{||\cdot||_2\le||\cdot||_1}{\le} \tfrac{\mu}{4d}\big(||A_{i,\star}||_1+b_i\big).
\end{aligned}\tag{2.7}$$

Furthermore, we have

$$\begin{aligned}
\tfrac{\mu}{2d}\big(||A_{i,\star}||_1+b_i\big) &\overset{||\cdot||_1\le d||\cdot||_2}{\le} \tfrac{\mu}{2d}\big(d||A_{i,\star}||_2+b_i\big)\\
&\overset{b_i\ge 0}{\le} \tfrac{\mu}{2}\big(||A_{i,\star}||_2+b_i\big)\\
&= \tfrac{\mu}{2}\sqrt{||A_{i,\star}||_2^2+b_i^2+2||A_{i,\star}||_2 b_i}\\
&\overset{2xy\le x^2+y^2}{\le} \tfrac{\mu}{2}\sqrt{2(||A_{i,\star}||_2^2+b_i^2)}\\
&\overset{4(1-\mu^2)\ge 4\frac{3}{4}=3\ge 2}{\le} \tfrac{\mu}{2}\sqrt{4(1-\mu^2)||A_{i,\star}||_2^2+2b_i^2}\\
&\le \mu\sqrt{(1-\mu^2)||A_{i,\star}||_2^2+b_i^2},
\end{aligned}\tag{2.8}$$

where $1-\mu^2\ge\frac{3}{4}$ since $\mu\le\frac{1}{2}$. Finally, let us prove (2.6), where we exploit the inequalities $\mu\le\frac{4db_i}{||A_{i,\star}||_1+b_i}$ for all $i\in[m]$.

$$\begin{aligned}
a_i(\tfrac{\mu}{4d}) &= \frac{b_i - \frac{\mu}{4d}\sqrt{\left(1-(\frac{\mu}{4d})^2\right)||A_{i,\star}||_2^2 + b_i^2}}{1-(\frac{\mu}{4d})^2} \\
&\overset{(2.7)}{\geq} \frac{b_i - \frac{\mu}{4d}\left(||A_{i,\star}||_1 + b_i\right)}{1-(\frac{\mu}{4d})^2} \\
&= \frac{1-\mu^2}{1-(\frac{\mu}{4d})^2}\cdot\underbrace{\frac{b_i - \frac{\mu}{4d}\left(||A_{i,\star}||_1 + b_i\right)}{1-\mu^2}}_{\geq 0} \\
&\overset{\frac{1}{4d}\geq\mu\geq 0}{\geq} \frac{1-\frac{\mu}{4d}}{1}\cdot\frac{b_i - \frac{\mu}{4d}\left(||A_{i,\star}||_1 + b_i\right)}{1-\mu^2} \\
&= \frac{b_i - \frac{\mu}{4d}b_i - (1-\frac{\mu}{4d})\frac{\mu}{4d}\left(||A_{i,\star}||_1 + b_i\right)}{1-\mu^2} \\
&\overset{\mu\geq 0}{\geq} \frac{b_i - \frac{\mu}{4d}b_i - \frac{\mu}{4d}\left(||A_{i,\star}||_1 + b_i\right)}{1-\mu^2} \\
&\overset{\mu\geq 0}{\geq} \frac{b_i - \frac{\mu}{2d}\left(||A_{i,\star}||_1 + b_i\right)}{1-\mu^2} = \widehat{a}_i(\mu) \\
&\overset{(2.8)}{\geq} \frac{b_i - \mu\sqrt{(1-\mu^2)||A_{i,\star}||_2^2 + b_i^2}}{1-\mu^2} \\
&= a_i(\mu)\,.
\end{aligned}$$

□

2.5 Algorithmic aspects of rock extensions

In this section we show how to compute rock extensions efficiently and how to utilize them in order to solve general linear programming problems. We first give an explicit algorithm for constructing a simple rock extension with linear diameter, assuming some prior information about the polytope. In the second part of this section we discuss a strongly polynomial time reduction of general (rational) linear programming to optimizing linear functions over rock extensions.

2.5.1 Computing rock extensions

The proof of Theorem 2.13 shows that for any rational simplex-containing non-degenerate system $Ax \leq b$ of m linear inequalities defining a (full-dimensional and simple) d-polytope P it is possible to construct a simple rational rock extension Q of P with diameter at most $2(m-d)$ in strongly polynomial time if the following additional information is available: an interior point o of P (with $\langle o\rangle$ bounded polynomially in $\langle A,b\rangle$) and a subsystem $A_{I,\star}x \leq b_I$ of $d+1$ inequalities defining a simplex containing P. Having that information at hand, we can shift the origin to o, scale the system to integrality, and then construct Q by choosing a-coefficients in accordance with the proof of Theorem 2.13. For that we explicitly state Algorithm 2.1. Note that it runs in strongly polynomial time.

We also need some $\epsilon > 0$ with encoding size polynomially bounded in $\langle A, b\rangle$ and such that $B^d_\epsilon(\mathbb{O}) \subseteq P$. We make the following explicit choice for ϵ. For $B^d_\epsilon(\mathbb{O}) \subseteq P$ to hold, ϵ should not exceed the minimum distance from $\mathbb{O}$ to a hyperplane corresponding to a facet of P. To achieve polynomial encoding size we bound this value from below and choose $\epsilon := \min_{i\in[m]} \frac{b_i}{d\Delta_{(A,b),1}} \leq \min_{i\in[m]} \frac{b_i}{||A_{i,\star}||_2} = \min_{i\in[m]} \operatorname{dist}\big(\mathbb{O}, H^{=}(A_{i,\star}, b_i)\big)$.

Algorithm 2.1 carries out the iterative construction of a rock extension described in the proof of Theorem 2.13. It starts with a pyramid over the given simplex $P^{\leq}(A_{I,\star}, b_I)$ and adds the inequalities indexed by $[m] \setminus I$ one by one. Note that we compute coefficients a_j in the reverse order of the iterative construction.

Algorithm 2.1 Computing a rock extension Q of P.

Input: A non-degenerate system $A \in \mathbb{Z}^{m\times d}, b \in \mathbb{Z}^m$ defining a polytope P with $\mathbb{O} \in \operatorname{int}(P)$ and a subset $I \subseteq [m], |I| = d+1$ with $P^{\leq}(A_{I,\star}, b_I)$ bounded.

Output: A vector $a \in \mathbb{Q}^m_{>0}$ with $\langle a\rangle$ polynomially bounded in $\langle A, b\rangle$ such that $Q = \{x \in \mathbb{R}^d \mid Ax + az \leq b, z \geq 0\}$ is a simple extension of P having diameter at most $2(m-d)$.

1: $a_I := b_I$
2: $D := 25d^3\Delta_{(A,b),1}2^{6\langle A,b\rangle}$
3: $\epsilon := \min_{i\in[m]} \frac{b_i}{d\Delta_{(A,b),1}}$
4: **for** $j \in [m] \setminus I$ **do**
5: $\quad \mu := \min\big\{\{\frac{4db_i}{||A_{i,\star}||_1+b_i}\}_{i\in[m]}, \frac{1}{4d}, \frac{\epsilon}{D}\big\}$
6: $\quad a_j := \frac{b_j - \frac{\mu}{2d}\left(||A_{j,\star}||_1+b_j\right)}{1-\mu^2}$
7: $\quad \epsilon := \frac{\mu}{4d}$
8: **end for**

What can we do if no interior point o of P is known (such that we could shift P to $P - o$ in order to have $\mathbb{O}$ in the interior), and neither is the set I? For now let us assume we are given a vertex x^U of a strongly non-degenerate polytope $P = P^{\leq}(A, b)$ with integral A and b, and let $U \subseteq [m]$ be the corresponding row basis of x^U. Then the point $o(\lambda) := x^U + \frac{\lambda}{||(A_{U,\star})^{-1}\mathbb{1}||_1}(A_{U,\star})^{-1}\mathbb{1}$ is an interior point of P for a small enough positive λ. This is due to the fact that P is simple and hence the extreme rays of the feasible cone of P at u are the columns of $(A_{U,\star})^{-1}$. Hence, the sum of the extreme rays emanating from x^U points into the interior of P. By choosing $\lambda := \frac{1}{2}(2^{2\langle A,b\rangle}d\Delta_{(A,b),1})^{-1} \leq \frac{1}{2}\delta_2$ (recall δ_2 from Definition 2.10 and the last inequality is due to (2.5) again), we guarantee that $o(\lambda) \in \operatorname{int}(P)$. Of course, before making this choice of λ one has to scale $Ax \leq b$ to integrality first.

The knowledge of x^U and U as above also enables us to come up with the set I required by Algorithm 2.1. Indeed, the inequalities $A_{U,\star}x \le b_U$ together with one additional redundant inequality $\mathbb{1}^T(A_{U,\star})^{-T}x \le d2^{2\langle A,b\rangle}||(A_{U,\star})^{-1}\mathbb{1}||_1 + 1$, denoted by $\alpha x \le \beta$, form a simplex containing P. In fact, $H^<(\alpha,\beta)$ contains any basic (feasible or infeasible) solution x^V of $Ax \le b$ (and hence the whole of P as well), since the sum of entries of x^V is at most $d\Delta_{(A,b),d}$. Hence $Ax \le b, \alpha x \le \beta$ is a non-degenerate system whenever $Ax \le b$ is one as well. Then I can be chosen as the union of U and the index of $\alpha x \le \beta$. Note that $o(\lambda)$ and the coefficients of the above inequality $\alpha x \le \beta$ have their encoding sizes polynomially bounded in $\langle A,b\rangle$.

Now, after shifting the origin to $o(\lambda)$ and scaling the system to integrality we can apply Algorithm 2.1 to construct a rock extension of P. Thus we have established the following.

Theorem 2.14. *Given $A \in \mathbb{Q}^{m\times d}$, $b \in \mathbb{Q}^m$, and $x^U \in \mathbb{Q}^d$ such that $Ax \le b$ is non-degenerate, $P = P^{\le}(A,b)$ is bounded, and x^U is a vertex of P, one can construct a matrix $A_Q \in \mathbb{Q}^{(m+2)\times(d+1)}$ and a vector $b_Q \in \mathbb{Q}^{m+2}$ in strongly polynomial time such that $Q = P^{\le}(A_Q, b_Q)$ is a simple rational rock extension of P with at most $m+2$ facets and diameter at most* $2(m-d+1)$.

2.5.2 Application to linear programming

Now we discuss rock extensions in the context of solving general (rational) linear programming problems. Since the previously described construction of rock extensions works only for the case of non-degenerate systems and requires knowing a vertex of the polytope, we introduce the following definition.

Definition 2.15. *We call a pair (S,u) a **strong input** if S is a rational non-degenerate system $Ax \le b$ defining a polytope P and u is a vertex of P.*

The following result provides the lever that allows us to use rock extensions in the theory of linear programming.

Theorem 2.16. *If there is a strongly polynomial time algorithm for finding optimal basic solutions for linear programs with strong inputs and rational objective functions, then all rational linear programs can be solved in strongly polynomial time.*

In order to prove the above theorem we first state and prove the following technical lemma.

Lemma 2.17. *For all $A \in \mathbb{Z}^{m\times d}$ with* $\operatorname{rank}(A) = d$, *$b \in \mathbb{Z}^m$, $c \in \mathbb{Z}^d$ such that $P := P^{\le}(A,b)$ is a pointed polyhedron and for every positive $\epsilon \le (3d||c||_1 2^{5\langle A,b\rangle})^{-1}$ the following holds for $P^\epsilon := P^{\le}(A, b+b^\epsilon)$, where $b^\epsilon_i := \epsilon^i$, $i \in [m]$.*

(1) *$P \neq \emptyset$ if and only if $P^\epsilon \neq \emptyset$. If P is non-empty, then P^ϵ is full-dimensional.*

(2) *For each row basis U feasible for $Ax \leq b + b^\epsilon$, the basic solution $A_U^{-1} b_U$ is a vertex of P.*

(3) *For each vertex v of P there is a row basis U of $Ax \leq b$ with $v = A_{U,\star}^{-1} b_U$ such that $A_{U,\star}^{-1}(b + b^\epsilon)_U$ is a vertex of P^ϵ.*

(4) *If W is an optimal feasible row basis for $\min\{c^T x \mid x \in P^\epsilon\}$, then W is an optimal feasible row basis for $\min\{c^T x \mid x \in P\}$ as well.*

(5) *The system of linear inequalities $Ax \leq b + b^\epsilon$ is non-degenerate.*

Proof. A proof for statement (1) can be found in (Schrijver, 1986, Chapter 13).

We start with the simple observation that U is a (feasible or infeasible) row basis for $Ax \leq b$ if and only if it is a row basis for $Ax \leq b + b^\epsilon$ since both systems have the same left-hand side matrix A. We will refer to any such U as a row basis of A. The following property $(\mathcal{P})$ turns out to be useful for the proof:

$(\mathcal{P})$ *If a basic (feasible or infeasible) solution $x^U := A_{U,\star}^{-1} b_U$ of $Ax \leq b$ with row basis U is contained in $H^<(A_{i,\star}, b_i)$ or $H^>(A_{i,\star}, b_i)$ for some $i \in [m]$, then the basic (feasible or infeasible) solution $x^{U,\epsilon} := A_{U,\star}^{-1}(b + b^\epsilon)_U$ of $Ax \leq b + b^\epsilon$ is contained in $H^<(A_{i,\star}, (b + b^\epsilon)_i)$ or $H^>(A_{i,\star}, (b + b^\epsilon)_i)$, respectively.*

We later show that $(\mathcal{P})$ holds for all small enough positive ϵ, but first let us observe how (2) and (3) follow from $(\mathcal{P})$.

For (2), assume $x^{U,\epsilon}$ is a feasible basic solution of $Ax \leq b + b^\epsilon$ with row basis U such that $x^U := A_{U,\star}^{-1} b_U$ is infeasible for $Ax \leq b$, i.e., there exists some $i \in [m]$ with $x^U \in H^>(A_{i,\star}, b_i)$. If $(\mathcal{P})$ holds, then the latter, however, contradicts the feasibility of $x^{U,\epsilon}$ for $Ax \leq b + b^\epsilon$. Thus $(\mathcal{P})$ implies (2).

In order to see that $(\mathcal{P})$ also implies (3), let $A_{E(v)} x \leq b_{E(v)}$ consist of all inequalities from $Ax \leq b$ that are satisfied with equality at a vertex v of P. Note that the set of feasible bases of $A_{E(v)} x \leq (b + b^\epsilon)_{E(v)}$ is non-empty, since $P^\leq(A_{E(v)}, (b + b^\epsilon)_{E(v)})$ is pointed. The latter holds because of $\operatorname{rank}(A_{E(v)}) = d$ (as v is a vertex of P) and since $P^\leq(A_{E(v)}, (b + b^\epsilon)_{E(v)})$ itself is non-empty with $v \in P^\leq(A_{E(v)}, (b + b^\epsilon)_{E(v)})$ (due to $b^\epsilon \geq \mathbb{O}$). We now can choose U as any feasible row basis of $A_{E(v)} x \leq (b + b^\epsilon)_{E(v)}$. We clearly have $v = A_{U,\star}^{-1} b_U$ and $A_{[m]\setminus E(v)} v < b_{[m]\setminus E(v)}$ by the definition of $E(v)$. Hence the basic solution $A_U^{-1}(b + b^\epsilon)_U$ is feasible for $Ax \leq b + b^\epsilon$ due to $(\mathcal{P})$.

Claim 2.18. *The property $(\mathcal{P})$ holds for $0 < \epsilon \leq (3d||c||_1 2^{5\langle A,b\rangle})^{-1}$ (we clearly can assume $c \neq \mathbb{O}$).*

Proof. Let $x^U = A^{-1}_{U,\star} b_U$ be a basic (feasible or infeasible) solution of $Ax \le b$ with a row basis $U \subseteq [m]$, and let $H^=(A_{i,\star}, b_i)$ with $i \in [m] \setminus U$ be a hyperplane with $x^U \notin H^=(A_{i,\star}, b_i)$. Furthermore, let $x^{U,\epsilon} := A^{-1}_{U,\star}(b + b^\epsilon)_U$ be the corresponding basic (feasible or infeasible) solution of the perturbed system. Consider the following expression:

$$\begin{aligned} A_{i,\star} x^{U,\epsilon} - (b+b^\epsilon)_i &= \sum_{j=1}^{d} A_{ij} x_j^{U,\epsilon} - (b+b^\epsilon)_i \\ &= \frac{\sum_{j=1}^{d} A_{ij} \det A_U^{j=b+b^\epsilon} - (b+b^\epsilon)_i \det A_{U,\star}}{\det A_{U,\star}} =: h_{U,i}(\epsilon)\,, \end{aligned} \tag{2.9}$$

where $A^{j=q}_{U,\star}$ denotes the square $d \times d$ matrix arising from $A_{U,\star}$ by replacing the j^{th} column by the vector q. Note that $h_{U,i}(\epsilon)$ is a univariate polynomial in ϵ with its free coefficient $\alpha_0 := h_{U,i}(0) = A_{i,\star}x^U - b_i \neq 0$ due to $x^U \notin H^=(A_i, b_i)$. Therefore the property $(\mathcal{P})$ holds if $\epsilon > 0$ is small enough, such that $h_{U,i}(\epsilon)$ has the same sign as α_0. We will need the following result on roots of univariate polynomials. See (Bienstock et al., 2023, Lemma 4.2); a proof can be found in (Basu et al., 2006, Theorem 10.2).

Lemma 2.19 (Cauchy)**.** *Let $f(x) = \alpha_n x^n + \cdots + \alpha_1 x + \alpha_0$ be a polynomial with real coefficients and $\alpha_0 \neq 0$. Let $\bar{x} \neq 0$ be a root of $f(x)$. Then $\frac{1}{\delta} \le |\bar{x}|$ holds with $\delta = 1 + \max\{|\frac{\alpha_1}{\alpha_0}|, \ldots, |\frac{\alpha_n}{\alpha_0}|\}$.*

Hence $(\mathcal{P})$ holds for all $0 < \epsilon < \frac{1}{\delta}$ (with δ chosen as in the lemma with respect to $h_{U,i}$) since there are no roots of $h_{U,i}(\epsilon)$ in the interval $(-\frac{1}{\delta}, +\frac{1}{\delta})$. To obtain an estimate on ϵ we have to bound δ from above. Therefore, let us take a closer look at the coefficients of $h_{U,i}(\epsilon)$. Due to Cramer's rule, the integrality of A and b, and since $|\det A_{U,\star}| \le \Delta_{(A,b),d}$, the absolute value of each non-vanishing coefficient of $h_{U,i}(\epsilon)$ is at least $\frac{1}{\Delta_{(A,b),d}}$. On the other hand, we claim that the magnitude of any coefficient $h^k_{U,i}(\epsilon)$ of $h_{U,i}(\epsilon)$ that is not the free coefficient is bounded from above by $\prod_{(i,j)\in[m]\times[d]}(1+|a_{ij}|)\prod_{i\in[m]}(1+|b_i|) \le 2^{\langle A,b\rangle}$. Observe that the latter product is the sum Σ over all subsets of indices of entries of (A,b) with each summand being the product of (absolute values) of the entries in question. Using column expansion for determinants in (2.9), one can come up with a formula for coefficients $h^k_{U,i}$, $k \in [m]$:

$$h^k_{U,i} = \begin{cases} \frac{1}{\det A_{U,\star}} \sum_{j=1}^{d} A_{ij} \det A^{j=\emptyset}_{U\setminus k}, & \text{for } k \in U \\ -1, & \text{for } k = i \\ 0, & \text{for } k \in [m] \setminus (U \cup \{i\}) \end{cases}$$

where $A^{j=\emptyset}_{U\setminus k}$ denotes the square matrix obtained by deleting the jth column and the kth row of $A_{U,\star}$. To upper bound $|h^k_{U,i}|$, $k \in [m]$ we use the Leibniz formula for determinants in the above equation, the triangle inequality for absolute value, and the fact $\frac{1}{|\det A_{U,\star}|} \leq 1$. One can convince themselves that the upper bounding expression obtained therein is, in fact, a subsum of Σ, and hence $|h^k_{U,i}| \leq 2^{\langle A,b\rangle}$ for each $k \in [m]$. Therefore $\delta \leq 1 + \Delta_{(A,b),d} 2^{\langle A,b\rangle} \leq 2 \cdot 2^{3\langle A,b\rangle}$ holds, where the last inequality follows from $\Delta_{(A,b),d} \leq 2^{2\langle A,b\rangle}$. For $0 < \epsilon \leq (3d||c||_1 2^{5\langle A,b\rangle})^{-1}$ we thus indeed have $\epsilon < \frac{1}{2}2^{-3\langle A,b\rangle} \leq \frac{1}{\delta}$ (as $c \neq \mathbb{O}$ is integral). Note that such a choice of ϵ is in fact determined by a later observation. □

Next, to show (5) $\big($before we establish (4)$\big)$ let us assume that $Ax \leq b + b^{\epsilon'}$ is not non-degenerate for some $\epsilon' \leq (3d||c||_1 2^{5\langle A,b\rangle})^{-1}$. Hence, there is a row basis $U \subseteq [m]$ of A with corresponding (feasible or infeasible) basic solutions $x^{U,\epsilon'}$ and x^U of the perturbed and the unperturbed system, respectively, such that there exists $i \in [m] \setminus U$ with $x^{U,\epsilon'} \in H^=(A_{i,\star}, (b + b^{\epsilon'})_i)$, thus $h_{U,i}(\epsilon') = 0$. Due to Lemma 2.19 (and the upper bound on ϵ') this implies $h_{U,i}(0) = 0$, thus $x^U \in H^=(A_{i,\star}, b_i)$. Since U is a row basis of A, there exists some $\lambda \in \mathbb{R}^d$ with $\lambda^T A_{U,\star} = A_{i,\star}$. We have $b_i = A_{i,\star}x^U = \lambda^T A_{U,\star}x^U = \lambda^T b_U$. Then $h_{U,i}(\epsilon) = A_{i,\star}x^{U,\epsilon} - (b+b^\epsilon)_i = \lambda^T A_{U,\star}(A_{U,\star})^{-1}(b+b^\epsilon)_U - (b+b^\epsilon)_i = \lambda^T b^\epsilon_U - \epsilon^i$ where we used $A_{i,\star} = \lambda^T A_{U,\star}$ and $x^{U,\epsilon'} = (A_{U,\star})^{-1}(b+b^\epsilon)_U$ for the first equality and $\lambda^T b_U = b_i$ and $b^\epsilon_i = \epsilon^i$ for the second one. Hence $h_{U,i}(\epsilon)$ is not the zero polynomial because of $i \notin U$. Consequently, there exists a polynomial $g_{U,i}(\epsilon)$ such that $h_{U,i}(\epsilon) = \epsilon^r g_{U,i}(\epsilon)$ with $r \geq 1$ and $g_{U,i}(0) \neq 0$. Applying Lemma 2.19 to $g_{U,i}(\epsilon)$ and bounding its coefficients in exactly the same way as for $h_{U,i}(\epsilon)$ yields that there are no roots of $g_{U,i}(\epsilon)$, and therefore no roots of $h_{U,i}(\epsilon)$, in the interval $(0, \frac{1}{2}2^{-3\langle A,b\rangle})$, thus contradicting $x^{U,\epsilon'} \in H^=(A_{i,\star}, (b + b^{\epsilon'})_i)$.

Finally, in order to show (4), we first prove the following claim.

Claim 2.20. *Let $U \subseteq [m]$ be a row basis of A with $x^{U,\epsilon}$ and x^U being the corresponding (feasible or infeasible) basic solutions of $Ax \leq b + b^\epsilon$ and $Ax \leq b$, respectively. Then $0 < \epsilon \leq (3d||c||_1 2^{5\langle A,b\rangle})^{-1}$ implies $|c^T x^U - c^T x^{U,\epsilon}| < \frac{1}{2\Delta^2_{(A,b),d}}$.*

Proof. By Cramer's rule and due to the triangle inequality, we have

$$|c^T(x^U - x^{U,\epsilon})| \leq \sum_{j=1}^d |c_j||x^U_j - x^{U,\epsilon}_j| = \sum_{j=1}^d |c_j| \Bigg| \underbrace{\frac{\det A^{j=b}_{U,\star} - \det A^{j=b+b^\epsilon}_{U,\star}}{\det A_{U,\star}}}_{=:f^j_U(\epsilon)} \Bigg|. \tag{2.10}$$

To prove the claim it suffices to show that for all $0 < \epsilon \leq (3d||c||_1 2^{5\langle A,b\rangle})^{-1}$ we have $|f^j_U(\epsilon)| < \frac{1}{2|c_j|d|\Delta^2_{(A,b),d}} =: \beta^j_0$ for each $j \in [d]$ with $c_j \neq 0$. In order to establish this, let $j \in [d]$ be an index with $c_j \neq 0$. Due to $f^j_U(0) = 0$ we have $f^j_U(\epsilon) = \alpha_l \epsilon^l + \cdots + \alpha_1 \epsilon$ with some $\alpha_1, \ldots, \alpha_l \in \mathbb{Q}$. For $\beta_0 := \frac{1}{2|c_j|d\Delta^2_{(A,b),d}}$ and $f^{j\pm}_U(\epsilon) := f^j_U(\epsilon) \pm \beta^j_0$ we have $f^{j-}_U(0) < 0 < f^{j+}_U(0)$. Due to Lemma 2.19, the polynomials $f^{j\pm}_U(\epsilon)$ thus have no roots in the interval $\big(-\frac{1}{\delta}, +\frac{1}{\delta}\big)$, where $\delta = 1 + \max\{|\frac{\alpha_1}{\beta_0}|, \dots, |\frac{\alpha_l}{\beta_0}|\}$. Hence in order to establish $|f_U^j(\epsilon)| < \beta_0^j$ it suffices to show $\epsilon < \frac{1}{\delta}$. In order to prove this we bound δ from above (thus $\frac{1}{\delta}$ from below) by upper bounding the coefficients α_k for all $k \in [l]$. Using the same line of arguments as above we once again conclude that $\alpha_k \le \prod_{(i,j)\in[m]\times[d]}(1+|a_{ij}|)\prod_{i\in[m]}(1+|b_i|) \le 2^{\langle A,b\rangle}$, $k \in [l]$. Hence $\frac{1}{\delta} \ge (1+2d|c_j|\Delta^2_{(A,b),d}|2^{\langle A,b\rangle})^{-1} > (3d||c||_1 2^{5\langle A,b\rangle})^{-1} \ge \epsilon$ as required. □

To complete the proof of claim (4) of Lemma 2.17, let $x^{W,\epsilon} := A^{-1}_{W,\star}(b+b^\epsilon)_W$ be an optimal feasible basic solution for $\min\{c^T x \mid x \in P^\epsilon\}$ with an optimal row basis W. Thus, due to (2), $x^W := A^{-1}_{W,\star} b_W$ is a feasible basic solution of $Ax \le b$. Furthermore, let v be an optimal vertex of P with respect to minimizing c and let U be a row basis of A with $v = x^U = A^{-1}_{U,\star} b_U$ such that $x^{U,\epsilon} := A^{-1}_{U,\star}(b+b^\epsilon)_U$ is a vertex of P^ϵ (such a row basis U exists by statement (3) of Lemma 2.17). Assume x^W is not optimal for $\min\{c^T x \mid x \in P\}$. Then we have $c^T(x^W - x^U) \ge \frac{1}{\Delta^2_{(A,b),d}}$, since c is integral and the least common denominator of the union of the coordinates of x^W and x^U is at most $\Delta^2_{(A,b),d}$ (as the least common denominator of entries of x^W is at most $\Delta_{(A,b),d}$ by Cramer's rule and so is the least common denominator of entries of x^U). But this contradicts

$$c^T(x^W - x^U) = \underbrace{c^T(x^W - x^{W,\epsilon})}_{<\frac{1}{2\Delta^2_{(A,b),d}}} + \underbrace{c^T(x^{W,\epsilon} - x^{U,\epsilon})}_{\le 0} + \underbrace{c^T(x^{U,\epsilon} - x^U)}_{<\frac{1}{2\Delta^2_{(A,b),d}}} < \frac{1}{\Delta^2_{(A,b),d}}, \tag{2.11}$$

where we used Claim 2.20 for bounding the first and the third term and the optimality of $x^{W,\epsilon}$ for bounding the second one. □

Now we can finally return to the proof of Theorem 2.16.

Proof of Theorem 2.16. We can assume without loss of generality that the polyhedron P of feasible solutions to a general linear program is pointed. Otherwise either the objective function is contained in the orthogonal complement of the lineality space of P (in which case one obtains a pointed feasible polyhedron by intersecting with the latter), or the problem is unbounded. Let $\mathcal{A}$ be a strongly polynomial time algorithm for finding optimal basic solutions for linear programs with strong inputs and rational objective functions. We first use $\mathcal{A}$ to devise a strongly polynomial time algorithm $\mathcal{A}^\star$ for finding optimal basic solutions for arbitrary rational linear programs $\min\{c^T x \mid Ax \le b\}$ if a *non-degenerate vertex* v of $P := P^{\le}(A,b)$ is specified within the input, i.e., a vertex for which there is a unique row basis $U \subseteq [m]$ with $x^U = v$.

In order to describe how $\mathcal{A}^\star$ works, we may assume that (after appropriate scaling) its input data A, b, c are integral. With $\epsilon := (3d||c||_1 2^{5\langle A,b\rangle})^{-1}$ let $P^\epsilon := \{x \in \mathbb{R}^d \mid Ax \le b + b^\epsilon\}$. Due to the uniqueness property of U and part (3) of Lemma 2.17, U is also a feasible row basis of the perturbed system. We scale that perturbed system to integrality, obtaining a non-degenerate (part (5) of Lemma 2.17) system $A'x \le b'$ with $P^\epsilon := \{x \in \mathbb{R}^d \mid A'x \le b'\}$

and a vertex $v' = x^{U,\epsilon}$. Then, as discussed in the context of Theorem 2.14, we add the inequality $\mathbb{1}^T(A'_{U,\star})^{-T}x \le d2^{2\langle A',b'\rangle}||(A'_{U,\star})^{-1}\mathbb{1}||_1 + 1$, denoted by $\alpha x \le \beta$, to $A'x \le b'$ and thus obtain a non-degenerate bounded system $\widetilde{A}x \le \widetilde{b}$ with a simplex-defining subsystem of $d+1$ inequalities. Let $\widetilde{P} := P^{\le}(\widetilde{A}, \widetilde{b})$. Note that the problem $\min\{c^T x \mid x \in P\}$ is unbounded if and only if $\min\{c^T x \mid x \in P^\epsilon\}$ is unbounded since the polyhedra P and P^ϵ have the same characteristic cone. Moreover, $\min\{c^T x \mid x \in P^\epsilon\}$ is unbounded if and only if an optimal row basis W (corresponding to any optimal vertex x^W) of $\min\{c^T x \mid x \in \widetilde{P}\}$ contains the added inequality $\alpha x \le \beta$ and the unique extreme ray of the feasible cone of $\widetilde{P}$ at x^W that is not contained in $H^=(\alpha, 0)$ has positive scalar product with c (recall that the polytope $\widetilde{P}$ is simple). Thus, in order to solve $\min\{c^T x \mid x \in P\}$ in strongly polynomial time, we can apply algorithm $\mathcal{A}$ to $\min\{c^T x \mid x \in \widetilde{P}\}$ (providing the algorithm with the vertex v' of $\widetilde{P}$). Any optimal row basis of the latter problem either proves that the former problem is unbounded or is an optimal row basis of the former problem due to part (4) of Lemma 2.17.

Finally, let us assume that we are faced with an arbitrary linear program in the form $\min\{c^T x \mid Ax \le b, x \ge 0\}$ with $A \in \mathbb{Z}^{m\times d}, b \in \mathbb{Z}^m$, and $c \in \mathbb{Z}^d$ (clearly, each rational linear program can be reduced to this form, for instance, by splitting the variables into x^+ and x^- and scaling the coefficients to integrality) and let $P := P^{\le}(A, b) \cap \mathbb{R}^d_{\ge 0}$. Due to parts (1) and (5) of Lemma 2.17 the perturbed system $Ax \le b + b^\epsilon, -x \le o^\epsilon$ with $b^\epsilon_i := \epsilon^i$ for all $i \in [m]$ and $o^\epsilon_j := \epsilon^{m+j}$ for all $j \in [d]$ is non-degenerate for $\epsilon := (3d||c||_1 2^{5(\langle A,b\rangle + \langle -\mathbb{I}_d, \mathbb{O}_d\rangle)})^{-1}$ with the polyhedron $P^\epsilon := \{x \in \mathbb{R}^d \mid Ax \le b + b^\epsilon, -x \le o^\epsilon\}$ being non-empty (in fact: full-dimensional) if $P \ne \emptyset$ and empty otherwise.

We follow a classical *Phase I* approach by first solving the auxiliary problem $\min\{\mathbb{1}_m^T s \mid (x,s) \in G\}$ with

$$G := \{(x,s) \in \mathbb{R}^{d+m} \mid Ax - s \le b + b^\epsilon, -x \le o^\epsilon, s \ge \mathbb{O}\}.$$

Note that $(\tilde{x}, \tilde{s})$ with $\tilde{x}_j = -\epsilon^{m+j}$, $j \in [d]$, and $\tilde{s}_i = \max\{\sum_{j\in[d]} A_{ij}\epsilon^{m+j} - b_i - \epsilon^i, 0\}$, $i \in [m]$, is a vertex of G, which is defined by a unique row basis $\widetilde{U}$, $|\widetilde{U}| = [m+d]$ (containing inequalities $-x \le o^\epsilon$ and for each $i \in [m]$ either $A_{i,\star}x - s_i \le b_i + \epsilon^i$ or $s_i \ge 0$) since $\sum_{j\in[d]} A_{ij}\epsilon^{m+j} - b_i - \epsilon^i \ne 0$ for any $i \in [m]$. The latter is due to the fact that P^ϵ is non-degenerate by Lemma 2.19, part (5). Hence we can apply algorithm $\mathcal{A}^\star$ in order to compute an optimal vertex $(x^\star, s^\star)$ of the auxiliary problem $\min\{\mathbb{1}_m^T s \mid (x,s) \in G\}$. If $\mathbb{1}^T s^\star \ne 0$ holds, then we can conclude $P^\epsilon = \emptyset$, thus $P = \emptyset$. Otherwise, $x^\star$ is a vertex of P^ϵ that clearly is non-degenerate (in fact, P^ϵ is simple). Thus we can solve $\min\{c^T x \mid x \in P^\epsilon\}$ by using algorithm $\mathcal{A}^\star$ once more. If the latter problem turns out to be unbounded, then so is $\min\{c^T x \mid x \in P\}$ (as P and P^ϵ have the same characteristic cone). Otherwise, the optimal row basis of $\min\{c^T x \mid x \in P^\epsilon\}$ found by $\mathcal{A}^\star$ is an optimal row basis for $\min\{c^T x \mid x \in P\}$ as well (due to part (4) of Lemma 2.17). □

It is well known that any strongly polynomial time algorithm for linear programming can be used to compute optimal basic solutions (if they exist) in strongly polynomial time (see, e.g., (Schrijver, 1986, Chapter 10) for more details). Hence Theorems 2.16 and 2.14 allow us to conclude the following.

Theorem 2.21. *If there exists a strongly polynomial time algorithm for linear programming with rational data over all simple polytopes whose diameters are bounded linearly in the numbers of inequalities in their descriptions, then all linear programs (with rational data) can be solved in strongly polynomial time.*

In fact, in order to come up with a strongly polynomial time algorithm for general linear programming problems (under the rationality assumption again) it is enough to devise a strongly polynomial time algorithm that optimizes linear functions over rock extensions with linear diameters.

2.6 Extensions with short monotone paths

The results of the previous sections showed for every d-polytope P described by a non-degenerate system of m linear inequalities the existence of a simple $(d+1)$-dimensional rock extension Q with at most $m+2$ facets, where each vertex admits a (z-increasing) "canonical" path of length at most $m-d+1$ to a distinguished vertex (the unique top vertex) of Q. For the rest of this chapter we will refer to such a Q as a *good* rock extension of P. Yet, no statement has been made so far regarding the potential monotonicity of the "canonical" paths in Q with respect to linear objective functions. In this section we build upon rock extensions in order allow for short monotone paths.

Without loss of generality we assume for the rest of the section that we are working with minimization problems. Note that our results similarly hold for maximization problems as well. For an objective function c we will call an optimal vertex of a polytope P *c-optimal.* A path in the graph of P is said to be *c-monotone* if the sequence of c-values of vertices along the path is strictly decreasing. To simplify our notation we further identify a vertex u of P with the corresponding base vertex $(u,0)$ of Q.

2.6.1 Changing the objective vector

It clearly does not hold that for every linear objective function $c \in \mathbb{Q}^d$ with w being a c-optimal vertex of a strongly non-degenerate d-polytope P and for any other vertex v of P, both the "canonical" path from v to the top vertex t of a good rock extension Q of P and the "canonical" path w-t traversed backward from t to w are c-monotone. Even the path from t to w itself is not always c-monotone, see Figure 2.6 for an example.

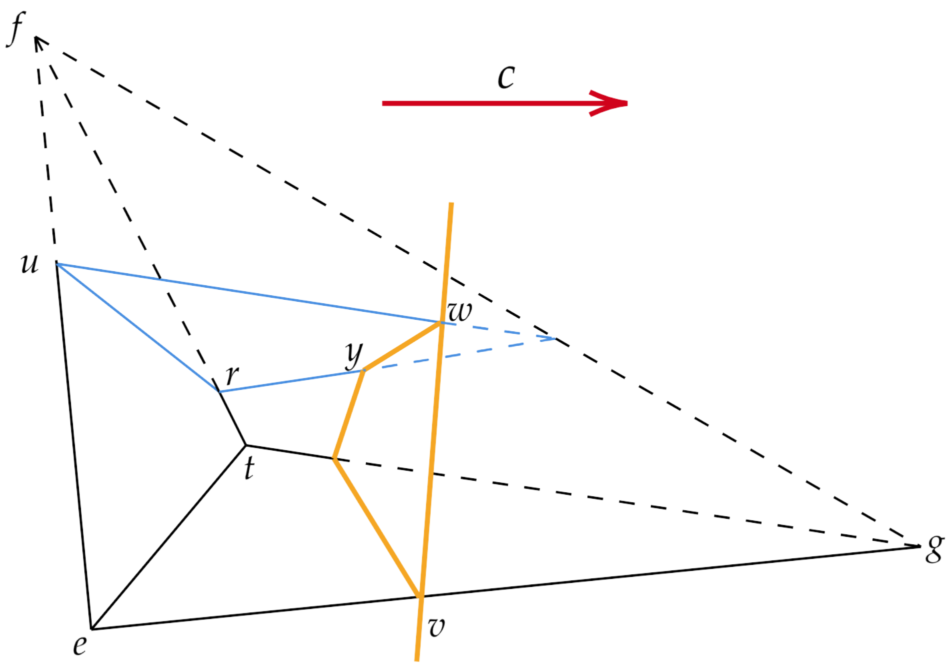

Fig. 2.6: A rock extension (top view) of the quadrangle $euwv$ constructed in the manner of Algorithm 2.1 from the rock extension $efgt$ of the two-dimensional simplex efg containing $euwv$ by consecutively adding the inequalities defining the blue the orange facets. Now w is the c-optimal vertex of the rock extension for the objective function c whose optimization direction points to the right in the picture. However, the (reversed) "canonical" path t-r-y-w is not c-monotone.

However, the latter issue can be handled by defining a new objective vector $\widetilde{c} := (c, c_z) \in \mathbb{R}^{d+1}$ with c_z being a large enough positive number, such that all "canonical" paths in Q, including the one for the $\widetilde{c}$-optimal vertex w, are $\widetilde{c}$-monotone when traversing them from the top. If follows from the result that we prove next and the fact, that z-coordinates of vertices of a top-down "canonical" path build a strictly decreasing sequence.

Lemma 2.22. *Let $Q \in \mathbb{R}^{d+1}$ be a rock extension defined by an integral system $A(x,z) \le b$ with $A \in \mathbb{Z}^{m\times(d+1)}, b \in \mathbb{Z}^m, x \in \mathbb{R}^d, z \in \mathbb{R}$. Then for every vector $c \in \mathbb{R}^d$ and for any number $c_z > 2||c||_1 2^{8\langle A,b\rangle}$ every path on Q with the next vertex having a strictly lower z-coordinate than its predecessor is (c, c_z)-monotone.*

Proof. Let u with row basis U and v with row basis V be two vertices of Q with $v_z < u_z$. The statement follows from a more general fact, that $(c, c_z)^T(v-u) < 0$, which we show next. On the one hand $c^T(v_x - u_x) \le |c^T(v_x - v_x)| \le \sum_{j=1}^{d} |c_j||v_j - u_j|$ by the triangle inequality. Using Cramer's rule as in the previous chapter we conclude that $|v_j - u_j| = |\frac{\det A^{j=b}_{V,\star}}{\det A_{V,\star}} - \frac{\det A^{j=b}_{U,\star}}{\det A_{U,\star}}| \le 2\Delta^2_{(A,b),d+1} \le 2 \times 2^{4\langle A,b\rangle}$ for any $j \in [d]$. The latter yields $c^T(v_x - u_x) \le 2||c||_1 2^{4\langle A,b\rangle}$. On the other hand, due to the choice of u and v, $v_z - u_z \le$

$-\frac{1}{\Delta^2_{A,d+1}} \leq -2^{-4\langle A,b\rangle}$ holds by the Cramer's rule again. Using the above bound for c_z we obtain $c_z(u_z - v_z) < -2||c||_1 2^{4\langle A,b\rangle}$ which concludes the proof. $\square$

Note, that by choosing c_z from the latter lemma as $2||c||_1 2^{8\left(\langle (A,a),b\rangle + \ell\right)} + 1$, where $\ell = \langle (\mathbb{O}_d, -1), 0\rangle$ is the encoding size of the system $(\mathbb{O}_d, -1)^T(x,z) \leq 0$, we ensure that the encoding size of the new objective vector $\widetilde{c}$ is polynomial in the encoding size of the input $\langle A, b, c\rangle$ with $Ax \leq b$ defining the polytope P, whose rock extension $Q = \{(x,z) \in \mathbb{R}^d \mid Ax + az \leq b, z \geq 0\}$ was constructed by Algorithm 2.1.

Obtaining the monotonicity property for all top-down "canonical" path in this way is justified by the fact that the top vertex of a rock extension constructed by Algorithm 2.1 is known (its row basis is defined by the $d+1$ inequalities indexed by I). Therefore one could start the simplex algorithm at the top vertex of the rock extension and it could in theory even follow the top-down "canonical" path of the c- and $\tilde{c}$-optimal vertex w.

2.6.2 Ridge extension

The sole change to the objective function, described in the previous subsection, however, does not offer a short monotone path from any vertex v to the optimal vertex w, since the "canonical" path from v to t is not $\widetilde{c}$-monotone (the sequence of $\widetilde{c}$-values along the path is, in fact, strictly increasing).

In order to incorporate it in a certain way we are going to spend one more dimension by building a crooked prism over the rock extension. Let P be a d-polytope defined by a system $Ax \leq b$ of m inequalities, and let $Q := \{(x,z) \in \mathbb{R}^{d+1} \mid Ax + az \leq b, z \geq 0\}$ be a rock extension of P ϵ-concentrated around $(o,1) \in \mathbb{R}^{d+1}$ for some $\epsilon > 0$ and $o \in P$. Consider the prism $Q \times [0,1]$. We now tilt the facets $Q \times \{0\}$ and $Q \times \{1\}$ toward each other such that the (Euclidean) distance between two copies of a point in Q is reduced by some factor that is proportional to its z-coordinate. We rigorously define the described polytope as follows

$$\widehat{Q} := \{(x,z,y) \in \mathbb{R}^{d+2} \mid Ax + az \leq b, z \geq 0, y - \frac{1}{3}z \geq 0, y + \frac{1}{3}z \leq 1\}.$$

We call $\widehat{Q}$ a *ridge* extension of P. See Figure 2.7 for an illustration.

We would like to note that the above construction is a special case of the deformed product ($\bowtie$) introduced by Amenta and Ziegler (1999). More precisely $\widehat{Q} = \left(Q, \phi\big((x,z)\big)\right) \bowtie \left([0,1], [\frac{1}{3}, \frac{2}{3}]\right)$ with $\phi\big((x,z)\big) = z$. Observe that if Q is simple, so is $\widehat{Q}$, since it is combinatorially equivalent to the prism over Q and prisms preserve simplicity. We will denote the two facets of $\widehat{Q}$ defined by inequalities $y - \frac{1}{3}z \geq 0$ and $y + \frac{1}{3}z \leq 1$ by Q^0 and Q^1, respectively. Note that both Q^0 and Q^1 are isomorphic to Q.

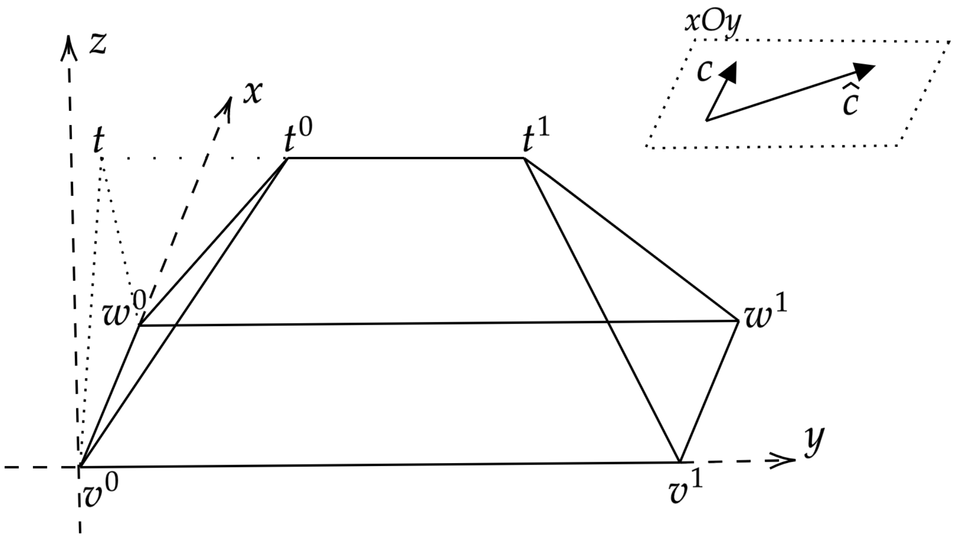

Fig. 2.7: The $d+2$-dimensional simple ridge extension $\widehat{Q}$ of a d-polytope P. P is represented here by the line segment $[v^0, w^0]$. Triangles tv^0w^0, $t^0v^0w^0$, and $t^1v^1w^1$ represent a rock extension Q of P with the top vertex t, and facets Q^0 and Q^1 of $\widehat{Q}$, both isomorphic to Q, respectively. Path v^0-t^0-t^1-w^1 is $\widehat{c}$-monotone for an auxiliary objective $\widehat{c}$ such that a $\widehat{c}$-optimal vertex w^1 is a preimage of a c-optimal vertex w^0 of P under the orthogonal projection onto x-axis.

Thus each vertex u of Q corresponds to two vertices of Q^0 and Q^1 denoted by u^0 and u^1, respectively.

Now, let Q be a good rock extension of P. Let $c \in \mathbb{R}^d$ be a linear objective function and w be a c-optimal vertex of P, and let v be some other vertex of P. Then for the "canonical" path from v to the top vertex t of Q there exists an isomorphic path from v^0 to t^0 of Q^0. Since the "canonical" v-t-path in Q is z-increasing and due to $Q^0 = f^0(Q)$ with $f^0 : (x,z) \mapsto (x, z, \frac{1}{3}z)$, the corresponding v^0-t^0-path in Q^0 is y-increasing. Similarly, there exists a y-increasing t^1-w^1-path in Q^1 isomorphic to the backward traversal of the z-increasing "canonical" w-t-path in Q, since $Q^1 = f^1(Q)$ with $f^1 : (x,z) \mapsto (x, z, 1 - \frac{1}{3}z)$. Together with the edge t^0t^1, the two aforementioned paths comprise a v^0-w^1-path of length at most $2(m-d+1)+1$ in $\widehat{Q}$ that is $\widehat{c}$-monotone for the objective function $\widehat{c} := (c, 0, c_y) \in \mathbb{R}^{d+2}$ with large enough positive c_y. Note that w^1 is a $\widehat{c}$-optimal vertex of $\widehat{Q}$ and a preimage of a c-optimal vertex w of P under the affine map $\pi_d : (x, z, y) \mapsto x$ projecting $\widehat{Q}$ down to P. In fact, similar to the proof of Lemma 2.22, it can be shown that choosing c_y as $6||c||_1 2^{8\langle\widehat{Q}\rangle} + 1$ (after scaling the system, defining $\widehat{Q}$ to integrality) is enough to guarantee $\widehat{c}$-monotonicity of any v^0-t^0-t^1-u^1-path of the above mentioned type for any two vertices u, v of Q. Thus, with the help of Theorem 2.14 we derive the following statement, where π_k denotes the orthogonal projection on the first k coordinates.

Theorem 2.23. *Let $A \in \mathbb{Q}^{m\times d}$ and $b \in \mathbb{Q}^m$ define a non-degenerate system of linear inequalities such that $P = P^{\leq}(A,b)$ is bounded. Then there exists a $(d+2)$-dimensional simple extension $\widehat{Q}$ with $\pi_d(\widehat{Q}) = P$ having at most $m+4$ facets such that for any linear objective function $c \in \mathbb{Q}^d$ there is a positive number c_y such that for any vertex v of P there exists a $(c,0,c_y)$-monotone path from the vertex $(v,0,0)$ to a $(c,0,c_y)$-optimal vertex w of $\widehat{Q}$ of length at most $2(m-d+1)+1$ with $\pi_d(w)$ being a c-optimal vertex of P. A system of linear inequalities defining $\widehat{Q}$ and the number c_y are computable in strongly polynomial time if a vertex of P is specified within the input.*

To state the concluding result of this chapter we introduce the following notion. The minimum length of a c-monotone path in the graph of a polytope P between a given vertex v and a c-optimal vertex of P is called the *monotone c-distance of v in the graph of P*. Then, combining Theorems 2.23 and 2.16 we conclude the following.

Theorem 2.24. *If there is a pivot rule for the simplex algorithm that requires only a number of steps (executable in strongly polynomial time) that is bounded polynomially in the monotone c-distance of v in the graph of P for every simple polytope P, objective function vector c, and starting vertex v, then the general (rational) linear programming problem can be solved in strongly polynomial time.*

2.7 Discussion

In this section we mention some open questions that appear to be interesting for future research.

The dimension of the rock extension exceeds that of the original polytope just by one. This leaves open the question whether it is possible to obtain extensions with better diameter bounds by using additional dimensions.

In Section 2.3 we showed that two- and three-dimensional (strongly non-degenerate) polytopes allow for rock extensions with logarithmic diameter and noted that the same argumentation does not work for some (strongly non-degenerate) four-dimensional polytope. The crucial property ($\mathcal{L}$) that we used in the two- and three-dimensional cases was the existence of a constant ratio of the polytope facets that are pairwise disjoint. Moreover, the new polytope obtained by removing the inequalities corresponding to these facets from the polytope description had to satisfy ($\mathcal{L}$) again. In fact, it was necessary to be able to iteratively repeat this procedure (of removing at least one l^{th} of the inequalities corresponding to pairwise disjoint facets of the polytope at once) until the remaining inequality system consist of $d+C$ inequalities where d is the dimension of the (original) polytope and C is a constant. An interesting question is whether there are known classes of polytopes in dimension four and higher (maybe originating from combinatorial optimization problems) that allow for such iterative constructions.

The most intriguing question, of course, is whether *for simple d-polytopes with n facets and diameter at most* $2(n-d)$ there exists a pivot rule for the simplex algorithm that is guaranteed to produce a (c-monotone) path to an optimal vertex of length bounded polynomially in d and n, as this would imply that the general linear programming problems can be solved in strongly polynomial time according to the results presented above. We would like to point out that Cardinal and Steiner (2023) showed that if $P \neq NP$ holds, then for every simplex pivot rule executable in polynomial time and for every constant $k \in \mathbb{N}$ there exists a linear program on a perfect matching polytope and a starting vertex of the said polytope such that the optimal solution can be reached in two c-monotone *non-degenerate* steps from the starting vertex, yet the pivot rule will require at least k *non-degenerate* steps to reach the optimal solution. This result, however, even under the assumption $P \neq NP$ does not rule out the existence of a pivot rule for which one can bound the number of steps by a polynomial in the diameter *plus* the number of facets, not even for general (rather than just simple) polytopes. Again, by the results presented above, such a pivot rule would imply a strongly polynomial time algorithm for general linear programming problems.

In an attempt to investigate the latter, we tried applying the celebrated shadow simplex algorithm (see, e.g., Bach and Huiberts (2025); Black et al. (2024); Borgwardt (1999); Dadush and Huiberts (2020); Megiddo (1986a); Spielman and Teng (2004); Vershynin (2009)) to rock extensions. Observe that starting the algorithm at the top vertex of a rock extension naturally calls for a shadow formed by the vectors $(\mathbb{O}_d, 1)$, $(c, 0) \in \mathbb{R}^{d+1}$ (optimizing the staring vertex and the optimal vertex, respectively), where $c \in \mathbb{R}^d$ is the original objective vector. However, we did not find a way to sensibly bound the number of vertices (of the c-monotone path of) the shadow due to lack of control over coordinates of the non-base vertices of a rock extension.

Alternatively, instead of trying to make the simplex algorithm follow a short path on a rock extension, one could investigate a corresponding path in the hyperplane arrangement associated with the LP in question (recall Subsection 2.2.3). Similarly to the previous case, given a vertex of the arrangement and an objective vector, it is highly unclear whether it is possible to create a simplex-type algorithm based on the hyperplane arrangement that would somewhat follow a short path on the arrangement to an optimal solution of the LP. The rest of this section aims to present some of the authors thoughts on the topic.

One line of reasoning is to allow such an algorithm to switch between the neighboring chambers of the arrangement under a certain condition (e.g., the algorithm always tries to perform the longest possible step along the c-monotone lines of the arrangement).

However, in this way, the algorithm might significantly drift away from the original chamber represented in the LP, especially if the reduction of the objective function value is disallowed. To enable "going back" one could alternate between the primal problem (1.4) and the dual (1.5). It might work as follows; once the primal simplex optimality criterion is satisfied at some vertex of the currently considered chamber of the arrangement (which might be different from the initial chamber), the framework would switch to the dual problem. One would repeat the latter, with the roles of the primal and the dual being switched. Perhaps some sort of dichotomy on the objective function value could provide control over the number of switches from the primal to the dual and back.

Alternating between the primal and the dual is also how the solvers for linear programming seem to work (Sophie Huiberts, personal communication). Moreover, a couple of algorithms for LP that make use of primal and dual information at the same time are already known (see e.g. Belahcene et al. (2018); Dantzig et al. (1956); Gabasov et al. (1979, 1981)). Therefore, this direction seems to be worth investigating.

3

Escaping degeneracy in linear time

3.1 Introduction

In the previous chapter we have shown how linear programming problems profit from perturbations and extended formulations, which help to guarantee the existence of short (monotone) paths for polyhedra of feasible solutions, while guarding the property of being simple.

However, even in the case where a short path between the starting vertex and an optimal one exists, it is not clear how the simplex algorithm would be bound to follow it (e.g., any two vertices of the famous cube of Klee and Minty (1972), which provides an exponential lower bound on the performance of the Dantzig pivot rule, can be connected by a path of linear length). Moreover, especially when dealing with combinatorial or 0/1 linear programming problems, a perturbation might break the underlying structure of the problem, which could otherwise assist in solving it. Therefore, efficient handling of degeneracy is of utmost importance when dealing with these types of problems.

In presence of degeneracy, an extreme point solution might have exponentially many distinct bases defining it, and so the simplex algorithm might perform an exponential number of consecutive degenerate pivots, staying in the same vertex, (see, e.g., Cunningham, 1979). Such behavior is referred to as *stalling*. In some pathological cases, certain pivot rules might even provide an infinite sequence of consecutive degenerate pivots at a degenerate vertex, a phenomenon called *cycling*. Although cycling can be easily avoided by employing, for instance, Bland's rule (due to Bland, 1977) or a lexicographic rule Terlaky, 2009, there is no known pivot rule that prevents stalling. As stated in several papers (e.g., Megiddo, 1986b; Murty, 2009), solving a general LP in strongly polynomial time can be reduced to finding a pivot rule that prevents stalling for general polyhedra. However, there are a few cases for which it is known that the issue of stalling can be handled. The most famous example is the class of transportation polytopes, for which pivot rules with polynomial bounds on the number of consecutive degenerate pivots were introduced by Cunningham (1979) and further developed by Ahuja et al.

(2002); Goldfarb et al. (1990); Rooley-Laleh (1981). See also the work of Orlin (1997) for a strongly polynomial version of the primal network simplex.

The results of this chapter are inspired to a great extent by the work of Kabadi and Punnen (2008). They show that, given an LP in standard equality form with a totally unimodular coefficient matrix $A \in \{-1, 0, 1\}^{m \times n}$, a vertex solution x and an improving feasible *circuit direction* for x, one can construct a pivot rule which performs at most m consecutive degenerate pivots. We generalize their proof, and show that one can employ *any* improving direction at a vertex x of a *general* polyhedron, to avoid stalling. Our result, proved in Section 3.2, is summarized in the following theorem.

Theorem 3.1. *Given any LP of the form* $\min\{c^T x : Ax = b, x \geq 0\}$ *with* $A \in \mathbb{R}^{m \times n}$*, there exists a pivot rule that limits the number of consecutive degenerate simplex pivots at any non-optimal extreme solution to at most* $n - m - 1$.

We stress here that the pivots considered in Theorem 3.1 are degenerate *simplex* pivots, meaning that each of them yields an *improving* direction at the given extreme point, though this direction is not feasible. It is important to point out that in general, given two adjacent extreme points x, x' of a linear program, one can easily perform a sequence of basis exchanges that yield x' from x by identifying the common tight linearly independent constraints, and introducing each of them to the current row basis in any order, until the direction $x' - x$ is seen. However, we here want a strategy that guarantees that *each* basis exchange is realizable by the simplex algorithm and hence defines an improving direction.

We then discuss some byproducts of our result in Section 3.3. In particular, we revise the analysis of the simplex algorithm by Kitahara and Mizuno (2013, 2014) who bound the number of non-degenerate pivots in terms of n, m and the maximum and the minimum non-zero coordinate of a basic feasible solution. Their analysis combined with our degeneracy-escaping technique show that the *total* number of simplex pivots (both degenerate and non-degenerate) can be bounded in a similar way. As a consequence, one can have a strongly-polynomial number of simplex pivots for several combinatorial LPs. In addition, we perform some computational experiments to evaluate the performance of the antistalling pivot rule in practice, reported in Section 3.4.

Of course the drawback of the whole machinery is that it requires an improving feasible direction at a given vertex. Though it is efficiently computable, this is in general as hard as solving a general linear programming problem, see Theorem 3.3 for a proof of the latter. For some classes of polytopes though (e.g., matching or flow polytopes) finding such a direction can be easier, thus making it worthwhile to apply our pivot rule. Most importantly, we think that the main importance of our result is from a theoretical perspective: it shows that, for several polytopes, not only a short path on the 1-skeleton exists, but a short sequence of *simplex pivots* always exists (and can also be efficiently computed). That is, a short sequence of *basis exchanges* that follow *improving* directions, when performing *both* degenerate and non-degenerate pivots.

3.2 Antistalling pivot rule

The goal of this section is to prove Theorem 3.1. Before doing that, we state some preliminaries, and also give a geometric intuition behind the result.

3.2.1 Foundations

In this chapter we use the algebraic view on the simplex algorithm. We will give a brief review of it here. The algebraic simplex algorithm works with LP in standard equality form (1.2) that we repeat here:

$$\begin{aligned} \min \quad & c^T x \\ s.t. \quad & Ax = b \\ & x \geq 0 \end{aligned} \tag{3.1}$$

Note that any LP can be put into this form (e.g., by splitting variables $x = x^+ - x^-$ and by introducing slack variables).

For the sake of completeness let us reiterate the following definition. A *basis* of (3.1) with an $m \times n$ real matrix A, $\operatorname{rank}(A) = m$ is a subset $B \subseteq [n]$ with $|B| = m$ and $A_{\star,B}$ being non-singular. The point x with $x_B = A_{\star,B}^{-1} b$, $x_N = \mathbb{O}$ where $N := [n] \setminus B$ is a *basic solution* of (3.1) with basis B. If additionally $x_B \geq 0$, both the basic solution x and the basis B are *feasible*. If $x_i > 0$ for all $i \in B$, then B and x are called *non-degenerate* and *degenerate* otherwise, i.e., if $x_i = 0$ for some $i \in B$. We let $\overline{A} := A_{\star,B}^{-1} A_{\star,N}$ and $\overline{c}_N^T := c_N^T - c_B^T \overline{A}$. In particular, $\overline{c}_N \in \mathbb{R}^N$ is the vector of so-called *reduced costs* for the basis B. The coordinates of the reduced cost vector will be addressed as $\overline{c}_{N,i}$, where the first subscript will be dropped if the basis $B = [n] \setminus N$ is clear from the context.

The simplex algorithm considers in each iteration a feasible basis B. If all elements of $\overline{c}_N$ are non-negative, then the basis B and the corresponding basic feasible solution x are known to be optimal. Otherwise, the algorithm *pivots* by choosing a non-basic coordinate with negative reduced cost to enter the basis, say f. It then performs a minimum ratio test to compute an index i that minimizes $x_i / \overline{A}_{if}$ among all indices $i \in B$ for which $\overline{A}_{if} > 0$. Such index i will be the one leaving the basis, and it corresponds to a basic coordinate which hits its bound first when moving along the direction given by the tight constraints indexed by $B \setminus \{f\}$. At each iteration there could be multiple candidate indices for entering the basis (all the ones with negative reduced cost), as well as multiple candidate indices to leave the basis (all the ones for which the minimum ratio test value is attained). The choice of the entering and the leaving coordinates is the essence of a pivot rule (see Terlaky and Zhang (1993) for a survey). A pivot is *degenerate* if the attained minimum ratio test value is 0 (hence, the extreme point solution x does not change), and *non-degenerate* otherwise. Note that only coordinates with negative reduced cost are considered for entering the basis, since only such a choice guarantees the simplex algorithm to make progress when a non-degenerate pivot occurs. To later emphasize that we only refer to that kind of (potentially) *improving* pivots, we will call them *simplex pivots.*

Finally, we let $\text{kern}(A) := \{y \in \mathbb{R}^n : Ay = 0\}$. Given (3.1), we call a vector $y \in \text{kern}(A)$ with $c^T y < 0$ an *improving direction.* Such a y is said to be *feasible* at a basic feasible solution x if $x + \epsilon y \geq 0$ for sufficiently small positive ϵ. Note that for any $y \in \text{kern}(A)$ and any basis B of A, the following holds:

$$c^T y = c_B^T y_B + c_N^T y_N = c_B^T(-\overline{A} y_N) + c_N^T y_N = \overline{c}_N^T y_N \tag{3.2}$$

3.2.2 A geometric intuition

Before providing a formal proof, we give a geometric intuition on how our pivot rule works. For this, it will be easier to abandon the standard equality form and go back to the inequality form (1.1). In particular, consider a degenerate vertex $x \in \mathbb{R}^d$ of a d-polytope P defined by a system $Ax \leq b$ of linear inequalities with $A \in \mathbb{R}^{n\times d}, b \in \mathbb{R}^n$. In this setting, degeneracy means that more than d inequalities are tight at x. Consider a subset N of the set of indices of all inequalities that are tight at x with $|N| = d$ and let $B := [n] \setminus N$. Note that here we intentionally redefine a notation used in the standard equality form, where **n**on basic coordinates always have their corresponding constraints tight. Since x is degenerate, there is at least one inequality with its index in B that is tight at x . We denote the set of such inequalities by $W \subseteq B$. Finally, assume x is not an optimal vertex of P when minimizing an objective function $c \in \mathbb{R}^d$ over P, and let $y^0 \in \mathbb{R}^d$ be an improving feasible direction at x, i.e., such that $x + \epsilon y^0 \in P$ for a sufficiently small positive ϵ and $c^T y^0 < 0$. In order to find an improving edge of P at x, we look at the directions given by the extremal rays of the basic cone $C(N) := \{x \in \mathbb{R}^n \mid A_{i,\star}x \leq 0, i \in N\}$. If there is an improving feasible direction among them, we are done. Otherwise, we reduce the dimension of the polytope in the following way. We pick an extremal ray $z \in C(N)$ such that $c^T z < 0$. Note that z is formed by $d-1$ inequalities from N. Let $f \in N$ be the only inequality index not used to define z. Consider a vector combination $y^0 + \alpha z$ where $\alpha \geq 0$ and note that it provides an improving direction for any non-negative α. Since y^0 is contained in the feasible cone $C(N \cup W) = \{x \in \mathbb{R}^n \mid A_{i,\star}x \leq 0, i \in N \cup W\}$ but z is not, the vector $y^0 + \alpha z$ leaves $C(N \cup W)$ for sufficiently large α. Hence there has to be a number $\alpha^1 \geq 0$ such that $y^0 + \alpha^1 z$ is contained in a facet of $C(N \cup W)$. Without loss of generality assume that the latter facet is defined by the g^{th} inequality. Note that $g \in W$ since y^0 and z are both contained in the basic cone $C(N)$ and so is $y^0 + \alpha z$ for any non-negative α. Then, we define a new improving feasible direction $y^1 := y^0 + \alpha^1 z$, $N = \big(N \setminus \{f\}\big) \cup \{g\}$, $B = \big(B \setminus \{g\}\big) \cup \{f\}$ and continue searching for an improving edge inside the facet defined by the g^{th} inequality. Since $x + \epsilon y^1$ belongs to this facet for an $\epsilon > 0$ small enough, the letter yields dimension reduction. After at most $d-1$ such steps we are bound to encounter a one-dimensional face of P containing $x + \epsilon y'$, where y' is an improving feasible direction at x. For an illustration see Figure 3.1.

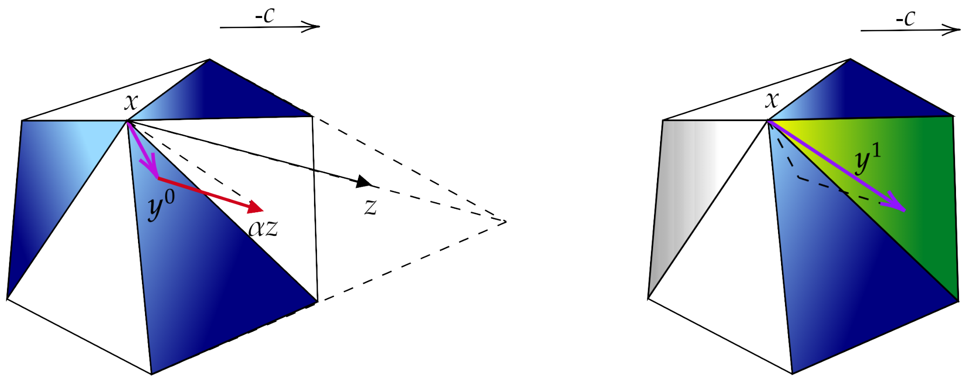

Fig. 3.1: To the left, facets corresponding to inequalities in N and W are colored in **n**avy and **w**hite, respectively. To the right, the gray **f**ading facet is defined by the f^{th} inequality and the **g**reen one corresponds to the g^{th}.

3.2.3 Proof of Theorem 3.1

Proof. Let x be a degenerate basic feasible solution of (3.1) with basis B. Without loss of generality, assume $B = [m]$. Then, $x_i = 0$ for $i \in N = [n] \setminus [m]$. Assume also that $x_i = 0$ for $i \in [k]$ with $1 \leq k \leq m$ and $x_i > 0$ for $i = k+1, \ldots, m$. Let $S(B) := \{i \in N \mid \overline{c}_i < 0\}$. Since x is not optimal, $S(B) \neq \emptyset$.

It is well known, that a basic feasible solution x of (3.1) with basis B is not optimal if and only if there exists an improving feasible direction, i.e., there exists $y^0 \in \mathbb{R}^n$ such that $Ay^0 = 0$, $c^T y^0 < 0$ and $y^0_i \geq 0$ for all $i \in [k] \cup \big([n] \setminus [m]\big)$. Consider any such y^0. Let $Q_1(y^0, B) := \{i \in [k] \mid y^0_i > 0\}$ and $Q_2(y^0, B) := \{i \in [n] \setminus [m] \mid y^0_i > 0\}$. Without loss of generality, let $Q_1(y^0, B) = [r]$ with $0 \leq r \leq k$ and $Q_2(y^0, B) = \{m+1, \ldots, m+t\}$ with $1 \leq t \leq n-m$. See Table 3.1 for an illustration.

Since $c^T y^0 = \overline{c}_N^T y^0_N < 0$ due to (3.2), it follows that $S(B) \cap Q_2(y^0, B) \neq \emptyset$. We choose the entering index to be a non-basic one with the most negative reduced cost[1] in the support of y^0, that is $f = \operatorname{argmin}_{i \in S(B) \cap Q_2(y^0, B)} \overline{c}_i$. To detect the leaving index, we

[1] We would like to point out that to prove the theorem, it suffices to choose *any* non-basic index with negative reduced cost in the support of y^0 as the entering index. The fact that f has the most negative reduced cost among such indices will only be used for the results in the next section.

	1		$\dots$	r		$\dots$	k		$\dots$	m	f		$\dots$	$m+t$		$\dots$	n
x	0	0	$\dots$	0	0	$\dots$	0	$+$	$\dots$	$+$	0	0	$\dots$	0	0	$\dots$	0
y^0	$+$	$+$	$\dots$	$+$	0	$\dots$	0	$\star$	$\dots$	$\star$	$+$	$+$	$\dots$	$+$	0	$\dots$	0
z	$-$	$\star$	$\dots$	$\star$	0_+	$\dots$	0_+	$\star$	$\dots$	$\star$	1	0	$\dots$	0	0	$\dots$	0
$y^0+\alpha z$	0	0_+	$\dots$	0_+	0_+	$\dots$	0_+	$\star$	$\dots$	$\star$	$+$	$+$	$\dots$	$+$	0	$\dots$	0

Table 3.1: An illustration of entries of x, y^0, z and $y^1 := y^0 + \alpha z$. A positive, a non-negative, a negative, and a sign-arbitrary entry is denoted by $+$, 0_+, $-$ and $\star$, respectively. Without loss of generality, we here assumed that the entering index f is $m+1$, and that an index g for which the minimum in **Case III** is attained is 1.

consider the following case distinction. Note that the case distinction depends on the basis B and the improving feasible direction y^0.

Case I: *There is no $i \in [k]$ with $\overline{A}_{if} > 0$.* In this case, the minimum ratio test for f is strictly positive, that is:

$$(*) \qquad \min_{i \in Eq^>(f)} \frac{x_i}{\overline{A}_{if}} > 0 \text{ where } Eq^>(f) := \{i \in [m] \mid \overline{A}_{if} > 0\}.$$

In particular, this means $Eq^>(f) \subseteq [m] \setminus [k]$. We perform a non-degenerate pivot, by selecting an index that minimizes $(*)$ as the leaving index.

Case II: $\overline{A}_{if} > 0$ *for some $i \in [k] \setminus [r]$.* In this case, we perform a degenerate pivot by selecting i as the leaving index. Let $B' := B \cup \{f\} \setminus \{i\}$. Because of degeneracy, the basic solution associated with B' is still x, and hence y^0 is still an improving feasible direction for x. Note that $|Q_2(y^0, B')| = |Q_2(y^0, B)| - 1$ since $y^0_f > 0$ but $y^0_i = 0$ by definition. Repeat the same for B' and y^0.

Case III: $\overline{A}_{if} \leq 0$ *for all $i \in [k] \setminus [r]$ and $\overline{A}_{if} > 0$ for some $i \in [r]$.* In this case, we are going to change our improving feasible direction. Consider the following vector $z \in \mathbb{R}^n$ which is, in fact, an improving (though not feasible) direction for the entering variable x_f:

$$z_i = \begin{cases} -\overline{A}_{if}, & \text{for } i \in [m] \\ 1, & \text{for } i = f \\ 0, & \text{otherwise} \end{cases} \tag{3.3}$$

Note that $Az = 0$ and $c^T z = \overline{c}_f < 0$. Moreover $z_i \geq 0$ for each $i \in \big([r] \setminus Eq^>(f)\big) \cup \big([k] \setminus [r]\big) \cup \big([n] \setminus [m]\big)$ and $z_i < 0$ for $i \in Eq^>(f)$ by definition. Set

$$y^1 := y^0 + \alpha z \text{ with } \alpha := \min_{i \in Eq^>(f) \cap [r]} \frac{y^0_i}{|z_i|} > 0\,.$$

Observe that y^1 is a feasible direction for x, since $A(y+\alpha z)=0$, and $y^1_i \geq 0$ for $i \in [k] \cup \big([n] \setminus [m]\big)$, because of the choice of z and α. Furthermore, y^1 is an improving direction, because $c^T z < 0$ and hence

$$c^T y^1 = c^T y^0 + \alpha c^T z < c^T y^0 < 0. \tag{3.4}$$

Note that $Q_2(y^1, B) = Q_2(y^0, B)$. Furthermore, let $g \in Eq^>(f) \cap [r]$ be the index for which the value of α is attained. We have $y^1_g = 0$ and $\overline{A}_{gf} > 0$. We now repeat the case distinction above for B and y^1.

The key observation is that when the third case of the above case distinction has occurred, repeating the same for B and y^1 falls into the second case (because the basis B has not changed, meanwhile $y^1_g = x_g = 0$ with $g \in B$ and $\overline{A}_{gf} > 0$). Therefore, after each degenerate pivot, the cardinality of the support of the improving feasible direction in the non-basic indices decreases. Hence, a sequence of $|Q_2(y^0, B)|$ degenerate pivots would yield an improving feasible direction y' and a basis B' of x with $Q_2(y', B') = 0$, which in turn implies $0 = \overline{c}^T_{N'} y'_{N'} = c^T y'$ due to (3.2), yielding a contradiction. Hence the number of consecutive degenerate pivots with the suggested pivot rule cannot exceed $n-m-1$. □

The next remark will be useful in the next section.

Remark 3.2. *Let x be the currently considered basic feasible solution with basis B. If y^0 is chosen to be equal to $\tilde{x} - x$ for some basic feasible solution $\tilde{x}$ of our LP with $c^T \tilde{x} > c^T x$, then one can observe that $|Q_2(y^0, B)| \leq m$ because $\tilde{x}$ has at most m non-zero coordinates. By the proof of Theorem 3.1 above, the latter value strictly decreases with each degenerate step produced by the antistalling pivot rule. Hence, the total number of consecutive degenerate pivots at the vertex x can be strengthened to be at most* $\min\{n-m-1, m-1\}$.

We finish this section with the previously announced result on the hardness of finding improving feasible directions.

Theorem 3.3. *Finding an improving feasible direction at a given vertex of a feasible polyhedron is as hard as solving a general linear programming problem.*

Proof. Let us consider a pair of feasible primal and dual problems (1.4, 1.5) and build the following LP by combining both of them:

$$\max \lambda \tag{3.5}$$
$$s.t. \quad c^T x = y^T b \tag{3.6}$$
$$Ax \leq \lambda b \qquad A^T y = \lambda c \tag{3.7}$$
$$y \geq 0 \tag{3.8}$$

$$0 \leq \quad \lambda \quad \leq 1 \tag{3.9}$$

with $A \in \mathbb{R}^{m\times n}$, $\text{rank}(A) = n, b \in \mathbb{R}^m, c \in \mathbb{R}^n$ and variables $x \in \mathbb{R}^n, y \in \mathbb{R}^m$, and $\lambda \in \mathbb{R}$. The point $z^o := (x^o, y^o, \lambda^o) = \mathbb{O}_{n+m+1}$ is a vertex of the feasible polyhedron of (3.5-3.9), since all inequities but $\lambda \leq 1$ are tight at z^o and it is not hard to construct a corresponding (row) basis. Note that every improving feasible direction $\Delta z = (\Delta x, \Delta y, \Delta\lambda)$ at z^o satisfies $\Delta\lambda > 0$ and (3.6-3.8). Therefore, the point $\tilde{z} := (\tilde{x}, \tilde{y}, \tilde{\lambda}) = z^o + \frac{1}{\Delta\lambda}\Delta z$ is feasible for (3.5-3.9). Due to feasibility of $\tilde{z}$ and since $\tilde{\lambda} = 1$, $(\tilde{x}, \tilde{y}) = \frac{1}{\Delta\lambda}(\Delta x, \Delta y)$ comprise a pair of optimal solutions for the primal and the dual problems (1.4, 1.5). Therefore, finding such Δz is at least as hard as solving both for the primal and the dual problems. □

3.3 Exploiting the antistalling rule for general bounds

Here we combine the result of the previous section with the analysis of Kitahara and Mizuno (2013, 2014). The authors of the latter works give a bound on the number of non-degenerate simplex pivots that depends on n, m and the maximum and the minimum non-zero coordinate of basic solutions. The combined analysis yields a similar bound on the total number of (both degenerate and non-degenerate) simplex pivots.

We need a few additional notations. For any vector x, we let $\text{supp}(x)$ denote its support. We denote the smallest and the largest non-zero coordinate of any basic feasible solution of (3.1) by δ and Δ, respectively. We let $x^\star$ be an optimal basic feasible solution of (3.1). Finally, for a generic iteration q of the simplex algorithm with basis B^q and an improving feasible direction y^q, we let $\Delta_{\bar{c}}^q := \max_{\{i\in N^q | y_i^q > 0\}} -\bar{c}_{N^q,i}$. Note that $\Delta_{\bar{c}}^q$ is the absolute value of the reduced cost of the entering variable according to the antistalling pivot rule defined in the previous section when using y^q.

We make use of the following result from Kitahara and Mizuno (2014).

Lemma 3.4 (Lemma 4 of Kitahara and Mizuno, 2014)**.** *If there exists a constant $\lambda > 0$ such that*

$$c^T(x^{q+1} - x^\star) \leq \big(1 - \frac{1}{\lambda}\big)c^T(x^q - x^\star) \tag{3.10}$$

holds for any consecutive distinct basic feasible solutions $x^q \neq x^{q+1}$ generated by the simplex algorithm (with any pivot rule), the total number of distinct basic feasible solutions encountered is at most

$$(n - m)\left\lceil \lambda \log_e \frac{m\Delta}{\delta} \right\rceil.$$

Given (3.1), apply the simplex algorithm with the antistalling pivot rule described in the previous section. In particular, at a general iteration t of the algorithm, let B^t, x^t, and

y^t be respectively the basis, the basic solution, and the improving feasible direction considered by the algorithm. Let $N^t := [n] \setminus B^t$. If $x^t \neq x^{t-1}$ (i.e., we encounter x^t for the first time, as a result of a non-degenerate pivot) let $y^t := x^\star - x^t$. Observe that, once this is specified, the improving feasible direction is determined by the antistalling rule for all remaining degenerate pivots at x^t.

Lemma 3.5. *Let $x^t = x^{t+k}$ be the basic feasible solution associated with bases B^t and B^{t+k}, and assume $x^{t+k} \neq x^{t+k+1}$. The following holds:*

(a) $c^T y^{t+k} \leq c^T y^t$.

(b) $y_i^{t+k} = y_i^t$ *for all* $i \in N^{t+k} \cap \operatorname{supp}(y^{t+k})$.

Proof. The statement (a) follows from (3.4). The statement (b) follows from the construction of our antistalling rule by induction: if Case III never occurs during the k degenerate pivots, then $y^{t+k} = y^t$ and the statement holds trivially. Suppose that the situation of Case III appears for B^{t+j}, y^{t+j}. The improving feasible direction changes (as $y^{t+j+1} = y^{t+j} + \alpha z$), but among the non-basic coordinates this only affects the value of y_f^{t+j+1} (where f is the entering index). Immediately after, Case II occurs and f gets pivoted in, while a basic index i with $y_i^{t+j+1} = 0$ gets pivoted out, i.e., $B^{t+j+1} = \big(B^{t+j} \setminus \{i\}\big) \cup \{f\}$ and $N^{t+j+1} = \big(N^{t+j} \setminus \{f\}\big) \cup \{i\}$. Hence, the statement holds. □

The next result, which establishes (3.10) with $\lambda := \frac{(n-m)\Delta}{\delta}$ for the simplex algorithm with the antistalling pivot rule, is inspired by Kitahara and Mizuno (2014)[Theorem 3].

Lemma 3.6. *Let $x^t = x^{t+k}$ be the basic feasible solution associated with bases B^t and B^{t+k}. Assume $x^{t-1} \neq x^t$ and $x^{t+k} \neq x^{t+k+1}$. We have*

$$c^T x^{t+k+1} - c^T x^\star \leq \big(1 - \frac{\delta}{(n-m)\Delta}\big)(c^T x^{t+k} - c^T x^\star).$$

Proof. The optimality gap for x^{t+k} can be bounded as follows:

$$\begin{aligned} c^T x^{t+k} - c^T x^\star &= c^T x^t - c^T x^\star \\ &= -c^T y^t \\ &\leq -c^T y^{t+k} \\ &= -\overline{c}^T_{N^{t+k}} y^{t+k}_{N^{t+k}} \\ &= \textstyle\sum_{i \in N^{t+k}} -\overline{c}^T_{N^{t+k},i} y_i^{t+k} \\ &= \textstyle\sum_{i \in N^{t+k} | y_i^{t+k} > 0} -\overline{c}^T_{N^{t+k},i} y_i^{t+k} \\ &\leq (n-m)\Delta_{\overline{c}}^{t+k} \Delta , \end{aligned} \tag{3.11}$$

where we used Lemma 3.5(a) for the first inequality, and (3.2) for the third equality. The last equality follows from $y_i^{t+k} \geq 0, i \in N^{t+k}$ which is implied by feasibility of

y^{t+k} at x^{t+k}. The last inequality in turn follows from $\max_{\{i\in N^{t+k} | y_i^{t+k}>0\}} -\overline{c}_{N^{t+k},i} = \Delta_{\overline{c}}^{t+k}$, the fact that $|N^{t+k}| = n-m$, and since $y_i^{t+k} = y_i^t = x_i^\star - x_i^t \leq \Delta$ holds for any $i \in N^{t+k} \cap \operatorname{supp}(y^{t+k})$, using Lemma 3.5(b).

Let x_f^{t+k} denote the entering variable at $(t+k)^{\text{th}}$ iteration. Note that $x_f^{t+k+1} \neq 0$ since $x^{t+k} \neq x^{t+k+1}$. Then,

$$
\begin{aligned}
c^T x^{t+k} - c^T x^{t+k+1} &= \overline{c}_{N^{t+k}}^T (x^{t+k} - x^{t+k+1})_{N^{t+k}} \\
&= -\overline{c}_{N^{t+k},f}\, x_f^{t+k+1} \\
&\geq \Delta_{\overline{c}}^{t+k} \delta \\
&\geq \tfrac{\delta}{(n-m)\Delta} c^T (x^{t+k} - x^\star)
\end{aligned}
$$

hold, where we used (3.2) for the first equality, $\Delta_{\overline{c}}^{t+k} = -\overline{c}_{N^{t+k},f}$ and $\delta \leq x_f^{t+k+1}$ for the second inequality, and (3.11) for the last inequality. Rearranging the terms yields the lemma statement. □

Now, combining Lemma 3.4 and Lemma 3.6 allows to conclude that the simplex algorithm with the antistalling pivot rule described in Section 3.2 encounters at most $(n-m)\lceil \frac{(n-m)\Delta}{\delta} \log_e \big(m\frac{\Delta}{\delta}\big) \rceil$ distinct basic feasible solutions. Since Theorem 3.1 and Remark 3.2 in turn bound the number of consecutive degenerate steps by $\min\{n-m-1, m-1\}$ (again, due to the choice of the improving feasible direction), combining it with the latter result yields a bound on the total number of pivots required to reach an optimal vertex. However, this bound does not take into account the number of (degenerate) pivots that might have to be performed at an optimal vertex to encounter a basis satisfying the optimality criterion. This is due to the fact that the antistalling pivot rule requires an improving feasible direction, which we do not have at an optimal vertex. Hence, we have to handle this case separately.

Let B and $B^\star$ be a non-optimal and an optimal basis, both associated with an optimal basic feasible solution $x^\star$ of (3.1). Assume B is the basis of $x^\star$ obtained by the antistalling pivot rule. Since B is not optimal, there exists $f \in N$ with $\overline{c}_{N,f} = c^T z < 0$ with z as in (3.3). Observe that there exists $i \in B \cap N^\star$ with $\overline{A}_{if} > 0$ and $x_i = 0$: otherwise, all coordinates of $z_{N^\star}$ would be non-negative and hence $z \in C_{N^\star} := \{x \in \mathbb{R}^n \mid Ax = 0, x_{N^\star} \geq 0\}$. The latter however contradicts the fact that all extreme rays of the above cone $C_{N^\star}$ have non-negative scalar products $\overline{c}_{N^\star}$ with c due to optimality of $B^\star$. Hence one could perform a simplex pivot on B with entering variable f and leaving variable i. Let $B' := \big(B \setminus \{i\}\big) \cup \{f\}$ and $N' := \big(N \setminus \{f\}\big) \cup \{i\}$. Note that $\overline{c}_{N',i} \geq 0$ and $i \in N' \cap N^\star$. Either B' is an optimal basis and we stop, or there exists $j \in N' \setminus \{i\}$ with $\overline{c}_{N',j} < 0$. In the latter case, however, we can enforce the constraint $x_i = 0$ by removing the variable x_i together with the corresponding column of A and entry c_i of c from (3.1). By doing so we obtain a new LP with the number of variables smaller by one that has B' and $B^\star$ as a non-optimal and an optimal basis, respectively, since $\overline{c}_{N'\setminus\{i\},j} < 0$ and $\overline{c}_{N^\star\setminus\{i\}} \geq 0$.

We can set $B = B'$ and repeat the process. Since at each iteration a variable with index from $N^\star$ gets removed, after at most $n-m$ iterations $B = B^\star$. Since the number of degenerate pivots at an optimal vertex is bounded by $n-m$, we can state the following result.

Theorem 3.7. *Given an LP* (3.1) *and an initial feasible basis, there exists a pivot rule that makes the simplex algorithm reach an optimal basis in at most*

$$\min\{n-m, m\}(n-m)\left\lceil \frac{(n-m)\Delta}{\delta} \log_e\left(m\frac{\Delta}{\delta}\right)\right\rceil$$

simplex pivots.

In general, the value of δ is NP-hard to compute (Kuno et al., 2018). However, observe that for integral A and b (which can be assumed without loss of generality for rational LPs), one can bound $\Delta \leq ||b||_1 \Delta_A$ and $\delta \geq \frac{1}{\Delta_A}$ due to Cramer's rule (recall that Δ_A is the largest absolute value of a sub-determinant of A). Then the next statement is a straightforward corollary of the above theorem.

Corollary 3.8. *For any basic feasible solution of an LP* (3.1) *with integral A and b, there exists a sequence of at most*

$$\min\{n-m, m\}(n-m)\lceil (n-m)\Delta_A^2 ||b||_1 \log_e\left(m\Delta_A^2||b||_1\right)\rceil$$

simplex pivots leading to an optimal basis.

3.3.1 Application to combinatorial LPs

We conclude this section by observing that, using the latter corollary, one can prove the existence of short sequences of simplex pivots (that is, of length strongly-polynomial in n, m) between any two extreme points of several combinatorial LPs, that is, LPs modeling the set of feasible solutions of famous combinatorial optimization problems. We report a few examples below.

(a) LPs modeling matching/vertex-cover/edge-cover/stable-set problems in bipartite graphs. For matchings, the LP maximizes a linear function over a set of constraint of the form $\{A'x \leq 1, x \geq 0\}$, where the coefficient matrix A' is the node-edge incidence matrix of an undirected bipartite graph. After putting the LP in standard equality form by adding slack variables, we get constraints of the form $\{Ax = 1, x \geq 0\}$, where A is a totally unimodular matrix (and so $\Delta_A = 1$). The result then follows from Corollary 3.8. The same holds for vertex-cover (minimizing a linear function over constraints of the form $\{A'^T x \geq 1, x \geq 0\}$), edge-cover (minimizing a linear function over constraints of the form $\{A'x \geq 1, x \geq 0\}$), stable-set (maximizing a linear function over constraints of the form $\{A'^T x \leq 1, x \geq 0\}$).

(b) LPs modeling optimization over the *fractional* matching/vertex-cover/edge-cover/ stable-set polytopes. These correspond to the natural LP relaxations of the problems discussed in (a), for general graphs. The same LPs as in (a), in non-bipartite graphs, have a constraint matrix A' (resp. A'^T) that is not totally unimodular. However, the set of constraints defines half-integral polytopes (see Appa and Kotnyek, 2006; Nemhauser and Trotter, 1974). Note that, after putting the LPs in standard equality form, the slack variables can be loosely bounded from above by the number of variables n. Hence, $\Delta \leq n$ and $\delta \geq \frac{1}{2}$, and the result follows from Theorem 3.7.

(c) LP for the stable marriage problem. The classical stable marriage problem is defined on a bipartite graph where each node has a strict preference order over the set of its neighbours. One looks for a matching that does not contain any blocking pair, that is, a pair of nodes that mutually prefer each other with respect to their matched neighbour. There is an (exact) LP formulation for the problem (Rothblum, 1992; Vande Vate, 1989), that has constraints of the form $\{A'x \leq 1, B'x \geq 1, x \geq 0\}$. Here A' is again the node-edge incidence matrix of an undirected bipartite graph, while B' is a matrix that stems from imposing an additional constraint for each edge $\{uv\}$, that essentially prevents $\{uv\}$ from being a blocking pair:

$$x_{uv} + \sum_{w: w >_u v} x_{uw} + \sum_{w: w >_v u} x_{vw} \geq 1$$

In the above expression, $w >_u v$ (resp. $w >_v u$) means that u prefers w over v (resp. v prefers w over u). After putting the LP in standard equality form, the slack variables can be bounded by 1. Hence, $\Delta = \delta = 1$, and the result follows from Theorem 3.7.

(d) LPs modeling various flow problems (such as max flows, min cost flows, flow circulations) with unit (or bounded) capacity values. Flow problems in capacitated graphs are modeled using LPs with constraints of the form $\{A'x = b, l \leq x \leq u\}$ where the constraint matrix A' is here a node-arc incidence matrix of a directed graph, which is totally unimodular (see Schrijver, 2003). The right-hand side vectors (b, l, u) can be bounded in terms of the total capacity values. Therefore, assuming these are bounded integers, one can rely on Corollary 3.8 to get the result, similarly to (a).

One can compute the corresponding pivot sequence for the problems mentioned in (a)-(d) by running the simplex with the antistalling pivot rule. However, one needs to presolve the LP and obtain an optimal basic solution in order to use it to guide the antistalling pivot rule.

We remark that the existence of a strongly polynomial sequence of pivots for general min-cost flows (hence for problems in (d)), was already known, as it follows from the fundamental work of Orlin (1997), which applies even in the case of large capacity values. In comparison, our bound is weaker, as it requires the capacity values to be bounded.

On the other hand, despite being known to be solvable in strongly polynomial time, to the best of our knowledge, the problems in (a)-(c) were not previously known to admit strongly polynomial bounds on the number of simplex pivots for their natural LP formulations.

3.4 Computational experiments

In this section we report the computational experiments that we conducted to evaluate the performance of the antistalling pivot rule when applied to actual linear programs.

For the implementation, we used the python package CyLP by Towhidi and Orban (2016) which wraps COIN-OR's CLP solver and provides tools for implementing a preferred pivot rule in python to substitute CLP's built-in ones. For the tests, we used the benchmark Netlib LP dataset containing 93 LPs of various dimensions and sparsity. All our experiments were conducted on a laptop with Intel Core i7-13700H 2.40GHz CPU and 32GB RAM.

For each of the aforementioned LPs, we ran the simplex algorithm with our antistalling pivot rule, and compare it with other 5 well-known pivot rules (which Towhidi and Orban (2016) provide implementations for). These are:

- Dantzig's rule: the entering variable is the one with the most negative reduced cost;
- Steepest edge: the entering variable is the one which yields a direction z maximizing $\frac{-c^T z}{||z||_2}$ (i.e., maximizing the improvement normalized according to the 2-norm);
- LIFO: the entering variable is the one that minimizes the number of iterations that have past since the variable left the basis;
- Most Frequent: the entering variable is the one that maximizes the number of times it was previously chosen as the entering variable;
- Bland's rule: the entering variable is the one with the lowest index.

We tracked the number of pivots required by each of the above rules. For our antistalling pivot rule, we guided it using the feasible direction $y = x^{\star} - x$ at any non-optimal vertex x of the LP, for a pre-computed optimal basic solution x^{*} (in particular, the one founded by Dantzig's pivot rule). Intuitively, this represents the best direction that can possibly guide our rule, as it is the direction leading immediately to an optimal solution. Our objective with the computational experiments was to see whether this choice translates into fewer pivots in practice, since y might change during the degenerate steps and therefore in reality we do not have control on the actual edge-direction we end up moving along.

Among the 93 instances in the dataset, there were 84 LPs for which each of the pivot rules was able to find an optimal solution with 30 minutes of timeout. We restrict our report to these LPs. For the sake of graphical representation, we further divided these 84 problems into 2 groups of 42 LPs each, by considering the maximum number of iterations required by any of the 6 tested pivot rules. In particular, the first group contains 42 LPs for which all pivot rules were able to compute an optimal solution within 2369 pivots. The second group contains the remaining 42 LPs. Figure 3.2 shows the results.

The computational experiments showed that the antistalling pivot rule, guided by a known optimal solution, actually performs quite well, and most often manages to find a relatively short sequence of pivots to an optimal basis compared to other pivot rules. We highlight in particular the problem 25fv47, where any other pivot rule required at least 7892 pivots whereas the antistalling only needed 1468. On the other hand, there are instances where the antistalling pivot rule actually showed the worse performance compared to all its competitors. Those are the LPs fit2p and seba: 69711 and 976 pivots against at most 61925 and 873, respectively.

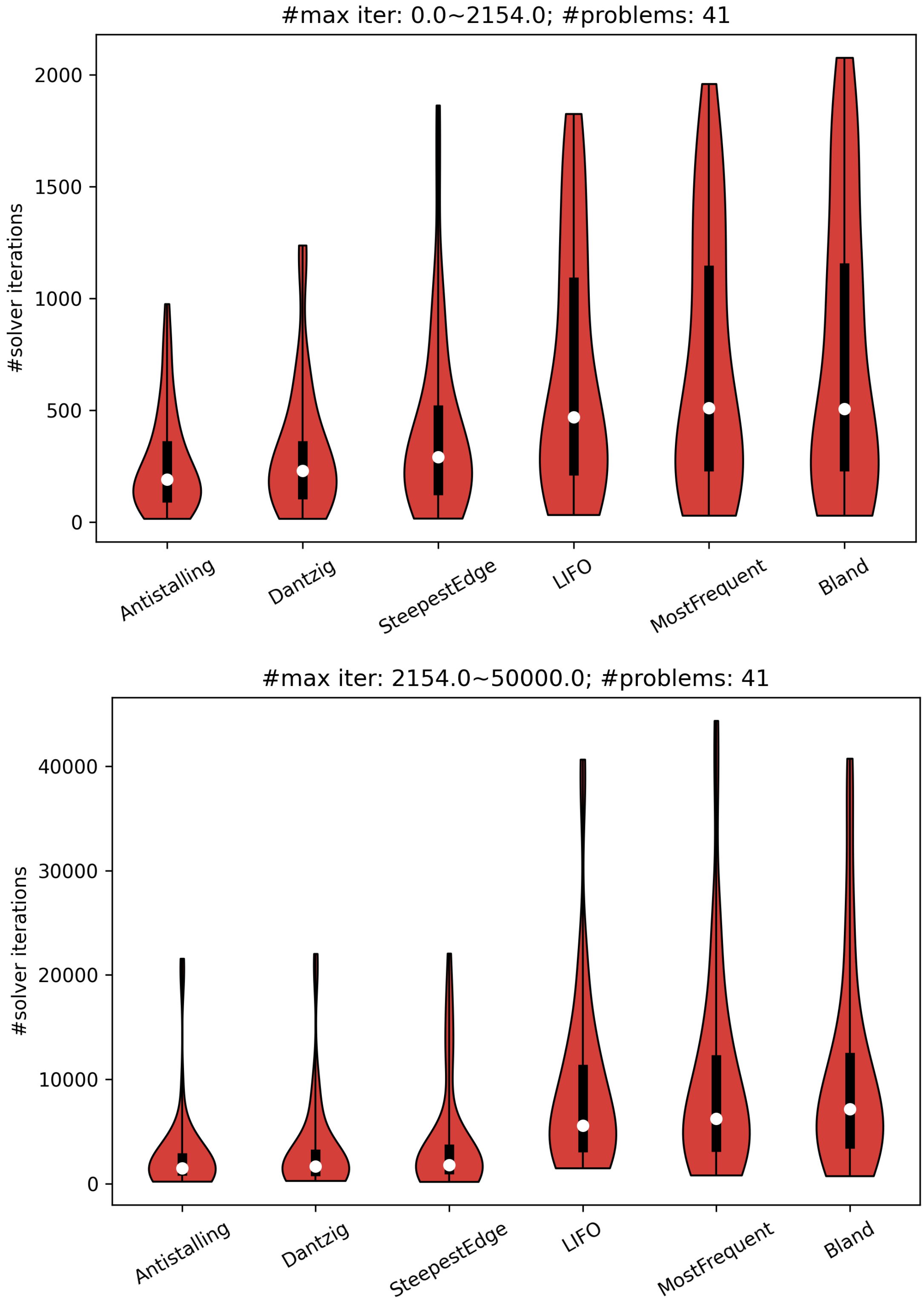

Fig. 3.2: Both the upper and the lower plots show the number of iterations required by each of the 6 considered pivot rules to solve the LPs. The 42 LPs in the upper plot are the ones solved within 2369 pivots by all pivot rules. The remaining 42 LPs are considered in the lower plot. Each of the so-called *violin* plot visualizes the distribution of the corresponding 42 numbers. White dots depict means, thick black dashes represent the intervals between 0.25- and 0.75-quantiles.

3.5 Discussion

We conclude this chapter with some remarks and potential research directions that originate from our findings.

Our results show the existence of a short sequence of degenerate simplex pivots that leads out of degeneracy. It would be interesting to show that similar results can be achieved through perturbations. Informally, perturbing the right-hand side of the constraints has the effect of "splitting" a degenerate vertex into a set of (possibly exponentially many) distinct vertices, each associated with a distinct basis. It is legitimate to ask whether every new vertex allows for a short monotone path (where monotonicity is with respect to the linear objective function of the LP), which leads to a vertex outside of this set. Note that the existence of the short degeneracy-escaping sequence of simplex pivots showed in this work, generally does not answer the latter question since some bases used in the sequence might become infeasible after perturbation.

At this point, we would like to connect to the results of Chapter 2. Given a basis B defining a degenerate vertex v, one could imagine to separate v from all its neighbouring vertices by a hyperplane: this way, one obtains a pyramid (with v at its apex). Suppose this pyramid is full-dimensional and is described by some $Ax \le b$: then, the above question is equivalent to perturbing the side facets of the pyramid to allow for a short monotone path from the new apex defined by the given basis B, to any of the vertices in the base of the pyramid. Observe that if the monotonicity assumption is relaxed, the letter can be done relying on the results of Section 2.5 (when viewing the base of the pyramid as the $d-1$-dimensional original polytope, and the pyramid as its d-dimensional extension). By Theorem 2.14, such rock extension has the property that the apex is connected to any base vertex by a path of at most linear length. However, these paths are not necessarily monotone for a given linear objective function, as discussed in Section 2.6, and moreover, the construction of the rock extension requires not only perturbing the right-hand side b but the left-hand side A as well.

Another future research direction, that seems to be supported by the performed computational experiments, would be to analyze the relation between the improving feasible direction y that is required by our antistalling pivot rule, and the actual edge-direction z along which one ends up moving after all the degenerate pivots. It would be very interesting to identify some conditions which ensure that z is a good approximation of y, e.g., in terms of the objective function's improvement.

Furthermore, as mentioned in Section 3.2, in order for our bound on the number of consecutive degenerate pivots to hold, the only requirement we need to impose when performing a pivot is that the entering variable lies in the support of the improving feasible direction y (besides, of course, having negative reduced cost). In order to obtain the results from Section 3.3, where both degenerate and non-degenerate pivots have to be taken into account, we chose the entering variable to be the one with the most negative reduced cost, i.e., we used Dantzig's rule just restricted to the coordinates in $\mathrm{supp}(y)$. Of course, one could think of analyzing the performance of the antistalling pivot rule with other choices of entering coordinates from $\mathrm{supp}(y)$, e.g., according to steepest edge or shadow vertex. Variations of these latter two rules, in particular, have been shown to play a key role in the context of $0/1$ polytopes (Black et al., 2024).

Finally, we are curious if the presented antistalling pivot rule can be applied in practice in any way. Clearly, the dependence on an improving feasible direction y, which can be quite hard to compute, is the main obstacle for that. One line of thought is to run the simplex algorithm with the antistalling pivot rule in parallel to the interior point method. Whenever the current interior point method solution x^i has better objective function value than the current simplex solution x^s, one could compute y needed for the antistalling pivot rule as $x^i - x^s$. The hope is, that if the interior point method slows down in the neighborhood of an optimal vertex (or an optimal face), the simplex can maybe reach it faster along the edges.

Bibliography

Ahuja, R. K., Orlin, J. B., Sharma, P., and Sokkalingam, P. (2002). A network simplex algorithm with O(n) consecutive degenerate pivots. *Operations Research Letters*, 30(3):141–148.

Amenta, N. and Ziegler, G. (1999). Deformed products and maximal shadows of polytopes. In Chazelle, B., Goodman, J. E., and Pollack, R., editors, *Advances in Discrete and Computational Geometry (South Hadley, MA, 1996)*, volume 223 of *Contemporary Mathematics*, pages 57–90. American Mathematical Society, Providence, RI.

Appa, G. and Kotnyek, B. (2006). A bidirected generalization of network matrices. *Networks*, 47(4):185–198.

Appel, K. and Haken, W. (1977). Every planar map is four colorable. Part I: Discharging. *Illinois Journal of Mathematics*, 21(3):429 – 490.

Appel, K., Haken, W., and Koch, J. (1977). Every planar map is four colorable. Part II: Reducibility. *Illinois Journal of Mathematics*, 21(3):491 – 567.

Avis, D. and Friedmann, O. (2017). An exponential lower bound for Cunningham's rule. *Mathematical Programming*, 161(1):271–305.

Bach, E. and Huiberts, S. (2025). Optimal smoothed analysis of the simplex method. arXiv preprint.

Basu, S., Pollack, R., and Roy, M.-F. (2006). *Algorithms in Real Algebraic Geometry*. Springer, Berlin, Heidelberg.

Belahcene, S., Marthon, P., and Aidene, M. (2018). Adaptive method for solving linear programming problems. *American Journal of Operations Research*, 8:92–111.

Bienstock, D., Pia, A. D., and Hildebrand, R. (2023). Complexity, exactness, and rationality in polynomial optimization. *Mathematical Programming Series B*, 197:661–692.

Black, A., Loera, J. D., Kafer, S., and Sanità, L. (2024). On the simplex method for 0/1 polytopes. *Mathematics of Operations Research*.

Black, A. E. (2024). Exponential lower bounds for many pivot rules for the simplex method. arXiv preprint.

Bland, R. G. (1977). New finite pivoting rules for the simplex method. *Mathematics of Operations Research*, 2(2):103–107.

Bonifas, N., di Summa, M., Eisenbrand, F., Hähnle, N., and Niemeier, M. (2014). On subdeterminants and the diameter of polyhedra. *Discrete & Computational Geometry*, 52:102–115.

Borgwardt, K. H. (1987). *The Simplex Method: A Probabilistic Analysis.* Springer, Berlin, Heidelberg.

Borgwardt, K. H. (1999). A sharp upper bound for the expected number of shadow vertices in LP-polyhedra under orthogonal projection on two-dimensional planes. *Mathematics of Operations Research*, 24(3):544–603.

Cardinal, J. and Steiner, R. (2023). Inapproximability of shortest paths on perfect matching polytopes. *Mathematical Programming Series B.*

CLP (Online). COIN-OR linear programming. *Accessed 11 Jul 2024.*

Cunningham, W. H. (1979). Theoretical properties of the network simplex method. *Mathematics of Operations Research*, 4(2):196–208.

Dadush, D. and Hähnle, N. (2016). On the shadow simplex method for curved polyhedra. *Discrete & Computational Geometry*, 59:882–909.

Dadush, D. and Huiberts, S. (2020). A friendly smoothed analysis of the simplex method. *SIAM Journal on Computing*, 49(5):STOC18–449–STOC18–499.

Dantzig, G., Ford, L., and Fulkerson, D. (1956). A primal-dual algorithm for linear programs. In Kuhn, H. and Tucker, A., editors, *Linear Inequalities and Related Systems Annals of Mathematical Studies*, pages 171–181. Princeton University Press, Princeton, NJ.

Dantzig, G. B. (1951). Maximization of a linear function of variables subject to linear inequalities. In Koopmans, T. C., editor, *Activity Analysis of Production and Allocation*, pages 339–347. John Wiley & Sons Inc.

Dantzig, G. B. (1963). *Linear programming and extensions.* Princeton University Press.

Disser, Y. and Hopp, A. V. (2019). On Friedmann's subexponential lower bound for Zadeh's pivot rule. In Lodi, A. and Nagarajan, V., editors, *IPCO 2019*, volume 11480 of *LNCS*, pages 168–180, Cham. Springer.

Gabasov, R., Kirillova, F., and Kostyukova, O. (1979). A method of solving general linear programming problems. *Doklady AN BSSR*, 23:197–200. In russian.

Gabasov, R., Kirillova, F., and Kostyukova, O. (1981). A modification of the adaptive method for solving general linear programming problems. *Vestnik BSU*, 1(2):37–41. In russian.

Garey, M. R. and Johnson, D. S. (1990). *Computers and Intractability; A Guide to the Theory of NP-Completeness.* W. H. Freeman & Co., USA.

Goldfarb, D. (1994). On the complexity of the simplex algorithm. In Gomez, S. and Hennart, J., editors, *Advances in Optimization and Numerical Analysis*, volume 275 of *MAIA*, pages 25–38. Springer, Dordrecht.

Goldfarb, D., Hao, J., and Kai, S.-R. (1990). Anti-stalling pivot rules for the network simplex algorithm. *Networks*, 20(1):79–91.

Goldfarb, D. and Sit, W. Y. (1979). Worst case behavior of the steepest edge simplex method. *Discrete Applied Mathematics*, 1(4):277 – 285.

Hansen, T. and Zwick, U. (2015). An improved version of the random-facet pivoting rule for the simplex algorithm. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing, STOC 2015*, pages 209–218, Portland. ACM.

Kabadi, S. N. and Punnen, A. P. (2008). Anti-stalling pivot rule for linear programs with totally unimodular coefficient matrix. In Neogy, S. K., Bapat, R. B., Das, A. K., and Parthasarathy, T., editors, *Mathematical Programming and Game Theory for Decision Making*, volume 1 of *Statistical Science and Interdisciplinary Research*, pages 15–20. World Scientific Publishing Co. Pte. Ltd., Singapore.

Kaibel, V. and Kukharenko, K. (2024). Rock extensions with linear diameters. *SIAM Journal on Discrete Mathematics*, 38(4):2982–3003.

Kaibel, V. and Walter, M. (2015). Simple extensions of polytopes. *Mathematical Programming Series B*, 154(1-2):381–406.

Kalai, G. and Kleitman, D. (1992). A quasi-polynomial bound for the diameter of graphs of polyhedra. *Bulletin of the American Mathematical Society*, 26:315–316.

Kitahara, T. and Mizuno, S. (2013). A bound for the number of different basic solutions generated by the simplex method. *Mathematical Programming*, 137:579–586.

Kitahara, T. and Mizuno, S. (2014). On the number of solutions generated by the simplex method for LP. In Xu, H., Teo, K. L., and Zhang, Y., editors, *Optimization and Control Techniques and Applications*, volume 86 of *Springer Proceedings in Mathematics & Statistics*, pages 75–90. Springer, Berlin, Heidelberg.

Klee, V. and Minty, G. J. (1972). How good is the simplex algorithm? In Shisha, O., editor, *Inequalities, III*, pages 159–175. Academic Press, New York.

Kukharenko, K. and Sanità, L. (2024). On the number of degenerate simplex pivots. In Vygen, J. and Byrka, J., editors, *Integer Programming and Combinatorial Optimization*, pages 252–264, Cham. Springer Nature Switzerland.

Kuno, T., Sano, Y., and Tsuruda, T. (2018). Computing Kitahara–Mizuno's bound on the number of basic feasible solutions generated with the simplex algorithm. *Optimization Letters*, 12:933–943.

Matschke, B., Santos, F., and Weibel, C. (2015). The width of five-dimensional prismatoids. *Proceedings of the London Mathematical Society*, 110(3):647–672.

Megiddo, N. (1986a). Improved asymptotic analysis of the average number of steps performed by the self-dual simplex algorithm. *Mathematical Programming*, 35:140–192.

Megiddo, N. (1986b). A note on degeneracy in linear programming. *Mathematical Programming*, 35(3):365–367.

Murty, K. G. (2009). Complexity of degeneracy. In Floudas, C. A. and Pardalos, P. M., editors, *Encyclopedia of Optimization*, pages 419–425. Springer, Boston.

Nemhauser, G. L. and Trotter, L. E. (1974). Properties of vertex packing and independence system polyhedra. *Mathematical Programming*, 6:48–61.

Netlib (Online). *Accessed 11 Jul 2024*.

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

Robertson, N., Sanders, D., Seymour, P., and Thomas, R. (1997). The four-colour theorem. *Journal of Combinatorial Theory, Series B*, 70(1):2–44.

Rooley-Laleh, E. (1981). *Improvements to the Theoretical Efficiency of the Simplex Method*. PhD thesis, University of Carleton, Ottawa.

Rothblum, U. G. (1992). Characterization of stable matchings as extreme points of a polytope. *Mathematical Programming*, 54(1):57–67.

Santos, F. (2012). A counterexample to the Hirsch conjecture. *Annals of Marthematics*, 176:383–412.

Schrijver, A. (1986). *Theory of Linear and Integer Programming*. John Wiley & Sons, Inc., New York, NY.

Schrijver, A. (2003). *Combinatorial Optimization - Polyhedra and Efficiency*. Springer, Berlin, Heidelberg.

Shamir, R. (1987). The efficiency of the simplex method: A survey. *Management Science*, 33:301–334.

Smale, S. (1998). Mathematical problems for the next century. *The Mathematical Intelligencer*, 20:7–15.

Spielman, D. A. and Teng, S.-H. (2004). Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *J. ACM*, 51(3):385–463.

Sukegawa, N. (2019). An asymptotically improved upper bound on the diameter of polyhedra. *Discrete & Computational Geometry*, 62:690–699.

Terlaky, T. (2009). Lexicographic pivoting rules. In Floudas, C. A. and Pardalos, P. M., editors, *Encyclopedia of Optimization*, pages 1870–1873. Springer, Boston.

Terlaky, T. and Zhang, S. (1993). Pivot rules for linear programming: A survey on recent theoretical developments. *Annals of Operations Research*, 46-47(1):203–233.

Todd, M. J. (2014). An improved Kalai–Kleitman bound for the diameter of a polyhedron. *SIAM Journal on Discrete Mathematics*, 28:1944–1947.

Towhidi, M. and Orban, D. (2016). Customizing the solution process of COIN-OR's linear solvers with Python. *Mathematical Programming Computation*, 8:377–391.

Vande Vate, J. H. (1989). Linear programming brings marital bliss. *Operations Research Letters*, 8(3):147–153.

Vershynin, R. (2009). Beyond Hirsch conjecture: Walks on random polytopes and smoothed complexity of the simplex method. *SIAM Journal on Computing*, 39(2):646–678.

Zadeh, N. (2009). What is the worst case behavior of the simplex algorithm? In Avis, D., Bremner, D., and Deza, A., editors, *Polyhedral computation*, volume 48 of *CRM Proc. Lecture Notes*, pages 131–143. American Mathematical Society, Providence.

Ziegler, G. M. (1994). *Lectures on Polytopes*. Graduate texts in mathematics. Springer, New York, NY.