6 Ellipsoid Method

The Ellipsoid Method is a procedure that can be used to solve in particular linear and, more generally, convex optimization problems. We will focus on linear optimization problems with a bounded feasible region, i.e., the feasible region is a polytope P, as this covers the most important use cases of the Ellipsoid Method in the context of Combinatorial Optimization.

A crucial advantage of the Ellipsoid Method compared to other approaches for linear programming, like the Simplex Method or interior-point methods, is that one does not need to process an explicit inequality description of P to solve the linear program. More precisely, the way the Ellipsoid Method gains information about P is through a method that solves an arguably simpler sub-problem, known as the *separation problem*.

6.1 Separation problem

The separation problem for a polyhedron $P \subseteq \mathbb{R}^n$ is defined as follows.

Definition 6.1: Separation problem & separation oracle

Given a point $y \in \mathbb{R}^n$:

- Decide whether $y \in P$, and if this is not the case,
- find $c \in \mathbb{R}^n$ such that $P \subseteq \{x \in \mathbb{R}^n \colon c^\top x < c^\top y\}$.

A procedure that solves the separation problem (for P) is often called a *separation oracle* (for P).

The second point in the above definition, where a vector $c \in \mathbb{R}^n$ has to be returned, is often called *separating y from p*.

In the context of the Ellipsoid Method, we call a hyperplane $H = \{x \in \mathbb{R}^n \colon c^\top x = \alpha\}$ a *separating hyperplane*, or more precisely a *y-separating hyperplane*, if the following holds: Let

$$H_\leq := \{x \in \mathbb{R} : c^\top x \leq \alpha\}, \text{ and}$$
$$H_\geq := \{x \in \mathbb{R} : c^\top x \geq \alpha\}$$

be the two closed half-spaces defined by H; then H is *separating* if P is contained in one of the above half-spaces, y in the other one, and H does not simultaneously contain y and is touching P, i.e., either $y \notin H$ or $P \cap H = \emptyset$. Hence, if $H = \{x \in \mathbb{R}^n : c^\top x = \alpha\}$ is a separating hyperplane, then c is a vector that solves the separation problem for y. Vice-versa, if $c \in \mathbb{R}^n$ is a vector solving the separation problem for y, then there exists a separating hyperplane with normal vector c, for example, $H = \{x \in \mathbb{R}^n : c^\top x \leq c^T y\}$.

The fact that the Ellipsoid Method is based on a separation oracle and does not need to process an explicit inequality description of P, allows it to solve in polynomial time many linear programs with an exponential number of constraints. For example, one can use the Ellipsoid Method to optimize any linear function over the matching polytope, and thus, in particular, one can efficiently compute a maximum weight matching via the Ellipsoid Method. The reason for this is that even though the matching polytope has an exponential number of constraints, one can construct a polynomial time separation oracle for it.

6.2 Optimization results based on Ellipsoid Method

Before formally introducing the Ellipsoid Method, we first highlight what kind of results can be obtained through it. The Ellipsoid Method is a very general and versatile technique, and a thorough coverage of its implications is beyond the scope of this course. We therefore restrict ourselves to a statement about optimizing over $\{0,1\}$-polytopes, which is of particular interest in the context of Combinatorial Optimization, where we often deal with $\{0,1\}$-polytopes.

Theorem 6.2

Let $P \subseteq \mathbb{R}^n$ be a $\{0,1\}$-polytope for which we are given a separation oracle. Furthermore, let $w \in \mathbb{Z}^n$. Then the Ellipsoid Method allows for finding an optimal vertex solution to the linear program $\max\{w^\top x \colon x \in P\}$ using a polynomial number (in n) of operations and calls to the separation oracle for P.

Hence, Theorem 6.2 implies that an efficient separation oracle for a $\{0,1\}$-polytope P is all we need to efficiently solve any linear program over P.

6.3 Example applications

We showcase the construction of a separation oracle on the example of the dominant of the r-arborescence polytope, which has an exponential number of facets. Hence, the Simplex Method, as we have discussed it, is not suitable for such a problem, because even just reading in the constraints to construct the simplex tableau would take exponential time, before we even start to do any computations.

In the problem sets, we will see further applications of how the Ellipsoid Method can be used to solve non-trivial problems through the construction of an appropriate separation oracle.

Minimum weight r-arborescence

Consider the problem of finding a minimum weight r-arborescence in a directed graph $G = (V, A)$ with non-negative arc weights $w \colon A \to \mathbb{Z}_{\geq 0}$. Clearly, this problem is equivalent to minimizing the linear function w over the dominant P of the r-arborescence polytope which, by Theorem 5.20, is described by

$$P = \left\{x \in \mathbb{R}^A_{\geq 0} \colon x(\delta^-(S)) \geq 1 \quad \forall S \subseteq V \setminus \{r\}, S \neq \emptyset \ \right\} \ .$$

Strictly seen, we cannot apply Theorem 6.2 to P, because P is not a $\{0,1\}$-polytope. More precisely, although P is integral, it is unbounded. This can easily be remedied by intersecting P with the hypercube to obtain

$$Q = P \cap [0,1]^A \ .$$

In the problem sets we show that the intersection of the dominant of any $\{0,1\}$-polytope with the unit hypercube results in a $\{0,1\}$-polytope. Hence, Q is a $\{0,1\}$-polytope. Moreover, because P is the dominant of a $\{0,1\}$-polytope and we are minimizing a non-negative linear function over it, there is an optimal solution x^* to $\min\{w^\top x \colon x \in P\}$ with $x^* \in \{0,1\}^n$. Hence, $x^* \in Q$. Thus, a minimum r-arborescence can be found by finding a vertex solution to

$$\min\{w^\top x \colon x \in Q\} \ . \tag{6.1}$$

By Theorem 6.2, it suffices to find a polynomial-time separation oracle for Q to find an optimal vertex solution to (6.1). It therefore remains to design a polynomial-time separation oracle for Q.

Hence, let $y \in \mathbb{R}^A$. We first check whether $y \in [0,1]^A$. This is easy to check since the unit hypercube is defined by $2|A|$ constraints. If one of these constraints is violated, then it immediately leads to a separating hyperplane. Hence, assume $y \in [0,1]^A$. It remains to check whether one of the constraints $x(\delta^-(S)) \geq 1$ for some $S \subseteq V \setminus \{r\}, S \neq \emptyset$ is violated. This can be checked via minimum s-t cut computations. For each $v \in V \setminus \{r\}$, we compute a minimum r-v cut $S_v \subseteq V$, where we use y as the capacities on the arcs. If one of these cuts S_v has a value strictly less than 1, then the constraint corresponding to $V \setminus S_v$ is violated and leads to a separating hyperplane. Otherwise, if $y(\delta^+(S_v)) \geq 1$ for all $v \in V$, then y satisfies all constraints of P and therefore $y \in Q$, because of the following. Assume that there was a violated constraint, i.e., $y(\delta^-(S)) < 1$ for some $S \subseteq V \setminus \{r\}, S \neq \emptyset$. Then, for any $v \in S$ we have

$$y(\delta^+(S_v)) \leq y(\delta^+(V \setminus S)) = y(\delta^-(S)) < 1 \ ,$$

where the first inequality follows from the fact that $V \setminus S$ is an r-v cut, and its cut value is thus at least as large as the value of the cut S_v, which is by definition a minimum r-v cut.

This shows that we can solve the separation problem over Q, and thus, by Theorem 6.2, we can find an optimal vertex solution to (6.1) in polynomial time through the Ellipsoid Method. Furthermore, as discussed above, such a vertex solution corresponds to a minimum weight r-arborescence.

6.4 Ellipsoid Method for finding point in full-dimensional $\{0,1\}$-polytope

Our final goal is to show how to do linear optimization via the Ellipsoid Method. However, to approach this goal, we start with the following simpler problem, which will lead us to linear optimization later on.

$$\text{Given a full-dimensional polytope } P \subseteq \mathbb{R}^n, \text{ find a point } x \in P. \tag{6.2}$$

As we will discuss later, a linear programming problem over some (full-dimensional) polytope Q, i.e.,

$$\begin{aligned} \max \quad & w^\top x \\ & x \in Q \ , \end{aligned} \tag{6.3}$$

can be reduced to finding points in $P_b = Q \cap \{w^\top x \geq b\}$ for varying right-hand sides $b \in \mathbb{R}$. For the time being, we do not assume that P is a $\{0,1\}$-polytope. We will add further conditions later on when we need them.

As the name suggests, the Ellipsoid Method works with ellipsoids. To be precise, additional to a separation oracle for P, it needs an ellipsoid E_0 that contains P, which is often called the *starting ellipsoid*. (In turns out that in case of a $\{0,1\}$-polytope P, one can choose a trivial ellipsoid, namely the smallest ball containing $[0,1]^n$. This is the reason why Theorem 6.2 does not require a starting ellipsoid to be given.)

We recall that an ellipsoid is the image of the unit ball under an affine bijection, which can be defined as follows.

Definition 6.3: Ellipsoid

An ellipsoid in $\mathbb{R}^n$ is a set

$$E(a, A) := \{x \in \mathbb{R}^n \colon (x-a)^\top A^{-1}(x-a) \leq 1\} \ ,$$

where $a \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$ is a positive definite matrix. The point a is called the *center* of the ellipsoid $E(a, A)$.[a]

[a] We use the common convention that a positive definite matrix is by definition symmetric.

In particular, an ellipsoid is always full-dimensional. Notice that the above definition of ellipsoid indeed corresponds to the image of the unit ball under an affine bijection. This can be seen as follows. A matrix $A \in \mathbb{R}^{n\times n}$ is positive definite if and only if there is a full-rank matrix $Q \in \mathbb{R}^{n\times n}$ such that $A = QQ^\top$. Hence, $A^{-1} = (Q^\top)^{-1}(Q^{-1}) = (Q^{-1})^\top Q^{-1}$, and therefore

$$\begin{aligned} E(a, A) &= \{x \in \mathbb{R}^n \colon \|Q^{-1}(x-a)\|_2 \leq 1\} \\ &= \{y + a \colon y \in \mathbb{R}^n, \|Q^{-1}y\|_2 \leq 1\} && \text{(Substitution with } y = x - a.\text{)} \\ &= \{Qz + a \colon z \in \mathbb{R}^n, \|z\|_2 \leq 1\} \ . && \text{(Substitution with } z = Q^{-1}y.\text{)} \end{aligned}$$

Figure 6.1 shows an example of an ellipsoid for $n = 2$.

The separation oracle for P and the ellipsoid $E_0 \supseteq P$ is all that is needed to run the Ellipsoid Method and obtain a point $x \in P$. However, to make sure that the Ellipsoid Method runs in polynomial, time we need a further condition. One sufficient condition for the Ellipsoid Method to run in polynomial time is that $\log\left({}^{\mathrm{vol}(E_0)}/_{\mathrm{vol}(P)}\right)$ is polynomially bounded by the input. We will see later that this holds for most cases we are interested in.

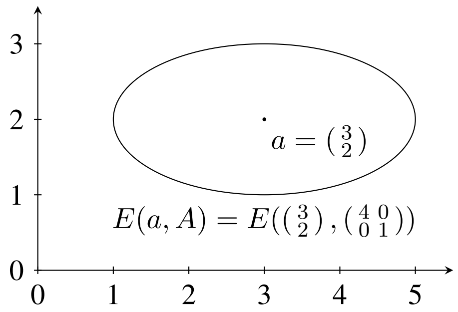

Figure 6.1: An example of an axis-parallel ellipsoid $E(a, A)$ in two dimensions. Notice that the eigenvectors of A correspond to the axes of the ellipsoid, and the square roots of the eigenvalues correspond to the radii of the corresponding axes.

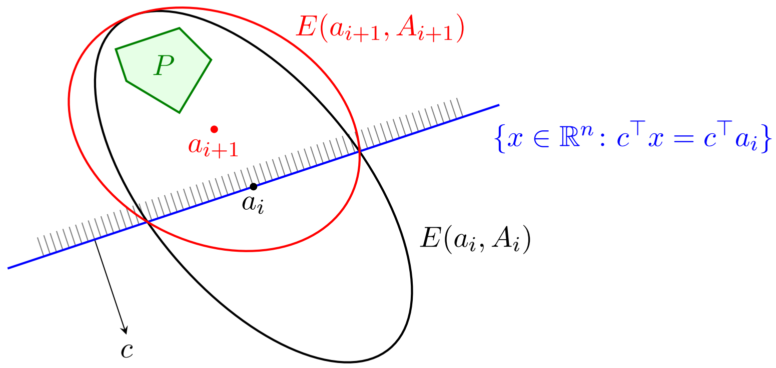

Figure 6.2: Illustration of a single iteration of the Ellipsoid Method. The polytope P is inside each ellipsoid considered by the Ellipsoid Method.

6.4.1 Description of Ellipsoid Method

The Ellipsoid Method to solve (6.2) is described in Algorithm 8.

Algorithm 8: Ellipsoid Method

Input : Separation oracle for a polytope $P \subseteq \mathbb{R}^n$ with $\dim(P) = n$, and an ellipsoid $E_0 = E(a_0, A_0)$ with $P \subseteq E_0$.
Output: A point $y \in P$.
$i = 0$.
while $a_i \notin P$ *(checked with separation oracle)* **do**
  Get $c \in \mathbb{R}^n$ such that $P \subseteq \{x \in \mathbb{R}^n : c^\top x < c^\top a_i\}$, using separation oracle.
  Find min. volume ellipsoid $E_{i+1} = E(a_{i+1}, A_{i+1})$ containing $E_i \cap \{x \in \mathbb{R}^n : c^\top x \leq c^\top a_i\}$.
  $i = i + 1$.
return a_i.

We will soon give some more details on how to compute E_{i+1}. It turns out that there is a relatively simple way to describe E_{i+1} in terms of E_i and c. Unfortunately, as we will see soon, this description of E_{i+1} involves taking a square root, which is an operation that we cannot perform up to arbitrary precision under the typical computational assumptions. Hence, in practice, one only computes an ellipsoid E'_{i+1} that approximates E_{i+1} and has a polynomial-size description. To simplify the exposition we will not go into these numerical details and assume that we use an exact description of E_{i+1}.

Figure 6.2 illustrates one iteration of the Ellipsoid Method. Notice that we have $P \subseteq E_i$ for each ellipsoid considered in the Ellipsoid Method. This can easily be verified by induction. The first ellipsoid E_0 contains P by assumption. Furthermore, for each iteration i, $P \subseteq \{x \in \mathbb{R}^n \colon c^\top x \leq c^\top a_i\}$, and hence $P \subseteq E_i \cap \{x \in \mathbb{R}^n \colon c^\top x \leq c^\top a_i\}$, as $P \subseteq E_i$. We thus obtain $P \subseteq E_{i+1}$ because $E_i \cap \{x \in \mathbb{R}^n \colon c^\top x \leq c^\top a_i\} \subseteq E_{i+1}$ by definition of E_{i+1}.

6.4.2 Getting a bound on the number of iterations

The key property of the constructed ellipsoids, besides the fact that they contain P, is that they shrink in terms of volume. More precisely, we have the following.

Lemma 6.4

$$\frac{\mathrm{vol}(E_{i+1})}{\mathrm{vol}(E_i)} < e^{-\frac{1}{2(n+1)}} \quad .$$

Before proving Lemma 6.4, we observe that it immediately implies an upper bound on the number of iterations that the Ellipsoid Method performs.

Lemma 6.5

The Ellipsoid Method will stop after at most $2(n+1)\ln\left(\frac{\mathrm{vol}(E_0)}{\mathrm{vol}(P)}\right)$ iterations.

Proof. Let $L \in \mathbb{Z}_{\geq 0}$ be the last iteration of the Ellipsoid Method, i.e., the value of the variable i in Algorithm 8 when it terminates. Since E_L contains P, we must have $\mathrm{vol}(P) \leq \mathrm{vol}(E_L)$, which, combined with Lemma 6.4, leads to

$$\mathrm{vol}(P) \leq \mathrm{vol}(E_L) \leq \mathrm{vol}(E_0) e^{-\frac{L}{2(n+1)}} \quad ,$$

and thus

$$L \leq 2(n+1)\ln\left(\frac{\mathrm{vol}(E_0)}{\mathrm{vol}(P)}\right) \quad .$$

□

In what follows, we will prove Lemma 6.4 and give an explicit description of the ellipsoid E_{i+1} based on c and $E_i = E(a_i, A_i)$. Later, we will show how the bound on the number of iterations given by Lemma 6.5 can be used to prove that the Ellipsoid Method runs in polynomial time for any full-dimensional $\{0,1\}$-polytope. To this end, we will first make the link between checking feasibility of a polytope and optimizing an LP more explicit.

Proof of Lemma 6.4 and explicit description for E_{i+1}

A key simplification for proving Lemma 6.4 is to observe that it suffices to consider the special case where $E_i = E(0, I)$ is the unit ball and $H_i = \{x \in \mathbb{R}^n : x_1 \geq 0\}$. As we show next, one can reduce to this case by an appropriate affine transformation.

Lemma 6.6

Let $E_i = E(a_i, A_i)$ be an ellipsoid and $H_i = \{x \in \mathbb{R}^n : c^\top x \leq c^\top a_i\}$ with $c^\top A_i c = 1$ (this property of c can be achieved without loss of generality by scaling c). Let $H_B = \{x \in \mathbb{R}^n : x_1 \geq 0\}$, and let E_B be a minimum volume ellipsoid containing $E(0, I) \cap H_B$. Let $\rho \colon \mathbb{R}^n \to \mathbb{R}^n$ be an affine bijection defined as follows:

$$\rho(x) = Q_i R x + a_i \ ,$$

where $Q_i \in \mathbb{R}^{n \times n}$ is any matrix such that $A_i = Q_i Q_i^\top$, and R is any orthogonal matrix satisfying $R^\top Q_i^\top c = -e_1$, where $e_1 = (1, 0, \ldots, 0) \in \{0,1\}^n$ is the vector with a single one in the first coordinate and zeros everywhere else.
Then, $\rho(E(0, I)) = E_i$, $\rho(H_B) = H_i$, and $\rho(E_B)$ is a minimum volume ellipsoid containing $E_i \cap H_i$.

Hence, Lemma 6.6 shows that to prove Lemma 6.4, it suffices to consider the unit ball case cut by the hyperplane H_B. Indeed, the ratio between a minimum volume ellipsoid E_{i+1} containing $E_i \cap H_i$, and $\mathrm{vol}(E_i) = \mathrm{vol}(\rho(E(0, I)))$ is equal to

$$\frac{\mathrm{vol}(E_{i+1})}{\mathrm{vol}(E_i)} = \frac{\mathrm{vol}(\rho(E_B))}{\mathrm{vol}(\rho(E(0, I)))} = \frac{\mathrm{vol}(E_B)}{\mathrm{vol}(E(0, I))} \ , \tag{6.4}$$

where the second equality follows by the fact that ρ scales the volumes of all measurable sets by the same factor.

Also, the lemma provides an explicit transformation ρ that allows for transforming a minimum volume ellipsoid for this special case into a minimum volume ellipsoid for the general case. Later, we will use this transformation to give an explicit description of the center a_{i+1} and defining matrix A_{i+1} for the next ellipsoid $E_{i+1} = E(a_{i+1}, A_{i+1})$ in the Ellipsoid Method.

Notice that $c^\top A_i c = 1$ implies $\|Q_i^\top c\|_2 = 1$, and hence, there is an orthogonal matrix R that maps $Q_i^\top c$ to $-e_1$. (Because an orthogonal transformation does not change lengths, $\|Q_i^\top c\|_2 = 1$ is necessary for R to exist.)

Proof of Lemma 6.6. The proof plan is as follows. We start by observing that ρ is a bijection, $\rho(E(0, I)) = E_i$, and $\rho(H_i) = H_B$. Furthermore, applying ρ to any measurable set changes its volume by a factor that only depends on ρ (a property that is well-known for affine transformations). From these properties we can interpret the special case with ball $E(0, I)$ and half-space H_B as the preimage with respect to ρ of the original problem. Because all volumes are scaled by the same factor, a minimum volume ellipsoid in the special case thus corresponds to a minimum volume ellipsoid for the original problem.

The transformation ρ is indeed a bijection since R is full-rank and Q_i is full-rank because $A_i = Q_i Q_i^\top$ and A_i is full-rank. Furthermore,

$$\begin{aligned}
\rho(E(0,I)) &= \{\rho(x) \colon x^\top x \leq 1\} \\
&= \{Q_i R x + a_i \colon x \in \mathbb{R}^n, x^\top x \leq 1\} \\
&= \left\{y \in \mathbb{R}^n \colon (y-a_i)^\top (Q_i^{-1})^\top R R^{-1} Q_i^{-1}(y-a_i) \leq 1\right\} && (y = Q_i R x + a_i) \\
&= \left\{y \in \mathbb{R}^n \colon (y-a_i)^\top A_i^{-1}(y-a_i) \leq 1\right\} && (Q_i Q_i^\top = A_i) \\
&= E_i \enspace .
\end{aligned}$$

Similarly, the hyperplane H_B gets mapped to H_i:

$$\begin{aligned}
\rho(H_B) &= \{\rho(x) \colon x \in \mathbb{R}^n, x_1 \geq 0\} \\
&= \{Q_i R x + a_i \colon x \in \mathbb{R}^n, x_1 \geq 0\} \\
&= \{y \in \mathbb{R}^n \colon (R^{-1} Q_i^{-1}(y-a_i))^\top e_1 \geq 0\} && (y = Q_i R x + a_i) \\
&= \{y \in \mathbb{R}^n \colon (y-a_i)^\top (Q_i^{-1})^\top R e_1 \geq 0\} && (R^{-1} = R^\top \text{ by orthogonality of } R) \\
&= \{y \in \mathbb{R}^n \colon -(y-a_i)^\top c \geq 0\} && (R^\top Q_i^\top c = -e_1) \\
&= \{y \in \mathbb{R}^n \colon c^\top y \leq c^\top a_i\} \\
&= H_i \enspace .
\end{aligned}$$

Notice that the above relations combined with $E_B \supseteq E(0,I) \cap H_B$ imply that $\rho(E_B)$ is indeed an ellipsoid that contains $\rho(E(0,I)) \cap \rho(H_B) = E_i \cap H_i$. It remains to show that $\rho(E_B)$ has minimum volume among all ellipsoid containing $E_i \cap H_i$. To prove this, let E_{i+1} be a minimum volume ellipsoid containing $E_i \cap H_i$ and we will show $\mathrm{vol}(E_{i+1}) \geq \mathrm{vol}(\rho(E_B))$ to finish the proof.

We recall a well-known property of affine functions $x \mapsto Ax + v$, namely that they scale any volume by the same factor $|\det A|$. For ρ this implies that for any measurable set $U \subseteq \mathbb{R}^n$ we have

$$\mathrm{vol}(\rho(U)) = \mathrm{vol}(U) \cdot |\det(Q_i R)| = \mathrm{vol}(U) \cdot |\det Q_i||\det R| = \mathrm{vol}(U) \cdot |\det Q_i| \enspace . \tag{6.5}$$

Notice that $\rho^{-1}(E_{i+1})$ is an ellipsoid containing $\rho^{-1}(E_i) \cap \rho^{-1}(H_i) = E(0,I) \cap H_B$, and hence

$$\mathrm{vol}(\rho^{-1}(E_{i+1})) \geq \mathrm{vol}(E_B) \enspace , \tag{6.6}$$

because E_B has minimum volume among all ellipsoid containing $E(0,I) \cap H_B$. We thus obtain

$$\begin{aligned}
\mathrm{vol}(E_{i+1}) &= \mathrm{vol}(\rho^{-1}(E_{i+1})) \cdot |\det Q_i| && (\text{by } 6.5) \\
&\geq \mathrm{vol}(E_B) \cdot |\det Q_i| && (\text{by } 6.6) \\
&= \mathrm{vol}(\rho(E_B)) \enspace , && (\text{by } 6.5)
\end{aligned}$$

thus completing the proof. □

Lemma 6.7

Let $H_B = \{x \in \mathbb{R}^n \colon x_1 \geq 0\}$. Then the ellipsoid

$$E_B = \left\{ x \in \mathbb{R}^n \;\middle|\; \left(\frac{n+1}{n}\right)^2 \left(x_1 - \frac{1}{n+1}\right)^2 + \frac{n^2-1}{n^2} \sum_{j=2}^{n} x_j^2 \leq 1 \right\} \tag{6.7}$$

contains $E(0,I) \cap H_B$.

The ellipsoid E_B is actually even the minimum volume ellipsoid containing $E(0,I) \cap H_B$. We do not show this fact in these notes. (We will prove this in one of the problem sets.) However, we will show that the ratio ${}^{\mathrm{vol}(E_B)}/_{\mathrm{vol}(E(0,I))}$ satisfies the inequality of Lemma 6.4. Of course, even without accepting that E_B is the smallest ellipsoid containing $E(0,I) \cap H_B$, this shows Lemma 6.4. Furthermore, also in the analysis that follows, we never need to prove that E_B is the smallest ellipsoid containing $E(0,I) \cap H_B$, because we can simply assume that, in the Ellipsoid Method, we work with the description of E_B given by (6.7) without assuming that it has minimum volume. We will later generalize the description of E_B given in (6.7) to the general case when E_i is not necessarily the unit ball and H_B is replaced by a general half-space going through the center of E_i.

Proof of Lemma 6.7. Let $x \in E(0,I) \cap H_B$. We have

$$\begin{aligned}
&\left(\frac{n+1}{n}\right)^2 \left(x_1 - \frac{1}{n+1}\right)^2 + \frac{n^2-1}{n^2} \sum_{j=2}^{n} x_j^2 \\
&\quad = \frac{n^2+2n+1}{n^2} x_1^2 - \left(\frac{n+1}{n}\right)^2 \frac{2x_1}{n+1} + \frac{1}{n^2} + \frac{n^2-1}{n^2} \sum_{j=2}^{n} x_j^2 \\
&\quad = \frac{2n+2}{n^2} x_1^2 - \frac{2n+2}{n^2} x_1 + \frac{1}{n^2} + \frac{n^2-1}{n^2} \underbrace{\sum_{j=1}^{n} x_j^2}_{\leq 1} && (x \in E(0,I)) \\
&\quad \leq \frac{2n+2}{n^2} \underbrace{x_1(x_1-1)}_{\leq 0} + 1 && (0 \leq x_1 \leq 1) \\
&\quad \leq 1 \enspace ,
\end{aligned}$$

and thus $x \in E_B$. $\square$

Proof of Lemma 6.4. As discussed, we can assume due to Lemma 6.6 that $E_i = E(0,I)$ is the unit ball, and the separating hyperplane used in the Ellipsoid Method at iteration i, which goes through the center of E_i, is given by $H_B = \{x \in \mathbb{R}^n \colon x_1 \geq 0\}$. From (6.7) we can read off the matrix A_{i+1} and the center point a_{i+1}, which define the ellipsoid E_{i+1}, i.e.,

$$E_{i+1} = \{x \in \mathbb{R}^n \colon (x - a_{i+1})^\top A_{i+1}^{-1} (x - a_{i+1}) \leq 1\} \enspace .$$

We have

$$a_{i+1} = \left(\frac{1}{n+1}, 0, 0, \dots, 0\right)^\top ,$$

$$A_{i+1}^{-1} = \begin{pmatrix} (\frac{n+1}{n})^2 & 0 & 0 & \dots & 0 \\ 0 & \frac{n^2-1}{n^2} & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ & & & \ddots & 0 \\ 0 & \dots & & 0 & \frac{n^2-1}{n^2} \end{pmatrix} ,$$

and thus

$$A_{i+1} = \begin{pmatrix} (\frac{n}{n+1})^2 & 0 & 0 & \dots & 0 \\ 0 & \frac{n^2}{n^2-1} & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ & & & \ddots & 0 \\ 0 & \dots & & 0 & \frac{n^2}{n^2-1} \end{pmatrix} .$$

As discussed, an ellipsoid $E(a, A)$ is the image of the unit ball with respect to the affine transformation $\phi(x) = Qx + a$, where $Q \in \mathbb{R}^{n \times n}$ is such that $A = QQ^\top$. Thus

$$\mathrm{vol}(E(a, A)) = \mathrm{vol}(E(0, I)) \cdot |\det(Q)| = \mathrm{vol}(E(0, I)) \cdot \sqrt{\det(A)} .$$

Hence,

$$\begin{aligned} \frac{\mathrm{vol}(E_{i+1})}{\mathrm{vol}(E_i)} &= \frac{\sqrt{\det(A_{i+1})}}{\sqrt{\det(I)}} = \sqrt{\det(A_{i+1})} \\ &= \frac{n}{n+1}\left(\frac{n^2}{n^2-1}\right)^{\frac{n-1}{2}} \\ &= \left(1 - \frac{1}{n+1}\right)\left(1 + \frac{1}{n^2-1}\right)^{\frac{n-1}{2}} \\ &< e^{-\frac{1}{n+1}} e^{\frac{n-1}{2(n^2-1)}} = e^{-\frac{1}{n+1}} e^{\frac{1}{2(n+1)}} = e^{-\frac{1}{2(n+1)}} \quad , \quad (1 + x < e^x \;\; \forall x \in \mathbb{R} \setminus \{0\}) \end{aligned}$$

as desired.

□

6.4.3 From the unit ball to the general case

In this section we discuss briefly how to derive an explicit description of the ellipsoid E_{i+1} in the general case, where E_i is not necessarily the unit ball and the half-space cutting E_i is a general half-space going through the origin a_i of E_i. Having such a description of E_{i+1} allows us to make the step of computing E_{i+1} in Algorithm 8 explicit.

We perform the generalization in two steps. First, we still assume that E_i is the unit ball. However, we consider an arbitrary half-space cutting it. Then, we extend this case to the general case.

An alternative way to get a description of E_{i+1} and a_{i+1} is to derive an explicit affine transformation that directly transforms the unit ball and half-space $\{x \in \mathbb{R}^n \colon x_1 \geq 0\}$ to a general ellipsoid and general half-space going through its center.

General half-space cutting $\boldsymbol{E(0, I)}$

Assume that $E_i = E(0, I)$ is still the unit ball, and consider a general half-space $H_i = \{x \in \mathbb{R}^n \colon c^\top x \leq 0\}$, where $\|c\|_2 = 1$, which can easily be achieved by scaling c. One can use an orthogonal transformation to transform this case to the one where the half-space is given by $\{x \in \mathbb{R}^n \colon x_1 \geq 0\}$. One can verify (and we do this in the problem sets) that this leads to an ellipsoid

$$
\begin{aligned}
E_{i+1} &= E(a_{i+1}, A_{i+1}) \ \text{, where} \\
a_{i+1} &= -\frac{1}{n+1} c \ \text{, and} \\
A_{i+1} &= \frac{n^2}{n^2-1}\left(I - \frac{2}{n+1} c c^\top\right) \ .
\end{aligned}
$$

General case

Now let $E_i = E(a_i, A_i)$ be a general ellipsoid, and let $H_i = \{x \in \mathbb{R}^n \colon c^\top x \leq c^\top a_i\}$ be a general hyperplane going through the center a_i of E_i. Since A_i is positive definite, there is a matrix $Q_i \in \mathbb{R}^{n\times n}$ such that $A_i = Q_i Q_i^\top$. Consider the affine bijective transformation $\phi(x) = Q_i x + a_i$ and its inverse $\phi^{-1}(x) = Q_i^{-1}(x - a_i)$. The function ϕ transforms $E(0, I)$ to $E_i = E(a_i, A_i)$. Hence, we can first apply ϕ^{-1} to E_i and H_i to obtain $E_i' = E(0, I)$ and H_i', then we can use our description of E_{i+1} for the unit ball case, and finally, we transform the found ellipsoid back using ϕ.

We start by describing the image of H_i under ϕ^{-1} to obtain H_i'.

$$
\begin{aligned}
H_i' = \phi^{-1}(H_i) &= \{Q_i^{-1}(x - a_i) \colon x \in \mathbb{R}^n, c^\top x \leq c^\top a_i\} \\
&= \{y \in \mathbb{R}^n \colon c^\top(Q_i y + a_i) \leq c^\top a_i\} && (y = Q_i^{-1}(x - a_i)) \\
&= \{y \in \mathbb{R}^n \colon c^\top Q_i y \leq 0\} \ .
\end{aligned}
$$

Hence, we have

$$\begin{aligned} H'_i &= \{x \in \mathbb{R}^n \colon d^\top x \leq 0\} \ \text{, where} \\ d &= \frac{Q_i^\top c}{\sqrt{c^\top Q_i Q_i^\top c}} = \frac{Q_i^\top c}{\sqrt{c^\top A_i c}} \ . \end{aligned}$$

By the unit ball case discussed previously, the minimum volume ellipsoid E'_{i+1} that contains $E(0, I) \cap H'_i$ is given by

$$\begin{aligned} E'_{i+1} &= E(a'_{i+1}, A'_{i+1}) \ \text{, where} \\ a'_{i+1} &= -\frac{1}{n+1} d \ \text{, and} \\ A'_{i+1} &= \frac{n^2}{n^2-1}\left(I - \frac{2}{n+1} dd^\top\right) \ . \end{aligned}$$

Hence, the ellipsoid E_{i+1} is given by

$$\begin{aligned} E_{i+1} = \phi(E'_{i+1}) &= \{\phi(x) \colon x \in \mathbb{R}^n, (x - a'_{i+1})^\top A'^{-1}_{i+1}(x - a'_{i+1}) \leq 1\} \\ &= \{y \in \mathbb{R}^n \colon (Q_i^{-1}(y - a_i) - a'_{i+1})^\top A'^{-1}_{i+1}(Q_i^{-1}(y - a_i) - a'_{i+1}) \leq 1\} \\ &= \{y \in \mathbb{R}^n \colon (y - a_i - Q_i a'_{i+1})^\top (Q_i^{-1})^\top A'^{-1}_{i+1} Q_i^{-1}(y - a_i - Q_i a'_{i+1}) \leq 1\} \ . \end{aligned}$$

where the second equality follows by using the substitution $y = \phi(x) = Q_i x + a_i$.

From this description we can derive a description of the center a_{i+1} and defining positive definite matrix A_{i+1} of $E_{i+1} = E(a_{i+1}, A_{i+1})$. For simplicity of notation we define

$$b = \frac{A_i c}{\sqrt{c^\top A_i c}} = Q_i d \ .$$

We obtain,

$$a_{i+1} = a_i + Q_i a'_{i+1} = a_i - \frac{1}{n+1} Q_i d = a_i - \frac{1}{n+1} b \ ,$$

and

$$A^{-1}_{i+1} = (Q_i^{-1})^\top A'^{-1}_{i+1} Q_i^{-1} \ ,$$

which implies

$$A_{i+1} = Q_i A'_{i+1} Q_i^\top = \frac{n^2}{n^2-1}\left(A_i - \frac{2}{n+1} bb^\top\right) \ .$$

We can now restate the Ellipsoid Method as described in Algorithm 8 in a more explicit form, by giving explicit formulas for how to compute E_{i+1}. Notice that there is no need to compute a Cholesky factorization $A_i = Q_i Q_i^\top$ of A_i.

Algorithm 9: Ellipsoid Method

Input : Separation oracle for a polytope $P \subseteq \mathbb{R}^n$ with $\dim(P) = n$, and an ellipsoid $E_0 = E(a_0, A_0)$ with $P \subseteq E_0$.
Output: A point $y \in P$.
$i = 0$.
while $a_i \notin P$ *(checked with separation oracle)* **do**
  Get $c \in \mathbb{R}^n$ such that $P \subseteq \{x \in \mathbb{R}^n : c^\top x < c^\top a_i\}$, using separation oracle.
  Let $b = \frac{A_i c}{\sqrt{c^\top A_i c}}$.
  Let $a_{i+1} = a_i - \frac{1}{n+1} b$.
  Let $A_{i+1} = \frac{n^2}{n^2-1}(A_i - \frac{2}{n+1} bb^\top)$.
  $i = i + 1$.
return a_i.

We recall that the square root that has to be taken for the calculation of b is an operation that we cannot perform exactly under the usual computational assumptions. Therefore, to obtain a working polynomial time version of the Ellipsoid Method, one typically only computes an approximate version of E_{i+1} that still contains $E_i \cap H_i$ but is slightly bigger than the minimum volume ellipsoid containing $E_i \cap H_i$.

6.4.4 From checking feasibility to optimization over $\{0,1\}$-polytopes

Even though the Ellipsoid Method works in a much more general context, we focus from now on on the problem of maximizing a linear function w over a $\{0,1\}$-polytope P. This case is highly relevant in combinatorial optimization, and allows us to show nicely how one can derive that the Ellipsoid Method runs in polynomial time on an interesting class of LPs.

Without loss of generality we can assume $w \in \mathbb{Z}^n$, since any $w \in \mathbb{Q}^n$ can be transformed into an integral vector through scaling. We start by discussing how to obtain the optimal value $\nu^* = \max\{w^\top x \colon x \in P\}$ of the LP in polynomial time, before we provide details of how this procedure can be used to also obtain an optimal vertex solution $x^* \in P$.

Again, we assume that P is full-dimensional.

Getting the optimal LP value ν^*

Let $w_{\max} = \max\{|w_k| \colon k \in [n]\}$. Since $P \subseteq \mathbb{R}^n$ is a $\{0,1\}$-polytope, there is an optimal solution $x^* \in \{0,1\}^n$, and since $\nu^* = w^\top x^*$, we have

$$\nu^* \in [-n w_{\max}, n w_{\max}] \cap \mathbb{Z} \enspace .$$

To determine ν^*, we check the non-emptiness of a series of polytopes of the form

$$P(\nu) = P \cap \left\{ x \in \mathbb{R}^n \colon w^\top x \geq \nu - \frac{1}{2} \right\} \enspace ,$$

where $\nu \in [-n w_{\max}, n w_{\max}] \cap \mathbb{Z}$. Notice that ν^* is the largest $\nu \in [-n w_{\max}, n w_{\max}] \cap \mathbb{Z}$ for which $P(\nu)$ is non-empty. Hence, we can find ν^* by performing a binary search over $\nu \in$

$[-nw_{\max}, nw_{\max}] \cap \mathbb{Z}$ and checking non-emptiness of $P(\nu)$. This takes $O(\log(nw_{\max})) = O(\log n + \log w_{\max})$ iterations, which is polynomial in the input.

Notice that in the definition of $P(\nu)$, we are looking for solutions of value at least $\nu - \frac{1}{2}$, instead of simply looking for solutions of value ν. Adding the $-\frac{1}{2}$ term guarantees that $P(\nu)$ is either empty or full-dimensional. If we drop the $-\frac{1}{2}$ term, then the polytope $P(\nu^*)$ is not full-dimensional anymore.

To show that ν^* can be computed in polynomial time via the Ellipsoid Method we still need to prove that we can check non-emptiness of $P(\nu)$ in polynomial time for any $\nu \in [-nw_{\max}, nw_{\max}] \cap \mathbb{Z}$. For this we have to specify what starting ellipsoid E_0 we choose, and we need to show that the Ellipsoid Method finds a point in $P(\nu)$ in a polynomial number of iterations if $P(\nu) \neq \emptyset$.

Starting ellipsoid E_0 As the starting ellipsoid E_0, we choose the ball centered at $(\frac{1}{2}, \dots, \frac{1}{2})$ of radius $\frac{1}{2}\sqrt{n}$, which is the smallest ball that contains $[0,1]^n$. Hence, $P(\nu) \subseteq [0,1]^n \subseteq E_0$. The volume of E_0 is

$$\mathrm{vol}(E_0) = \frac{1}{2^n}(\sqrt{n})^n \,\mathrm{vol}(E(0,I)) \ .$$

We recall that the volume of the unit ball $E(0,I)$ is given by

$$\mathrm{vol}(E(0,I)) = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)} \leq \pi^{n/2} \leq 2^n \ ,$$

where Γ is the gamma function, which extends factorials to non-integers. It is given by $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$. However, the crude upper bound $\mathrm{vol}(E(0,I)) \leq 2^n$ also immediately follows by observing that $E(0,I) \subseteq [-1,1]^n$, and hence $\mathrm{vol}(E(0,I)) \leq \mathrm{vol}([-1,1]^n) = 2^n$. Thus,

$$\log(\mathrm{vol}(E_0)) = O(n \log n) \ .$$

Bounding the number of iterations To bound the number of iterations of the Ellipsoid Method when applied to $P(\nu)$ for $\nu \in \mathbb{Z}$ we leverage Lemma 6.5, for which we need to show that if $P(\nu) \neq \emptyset$, then its volume is not too small. Hence, assume $P(\nu) \neq \emptyset$, which implies $\exists y_0 \in P(\nu) \cap \{0,1\}^n$. (Actually, any optimal vertex solution to $\max\{w^\top x \colon x \in P\}$ can be chosen as y_0.) Since P is assumed to be a full-dimensional $\{0,1\}$-polytope, there are furthermore points $y_1, \dots, y_n \in P \cap \{0,1\}^n$ such that the simplex $\Delta = \mathrm{conv}(\{y_0, \dots, y_n\})$ is full-dimensional. The points $y_1, \dots, y_n$ are not necessarily contained in $P(\nu)$. Therefore we shrink the simplex Δ towards the vertex y_0, to obtain a scaled-down simplex Δ' which is fully contained in $P(\nu)$. We will then use $\mathrm{vol}(\Delta')$ as a lower bound for $\mathrm{vol}(P(\nu))$. More precisely, we show that it suffices to shrink Δ by a factor of

$$\alpha = \frac{1}{2nw_{\max}} \ .$$

The simplex Δ' has as vertices y_0, and the following points $z_1, \dots, z_n$:

$$z_i = y_0 + \alpha(y_i - y_0) \quad \forall i \in [n] \ .$$

To show $\Delta' = \operatorname{conv}(\{y_0, z_1, \ldots, z_n\}) \subseteq P(\nu)$, we have to show that $z_i \in P(\nu)$ for $i \in [n]$, which reduces to checking $w^\top z_i \geq \nu - \frac{1}{2}$ because $z_i \in P$. This indeed holds:

$$\begin{aligned} w^\top z_i &= \underbrace{w^\top y_0}_{\geq \nu} + \alpha w^\top (y_i - y_0) \geq \nu + \alpha w^\top (y_i - y_0) \quad (\text{since } w^\top y_0 \geq \nu - 1/2 \text{ and } w^\top y_0 \in \mathbb{Z}) \\ &\geq \nu - \alpha n w_{\max} = \nu - \frac{1}{2} \ . \end{aligned}$$

Hence, $\Delta' \subseteq P(\nu)$. Furthermore, since Δ' is a scaled-down version of Δ, with scaling factor α, we obtain

$$\operatorname{vol}(\Delta') = \alpha^n \operatorname{vol}(\Delta) \ .$$

Recall that the volume of a simplex in $\mathbb{R}^n$ is $\frac{1}{n!}$ times the volume of the corresponding parallelepiped. In particular, we obtain

$$\operatorname{vol}(\Delta) = \frac{1}{n!} \cdot |\det(y_1 - y_0, \ldots, y_n - y_0)| \geq \frac{1}{n!} \ ,$$

since the matrix with columns $y_i - y_0$ is integral and non-singular because Δ is full-dimensional. Hence,

$$\operatorname{vol}(\Delta') = \alpha^n \operatorname{vol}(\Delta) \geq \left(\frac{1}{2nw_{\max}}\right)^n \frac{1}{n!} \ ,$$

and therefore

$$-\log(\operatorname{vol}(P(\nu))) \leq -\log(\operatorname{vol}(\Delta')) = O(n \log n + n \log w_{\max}) \ .$$

Thus, by Lemma 6.5 the number of iterations needed by the Ellipsoid Method to find a point in $P(\nu)$ is at most

$$\begin{aligned} 2(n+1) \ln\left(\frac{\operatorname{vol}(E_0)}{\operatorname{vol}(P(\nu))}\right) &= O\big(n \cdot [\log(\operatorname{vol}(E_0)) - \log(\operatorname{vol}(P(\nu)))]\big) \\ &= O(n \cdot (n \log n + n \log w_{\max})) \\ &= O(n^2 \cdot (\log n + \log w_{\max})) \ , \end{aligned}$$

which is polynomial.

Consequently, to check non-emptiness of $P(\nu)$ for an arbitrary $\nu \in \mathbb{Z}$, it suffices to run the Ellipsoid Method for $O(n^2 \cdot (\log n + \log w_{\max}))$ iterations. It will either find a point in $P(\nu)$, which obviously shows non-emptiness of $P(\nu)$, or if no point in $P(\nu)$ is found over this many iterations, then $P(\nu) = \emptyset$.

Determining an optimal $\{0,1\}$-solution x^*

The above discussion shows that we can find with the Ellipsoid Method the optimal value ν^* of any LP over a full-dimensional $\{0,1\}$-polytope P. Furthermore, we also get a point $y \in P$ with $w^\top y \geq \nu^* - \frac{1}{2}$. However, especially in Combinatorial Optimization settings we are interested in getting an optimal vertex solution $x^* \in P \cap \{0,1\}^n$, which typically corresponds to some underlying discrete structure we are interested in. There are several ways how this can be done.

One procedure that often works very efficiently is to start with y and do local operations to obtain a vertex x with $w^\top x \geq w^\top y$, which implies that x is an optimal vertex solution. We do not go into details of this procedure, and present a conceptually simpler, though often slower, approach that works in our setting and is efficient.

The idea of this approach is to reduce the problem of finding an optimal vertex solution to a sequence of problems that ask to find only the optimal value of an LP. Let $S \subseteq [n]$. Consider a modified objective function w^S defined by

$$w_i^S = \begin{cases} w_i + 1 & \text{if } i \in S \ , \\ w_i & \text{if } i \in [n] \setminus S \ . \end{cases}$$

Optimizing with respect to w^S allows for obtaining insights into an optimal solution using the following result.

Lemma 6.8

Let $S \subseteq [n]$. The following two statements are equivalent.

(i) There is an optimal solution x^* to $\max\{w^\top x \colon x \in P\}$ with $x_i^* = 1$ for $i \in S$.
(ii) $\nu^* + |S| = \max\{(w^S)^\top x \colon x \in P\}$.

Proof. By definition of w^S, we have for any point $y \in \{0,1\}^n$,

$$(w^S)^\top y = w^\top y + |\{i \in S \colon y_i = 1\}| \ .$$

Clearly, for any vertex $y \in P$, we have $w^\top y \leq \nu^*$ with equality if and only if y^* is a maximizer of $\max\{w^\top x \colon x \in P\}$. Furthermore, we also have $|\{i \in S \colon y_i = 1\}| \leq |S|$. This shows that $\max\{(w^S)^\top x \colon x \in P\} \leq \nu^* + |S|$, with equality if and only if there is a vertex $y \in P$ that maximizes $\max\{w^\top x \colon x \in P\}$ and satisfies $y_i = 1$ for $i \in S$, thus proving the lemma. $\square$

Based on Lemma 6.8, we can construct an optimal vertex solution coordinate by coordinate. More precisely, we will determine among all optimal vertex solutions of $\max\{w^\top x \colon x \in P\}$ the *lexicographically largest* one $x^* = (x_1^*, \dots, x_n^*)$, i.e., for any other optimal vertex solution y, there exists $k = k(y) \in \{0, \dots, n-1\}$ such that $y_i = x_i^*$ for $i \in [k]$, and $x_{k+1}^* = 1$ whereas $y_{k+1} = 0$.

We construct x^* iteratively. In the ith iteration we determine the value of x_i^*. Assume that we know the values of $x_1^*, \dots, x_{i-1}^*$ for some $i \in [n]$. Let $U = \{j \in [i-1] \colon x_j^* = 1\}$, and define $S = U \cup \{i\}$. Using Lemma 6.8 we can decide whether there is an optimal solution z^* to $\max\{w^\top x \colon x \in P\}$ with $z_j^* = 1$ for $j \in S$. If this is the case, then $x_i^* = 1$, otherwise, $x_i^* = 0$. Hence, after n iterations we obtain x^*.

In summary, we obtain the following theorem.

Theorem 6.9

Let $P \subseteq \mathbb{R}^n$ be a full-dimensional $\{0,1\}$-polytope for which we are given a separation oracle. Furthermore, let $w \in \mathbb{Z}^n$. Then the Ellipsoid Method allows for finding an optimal vertex solution to the linear program $\max\{w^\top x \colon x \in P\}$ using a polynomial number of operations and calls to the separation oracle for P.

6.5 Comments on the non-full-dimensional case

As stated in Theorem 6.2 we can also apply the Ellipsoid Method to optimize a linear function over a $\{0,1\}$-polytope P that is not full-dimensional. We only give a rough outline how this generalization can be achieved, without providing full details.

To simplify the explanation, assume that P is $n-1$ dimensional. Hence, there is a unique hyperplane $H = \{x \in \mathbb{R}^n \colon g^\top x = b\}$ such that $P \subseteq H$.

The ellipsoids constructed during the Ellipsoid Method become more and more flat over the iterations, because all of them contain P and their volumes shrink. After a well-chosen number of iterations $k = O(\mathrm{poly}(n))$ of the Ellipsoid Method, we consider the shortest axis of the current ellipsoid $E_k = E(a_k, A_k)$, which we denote by $d \in \mathbb{R}^n$. Hence, d is the eigenvector of A_k that corresponds to the smallest eigenvalue. Consider the hyperplane $F = \{x \in \mathbb{R}^n \colon d^\top x = d^\top a_k\}$. Intuitively, we expect F to be "close" to H. Indeed, if k is chosen well, one can round the coefficients of the hyperplane F to obtain the hyperplane H. However, this rounding procedure is not straightforward, and is based on an approximation algorithm to solve Simultaneous Diophantine Approximations. Once H is obtained, one can reduce the problem to the affine subspace defined by H by eliminating one of the coordinates using the equation provided by H. We therefore consider P to be a polytope in the affine subspace defined by H. Since P is $(n-1)$-dimensional in its original description in $\mathbb{R}^n$, we thereby obtain a full-dimensional polytope when removing one variable of P. A similar argument as above, applied repeatedly, works when P has dimension that is smaller than $n-1$.

Formalizing these arguments leads to a proof of Theorem 6.2.