6 The Complexity of Linear Programming

6.1 A Lower Bound for the Simplex Method

The complexity of the simplex method depends on two factors, the number of steps in each round and the number of iterations. It is not hard to see that the tableau form requires $O(mn)$ arithmetic operations in each round. We will now describe an instance of the linear programming problem, and a specific pivot rule, such that the simplex method requires an exponential number of iterations to find the optimal solution. For this, we construct a polytope with an exponential number of vertices, and a so-called *spanning path* that traverses all of the vertices, in such a way that consecutive vertices are adjacent and a certain linear objective strictly increases along the path. This shows that the simplex method requires an exponential number of iterations in the worst case, for the specific pivoting rule that follows the spanning path.

Consider the unit cube in $\mathbb{R}^n$, given by the constraints

$$0 \leqslant x_i \leqslant 1 \quad \text{for } i = 1, \ldots, n.$$

The unit cube has 2^n vertices, because either one of the two constraints $0 \leqslant x_i$ and $x_i \leqslant 1$ can be active for each dimension i. Further consider a spanning path of the unit cube constructed inductively as follows. In dimension 1, the path moves from $x_1 = 0$ to $x_1 = 1$. In dimension k, the path starts with $x_k = 0$ and traverses the spanning path for dimensions x_1 to x_{k-1}, which exists by the induction hypothesis. It then moves to the adjacent vertex with $x_1 = 1$, and traverses the spanning path for dimensions x_1 to x_{k-1} in the reverse direction. This construction is illustrated of the left of Figure 6.1.

Now assume that we are trying to minimize the objective $-x_n$, and observe that so far it decreases only once, namely in the middle of the path. This can easily be fixed. Let $\epsilon \in (0, 1/2)$, and consider the perturbed unit cube with constraints

$$\begin{aligned} \epsilon \leqslant x_1 &\leqslant 1, \\ \epsilon x_{i-1} \leqslant x_i &\leqslant 1 - \epsilon x_{i-1} \quad \text{for } i = 2, \ldots, n \end{aligned} \tag{6.1}$$

An example is shown on the right of Figure 6.1. It is easily verified that x_n now increases strictly along the path described above. We obtain the following result.

THEOREM 6.1. *Consider the linear programming problem of minimizing $-x_n$ subject to* (6.1). *Then there exists a pivoting rule and an initial basic feasible solution such that the simplex method requires $2^n - 1$ iterations before it terminates.*

Observe that each of the numbers in the description of the perturbed unit cube can be represented using $O(\log \epsilon^{-n}) = O(n)$ bits, the number of iterations is therefore also exponential in the input size.

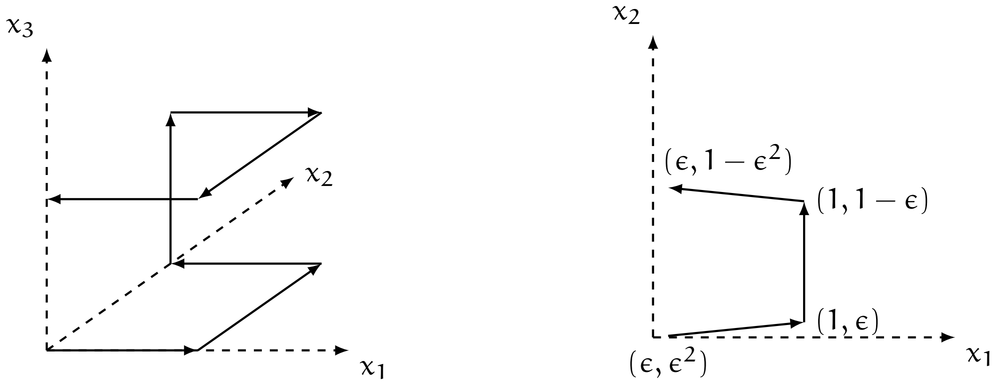

Figure 6.1: Spanning paths of the three-dimensional unit cube (left) and of the perturbed two-dimensional unit cube with $\epsilon = 1/10$ (right)

Interestingly, the first and last vertices of the spanning paths constructed above are adjacent, which means that a different pivoting rule could reach the optimal solution in a single step. However, similar worst-case instances have been constructed for many other pivot rules, and no pivot rule is known to guarantee a polynomial worst-case running time. The *diameter* of a polytope, i.e., the maximum number of steps necessary to get from any vertex to any other vertex, provides a lower bound of the number of iterations of the simplex method that is independent of the pivoting rule. The Hirsch conjecture, which states that the diameter of a polytope in dimension d with n facets cannot be greater than $n-d$, was disproved in 2010. Whether the diameter is bounded by a polynomial function of n and d remains open.

In practice, the performance of the simplex method is often much better, usually linear in the number of constraints. However, it is not clear how the intuition of a good average-case performance could be formalized, because this would require a natural probability distribution over instances of the linear programing problem. This is a problem that applies more generally to the average-case analysis of algorithms.

6.2 The Idea for a New Method

Again consider the linear program (2.2) and its corresponding dual:

$$\min\{\, c^T x : Ax = b, x \geqslant 0 \,\}$$
$$\max\{\, b^T \lambda : A^T \lambda \leqslant c \,\}.$$

By strong duality, each of these problems has a bounded optimal solution if and only if the following set of linear constraints is feasible:

$$c^T x = b^T \lambda, \quad Ax = b, \quad x \geqslant 0, \quad A^T \lambda \leqslant c.$$

We can thus concentrate on the following decision problem: given a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$, is the set $\{x \in \mathbb{R}^n : Ax \geqslant b\}$ non-empty? We will now consider a method for solving this problem, known as the *ellipsoid method*.

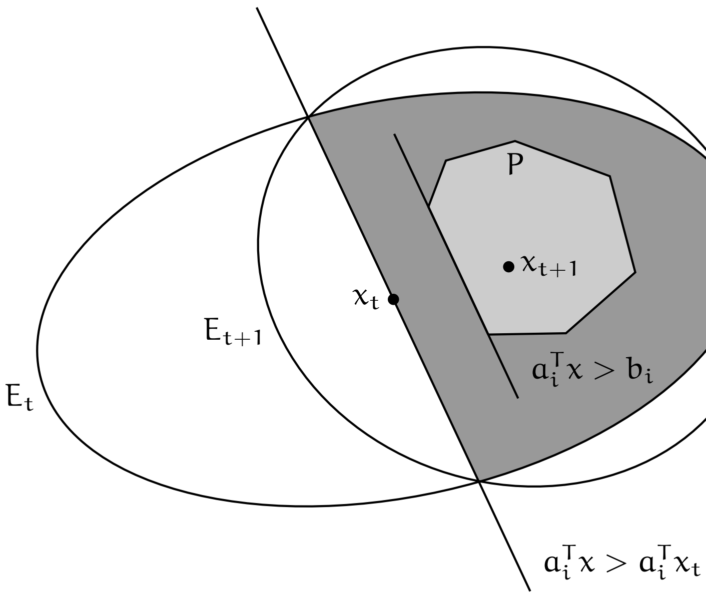

Figure 6.2: A step of the ellipsoid method where $x_t \notin P$ but $x_{t+1} \in P$. The polytope P and the half-ellipsoid that contains it are shaded.

We need some definitions. A symmetric matrix $D \in \mathbb{R}^{n\times n}$ is called *positive definite* if $x^T D x > 0$ for all non-zero $x \in \mathbb{R}^n$. A set of vectors $E \subseteq \mathbb{R}^n$ given by

$$E = E(z, D) = \{\, x \in \mathbb{R}^n : (x - z)^T D^{-1}(x - z) \leqslant 1 \,\}$$

for a positive definite symmetric matrix $D \in \mathbb{R}^{n\times n}$ and a vector $z \in \mathbb{R}^n$ is called an *ellipsoid* with center z. If $D \in \mathbb{R}^{n\times n}$ is non-singular and $b \in \mathbb{R}^n$, then the mapping $S : \mathbb{R}^n \to \mathbb{R}^n$ given by $S(x) = Dx + b$ is called an *affine transformation*. We further write $S(L)$ for the *image* of $L \subseteq \mathbb{R}^n$ under S, i.e., $S(L) = \{y \in \mathbb{R}^n : y = S(x)$ for some $x \in \mathbb{R}^n\}$. The *volume* of a set $L \subseteq \mathbb{R}^n$ if finally defined as $\mathrm{Vol}(L) = \int_{x\in L} dx$.

Let $P = \{x \in \mathbb{R}^n : Ax \geqslant b\}$ for some $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^m$. To decide whether P is non-empty, the ellipsoid method generates a sequence $\{E_t\}$ of ellipsoids E_t with centers x_t. If $x_t \in P$, then P is non-empty and the method stops. If $x_t \notin P$, then one of the constraints is violated, i.e., there exists a row j of A such that $a_j^T x_t < b_j$. Therefore, P is contained in the half-space $\{x \in \mathbb{R}^n : a_j^T x \geqslant a_j^T x_t\}$, and in particular in the intersection of this half-space with E_t, which we will call a *half-ellipsoid*.

The following is the key result underlying the ellipsoid method. It states that there exists a new ellipsoid E_{t+1} that contains the half-ellipsoid and whose volume is only a fraction of the volume of E_t. This situation is illustrated in Figure 6.2.

Theorem 6.2. *Let* $E = E(z, D)$ *be an ellipsoid in* $\mathbb{R}^n$ *and* $a \in \mathbb{R}^n$ *non-zero. Consider the half-space* $H = \{x \in \mathbb{R}^n : a^T x \geqslant a^T z\}$, *and let*

$$z' = z + \frac{1}{n+1}\frac{Da}{\sqrt{a^T D a}},$$
$$D' = \frac{n^2}{n^2 - 1}\left(D - \frac{2}{n+1}\frac{Daa^T D}{a^T D a}\right).$$

Then D' *is symmetric and positive definite, and therefore* $E' = E(z', D')$ *is an ellipsoid. Moreover,* $E \cap H \subseteq E'$ *and* $\mathrm{Vol}(E') < e^{-1/(2(n+1))}\mathrm{Vol}(E)$.

If the procedure is repeated, it will either find a point in P or generate smaller and smaller ellipsoids containing P. In the next lecture, this procedure will be turned into an algorithm by observing that the volume of P must either be zero or larger than a certain threshold that depends on the size of the description of P.

We now sketch the proof of Theorem 6.2. We use the following lemma about affine transformations, which is not hard to prove.

LEMMA 6.3. *Let* $S : \mathbb{R}^n \to \mathbb{R}^n$ *be an affine transformation given by* $S(x) = Dx + b$ *and let* $L \subseteq \mathbb{R}^n$. *Then,* $\mathrm{Vol}(S(L)) = |\det(D)|\mathrm{Vol}(L)$.

Proof sketch of Theorem 6.2. We prove the theorem for $E = \{x \in \mathbb{R}^n : x^T x \leqslant 1\}$ and $H = \{x \in \mathbb{R}^n : x_1 \geqslant 0\}$. Since every pair of an ellipsoid and a hyperplane as in the statement of the theorem can be obtained from E and H via some affine transformation, the general case then follows by observing that affine transformations preserve inclusion and, by Lemma 6.3, relative volume of sets.

Let $e_1 = (1, 0, \ldots, 0)^T$. Then,

$$\begin{aligned} E' &= E\left(\frac{e_1}{n+1}, \frac{n^2}{n^2-1}\left(I - \frac{2}{n+1}e_1e_1^T\right)\right) \\ &= \left\{x \in \mathbb{R}^n : \frac{n^2-1}{n^2}\sum_{i=1}^{n} x_i^2 + \frac{1}{n^2} + \frac{2(n+1)}{n^2}x_1(x_1-1) \leqslant 1\right\}. \end{aligned}$$

Consider an arbitrary $x \in E \cap H$, and observe that $0 \leqslant x_1 \leqslant 1$ and $\sum_{i=1}^{n} x_i^2 \leqslant 1$. It is easily verified that $x \in E'$ and thus $E \cap H \subseteq E'$.

Now consider the affine transformation $F : \mathbb{R}^n \to \mathbb{R}^n$ given by

$$F(x) = \frac{e_1}{n+1} + \left(\frac{n^2}{n^2-1}\left(I - \frac{2}{n+1}e_1e_1^T\right)\right)^{\frac{1}{2}} x.$$

It is not hard to show that $E' = F(E)$. Therefore, by Lemma 6.3,

$$\begin{aligned} \frac{\mathrm{Vol}(E')}{\mathrm{Vol}(E)} &= \sqrt{\det\left(\frac{n^2}{n^2-1}\left(I - \frac{2}{n+1}e_1e_1^T\right)\right)} \\ &= \left(\frac{n^2}{n^2-1}\right)^{\frac{n}{2}}\left(1 - \frac{2}{n+1}\right)^{\frac{1}{2}} = \frac{n}{n+1}\left(\frac{n^2}{n^2-1}\right)^{\frac{n-1}{2}} \\ &= \left(1 - \frac{1}{n+1}\right)\left(1 + \frac{1}{n^2-1}\right)^{\frac{n-1}{2}} < e^{-\frac{1}{n+1}}\left(e^{\frac{1}{n^2-1}}\right)^{\frac{n-1}{2}} = e^{-\frac{1}{2(n+1)}}, \end{aligned}$$

where the strict inequality follows by using twice that $1 + a < e^a$ for all $a \neq 0$. □

A more detailed description of the ellipsoid method and an overview of the proof of correctness will be given in the next lecture.