Optimization 1

Lecture notes, University of Technology, Graz
based on the lecture by Bettina Klinz

Lukas Prokop

April 1, 2019

Contents

0 Course

↓ *This lecture took place on 2019/03/04.*

- Lecture
  - Monday, 12:15–14:00
  - Tuesday, 16:15–18:00
- First week, the practicals session will be used for the lecture
- Practicals will take place usually on Wednesday, 16:15–18:00
  in exceptional cases on Thursday, 16:15–18:00
- 2 websites (work in progress):
  - http://www.math.tugraz.at/~klinz/optimvo (list of literature)
  - http://www.math.tugraz.at/~klinz/optimue (practicals mode, practicals exercises, additional content)
- Two large topics in this lecture
  - Linear optimization (linear target function, linear side conditions)
  - Non-linear optimization without sub conditions (unconstrained non-linear optimization)
    where non-linear optimization denotes that the target function is non-linear
- Be aware, that this class might be the lecture requiring previous results of classes the most.
- Advanced lecture in masters
  - Lecture "Non-linear optimization" (includes nonlinear optimization with sub conditions)
- Exam
  - written + orally, in case of negotiation and few candidates only orally

- 1st date will be at the end of the semester, optionally in summer holidays
- 2 written exams for the practicals

0.1 Linear optimization

We have already seen optimization problems in high school or previous semesters. But, for example, handling constraints consisting of inequalities was tedious or trivial. We consider more sophisticated techiques here. In practice, linear models occur rarely. But they often provide a sufficient heuristic.

0.2 Introduction and some examples for linear optimization models

Example 1 (Production planning model)**.** *A factory can produce n goods. The revenue per unit of good j is given by c_j units of money. Production is limited by restrictions, that result from constrained availability of staff, equipment and raw materials. Let m denote the number of these resources and b_i is the maximum availability of resource i ($i = 1, \dots, m$). Let a_{ij} with $1 \leq i \leq m$ and $1 \leq j \leq n$ denote the quantity of resource i required to produce 1 unit of good j. Our goal is to maximize revenue with respect to the given constraints.*

Decision variable: X_j *is the quantity of good j*

$$\textit{target function: } \max \sum_{j=1}^{n} C_j X_j$$

$$\textit{subject to (constraints)} \sum_{j=1}^{n} a_{ij} x_j \leq b_i \qquad i \in \{1, \dots, m\}$$

$$\textit{with } x_j \geq 0 \qquad \textit{sign condition}$$

Pay attention! We do not require $x_j \in \mathbb{Z}$. For $x_j \in \mathbb{Z}$ we get an integral linear program which is not *a linear programm! (this is a difficult subproblem of optimization)*

Example 2 (Mixture problem)**.** *Consider n kinds of raw materials. Our goal: We want to create a new material by mixing existing raw materials to reduce costs.*

Example (Alloys)**.** *Consider n different base alloys $L_1, \dots, L_n$. For each material we have the lead content in percent $(a_1, \dots, a_n)$ and the costs per unit of weight $(c_1, \dots, c_n)$. We have to produce alloys with lead content b%. The decision variable X_j is given by the ratio of L_j.*

Formally, the problem can be defined as

$$\min \sum_{j=1}^{n} c_j x_j \;\; s.t. \;\; \sum_{j=1}^{n} x_j = 1, \sum_{j=1}^{n} a_j x_j = b, x_j \geq 0$$

These are typical mixture constraints and they ensure a given lead content.

Example 3 (Linear transportation problem)**.** *Given m firms, n customers, a_i is the offer by firm i and b_j is the demand by customer j. c_{ij} are the transportation costs from firm i to customer j.*

Find an admissible transportation plan with minimal costs. The decision variable X_{ij} is the quantity of goods transported from firm i to customer j.

Formally,

$$\min \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} X_{ij} \;\; s.t. \;\; \sum_{i=1}^{m} X_{ij} = b_j, \sum_{j=1}^{n} X_{ij} = a_i, X_{ij} \geq 0 \quad j \in \{1, \ldots, n\}, i \in \{1, \ldots, m\}$$

Example (Diet problem)**.** *The following problem has a strong historical background in optimization sciences: Stigler diet problem (in the year 1939) by Georg Stigler.*

Given a list of 77 ingredients. Per ingredient we are given features such as calories and proteins. Find an optimal combination to minimize costs.

His heuristic results were confirmed as almost optimal in 1947.

In the following lectures, our goal will be:

- Theory of linear optimization (Knowledge about fundamentals and background)
- Algorithmic solutions procedures: in the lecture we will discuss two procedures:
  - Simplex process (G. Dantzig, 1947, in practice useful, no polynomial runtime)
  - Inner point method (in practice useful, polynomial runtime)

  Outside this lecture:

  - Ellipsoid method: polynomial runtime, but in practice useless

1 Geometrical considerations of linear optimization

Definition. *Standard form of a linear program (canonical representation)*

$$\max z(x) = z_0 + \sum_{j=1}^{n} c_j x_j \text{ such that } \sum_{j=1}^{n} a_{ij} x_j \leq b_i, X_j \geq 0 \qquad i \in \{1, \dots, m\}$$

In compact notation:

$$\max z_0 + c^t x \text{ s.t. } Ax \leq b, x \geq 0$$

typically $\max c^t x$. z_0 *does not influence the result.*

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \qquad b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \qquad a_{ij} \in \mathbb{R}, b_i \in \mathbb{R}, c_j \in \mathbb{R}$$

↓ *This lecture took place on 2019/03/05.*

Revision. *The canonical/standard form of a linear program is given by*

$$\max c^t x (+ \text{ optionally } z_0) \text{ such that } Ax \leq b \qquad x \geq 0$$

Remark (Observation)**.** *Every arbitrary linear program (min/max of an affine linear function over a linear constraint) can be transformed into the canonical form above.*

1. *The minimization over* $c^t x$ *corresponds to* $-c^t x$ *as maximization problem.*
2. *Constraints of form* $\alpha^t x \geq b$ *can be written as* $-\alpha^t x \leq -\beta$.
3. *A constraint of form* $\alpha^t x = \beta$ *can be written as* $\alpha^t x \leq \beta$ *with* $\alpha^t x \geq \beta$. *Or equivalently as* $\alpha^t x \leq \beta$ *with* $-\alpha^t x \leq -\beta$.

   Remark.

   - *Disadvantage: One constraint is transformed into two.*
   - *In practice, the explicit handling of equality constraints should be preferred.*

4. *Let* x_j *not be restricted w.r.t. the sign. Write* x_j *as* $x_j = x_j^+ - x_j^-$ *with* $x_j^+ \geq 0$ *and* $x_j^- \geq 0$. x_j *will be replaced by* $x_j^+ - x_j^-$ *with* $x_j^+ \geq 0$ *and* $x_j^- \geq 0$.

Remark. *Disadvantage: Number of variables in increased.*

Remark (Terminology)**.**

- *2 points* $u, v \in \mathbb{R}^n$ *define a* line $G(u, v) = \{u + \lambda(v - u) \mid \lambda \in \mathbb{R}\}$.
- *A* halfline *is formally given by* $\{u + \lambda(v - u) \mid \lambda \geq 0\}$
- *A* closed interval *(or* segment*) is defined as* $[u, v] = \{u + \lambda(v - u) \mid \lambda \in [0, 1]\}$ *and an* open interval *is defined as* $(u, v) = \{u + \lambda(v - u) \mid \lambda \in (0, 1)\}$.

Lemma 1.1.
- *An affine linear function* $z_0 + c^t x$ *takes up its maximum/minimum in segment* $[u, v]$ *in its end points* u *or* v.
- *An affine linear function* $z_0 + c^t x$ *takes up its maximum/minimum on a half-line in the end points.*

Proof. Left as an exercise to the reader (use parameter representation and insert it into the function) □

1.1 Hyperplane and Halfspaces

By the linear inequality $\alpha^t x \leq \beta$ ($\alpha \in \mathbb{R}^n, \beta \in \mathbb{R}$) with $\alpha \neq 0$ (zero vector) we define a *closed halfspace.*

$$H_{\leq} := \left\{x \in \mathbb{R}^n \mid \alpha^t x \leq \beta\right\}$$

Analogously we can define open halfspaces:

$$H_{<} := \left\{x \in \mathbb{R}^n \mid \alpha^t x < \beta\right\}$$

Hyperplane:

$$H_{=} := \left\{x \in \mathbb{R}^n \mid \alpha^t x = \beta\right\}$$

In $\mathbb{R}^2$, hyperplanes are halflines. In $\mathbb{R}^3$, hyperplanes are halfplanes.

Possible cases for the position of lines $G(u, v)$ w.r.t. to the hyperplane $H = \{x \mid \alpha^t x = \beta\}$.

Case 1 Line G is contained in H (denoted $G \subseteq H$)
$\alpha^t u = \beta$, $\alpha^t(u - v) = 0$

Case 2 G is parallel to H
$\alpha^t u \neq \beta$, $\alpha^t(u - v) = 0$

Case 3 G intersects H in one point c
$\alpha^t(u - v) \neq 0$

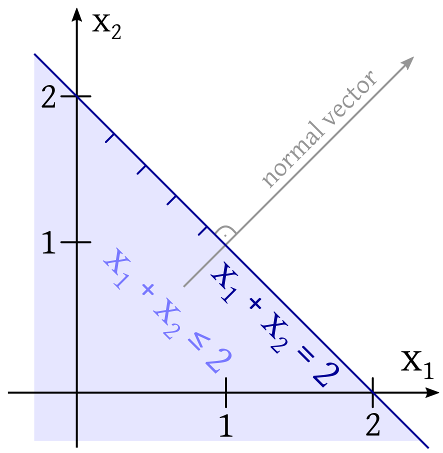

Figure 1: Constraint $x_1 + x_2 \leq 2$ visualized in $\mathbb{R}^2$. To determine the halfplane (bright blue) resulting from the constraint, it helps to express the constraint with x_2 on the left-hand side: $x_2 \leq 2 - x_1$. Then determine x_2 for two different x_1 assuming an equality operator, e.g. $x_1 = 0$ with $x_2 = 2$ and $x_1 = 1$ with $x_2 = 1$. Then choose some large value x_1 and x_2 (e.g. $x_1 = 10$ and $x_2 = 10$). Does it satisfy $x_1 + x_2 \leq 2$? No, the halfplane containing $(10, 10)$ is not the one, we are looking for. Some people prefer to consider the normal vector. Sometimes dashes are used to mark the side of the considered halfplane.

Lemma 1.2. *If the line $G(u, v)$ is neither contained in halfplane $H = \{x \mid \alpha^t x = \beta\}$ nor in some open halfspace constrained by H, G intersects the halfplane H in one point.*

Remark (Observation)**.** *The parameter representation of a line is ambiguous. We can choose u und v wisely: $G(u, v), \overline{x} \in G$. We can always choose u and v such that $\overline{x} \in (u, v)$.*

Let $\overline{x} \in H_< = \{x \in \mathbb{R}^n \mid \alpha^t x < \beta\}$. If $G \subseteq H_<$, then $u, v \in H_<$. Otherwise, due to $\overline{x} \in G \cap H_<$, G intersects the hyperplane $H = \{x \in \mathbb{R}^n \mid \alpha^t x = \beta\}$.

Furthermore we can choose a representation (u, v) such that $\overline{x} \in (u, v) \subseteq H_<$ and $v \in H_=$. This can be generalized to multiple halfspaces.

Definition. *A* polyhedron *(dt. "Polyeder") is the intersection of finitely many halfspaces. A bounded polyhedron is also called* polytope.

Remark (Observation)**.** *The admissible set $P(A, b)$ of the linear program (given in the previous revision) is a polyhedron.*

Lemma 1.3. *Halfspaces and thus also polyhedrons are convex sets.*

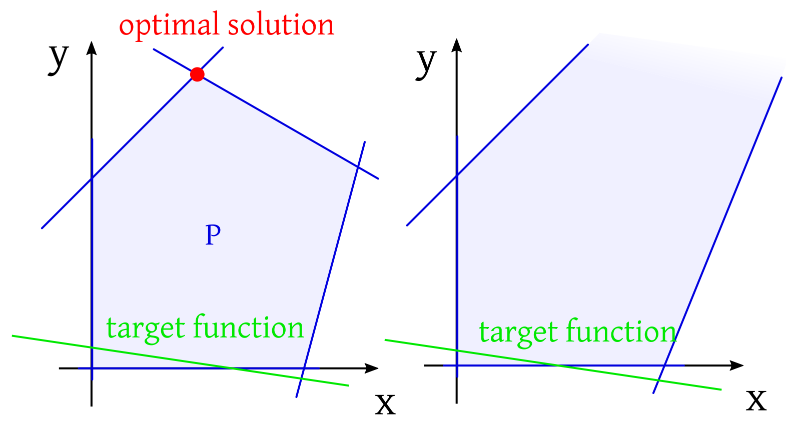

Figure 2: Bounded (left) and unbounded (right) sets. The unbounded set does not have an optimal solution.

Furthermore affine-linear functions are convex *and* concave. Linear optimization is a special case of convex optimization minimizing a convex function over a convex set. The neat property of such convex optimization tasks is that local and global minima/maxima collapse (just like in the linear case).

Geometric illustration for $n = 2$

$n = 2$ means that we consider 2 variables.

$$\max c_1 x_1 + c_2 x_2 \qquad a_{i1} x_1 + a_{i2} x_2 \leq b_i \quad i = 1, \ldots, m \qquad x_1, x_2 \geq 0$$

Remark (Observation)**.** *Optimal solutions occur at vertices of the set (intersection of two constraints). Compare with Figure 2.*

This approach provides a graphical solution method for linear programs with 2 variables.

The generic geometric consideration of a polyhedron is

$$P = \{x \in \mathbb{R}^n \mid Ax \leq b\} \qquad A \in \mathbb{R}^{m \times n}$$

Without loss of generality, we assume all row vectors a_i of A are non-zero.

$$H_i := \left\{x \in \mathbb{R}^n \mid a_i^t x = b_i\right\} \qquad i\text{-th halfplane} \quad i = 1, \ldots, m$$

Let $I \subseteq \{1, \ldots, m\}$. Consider

$$\hat{H}_I := \bigcap_{i \in I} H_i = \left\{x \in \mathbb{R}^n \mid a_i^t x = b_i \text{ for } i \in I\right\} \qquad \text{affine subspace of } \mathbb{R}^n$$

If $\hat{H}_I \cap P \neq P \neq \emptyset$, then this set is called *face* of P. The face of P is called minimal, if it does not contain any other face properly.

Example. *A cube in $\mathbb{R}^3$ has 27 faces:*

- *The cube itself, $I = \emptyset$*
- *6 faces, $I = \{1\}, I = \{2\}, \dots, I = \{6\}$*
- *12 face edges, $|I| = 2$*
- *8 vertices, $|I| = 3$*

Example (Faces in the left subfigure in Figure 2)**.** *11 faces in total (5 vertices of dimension 0, 5 with dimension 1, 1 with dimension 2)*

More formally: Let P be described by a hyperplane H_i ($a_i^t x = b_i$). Let $S \subseteq \mathbb{R}^n, S \neq 0$.

Definition. $I(S) \coloneqq \{i \in \{1, \dots, m\} \mid S \subseteq H_i\}$

For $S = \{x_0\}$, we write $I(x_0)$ instead of $I(\{x_0\})$.

Definition.

$$L(S) \coloneqq \left\{x \mid a_i^t x = b_i, i \in I(S)\right\}$$

is the smallest affine subspace containing S. For a non-empty S, S is a face iff $S = L(S) \cap P$.

If S is a minimal face, then $S = L(S)$ ($L(S)$ is then a part of the polyhedron).

Definition (Dimension of a face)**.**

$$\dim S \coloneqq \dim L(S)$$

so the dimension of the smallest affine subspace containing S.

Definition (Vertex)**.** *A vertex is a face of dimension* 0.

Remark. *For polyhedron, the vertex term from above is an alternative definition for the vertex term for convex sets (the terms correspond).*

Remark. *A circle has infinitely many vertices; not none.*

Let S be convex set in $\mathbb{R}^n$. x is called vertex of S if it is not possible to represent x as $x = y + (1 - \lambda)z$ with $y, z \in S, y \neq z, \lambda \in [0, 1]$.

↓ *This lecture took place on 2019/03/06.*

Not every polyhedron has vertices. For example, if the boundaries are given as two parallels, then no vertices can be identified.

Remark. *The problem with two parallels is not representable in canonical form.*

Definition. *A polyhedron* with *vertices is called* acute.

A vertex is a minimal face. A face results as unique solution of the corresponding equation system.

$$a_i^t x = b_i \qquad \forall i \in I(z)$$

If face $S = L(\overline{x}) \cap P$ for $\overline{x} \in \overline{P(A,b)} =: P$ has dimension ≥ 1 (hence no vertex), then you can let some line G pass through $\overline{x}$ that lies in $L(\overline{x})$.

$\overline{x}$ lies in the intersection of open halfspaces $\left\{x \mid a_i^t x < b_i, i \notin I(\overline{x})\right\}$. Thus we can provide a representation of line G passing through two points c and d with $c, d \in S, \overline{x} \in (c, d)$.

Assume the halfspace with end point $\overline{x}$ in direction d is bounded by some of the constraining hyperplanes of these open halfspaces. Then d can be chosen as the closest intersection point of the line with one of these hyperplanes H_i for $i \notin I(\overline{x})$.

$$\Longrightarrow \left|I(d)\right| > \left|I(\overline{x})\right| \qquad \Longrightarrow \dim L(d) < \dim L(\overline{x})$$

Analogously, the same applies to the constraint by the halfline with end point $\overline{x}$ and in direction c. $\dim L(c) < \dim L(\overline{x})$. Step by step, we can reduce the dimension to end up with a vertex.

Theorem 1.4 (Statement about acute-angled polyhedrons)**.** *1. A non-empty polyhedron is acute iff it does not contain any line.*

2. Every face of an acute polyhedron contains one vertex.

3. A polyhedron has (at most) finitely many vertices.

Proof. 1a. Assume P does not contain any lines. Let $x_0 \in P$ ($P \neq 0$). If x_0 is a vertex, then P is acute. If x_0 is not a vertex, then consider the face $S := L(x_0) \cap P$. Then $\dim S \geq 1$ holds true. Now we use the idea, that was sketched above right before the theorem.

We choose a line $G(c, d) \subseteq L(x_0)$ where c and d are chosen as described before. Because P does not contain a line, S also does not contain any. Without loss of generality, we assume that the halfline with end point x_0 in direction d is bounded and thus $\dim L(d) < \dim L(x_0)$. If d is not a vertex, repeat this construction with some new $x_0 = d$.

In every step, we are losing at least one dimension. After at most n steps, we are going to have a vertex.

1b. Let P be acute, then P has a vertex. Choose a vertex $\overline{x}$ and n inequalities with maximum row rank. $I \subseteq I(\overline{x})$, submatrix A_I of A has $\text{rank}(A_I) = n$.

Assume there exists a line $G(u,v)$ with $u \neq v$, that in entirely contained in P. The inequalities $A_I u + \lambda \cdot A_I(v-u) \leq b_I$ must be true for all $\lambda \in \mathbb{R}$.

Because λ is unbounded, it should be true that $A_I(v-u) = 0$ where A_I is a matrix of full rank. So $v - u = 0 \implies v = u$. This is a contradiction.

2. A face S of an acute polyhedron P cannot contain any line because $S \subseteq P$ und S is itself a polyhedron. By the first statement, S has one vertex.

3. Let P be described by some $m \times n$ matrix A. $\{1, \dots, m\}$ has only finitely many subsets. Thus we have only finitely many faces.

$$\leq \binom{m}{n} \text{ vertices}$$

□

Remark. *By Theorem 1.4, it is immediate that non-empty polyhedrons, that result from linear programs in canonical form, $P = \{x \in \mathbb{R}^n \mid Ax \leq b, x \geq 0\}$ are always acute. So it has at least one vertex (because the set $\{x \in \mathbb{R}^n \mid x_i \geq 0, i \in \{1, \dots, n\}\}$ does not contain a line).*

1.1.1 Fundamental theorem of Linear Optimization

Theorem 1.5 (Fundamental theorem of Linear Optimization)**.** *1. If an affine-linear function $z(x)$ takes up its maximum/minimum in a polyhedron P in $\overline{x} \in P$, then also in all points of face $S = L(\overline{x}) \cap P$.*

2. Especially the optimum is taken up in a vertex of P, if P is acute.

Proof. 1. Let $\overline{x}$ be a maximum (analogously for minima) and let $y \neq \overline{x}$ be another point at S. Then the line $G := G(\overline{x}, y)$ in $L(\overline{x})$. For G there exists a representation $G = G(c,d)$ with $c, d \in P, \overline{x} \in (c,d)$. By Lemma 1.1 the affine-linear function $z(x)$ takes up its maximum in line segment $[c,d]$ in c or d. Without loss of generality, we assume its maximum in c. Thus $z(c) \geq z(\overline{x})$, because $\overline{x} \in (c,d)$.

On the other hand, we have $z(\overline{x}) \geq z(c)$, because $c \in P$ and the maximum is reached in $\overline{x}$.

Thus $z(c) = z(\overline{x})$. Hence, $z(x)$ is constant in G and therefore $z(y) = z(\overline{x})$. y was chosen arbitrarily, then z is constant at face S.

2. Follows by Theorem 1.4 (b).

□

The polyhedron for linear programs in canonical form are empty or acute (have vertices). The Fundamental theorem of Linear Optimization followingly states that for such linear programs (and thus any linear program becausc every linear program can be represented in canonical form) it suffices to investigate all vertices.

Corollary 1.6. *If* $\max\{c^{\perp}x : x \in P\}$ *has a linear optimization solution* x^**, then* $c^t x^* = \max\{c^t x : x \in V(P)\}$ *where* $V(P)$ *is the set of vertices of* P*.*

Thus we retrieve a finite method for linear programs: Determine all vertices and filter the vertex optimizing the target function.

Remark (Disadvantage)**.** *Because there are exponentially (in* n *and* m*) many vertices in general, there is no practically useful method of this idea.*

2 The generic Simplex Method

The Simplex Method goes back to George Dantzig (1947). The method relies on the Fundamental theorem of Linear Optimization and tries to find an optimal vertex. It utilizes convexity to claim a local minimum as global one. Thus for a given vertex x^*, any adjacent vertex (reachable by one edge) has a worse target function value.

The basic idea is:

1. Determine an initial vertex x (if none exists, the polyhedron is empty because we utilize the canonical form).
2. Test whether x is a local optimum. Consider the edges starting from x (they are either unbounded or lead to adjacent vertices).
3. If x is a local optimum, then stop.
4. Otherwise either an unbounded problem is given or we replace x by some adjacent vertex with a better target function value.
5. Iterate this process.

This process is necessarily finite, because there are only finitely many vertices. This process gives rise to the generic Simplex algorithm:

1. Choose an arbitrary vertex x of P as initial vertex. If none exists ($P = \emptyset$), then stop.
2. While there exists some edge k starting from x increasing along the target function value, do

(a) Choose such an edge

(b) If k is not a halfline of our polyhedron P then

i. substitute x by edge $\tilde{x}$ at the other end of k

else stop, as the problem is unbounded

3. Return vertex x

↓ *This lecture took place on 2019/03/11.*

Our next goal is to implement of this algorithmic idea algebraically.

Remark (Observation)**.** *It is difficult to transform inequality systems of form* $Ax \leq b$*.*

Transformation of an inequality system:
Canonical form $\max\{c^\perp x \text{ s.t. } Ax \leq b_1, x \geq 0\}$ with $x \in \mathbb{R}^n, b \in \mathbb{R}^m, c \in \mathbb{R}^n, A \in \mathbb{R}^{m\times n}$.

Polyhedron $P = \{x \in \mathbb{R}^n \mid Ax \leq b, x > 0\}$.

Introduction of auxiliary variables y_i (slack variables) (dt. "Schlupfvariable"). $y = b - Ax$ in vector notation. $y_i = b_i - a_{i1}x_1 - \cdots - a_{in}x_n \quad i = 1, \ldots, m$.

Every point $x \in \mathbb{R}^n$ corresponds to exactly one point $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^{n+m}$.

Polyhedron $P \to$ polyhedron $\tilde{P} = \left\{\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^{m+n} \mid Ax + y = b, x, y \geq 0\right\}$. The polyhedron structure is retained in such a way that dimensions of faces are preserved and vertices will become vertices.

The following correspondence will become useful:

$$x_{n+1} := y_1 \qquad x_{n+2} := y_2 \qquad \ldots \qquad x_{n+m} := y_m$$

This provides a uniform naming of variables.

This results in the following representation, we call *normal form*

$$\max c_1x_1 + \cdots + c_nx_n + c_{n+1}x_{n+1} + \cdots + c_{n+m}x_{n+m}$$

subject to

$$\begin{array}{ccccccc}
a_{11}x_1 & +\ldots & +a_{1n}x_n & +x_{n+1} & & = b_1 \\
a_{11}x_1 & +\ldots & +a_{1n}x_n & & +x_{n+1} & = b_2 \\
 & & & & \vdots & \\
a_{11}x_1 & +\ldots & +a_{1n}x_n & & & \\
 & & & & \vdots & \\
a_{m1}x_1 & +\ldots & +a_{mn}x_n & & & = b_m \\
x_1 & ,\ldots, & x_{m+n} & & & \geq 0
\end{array}$$

We agree on $c_{n+1} = \cdots = c_{n+m} = 0$.

In the following, we will also denote the previous coefficient matrix with A. This A results from the canonical form and a $m \times n$ unit matrix I.

$$\left(\begin{array}{c|c} A_{\text{canonical}} & I \end{array} \right)$$

Example (Canonical form)**.**

$$\max x_1 + x_2$$

s.t.

$$\begin{aligned} x_1 + 2x_2 &\leq 4 \\ 2x_1 - x_2 &\leq 3 \\ x_2 &\leq 1 \\ x_1, x_2 &\geq 0 \end{aligned}$$

Example (Normal form)**.**

$$\max x_1 + x_2$$

s.t.

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 4 \\ 2x_1 - x_2 + x_4 &= 3 \\ x_2 + x_5 &\leq 1 \\ x_1, x_2, \ldots, x_5 &\geq 0 \end{aligned}$$

Remark. *In the following, we assume a linear program in normal form. A is a $m \times (m+n)$ matrix for which we assume that it has full row rank* $\operatorname{rank}(A) = m$.

In a similar way, P denotes the polyhedron corresponding to our system $Ax = b, x > 0$. Let $J \subseteq \{1, \ldots, m+n\} \to (J(1), J(2), \ldots, J(k))$ be a map to index vectors where J is an index set $|J| = K$.

Be aware that we implicitly switch between sets and tuples.

Our next goal is to introduce the terms basis, basis solution, non-basis.

Definition. *A submatrix A_B of A with $A_B = (A_{B(1)}, \ldots, A_{B(m)})$ and* $\operatorname{rank} A_B = m$ *(thus the m columns of A_B are linear independent) is called* basis matrix *and B is called* basis. *Here we assume that A_B is regular.*

The remaining columns of A are summed up in index vector N.

$$\text{matrix } A_N = (A_{N(1)}, \ldots, A_{N(n)}) \text{ where } N = \{1, \ldots, m+n\} \setminus B$$

N is called *non-basis* and considered as set. A_N is called *non-basis matrix.*

We call x_j with $j \in B$ *basis variable* and x_j with $j \in N$ *non-basis variable.*

The following compact notations are practical:

x_B	...	vector of basis variables
x_N	...	vector of non-basis variables
c_B	...	vector of cost-coefficients c_j for $j \in B$ (basis variable)
c_N	...	vector of cost-coefficients c_j for $j \in N$ (non-basis variable)

$$\max c^t x \qquad Ax = b, x \geq 0$$

$$\implies \max c_B^t x_B + c_N^t x_N \qquad A_B x_B + A_N x_N = b; x_B, x_N \geq 0$$

We can write it as,

$$c = (c_B, c_N) \quad x = (x_B, x_N) \quad A = (A_B, A_N)$$

Definition. *A vector $x \in \mathbb{R}^{m+n}$ is called* basis solution *of a linear optimization problem in normal form ($\max\{c^t x \mid Ax = b, x \geq 0\}$), if there exists some basis B with $A_B x_B = b$ and $x_N = 0$ (remark: $x_B = A_B^{-1} b$).*

A basis solution is called admissible *if $x_B \geq 0$. In this case, B is called* admissible basis.

A basis solution is called *degenerate*, if there exists some i with $x_{B(i)} = 0$. Otherwise x_B is called non-degenerate. Analogously we define *degenerate bases* and *non-degenerate bases*.

Remark. *A basis solution x is in polyhedron P, if it is admissible.*

For the generic Simplex method, we go from vertex to vertex and thus from admissible solution to admissible solution.

Example. $N = (1, 2)$*. So, non-basis variables are* x_1, x_2
$B = (3, 4, 5)$*. So, basis variables are* x_3, x_4, x_5.

The corresponding basis solution $(0, 0, 4, 3, 1)$.

$$N = (1, 5) \qquad B = (2, 3, 4)$$

is the corresponding basis solution.

Solve the system:

$$\begin{pmatrix} 2 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \\ 1 \end{pmatrix}$$

$$\begin{aligned} 2x_2 + x_3 &= 4 \\ -x_2 + x_4 &= 3 \\ x_2 &= 1 \end{aligned}$$

So $x_3 = 2$ and $x_4 = 4$

$$(0, 1, 2, 4, 0)$$

is admissible and non-degenerated.

$$B = (1, 2, 3) \qquad N = (4, 5)$$
$$B = (1, 2, 4) \qquad N = (3, 5)$$
$$B = (1, 2, 5) \qquad N = (3, 4)$$

all lead to $x = (2, 1, 0, 0, 0)^t$ (admissible, degenerated).

Remark (Here be dragons)**.** *To some non-degenerate basis solution, there exists exactly one basis. This does not hold true for degenerated basis solutions.*

The same vertex of the polyhedron corresponds to several bases in case of degeneration.

Theorem 2.1. *The admissible basis solutions correspond to the vertices of the polyhedron, vice versa. If the basis solution is non-degenerated, then the corresponding basis is uniquely determined.*

Proof.

1. The basis solution $\tilde{x}$ (for basis B and non-basis N) maps [by the definition of the basis solution] to $\tilde{x}_N = 0$ and $\tilde{x}_B$ is the unique solution of $A_B \tilde{x}_B = b$ ($Ax = b$).

   $$\{\tilde{x}\} = \{x \mid x_N = 0\} \cap \{x \mid Ax = b\}$$

   If $\tilde{x}$ is an admissible basis solution, then $\tilde{x}_B \geq 0$ and thus $\tilde{x} \geq 0$ and thus $\tilde{x} \in P$. Hence $\tilde{x}$ is a vertex of P.

2. Let $\hat{x}$ be a vertex of P. Then the sign conditions must be satisfied and $\hat{x}$ is not uniquely defined by $m + n$ equations. Hence $\left\{ \begin{matrix} n \\ x \end{matrix} \right\} = \{x \mid x_N = 0\} \cap \{x \mid Ax = b\}$ with $N \subseteq \{1, \ldots, m + n\}$. A_B must be regular. $\hat{x}$ is basis solution.

3. In some non-degenerated basis solution, there are exactly m components $\neq 0$. B is uniquely defined and thus N.

□

Theorem 2.1 allows us to use basis solutions to implement the Simplex method numerically/algebraically.

Remark. *Let a linear program in normal form be given. We assume it was transformed from the canonical representation. With $N = (1, \ldots, n)$ and $B = (n+1, \ldots, n+m)$ for $b \geq 0$, we always get one admissible basis solution $x \begin{pmatrix} x_N = 0 \\ x_B = b \end{pmatrix}$.*

This corresponds to the origin in the coordinate system of the canonical form. It can be used as initial guess in the Simplex method. For the other cases, we are still looking for an approach.

Now let B be a fixed basis and N is the corresponding non-basis.

$$\begin{aligned} Ax = b \iff A_B x_B + A_N x_N = b \qquad A_B \text{ regular} \\ x_B = A_B^{-1}(bA_N x_N) = \underbrace{A_B^{-1} b}_{:=\tilde{b}} - \underbrace{A_B^{-1} A_N}_{\tilde{A}_N} x_N \\ \tilde{b} := A_B^{-1} b \qquad \tilde{A}_N = A_B^{-1} A_N \\ x_B = \tilde{b} - \tilde{A}_N x_N \end{aligned}$$

Polyhedron P:

$$x_B = \tilde{b} - A_N x_N \qquad x_B, x_N \geq 0$$

where the basis variables are represented by the non-basis variables. This is the reduced representation wrt. (BN).

Representation of form:

$$x_{B(i)} = t_{i_0} + \sum_{j=1}^{n} t_{ij} x_{N(j)} \qquad i = 1, \dots, m$$

where t_{ij} are the representation coefficients t_{ij} with $i = 1, \dots, m$ and $j = 0, \dots, n$.

Projection of the polyhedron in the space of independent variables (non-basis variables).

We retrieve the canonical form representation in this space

$$\tilde{A}_N x_N \leq \tilde{b} \qquad x_N \geq 0$$

Example (continued)**.**

$$N = (1, 5) \qquad B = (2, 3, 4)$$

$$\begin{aligned} x_3 &= 2 - x_1 + 2x_5 \\ x_4 &= 4 - 2x_1 - x_5 \\ x_2 &= 1 - x_5 \end{aligned}$$

We want to insert this new representation into the target function.

$$\begin{aligned} z(x) = z &= z_0 + c^t x = z_0 + c_B^t x_B + c_N^t x_N \\ &= \underbrace{(z_0 + \underbrace{c_B^t A_B^{-1} b}_{\text{constant}})}_{\tilde{z}_0} - \underbrace{(c_B^t A_B^{-1} A_N X_N + C_N^t x_N)}_{+(c_N^t - c_B^t A_B^{-1} A_N) x_N} \\ &= \tilde{z}_0 + \tilde{c}_N^t x_N \end{aligned}$$

with $\tilde{c}_N^t = c_N^t - c_B^t \underbrace{A_B^{-1} A_N}_{} \tilde{A}_N$. $\tilde{c}_N$ is called *reduced cost coefficients*. Later, $\tilde{z}_0$ and $\tilde{c}_N$ will become the 0-th row of the coefficient tableau.

> ↓ *This lecture took place on 2019/03/12.*

Revision.

$$z \coloneqq t_{00} + \sum_{j=1}^{n} t_{0j} X_{N(j)}$$

as 0-th row of the tableau.

Example.

$$N = (1, 5) \qquad B = (2, 3, 4)$$

$$\max x_1 + x_2 = 1 + x_1 - x_5$$

Let $x_2 = 1 - x_5$. Representation of the target function in the space of non-basis variables.

2.1 Sufficient optimality criterion for basis solutions

Basis solutions correspond to vertices.

A sufficient basis solution x for basis B (non-basis N) is optimal for the given linear program in normal form if $\tilde{c}_n \leq 0$ where $\tilde{c}_n$ is the vector of reduced cost coefficients.

> Reduced form
>
> $$\max \left\{ \tilde{c}_N^t x_n / \tilde{A}_N x_N \leq \tilde{b}_n, x_n \geq 0 \right\}$$
>
> with $\tilde{c}_N, \tilde{A}_N, \tilde{b}$ as established in the last lecture.

Remark. *The criterion is not necessary.*

Example (Continued)**.** $\tilde{c}_N = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \not\geq 0$. *Criterion not satisfied.*

Remark (Research question)**.** *How can we potentially improve the target function value if the optimality criterion is not satisfied?*

Currently we are in a vertex. Basis solution und non-basis variables with $\tilde{c}_{N(j)} = t_{0j} > 0$ have potential to give a better target function value if we increase $X_{N(j)}$ from 0 to some value > 0. We can only increase $X_{N(j)}$ such that admissibility is preserved.

Example (Continued)**.**

$$\max 1 + x_1 - x_5$$

$$\begin{aligned}
x_2 &= 1 - x_5 && \textit{no constraint}\\
x_3 &= 2 - x_1 + 2x_5 && \implies x_1 \leq 2\\
x_4 &= 4 - 2x_1 - x_5 && \implies x_1 \leq 2\\
x_i &\geq 0 \forall i \in \{2,3,4\}
\end{aligned}$$

We want to increase x_1! By how much is it admissible? The answer in this example is 2.

Which variable leaves the basis and becomes a non-basis variable instead of x_1? We have two options here: x_3 or x_4 (both values are 0 if $x_1 = 2$).

Assume we chose x_4. The new basis is $(1,2,3)$ and the new non-basis is $(4,5)$. We get a new reduced representation. And so on and so forth.

Step of improvement, general description

Let s chosen[1] such that $\tau_{N(s)} = t_{0s} > 0$ (optimality criterion is not satisfied). Our goal is to make $X_{N(s)}$ as large as possible. All other non-basis variables are fixed to be zero.

Case distinction:

Case 1: all $t_{is} \geq 0$ for all i $X_{N(s)}$ can be arbitrary large. The linear program is unbounded.

Case 2: there exists some i with $t_{is} < 0$ Then we determine

$$\varepsilon := \min \left\{ \frac{\overbrace{t_{i0}}^{\tilde{b}_i}}{-t_{is}} \mid t_{is} < 0 \right\}$$

Let r be such that $\varepsilon = \frac{t_{r0}}{-t_{rs}}$. $X_{B(r)}$ takes up the value 0.

We substitute the variable in $N(s)$ with the variable in $B(r)$.

Remark. *r is not necessarily unique. Currently, the choice in such cases is arbitrary.*

New basis:

$$\overline{B}(i) := \begin{cases} B(i) & i \neq r \\ N(s) & i = r \end{cases}$$

[1] the selection criteria will be discussed later

New non-basis:

$$\overline{N}(j) := \begin{cases} N(j) & j \neq s \\ B(r) & j = s \end{cases}$$

In the following, r will be called *pivot row*, s will be called *pivot column* and t_{rs} will be called *pivot element*. The transition from (B, N) to $(\overline{B}, \overline{N})$ is called *pivot step*.

So the basis solution for (B, N) becomes the basis solution for $(\overline{B}, \overline{N})$. The vertex x becomes the adjacent vertex $\overline{x}$.

Implementation of the basis exchange

Exchange $B(r) \leftrightarrow N(s)$. We solve the constraint belonging to pivot row r (to $x_{B(r)} = \dots$) by $X_{N(s)}$ and insert it into the remaining constraints.

$$-t_{rs}x_{N(s)} = t_{r0} - x_{B(r)} + \sum_{j \neq s} t_{rj}x_{N(j)}$$

$$\Longrightarrow \underbrace{x_{N(s)}}_{=x_{\overline{B}(r)}} = -\frac{t_{r0}}{t_{rs}} + \frac{1}{t_{rs}}x_{B(r)} + \sum_{j \neq s} \frac{t_{rj}}{t_{rs}}x_{N(j)}$$

This constraint is finished.

For $i \neq r$ we get

$$x_{\overline{B}(i)} = x_{B(i)} = t_{i0} + t_{is}\left(\frac{t_{r0}}{-t_{rs}} + \frac{1}{t_{rs}}x_{B(r)} + \sum_{j \neq s} \frac{t_{rj}}{-t_{rs}}x_{N(j)}\right) + \sum_{j \neq s} t_{ij}x_{N(j)}$$

$$= \underbrace{(t_{i0} - \frac{t_{is}}{t_{rs}}t_{r0})}_{\bar{t}_{i0}} + \underbrace{\frac{t_{is}}{t_{rs}}}_{\bar{z}_{is}} x_{\overline{N}(s)} + \sum_{j \neq s} \underbrace{\left(t_{ij} - \frac{t_{is}}{t_{rs}}t_{rj}\right)}_{\bar{t}_{ij}} x_{\overline{N}(j)}$$

$$x_{\overline{B}(i)} = \bar{t}_{i0} + \bar{t}_{is}x_{\overline{N}(s)} + \sum_{j \neq s} \bar{t}_{ij}x_{\overline{N}(j)}$$

$i \neq r$, constraint is finished.

Analogously, for the target function row (case $i = 0$).

Summary in tableau form

Tableau T is transformed into tableau F by a simplex step. The tableau is given as a table with highlighted 0th row (target function). The value t_{rs} is given by pivot row r and pivot column s.

	s	j
r	A	B
i	C	D

new x-element = old x-element minus $\frac{C \cdot B}{A}$.
Pivot element $\rightarrow$ use reciprocal.
Remaining pivot column: $\cdot - 1$ divided by pivot element
Remaining pivot row: divide by pivot element

The transformation laws for the pivot step are given by

$$\overline{t_{rs}} := \frac{1}{t_{rs}}$$

$$\overline{t_{rj}} := -\frac{t_{rj}}{t_{rs}} \qquad j = 0, \dots, n; j \neq s$$

$$\overline{t_{is}} := -\frac{t_{is}}{t_{rs}} \qquad i = 0, \dots, m; j \neq r$$

Remark. *Internalize these rules by heart!*

In more detail:

Example (Our standard example)**.**

$$\max x_1 + x_2$$

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 4 \\ 2x_1 - x_2 + x_4 &= 3 \\ x_2 + x_5 &= 1 \\ x_1, x_2, x_3, x_4, x_5 &\geq 0 \end{aligned}$$

Initial tableau for $B = (3, 4, 5)$ and $N = (1, 2)$.

	x_1	x_2	
0	1	1	
4	1	2	x_3
3	2	-1	x_4
1	0	1	x_5

belongs to solution $x_1 = x_2 = 0, x_3 = 4, x_4 = 3$ and $x_5 = 1$.

Non-optimal, x_1 and x_2 can be considered as new basis variables. Assume we choose $s = 1$.

$$\varepsilon = \min \left\{ \frac{4}{1}, \frac{3}{2} \right\} = \frac{3}{2}$$

$r = 2$, so x_4 is removed from the basis. This corresponds to $x_1 = \frac{3}{2}$ and $x_2 = 0$, so target function value is $\frac{3}{2}$.

$$\begin{array}{c|ccc} & x_4 & x_2 & \\ -\frac{3}{2} & -\frac{1}{2} & \frac{3}{2} & \\ \hline \frac{5}{2} & -\frac{1}{2} & \frac{5}{2} & x_3 \\ \frac{3}{2} & \frac{1}{2} & -\frac{1}{2} & x_1 \\ 1 & 0 & 1 & x_5 \end{array}$$

New basis: $(1, 3, 5)$
New non-basis: $(2, 4)$

The current values of the basis are retrievable.

It is not yet optimal. x_2 should be removed from the non-basis. $s > 2$ (second column)

$$\varepsilon = \min\left\{\frac{\frac{5}{2}}{\frac{5}{2}}, \frac{1}{1}\right\} = 1$$

Two options. Let's choose $r = 3$.

Tableau:

$$\begin{array}{c|ccc} & x_4 & x_5 & \\ -3 & -\frac{1}{2} & -\frac{3}{2} & \\ \hline 0 & -\frac{1}{2} & -\frac{5}{2} & x_3 \\ 2 & \frac{1}{2} & \frac{1}{2} & x_1 \\ 1 & 0 & 1 & x_2 \end{array}$$

is optimal. $x_1 = 2$ and $x_2 = 1$ with target function value 3.

Next week: How can we determine an admissible solution?

↓ *This lecture took place on 2019/03/18.*

Revision. *It remains to discuss:*

- *Which solution is necessary to begin with if b is not ≥ 0?*
- *What about finiteness of the algorithm if the basis solution is degenerated?*

To determine admissible basis solutions, we have 2 approaches: The first one is called "2-phase method by Dantzig".

2-*phase method by Dantzig*

Given $Ax = b$ with $x \geq 0$ and $\exists i : b_i < 0$.

We sort the rows (= constraints) of A and b such that in the first rows are those with negative right-hand side.

$$b = \begin{pmatrix} \hat{b} \\ \hat{\hat{b}} \end{pmatrix} \text{ with } \hat{b} < 0, \hat{\hat{b}} \geq 0 \text{ and } A = \begin{pmatrix} \hat{A} \\ \hat{\hat{A}} \end{pmatrix}$$

System in normal form (block remains unchanged):

$$\hat{A}x + \hat{y} = \hat{b} \qquad \hat{\hat{A}}x + \hat{\hat{y}} = \hat{\hat{b}}$$

where $\hat{\hat{y}}$ is the vector of slack variables for the remainder with $b_i < 0$ and $b_i \geq 0$. By multiplication with -1:

$$\begin{aligned} -\hat{A}x - \hat{y} + u &= -\hat{b} \\ \hat{\hat{A}}x + \hat{\hat{y}} &= \hat{\hat{b}} \end{aligned}$$

Choose u and $\hat{\hat{y}}$ as basis variable.

Resolve by the basis variables

$$\begin{aligned} u &= -\hat{b} + \hat{A}x + \hat{y} \\ \hat{\hat{y}} &= \hat{\hat{b}} - \hat{\hat{A}}x \end{aligned}$$

Gives an admissible basis solution (of the new system). The remainders are non-basis variables.

$$u = -\hat{\hat{b}} \qquad \hat{\hat{y}} = \hat{\hat{b}}$$

Solutions with $u \neq 0$ are non-admissible solutions for our original problem. Our goal is to find solutions with $u = 0$ if such a solution exists.

The implementation is done by introducing an auxiliary problem

$$\min e^t u \text{ with } u \geq 0 \qquad \text{where } e = (1, \ldots, 1)$$

$$\iff \max \underbrace{-e^t u}_{Z_H}$$

means that we find a solution with $u = 0$, if possible.

$$Z_H = -e^t u = e^t(\hat{b} - \hat{A}x - \hat{y}) = e^t\hat{b} - e^t\hat{A}x - e^t\hat{y}$$

Representation in the space of non-basis variables:

Example.

$$\max -x_1 - 2x_2$$

subject to

$$
\begin{aligned}
x_1 + x_2 &\geq 3 \\
x_2 &\geq 2 \\
-x_1 + x_2 &\leq 3 \\
x_1 - x_2 &\leq 3 \\
x_1, x_2 &\geq 0
\end{aligned}
$$

$$
\begin{aligned}
\implies -x_1 - x_2 &\leq -3 \\
-x_2 &\leq -2 \\
-x_1 + x_2 &\leq 3 \\
x_1 - x_2 &\leq 3 \\
x_1, x_2 &\geq 0
\end{aligned}
$$

$$
\begin{aligned}
\implies x_1 + x_2 - y_1 + u_1 &= 3 \\
x_2 - y_2 + u_2 &= 2 \\
-x_1 + x_2 + y_3 &= 3 \\
x_1 - x_2 + y_4 &= 3 x_1, x_2, y_1, y_2, y_3, y_4, u_1, u_2 \qquad \geq 0
\end{aligned}
$$

Here the variables before the equality sign are the basis variables to begin with.

The auxiliary problem is given by

$$\max -u_1 - u_2 = \max -5 - y_1 - y_2 + x_1 + 2x_2$$

$$
\begin{array}{r|rrrrl}
 & x_1 & x_2 & y_1 & y_2 & \\
5 & 1 & 2 & -1 & -1 & \\
\hline
0 & -1 & -2 & 0 & 0 & \\
\hline
3 & 1 & 1 & -1 & 0 & u_1 \\
2 & 0 & 1 & 0 & -1 & u_2 \\
3 & -1 & 1 & 0 & 0 & y_3 \\
3 & 1 & -1 & 0 & 0 & y_4
\end{array}
$$

The initial solution is given by (is non-optimal):

$$u_1 = 3 \qquad u_2 = 2 \qquad y_3 = 3 \qquad y_4 = 3$$

Choose $r = 2$ *and* $s = 2$ *and apply the pivot step method.*

$$
\begin{array}{c|cccc|l}
 & x_1 & u_1 & y_1 & y_2 & \\
1 & 1 & -2 & -1 & 1 & \\
\hline
4 & -1 & 2 & 0 & -2 & \\
\hline
1 & 1 & -1 & -1 & 1 & u_1 \\
2 & 0 & 1 & 0 & -1 & x_2 \\
1 & -1 & -1 & 0 & 1 & y_3 \\
5 & 1 & 1 & 0 & -1 & y_4
\end{array}
$$

u_2 became non-basis variables and we thus we can remove (i.e. ignore) the column corresponding to u_2. This solution is non-optimal (choose $s = 1$ and $r = 1$).

$$
\begin{array}{c|ccc|l}
 & u_1 & y_1 & y_2 & \\
0 & 1 & 0 & 0 & \\
5 & +1 & -1 & -1 & \\
\hline
1 & 1 & -1 & 1 & x_1 \\
2 & 0 & 0 & -1 & x_2 \\
2 & 1 & -1 & 2 & y_3 \\
4 & -1 & 1 & -2 & y_4
\end{array}
$$

optimal for the auxiliary problem. We remove the second column from left.

$$x_1 = 1 \qquad x_2 = 2$$

is an admissible solution for the original problem (is already optimal for the original problem otherwise continue with the remaining tableau in the second phase). The target function value is 0 for the auxiliary problem. $u_1 = u_2 = 0$.

Thus the algorithm looks as follows:

1. Continue until the first auxiliary problem is solved.

   Case 1 The auxiliary problem has optimal value 0, then second phase

   Case 2 The auxiliary problem has optimal value $\neq 0$, then stop because the original problem does not have an admissible solution

2. Continue until the original problem is optimally solved.

Remark. *Once variable u_i end up in the non-basis, the corresponding column in tableau can be removed. At the end of the first phase, the auxiliary row can be removed. Attention! If some u_i is a basis variable at the end of the first phase by degeneration, this variable must not be removed for the second phase.*

2.2 M-method

We consider the new problem

$$\tilde{e}_i = \begin{cases} & b_i < 0 \\ 0 & b_i \geq 0 \end{cases}$$

thus the new variable $\tilde{x}$ is substracted with $b_i < 0$ in the constraints.

$$\max c^t x - M\tilde{x}$$

subject to

$$\begin{aligned} Ax + y - \tilde{e}^t\tilde{x} &= b \\ x, y, \tilde{x} &\geq 0 \end{aligned}$$

with M sufficiently large such that $\tilde{x}$ takes up value 0 in an optimal solution ($\exists$ a solution with $\tilde{x} = 0 \iff$ original problem has an admissible solution).

Remark. *Or we can consider*

$$Ax + y - e^t\tilde{x} = b \qquad x, y, \tilde{x} \geq 0$$

$\tilde{x}$ occurs in all constraints.

To get an admissible solution for the auxiliary problem (the problem extended by $\tilde{x}$), wlog. $b_m = \min_{x \leq 1 \leq m} b_i$ with $b_m < 0$ assuming A has m rows. Subtract the m-th row from all the others (for the variant, where $-\tilde{x}$ occurs in all constraints)

$$\begin{array}{lllllll} (a_{11} - a_{m1})x_1 + \dots & +(a_{11} - a_{mn})x_n + y_1 & & & - y_m & = b_1 - b_m & (1) \\ (a_{21} - a_{m1})x_1 + \dots & +(a_{21} - a_{mn})x_n & +y_1 & & - y_m & = b_2 - b_m & \\ \vdots & & & \ddots & & \vdots & \\ (a_{m-1,1} - a_{m1})x_1 + \dots & +(a_{m-1,r} - a_{mn})x_n & & & -y_{m-1} - y_m & = b_{m-1} - b_m & \\ -a_{m,1}x_1 + \dots & + - a_{mn}x_n + \tilde{x} & & & - y_m & = \underbrace{-b_m}_{>0} & (2) \end{array}$$

m-th row multiplied with -1. With $x_{n+1}, \dots, x_{n+m-1}, \tilde{x}$ we are given an admissible solution.

Remark (Problem in practice)**.** *One problem in practice is the choice of M.*

Our workaround is not to choose M explicitly. Instead we split the target function into two parts (the auxiliary part including $\tilde{x}$ and the remainder). This represents the lexicographic ordering for vectors.

Example.

	x_1	x_2	$\tilde{x}$	
0	0	0	−1	
0	−1	−2	0	
−3	−1	−1	−1	x_3
−2	0	−1	−1	x_4
3	−1	1	−1	x_5
3	1	−1	−1	x_6

must be converted into a correct initial tableau. Either by (1) *or alternatively bring* $\tilde{x}$ *into the basis. Choose the variable, that leaves the basis as the one, where the constraints take up* $\min b_i$ *(this corresponds to* (1)*).*

Remark. *Here the constraints are written in the original order. They must be modified with* (1)*.*

$$\begin{array}{r|rrrl}
 & x_1 & x_2 & x_3 & \\
3 & 1 & 1 & -1 & \\
0 & -1 & -2 & 0 & \\
\hline
3 & 1 & 1 & -1 & \tilde{x} \\
1 & 1 & 0 & -1 & x_4 \\
6 & 0 & 2 & -1 & x_5 \\
6 & 2 & 0 & -1 & x_6
\end{array}$$

The auxiliary problem is not yet optimal. Optimize the auxiliary problem first.

$$\begin{array}{r|rrrl}
 & x_4 & x_2 & x_3 & \\
2 & -1 & 1 & 0 & \\
1 & 1 & -2 & -1 & \\
\hline
2 & -1 & 1 & 0 & \tilde{x} \\
1 & 1 & 0 & -1 & x_1 \\
6 & 0 & 2 & -1 & x_5 \\
6 & -2 & 0 & 1 & x_6
\end{array}$$

Still not yet optimal for the auxiliary problem. $s = 2, r = 1$.

$$\begin{array}{r|rrrl}
 & x_4 & \tilde{x} & x_3 & \\
0 & 0 & -1 & 0 & \\
5 & -1 & 2 & -1 & \\
\hline
2 & -1 & 1 & 0 & x_2 \\
1 & 1 & 0 & -1 & x_1 \\
2 & 2 & -2 & -1 & x_5 \\
4 & -2 & 0 & 1 & x_6
\end{array}$$

The auxiliary problem is solved and $\tilde{x} = 0$ *(optimal value* 0*). So, there exists an admissible solution for the original problem.*

If $\tilde{x}$ *is in the non-basis (as in the example), this column can now be removed. Usually we continue with the remaining problem with the original target function. Here we stop, because optimality was reached.*

On finiteness

Remark (Obvious observation)**.** *If no degenerated basis solutions occur, the simplex process is finite (in every iteration the target function value increases, because* $\tilde{c}_{N(s)} > 0$, $a_{rs} > 0$ *and* $\tilde{b}_{B(r)} > 0$*).*

Question: Can it happen that, in the case of degenerated basis solutions, we traverse a cycle?
Answer: Yes, if we do not take proper prerequisites.

Such an example goes back to Gass (see practicals, exercise 45)

$$\max \frac{3}{4}x_1 - 150x_2 + \frac{1}{50}x_3 - 6x_4$$

subject to

$$\begin{aligned} \frac{1}{4}x_1 - 60x_2 - \frac{1}{25}x_3 + 9x_4 &\leq 0 \\ \frac{1}{2}x_1 - 90x_2 - \frac{1}{50}x_3 + 3x_4 &\leq 0 \\ x_3 &\leq 1 \\ x_1, x_2, x_3, x_4 &\geq 0 \end{aligned}$$

Question: Which prerequisites can we make to avoid cycles?

In the following, there are two approaches:

- Rule by Bland
- lexicographical row selection rule

Remark (About the lexicographical approach)**.** *Recall that in the case of degenerated basis solutions the choice of the pivot row is ambiguous (with the previous approach).*

Idea: *Introduce an extended criterion to choose a pivot row.*
Remark: *On constraint on the choico of a pivot column!*

A new rule for choice of the pivot columns considers the lexicographical minimum over row vectors instead of the minimum over scalars.

↓ *This lecture took place on 2019/03/19.*

2.3 Rules to avoid cycles

Today, we will discuss the lexicographical row selection rule.

We need appropriate measures to ensure in every pivot step a certain kind of progress. We already know that considering the target function value itself does not suffice.

In case of degenerated basis solutions, the minimum is taken up in computation of ε^* for ≥ 2 rows.

$$\begin{array}{c|ccc}
 & & & \\ \hline
* & \dots & \square & \dots \\
* & \dots & \square & \dots
\end{array}$$

So far, we ignored the columns filled with dots.

For the compact representation of the row selection rule, we start with the system $(A|I)x = b$ that results from the inequality system by introduction of slack variables.

Definition. *Let v be a vector in $\mathbb{R}^k$. v is called* lexicographically positive *(compact notation: $v \succ 0$) if its first non-zero component is positive.*

$$v \in \mathbb{R}^k \textit{ is lexicographically } \begin{array}{cll} \textit{smaller-equal} & \textit{than } u \in \mathbb{R}^k, \textit{ if} & v = u \textit{ or } u - v \succ 0 \\ \textit{smaller} & \textit{than } u \in \mathbb{R}^k, \textit{ if} & u - v \succ 0 \\ \textit{greater-equal} & \textit{than } u \in \mathbb{R}^k, \textit{ if} & v = u \textit{ or } v - u \succ 0 \\ \textit{greater} & \textit{than } u \in \mathbb{R}^k, \textit{ if} & v - u \succ 0 \end{array}$$

In the following, we call lexmin *the minimum with respect to this lexicographical order of vectors.*

Example. *Thus,* $(0, 0, 0, 4, 1, -7)$ *is lexicographically positive.* $(0, -1, 2, 3)$ *is lexicographically negative.*

Example. *Compare this example with the practicals exercise 45.*

$$\max \frac{3}{4}x_1 - 150x_2 + \frac{1}{50}x_2 - 6x_4$$

subject to

$$\begin{aligned}
\frac{1}{4}x_1 - 60x_2 - \frac{1}{25}x_3 + 9x_4 + x_5 &\leq 0 \\
\frac{1}{2}x_1 - 90x_2 - \frac{1}{50}x_3 + 3x_4 + x_6 &\leq 0 \\
x_3 + x_7 &\leq 1 \\
x_1, x_2, x_3, x_4 &\geq 0
\end{aligned}$$

Choose $s = 1$. The classical row selection rule provides no distinction between row 1 and 2 ($\frac{0}{\frac{1}{4}}$ versus $\frac{0}{\frac{1}{2}}$). Here the lexicographic rule chooses $r = 1$. $u \prec v$.

$$\begin{array}{c|cccc|ccc}
 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\
0 & \frac{3}{4} & -150 & \frac{1}{50} & 6 & 0 & 0 & 0 \\ \hline
0 & \frac{1}{4} & -60 & -\frac{1}{25} & 9 & 1 & 0 & 0 \\
0 & \frac{1}{2} & -90 & -\frac{1}{50} & 3 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1
\end{array}$$

with first and second row as,

$$\begin{array}{c|cccc|ccc} 0 & 1 & -240 & -\frac{4}{25} & 36 & 4 & 0 & 0 \\ 0 & 1 & -180 & -\frac{1}{25} & 6 & 0 & 2 & 0 \end{array}$$

In the following, we need the lexicographical sorting for vectors. Without loss of generality we assume that the rows of our tableau in extended form is lexicographically positive (otherwise apply column exchanges).

The remaining example is left for the practicals.

Definition (Lexicographical row seletion rule)**.** *Let s be the chosen pivot column. The pivot row r is chosen as the lexicographically smallest of the weighted row vectors of rows i with $t_{is} > 0$. t_{is} is the tableau entry in row i and column s.*

$$\operatorname{lexmin}_{i \in \{1,\ldots,n\}} \left\{ \frac{t_i}{t_{is}} \mid t_{is} > 0 \right\}$$

where t_i is the vector of the i-th tableau row.

Remark. *The choice above provides a unique solution. (Assumption: the choice is ambiguous.)*

$$\frac{t_i}{t_{is}} = \frac{t_k}{t_{ks}} \text{ for } i \neq k \implies \text{ the } i\text{-th row is a multiple of the } k\text{-th row}$$

$$\implies x \neq 0 \text{ with } t_i = \lambda \cdot t_k$$

This is a contradiction with $\operatorname{rank}(A) = m$.

It remains to show that the lexicographical row selection rule satisfies its purpose.

Theorem 2.2. *If you choose the pivot row by the lexicographical row selection rule, then*

1. *The vector in the target function row decreases strictly lexicographically.*
2. *All row vectors stay positive.*

Proof. 1. The new target function coefficients result from

$$\bar{t}_{vj} := t_{oj} - t_{rj} \underbrace{\frac{t_{os}}{t_{rs}}}_{>0} \qquad j \in \{1, \ldots, n+m\}$$

Because the first non-vanishing $t_{rj} > 0$, we get $T_0 \prec t_0$ where T_0 is the new vector (in the tableau in the target function row) and t_0 is the old vector. Thus, we get lexicographical decline.

2. For $t_{is} > 0$, due to choice of r,

$$\frac{1}{t_{is}} t_i \succ \frac{1}{t_{rs}} t_r$$

$$\implies \overline{t_i} = t_i - \frac{t_{is}}{t_{rs}} \cdot t_r \succ 0$$

where $\overline{t_i}$ is the i-th row. For $t_{is} \le 0$ due to $t_{rs} > 0$, we have

$$\overline{t_i} = t_i - \frac{t_{is}}{t_{rs}} \cdot t_r \succ t_i$$

Thus all vectors retain lexicographically positive.

□

Theorem 2.3. *The simplex method with lexicographical row selection rule is finite.*

Proof. Immediate by Theorem 2.2 (1). No basis solution can occur twice and there are only finitely many basis solutions. □

Remark (Where does the lex. selection rule come from?)**.** *Perturbation of the polyhedron/equation system:*

$$b \to b + \varepsilon \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix} \qquad 0 < \varepsilon_m \ll \varepsilon_{m-1} \ll \cdots \ll \varepsilon_1 \ll \text{ all other data}$$

Geometrically, a small deviation (perturbation) is introduced to the polyhedron.

In practice this cannot be implemented, because proper choice of ε is unknown. The lexicographical rule is an implicit implementation of this idea.

Remark (About the Rule by Bland)**.** *We constrain the choice of the pivot column and the pivot row. We choose the pivot column s as follows:*

$$N(s) := \left\{ N(j) \mid t_{0j} > 0 \quad j \in \{0, \ldots, n\} \right\}$$

(among the non-basis variables that violate the optimization condition, choose the one with smallest index).

Choose the pivot row r as follows:

$$B(r) := \min \left\{ B(q) \mid \tilde{b}_q = \min_{\substack{i \in \{1,\ldots,m\} \\ t_{is} > 0}} \tilde{b}_i \right\}$$

with $\tilde{b}_q := \frac{t_{q0}}{t_{qs}}$ and $\tilde{b}_i := \frac{t_{i0}}{t_{is}}$. So this chooses the variable with the smallest index *among all the variables we consider for leaving the basis.*

Theorem 2.4. *The simplex method enforced with the Rule by Bland is finite.*

Proof. Compare with appropriate literature. □

In the previous example, we need to choose $s = 1$ (due to the lexicographical rule, $s = 3$ would be possible as well). $r = 1$ (x_5 is the smallest index).

This conclude the generic simplex method.

3 Three extensions to the simplex method

subtitled "algorithmic aspects of the simplex method".

In this chapter, we are going to discuss extensions as well as efficiency in the algorithmic implementation of the simplex method.

3.1 Possible rule for choice of the pivot column

If we do not use the rule by Bland, we can choose every column that violates the optimality condition (thus $t_{0j} > 0$ with $t_{0j} = \tilde{c}_{N(j)}$). In literature the following rules are recommended:

Method of steepest slope in the space of non-basis variables. Also called *Rule by Dantzig*. Choose column s with $t_{0s} = \max t_{0j}$, i.e. $\max \left\{ t_{0j} \mid t_{0j} > 0 \right\}$.

Method of largest absolute increase of the target function. For every possible choice of j of the pivot column (hence $t_{0j} > 0$), we determine the corresponding pivot row $r(j)$. For every possible choice of j, this results in one pivot element $t_{r(j)j}$. Choose j (pivot column s) only such that we maximize

$$\frac{t_{0j} t_{r(j)0}}{t_{r(j)j}}$$

where t_{0j} corresponds to $(\tilde{c}_n)_j$ and $t_{r(j)0}$ corresponds to $\tilde{b}_{r(j)}$ and the fraction represents the increase of the target function.

Remark. *Disadvantage: determination is computationally intense.*

Method of steepest slope in the space of all variables. If we choose $x = (x_B, x_N)^t$ in some vertex, then the solution is given by $(\tilde{b}, 0)^t$.

If $x_{N(j)}$ increases from 0 to 1 (by 1 unit), then the target function value increases by $(\tilde{c}_N)_j$ (corresponds to t_{0j}), by the one hand, and x_B changes to

$$x_B = \underbrace{A_B^{-1} b}_{\tilde{b} - \tilde{a}_j} - A_B^{-1} A_N X_{N(j)}$$

on the other hand. Here j denotes the j-th column of $\tilde{A}_N$ (column corresponding to $X_{N(j)}$).

- Change in $\mathbb{R}^{m+n}$ by $(-\tilde{a}_{1j}, \ldots, -\tilde{a}_{mj}, 0, \ldots, 0, 1, 0, \ldots, 0)$ where 1 is given for $X_N(j)$
- for $N(j)$ corresponding component of the gradient of the target function in $\mathbb{R}^{m+k}$
- Row selection rule: Choose s such that $\frac{t_{0j}}{\sqrt{1+\sum_{i=1}^{m} t_{ij}^2}}$ (where t_{0j} corresponds to $(\tilde{c}_N)_j$ and t_{ij}^2 corresponds to $\tilde{a}_{ij}^2$) becomes maximal for $s = j$ and below the columns j with $(\tilde{c}_N)_j > 0$.

↓ *This lecture took place on 2019/03/25.*

3.2 Extensions to the Simplex method

Ideas: Direct handling of equations, no sign-restricted variables, upper bounds.

Basically this is not required, because every linear program can be transformed to canonical form. This happens to the disadvantage on the number of restrictions or the number of variables.

1st case: no sign-restricted variables Adjustment of the optimality criterion, adjustments to determine the pivot row. The details are given in the practicals.

2nd case: equations The approach is similar to the 2-phases method. Introduce one auxiliary variable for each equation and auxiliary target function. The minimum number of auxiliary variables corresponds to the maximum minus the sum of auxiliary variables. Compare this with the practicals.

3rd case: lower bound

$$l_j \leq x_j$$

Transformation to $\tilde{x}_j \geq 0$ by $\tilde{x}_j = x_j - l_j$. Compare this with the practicals.

4th case: upper bound

$$x_j \leq u_j \qquad (0, \ldots, 0, 1, 0, \overbrace{\ldots}^{x_j}, 0)$$

Goal: implicit handling of such residue classes without maintaining them in the tableau.

Idea: $x_j + \overline{x_j} = u_j$. We call $\overline{x_j}$ *complementary variables* to x_j and x_j complementary to $\overline{x_j}$. The value of x_j results from value $\overline{x_j}$; vice versa with $\overline{x_j}$ and x_j. It suffice to maintain one of the two variables; the value of the other can be easily determined.

Determine K, the index set of the variables bounded by above. We have $0 \leq x_j \leq u_j$ for $j \in K$ and $0 < u_j < \infty$. N denotes the index vector of the current non-basis and B is the index vector of the current basis. S is the index of the pivot column.

We need to distinguish three cases:

Case 1 $x_{N(s)}$ is bounded, so $x_{N(s)} \leq u_{N(s)}$

Case 2 $x_{B(i)} \leq 0 \to x_{B(i)} = \tilde{b}_i - \tilde{a}_{is} x_{N(s)} \geq 0 \implies x_{N(s)} \leq \frac{\tilde{b}_i}{\tilde{a}_{is}}$ for $\tilde{a}_{is} > 0$

Case 3 $x_{B(i)}$ is bounded by $u_{B(i)}$

$$x_{B(i)} = \tilde{b}_i - \tilde{a}_{is} x_{N(s)} \leq u_{B(i)} \implies x_{N(s)} \leq \frac{\tilde{b}_i - u_{B(i)}}{\tilde{a}_{is}} \text{ for } \tilde{a}_{is} < 0$$

By the choice of the pivot row, we need to determine the following minimum.

$$\min \left\{ \underbrace{U_{N(s)}}_{\text{Case 1}}, \underbrace{\min \left\{ \frac{\tilde{b}_i}{\tilde{a}_{is}} \mid \tilde{a}_{is} > 0 \right\}}_{\text{Case 2}}, \underbrace{\min \left\{ \frac{\tilde{b}_i - u_{B(i)}}{\tilde{a}_{is}} \mid \tilde{a}_{is} < 0 \text{ and } B(i) \in K \text{ i.e. } x_{B(i)} \text{ is bounded by above} \right\}}_{\text{Case 3}} \right\}$$

It remains to discuss what happens if the minimum results from the expression in case 1 or 3.

If it results from ...

Case 1, then
- $x_{N(s)}$ turns from 0 to $\tilde{U}_{N(s)}$
- $\overline{x_{N(s)}}$ becomes zero
- We replace $x_{N(s)}$ by $\bar{x}_{N(s)}$
- So a new non-basis variable is introduced, the basis retains unchanged

$$\overline{x_{N(s)}} = u_{N(s)} - x_{N(s)} \qquad \text{insertion into tableau representation}$$

$$\tilde{a}_{i1} x_{N(1)} + \cdots + \tilde{a}_{is} \underbrace{x_{N(s)}}_{(u_{N(s)} - \bar{x}_{N(s)})} + \cdots + \tilde{a}_{in} x_{N(n)} + x_{B(i)} = \tilde{b}_i$$

Transformation $T(s)$ s-th column with -1 and substitute the right hand side $\tilde{b}_i$ by $\tilde{b}_i - \tilde{a}_{is} u_{N(s)}$.

Case 2, then classical pivot operation with $\tilde{a}_{rc} > 0$.

Case 3, then combination of exchange of basis variables with non-basis variables and transition to complementary variables

Computational implementation:

- first, make a pivot step with $\tilde{a}_{rs}$ (consider $\tilde{a}_{rs} = 0$)
- second, apply transformation $T(s)$

Remark (Compact representation)**.** *Use negative indices for complementary variables $\overline{x_3}$ in basis, so use -3 instead of 3 as index in B.*

Example 4.

$$\max -x_1 + 4x_2$$

subject to

$$\begin{aligned} x_1 - x_2 &\leq 2 \\ -x_1 + x_2 &\leq 3 \\ x_2 &\leq 4 x_1, x_2 \qquad\qquad \geq 0 \end{aligned}$$

...becomes ...

$$\max -x_1 + 4x_2$$

subject to

$$\begin{aligned} x_1 - x_2 + x_3 &= 2 \\ x_2 &\leq 4 \\ x_3 &\leq 5 x_1, x_2, x_3 \qquad\qquad \geq 0 \end{aligned}$$

where $x_2 \leq 4$ and $x_3 \leq 5$ are two variables with upper bounds.

$$\begin{array}{r|rrl} & x_1 & x_2^* & \\ 0 & -1 & 4 & \\ \hline 2 & 1 & -1 & x_3 \end{array}$$

$$s = 2 \qquad \min \left\{ \underbrace{4}_{\text{Case 1}}, \underbrace{\frac{2-5}{-1}}_{\text{Case 3}} \right\} = 3$$

(a):

$$\begin{array}{r|rrl} & x_1 & x_3 & \\ 8 & 3 & 4 & \\ \hline -2 & -1 & -1 & x_2 \end{array}$$

(b): $T(2)$

$$\begin{array}{r|rrl} & x_1 & \overline{x}_3 & \\ -12 & 3 & -4 & \\ \hline -3 & -1 & 1 & x_2 \end{array}$$

$-1 \implies$ admissible tableau.

Now $s = 1$.

$$\min \left\{ \underbrace{\frac{3-4}{-1}}_{\text{Case 3}} \right\}$$

$$\begin{array}{r|rrl} & x_2 & \overline{x_3} & \\ -3 & 3 & -1 & \\ \hline -3 & -1 & -1 & x_1 \end{array}$$

$T(1)$

$$\begin{array}{r|rrl} & \overline{x}_2 & \overline{x_3} & \\ -15 & 3 & -1 & \\ \hline 1 & 1 & -1 & x_1 \end{array}$$

$$\implies x_1 = 1 \qquad \overline{x_2} = 0 \implies x_2 = 4 \qquad \overline{x_3} = 0 \implies x_3 = 5$$

3.3 The revised Simplex method

Recall:

- $x_B = A_B^{-1} b$
- $\tilde{C}_n^t = C_N^t - C_B^t A_B^{-1} A_N$
- $\tilde{A}_N = A_B^{-1} A_N$

One interpretation of the steps in the context of pivot operations of the Simplex method is, that 3 linear equation systems are solved.

- $A_B x_B = b$ to retrieve the current basis solution, $x_B = \tilde{b}$
- $A_B^t \Pi = c_B, \rightarrow \tilde{c}_N^t = c_N^t - \Pi_t A_n$
- $A_B \tilde{a}_s = a_s$ where a_s is the s-th column of A and $\tilde{a}_s$ is the s-th column of $\tilde{A}$

Regarding dimensions:

$$A_B, A_B^t \ldots m \times n \text{ matrices} \qquad \text{tableau size } O(mn)$$

If we solve the 3 equation systems in every step manually, we need $O(m^3)$ computational resources (independent of n). In the tableau $O(nm)$, so only useful for n at least $O(m^2)$.

In practice, we use the fact that A_B changes in every step only marginally (1 column!). The update formulas can be given explicitly (also for corresponding basis inverse). From a numerical perspective, it is advantageous to combine the process with known methods of linear algebra (for example, LU decomposition).

Conclusion: There exists numerically stable implementations of the Simplex method. Without proper prerequisites the rounding errors accumulate, cancellation effects occur such that results for larger ill-conditioned problems become unuseful.

3.4 Brief consideration of the runtime of the Simplex process

Obviously, the application of one pivot step can be done in polynomial time. The essential question is, does a polynomial boundary exist for the maximum number of pivot steps? A partial answer can be given: No such boundary is known and for all known column selection rules pathological examples have been found such that exponentially many pivot steps are required.

By current research, the simplex method does not provide a polynomial solution for the linear program. We are going to cover one such method in the section about the inner point method. But in practice, the simplex method works very good. Empirically, $O(n)$ pivot steps suffice.

Further analysis was done:

- Average case analysis by Borgwardt
- Smoothed analysis by Spielman, et al.

4 Duality for linear optimization

We will cover the definition of duality, duality theorems and alternative theorems.

4.1 A motivational example

Example (Transportation example)**.** *Linear transportation problem: transportation costs c_{ij}, m firms, offer a_i, n customers and demand b_j.*

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

subject to

$$\begin{aligned}
\sum_{j=1}^{n} x_{ij} &= a_i \text{ for } i = 1, \dots, m \\
\sum_{i=1}^{m} x_{ij} &= b_j \text{ for } j = 1, \dots, n \\
x_{ij} &\geq 0 \text{ for } i = 1, \dots, m; j = 1, \dots, n
\end{aligned}$$

↓ *This lecture took place on 2019/03/26.*

External transportation service: Purchase at firm with price u_i, sale at customer j with price v_j.

$$\begin{aligned}
\max \left(\sum_{i=1}^{n} b_j v_j - \sum_{i=1}^{m} a_i u_i \right) & \qquad \text{models the profit} \\
v_j - u_i \leq c_{ij} & \qquad i = 1, \dots, m; j = 1, \dots, n \\
v_j + \overline{u_i} \leq c_{ij} & \qquad i = 1, \dots, m; j = 1, \dots, n
\end{aligned}$$

or correspondingly with $\overline{u}_i = -u_i$. More examples in an economical context can be found.

4.2 Definition of a dual linear program

Definition. *Let (P) be a linear problem* $\max\{c^t x \mid Ax \leq b, x \geq 0\}$*. The dual linear problem (D) has form* $\min\{b^t y \mid A^t y \geq c, y \geq 0\}$*.*

Remark. *The problem (P) is also called* primal problem.

Example 5. *We compare the models side-by-side.*

$$\max x_1 + 2x_2 \qquad (P)$$

subject to

$$\begin{aligned} x_1 &\le 4 & y_1 \\ 2x_1 + x_2 &\le 10 & y_2 \\ -x_1 + x_2 &\le 5 & y_3 \\ x_1, x_2 &\ge 0 \end{aligned}$$

$$\min 4y_1 + 10y_2 + 5y_3 \qquad (D)$$

subject to

$$\begin{aligned} y_1 + y_2 - y_3 &\ge 1 \\ y_2 + y_3 &\ge 2 \\ y_1, y_2, y_3 &\ge 0 \end{aligned}$$

Example 6.

$$\min 4x_1 + 3x_2$$

subject to

$$x_1 + 2x_2 \ge 7 \tag{3}$$
$$2x_1 - x_2 \ge 5 \tag{4}$$
$$3x_1 + x_2 \ge -2 \tag{5}$$

First, we model it canonically:

$$\max -4x_1 - 3x_2$$

$$\begin{aligned} -x_1 - 2x_2 &\le -7 \\ -2x_1 + x_2 &\le -5 \\ -3x_1 - x_2 &\le 2 \end{aligned}$$

Secondly, we create sign constraints with $x_1 = x_1^+ - x_1^-$ *and* $x_2 = x_2^+ - x_2^-$.

$$\max -4x_1^+ + 4x_1^- - 3x_2^+ + 3x_1^-$$

subject to

$$\begin{aligned} -x_1^+ + x_1^- - 2x_2^+ + 2x_2^- &\le -7 & y_1 \\ -2x_1^+ + 2x_1^- + x_2^+ - x_2^- &\le -5 & y_2 \\ -3x_1^+ + 3x_1^- - x_2^+ + x_2^- &\le 2 & y_3 \\ x_i^\pm &\ge 0 \end{aligned}$$

Thus we get the dual problem:

$$\min -7y_1 - 5y_2 + 2y_3$$

subject to

$$\begin{aligned} -y_1 - 2y_2 - 3y_3 &\ge -4 \\ y_1 + 2y_2 + 3y_3 &\ge 4 \\ -2y_1 + y_2 - y_3 &\ge -3 \\ 2y_1 - y_2 + y_3 &\ge 3 \\ y_1, y_2, y_3 &\ge 0 \end{aligned}$$

$$\iff \min -7y_1 - 5y_2 + 2y_3$$

subject to

$$\begin{aligned} y_1 + 2y_2 + 3y_3 &= 4 \\ 2y_1 - y_2 + y_3 &= 3 \\ y_1, y_2, y_3 &\geq 0 \end{aligned}$$

Example 7. *Take the resulting dual problem in Example 2 and determine its dual problem.*

Remark (Observation)**.** *The dual problem of the primal problem is the dual problem. The dual problem of the dual problem is the primal problem.*

In general:

1. An equation in the primal problem corresponds to one unbounded (non-sign-restricted) variable in the dual problem
2. A non-sign-restricted variable in the primal problem corresponds to one equation in the dual problem
3. The dual problem of the dual problem is the primal problem

Primal problem	Dual problem
Maximization of the target function	Minimization of the target function
target function coefficients (c)	RHS vector
RHS vector (b)	target function vector
coefficients matrix (A)	transposed coefficients matrix
i-th constraint is $\leq$ constraint	i-th dual variable $y_i \geq 0$
i-th constraint is $=$ constraint	i-th dual variable y_i is not sign-restricted
$x_j \geq 0$	j-th constraint $\geq$ constraint
x_j is not sign-restricted	j-th constraint $=$ constrtaint (equation)

Example 8. *Consider the transportation problem. Begin with the minimization problem (the dual maximization problem).*

$$\max \sum_{i=1}^{m} a_i \cdot \alpha_i + \sum_{j=1}^{n} b_j \beta_j$$

subject to

$$\alpha_i + \beta_j \leq c_{ij}$$

Remark. *α_i and β_j are not sign-restricted, because in the primal problem we have equations. Compare with the transportation problem.*

Original constraints:

$$(D) \leftarrow (P) \left(\begin{array}{ccccccccc} 1 & \dots & 1 & 0 & \dots & & & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & 0 \\ & & & & \vdots & & & & \\ 0 & \dots & & & & 0 & 1 & \dots & 1 \\ \hline 1 & \ddots & 0 & 1 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & & 1 & 0 & & 1 & 1 & & 1 \end{array}\right)$$

The horizontal separates the firms model (top m rows) from the customers model (bottom n rows) giving $m + n$ rows in total.

Remark (Semantics of dual variables)**.**

- *β_j correponds to v_j in the introductory transportation problem (where v_j represents the price to be paid for transportation at customer j).*
- *α_i corresponds to u_i (where u_i represents the price to be paid at firm i)*

Remark. *Do* not *mix models! So either model a linear program*

- *as maximization problem with constraints $\leq$ or $=$*
- *as minimization problem with constraints $\geq$ or $=$*

Then apply the rules.

Remark. *Consider Example 2. Every point satisfying* (3)–(5) *also satisfies the corresponding linear combinations. e.g. 2 times* (3) + (4)*. Searching for the best boundaries exactly leads to the dual linear program.*

4.3 Duality- and alternative theorems

2 central duality theorems:

- strong duality theorem

  $c^t x^* = b^t y^* \qquad x^*$ optimal for primal problem, y^* optimal for dual problem

- weak duality theorem

  $c^t x \leq b^t y \qquad$ for $x \in M_P = \{x \mid Ax \leq b, x \geq 0\}, y \in M_D = \left\{y \mid A^t y \geq c, y \geq 0\right\}$

and the theorem of the complementary slackness (optimality check).

Remark. *In the following, let M_D denote the admissible set of the dual problem. Let M_P denote the admissible set of the primal problem.*

Theorem 4.1 (Weak duality theorem). *1. $x \in M_P, y \in M_D$. Then $c^t x \leq b^t y$*

2. *Let $M_P \neq \emptyset$ (hence the primal problem has an admissible solution) ($c^t x$ over M_P). Then M_P is empty, if the primal theorem is unbounded.*

3. *Let $M_D \neq \emptyset$ (hence the dual problem has an admissible solution). Then M_P is empty iff the dual problem is unbounded.*

Proof. 1. $x \in M_P$, so $Ax \leq b$ with $x \geq 0$. $y \in M_D$, so $A^t y \geq c$, $y \geq 0$.

$$c^t x \leq (A^t y)^t x = y^t A x \leq y^t b = (b^t y)$$

2. It suffices to prove (2) or (3).

   We will prove (3) und for this purpose, we are going to prove an auxiliary result first. And we are going to show this later.

□

We are going to briefly cover the topic of alternative theorems. The following is an introductory example:

Theorem 4.2. *A is an $m \times n$ matrix over $\mathbb{R}$. Let $b \in \mathbb{R}^m$. Exactly one of the following alternatives holds true:*

1. *$\exists x \in \mathbb{R}^n : Ax = b$*
2. *$\exists y \in \mathbb{R}^m : y^t A = 0, y^t b = 1$*

Hence, either the linear equation system $Ax = b$ has a linear solution or the linear equation system $y^t A = 0$ and $y^t b = 1$ has a linear solution.

Proof. Assume both alternatives are simultaneously true. This immediately gives a contradiction because $0 = g^t A x = y^t b = 1$.

The first case has no solution iff b is not in the space spanned by columns of A, hence $\operatorname{rank}([A|b]) = \operatorname{rank}(A) + 1$

$$\to \operatorname{rank}\left(\begin{pmatrix} A & b \\ 0 & 1 \end{pmatrix}\right) = \operatorname{rank}(A) + 1$$

if and only if the last row of the matrix above is linear dependent on the rows of $\begin{pmatrix} A & b \end{pmatrix}$, hence

$$y^t A = 0, y^t b = 1$$

has a solution. □

Theorem 4.3. *Exactly one of the following alternatives holds true:*

1. $\exists x \in \mathbb{R}^n : Ax = b, x \geq 0$
2. $\exists y \in \mathbb{R}^m : y^t A \geq 0, y^t b < 0$

Proof. If both alternatives hold true, then $0 \leq y^t Ax = y^t b < 0$. This gives a contradiction

Without loss of generality, we can assume that $b \geq 0$. Otherwise, we change the signs in the i-th row of $Ax = b$ and y_i simultaneously.

If the first alternative is not solvable, then the linear optimization problem

$$\begin{aligned}\gamma &:= \max\left\{-e^t u \mid Ax + u = b, x \geq 0, u \geq 0\right\} \text{ where } e \text{ is the zero-vector} \\ &:= \max\left\{-\sum u_i \mid Hz = b, z \geq 0, u \geq 0\right\}\end{aligned}$$

has only solutions with negative target function value (because no solution with $u = 0$ exists).

TODO (o)

So $\gamma < 0$. Now we apply the fundamental theorem of linear optimization. Thus there exists an optimal (finite) basis solution for (o) with target function value γ. □

↓ *This lecture took place on 2019/04/01.*

Revision 1 (Theorem 4.3)**.** *1. either $Ax = b$ with $x \geq 0$ has a solution*

2. or $y^t A \geq 0, y^t b < 0$ has a solution

Corollary 4.4 (Farkas Lemma, 1894)**.** *TFAE:*

- $\exists x \in \mathbb{R}^n : Ax = b, x \geq 0$
- $\forall y \in \mathbb{R}^m : y^t A \geq 0 \implies y^t b \geq 0$

Also has a geometric interpretation: The halfspace $\{y \mid y^t b \geq 0\}$ contains a polyedric cone $\{y \mid y^t A \geq 0\}$ iff $b \in \{Ax \mid x \geq 0\}$ where $\{Ax \mid x \geq 0\}$ is the cone $K(A)$.

Now we can complete the proof of Theorem 4.1. Thus we need to show statement (3) of Theorem 4.1.

Revision 2. *Let $M_D \neq \emptyset$. Then M_P is empty iff $y^t b$ is unbounded by below.*

Proof.
- Let $M_p \neq \emptyset$. By (1), we have that $c^t x \leq b^t y \forall x \in M_p$, $y \in M_D$, thus every $x \in M_p$ provides a lower bound for the target function values of the primal problem.

- Let $M_p = \emptyset$, thus $Ax = b$ with $x \geq 0$ has no solution. P has no admissible solution, so $y^tA \geq 0, y^tb < 0, y \geq 0$ has a solution (why? we will discuss it later). Let $\hat{y}$ be such a solution. Let $\overline{y} \in M_D$. Consider $\overline{y} + \lambda\hat{y} \in M_D$ for $\lambda \geq 0$ (half-line).

Target function value: $(\overline{y} + \lambda\hat{y})^tb = \overline{y}^tb + \lambda\hat{y}^tb$ is unbounded by below. □

$$(P) \qquad z_P = \max\left\{c^tx \mid Ax \leq b, x \geq 0\right\} \qquad M_p = \{x \mid Ax \leq b, x \geq 0\}$$
$$(D) \qquad p_D = \min\left\{b^ty \mid A^ty \geq c, y \geq 0\right\} \qquad M_D = \left\{y \mid A^ty \geq c, y \geq 0\right\}$$

If $M_P \neq \emptyset$ and $M_D \neq \emptyset$, then $z_P \leq z_D$ by Theorem 4.1.

Question: In this case, can $z_P < z_D$?
Answer: No, see Theorem 4.6.

Lemma 4.5. *$P_z = \{x \mid z \leq c^tx, Ax \leq b, x \geq 0\}$. Let $M_p \neq \emptyset$. Then $P_z = \emptyset \iff y^tb < z, A^ty \geq c, y \geq 0$ solvable.*

Theorem 4.6 (Strong duality theorem)**.** *If one of the linear problems, dual to each other, has a finite optimal solution, then also the other and the optimal target function values correspond (thus, $z_P = z_D$).*

Proof. There are various, different proofs, including one directly over the Simplex method. It suffices to prove one of the two statements.

Let (D) have a finite optimal solution y^*, so $b^ty^* = z_D$. By Theorem 4.1 (3), we have $M_p \neq \emptyset$ (b^ty is bounded by below). Because there is no admissible solution for (D) with target function value $< z_D$, the system $y^tb < z_D$, $A^ty \geq c, y >\geq 0$ is not solvable. Now consider P_{z_D} with $P_z = \{x \mid x \leq c^tx, Ax \leq b, x \geq 0\}$. We can conclude $P_{z_D} \neq \emptyset$ (represented in Lemma 4.5), thus $\exists x^*$ with $z_D \leq c^tx^*$, $Ax^* \leq b, x^* \geq 0$ with $x^* \in M_P$. Now we apply the weak duality theorem (Theorem 4.1), so $c^tx^* \leq b^ty^*$ where $b^ty^* = z_D$

$$\implies c^tx^* = b^ty^* = z_D$$

hence x^* is optimal for the primal problem and y^* is optimal for the dual problem. □

Proof of Lemma 4.5. $M_P \neq \emptyset$ (equivalently $\exists x : Ax = b, x \geq 0$) iff $y^tb < 0, A^ty \geq 0, y \geq 0$ has no solution (remember this as "the criterion").

$$P_z = \emptyset \iff \begin{array}{l} (y_0, y^t)\binom{-z}{y} < 0 \\ (y_0, y)^t\binom{-c^t}{A} \geq 0 \end{array} \text{ is solvable}$$

A solution with $y_0 = 0$ is not possible due to "the criterion". So $y_0 > 0$. We can substitute $y_0 \geq 0$ by $y_0 = 1$ (because the set of solutions creates a cone) without restricting the set of solutions. Thus

$$P_z = \emptyset \iff y^t b < z, y^t A \geq c^t, y \geq 0 (A^t y \geq c) \text{ solvable}$$

□

Remark (P has no admissible solution, so $y^t A \geq 0$ has a solution)**.** *It is trivial to see that the alternative theorem 4.3 can be generalized to a combination of equations and inequalities.*

Theorem 4.7 (Theorem of complementary slackness)**.** *Let $x \in M_p$ and $y \in M_D$. Then x is optimal for the primal problem and y is optimal for the dual problem if and only if (1) $x^t(A^t y - c) = 0$ and (2) $y^t(Ax - b) = 0$. Then (x, y) is called* optimal pair*.*

Remark (Interpretation of $x^t(A^t y - c) = 0$ and $y^t(Ax - b) = 0$)**.**

- *$x_i = 0$ (i-th primal variable is 0) or $(A^t y - c)_i = 0$ (the i-th restriction of the dual problem is fulfilled with equality) (corresponds to the i-th slack variable in the dual problem).*
- *$y_i = 0$ (i-th primal variable is 0) or $(Ax - b)_i = 0$ (the i-th primal restriction is fulfilled with equality).*

This follows because

$$x \geq 0 \qquad (A^t y - c) \geq 0 \textit{ because } A^t y \geq c$$

hence every summand must be contained in $x^t(A^t y - c) \geq 0$ and thus equals zero.

$y \geq 0$ $(A^t x - b) \leq 0$, thus every summand is in $y^t(Ax - b) \leq 0$ und thus equals zero.

Remark. *The interpretation above plays a vital role in optimality tests (examples will be provided in the practicals). Also they can be used to derive an optimal solution of the primal problem given the solution of the dual problem; vice versa.*

Proof of Theorem 4.7.

- Let (x, y) be an optimal pair, thus $Ax \leq b$, $x \geq 0$, $A^t y \geq c$, $y \geq 0$, $c^t x = b^t y$.

$$(Ax)^t y \leq b^t y = c^t x \implies \underbrace{x^t}_{\geq 0} \underbrace{(A^t y - c)}_{\geq 0} \leq 0 \implies x^t(A^t y - c) = 0$$

Analogously, $(A^t y)^t x \geq c^t x = b^t y$,

$$\underbrace{y^t}_{\geq 0} \underbrace{(Ax - b)}_{\leq 0} \geq 0 \qquad \to \qquad y^t(Ax - b) = 0$$

- If $x \in M_P$ and $y \in M_D$ satisfy $x^t(A^t y - c) = 0$ and $y^t(Ax - b) = 0$, then

$$x^t(A^t y - c) = y^t(Ax - b) = 0 \implies x^t A^t y - \underbrace{x^t c}_{c^t x} = y^t A x - \underbrace{y^t b}_{b^t y}$$

  hence $c^t x = b^t y$ so (x, y) is an optimal pair.

□

Remark. *We can even show sometimes more restrictive: There always exists an optimal pair (x^*, y^*) that is strictly complementary, thus*

$$\begin{aligned} x_j^* = 0 &\iff (A^t y^* - c)_j > 0 \qquad j = 1, \ldots, n \\ y_i^* = 0 &\iff (Ax - b)_i < 0 \qquad i = 1, \ldots, m \end{aligned}$$

Attention! *Not all optimal pairs (x, y) are strictly complementary. (A proof is not given at this point)*

4.4 The dual Simplex method

In contrast to the primal problem, we now consider ...

$$(P)\ \max\left\{c^t x \mid Ax \le b, x \ge 0\right\} \qquad (D)\ \min\left\{b^t y \mid A^t y \ge c, y \ge 0\right\}$$

$$(D')\ \max\left\{-b^t y \mid -A^t y \le -c, y \ge 0\right\}$$

This gives an equivalent problem to (D) in canonical form.

Idea: Solve (D') with the primal Simplex method and we consider this as a method to solve the original problem (*dual Simplex method*, Lemke, 1954)

Again, we use the tableau form. A Simplex tableau T is called *dual admissible*, if

Index