Material:

- Script Nr. 527
- Sheets of the course (on internet)

Use it as a rough guide; for motivation, geometric illustration, and proofs join the courses.
Chapter 1: Real vector spaces  linear spaces, inner products, differentiable functions. By “Self-instruction”

Chapter 2: Linear equations, - inequalities
Gaussian elimination, least square approximation, Fourier-Motzkin algorithm, Farkas lemma

Chapter 3: Linear programs
primal-dual linear programs, optimality conditions, matrix games

Chapter 4: Convex analysis
separating hyperplanes, convex sets, convex functions, differential theory

Chapter 5: Unconstrained optimization
optimality conditions, minimizing convex functions, descent methods, conjugate direction method, line search, Newton’s method, Gauss-Newton method, Quasi-Newton methods, minimization of nondifferentiable functions
We start with some definitions:

**Definitions in matrix theory**

- $M = (m_{ij})$ is said to be
  - *lower triangular*: if $m_{ij} = 0$ for $i < j$,
  - *upper triangular*: if $m_{ij} = 0$ for $i > j$.

- $P = (p_{ij}) \in \mathbb{R}^{m \times m}$ is a *permutation matrix*
  - if $p_{ij} \in \{0, 1\}$ and each row and each column of $P$ contains exactly one coefficient 1.

**Note that** $P^T P = I$, implying $P^{-1} = P^T$ for the inverse $P^{-1}$ of $P$. 

\[ Q \in \mathbb{R}^{n \times n}, \text{ } Q \text{ symmetric, is called positive semi-definite (not. } Q \geq 0) \text{ if:} \]

\[ x^T Q x \geq 0 \text{ for all } x \in \mathbb{R}^n, \]

positive definite (not. } Q > 0) if:

\[ x^T Q x > 0 \text{ for all } x \in \mathbb{R}^n, \ x \neq 0. \]

---

2.1 Gauss-elimination (for solving \( Ax = b \))

Motivation: We show by a simple example that “successive elimination” is equivalent with Gauss-algorithm.
General Idea: To eliminate $x_1, x_2 \ldots$ is equivalent with transforming $Ax = b$ or $(A | b)$ to “triangular” normal form $(\tilde{A} | \tilde{b})$ (with same solution set). Then solve $\tilde{A}x = \tilde{b}$, recursively:

$$
\begin{align*}
  a_{11} & \quad a_{12} & \ldots & \quad a_{1n} & \quad | & \quad b_1 \\
  a_{21} & \quad a_{22} & \ldots & \quad a_{2n} & \quad | & \quad b_2 \\
  \vdots & \quad \vdots & \ddots & \quad \vdots & \quad | & \quad \vdots \\
  a_{m1} & \quad a_{m2} & \ldots & \quad a_{mn} & \quad | & \quad b_m
\end{align*}
$$

Transformation into form $(\tilde{A} | \tilde{b})$:

$$
\begin{align*}
  \tilde{a}_{1j_1} & \quad \tilde{a}_{1j_2} & \ldots & \quad \tilde{a}_{1j_{r-1}} x_{j_{r-1}} & \ldots & \quad \tilde{a}_{1j_r} & \quad | & \quad \tilde{b}_1 \\
  \tilde{a}_{2j_2} & \quad \tilde{a}_{2j_2} & \ldots & \quad \tilde{a}_{2j_{r-1}} & \quad \tilde{a}_{2j_r} & \quad | & \quad \tilde{b}_2 \\
  \vdots & \quad \vdots & \ddots & \quad \vdots & \quad \vdots & \quad | & \quad \vdots \\
  \tilde{a}_{r-1j_{r-1}} & \quad \tilde{a}_{r-1j_r} & \ldots & \quad \tilde{a}_{r-1j_r} & \quad | & \quad \tilde{b}_{r-1} \\
  \tilde{a}_{rj_r} & \quad \tilde{a}_{rj_r} & \ldots & \quad \tilde{a}_{rj_r} & \quad | & \quad \tilde{b}_r \\
  \vdots & \quad \vdots & \ddots & \quad \vdots & \quad \vdots & \quad | \quad \vdots \\
  \tilde{a}_{rj_1} & \quad \tilde{a}_{rj_1} & \ldots & \quad \tilde{a}_{rj_1} & \quad | & \quad \tilde{b}_m
\end{align*}
$$
This “Gauss elimination” uses 2 types of row operations:

**(G1)** \((i, j)\)-pivot: for \(k > i\),

\[ \text{add } \lambda \times \text{row } i \text{ to row } k \]; \quad \text{with } \lambda = -\frac{a_{kj}}{a_{ij}} \]

**(G2)** interchange row \(i\) with row \(k\)

The “matrix form” of these operations are:

**Ex.2.3** The matrix form of (G1): \((A | b) \rightarrow (\tilde{A} | \tilde{b})\)

is given by \((\tilde{A} | \tilde{b}) = M (A | b)\)

with a nonsingular lower triangular \(M \in \mathbb{R}^{m \times m}\)

**Ex.2.4** The matrix form of (G2): \((A | b) \rightarrow (\tilde{A} | \tilde{b})\)

is given by \((\tilde{A} | \tilde{b}) = P (A | b)\)

with a permutation matrix \(P \in \mathbb{R}^{m \times m}\)
Implications of the Gauss algorithm:

**Th. 2.1** For every $A \in \mathbb{R}^{m \times n}$, there exists an $(m \times m)$-permutation matrix $P$ and an invertible lower triangular matrix $M \in \mathbb{R}^{m \times m}$ such that

$$U = MPA$$
is upper triangular.

**Cor. 2.1** [LU-factorization]
For $A \in \mathbb{R}^{m \times n}$, there exists an $(m \times m)$-permutation matrix $P$, an invertible, lower triang. $L \in \mathbb{R}^{m \times m}$ and an upper triang. $U \in \mathbb{R}^{m \times n}$ such that $LU = PA$.

**Rem.:** Solve $Ax = b$ by using the decomposition $PA = LU$! (How?)
Cor. 2.2  [Gale’s Theorem]
Exactly one of the following statements is true:

(a) The system $Ax = b$ has a solution $x$.
(b) There exists $y \in \mathbb{R}^m$ such that: $y^T A = 0^T$ and $y^T b \neq 0$.

Remark: In “normal form” $A \rightarrow \tilde{A}$, the number $r$ gives dimension of the space spanned by the rows of $A$.
This equals the dimension of the space spanned by the columns of $A$. 
Note: “Gauss row operations” destroy symmetry. So we modify “Gauss” in order to maintain symmetry.

Perform row and “same” column-operations:

- use (G1’): \[ A \rightarrow MAM^T \]
- instead of (G2) use (G2’):
  - if \( a_{ii} = 0, \ a_{kk} \neq 0, \ k > i \):
    - interchange row \( i \) and row \( k \)
    - interchange col. \( i \) and col. \( k \)
  - if \( a_{ii} = 0, \ a_{kk} = 0 \forall k > i, \ a_{ki} \neq 0, \ k > i \):
    - add row \( k \) to row \( i \) and
    - add col. \( k \) to col. \( i \)

G2’ transforms: \[ A \rightarrow BAB^T \ (B \text{ non-singular}) \]
Note: By “symmetric Gauss” the solution set of $Ax = b$ is destroyed!!! But it is useful to get the following results.

**Implications of the symmetric Gauss algorithm**

**Th. 2.2.** Let $A \in \mathbb{R}^{n \times n}$ be symmetric. Then with some nonsingular $Q \in \mathbb{R}^{n \times n}$

$$QAQ^T = D = \text{diag}(d_1, \ldots, d_n)$$

Recall: A symmetric $Q \in \mathbb{R}^{n \times n}$ is called positive semi-definite (not. $Q \geq 0$) if:

$$x^TQx \geq 0 \quad \text{for all} \quad x \in \mathbb{R}^n.$$
Cor. 2.3. Let $A$ be symmetric, $Q$ nonsingular such that $QAQ^T = \text{diag}(d_1, \ldots, d_n)$. Then

(a) $A \geq 0 \iff d_i \geq 0$, $i = 1, \ldots, n$
(b) $A > 0 \iff d_i > 0$, $i = 1, \ldots, n$

**Implication:** The check $A \geq 0$ (positive semidefinite) can be done by the Gauss-algorithm (polynomial).

Cor. 2.4. Let $S \in \mathbb{R}^{n \times n}$ be symmetric. Then

(a) $S \geq 0 \iff S = AA^T$ with some $A \in \mathbb{R}^{n \times m}$
(b) $S > 0 \iff S = AA^T$ with some nonsingular $A$

**Complexity of Gauss algorithm**

For $a$, the number of “\(\pm, \cdot, /\) flop’s” (floating point operations) needed to solve $Ax = b$ with $A \in \mathbb{R}^{n \times n}$:

\[a \leq n^3\]
2.2. Orthogonal projection, Least Square

**Assumption:** $V$ is a linear vectorspace over $\mathbb{R}$ with inner product $\langle x|y \rangle$ and (induced) norm $\|x\| = \sqrt{\langle x|x \rangle}$.

**Minimization Problem:** Given $x \in V$, subspace $W \subset V$ find $\hat{x} \in W$ such that:

$$\|x - \hat{x}\| = \min_{y \in W} \|x - y\| \quad (2.13)$$

The vector $\hat{x}$ is called the *projection of x onto W*.

**L 2.1. (sufficient condition)** Assume $\hat{x} \in W$ is such that

$$\langle x - \hat{x}|w \rangle = 0 \; \forall w \in W.$$ 

Then $\hat{x}$ is unique solution of (2.13).

To solve (2.13) we “construct” a solution via L.2.1:
We construct a solution $\hat{x} \in W$ satisfying $
abla x - \hat{x} \in W$ as follows (assuming that $W$ has a basis $a_1, \ldots, a_m$, i.e., $W = \text{span} \{a_1, \ldots, a_m\}$): Write

$$\hat{x} := \sum_{i=1}^{m} z_i a_i$$

Then $\langle x - \hat{x} | w \rangle = 0$, $\forall w \in W$ is equivalent with

$$\langle x - \sum_{i=1}^{m} z_i a_i | a_j \rangle = 0, \quad j = 1, \ldots, m$$

or

$$\sum_{i=1}^{m} \langle a_i | a_j \rangle z_i = \langle x | a_j \rangle, \quad j = 1, \ldots, m$$

Defining the Gram-matrix $G := (\langle a_i | a_j \rangle)$, $b \in \mathbb{R}^m$, $b_j = \langle x | a_j \rangle$ this leads to the linear equation (for $z$)

$$(2.16) \quad Gz = b \quad \text{with solution} \; \hat{z} = G^{-1}b$$
Ex. The Gram-matrix is positive definite, thus non-singular (under our assumption) \textit{Proof!}

\underline{Special case 1: } \quad V = \mathbb{R}^n, \quad \langle x \mid y \rangle = x^T y \quad \text{and} \quad W = \text{span} \{a_1, \ldots, a_m\}. \quad \text{Then with } A := [a_1, \ldots, a_m] \text{ the projection of } x \text{ onto } W \text{ is given by}

\hat{x} = A(A^T A)^{-1} A^T x

\underline{Special case 2: } \quad V = \mathbb{R}^n, \quad \langle x \mid y \rangle = x^T y, \quad a_1, \ldots, a_m \in \mathbb{R}^n \text{ lin. independent and } W' = \{w \in \mathbb{R}^n \mid a_i^T w = 0, \ i = 1, \ldots, m\}. \quad \text{Then the projection of } x \text{ onto } W' \text{ is given by}

\hat{x}' = x - A(A^T A)^{-1} A^T x

\underline{Special case 3: } \quad W = \text{span} \{a_1, \ldots, a_m\} \text{ with } \{a_i\}, \text{ an orthonormal basis, i.e., } \langle a_i \mid a_j \rangle = 0, \ i \neq j; = 1, i = j). \quad \text{Then the projection of } x \text{ onto } W \text{ is given by}

\hat{x} = \sum_{j=1}^{m} z_j a_j \quad \text{with} \quad z_j = \langle a_j \mid x \rangle \quad \forall j \quad \text{“Fouriercoefficients”}.
Problem: Given $W = \text{span}\ \{a_1, \ldots, a_m\}$, find an orthogonal basis $W = \text{span}\ \{b_1, \ldots, b_m\}$ (i.e., $\langle b_i, b_j \rangle = 0$, $i \neq j$ ($>0$, $i = j$)).

Recall the Gram-Schmidt algorithm for solving this Problem: start with $b_1 := a_1$ and iterate

$$b_k = a_k - \sum_{i=1}^{k-1} \frac{\langle b_i, a_k \rangle}{\langle b_i, b_i \rangle} b_i$$

Gram-Schmidt in matrix form: With $W \subset V := \mathbb{R}^n$. Put

$$A = \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix}, \quad B = \begin{pmatrix} b_1^T \\ \vdots \\ b_m^T \end{pmatrix}.$$

Then the Gram-Schmidt-steps are equivalent with:

- add multiple of row $j < k$ to row $k$
- multiply row $k$ by scalar (in case of normalisation)
Matrix form of “Gram-Schmidt”: Given \( A \in \mathbb{R}^{m \times n} \), there is a decomposition

\[ B = LA \]

with lower triangular nonsingular matrix \( L (l_{ii} = 1) \) and the rows \( b_j \) of \( B \) are orthogonal, i.e. \( \langle b_i | b_j \rangle = 0, \; i \neq j \).

A corollary of this fact:

Prop. 2.1 (Hadamard’s inequality) Let \( A \in \mathbb{R}^{m \times n} \) with rows \( a_i^T \). Then

\[
0 \leq \det (AA^T) \leq \prod_{i=1}^{m} a_i^T a_i
\]
**Definition.** \( \lambda \in \mathbb{C} \) is an eigenvalue of \( A \in \mathbb{R}^{n \times n} \) if there is an (eigenvector) \( 0 \neq x \in \mathbb{C}^n \) with \( Ax = \lambda x \).

The results above (together with the Theorem of Weierstrass) allow a proof of:

**Th. 2.3** *(Spectral theorem for symmetric matrices)*

Let \( A \in \mathbb{R}^{n \times n} \) be symmetric. Then there exists an orthogonal matrix \( Q (Q^TQ = I) \) and eigenvalues \( \lambda_1, \ldots, \lambda_n \in \mathbb{R} \) such that

\[
Q^T AQ = D = \text{diag} (\lambda_1, \ldots, \lambda_n)
\]
2.3 Integer Solutions of Linear Equations \((x_j \in \mathbb{Z})\)

Example  The equation \(3x_1 - 2x_2 = 1\) has a solution \(x = (1, 1) \in \mathbb{Z}^2\). But the equation \(6x_1 - 2x_2 = 1\) does not allow an entire solution \(x\).

Key remark: Let \(a_1, a_2 \in \mathbb{Z}\) and let \(a_1 x_1 + a_2 x_2 = b\) have a solution \(x_1, x_2 \in \mathbb{Z}\). Then \(b = \lambda c\) with \(\lambda \in \mathbb{Z}\), \(c = \text{gcd} (a_1, a_2)\)

Here: \(\text{gcd} (a_1, a_2)\) denotes the greatest common divisor of \(a_1, a_2\).

Lem.2.2 [Euclid’s Algorithm] Let \(c = \text{gcd} (a_1, a_2)\). Then

\[ L(a_1, a_2) := \{a_1 \lambda_1 + a_2 \lambda_2 \mid \lambda_1, \lambda_2 \in \mathbb{Z}\} = \{c \lambda \mid \lambda \in \mathbb{Z}\} =: L(c). \]

(The proof of) this result allows to

“solve \(a_1 x_1 + a_2 x_2 = b\) (in \(\mathbb{Z}\))."
Algorithm to solve, \( a_1 x_1 + a_2 x_2 = b \) (in \( \mathbb{Z} \))

- Compute \( c = \gcd (a_1, a_2) \). If \( \lambda := b/c \notin \mathbb{Z} \), no entire solution exists.
- If \( \lambda := b/c \in \mathbb{Z} \), compute solutions \( \lambda_1, \lambda_2 \in \mathbb{Z} \) of \( \lambda_1 a_1 + \lambda_2 a_2 = c \). Then
  \[
  (\lambda_1 \lambda)a_1 + (\lambda_2 \lambda)a_2 = b.
  \]

General problem: Given \( a_1, \ldots, a_n, b \in \mathbb{Z}^m \), find \( x = (x_1, \ldots, x_n) \in \mathbb{Z}^n \) such that

\[
(\star) \quad a_1 x_1 + a_2 x_2 + \ldots + a_n x_n = b \quad \text{or} \quad Ax = b
\]

where \( A := [a_1, \ldots, a_n] \).

Def. We introduce the lattice generated by \( a_1, \ldots, a_n \),

\[
L = L(a_1, \ldots, a_n) = \left\{ \sum_{j=1}^n a_j \lambda_j \mid \lambda_j \in \mathbb{Z} \right\} \subseteq \mathbb{R}^m.
\]
Assumption 1: \( \text{rank } A = m \) \((m \leq n)\); \(wlog., a_1, \ldots, a_m\) are linearly independent.

To solve the problem: Find \( C = [c_1 \ldots c_m] \in \mathbb{Z}^{m \times m} \) such that

\[
(\ast \ast) \quad L(c_1, \ldots, c_m) = L(a_1, \ldots, a_n).
\]

Then \((\ast)\) has a solution \( x \in \mathbb{Z}^n \) iff \( \lambda := C^{-1}b \in \mathbb{Z}^n \).

Bad news: As in the case of one equation: in general

\[
L(a_1, \ldots, a_m) \neq L(a_1, \ldots, a_n).
\]

Lem.2.3 Let \( c_1, \ldots, c_m \in L(a_1, \ldots, a_n) \). Then

\[
L(c_1, \ldots, c_m) = L(a_1, \ldots, a_n) \text{ iff for all } j = 1, \ldots, n, \text{ the system}
\]

\[
C\lambda = a_j \quad \text{has an integral solution.}
\]

Last step: Find such \( c_i \)'s
Main Result: The algorithm

Lattice Basis

INIT: \( \mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_m] = [\mathbf{a}_1, \ldots, \mathbf{a}_m] \);
ITER: Compute \( \mathbf{C}^{-1} \);

If \( \mathbf{C}^{-1}\mathbf{a}_j \in \mathbb{Z}^m \) for \( j = 1, \ldots, n \), then STOP;
If \( \lambda = \mathbf{C}^{-1}\mathbf{a}_j \notin \mathbb{Z}^m \) for some \( j \), then

Let \( \mathbf{a}_j = \mathbf{C}\lambda = \sum_{i=1}^{m} \lambda_i \mathbf{c}_i \) and compute
\( \mathbf{c} = \sum_{i=1}^{m} (\lambda_i - [\lambda_i])\mathbf{c}_i = \mathbf{a}_j - \sum_{i=1}^{m} [\lambda_i]\mathbf{c}_i \);
Let \( k \) be the largest index \( i \) such that \( \lambda_i \notin \mathbb{Z} \);
Update \( \mathbf{C} \) by replacing \( \mathbf{c}_k \) with \( \mathbf{c} \) in column \( k \);

NEXT ITERATION stops after at most \( K = \log_2(\text{det}[\mathbf{a}_1, \ldots, \mathbf{a}_m]) \) steps with a matrix \( \mathbf{C} \) satisfying (**).
Th. 2.4  Let $A \in \mathbb{Z}^{m \times n}$ and $b \in \mathbb{Z}^m$ be given. Then exactly one of the following statements is true:

(a) There exists some $x \in \mathbb{Z}^n$ such that $Ax = b$.

(b) There exists some $y \in \mathbb{R}^m$ such that $y^T A \in \mathbb{Z}^n$ and $y^T b \notin \mathbb{Z}$. 
2.4 Linear Inequalities $Ax \leq b$

Fourier-Motzkin algorithm for solving $Ax \leq b$. Eliminate $x_1$:

$$a_{r1}x_1 + \sum_{j=2}^{n} a_{rj}x_j \leq b_r \quad r = 1, \ldots, k$$

$$a_{s1}x_1 + \sum_{j=2}^{n} a_{sj}x_j \leq b_s \quad s = k + 1, \ldots, \ell$$

$$\sum_{j=2}^{n} a_{tj}x_j \leq b_t \quad t = \ell + 1, \ldots, m$$

with $a_{r1} > 0$, $a_{s1} < 0$. Divide by $a_{r1}$, $|a_{s1}|$, giving (for $r$ and $s$)

$$x_1 + \sum_{j=2}^{n} a'_{rj}x_j \leq b'_r \quad r = 1, \ldots, k$$

$$-x_1 + \sum_{j=2}^{n} a'_{sj}x_j \leq b'_s \quad s = k + 1, \ldots, \ell$$
So $Ax \leq b$ has a solution $x = (x_1, \ldots, x_n)$ if and only if there is a solution $x' = (x_2, \ldots, x_n)$ of

$$
\sum_{j=2}^{n} (a'_{sj} + a'_{rj})x_j \leq b'_r + b'_s \quad r = 1, \ldots, k; \quad s = k + 1 \ldots, \ell
$$

$$
\sum_{j=2}^{n} a_{tj}x_j \leq b_t \quad t = \ell + 1, \ldots, m.
$$

In matrix form: $Ax \leq b$ has a solution $x = (x_1, \ldots, x_n)$ if and only if there is a solution of the transformed system:

$$
A'x' \leq b' \quad \text{or} \quad (0 \ A')x \leq b'
$$

Remark: Any row of $(0 \ A'|b')$ is a positive combination of rows of $(A|b)$:

any row is of the form $y^T(A|b)$, $y \geq 0$
By eliminating $x_1, x_2, \ldots, x_n$ in this way we finally obtain an “equivalent” system

$$\tilde{A}(n)x \leq \tilde{b} \quad \text{where} \quad \tilde{A}(n) = 0$$

which is (recursively) solvable iff $0 \leq \tilde{b}_i, \forall i$.

**Th. 2.5 [Projection Theorem]** Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$. Then all for $k = 1, \ldots, n$, the projection

$$P^{(k)} = \{(x_{k+1}, \ldots, x_n) \mid (x_1, \ldots, x_k, x_{k+1}, \ldots, x_n) \in P$$

for suitable $x_1, \ldots, x_k \in \mathbb{R}\}

is the solution set of a linear system $A^{(k)}x^{(k)} \leq b^{(k)}$ in $n - k$ variables $x^{(k)} = (x_{k+1}, \ldots, x_n)$.

In principle: Linear inequalities can be solved by FM. However this might be inefficient! (Why?)
2.4.1. Solvability of linear systems

We consider so-called Farkas lemmata. They are the basis of optimality and duality results in LP.

**Th. 2.6  [Lemma of Farkas]** Exactly one of the following statements is true:

(I) \( Ax \leq b \) has a solution \( x \in \mathbb{R}^n \).

(II) There exists \( y \in \mathbb{R}^m \) such that 

\[
y^T A = 0^T , \quad y^T b < 0 \quad \text{and} \quad y \geq 0.
\]

**Ex. 2.24  (more general)** Let \( A \in \mathbb{R}^{m \times n} \), \( C \in \mathbb{R}^{k \times n} \), \( b \in \mathbb{R}^m \), \( c \in \mathbb{R}^k \). Then precisely one of the alternatives is valid.

(I) There is a solution \( x \) of: \( Ax \leq b \), \( Cx = c \)

(II) There is a solution \( \mu \in \mathbb{R}^m \), \( \mu \geq 0 \), \( \lambda \in \mathbb{R}^k \) of :

\[
\begin{pmatrix}
A^T \\
b^T
\end{pmatrix} \mu + \begin{pmatrix}
C^T \\
c^T
\end{pmatrix} \lambda = \begin{pmatrix}
0 \\
-1
\end{pmatrix}
\]
Cor.2.5 [Gordan] Given $A \in \mathbb{R}^{m \times n}$, exactly one of the following alternatives is true:

(I) $Ax = 0, x \geq 0$ has a solution $x \neq 0$.

(II) $y^T A < 0^T$ has a solution $y$.

Remark: As we shall see in Chapter 3, the Farkas Lemma in the following form is the strong duality of LP in disguise.

Cor.2.6 [Farkas, implied inequalities] Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, $z \in \mathbb{R}$. Assume that $Ax \leq b$ is feasible. Then the following are equivalent:

(a) $Ax \leq b \Rightarrow c^T x \leq z$

(b) $y^T A = c^T$, $y^T b \leq z$, $y \geq 0$ has a solution $y$. 
Application: Markov chains \((Existence \ of \ a \ steady \ state)\)

Def. A vector \(\pi = (\pi_1, \ldots, \pi_n)\) with \(\pi_i \geq 0, \sum_i \pi_i = 1\) is called a \textit{probability distribution} on \(\{1, \ldots, n\}\).

A matrix \(P = (p_{ij})\) where each row \(P_i\) is a probability distribution is called a \textit{stochastic matrix}.

In a stochastic process:
- \(\pi_i\%\) of population is in state \(i\)
- \(p_{ij}\) is probability of transition from state \(i \to j\)
- So the transition step \(k \to k + 1\) is: \(\pi^{(k+1)} = P^T \pi^{(k)}\)

Probability distribution \(\pi\) is called \textit{steady state} if \(\pi = P^T \pi\)

As a corollary of Gordan’s result:
Each stochastic matrix \(P\) has a steady state \(\pi\).
Given $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$.

\[
\text{LP}_p : \quad \max_{x \in \mathbb{R}^n} c^T x \quad \text{s.t.} \quad Ax \leq b
\]

\[
\text{LP}_d : \quad \min_{y \in \mathbb{R}^m} b^T y \quad \text{s.t.} \quad A^T y = c, \quad y \geq 0,
\]

is the pair of primal and dual programs.

Notation.

- $F_p = \{ x \mid Ax \leq b \}$ feasible set of $\text{LP}_p$
- $F_d = \{ y \mid A^T y = c, \ y \geq 0 \}$ feas. set of $\text{LP}_d$
- $z^*_p := \max_{x \in F_p} c^T x$ max. value of $\text{LP}_p$
- $z^*_d := \min_{y \in F_d} b^T y$ min. value of $\text{LP}_d$
- $\bar{x} \in F_p$ is optimal (maximizer of $\text{LP}_p$) if $c^T \bar{x} = z^*_p$.  

Math. Prog., Chapter 2, 3
Weak duality is easy to prove:

L.3.1 (Weak Duality) Let $Ax \leq b$, $A^T y = c$, $y \geq 0$. Then,

$$c^T x \leq b^T y \quad \text{and thus} \quad z_p^* \leq z_d^*$$

If we have $c^T x = b^T y$, then $x$, $y$ are optimal solutions of $LP_p$, $LP_d$ resp.

Strong duality is a direct consequence of Farkas’ lemma:

Th.3.1 (Strong Duality)

If either $LP_p$ or $LP_d$ is feasible then: $z_p^* = z_d^*$

If both are feasible, then optimal solutions $x$ and $y$ of $LP_p$ and $LP_d$ exists (satisfying $c^T x = b^T y$).
Th.3.1 implies: x and y are optimal solutions of LP$_p$ and LP$_d$, resp., if and only if they solve the system (of lin.=, $\leq$)

\[
\begin{align*}
Ax & \leq b \\
A^T y & = c \\
c^T x - b^T y & = 0 \\
y & \geq 0.
\end{align*}
\]

Note that: for $x$, $y$ satisfying ($\ast$) we have

\[
b^T y - c^T x = y^T (b - Ax) = 0
\]

The relation

\[
y^T (b - Ax) = 0
\]

is called *complementarity condition*. 

Math. Prog., Chapter 2,3
Cor. : (Optimality conditions)

Let $x \in F_p$: then $x$ solves $\text{LP}_p \iff$

there ex. $y \in F_d$ such that $y^T(b - Ax) = 0$

Let $y \in F_d$: Then $y$ solves $\text{LP}_d \iff$

there ex. $x \in F_p$ such that $y^T(b - Ax) = 0$
3.1.2 Equivalent LP’s

LP’s in other forms can be transformed into the given “standard forms” For example: The program:

\[
\max_{x \in \mathbb{R}^n} \quad c^T x \quad \text{s.t.} \quad Ax \leq b, \quad x \geq 0
\]

has the dual:

\[
\min_{y \in \mathbb{R}^n} \quad b^T y \quad \text{s.t.} \quad A^T y \geq c, \quad y \geq 0
\]

Rules for primal dual pairs:

<table>
<thead>
<tr>
<th>Primal Problem</th>
<th>Dual Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>max</td>
<td>min</td>
</tr>
<tr>
<td>free variable</td>
<td>equality constraint</td>
</tr>
<tr>
<td>non-negative variable</td>
<td>$\geq$ constraint</td>
</tr>
<tr>
<td>equality constraint</td>
<td>free variable</td>
</tr>
<tr>
<td>$\leq$ constraint</td>
<td>non-negative variable</td>
</tr>
</tbody>
</table>
3.1.3. Shadow prices

Production model: (n products, m resources)

\[ c_j \quad \text{prices per unit for product } j \quad \rightarrow \quad c \]
\[ b_i \quad \text{bounds for resource } R_i \quad \rightarrow \quad b \]
\[ a_{ij} \quad \text{units of resource } R_i \text{ needed for unit of product } j \quad \rightarrow \quad A \]
\[ x_j \quad \text{production of product } j \quad \rightarrow \quad x \]

**primal program:** \((\max \text{ profit } c^T x)\)

\[
\max_{x \in \mathbb{R}^n} \quad c^T x \quad \text{s.t.} \quad Ax \leq b, \quad x \geq 0
\]

**optimal solution:** \(\bar{x} \quad \rightarrow \quad \text{profit } \bar{z} = c^T \bar{x}\)

**dual:** \(\min_{y \in \mathbb{R}^n} \quad b^T y \quad \text{s.t.} \quad A^T y \geq c, \quad y \geq 0\)

**optimal solution:** \(\bar{y} \quad \rightarrow \quad \text{value } \bar{z} = b^T \bar{y}\)
Question: Can we obtain a higher profit if we buy additional amount of resource \( b_{i_0} \)?

If we change: \( b_{i_0} \rightarrow \tilde{b}_{i_0} := b_{i_0} + t, \ (t > 0) \) then \( \tilde{z} \rightarrow \bar{z} + ?? \)

We find: \[ \tilde{z} \leq \bar{z} + t \cdot \bar{y}_{i_0} \] \((\bar{y}_{i_0} \text{ is shadow price})\)

Answer: Yes, if the price per unit for \( R_{i_0} \) is smaller than \( \bar{y}_{i_0} \) (shadow price).
MG is an example of a non-cooperative game with 2 players (using a pure or a mixed strategy).

Given $A \in \mathbb{R}^{m \times n}$ and row-players R and (column)-player C

Game with pure strategy

- R chooses row $i$: if $a_{ij} > 0$ R wins $a_{ij}$
- C chooses col. $j$: if $a_{ij} < 0$ C wins $|a_{ij}|$

For this pure strategy game a so-called Nash equilibrium need not exist.
**Game** with mixed strategies: \( x \in \mathbb{R}^m, \ y \in \mathbb{R}^n \)

- R chooses row \( i \) with probability \( x_i \); \( x_i \geq 0, \sum_i x_i = 1 \)
- C chooses col. \( j \) with probability \( y_j \); \( y_j \geq 0, \sum_j y_j = 1 \)

The expected gain for R (loss for C):

\[
x^T A y = \sum_i x_i \left( \sum_j a_{ij} y_j \right) = \sum_j y_j \left( \sum_i a_{ij} x_i \right)
\]

Strategies:

- given \( \bar{y} \): R plays \( \bar{x} \) as solution of:
  \[
  \max_x x^T A \bar{y} = \max_i \sum_j a_{ij} \bar{y}_j
  \]

- given \( \bar{x} \): C plays \( \bar{y} \) as solution of:
  \[
  \min_y \bar{x}^T A y = \min_j \sum_i a_{ij} \bar{x}_i
  \]

Math. Prog., Chapter 2, 3
Best strategy against best of opponent:

for R: \[ \max_x \min_y x^T Ay = \max_x \min_j \sum_i a_{ij} x_i \]

for C: \[ \min_y \max_x x^T Ay = \min_y \max_i \sum_j a_{ij} y_j \]

Th.3.2 [\text{minmax-theorem}] There exist feasible \( \overline{x}, \overline{y} \) such that
\[ \min_y \overline{x}^T Ay = \max_x \overline{x}^T Ay \]

This implies:
\[ \max_x \min_y x^T Ay = \min_y \max_x x^T Ay = \overline{x}^T \overline{y} \]
\( \overline{x}, \overline{y} \) represent a Nash equilibrium of the mixed strategy matrix game.

Def. A game is fair if \( \overline{z} = \overline{w} = \overline{x}^T \overline{y} = 0 \) holds.
Methods for linear programs

1. Simplex method

\[ \text{LP}_p: \max_{x \in \mathbb{R}^n} c^T x \text{ s.t. } a_i^T x \leq b_i, \ i = 1, \ldots, m. \]

proceeds ’from vertex to vertex’ of the feasible set \( F_p \) until we have found a vertex \( x \) such that (with suitable \( y \)) the sufficient optimality condition holds:

\[ A^T y = c, \ y \geq 0, \ y^T (b - Ax) = 0 \]
2. Interior point method: Consider the system of equations

\[ P(t) : \begin{align*}
Ax + s &= b \\
A^Ty &= c \\
y_i(b - Ax)_i &= t > 0 \quad \forall i
\end{align*} \]

with \( y, (b - Ax) > 0 \). Here \( t > 0 \) is a parameter.

Idea: Compute (by ’Newton’) solutions \( x(t), y(t), s(t), t > 0 \) of \( P(t) \). Then for \( t \downarrow 0 \) (hopefully)

\[ x(t), y(t), s(t) \longrightarrow x, y, s \]

With solutions \( x, y \) of the primal-dual pair of LP’s

Remark: The “worst case behavior” of the Simplex algorithm is not “polynomial”.

The interior point method can be implemented as a “polynomial” algorithm for solving LP.