Mathematical Programming I,
Sheets for the course, version: 04-04-2012
Georg Still

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 triangluar: **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**.**

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

$$\mathbf{x}^T Q\mathbf{x} \geq 0 \qquad \text{for all } \mathbf{x} \in \mathbb{R}^n,$$

positive definite **(not.** $Q > 0$**) if:**

$$\mathbf{x}^T Q\mathbf{x} > 0 \qquad \text{for all } \mathbf{x} \in \mathbb{R}^n, \ \mathbf{x} \neq \mathbf{0}\ .$$

2.1 Gauss-elimination (for solving $\boldsymbol{Ax} = \boldsymbol{b}$)

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 $A\mathbf{x} = \mathbf{b}$ or $(A \mid \mathbf{b})$ to "triangular" normal form $(\tilde{A} \mid \tilde{b})$ (with same solution set). Then solve $\tilde{A}x = \tilde{b}$, recursively:

$$\begin{array}{cccc|c} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & & & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{array}$$

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

$$\begin{array}{cccccccc|c} \underline{\tilde{a}_{1j_1}} & .. & \tilde{a}_{1j_2} & .. & \tilde{a}_{1j_{r-1}} x_{j_{r-1}} & .. & \tilde{a}_{1j_r} & .. & \tilde{b}_1 \\ & & \underline{\tilde{a}_{2j_2}} & .. & \tilde{a}_{2j_{r-1}} & .. & \tilde{a}_{2j_r} & .. & \tilde{b}_2 \\ & & & \ddots & & & & & \vdots \\ & & & & \underline{\tilde{a}_{r-1j_{r-1}}} & .. & \tilde{a}_{r-1j_r} & .. & \tilde{b}_{r-1} \\ & & & & & & \underline{\tilde{a}_{rj_r}} & .. & \tilde{b}_r \\ & & & & & & & & \vdots \quad \tilde{b}_m \end{array}$$

This "Gauss elimination" uses 2 types of row operations:

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

add $\lambda \times$ **row** i **to row** k ; **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):** $(\mathbf{A} \mid \mathbf{b}) \to (\tilde{\mathbf{A}} \mid \tilde{\mathbf{b}})$

is given by $(\tilde{\mathbf{A}} \mid \tilde{\mathbf{b}}) = \mathbf{M}\, (\mathbf{A} \mid \mathbf{b})$

with a nonsingular lower triangular $\mathbf{M} \in \mathbb{R}^{m \times m}$

Ex.2.4 **The matrix form of (G2):** $(\mathbf{A} \mid \mathbf{b}) \to (\tilde{\mathbf{A}} \mid \tilde{\mathbf{b}})$

is given by $(\tilde{\mathbf{A}} \mid \tilde{\mathbf{b}}) = \mathbf{P}\, (\mathbf{A} \mid \mathbf{b})$

with a permutation matrix $\mathbf{P} \in \mathbb{R}^{m \times m}$

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

$$\mathbf{U} = \mathbf{MPA} \quad \text{is upper triangular.}$$

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 $A\mathbf{x} = \mathbf{b}$ has a solution x.**

(b) **There exists $\mathbf{y} \in \mathbb{R}^m$ such that: $\mathbf{y}^T A = \mathbf{0}^T$ and $\mathbf{y}^T \mathbf{b} \neq 0$.**

Remark: **In "normal form" $A \to \widetilde{A}$, 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.

2.1.3 "Gauss-Algoritm" for symmmetric **A**

Note: *"Gauss row operations" destroy symmetry. So we modify "Gauss" in order to maintain symmetry.*

Perform row and "same" column-operations:

- **use (G1'):** $A \to MAM^T$
- **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: $A \to BAB^T$ **(**B **nonssingular)**

Note: *By "symmetric Gauss" the solution set of* $\boldsymbol{Ax} = \boldsymbol{b}$ *is destroyed!!!* *But it is useful to get the followig results.*

Implications of the symmetric Gauss algorithm

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

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

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

$$\mathbf{x}^T\mathbf{Qx} \geq 0 \qquad \textbf{for all } \mathbf{x} \in \mathbb{R}^n.$$

Cor. 2.3. **Let A be symmetric, Q nonsingular such that** $\mathbf{QAQ}^T = diag(d_1, \ldots, d_n)$**. Then**

(a) $\mathbf{A} \geq 0 \quad \Leftrightarrow \quad d_i \geq 0, \; i = 1, \ldots, n$

(b) $\mathbf{A} > 0 \quad \Leftrightarrow \quad d_i > 0, \; i = 1, \ldots, n$

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

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

(a) $\mathbf{S} \geq 0 \quad \Leftrightarrow \quad \mathbf{S} = \mathbf{AA}^T$ **with some** $\mathbf{A} \in \mathbb{R}^{n \times m}$

(b) $\mathbf{S} > 0 \quad \Leftrightarrow \quad \mathbf{S} = \mathbf{AA}^T$ **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** $A \in \mathbb{R}^{n \times n}$**:**

$$\boxed{a \leq n^3}$$

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

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

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

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

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

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

To solve (2.13) we "construct" a solution via L.2.1:

<u>We construct a solution</u> $\hat{\mathbf{x}} \in W$ satisfying $\langle \mathbf{x} - \hat{\mathbf{x}} | \mathbf{w} \rangle = 0 \;\; \forall \mathbf{w} \in W$ as follows *(assuming that W has a basis $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$, i.e., $\boldsymbol{W} = \mathrm{span}\,\{a_1, \ldots, a_m\}$)*: **Write**

$$\hat{\mathbf{x}} := \sum_{i=1}^{m} z_i \mathbf{a}_i$$

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

$$\langle \mathbf{x} - \sum_{i=1}^{m} z_i \mathbf{a}_i \mid \mathbf{a}_j \rangle = 0 \;, \qquad j = 1, \ldots, m$$

$$\text{or} \quad \sum_{i=1}^{m} \langle \mathbf{a}_i | \mathbf{a}_j \rangle z_i = \langle \mathbf{x} | \mathbf{a}_j \rangle \;, \qquad j = 1, \ldots, m$$

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

$$(2.16) \qquad G\mathbf{z} = \mathbf{b} \quad \text{with solution } \hat{z} = G^{-1}\mathbf{b}$$

Ex. The Gram-matrix is positive definite, thus non-singular (under our assumption) *Proof!*

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

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

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

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

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

$$\hat{\mathbf{x}} = \sum_{j=1}^{m} z_j a_j \quad \text{with} \quad z_j = \langle a_j|x\rangle\ \forall j \quad \text{“}Fouriercoefficients\text{''}.$$

Problem: Given $W = \text{span}\{a_1, \ldots, a_m\}$, find an orthogonal basis $W = \text{span}\{b_1, .., b_m\}$ (*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

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

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

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1^T \\ \ldots \\ \mathbf{a}_m^T \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \mathbf{b}_1^T \\ \ldots \\ \mathbf{b}_m^T \end{pmatrix}.$$

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 $\mathbf{A} \in \mathbb{R}^{m \times n}$, there is a decomposition

$$\mathbf{B} = \mathbf{L}\mathbf{A}$$

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

A corollary of this fact:

Prop. 2.1 (*Hadamard's inequality*) **Let** $\mathbf{A} \in \mathbb{R}^{m \times n}$ **with rows** $\mathbf{a}_i^T$. **Then**

$$0 \leq \det(\mathbf{A}\mathbf{A}^T) \leq \prod_{i=1}^{m} \mathbf{a}_i^T \mathbf{a}_i$$

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

$$Q^TAQ = D = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)$$

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

Key remark: **Let** $a_1, a_2 \in \mathbb{Z}$ **and let** $a_1x_1 + a_2x_2 = b$ **have a solution** $x_1, x_2 \in \mathbb{Z}$**. Then** $b = \lambda c$ *with* $\lambda \in \mathbb{Z},\ c =$ **gcd** (a_1, a_2)

Here: **gcd** (a_1, a_2) **denotes the greatest common divisor of** a_1, a_2**.**

Lem.2.2 *[Euclid's Algorithm]* **Let** $c =$ **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 $a_1x_1 + a_2x_2 = b$ **(in** $\mathbb{Z}$**)".**

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

- **Compute** $c = \textbf{gcd}\,(a_1, a_2)$**. 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** $\mathbf{a}_1, \ldots, \mathbf{a}_n, \mathbf{b} \in \mathbb{Z}^m$**, find** $\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{Z}^n$ **such that**

$$(\star) \quad \mathbf{a}_1 x_1 + \mathbf{a}_2 x_2 + \ldots + \mathbf{a}_n x_n = \mathbf{b} \quad \text{or} \quad \mathbf{A}\mathbf{x} = \mathbf{b}$$

where $\mathbf{A} := [\mathbf{a}_1, \ldots, \mathbf{a}_n]$**.**

Def. We introduce the *lattice* generated by $\mathbf{a}_1, \ldots, \mathbf{a}_n$**,**

$$L = L(\mathbf{a}_1, \ldots, \mathbf{a}_n) \ = \ \left\{ \sum_{j=1}^{n} \mathbf{a}_j \lambda_j \mid \lambda_j \in \mathbb{Z} \right\} \ \subseteq \mathbb{R}^m\ .$$

Assumption 1: rank $\mathbf{A} = m$ ($m \leq n$); *wlog.*, $\mathbf{a}_1, \ldots, \mathbf{a}_m$ are linearly independent.

To solve the problem: Find $\mathbf{C} = [\mathbf{c}_1 \ldots \mathbf{c}_m] \in \mathbb{Z}^{m \times m}$ such that

$$(\star\star) \qquad L(\mathbf{c}_1, \ldots, \mathbf{c}_m) \;=\; L(\mathbf{a}_1, \ldots, \mathbf{a}_n) \;.$$

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

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

$$L(\mathbf{a}_1, \ldots, \mathbf{a}_m) \;\neq\; L(\mathbf{a}_1, \ldots, \mathbf{a}_n) \;.$$

Lem.2.3 Let $\mathbf{c}_1, \ldots, \mathbf{c}_m \in L(\mathbf{a}_1, \ldots, \mathbf{a}_n)$. Then $L(\mathbf{c}_1, \ldots, \mathbf{c}_m) = L(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ iff for all $j = 1, \ldots, n$, the system

$$\mathbf{C}\lambda = \mathbf{a}_j \quad \text{has an integral solution.}$$

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 $\boldsymbol{\lambda} = \mathbf{C}^{-1}\mathbf{a}_j \notin \mathbb{Z}^m$ for some j, then
Let $\mathbf{a}_j = \mathbf{C}\boldsymbol{\lambda} = \sum_{i=1}^{m} \lambda_i \mathbf{c}_i$ 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 $K = \log_2(\det[\mathbf{a}_1, \ldots, \mathbf{a}_m])$ steps with a matrix C satisfying ($\star\star$).

Th. 2.4 **Let $\mathbf{A} \in \mathbb{Z}^{m \times n}$ and $\mathbf{b} \in \mathbb{Z}^m$ be given. Then exactly one of the following statements is true:**

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

(b) **There exists some $\mathbf{y} \in \mathbb{R}^m$ such that $\mathbf{y}^T\mathbf{A} \in \mathbb{Z}^n$ and $\mathbf{y}^T\mathbf{b} \notin \mathbb{Z}$.**

Fourier-Motzkin algorithm *for solving* $\boldsymbol{Ax} \leq \boldsymbol{b}$. **Eliminate** x_1**:**

$$
\begin{array}{rll}
\underline{a_{r1}} x_1 + & \sum_{j=2}^{n} a_{rj} x_j \leq b_r & r = 1, \ldots, k \\
\underline{a_{s1}} x_1 + & \sum_{j=2}^{n} a_{sj} x_j \leq b_s & s = k+1, \ldots, \ell \\
 & \sum_{j=2}^{n} a_{tj} x_j \leq b_t & t = \ell+1, \ldots, m
\end{array}
$$

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

$$
\begin{array}{rll}
x_1 + & \sum_{j=2}^{n} a'_{rj} x_j \leq b'_r & r = 1, \ldots, k \\
-x_1 + & \sum_{j=2}^{n} a'_{sj} x_j \leq b'_s & s = k+1, \ldots, \ell
\end{array}
$$

So $\mathbf{Ax} \le \mathbf{b}$ **has a solution** $\mathbf{x} = (x_1, .., x_n)$ **if and only if there is a solution** $\mathbf{x}' = (x_2, .., x_n)$ **of**

$$\begin{aligned} \sum_{j=2}^{n} (a'_{sj} + a'_{rj}) x_j &\le b'_r + b'_s \qquad r = 1, \dots, k;\ s = k+1 \dots, \ell \\ \sum_{j=2}^{n} a_{tj} x_j &\le b_t \qquad t = \ell + 1, \dots, m. \end{aligned}$$

In matrixform: $\mathbf{Ax} \le \mathbf{b}$ **has a solution** $\mathbf{x} = (x_1, .., x_n)$ **if and only if there is a solution of the transformed system:**

$$\mathbf{A}'\mathbf{x}' \le \mathbf{b}' \qquad \text{or} \qquad (\ \mathbf{0}\ \mathbf{A}'\ )\mathbf{x} \le \mathbf{b}'$$

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

$$\text{any row is of the form} \quad \mathbf{y}^T(\mathbf{A}|\mathbf{b}), \quad \mathbf{y} \ge \mathbf{0}$$

By eliminating $x_1, x_2, \ldots, x_n$ in this way we finally obtain an "equivalent" system

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

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

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

$$P^{(k)} = \{(x_{k+1}, .., x_n) \mid (x_1, .., x_k, x_{k+1}, .., x_n) \in P \text{ for suitable } x_1, ..., x_k \in \mathbb{R}\}$$

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

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) **$\mathbf{Ax} \leq \mathbf{b}$ has a solution $\mathbf{x} \in \mathbb{R}^n$.**

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

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

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

(I) **There is a solution x of: $\mathbf{Ax} \leq \mathbf{b}$, $\mathbf{Cx} = \mathbf{c}$**

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

$$\begin{pmatrix} \mathbf{A}^T \\ \mathbf{b}^T \end{pmatrix} \boldsymbol{\mu} + \begin{pmatrix} \mathbf{C}^T \\ \mathbf{c}^T \end{pmatrix} \boldsymbol{\lambda} = \begin{pmatrix} \mathbf{0} \\ -1 \end{pmatrix}$$

Cor.2.5 [*Gordan*] **Given $\mathbf{A} \in \mathbb{R}^{m \times n}$, exactly one of the following alternatives is true:**

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

(II) $\mathbf{y}^T\mathbf{A} < \mathbf{0}^T$ **has a solution $\mathbf{y}$.**

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 $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{b} \in \mathbb{R}^m, \mathbf{c} \in \mathbb{R}^n, z \in \mathbb{R}$. Assume that $\mathbf{Ax} \leq \mathbf{b}$ is feasible. Then the following are equivalent:**

(a) $\mathbf{Ax} \leq \mathbf{b} \quad \Rightarrow \quad \mathbf{c}^T\mathbf{x} \leq z$

(b) $\mathbf{y}^T\mathbf{A} = \mathbf{c}^T \ , \ \mathbf{y}^T\mathbf{b} \leq z \ , \ \mathbf{y} \geq \mathbf{0}$ **has a solution $\mathbf{y}$.**

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 ***probability distribution*** on $\{1, .., n\}$.

A matrix $P = (p_{ij})$ where each row $P_{i\cdot}$ is a probability distribution is called a *stochastic matrix*.

In a stochastic proces:

- π_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 π is called *steady state* if $\pi = P^T \pi$

As a corollary of Gordan's result:

Each stochastic matrix P has a steady state π.

3. Linear Programs

Given $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{c} \in \mathbb{R}^n$.

$$\textbf{LP}_p: \quad \max_{\mathbf{x} \in \mathbb{R}^n} \ \mathbf{c}^T\mathbf{x} \quad \textbf{s.t.} \quad \mathbf{A}\mathbf{x} \le \mathbf{b}$$

$$\textbf{LP}_d: \quad \min_{\mathbf{y} \in \mathbb{R}^m} \ \mathbf{b}^T\mathbf{y} \quad \textbf{s.t.} \quad \mathbf{A}^T\mathbf{y} = \mathbf{c}\,, \ \mathbf{y} \ge \mathbf{0}\,,$$

is the pair of primal and dual programs.

Notation.

- $F_p = \{\mathbf{x} \mid \mathbf{A}\mathbf{x} \le \mathbf{b}\}$ **feasible set of LP$_p$**
- $F_d = \{\mathbf{y} \mid \mathbf{A}^T\mathbf{y} = c, \mathbf{y} \ge \mathbf{0}\}$ **feas.set of LP$_d$**
- $z_p^* := \max_{\mathbf{x} \in F_p} \ \mathbf{c}^T\mathbf{x}$ **max. value of LP$_p$**
- $z_d^* := \min_{\mathbf{y} \in F_d} \ \mathbf{b}^T\mathbf{y}$ **min. value of LP$_d$**
- $\overline{x} \in F_p$ **is optimal (maximizer of LP$_p$) if $c^T\overline{x} = z_p^*$.**

Weak duality *is easy to prove:*

L.3.1 *(Weak Duality)* **Let** $\mathbf{Ax} \le \mathbf{b}$, $\mathbf{A}^T\mathbf{y} = \mathbf{c}$, $\mathbf{y} \ge \mathbf{0}$. **Then,**

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

If we have $\mathbf{c}^T\mathbf{x} = \mathbf{b}^T\mathbf{y}$**, then** $\mathbf{x}, \mathbf{y}$ **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:** $\boxed{z_p^* = z_d^*}$

If both are feasible, then optimal solutions x and y of LP_p **and** LP_d **exists (satisfying** $\mathbf{c}^T\mathbf{x} = \mathbf{b}^T\mathbf{y}$**).**

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.*$=, \leq$)

$$(\star) \qquad \begin{array}{rcrcl} \mathbf{Ax} & & & \leq & \mathbf{b} \\ & & \mathbf{A}^T\mathbf{y} & = & \mathbf{c} \\ \mathbf{c}^T\mathbf{x} & - & \mathbf{b}^T\mathbf{y} & = & \mathbf{0} \\ & & \mathbf{y} & \geq & \mathbf{0}. \end{array}$$

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

$$\mathbf{b}^T\mathbf{y} - \mathbf{c}^T\mathbf{x} = \mathbf{y}^T(\mathbf{b} - \mathbf{Ax}) = \mathbf{0}$$

The relation

$$\mathbf{y}^T(\mathbf{b} - A\mathbf{x}) = 0$$

is called ***complementarity condition***.

Cor. : *(Optimality conditions)*

- **Let $\mathbf{x} \in F_p$: then x solves LP$_p$** $\Longleftrightarrow$

  there ex. $\mathbf{y} \in F_d$ **such that** $\mathbf{y}^T(\mathbf{b} - A\mathbf{x}) = 0$

- **Let $\mathbf{y} \in F_d$: Then y solves LP$_d$** $\Longleftrightarrow$

  there ex. $\mathbf{x} \in F_p$ **such that** $\mathbf{y}^T(\mathbf{b} - A\mathbf{x}) = 0$

3.1.2 Equivalent LP's

LP's in other forms can be transformed into the given "standard forms" **For example: The program:**

$$\max_{\mathbf{x}\in\mathbb{R}^n} \ \mathbf{c}^T\mathbf{x} \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} \le \mathbf{b}\,, \quad \mathbf{x} \ge \mathbf{0}$$

has the dual:

$$\min_{\mathbf{y}\in\mathbb{R}^n} \ \mathbf{b}^T\mathbf{y} \quad \text{s.t.} \quad \mathbf{A}^T\mathbf{y} \ge \mathbf{c}, \quad \mathbf{y} \ge \mathbf{0}$$

Rules for primal dual pairs:

Primal Problem	Dual Problem
max	min
free variable	equality constraint
non-negative variable	$\geq$ constraint
equality constraint	free variable
$\leq$ constraint	non-negative variable

3.1.3. Shadow prices

Production model: (n products, m resources)

c_j	**prices per unit for product** j	$\to \boldsymbol{c}$
b_i	**bounds for resource** R_i	$\to \boldsymbol{b}$
a_{ij}	**units of resource** R_i **needed for unit of product** j	$\to \boldsymbol{A}$
x_j	**production of product** j	$\to \boldsymbol{x}$

primal program: *(max profit $c^T x$)*

$$\max_{\mathbf{x}\in\mathbb{R}^n} \ \mathbf{c}^T\mathbf{x} \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} \le \mathbf{b}\,, \quad \mathbf{x} \ge \mathbf{0}$$

optimal solution: $\overline{\boldsymbol{x}} \quad \to \quad$ *profit* $\overline{\boldsymbol{z}} = \mathbf{c}^T\overline{\boldsymbol{x}}$

dual: $\min_{\mathbf{y}\in\mathbb{R}^n} \ \mathbf{b}^T\mathbf{y} \quad \text{s.t.} \quad \mathbf{A}^T\mathbf{y} \ge \mathbf{c}, \quad \mathbf{y} \ge \mathbf{0}$

optimal solution: $\overline{\boldsymbol{y}} \quad \to \quad$ *value* $\overline{\boldsymbol{z}} = \mathbf{b}^T\overline{\boldsymbol{y}}$

Question: Can we obtain a higher profit if we buy additional amount of resource b_{i_0}?

If we change: $b_{i_0} \to \tilde{b}_{i_0} := b_{i_0} + t, \ (t > 0 \ )$ **then** $\tilde{z} \to \overline{\boldsymbol{z}} + ??$

We find: $\boxed{\tilde{z} \leq \overline{\boldsymbol{z}} + t \cdot \overline{y}_{i_0}}$ ($\overline{y}_{i_0}$ is shadow price)

Answer: Yes, if the price per unit for R_{i_0} is smaller than $\overline{y}_{i_0}$ (*shadow price*).

3.1.4 Matrix games:

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:* $\mathbf{x} \in \mathbb{R}^m, \ \mathbf{y} \in \mathbb{R}^n$

- **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** $\overline{y}$**: R plays** $\overline{x}$ **as solution of:**

$$\max_x x^T A \overline{y} = \max_i \sum_j a_{ij} \overline{y}_j$$

- **given** $\overline{x}$**: C plays** $\overline{y}$ **as solution of:**

$$\min_y \overline{x}^T A y = \min_j \sum_i a_{ij} \overline{x}_i$$

Best strategy against best of opponent:

$$\text{for R:} \quad \max_x \min_y x^T A y = \max_x \min_j \sum_i a_{ij} x_i$$

$$\text{for C:} \quad \min_y \max_x x^T A y = \min_y \max_i \sum_j a_{ij} y_j$$

Th.3.2 [*minmax-theorem*] **There exist feasible $\overline{x}, \overline{y}$ such that**

$$\min_y \overline{x}^T A y = \max_x x^T A \overline{y}$$

This implies: $\boxed{\max_x \min_y x^T A y = \min_y \max_x x^T A y = \overline{x}^T A \overline{y}}$

$\overline{x}, \overline{y}$ represent a Nash equilibrium of the mixed strategy matrix game.

Def. **A game is fair if $\overline{z} = \overline{w} = \overline{x}^T A \overline{y} = 0$ holds.**

Methods for linear programs

1. Simplex method

$$\textbf{LP}_p: \quad \max_{\mathbf{x}\in\mathbb{R}^n} \ \mathbf{c}^T\mathbf{x} \quad \textbf{s.t.} \quad \mathbf{a}_i^T\mathbf{x} \le b_i, \ \ i = 1, .., m.$$

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:

$$A^T\mathbf{y} = \boldsymbol{c}, \ \mathbf{y} \ge \mathbf{0}, \ \ \mathbf{y}^T(\mathbf{b} - A\mathbf{x}) = 0$$

2. Interior point method: Consider the system of equations

$$P(t): \quad \begin{array}{rl} Ax \quad + s & = b \\ A^T y & = c \\ y_i(b - Ax)_i & = \underline{t > 0} \quad \forall i \end{array}$$

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

Idea: **Compute (by 'Newton') solutions** $\mathbf{x}(t), \mathbf{y}(t), \mathbf{s}(t),\ t > 0$ **of P(t). Then for** $t \downarrow 0$ **(hopefully)**

$$\mathbf{x}(t), \mathbf{y}(t), \mathbf{s}(t) \longrightarrow \mathbf{x}, \mathbf{y}, \mathbf{s}$$

With solutions $\mathbf{x}, \mathbf{y}$ **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.