# NMSA403 Optimization Theory – Exercises

**Martin Branda**

Charles University, Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics

Version: 17/12/2021

Symbol **(*)** denotes examples which are left to the readers as an exercise.

# Contents

# 1 Introduction to optimization

Please repeat the topics which were contained in Introduction to optimization or similar lectures:

- Polyhedral sets
- Cones
- Extreme points and directions
- Farkas theorem
- Convexity of sets and functions
- Symmetric Local Optimality Conditions (SLPO)

**Example 1.1.** *Derive the sets of extreme points and extreme directions for the polyhedral set*

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

*Use*

- *the picture,*
- *the computational approach (direct method).*

**Solution.** Picture is left to readers. We will focus on the computational approach. First, we transform the constraints into the standard (equality) form using nonnegative slack variables[1], i.e.

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

So we have matrix and rhs vector

$$A = \begin{pmatrix} -1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \; b = \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

Now, to get the extreme points we must solve all systems of linear equalities with square $2 \times 2$ regular submatrices of $A$ and nonengative solutions, e.g. by solving

$$\begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix},$$

we obtain $(2, 3, 0, 0)$ where the elements corresponding to the omitted columns are substituted by 0.

[1] $+$ with $\leq$, $-$ with $\geq$

Similar approach can be used to get the extreme directions where we solve all homogenous systems with rectangle $2\times 3$ submatrices of $A$ and we look for nonnegative solutions, e.g. by solving

$$\begin{pmatrix} -1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

we get $(1, 0, 1, 0)$ where again the elements corresponding to the omitted columns are set to 0.

After going through all possibilities, we obtain

$$\operatorname{ext}(M) = \left\{ \begin{pmatrix} 2 \\ 3 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 3 \end{pmatrix} \right\}, \ \operatorname{extd}(M) = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\}.$$

**Example 1.2.** *Apply the graphical method to solve the dual problem to the following problem*

$$\begin{array}{lccccccccc} \min & -6x_1 & & & + & 3x_3 & + & 5x_4 & & \\ \text{s.t.} & 3x_1 & + & x_2 & - & x_3 & - & x_4 & = & 3, \\ & 2x_1 & - & x_2 & + & 2x_3 & + & 3x_4 & \leq & 2, \\ & x_1 \geq 0, & & x_2 \geq 0, & & x_3 \geq 0, & & x_4 \geq 0. & & \end{array}$$

*Derive the optimal solution(s) of the primal problem using the complementarity conditions. Identify the extreme points and directions of the dual problem.*

**Solution:** We can derive the dual problem using the table. First the primal problem is coded, then the dual problem is derived:

| | | $x_1$ $\geq 0$ | $x_2$ $\geq 0$ | $x_3$ $\geq 0$ | $x_4$ $\geq 0$ | | |
|---|---|---|---|---|---|---|---|
| | | 3 | 1 | -1 | -1 | $=$ | 3 |
| | | 2 | -1 | 2 | 3 | $\leq$ | 2 |
| | | -6 | 0 | 3 | 5 | min | |

| | | $x_1$ $\geq 0$ | $x_2$ $\geq 0$ | $x_3$ $\geq 0$ | $x_4$ $\geq 0$ | | |
|---|---|---|---|---|---|---|---|
| $y_1$ | $\in \mathbb{R}$ | 3 | 1 | -1 | -1 | $=$ | 3 |
| $y_2$ | $\leq 0$ | 2 | -1 | 2 | 3 | $\leq$ | 2 |
| | | $\leq$ | $\leq$ | $\leq$ | $\leq$ | | max |
| | | -6 | 0 | 3 | 5 | min | |

We can formulate the dual problem:

$$\begin{aligned}
\max\ & 3y_1 + 2y_2 \\
s.t.\ & 3y_1 + 2y_2 \le -6, \\
& y_1 - y_2 \le 0, \\
& -y_1 + 2y_2 \le 3, \\
& -y_1 + 3y_2 \le 5, \\
& y_2 \le 0.
\end{aligned}$$

As a solution we obtain line segment connecting two extreme points: $(-2,0), (-6/5,-6/5)$, which can be written as

$$(y_1, y_2) \in \left\{ \left( t, -3 - \frac{3}{2}t \right),\ t \in \left[ -2, -\frac{6}{5} \right] \right\}.$$

The optimal value is equal to $-6$. Note that there is another extreme point $(-3,0)$ of the feasibility set and the (non-normalized) extreme directions are: $(-1,-1)$, $(-2,-1)$. So the set is polyhedral, but not a polyhedron.

Now, using the complementarity conditions

$$\begin{aligned}
& x_1(3y_1 + 2y_2 + 6) = 0, \\
& x_2(y_1 - y_2) = 0, \\
& x_3(-y_1 + 2y_2 - 3) = 0, \\
& x_4(-y_1 + 3y_2 - 5) = 0,
\end{aligned}$$

we obtain that $x_2 = x_3 = x_4 = 0$. Using the second set of complementarity conditions

$$\begin{aligned}
& y_1(3x_1 + x_2 - x_3 - x_4 - 3) = 0 \\
& y_2(2x_1 - x_2 + 2x_3 + 3x_4 - 2) = 0,
\end{aligned}$$

we have $x_1 = 1$ with optimal solution of the primal problem equal to $-6$, which corresponds to the strong duality.

**Example 1.3.** *Consider the problem*

$$\begin{aligned}
\min\ & x_1 \\
s.t.\ & (x_1 - 1)^2 + (x_2 - 1)^2 \le 1, \\
& (x_1 - 1)^2 + (x_2 + 1)^2 \le 1, \\
& x_1, x_2 \ge 0.
\end{aligned}$$

*Using the picture find an optimal solution. Formulate the symmetric local optimality conditions and verify if they are fulfilled for the optimal solution. Discuss the constraints qualification conditions.*

**Solution:** The optimal solution is obviously the only feasible point $(1,0)$. Note that the functions (in the objective function as well as in the constraints) are convex. Write the Lagrange function

$$L(x_1, x_2 y_1, y_2) = x_1 + y_1\big((x_1-1)^2 + (x_2-1)^2 - 1\big) + y_2\big((x_1-1)^2 + (x_2+1)^2 - 1\big),\ y_1, y_2 \ge 0.$$

Derive the SLPO optimality conditions

$$
\begin{aligned}
\text{i)}\ & \frac{\partial L}{\partial x_1} = 1 + 2y_1(x_1-1) + 2y_2(x_1-1) \ge 0,\ x_1\frac{\partial L}{\partial x_1} = 0,\ x_1 \ge 0,\\
& \frac{\partial L}{\partial x_2} = 2y_1(x_2-1) + 2y_2(x_2+1) \ge 0,\ x_2\frac{\partial L}{\partial x_2} = 0,\ x_2 \ge 0,\\
\text{ii)}\ & \frac{\partial L}{\partial y_1} = (x_1-1)^2 + (x_2-1)^2 - 1 \le 0,\ y_1\big((x_1-1)^2 + (x_2-1)^2 - 1\big) = 0,\ y_1 \ge 0,\\
& \frac{\partial L}{\partial y_2} = (x_1-1)^2 + (x_2+1)^2 - 1 \le 0,\ y_2\big((x_1-1)^2 + (x_2+1)^2 - 1\big) = 0,\ y_2 \ge 0.
\end{aligned} \tag{1}
$$

Now, by substituting point (1,0) into the conditions, since $x_1 = 1 \neq 0$ we obtain

$$\frac{\partial L}{\partial x_1}(1,0) = 1 \neq 0,$$

i.e. there are no Lagrange multipliers $y_{1,2} \ge 0$ such that $(1,0,y_1,y_2)$ fulfills the SLPO conditions. The explanation is easy: No Constraint Qualification condition is fulfilled. We can quickly check the basic (strongest) ones:
Slater CQ: There is no $x \ge 0$ such that

$$(x_1-1)^2 + (x_2-1)^2 < 1 \ \&\ (x_1-1)^2 + (x_2+1)^2 < 1.$$

LI CQ: We can compute the gradients

$$\nabla g_1(1,0) = \begin{pmatrix} 2(x_1-1) \\ 2(x_2-1) \end{pmatrix}\Big|_{(1,0)} = \begin{pmatrix} 0 \\ -2 \end{pmatrix},\ \nabla g_2(1,0) = \begin{pmatrix} 2(x_1-1) \\ 2(x_2+1) \end{pmatrix}\Big|_{(1,0)} = \begin{pmatrix} 0 \\ 2 \end{pmatrix},$$

which are obviously not linearly independent. A more detailed explanation will be given in the second part of the semester.

# 2 Separating hyperplane theorems

Please see Lecture notes, Section 1.5: theorem about projection of a point to a convex set (obtuse angle), separation of a point and a convex set, proper and strict separability.

Using the theorem about the separation of a point and a convex set, prove the following lemma about the existence of a supporting hyperplane.

**Lemma 2.1.** *Let $\emptyset \neq K \subset \mathbb{R}^n$ be a convex set and $x \in \partial K$. Then, there is $\gamma \in \mathbb{R}^n$, $\gamma \neq 0$ such that*

$$\inf\{\langle \gamma, y\rangle : \ y \in K\} \geq \langle \gamma, x\rangle .$$

Hint: separate a sequence $x_n \notin K$ which converge to the point $x$ on the boundary, show the convergence of separating hyperplanes characterized by $\gamma_n \neq 0$. See Theorem 1.36 in Lecture notes.

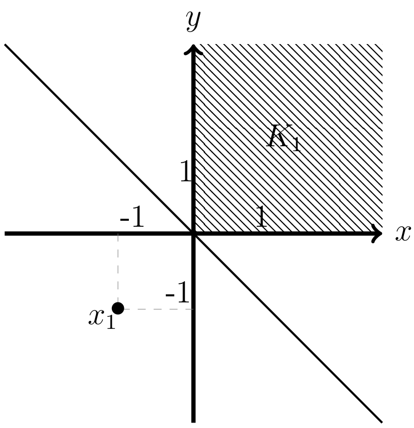
From the following examples, where we can draw a picture, you have to get an idea of the separability of sets and points. Although, in general (in proofs), we usually have to rely on existential theorems only.

**Example 2.2.** *Find a separating or supporting hyperplane for the following sets and points:*

$$\begin{array}{ll} x_1 = (-1,-1) & K_1 = \{(x,y);\ x \geq 0, y \geq 0\}, \\ x_2 = (3,1) & K_2 = \{(x,y);\ x^2 + y^2 < 10\}, \\ x_3 = (3,0,0) & K_3 = \{(x,y,z);\ x^2 + y^2 + z^2 \leq 9\}, \\ x_4 = (0,2,0) & K_4 = \{(x,y,z);\ x + y + z \leq 1\}. \end{array}$$

**Solution:** Use pictures and realize that $\gamma$ is *the normal vector* of the separating/supporting hyperplane.

**1.** We start with the picture

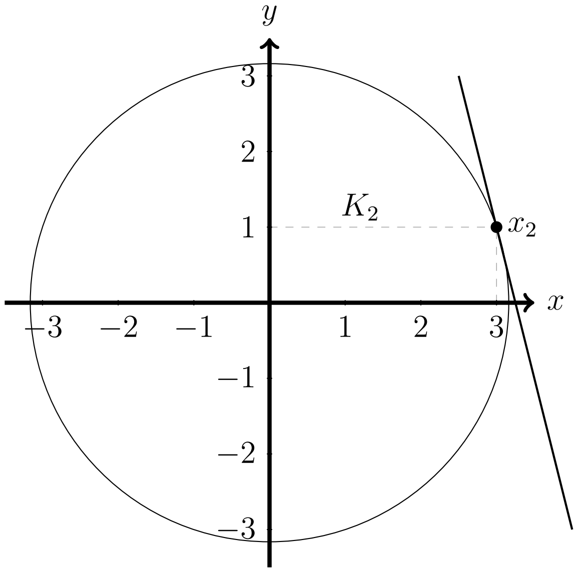



For $x_1, K_1$ we can use $\gamma = (1,1)$ to construct the separating hyperplane, then

$$\min_{(x,y)\in K_1} x + y = 0 > -1 - 1 = -2.$$

Note that other choices are also possible, in particular $(\gamma_1, \gamma_2) \neq 0$ with $\gamma_1, \gamma_2 \geq 0$.

**2.** Picture




Since $x_2 \in \partial(\mathrm{cl}(K_2))$ (on the boundary of the closure) and $K_2$ is a convex set, we are going to construct the supporting hyperplane. We set $\gamma = (-3, -1)$ and verify the property

$$\inf_{(x,y)\in K_2} -3x - y = -10 \geq -3 \cdot 3 - 1 \cdot 1 = -10.$$

**3.** Since $x_3 \in \partial K_3$ and $K_3$ is a convex closed set, we are going to construct the supporting hyperplane. We set $\gamma = (1, 0, 0)$ and verify the property

$$\min_{(x,y,z)\in K_3} x + 0 \cdot y + 0 \cdot z = -3 < 3 = 3 + 0 + 0.$$

So we try $\gamma = (-1, 0, 0)$:

$$\min_{(x,y,z)\in K_3} -x + 0 \cdot y + 0 \cdot z = -3 = -3 + 0 + 0,$$

which works.

**4.** Since $x_4 \notin \mathrm{cl}(K_4)$ and $K_4$ is a convex closed set, we can construct the separating hyperplane using $\gamma = (-1, -1, -1)$

$$\min_{(x,y,z)\in K_4} -x - y - z = -1 >= -2 - 0 - 0.$$

**Example 2.3.** *Let $K \subseteq \mathbb{R}^n$, $K \neq \emptyset$. Show that $K$ is a closed convex set if and only if it is an intersection of all closed half-spaces which contain $K$.*

Hint: Show that if $y \notin K$, then it is not contained in the intersection using the theorem about separation of a point and a convex set. See Theorem 1.37 in Lecture notes.

**Example 2.4.** *Provide a description of the circle in $\mathbb{R}^2$ and ball in $\mathbb{R}^3$ as a intersection of supporting halfspaces.*

**Solution** ($\mathbb{R}^2$ case): WLOG consider $K = \{(x_1, x_2) : \; x_1^2 + x_2^2 \le 1\}$. Then in each point of the boundary $\partial(K)$ we can construct the supporting hyperplane. In fact, the point $x \in \partial(K)$ already corresponds to the normal vector $\gamma$, i.e. the supporting hyperplane is

$$H_x = \{y \in \mathbb{R}^2 : \; x^T y = c\},$$

where the choice of $c$ is obvious from the fact that $x \in H_x \cap K$, i.e. $c = 1$. So, we obtain

$$K = \bigcap_{x \in \partial(K)} \{y \in \mathbb{R}^2 : \; x^T y \le 1\}.$$

Prove the following theorem which gives a sufficient condition for proper separability of two convex sets.

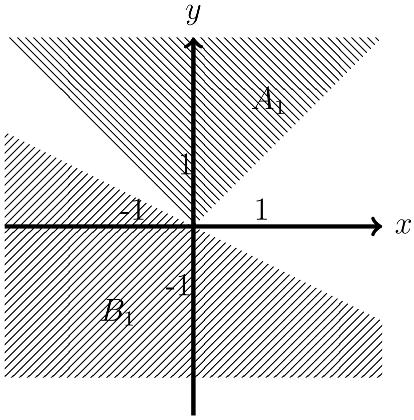
**Theorem 2.5.** *Let $A, B \subset \mathbb{R}^n$ be non-empty convex sets. If $\operatorname{rint}(A) \cap \operatorname{rint}(B) = \emptyset$, then $A$ and $B$ can be properly separated.*

Hint: Separate set $K = A - B$ and point 0. First, show that $0 \notin \operatorname{rint}(K)$. See Theorem 1.45 in Lecture notes.

**Example 2.6.** *Verify whether the following pairs of (convex ?) sets are properly or strictly separable or not. If they are separable, suggest a possible value of $\gamma$ and verify the property.*

$$\begin{array}{ll}
A_1 = \{(x, y); \; y \ge |x|\}, & B_1 = \{(x, y); \; 2y + x \le 0\}, \\
A_2 = \{(x, y); \; xy \ge 1, x > 0\}, & B_2 = \{(x, y; \; x \le 0, y \le 0\}, \\
A_3 = \{(x, z); \; x + y + z \le 1\}, & B_3 = \{(x, y, z); \; (x-2)^2 + (y-2)^2 + (z-2)^2 \le 1\}, \\
A_4 = \{(x, y, z); \; 0 \le x, y, z \le 1\}, & B_4 = \{(x, y, z); \; (x-2)^2 + (y-2)^2 + (z-2)^2 \le 3\}. \\
A_5 = \{(x, y); \; e^{-x} \le y\}, & B_5 = \{(x, y); \; -e^{-x} \ge y\}.
\end{array}$$

**Solution:** Use general theorems and pictures.
**1.** We start with the picture




Sets $A_1, B_1$ are properly separable using vector $\gamma = (1, 2)$:

$$\min_{(x,y) \in A_1} x + 2y = 0 \ge 0 = \max_{(x,y) \in B_1} x + 2y.$$

Note that the assumptions of general theorem about proper separability are fulfilled (both sets are convex and $\mathrm{int}(A_1) \cap \mathrm{int}(B_1) = \emptyset$.

**2.** From a picture (that you can draw yourself), we can see that sets $A_2, B_2$ can be strictly separated even though they do not fulfill the assumption of the general theorem. We can use $\gamma = (1, 1)$ and verify

$$\min_{(x,y)\in A_2} x + y = 2 > 0 = \max_{(x,y)\in B_2} x + y,$$

where the minimum is attained at (1,1) and maximum at (0,0).

**3.** If you use your spatial imagination or a picture, you can realize that $A_3 \cap B_3 = \emptyset$. The assumptions of the theorem about strict separability are fulfilled and the sets can be strictly separated using $\gamma = (-1, -1, -1)$:

$$\min_{(x,y,z)\in A_3} -x - y - z = -1 > -6 + \sqrt{3} = \max_{(x,y,z)\in B_3} -x - y - z.$$

The closest point from $B_3$ to $A_3$, where the maximum over $B_3$ is attained, is $(2 - \frac{\sqrt{3}}{3}, 2 - \frac{\sqrt{3}}{3}, 2 - \frac{\sqrt{3}}{3})$.

**4.** In this case, you can realize that $A_4 \cap B_4 = (1, 1, 1)$. The assumptions of the theorem about proper separability are fulfilled and the sets can be properly separated using $\gamma = (-1, -1, -1)$:

$$\min_{(x,y,z)\in A_4} -x - y - z = -3 = \max_{(x,y,z)\in B_4} -x - y - z.$$

**5.** This example is quite interesting, because both sets are closed convex and has no intersection, but cannot be strictly separated: general theorem about strict separability requires that at least one set is compact which is not fulfilled here. The sets can be properly separated using $\gamma = (0, 1)$, i.e.

$$\inf_{(x,y)\in A_5} 0 \cdot x + y = 0 = \sup_{(x,y)\in B_5} 0 \cdot x + y.$$

**Example 2.7.** (*) *Discuss the proof of the Farkas theorem.*

Hint: Use an alternative formulation of the FT: Denote the columns of $A$ by $a_{\bullet,i}$, $i = 1, \ldots, n$. Then $Ax = b$ has a nonnegative solution if and only if $b \in \mathrm{pos}(\{a_{\bullet,1}, \ldots, a_{\bullet,n}\})$.

**Example 2.8.** *Reformulate the Farkas theorem for the sets*

$$\begin{aligned} M_1 &= \{x : Ax \geq b\}, \\ M_2 &= \{x : Ax \geq b,\ x \geq 0\}, \\ M_3 &= \{x : Ax \leq b,\ x \leq 0\}, \\ M_4 &= \{x : Ax = b\}. \end{aligned}$$

**Solution:** Consider $M_1$. We can transform the constraints to the standard form, i.e. we use $x = x^+ - x^-$, $x^+, x^- \geq 0$, and slack variables $y \geq 0$ such that

$$A(x^+ - x^-) - Iy = b,$$

i.e. we can set $\tilde{A} = (A| - A| - I)$ and $\tilde{x}^T = (x^{+T}|x^{-T}|y^T) \geq 0$. Now we can apply the Farkas theorem to

$$\tilde{A}\tilde{x} = b, \; \tilde{x} \geq 0.$$

We can simplify $A^T u \geq 0$, $-A^T u \geq 0$, and $-Iu \geq 0$ to $A^T u = 0$ and $u \leq 0$. The modified Farkas theorem is:
$Ax \geq b$ has a solution if and only if for all $u \leq 0$, $A^T u = 0$ it holds $b^T u \geq 0$.

# 3 Convex sets and functions

Repeat the rules for estimating convexity of functions and sets:

- intersection of convex sets is a convex set,
- level sets of convex functions are convex sets,
- nonnegative combination of convex functions is a convex function,
- maximum of convex functions is a convex function,
- function composition when
  - inner function is linear and outer convex,
  - inner function is convex and outer convex and nondecreasing

  is a convex function,
- once differentiable univariate function is convex iff the first order derivative is nondecreasing,
- twice differentiable univariate function is convex iff the second order derivative is nonnegative,
- twice continuously differentiable multivariate function is convex iff Hessian matrix is positive semidefinite.

**Definition 3.1.** *For a function $f : \mathbb{R}^n \to \mathbb{R}^*$, we define its* ***epigraph***

$$\mathrm{epi}(f) = \left\{(x, \nu) \in \mathbb{R}^{n+1} : \ f(x) \le \nu\right\}.$$

**Example 3.2.** *Prove the equivalence between the possible definitions of convex functions $f : \mathbb{R}^n \to \mathbb{R}^*$:*

1. $\mathrm{epi}(f)$ *is a convex set,*
2. $\mathrm{Dom}(f)$ *is a convex set and for all $x, y \in \mathrm{Dom}(f)$ and $\lambda \in (0, 1)$ we have*

$$f(\lambda x + (1 - \lambda)y) \le \lambda f(x) + (1 - \lambda)f(y).$$

**Solution:** Let the epigraph be convex, $x, y \in \mathrm{Dom}(f)$, and choose $\lambda \in (0, 1)$. We know that $(x, f(x)) \in \mathrm{epi}(f)$ as well as $(y, f(y)) \in \mathrm{epi}(f)$. Since the epigraph is a convex set, we know that the convex combination of the points belongs also to the epi, i.e.

$$\lambda(x, f(x)) + (1 - \lambda)(y, f(y)) \in \mathrm{epi}(f).$$

This means that

$$f(\lambda x + (1 - \lambda)y) \le \lambda f(x) + (1 - \lambda)f(y),$$

which we wanted to show.
Now let 2. is valid. Choose $\lambda \in (0,1)$, and $(x,\nu)$, $(y,\eta) \in \text{epi}(f)$, i.e. $\nu \geq f(x)$ and $\eta \geq f(y)$. From 2. we know that

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \leq \lambda \nu + (1-\lambda)\eta,$$

i.e.

$$\big(\lambda x + (1-\lambda)y, \lambda\nu + (1-\lambda)\eta\big) \in \text{epi}(f),$$

and we can conclude that the epigraph is a convex set.

**Remark 3.3.** *You can restrict domain of any function and define its outside the domain using $+\infty$. This influences also the epigraph. For example consider $f(x) = x^3$ on $\mathbb{R}$, and $g(x) = x^3$ on $\mathbb{R}_+$ and $g(x) = +\infty$ for $x \in (-\infty, 0)$. Realize how their epigraphs look. Then $g$ is obviously convex, but $f$ not.*

**Example 3.4.** *Decide whether the following sets are convex:*

$$\begin{aligned}
M_1 &= \left\{(x,y) \in \mathbb{R}^2_+ :\ ye^{-x} - x \geq 1\right\}, && (2)\\
M_2 &= \left\{(x,y) \in \mathbb{R}^2 :\ x \geq 2 + y^2\right\}, && (3)\\
M_3 &= \left\{(x,y) \in \mathbb{R}^2 :\ x^2 + y\log y^4 \leq 139,\ y \geq 2\right\}, && (4)\\
M_4 &= \left\{(x,y) \in \mathbb{R}^2 :\ \log x - y^2 \geq 1,\ x \geq 1,\ y \geq 0\right\}, && (5)\\
M_5 &= \left\{(x,y) \in \mathbb{R}^2 :\ (x^3 + e^y)\log(x^3 + e^y) \leq 49,\ x \geq 0,\ y \geq 0\right\}, && (6)\\
M_6 &= \left\{(x,y) \in \mathbb{R}^2 :\ x\log x + xy \geq 0,\ x \geq 1\right\}, && (7)\\
M_7 &= \left\{(x,y) \in \mathbb{R}^2 :\ 1 - xy \leq 0,\ x \geq 0\right\}, && (8)\\
M_8 &= \left\{(x,y,z) \in \mathbb{R}^3 :\ \frac{1}{2}(x^2 + y^2 + z^2) + yz \leq 1,\ x \geq 0,\ y \geq 0\right\}, && (9)\\
M_9 &= \left\{(x,y,z) \in \mathbb{R}^2 :\ 3x - 2y + z = 1\right\}. && (10)
\end{aligned}$$

**Solution:**

2. $M_2$ is an epigraph of function $f(y) = 2 + y^2$, which is obviously convex, so a convex set.

3. We can show that the function which defines the set is convex, i.e.

$$f(x,y) = x^2 + y\log y^4 = x^2 + 4y\log y \text{ on } \mathbb{R} \times [2, \infty).$$

Function $x^2$ is obviously convex and $g(y) := y\log y$ is convex on $(0,\infty)$ – we can verify it using derivatives

$$g'(y) = \log y + 1,\ g''(y) = \frac{1}{y}.$$

So $M_3$ is a level set of the convex function.

7. We cannot use the level set approach because the function is not convex, but from the picture it is obvious that the set $M_7$ is a convex set.

9. $M_9$ is a hyperplane, thus convex set.

**Example 3.5.** *Establish conditions under which the following sets are convex:*

$$
\begin{aligned}
M_{10} &= \left\{x \in \mathbb{R}^n : \ \alpha \leq a^T x \leq \beta\right\}, \ \text{for some } a \in \mathbb{R}^n, \alpha, \beta \in \mathbb{R}, && (11)\\
M_{11} &= \left\{x \in \mathbb{R}^n : \ \left\|x - x^0\right\| \leq \|x - y\|, \ \forall y \in S\right\}, \ \text{for some } S \subseteq \mathbb{R}^n, && (12)\\
M_{12} &= \left\{x \in \mathbb{R}^n : \ x^T y \leq 1, \ \forall y \in S\right\}, \ \text{for some } S \subseteq \mathbb{R}^n. && (13)
\end{aligned}
$$

**Solution:**

12. If $S = \emptyset$, then we have no condition and $M_{12} = \mathbb{R}^n$, which is obviously convex. Now consider $S \neq \emptyset$, then we can use the definition: let $\lambda \in (0,1)$ and $x_1, x_2 \in M_{12}$, which means that

$$x_1^T y \leq 1 \text{ and } x_2^T y \leq 1, \ \forall y \in S.$$

We can obtain

$$(\lambda x_1 + (1-\lambda)x_2)^T y = \lambda\, x_1^T y + (1-\lambda)\, x_2^T y \leq \lambda + (1-\lambda) = 1, \ \forall y \in S,$$

which implies that $\lambda x_1 + (1-\lambda)x_2 \in M_{12}$ and therefore the considered set is convex. Note that no restrictions (such as convexity) on the set $S$ are necessary.

**Example 3.6.** *Verify if the following functions are convex:*

$$
\begin{aligned}
f_1(x,y) &= x^2 y^2 + \frac{x}{y}, \ x > 0, y > 0, && (14)\\
f_2(x,y) &= xy, && (15)\\
f_3(x,y) &= \log(e^x + e^y) - \log x, \ x > 0, && (16)\\
f_4(x,y) &= \exp\{x^2 + e^{-y}\}, \ x > 0, \ y > 0, && (17)\\
f_5(x,y) &= -\log(x+y), \ x > 0, \ y > 0, && (18)\\
f_6(x,y) &= \sqrt{e^x + e^{-y}}, && (19)\\
f_7(x,y) &= x^3 + 2y^2 + 3x, && (20)\\
f_8(x,y) &= -\log(cx + dy), \ c, d \in \mathbb{R}, && (21)\\
f_9(x,y) &= \frac{x^2}{y}, \ y > 0, && (22)\\
f_{10}(x,y) &= xy \log xy, \ x > 0, y > 0, && (23)\\
f_{11}(x,y) &= |x+y|, && (24)\\
f_{12}(x) &= \sup_{y \in \mathrm{Dom}(f)} \{x^T y - f(y)\} = f^*(x), \ f : \mathbb{R}^n \to \mathbb{R}, && (25)\\
f_{13}(x) &= \|Ax - b\|_2^2. && (26)
\end{aligned}
$$

**Solution:**
**6.** Consider $f_6(x,y) = \sqrt{e^x + e^{-y}}$. The first order partial derivatives are equal to

$$\begin{aligned} \frac{\partial f_6}{\partial x}(x,y) &= \frac{e^x}{2\sqrt{e^x + e^{-y}}}, \\ \frac{\partial f_6}{\partial y}(x,y) &= \frac{-e^{-y}}{2\sqrt{e^x + e^{-y}}}, \end{aligned}$$

and the second order derivatives are

$$\begin{aligned} \frac{\partial^2 f_6}{\partial x^2}(x,y) &= \frac{e^{2x} + 2e^{x-y}}{4(e^x + e^{-y})^{\frac{3}{2}}}, \\ \frac{\partial^2 f_6}{\partial y^2}(x,y) &= \frac{e^{-2y} + 2e^{x-y}}{4(e^x + e^{-y})^{\frac{3}{2}}}, \\ \frac{\partial^2 f_6}{\partial y \partial x}(x,y) = \frac{\partial^2 f_6}{\partial x \partial y}(x,y) &= \frac{e^{x-y}}{4(e^x + e^{-y})^{\frac{3}{2}}} \end{aligned}$$

To verify that the Hessian matrix is positive definite, it is sufficient to look on the numerators, because the common denominator $4(e^x + e^{-y})^{\frac{3}{2}}$ is always positive. Obviously $e^{2x} + 2e^{x-y}$ is positive, thus it remains to verify that

$$(e^{2x} + 2e^{x-y})(e^{-2y} + 2e^{x-y}) - (e^{x-y})^2 > 0.$$

**11.** Inner function $x + y$ is linear and the outer function $|\cdot|$ is convex, which is sufficient to prove the convexity of $f_{11}$.
**13.** Realize that

$$f_{13}(x) = \|Ax - b\|_2^2 = (Ax - b)^T(Ax - b) = x^T A^T Ax - 2b^T Ax + b^T b.$$

The first term is convex because the matrix $A^T A$ is positive semidefinite, the second term is linear (affine) and the last one is a constant. So the function as a sum of two convex functions (and a constant) is convex.

**Example 3.7.** *Show that*

$$f_{14}(x) = -\sum_{i=1}^{n} \ln(b_i - a_i^T x). \tag{27}$$

*is convex on its domain* $\mathrm{dom}(f_{14}) = \{x \in \mathbb{R}^n : a_i^T x < b_i,\, i = 1, \dots, n\}$.

**Example 3.8.** *Let* $f(x,y) : \mathbb{R}^{n+m} \to \mathbb{R}$ *be a convex function (jointly in* $x$ *and* $y$*) and* $C \subseteq \mathbb{R}^m$ *be a nonempty convex set. Show that*

$$g(x) = \inf_{y \in C} f(x,y). \tag{28}$$

*is convex.*

**Solution:** Consider $\lambda \in (0,1)$ and points $x_1, x_2$ from the domain of the function $g$, i.e. for each $\varepsilon > 0$ there exist $y_1, y_2 \in C$ such that

$$\begin{aligned} g(x_1) + \varepsilon &\geq f(x_1, y_1), \\ g(x_2) + \varepsilon &\geq f(x_2, y_2), \end{aligned}$$

because the infimum need not to be attained but the function value must be arbitrary close. Then, we can obtain the following relations

$$f(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \leq \lambda f(x_1, y_1) + (1-\lambda)f(x_2, y_2) \leq \lambda g(x_1) + (1-\lambda)g(x_2) + \varepsilon.$$

It remain to show that

$$g(\lambda x_1 + (1-\lambda)x_2) \leq f(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2),$$

which follows from the fact that $C$ is a convex set, therefore the convex combination of $y_1, y_2$ is feasible, i.e.

$$\lambda y_1 + (1-\lambda)y_2 \in C.$$

Remind that $g$ is defined as the infimum of $f$ over $C$. By letting $\varepsilon \to 0_+$ and combining the above inequalities, we can conclude that $g$ is convex on its domain.

**Example 3.9.** *(Vector composition) Let $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, k$ and $h : \mathbb{R}^k \to \mathbb{R}$ be convex functions. Moreover let $h$ be nondecreasing in each argument. Then*

$$f(x) = h\big(g_1(x), \ldots, g_k(x)\big). \tag{29}$$

*is convex.*
*Apply to*

$$f(x) = \log\left(\sum_{i=1}^{k} e^{g_i(x)}\right),$$

*where $g_i$ are convex.*
*Hint: the first part can be verified using the definition of convexity, in the second part compute the Hessian matrix $\mathcal{H}(x)$ and use the Cauchy-Schwarz inequality $(a^T a)(b^T b) \geq (a^T b)^2$ to verify that $v^T \mathcal{H}(x) v \geq 0$ for all $v \in \mathbb{R}^k$.*

**Solution:** We can prove the convexity using the definition and the assumed properties. Consider $x_1, x_2 \in \mathbb{R}^n$ and $\lambda \in (0,1)$, then

$$\begin{aligned} f(\lambda x_1 + (1-\lambda)x_2) &= h\big(g_1(\lambda x_1 + (1-\lambda)x_2), \ldots, g_k(\lambda x_1 + (1-\lambda)x_2)\big) \\ &\leq h\Big(\lambda g_1(x_1) + (1-\lambda)g_1(x_2), \ldots, \lambda g_k(x_1) + (1-\lambda)g_k(x_2)\Big) \\ &\leq \lambda h\big(g_1(x_1), \ldots, g_k(x_1)\big) + (1-\lambda)h\big(g_1(x_2), \ldots, g_k(x_2)\big) \\ &= \lambda f(x_1) + (1-\lambda)f(x_2), \end{aligned}$$

where the first inequality follows from that $g_1, \ldots, g_k$ are convex and $h$ is nondecreasing, the second one from the convexity of $h$.

Now, consider the function

$$h(z) = \log\left(\sum_{i=1}^{k} e^{z_i}\right).$$

Obviously it is nondecreasing in each argument. We can show that it is also convex. We compute its second order partial derivatives

$$\begin{aligned}
\frac{\partial^2 h}{\partial z_j^2}(z) &= \frac{e^{z_j}(\sum_{i=1}^{k} e^{z_i}) - e^{z_j}e^{z_j}}{(\sum_{i=1}^{k} e^{z_i})^2},\ j = 1, \dots, k, \\
\frac{\partial^2 h}{\partial z_j \partial z_l}(z) &= \frac{-e^{z_j}e^{z_l}}{(\sum_{i=1}^{k} e^{z_i})^2},\ j \neq l.
\end{aligned}$$

If we use the notation $y = (e^{z_1}, \dots, e^{z_k})^T$ and $\mathbb{I} = (1, \dots, 1)^T \in \mathbb{R}^k$, we can write the Hessian matrix in the form

$$\mathcal{H}_h(y) = \frac{1}{(\mathbb{I}^T y)^2}\left(\mathrm{diag}(y)(\mathbb{I}^T y) - yy^T\right),$$

where $\mathbb{I}^T y = \sum_{i=1}^{k} y_i = \sum_{i=1}^{k} e^{z_i}$ and $\mathrm{diag}(y)$ denotes the diagonal matrix with elements $y$. We would like to verify that $v^T \mathcal{H}_h(y) v \geq 0$ for arbitrary $v \in \mathbb{R}^k$. We can compute

$$v^T \mathcal{H}_h(y) v = \frac{(\sum_{i=1}^{k} y_i v_i^2)(\sum_{i=1}^{k} y_i) - (\sum_{i=1}^{k} y_i v_i)^2}{(\sum_{i=1}^{k} y_i)^2}.$$

By setting $a_i = \sqrt{y_i} v_i$ and $b_i = \sqrt{y_i}$ and using the Cauchy-Schwarz inequality in the form $(a^T a)(b^T b) - (a^T b)^2 \geq 0$, we can obtain that the numerator is nonnegative, i.e. $v^T \mathcal{H}_h(y) v \geq 0$.

**Example 3.10.** (*) *Verify that the geometric mean is concave*[2]*:*

$$f(x) = \left(\prod_{i=1}^{n} x_i\right)^{1/n},\ x \in (0, \infty)^n. \tag{30}$$

*Hint: compute the Hessian matrix and use the Cauchy-Schwarz inequality* $(a^T a)(b^T b) \geq (a^T b)^2$.

[2] $f$ is concave $\Leftrightarrow$ $-f$ is convex.

# 4 Subdifferentiability and subgradient

From Introduction to optimization (or similar course), you should remember the following property which holds for any differentiable convex function $f : X \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$:

$$\forall x, y \in X \;\; f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle .$$

This property can be generalized for nondifferentiable convex function by the notation of subdifferentiability. Any subgradient $a \in \mathbb{R}^n$ of function $f$ at $x \in X$ fulfills

$$f(y) - f(x) \geq \langle a, y - x \rangle \;\; = a^T(y - x) \; \forall y \in X.$$

Set of all subgradients at $x$ is called subdifferential of $f$ at $x$ and denoted by $\partial f(x)$.

Optimality condition

$$0 \in \partial f(x^*)$$

is sufficient for $x^* \in X$ being a global minimum of convex function $f$.

**Example 4.1.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then, $a$ is a subgradient of $f$ at $x$ if and only if $(a, -1)$ supports* $\mathrm{epi}(f)$ *at $(x, f(x))$.*

**Solution:** Apply the definition of the supporting hyperplane to an epigraph, i.e. use $\gamma = (a, -1)$:

$$\max_{(y,z) \in \mathrm{epi}(f)} a^T y - z \leq a^T x - f(x).$$

Now, realize that $(y, f(y)) \in \mathrm{epi}(f)$ and $f(y)$ is the smallest value of $z$ leading to

$$\forall y \in \mathrm{dom}(f) \;\; a^T y - f(y) \leq a^T x - f(x).$$

Finally, it is sufficient to reorganize the formula to get the definition of subgradient

$$\forall y \in \mathrm{dom}(f) \;\; f(y) - f(x) \geq a^T y - a^T x.$$

**Example 4.2.** (*) *Consider (do not necessarily prove, rather think about) the following properties of subgradient:*

1. *if $f$ is convex, then $\partial f(x) \neq \emptyset$ for all $x \in$* rint dom*$f$.*
2. *if $f$ is convex and differentiable, then $\partial f(x) = \{\nabla f(x)\}$.*
3. *if $\partial f(x) = \{g\}$ (is singleton), then $g = \nabla f(x)$.*
4. *$\partial f(x)$ is a closed convex set.*

**Example 4.3.** *Derive the subdifferential for the following functions:*

$$\begin{array}{rcll} f_1(x) & = & |x|, & \\ f_2(x) & = & x^2 & \text{if } x \leq -1, \\ & & -x & \text{if } x \in [-1, 0], \\ & & x^2 & \text{if } x \geq 0, \\ f_3(x, y) & = & |x + y| & \text{at } (0, 0). \end{array}$$

**Solution:**

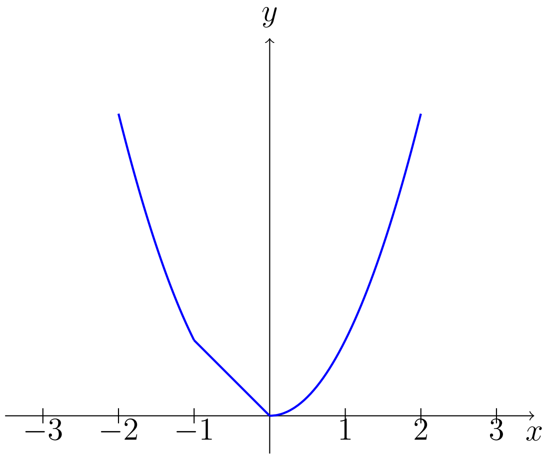
**1.** The function $f_1$ is convex and nondifferentiable at $x = 0$. If we write the definition of subgradient in that point

$$|x| - |0| \geq a(x - 0),$$

we can see that the possible values of $a$ are in $[-1, 1]$. In all other points, the subgradient corresponds to the derivative. So we have

$$\begin{array}{rcll} \partial f_1(x) & = & \{-1\} & \text{if } x \in (-\infty, 0), \\ & & [-1, 1] & \text{if } x = 0, \\ & & \{1\} & \text{if } x \in (0, \infty). \end{array}$$

**2.** We can use the plot of $f_2$




We can see that the function is differentiable almost everywhere with the exception of two points $x \in \{-1, 0\}$, i.e. we have

$$\begin{array}{rcll} \partial f_2(x) & = & \{2x\} & \text{if } x \in (-\infty, -1), \\ & & [-2, -1] & \text{if } x = -1, \\ & & \{-1\} & \text{if } x \in (-1, 0), \\ & & [-1, -0] & \text{if } x = 0, \\ & & \{2x\} & \text{if } x \in (0, \infty). \end{array}$$

Note that $\partial f_2(0) = [f'_{2-}(0), f'_{2+}(0)]$, i.e. it correspond to the interval bounded by one-sided derivatives.

**3.** We can start with the definition and by elaborating possible values we get the explicit formula for $\partial f_3(0, 0)$, i.e.

$$|x + y| - |0 + 0| \geq a_1(x - 0) + a_2(y - 0), \ (x, y) \in \mathbb{R}^2,$$

or simply

$$|x + y| \geq a_1 x + a_2 y, \ (x, y) \in \mathbb{R}^2.$$

Obviously $a_1, a_2 \in [-1, 1]$, otherwise we will get a contradiction immediately (take, e.g., $(x, y) = (1, 0)$). Now consider the case $a_1 \neq a_2$, WLOG $a_1 > a_2$. Consider $x > 0$ and set $y = -x$. Then we have

$$0 = |x - x| < a_1 x - a_2 x = x(a_1 - a_2) > 0,$$

which is a contradiction with the definition of subgradient. We can conclude that

$$\partial f_3(0,0) = \{(a,a) : \ a \in [-1,1]\}.$$

**Lemma 4.4.** *Let* $f_1, \ldots, f_k : \mathbb{R}^n \to \mathbb{R}$ *be convex functions and let*

$$f(x) := f_1(x) + \cdots + f_k(x).$$

*Then it holds*

$$\partial f_1(x) + \cdots + \partial f_k(x) \subseteq \partial f(x).$$

**Solution:** Let $a_1 \in \partial f_1(x), \ldots, a_k \in \partial f_k(x)$. We would like to show that

$$a_1 + \cdots + a_k \in \partial f(x).$$

We can the definition of subgradient: Consider $y \in \mathbb{R}^n$ and write

$$\begin{aligned} f(y) - f(x) &= \sum_{i=1}^{k} f_i(y) - \sum_{i=1}^{k} f_i(x) \\ &= \sum_{i=1}^{k} f_i(y) - f_i(x) \\ &\geq \sum_{i=1}^{k} a_i^T (y - x) \\ &= \left( \sum_{i=1}^{k} a_i \right)^T (y - x), \end{aligned}$$

which confirms that $\sum_{i=1}^{k} a_i \in \partial f(x)$.

# 5 Generalizations of convex functions

We can be faced some optimization problems when the functions in the objective and constraints are not convex, but the problem still posses some global optimality properties. For example, the set of the feasible solutions can still be convex in the case of quasiconvex constraints or the objective may be pseudoconvex. Many optimality conditions, which are sufficient to prove optimality under convexity assumptions, can be valid under some of the generalizations, see Bazaraa et al. (2006).

## 5.1 Quasiconvex functions

**Definition 5.1.** *We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is quasiconvex, if all its level sets are convex, i.e. $\{x \in \mathrm{dom}(f) : \ f(x) \le \alpha\}$ is convex for all $\alpha \in \mathbb{R}$.*

Obviously all convex functions are quasiconvex.

**Example 5.2.** *Find several examples of functions which are quasiconvex, but they are not convex. Try to find an example of function which is not continuous on the interior of its domain (thus it cannot be convex).*

**Solution**: Consider logarithm, which is not convex (it is even concave), but it is quasiconvex, because the level sets $\{x \in (0,\infty) : \ \ln(x) \le \alpha\} = (0, e^\alpha]$ are intervals, thus convex sets.

Consider the cumulative distribution function of a discrete random variable with equiprobable realizations $\xi_1, \xi_2, \xi_3$:

$$F(x) = \frac{1}{3}\sum_{i=1}^{3} I(\xi_i \le x).$$

CDF is not convex, but it is quasiconvex because the $\alpha$-level sets are intervals

$$\{x \in \mathbb{R} : F(x) \le \alpha\} = (-\infty, F^{-1}(\alpha)], \ \alpha \in [0,1],$$

where the generalized quantile (inverse of cdf) is defined as

$$F^{-1}(\alpha) = \min x : \ F(x) \ge \alpha.$$

**Example 5.3.** *Show that the following property is equivalent to the definition of quasiconvexity:*

$$f(\lambda x + (1-\lambda)y) \le \max\{f(x), f(y)\}$$

*for all $x, y \in \mathrm{dom}(f)$ and $\lambda \in [0,1]$.*

**Solution**: Let all level sets of the function be convex. Consider arbitrary $x, y \in \mathrm{dom}(f)$, $\lambda \in (0,1)$ and set $\alpha := \max\{f(x), f(y)\}$. This means that points $x, y$ belong to the $\alpha$-level set. Since the $\alpha$-level set is convex, we have also that the convex combination $\lambda x + (1-\lambda)y$ belongs to it too. This means that

$$f(\lambda x + (1-\lambda)y) \le \alpha = \max\{f(x), f(y)\}, \tag{31}$$

which we wanted to show.

Now let the property (31) be fulfilled. Choose $\alpha$ such that the corresponding level set is nonempty. Take two arbitrary $x, y$ from the $\alpha$-level set and $\lambda \in (0,1)$. Since the property (31) is fulfilled, we have that

$$f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\} \leq \alpha,$$

i.e. the convex combination $\lambda x + (1-\lambda)y$ belongs to the $\alpha$-level set and the set is convex.

**Example 5.4.** *Verify that the following functions are quasiconvex on given sets:*

$$\begin{aligned} f_1(x,y) &= xy \text{ for } (x,y) \in \mathbb{R}_+ \times \mathbb{R}_-, \\ f_2(x) &= \frac{a^T x + b}{c^T x + d} \text{ for } c^T x + d > 0. \end{aligned}$$

**Solution**:
**1.** We use the original definition and investigate the convexity of the level sets:

$$\{(x,y) \in \mathbb{R}_+ \times \mathbb{R}_- : \ xy \leq \alpha\}.$$

For $\alpha \geq 0$ the level set equals to the whole quadrant $\mathbb{R}_+ \times \mathbb{R}_-$, i.e. it is convex. When $\alpha < 0$, we have

$$\left\{(x,y) \in (0,\infty) \times (-\infty, 0) : \ y \leq \frac{\alpha}{x}\right\},$$

which is also a convex set (use a plot, if necessary). Thus the function is quasiconvex.
**2.** It is easier to use the first definition and consider a nonempty $\alpha$-level set

$$\left\{x \in \mathbb{R}^n : \ c^T x + d > 0, \ \frac{a^T x + b}{c^T x + d} \leq \alpha\right\}.$$

The constraint can be rewritten as

$$a^T x + b \leq \alpha(c^T x + d),$$

or

$$(a - \alpha c)^T x + b - \alpha d \leq 0,$$

which is a halfspace, same as $c^T x + d > 0$, thus a convex set.

**Example 5.5.** *Continuous function $f : \mathbb{R} \to \mathbb{R}$ is quasiconvex if and only if one of the following conditions holds*

- *$f$ is nondecreasing,*
- *$f$ is nonincreasing,*
- *there is a $c \in \mathbb{R}$ such that $f$ is nonincreasing on $(-\infty, c]$ and nondecreasing on $[c, \infty)$.*

**Solution**: Realize that the level sets of the above described cases are intervals which are the only convex subsets of $\mathbb{R}$. In any other case you arrive for some $\alpha$ to a union of disjoint intervals which is not a convex set.

**Example 5.6.** (*) *Let $f$ be differentiable. Show that $f$ is quasiconvex if and only if it holds*

$$f(y) \leq f(x) \implies \nabla f(x)^T (y - x) \leq 0.$$

**Example 5.7.** (*) *Let $f$ be a differentiable quasiconvex function. Show that the condition*

$$\nabla f(\overline{x}) = 0$$

*does not imply that $\overline{x}$ is a local minimum of $f$. Find a counterexample.*

**Example 5.8.** *Let $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ be quasiconvex functions, $g : \mathbb{R} \to \mathbb{R}$ be a nondecreasing function and $t \geq 0$ be a scalar. Prove that the following operations preserve quasiconvexity:*

1. $t\, f_1$,
2. $\max\{f_1, f_2\}$,
3. $g \circ f_1$.

**Solution**:
**1.** Consider the alternative definition (31), i.e. it holds

$$f_1(\lambda x + (1 - \lambda) y) \leq \max\{f_1(x), f_1(y)\}$$

for all $x, y \in \mathrm{dom}(f_1)$ and $\lambda \in [0, 1]$. If we multiply the inequality by nonnegative scalar $t$, we obtain

$$t\, f_1(\lambda x + (1 - \lambda) y) \leq t\, \max\{f_1(x), f_1(y)\} = \max\{t\, f_1(x), t\, f_1(y)\},$$

i.e. the function $t\, f_1$ is also quasiconvex.
**2.** Left to the readers.
**3.** We can again use that

$$f_1(\lambda x + (1 - \lambda) y) \leq \max\{f_1(x), f_1(y)\}$$

for all $x, y \in \mathrm{dom}(f_1)$ and $\lambda \in [0, 1]$ and apply nondecreasing function $g$

$$g\Big(f_1(\lambda x + (1 - \lambda) y)\Big) \leq g\Big(\max\{f_1(x), f_1(y)\}\Big) = \max\Big\{g\big(f_1(x)\big), g\big(f_1(y)\big)\Big\},$$

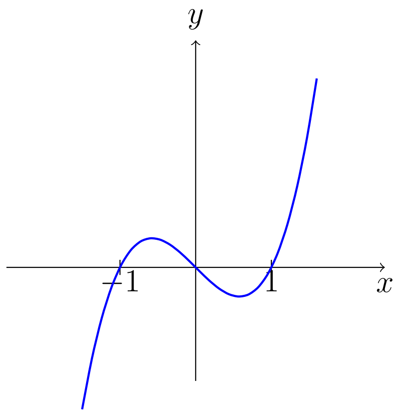
which confirms the quasiconvexity of the composition.

**Example 5.9.** *Let $f_1, f_2$ be quasiconvex functions. Find counterexamples that the following operations DO NOT preserve quasiconvexity:*

*1.* $f_1 + f_2$,

*2.* $f_1 f_2$.

**Solution**:
**1.** Consider $f_1(x) = x^3$ and $f_2(x) = -x$, which are obviously both quasiconvex. However, if we plot their sum, we obtain




which is not a quasiconvex function, because it does not fulfill Lemma 5.5.
**2.** Left to the readers.

**Example 5.10.** *Verify that the following functions are quasiconvex on given sets:*

$$\begin{aligned}
f_1(x,y) &= \frac{1}{xy} \text{ on } \mathbb{R}^2_{++},\\
f_2(x,y) &= \frac{x}{y} \text{ on } \mathbb{R}^2_{++},\\
f_3(x,y) &= \frac{x^2}{y} \text{ on } \mathbb{R} \times \mathbb{R}_{++},\\
f_4(x,y) &= \sqrt{|x+y|} \text{ on } \mathbb{R}^2.
\end{aligned}$$

**Solution**:
**1.** Use one of the definitions.
**2.** Use one of the definitions.
**3.** Use one of the definitions.
**4.** Since $|x+y|$ is convex, therefore quasiconvex, and $\sqrt{\cdot}$ is nondecreasing, we can use the rule and conclude that $f_4$ is quasiconvex.

**Example 5.11.** *Let $S$ be a nonempty convex subset of $\mathbb{R}^n$, $g : S \to \mathbb{R}_+$ be convex and $h : S \to (0, \infty)$ be concave. Show that the function defined by*

$$f(x) = \frac{g(x)}{h(x)}$$

*is quasiconvex on $S$.*

**Solution**: We can show that the $\alpha$-level sets are convex: for $\alpha \geq 0$ the constraint can be rewritten as

$$\left\{x \in S: \ \frac{g(x)}{h(x)} \leq \alpha\right\} = \left\{x \in S: \ g(x) - \alpha\, h(x) \leq 0\right\},$$

where $g(x)$ is assumed to be convex and $-\alpha\, h(x)$ is also convex. Therefore their sum is a convex function and we have its 0-level set. Note that the attempt to verify the alternative definition was usually not successful.

### 5.1.1 Additional examples: strictly quasiconvex functions

**Definition 5.12.** *We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is strictly quasiconvex if*

$$f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\}$$

*for all $x, y$ with $f(x) \neq f(y)$ and $\lambda \in (0,1)$.*

**Lemma 5.13.** (*) *Let $f$ be strictly quasiconvex and $S$ be a convex set. Then any local minimum $\overline{x}$ of $\min_{x \in S} f(x)$ is also a global minimum.*

## 5.2 Pseudoconvex functions

**Definition 5.14.** *Consider $S \subset \mathbb{R}^n$ a nonempty open set. We say that differentiable function $f : S \to \mathbb{R}$ is pseudoconvex with respect to $S$ if it holds*

$$\nabla f(x)^T (y - x) \geq 0 \implies f(y) \geq f(x)$$

*for all $x, y \in S$.*

**Remark 5.15.** *Sometimes it is useful to employ an alternative expression of the definition*

$$f(y) < f(x) \implies \nabla f(x)^T (y - x) < 0$$

*for all $x, y \in S$.*

**Example 5.16.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Show that if $f$ is convex, then it is also pseudoconvex on its domain.*

**Solution:** We know that for every differentiable convex function it holds

$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \ x, y \in \mathrm{dom}(f).$$

Then obviously $\nabla f(x)^T (y - x) \geq 0$ implies that $f(y) \geq f(x)$.

**Example 5.17.** (*) *Find a pseudoconvex function which is not convex.*

Hint: Consider increasing functions.

**Example 5.18.** *Use the definition to show that the following fractional linear function is pseudoconvex:*

$$f(x) = \frac{a^T x + b}{c^T x + d} \text{ for } c^T x + d > 0.$$

**Solution:** We focus on the univariate case, i.e. $x \in \mathbb{R}$ and

$$f(x) = \frac{ax + b}{cx + d} \text{ for } cx + d > 0.$$

We compute the derivative

$$f'(x) = \frac{a(cx + d) - (ax + b)c}{(cx + d)^2}.$$

We would like to show that $f'(x)(y - x) \geq 0$ implies $f(y) \geq f(x)$, i.e.

$$\Big(a(cx + d) - (ax + b)c\Big)(y - x) = ady + bcx - adx - bcy \geq 0$$

implies

$$\frac{ay + b}{cy + d} \geq \frac{ax + b}{cx + d},$$

which is on the domain equivalent to

$$(ay + b)(cx + d) - (ax + b)(cy + d) = ady + bcx - adx - bcy \geq 0.$$

The conditions are obviously the same.

**Example 5.19.** (*) *Consider fractional function $f$ as defined in Example 5.11. Moreover, let $S$ be open and $g$, $h$ be differentiable on $S$. Show that $f$ is pseudoconvex.*

**Example 5.20.** *Let $f : S \to \mathbb{R}$ be a differentiable function. Show that if $f$ is pseudoconvex, than it is also quasiconvex.*

**Solution:** We will verify the alternative definition from Example 5.3. Take $x, y \in S$ and $\lambda \in (0, 1)$. Set $z = \lambda x + (1 - \lambda)y$. If $f(z) \leq f(x)$, we are finished. So assume that $f(z) > f(x)$. Since the function is quasiconvex, we have

$$\nabla f(z)^T (x - z) < 0.$$

Consider

$$\begin{aligned} x - z &= x - \lambda x - (1 - \lambda)y = (1 - \lambda)(x - y), \\ y - z &= y - \lambda x - (1 - \lambda)y = -\lambda(x - y), \end{aligned}$$

hence

$$y - z = \frac{-\lambda}{1 - \lambda}(x - z),$$

and using the properties of the scalar product we obtain

$$\nabla f(z)^T(y-z) = \frac{-\lambda}{1-\lambda}\nabla f(z)^T(x-z) > 0.$$

Function $f$ is pseudoconvex, therefore $f(z) \leq f(y)$. Hence we have

$$f(z) \leq \max\{f(x), f(y)\}.$$

**Example 5.21.** *Show that if $\nabla f(\overline{x}) = 0$ for a pseudoconvex $f$, then $\overline{x}$ is a global minimum of $f$.*

**Solution:** The condition $\nabla f(\overline{x})^T(y-\overline{x}) = 0 \geq 0$ is fulfilled which implies that $f(y) \geq f(\overline{x})$ for all $y$ from the domain.

**Example 5.22.** (*) *The following table summarizes relations between the stationary points and minima of a differentiable function $f$:*

| | | | | | |
|---|---|---|---|---|---|
| *$f$ general:* | *$\overline{x}$ global min.* | $\Longrightarrow$ | *$\overline{x}$ local min.* | $\Longrightarrow$ | $\nabla f(\overline{x}) = 0$ |
| *$f$ quasiconvex:* | *$\overline{x}$ global min.* | $\Longrightarrow$ | *$\overline{x}$ local min.* | $\Longrightarrow$ | $\nabla f(\overline{x}) = 0$ |
| *$f$ strictly quasiconvex:* | *$\overline{x}$ global min.* | $\Longleftrightarrow$ | *$\overline{x}$ local min.* | $\Longrightarrow$ | $\nabla f(\overline{x}) = 0$ |
| *$f$ pseudoconvex:* | *$\overline{x}$ global min.* | $\Longleftrightarrow$ | *$\overline{x}$ local min.* | $\Longleftrightarrow$ | $\nabla f(\overline{x}) = 0$ |
| *$f$ convex:* | *$\overline{x}$ global min.* | $\Longleftrightarrow$ | *$\overline{x}$ local min.* | $\Longleftrightarrow$ | $\nabla f(\overline{x}) = 0.$ |

*For details see Bazaraa et al. (2006).*

# 6 Optimality conditions

## 6.1 Optimality conditions based on directions

**Example 6.1.** *Consider the global optimization problem*

$$\min 2x_1^2 - x_1x_2 + x_2^2 - 3x_1 + e^{2x_1+x_2}.$$

*Find a descent direction at point (0,0) and formulate a minimization problem in that direction.*

**Solution:** Compute the gradient, i.e. denote the objective function by $f$ and

$$\begin{aligned}\frac{\partial f}{\partial x_1} &= 4x_1 - x_2 - 3 + 2e^{2x_1+x_2},\\ \frac{\partial f}{\partial x_2} &= -x_1 + 2x_2 + e^{2x_1+x_2}.\end{aligned}$$

We have that

$$\nabla f(0,0) = \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

which is not zero vector, i.e. $-\nabla f(0,0)$ is a (local) descent direction. Below we will describe the whole set of descent directions. Using step length $t \geq 0$ we move to a new point in the descent direction as

$$\begin{pmatrix} x_1^t \\ x_2^t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Then we can solve the following one dimensional problem with decision variable $t$

$$\min_{t\geq 0} f(x_1^t, x_2^t).$$

Note that many algorithms for solving nonlinear programming problems work in this way and they differ by the choices of the descent direction. In our case, it corresponds to the gradient (first order) algorithm.

**Example 6.2.** *Verify the optimality conditions at point* $(2,4)$ *for problem*

$$\begin{aligned}\min\ & (x_1-4)^2 + (x_2-6)^2\\ \text{s.t. } & x_1^2 \leq x_2,\\ & x_2 \leq 4.\end{aligned}$$

*Consider the same point for the problem with the second inequality constraint in the form*

$$x_2 \leq 5.$$

**Solution:** Use the basic optimality conditions derived for a convex objective function and a convex set of feasible solutions, see Theorem 3.3 in Lecture notes. Denote the objective function by $f$ and compute its gradient

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= 2(x_1 - 4), \\ \frac{\partial f}{\partial x_2} &= 2(x_2 - 6), \end{aligned}$$

hence $\nabla f(2,4) = (-4,-4)^T$. The optimality condition stands

$$\nabla f(2,4)^T \begin{pmatrix} x_1 - 2 \\ x_2 - 4 \end{pmatrix} = -4(x_1 - 2) - 4(x_2 - 4) \geq 0,$$

for all feasible solutions, which is fulfilled because $|x_1| \leq 2$ and $x_2 \leq 4$, i.e. $(2,4)$ is a point of global minima.

If we change the last constraint, the optimality condition remains valid, however the set of feasible solutions is $|x_1| \leq \sqrt{5}$ and $x_2 \leq 5$, i.e. by choosing $x_1 = 2$ and $x_2 = 5$, we obtain

$$-4(2-2) - 4(5-4) < 0,$$

i.e. the condition is not fulfilled.

**Example 6.3.** *Consider open* $\emptyset \neq S \subseteq \mathbb{R}^n$, $f : S \to \mathbb{R}$, *and define set of improving directions of* $f$ *at* $x \in S$

$$F_f(x) = \{s \in \mathbb{R}^n : \; s \neq 0, \; \exists \delta > 0 \; \forall 0 < \lambda < \delta : \; f(x + \lambda s) < f(x)\}.$$

1. *For a differentiable* $f$, *define the inner approximation*

$$F_{f,0}(x) = \{s \in \mathbb{R}^n : \; \langle \nabla f(x), s \rangle < 0\}.$$

*Show that it holds*

$$F_{f,0}(x) \subseteq F_f(x).$$

2. *Moreover, if* $f$ *is pseudoconvex at* $x$ *with respect to a neighborhood of* $x$, *then*

$$F_{f,0}(x) = F_f(x).$$

3. *If* $f$ *is convex, then*

$$F_f(x) = \{\alpha(y - x) : \; \alpha > 0, \; f(y) < f(x), \; y \in S\}.$$

4. *For a differentiable* $f$, *define the outer approximation*

$$F'_{f,0}(x) = \{s \in \mathbb{R}^n : \; s \neq 0, \; \langle \nabla f(x), s \rangle \leq 0\}.$$

*Show that it holds*

$$F_f(x) \subseteq F'_{f,0}(x).$$

**Solution:** We will use the scalarization function:

$$\varphi_{x,s}(\lambda) = f(x + \lambda s),$$

which is defined on $D_{x,s} = \{\lambda \in \mathbb{R} : \ x + \lambda s \in S\}$. If $f$ is differentiable at $x$, then $\varphi_{x,s}$ is differentiable at 0 and it holds

$$\varphi'_{x,s}(0) = \nabla f(x)^T s.$$

**1.** Take $s \in F_{f,0}(x)$, hence

$$\varphi'_{x,s}(0) = \nabla f(x)^T s < 0,$$

i.e. $\varphi_{x,s}(\lambda)$ is decreasing on some $\delta$-neighborhood of 0 for some $\delta > 0$, which means that

$$f(x + \lambda s) < f(x), \ 0 < \lambda < \delta.$$

This means that $s \in F_f(x)$.
**2.** Let $f$ be pseudoconvex with respect to $\delta'$-neighborhood of $x$ for some $\delta' > 0$. Consider $s \in F_f(x)$, which means that

$$f(x + \lambda s) < f(x), \ 0 < \lambda < \delta, \ \text{for some } \delta > 0.$$

Together with psedoconvexity this implies that

$$\nabla f(x)^T (x + \lambda s - x) = \lambda \nabla f(x)^T s < 0, \ \text{for } \lambda \in (0, \delta''),$$

where $\delta'' = \min\{\delta / \|s\|, \delta'\}$. This means that $\nabla f(x)^T s < 0$ and $s \in F'_{f,0}(x)$.
**3.** Left to the readers.
**4.** The proof is in the reverse order of 1. when

$$f(x + \lambda s) < f(x), \ 0 < \lambda < \delta$$

implies that

$$\varphi'_{x,s}(0) = \nabla f(x)^T s \leq 0.$$

**Example 6.4.** *Consider the global optimization problem from Example 6.1*

$$\min 2x_1^2 - x_1 x_2 + x_2^2 - 3x_1 + e^{2x_1 + x_2}.$$

*Derive the set of improving directions at (0,0).*

**Solution:** We know that

$$\begin{aligned} \frac{\partial f}{\partial x_1}(0,0) &= -1, \\ \frac{\partial f}{\partial x_2}(0,0) &= 1, \end{aligned}$$

i.e. we get

$$\begin{aligned} F_{f,0}(0,0) &= \{s \in \mathbb{R}^2 : \ s_1 > s_2\}, \\ F'_{f,0}(0,0) &= \{s \in \mathbb{R}^2 : \ s \neq 0, \ s_1 \geq s_2\}. \end{aligned}$$

Since $f$ is convex, thus pseudoconvex, we have that

$$F_f(x) = F_{f,0}(x).$$

**Example 6.5.** *Consider the global optimization problem*

$$\min 2x_1^2 - 3x_1x_2^2 + x_2^4.$$

*Derive the set of improving directions at (0,0).*

**Solution:** Compute the gradient of the objective function

$$\begin{aligned}\frac{\partial f}{\partial x_1} &= 4x_1 - 3x_2^2|_{(0,0)} = 0,\\ \frac{\partial f}{\partial x_2} &= -6x_1x_2 + 4x_2^3|_{(0,0)} = 0.\end{aligned}$$

Since we obtained zero gradient, realize that the approximating sets are trivial

$$\begin{aligned}&F_{f,0}(0,0) = \emptyset,\\ &F'_{f,0}(0,0) = \mathbb{R}^2 \setminus \{(0,0)\}.\end{aligned}$$

We should verify whether $(0,0)$ is a point of local minima. However, the Hessian matrix does not answer the question and higher order derivatives are necessary.

We focus on local approximation of the set of feasible solutions.

**Example 6.6.** *Consider open* $\emptyset \neq S \subseteq \mathbb{R}^n$*, functions* $g_i : S \to \mathbb{R}$*, and the set of feasible solutions*

$$M = \{x \in S : \ g_i(x) \leq 0, \ i = 1, \ldots, m\}.$$

*Define the set of feasible directions of* $M$ *at* $x \in M$

$$D_M(x) = \{s \in \mathbb{R}^n : \ s \neq 0, \ \exists\delta > 0 \ \forall 0 < \lambda < \delta : \ x + \lambda s \in M\}.$$

1. *If* $M$ *is a convex set, then*

$$D_M(x) = \{\alpha(y - x) : \ \alpha > 0, \ y \in M, \ y \neq x\}.$$

2. *For differentiable* $g_i$ *and* $x \in M$ *define*[3]

$$\begin{aligned}G_{g,0}(x) &= \{s \in \mathbb{R}^n : \ \langle \nabla g_i(x), s \rangle < 0, \ i \in I_g(x)\},\\ G'_{g,0}(x) &= \{s \in \mathbb{R}^n : \ s \neq 0, \ \langle \nabla g_i(x), s \rangle \leq 0, \ i \in I_g(x)\}.\end{aligned}$$

*In general, it holds*[4]

$$G_{g,0}(x) \subseteq D_M(x) \subseteq G'_{g,0}(x).$$

**Solution:**
**1.** Left to the readers.
**2.** We will use the scalarization functions: for $i \in I_g(x)$ define

$$\varphi_{i,x,s}(\lambda) = g_i(x + \lambda s),$$

[3]Set of indices of active constraints: $I_g(x) = \{i : \ g_i(x) = 0\}$

[4]Note that there are sufficient conditions stating which approximation is tight, see Bazaraa et al. (2006) for details.

which is defined on $D_{x,s} = \{\lambda \in \mathbb{R} : \; x + \lambda s \in S\}$. If $g_i$ is differentiable at $x$, then $\varphi_{i,x,s}$ is differentiable at 0 and it holds

$$\varphi'_{i,x,s}(0) = \nabla g_i(x)^T s.$$

Let $x \in M$ and $s \in G_{g,0}(x)$. This means that for $i \in I_g(x)$ with $g_i(x) = 0$ we have

$$\varphi'_{i,x,s}(0) = \nabla g_i(x)^T s < 0,$$

i.e. there is $\delta_i > 0$ such that $g_i$ is decreasing in direction $s$ from $x$ using step lengths $\lambda \in (0, \delta_i)$, thus $g_i(x + \lambda s) \leq 0$ for $\lambda \in (0, \delta_i)$. For $i \notin I_g(x)$ with $g_i(x) < 0$, differentiability (hence continuity) implies that there is $\delta_i > 0$ such that $g_i(x+\lambda s) \leq 0$ for $\lambda \in (0, \delta_i)$. It is enough to consider $\delta = \min_i \delta_i$ to show that $s \in D_M(x)$.
Proof of $D_M(x) \subseteq G'_{g,0}(x)$ is left to the readers.

**Example 6.7.** *Derive the above defined sets of directions for the sets*

$$M_1 = \{(x, y) : \; -(x-2)^2 \geq y-2, \; -(y-2)^2 \geq x-2\},$$
$$M_2 = \{(x, y) : \; (x-2)^2 \geq y-2, \; (y-2)^2 \geq x-2\},$$

*at point (2,2). Use the pictures to decide which of the approximations are tight.*

**Solution:** Consider $M_1$ and compute

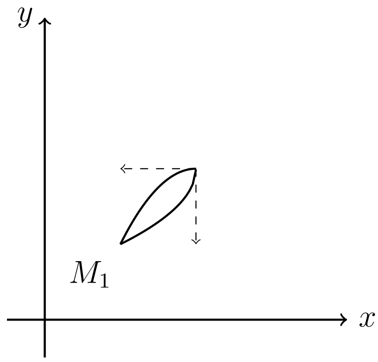
$$\nabla g_1(2,2) = \begin{pmatrix} 2(x-2) \\ 1 \end{pmatrix} |_{(2,2)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \; \nabla g_2(2,2) = \begin{pmatrix} 1 \\ 2(y-2) \end{pmatrix} |_{(2,2)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

So we have

$$G_{g,0}(2,2) = \{s \in \mathbb{R}^2 : \; s_1 < 0, s_2 < 0\},$$
$$G'_{g,0}(2,2) = \{s \in \mathbb{R}^2 : \; s \neq 0, \; s_1 \leq 0, s_2 \leq 0\}.$$

Using the picture, we can conclude that the tangent directions do not belong to the set of feasible directions, i.e.

$$D_{M_1}(2,2) = G_{g,0}(2,2).$$




Consider $M_2$ and compute

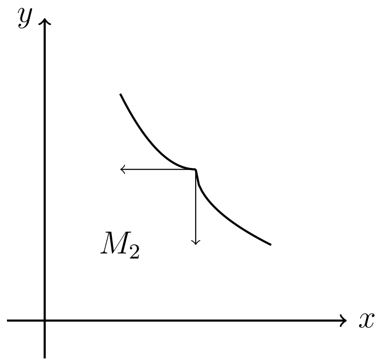
$$\nabla g_1(2,2) = \begin{pmatrix} -2(x-2) \\ 1 \end{pmatrix} |_{(2,2)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \; \nabla g_2(2,2) = \begin{pmatrix} 1 \\ -2(y-2) \end{pmatrix} |_{(2,2)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

So we have

$$G_{g,0}(x) = \{s \in \mathbb{R}^2 : \ s_1 < 0, s_2 < 0\},$$
$$G'_{g,0}(x) = \{s \in \mathbb{R}^2 : \ s \neq 0, \ s_1 \leq 0, s_2 \leq 0\}.$$

Using the picture, we can conclude that the tangent directions belong to the set of feasible directions, i.e.

$$D_{M_2}(2,2) = G'_{g,0}(2,2).$$




**Example 6.8.** (*) *Derive the above defined sets of directions for a polyhedral set*

$$M = \{x \in \mathbb{R}^n : \ Ax \leq b\}.$$

**Example 6.9.** *Derive the above defined sets of directions for the problem*

$$\begin{aligned} \min\ & (x_1-3)^2 + (x_2-2)^2 \\ \text{s.t. } & x_1^2 + x_2^2 \leq 5, \\ & x_1 + x_2 \leq 3, \\ & x_1 \geq 0, \ x_2 \geq 0, \end{aligned}$$

*at point (2,1). Discuss the intersection of the sets of directions which is an optimality condition. Apply the Farkas theorem to the conditions on directions.*

**Solution:** Denote

$$\begin{aligned} g_1(x) &= x_1^2 + x_2^2 - 5, \\ g_2(x) &= x_1 + x_2 - 3, \\ g_3(x) &= -x_1, \\ g_4(x) &= -x_2, \end{aligned}$$

and realize that $I_g(2,1) = \{1,2\}$. Now compute

$$\nabla g_1(2,1) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}|_{(2,1)} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}, \ \nabla g_2(2,1) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

So we have

$$G_{g,0}(x) = \{s \in \mathbb{R}^2 : \ 2s_1 + s_2 < 0, s_1 + s_2 < 0\},$$
$$G'_{g,0}(x) = \{s \in \mathbb{R}^2 : \ s \neq 0, \ 2s_1 + s_2 \leq 0, s_1 + s_2 \leq 0\}.$$

Using the picture, we can conclude that

$$D_M(2,1) = \{s \in \mathbb{R}^2 : \ s \neq 0, \ 2s_1 + s_2 < 0, s_1 + s_2 \leq 0\}.$$

Note that this means that

$$G_{g,0}(x) \subsetneq D_M(x) \subsetneq G'_{g,0}(x).$$

Denote $f(x) = (x_1 - 3)^2 + (x_2 - 2)^2$ and

$$\nabla f(2,1) = \begin{pmatrix} 2(x_1 - 3) \\ 2(x_2 - 2) \end{pmatrix} |_{(2,1)} = \begin{pmatrix} -2 \\ -2 \end{pmatrix}.$$

So we have

$$\begin{aligned} F_{f,0}(2,1) &= \{s \in \mathbb{R}^2 : \ s_1 + s_2 > 0\}, \\ F'_{f,0}(2,1) &= \{s \in \mathbb{R}^2 : \ s \neq 0, \ s_1 + s_2 \geq 0\}. \end{aligned}$$

Since the objective function is convex, thus pseudoconvex, we know that

$$F_f(2,1) = F_{f,0}(2,1).$$

We can see that it holds

$$F_f(2,1) \cap D_M(2,1) = \emptyset,$$

which is necessary and under convexity also sufficient optimality condition based on the sets of directions, i.e. there is no feasible direction from $(2,1)$ to the set in which the objective function decreases.

Now we can apply the Farkas theorem to

$$\forall s \in \mathbb{R}^2 \ \begin{pmatrix} -\nabla g_1(2,1)^T \\ -\nabla g_2(2,1)^T \end{pmatrix} s \geq 0 \ \Rightarrow \ \nabla f(2,1)^T s \geq 0,$$

i.e. any feasible direction (with respect to the outer approximation) is not an improving direction. The implication is fulfilled if and only if system

$$-\nabla g_1(2,1)u_1 - \nabla g_2(2,1)u_2 = \nabla f(2,1)$$

has a nonnegative solution. After reorganizing the terms, we obtain Karush-Kuhn-Tucker (KKT) conditions:

$$\nabla f(2,1) + \nabla g_1(2,1)u_1 + \nabla g_2(2,1)u_2 = 0, \ u_{1,2} \geq 0.$$

We emphasize that this is just an outline and we refer the readers to the lecture notes for a precise derivation. We can compute the Lagrange multipliers $u_1, u_2$ by solving

$$\begin{aligned} &-2 + 4u_1 + u_2 = 0, \\ &-2 + 2u_1 + u_2 = 0. \end{aligned}$$

We obtain $u_1 = 0, u_2 = 2$ which are both nonnegative.

**Example 6.10.** *Derive the sets of directions for the problem*

$$\begin{aligned} &\min\ (x_1-3)^2+(x_2-3)^2 \\ &\text{s.t. } x_1^2+x_2^2=4, \end{aligned}$$

*at point* $(\sqrt{2},\sqrt{2})$. *Consider the set of feasible directions for equality constraints* $h_j(x)=0,\ j=1,\dots,l$, *where* $h_j: S\to\mathbb{R}$ *are differentiable:*

$$H_{h,0}(x)=\{s\in\mathbb{R}^n:\ s\neq 0,\ \langle\nabla h_j(x),s\rangle=0,\ j=1,\dots,l\}.$$

**Solution:** We have

$$\begin{aligned} \nabla f(\sqrt{2},\sqrt{2}) &= \begin{pmatrix} 2(x_1-3) \\ 2(x_2-3) \end{pmatrix}\Big|_{(\sqrt{2},\sqrt{2})} = \begin{pmatrix} 2(\sqrt{2}-3) \\ 2(\sqrt{2}-3) \end{pmatrix}, \\ \nabla h(\sqrt{2},\sqrt{2}) &= \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}\Big|_{(\sqrt{2},\sqrt{2})} = \begin{pmatrix} 2\sqrt{2} \\ 2\sqrt{2} \end{pmatrix} \end{aligned}$$

Then obviously

$$\begin{aligned} F_f(\sqrt{2},\sqrt{2}) &= \{s\in\mathbb{R}^2:\ s_1+s_2>0\}, \\ H_{h,0}(\sqrt{2},\sqrt{2}) &= \{s\in\mathbb{R}^2:\ s\neq 0,\ s_1+s_2=0\}, \end{aligned}$$

and

$$F_f(\sqrt{2},\sqrt{2})\cap H_{h,0}(\sqrt{2},\sqrt{2})=\emptyset,$$

i.e. the optimality condition is fulfilled.

## 6.2 Karush–Kuhn–Tucker optimality conditions

### 6.2.1 A few pieces of the theory

We emphasize that this section contains just a basic summary and we refer the readers to the lecture notes for formal definitions and propositions.

Consider a **nonlinear programming problem** with inequality and equality constraints:

$$\begin{aligned} \min\ & f(x) \\ \text{s.t. } & g_i(x) \le 0,\ i = 1, \dots, m, \\ & h_j(x) = 0,\ j = 1, \dots, l, \end{aligned} \tag{32}$$

where $f, g_i, h_j : \mathbb{R}^n \to \mathbb{R}$ are differentiable functions. We denote by $M$ the set of feasible solutions.

Define the **Lagrange function** by

$$L(x, u, v) = f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{j=1}^{l} v_j h_j(x),\ u_i \ge 0. \tag{33}$$

The **Karush–Kuhn–Tucker optimality conditions** are then (feasibility, complementarity and optimality):

$$\begin{aligned} \text{i) } & g_i(x) \le 0,\ i = 1, \dots, m,\ h_j(x) = 0,\ j = 1, \dots, l, \\ \text{ii) } & u_i g_i(x) = 0,\ u_i \ge 0,\ i = 1, \dots, m, \\ \text{iii) } & \nabla_x L(x, u, v) = 0, \end{aligned} \tag{34}$$

Any point $(x, u, v)$ which fulfills the above conditions is called a KKT point.

If a Constraint Qualification (CQ) condition is fulfilled, then the KKT conditions are necessary for local optimality of a point. Basic CQ conditions are:

- **Slater CQ:** $\exists \tilde{x} \in M$ such that $g_i(\tilde{x}) < 0$ for all $i$ and the gradients $\nabla_x h_j(\tilde{x})$, $j = 1, \dots, l$ are linearly independent.
- **Linear independence CQ** at $\hat{x} \in M$: all gradients

  $$\nabla_x g_i(\hat{x}),\ i \in I_g(\hat{x}),\ \nabla_x h_j(\hat{x}),\ j = 1, \dots, l$$

  are linearly independent.

These conditions are quite strong and are sufficient for weaker CQ conditions, e.g. the Kuhn–Tucker condition (Mangasarian–Fromovitz CQ, Abadie CQ, ...).

Consider the set of active (inequality) constraints and its partitioning

$$\begin{aligned} I_g(x) &= \{i :\ g_i(x) = 0\}, \\ I_g^0(x) &= \{i :\ g_i(x) = 0,\ u_i = 0\}, \\ I_g^+(x) &= \{i :\ g_i(x) = 0,\ u_i > 0\}, \end{aligned} \tag{35}$$

i.e.

$$I_g(x) = I_g^0(x) \cup I_g^+(x).$$

Let all functions be twice differentiable. We say that the **second-order sufficient condition** (SOSC) is fulfilled at a KKT point $(x, u, v)$ if for all $0 \neq z \in \mathbb{R}^n$ such that

$$\begin{aligned} &z^T \nabla_x g_i(x) = 0,\ i \in I_g^+(x), \\ &z^T \nabla_x g_i(x) \leq 0,\ i \in I_g^0(x), \\ &z^T \nabla_x h_j(x) = 0,\ j = 1, \ldots, l, \end{aligned} \tag{36}$$

it holds

$$z^T\, \nabla^2_{xx} L(x, u, v)\, z > 0. \tag{37}$$

Then $x$ is a strict local minimum of the nonlinear programming problem (32).

**To summarize**, we are going to practice the following relations:

1. KKT point and convex problem $\rightarrow$ global optimality at $x$.
2. KKT point and SOSC $\rightarrow$ (strict) local optimality at $x$.
3. Local optimality at $x$ and a constraint qualification (CQ) condition $\rightarrow \exists(u, v)$ such that $(x, u, v)$ is a KKT point.

### 6.2.2 Karush–Kuhn–Tucker optimality conditions

**Example 6.11.** *Consider the nonlinear programming problems from Example 6.2. Compute the Lagrange multipliers at given points.*

**Solution:** Define the Lagrange function

$$L(x_1, x_2, u_1, u_2) = (x_1 - 4)^2 + (x_2 - 6)^2 + u_1(x_1^2 - x_2) + u_2(x_2 - 4),\ u_{1,2} \geq 0.$$

Write the KKT optimality conditions:

i) feasibility,

ii) complementarity

$$u_1(x_1^2 - x_2) = 0,\ u_2(x_2 - 4) = 0,\ u_{1,2} \geq 0,$$

iii) optimality

$$\begin{aligned} \frac{\partial L}{\partial x_1} &= 2(x_1 - 4) + 2u_1 x_1 = 0, \\ \frac{\partial L}{\partial x_2} &= 2(x_2 - 6) - u_1 + u_2 = 0. \end{aligned}$$

Now, consider point $(2,4)$ and compute the Lagrange multipliers. Since both constraints are active, i.e. $I_g(2,4) = \{1,2\}$, we must use the optimality iii) and we get

$$\begin{aligned} &-4 + 4u_1 = 0, \\ &-4 - u_1 + u_2 = 0, \end{aligned}$$

and obtain $u_1 = 1 \geq 0$ and $u_2 = 5 \geq 0$. We have obtained KKT point $(2,4,1,5)$ and $(2,4)$ is a global solution, because the problem is convex.

**Example 6.12.** (*) *Consider the nonlinear programming problems from Example 6.9. Compute the Lagrange multipliers at given points.*

**Example 6.13.** *Using the KKT conditions find the closest point to (0,0) in the set defined by*

$$M = \{x \in \mathbb{R}^2 : \ x_1 + x_2 \geq 4, \ 2x_1 + x_2 \geq 5\}.$$

*Can several points (solutions) exist?*

**Solution:** We formulate a nonlinear programming problem using the Euclidean distance in the objective[5]:

$$\begin{aligned} \min\ & x_1^2 + x_2^2 \\ \text{s.t.}\ & -x_1 - x_2 + 4 \leq 0, \\ & -2x_1 - x_2 + 5 \leq 0. \end{aligned}$$

Write the Lagrange function

$$L(x_1, x_2, u_1, u_2) = x_1^2 + x_2^2 + u_1(-x_1 - x_2 + 4) + u_2(-2x_1 - x_2 + 5), \ u_1, u_2 \geq 0.$$

Derive the KKT conditions

$$\begin{aligned} \text{i)}\ & \text{feasibility}, \\ \text{ii)}\ & u_1(-x_1 - x_2 + 4) = 0, \ u_1 \geq 0, \\ & u_2(-2x_1 - x_2 + 5) = 0, \ u_2 \geq 0, \\ \text{iii)}\ & \frac{\partial L}{\partial x_1} = 2x_1 - u_1 - 2u_2 = 0, \\ & \frac{\partial L}{\partial x_2} = 2x_2 - u_1 - u_2 = 0. \end{aligned} \tag{38}$$

Now, we will try to find the KKT point by analyzing the optimality conditions, where we proceed according to the complementarity conditions:
**1.** Set $u_1 = 0$, $u_2 = 0$: We have from iii) that $x_1 = 0$, $x_2 = 0$ which is not feasible point.
**2.** Set $x_1 + x_2 = 4$, $u_2 = 0$: Together with iii) we solve

$$\begin{aligned} 2x_1 - u_1 &= 0, \\ 2x_2 - u_1 &= 0, \\ x_1 + x_2 &= 4. \end{aligned} \tag{39}$$

[5]The square root can be omitted.

We obtain $x_1 = 2$, $x_2 = 2$, $u_1 = 4 > 0$, i.e. we have KKT point $(2, 2, 4, 0)$.
**3.** Set $u_1 = 0$, $2x_1 + x_2 = 5$: Solve

$$\begin{aligned} 2x_1 - 2u_2 &= 0, \\ 2x_2 - u_2 &= 0, \\ 2x_1 + x_2 &= 5. \end{aligned} \tag{40}$$

We get $x_1 = 2$, $x_2 = 1$, $u_2 = 2$, which is not feasible point.
**4.** Set $x_1 + x_2 = 4$, $2x_1 + x_2 = 5$: We get $x_1 = 1$, $x_2 = 3$ and compute the Lagrange multipliers by solving

$$\begin{aligned} u_1 + 2u_2 &= 2, \\ u_1 + u_2 &= 6. \end{aligned} \tag{41}$$

We obtain $u_1 = 10$, $u_2 = -4 < 0$, i.e. the Lagrange multipliers are not nonnegative and (1,3,10,-4) is not KKT point.

Since the set $M$ is convex, the closest point corresponding to the projection $(2, 2)$ must be unique.

**Example 6.14.** *Verify that the point $(x_1, x_2) = (\frac{4}{5}, \frac{8}{5})$ is a local/global solution of the problem*

$$\begin{aligned} \min\ & x_1^2 + x_2^2, \\ s.t.\ & x_1^2 + x_2^2 \leq 5, \\ & x_1 + 2x_2 = 4, \\ & x_1, x_2 \geq 0. \end{aligned}$$

**Solution:** Write the Lagrange function

$$L(x_1, x_2, u_1, u_2) = x_1^2 + x_2^2 + u_1(x_1^2 + x_2^2 - 5) - u_2 x_1 - u_3 x_2 + v(x_1 + 2x_2 - 4),\ u_1, u_2, u_3 \geq 0.$$

Derive the KKT conditions

$$\begin{aligned} &\text{i) feasibility,} \\ &\text{ii) } u_1(x_1^2 + x_2^2 - 5) = 0,\ u_1 \geq 0, \\ &\quad u_2 x_1 = 0,\ u_2 \geq 0, \\ &\quad u_3 x_2 = 0,\ u_3 \geq 0, \\ &\text{iii) } \frac{\partial L}{\partial x_1} = 2x_1 + 2u_1 x_2 - u_2 + v = 0, \\ &\quad \frac{\partial L}{\partial x_2} = 2x_2 + 2u_1 x_2 - u_3 + 2v = 0. \end{aligned} \tag{42}$$

For point $(x_1, x_2) = (\frac{4}{5}, \frac{8}{5})$, we have that $u_{1,2,3} = 0$ (from complementarity conditions, i.e. none of the inequality constraints is active) and $v = -\frac{8}{5}$ which is feasible value for Lagrange multiplier corresponding to equality constraint. So we have obtained KKT point $(\frac{4}{5}, \frac{8}{5}, 0, 0, 0, -\frac{8}{5})$.

Since the objective function and inequality constraints are convex, and the equality constraint is linear (affine), $(x_1, x_2) = (\frac{4}{5}, \frac{8}{5})$ is a global solution.

**Example 6.15.** *Consider the problem*

$$\begin{aligned} \min\ & 2e^{x_1-1} + (x_2 - x_1)^2 + x_3^2 \\ \text{s.t. } & x_1x_2x_3 \le 1, \\ & x_1 + x_3 \ge c, \\ & x \ge 0. \end{aligned}$$

*For which values of $c$ does $\bar{x} = (1, 1, 1)$ with multipliers fulfill the KKT conditions?*

**Solution:** Write the Lagrange function

$L(x_1, x_2, x_3, u_1, u_2) = 2e^{x_1-1}+(x_2-x_1)^2+x_3^2+u_1(x_1x_2x_3-1)+u_2(-x_1-x_3+c),\ u_1, u_2 \ge 0.$

Since we are given point $\bar{x} = (1, 1, 1) > 0$, we can skip corresponding multipliers $u_{3,4,5}$ and related terms, because all these multipliers must be equal to zero (from complementarity). Derive the KKT conditions

$$\begin{aligned} \text{i) } & \text{feasibility,} \\ \text{ii) } & u_1(x_1x_2x_3 - 1) = 0,\ u_1 \ge 0, \\ & u_2(-x_1 - x_3 + c) = 0,\ u_2 \ge 0, \\ \text{iii) } & \frac{\partial L}{\partial x_1} = 2e^{x_1-1} - 2(x_2 - x_1) + u_1x_2x_3 - u_2 = 0, \\ & \frac{\partial L}{\partial x_2} = 2(x_2 - x_1) + u_1x_1x_3 = 0, \\ & \frac{\partial L}{\partial x_3} = 2x_3 + u_1x_1x_2 - u_2 = 0. \end{aligned} \tag{43}$$

So for $\bar{x} = (1, 1, 1)$, we solve

$$\begin{aligned} & 2 + u_1 - u_2 = 0, \\ & u_1 = 0, \\ & 2 + u_1 - u_2 = 0, \end{aligned} \tag{44}$$

and get multipliers $u_1 = 0$, $u_2 = 2$. We must not forget for the feasibility. Obviously first constraint is fulfilled and for the second one it must hold $c \le 2$. However, if $c < 2$, then the second complementarity condition is not fulfilled $u_2(-x_1-x_3+c) \ne 0$. Therefore we have obtained KKT point (1,1,1,0,2) only if $c = 2$ when the second constraint is active.

**Example 6.16.** (*) *Consider the problem*

$$\begin{aligned} \min\ & \frac{x_1 + 3x_2 + 3}{2x_1 + x_2 + 6} \\ \text{s.t. } & 2x_1 + x_2 \le 12, \\ & -x_1 + 2x_2 \le 4, \\ & x_1, x_2 \ge 0. \end{aligned}$$

*Verify that the KKT conditions are fulfilled for all points on the line between (0,0) and (6,0). Are the KKT conditions sufficient for global optimality?*

**Example 6.17.** *Consider the (water-filling[6]) problem*

$$\begin{aligned} \min \ & -\sum_{i=1}^{n} \log(\alpha_i + x_i) \\ \text{s.t.} \ & \sum_{i=1}^{n} x_i = 1 \\ & x_i \geq 0, \end{aligned}$$

*where $\alpha_i > 0$ are parameters. Using the KKT conditions find the solutions.*

**Solution:** First realize that the problem is convex, i.e. the objective is convex and the constraints are linear. Consider the Lagrange function

$$L(x, u, v) = -\sum_{i=1}^{n} \log(\alpha_i + x_i) - \sum_{i=1}^{n} u_i x_i + v \left( \sum_{i=1}^{n} x_i - 1 \right), \ u_i \geq 0, v \in \mathbb{R}.$$

The KKT conditions are:

$$\begin{aligned} \text{i)} \ & \sum_{i=1}^{n} x_i = 1, \ x_i \geq 0, \ i = 1, \dots, n \\ \text{ii)} \ & u_i x_i = 0, \ u_i \geq 0, \ i = 1, \dots, n, \\ \text{iii)} \ & -\frac{1}{\alpha_i + x_i} - u_i + v = 0, \ i = 1, \dots, n. \end{aligned}$$

We will proceed in several steps:

1. Since it holds
$$v = \frac{1}{\alpha_i + x_i} + u_i, \ \forall i,$$
and $\alpha_i > 0$ and $u_i \geq 0$, multiplier $v$ must be positive.

2. Now we can elaborate the complementarity conditions ii) for arbitrary $i \in \{1, \dots, n\}$, i.e. $u_i = 0$ or $x_i = 0$:

   2.a. Let $u_i = 0$, then using iii) and 1. we obtain
   $$x_i = \frac{1}{v} - \alpha_i,$$
   which is nonnegative if and only if $v \leq 1/\alpha_i$.

   2.b. Let $x_i = 0$, then using iii) and 1. we obtain
   $$u_i = -1/\alpha_i + v,$$
   which is nonnegative if and only if $v \geq 1/\alpha_i$. Now realize that if $v \geq 1/\alpha_i$, then corresponding $x_i$ cannot be positive because from iii) it would hold
   $$-\frac{1}{\alpha_i + x_i} + v = u_i > 0,$$
   which violates the complementarity condition ($x_i$ and $u_i$ cannot be both positive). In other words, $x_i$ is positive if and only if $v \in (0, 1/\alpha_i)$.

[6] See Boyd and Vandenberghe (2004).

We have obtained two cases which are distinguished by relation between $v$ and $1/\alpha_i$. Then we can write

$$x_i = \max\left\{\frac{1}{v} - \alpha_i, 0\right\}.$$

3. It remains to determine the value of Lagrange multiplier $v$ using the equality constraint

$$\sum_{i=1}^{n} \max\left\{\frac{1}{v} - \alpha_i, 0\right\} = 1,$$

which has a unique solution since the function of $\sum_{i=1}^{n} \max\{\cdot - \alpha_i, 0\}$ is piecewise-linear, continuous and increasing with breakpoints at points $\alpha_i$. Note that there is no closed-form formula for $v$, we are satisfied with its existence.

**Example 6.18.** *Let $n \geq 2$. Consider the problem*

$$\begin{aligned} &\min\ x_1 \\ &\text{s.t.}\ \sum_{i=1}^{n}\left(x_i - \frac{1}{n}\right)^2 \leq \frac{1}{n(n-1)}, \\ &\sum_{i=1}^{n} x_i = 1. \end{aligned}$$

*Show that*

$$\left(0, \frac{1}{n-1}, \ldots, \frac{1}{n-1}\right)$$

*is an optimal solution.*

**Solution:** First, realize that the considered point is feasible. Write the Lagrange function

$$L(x_1, \ldots, x_n, u, v) = x_1 + u\left(\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\right)^2 - \frac{1}{n(n-1)}\right) + v\left(\sum_{i=1}^{n} x_i - 1\right),$$

where $u \geq 0$ and $v \in \mathbb{R}$. The KKT conditions (feasibility, complementarity and optimality) are

$$\begin{aligned} &\text{i)}\ \sum_{i=1}^{n}\left(x_i - \frac{1}{n}\right)^2 \leq \frac{1}{n(n-1)},\ \sum_{i=1}^{n} x_i = 1, \\ &\text{ii)}\ u\left(\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\right)^2 - \frac{1}{n(n-1)}\right) = 0,\ u \geq 0, \\ &\text{iii)}\ \frac{\partial L}{\partial x_1} = 1 + 2u\left(x_1 - \frac{1}{n}\right) + v = 0, \\ &\frac{\partial L}{\partial x_i} = 2u\left(x_1 - \frac{1}{n}\right) + v = 0,\ i \neq 1. \end{aligned} \tag{45}$$

Realize that the inequality constraint is active at the considered point, i.e.

$$\left(0-\frac{1}{n}\right)^2+\sum_{i=2}^{n}\left(\frac{1}{n-1}-\frac{1}{n}\right)^2=\frac{1}{n(n-1)}.$$

To obtain the values of Lagrange multipliers, we solve the optimality conditions

$$\begin{aligned}
&1-\frac{2u}{n}+v=0,\\
&2u\left(\frac{1}{n-1}-\frac{1}{n}\right)+v=0,\ (\forall i\neq 1).
\end{aligned}\tag{46}$$

By solving this linear system for $u$ and $v$, we obtain the values

$$\begin{aligned}
u&=\frac{n-1}{2}\geq 0,\\
v&=\frac{-1}{n}\in\mathbb{R}.
\end{aligned}\tag{47}$$

Thus, we have obtained a KKT point

$$(x,u,v)=\left(0,\frac{1}{n-1},\ldots,\frac{1}{n-1},\frac{n-1}{2},\frac{-1}{n}\right),$$

Since the objective function is convex (linear), the inequality constraint is convex and the equality constraint is linear, the considered point is a global solution (minimum) of the problem.

**Example 6.19.** *Let $n\geq 2$. Consider the problem*

$$\begin{aligned}
\min\ &\sum_{j=1}^{n}\frac{c_j}{x_j}\\
\text{s.t.}\ &\sum_{j=1}^{n}x_j=1,\\
&x_j\geq\varepsilon,\ j=1,\ldots,n,
\end{aligned}$$

*where $c_j>0,\forall j$, $\varepsilon>0$ are parameters. Using the KKT conditions find an optimal solution.*

**Solution:** Write the Lagrange function

$$L(x,u,v)=\sum_{j=1}^{n}\frac{c_j}{x_j}+\sum_{j=1}^{n}u_j(\varepsilon-x_j)+v\left(\sum_{j=1}^{n}x_j-1\right),\ u_j\geq 0,v\in\mathbb{R}.$$

The KKT condition are then

i) feasibility,

ii) complementarity:

$$u_j(\varepsilon - x_j) = 0,\ u_j \geq 0\ j = 1, \ldots, n,$$

iii) optimality:

$$\frac{\partial L(x,u,v)}{\partial x_j} = \frac{-c_j}{(x_j)^2} - u_j + v = 0,\ j = 1, \ldots, n.$$

We can split the elaboration of the KKT conditions into three cases:

1. $x_j = \varepsilon$ for all $j$,
2. $u_j = 0$ for all $j$,
3. $x_j = \varepsilon$ for $j \in J_1$ and $u_j = 0$ for $j \in J_2$, where $J_1 \cup J_1 = \{1, \ldots, n\}$ and $J_1 \cap J_1 = \emptyset$. (This covers many cases which can be resolved at once.)

1. $x_j = \varepsilon$ for all $j$: the obtained point can be feasible only if $\varepsilon = 1/n$. Then we can calculate the Lagrange multipliers using the optimality conditions

$$\frac{-c_j}{\varepsilon^2} - u_j + v = 0,\ j = 1, \ldots, n.$$

If all $u_j$ are nonnegative, we have a KKT point.

2. $u_j = 0$ for all $j$ leads to

$$x_j = \sqrt{\frac{c_j}{v}},\ \forall j.,$$

which is well defined if $v > 0$. We can calculate $v$ using the constraint

$$\sum_{j=1}^{n} x_j = \sum_{j=1}^{n} \sqrt{\frac{c_j}{v}} = 1,$$

leading to

$$v = \left( \sum_{i=1}^{n} \sqrt{c_i} \right)^2 > 0,$$

and

$$x_j = \frac{\sqrt{c_j}}{\sum_{i=1}^{n} \sqrt{c_i}}.$$

What is not obvious, is the feasibility, i.e. $x_j \geq \varepsilon,\ \forall j$. If it holds, then

$$\left( \frac{\sqrt{c_1}}{\sum_{i=1}^{n} \sqrt{c_i}}, \ldots, \frac{\sqrt{c_n}}{\sum_{i=1}^{n} \sqrt{c_i}}, 0, \ldots, 0, \left( \sum_{i=1}^{n} \sqrt{c_i} \right)^2 \right)$$

is a KKT point.

3. $u_j = 0$ for $j \in J_2$ leads to

$$x_j = \sqrt{\frac{c_j}{v}},\ j \in J_2,$$

which is well defined if $v > 0$. Together with $x_j = \varepsilon, j \in J_1$, using the feasibility condition, we have

$$\sum_{j=1}^{n} x_j = \sum_{j \in J_1} \varepsilon + \sum_{j \in J_2} \sqrt{\frac{c_j}{v}} = 1.$$

Let $n_1 = |J_1|$ be the cardinality of set $J_1$. Then we obtain

$$v = \left( \frac{\sum_{j \in J_2}^{n} \sqrt{c_j}}{1 - n_1 \varepsilon} \right)^2,$$

which is well defined if $n_1 \varepsilon < 1$, and then also positive. The remaining Lagrange multipliers are equal to

$$u_j = \frac{-c_j}{\varepsilon^2} + v,\ j \in J_1.$$

If these multipliers are nonnegative and $x_j \geq \varepsilon,\ \forall j \in J_2$, we have obtained a KKT point.

We can easily realize that the problem is convex, because the objective function is convex for $x_j > 0$ (realize that $c_j > 0$) and the constraints are linear. Thus, any KKT point corresponds to a global solution.

**Example 6.20.** (*) *Consider the problem*

$$\begin{aligned} \min \ & \sum_{j=1}^{n} \frac{c_j}{x_j} \\ \text{s.t.} \ & \sum_{j=1}^{n} a_j x_j = b, \\ & x_j \geq \varepsilon, \end{aligned}$$

*where $a_j, b, c_j, \varepsilon > 0$ are parameters. Using the KKT conditions find an optimal solution.*

**Example 6.21.** (*) *Write the KKT conditions for a linear programming problem.*

**Example 6.22.** (*) *Derive the least square estimate for coefficients in the linear regression model under linear constraints, i.e. solve the problem*

$$\begin{aligned} \min_{\beta} \ & \|Y - X\beta\|^2, \\ \text{s.t.} \ & A\beta = b. \end{aligned}$$

### 6.2.3 Constraint qualification conditions

**Example 6.23.** *Consider the problem*

$$\begin{aligned} &\min\ x_1 \\ &s.t.\ (x_1-1)^2+(x_2-1)^2 \le 1 \\ &\qquad (x_1-1)^2+(x_2+1)^2 \le 1. \end{aligned}$$

*The optimal solution is obviously the only feasible point* $(1,0)$. *Why are not the KKT conditions fulfilled? Discuss the Constraint Qualification conditions.*

**Solution:** Write the Lagrange function

$$L(x_1,x_2u_1,u_2) = x_1+u_1\big((x_1-1)^2+(x_2-1)^2-1\big)+u_2\big((x_1-1)^2+(x_2+1)^2-1\big),\ u_1,u_2 \ge 0.$$

Derive the KKT conditions

$$\begin{aligned} \text{i) }& \text{feasibility,} \\ \text{ii) }& u_1\big((x_1-1)^2+(x_2-1)^2-1\big)=0,\ u_1\ge 0, \\ & u_2\big((x_1-1)^2+(x_2+1)^2-1\big)=0,\ u_2\ge 0, \\ \text{iii) }& \frac{\partial L}{\partial x_1} = 1+2u_1(x_1-1)+2u_2(x_1-1)=0, \\ & \frac{\partial L}{\partial x_2} = 2u_1(x_2-1)+2u_2(x_2+1)=0. \end{aligned} \tag{48}$$

Now, by substituting point (1,0) into the equation we obtain

$$\frac{\partial L}{\partial x_1}(1,0) = 1 \ne 0,$$

which is a contradiction with the optimality condition. In other words, there are no multipliers $u_{1,2} \ge 0$ such that $(1,0,u_1,u_2)$ is a KKT point. The explanation is easy: No Constraint Qualification condition is fulfilled. We can quickly check the basic (strongest) ones:
Slater CQ: There is no $x$ such that

$$(x_1-1)^2+(x_2-1)^2 < 1\ \&\ (x_1-1)^2+(x_2+1)^2 < 1.$$

LI CQ: We can compute the gradients

$$\nabla g_1(1,0) = \begin{pmatrix} 2(x_1-1) \\ 2(x_2-1) \end{pmatrix}\Big|_{(1,0)} = \begin{pmatrix} 0 \\ -2 \end{pmatrix},\ \nabla g_2(1,0) = \begin{pmatrix} 2(x_1-1) \\ 2(x_2+1) \end{pmatrix}\Big|_{(1,0)} = \begin{pmatrix} 0 \\ 2 \end{pmatrix},$$

which are not linearly independent.

To summarize, the basic (strongest) CQ conditions are not fulfilled, but since we were not able to find the Lagrange multipliers, even weaker CQ conditions cannot be fulfilled.

**Example 6.24.** *Consider the problem with real parameter $a$*

$$\begin{aligned} &\min\ x_1 \\ &s.t.\ (x_1-1)^2+(x_2-1)^2 \le 1 \\ &\qquad (x_1-1)^2+(x_2+a)^2 \le 1. \end{aligned}$$

*Discuss the Slater and LI Constraint Qualification conditions.*

**Solution:** First realize that the set of feasible solutions is nonempty only if $a \in [-3,1]$. We can verify the CQ conditions:
Slater CQ: the functions defining the set of feasible solutions are convex and

$$(x_1-1)^2+(x_2-1)^2 < 1 \ \&\ (x_1-1)^2+(x_2+a)^2 < 1$$

is nonempty if $a \in (-3,1)$.
LI CQ: Obviously, the optimal solution is the left intersection of the circles, which can be expressed as

$$(x_1,x_2) = \left(1-\sqrt{1-\left(\frac{1-a^2}{2(a+1)}-1\right)^2}, \frac{1-a^2}{2(a+1)}\right),\ a \in [-3,1].$$

We can compute the gradients in these points

$$\nabla g_1(1,0) = \begin{pmatrix} 2(x_1-1) \\ 2(x_2-1) \end{pmatrix},\ \nabla g_2(1,0) = \begin{pmatrix} 2(x_1-1) \\ 2(x_2+a) \end{pmatrix},$$

which are linearly independent if $a \in (-3,1)\backslash\{-1\}$. Note that $a=-1$ corresponds to the case when the constraints are identical. Therefore it can be resolved by forgetting of one of the constraints. However, for $a=-3$ we have $(x_1,x_2)=(1,2)$ and the gradients are linearly dependent. Case $a=1$ was discussed in the previous example.

**Example 6.25.** *Consider the problem*

$$\begin{aligned} &\min\ (x_1-2)^2+x_2^2 \\ &s.t.\ -(1-x_1)^3+x_2 \le 0, \\ &\qquad x_2 \ge 0. \end{aligned}$$

*Use the picture to show that $(1,0)$ is the global optimal solution. Why are not the KKT conditions fulfilled? Discuss the Constraint Qualification conditions.*

**Solution:** Write the Lagrange function

$$L(x_1,x_2u_1,u_2) = (x_1-2)^2+x_2^2+u_1\big(-(1-x_1)^3+x_2\big)-u_2\big(x_2\big),\ u_1,u_2 \ge 0.$$

Derive the KKT conditions

$$\begin{aligned} &\text{i) feasibility}, \\ &\text{ii) } u_1\big(-(1-x_1)^3+x_2\big)=0,\ u_1 \ge 0, \\ &\quad u_2\big(x_2\big)=0,\ u_2 \ge 0, \\ &\text{iii) } \frac{\partial L}{\partial x_1} = 2(x_1-2)+3u_1(x_1-1)^2=0, \\ &\quad \frac{\partial L}{\partial x_2} = 2x_2+u_1-u_2=0. \end{aligned} \tag{49}$$

Now, by substituting point (1,0) into the equation we obtain

$$\frac{\partial L}{\partial x_1}(1,0) = -2 \neq 0,$$

which is a contradiction with the optimality condition. In other words, there are no multipliers $u_{1,2} \geq 0$ such that $(1, 0, u_1, u_2)$ is a KKT point. The explanation is easy: No Constraint Qualification condition is fulfilled. Note that the interior of the set of feasible solution is nonempty, however the function $g_1$ which defines the set is not convex. Therefore the Slater CQ is not fulfilled.

### 6.2.4 Second Order Sufficient Condition (SOSC)

When the problem is not convex, then the solutions of the KKT conditions need not to correspond to global optima. The Second Order Sufficient Condition (SOSC) can be used to verify if the KKT point (its $x$ part) is at least a local minimum.

**Example 6.26.** *Consider the problem*

$$\begin{aligned} \min\ & -x \\ s.t.\ & x^2 + y^2 \leq 1 \\ & (x-1)^3 - y \leq 0. \end{aligned}$$

*Using the KKT optimality conditions find all stationary points. Using the SOSC verify if some of the points corresponds to a (strict) local minimum.*

**Solution:** Write the Lagrange function

$$L(x, y, u_1, u_2) = -x + u_1 \left(x^2 + y^2 - 1\right) + u_2 \left((x-1)^3 - y\right),\ u_1, u_2 \geq 0.$$

Derive the KKT conditions

$$\begin{aligned} \text{i)}\ & \text{feasibility}, \\ \text{ii)}\ & u_1 \left(x^2 + y^2 - 1\right) = 0,\ u_1 \geq 0, \\ & u_2 \left((x-1)^3 - y\right),\ u_2 \geq 0, \\ \text{iii)}\ & \frac{\partial L}{\partial x} = -1 + 2u_1 x + 3u_2 (x-1)^2 = 0, \\ & \frac{\partial L}{\partial y} = 2u_1 y - u_2 = 0. \end{aligned} \tag{50}$$

Now, we will try to find the KKT point by analyzing the optimality conditions, where we proceed according to the complementarity conditions:
**1.** Set $u_1 = 0$, $u_2 = 0$: We have from iii) that $-1 = 0$ which is a contradiction.

**2.** Set $u_1 = 0$, $(x-1)^3 - y = 0$: We have from second equality of iii), that $u_2 = 0$ which is again a contradiction with first equality of iii) $-1 \neq 0$.

**3.** Set $x^2 + y^2 = 1$, $u_2 = 0$: Second equality of iii) reduces to

$$2u_1 y = 0.$$

Case $u_1 = 0$ leads again to a contradiction, so the only possibility is $y = 0$. Using $x^2 + y^2 = 1$ we get $x \in \{-1, 1\}$. Using first equality of iii) we obtain the Lagrange multipliers, for $x = 1$ we get $u_1 = \frac{1}{2}$. However, for $x = -1$ we obtain $u_1 = -\frac{1}{2} < 0$, which is not feasible Lagrange multiplier. We have found one KKT point

$$(x, y, u_1, u_2) = \left(1, 0, \frac{1}{2}, 0\right).$$

**4.** Set $x^2 + y^2 = 1$, $(x-1)^3 - y = 0$: Using a picture, we can find two intersections of the curves: $(1, 0)$ and $(0, -1)$. The first point has been resolved by the previous point 3. Using iii) for $(0, -1)$ we can get $u_1 = -\frac{1}{6} < 0$ and $u_2 = \frac{1}{3}$, which are not feasible values of Lagrange multipliers.

Since the problem is non-convex, we can apply SOSC (36), (37). We have $I_g(1, 0) = \{1, 2\}$, $I_g^+(1, 0) = \{1\}$ and $I_g^0(1, 0) = \{2\}$. We can cumpute the gradients

$$\nabla g_1(1,0) = \begin{pmatrix} 2x \\ 2y \end{pmatrix} |_{(1,0)} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \ \nabla g_2(1,0) = \begin{pmatrix} 3(x-1)^2 \\ -1 \end{pmatrix} |_{(1,0)} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

so the conditions on $0 \neq z \in \mathbb{R}^2$ are:

$$\begin{aligned} 2z_1 &= 0, \\ -z_2 &\leq 0. \end{aligned}$$

So we have

$$Z(1,0) = \left\{z \in \mathbb{R}^2 : \ z_1 = 0, \ z_2 > 0\right\} \neq \emptyset.$$

We must compute the Heassian matrix of the Lagrange function with respect to the decision variables

$$\nabla^2_{xx} L\left(1, 0, \frac{1}{2}, 0\right) = \begin{pmatrix} 2u_1 + 6u_2(x-1) & 0 \\ 0 & 2u_1 \end{pmatrix} |_{(1,0,\frac{1}{2},0)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus we have that $z^T \nabla^2_{xx} L(1, 0, \frac{1}{2}, 0)\, z > 0$ for any $z \in Z(1, 0)$, which implies that $(1, 0)$ is a strict local minimum of the problem.

**Example 6.27.** *Consider the problem*

$$\begin{aligned} &\min \ x^2 - y^2 \\ &s.t. \ x - y = 1 \\ &\qquad x, y \geq 0. \end{aligned}$$

*Using the KKT optimality conditions find all stationary points. Using the SOSC verify if some of the points corresponds to a (strict) local minimum.*

**Solution:** Write the Lagrange function

$$L(x, y, u_1, u_2, v) = x^2 - y^2 - u_1 x - u_2 y + v(x - y - 1), \ u_1, u_2 \geq 0.$$

Derive the KKT conditions

$$
\begin{aligned}
\text{i)}\ & \text{feasibility,} \\
\text{ii)}\ & -u_1 x = 0,\ \ u_1 \geq 0, \\
& -u_2 y = 0,\ \ u_2 \geq 0, \\
\text{iii)}\ & \frac{\partial L}{\partial x} = 2x - u_1 + v = 0, \\
& \frac{\partial L}{\partial y} = -2y - u_2 - v = 0.
\end{aligned}
\tag{51}
$$

Solving this conditions together with feasibility leads to one feasible KKT point

$$(x, y, u_1, u_2, v) = (1, 0, 0, 2, -2).$$

Since the problem is non-convex, we can apply SOSC (36), (37). We have $I_g(1,0) = I_g^+(1,0) = \{2\}$ and $I_g^0(1,0) = \emptyset$, so the conditions on $0 \neq z \in \mathbb{R}^2$ are:

$$
\begin{aligned}
z_1 - z_2 &= 0, \\
-z_2 &= 0.
\end{aligned}
$$

Since no $z \neq 0$ exists, the SOSC is fulfilled. (It is not necessary to compute $\nabla^2_{xx} L$.)

**Example 6.28.** *Consider the problem*

$$
\begin{aligned}
&\min\ -x^2 - 4xy - y^2 \\
&s.t.\ x^2 + y^2 = 1.
\end{aligned}
$$

*Using the SOSC verify that point $(\sqrt{2}/2, \sqrt{2}/2)$ corresponds to a (strict) local minimum.*

**Solution:** Write the Lagrange function

$$L(x, y, v) = -x^2 - 4xy - y^2 + v\left(x^2 + y^2 - 1\right).$$

Derive the KKT conditions

$$
\begin{aligned}
\text{i)}\ & \text{feasibility,} \\
\text{ii)}\ & - \\
\text{iii)}\ & \frac{\partial L}{\partial x} = -2x - 4y + 2vx = 0, \\
& \frac{\partial L}{\partial y} = -2y - 4x + 2vy = 0.
\end{aligned}
\tag{52}
$$

We can compute the Lagrange multiplier and obtain the KKT point

$$(x, y, v) = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 3\right).$$

Since the problem is non-convex, we can apply SOSC (36), (37). We have

$$\nabla h(\sqrt{2}/2, \sqrt{2}/2) = \begin{pmatrix} 2x \\ 2y \end{pmatrix} |_{(\sqrt{2}/2, \sqrt{2}/2)} = \begin{pmatrix} \sqrt{2} \\ \sqrt{2} \end{pmatrix},$$

so we have

$$Z(\sqrt{2}/2, \sqrt{2}/2) = \left\{ z \in \mathbb{R}^2 : \ z_1 + z_2 = 0, \ z \neq 0 \right\} = \{(z_1, -z_1) : \ z_1 \in \mathbb{R} \setminus \{0\}\}.$$

We must compute the Hessian matrix

$$\nabla^2_{xx} L\left(\sqrt{2}/2, \sqrt{2}/2, 3\right) = \begin{pmatrix} -2+2v & -4 \\ -4 & -2+2v \end{pmatrix} |_{(\sqrt{2}/2, \sqrt{2}/2, 3)} = \begin{pmatrix} 4 & -4 \\ -4 & 4 \end{pmatrix}.$$

Thus we have that $z^T \nabla^2_{xx} L(\sqrt{2}/2, \sqrt{2}/2, 3)\, z = 16 z_1^2 > 0$ for any $z_1 \in \mathbb{R} \setminus \{0\}$, which implies that $(\sqrt{2}/2, \sqrt{2}/2)$ is a strict local minimum of the problem.

**Example 6.29.** (*) *Consider the problem*

$$\begin{aligned} \min \ & -(x-2)^2 - (y-3)^2 \\ s.t. \ & 3x + 2y \geq 6, \\ & -x + y \leq 3, \\ & x \leq 2. \end{aligned}$$

*Using the KKT optimality conditions find all stationary points. Using the SOSC verify if some of the points corresponds to a (strict) local minimum.*

# 7 Motivation and real-life applications

## 7.1 Operations Research/Management Science and Mathematical Programming

**Goal**: improve/stabilize/set of a system. You can reach the goal in the following steps:

- Problem understanding
- Problem description – probabilistic, statistical and econometric models
- Optimization – mathematical programming (formulation and solution)
- Verification – backtesting, stresstesting
- Implementation (Decision Support System)
- Decisions

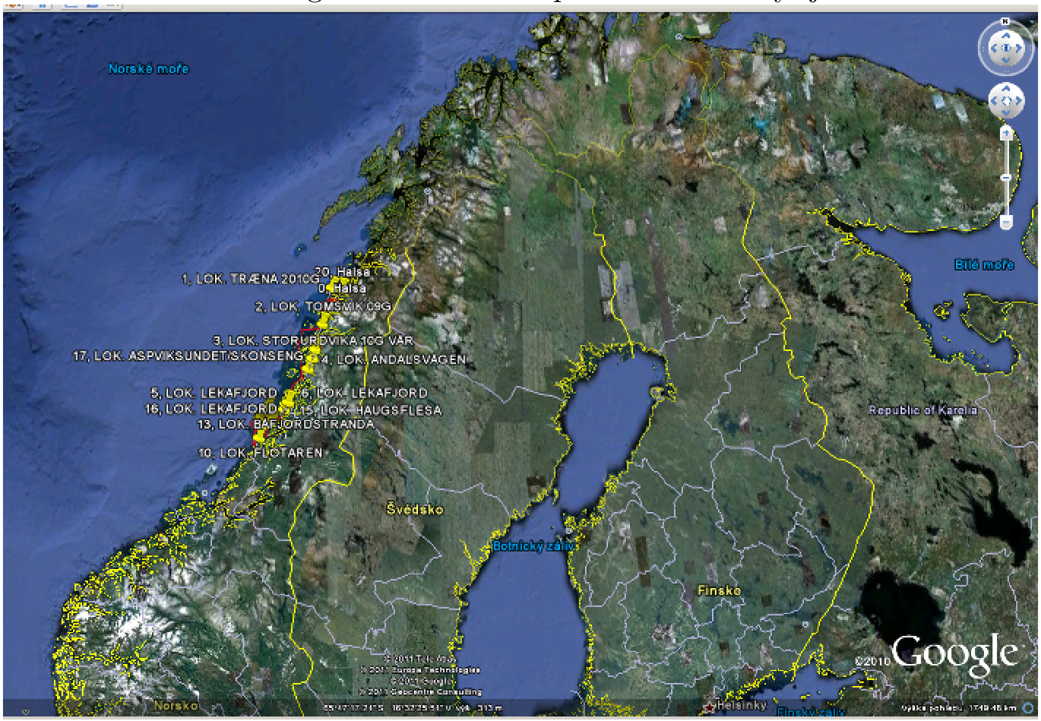
## 7.2 Marketing – Optimization of advertising campaigns

- **Goal** – maximization of the effectiveness of a advertising campaign given its costs or vice versa
- **Data** – "peoplemeters", public opinion poll, historical advertising campaigns
- **Target group** – (potential) customers (age, region, education level ...)
- **Effectiveness criteria**
  - GRP (TRP) – rating(s)
  - Effective frequency – relative number of persons in the target group hit k-times by the campaign
- Nonlinear (nonconvex) or integer programming

## 7.3 Logistic – Vehicle routing problems

- **Goal** – maximize *filling rate* of the ships (operation planning), fleet composition, i.e. capacity and number of ships (strategic planning)
- **Rich Vehicle Routing Problem**
  - time windows
  - heterogeneous fleet
  - several depots and inter-depot trips
  - several trips during the planning horizon
  - *non-Euclidean distances* (fjords)
- Integer programming :-(, constructive heuristics and tabu search

Figure 1: A boat trip around Norway fjords




Literature: M.B., K. Haugen, J. Novotný, A. Olstad (2017).

Related problems with increasing complexity:

- Traveling Salesman Problem
- Uncapacitated Vehicle Routing Problem (VRP)
- Capacitated VRP
- VRP with Time Windows
- ...

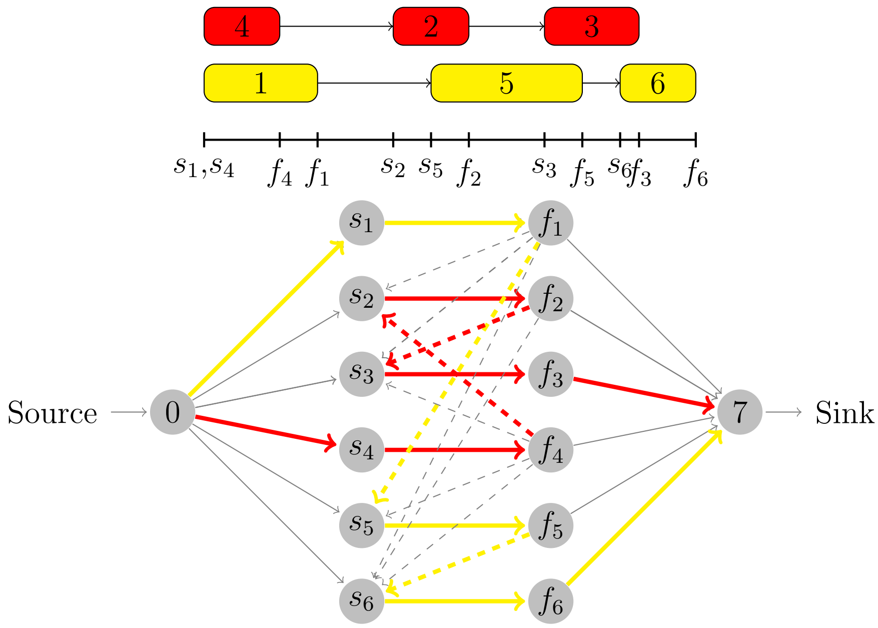
OR approach to real-life problem solution:

1. Mathematical formulation
2. Solving using GAMS based on historical data
3. Heuristic(s) implementation
4. Implementation to a Decision Support System

## 7.4 Scheduling – Reparations of oil platforms

- **Goal** – send the right workers to the oil platforms taking into account uncertainty (bad weather – helicopter(s) cannot fly – jobs are delayed)
- **Scheduling** – jobs = reparations, machines = workers (highly educated, skilled and costly)

Figure 2: Fixed interval schedule (assignment of 6 jobs to 2 machines) and corresponding network flow




- Integer and stochastic programming

Literature: M. Branda, J. Novotný, A. Olstad (2016), M. Branda, S. Hájek (2017)

## 7.5 Insurance – Pricing in nonlife insurance

- **Goal** – optimization of prices in MTPL/CASCO insurance taking into account riskiness of contracts and competitiveness of the prices on the market
- **Risk** – compound distribution of random losses over 1y (Data-mining & GLM)
- Nonlinear stochastic optimization (probabilistic or expectation constraints)
- See Table 1

Literature and detailed information: M.B. (2012, 2014)

## 7.6 Power industry – Bidding, power plant operations

**Energy markets**

- **Goal** – profit maximization and risk minimization
- **Day-ahead bidding** from wind (power) farm
- Nonlinear stochastic programming

**Power plant operations**

Table 1: Multiplicative tariff of rates: final price can be obtained by multiplying the coefficients

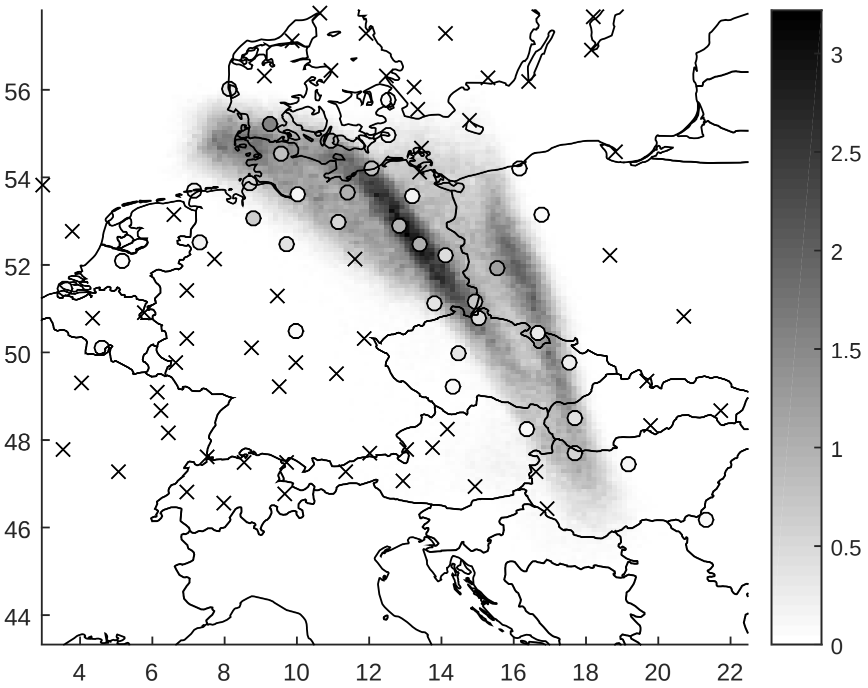
| | | GLM | SP model (ind.) | SP model (col.) |
|---|---|---|---|---|
| TG | up to 1000 ccm | 3 805 | 9 318 | 5 305 |
| TG | 1000–1349 ccm | 4 104 | 9 979 | 5 563 |
| TG | 1350–1849 ccm | 4 918 | 11 704 | 6 296 |
| TG | 1850–2499 ccm | 5 748 | 13 380 | 7 125 |
| TG | over 2500 ccm | 7 792 | 17 453 | 9 169 |
| Region | Capital city | 1.61 | 1.41 | 1.41 |
| Region | Large towns | 1.16 | 1.18 | 1.19 |
| Region | Small towns | 1.00 | 1.00 | 1.00 |
| Region | Others | 1.00 | 1.00 | 1.00 |
| Age | 18–30y | 1.28 | 1.26 | 1.27 |
| Age | 31–65y | 1.06 | 1.11 | 1.11 |
| Age | over 66y | 1.00 | 1.00 | 1.00 |
| DL less that 5y | NO | 1.00 | 1.00 | 1.00 |
| DL more that 5y | YES | 1.19 | 1.13 | 1.12 |

- **Goal** – profit maximization and risk minimization
- **Coal power plants** – demand seasonality, ...
- Stochastic linear programming (multistage/multiperiod)

## 7.7 Environment – Inverse modelling in atmosphere

- **Goal** – identification of the source and the amount released into the atmosphere
- **Standard approach** – dynamic Bayesian models
- **New approach** – Sparse optimization – Nonlinear/quadratic integer programming (weighted least squares with nonnegativity and sparsity constraints)
- **Applications**: nuclear power plants accidents, volcano accidents, nuclear tests, emission of pollutants ...
- **Project homepage:** http://stradi.utia.cas.cz/

Literature and detailed information: L. Adam, M.B. (2016).



56
54
52
50
48
46
44
4
6
8
10
12
14
16
18
20
22
3
2.5
2
1.5
1
0.5
0

# Literature


- L. Adam, M. Branda (2016). **Sparse optimization for inverse problems in atmospheric modelling.** Environmental Modelling & Software 79, 256–266. (free Matlab codes available)

- Bazaraa, M.S., Sherali, H.D., and Shetty, C.M. (2006). **Nonlinear programming: theory and algorithms**, Wiley, Singapore, 3rd edition.

- Boyd, S., Vandenberghe, L. (2004). **Convex optimization**, Cambridge University Press, Cambridge.

- M. Branda (2012). **Underwriting risk control in non-life insurance via generalized linear models and stochastic programming**. Proceedings of the 30th International Conference on MME 2012, 61–66.

- M. Branda (2014). **Optimization approaches to multiplicative tariff of rates estimation in non-life insurance**. Asia-Pacific Journal of Operational Research 31(5), 1450032, 17 pages, 2014.

- M. Branda, S. Hájek (2017). **Flow-based formulations for operational fixed interval scheduling problems with random delays**. Computational Management Science 14 (1), 161–177.

- M. Branda, J. Novotný, A. Olstad (2016). **Fixed interval scheduling under uncertainty - a tabu search algorithm for an extended robust coloring formulation**. Computers & Industrial Engineering 93, 45–54.

- M. Branda, K. Haugen, J. Novotný, A. Olstad (2017). **Downstream logistics optimization at EWOS Norway**. Mathematics for Applications 6 (2), 127-141.

- P. Lachout (2020). **Optimization Theory.** Lecture notes.

- R.T. Rockafellar (1972). **Convex analysis**, Princeton University Press, New Jersey.