EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

NMSA403 Optimization Theory – Exercises

Collection of Examples[1]

Martin Branda

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

Version: 31/12/2022

[1]Materials have been produced within and supported by the project "*Increasing the quality of education at Charles University and its relevance to the needs of the labor market*" kept under number CZ.02.2.69/0.0/0.0/16$_0$15/0002362.

Contents

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

1 Formulations of optimization problems

Consider real functions $f : \mathbb{R}^n \to \mathbb{R}$, $g_k : \mathbb{R}^n \to \mathbb{R}$, $k = 1, \ldots, m$, $h_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, p$. Then general mathematical programming/optimization problem can be formulated as

$$\begin{aligned} \min\ & f(x) \\ \text{s.t. } & g_k(x) \le 0,\ k = 1, \ldots, m, \\ & h_j(x) = 0,\ j = 1, \ldots, p, \\ & x \in \mathbb{R}^n, \end{aligned}$$

where we consider

- $f(x)$ – objective function,
- $g_k(x) \le 0$ – inequality constraints,
- $h_j(x) = 0$ – equality constraints,
- $x \in \mathbb{R}^n$ – continuous/real decision variables,
- set of feasible solutions $M \subset \mathbb{R}^n$

$$M = \{x \in \mathbb{R}^n :\ g_k(x) \le 0,\ k = 1, \ldots, m,\ h_j(x) = 0,\ j = 1, \ldots, p\}.$$

- We say that $\hat{x} \in M$ is optimal solution of the above problem if $f(\hat{x}) \le f(x)$ for all $x \in M$. Then $f(\hat{x})$ is optimal value.

If for some/all decision variables it holds $x_i \in \mathbb{N}$, then we speak about integer variables and (mixed-)integer programming problems. If a variable is restricted to values 0 or 1, i.e. $x_i \in \{0, 1\}$, we call it binary variable. We can also classify the problems according to the functions. If all functions f, g_k, h_j are linear, then we speak about linear programming problem (LP). If at least one function is nonlinear, then we face a nonlinear programming problem (NLP). If all functions are polynomial, but at most quadratic, then the problem is called quadratic (QP).

Example 1.1. Consider the linear regression model

$$y_i = \beta^T x_i + \varepsilon_i,\ i = 1, \ldots, n,$$

where $\beta \in \mathbb{R}^m$ are unknown coefficients, $x_i \in \mathbb{R}^m$ is vector of regressors (explanatory variables), y_i is dependent variable and $\varepsilon_i \sim (0, \sigma^2)$ are i.i.d. error terms. Derive the l_2 and l_1 estimates of its parameters for the following cases:

1. without additional constraints,
2. with linear constraints on coefficients, e.g., the sum of all coefficients is at most 1,
3. the number of nonzero coefficients is restricted by $\kappa < m$, i.e. so called cardinality constraint or sparsity is applied.

Discuss the classification of the resulting optimization problems.

Solution. 1. l_2 estimate called also *least square estimate* can be obtained as

$$\min_\beta \sum_{i=1}^n (y_i - \beta^T x_i)^2.$$

If we set

$$X = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix}$$

and

$$Y^T = (y_1, \ldots, y_n),$$

we can rewrite the problem in the matrix form

$$\min_\beta (Y - X\beta)^T (Y - X\beta) = \min_\beta Y^T Y - 2Y^T X\beta + \beta^T X^T X\beta.$$

Obviously the problem is continuous quadratic (without constraints).

Getting l_1 estimate leads to problem

$$\min_\beta \sum_{i=1}^n |y_i - \beta^T x_i|.$$

Before solving the problem, we would like to avoid the absolute value since it is nondifferentiable and can cause problem even to professional software tools. We can use two new nonnegative variables and set

$$z_i^+ - z_i^- = y_i - \beta^T x_i,\ z_i^+, z_i^- \geq 0.$$

Then, when we minimize the absolute value, we get the equality

$$\min_\beta \sum_{i=1}^n |y_i - \beta^T x_i| = \min_{\beta, z^+, z^-} \sum_{i=1}^n z_i^+ + z_i^-$$

subject to (s.t.) the constraints

$$z_i^+ - z_i^- = y_i - \beta^T x_i,\ z_i^+, z_i^- \geq 0,\ i = 1 \ldots, n.$$

Obviously the resulting problem is now continuous linear (with constraints).

2. Now, we add linear constraints on the coefficient which can be motivated by econometric applications. The l_2 estimate leads to problem

$$\min_\beta \sum_{i=1}^n (y_i - \beta^T x_i)^2$$
$$\text{s.t. } \sum_{j=1}^m \beta_j \leq 1,$$

which a continuous quadratic program with a constraint. More linear constraints can be added, e.g., in the matrix form

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

where $A \in \mathbb{R}^{k\times m}$ and $b \in \mathbb{R}^k$, i.e. we have added k linear constraints. The l_1 case is analogous.

3. The cardinality constraint, i.e. a restriction on the number of nonzero coefficients or in general elements of the decision vector, can be modelled using binary variables. We consider additional binary variables $z_j \in \{0,1\}$, which must be equal to one if the corresponding coefficient β_j is nonzero. This can be ensured by the following constraints

$$
-M\, z_j \le \beta_j \le M\, z_j,\ \ j = 1,\dots,m,
$$

where M is a sufficiently large constant. Realize that we assume that $|\beta_j| \le M$. The *sparse* l_2 estimate is then obtained by solving

$$
\begin{aligned}
\min_{\beta,z}\ & \sum_{i=1}^{n} (y_i - \beta^T x_i)^2 \\
\text{s.t. } & -M\, z_j \le \beta_j \le M\, z_j,\ \ j = 1,\dots,m, \\
& \sum_{j=1}^{m} z_j \le \kappa, \\
& z_j \in \{0,1\},\ \ j = 1,\dots,m,
\end{aligned}
$$

which is a mixed-integer (mixed-binary) quadratic programming problem. In the case of *sparse* l_1 estimate we obtain problem

$$
\begin{aligned}
\min_{\beta,z,z^+,z^-}\ & \sum_{i=1}^{n} z_i^+ + z_i^- \\
\text{s.t. } & z_i^+ - z_i^- = y_i - \beta^T x_i,\ \ i = 1\dots,n, \\
& -M\, z_j \le \beta_j \le M\, z_j,\ \ j = 1,\dots,m, \\
& \sum_{j=1}^{m} z_j \le \kappa, \\
& z_i^+, z_i^- \ge 0,\ \ i = 1\dots,n, \\
& z_j \in \{0,1\},\ \ j = 1,\dots,m,
\end{aligned}
$$

which is a mixed-integer linear programming problem.

Example 1.2. The company produces 3 types of products V_1, V_2, V_3. In production, raw materials S_1 and S_2 and machine time Z_1 are consumed. To produce 1 kg of product V_1, 2 kg of raw material S_1 and 6 kg of raw material S_2 are consumed, the production time is 13 hours. To produce 1 kg of product V_2, 3 kg of raw material S_1 and 8 kg of raw material S_2 are needed for 17 hours. The production time of V_3 is

15 hours and 5 kg of raw material S_1 and 1 kg of raw material S_2 are consumed. For 1 year, we have 200 kg of S_1, 150 kg of S_2 and 1800 hours on the Z_1 device. When selling, the company will receive 2 CZK for 1 kg V_1, 4 CZK for 1 kg V_2 and 3 CZK for 1 kg V_3. Determine the optimal production schedule, i.e. determine how many kg of which product should be produced per year to maximize profit.

Solution. The solution includes the following steps:

- introducing the variables used in the problem,
- formulation of the objective function,
- introducing restrictions on variables,
- variable value range definition.

Decision variables x_1, x_2, x_3 express the number of kg of products V_1, V_2, V_3 that will be produced during a year. The objective is then

$$\max \quad 2x_1 + 4x_2 + 3x_3$$

under the constraints (s.t.)

$$\begin{aligned} 2x_1 + 3x_2 + 5x_3 &\leq 200, \\ 6x_1 + 8x_2 + x_3 &\leq 150, \\ 13x_1 + 17x_2 + 15x_3 &\leq 1800, \end{aligned}$$

together with nonnegativity constraints

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0.$$

2 Extremes of functions

Example 2.1. Verify the inequality between the arithmetic mean and the geometric mean

$$\frac{1}{n}(x_1 + \cdots + x_n) \geq \sqrt[n]{x_1 \cdots x_n}$$

for positive $x_i > 0$.

Solution. We rewrite the task of finding extrema of the function $\sqrt[n]{x_1 \cdots x_n}$ on the set $\frac{1}{n}(x_1 + \cdots + x_n) = c$ i.e.

$$\begin{aligned} \max \quad & (x_1 \cdots x_n)^{1/n} \\ \text{s.t.} \quad & \frac{1}{n}\sum_i x_i = c, \end{aligned}$$

we express the Lagrange function

$$\mathcal{L}(\boldsymbol{x}, \lambda) = (x_1 \cdots x_n)^{1/n} + \lambda \left(c - \frac{1}{n}\sum_i x_i \right),$$

and derive the optimality/stationary conditions:

$$\begin{aligned} \frac{\partial \mathcal{L}(\boldsymbol{x}, \lambda)}{\partial x_i} : \quad & \frac{1}{n}(x_1 \cdots x_n)^{1/n-1} \cdot \frac{x_1 \cdots x_n}{x_i} - \frac{\lambda}{n} = 0, \quad i = 1, \ldots, n, \\ \frac{\partial \mathcal{L}(\boldsymbol{x}, \lambda)}{\partial \lambda} : \quad & \frac{1}{n}\sum_i x_i = c. \end{aligned}$$

modifying the first equations, we get

$$\frac{1}{n}(x_1 \cdots x_n)^{1/n} = \frac{\lambda}{n} x_i,$$

and if we sum for all i, we get

$$(x_1 \cdots x_n)^{1/n} = \frac{\lambda}{n}\sum_i x_i.$$

We can substitute this back into the equation

$$\lambda x_i = (x_1 \cdots x_n)^{1/n},$$

and get the necessary condition

$$x_i = \frac{1}{n}\sum_i x_i = c, \quad \forall i.$$

In other words, the maximum of this problem is c. The Lagrange multiplier can be obtained from

$$\lambda c = (c \cdots c)^{1/n} = c,$$

i.e. $\lambda = 1$.

Example 2.2. Find the extrema of the following function

$$f(x, y) = \sqrt{3}x - y + 2$$

on the set $x^2 + 2x + y^2 = 0$.

Example 2.3. Find the global extrema of the following function

$$f(x, y) = \frac{1}{2}x^2y - x^2 + xy - 2x + y^2 - \frac{7}{2}y + 1$$

on the set $[-3, 1] \times [0, 4]$.

Solution. Denote by $M = [-3, 1] \times [0, 4]$ the set of feasible solutions. First we look at the free extreme:

$$\frac{\partial f}{\partial x}(x, y) = (x + 1)(y - 2)$$
$$\frac{\partial f}{\partial y}(x, y) = \frac{1}{2}x^2 + x + 2y - \frac{7}{2}.$$

The necessary condition for a local extremum is met at the point $[-1, 2] \in M$, which is a stationary point. We must compute Hessian matrix

$$H(-1, 2) = \begin{pmatrix} y-2 & x+1 \\ x+1 & 2 \end{pmatrix}|_{(x,y)=[-1,2]} = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$$

According to the rule about major minors, this is a positive semidefinite quadratic matrix, so we cannot draw any conclusions from it.

Alternatively, we can modify the function to

$$f(x, y) = \frac{1}{2}(x + 1)^2(y - 2) + (y - 2)^2 - 2$$

and thus verifying whether in at the point $[-1, 2]$ the local minimum is equivalent to the fact that the function $g(x, y) = \frac{1}{2}x^2y + y^2$ has at the point $[0, 0]$ local minimum. But let's consider the parameterization $x = 2t, y = -t^2$, we get $g(2t, -t^2) = -2t^4 + t^4 = -t^4$. But this function has in $t = 0$ a local maximum and therefore the functions $g(x, y)$ and $f(x, y)$ do not have a local minimum at the suspicious points.

Global extremes can be found by examining the stationary points and extreme points identified on the boundary of the set M. Let's examine the line segments first. We substitute for $x = -3$ and get the function

$$f(-3, y) = y^2 - 2y - 2.$$

This quadratic function takes an extremum at $y = 1$. This leads to the suspicious point $[-3, 1]$. Similarly, successively substituting after $x = 1, y = 0$ and $y = 4$ we get other suspicious points $[1, 1], [-1, 0], [-1, 4]$). To these we still need to add the corners: $[-3, 0], [-3, 4], [1, 0], [1, 4]$. Examining the function value, we find that the global maximum of the function takes place in [-3, 4] and [1,4] and that is 6, while the global minimum then in [-3,1] and [1,1] and that is -3.

Note that later we will be able to solve the problem as a nonlinear programming one in the form

$$\begin{aligned} \min\ & \frac{1}{2}x^2y - x^2 + xy - 2x + y^2 - \frac{7}{2}y + 1 \\ \text{s.t.}\ & -3 \le x \le 1, \\ & 0 \le y \le 4, \end{aligned}$$

c.f. Section 7.

3 Convex sets

Definition 3.1. We say that a set A is convex if, with every two points, it also contains their convex linear combination, i.e. $\forall x, y \in A$ and $\lambda \in (0,1)$ is also $\lambda x + (1-\lambda)y \in A$.

Example 3.2. $A, B \in \mathbb{R}^n$ convex. Decide whether or not the following sets must also be convex.

1. $A \cup B$,
2. $A \cap B$,
3. $A \setminus B$,
4. A^C,
5. $\mathrm{cl}(A) = \bar{A}$,
6. $\mathrm{int}(A)$,
7. αA,
8. Minkowski sum $A + B = \{x \in \mathbb{R}^n, x = a + b, a \in A, b \in B\} = \bigcup_{b \in B}(A + b)$,
9. Minkowski difference $A \ominus B = \{x \in \mathbb{R}^n, x + (-1)B \subset A\} = \bigcap_{b \in B}(A - b)$,
10. Cartesian product $A \times B$.

Solution. In turn, either by proof or counterexample:

1. no, e.g.: $A = \{0\}, B = \{1\}, x = 0, y = 1, \lambda = 1/2$.
2. yes, let $x, y \in A \cap B, \lambda \in [0,1]$, then $x, y \in A, x, y \in B$, these are convex, so it holds and that $\lambda x + (1-\lambda)y \in A, \lambda x + (1-\lambda)y \in B$ and therefore $\lambda x + (1-\lambda)y \in A \cap B$
3. no, e.g.: $A = [0,1], B = \{1/2\}, x = 0, y = 1, \lambda = 1/2$.
4. no, e.g.: $A = \{0\}, x = -1, y = 1, \lambda = 1/2$.
5. yes, $x, y \in \bar{A}$, $\exists x_n, y_n$ such that $x_n \to x$ and $y_n \to y$ and at the same time $\lambda x_n + (1-\lambda)y_n \in A$. From the assumptions $\lambda x_n + (1-\lambda)y_n \to \lambda x + (1-\lambda)y \in \bar{A}$.
6. yes, $x, y \in A \Rightarrow \exists \varepsilon > 0 : U_\varepsilon(x) \subset A, U_\varepsilon(y) \subset A$. Choose $\lambda \in (0,1)$ and $z \in U_\varepsilon(\lambda x + (1-\lambda)y)$. $x + z - (\lambda x + (1-\lambda)y) \in U_\varepsilon(x)$ and $y + z - (\lambda x + (1-\lambda)y) \in U_\varepsilon(y)$. The convex combination of these points is z, and $z \in A$ follows from the convexity of A, which gives us that $\lambda x + (1-\lambda)y \in \mathrm{int}(A)$.
7. yes, for $\alpha = 0$ trivial, let $\alpha \neq 0$ and $x, y \in \alpha A, \lambda \in [0,1]$, then $x/\alpha, y/\alpha \in A$ and it is convex, so it is also true that $(\lambda x/\alpha + (1-\lambda)y/\alpha) \in A$ and therefore $\lambda x/ + (1-\lambda)y \in \alpha A$.
8. left to the readers,

9. left to the readers,

10. directly from the definition.

Example 3.3. Let $M \subset \mathbb{R}^n$ be a closed convex that contains the line p. Show that for $b \in M$ and $b \notin p$, the set M contains a line parallel to p passing through the point b.

Solution. We know that there exist $a \in M$ and $u \in \mathbb{R}^n$, that $a + tu \in M$, $t \in \mathbb{R}$, where the vector u determines the direction of the line p. From the convexity of M it follows, $\lambda(a + tu) + (1 - \lambda)b = b - \lambda(b - a) + \lambda tu$, for each $\lambda \in [0, 1], t \in \mathbb{R}$. This means that $b - \lambda(b - a) + su \in M$ for every $s \in \mathbb{R}$ and $\lambda > 0$. The limit transition for $\lambda \to 0+$ is $b + su \in \text{clo}(M) = M$.

Example 3.4. Let A have the property that $\forall x, y \in A$ also $x/2 + y/2 \in A$.

1. is A convex?

2. is A convex if we know that A is closed?

3. is A convex if we know that A is open?

Solution. We realize that it is possible to assemble combinations of the type

$$\frac{k_n}{2^n}x + (1 - \frac{k_n}{2^n})y,$$

with $k_n \in \mathbb{N}, n \in \mathbb{N}$ and $k_n \leq 2^n$, e.g.

$$\frac{1}{2}\left(\frac{x}{2} + \frac{y}{2}\right) + \frac{y}{2} = \frac{1}{4}x + \frac{3}{4}y \in A.$$

1. no, e.g. $\mathbb{Q}$ satisfies the given requirement but is not convex.
2. yes, since $\forall \lambda \in [0, 1]$ there exist $\lambda_n = \frac{k_n}{2^n}$ such that $\lambda_n \to \lambda$. We know that $\lambda_n x + (1 - \lambda_n)y \in A$ and since A is also closed $\lambda_n x + (1 - \lambda_n)y \to \lambda x + (1 - \lambda)y \in \text{cl}(A) = A$.
3. yes, choose $x, y \in A, \lambda \in [0, 1]$ and $\varepsilon > 0 : U_\varepsilon(x) \subset A, U_\varepsilon(y) \subset A$. Then there exists $\lambda_n = \frac{k_n}{2^n} : |1 - \lambda/\lambda_n| < \varepsilon/(|x| + |y|)$. Let us define $x_n = x \cdot \lambda/\lambda_n + y - y\lambda/\lambda_n$. Applies to:

$$|x - x_n| \leq |x(1 - \lambda/\lambda_n) - y(1 - \lambda/\lambda_n)| \leq |1 - \lambda/\lambda_n| \cdot (|x| + |y|) \leq \varepsilon.$$

In other words, $x_n \in A$, then $\lambda_n x_n + (1 - \lambda_n)y = \lambda x + (1 - \lambda)y \in A$.

3.1 Geometry for linear programming problems

Remind terms from the lectures: *convex hull, non-negative (positive) hull, extreme points, direction of a set, extreme directions, $P + K$ decomposition for convex polyhedral set (not containing a straight line):*

$$A = \text{conv}(\text{ext}(A)) + \text{pos}(\text{extd}(A)),$$

where $P = \text{conv}(\text{ext}(A))$ is a convex polyhedron generated by a finite number of extreme points $\text{ext}(A)$, and $K = \text{pos}(\text{extd}(A))$ is a convex polyhedral cone generated by a finite number of extreme directions $\text{extd}(A)$. The theory and algorithms for linear programming problems are based on these terms and properties.

Definition 3.5. We say that $A \subseteq \mathbb{R}^n$ is a convex polyhedral set if it is the intersection of finitely many closed half-spaces, i.e. sets of the form $\{x \in \mathbb{R}^n, a^T x \leq c\}$. A convex polyhedron (= polytope) is a bounded convex polyhedral set.

Definition 3.6. A cone is a set $K \subseteq \mathbb{R}^n$ for which $0 \in K$ and $\forall x \in K, \alpha \geq 0$ also $\alpha x \in K$. A convex cone is a convex set that satisfies the definition of a cone.

Example 3.7. Decide what type the following sets are

1.
$$\begin{aligned} x + 2y &\geq 2 \\ -2x + 3y &\leq 6 \\ x, y &\geq 0 \end{aligned}$$

2.
$$\begin{aligned} x + 2y &\geq 2 \\ -2x + 3y &\leq 6 \\ x + y &\leq 4 \end{aligned}$$

Solution. With the help of the graph it can be seen that

1. convex polyhedral set:
$$A_1 = \text{conv}\big(\{(2,0), (0,1), (0,2)\}\big) + \text{pos}\left(\{(1,0), \frac{1}{\sqrt{13}}(3,2)\}\right),$$
2. convex polyhedron:
$$A_2 = \text{conv}\left(\left\{\left(\frac{-6}{7}, \frac{10}{7}\right), (6,-2), \left(\frac{6}{5}, \frac{14}{5}\right)\right\}\right) + \text{pos}\,(\emptyset)\,.$$

Example 3.8. Decide which set is a convex polyhedron and which one is not:

1. $A_1 = \{(x,y) \in \mathbb{R}^2 : |x| + |y| \leq 1\}$,

2. $A_2 = \{(x, y) \in \mathbb{R}^2 : -1 \le x \le y \le 1\}$,
3. $A_3 = \{(x, y) \in \mathbb{R}^2 : |x + y| \le 1\}$.

Solution.

1. convex polyhedron (L_1 circle):

$$A_1 = \operatorname{conv}\{(1, 0), (-1, 0), (0, 1), (0, -1)\},$$

2. convex polyhedron:

$$A_2 = \operatorname{conv}\{(-1, 1), (-1, -1), (1, 1)\},$$

3. convex polyhedral set, not polyhedron:

$$A_3 = \left\{(x, y) \in \mathbb{R}^2 : x + y \le 1,\ x + y \ge -1\right\}.$$

Since A_3 contains a straight line, it has no $P + K$ decomposition.

Example 3.9. Verify that the set $K = \{(x, y) \in \mathbb{R}^2 : 0 \le x \le y\}$ is a convex cone, which does not contain a straight line.

Solution. Choose $(x_1, y_1), (x_2, y_2) \in K, \lambda \in [0, 1]$. Apparently also $0 \le \lambda x_1 + (1 - \lambda)x_2 \le \lambda y_1 + (1 - \lambda)y_2$.

Example 3.10. Decide which set is a convex polyhedral set, a polyhedron or a cone:

1. $A_1 = \{(x, y, z) \in \mathbb{R}^3 : |x| \le 1,\ |y| \le 1,\ |z| \le 1\}$,
2. $A_2 = \{(x, y, z) \in \mathbb{R}^3 : |x| + |y| + |z| \le 1\}$,
3. $A_3 = \{(x, y, z) \in \mathbb{R}^3 : |x + y + z| \le 1\}$,
4. $A_4 = \{(x, y, z) \in \mathbb{R}^3 : |x| \le |y| \le |z|\}$,
5. $A_5 = \{(x, y, z) \in \mathbb{R}^3 : x + y + z \le 1, x \ge 0,\ y \ge 0,\ z \ge 0\}$.

Solution.

1. convex polyhedron:

$$\begin{aligned} A_1 = \operatorname{conv}\{&(-1, -1, -1), (-1, -1, 1), (-1, 1, 1), (-1, 1, -1), \\ &(1, -1, -1), (1, -1, 1), (1, 1, -1), (1, 1, 1)\}, \end{aligned}$$

2. convex polyhedron:

$$A_2 = \operatorname{conv}\{(1, 0, 0), (-1, 0, 0), (0, 1, 0), (0, -1, 0), (0, 0, 1), (0, 0, -1)\},$$

3. convex polyhedral set (containing straight line), not convex polyhedron, nor cone,
4. nonconvex cone, consider points $(0.5, 0.5, 1)$ and $(0.5, -0.5, 1)$ and $\lambda = 0.5$.
5. convex polyhedron:

$$A_5 = \operatorname{conv}\{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1)\}.$$

4 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 4.1. 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

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

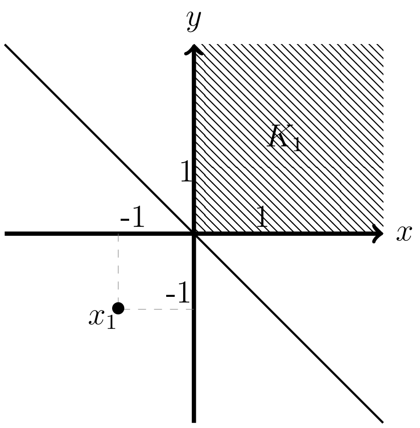
Hint: separate a sequence $x_n \notin K$ which converge to the point x on the boundary, show the convergence of separating hyperplanes characterized by $\gamma_n \neq 0$. See Theorem 1.45 in Lecture notes.

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 4.2. 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 γ is *the normal vector* of the separating/supporting hyperplane.
1. We start with the picture

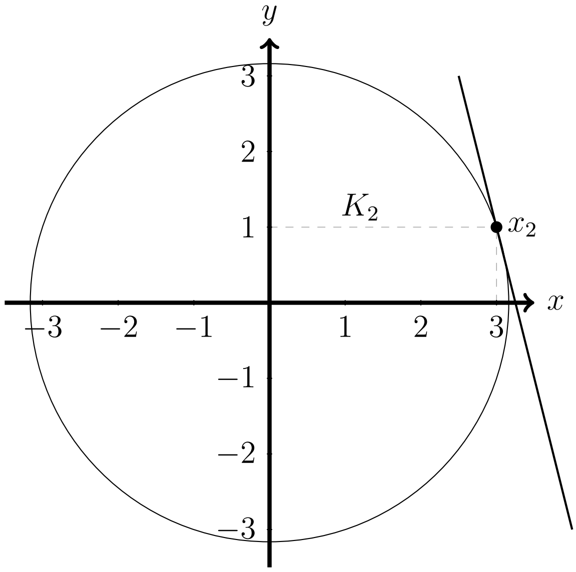


For x_1, K_1 we can use $\gamma = (1,1)$ to construct the separating hyperplane, then

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

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 4.3. 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 4.4. 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 γ, 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.

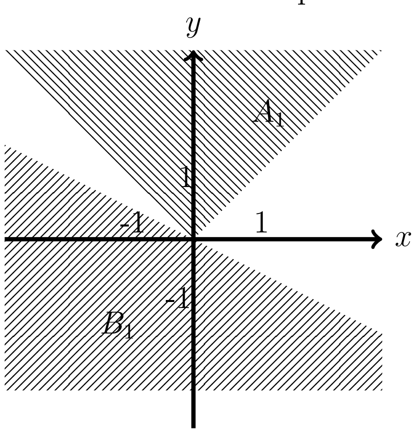
Theorem 4.5. Let $A, B \subset \mathbb{R}^n$ be non-empty convex sets. If $\mathrm{rint}(A) \cap \mathrm{rint}(B) = \emptyset$, then A and B can be properly separated.

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

Example 4.6. Verify whether the following pairs of (convex ?) sets are properly or strictly separable or not. If they are separable, suggest a possible value of γ and verify the property.

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

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.

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

4.1 Farkas theorem

Example 4.7. (*) 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 4.8. Reformulate the Farkas theorem for the sets

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

Solution: Consider M_1. We can transform the constraints to the standard form, i.e. we use $x = x^+ - x^-$, $x^+, x^- \ge 0$, and slack variables $y \ge 0$ 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 $A^Tu \geq 0$, $-A^Tu \geq 0$, and $-Iu \geq 0$ to $A^Tu = 0$ and $u \leq 0$. The modified Farkas theorem is:
$Ax \geq b$ has a solution if and only if for all $u \leq 0$, $A^Tu = 0$ it holds $b^Tu \geq 0$.

Example 4.9. Decide whether the system $\boldsymbol{Ax} = \boldsymbol{b}$ has a non-negative solution, where

$$\boldsymbol{A} = \begin{pmatrix} 2 & 1 & -1 & 0 \\ -3 & 2 & 0 & 4 \end{pmatrix},\ \boldsymbol{b} = \begin{pmatrix} 2 \\ 6 \end{pmatrix}.$$

Solution: The example is directly ready for the use of Farkas' theorem. We will therefore examine whether $b^Tu \geq 0$ holds for (u_1, u_2) such that $A^Tu \geq 0$. Let's first write down what $A^T \geq 0$ means.

$$\begin{aligned} 2u_1 \quad -3u_2 &\geq 0 \\ u_1 \quad +2u_2 &\geq 0 \\ -u_1 \qquad\quad &\geq 0 \\ 4u_2 &\geq 0 \end{aligned}$$

And now we will find out if $b^T \geq 0$ holds, we will modify the expression:

$$2u_1 + 6u_2 = 2 \cdot (u_1 + 2u_2) + 2u_2 \geq 0.$$

The last inequality applies from the second and fourth assumptions equations, i.e. the system has a non-negative solution.

Example 4.10. Use Farkas' theorem to determine whether the set is nonempty:

$$\left\{(x, y, z) \in \mathbb{R}^3 :\ 2x + y - z \leq 3,\ x + 2y + 2z = 4,\ y \geq 0,\ z \geq 0\right\}.$$

Solution: We convert the system into standard form by dividing x into positive and negative parts ($x = x^+ - x^-, x^+ \geq 0, x^- \geq 0$) and by adding a slip variable we convert inequality into equality. Then we can use Farkas' theorem as in the previous case. Let's write down the form of the matrix A and the vector b.

$$A = \begin{pmatrix} 2 & -2 & 1 & -1 & 1 \\ 1 & -1 & 2 & 2 & 0 \end{pmatrix}, b = \begin{pmatrix} 3 \\ 4 \end{pmatrix}.$$

Now we proceed in the same way, we write down the equations of assumptions $A^Tu \geq 0$ and find the non-negativity of b^Tu.

$$\begin{aligned} 2u_1 \quad +u_2 &\geq 0 \\ -2u_1 \quad -u_2 &\geq 0 \\ u_1 \quad +2u_2 &\geq 0 \\ -u_1 \quad +2u_2 &\geq 0 \\ u_1 \qquad\quad &\geq 0 \end{aligned}$$

$$3u_1 + 4u_2 = u_1 + 2 \cdot (u_1 + 2u_2) \geq 0.$$

The last inequality holds from the third and fifth equations of assumptions, i.e. the system has a non-negative solution and the set is non-empty.

5 Linear programming

5.1 Graphical approach

Example 5.1. Use graphical method to solve the following problem:

$$\begin{aligned} \min 4x_1 + 5x_2 & & \\ x_1 + 4x_2 &\geq& 5, \\ 3x_1 + 2x_2 &\geq& 7, \\ x_1, x_2 &\geq& 0. \end{aligned}$$

Denote by M the set of feasible solutions. Modify the problem and discuss the optimal solution

1.
$$\min_{x \in M} x_1 + 4x_2$$

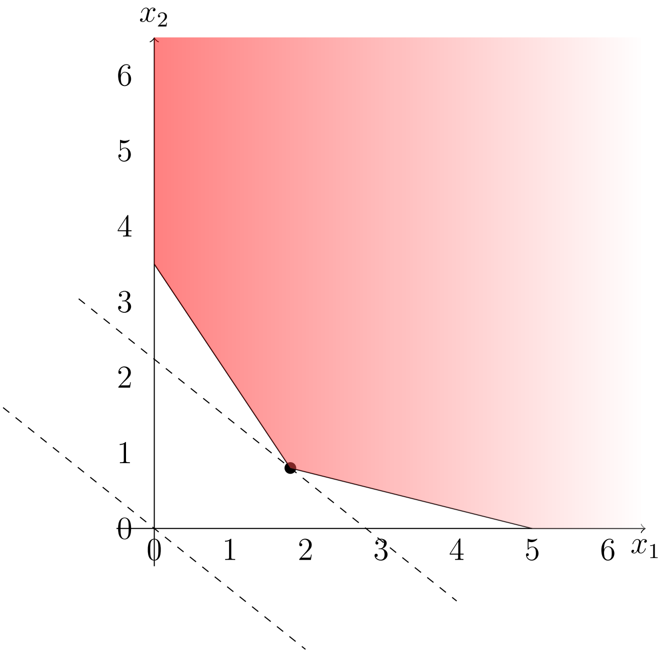
2.
$$\min_{x \in M} x_2$$

3.
$$\min_{x \in M'} 4x_1 + 5x_2,$$
where $M' = M \cap \{x_1 + x_2 \leq 1\}$

4.
$$\max_{x \in M} 4x_1 + 5x_2$$

5.
$$\max_{x \in M''} 4x_1 + 5x_2,$$
where $M'' = M \cap \{x_1 + x_2 \leq 6\}$

Solution.



We can conclude that the optimal solution is $(1.8, 0.8)$.

1. The optimal solution is the line segment connecting points $(1.8, 0.8)$ and $(5, 0)$.
2. The optimal solution is the half line and $(5, 0) + t(1, 0), t \geq 0$.
3. The problem is infeasible, the set of feasible solutions is empty.
4. The problem is unbounded.
5. The optimal solution is $(0, 6)$.

5.2 Direct approach

Example 5.2. Reformulate the following linear programming problem in a standard form:

$$\begin{aligned} \max\ & 2x_1 - 4x_2 - 5x_3 \\ \text{s.t.}\ & -x_1 + 2x_2 - 3x_3 \leq 1, \\ & 4x_1 - 5x_2 - 6x_3 \geq -2, \\ & 7x_1 + 8x_2 - 9x_3 = 3, \\ & x_1 \geq 0,\ x_2 \leq 0,\ x_3 \in \mathbb{R}. \end{aligned}$$

Solution. Use the split $x_3 = x_3^+ - x_3^-$, $x_3^+, x_3^- \geq 0$, substitution $\tilde{x}_2 = -x_2$ and two slack variables $y_1, y_2 \geq 0$ together with the standard equality

$$\max_{x \in M} f(x) = -\min_{x \in M} -f(x)$$

to get the equivalent reformulation in the standard form

$$\begin{aligned} -\min\ & -(2x_1 + 4\tilde{x}_2 - 5x_3^+ + 5x_3^-) \\ \text{s.t.}\ & -x_1 - 2\tilde{x}_2 - 3x_3^+ + 3x_3^- + y_1 = 1, \\ & 4x_1 + 5\tilde{x}_2 - 6x_3^+ + 6x_3^- - y_2 = -2, \\ & 7x_1 - 8\tilde{x}_2 - 9x_3^+ + 9x_3^- = 3, \\ & x_1, \tilde{x}_2, x_3^+, x_3^-, y_1, y_2 \geq 0. \end{aligned}$$

Example 5.3. Derive the sets of extreme points and extreme directions for the convex polyhedral set M defined by

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

Use

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

Reformulate the set as a sum of a convex polyhedron and a convex polyhedral cone. Use the representation to solve the linear programming problems

$$\min_{x \in M} 2x_1 - 3x_2$$

and

$$\min_{x \in M} -3x_1 + 2x_2$$

.

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[2], 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 2×2 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.

Similar approach can be used to get the extreme directions where we solve all homogenous systems with rectangle 2×3 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

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

Then we have

$$M = \operatorname{conv}(\operatorname{ext}(M)) + \operatorname{pos}(\operatorname{extd}(M)).$$

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

Compare the computed points and vectors with the ones obtained from the picture.

Set

$$f_1(x) = 2x_1 - 3x_2,\ \ f_2(x) = -3x_1 + 2x_2.$$

We apply the direct approach/method to solve the LP problems:

1. First, we evaluate the objective function(s) in all extreme points. For the first objective f_1 we obtain: -5, -3, 0. So the candidate for the optimal solution is the extreme point (2,3,0,0). For the second objective f_2 we obtain values: 0, 2, 0. So we have two candidates: (2,3,0,0) and (0,0,1,3).
2. Now, we verify the optimality condition by evaluating the objectives in all extreme directions which must be nonnegative. We obtain $f_1(1,0,1,0) = 2 \geq 0$ and $f_2(1,0,1,0) = -3 < 0$, i.e. for the first objective we verified the optimality condition and the point (2,3,0,0) is optimal solution with optimal value equal to -5, whereas the problem with objective function f_2 is unbounded from below in the direction (1,0,1,0).

Example 5.4. Using the direct approach solve the following LP problem:

$$\begin{aligned} \min\ & 2x_1 + 2x_2 - x_3 - 2x_4 \\ \text{s.t.}\ & 2x_1 - x_2 - x_3 = -2 \\ & -x_1 + 2x_2 - x_4 = -1, \\ & x_1, x_2, x_3, x_4 \geq 0. \end{aligned}$$

Solution. We can obtain

$$\text{ext}(M) = \left\{ \begin{pmatrix} 1 \\ 0 \\ 4 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 0 \\ 5 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 2 \\ 1 \end{pmatrix} \right\},\ \text{extd}(M) = \left\{ \begin{pmatrix} 2 \\ 1 \\ 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 0 \\ 3 \end{pmatrix} \right\}.$$

The values of the objective function in the extreme point are -2, -6, -4, and in the extreme directions 3, 0. Note that the objective function is constant in the direction (1,2,0,3). This implies that the optimal solution exists and is attained on the whole half-line $(0,2,0,5) + t(1,2,0,3)$, $t \geq 0$ with optimal value -6.

5.3 Duality

Example 5.5. Formulate the dual problem to the following linear programming problem:

$$\begin{aligned} \max\ & 2x_1 - 4x_2 - 5x_3 \\ \text{s.t.}\ & -x_1 + 2x_2 - 3x_3 \leq 1, \\ & 4x_1 - 5x_2 - 6x_3 \geq -2, \\ & 7x_1 + 8x_2 - 9x_3 = 3, \\ & x_1 \geq 0,\ x_2 \leq 0,\ x_3 \in \mathbb{R}. \end{aligned}$$

	x_1 ≥ 0	x_2 ≤ 0	x_3 $\in \mathbb{R}$		
	-1	2	-3	$\leq$	1
	4	-5	-6	$\geq$	-2
	7	8	-9	$=$	3
	2	-4	-5	max	

Solution: The initial LP problem (P) can be filled to the following table

The table can be used to formulate the dual problem

		x_1 ≥ 0	x_2 ≤ 0	x_3 $\in \mathbb{R}$		
y_1	≥ 0	-1	2	-3	$\leq$	1
y_2	≤ 0	4	-5	-6	$\geq$	-2
y_3	$\in \mathbb{R}$	7	8	-9	$=$	3
		$\geq$	$\leq$	$=$		min
		2	-4	-5	max	

Now, we can formulate the dual problem (D)

$$\begin{aligned} &\min\ y_1 - 2y_2 + 3y_3 \\ &\text{s.t.}\ -y_1 + 4y_2 + 7y_3 \geq 2, \\ &\quad 2y_1 - 5y_2 + 8y_3 \leq -4, \\ &\quad -3y_1 - 6y_2 - 9y_3 = -5, \\ &\quad y_1 \geq 0,\ y_2 \leq 0,\ y_3 \in \mathbb{R}. \end{aligned}$$

Example 5.6. 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.

Discuss the solution in the case when the right-hand side vector is changed to $(-3, -2)$ or the constraint on variable x_4 to $x_4 \leq 0$.

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

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

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

We can formulate the dual problem:

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

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.

Changing the right-hand side to $(-3, -2)$ leads to a dual problem which is unbounded, which means that the primal problem is infeasible.

Changing the constraint $x_4 \leq 0$ leads to a dual problem which is infeasible, which means that the primal problem is unbounded.

Example 5.7. Apply the graphical method to solve the dual problem to the following problem

$$\begin{aligned} \min\ & 5x_2 + 6x_3 + 2x_4 \\ \text{s.t.}\ & x_1 + x_2 + 2x_3 + x_4 = 2, \\ & x_1 - x_2 - x_3 \le 1, \\ & x_1 \ge 0,\ x_2 \ge 0,\ x_3 \ge 0,\ x_4 \ge 0. \end{aligned}$$

Derive the optimal solution(s) of the primal problem using the complementarity conditions.

5.4 Simplex algorithm

Example 5.8. Solve the example by the simplex method:

$$\begin{array}{lccccc} \min & 2x_1 & -x_2 & & \\ \text{s.t.} & -x_1 & +x_2 & \le & 1 \\ & & x_2 & \le & 3 \\ & x_1, x_2 \ge 0 & & & \end{array}$$

Change the objective function to $-3x_1 + 2x_2$.

Solution: By adding slack variables, we convert the problem to the standard form:

$$\begin{array}{lccccccc} \min & 2x_1 & -x_2 & & & & \\ \text{s.t.} & -x_1 & +x_2 & +\boxed{x_3} & & = & 1 \\ & & x_2 & & +\boxed{x_4} & = & 3 \\ & x_1, x_2, x_3, x_4 \ge 0 & & & & & \end{array}$$

We select the variables x_3 and x_4 as bases and write them into the simplex table, which has the following form for the problem in the standard form and basis B:

			c^T
			x^T
c_B	x_B	$B^{-1}b$	$B^{-1}A$
		$c_B^T B^{-1}b$	$c_B^T B^{-1}A - c^T$

In our case, B is a unit matrix, which enables the expressions of the formulea:

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

We see that we have a feasible basic solution $B^{-1}b = (1, 3) \ge 0$, but the optimality condition $c_B^T B^{-1}A - c^T = (-2, 1, 0, 0) \le 0$ is not satisfied, so we proceed to the simplex iteration. The optimality criterion is violated for the variable x_2, so we include it in the base. Now we look for components on the shares of the current solution $(1,3)$ and the column corresponding to the variable we are currently adding in the constraint matrix $(1,1)$. We consider the ratio only if we are dividing by a positive number. We get ratios of $(1,3)$. We select the minimum ratio and eliminate the corresponding variable x_3 from the basis. Now, using Gauss-Jordan elimination, it adjusts the constraint matrix together with the left side to such a form that B is again a unit matrix.

			2	-1	0	0
			x_1	x_2	x_3	x_4
-1	x_2	1	-1	1	1	0
0	x_4	2	1	0	-1	1
		-1	-1	0	-1	0

We now have a feasible basic solution that meets the optimality condition. So we found the optimal solution $(x_1, x_2, x_3, x_4) = (0, 1, 0, 2)$ with optimal value -1.

If we change the objective function, we obtain:

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

Obviously, the optimality condition is not fulfilled and x_1 should be included into the basis. However, the corresponding column does not contain any positive number, hence there is no candidate to be included into the basis. This means that the problem is unbounded from below. Moreover, we can identify the direction of decrease (1,0,1,0).

Example 5.9. Solve the LP problem using the simplex method

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

Solution: First we add the slack variables and create tthe simplex table.

			3	-1	6	0	0
			x_1	x_2	x_3	x_4	x_5
0	x_4	6	1	3	-2	1	0
0	x_5	2	-2	0	1	0	1
		0	-3	1	-6	0	0

We see that we must add the variable x_2 to the base. Now, when we calculate the ratios, only one comes into consideration, because there is only one positive number in the column in the constraint matrix corresponding to x_2. So we eliminate x_4 from the base.

			3	-1	6	0	0
			x_1	x_2	x_3	x_4	x_5
-1	x_2	2	$\frac{1}{3}$	1	$-\frac{2}{3}$	$\frac{1}{3}$	0
0	x_5	2	-2	0	1	0	1
		-2	$-\frac{10}{3}$	0	$-\frac{16}{3}$	$-\frac{1}{3}$	0

The feasibility and optimality conditions are met, so we have the optimal solution $(0, 2, 0, 0, 2)$ with optimal value -2.

Example 5.10. Using two-phase simplex algorithm solve the LP problem:

$$\begin{array}{lrcrcrcl} \min & 6x_1 & - & 2x_2 & + & 8x_3 & & \\ \text{s.t.} & x_1 & - & 2x_2 & & & \leq & -6 \\ & x_1 & - & 4x_2 & - & 2x_3 & \leq & 8 \\ & & - & 2x_2 & + & x_3 & \geq & 7 \\ & x_1 & & & & & \geq & 0 \\ & & & x_2 & & & \geq & 0 \\ & & & & & x_3 & \geq & 0. \end{array}$$

Discuss the steps if we change the objective function to

$$6x_1 - 2x_2 - 8x_3.$$

Solution: Now we cannot use the same approach, adding slack variables does not give us a unit matrix like in the previous cases. We can multiply the inequalities by -1 to get the identity matrix, but then we will not have an initial feasible solution, i.e., $b \geq 0$ does not hold. Therefore, we will use a two-phase simplex, where in the first phase we are looking for a feasible basis. Let's start by tabulating the problem, adding slack variables, and adjusting to make the right-hand side non-negative.

			6	-2	-8	0	0	0
			x_1	x_2	x_3	x_4	x_5	x_6
?	?	6	-1	2	0	-1	0	0
?	?	8	1	-4	-2	0	1	0
?	?	7	0	-2	1	0	0	-1
		?	?	?	?	?	?	?

Thus, the unit matrix cannot be created from the columns of the constraint matrix. Therefore, we add as many new variables as necessary to create a unit matrix. One column is already in the canonical form, so it is enough to add 2 variables z_1 and z_2 (and thereby create the remaining 2 canonical columns) and solve the auxiliary problem, where all coefficients in the objective function for the variables x and for the variables z. We set the coefficient to 1. Thus, we minimize the sum of z with the goal to eliminate them.

Now, we can solve the LP problem using the simplex.

			0	0	0	0	0	0	1	1
			x_1	x_2	x_3	x_4	x_5	x_6	z_1	z_2
1	z_1	6	-1	2	0	-1	0	0	1	0
0	x_5	8	1	-4	-2	0	1	0	0	0
1	z_2	7	0	-2	**1**	0	0	-1	0	1
		13	-1	0	1	-1	0	-1	0	0
1	z_1	6	-1	**2**	0	-1	0	0	1	0
0	x_5	22	1	-8	0	0	1	-2	0	2
0	x_3	7	0	-2	1	0	0	-1	0	1
		6	-1	2	0	-1	0	0	0	-1
0	x_2	3	-1/2	1	0	-1/2	0	0	1/2	0
0	x_5	46	-3	0	0	-4	1	-2	4	2
0	x_3	13	-1	0	1	-1	0	-1	1	1
		0	0	0	0	0	0	0	-1	-1

- We add x_3 and remove z_2 from the basis.
- We add x_2 and remove z_1.

We have now obtained a feasible solution for the original problem, so we get rid of the auxiliary variables z and solve the original problem.

			6	-2	8	0	0	0
			x_1	x_2	x_3	x_4	x_5	x_6
-2	x_2	3	-1/2	1	0	-1/2	0	0
0	x_5	46	-3	0	0	-4	1	-2
8	x_3	13	-1	0	1	-1	0	-1
		98	-13	0	0	-7	0	-8

The optimality condition is satisfied. Hence, we have obtained an optimal solution (without additional iterations):

$$(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4, \hat{x}_5, \hat{x}_6) = (0, 3, 13, 0, 46, 0)$$

with optimal value -110.

			6	-2	**-8**	0	0	0
			x_1	x_2	x_3	x_4	x_5	x_6
-2	x_2	3	-1/2	1	0	-1/2	0	0
0	x_5	46	-3	0	0	-4	1	-2
-8	x_3	13	-1	0	1	-1	0	-1
		-110	**3**	0	0	9	0	8

If we change the objective function, i.e. set $c_3 = -8$, the optimality condition is not satisfied, for example, for x_1, and the column belonging to x_1 has no positive component, so the problem is unbounded from below. We can identify the direction in which the objective function decreases:

$$(1, 1/2, 1, 0, 3, 0).$$

6 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 6.1. 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 6.2. 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, ν), $(y, \eta) \in \mathrm{epi}(f)$, i.e. $\nu \ge f(x)$ and $\eta \ge f(y)$. From 2. we know that

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

i.e.

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

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

Remark 6.3. 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 6.4. Decide whether the following sets are convex:

$$
\begin{aligned}
M_1 &= \left\{(x,y) \in \mathbb{R}^2_+ :\ ye^{-x} - x \geq 1\right\}, && (6.1)\\
M_2 &= \left\{(x,y) \in \mathbb{R}^2 :\ x \geq 2 + y^2\right\}, && (6.2)\\
M_3 &= \left\{(x,y) \in \mathbb{R}^2 :\ x^2 + y \log y^4 \leq 139,\ y \geq 2\right\}, && (6.3)\\
M_4 &= \left\{(x,y) \in \mathbb{R}^2 :\ \log x - y^2 \geq 1,\ x \geq 1,\ y \geq 0\right\}, && (6.4)\\
M_5 &= \left\{(x,y) \in \mathbb{R}^2 :\ (x^3 + e^y) \log(x^3 + e^y) \leq 49,\ x \geq 0,\ y \geq 0\right\}, && (6.5)\\
M_6 &= \left\{(x,y) \in \mathbb{R}^2 :\ x \log x + xy \geq 0,\ x \geq 1\right\}, && (6.6)\\
M_7 &= \left\{(x,y) \in \mathbb{R}^2 :\ 1 - xy \leq 0,\ x \geq 0\right\}, && (6.7)\\
M_8 &= \left\{(x,y,z) \in \mathbb{R}^3 :\ \frac{1}{2}(x^2 + y^2 + z^2) + yz \leq 1,\ x \geq 0,\ y \geq 0\right\}, && (6.8)\\
M_9 &= \left\{(x,y,z) \in \mathbb{R}^2 :\ 3x - 2y + z = 1\right\}. && (6.9)
\end{aligned}
$$

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 6.5. Establish conditions under which the following sets are convex:

$$
\begin{aligned}
M_{10} &= \left\{x \in \mathbb{R}^n :\ \alpha \leq a^T x \leq \beta\right\}, \text{ for some } a \in \mathbb{R}^n, \alpha, \beta \in \mathbb{R}, && (6.10)\\
M_{11} &= \left\{x \in \mathbb{R}^n :\ \left\|x - x^0\right\| \leq \|x - y\|,\ \forall y \in S\right\}, \text{ for some } S \subseteq \mathbb{R}^n, && (6.11)\\
M_{12} &= \left\{x \in \mathbb{R}^n :\ x^T y \leq 1,\ \forall y \in S\right\}, \text{ for some } S \subseteq \mathbb{R}^n. && (6.12)
\end{aligned}
$$

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 6.6. Verify if the following functions are convex:

$$\begin{aligned}
f_1(x,y) &= x^2y^2 + \frac{x}{y}, \ x>0, y>0, && (6.13)\\
f_2(x,y) &= xy, && (6.14)\\
f_3(x,y) &= \log(e^x + e^y) - \log x, \ x>0, && (6.15)\\
f_4(x,y) &= \exp\{x^2 + e^{-y}\}, \ x>0, \ y>0, && (6.16)\\
f_5(x,y) &= -\log(x+y), \ x>0, \ y>0, && (6.17)\\
f_6(x,y) &= \sqrt{e^x + e^{-y}}, && (6.18)\\
f_7(x,y) &= x^3 + 2y^2 + 3x, && (6.19)\\
f_8(x,y) &= -\log(cx+dy), \ c,d \in \mathbb{R}, && (6.20)\\
f_9(x,y) &= \frac{x^2}{y}, \ y>0, && (6.21)\\
f_{10}(x,y) &= xy\log xy, \ x>0, y>0, && (6.22)\\
f_{11}(x,y) &= |x+y|, && (6.23)\\
f_{12}(x) &= \sup_{y \in \mathrm{Dom}(f)} \{x^T y - f(y)\} = f^*(x), \ f: \mathbb{R}^n \to \mathbb{R}, && (6.24)\\
f_{13}(x) &= \|Ax - b\|_2^2. && (6.25)
\end{aligned}$$

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, bacause

$$x^T A^T Ax = (Ax)^T Ax = \sum_{i=1}^{n} (a_{i\cdot}^T x)^2 \geq 0,$$

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 6.7. Show that

$$f_{14}(x) = -\sum_{i=1}^{n} \ln(b_i - a_i^T x). \tag{6.26}$$

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

Example 6.8. 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

$$g(x) = \inf_{y \in C} f(x, y). \tag{6.27}$$

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

$$g(x_1) + \varepsilon \geq f(x_1, y_1),$$
$$g(x_2) + \varepsilon \geq f(x_2, y_2),$$

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

$$f(x) = h\big(g_1(x), \dots, g_k(x)\big). \tag{6.28}$$

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

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

where the first inequality follows from that $g_1, \dots, g_k$ 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 6.10. (*) Verify that the geometric mean is concave[3]:

$$f(x) = \left(\prod_{i=1}^{n} x_i \right)^{1/n}, \ x \in (0, \infty)^n. \tag{6.29}$$

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

6.1 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 6.11. 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 epi(f) at $(x, f(x))$.

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

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

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

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

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

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

Example 6.12. (*) Consider (do not necessarily prove, rather think about) the following properties of subgradient:

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

1. if f is convex, then $\partial f(x) \neq \emptyset$ for all $x \in \text{rint dom} f$.
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 6.13. Derive the subdifferential for the following functions:

$$\begin{array}{rcll} f_1(x) & = & |x|, & \\ f_2(x) & = & x^2 & \text{if } x \leq -1, \\ & & -x & \text{if } x \in [-1, 0], \\ & & x^2 & \text{if } x \geq 0, \\ f_3(x,y) & = & |x+y| & \text{at } (0,0), \\ f_4(x) & = & \max\{x, x^2\}, & x \in \mathbb{R}. \end{array}$$

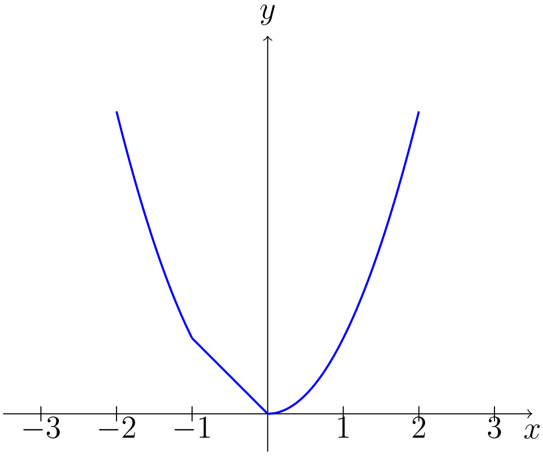
Solution:
1. The function f_1 is convex and nondifferentiable at $x = 0$. If we write the definition of subgradient in that point

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

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

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

4. We can see that the function f_4 is differentiable almost everywhere with the exception of two points $x \in \{0,1\}$, i.e. we have

$$\begin{array}{rcll} \partial f_4(x) & = & \{2x\} & \text{if } x \in (-\infty, 0), \\ & & [0,1] & \text{if } x = 0, \\ & & \{1\} & \text{if } x \in (0,1), \\ & & [1,2] & \text{if } x = 1, \\ & & \{2x\} & \text{if } x \in (1,\infty). \end{array}$$

Lemma 6.14. 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 $\sum_{i=1}^k a_i \in \partial f(x)$.

7 Nonlinear programming problems: Karush–Kuhn–Tucker Optimality conditions

7.1 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:

$$\begin{aligned} \min\ & f(x) \\ \text{s.t. } & g_i(x) \le 0,\ i = 1, \dots, m, \\ & h_j(x) = 0,\ j = 1, \dots, l, \end{aligned} \tag{7.1}$$

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

We say that the **problem is convex** if functions f, $g_i, \forall_i$ are convex and $h_j, \forall_j$ are affine.

Define the **Lagrange function** by

$$L(x, u, v) = f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{j=1}^{l} v_j h_j(x),\ u_i \ge 0. \tag{7.2}$$

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

$$\begin{aligned} \text{i) } & g_i(x) \le 0,\ i = 1, \dots, m,\ h_j(x) = 0,\ j = 1, \dots, l, \\ \text{ii) } & u_i g_i(x) = 0,\ u_i \ge 0,\ i = 1, \dots, m, \\ \text{iii) } & \nabla_x L(x, u, v) = 0, \end{aligned} \tag{7.3}$$

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, ...).

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

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

7.2 Karush–Kuhn–Tucker optimality conditions

Example 7.1. Consider the nonlinear programming problem

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

Compute the Lagrange multipliers at point $(2, 4)$.

Solution: Define the Lagrange function

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

Write the KKT optimality conditions:

i) feasibility,

ii) complementarity

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

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 $u_1 = 1 \ge 0$ and $u_2 = 5 \ge 0$. We have obtained KKT point $(2, 4, 1, 5)$ and $(2, 4)$ is a global solution, because the problem is convex (the objective function is a sum of two one-dimensional quadratic functions, the first constraint is a difference of one-dimensional quadratic and linear function, the second constraint is linear).

Example 7.2. (*) Consider the nonlinear programming problem

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

Compute the Lagrange multipliers at point (2,1).

Example 7.3. 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[4]:

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

The problem is obviously convex (sum of one-dimensional quadratic functions in the objective, linear constraints). We can 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

$$\begin{aligned} \text{i)} \ & \text{feasibility}, \\ \text{ii)} \ & u_1(-x_1 - x_2 + 4) = 0, \ u_1 \geq 0, \\ & u_2(-2x_1 - x_2 + 5) = 0, \ u_2 \geq 0, \\ \text{iii)} \ & \frac{\partial L}{\partial x_1} = 2x_1 - u_1 - 2u_2 = 0, \\ & \frac{\partial L}{\partial x_2} = 2x_2 - u_1 - u_2 = 0. \end{aligned} \tag{7.4}$$

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

$$\begin{aligned} 2x_1 - u_1 &= 0, \\ 2x_2 - u_1 &= 0, \\ x_1 + x_2 &= 4. \end{aligned} \tag{7.5}$$

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

$$\begin{aligned} 2x_1 - 2u_2 &= 0, \\ 2x_2 - u_2 &= 0, \\ 2x_1 + x_2 &= 5. \end{aligned} \tag{7.6}$$

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

$$\begin{aligned} u_1 + 2u_2 &= 2, \\ u_1 + u_2 &= 6. \end{aligned} \tag{7.7}$$

[4] The square root can be omitted.

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 7.4. 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

$$L(x_1, x_2, u_1, u_2) = x_1^2+x_2^2+u_1(x_1^2+x_2^2-5)-u_2x_1-u_3x_2+v(x_1+2x_2-4),\ u_1, u_2, u_3 \geq 0.$$

Derive the KKT conditions

$$\begin{aligned} \text{i)}\ & \text{feasibility}, \\ \text{ii)}\ & u_1(x_1^2 + x_2^2 - 5) = 0,\ u_1 \geq 0, \\ & u_2x_1 = 0,\ u_2 \geq 0, \\ & u_3x_2 = 0,\ u_3 \geq 0, \\ \text{iii)}\ & \frac{\partial L}{\partial x_1} = 2x_1 + 2u_1x_2 - u_2 + v = 0, \\ & \frac{\partial L}{\partial x_2} = 2x_2 + 2u_1x_2 - u_3 + 2v = 0. \end{aligned} \tag{7.8}$$

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 7.5. Consider the problem

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

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

Solution: Write the Lagrange function

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

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

$$\begin{aligned}
\text{i) } & \text{feasibility}, \\
\text{ii) } & u_1(x_1x_2x_3 - 1) = 0,\ u_1 \geq 0, \\
& u_2(-x_1 - x_3 + c) = 0,\ u_2 \geq 0, \\
\text{iii) } & \frac{\partial L}{\partial x_1} = 2e^{x_1-1} - 2(x_2 - x_1) + u_1x_2x_3 - u_2 = 0, \\
& \frac{\partial L}{\partial x_2} = 2(x_2 - x_1) + u_1x_1x_3 = 0, \\
& \frac{\partial L}{\partial x_3} = 2x_3 + u_1x_1x_2 - u_2 = 0.
\end{aligned} \tag{7.9}$$

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

$$\begin{aligned}
& 2 + u_1 - u_2 = 0, \\
& u_1 = 0, \\
& 2 + u_1 - u_2 = 0,
\end{aligned} \tag{7.10}$$

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 $c \leq 2$. However, if $c < 2$, then the second complementarity condition is not fulfilled $u_2(-x_1 - x_3 + c) \neq 0$. Therefore we have obtained KKT point (1,1,1,0,2) only if $c = 2$ when the second constraint is active.

Example 7.6. 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

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

is an optimal solution.

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

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

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

$$\begin{aligned}
&\text{i)} \sum_{i=1}^{n} \left(x_i - \frac{1}{n} \right)^2 \leq \frac{1}{n(n-1)}, \ \sum_{i=1}^{n} x_i = 1, \\
&\text{ii)} \ u \left(\sum_{i=1}^{n} \left(x_i - \frac{1}{n} \right)^2 - \frac{1}{n(n-1)} \right) = 0, \ u \geq 0, \\
&\text{iii)} \ \frac{\partial L}{\partial x_1} = 1 + 2u \left(x_1 - \frac{1}{n} \right) + v = 0, \\
&\frac{\partial L}{\partial x_i} = 2u \left(x_1 - \frac{1}{n} \right) + v = 0, \ i \neq 1.
\end{aligned} \tag{7.11}$$

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

$$\begin{aligned}
&1 - \frac{2u}{n} + v = 0, \\
&2u \left(\frac{1}{n-1} - \frac{1}{n} \right) + v = 0, \ (\forall i \neq 1).
\end{aligned} \tag{7.12}$$

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

$$\begin{aligned}
u &= \frac{n-1}{2} \geq 0, \\
v &= \frac{-1}{n} \in \mathbb{R}.
\end{aligned} \tag{7.13}$$

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 7.7. 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:

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

iii) optimality:

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

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 $J_1 \cup J_1 = \{1, \dots, n\}$ and $J_1 \cap J_2 = \emptyset$. (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

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

i.e. we have

$$u_j = -c_j\, n^2 + v, \ j = 1, \dots, n.$$

If all u_j are nonnegative, we have a KKT point. It is enough to select

$$v \geq \max_j \{c_j\, n^2\}.$$

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

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

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 7.8. (*) 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 7.9. (*) Write the KKT conditions for a linear programming problem.

Example 7.10. Consider the (water-filling[5]) 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:

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

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 $i \in \{1, \ldots, n\}$, 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 $v \in (0, 1/\alpha_i)$.

[5] 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 α_i. Note that there is no closed-form formula for v, we are satisfied with its existence.

Example 7.11. (*) Derive the least square estimate for coefficients in the linear regression model under linear constraints, i.e. solve the problem

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

Example 7.12. (*) Consider the problem

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

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?

7.3 Constraint qualification conditions

Example 7.13. Consider the problem

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

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

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

Derive the KKT conditions

$$\begin{aligned} &\text{i) feasibility,}\\ &\text{ii) } u_1\big((x_1-1)^2+(x_2-1)^2-1\big) = 0,\ u_1 \geq 0,\\ &\quad u_2\big((x_1-1)^2+(x_2+1)^2-1\big) = 0,\ u_2 \geq 0,\\ &\text{iii) } \frac{\partial L}{\partial x_1} = 1 + 2u_1(x_1-1) + 2u_2(x_1-1) = 0,\\ &\quad \frac{\partial L}{\partial x_2} = 2u_1(x_2-1) + 2u_2(x_2+1) = 0. \end{aligned} \tag{7.14}$$

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

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

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. 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 7.14. Consider the problem with real parameter a

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

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 7.15. Consider the problem

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

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

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

Derive the KKT conditions

$$\begin{aligned} \text{i) } & \text{feasibility}, \\ \text{ii) } & u_1\big(-(1 - x_1)^3 + x_2\big) = 0, \ u_1 \ge 0, \\ & u_2\big(x_2\big) = 0, \ u_2 \ge 0, \\ \text{iii) } & \frac{\partial L}{\partial x_1} = 2(x_1 - 2) + 3u_1(x_1 - 1)^2 = 0, \\ & \frac{\partial L}{\partial x_2} = 2x_2 + u_1 - u_2 = 0. \end{aligned} \tag{7.15}$$

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 $u_{1,2} \ge 0$ 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.

8 Appendix

8.1 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 8.1. Consider the problem

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

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

$$\begin{aligned} &\text{i)}\ \frac{\partial L}{\partial x_1} = 1 + 2y_1(x_1-1) + 2y_2(x_1-1) \ge 0,\ x_1\frac{\partial L}{\partial x_1} = 0,\ x_1 \ge 0, \\ &\quad \frac{\partial L}{\partial x_2} = 2y_1(x_2-1) + 2y_2(x_2+1) \ge 0,\ x_2\frac{\partial L}{\partial x_2} = 0,\ x_2 \ge 0, \\ &\text{ii)}\ \frac{\partial L}{\partial y_1} = (x_1-1)^2+(x_2-1)^2-1 \le 0,\ y_1\big((x_1-1)^2+(x_2-1)^2-1\big) = 0,\ y_1 \ge 0, \\ &\quad \frac{\partial L}{\partial y_2} = (x_1-1)^2+(x_2+1)^2-1 \le 0,\ y_2\big((x_1-1)^2+(x_2+1)^2-1\big) = 0,\ y_2 \ge 0. \end{aligned} \tag{8.1}$$

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

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

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 $x \geq 0$ such that

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

LI CQ: We can compute the gradients

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

which are obviously not linearly independent.

8.2 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).

8.2.1 Quasiconvex functions

Definition 8.2. We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is quasiconvex, if all its level sets are convex, i.e. $\{x \in \text{dom}(f) : \; f(x) \leq \alpha\}$ is convex for all $\alpha \in \mathbb{R}$.

Obviously all convex functions are quasiconvex.

Example 8.3. 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 $\{x \in (0,\infty) : \; \ln(x) \leq \alpha\} = (0, e^{\alpha}]$ are intervals, thus convex sets.

Consider the cumulative distribution function of a discrete random variable with equiprobable realizations ξ_1, ξ_2, ξ_3:

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

CDF is not convex, but it is quasiconvex because the α-level sets are intervals

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

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

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

Example 8.4. Show that the following property is equivalent to the definition of quasiconvexity:

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

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

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

$$f(\lambda x + (1-\lambda)y) \leq \alpha = \max\{f(x), f(y)\}, \tag{8.2}$$

which we wanted to show.

Now let the property (8.2) be fulfilled. Choose α such that the corresponding level set is nonempty. Take two arbitrary x, y from the α-level set and $\lambda \in (0,1)$. Since the property (8.2) 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 α-level set and the set is convex.

Example 8.5. 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 α-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 8.6. 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 α to a union of disjoint intervals which is not a convex set.

Example 8.7. (*) 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 8.8. (*) 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 8.9. 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 (8.2), i.e. it holds

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

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

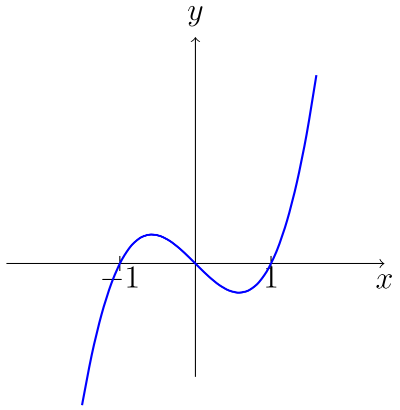
$$g\Big(f_1(\lambda x + (1-\lambda)y)\Big) \leq g\Big(\max\{f_1(x), f_1(y)\}\Big) = \max\Big\{g\big(f_1(x)\big), g\big(f_1(y)\big)\Big\},$$

which confirms the quasiconvexity of the composition.

Example 8.10. 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 8.6.
2. Left to the readers.

Example 8.11. Verify that the following functions are quasiconvex on given sets:

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

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 8.12. 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 α-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.

8.2.2 Additional examples: strictly quasiconvex functions

Definition 8.13. 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 8.14. (*) 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.

8.2.3 Pseudoconvex functions

Definition 8.15. 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 8.16. 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 8.17. 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 8.18. (*) Find a pseudoconvex function which is not convex.

Hint: Consider increasing functions.

Example 8.19. 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 8.20. (*) Consider fractional function f as defined in Example 8.12. Moreover, let S be open and g, h be differentiable on S. Show that f is pseudoconvex.

Example 8.21. 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 8.4. 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 8.22. 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 8.23. (*) 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).

8.3 Optimality conditions based on directions

Example 8.24. 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 8.25. Verify the optimality conditions at point $(2, 4)$ for problem

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

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

$$x_2 \le 5.$$

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

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

for all feasible solutions, which is fulfilled because $|x_1| \le 2$ and $x_2 \le 4$, 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 $|x_1| \le \sqrt{5}$ and $x_2 \le 5$, 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 8.26. Consider open $\emptyset \ne S \subseteq \mathbb{R}^n$, $f : S \to \mathbb{R}$, and define set of improving directions of f at $x \in S$

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

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 δ-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 δ'-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 8.27. Consider the global optimization problem from Example 8.24

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

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 8.28. 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 8.29. Consider open $\emptyset \neq S \subseteq \mathbb{R}^n$, functions $g_i : S \to \mathbb{R}$, and the set of feasible solutions

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

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[6]

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

In general, it holds[7]

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

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

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

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 8.30. 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

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

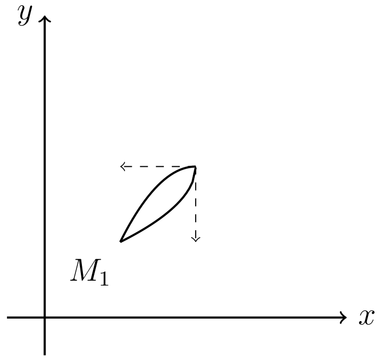
So we have

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

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

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

[6]Set of indices of active constraints: $I_g(x) = \{i : \ g_i(x) = 0\}$
[7]Note that there are sufficient conditions stating which approximation is tight, see Bazaraa et al. (2006) for details.



Consider M_2 and compute

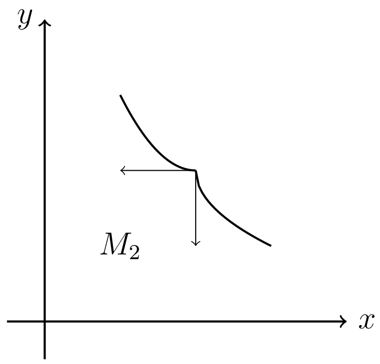
$$\nabla g_1(2,2) = \begin{pmatrix} -2(x-2) \\ 1 \end{pmatrix}|_{(2,2)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix},\ \nabla g_2(2,2) = \begin{pmatrix} 1 \\ -2(y-2) \end{pmatrix}|_{(2,2)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

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

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

Example 8.32. Derive the above defined sets of directions for the problem

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

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

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

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 8.33. 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 $h_j(x) = 0,\ j = 1, \ldots, l$, where $h_j : S \to \mathbb{R}$ are differentiable:

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

Solution: We have

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

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.

8.4 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.

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

$$\begin{aligned} I_g(x) &= \{i :\ g_i(x) = 0\}, \\ I_g^0(x) &= \{i :\ g_i(x) = 0,\ u_i = 0\}, \\ I_g^+(x) &= \{i :\ g_i(x) = 0,\ u_i > 0\}, \end{aligned} \tag{8.3}$$

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

$$\begin{aligned} &z^T \nabla_x g_i(x) = 0, \ i \in I_g^+(x), \\ &z^T \nabla_x g_i(x) \leq 0, \ i \in I_g^0(x), \\ &z^T \nabla_x h_j(x) = 0, \ j = 1, \dots, l, \end{aligned} \tag{8.4}$$

it holds

$$z^T \, \nabla_{xx}^2 L(x, u, v) \, z > 0. \tag{8.5}$$

Then x is a strict local minimum of the nonlinear programming problem (7.1).

Example 8.34. Consider the problem

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

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

$$\begin{aligned} &\text{i) feasibility,} \\ &\text{ii) } u_1 \left(x^2 + y^2 - 1 \right) = 0, \ u_1 \geq 0, \\ &\quad u_2 \left((x-1)^3 - y \right), \ u_2 \geq 0, \\ &\text{iii) } \frac{\partial L}{\partial x} = -1 + 2u_1 x + 3u_2 (x-1)^2 = 0, \\ &\quad \frac{\partial L}{\partial y} = 2u_1 y - u_2 = 0. \end{aligned} \tag{8.6}$$

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 (8.4), (8.5). 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 8.35. Consider the problem

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

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

$$\begin{aligned} &\text{i) feasibility,} \\ &\text{ii) } -u_1 x = 0,\ u_1 \geq 0, \\ &\quad\ -u_2 y = 0,\ u_2 \geq 0, \\ &\text{iii) } \frac{\partial L}{\partial x} = 2x - u_1 + v = 0, \\ &\quad\ \frac{\partial L}{\partial y} = -2y - u_2 - v = 0. \end{aligned} \tag{8.7}$$

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 (8.4), (8.5). 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 8.36. 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

$$\begin{aligned} &\text{i) feasibility,} \\ &\text{ii) } - \\ &\text{iii) } \frac{\partial L}{\partial x} = -2x - 4y + 2vx = 0, \\ &\quad\ \frac{\partial L}{\partial y} = -2y - 4x + 2vy = 0. \end{aligned} \tag{8.8}$$

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 (8.4), (8.5). 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 8.37. (*) 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.

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.