Exam Optimization 1 - Spring 2022

Problem 1 **(Steepest Descent with Exact Line Search)** [20 points]

a) In the lectures, we have proven the following results: *Let x^k be the sequence generated by the steepest descent algorithm with exact line search. Then, for all k, $x^{k+1} - x^k$ is orthogonal to $x^{k+2} - x^{k+1}$.* Show that this assertion

1. does in general not hold if the descent direction is different from steepest descent (but exact line search is maintained) by providing an explicit example;
2. does in general not hold if the step length is not chosen according to the exact line search by providing explicit example.

b) Consider the quadratic optimization problem $\min_{x\in\mathbb{R}^n} f(x)$ with $f(x) = x^T Qx + b^T x$.

1. Sketch the level curves of f in the case $n = 2$ and with a symmetric positive definite matrix $\mathbb{R}^{n,n} \ni Q \neq I$ (I denotes the identity matrix) and a vector $b \in \mathbb{R}^n$ that you may choose yourself. Include a starting point x^0 and two further iteration points x^1, x^2 for the steepest descent method with exact line search. Choose x^0 in such a way that the algorithm does not terminate after these two iterations.
2. Let $Q = I$ (identity matrix). In another coordinate system (i.e., do not use the same coordinate system as in 1.), sketch the level curves and a starting point x^0 for the steepest descent method with exact line search. Explain how many iterations are needed until the method converges.

Solution:

a) 1. Let the search direction p^k be s.t. $p^k \neq \nabla f(x^k)$. For example, let $f(x) = x_1^2 + x_2^2$, thus $\nabla f(x) = (2x_1, 2x_2)^T$ and $x^0 = (1, 0)^T$. We choose $p^0 = (-1, 1)^T$. We have $\nabla f(x^0)^T p^0 = (2, 0)^T(-1, 1) = -2 < 0$ and thus, p^0 is indeed a descent direction. Then $x^1 = x^0 + \alpha_0 p^0$. In order to obtain the step size α_0, we solve the convex problem

$$\min_{\alpha\geq 0} f(x^0 + \alpha p^0),$$

and we get

$$\frac{d}{d\alpha}(x_1^0 + \alpha p_1^0)^2 + (x_2^0 + \alpha p_2^0)^2 = 2p_1^0(x_1^0 + \alpha p_1^0) + 2p_2^0(x_2^0 + \alpha p_2^0) = -2 + 4\alpha \stackrel{!}{=} 0,$$

which leads to $\alpha = \frac{1}{2}$. Therefore,

$$x^1 = x^0 + \alpha_0 p^0 = \left(\frac{1}{2}, \frac{1}{2}\right)^T.$$

Let us now choose the next search direction, for example, $p^1 = (0, -1)^T$. We have $\nabla f(x^1)^T p^1 = (1, 1)^T(0, -1) = -1 < 0$ and thus, p^1 is indeed a descent direction. Moreover, we have ${p^1}^T p^0 = (0, -1)^T(-1, 1) = -1 \neq 0$, and so, p^1 and p^0 are not orthogonal.

a) 2. Let us choose $f(x) = x^2$, $x^0 = 1$. Then $f'(x) = 2x$ and $f'(x^0) = 2$. Then $p^0 = -f'(x^0) = -2$ is the steepest descent direction. Choose $\alpha = \frac{1}{4}$, and thus, $x^1 = x^0 + \alpha p^0 = 1 + \frac{1}{4} \cdot (-2) = \frac{1}{2}$. We have $\nabla f(x^1) = 1$ and let $p^1 = -1$. Then we have

$${p^1}^T p^0 = (-1) \cdot (-2) = 2 \neq 0,$$

and thus, p^1 and p^0 are not orthogonal.
b) 1.

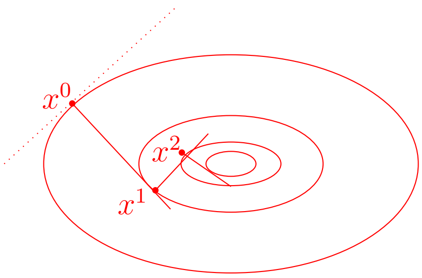


Figure 1: Steepest descent with exact line search: Zick-zack lines.

2. See Figure 2.
Grading remark: In general, the questions have been quite well answered. Still, sometimes explanations were missing, and there was just some graphical representation, particularly for point 1.b.2. For this point, an explanation was necessary to get the 5 points. Some students did not provide explicit examples but presented some theoretical derivation. In this case, most of the points were usually still assigned but not the maximum amount. The points have been evenly distributed among the 4 parts 1.a.1, 1.a.2, 1.b.1 and 1.b.2.

Problem 2 **(Lower Semi-Continuity, Convexity)** [20 points] Let $S \subset X = \mathbb{R}^n, S \neq \mathbb{R}^n$, be nonempty, convex and closed. We consider the function $\sigma_S : X \to \mathbb{R} \cup \{+\infty\}$ defined by

$$\sigma_S(z) := \sup_{x \in S} \langle z \mid x \rangle, \; z \in X,$$

where $\langle \cdot \mid \cdot \rangle$ denotes the duality product.

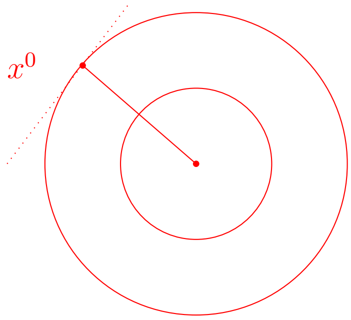


Figure 2: Steepest descent with exact line search: Zick-zack lines. Because $Q = I$, the algorithm stops at the minimal solution after one iteration.

a) Show that σ_S is proper, i.e., $\sigma_S(z) > -\infty$ for every $z \in \mathbb{R}^n$ and $\sigma_S(z) < +\infty$ for at least one $z \in \mathbb{R}^n$.

b) Show that σ_S is lower semi-continuous.

c) Show that σ_S is convex.

d) Derive the function σ_S for the unit ball $S = B(0, 1) := \{x \in \mathbb{R}^n | \; ||x||_2 \leq 1\}$, i.e., derive a simplified representation of the function σ_S that does not use maximum/supremum, but consists of elementary operations like $+, -, \cdot, ||.||$.

Solution:
a) As S is not empty, we cannot have $\sigma_S(z) = -\infty$ for all $z \in \mathbb{R}^n$. Since $S \neq \mathbb{R}^n$, $\sigma_S(z)$ cannot be everywhere $+\infty$.
b) A function f is lower semi-continuous if its epigraph $\{(z, r) \in \mathbb{R}^n \times \mathbb{R} | f(z) \leq r\}$ is closed. We show that

$$\forall r \in \mathbb{R} : \quad S_r := \{z \in \mathbb{R}^n | \sigma_S(z) \leq r\}$$

is closed. Let $r \in \mathbb{R}$ be arbitrary. If $S_r = \emptyset$, then S_r is closed. Otherwise, we consider a convergent sequence $\{z^k\} \subseteq S_r$ $(k = 1, 2, \ldots)$ with limit value $y \in \mathbb{R}^n$. We need to show that $y \in S_r$. From the definition of σ_S we get

$$\forall k : \sup_{x \in S} \langle z^k \mid x \rangle \leq r,$$

and

$$\forall k, \forall x \in S : \langle z^k \mid x \rangle \leq r.$$

Because of the continuity of $\langle \cdot \mid x\rangle$, we get

$$\forall x \in S : \langle z^k \mid x\rangle \to \langle y \mid x\rangle,$$

and therefore,

$$\forall x \in S : \langle y \mid x\rangle \le r$$

and

$$\sup_{x\in S}\langle y \mid x\rangle \le r.$$

Hence, $y \in S_r$.

c) σ_S is defined as the pointwise supremum of certain linear, i.e. convex functions and is thus itself convex.

d) $\sigma_{B(0,1)}$ is finite for every $z \in \mathbb{R}^n$, as $B(0,1)$ is bounded. For fixed $z \in \mathbb{R}^n$, we have

$$\forall x \in B(0,1) : |\langle z \mid x\rangle| \le ||z|| \cdot ||x|| \le ||z||$$

because $||x|| \le 1$. Therefore $\sigma_{B(0,1)}(z) \le ||z||$, and $\sigma_{B(0,1)}(z)$ is therefore finite for every $z \in \mathbb{R}^n$. The last inequality yields

$$\forall z \in \mathbb{R}^n : \sigma_{B(0,1)}(z) = ||z||.$$

Because for $z = 0$, we have $\sigma_{B(0,1)}(0) = 0$ and for a $z \neq 0$ and

$$x_1 := \frac{z}{||z||} \in B(0,1),$$

we have $\langle z \mid x_1\rangle = ||z||$, therefore $\sigma_{B(0,1)}(z) = ||z||$.

Grading remark: Almost everyone answered the convexity point correctly, but there have been some problems in parts (a) and (b). In (a) was essential to remark that S is non-empty. Some students assumed the boundedness of S, which is more than what is given by the exercise. Ok, in general, part (d). The 20 points are evenly distributed among the four parts of the exercise.

Problem 3 **(Linear Programs - Duality)** [20 points] Consider the following linear program

$$
\begin{array}{llrrrrl}
\text{(LP)} & \min\limits_{x\in\mathbb{R}^4} & -4x_1 & +3x_2 & & & \\
 & \text{subject to} & -x_1 & +x_2 & -x_3 & & = 1, \\
 & & 3x_1 & -x_2 & & +x_4 & \ge 8, \\
 & & x_1 & & & & \ge 0, \\
 & & & & x_3 & & \ge 0 \\
 & & & & & x_4 & \ge 0.
\end{array}
$$

1. Write down (LP) as an (LP3), i.e., in the form

$$\min c^\top x$$
$$\text{subject to } Ax = b,\ x \ge 0.$$

2. Write down the corresponding dual program of the (LP3). Show that this dual problem does not have any feasible points.

3. What does your result in **b)** mean for the solution of the primal problem (LP)? Explain your reasoning.

Solution:

1. We formulate (LP) as (LP3). Because there are no nonnegativity constraints on $x_2 \in \mathbb{R}$, we introduce the following variables for $i = 2$:

$$x_i^+ = \begin{cases} x_i & (x_i \geq 0) \\ 0 & (x_i < 0) \end{cases}$$

$$x_i^- = \begin{cases} 0 & (x_i > 0) \\ -x_i & (x_i \leq 0). \end{cases}$$

This means $x_i = x_i^+ - x_i^-$ with $x_i^+, x_i^- \geq 0$ for $i = 2$. To transform $3x_1 - x_2 + x_4 \geq 8$ to an equality-constraint problem, we further introduce the slack variable $y_1 \geq 0$. So we have the following problem which is equivalent to (LP):

$$\begin{aligned} \min_x \; & -4x_1 + 3x_2^+ - 3x_2^- \\ \text{subject to} \quad & -x_1 + x_2^+ - x_2^- - x_3 = 1, \\ & 3x_1 - x_2^+ + x_2^- + x_4 - y_1 = 8, \\ & x_1, x_2^+, x_2^-, x_3, x_4, y_1 \geq 0, \end{aligned}$$

or

$$\min_x c^\top x \text{ subject to } \; Ax = b, \; x \geq 0$$

with $x = (x_1, x_2^+, x_2^-, x_3, x_4, y_1)^\top$, $A = \begin{pmatrix} -1 & 1 & -1 & -1 & 0 & 0 \\ 3 & -1 & 1 & 0 & 1 & -1 \end{pmatrix}$, $b = (1, 8)^\top$, $c = (-4, 3, -3, 0, 0, 0)^\top$.

2. The dual problem now reads

$$b^\top u \to \max \text{ subject to } \; A^\top u \leq c.$$

Therefore, we have for the dual problem:

$$\begin{aligned} & u_1 + 8u_2 \to \max \\ \text{subject to} \quad & -u_1 + 3u_2 \le -4, \\ & u_1 - u_2 \le 3, \\ & -u_1 + u_2 \le -3 \\ & -u_1 \le 0 \\ & u_2 \le 0 \\ & -u_2 \le 0 \\ & u_1, u_2 \in \mathbb{R}, \end{aligned}$$

or, equivalently

$$\begin{aligned} & u_1 + 8u_2 \to \max \\ \text{s.t.} \quad & -u_1 + 3u_2 \le -4, \\ & u_1 - u_2 = 3, \\ & u_1 \ge 0 \\ & u_2 = 0. \end{aligned}$$

Because $u_2 = 0$ and $u_1 - u_2 = 3$, we get $u_1 = 3$. However, this point $\bar{u} := (3, 0)^\top$ does not satisfy the first constraint. Therefore, the set of feasible solutions of the dual problem is empty.

3. Since the dual problem does not have any feasible points, the primal problem does not have a minimizer, i.e., the finite problem is either unbounded or does not have any feasible points (here: it is unbounded).

Grading remark: This question was answered sufficiently by almost everyone. Note that the infeasibility of the dual problem does not necessarily mean that the primal problem is unbounded (it does in this example), but it can also mean that the primal problem is infeasible. 10 points for part 1., 8 points for 2., and 2 points for 3.

Problem 4 **(Constrained Optimization)** [15 points] Show that any constrained optimization problem

$$\min_x z(x) \quad \text{s. t. } f(x) \le 0$$

with $z : \mathbb{R}^n \to \mathbb{R}$ and $f \colon \mathbb{R}^n \to \mathbb{R}^m$ can be reformulated as an optimization problem with a linear objective $\tilde{z}(\tilde{x}) = \tilde{c}^T \tilde{x}$. (Hint: Note that we use the notation "$\tilde{x}$" here. That means that the arguments $\tilde{x}$ do not have to have the same meaning and/or dimension as x.)

Solution: Let $x \in \mathbb{R}^n$ and $f\colon \mathbb{R}^n \to \mathbb{R}^m$. The idea is to introduce a new variable $\tilde{x}_{n+1} := z(x)$. Then the new variable vector reads

$$\tilde{x} = \begin{pmatrix} x \\ z(x) \end{pmatrix} = \begin{pmatrix} x \\ \tilde{x}_{n+1} \end{pmatrix}.$$

This relation is added to the constraints of the problem:

$$\tilde{x}_{n+1} \leq z(x)\,, \quad -\tilde{x}_{n+1} \leq -z(x)\,.$$

The minimization of z is now equivalent to minimizing the component $\tilde{x}_{n+1}$. The ingredients of the problem therefore read

$$\tilde{x} := \begin{pmatrix} x \\ \tilde{x}_{n+1} \end{pmatrix}, \quad \tilde{z}(\tilde{x}) = \tilde{c}^T\tilde{x}\,, \quad \tilde{c} := \begin{pmatrix} 0 \\ \pm 1 \end{pmatrix}, \quad \tilde{f} := \begin{pmatrix} f(\tilde{x}_{1:n}) \\ z(\tilde{x}_{1:n}) \\ -z(\tilde{x}_{1:n}) \end{pmatrix},$$

where the relation in $\tilde{c}$ needs to be chosen depending on whether $\tilde{z}$ is to be minimized or maximized.

Grading remark: Some students suggested to replace the objective function by another (scalar) variable t and add the constraint $z(x) \leq t$ (for all x) to the constraints. This is also correct; however, the question asks for a specific $\tilde{c}$.

Problem 5 **(Vector Optimization)** [25 points]

a) We consider a bi-objective optimization problem where the image set $S \subset \mathbb{R}^2$ is closed and bounded. Your task is to illustrate the difference between optimality w.r.t. different cones, in particular with respect to

- the natural ordering by means of $C_1 = \mathbb{R}^2_+ := \{x \in \mathbb{R}^2 | x_i \geq 0,\ i = 1, 2\}$, and
- a closed halfspace $C_2 := \{x \in \mathbb{R}^2 | l^T x \geq 0\}$ for some $l \in \mathbb{R}^2$, $l \neq 0$.

Choose a set S and a vector l such that

1. the set of minimal elements with respect to C_1 is non-connected[1].
2. the set of minimal elements with respect to C_2 is nonempty, non-connected, and does not consist only of isolated points.

[1] A *connected set* is a set which cannot be partitioned into two nonempty subsets such that each subset has no points in common with the set closure of the other.

Sketch the chosen set $S \subset \mathbb{R}^2$ for 1. and 2. and highlight the set of minimal elements with respect to C_1 and C_2 in your sketch (the set S does not need to be the same in 1. and 2., but it can be). *Remark:* You do not need to specify an algebraic expression that defines S or l; a clear sketch is sufficient. In your sketch, the sets of minimal elements should be clearly marked.

b) Let $X = \mathbb{R}^2$, S a nonempty subset of X, and $C = \mathbb{R}^2_+$. Show that if there exists a weight vector $w \in \mathbb{R}^2_+$, $w \neq (0,0)^T$, with

$$\forall x = (x_1, x_2)^T \in S \setminus \{x^*\} : \quad w_1 x_1^* + w_2 x_2^* < w_1 x_1 + w_2 x_2, \tag{1}$$

then x^* is a minimal element in S w.r.t. C.

c) Let $X = \mathbb{R}^2$, $C = \mathbb{R}^2_+$. Give an example (either by a drawing or numerically) of a set S where minimal elements of S w.r.t. C cannot be found by means of (1).

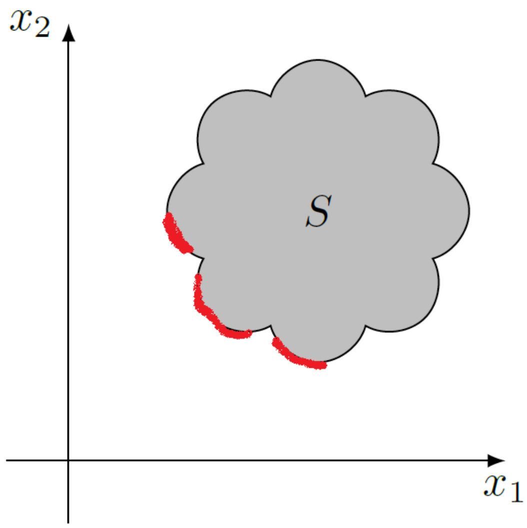
d) In the lecture, we have shown the following lemma:
Let S be a nonempty subset of a partially ordered linear space with an ordering cone C. Then every minimal element of the set S is also a minimal element of the set $S + C$.
Let $X = \mathbb{R}^2$, $C \subset \mathbb{R}^2$. Give an example (either by a drawing or numerically) of sets S and C which shows the converse assertion of the above lemma in general does not hold true.

e) In the lecture, we have shown the following theorem:
If there exists an $\ell \in C^* \setminus \{0\}$ with

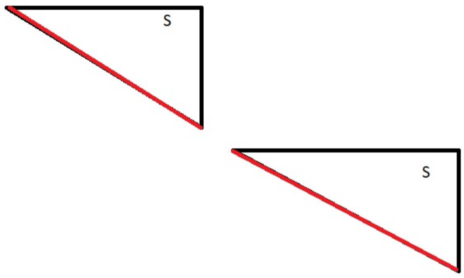
$$\forall x \in S \setminus \{x^*\} : \quad \ell(x^*) < \ell(x), \tag{2}$$

then $x^* \in S^C_{\min}$, i.e., x^* is a minimal element in S.
Now show that if there exists some $\ell \in C^\# := \{\ell \in X^* \mid \forall\, c \in C \setminus \{0\} : \ \ell(c) > 0\}$ (the so-called quasi-interior of C^*) such that the assertion (2) holds true, then $x^* \in S^C_{\min}$.

Solution:
Grading remark: Here we were mainly looking for an intuition of the problem formulation and optimality notions. So, we did not pay attention to small inaccuracies in illustrations. Moreover, the answers to the questions can be found by following the last two lectures on vector optimization. What is asked here is not really new, as problem 5 is mainly intended to check the main understanding of the concepts.
a) 1. The set of minimal elements of S w.r.t. C_1 is illustrated in red color.



2. The set of minimal elements of S w.r.t. C_2 consists of two line segments:



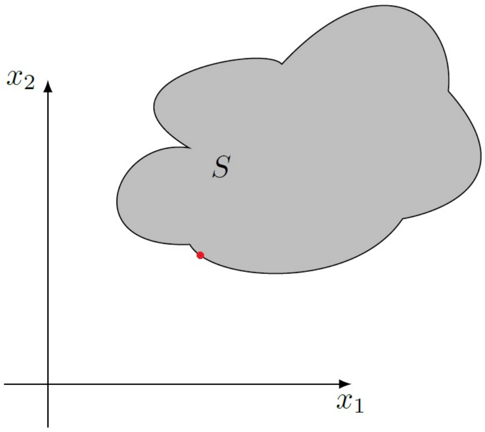
Grading remark: The set S can be the same in 1. and 2. 5 points for **a)**.
b) We use a proof by contradiction. Assume that x^* is not minimal w.r.t. C. Then

$$(\{x^*\} - C) \cap S \neq \{x^*\}, \tag{3}$$

as C is pointed. This means that there exists some $\bar{x} \in S \setminus \{x^*\}$ such that $\bar{x} \in \{x^*\} - C$. Multiplication with the weight vector w from (1) gives

$$w_1\bar{x}_1 + w_2\bar{x}_2 \leq w_1x_1^* + w_2x_2^*.$$

Using the assumption (1), we obtain therefore a contradiction.
Grading remark: The proof has been done more generally in higher dimension than $\mathbb{R}^2$ (namely in a real vector space) in the lecture. 5 points for **b)**.
c) Here, we have a set S that is nonconvex, and the illustrated red point cannot be found by



means of linear scalarization (1).

Grading remark: Some students drew a set that is unbounded, or even suggested $S = \mathbb{R}^2$, where minimal elements do not exist. This is also correct. 5 points for **c)**.

d) The converse direction of the above lemma holds true if C is pointed, see [Jahn, Lemma 4.7, p. 106]. Therefore, we can choose, for example, the cone C_2 from **a)** for any $l \neq (0,0)^T$ and S from **a)**.

Grading remark: 5 points for **d)**. Almost everyone who attempted problem 5 managed part **d)**. We noticed this property also in the lecture.

e) Assume that x^* is not minimal. Then, from

$$(\{x^*\} - C) \cap S \not\subseteq \{x^*\} + C, \tag{4}$$

there exists an $\overline{x} \in S$ with $\overline{x} \in \{x^*\} - C$, but $\overline{x} \notin \{x^*\} + C$. The first assertion implies $x^* - \overline{x} \in C$, while the second leads to $x^* - \overline{x} \neq \{0\}$, and therefore $\ell(\overline{x}) < \ell(x^*)$ for any $\ell \in C^{\#}$. This contradicts our assumption.

Grading remark: The proof is very similar to the one given in the lecture, the main difference is the use of the algebraic interior of the dual cone C^* instead of C^*. 5 points for **e)**.