Exam Optimization 1 - Spring 2021

Problem 1 **(Existence of Minimizers)** [5 points]

Let S be a nonempty, closed subset of $\mathbb{R}^n$, and let $f : S \to \mathbb{R}$ be continuous. Show that if the set $\hat{S} := \{x \in S \mid f(x) \leq f(\hat{x})\}$ is bounded for an $\hat{x} \in S$, then the optimization problem $\min_{x \in S} f(x)$ has a minimizer.

Solution: Because f is continuous, S closed and because of $\leq$ in the definition of $\hat{S} \Rightarrow \hat{S}$ closed
Therefore: $\hat{S}$ is compact. Moreover, f is continuous.
By Weierstrass, there exists $\bar{x} \in \hat{S} : f(\bar{x}) \leq f(x) \quad \forall x \in \hat{S}$.
Note that $\bar{x} \in S$.
Furthermore, $\forall x \in S \setminus \hat{S} \Rightarrow f(x) > f(\hat{x}) \geq f(\bar{x})$, and the claim follows.

Problem 2 **(Karush-Kuhn-Tucker-Conditions I)** [10 points]

Calculate all points (with according Lagrange-multipliers) which satisfy the Karush-Kuhn-Tucker-conditions for the following optimization problem:

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \quad & -x_1 - x_2 \\ \text{subject to} \quad & x_2 \geq x_1^2, \\ & x_2 \leq x_1 + 2. \end{aligned}$$

Solution: The Lagrange function reads $\mathcal{L}(x, \lambda) = -x_1 - x_2 + \lambda_1(x_1^2 - x_2) + \lambda_2(x_1 - 2 - x_2)$. The KKT conditions are

$$\nabla_x \mathcal{L}(x, \lambda) = \begin{pmatrix} -1 + 2\lambda_1 x_1 - \lambda_2 \\ -1 - \lambda_1 + \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{1}$$

$$\lambda_1(x_1^2 - x_2) = 0 \tag{2}$$

$$\lambda_2(-x_1 - 2 + x_2) = 0 \tag{3}$$

$$\lambda_1, \lambda_2 \geq 0 \tag{4}$$

$$x_2 \geq x_1^2$$

$$x_2 \leq x_1 + 2.$$

- $\lambda_1 = \lambda_2 = 0$: This contradicts (1).
- $\lambda_1 = 0$, $\lambda_2 > 0$: By (1), we get $\lambda_2 = -1$, a contradiction.
- $\lambda_2 = 0$, $\lambda_1 > 0$: By the second line in (1), we get $\lambda_1 = -1$, a contradiction.
- $\lambda_1 > 0$, $\lambda_2 > 0$: (3) yields $-x_1 - 2 + x_2 = 0$, thus $x_1 = x_2 - 2$. Together with (2), we get $x_2 = x_1^2 = (x_2 - 2)^2$ and thus $x_2 = 4$ and $x_2 = 2$. If $x_2 = 2$, (2) implies that $x_1 = 1$, but this contradicts (3). If $x - 2 = 4$, we get from (3) that $x_1 = -2$. From the first line in (1) we get that $\lambda_1 = \frac{1}{3}$, and from the second line in (1), it follows that $\lambda_2 = \frac{4}{3}$. So, the pair (x, λ) with $x = (-2, 4)^\top$ and $\lambda = (\frac{1}{3}, \frac{4}{3})^\top$ is the only point that satisfies the KKT conditions.

Problem 3 (Karush-Kuhn-Tucker-Conditions II) [15 points] Here, we randomize with $c = 1, 2, 3, 4, 5$.

The point $x = (0, 0)^\top \in \mathbb{R}^2$ is a global minimizer of the problem

$$\begin{aligned} \min_{x=(x_1,x_2)^\top \in \mathbb{R}^2} \quad & x_1 \\ \text{subject to } & cx_2 - x_1^3 \leq 0 \\ & -cx_2 - x_1^3 \leq 0. \end{aligned} \tag{5}$$

a) Show that the Karush-Kuhn-Tucker conditions of problem (5) have no solution.

b) Explain why the first-order necessary optimality conditions for constrained optimization problems are not applicable for $x = (0, 0)^\top \in \mathbb{R}^2$.

Solution:
(a) The Lagrange function is $\mathcal{L}(x, \lambda) = x_1 + \lambda_1(cx_2 - x_1^3) + \lambda_2(-cx_2 - x_1^3)$. We get

$$\nabla_x \mathcal{L}(x, \lambda) = \begin{pmatrix} 1 - 3c\lambda_1 x_1^2 - 3c\lambda_2 x_1^2 \\ c\lambda_1 - c\lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{6}$$

$$\lambda_1(cx_2 - x_1^3) = 0 \tag{7}$$

$$\lambda_2(-cx_2 - x_1^3) = 0 \tag{8}$$

$$\lambda_1, \lambda_2 \geq 0 \tag{9}$$

$$x_2 \geq x_1^2$$

$$x_2 \leq x_1 + 2$$

The second line in (6) implies $\lambda_1 = \lambda_2 \overset{(9)}{\geq} 0$. If $\lambda_1 = \lambda_2 = 0$, then the first line in (6) does not hold. If $\lambda_1 = \lambda_2 > 0$, then we have by (7) that $cx_2 - x_1^3 = 0$ and therefore $cx_2 = x_1^3$. By (8),

we get $-cx_2 - x_1^3$ and thus $-cx_2 = x_1^3$. Together, $x_1 = x_2 = 0$, in contradiction to the first line in (6). Therefore, the KKT conditions have no solution.
(b) Note that the linear independence constraint qualification (LICQ) does not hold:

$$\nabla c_1(x) = \begin{pmatrix} -3x_1^2 \\ c \end{pmatrix}\Bigg|_{x=(0,0)^\top} = \begin{pmatrix} 0 \\ c \end{pmatrix}$$
$$\nabla c_2(x) = \begin{pmatrix} -3x_1^2 \\ -c \end{pmatrix}\Bigg|_{x=(0,0)^\top} = \begin{pmatrix} 0 \\ -c \end{pmatrix}.$$

$\left\{(0,c)^\top, (0,-c)^\top\right\}$ is **not** linearly independent. Therefore, the first-order necessary optimality conditions for constrained optimization problems are not applicable here.

Problem 4 **(Linear Programs - Duality)** [15 points]

Consider the following linear program

$$\begin{array}{llrrrrl} \text{(LP)} & \min\limits_{x\in\mathbb{R}^4} & -2x_1 & +3x_2 & & & \\ & \text{subject to} & -x_1 & +x_2 & -x_3 & & = 1, \\ & & 3x_1 & -x_2 & & +x_4 & \geq 8, \\ & & x_1 & & & & \geq 0, \\ & & & & x_3 & & \geq 0. \end{array}$$

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

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

2. Write down the corresponding dual program of the (LP3) and compute a solution of the dual program.

3. Use a duality result to compute all minimizers of (LP).

Solution:

1. We formulate (LP) as (LP3). Because there are no nonnegativity constraints on $x_2, x_4 \in \mathbb{R}$, we introduce the following variables for $i = 2, 4$:

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

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

This means $x_i = x_i^+ - x_i^-$ with $x_i^+, x_i^- \ge 0$ for $i = 2, 4$. So we have the following problem which is equivalent to (LP):

$$\begin{aligned} \min_x -2x_1 + 3x_2^+ - 3x_2^- \\ \text{subject to} \quad -x_1 + x_2^+ - x_2^- - x_3 = 1, \\ 3x_1 - x_2^+ + x_2^- + x_4^+ - x_4^- - y_1 = 8, \\ x_1, x_2^+, x_2^-, x_3, x_4^+, x_4^-, y_1 \ge 0, \end{aligned}$$

or

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

with $x = (x_1, x_2^+, x_2^-, x_3, x_4^+, x_4^-, y_1)^\top$, $A = \begin{pmatrix} -1 & 1 & -1 & -1 & 0 & 0 & 0 \\ 3 & -1 & 1 & 0 & 1 & -1 & -1 \end{pmatrix}$, $b = (1, 8)^\top$, $c = (-2, 3, -3, 0, 0, 0, 0)^\top$.

2. The dual problem now reads

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

Therefore, we have for the dual problem:

$$\begin{aligned} u_1 + 8u_2 &\to \max \\ \text{subject to} \quad -u_1 + 3u_2 &\le -2, \\ u_1 - u_2 &\le 3, \\ -u_1 + u_2 &\le -3 \\ -u_1 &\le 0 \\ u_2 &\le 0 \\ -u_2 &\le 0 \\ -u_2 &\le 0 \\ u_1, u_2 &\in \mathbb{R}, \end{aligned}$$

or, equivalently

$$\begin{aligned} u_1 + 8u_2 &\to \max \\ \text{bei} \quad -u_1 + 3u_2 &\le -2, \\ u_1 - u_2 &= 3, \\ u_1 &\ge 0 \\ u_2 &= 0. \end{aligned}$$

Because $u_2 = 0$ and $u_1 - u_2 = 3$, we get $u_1 = 3$. Setting $\bar{u} := (3, 0)^\top$. $\bar{u}$ satisfies all constraints and $\{\bar{u}\}$ is therefore the set of all feasible points of the corresponding dual problem of (LP). Therefore, $\bar{u}$ is the optimal solution of the dual problem, with optimal function value 3.

3. By our duality results, we get for the minimizer of (LP) that $-2x_1 + 3x_2 = 3$. Together with the constraints we get the system

$$-2x_1 + 3x_2 = 3, \tag{10}$$
$$-x_1 + x_2 - x_3 = 1, \tag{11}$$
$$3x_1 - x_2 + x_4 \geq 8, \tag{12}$$
$$x_1, x_3 \geq 0. \tag{13}$$

(10) implies $x_2 = \frac{2}{3}x_1 + 1$. Plugging this into (11) yields $-x_1 + \frac{2}{3}x_1 + 1 - x_3 = 1$. Therefore,

$$x_3 = -\frac{1}{3}x_1 \tag{14}$$

Putting this into (12) yields $3x_1 - \frac{2}{3}x_1 - 1 + x_4 \geq 8$, leading to

$$x_4 \geq -\frac{7}{3}x_1 + 9 \tag{15}$$

Because of (14) and $x_1, x_3 \geq 0$, we have $x_1 = x_3 = 0$.

Putting $x_1 = 0$ in (15) implies $x_4 \geq 9$.

Setting $x_1 = x_3 = 0$ in (11) yields $x_2 = 1$. Therefore, we can conclude that $\bar{x}$ is a minimizer for (LP) if and only if $\bar{x}$ has the form $\bar{x} = (0, 1, 0, x_4)^\top$ with $x_4 \geq 9$.

Problem 5 (Linear Programs) [10 points]

Consider the linear optimization problem:

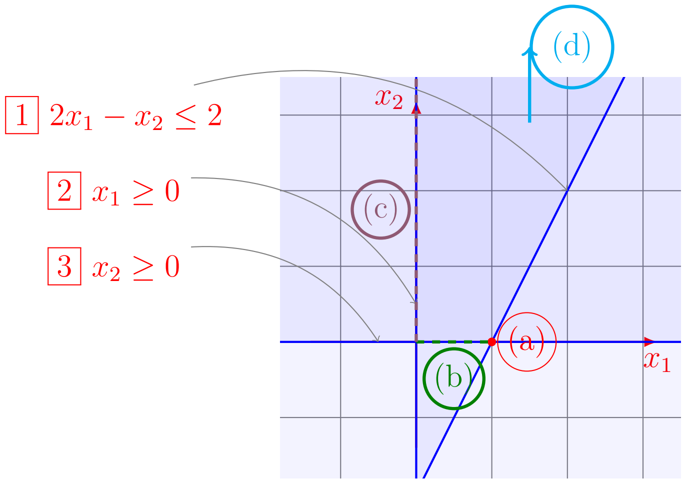
$$\begin{array}{lll} \min\limits_{x=(x_1,x_2)^\top} & -c_1x_1 - c_2x_2 & \\ \text{subject to} & 2x_1 - x_2 & \leq 2 \\ & x_1, x_2 & \geq 0 \end{array}$$

Determine coefficients (c_1, c_2) of the objective function such that

1. the problem has a unique minimizer.

2. the problem has multiple minimizers and the set of minimizers is bounded.
3. the problem has multiple minimizers and the set of minimizers is unbounded.
4. the problem has no minimizer.

Solution:



(a) $c = (1,\ -1)^\top$ opt value = 1 being attained at $(1, 0)^\top$

(b) $c = (0,\ -1)^\top$ opt value = 0, set is line segment from $(0, 0)^\top$ to $(1, 0)^\top$

(c) $c = (-1,\ 0)^\top$ opt value = 0, set goes along the x_2-axis with $x_2 \geq 0$

(d) $c = (0,\ 1)^\top$

Problem 6 (Steepest Descent) [15 points]
Consider the function $f : \mathbb{R}^n \to \mathbb{R}$,

$$f(x) = x^\top H x - b^\top x$$

with a symmetric positive definite matrix $H \in \mathbb{R}^{n,n}$. Let $x^0 - \tilde{x}$ be an eigenvector of H. Show that the iteration $x^{k+1} = x^k + \alpha_k p^k$ with the search direction $p^k = -\nabla f(x^k)$ finds the minimizer $\tilde{x}$ in one step. Explain how α_k needs to be chosen for this result to hold.
Solution: We have $\nabla f = 2Hx - b$, and so, for a minimizer we get $\nabla f = 2H\tilde{x} - b = 0$, such that $\tilde{x} = \frac{1}{2}H^{-1}b$ (H^{-1} exists because H is a symmetric positive definite matrix). We get for

the search direction: $p^k = -\nabla f(x^k) = -2Hx^k + b$, and so, $p^0 = b - 2Hx^0$. We want to show that $\tilde{x} = x^0 + \alpha p^0$, which is equivalent to

$$x^0 - \tilde{x} = -\alpha p^0. \tag{16}$$

We know: $-p^0 = 2Hx^0 - b = 2Hx^0 - 2H\tilde{x} = 2H(x^0 - \tilde{x})$. With this, our claim (16) is equivalent to

$$(I - 2\alpha_k H)(x^0 - \tilde{x}) = 0. \tag{17}$$

Since $x^0 - \tilde{x}$ is an eigenvector of H, we have

$$(x^0 - \tilde{x}) \cdot \underbrace{\lambda}_{\text{eigenvalue}} = H(x^0 - \tilde{x}),$$

or, equivalently,

$$(x^0 - \tilde{x}) \cdot (\lambda I - H) = 0.$$

Comparing this with (17), we see that $\alpha_k = \frac{1}{2\lambda}$ yields the desired result. Note that $\lambda > 0$, because H is symmetric positive definite, and so, α_k is well-defined.

Problem 7 (Wolfe Conditions) [15 points]
In the lecture, we have proven the following results: Let the function $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable, p^k be a descent direction at x^k , and assume that f is bounded below along the ray $\{x^k + \alpha p^k | \alpha > 0\}$. Then, if $0 < c_1 < c_2 < 1$, there exist intervals of step lengths satisfying the Wolfe conditions. Show that such intervals may no longer exist if $0 < c_2 < c_1 < 1$ by providing and discussing an explicit counterexample for a specific function f.

Solution: The Wolfe conditions are:

$$\Phi(\alpha) \leq \underbrace{f(x^k + c_1\alpha \nabla f^{k^\top} p^k)}_{\text{sufficient decrease}} \tag{18}$$

$$\underbrace{\Phi'(\alpha) \leq c_2 \Phi'(0)}_{\text{curvature condition}} \tag{19}$$

For example, let $f(x) = x^2$, $x^k = x = -1$, $p^k = p = 1$, $c_1 = 0.98$, $c_2 = 0.02$. Then $\Phi(\alpha) = f(x+\alpha p) = (x+\alpha p)^2 = x^2 + 2\alpha x p + \alpha^2 p^2 = \alpha^2 - 2\alpha + 1$. We also have $\Phi'(\alpha) = 2\alpha - 2$. The sufficient decrease condition (18) yields

$$\Phi(\alpha) \leq f(x) + c_1 \alpha f'(x) p = 1 - 2c_1\alpha.$$

Thus,

$$\Phi(\alpha) = \alpha^2 - 2\alpha - 1 \leq 1 - 1.98\alpha,$$

leading to $\alpha^2 - 0.02\alpha \leq 0$. This holds true only for $\alpha \in [0, 0.02]$. The curvature condition (19) leads to

$$\Phi'(\alpha) = 2\alpha - 2 \geq c_2 \cdot \Phi'(0) = -0.02,$$

which gives $\alpha \geq 0.99$, a contradiction.

Problem 8 (Trust-Region Method) [15 points]
Let $x = (x_1, x_2)^\top$,

$$f(x) = x_1^4 + 3x_1^2 + 4x_2^2.$$

Compute, explicitly and in detail, one step for the **trust region method** (that is, until you reach x^1) with the model function $m(p) = f(x^k) + {g^k}^\top p + \frac{1}{2} p^\top H_k p$, with $g^k = \nabla f(x^k)$, $H_k = \nabla^2 f(x^k)$, trust-region radius $\Delta = 1$, parameters $\eta_1 = 0.2$, $\eta_2 = 0.9$, $\sigma_1 = 0.5$, $\sigma_2 = 2$ and starting point

$$x^0 = (-2, 0)^\top.$$

Solution: It holds

$$\nabla f(x) = \begin{pmatrix} 4x_1^3 + 6x_1 \\ 8x_2, \end{pmatrix}$$

and

$$\nabla^2 f(x) = \begin{pmatrix} 12x_1^2 + 6 & 0 \\ 0 & 8 \end{pmatrix}.$$

Note that $\nabla^2 f(x)$ is symmetric and positive definite for all x:

$$\begin{aligned} x^\top \nabla^2 f(x) x &= (x_1, x_2) \begin{pmatrix} 12x_1^2 + 6 & 0 \\ 0 & 8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= (12x_1^2 + 6)x_1^2 + 8x_2^2 > 0 \text{ for all } x \in \mathbb{R}^2 \end{aligned}$$

(or, because the matrix is symmetric and the eigenvalues are the elements in the diagonal, it can be seen that the eigenvalues are positive for all $x \in \mathbb{R}^2$, and thus the matrix is symmetric positive definite). We have

$$g^0 = \nabla f(x^0) = \begin{pmatrix} 4 \cdot (-2)^3 - 6 \cdot 2 \\ 0 \end{pmatrix} = \begin{pmatrix} -44 \\ 0 \end{pmatrix}$$

and

$$H_0 = \nabla^2 f(x^0) = \begin{pmatrix} 12 \cdot (-2)^2 + 6 & 0 \\ 0 & 8 \end{pmatrix} = \begin{pmatrix} 54 & 0 \\ 0 & 8 \end{pmatrix}.$$

We get

$$H_0^{-1} = \begin{pmatrix} \frac{1}{54} & 0 \\ 0 & \frac{1}{8} \end{pmatrix}.$$

Furthermore,

$$||H_0^{-1}g^0|| = ||\begin{pmatrix} \frac{1}{54} & 0 \\ 0 & \frac{1}{8} \end{pmatrix} \cdot \begin{pmatrix} -44 \\ 0 \end{pmatrix}|| = ||\begin{pmatrix} -\frac{44}{54} \\ 0 \end{pmatrix}|| = ||\begin{pmatrix} -\frac{22}{27} \\ 0 \end{pmatrix}|| = \sqrt{\left(\frac{-22}{27}\right)^2} = \frac{22}{27} < 1 = \Delta.$$

Therefore, $p^0 = -H_0^{-1}g^0 = \begin{pmatrix} \frac{44}{54} \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{22}{27} \\ 0 \end{pmatrix}$. So we get $\hat{x}^0 = x^0 + p^0 = \begin{pmatrix} -2 \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{22}{27} \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{32}{27} \\ 0 \end{pmatrix}$. We compute the decrease:

$$\delta_0 = \frac{f(x^0) - f(\hat{x}^0)}{\underbrace{m(0)}_{=f(x^0)} - \underbrace{m(p^0)}_{=28+{g^0}^\top p^0 + \frac{1}{2}{p^0}^\top H_0 p^0 \approx -7.84}} \approx \frac{28 - 6.187}{28 + 7.84} \approx 0.6085.$$

We have $\delta_0 \approx 0.6085 \in [\underbrace{\eta_1}_{=0.2}, \underbrace{\eta_2}_{=0.9}]$. Therefore, we choose $\Delta_1 \in [\underbrace{\sigma_1}_{=0.5} \cdot \Delta, \Delta]$, which means that the trust-region radius will be shrinked. Moreover, $\delta_0 > \eta_1 = 0.2$, and so, $x^1 := \hat{x}^0$ is the new iteration point.