MAT 3007 — Optimization

Solutions 7

Problem 1 (Convex Sets): (approx. *25* points)

In this exercise, we study convexity of various sets.

a) Verify whether the following sets are convex or not and explain your answer!

$$\Omega_1 = \{x \in \mathbb{R}^n : \|x-a\|_2 \le \|x-b\|_2\}, \quad a, b \in \mathbb{R}^n,\ a \neq b,$$
$$\Omega_2 = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : x^\top x \le t\}.$$

b) Show that the hyperbolic set $\{x \in \mathbb{R}^2_+ : x_1x_2 \ge 1\}$ is convex, where $\mathbb{R}^2_+ = \{x \in \mathbb{R}^2 : x \ge 0\}$.

Hint: Rewrite the condition "$x_1x_2 \ge 1$" in a suitable way.

c) Decide whether the following statements are true or false. Explain your answer and either present a proof / verification or a counter-example.

- The union of two non-convex sets $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ is always a non-convex set.
- Let $\Omega \subset \mathbb{R}^n$ be a convex set and suppose that the set $S := \{(x,t) \in \Omega \times \mathbb{R} : f(x) \le t\} \subset \mathbb{R}^n \times \mathbb{R}$ is convex. Then, $f : \Omega \to \mathbb{R}$ is a convex function.

Solution :

a) The condition $\|x-a\| \le \|x-b\|$ in the definition of Ω_1 is equivalent to

$$\|x-a\|^2 \le \|x-b\|^2 \iff -2a^\top x + \|a\|^2 \le -2b^\top x + \|b\|^2 \iff 2(b-a)^\top x \le \|b\|^2 - \|a\|^2.$$

Hence, Ω_1 is a convex half space.

The set Ω_2 is convex. To see this, let us set $g(x,t) := x^\top x - t = \|x\|^2 - t$. Then, it holds that

$$\nabla g(x,t) = \begin{pmatrix} 2x \\ -1 \end{pmatrix} \quad \text{and} \quad \nabla^2 g(x,t) = \begin{pmatrix} 2I & 0 \\ 0 & 0 \end{pmatrix}.$$

The Hessian $\nabla^2 g$ is obviously positive semidefinite for all $(x,t) \in \mathbb{R}^n \times \mathbb{R}$. Hence, g is convex and Ω_2 defines a convex level set.

b) We first notice $\Omega := \{x \in \mathbb{R}^2_+ : x_1x_2 \ge 1\} = \{x \in \mathbb{R}^2_{++} : x_1x_2 \ge 1\}$, where $\mathbb{R}^2_{++} = \{x \in \mathbb{R}^2 : x > 0\}$. Since the (natural) logarithm $\log(\cdot)$ is monotonically increasing, the condition $x_1x_2 \ge 1$ is equivalent to

$$f(x_1,x_2) := -\log(x_1) - \log(x_2) \le 0.$$

The gradient and Hessian of f are given by:

$$\nabla f(x_1,x_2) = \begin{pmatrix} -x_1^{-1} \\ -x_2^{-1} \end{pmatrix}, \quad \nabla^2 f(x_1,x_2) = \begin{pmatrix} x_1^{-2} & 0 \\ 0 & x_2^{-2} \end{pmatrix}$$

Since $\nabla^2 f$ is positive definite on the open and convex set $\mathbb{R}^2_{++}$, it follows that f is convex on $\mathbb{R}^2_{++}$. This shows that the set Ω is convex.

c) The first statement is wrong: we can define $\Omega_1 := \{-1, 0, 1\}$ and $\Omega_2 := (-1, 0) \cup (0, 1)$. Both Ω_1 and Ω_2 are obviously not convex sets. However, $\Omega_1 \cup \Omega_2 = [-1, 1]$ is a convex interval.

   We verify the second statement briefly. Let $x, y \in \Omega$ and $\lambda \in [0, 1]$ be arbitrary. Then, we have $(x, f(x)), (y, f(y)) \in S$. By the convexity of S, we can infer $\lambda(x, f(x)) + (1 - \lambda)(y, f(y)) = (\lambda x + (1 - \lambda)y, \lambda f(x) + (1 - \lambda)f(y)) \in S$. However, by definition of S, this means

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

   Since $x, y \in \Omega$ and $\lambda \in [0, 1]$ are arbitrary, this implies that f is convex on Ω.

Problem 2 (Convex Compositions): (approx. *25* points)
Either prove or find a counterexample for each of the following statements (you can assume that all functions are twice continuously differentiable if needed):

a) If $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ are convex, then the composition $f \circ g : \mathbb{R}^n \to \mathbb{R}$, $(f \circ g)(x) = f(g(x))$ is convex.

b) Let $\Omega \subset \mathbb{R}^n$ be a convex set and suppose that $g : \Omega \to \mathbb{R}$ is convex and $f : I \to \mathbb{R}$ is convex and nondecreasing where $I \supseteq g(\Omega)$ is an interval containing $g(\Omega)$. Then, $f \circ g$ is convex.

c) If $f : \mathbb{R} \to \mathbb{R}$ is increasing and non-negative, then $x \mapsto xf(x)$ is a convex function on $\mathbb{R}_+$.

Solution :

a) *False.* Set $g(x) = x^2$ and $f(x) = -x$. Both functions are obviously convex, but $f(g(x)) = -x^2$ is a concave function.

b) *True.* We prove this result by using the basic definition of convexity. Let $x, y \in \Omega$ and $\lambda \in [0, 1]$ be arbitrary. Using the convexity of g, we have $g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)$. Since the interval I is convex and we have $g(\Omega) \subseteq I$, it follows $\lambda g(x) + (1 - \lambda)g(y) \in I$. Moreover, since f is nondecreasing, we have

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

   where we used the convexity of f in the last step. This shows that $f \circ g$ is convex.

c) *False.* Consider the following counterexample

$$f(x) = \begin{cases} 2x & x \in [0, 1], \\ \frac{1}{2}x + \frac{3}{2} & x > 1, \end{cases} \quad \text{and} \quad xf(x) = \begin{cases} 2x^2 & x \in [0, 1], \\ \frac{1}{2}x^2 + \frac{3}{2}x & x > 1. \end{cases}$$

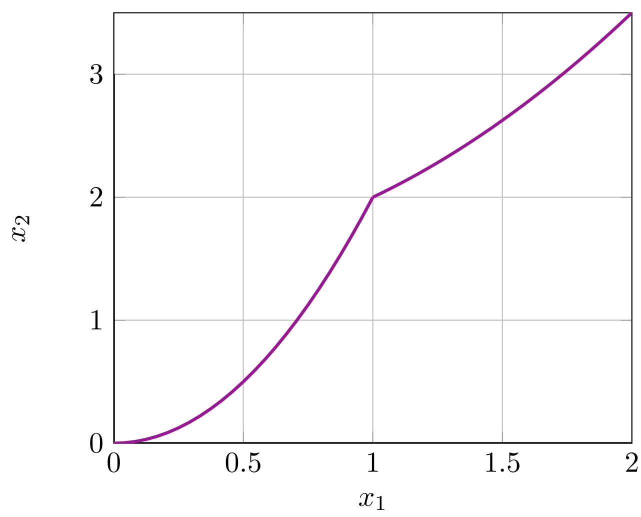
   From the sketch, we can see that the mapping $x \mapsto xf(x)$ can not be convex. Note that in this example $xf(x)$ is not differentiable at 1. However, it is possible to come up with a function which is smooth at 1 but still keeps the shape and characteristics of this example.

Problem 3 (Convex Functions): (approx. *25* points)
In this exercise, convexity properties of different functions are investigated.

a) Verify that the following functions are convex over the specified domain:

   – $f : \mathbb{R}_{++} \to \mathbb{R}$, $f(x) := \sqrt{1 + x^{-2}}$, where $\mathbb{R}_{++} := \{x \in \mathbb{R} : x > 0\}$.

   – $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) := \frac{1}{2}\|Ax - b\|^2 + \mu\|Lx\|_1$, where $A \in \mathbb{R}^{m \times n}$, $L \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^m$, and $\mu > 0$ are given and $\|y\|_1 := \sum_{i=1}^{p} |y_i|$, $y \in \mathbb{R}^p$.



– $f : \mathbb{R}^{n+1} \to \mathbb{R}$, $f(x, y) := \frac{\lambda}{2}\|x\|^2 + \sum_{i=1}^m \ln(1 + \exp(-b_i(a_i^\top x + y)))$, where $a_i \in \mathbb{R}^n$ and $b_i \in \{-1, 1\}$ are given data points for all $i = 1, \ldots, m$ and $\lambda > 0$ is a parameter.

b) Let $r(x) := \|x\|_q = (\sum_{i=1}^n |x_i|^q)^{1/q}$ be the ℓ_q-norm on $\mathbb{R}^n$ with $q \in [1, \infty)$. Show that r is a convex function.

Hint: As a norm, the mapping r satisfies certain properties that might be useful here.

Solution :

a) The mapping $f(x) = \sqrt{1 + x^{-2}}$ is twice continuously differentiable on the convex set $\mathbb{R}_{++}$ and it holds that

$$f'(x) = \frac{1}{2\sqrt{1 + x^{-2}}} \frac{-2}{x^3} = \frac{-1}{x^2\sqrt{1 + x^2}}, \quad f''(x) = \frac{2x\sqrt{1 + x^2} + \frac{x^3}{\sqrt{1+x^2}}}{x^4(1 + x^2)} = \frac{3x^2 + 2}{x^3(1 + x^2)\sqrt{1 + x^2}}.$$

Notice that we have $f''(x) > 0$ for all $x > 0$ and thus, f is convex on $\mathbb{R}_{++}$.

We now study $f(x) := \frac{1}{2}\|Ax - b\|^2 + \mu\|Lx\|_1$. Since $\frac{1}{2}\| \cdot \|^2$ is convex (the Hessian is the identity matrix), we know that $x \mapsto \frac{1}{2}\|Ax - b\|^2$ is convex as a composition of a linear and convex function. The ℓ_1-norm satisfies $\|\lambda x + (1 - \lambda)y\|_1 \le \lambda\|x\|_1 + (1 - \lambda)\|y\|_1$ for all $\lambda \in [0, 1]$ and $x, y \in \mathbb{R}^p$ (this is a simple consequence of the triangle inequality). Thus, $x \mapsto \|x\|_1$ is convex. Again $x \mapsto \|Lx\|_1$ is then a convex function. Together, this implies that f is convex.

Finally, let us define $g(x, y) = \frac{\lambda}{2}\|x\|^2$ and $g_i(x, y) = \ln(1 + \exp(-b_i(a_i^\top x + y)))$. Then, f can be interpreted as the sum of the functions g and g_i, $i = 1, ..., m$ and convexity follows if each of the functions g, g_i, $i = 1, ..., m$ is convex. The mapping g_i is the composition of the function $z \mapsto \gamma(z) = \ln(1 + \exp(z))$ and the affine-linear function $(x, y) \mapsto h_i(x, y) := -b_i(a_i^\top x + y)$. Since h_i is convex (as a linear mapping), the function g_i is convex if the mapping γ is convex. It holds that

$$\gamma'(z) = \frac{e^z}{1 + e^z} = 1 - \frac{1}{1 + e^z} \quad \text{and} \quad \gamma''(z) = \frac{e^z}{(1 + e^z)^2} > 0.$$

Hence, the functions g_i are all convex. Finally, the Hessian of g is given by

$$\mathbb{R}^{(n+1)\times(n+1)} \ni \nabla^2 g(x, y) = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \succeq 0.$$

This establishes convexity of f.

b) Let $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$ be given. Then the triangle inequality and the positive homogeneity of the norm function r imply

$$r(\lambda x + (1-\lambda)y) \le r(\lambda x) + r((1-\lambda)y) = |\lambda| r(x) + |1-\lambda| r(y) = \lambda r(x) + (1-\lambda) r(y).$$

This shows that r is convex.

Problem 4 (Weighted Least-Squares Problem): (approx. *25* points)
We consider the following least squares-type problem with variable weights:

$$\min_{x,w} \sum_{i=1}^{m} \frac{(a_i^\top x - b_i)^2}{1+w_i} + \delta^2 \cdot \mathbb{1}^\top w \quad \text{s.t.} \quad w \ge 0, \tag{1}$$

where $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$, $i = 1, \ldots, m$ are given data points, $\delta > 0$ is a parameter, and $\mathbb{1} \in \mathbb{R}^m$ is the vector of all-ones.

a) Show that problem (1) is a convex optimization problem.

b) Show that the optimization problem (1) can be simplified to the equivalent problem:

$$\min_{x} \sum_{i=1}^{m} \varphi_\delta(a_i^\top x - b_i) \quad \text{where} \quad \varphi_\delta(y) := \begin{cases} y^2 & \text{if } |y| \le \delta \\ \delta(2|y| - \delta) & \text{if } |y| > \delta. \end{cases} \tag{2}$$

Hint: Optimize over w in (1) assuming that x is fixed and establish a connection to problem (2). Explain your steps and derivations!

Solution :

a) The constraints are linear and defines a convex feasible set. The mapping $(x,w) \mapsto \delta^2 \cdot \mathbb{1}^\top w$ is linear and thus convex. We need to verify convexity of the first term appearing in the objective function of (1). Let us define $\gamma : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$, $\gamma(u,v) := u^2/(1+v)$. Then, we can write

$$\sum_{i=1}^{m} \frac{(a_i^\top x - b_i)^2}{1+w_i} = \sum_{i=1}^{m} \gamma(a_i^\top x - b_i, e_i^\top w),$$

where e_i is the i-th unit vector. If γ is convex, then $(x,w) \mapsto \gamma(a_i^\top x - b_i, e_i^\top w)$ is convex for all i — as the composition of a convex and linear mapping. This then establishes convexity of the problem (1). We compute the derivatives of γ:

$$\nabla \gamma(u,v) = \begin{pmatrix} \frac{2u}{1+v} \\ -\frac{u^2}{(1+v)^2} \end{pmatrix}, \quad \nabla^2 \gamma(u,v) = \begin{pmatrix} \frac{2}{1+v} & -\frac{2u}{(1+v)^2} \\ -\frac{2u}{(1+v)^2} & \frac{2u^2}{(1+v)^3} \end{pmatrix}.$$

It holds that $\mathrm{tr}(\nabla^2\gamma(u,v)) = \frac{2}{1+v}(1 + \frac{u^2}{(1+v)^2}) > 0$ and $\det(\nabla^2\gamma(u,v)) = \frac{4u^2}{(1+v)^4} - \frac{4u^2}{(1+v)^4} = 0$ for all $(u,v) \in \mathbb{R} \times \mathbb{R}_+$. This implies that $\nabla^2\gamma$ is positive semidefinite on $\mathbb{R} \times \mathbb{R}_+$ and proves convexity of γ and problem (1).

b) For fixed x, problem (1) is still a convex optimization problem in w (this follows easily from our previous discussion). The KKT conditions are given by:

$$-\frac{(a_i^\top x - b_i)^2}{(1+w_i)^2} + \delta^2 - \lambda_i = 0, \quad \lambda_i \ge 0, \quad w_i \ge 0, \quad \lambda_i w_i = 0 \quad \forall\ i.$$

In the case $w_i > 0$, we obtain $\lambda_i = 0$ and $1 + w_i = \delta^{-1}|a_i^\top x - b_i|$, i.e., $w_i = \delta^{-1}|a_i^\top x - b_i| - 1$. Notice that this requires $|a_i^\top x - b_i| > \delta$. In the case $w_i = 0$, it follows $\lambda_i = \delta^2 - (a_i^\top x - b_i)^2$ which can only occur if $|a_i^\top x - b_i| \leq \delta$. Merging these cases, we can infer

$$w_i = \begin{cases} 0 & \text{if } |a_i^\top x - b_i| \leq \delta \\ \delta^{-1}|a_i^\top x - b_i| - 1 & \text{if } |a_i^\top x - b_i| > \delta, \end{cases} \quad \forall\ i.$$

Since the problem is convex, we can further conclude that this KKT point is a globally optimal solution. Defining $\mathcal{A} := \{i : |a_i^\top x - b_i| \leq \delta\}$ and $\mathcal{I} := \{i : |a_i^\top x - b_i| > \delta\}$, we can now write:

$$\begin{aligned} &\min_{w \geq 0} \sum_{i=1}^m \frac{(a_i^\top x - b_i)^2}{1 + w_i} + \delta^2 \cdot \mathbb{1}^\top w \\ &\qquad\qquad = \sum_{i \in \mathcal{A}} (a_i^\top x - b_i)^2 + \sum_{i \in \mathcal{I}} \frac{(a_i^\top x - b_i)^2}{\delta^{-1}|a_i^\top x - b_i|} + \delta^2 \cdot (\delta^{-1}|a_i^\top x - b_i| - 1) \\ &\qquad\qquad = \sum_{i=1}^m \varphi_\delta(a_i^\top x - b_i). \end{aligned}$$

This finishes our derivations.