Lund University

Department of
AUTOMATIC CONTROL

Exam in Optimization for Learning

2019-10-28

Points and grading

All answers must include a clear motivation. Answers should be given in English. The total number of points is 20. The maximum number of points is specified for each subproblem. Preliminary grading scales:

Grade 3:	12 points on the exam
4:	17 points on exam plus extra-credit handin
5:	22 points on exam plus extra-credit handin

Accepted aid

Authorized *Cheat Sheet*.

Results

Solutions will be posted on the course webpage, and results will be registered in LADOK. Date and location for display of corrected exams will be posted on the course webpage.

1. Determine if the following sets are convex or not:

a. $S_1 = \{(x, y) \in \mathbb{R}^2 : x^2 = y\}$. (1 p)

b. $S_2 = \{x \in \mathbb{R}^n : \max_{i=1,\dots,n} x_i \le r\}$, where $r \in \mathbb{R}$. (1 p)

c. $S_3 = \{(x, t) \in \mathbb{R}^2 : |x|^2 \le t^2\}$. (1 p)

d. $S_4 = \mathrm{epi}(\exp)$ where exp is the exponential function. (1 p)

e. $S_5 = \{x \in \mathbb{R}^n : Ax \ge b\}$ where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. (1 p)

In each subproblem, you are allowed to assume that norms are convex and that the function $h_p : \mathbb{R} \to \mathbb{R}$ such that

$$h_p(z) = \max(0, z)^p \tag{1}$$

for each $z \in \mathbb{R}$ is convex and nondecreasing, for any $p \ge 1$.

Solution

a. S_1 is not convex. Take $(-1, 1) \in S_1$ and $(1, 1) \in S_1$. Then

$$\frac{1}{2}(-1, 1) + \frac{1}{2}(1, 1) = (0, 1) \notin S_1.$$

b. S_2 is convex. S_2 is the r-sublevel set of a convex function, since the function is a point-wise maximum of convex (in particular, linear) functions.

c. S_3 is not convex. Take $(1, 1) \in S_3$ and $(1, -1) \in S_3$. Then

$$\frac{1}{2}(1, 1) + \frac{1}{2}(1, -1) = (1, 0) \notin S_3.$$

d. S_4 is convex. The exponential function is convex, therefore its epigraph is convex.

e. S_5 is convex. S_5 is a polytope.

2. Determine whether or not the functions below are convex.

a. $f_1 : \mathbb{R} \to \mathbb{R}$ such that

$$f(x) = \begin{cases} x & \text{if } x > 0, \\ -1 & \text{if } x \le 0. \end{cases}$$

(1 p)

b. $f_2 : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ such that $f_2 = g^\star$, where $g : \mathbb{R} \to \mathbb{R}$ is such that

$$g(x) = a_4x^4 + a_3x^3 + a_2x^2 + a_1x^1 + a_0$$

for each $x \in \mathbb{R}$, where $a_0, a_1, a_2, a_3 \in \mathbb{R}$. (1 p)

c. $f_3 : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ such that

$$f_3(x) = \begin{cases} -\min(\log(x), -e^{-x}) & \text{if } x > 0, \\ \infty & \text{if } x \leq 0. \end{cases}$$

(1 p)

d. $f_4 : \mathbb{R} \to \mathbb{R}$ such that

$$f_4(x) = |x|^3$$

for each $x \in \mathbb{R}$. (1 p)

e. $f_5 : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ such that

$$f(x) = \begin{cases} \|x\|_2^2 & \text{if } Ax = b, \\ \infty & \text{if } Ax \neq b, \end{cases}$$

for each $x \in \mathbb{R}^n$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. (1 p)

Solution

a. f_1 is not convex. Take $x = 0$, $y = 1$, and $\theta = 0.5$. Then

$$\theta x + (1 - \theta)y = 0.5$$

and

$$f(\theta x + (1 - \theta)y) = 0.5 > 0 = \theta f(x) + (1 - \theta)f(y).$$

b. f_2 is convex. The conjugate of any function is a point-wise supremum of convex (in particular, affine) functions, and therefore itself convex.

c. f_3 is convex. Note that f_3 is the maximum of the two convex functions

$$x \mapsto \begin{cases} -\log(x) & \text{if } x > 0, \\ \infty & \text{if } x \leq 0, \end{cases} \qquad \text{from } \mathbb{R} \text{ to } \mathbb{R} \cup \{\infty\}$$

and

$$x \mapsto e^{-x} \qquad \text{from } \mathbb{R} \text{ to } \mathbb{R},$$

and is therefore itself convex. The convexity of these two functions can be checked using, e.g., the second-order condition for convexity.

d. f_4 is convex. Note that

$$f_4'(x) = 3x\,|x| \qquad \text{and} \qquad f_4''(x) = 6\,|x| \geq 0$$

for each $x \in \mathbb{R}$. The second-order condition for convexity gives that f_4 is convex.

e. f_5 is convex. Note that the mapping

$$x \mapsto h_2(\|x\|_2) = \|x\|_2^2$$

from $\mathbb{R}^n$ to $\mathbb{R}$ is convex since it is a composition of the convex and nondecreasing function h_2 defined in (1) and the convex function $\|\cdot\|_2$. Moreover,

$$x \mapsto \iota_{\{y \in \mathbb{R}^n : Ay = b\}}(x)$$

from $\mathbb{R}^n$ to $\mathbb{R}$ is convex, since the set $\{y \in \mathbb{R}^n : Ay = b\}$ is convex (in particular, a polytope). However, note that f_5 can be written as

$$f_5 = \|\cdot\|_2^2 + \iota_{\{y \in \mathbb{R}^n : Ay = b\}},$$

i.e., a sum of convex functions, and is therefore itself convex.

3. Consider the proper, closed and convex function $f : \mathbb{R} \to \mathbb{R}$ such that

$$f(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leq \frac{1}{2}, \\ \frac{1}{2}(|x| - \frac{1}{4}) & \text{if } |x| > \frac{1}{2}, \end{cases}$$

for each $x \in \mathbb{R}$, known as the *Huber loss*. Let $\gamma > 0$. Compute the proximal operator $\operatorname{prox}_{\gamma f}$. (1 p)

Solution

Let $z \in \mathbb{R}$ and

$$\begin{aligned} x &= \operatorname{prox}_{\gamma f}(z) \\ &= \underset{y \in \mathbb{R}}{\operatorname{argmin}} \left(f(y) - \frac{1}{2\gamma}(y - z)^2 \right). \end{aligned}$$

Fermat's rule gives that this holds if and only if

$$0 \in \partial f(x) + \gamma^{-1}(x - z). \tag{2}$$

Note that f is differentiable with derivative

$$f'(x) = \begin{cases} x & \text{if } |x| \leq \frac{1}{2}, \\ \frac{\operatorname{sgn} x}{2} & \text{if } |x| > \frac{1}{2}. \end{cases}$$

Since f also is convex, we know that $\partial f(x) = \{\nabla f(x)\}$, for each $x \in \mathbb{R}$. We consider two different cases:

- Suppose that $|x| \leq 1/2$. Then (2) gives that

  $$\begin{gathered} 0 = x + \gamma^{-1}(x - z) \\ \Leftrightarrow \\ x = (1 + \gamma)^{-1} z, \end{gathered}$$

  and $|z| \leq (1 + \gamma)/2$.

- Suppose that $|x| > 1/2$. Then (2) gives that

$$
\begin{aligned}
0 = \frac{\operatorname{sgn}(x)}{2} + \gamma^{-1}(x - z) \\
\Leftrightarrow \\
z = x + \frac{\gamma \operatorname{sgn}(x)}{2} \qquad [\text{note that } \operatorname{sgn} x = \operatorname{sgn} z] \\
\Leftrightarrow \\
z = x + \frac{\gamma \operatorname{sgn}(z)}{2} \\
\Leftrightarrow \\
x = z - \frac{\gamma \operatorname{sgn}(z)}{2}
\end{aligned}
$$

  and $|z| > (1+\gamma)/2$.

This covers all cases. We conclude that

$$
\operatorname{prox}_{\gamma f}(z) = \begin{cases} (1+\gamma)^{-1} z & \text{if } |z| \leq \dfrac{1+\gamma}{2}, \\ z - \dfrac{\gamma \operatorname{sgn}(z)}{2} & \text{if } |z| > \dfrac{1+\gamma}{2}. \end{cases}
$$

4. Consider the proper, closed and convex function $f : \mathbb{R}^n \to \mathbb{R}$ such that

$$
f(x) = \begin{cases} \|x\|_2 & \text{if } \|x\|_2 \leq 1, \\ \frac{1}{2}\left(\|x\|_2^2 + 1\right) & \text{if } \|x\|_2 > 1 \end{cases}
$$

for each $x \in \mathbb{R}$, known as the *reversed Huber* function. Compute ∂f. (1 p)

Solution

Note that f is differentiable everywhere except at 0. In particular,

$$
\nabla f(x) = \begin{cases} \dfrac{x}{\|x\|_2} & \text{if } x \in \mathbb{R}^n \setminus \{0\} \text{ and } \|x\|_2 \leq 1, \\ x & \text{if } x \in \mathbb{R}^n \text{ and } \|x\|_2 > 1. \end{cases}
$$

Since f is also convex, we know that $\partial f(x) = \{\nabla f(x)\}$ for each $x \in \mathbb{R}^n \setminus \{0\}$. It remains to find $\partial f(0)$. Note that

$$
\begin{aligned}
s \in \partial f(0) \\
\Leftrightarrow \\
f(y) \geq f(0) + s^T y, \quad \forall y \in \mathbb{R}^n \\
\Leftrightarrow \\
s^T y \leq \begin{cases} \|y\|_2\,, & \forall y \in \mathbb{R}^n : \|y\|_2 \leq 1, \\ \frac{1}{2}\left(\|y\|_2^2 + 1\right), & \forall y \in \mathbb{R}^n : \|y\|_2 > 1. \end{cases}
\end{aligned}
$$

The first requirement in the cases above holds if and only if $\|s\| \leq 1$, using the Cauchy-Schwarz inequality. In fact, if $\|s\| \leq 1$, then $s^T y \leq \|y\|_2$ for each $y \in \mathbb{R}^n$, by the Cauchy-Schwarz inequality. But if $\|s\| \leq 1$, the second requirement in the cases above holds automatically since

$$\begin{gathered} 0 \leq (\|y\|_2 - 1)^2, \quad \forall y \in \mathbb{R}^n \\ \Leftrightarrow \\ 0 \leq \|y\|_2^2 - 2\,\|y\|_2 + 1, \quad \forall y \in \mathbb{R}^n \\ \Leftrightarrow \\ \|y\|_2 \leq \frac{1}{2}\Big(\|y\|_2^2 + 1\Big), \quad \forall y \in \mathbb{R}^n, \end{gathered}$$

and

$$s^T y \leq \|y\|_2\,, \quad \forall y \in \mathbb{R}^n,$$

as argued above. We conclude that

$$\partial f(0) = \{s \in \mathbb{R}^n : \|s\|_2 \leq 1\}.$$

5. A proper and convex function $f : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ has the following properties: $f(-2) = 3,\ \partial f(-1) = \{-1\},\ \partial f(0) = [-1, 0]$. What can you conclude about the following properties?

a. Smoothness. (1 p)

b. Strong convexity. (1 p)

Solution

a. The function f is not smooth since it is not differentiable at 0 ($\partial f(0)$ is not a singleton).

b. The function f is not strongly convex. Note that

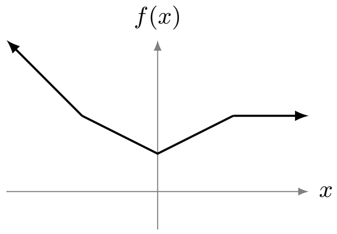
$$-1 \in \partial f(-1) \quad \text{and} \quad -1 \in \partial f(0).$$

Therefore, ∂f is not strongly monotone since

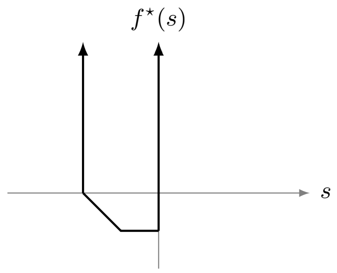
$$((-1) - (-1))((-1) - 0) = 0 < \sigma = \sigma\,|(-1) - 0|$$

for any $\sigma > 0$. Therefore, f can not be strongly convex.

6. Sketch the conjugate of the piecewise linear function showed below. Outside the plotted domain, assume the graph continues in the same direction as on the boundary. (1 p)



Solution



7. Let $\gamma > 0$, $b \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$.

a. Find $(\gamma f)^*$ expressed in terms of f^* and γ. (1 p)

b. Find $(f(\cdot - b))^*$ expressed in terms f^* and b. (1 p)

Solution
Let $s \in \mathbb{R}^n$.

a. Then

$$\begin{aligned}(\gamma f)^*(s) &= \sup_{x\in\mathbb{R}^n} \left(s^T x - \gamma f(x)\right) \\ &= \gamma \sup_{x\in\mathbb{R}^n} \left(\left(\gamma^{-1} s\right)^T x - f(x)\right) \\ &= \gamma f^*(\gamma^{-1} s).\end{aligned}$$

b. Note that

$$\begin{aligned}(f(\cdot - b))^*(s) &= \sup_{x\in\mathbb{R}^n} \left(s^T x - f(x-b)\right) \qquad [z = x - b] \\ &= \sup_{z\in\mathbb{R}^n} \left(s^T (z+b) - f(z)\right) \\ &= s^T b + \sup_{z\in\mathbb{R}^n} \left(s^T z - f(z)\right) \\ &= s^T b + f^*(s).\end{aligned}$$

8. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable, and that $g_i : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is proper, closed and convex for each $i = 1, \dots, n$. Consider the coordinate proximal-gradient method:

1. Pick an initial point $x^0 \in \mathbb{R}^n$ and step-lengths $\gamma_i > 0$ for each $i = 1, \dots, n$
2. For $k = 0, 1, 2, \dots$
   (a) Choose an index $i \in \{1, \dots, n\}$ (according to some schedule)
   (b) Update $x_i^{k+1} = \operatorname{prox}_{\gamma_i g_i}\left(x_i^k - \gamma_i(\nabla f(x^k))_i\right)$
   (c) Update $x_j^{k+1} = x_j^k$ for each $j = 1, \dots, n$ such that $j \neq i$

Now, suppose that $x^\star \in \mathbb{R}^n$ is a fixed point of the coordinate proximal-gradient method, i.e., if $x^k = x^\star$, then $x^{k+1} = x^\star$, regardless of which coordinate i was chosen. Show that $x^\star$ solves

$$\underset{x=(x_1,\dots,x_n)\in\mathbb{R}^n}{\text{minimize}}\ f(x) + \sum_{i=1}^n g_i(x_i).$$

(1 p)

Solution

First, define $g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ such that

$$g(x) = \sum_{i=1}^n g_i(x_i)$$

for each $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. The optimization problem can then be written as

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ f(x) + g(x).$$

Moreover, we know that

$$\partial g(x) = \{(s_1, \dots, s_n) \in \mathbb{R}^n : s_i \in \partial g_i(x_i) \text{ for each } i = 1, \dots, n\}.$$

Now, we have that

$$\begin{aligned} x_i^\star &= \operatorname{prox}_{\gamma_i g_i}(x_i^\star - \gamma_i(\nabla f(x^\star))_i) \\ &= \underset{x\in\mathbb{R}}{\operatorname{argmin}} \left(g_i(x) + \frac{1}{2\gamma_i}(x - (x_i^\star - \gamma_i(\nabla f(x^\star))_i))^2\right), \end{aligned}$$

for each $i = 1, \dots, n$. Fermat's rule gives that this is equivalent to that

$$\begin{aligned} 0 &\in \partial g_i(x_i^\star) + \gamma_i^{-1}(x_i^\star - (x_i^\star - \gamma_i(\nabla f(x^\star))_i)) \\ &= \partial g_i(x_i^\star) + (\nabla f(x^\star))_i, \end{aligned}$$

or

$$-(\nabla f(x^\star))_i \in \partial g_i(x_i^\star),$$

for each $i = 1, \ldots, n$. This implies that

$$-\nabla f(x^\star) \in \partial g(x^\star)$$

or

$$0 \in \nabla f(x^\star) + \partial g(x^\star) \subseteq \partial (f+g)(x^\star).$$

Fermat's rule implies that

$$x^\star \in \operatorname*{Argmin}_{x \in \mathbb{R}^n} (f(x) + g(x)),$$

as desired.

9. Consider the primal problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \, \|x\|_1 + \left|\mathbf{1}^T x\right|$$

and the dual problem

$$\underset{\mu \in \mathbb{R}}{\text{minimize}} \, \iota_{[-\mathbf{1},\mathbf{1}]}(-\mathbf{1}\mu) + \iota_{[-1,1]}(\mu)$$

where $\mathbf{1} \in \mathbb{R}^n$ is a vector of all ones. Here we used that

$$(\|\cdot\|_1)^\star = \iota_{[-\mathbf{1},\mathbf{1}]}$$

and

$$(|\cdot|)^\star = \iota_{[-1,1]}.$$

Suppose that $\mu^\star \in \mathbb{R}$ is a solution to the dual problem. Recover the primal solution $x^\star \in \mathbb{R}^n$ using $\mu^\star$. (2 p)

Solution

First, note that it must be the case that $-1 \le \mu^\star \le 1$ for a dual optimal solution, since if not, the objective function of the dual problem would be infinite, which clearly is not optimal.

Next, note that

$$\begin{aligned} \operatorname{relint} \operatorname{dom} \left(\iota_{[-\mathbf{1},\mathbf{1}]} \circ -\mathbf{1}\right) \cap \operatorname{relint} \operatorname{dom} \iota_{[-1,1]} &= \operatorname{relint}[-1,1] \cap \operatorname{relint}[-1,1] \\ &= (-1,1) \\ &\neq \emptyset, \end{aligned}$$

i.e., constraint qualification holds for the dual problem. Fermat's rule gives that $\mu^\star$ is a solution to the dual problem if and only if

$$\begin{aligned} 0 &\in \partial\Big(\left(\iota_{[-\mathbf{1},\mathbf{1}]} \circ -\mathbf{1}\right) + \iota_{[-1,1]}\Big)(\mu^\star) \\ &= \partial\left(\iota_{[-\mathbf{1},\mathbf{1}]} \circ -\mathbf{1}\right)(\mu^\star) + \partial \iota_{[-1,1]}(\mu^\star) \\ &= -\mathbf{1}^T \partial \iota_{[-\mathbf{1},\mathbf{1}]}(-\mathbf{1}\mu^\star) + \partial \iota_{[-1,1]}(\mu^\star). \end{aligned}$$

This is equivalent to that there exists a point $x^\star \in \mathbb{R}^n$ such that

$$\begin{cases} \mathbf{1}^T x^\star \in \partial\iota_{[-1,1]}(\mu^\star), \\ x^\star \in \partial\iota_{[-\mathbf{1},\mathbf{1}]}(-\mathbf{1}\mu^\star). \end{cases} \tag{3}$$

Note that

$$\partial\iota_{[-\mathbf{1},\mathbf{1}]}(z) = \left\{ (s_1,\ldots,s_n) \in \mathbb{R}^n : s_i \in \begin{cases} (-\infty, 0] & \text{if } z_i = -1, \\ \{0\} & \text{if } -1 < z_i < 1, \\ [0,\infty) & \text{if } z_i = 1, \\ \emptyset & \text{if } z_i \in \mathbb{R} \setminus [-1,1], \end{cases} \quad \forall i \in \{1,\ldots,n\} \right\}$$

for each $z = (z_1,\ldots,z_n) \in \mathbb{R}^n$, and that

$$\partial\iota_{[-1,1]}(\mu) = \begin{cases} (-\infty, 0] & \text{if } \mu = -1, \\ \{0\} & \text{if } -1 < \mu < 1, \\ [0,\infty) & \text{if } \mu = 1, \\ \emptyset & \text{if } \mu \in \mathbb{R} \setminus [-1,1], \end{cases}$$

for each $\mu \in \mathbb{R}$.

We consider three different cases:

- Suppose that $-1 < \mu^\star < 1$. The second condition in (3) gives that $x^\star = 0$.
- Suppose that $\mu^\star = 1$. The first condition in (3) gives that

  $$\mathbf{1}^T x^\star \geq 0$$

  and the second condition in (3) gives that

  $$x_i^\star \leq 0$$

  for each $i = 1,\ldots,n$. This is only possible if $x^\star = 0$.
- Suppose that $\mu^\star = -1$. The first condition in (3) gives that

  $$\mathbf{1}^T x^\star \leq 0$$

  and the second condition in (3) gives that

  $$x_i^\star \geq 0$$

  for each $i = 1,\ldots,n$. This is only possible if $x^\star = 0$.

This covers all cases. We conclude that $x^\star = 0$.

One can verify that $x^\star = 0$ is in fact the optimal solution to the primal problem simply by inspecting the objective function in the primal problem.