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. Which of the following sets S are convex?

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

b. $S = \{x \in \mathbb{R}^n : \max_i(x_i) \leq r\}$, where $r > 0$. (1 p)

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

d. $S = \operatorname{epi}(f)$ where $f(x) = e^x$ and $x \in \mathbb{R}$. (1 p)

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

Solution

a. Not convex. Take $(x_1, y_1) = (-1, 1)$ and $(x_2, y_2) = (1, 1)$, but $(0, 1)$ is not in the set.

b. Trivial from definition. Or since S is sub-levelset of convex function (max of convex functions is convex).

c. Not convex. Take $(x_1, y_1) = (1, 1)$ and $(x_2, y_2) = (1, -1)$, but $(1, 0)$ is not in the set.

d. f is convex, therefore its epigraph is convex.

e. S is an intersection of half-spaces and therefore convex.

2. Which of the following functions f are convex? Prove or disprove.

a. $f(x) = \begin{cases} x & \text{if } x > 0 \\ -1 & \text{if } x \leq 0 \end{cases}$ where $x \in \mathbb{R}$. (1 p)

b. $f(x) = g^*(x)$ where $g(x) = a_4x^4 + a_3x^3 + a_2x^2 + a_1x^1 + a_0$ where $x \in \mathbb{R}$. (1 p)

c. $f(x) = \begin{cases} -\min(\log(x), -e^{-x}) & \text{if } x > 0 \\ \infty & \text{if } x \leq 0 \end{cases}$ where $x \in \mathbb{R}$. (1 p)

d. $f(x) = |x|^3$ where $x \in \mathbb{R}$. (1 p)

e. $f(x) = \begin{cases} \|x\|_2^2 & \text{if } Ax = b \\ \infty & \text{otherwise} \end{cases}$ where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. (1 p)

Solution

a. Nonconvex. Take $x = 0, y = 1, \theta = 0.5$: $\theta f(x) + (1-\theta)f(y) = -0.5 + 0.5 \cdot 0.5 = -0.25 < f(\theta x + (1-\theta)y) = f(0.5) = 0.5$

b. The conjugate of any function is a sup over affine functions, hence convex.

c. Convex. The domain is convex and $-\min(\log(x), -e^{-x}) = \max(-\log(x), e^{-x})$ is convex. ($\log(x)$ is concave, so $-\log(x)$ is convex and so is e^{-x}. Thus it is max of convex functions.)

d. Convex. It is twice differentiable with non-negative second derivative, $f''(x) = 6|x|$.

e. Convex. Convex domain and convex on the domain. Alternative, sum of convex functions $f(x) = x^2 + \iota_{Ax=b}(x)$.

3. Compute the proximal operator $\text{prox}_{\gamma f}(x)$ for the Huber loss

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

where $f : \mathbb{R} \to \mathbb{R}$. (1 p)

Solution

We have that $x = \text{prox}_{\gamma f}(z)$ iff $0 = \partial f(x) + \frac{1}{2\gamma}(x - z)$.
Assume $x < -1/2$, then $0 = -1/2 + \frac{1}{\gamma}(x - z) \Leftrightarrow x = z + \gamma/2$, where $z < -(1 + \gamma)/2$.
Assume $-1/2 < x < 1/2$, then $0 = x + \frac{1}{\gamma}(x - z) \Leftrightarrow x = \frac{1}{1+\gamma}z$, where $0 < z < (1 + \gamma)/2$.
Assume $1/2 < x$, then $0 = 1/2 + \frac{1}{\gamma}(x - z) \Leftrightarrow x = z - \gamma/2$, where $z > (1 + \gamma)/2$.
From symmetry around 0 we get

$$\text{prox}_{\gamma f}(z) = \begin{cases} z + \gamma/2 & \text{if } z < -(1 + \gamma)/2 \\ \dfrac{1}{1 + \gamma}z & \text{if } -(1 + \gamma)/2 < z < (1 + \gamma)/2 \\ z - \gamma/2 & \text{if } (1 + \gamma)/2 < z \end{cases}$$

4. Compute the subdifferential to the berHu (reversed Huber) loss

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

(1 p)

Solution

f is convex and differentiable everywhere except in 0, so subdifferentials are given by the gradients:

$$\nabla f(x) = x/\|x\|_2, \quad \forall x : 0 \neq \|x\|_2 \leq 1,$$

and

$$\nabla f(x) = x, \quad \forall x : \|x\|_2 \geq 1.$$

At zero, the definition of subgradient s gives that for all y:

$$f(y) \geq f(0) + s^T y \Leftrightarrow \|y\|_2 \geq f(0) + s^T y = s^T y$$

Which holds if and only if $\|s\|_2 \leq 1$, i.e $\partial f(0) = \{s : \|s\|_2 \leq 1\}$.

5. A convex function f 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)

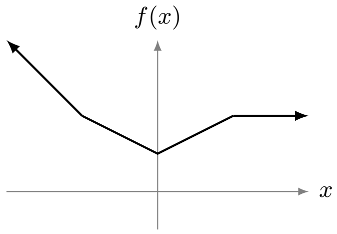
b. Strong convexity (1 p)

Solution

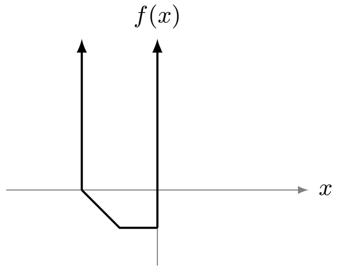
a. Not smooth since it is not differentiable at 0.

b. Not strongly convex since $-1 \in \partial f$ at -1 and 0. I.e the second derivative of f in this interval is 0.

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. Consider the functions $g(x) = \gamma f(x)$ and $h(x) = f(x - b)$ where $\gamma > 0$.

a. Find g^* expressed in f^* and γ. (1 p)

b. Find h^* expressed in f^* and b. (1 p)

Solution

a.

$$
\begin{aligned}
g^*(y) &= \sup_x y^T x - g(x) \\
&= \sup_x y^T x - \gamma f(x) \\
&= \gamma \sup_x \tfrac{1}{\gamma} y^T x - f(x) \\
&= \gamma \sup_x (\tfrac{1}{\gamma} y)^T x - f(x) \\
&= \gamma f^*(\tfrac{1}{\gamma} y)
\end{aligned}
$$

b.

$$
\begin{aligned}
h^*(y) &= \sup_x y^T x - h(x) \\
&= \sup_x y^T x - f(x-b) \\
&= [z = x - b] \\
&= \sup_z y^T (z+b) - f(z) \\
&= \sup_z y^T z + y^T b - f(z) \\
&= y^T b + \sup_z y^T z - f(z) \\
&= b^T y + f^*(y)
\end{aligned}
$$

8. Consider the coordinate proximal-gradient method (CPG)

$$
\begin{aligned}
&\text{Choose } i \\
&x_i^{k+1} = \operatorname{prox}_{\gamma_i g_i}(x_i^k - \gamma_i(\nabla f(x^k))_i) \\
&x_j^{k+1} = x_j^k \quad \forall j \neq i.
\end{aligned}
$$

Let $x^\star$ be a fixed point of CPG, i.e. $x^k = x^\star \implies x^{k+1} = x^\star$, regardless of which coordinate i was chosen. Show that $x^\star$ solves the problem

$$
\min f(x) + \sum_{i=1}^{n} g_i(x_i)
$$

given that f and g_i are closed convex and constraint qualification hold. (1 p)

Solution

For all i the following hold

$$
\begin{aligned}
x_i^\star &= \operatorname{prox}_{\gamma_i g_i}(x_i^\star - \gamma_i(\nabla f(x^\star))_i) \\
&= \operatorname{argmax}_{z_i} g_i(z_i) + \tfrac{1}{2\gamma_i}\|z_i - (x_i^\star - \gamma_i(\nabla f(x^\star))_i)\|^2.
\end{aligned}
$$

Using Fermat's rule gives

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

Gathering all i gives

$$
0 \in \begin{bmatrix} \partial g_1(x_1^\star) + (\nabla f(x^\star))_1 \\ \vdots \\ \partial g_n(x_n^\star) + (\nabla f(x^\star))_n \end{bmatrix} = \partial g(x^\star) + \nabla f(x^\star)
$$

which is Fermat's rule for the given problem, i.e. $x^\star$ solves the problem.

9. Consider primal problem

$$\min_{x \in \mathbb{R}^n} \|x\|_1 + |\mathbf{1}^T x|$$

and dual problem

$$\min_{y \in \mathbb{R}} \iota_{[-\mathbf{1},\mathbf{1}]}(-\mathbf{1}y) + \iota_{[-1,1]}(y)$$

where $\mathbf{1} \in \mathbb{R}^n$ is a vector of all ones. Then function $\iota_{[a,b]}$ is the indicator of the set $\{x \in \mathbb{R}^m : a \leq x \leq b\}$ where the inequality is applied element wise.

Let $y^\star$ be a solution to the dual problem and assume the primal solution $x^\star$ is unknown. Recover $x^\star$ from $y^\star$ using the primal-dual optimality conditions.

(2 p)

Solution

Using the primal-dual optimality conditions gives

$$\begin{cases} \mathbf{1}^T x \in \partial g^*(y) \\ \quad x \in \partial f^*(-\mathbf{1}y) \end{cases}$$

where $g^*(y) = \iota_{[-1,1]}(y)$ and $f^*(z) = \iota_{-\mathbf{1},\mathbf{1}}(z)$. The sub-differential of $f^*(z)$ is given by

$$(\partial f^*(z))_i = \begin{cases} 0 & \text{if } -1 < z_i < 1 \\ [0, \infty) & \text{if } z_i = 1 \\ (-\infty, 0] & \text{if } z_i = -1 \end{cases}.$$

We see that if $-1 < y < 1$ then $x = 0$.

If $y = 1$ then $\forall i, x_i \in (-\infty, 0]$ and ∂g^* needs to be used. Note that $y = 1 \implies \mathbf{1}^T x \leq 0$ and equality only hold when $x = 0$. Furthermore,$\partial g^*(1) = [0, \infty) \implies \mathbf{1}^T x \geq 0$ which gives that $\mathbf{1}^T x = 0$ and $y = 1 \implies x = 0$. An analogue argument can be made for the case when $y = -1$.

Finally we note that an optimal y can't take any other values, i.e. the dual solution $y^\star$ gives the (unsurprising) primal solution $x^\star = 0$.