Lund University

Department of
AUTOMATIC CONTROL

Exam in Optimization for Learning

Test exam

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

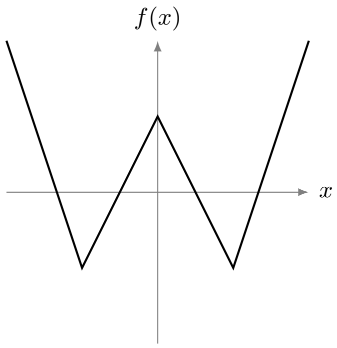
Authorized *Cheat Sheet*.

Results

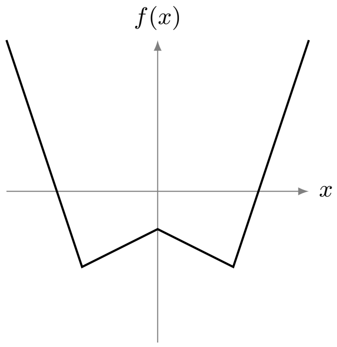
Solutions will be posted on the course webpage, and results will be registered in LADOK. Date and location for display of corrected exams will be posted on the course webpage.

1. Are the following sets C convex? You may, where applicable and without proving it, use that $\|x\|_\infty := \max_i(|x_i|)$ is a convex function and that $\{x \in \mathbb{R}^n : x \geq 0\}$ is a convex set.

a. Let $C = \{x \in \mathbb{R} : f(x) \leq 0\}$ where f is given below: (1 p)



b. $C = \{x \in \mathbb{R} : f(x) \leq 0\}$ where f is given below: (1 p)



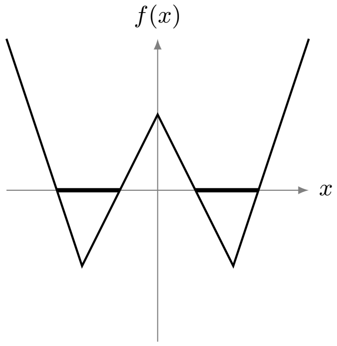
c. $C = \{x \in \mathbb{R}^n : x \geq 0 \text{ and } \|x\|_\infty \leq 1\}$. (1 p)

d. $C = \{x \in \mathbb{R}^n : x \geq 0 \text{ or } \|x\|_\infty \leq 1\}$. (1 p)

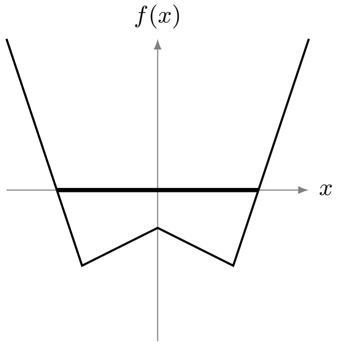
e. $C = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : \|x\|_\infty \leq t\}$. (1 p)

Solution

a. Nonconvex.



b. Convex.



c. Convex. $x \geq 0$ convex and $\|x\|_\infty - 1 \leq 0$ convex since sublevelset of convex function. Intersection of convex functions is convex.

d. Nonconvex. Consider $\mathbb{R}^2$ and the points $x_1 = (-1, 1)$ and $x_2 = (0, 2)$. Then $\|x_1\|_\infty \leq 1$ and $x_2 \geq 0$, but the convex combination $x_3 = \frac{1}{2}(x_1 + x_2) = (-1/2, 3/2)$ is in none of the sets.

e. Convex. The set can be written as $\{(x, t) : \|x\|_\infty - t \leq 0\}$, which is the sublevel set of convex function, hence convex.

2. Are the following functions f convex? Where applicable, you may use convexity preserving operations or graphical arguments.

a. $f(x) = \frac{1}{2}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ (1 p)

b. $f(x) = e^{\|x\|_1}$ where $x \in \mathbb{R}^n$ (1 p)

c. $f(x) = \left|e^{\|x\|_1}\right|$ where $x \in \mathbb{R}^n$ (1 p)

d. $f(x) = \max(f(x), g(x))$ where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}$ are both convex. (1 p)

e. $f(x) = \min(x^2, x)$ where $x \in \mathbb{R}$ (1 p)

f. $f(x) = \max(x^2, x, -x^2)$ where $x \in \mathbb{R}$ (1 p)

g. $f(x) = \max(x^2, x, 1 - x^2)$ where $x \in \mathbb{R}$ (1 p)

Solution

a. Hessian is constant and symmetric, but not positive semidefinite (eigenvalues -1 and 1). Hence nonconvex.

b. It is a composition between nondecreasing e^u and convex $\|x\|_1$, hence convex.

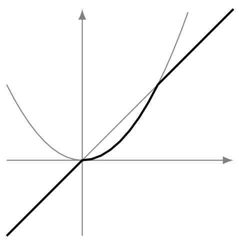
c. Convex since $|e^{\|x\|_1}| = e^{\|x\|_1}$ which is convex due to **b.**.

d. Supremum (maximum) of convex functions is convexity preserving operation, hence convex. The epigraph is convex since

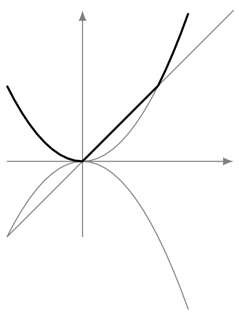
$$\begin{aligned} \mathrm{epi}(\max(f, g)) &= \{(x, t) : \max(f(x), g(x)) \leq t\} = \{(x, t) : f(x) \leq t \text{ and } g(x) \leq t\} \\ &= \{(x, t) : f(x) \leq t\} \cap \{(x, t) : g(x) \leq t\} = \mathrm{epi} f \cap \mathrm{epi} g \end{aligned}$$

i.e., intersection of individual *convex* epigraphs, hence convex.

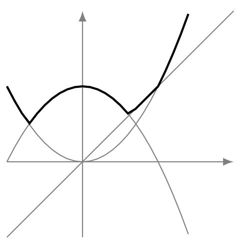
e. Not convex.



f. Convex.



g. Not convex.



3. Consider the following problem

$$\underset{u}{\text{minimize}} \max(h_1^T u, h_2^T u) + \tfrac{1}{2}\|u\|_2^2.$$

Show that u solves this if and only if (u, t) solves

$$\begin{array}{ll} \underset{u,t}{\text{minimize}} & t + \tfrac{1}{2}\|u\|_2^2 \\ \text{subject to} & h_1^T u \leq t \\ & h_2^T u \leq t \end{array}$$

(1 p)

Solution

Since $t \geq h_1^T u$ and $t \geq h_2^T u$ it must satisfy $t \geq \max(h_1^T u, h_2^T u)$ for all feasible (t, u). Now, assume that an optimal pair $(t^\star, u^\star)$ satisfies $t^\star = \max(h_1^T u^\star, h_2^T u^\star) + \epsilon$, where $\epsilon > 0$ (i.e., $t^\star > \max(h_1^T u^\star, h_2^T u^\star)$). Then $\tilde{t} = t^\star - \epsilon$ achieves a lower cost (since t is in the objective and we minimize over t) and satisfies both constraints: $h_1^T u^\star \leq \tilde{t}$ and $h_2^T u^\star \leq \tilde{t}$. We conclude that the optimal pair $(t^\star, u^\star)$ satisfies $t^\star = \max(h_1^T u^\star, h_2^T u^\star)$. Hence the problems are equivalent.

4. Consider the following problem (that satisfies constraint qualification):

$$\underset{x}{\text{minimize}}\ \underbrace{\|Ax-b\|_1}_{f(Ax)} + \underbrace{\sum_{i=1}^{n} g_i(x_i)}_{g(x)}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, all $g_i : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ are the same and satisfy

$$g_i(x_i) = \begin{cases} \frac{1}{2}x_i^2 & \text{if } x_i \in [-1,1] \\ \infty & \text{else} \end{cases}$$

and $f(y) = \|y-b\|_1$ with dual problem

$$\underset{\mu}{\text{minimize}}\ f^*(\mu) + g^*(-A^T\mu).$$

a. Compute the conjugate f^*. (1 p)

b. Compute the conjugate g^*. (1 p)

c. Can the primal problem be solved using the proximal gradient method? (1 p)

d. Can the dual problem be solved using the proximal gradient method? (You don't need to have solved **a.** or **b.** to answer this.) (1 p)

e. Assume a dual solution $\mu^\star$ is known. Provide *one* simple condition that recovers a primal solution. (You don't need to have solved **a.** or **b.** to answer this, you can phrase the recovery in terms of functions f or g, or conjugates f^* and g^*.) (1 p)

Solution

a. We first note that

$$f^*(s) = \sup_y (s^T y - \|y-b\|_1) = \sup_v (s^T(v+b) - \|v\|_1) = \sup_v (s^T v - \|v\|_1) + s^T b.$$

Therefore, we can compute the conjugate of $\|\cdot\|_1$ and then add $s^T b$.
We let $\hat{f}(y) = \sum_{i=1}^m |y_i|$, which implies $\hat{f}^*(s) = \sum_{i=1}^m \hat{f}_i^*(s)$ by separability of $\hat{f}$, where $\hat{f}_i(y_i) = |y_i|$. We have

$$\hat{f}_i^*(s_i) = \sup_{y_i}(s_i - |y_i|).$$

There are many ways to compute the conjugate of 1D functions. We provide a proof that directly optimizes the definition for different s.

- For $s_i < -1$, let $y_i = t_- \le 0$ with $t_- \to -\infty$, which gives

$$\hat{f}_i^*(s_i) = \sup_{y_i}(y_i s_i - |y_i|) \ge s_i t_- - |t_-| = (s_i+1)t_- \to \infty$$

  since $s_i < -1$.

- For $s_i > 1$, let $y_i = t_+ \leq 0$ with $t_+ \to \infty$, which gives

$$\hat{f}_i^*(s_i) = \sup_{y_i}(y_i s - |y_i|) \geq s_i t_+ - |t_+| = (s_i - 1)t_+ \to \infty$$

  since $s_i > 1$.
- For $s_i \in [-1, 1]$, we have by Cauchy-Schwarz that $s_i y_i \leq |y_i||s_i| \leq |y_i|$ for all y_i. Therefore $\hat{f}_i^*(s_i) = \sup_{y_i} s_i^T y_i - |y_i| \leq \sup_{y_i} |y_i| - |y_i| = 0$. Further, $\hat{f}_i^*(s_i) = \sup_{y_i} s_i y_i - |y_i| \geq \sup_{y_i} s_i 0 - 0 = 0$. Hence $\hat{f}_i^*(s_i) = 0$ for all $s_i \in [-1, 1]$.

The conjugate becomes

$$\hat{f}_i^*(s_i) = \begin{cases} 0 & \text{if } s_i \in [-1, 1] \\ \infty & \text{else} \end{cases}$$

i.e. $\hat{f}_i^*(s_i) = \iota_{[-1,1]}(s_i)$. This implies

$$f^*(s) = \hat{f}^*(s) + s^T b = \left(\sum_{i=1}^{m} \iota_{[-1,1]}(s_i)\right) + s^T b = \iota_{[-\mathbf{1},\mathbf{1}]}(s) + s^T b.$$

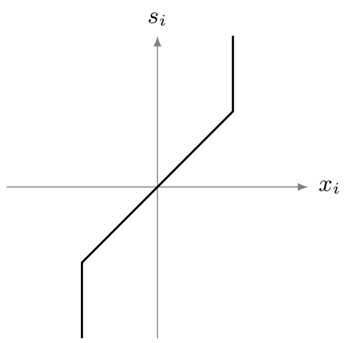
b. We have $g(x) = \sum_{i=1}^m g_i(x_i)$, which implies $g^*(s) = \sum_{i=1}^m g_i^*(s_i)$ by separability of g. We will compute the conjugate from the relation

$$g_i^*(s_i) = s_i x_i - g_i(x_i) \quad \text{if and only if} \quad s_i \in \partial g_i(x_i).$$

The subgradient of g_i is given by

$$\partial g_i(x_i) = \begin{cases} (-\infty, -1] & \text{if } x_i = -1 \\ x_i & \text{if } x_i \in (-1, 1) \\ [1, \infty) & \text{if } x_i = 1 \\ \emptyset & \text{else} \end{cases}$$

or in graphics



Therefore

- For $x_i = -1$: $g_i(x_i) = \frac{1}{2}$, $s_i \in (-\infty, -1]$, and

$$g_i^*(s_i) = s_i x_i - g_i(x_i) = -s_i - \tfrac{1}{2}$$

  for $s_i \in (-\infty, -1]$.

- For $x_i \in (-1, 1)$: $g_i(x_i) = \frac{1}{2}x_i^2$, $s_i = x_i$, and

$$g_i^*(s_i) = s_i x_i - g_i(x_i) = s_i^2 - \tfrac{1}{2}x_i^2 = s_i^2 - \tfrac{1}{2}s_i^2 = \tfrac{1}{2}s_i^2$$

  for $s_i \in (-1, 1)$.
- For $x_i = 1$: $g_i(x_i) = \frac{1}{2}$, $s_i \in [1, \infty)$, and

$$g_i^*(s_i) = s_i x_i - g_i(x_i) = s_i - \tfrac{1}{2}$$

  for $s_i \in [1, \infty)$.

To conclude:

$$g^*(s) = \sum_{i=1}^n g_i(s_i) = \sum_{i=1}^n \begin{cases} -s_i - \frac{1}{2} & \text{if } s_i \le -1 \\ \frac{1}{2}s_i^2 & \text{if } s_i \in [-1, 1] \\ s_i - \frac{1}{2} & \text{if } s_i \ge 1 \end{cases}$$

which is sum of Huber functions.

c. No. None of the functions are smooth.

d. Yes. Each g_i is strongly convex and so is g. Therefore g^* is smooth and so is $g^* \circ -A^T$. Therefore the proximal gradient method can be applied to the dual. In addition, f is separable and easy to prox on.

e. By primal-dual optimality condition

$$x \in \partial g^*(-A^T \mu^\star).$$

Since g is strongly convex, g^* is smooth, hence differentiable. Therefore $\partial g^*(-A^T \mu^\star)$ is a singleton and must be the primal solution.

5. Consider the following constrained optimization problem

$$\underset{x}{\text{minimize}}\, f(x) + \iota_C(x),$$

where constraint qualification holds so that $x^\star$ solves the problem if and only if $0 \in \partial f(x^\star) + \partial \iota_C(x^\star)$. Assume you have solved the dual problem

$$\underset{\mu}{\text{minimize}}(f^*(\mu) + g^*(-\mu))$$

and found the optimal dual variable $\mu^\star = 0$. Show that all primal optimal $x^\star$ are also solutions to the unconstrained minimization problem

$$\text{minimize}\, f(x)$$

using primal-dual optimality conditions and Fermat's rule. (1 p)

Solution

By primal-dual optimality conditions, we have that $x^\star$ is a solution if and only if

$$\mu^\star \in \partial f(x^\star) \qquad \text{and} \qquad -\mu^\star \in \partial \iota_C(x^\star)$$

Now, since $\mu^\star = 0$, this implies $0 \in \partial f(x^\star)$ which is Fermat's rule for minimizing $f(x)$.

6. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is β-smooth for some $\beta > 0$. Show that one step of the gradient method $x_{k+1} = x_k - \gamma \nabla f(x_k)$ gives the following function value decrease guarantee

$$f(x_{k+1}) - f(x_k) \leq -\gamma(1 - \tfrac{\beta\gamma}{2})\|\nabla f(x_k)\|_2^2.$$

Find the γ that gives the largest guaranteed function value decrease at any given iteration. (1 p)

Solution

Using descent lemma for points x_{k+1} and x_k and the gradient method iteration, we get

$$\begin{aligned}
f(x_{k+1}) &\leq f(x_k) + \nabla f(x_k)^T(x_{k+1} - x_k) + \tfrac{\beta}{2}\|x_{k+1} - x_k\|_2^2 \\
&= f(x_k) - \gamma \nabla f(x_k)^T \nabla f(x_k) + \tfrac{\beta\gamma^2}{2}\|\nabla f(x_k)\|_2^2 \\
&= f(x_k) - \gamma(1 - \tfrac{\beta\gamma}{2})\|\nabla f(x_k)\|_2^2
\end{aligned}$$

The function value decrease is upper bounded by

$$-\gamma(1 - \tfrac{\beta\gamma}{2})\|\nabla f(x_k)\|_2^2.$$

Since $\|\nabla f(x_k)\|_2^2$ is fixed, this amounts to minimizing $h(\gamma) := -\gamma(1 - \frac{\beta\gamma}{2}) = -\gamma + \frac{\beta}{2}\gamma^2$. Since h is differentiable and convex in γ (linear plus convex quadratic), the minimum is found at zero gradient: $\nabla h(\gamma) = -1 + \beta\gamma = 0$, i.e., the optimal $\gamma = \frac{1}{\beta}$.