LUND UNIVERSITY

Department of
AUTOMATIC CONTROL

Exam in Optimization for Learning

2020-10-26

Points and grading

All answers must include a clear motivation. Answers should be given in English. The total number of points is 25. The maximum number of points is specified for each subproblem. Preliminary grading scales:

Grade 3: 12 points
4: 17 points
5: 22 points

Accepted aid

All material form the course.

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 whether or not the functions below are convex.

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

$$f_1(x) = \frac{1}{g(x)}$$

for each $x \in \mathbb{R}^n$, where $g : \mathbb{R}^n \to \mathbb{R}$ is concave and satisfies $g(x) > 0$ for each $x \in \mathbb{R}^n$. (1 p)

b. $f_2 : \mathbb{R}^n \to \mathbb{R}$ such that

$$f(x) = \sqrt{x^T L^T L x}$$

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

c. $f_3 : \mathbb{R}^2 \to \mathbb{R}$ such that

$$f_3(x) = \begin{cases} \sqrt{x_1 x_2} & \text{if } x = (x_1, x_2) \in \mathbb{R}^2_{++}, \\ \infty & \text{if } x \in \mathbb{R}^2 \setminus \mathbb{R}^2_{++}. \end{cases}$$

(1 p)

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

$$f_4(x) = \sum_{i=1}^{r} x_{(i)} = x_{(1)} + \ldots + x_{(r)}$$

for each $x \in \mathbb{R}^n$, where r is an integer such that $1 \le r \le n$ and $x_{(i)}$ denote the ith largest component of x, i.e.,

$$x_{(1)} \ge \ldots \ge x_{(n)}.$$

(1 p)

Remark: In each subproblem, you are allowed to assume that norms are convex. This remark also holds for all other problems in this exam.

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

a. $S_1 = \{x \in \mathbb{R}^n : x_1 + \ldots + x_n = 1\}$. (1 p)

b. $S_2 = \{x \in \mathbb{R}^n : \|x - a\|_2 \le \|x - b\|_2\}$, where $a, b \in \mathbb{R}^n$ and $a \ne b$. (1 p)

c. $S_3 = \left\{x \in \mathbb{R}^3 : 2x_1 \ge \sqrt{x_2^2 + x_3^2}\right\}$. (1 p)

d. $S_4 = \left\{x \in \mathbb{R}^2 : 2 \le e^{x_1^2 + x_2^2} \le 4\right\}$. (1 p)

e. $S_5 = \left\{x \in \mathbb{R}^n : x^T y \le 1, \forall y \in C\right\}$, where $C \subseteq \mathbb{R}^n$. (1 p)

3. Consider the convex function $f : \mathbb{R}^n \to \mathbb{R}$ such that

$$f(x) = \left(a^T x - b\right)^2$$

for each $x \in \mathbb{R}^n$, where $a \in \mathbb{R}^n \setminus \{0\}$, $b \in \mathbb{R}$ and $n \ge 2$.

a. Prove or disprove that f is strongly convex. (1 p)

b. Find the conjugate function f^*. (2 p)

c. Let $\gamma > 0$. Find the proximal operator $\mathrm{prox}_{\gamma f}$. (1 p)

4. Let $C \subseteq \mathbb{R}^n$ be a nonempty, closed and convex set. Its support function $\sigma_C : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is defined as

$$\sigma_C(y) = \sup_{x \in C} y^T x$$

for each $y \in \mathbb{R}^n$.

a. Show that the support function σ_C is convex, independent of the convexity of the set C. (1 p)

b. Show that $\sigma_C^* = \iota_C$. (1 p)

c. Find an expression for $\mathrm{prox}_{\gamma \sigma_C}$, where $\gamma > 0$, that involves Π_C, i.e., the (Euclidean) projection onto the set C. (1 p)

5. Consider the problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \; \frac{1}{2}\|Ax - b\|_2^2$$

where $A \in \mathbb{R}^{n \times n}$ satisfies

$$A = \mathbf{diag}(a_1, \ldots, a_n), \quad a_i \neq 0, \quad \forall i \in \{1, \ldots, n\},$$

and $b = (b_1, \ldots, b_n) \in \mathbb{R}^n$. Let $f : \mathbb{R}^n \to \mathbb{R}$ such that

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2$$

for each $x \in \mathbb{R}^n$.

a. Give a closed-form expression of the solution. (0.5 p)

b. Show that $\beta = \max_{i \in \{1,\ldots,n\}} a_i^2$ is a smoothness constant for f. (1 p)

c. Show that $\beta_i = a_i^2$ is a coordinate-wise smoothness constants for f, for each coordinate $i = 1, \ldots, n$. (1 p)

d. Consider the gradient method with step-size $1/\beta$, where β is the smoothness constant in **b.**. Suppose you are given the iterate $x^k \in \mathbb{R}^n$, where $k \in \mathbb{N}_0$ is the iteration number. For each coordinate $i = 1, \ldots, n$, provide the update formula for x_i^k. Utilize that $A = \mathbf{diag}(a_1, \ldots, a_n)$. (1 p)

e. Let $b_i = 0$ for each $i = 1, \ldots, n$ and provide an exact linear convergence rate for each of the coordinates for the gradient method in **d.**. This means, find the $\rho_i \in [0, 1)$ such that

$$\left\|x_i^{k+1}\right\|_2 = \rho_i \left\|x_i^k\right\|_2,$$

for each coordinate $i = 1, \ldots, n$. (Each coordinate will converge linearly to $x_i^\star = 0$ in this case.) (1 p)

f. Now drop the assumption that $b_i = 0$ for each $i = 1, \ldots, n$. Consider the coordinate gradient method (i.e., no proximal operator) with step-sizes $1/\beta_i$, where β_i are the coordinate smoothness constants in **c.**. Provide an update formula for each coordinate $i = 1, \ldots, n$. Utilize that $A = \mathbf{diag}(a_1, \ldots, a_n)$. Show that $x_i^{k+1} = x_i^\star$ with $x_i^\star$ from **a.**, independent on $x^k \in \mathbb{R}^n$. (1 p)

6. Consider minimizing a function $f : \mathbb{R}^n \to \mathbb{R}$, with minimizer $x^\star \in \mathbb{R}^n$, using a stochastic optimization algorithm, starting at some predetermined (deterministic) point $x_0 \in \mathbb{R}^n$. Analysis of the algorithm resulted in the following inequality

$$\mathbb{E}\left[\|x_{k+1} - x^\star\|_2^2 \mid x_k\right] \leq \|x_k - x^\star\|_2^2 - 2\gamma(f(x_k) - f(x^\star)) + \gamma^2 G, \quad \forall k \in \mathbb{N}_0,$$

where G is a deterministic positive constant and γ is a deterministic fixed positive step-size of the algorithm. In particular, $(x_k)_{k \in \mathbb{N}_0}$ is a stochastic process.

a. Apply an expectation to the above inequality to derive a Lyapunov inequality for the algorithm. (1 p)

b. Use the obtained Lyapunov inequality to show that

$$\sum_{i=0}^{k} \mathbb{E}[f(x_i) - f(x^\star)] \leq \frac{\|x_0 - x^\star\|_2^2 + G(k+1)\gamma^2}{2\gamma}, \quad \forall k \in \mathbb{N}_0. \tag{1}$$

(1.5 p)

c. The upper bound (1) goes to infty as $k \to \infty$ unless $G = 0$. Consider the step-size $\gamma = \theta/\sqrt{K+1}$, where $K \in \mathbb{N}_0$ is the total number of iterations we wish to run the algorithm and $\theta > 0$. Show that we get a $\mathcal{O}(1/\sqrt{K+1})$ convergence bound. In particular, show that

$$\min_{i \in \{0,\ldots,K\}} \mathbb{E}[f(x_i) - f(x^\star)] \leq \frac{\|x_0 - x^\star\|_2^2 + G\theta^2}{2\theta\sqrt{K+1}}.$$

(1 p)