LUND UNIVERSITY

Department of
AUTOMATIC CONTROL

Exam in Optimization for Learning

2021-10-26

Grading and points

All answers must include a clear motivation. Answers should be given in English. Number all your solution sheets and indicate the total number of sheets, e.g., 1/12, 2/12 and so on.

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

You are allowed to bring lecture slides. You may use the results in the slides unless the opposite is explicitly stated.

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. $S_1 = \{x \in \mathbb{R} : x \text{ is integer and } x \geq 5\}$. (1 p)

b. $S_2 = \{x \in \mathbb{R}^m : \|x\|_2 \leq 1\}$. (1 p)

c. $S_3 = \{x \in \mathbb{R}^n : Ax + b \in D\}$ where

$$D = \{y \in \mathbb{R}^m : \|y\|_2 \leq 1\},$$

$A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. (1 p)

d. $S_4 = \{x \in \mathbb{R} : f(x) \leq 1\}$ where

$$f(x) = \begin{cases} \cos x & \text{if } 0 \leq x \leq 2\pi, \\ \infty & \text{otherwise} \end{cases}$$

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

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

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

$$f_1(x) = \log\left(1 + \mathrm{e}^{-x^2}\right)$$

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

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

$$f_2(X) = \lambda_{\max}(X)$$

for each $X \in \mathbb{S}^n$, where $\lambda_{\max}$ denotes the largest eigenvalue. (1 p)

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

$$f_3(x) = \sum_{i=1}^{r} \left|x_{\langle i \rangle}\right|$$

for each $x \in \mathbb{R}^n$ where $1 \leq r \leq n$ is an integer and $x_{\langle i \rangle}$ is the component of x with the ith largest absolute value, meaning that

$$\left|x_{\langle 1 \rangle}\right| \geq \ldots \geq \left|x_{\langle n \rangle}\right|.$$

(1 p)

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

$$f_4 = \iota_S$$

where $S \subseteq \mathbb{R}^n$ is given by

$$S = \{x \in \mathbb{R}^n : \|x\|_0 = r\}$$

where $1 \leq r \leq n$ is a fixed integer and

$$\|x\|_0 = \text{ number of nonzero elements in the vector } x$$

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

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

$$f(x) = \begin{cases} -0.5x & \text{if } x \leq 0, \\ 3x & \text{if } x > 0. \end{cases}$$

See Figure 1.

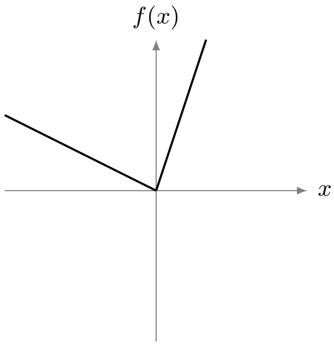

Figure 1 Function f in Problem 3.

a. Compute the subdifferential ∂f (1 p)

b. Compute prox_f (1 p)

c. Compute f^* (1 p)

d. Compute prox_{f^*} (1 p)

4. We will consider the problem of selecting an optimal portfolio of stocks using a mean-variance model. Suppose you wish to invest W SEK, for some $W > 0$, by picking among $n \in \mathbb{N}$ different stocks. Your portfolio of stocks is constructed at present time by purchasing x_i SEK worth of stock i, for each $i = 1, \ldots, n$. The portfolio can be represented by the vector $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Naturally, there is the budget constraint that

$$\mathbf{1}^T x = W,$$

i.e., the sum of the investments equals the investment budget W. We denote the set of feasible portfolios by

$$B = \{x \in \mathbb{R}^n : \mathbf{1}^T x = W\}.$$

Note that we allow x to have negative components. A negative component x_i corresponds to short-selling stock i, i.e., borrowing the stock and immediately selling it. The portfolio of stocks is held constant until some predetermined time in the future when all investments are liquidated (sold). This corresponds to a one-period investment problem. Let r be n-dimensional, where r_i is the return of stock i over the period. In order to model our uncertainty of the future stock returns, we let r be a n-dimensional random variable with known expected value $\mathbb{E}[r] = \mu \in \mathbb{R}^n$ and known covariance matrix $\text{Var}[r] = \Sigma \in \mathbb{S}^n_{++}$, i.e., Σ is a real-valued positive definite $n \times n$ matrix. The return of the portfolio, the expected return of the portfolio, and the variance of the return of the portfolio are given by

$$r^T x, \qquad \mathbb{E}\left[r^T x\right] = \mu^T x \quad \text{and} \quad \mathrm{Var}\left[r^T x\right] = x^T \Sigma x,$$

respectively. In the mean-variance model we seek the portfolio x that solves the optimization problem

$$\underset{x \in B}{\text{minimize}} \, -\mu^T x + \gamma x^T \Sigma x = \underset{x \in \mathbb{R}^n}{\text{minimize}} \, -\mu^T x + \gamma x^T \Sigma x + \iota_B(x) \tag{1}$$

where $\gamma > 0$ is given. The variance of the return of the portfolio $x^T \Sigma x$ is a proxy for the risk inherent in the investment. Therefore, γ is usually called the risk aversion parameter and is an inverse measure of an investors risk appetite. For future reference, we define the function $f : \mathbb{R}^n \to \mathbb{R}$ such that

$$f(x) = -\mu^T x + \gamma x^T \Sigma x$$

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

a. Prove that f is strongly convex. (0.5 p)

b. Prove that B is convex. What does this imply for ι_B? (0.5 p)

c. Why does optimization problem (1) have a unique minimizer? (0.5 p)

d. Compute the subdifferential ∂f. (0.5 p)

e. Show that

$$\partial \iota_B(x) = \begin{cases} \{\alpha \mathbf{1} : \alpha \in \mathbb{R}\} & \text{if } x \in B, \\ \emptyset & \text{if } x \notin B \end{cases}$$

for each $x \in \mathbb{R}^n$. (1.5+0.5 p)

f. Using the subdifferentials in **d.** and **e.**, find the optimal portfolio according to the mean-variance model (1).
(You may assume that the expression for $\partial \iota_B$ in **e.** holds.) (2 p)

g. Show that the conjugate functions of f and ι_B satisfy

$$f^*(s) = \frac{1}{4\gamma}(s + \mu)^T \Sigma^{-1} (s + \mu)$$

for each $s \in \mathbb{R}^n$ and

$$\iota_B^*(s) = \begin{cases} \alpha W & \text{if } s = \alpha \mathbf{1} \text{ for some } \alpha \in \mathbb{R}, \\ \infty & \text{otherwise} \end{cases}$$

for each $s \in \mathbb{R}^n$, respectively. (1+1 p)

h. State the dual problem

$$\underset{s \in \mathbb{R}^n}{\text{minimize}} \, f^*(-s) + \iota_B^*(s) \tag{2}$$

to problem (1) and express it as an optimization problem over a single real variable. (You may assume that the expressions for f^* and ι_B^* in **g.** hold.)

Solve the dual problem (2) over that single variable and relate it to the optimal α in **f.** that comes from the subdifferential in **e.**. Give the dual optimal point $s^* \in \mathbb{R}^n$. (1 p)

i. Given the optimal point $s^* \in \mathbb{R}^n$ of the dual problem (2) in **h.**, show how to recover the primal solution, i.e., the solution to (1). You are allowed to directly use any one of the primal dual necessary and sufficient optimality conditions. Show that the recovered primal solution is the same as in **f.**. (1 p)

5. Consider the 1-norm regularized SVM problem

$$\underset{w \in \mathbb{R}^n}{\text{minimize}} \sum_{i=1}^{N} \underbrace{\max\left(0, 1 - y_i w^T x_i\right)}_{=f_i(w)} + \lambda \|w\|_1 \tag{3}$$

given the labeled training data set $\{(x_i, y_i)\}_{i=1}^N$, where $x_i \in \mathbb{R}^n$ and $y_i \in \{-1, 1\}$ are training data and labels, respectively.

a. Find the smallest nonnegative constant $\lambda_0 \in \mathbb{R}$ such that if $\lambda \geq \lambda_0$, then

$$w = 0$$

is an optimal point for (3). (2 p)

b. Is the proximal gradient method applicable to find a solution of problem (3)? Is it applicable to solve a corresponding Fenchel dual problem? (1 p)