ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

IFOR
Institute for Operations Research
Department of Mathematics

Convex Optimization

Mock Exam

May 22nd, 2022

Student ID:	

General remarks.

- Similar to the below remarks will also appear on the actual exam and do not apply to this mock exam.
- Place your student identity card on the table.
- This is a closed book exam. No extra material is allowed (except for a subject-neutral dictionary).
- Use a pen to write down your solutions. Pencils and red pens are not allowed.
- Turn off all your technical devices and put them away.
- Write your solutions directly after each exercise, if you need more space use the extra sheets at the end. Extra paper will be distributed if needed.
- You have 3 hours to solve the exam. 15 minutes before the end of the exam, no premature submission is allowed anymore.
- For all problems, provide a complete solution in English, including all explanations and implications in a mathematically clear and well-structured way. Please hand in a readable and clean solution. Cross out any invalid solution attempts.
- Unless you are explicitly asked to prove them, you may use results from the lecture without proof (provided that you reference or state them clearly). Results from the problem sets may only be used if you prove them.
- Ask any question that you might have immediately and during the exam.

Good luck!

Problem:	1	2	3	4	5	6	7	Total
Points:	/ 6	/ 10	/ 10	/ 10	/ 6	/ 10	/ 8	/ 60

Exercise 1 (6 points)

Prove that the nonnegative orthant in $\mathbb{R}^n$,

$$K = \{\, x \in \mathbb{R}^n \mid x_i \geq 0 \text{ for } i \in \{\, 1, \ldots, n \,\} \,\}$$

is a self dual cone.

Exercise 2 (10 points)

Handwritten digit recognition with the MNIST dataset is nowadays mainly known as an introductory example for Convolutional Neural Networks. A basic approach to tackle it, is however based on Convex Programs.
The general goal in handwritten digit recognition is: given a greyscale image (28×28 pixels in this case) that represents a handwritten digit, decide which digit $d \in \{ 0, \dots, 9 \}$ it represents.
In order to learn how numbers are written, we are given a finite training set $P = \{ (x, l) \} \subseteq \mathbb{R}^{784} \times \{ 0, \dots, 9 \}$ where x is a given training image and l the digit it represents. Here we interpret each 28×28 image as a 784-dimensional vector by ordering the pixels in a fixed way. Now the goal is to find a matrix $D \in \mathbb{R}^{10x784}$ for which $y = Dx$ tells us which image x represents. An approach commonly used is to interpret y_i, for $i \in \{ 0, \dots, 9 \}$, as the probability of x representing i. This does however not work immediately since y may have negative entries and the sum of entries might not be 1. Hence we need to normalize y with a function $z_i(y)$ such that $z_i(y) > 0$ and $\sum_{i=0}^{9} z_i(y) = 1$. To measure how *well* our matrix D is at recognizing digits, we use the loss function

$$\ell(D) = - \sum_{(x,l) \in P} \ln \left(z_l(Dx) \right).$$

which penalizes assigning a low probability to the correct value l.

Given this information,

1. prove that $z_j : \mathbb{R}^{10} \to \mathbb{R}, z_j(y) := \frac{e^{y_j}}{\sum_{i=0}^{9} e^{y_i}}$ satisfies $z_j(y) \in (0, 1)$ and $\sum_{i=0}^{9} z_i(y) = 1$ for all $y \in \mathbb{R}^{10}$ and $i \in \{ 0, \dots, 9 \}$ **(2 points)**.
2. prove that ℓ as a function $\mathbb{R}^{7840} \to (0, \infty)$ is convex for your choice of z_i or adapt z_i accordingly. **(6 points)**
3. sum up that the problem can be solved with a Convex Optimization Problem **(2 points)**.

Exercise 3 (10 points)

Consider the mathematical problem

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \ & f(x) \\ \text{s.t. } & Ax \leq b, \end{aligned}$$

where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. Derive the explicit dual problem for the given $f\colon \mathbb{R}^n \to \mathbb{R}$.

1. $f(x) = \frac{1}{2}x^\top Cx$ for a given symmetric positive definite $C \in \mathbb{R}^{n\times n}$ (Quadratic Program) **(5 points)**.
2. $f(x) = c^\top x + \epsilon \operatorname{lb}(x)$ where $c \in \mathbb{R}^n$ and $\epsilon > 0$. Here the function $\operatorname{lb}\colon \mathbb{R}^n \to (-\infty, \infty]$ is the so called log barrier,

$$\operatorname{lb}(x) := \begin{cases} -\sum_{j=1}^n \log(x_j) & \text{, if } x > 0 \\ +\infty & \text{, otherwise} \end{cases}$$

(5 points).

Exercise 4 (10 points)

Consider the general descent method of the form as introduced in the lecture, for continuous differentiable

Input : Starting point $x^{(0)} \in \mathbb{R}^n$
Output: A sequence of iterations $\left(x^{(k)}\right) \subseteq \mathbb{R}^n$
$k \leftarrow 0$
repeat
  Choose a descent direction $\Delta x^{(k)}$, i.e. $\nabla f\left(x^{(k)}\right)^\top \Delta x^{(k)} < 0$
  Choose a descent step $t^{(k)}$ such that $f\left(x^{(k)} + t^{(k)}\Delta x^{(k)}\right) < f\left(x^{(k)}\right)$
  $x^{(k+1)} \leftarrow x^{(k)} + t^{(k)}\Delta x^{(k)}$
  $k \leftarrow k+1$
until $\nabla f\left(x^{(k)}\right) = 0$;
return $(x^{(k)})$,

Algorithm 1: Descent method

$f\colon \mathbb{R}^n \to \mathbb{R}$. Our goal will be to find stationary points of f which in some cases (e.g. when f is convex) are also minimas. It is clear that this is achieved when Algorithm 1 stops after a finite number of iterations. So we will suppose that $\left(x^{(k)}\right)_{k\in\mathbb{N}}, \left(\Delta x^{(k)}\right)_{k\in\mathbb{N}}, \left(t^{(k)}\right)_{k\in\mathbb{N}}$ are infinite sequences. In this case we call a subsequence $\left(t^{(k)}\right)_{k\in K}$ of stepsizes *admissible*, if

$$i)\ \forall k \in \mathbb{N}\colon f\left(x^{(k)} + t^{(k)}\Delta x^{(k)}\right) \le f\left(x^{(k)}\right),$$

$$ii)\ f\left(x^{(k)} + t^{(k)}\Delta x^{(k)}\right) - f\left(x^{(k)}\right) \to 0 \ \Rightarrow\ \left(\frac{\nabla f\left(x^{(k)}\right)^\top \Delta x^{(k)}}{\left\|\Delta x^{(k)}\right\|}\right)_{k\in K} \to 0.$$

A subsequence $\left(\Delta x^{(k)}\right)_{k\in K}$ of descent directions is called *admissible*, if

$$i)\ \forall k \in \mathbb{N}\colon \nabla f\left(x^{(k)}\right)^\top \Delta x^{(k)} < 0,$$

$$ii)\ \left(\frac{\nabla f\left(x^{(k)}\right)^\top \Delta x^{(k)}}{\left\|\Delta x^{(k)}\right\|}\right)_{t\in K} \to 0 \ \Rightarrow\ \left(\nabla f\left(x^{(k)}\right)\right)_{t\in K} \to 0.$$

Prove to following statement.

a) Suppose $\bar{x}$ is a limit point of $\left(x^{(k)}\right)$, i.e. there exists a subsequence $\left(x^{(k)}\right)_{k\in K} \to \bar{x}$. Furthermore suppose that the subsequences $\left(t^{(k)}\right)_{k\in K}, \left(\Delta x^{(k)}\right)_{k\in K}$ indexed by the same index set are admissible. Then $\bar{x}$ is a stationary point **(6 points)**.

Now we want to find rules which imply admissibility. Therefore we say $\left(\Delta x^{(k)}\right)_{k\in K}$ fulfils an *angle inequality*, if there exists a $\eta \in (0,1)$ such that

$$\cos\angle\left(-\nabla f\left(x^{(k)}\right), \Delta x^{(k)}\right) = \frac{-\nabla f\left(x^{(k)}\right)^\top \Delta x^{(k)}}{\left\|\nabla f\left(x^{(k)}\right)\right\| \left\|\Delta x^{(k)}\right\|} \ge \eta$$

for all $k \in K$.

b) Prove that $\left(\Delta x^{(k)}\right)_{k\in K}$ is admissible if it is generated by Algorithm 1 and fulfils the angle inequality **(2 points)**.

Finally we call stepsizes $\left(t^{(k)}\right)_{k\in K}$ *efficient*, if there exists a $\theta > 0$ such that

$$f\left(x^{(k)} + t^{(k)}\Delta x^{(k)}\right) \leq f\left(x^{(k)}\right) - \theta\left(\frac{\nabla f\left(x^{(k)}\right)^{\top}\Delta x^{(k)}}{\left\|\Delta x^{(k)}\right\|}\right)^2$$

for all $k \in K$.

c) Prove that $\left(t^{(k)}\right)_{k\in K}$ is admissible if it is generated by Algorithm 1 and efficient **(2 points)**.

Exercise 5 (6 points)

Consider a problem of managing a portfolio of $n \in \mathbb{N}_{\geq 1}$ assets over a given period of time (e.g., for one month). Given a budget $B > 0$ the goal is to decide which assets we should buy for this period. Consider a setting where the change in price of asset $i \in [n]$ is a random variable $p_i \in \mathbb{R}$ (i.e. if the actual price triples during our period, the corresponding realization of p_i would be 3). Let the variable $x_i \in [0, 1]$ describe the fraction of the budget B devoted to buying asset i. Then the mean value of the return is $\overline{p}^\top x$ and the variance is $x^\top \Sigma x$. The choice of the assets to buy requires a balance between maximizing the expected return and minimizing the risk, that is, the variance. In the classical Markowitz portfolio optimization problem, the objective is to minimize the risk for a given fixed return value r at the end of the period. This can be modelled with the following program:

$$
\begin{aligned}
\min\ & x^\top \Sigma x \\
& \overline{p}^\top x \geq r \\
& \frac{1}{B}\mathbf{1}^\top x = 1 \\
& x \geq 0 \qquad \text{for } i \in [n]
\end{aligned}
$$

which is a quadratic program.

Extend this to take transaction costs into account. Let $s \in \mathbb{R}_{\geq 0}$ and $b \in \mathbb{R}_{\geq 0}$ be constants describing the selling and buying fee of a single unit of an asset, respectively. Given an initial portfolio, $x_0 \in \mathbb{R}^n$ and the transaction costs s and b (we only sell and buy assets at the beginning of the period), build a QP that minimizes the risk for a given return $r \in \mathbb{R}$ after the end of the period.

Exercise 6 (10 points)

Let $A_0, \dots, A_n \in \mathcal{S}^m$ and for $x \in \mathbb{R}^n$ define

$$A(x) \coloneqq A_0 + x_1 A_1 + \cdots + x_n A_n \in \mathcal{S}^m.$$

We order the eigenvalue of $A(x)$ in decreasing order, i.e. $\lambda_1(x) \geq \cdots \geq \lambda_m(x)$. Decide if the following problems can be formulated as a SDP for any $A_0, \dots, A_n$ and give a short proof or counterexample respectively.

1. **(3 points)** Minimize the largest eigenvalue $\lambda_1(x)$, i.e.

$$\min_{x \in \mathbb{R}^n} \lambda_1(x).$$

2. **(3 points)** Minimize the smallest eigenvalue $\lambda_m(x)$, i.e.

$$\min_{x \in \mathbb{R}^n} \lambda_m(x).$$

3. **(4 points)** Minimize the spread between the largest and smallest eigenvalue, i.e.

$$\min_{x \in \mathbb{R}^n} \lambda_1(x) - \lambda_m(x).$$

Exercise 7 (8 points)

Decide which of the following functions $f_i \colon \mathbb{R} \to \mathbb{R}_{>0}$ are log-convex (**T**) or not (**F**).

T	**F**	
☐	☐	$f_1(x) = \log(p) + \lvert x\rvert^p$ for $p \in [2, \infty)$ **(2 points)**.
☐	☐	$f_2(x) = \frac{1}{\lvert x\rvert^p}$ for $x \in (0, \infty)$ and $p > 0$ **(2 points)**.

Decide which of the following functions $f_i \colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ are log-log convex (**T**) or not (**F**).

T	**F**	
☐	☐	$f_3(x) = \max\left\{ x, \frac{1}{x} \right\}$ **(2 points)**.
☐	☐	$f_4(x) = \tan^{-1}(x)$ **(2 points)**.

Extra sheets you may need