Exam Continuous Optimization

18 January 2021, 14.00–17.00

The exam consists of 4 questions. In total you can obtain 90 points. The final grade is $1 + \#\text{points}/10$ rounded to the nearest integer.

This is an open-book exam. It is NOT allowed to discuss with anyone else. If you have any questions regarding the exam, or technical questions regarding uploading of your answer, please contact David de Laat at d.delaat@tudelft.nl.

Please review the instructions posted on the announcement page for the course. The most important points are repeated below:

- Write your answers **by hand** and start each exercise on a new sheet.
- On your first answer sheet, you should write the following statement: "This exam will be solely undertaken by myself, without any assistance from others, and without use of sources other than those allowed."
- When scanning your work place your student ID on the first page. If you do not have a student ID please use some other form of identification but in that case make sure only your name and photo are visible.
- Scan your work and submit it as **one single pdf-file** at 17.00.
- You should keep an eye on your email from 17.00-18.00 because you can be asked to join the zoom call for a random check.

Good luck!

1. Consider the optimization problem

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le 0 \text{ for } i = 1, \ldots, m, \\ & Ax = b. \end{aligned}$$

where $f_0, \ldots, f_m$ are convex and twice-continuously differentiable on $\mathbb{R}^n$.

(a) (6 points) Show Slater's condition holds if there exist feasible points $x_1, \ldots, x_m \in \mathbb{R}^n$ with $f_i(x_i) < 0$ for $i = 1, \ldots, m$.

(b) (6 points) Use the second-order condition for convexity to show that the barrier functional

$$\phi(x) = -\sum_{i=1}^{m} \log(-f_i(x))$$

is convex.

(c) (6 points) The barrier (or centralizer) problem for a given t is defined as

$$\begin{aligned} \text{minimize} \quad & t f_0(x) + \phi(x) \\ \text{subject to} \quad & Ax = b. \end{aligned}$$

Write down the Lagrangian and the KKT conditions for this problem.

(d) (6 points) The optimal solution to the barrier (or centralizer) problem is denoted by $x^*(t)$ and for $t > 0$ these solutions form a path called the central path. Explain how the tangent vector $dx^*(t)/dt$ to the central path can be computed. (Hint: use the KKT conditions form (c).)

2. Consider the unconstrained optimization problem

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

where $f \colon \mathbb{R}^n \to \mathbb{R}$ is strongly convex and continuously differentiable on $\mathbb{R}^n$.

(a) (6 points) Give an example of a strongly convex function f with $n = 2$ for which gradient descent performs badly but Newton's method works well. Why does gradient descent perform badly for this example? Why does Newton's method work well for this example?

(b) (6 points) Consider the norm $\| \cdot \|$ defined by

$$\|x\| = 2\|x\|_2.$$

Express the steepest descent direction Δx_{sd} and the normalized steepest descent direction Δx_{nsd} explicitly in terms of the gradient $\nabla f(x)$.

(c) (6 points) Explain whether or not steepest descent using the above norm is the same as gradient descent when exact line search is used. And what about when backtracking line search is used?

(d) (6 points) Suppose $n = 2$ and $f(x) = x_1^2 + x_2^2 - \cos(x_1)$. Find the gradient, Hessian, and Newton step Δx_{nt} at the point $(0, 1)$.

3. Consider the 1-dimensional optimization problem

$$\begin{aligned} \text{minimize} \quad & (x-2)^2 \\ \text{subject to} \quad & 0 \le x \le 5. \end{aligned}$$

(a) (6 points) Explain why the objective function is unimodal (according to the definition of unimodal we used in the lecture).

(b) (6 points) Suppose we apply Fibonacci line search with initial bracket $[0, 5]$. What is the bracket after 4 function evaluations? Make a sketch to support your answer.

(c) (3 points) How many iterations does quadratic fit search need to find the minimum? Explain your answer.

(d) (9 points) Find the Lagrangian, Lagrange dual function, and Lagrange dual problem.

4. Let $(x_1, y_1), \ldots, (x_N, y_N) \in \mathbb{R}^n \times \{-1, 1\}$ be a training set and $\gamma > 0$ a parameter. Consider the support vector problem

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|a\|_2^2 + \gamma 1^\mathsf{T} u \\ \text{subject to} \quad & y_i(a^\mathsf{T} x_i + b) \geq 1 - u_i \text{ for } i = 1, \ldots, N, \\ & u \geq 0 \end{aligned}$$

with optimal solution (a^*, b^*, u^*).

(a) (3 points) Does strong duality hold? Explain why or why not.

(b) (3 points) How do we use the solution (a^*, b^*, u^*) to decide to which class ($+1$ or -1) a new point $z \in \mathbb{R}^n$ belongs?

(c) (6 points) Explain why we have the terms $\frac{1}{2}\|a\|_2^2$ and $\gamma 1^\mathsf{T} u$ in the objective. What are these terms achieving in relation to the hyperplane and slab around the hyperplane defined by a and b? What happens when the parameter γ is very large?

(d) (6 points) Suppose that for a given i and given dual optimal solution, the dual variables corresponding to the constraints $y_i(a^\mathsf{T} x_i + b) \geq 1 - u_i$ and $u_i \geq 0$ are both nonzero. What does this say about x_i in relation to the hyperplane and slab around the hyperplane defined by (a^*, b^*)?

End of test