Re-exam Continuous Optimization

22 February 2021, 14.00–17.00

The exam consists of 5 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 17.00-17.30 because you can be asked to join the zoom call for a random check.

Good luck!

1. Let $\alpha > 0$ and consider the function f defined by

$$f(x_1, x_2) = x_1^2 + 2x_2^2 + \alpha \sin(x_2).$$

(a) (6 points) Find the gradient descent search direction Δx at $x = (1, 0)$.

Solution: We have

$$\nabla f(x) = \begin{pmatrix} 2x_1 \\ 4x_2 + \alpha \cos(x_2) \end{pmatrix}$$

so

$$\Delta x = -\nabla f(1, 0) = -\begin{pmatrix} 2 \\ \alpha \end{pmatrix} = \begin{pmatrix} -2 \\ -\alpha \end{pmatrix}.$$

(b) (6 points) For which values of α is the function f convex?

Solution: Since the domain of f is all of $\mathbb{R}^2$, by the second-order condition for convexity f is convex if and only if $\nabla^2 f(x)$ is positive semidefinite for all x. We have

$$\nabla^2 f(x) = \begin{pmatrix} 2 & 0 \\ 0 & 4 - \alpha \sin(x_2) \end{pmatrix}$$

so

$$\nabla^2 f(x) \succeq 0$$

for all $x \in \mathbb{R}^2$ if and only if $-4 \leq \alpha \leq 4$.

(c) (6 points) Suppose $\alpha = 1$. What is the next iterate after one step of Newton's method starting from $x = (1, 0)$?

Solution: We have

$$\Delta x_{\text{nt}} = -\nabla^2 f(1, 0)^{-1} \nabla f(1, 0) = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}^{-1} \begin{pmatrix} -2 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1/4 \end{pmatrix}.$$

Afrer applying one step of the (pure) Newton method we get

$$x = (1, 0) + (-1, -1/4) = (0, -1/4).$$

(With backtracking we get the same solution.)

2. Let n be a positive integer and consider the optimization problem

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{n} x_i^2 \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i \geq 1. \end{aligned}$$

(a) (12 points) Derive the Lagrangian, Lagrange dual function, and Lagrange dual problem.

Solution: The Lagrangian is

$$L(x, \lambda) = \sum_{i=1}^{n} x_i^2 + \lambda \left(1 - \sum_{i=1}^{n} x_i\right).$$

We have

$$\nabla_x L(x, \lambda) = 2x - \lambda 1$$

so the minimum of $L(x, \lambda)$ over x is attained at $x = \frac{\lambda}{2} 1$, and

$$g(\lambda) = \inf_x L(x, \lambda) = \frac{n\lambda^2}{4} + \lambda(1 - \frac{n\lambda}{2}) = \lambda - \frac{n\lambda^2}{4}.$$

The Lagrange dual problem is

$$\begin{aligned} \text{maximize} \quad & \lambda - \frac{n\lambda^2}{4} \\ \text{subject to} \quad & \lambda \geq 0. \end{aligned}$$

(b) (6 points) Give optimal primal and dual solutions and show they are optimal.

Solution: Let $x = \frac{1}{n} 1$ and $\lambda = 2/n$. Then the primal and dual objective values are both $1/n$, so by weak duality both solutions are optimal.

3. (a) (12 points) Suppose

$$F(x) = f(g(x) + h(x))$$

with $f(x) = x^2$, $g(x) = 1/x$, and $h(x) = x^3$. Show how $F'(1)$ can be computed using reverse-mode automatic differentiation (a.k.a. backpropagation). Show two separate diagrams for the forward and backward phases.

(b) (6 points) Explain why reverse-mode automatic differentiation can be much faster than computing and evaluating the symbolic derivative even for problems with only 1 variable.

Solution: Consider for instance the function $F(x) = f_1(\ldots(f_n(x))\ldots)$. Evaluating $F'(1)$ when $F'(x)$ is expressed by its symbolic derivative needs roughly $\sum_{i=1}^{n} i \approx n^2$ evaluations of the functions $f_1, \ldots, f_n$ and their derivatives. Reverse-mode automatic differentiation needs only $2n$ such evaluations.

4. Consider the primal-dual interior point method as discussed in class (and in the book) with the parameter $\mu = 2$. We apply this to the problem

$$\begin{aligned} \text{minimize} \ & x_1^4 + x_2 \\ \text{subject to} \ & 1 - x_1 - x_2 \leq 0. \end{aligned}$$

Suppose the the current primal-dual iterate is (x, λ) with $x = (1, 1)$ and $\lambda = 1$.

(a) (6 points) Compute the surrogate duality gap. Explain why this implies x and λ as given above cannot both be optimal.

Solution: We have

$$\hat{\eta}(x, \lambda) = -(1 - x_1 - x_2)\lambda = 1.$$

Since strong duality holds (easily checked using Slater's condition), complementary slackness implies that if x and λ would be optimal, then $\hat{\eta}(x, \lambda) = 0$.

(b) (9 points) Compute the primal-dual search direction $\Delta y_{\text{pd}} = (\Delta x_{\text{pd}}, \Delta \lambda_{\text{pd}})$.

Solution: We have

$$t = \frac{\mu m}{\hat{\eta}} = \frac{2}{1} = 2,$$

so

$$r_{\text{dual}} = \nabla f_0(x) + Df(x)^{\mathsf{T}} \lambda = \begin{pmatrix} 4 \\ 1 \end{pmatrix} + \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

and

$$r_{\text{cent}} = -\text{diag}(\lambda) f(x) - \frac{1}{t} 1 = -\lambda(1 - x_1 - x_2) - \frac{1}{t} = 1 - \frac{1}{2} = \frac{1}{2}.$$

The system

$$\begin{pmatrix} \nabla^2 f_0(x) + \lambda \nabla^2 f_1(x) & Df(x)^{\mathsf{T}} \\ -\lambda Df(x) & -f_1(x) \end{pmatrix} \begin{pmatrix} \Delta x_{\text{pd}} \\ \Delta \lambda_{\text{pd}} \end{pmatrix} = -\begin{pmatrix} r_{\text{dual}} \\ r_{\text{cent}} \end{pmatrix}$$

reduces to

$$\begin{pmatrix} 12 & 0 & -1 \\ 0 & 0 & -1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \Delta x_{\text{pd}} \\ \Delta \lambda_{\text{pd}} \end{pmatrix} = -\begin{pmatrix} 3 \\ 0 \\ 1/2 \end{pmatrix},$$

with solution

$$\Delta x_{\text{pd}} = (-1/4, -1/4) \quad \text{and} \quad \Delta \lambda_{\text{pd}} = 0.$$

(c) (3 points) Show that Δx_{pd} is a primal descent direction and explain the value of $\Delta \lambda_{\text{pd}}$.

Solution: We have

$$-\nabla f_0(x)^{\mathsf{T}} \Delta x_{\text{pd}} = -\begin{pmatrix} 4 & 1 \end{pmatrix} \begin{pmatrix} -1/4 \\ -1/4 \end{pmatrix} = 5/4 > 0.$$

Furthermore, $\Delta \lambda_{\text{pd}} = 0$ because $\lambda = 1$ is already dual optimal (which follows immediately from the stationarity condition).

5. Let $Q \in \mathbb{R}^{n\times n}$ be a symmetric matrix and consider the nonconvex optimization problem

$$\begin{aligned} \text{minimize} \quad & x^\mathsf{T} Q x \\ \text{subject to} \quad & \|x\|_2^2 = 1. \end{aligned}$$

(a) (6 points) Express the optimal value of this problem in terms of the eigenvalue(s) of Q.

> **Solution:** This problem can be solved in many different ways. Easiest way is probably using the stationarity condition.

(b) (6 points) We apply the penalty method to obtain the unconstrained problem

$$\text{minimize } x^\mathsf{T} Q x + \alpha(\|x\|_2^2 - 1)^2$$

where $\alpha > 0$ is the penalty parameter. Show the optimal solution is an eigenvector of Q.

> **Solution:** Let $f(x) = x^\mathsf{T} Q x + \alpha(\|x\|_2^2 - 1)^2$. Since $f(x) \to \infty$ as $\|x\|_2 \to \infty$, the minimum is attained at a critical point. We have
>
> $$\nabla f(x) = 2Qx + 4\alpha(\|x\|_2^2 - 1)x,$$
>
> so the critical points are eigenvectors.

(c) (6 points) Show that as $\alpha \to \infty$, the vector x_α^* converges to a feasible solution of the original constrained problem by showing that the following inequalities hold:

$$\frac{\lambda_{\min}(Q)}{2\alpha} \leq 1 - \|x_\alpha^*\|_2 \leq \frac{\lambda_{\max}(Q)}{2\alpha}.$$

Here $\lambda_{\max}(Q)$ and $\lambda_{\min}(Q)$ are the largest and smallest eigenvalues of Q.

> **Solution:** We have
>
> $$Qx_\alpha^* = -2\alpha(\|x_\alpha^*\|_2^2 - 1)x_\alpha^* = 2\alpha(1 - \|x_\alpha^*\|_2^2)x_\alpha^*.$$
>
> This shows
>
> $$\lambda_{\min}(Q) \leq 2\alpha(1 - \|x_\alpha^*\|_2^2) \leq \lambda_{\max}(Q)$$
>
> so
>
> $$\frac{\lambda_{\min}(Q)}{2\alpha} \leq 1 - \|x_\alpha^*\|_2 \leq \frac{\lambda_{\max}(Q)}{2\alpha}.$$