Math 164-1: Optimization

Instructor: Alpár R. Mészáros

Second Midterm, May 18, 2016

Name (use a pen):

Student ID (use a pen):

Signature (use a pen):

Rules:

- Duration of the exam: **50 minutes**.
- By writing your name and signature on this exam paper, you attest that you are the person indicated and will adhere to the UCLA Student Conduct Code.
- You may use either a pen or a pencil to write your solutions. However, if you use a pencil I will withheld your paper for **two** weeks after grading it.
- **No** calculators, computers, cell phones (all the cell phones should be turned off during the exam), notes, books or other outside material are permitted on this exam. If you want to use scratch paper, you should ask for it from one of the supervisors. Do not use your own scratch paper!
- Please justify all your answers with mathematical precision and write rigorous and clear proofs. You may loose points in the lack of justification of your answers.
- Theorems from the lectures and homework may be used in order to justify your solution. In this case state the theorem you are using.
- This exam has 3 problems and is worth **20 points**. Adding up the indicated points you can observe that there are **28 points**, which means that there are **8 "bonus" points**. This permits to obtain the highest score 20, even if you do not answer some of the questions. On the other hand nobody can be bored during the exam. All scores higher than 20 will be considered as 20 in the gradebook.
- I wish you success!

Problem	Score
Exercise 1	
Exercise 2	
Exercise 3	
Total	

Exercise 1 (11 points)**.**

Let $n \geq 1$ be an integer and let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix (non necessarily positive definite) for which all of its eigenvalues are non-zero. Let $a \in \mathbb{R}^n$ be a given vector and we consider the function $f : \mathbb{R}^n \to \mathbb{R}$, defined as

$$f(x) = \frac{1}{2}(x-a)^\top A^2(x-a),$$

where $A^2 = AA$.

(1) Using first and second order optimality conditions show that f has a unique global minimizer on $\mathbb{R}^n$ and determine this optimizer. Denote it by x^*.

(2) Write the updates in the steepest descent algorithm (i.e. gradient descent with optimal step size) starting from a point $x^0 \in \mathbb{R}^n$ to approximate the optimizer x^* of f that has been determined in (1). Determine the step size α_k in each step.

(3) Imagine that one wants to use a fixed step gradient algorithm too, to approximate x^*. Which is maximal range for the step size α in terms of the eigenvalues of A that ensures global convergence for the algorithm?

(4) Give an example of $A \in \mathbb{R}^{2\times 2}$ diagonal matrix that has a zero and a non-zero eigenvalue. Take $a \in \mathbb{R}^2$. Determine the global minimizers of f in $\mathbb{R}^2$ in this case. What can we say about the uniqueness of them?

(5) Explain what will happen if we want to proceed with a fixed step size gradient algorithm for (4). Does an algorithm like this converge globally? If yes, for which values of the step size α and to which limit point x^*?

(6) Explain what is the major difference between the cases when A has at least one zero eigenvalue and when it does not, from the point of view of the gradient descent algorithms.

Exercise 2 (8 points).

Let us consider the function $f : (0, +\infty) \to \mathbb{R}$ defined as

$$f(x) = x - \ln(x),$$

where ln denotes the natural logarithm of base e.

(1) Using eventually first and second order optimality conditions, show that f has a unique minimizer on $(0, +\infty)$. Denote this by x^*.

In what follows, we are aiming to approximate x^* from (1) using Newton's algorithm.

(2) Write the updates in Newton's algorithm used to approximate the minimizer of f above. Denote the sequence of iterates by $(x^k)_{k\geq 0}$. Determine the biggest range for the initial guess $x^0 > 0$ for which one has after one iteration that $x^1 > 0$. Denote this range by I.

(3) Let $\varepsilon > 0$ be a given error term. Explain why is the condition $|1-x^k| \leq \varepsilon$ a good stopping condition for Newton's algorithm approximating x^*.

(4) Show that for all $x^0 \in I$ (where I is determined in (2)) the sequence $(x^k)_{k\geq 0}$ is converging to x^*. *Hint:* compute for instance the error $1-x^1$ in terms of x^0, then write this relation also for x^{k+1} and x^k. Other possibilities, like showing directly that $|x^k - x^*| \to 0$ as $k \to +\infty$ can be also considered.

(5) Propose a modification of the above algorithm that will ensure that it is converging also if $x^0 \in (0, +\infty) \setminus I$. *Hint:* you may think to introduce a step size in the algorithm, which is exactly 1 in the usual Newton algorithm.

Exercise 3 (9 points)**.**

We aim to compute an approximation of $\sqrt{2}$. For this, we construct a sequence $(x^k)_{k\geq 0}$ that converges to one of the solutions of the equation $x^2 = 2$.

(1) Suppose we have two initial guesses $x^0, x^1 \in \mathbb{R}$. Write down the definition of the sequence $(x^k)_{k\geq 0}$ constructed by the *secant method*. Write the formula in a compact form.

(2) Setting $x^0 = 0$ and $x^1 = 1$, compute x^2, x^3 and x^4 using the algorithm given in (1). What do you observe?

(3) Give two initial guesses x^0 and x^1 for which the sequence constructed in (1) tends to converge to $-\sqrt{2}$ instead. Justify your choice.

(4) Explain analytically and geometrically the behavior of the algorithm described in (1), when one chooses $x^0 = a$ and $x^1 = -a$ for some $a \in \mathbb{R}$ as initial guesses.

(5) Give a sufficient condition for the initial guesses x^0 and x^1 (discuss also the geometrical intuition behind it) that guarantees that the algorithm described in (1) has a tendance to converge to $\sqrt{2}$.

Notice: by the notion of *tendance of convergence* we mean that we have a strong belief that the algorithm converges and this is supported by a couple of iterations and the geometrical intuition behind. You *do not* need to show actual convergences in this exercise!