Deadline: Sun Nov 19, 2023, 8:00 am Submit single unzipped PDF file on learn-web course "SoSe 2021: 3104 Modern Optimization Techniques

Instructions

Please following these instructions for solving and submitting the exercise sheet.

1. Student should clearly write his/her name, matriculation number and tutorial group number (i.e. "Group 1: Tuesday Tutorial", "Group 2: Wednesday Tutorial").
2. The submission should be made before the deadline, only through learnweb to your group submission link.
3. Should be submitted as a single unzipped PDF file on learn-web course "SoSe 2023: 3104 Modern Optimization Techniques".
4. Each student must submit an individual solution in-order to be eligible for bonus points.
5. Group submission are acceptable but will not contribute towards bonus points.

1 Gradient Descent (11 points)

Using the GD algorithm reviewed in class, minimize the function showed bellow:

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R}$$
$$f(x_1, x_2) = x_1^2 + 3x_2^2 + 2x_1 + 0.5x_2$$

(a) Compute the optimal solution analytically. i.e. the minimum $x^* = (x_1^*, x_2^*)$ and the associated p^*

(b) Perform 3 iterations and evaluate the function at the end of each iteration. Use an initial point $x_0 = (3, -1)$ and a step size $\mu = 0.2$. Is the algorithm minimizing it?

(c) What happen if you change the step size to 0.5? Perform 3 iterations and evaluate the function at the end of each iteration as well.

(d) What is the function being minimized in a non-regularized least squares linear regression? Show how to derive its closed form solution.

2 Backtracking Line Search (9 points)

Let us define a function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ through:

$$f(x_1, x_2) = x_1^2 + x_2^2$$

a) Suppose you want to do a backtracking line search using the negative gradient $\Delta x = -\nabla f(x)$ as descent direction. Suppose you are in a current point $x' = (x_1', x_2')$, write down the backtracking condition

$$f(x + \mu\Delta x) > f(x) + a\mu\nabla f(x)\Delta x$$

for these special settings.

b) We pick $a = 0.5$, $b = 0.1$ and start with a rather high initial step size $\mu = 10$. How small does μ have to become for the backtracking condition to be false? How many backtracking iterations will be done until this happens?
[NOTE:] you can use $x = (0.5, 1)$ to show your working for this question.