OPTIMITZACIO
Fall 2023
Exercises: Gradient Methods

Due: **5.12.2023 , 23:59h**, in the virtual campus

Exercise 3.1: Consider the function

$$f(x,y) = x^2 + xy + y^2 + 5.$$

- Starting at (1,1), write down 2 steps of conjugate gradient descent method for f.
- For the same starting point and function, do 2 steps in the hypergradient descent method.

Exercise 3.2: Write down 1 step in the classical Newton method for the function

$$f(x,y) = (x+1)^2 + (y+3)^2 + 4$$

starting at (0,0).

Exercise 3.3: Recall that the mini-batch modification of stochastic gradient descent in the optimization problem

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x) \to min$$

is an update of the following form

$$x^{(k+1)} = x^{(k)} - \frac{\alpha^{(k)}}{|I_k|}\sum_{i\in I_k} \nabla f_i\big(x^{(k)}\big),$$

where I_k is a subset of $\{1, \dots, n\}$ selected uniformly at random at step k ('a mini-batch'), and $|I_k|$ is the number of elements in I_k ('size of mini-batch'). Suppose we choose the size of minibatch equal to 2 (that is, I_k has two elements). Show that:

- $\sum_{i\in I_k} \nabla f_i(x)$ is a stochastic gradient.
- the variance of this gradient is smaller than the variance of the standard stochastic gradient $\nabla_{i(k)}(x)$ (i.e., when the size of the mini-batch is equal to 1).
- Does your argument work for other sizes of the mini-batches?