

# Mathematics 2

## Homework GD.2

March 19, 2025

The homework consists of six theoretical and practical problems. The solutions are to be submitted as **one .zip file** to the appropriate mailbox on ucilnica. The submission should contain a .pdf file containing clear and well described solutions to each of the problems: procedure, proofs and explicit computations (for theoretical problems), code, explanation of choices of parameters, numerical results, etc.

### 1 Theoretical problems

1. A sequence  $\{x_i\}_i$  is eventually  $p$ -periodic if for some  $N, p \in \mathbb{N}$  we have  $x_{N+j} = x_{N+j+p}, \forall j \in \mathbb{N}$ . Which of the following procedures on a strictly convex function may result in non-constant 2- or 3-periodic sequences for an appropriate selection of fixed positive parameters  $\gamma, \mu$  and starting point  $x_1$ : GD, Polyak GD, Nesterov GD? Explain/prove your answer.
2. Determine the optimal learning rates  $\gamma, \mu$  for the Polyak GD for function

$$f(x, y, z) = x^2 + 2y^2 - 2yz + 4z^2 + 3x - 4y + 5z.$$

### 2 Programming problems

3. Implement GD, Polyak GD, Nesterov GD, and AdaGrad GD.
4. Implement the Newton method and BFGS.
5. Compare the methods of 3. and 4. on:
  - (a)  $f(x, y, z) = (x - z)^2 + (2y + z)^2 + (4x - 2y + z)^2 + x + y$  for starting points  $(0, 0, 0)$  and  $(1, 1, 0)$ .
  - (b)  $f(x, y, z) = (x-1)^2 + (y-1)^2 + 100(y-x^2)^2 + 100(z-y^2)^2$  for starting points  $(1.2, 1.2, 1.2)$  and  $(-1, 1.2, 1.2)$ .
  - (c)  $f(x, y) = (1.5 - x + x y)^2 + (2.25 - x + x y^2)^2 + (2.625 - x + x y^3)^2$  with starting points  $(1, 1)$  and  $(4.5, 4.5)$ .

Describe:

- (a) which one performs best in 2, 5, 10, 100 steps;
  - (b) which one performs best in .1, 1, 2 seconds.
6. Linear regression. Generate  $N$  points of the form  $\{(i, i + \nu) \mid i = 1, 2, \dots, N\}$ , where  $\nu$  is a random noise on the interval  $[0, 1]$ . Using the least squares fitting try to fit a linear function  $g(x) = kx + n$  to these points. Use GD, SGD, Newton, BFGS and L-BFGS for  $N=50, 100, 1000, 10\ 000, 100\ 000, 1\ 000\ 000$  and compare the performances.
  7. Extra problem: Suppose  $H$  is a symmetric positive definite matrix, which is also block diagonal. Can you think of an algorithm that will improve the known GD algorithms for functions of the form  $f(x) = x^T H x$ ? If so, describe the solution and justify it as well as possible. Then demonstrate it on an example.