Homework 8: Gradient descent

Hand in: 09.12.2025 (Tuesday)
Please follow the submission instructions from the webpage of the course.

Correction: tutorial session on 11.12.2025 (Thursday)

Exercise 1: Gradient descent on overparameterized linear least squares problems (10 points)

This exercise is inspired from Exercise 7 in Chapter 3 of the book "Optimization for Data Analysis", by Stephen Wright and Benjamin Recht.

Consider the linear least squares problem:

$$\min_{x \in \mathbb{R}^n} \; f(x) = \frac{1}{2} \|Ax - b\|^2 \tag{1}$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

Assume that $m < n$ and that A has full column rank.
We will note $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m$ the eigenvalues of the matrix $AA^\top$.
Also, we assume that there exists (at least) one solution z to the linear system $Az = b$.

Remark: *You can use without proof that the nonzero eigenvalues of $A^\top A$ are the same as the nonzero eigenvalues of $AA^\top$.*

1. Characterize the stationary points, local minima and global minima of the optimization problem (1).
2. Write down the steepest gradient descent update rule for the optimization problem (1), with the choice of step size $\alpha = \frac{1}{L}$ (you also have to express L explicitly).
3. Let $x_0, \dots, x_k$ be the iterates of the steepest gradient descent method with $x_0 = 0$.
   Using the results from the lecture, can we derive an inequality of the form:
   $$\frac{1}{2} \|Ax_k - b\|^2 \leq C\rho^k, \tag{2}$$
   for some $C > 0$ and $\rho \in (0, 1)$?
4. Define $r_k := Ax_k - b$. Show that:
   $$r_{k+1} = Mr_k \tag{3}$$
   for some symmetric matrix M.
   Also, provide the eigenvalues of the matrix M.
5. Conclude that we actually have the inequality:
   $$\frac{1}{2} \|Ax_k - b\|^2 \leq C\rho^k \tag{4}$$
   for some $C > 0$ and $\rho \in (0, 1)$.

Exercise 2: Gauss-Southwell method (12 points)

This exercise is inspired from Exercise 4 in Chapter 3 of the book "Optimization for Data Analysis", by Stephen Wright and Benjamin Recht.

The Gauss-Southwell method is the following iterative method:

$$\text{For } k = 0, 1, 2, \cdots :$$

$$x_{k+1,i} = \begin{cases} x_{k,i} - \alpha \nabla f(x_k)_{i_k} & \text{if } i = \arg\max_j |\nabla f(x_k)_j| \\ x_{k,i} & \text{otherwise} \end{cases} \tag{5}$$

where $x_{k,i}$ is the i-th component of the vector x_k, and α is a step size.

1. Rewrite the Gauss-Southwell method in the standard form:

$$\text{For } k = 0, 1, 2, \cdots :$$

$$x_{k+1} = x_k + \alpha \varphi(x_k) \tag{6}$$

   where the function φ has to be explicitly given.

2. Prove that the function $\varphi(x)$ verifies the two following inequalities for all $x \in \mathcal{X}$:

$$\|\varphi(x)\| \leq \|\nabla f(x)\| \tag{7a}$$

$$-\nabla f(x)^\top \varphi(x) \geq \frac{1}{n} \|\nabla f(x)\|^2 \tag{7b}$$

3. Now assume that f is L-smooth. Prove that the iterates of the Gauss-Southwell method satisfy the following inequality:

$$f(x_{k+1}) \leq f(x_k) - C(\alpha) \|\nabla f(x_k)\|^2 \tag{8}$$

   where $C(\alpha)$ is a function of α that you have to find.

   Hint: *Use the following inequality for L-smooth functions:*

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2} \|y - x\|^2$$

   and the inequalities from the previous questions.

4. Find the value $\bar{\alpha}$ that minimizes the value of $C(\alpha)$.

5. Now assume that f is μ-strongly convex. Proves that the following holds for any x:

$$f(x) - f(x^\star) \leq \frac{1}{2\mu} \|\nabla f(x)\|^2 .$$

   Use this result to prove the following inequality for the iterates of the Gauss-Southwell with $\alpha = \bar{\alpha}$:

$$f(x_k) - f(x^\star) \leq \rho^k \left(f(x_0) - f(x^\star)\right), \tag{9}$$

   where $\rho \in (0, 1)$ has to be explicitly found.

6. Conclude regarding the convergence of the method for strongly convex functions (and L-smooth functions).

Exercise 3: L2 Penalization for Almost Convex Functions (10 points)

This exercise is inspired from Exercise 6 in Chapter 3 of the book "Optimization for Data Analysis", by Stephen Wright and Benjamin Recht.

Consider the optimization problem of minimizing a function $g(x)$ that is continuously twice differentiable, but not convex. However, it is *almost convex*, and L-smooth, i.e. the following holds:

$$-\varepsilon I_n \prec \nabla^2 g(x) \prec M I_n \tag{10}$$

where $\varepsilon \geq 0$ is rather small.
Also, assume that a good guess $\bar{x}$ of the solution $x^\star$ is available:

$$\|x^\star - \bar{x}\| \leq r \tag{11}$$

for some $r > 0$.
We choose to apply the gradient descent method to a modified version of the problem:

$$\min_{x \in \mathbb{R}^n} f_\lambda(x) = g(x) + \frac{\lambda}{2} \|x - \bar{x}\|^2 \tag{12}$$

where $\lambda \geq 0$ is a regularization parameter.

1. Let $x^\star_\lambda$ be the solution of the optimization problem (12). Prove that the following holds for all $x \in \mathbb{R}^n$:

$$g(x) - g(x^\star) \leq f_\lambda(x) - f_\lambda(x^\star_\lambda) + \frac{\lambda r^2}{2} \tag{13}$$

2. Assume that $\lambda > \varepsilon$. Then show that f_λ is μ-strongly convex for some $\mu > 0$ that you should specify.

3. Write down the steepest gradient descent update rule for the optimization problem (12), with the optimal choice of step size.

   <u>*Hint:*</u> *For an L-smooth function; the optimal step size choice is* $\alpha = \frac{1}{L}$.

4. Prove, using the results from the lecture, that the following holds:

$$f(x_k) - f(x^\star_\lambda) \leq C_\lambda \rho_\lambda^k \tag{14}$$

   where $\rho_\lambda \in (0, 1)$ and $C_\lambda > 0$ have to be explicitly given.

5. Conclude that for a specific choice of of λ (that you should specify), and $k \geq \bar{k}$ (where $\bar{k}$ is a number that you have to specify), the following holds:

$$g(x_k) - g(x^\star) \leq 2\varepsilon r^2 \tag{15}$$

Programming tasks (4 bonus points)

Open the jupyter notebook `programming_exercise3.ipynb`, and fill in the missing parts of the code.

If you are struggling with downloading Jupyter notebook, you can also use it online via

`https://jupyter.org/try-jupyter/lab`.