

Artificial Intelligence and Machine Learning

Exercises – Numeric Optimization Techniques

Question 1 (Gradients) ☼

Compute the gradients of the following functions:

$$1. \begin{cases} f_1 : \mathbb{R}^3 \rightarrow \mathbb{R} \\ (x, y, z) \mapsto 3x^2 - 5y^2 + 2z^2 \end{cases}$$

$$4. \begin{cases} f_4 : \mathbb{R}^2 \rightarrow \mathbb{R} \\ (x, y) \mapsto \frac{xy}{3x+y^2} \end{cases}$$

$$2. \begin{cases} f_2 : \mathbb{R}^2 \rightarrow \mathbb{R} \\ (x, y) \mapsto \ln(\sqrt{xy^3}) \end{cases}$$

$$5. \begin{cases} f_5 : \mathbb{R}^2 \rightarrow \mathbb{R} \\ (x, y) \mapsto \ln(e^{-x} + e^{2y}) \end{cases}$$

$$3. \begin{cases} f_3 : \mathbb{R}^M \rightarrow \mathbb{R} \\ \mathbf{x} \mapsto \|\mathbf{x}\|^2 \end{cases}$$

$$6. \begin{cases} f_6 : \mathbb{R}^2 \rightarrow \mathbb{R} \\ (x, y) \mapsto \sqrt{x+y} \sin(xy) \end{cases}$$

Hint: $\|\mathbf{x}\| := \sqrt{\sum_{m=1}^M x_m^2}$ denotes the **Euclidean norm** of the vector $\mathbf{x} \in \mathbb{R}^M$.

Question 2 (Gradient descent for the Rosenbrock function)

Let the function

$$\begin{cases} f : \mathbb{R}^2 \rightarrow \mathbb{R} \\ (x_1, x_2) \mapsto \left(100 \cdot (x_2 - x_1^2)^2 + (x_1 - 1)^2 \right) \end{cases}$$

be given. This function is known as the **Rosenbrock function** whose graph is shown in the following figure 1:

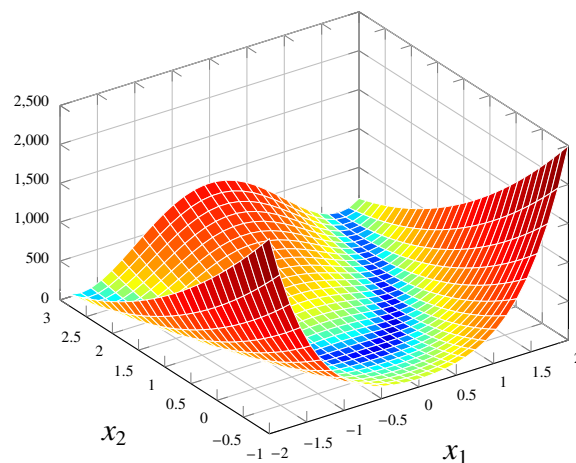


Figure 1: Plot of the two-dimensional Rosenbrock function.

1. Compute the gradient of f and perform five iterations of gradient descent using the learning rate $\alpha := 0.0005$. Start at the point $\mathbf{x}^0 = (0.85 \ 1.10)^\top$.
2. Now perform three iterations of gradient descent using the same starting point and the learning rate $\alpha := 0.005$. What phenomenon do you observe?



Question 3 (Gradient descent for the Himmelblau function) ✳

Let the following function

$$\begin{cases} f: \mathbb{R}^2 \rightarrow \mathbb{R} \\ (x, y) \mapsto (x^2 + y - 11)^2 + (x + y^2 - 7)^2 \end{cases}$$

be given. This function is known as the **Himmelblau function**. Perform two iterations of gradient descent using the learning rate $\alpha := 0.02$. Start at the coordinates $\mathbf{x}^0 := \mathbf{0}$ (zero vector). Please answer the following questions:

1. What is the value of \mathbf{x}^2 (position after two iterations)?
2. What is the function value $f(\mathbf{x}^2)$ compared to $f(\mathbf{x}^0)$?
3. Is f a convex function? **Hint:** Figure 2 might be helpful. (Please justify your answer!)
4. In which points will you eventually end up when initializing the algorithm at the coordinates $\mathbf{x}^0 := \mathbf{0}$ and $\tilde{\mathbf{x}}^0 := (-1, 1)^\top$, respectively? Write a Python program and perform a sufficient amount of iterations!

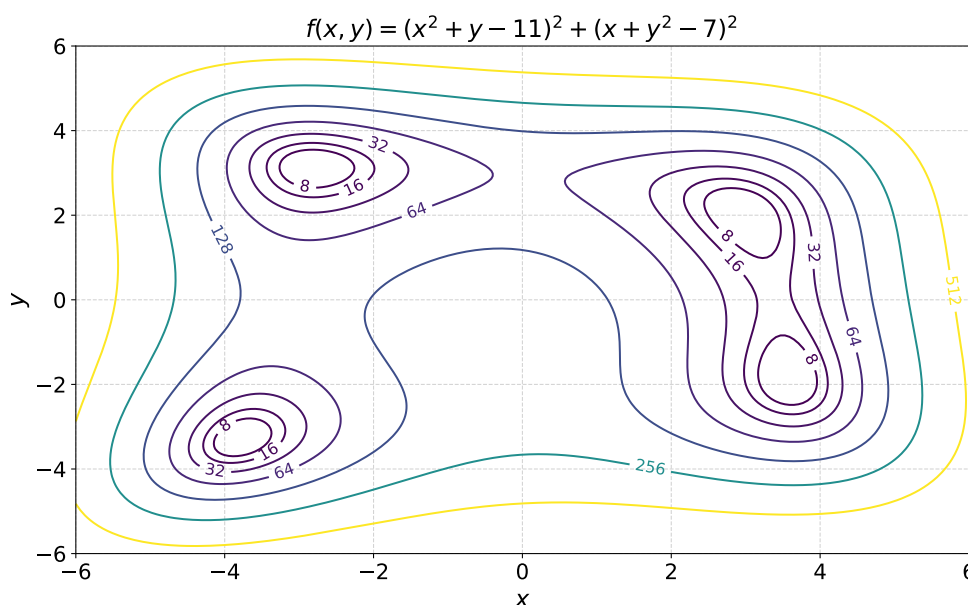


Figure 2: Contour plot of the Himmelblau function.

Question 4 (Gradient descent learning rate) ⌘

What is a suitable value for the learning rate α ? What problems do you face when choosing it too low or too high?

Question 5 (Gradient ascent) ⌘

Suppose you want to find a maximum of a given function f . How would you have to alter the gradient descent update rule to achieve your goal?

Remark: The resulting algorithm is called **gradient ascent**.

Question 6 (Gradient descent update rule) ⌘

Tick the correct parameter update rule used in gradient descent for the function f .

- ☐ $\theta^{t+1} \leftarrow \theta^t + \alpha \nabla f(\theta^t)$
- ☐ $\theta^{t+1} \leftarrow \theta^t - \alpha \nabla f(\theta^t)$
- ☐ $\theta^{t+1} \leftarrow \alpha \nabla f(\theta^t)$
- ☐ All options are incorrect.

Question 7 (Newton's method)

Compute the Hessian matrix of the Rosenbrock function (*see question 2*) and perform two iterations of Newton's method starting from the coordinates $x^0 := (0.85 \quad 1.10)^\top$.

Question 8 (Constrained optimization)

Let the function

$$f(x_1, x_2) = x_1^2 + \frac{1}{2}x_2^2 + 3$$

be given. Work through the following tasks:

1. Write f in the form

$$f(x) = \frac{1}{2}x^\top Qx + c^\top x + \alpha,$$

where $Q \in \mathbb{R}^{2 \times 2}$ is a symmetric matrix, $c \in \mathbb{R}^2$ and $\alpha \in \mathbb{R}$.
Is f a convex function? (*Please justify your answer!*)

2. Compute the (unconstrained) minimum of f analytically.
3. Let the constraint

$$h(x_1, x_2) := x_1 + x_2 - 1 = 0$$

be given. What is the (constrained) solution given this additional constraint?

Question 9 (Dual optimization problem)

Consider the following primal optimization problem

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && \mathbf{x}^\top \mathbf{x} \\ & \text{subject to} && \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned}$$

where $\mathbf{A} \in \mathbb{R}^{K \times M}$. Thus, this minimization problem contains K (linear) equality constraints and no inequality constraints. Please work through the following tasks:

1. Derive the dual optimization problem!
2. Consider the above optimization problem for the case $K = M = 2$. Let the equality constraints be given by

$$h_1(x_1, x_2) := 2x_1 - 1 = 0$$

$$h_2(x_1, x_2) := x_1 + x_2 = 0.$$

Solve this optimization problem

- (a) via the primal formulation, and
- (b) via the dual formulation.

Convince yourself that the solutions are indeed identical.