4 Unconstrained optimization

Exercise 4.1
Find all critical points of the function

$$f(x, y) = x^2 + y^2$$

and classify each as a relative maximum, a relative minimum, or a saddle point.

Exercise 4.2
Find all critical points of the function

$$f(x, y) = x^3 - y^3 + 6xy$$

and classify each as a relative maximum, a relative minimum, or a saddle point.

Exercise 4.3
Write the function

$$f(x, y) = x^2 + y^2 + xy + x - 2y + 3$$

under the matrix form

$$f(x, y) = \frac{1}{2} u^T A u + b^T u + c$$

where $u = (x, y)$, A is a symmetric square matrix, b is a vector and c is a scalar. Is this function convex ?

Exercise 4.4
Consider the function $f : \mathbb{R} \mapsto \mathbb{R}$ defined by

$$f(x) = ax^2 + bx + c$$

with $a > 0$, $b \in \mathbb{R}$ and $c \in \mathbb{R}$. We want to prove the convergence of the gradient descent algorithm with constant step which consists in a sequence of values $x^{(k)}$ such that, given $x^{(0)} \in \mathbb{R}$,

$$x^{(k+1)} = x^{(k)} - \eta f'(x^{(k)})$$

with $\eta > 0$.

1. Find the critical point x^* of f and show it is a global minimum by studying the sign of f'.
2. Show that, for all $(x, y) \in \mathbb{R}^2$,

$$f(y) - f(x) - (y - x) f'(x) = a(y - x).$$

3. Show that

$$f(x^{(k+1)}) - f(x^{(k)}) = -\eta \left(f'(x^{(k)})\right)^2 (1 - a\eta).$$

4. Show that

$$\sum_{k=0}^{N} \left(f'(x^{(k)})\right)^2 = \frac{1}{\eta(1 - a\eta)} \left(f(x^{(0)}) - f(x^{(N+1)})\right).$$

5. Assume that $\eta < \frac{1}{a}$, then show that

$$\sum_{k=0}^{N} \left(f'(x^{(k)})\right)^2 \leq \frac{1}{\eta(1-a\eta)} \left(f(x^{(0)}) - f(x^*)\right).$$

6. Assuming that $\eta < \frac{1}{a}$, deduce that

$$\lim_{k \to +\infty} f'(x^{(k)}) = 0,$$

which involves that

$$\lim_{k \to +\infty} x^{(k)} = x^*.$$

Exercise 4.5

We want to code with Jupyter notebook the gradient descent algorithm with constant step in order to compute the minimum of the function :

$$f(x) = \frac{1}{10}x^2 - 8x + 10.$$

1. Plot the function $f(x)$ for $n = 1000$ equally spaced samples x_i satisfying $-60 \leq x_i \leq 140$.
2. Compute the global minimum x^* of $f(x)$.
3. Code the gradient descent algorithm with constant step $\eta = 2$. The number of iterations is $N = 20$ and $x^{(0)} = -50$. Save the values $x^{(k)}$, $f'(x^{(k)})$ and $e_k = f(x^{(k)}) - f(x^*)$ as some sequences depending on the iteration step k.
4. Use a subplot figure to plot the sequence $f'(x^{(k)})$ with respect to k and to plot the sequence e_k with respect to k.
5. Plot the function $f(x)$ with all the steps $x^{(k)}$ appearing as yellow circles on the curve. The goal of this curve is to show the convergence of the gradient descent algorithm.
6. Try the constant step $\eta = 8$ and next $\eta = 12$. Discuss the impact of η on the algorithm convergence.
7. Come back to the constant step $\eta = 2$ and try the initialization point $x^{(0)} = 10$ and next $x^{(0)} = 100$. Discuss the impact of $x^{(0)}$ on the algorithm convergence.