1. Consider the function
$$f(x, y) = x^2 + 2y^2 + xy + x - y + 30.$$
Show that f satisfies the convergence theorem of the gradient descent with constant step size. Give the value of the step that guarantees the fastest rate of convergence. Calculate by hand the first 3 iterates of the algorithm with $(x_0, y_0) = (3, 3)$.

2. You are given the function
$$f(x, y) = 4x^2 - 4xy + 2y^2.$$
Write the algorithm of gradient descent with optimal step and apply by hand three iterations. Give the values of $f(x_0), f(x_1), f(x_2), f(x_3)$ and make sure that the values decrease at every step.

3. Let A be a symmetric positive definite matrix of dimension $d \times d$. Let
$$\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_d,$$
be its eigenvalues and consider the iterative method
$$X^{(n+1)} = X^{(n)} - a(AX^{(n)} - b),$$
with $X^{(0)} \in \mathbb{R}^d$ given.

   a) Write it in the form
   $$X^{(n+1)} = BX^{(n)} + c,$$
   where B, c must be precised.

   b) Assume that $a\lambda_d < 2$. Show that all eigenvalues of B are in $(-1, 1)$. Hence conclude that $X^{(n)}$ converges. What is the value of the limit?

   c) For what value of a is the convergence fastest?

   d) Write a program in Python that applies the above iterative method for A the discretization matrix of the Laplacian defined in the first TP. Take as b the constant vector of $1's$ and $X^{(0)} = 0$, $d = 10$ and $a = 10^{-2}$. Verify that the seventh coordinate of $X^{(10000)}$ is 13.9957.

4. The following result is known as the eigenvalue power method and is used as a way to compute the largest eigevalue of a matrix.

   Theorem 1. *Let A be a $d \times d$ real matrix with eigenvalues $\lambda_1, .., \lambda_p$, $p \leq d$ disitnct with*
   $$|\lambda_1| \leq .. \leq |\lambda_{p-1}| < |\lambda_p|,$$
   and construct the iterates
   $$x^0 \in \mathbb{R}^d, y^{n+1} = Ax^n, x^{n+1} = \frac{y^{n+1}}{||y^{n+1}||}.$$

If x^0 is not chosen in the space generated by the eigenvectors associated to the first $p-1$ eigenvalues, then

$$< Ax^n, x^n >$$

converges to λ_p.

a) Assume that a given matrix has d distinct real eigevalues. Show the above convergence result without the normalization assumption $\frac{y^{n+1}}{||y^{n+1}||}$, that is, by writing any vector in the basis formed by the eigenvectors show that

$$\frac{< Ax^n, x^n >}{||x^n||}$$

converges to the largest eigenvalue.

b) Consider the optimization problem

$$max < Ax, x >$$

under the constraint $x \in B = \{x \in \mathbb{R}^d, ||x||_2 \leq 1\}$. Write the expression of the projection on B.

c) Write the gradient projected algorithm for this problem. Can we conclude for the convergence? Justify.

d) Write the algorithm using the eigenvalue power method and with the above theorem conclude about the convergence.

5. The following problem will be solved using gradient with penalty. Consider a cable with length $L = 2$ fixed at points $(0, 1)$ and $(1, 3/2)$. It is discretized with N distinct points $(x_1, y_1), ..., (x_N, y_N)$. Define the random vector

$$u = (x_1, ..., x_N, y_1, ..., y_N) \in \mathbb{R}^{2N}.$$

We want to minimize the potential energy of the cable $E(u)$ under the constraints of length and fixed points. Mathematically we are looking for a solution $u \in \mathbb{R}^{2N}$

$$min_{u \in K} E(u), \quad with \quad E(u) = \frac{1}{N+1} \sum_{i=1}^{N} y_i$$

and $K = \{u \in \mathbb{R}^{2N}, \phi_i(u) = 0, i = 1, ..., N+1\}$, where the constraints of length of each segment are $\phi_i(u) = l_i(u) - h$ with

$$l_i(u) = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}, \quad i = 1, .., N+1$$

and $h = \frac{L}{N+1}$, $(x_0, y_0) = (0, 1)$ and $(x_{N+1}, y_{N+1}) = (1, 3/2)$. For penalization we suggest the function

$$P(u) = \frac{1}{2} \sum_{i=1}^{N+1} \phi_i^2(u).$$

a) Write a function called $energy(u)$ which calculates the value of E and its gradient at u.

b) Write a function called $penalty(u)$ which calculates the value and the gradient of P. As a first step prove that

$$\frac{\partial P}{\partial x_i}(u) = (1 - \frac{h}{l_i})(x_i - x_{i-1}) + (1 - \frac{h}{l_{i+1}})(x_i - x_{i+1}),$$

and

$$\frac{\partial P}{\partial y_i}(u) = (1 - \frac{h}{l_i})(y_i - y_{i-1}) + (1 - \frac{h}{l_{i+1}})(y_i - y_{i+1}).$$

c) Apply the gradient descent with fixed step for

$$min_{x \in \mathbb{R}^{2N}} E(u) + \frac{1}{\epsilon} P(u)$$

take u^0 random, $N = 20$, $\epsilon = 0.1$ then 0.01, constant step size $a = \frac{\epsilon}{3}$ and stopping criterion

$$||\nabla E(u) + \frac{1}{\epsilon} \nabla P(u)||_\infty \leq 0.001.$$

The result should be the position (x, y).