Optimization & Optimal Research
Practical Session

Abstract

The aim of this practical session is to write the algorithm we have studied during the class. I will ask you to study the convexity of the function, find the global minimum or local minima. We will also so the influence of the learning rate for the gradient descent with a fix step and see a condition of convergence of this algorithm. You can use the language you want, but I will use R for the correction and I will not be able to help you if you are using an other one (maybe if you are using Matlab).

Introduction (20 minuts)

We will use the following functions :

$$f_1(x,y) = x^2 + \frac{y^2}{20},$$

$$f_2(x,y) = \frac{x^2}{2} + \frac{y^2}{2},$$

$$f_3(x,y) = (1-x)^2 + 10(y-x^2)^2,$$

$$f_4(x,y) = \frac{x^2}{2} + x\cos(y).$$

For the function f_3 we will take $(x_0, y_0) = (-1, 1)$ as the initialization of our algorithms.

1. Compute the gradient of each functions
2. Which of the functions ar convex ? Why ? You can use what we have studied during the lessons or in exercises.
3. Plot these functions.
4. What is the global minimum of each functions ?

We now want to solve the following optimization problem

$$\underset{(x,y)\in\mathbb{R}^2}{min} f(x,y),$$

for each function f using the different algorithm studied in class.

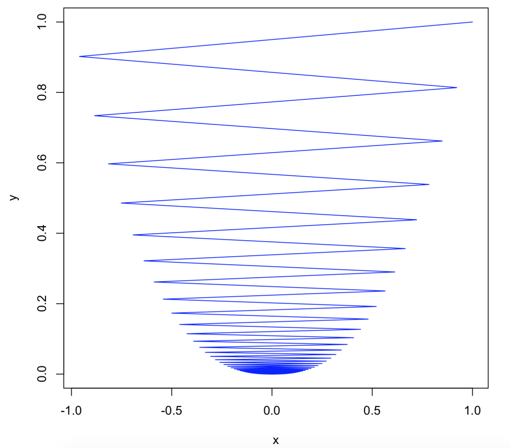
1 Gradient descent with a constant learning rate (40 minuts)

We recall that the gradient descent algorithm with constant learning rate $\eta > 0$ updates the weights at each iteration as follows :

$$u_{k+1} \leftarrow u_k - \eta \nabla f(u_k).$$

1. Define a function called *gradientdescent* using this gradient descent algorithm.

2. Use your function for the different values of η with the function (f_1, f_2, f_3, f_4). What do you notice ? Represent the convergence of (x, y) in a graph such as the one follows :



3. Now we consider the function f_2, if you have not done it, test the algorithm for $\eta = 1.9$ and $\eta = 2.1$. Give a condition on η so that the algorithm converges.

   We suppose that the function f is $\alpha-$elliptical and the gradient function ∇f is $L-$lipschitzien. We can show that this algorithm converges if we take η such that : $0 < \eta < \dfrac{2\alpha}{L^2}$.

   (a) Find the value of L such that :

   $$\|\nabla f(x_1, y_1) - \nabla f(x_2, y_2)\| \leq L\|(x_1, y_1) - (x_2, y_2)\|$$

   (b) Compute λ_{min} the smallest eigenvalue of f. We admit that $\alpha = \lambda_{min}$ and conclude.

   (c) Try to do the same for the function f_1.

4.

2 Gradient descent with optimal step (30 minuts)

Now the learning rate is no more constant, it is determined by solving the problem :

$$\eta^{(k)} = \underset{eta>0}{Argmin}\ f(u_k - \eta \nabla f(u_k)).$$

1. Give an explicit expression of η for the first and/or second function(s).

2. Implement the algorithm.

3. Solve the problem of minimization of function f_1 and compare to the previous algorithm.

4. Do the same for the function f_3.

3 Newton's Method (30 minuts)

The Newton's Method is solving Euler's Equation

$$\nabla f(u) = 0.$$

An iteration of the Newton's algorithm is given by :

$$u_{k+1} \leftarrow u_k - (H_f(u))^{-1} \nabla f(u),$$

where H_j refers to the Hessian matrix of the function f.

Remark : It is possible to improve this method combining it with a line search algorithm, setting :

$$u_{k+1} \leftarrow u_k - \eta (H_f(u))^{-1} \nabla f(u),$$

where η is constant or satisfies the Wolfe's condition.

1. Compute the Hessian matrix for the functions f_1, f_3 and f_4.
2. Implement the Newton's method for the function f_3 and f_4 to find the global minimum.
3. For the function f_2 and/or f_3, compare the convergence of the three algorithms.
4. What are the avantage(s) and drawback(s) of this method ?