Homework nr. 8

Consider $F : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ a real function. Approximate a (local or global) minimum point of function F using the gradient descendent method. Test different methods to calculate the learning rate. Compute the gradient of the function F using the analytic formula and the approximation formula. Compare the solutions obtained by using the two computing methods of the function's F gradient, in terms of number of iterations used by the two methods (for the same precision $\epsilon > 0$).

Functions' Minimization

Let $F : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ be a real function, twice differentiable, $F \in C^2(\mathbb{R} \times \mathbb{R})$, for which we want to approximate the solution x^* of the minimization problem:

$$\min\{F(x,y); (x,y) \in V\} \quad \longleftrightarrow \quad F(x^*,y^*) \leq F(x,y) \quad \forall (x,y) \in V \tag{1}$$

where V is either $V = \mathbb{R} \times \mathbb{R}$ (where (x^*,y^*) is a global minimum point) or $V = S((\bar{x},\bar{y}),r)$, is a sphere with the center $(\bar{x},\bar{y})$ and the radius r (which is a local minimum point).

A point $(\tilde{x},\tilde{y})$ is a *critical point* for function F, if it is the solution of the next system of equations:

$$\nabla F(\tilde{x},\tilde{y}) = 0 \quad , \quad \nabla F(x,y) = \begin{pmatrix} \dfrac{\partial F}{\partial x}(x,y) \\ \\ \dfrac{\partial F}{\partial y}(x,y) \end{pmatrix}. \tag{2}$$

It is known that, for twice differentiable functions, the minimum points of function F are among the critical points. A critical point is a minimum point if the Hessian matrix is positively semi-definable:

$$H(x,y) = \begin{pmatrix} \dfrac{\partial^2 F}{\partial x^2} & \dfrac{\partial^2 F}{\partial x \partial y} \\ \\ \dfrac{\partial^2 F}{\partial y \partial x} & \dfrac{\partial^2 F}{\partial y^2} \end{pmatrix} \quad , \quad \big(H(\tilde{x},\tilde{y})z, z\big)_{\mathbb{R}^2} \geq 0 \quad \forall z \in \mathbb{R}^2$$

Gradient Descendent Method

The minimum point of a function F is approximated by constructing the string $\{(x_k, y_k)\}$ which, under certain conditions, converges to the minimum point (x^*, y^*). The convergence of this string depends on the choice of the first element of the string, i.e., (x_0, y_0).

The $k+1$-element of the string (x_{k+1}, y_{k+1}), is constructed from the previous one, (x_k, y_k), as follows:

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \end{pmatrix} - \eta_k \nabla F(x_k, y_k)\ ,\ k = 0, 1, \dots\ , \qquad \text{where } \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - \text{is randomly chosen.} \tag{3}$$

The element η_k is called the learning rate, or the iteration step.

Strategies for choosing the learning rate

1. $\eta_k = \eta\ ,\ \forall k\ (\eta = 10^{-3}, 10^{-4}, \dots)$. A constant learning rate with a too big value makes the minimum point hard to be found, while a too small value for the learning rate has the disadvantage of a too costing computation.

2. A possibility to solve problems with a constant learning rate is to consider a variable value, depending on the local context. The method described below is called *backtracking* adjustment of the step length/learning rate (or *backtracking line search*). This method works for convex functions.

   Consider $\beta \in (0, 1)$ a constant value (usually we take $\beta = 0.8$). At each step the learning rate is computed as follows:

   $\eta = 1;$
   $p = 1;$
   while $F((x_k, y_k) - \nabla F(x_k, y_k)) > F(x_k, y_k) - \frac{\eta}{2}\|\nabla F(x_k, y_k)\|^2$ && $p < 8$
   $\quad \eta = \eta\,\beta;$
   $\quad p{+}{+}\,;$

Important remark: The way in which the initial element, (x_0, y_0) is chosen may cause the convergence or divergence of the string (x_k, y_k) to

(x^*, y^*). Usually, a choice of the initial data in the proximity of (x^*, y^*) assures the convergence $(x_k, y_k) \longrightarrow (x^*, y^*)$ for $k \to \infty$.

It is not necessary to memorize all elements of the string $\{(x_k, y_k)\}$, but only the 'last' computed element (x_{k_0}, y_{k_0}). We say that an element (x_{k_0}, y_{k_0}) approximates a minimum point, (x^*, y^*), denoted by $(x_{k_0}, y_{k_0}) \approx (x^*, y^*)$ (where (x_{k_0}, y_{k_0}) is the last element of the string that we want to compute), if the difference between two successive elements of the string is small enough, i.e.,

$$\| \begin{pmatrix} x_{k_0} \\ y_{k_0} \end{pmatrix} - \begin{pmatrix} x_{k_0-1} \\ y_{k_0-1} \end{pmatrix} \| \leq \epsilon \tag{4}$$

where ϵ is the precision with which we want to approximate the solution (x^*, y^*).

Therefore, a possible approximation scheme of the solution (x^*, y^*), is the following one:

Computing Scheme

randomly choose the initial values of the string, x, y ;
$k = 0$;
do
{
- compute $\nabla F(x,y)$;
- compute the learning rate η using one of the two methods;
- $x = x - \eta \dfrac{\partial F}{\partial x}(x,y)$;
- $y = y - \eta \dfrac{\partial F}{\partial y}(x,y)$;
- $k = k + 1$;

}
while $(\eta \, \|\nabla \, F(x,y)\| \geq \epsilon$ and $k \leq k_{\max}$ and $\eta \, \|\nabla \, F(x,y)\| \leq 10^{10}$)
if ($\eta \, \|\nabla \, F(x,y)\| \leq \epsilon$) $(x,y) \approx (x^*, y^*)$;
else *"divergence"* ; //(try to change the initial data)

A possible value for $k_{\max}$ is 30000 and $\epsilon > 10^{-5}$.

To compute the value of function's F gradient in a certain point, the analytical gradient formula must be used (where the function is declared in the program). Also use the following approximation formula:

$$\nabla F(x,y) \approx \begin{pmatrix} G_1(x,y,h) \\ G_2(x,y,h) \end{pmatrix}$$

where

$$\frac{\partial F}{\partial x}(x,y) \approx G_1(x,y,h) = \frac{3F(x,y) - 4F(x-h,y) + F(x-2h,y)}{2h}$$

$$\frac{\partial F}{\partial y}(x,y) \approx G_2(x,y,h) = \frac{3F(x,y) - 4F(x,y-h) + F(x,y-2h)}{2h}$$

with $h = 10^{-5}$ or 10^{-6} (may be considered as an input parameter).

Examples

$$F(x,y) = x^2+y^2-2x-4y-1\,, \quad \nabla F(x,y) = \begin{pmatrix} 2x-2 \\ 2y-4 \end{pmatrix}\,, \quad x^* = 1\,,\ y^* = 2$$

$$F(x,y) = 3x^2-12x+2y^2+16y-10\,, \quad \nabla F(x,y) = \begin{pmatrix} 6x-12 \\ 4y+16 \end{pmatrix}\,, \quad x^* = 2\,,\ y^* = -4$$

$$F(x,y) = x^2-4xy+5y^2-4y+3\,, \quad \nabla F(x,y) = \begin{pmatrix} 2x-4y \\ -4x+10y-4 \end{pmatrix}\,, \quad x^* = 4\,,\ y^* = 2$$

$$F(x,y) = x^2y-2xy^2+3xy+4\,, \quad \nabla F(x,y) = \begin{pmatrix} 2xy-2y^2+3y \\ x^2-4xy+3x \end{pmatrix}\,, \quad x^* = -1\,,\ y^* = 0.5$$