Tema nr. 8

Fie $F : \mathbb{R}\times\mathbb{R} \longrightarrow \mathbb{R}$ o funcţie reală. Să se aproximeze un punct de minim (local sau global) al funcţiei F folosind metoda gradientului descendent. Să se testeze diversele metode de calcul a ratei de învăţare. Să se calculeze gradientul funcţiei F folosind formula analitică şi formula aproximativă. Să se compare soluţiile obţinute folosind cele două moduri de calcul a gradientului funcţiei F, din punctul de vedere al numărului de iteraţii efectuate pentru obţinerea soluţiilor (pentru aceeaşi precizie $\epsilon > 0$).

Minimizarea funcţiilor

Fie $F : \mathbb{R}\times\mathbb{R} \longrightarrow \mathbb{R}$ o funcţie reală de două ori derivabilă , $F \in C^2(\mathbb{R}\times\mathbb{R})$, pentru care vrem să aproximăm soluţia x^* a problemei de minimizare:

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

unde $V = \mathbb{R}\times\mathbb{R}$ ((x^*,y^*) este punct de minim global) sau $V = S((\bar{x},\bar{y}),r)$, sfera de centru $(\bar{x},\bar{y})$ şi rază r (punct de minim local). Se numeşte *punct critic* pentru funcţia F, un punct $(\tilde{x},\tilde{y})$ care este rădăcină a sistemului de ecuaţii:

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

Se ştie că pentru funcţiile de două ori derivabile, punctele de minim ale funcţiei F se găsesc printre punctele critice. Un punct critic este punct de minim dacă matricea hessiană este pozitiv semidefinită:

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

Metoda gradientului descendent

Punctul de minim al funcţiei F se aproximează construind un şir $\{(x_k, y_k)\}$ care, în anumite condiţii, converge la punctul de minim (x^*, y^*) căutat. Convergenţa şirului depinde de alegerea primului element ale şirului, (x_0, y_0).

Elementul $k+1$ al şirului, (x_{k+1}, y_{k+1}), se construieşte pornind de la elementul precedent, (x_k, y_k), astfel:

$$\begin{aligned} \begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \end{pmatrix} \ - \ \eta_k \nabla F(x_k, y_k)\ ,\ k = 0, 1, \dots\ , \\ \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} - \ \text{daţi random} \end{aligned} \tag{3}$$

Elementul η_k poartă numele de rată de învăţare sau pasul iteraţiei.

Strategii de alegere a ratei de învăţare

1. $\eta_k = \eta\ ,\ \forall k\ (\eta = 10^{-3}, 10^{-4}, ...)$. O rată de învăţare constantă prea mare poate face ca punctul de minim să nu poată fi găsit, iar o valoare prea mică pentru rata de învăţare are dezavantajul unui cost de calcul mare.

2. Un mod de a rezolva problemele care apar în cazul ratei de învăţare constante este de a considera o valoare variabilă, în funcţie de contextul local. Metoda descrisă mai jos poartă denumirea de ajustare de tip *backtracking* a lungimii pasului/ratei de învăţare (*backtracking line search*). Această metodă funcţionează pentru funcţii convexe.

   Se considera $\beta \in (0, 1)$ fixat (de obicei se alege $\beta = 0.8$). La fiecare pas rata de învăţare se calculează astfel:

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

Observaţie importantă: Alegerea elementelor iniţiale, (x_0, y_0) poate determina convergenţa sau divergenţa şirului (x_k, y_k) la (x^*, y^*). De obicei, o alegere a datelor iniţiale în vecinătatea lui (x^*, y^*) asigură convergenţa $(x_k, y_k) \longrightarrow (x^*, y^*)$ pentru $k \to \infty$.

Nu este necesară memorarea întregului şir $\{(x_k, y_k))\}$ ci avem nevoie doar de 'ultimul' element (x_{k_0}, y_{k_0}) calculat. Se consideră că un element (x_{k_0}, y_{k_0}) aproximează punctul de minim căutat, (x^*, y^*), $(x_{k_0}, y_{k_0}) \approx (x^*, y^*)$ ((x_{k_0}, y_{k_0}) este ultimul element al şirului care se calculează) atunci când diferenţa dintre două elemente succesive ale şirului devine suficient de mică, i.e.,

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

unde ϵ este precizia cu care vrem să aproximăm soluţia (x^*, y^*).

Prin urmare, o schemă posibilă de aproximare a soluţiei (x^*, y^*) este următoarea:

Schema de calcul

se aleg random valorile iniţiale ale şirului, x, y ;
$k = 0$;
do
{
- calculează $\nabla F(x,y)$;
- calculează rata de învăţare η folosind una din cele 2 metode;
- $x = x - \eta \dfrac{\partial F}{\partial x}(x,y)$;
- $y = y - \eta \dfrac{\partial F}{\partial y}(x,y)$;
- $k = k + 1$;
}
while ($\eta \|\nabla F(x,y)\| \geq \epsilon$ şi $k \leq k_{\max}$ şi
$\eta \|\nabla F(x,y)\| \leq 10^{10}$)
if ($\eta \|\nabla F(x,y)\| \leq \epsilon$) $(x,y) \approx (x^*,y^*)$;
else *"divergenţă"* ; //(de încercat schimbarea datelor iniţiale)

O valoare posibilă pentru $k_{\max}$ este 30000 şi $\epsilon > 10^{-5}$.

Pentru a calcula valoarea gradientului funcţiei F într-un punct oarecare se va folosi formula analitică de calcul a gradientului (funcţie declarată în program) şi de asemenea se va folosi următoarea formulă aproximativă:

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

unde

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

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

cu $h = 10^{-5}$ sau 10^{-6} (poate fi considerat parametru de intrare).

Exemple

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

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

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

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