LUCRAREA 6

PROGRAMAREA NELINIARĂ. METODE DE ORDINUL 2

6.1. Aspecte generale

Utilizarea gradientului ca direcţie de explorare este conformă cu aproximarea liniară a funcţiei obiectiv în dezvoltarea acesteia în serie Taylor. Totuşi, o aproximare pătratică de forma:

$$F(X) = F\left(X^{(k)}\right) + \left(g^{(k)}, \Delta X^{(k)}\right) + \frac{1}{2}\left(d^{(k)}, H^{(k)} d^{(k)}\right) \tag{6.1}$$

este mai aproape de realitate. Minimul funcţiei se obţine prin derivare în raport cu componentele lui $d^{(k)}$ şi anularea derivatelor respective, de unde rezultă:

$$d^{(k)} = -H^{(k)^{-1}} g^{(k)} \tag{6.2}$$

Astfel, dacă se introduce (6.2) în relaţia iterativă,

$$X^{(k+1)} = X^{(k)} + \lambda^{(k)} d^{(k)}, \quad k = 1,\ 2,\ 3, \ldots \tag{6.3}$$

rezultă:

$$X^{(k+1)} = X^{(k)} - \lambda^{(k)} \cdot H^{(k)^{-1}} g^{(k)} \tag{6.4}$$

unde $\lambda^{(k)}$ poate lua valori arbitrare sau optime. În cazul unor funcţii obiectiv pătratice, minimul va fi atins încă din prima iteraţie, pentru $\lambda^{(0)} = 1$.

Aplicarea metodelor de ordinul 2, denumite şi metode Newton, cere ca matricea $H^{(k)}$ să fie pozitiv definită la fiecare iteraţie. Pentru că aproximarea pătratică este superioară celei liniare, convergenţa metodei în apropierea optimului va fi sensibil mai bună decât a metodelor de ordinul 1. În plus, disponibilitatea derivatelor de ordinul doi permite verificarea condiţiilor de suficienţă a optimului.

Totuşi, aceste avantaje se pierd datorită volumului de calcul necesar obţinerii derivatelor de ordinul doi şi pentru inversarea matricei $H^{(k)}$. În plus, din cauza condiţiei $H^{(k)} > 0$ pot apărea situaţii în care metoda nu poate fi aplicată (cazurile în care $F(X)$ este liniară pe anumite porţiuni, iar $H^{(k)} = 0$).

De asemenea, pot apărea unele cazuri în care $F(X^{(k+1)}) > F(X^{(k+1)})$, deşi $H^{(k)} > 0$, aproximaţia pătratică fiind valabilă doar într-o zonă foarte limitată. Pentru a evita aceste situaţii s-au propus diferite metode care constrâng matricea $H^{(k)}$ să fie pozitiv definită la fiecare iteraţie.

6.2. Metode de tip quasi-Newton

Un loc special în cadrul metodelor Newton îl ocupă aşa-zisele metode quasi-Newton. Acestea din urmă au fost elaborate în scopul evitării principalelor dezavantaje ale metodelor Newton referitoare la inversarea matricei hessiene în fiecare iteraţie. Metodele quasi-Newton au la bază relaţia de recurenţă,

$$X^{(k+1)} = X^{(k)} - \lambda^k \cdot \widetilde{H}^{(k)} g^{(k)} \tag{6.5}$$

unde matricea $\widetilde{H}^{(k)}$ reprezintă o aproximaţie a lui ${H^{(k)}}^{-1}$.

Relaţia (6.5) poate fi considerată ca fiind relaţia fundamentală care stă la baza metodelor ce utilizează derivatele. Astfel, în cazul în care matricea $\widetilde{H}^{(k)}$ se identifică cu matricea unitate I_n, se obţin metodele de gradient, iar pentru cazul în care matricea $\widetilde{H}^{(k)}$ se identifică cu inversa matricei hessiene ${H^{(k)}}^{-1}$, se obţin metodele de tip Newton.

Cel mai performant algoritm din această categorie de metode este *algoritmul Davidson-Fletcher-Powell.* Caracteristic acestui algoritm este faptul că modificarea matricei $\widetilde{H}^{(k)}$ la fiecare iteraţie este facută astfel încât pentru o funcţie pătratică de n variabile, în limita a n iteraţii, matricea $\widetilde{H}^{(k)}$ să devină egală cu matricea ${H^{(k)}}^{-1}$. Iniţial, matricea $\widetilde{H}$ se alege ca fiind matricea unitate I_n, astfel încât prima iteraţie decurge după metoda gradientului. În continuare se face o modificare treptată de la direcţia gradientului la cea a metodei Newton, reuşindu-se astfel combinarea avantajelor aferente celor două metode. De fapt într-o regiune îndepărtată de optim metoda gradientului este mai rapidă, în timp ce în apropierea optimului, metoda Newton este superioară.

Convergenţa metodei Davidon-Fletcher-Powell a fost demonstrată doar pentru funcţii obiectiv pătratice, având matricea hessiană pozitiv definită. Relaţia de recurenţă care stă la baza metodei este următoarea:

$$\begin{aligned}\widetilde{H}^{(k+1)} &= \widetilde{H}^{(k)} + A^{(k)} - B^{(k)} = \widetilde{H}^{(k)} + \frac{[\Delta X^{(k)}][\Delta X^{(k)}]^t}{(\Delta X^{(k)}, \Delta g^{(k)})} - \\ &- \frac{[\widetilde{H}^{(k)}][\Delta g^{(k)}][\Delta g^{(k)}]^t [\widetilde{H}^{(k)}]^t}{(\Delta g^{(k)}, H^{(k)} \Delta g^{(k)})}\end{aligned} \tag{6.6}$$

unde:

$$\Delta X^{(k)} = X^{(k+1)} - X^{(k)}, \ \ \Delta g^{(k)} = g^{(k+1)} - g^{(k)} \tag{6.7}$$

6.3. Exemple numerice

Exemplul 1

Să se determine minimul funcţiei:

$$F(X) = \min(x_1^2 + 4 \cdot x_2^2 - 4)$$

folosind metoda Newton, indicându-se ca punct de pornire $X^{(0)} = [5 \quad 4]^t$, iar $\lambda^{(0)} = 1$.

Se calculează gradientul şi hesianul funcţiei $F(X)$:

$$g = \begin{bmatrix} \dfrac{\partial F(X)}{\partial x_1} & \dfrac{\partial F(X)}{\partial x_2} \end{bmatrix}_t = [2x_1 \quad 8x_2]^t$$

$$H = C = \begin{bmatrix} \dfrac{\partial^2 F(X)}{\partial x_1^2} & \dfrac{\partial^2 F(X)}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 F(X)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 F(X)}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$$

Aplicând relaţia (6.4) rezultă:

$$X^{(1)} = X^{(0)} - H^{(0)^{-1}} \cdot g^{(0)} = \begin{bmatrix} 5 \\ 4 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}^{-1} \begin{bmatrix} 10 \\ 32 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Exemplul 2

1. Să se determine minimul funcţiei:

$$F(X) = \min\left(4 \cdot (x_1 - 2)^2 + (x_2 - 3)^2\right)$$

folosind metoda Davidon-Fletcher-Powell, indicându-se ca punct de pornire $X^{(0)} = [6 \quad 5]^t$.

Se calculează gradientul funcţiei obiectiv:

$$g^{(k)} = \begin{bmatrix} 8(x_1 - 2) \\ 2(x_2 - 3) \end{bmatrix}_{X = X^{(k)}}, \quad g^{(0)} = \begin{bmatrix} 32 \\ 4 \end{bmatrix}$$

Iteraţia 1

Se consideră $\widetilde{H}^{(0)} = I_n$, astfel încât relaţia iterativă are următoarea formă:

$$X^{(1)} = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \end{bmatrix} = \begin{bmatrix} 6 \\ 5 \end{bmatrix} - \lambda^{(0)} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 32 \\ 4 \end{bmatrix} = \begin{bmatrix} 6 - 32\lambda^{(0)} \\ 5 - 4\lambda^{(0)} \end{bmatrix}$$

Pentru determinarea lungimii pasului de deplasare $\lambda^{(0)}$, se introduc valorile $x_1^{(1)}$ şi $x_2^{(1)}$ în expresia funcţiei obiectiv F, se face derivata în raport cu $\lambda^{(0)}$, după care se anulează:

$$\frac{\partial F(x^{(1)})}{\partial \lambda^{(0)}} = 65 - 514\lambda^{(0)} = 0 \rightarrow \lambda^{(0)} = 0{,}1264$$

Prin urmare, valoarea noii aproximaţii este:

$$X^{(1)} = \begin{bmatrix} 6 \\ 5 \end{bmatrix} - 0{,}1264 \cdot \begin{bmatrix} 32 \\ 4 \end{bmatrix} = \begin{bmatrix} 1{,}955 \\ 4{,}494 \end{bmatrix}.$$

Iteraţia 2

Se calculează ΔX şi Δg:

$$g^{(1)} = \begin{bmatrix} -0{,}36 \\ 2{,}988 \end{bmatrix}, \ \Delta X^{(1)} = \begin{bmatrix} -4{,}045 \\ -0{,}506 \end{bmatrix} \text{ şi } \Delta g^{(1)} = \begin{bmatrix} -32{,}36 \\ -1{,}012 \end{bmatrix}$$

Folosind relaţia (6.6) se calculează matricea $\widetilde{H}^{(1)}$:

$$H^{(1)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{\begin{bmatrix} -4{,}045 & 0 \\ -0{,}506 & 0 \end{bmatrix} \cdot \begin{bmatrix} -4{,}045 & -0{,}506 \\ 0 & 0 \end{bmatrix}}{\begin{bmatrix} -4{,}045 & -0{,}506 \end{bmatrix} \cdot \begin{bmatrix} -32{,}36 \\ -1{,}012 \end{bmatrix}} +$$

$$+ \frac{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} -32{,}36 & 0 \\ -1{,}012 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}{\begin{bmatrix} -32{,}36 & -1{,}012 \end{bmatrix} \cdot \begin{bmatrix} -32{,}36 \\ -1{,}012 \end{bmatrix}} = \begin{bmatrix} 0{,}1554 & -0{,}0147 \\ -0{,}0147 & 1{,}001 \end{bmatrix}$$

În continuare se poate calcula noua aproximaţie $X^{(2)}$:

- $$X^{(2)} = \begin{bmatrix} X_1^{(2)} \\ X_2^{(2)} \end{bmatrix} = \begin{bmatrix} 1{,}955 \\ 4{,}494 \end{bmatrix} - \lambda^{(1)} \begin{bmatrix} 0{,}1554 & -0{,}0147 \\ -0{,}0147 & 1{,}0010 \end{bmatrix} \cdot \begin{bmatrix} -0{,}36 \\ 4{,}494 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

unde $\lambda^{(1)} = 0{,}4984$ a fost determinat în mod similar ca $\lambda^{(0)}$.

Deoarece $g^{(2)} = \begin{bmatrix} 0 & 0 \end{bmatrix}^t$, calculele sunt terminate, punctul $X^{(2)} = \begin{bmatrix} 2 & 3 \end{bmatrix}^t$ fiind soluția optimală căutată.

6.4. Desfăşurarea lucrării

1. Se studiază textul lucrării.
2. Să se determine minimul funcţiei:

   $$F(X) = \min(x_1^2 + 2 \cdot x_2^2 - 1)$$

   folosind metoda Newton, indicându-se ca punct de pornire $X^{(0)} = \begin{bmatrix} 3 & 1 \end{bmatrix}^t$. Rezultatele obținute vor fi comparate cu cele obținute folosind funcția Matlab **fminunc**.
3. Să se determine minimul funcţiei:

   $$F(X) = \min(x_1^2 + 3 \cdot x_2^2 - 2)$$

   folosind metoda cvasi Newton, indicându-se ca punct de pornire $X^{(0)} = \begin{bmatrix} 2 & 1 \end{bmatrix}^t$. Rezultatele obținute vor fi comparate cu cele obținute folosind funcția Matlab **fminunc**.