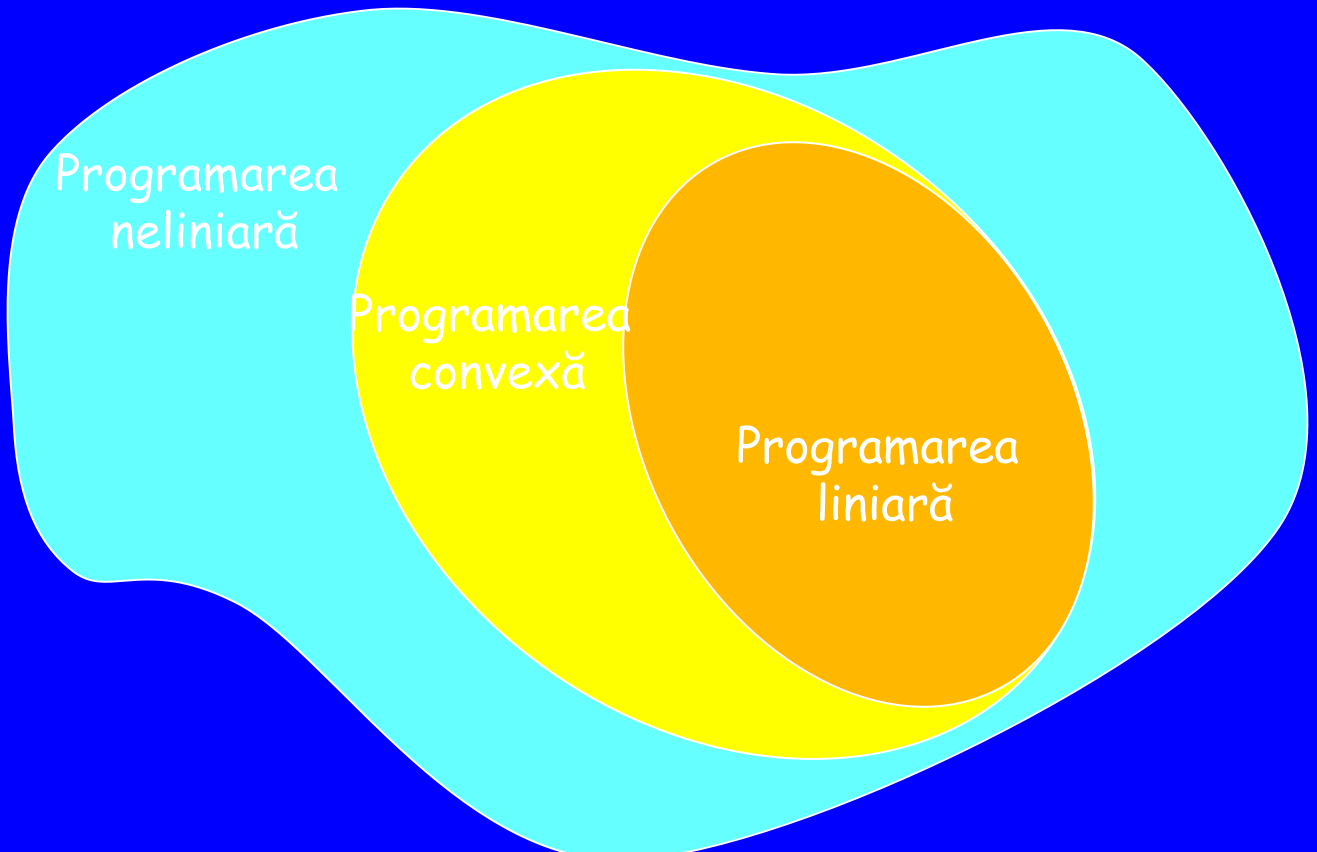
Programarea neliniară

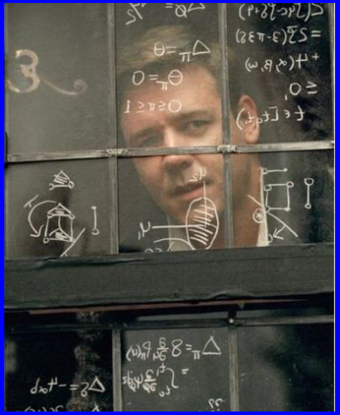
Ierarhizarea problemelor de optimizare



1. Introducere

Instoric:

- **Harold William Kuhn (1925 -)**
  (Condiţiile necesare şi suficiente pentru soluţia optimală a problemelor de programare neliniară - 1951, teoria jocurilor)



- **Albert William Tucker (1905-1995)**
  (Condiţiile necesare şi suficiente pentru soluţia optimală a problemelor de programare neliniară – 1951, teoria jocurilor)

Modelul matematic general al unei probleme de programare liniară

min (max) $F(X)$

$g_i(X)\ (\le, \ge, =)\ b_i,\quad i = 1,\ldots,m$

$X = (x_1,\ldots, x_n)^t$ vectorul n-dimensional al variabilelor de optimizare

$F(X)$ – funcţia obiectiv;

$g_i(X)$ restricţiile modelului matematic;

b_i – vectorul termenilor liberi.

Programarea neliniară fără restricţii

- Pentru funcţiile obiectiv cu o singură variabilă pentru determinarea soluţiei optime se folosesc derivatele de ordinul 1 şi 2.

- Pentru funcţiile cu mai multe variabile pentru determinarea soluţiei optime se folosesc gradientul şi matricea Hessiană.

- Gradientul unei funcţii reprezintă derivata de ordinul 1 în raport cu toate variabilele, iar matricea Hessiană este reprezentată de derivatele de ordinul 2.

Gradientul unei funcţii

Pentru o funcţie *F* definită în raport cu variabilele x_1, x_2, …, x_n, gradientul este definit astfel:

$$\nabla F = \left[\frac{\partial F}{\partial x_1} \quad \frac{\partial F}{\partial x_2} \quad \cdots \quad \frac{\partial F}{\partial x_n} \right]$$

Exemplu: $F = 15x_1 + 2(x_2)^3 - 3x_1(x_3)^2$

$$\nabla F = \left[15 - 3(x_3)^2 \quad 6(x_2)^2 \quad -6x_1x_3 \right]$$

Matricea Hessiană

- Matricea Hessiană (∇^2) a unei funcţii $F(x_1, x_2, \ldots, x_n)$ este definită astfel:

$$\nabla^2 F = \begin{bmatrix} \dfrac{\partial F^2}{\partial x_1^{\ 2}} & \dfrac{\partial F^2}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial F^2}{\partial x_1 \partial x_n} \\ \dfrac{\partial F^2}{\partial x_2 \partial x_1} & \dfrac{\partial F^2}{\partial x_2^{\ 2}} & \cdots & \dfrac{\partial F^2}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial F^2}{\partial x_n \partial x_1} & \dfrac{\partial F^2}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial F^2}{\partial x_n^{\ 2}} \end{bmatrix}$$

Matricea Hessiană

- **Exemplu:**

$$F = 15x_1 + 2(x_2)^3 - 3x_1(x_3)^2$$

$$\nabla F = \begin{bmatrix} 15 - 3(x_3)^2 & 6(x_2)^2 & -6x_1x_3 \end{bmatrix}$$

$$\nabla^2 F = \begin{bmatrix} 0 & 0 & -6x_3 \\ 0 & 12x_2 & 0 \\ -6x_3 & 0 & -6x_1 \end{bmatrix}$$

Condiţiile de optimalitate

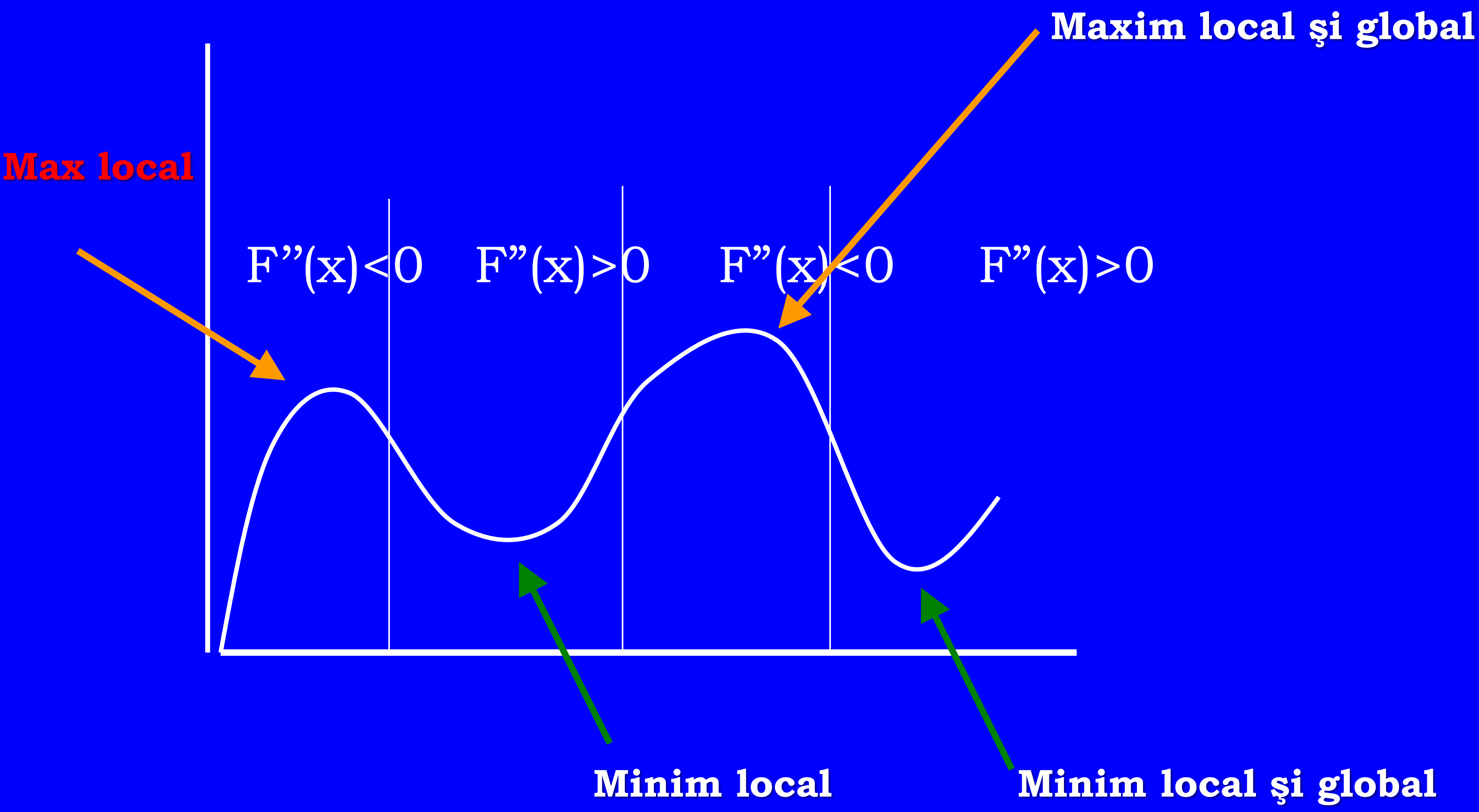
- Optimizarea fără restricţii – problemă de calcul cu mai multe variabile. Pentru Y=F(X), punctul de optim va fi punctul în care F '(X) =0 iar F''(X) îndeplinește condițiile privind derivata de ordinul al II-lea.

- Punctul de minim îndeplineşte condiţiile:

  F'(X) =0 şi F''(X) >0.
- Punctul de maxim îndeplineşte condiţiile:

  F '(X) =0 and F''(X) <0.

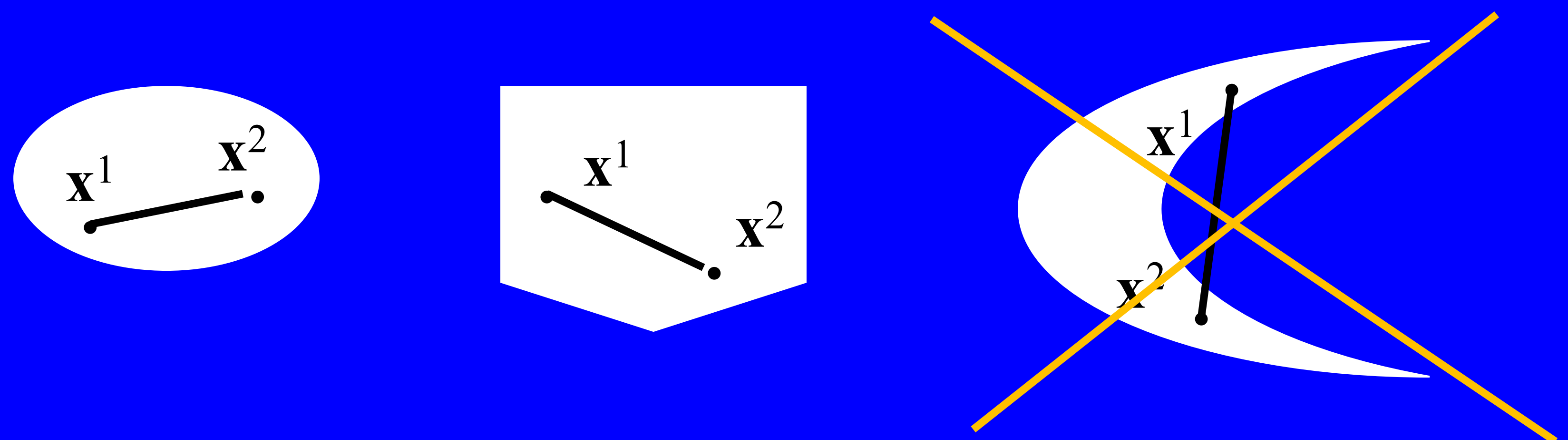
Concavitatea şi derivata de ordinul II



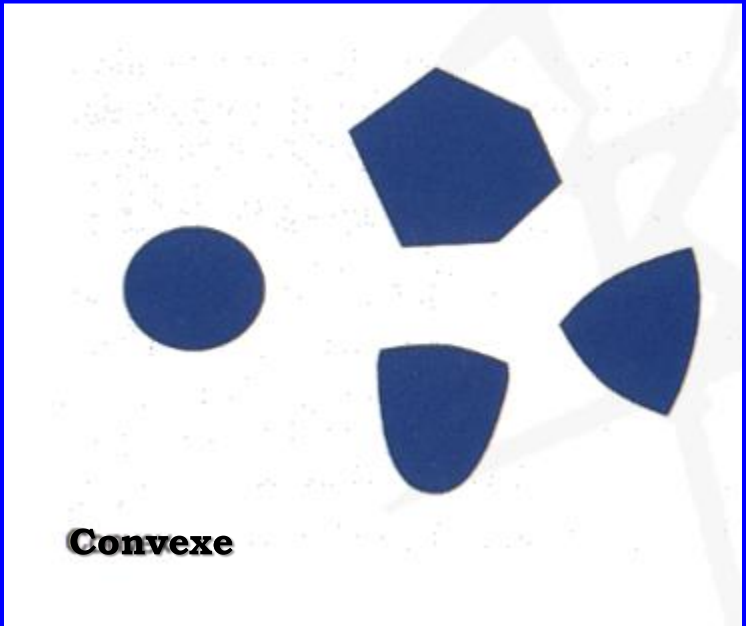
Mulţimi convexe

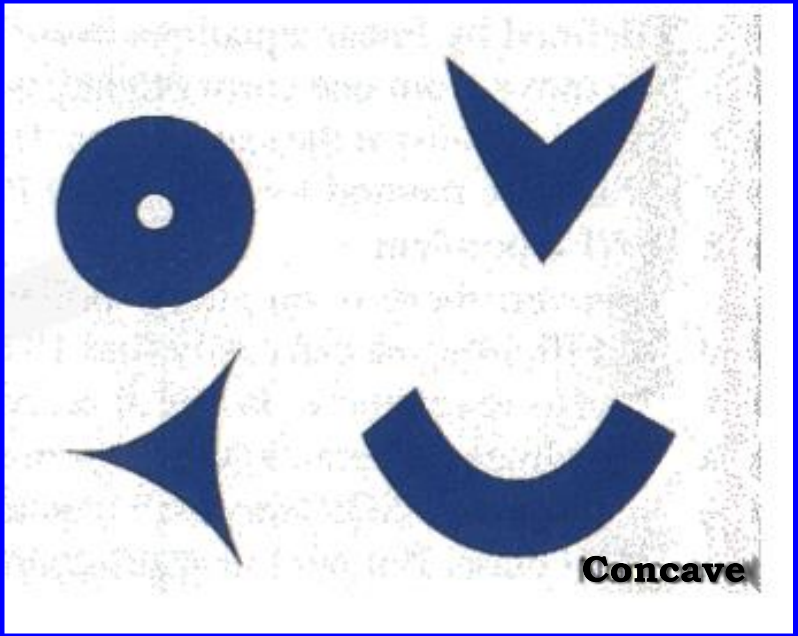
Definiţie: O mulţime $S \subseteq \Re^n$ este convexă dacă orice punct de pe segmentul de dreaptă care uneşte oricare două puncte x^1, $x^2 \in S$ aparţine mulţimii S. Din punct de vedere matematic putem scrie:

$$x^0 = \lambda x^1 + (1-\lambda)x^2 \in S \text{ pentru } \lambda \in [0, 1].$$

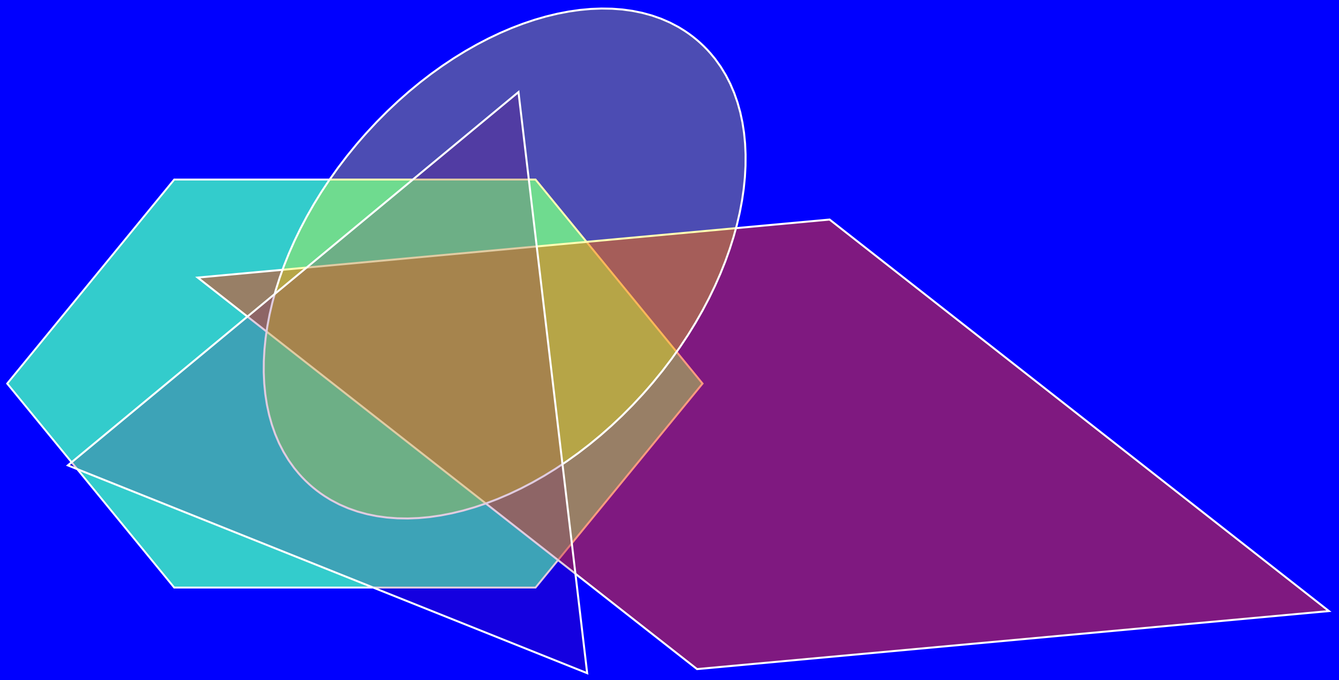


Mulţimi convexe şi concave





Mulţimi convexe



Intersecţia unui număr oricât de mare de mulţimi convexe este o mulţime covexă.

Exemplu:

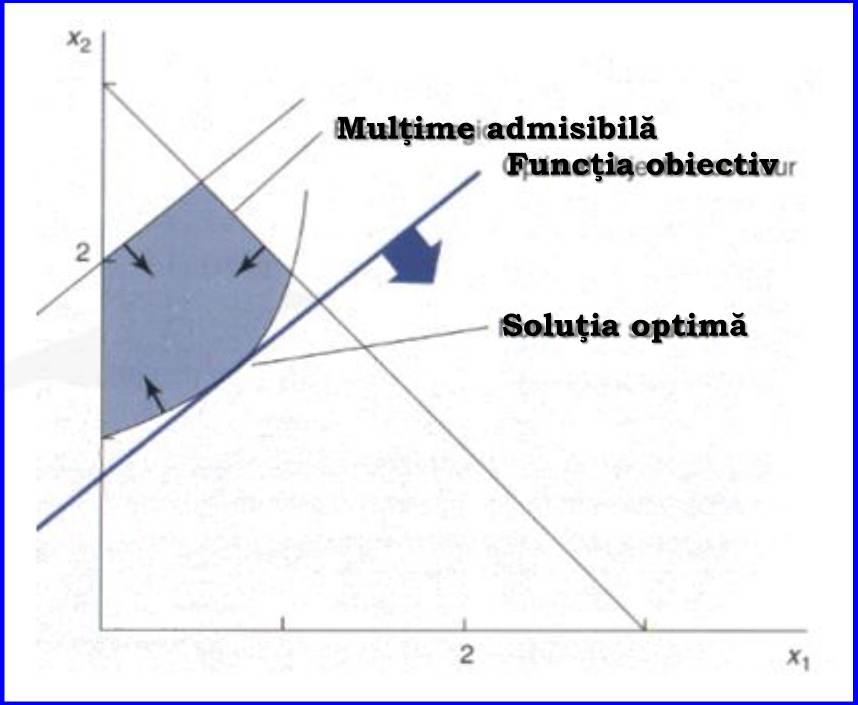
max $(x_1 - x_2)$

$$-x_1^2 - x_2 \geq 1$$

$$x_1 + x_2 \leq 3$$

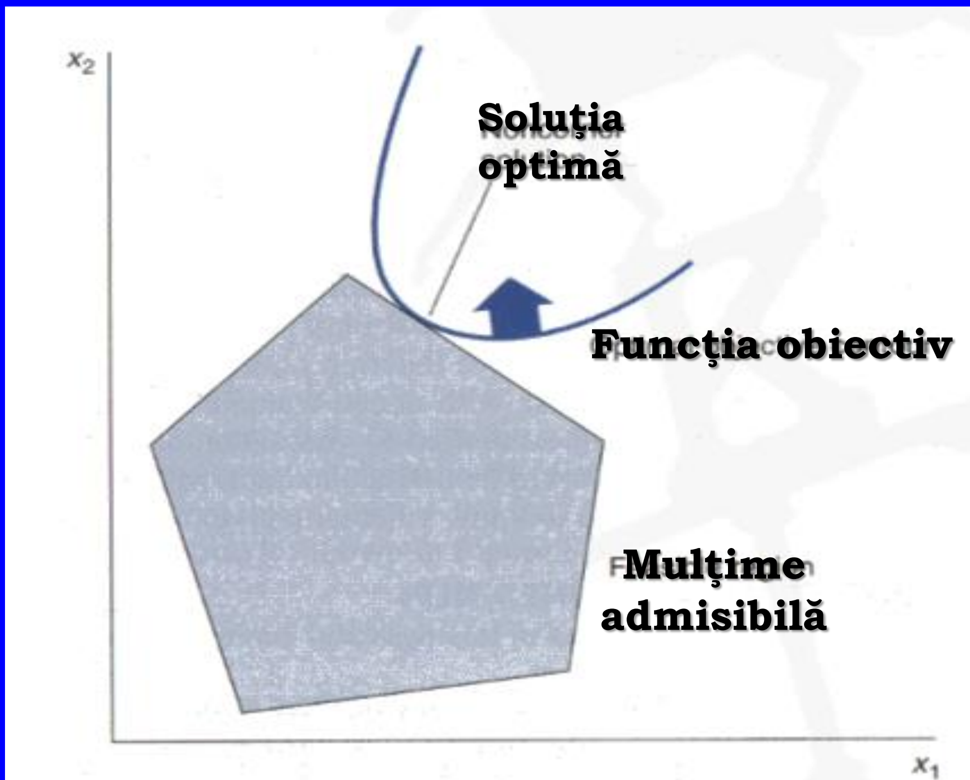
$$-x_1 + x_2 \geq 2$$

$$x_1 \geq 0,\ x_2 \geq 0$$



- **Muţimea admisibilă a soluţiilor nu este un poliedru convex.**
- **Soluţia optimă nu se găseşte în vârful unui poliedru convex**

Funcţie obiectiv neliniară cu restricţii liniare.



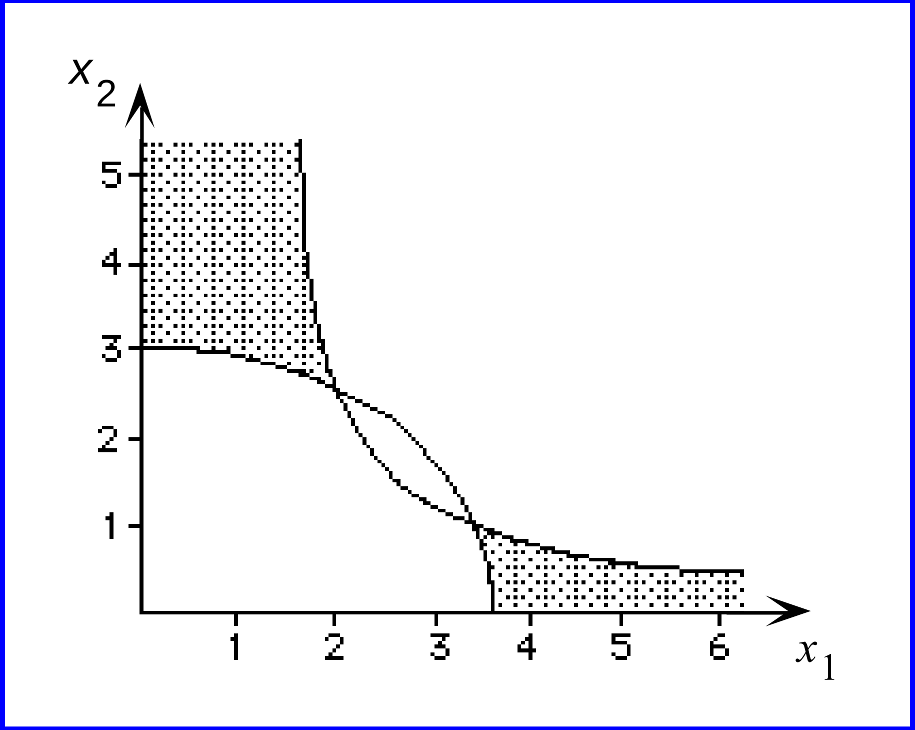
Deşi mulţimea admisibilă este un poliedru convex, soluția optimă nu se găseşte într-unul din vârfurile acestui poliedru.

Există şi situaţii în care soluția optimă se găseşte în interiorul Mulţimii admisibile.

Exemplu de mulțime admisibilă neconvexă

$$S = \{(x_1, x_2) : (0.5x_1 - 0.6)x_2 \le 1$$

$$2(x_1)^2 + 3(x_2)^2 \ge 27;\ x_1, x_2 \ge 0\}$$



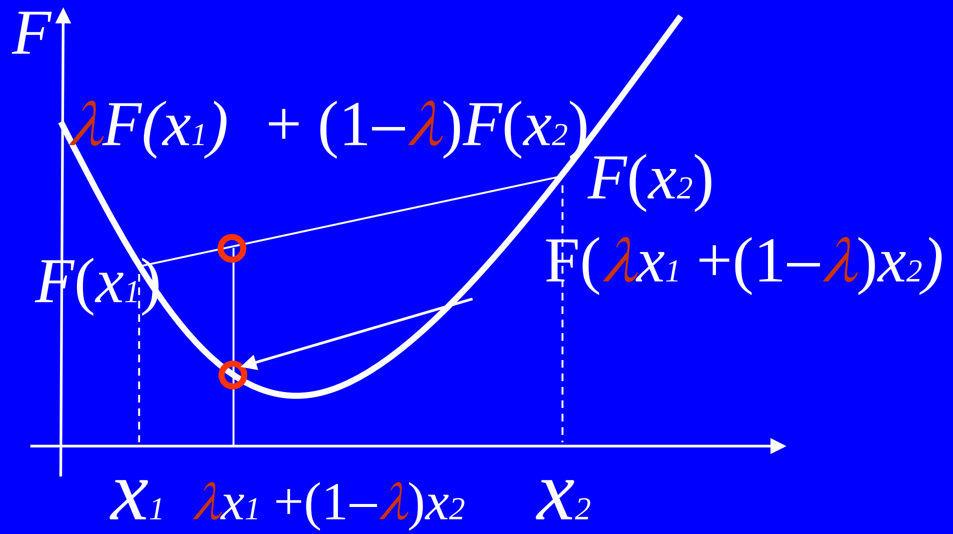
Funcţie convexă

$S \subseteq R^n$ —— **Mulţime convexă**

$F: S \to R$ —— **este o funcţie convexă dacă**

$$F(\lambda x_1 + (1-\lambda)x_2) \le \lambda F(x_1) + (1-\lambda)F(x_2),\ 0 \le \lambda \le 1$$

Dacă este valabilă inegaliatatea atunci funcţia F(X) este strict convexă.



Programarea convexă

Fie $S = \{ X \in \Re^n : g_i(X) \leq b_i,\ i = 1,\ldots,m \}$

Definiţie: Dacă $g_i(X)$ este o funcţie convexă pentru orice $i = 1,\ldots,m$ atunci S este o mulţime convexă.

Teorema programării convexe: Fie $x \in \Re^n$ şi $F(X)$ o funcţie convexă definită pe o muţime convexă de restricţii S. Dacă există o soluţie finită pentru problema:

$$\min \{ F(X) : X \in S \}$$

Atunci toate punctele de optim local sunt şi puncte de optim global. Dacă F(X) este strict convexă, punctul de optim este unic.

Programarea convexă

Min $F(x_1,\ldots,x_n)$

$$g_i(x_1,\ldots,x_n) \le b_i$$
$$i = 1,\ldots,m$$
$$x_1 \ge 0,\ldots,x_n \ge 0$$

este o problemă de programare convexă dacă F este convexă şi fiecare restricţie g_i este convexă.

Max F $(x_1,\ldots,x_n)$

$$g_i(x_1,\ldots,x_n) \le b_i$$
$$i = 1,\ldots,m$$
$$x_1 \ge 0,\ldots,x_n \ge 0$$

Este o problemă de programare convexă dacă F este concavă şi fiecare restricţie g_i este convexă.

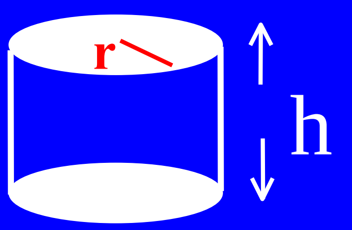
Exemplu:

Să se construiască un cilindru (având şi părţile superioară şi inferioară) având un volum maxim, astfel încât suprafaţa sa să nu depăşească o coală de tablă de dimensiune s.

$$\max \quad V(r,h) = \pi r^2 h$$

$$2\pi r^2 + 2\pi r h = s$$

$$r \geq 0,\ h \geq 0$$



Există mai multe moduri de abordare a acestei probleme. Unul dintre acestea determină înăţimea h din restricţia de suprafaţă şi înlocuişte expresia obţinută în funcția obiectiv.

Soluţia

$$h = \frac{s - 2\pi r^2}{2\pi r} \rightarrow \text{Volumul} = V = \pi r^2 [\frac{s - 2\pi r^2}{2\pi r}] = \frac{rs}{2} - \pi r^3$$

$$\frac{dV}{dr} = 0 \Rightarrow \quad r = (\frac{s}{6\pi})^{1/2} \quad h = \frac{s}{2\pi r} - r = 2(\frac{s}{6\pi})^{1/2}$$

$$V = \pi r^2 h = 2\pi (\frac{s}{6\pi})^{3/2} \qquad r = (\frac{s}{6\pi})^{1/2} \qquad h = 2(\frac{s}{6\pi})^{1/2}$$

Este soluţia obţinută soluţie optimă?

Testul de convexitate

$$V(r) = \frac{rs}{2} - \pi r^3 \rightarrow \frac{\mathrm{d}V(r)}{\mathrm{d}r} = \frac{s}{2} - 3\pi r^2 \rightarrow \frac{\mathrm{d}^2 V(r)}{\mathrm{d}r^2} = -6\pi r$$

$\frac{\mathrm{d}^2 V}{\mathrm{d}r^2} \leq 0$ pentru orice $r \geq 0$

$V(r)$ este concavă pentru orice $r \geq 0$ astfel încât soluția obținută reprezintă maximul global.

Teoreme

Teorema 1.

Pentru o funcţie $F(X)$, derivabilă pe o mulţime convexă, sunt valabile următoarele afirmaţii echivalente:

1. $F(X)$ este convexă;

2. $F(X_2) - F(X_1) \geq (\nabla F(X_1), X_2 - X_1)$, oricare ar fi $X_2, X_1 \in S$
adică valoarea funcţiei $F(X)$ creşte pe direcţia gradientului $\nabla F(X_1)$.
(X, Y) este produsul scalar al vectorilor X şi Y.

3. Produsul scalar $(\nabla F(X+\lambda d), d)$ este nedescrescător în raport cu λ.

4. Matricea Hessiană este semi-pozitiv definită pentru toate punctele $X \in S$, adică:

$$(d, H(X) \cdot d) \geq 0$$

În cazul funcţiilor strict convexe, pentru punctele 2 şi 3 sunt valabile doar inegalităţile

Teorema 2.

Orice punct de minim local este şi punct de minim global. Dacă funcţia F(X) este strict convexă atunci soluţia optimală (dacă există) este unică.

Teorema 3.

Se consideră următoarea problemă de programare convexă:

min (F(X)), $X \in S$

unde: S este mulţimea admisibilă de soluţii (este o mulţime convexă);
F(X) – derivabilă şi convexă;

X* reprezintă o soluţie a problemei convexe de mai sus dacă şi numai dacă sunt adevărate următoarele :

$(\nabla F(X^*), X - X^*) \geq 0$, oricare ar fi $X \in S$

In cazul în care $X \in R$, iar restricţiile lipsesc relaţia devine:

$\nabla F(X^*) \geq 0$

Dacă F(X) este derivabilă şi convexă iar muţimea este strict convexă, atunci X* este unic.

Teorema 3 prezintă condiţiile necesare şi suficiente pentru un punct de minim.

Programarea neliniară convexă fără restricţii

Se consideră următoarea problemă de programare convexă:

$$\min (F(X)),\ \ X \in S \equiv R^n$$

Condiţiile necesare şi suficiente ca un punct X* să fie un minim local şi global sunt:

$$\nabla F(X^*) = 0 \ \text{ şi } \ (d, H(X)d) \geq 0$$

iar pentru un minim global unic, acestea sunt:

$$\nabla F(X^*) = 0 \ \text{ şi } \ (d, H(X)d) > 0$$

Notaţii: $g^{(k)} = \nabla F(X^{(k)})$, $H^{(k)} = \nabla^2 F(X^{(k)})$

Programarea neliniară convexă fără restricţii

În general, pentru rezolvarea unei probleme de programare neliniară se foloseşte relaţia iterativă:

$$X^{(k+1)} = X^{(k)} + \lambda^{(k)} \cdot d^{(k)}$$

unde:

$d^{(k)} \in R^n$ – reprezintă o direcţie de deplasare în iteraţia k din punctul curent $X^{(k)}$;

$\lambda^{(k)}$ – scalar ce reprezintă lungimea pasului de deplasare în iteraţia curentă.

OBSERVAŢII:

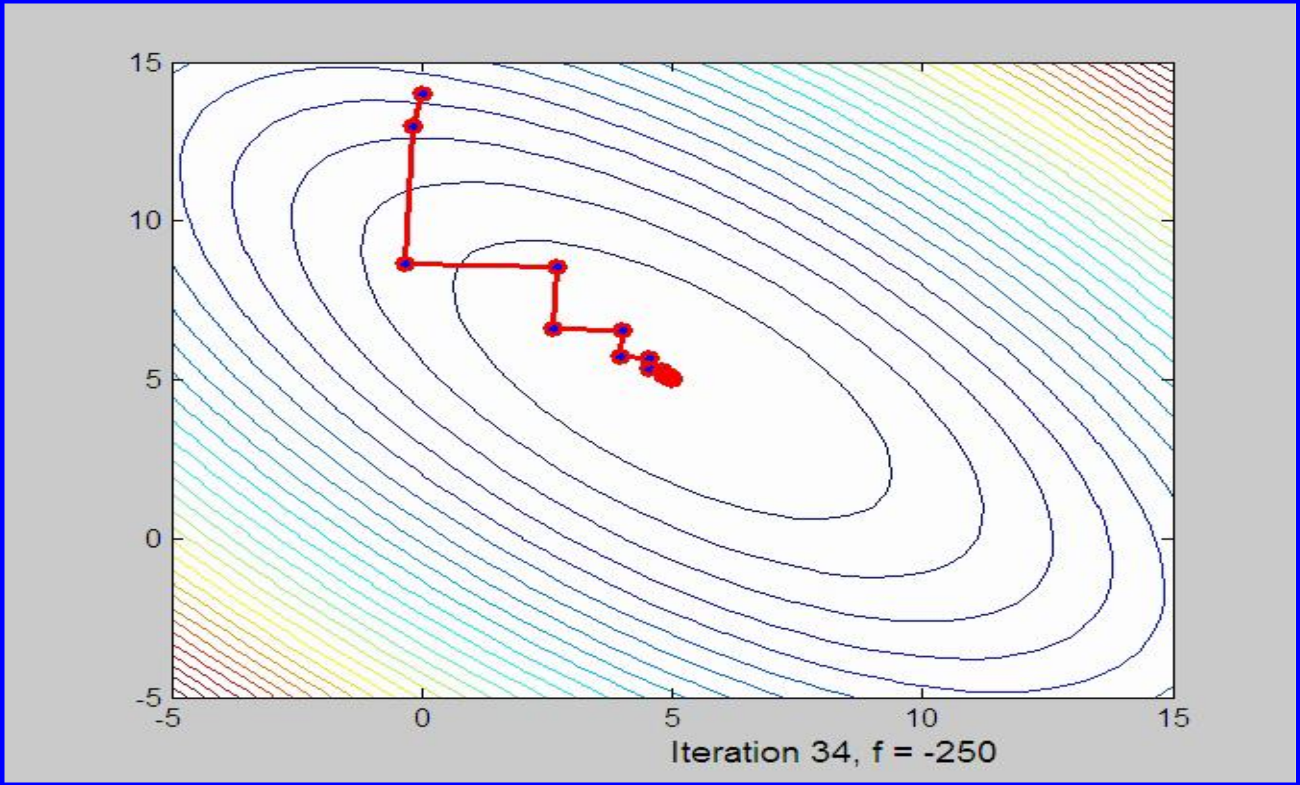
1. Metodele de optimizare diferă prin procedurile concrete de alegere a parametrilor $d^{(k)}$ şi $\lambda^{(k)}$.

2. Având în vedere reducerea efortului de calcul privind implementarea procedurii iterative, în fiecare etapă, informaţia disponibilă pentru obţinerea direcţiei $d^{(k)}$ şi a pasului $\lambda^{(k)}$ este strict limitată la valorile funcţiei şi a primei sale derivate.

Proceduri pentru alegerea pasului de deplasare

Determinarea pasului de deplasare $\lambda^{(k)}$ constituie o parte critică a metodelor de optimizare din cadrul programării neliniare.

O valoare prea mică a pasului conduce la un număr mare de iteraţii în procesul de optimizare, deci la creşterea timpului de calcul, iar o valoare prea mare poate conduce la oscilaţii în jurul minimului şi deci, din nou, la creşterea timpului de calcul.



- Gradientul este întotdeauna perpendicular pe normala la suprafaţa funcţiei obiectiv.
- După fiecare iteraţie noua direcţie a gradientului este întotdeauna perpendiculară pe direcția din iteraţia anterioară.
- În consecință, iteraţiile au tendința de a asigura o deplasare în zig-zag într-un mod foarte ineficient.

Metode de alegere optimală a pasului de deplasare

Aceste metode urmăresc determinarea pasului $\lambda^{(k)}$ Din condiţia realizării minimului funcţiei F(X) pe direcţia $d^{(k)}$:

$$F\left(X^{(k)}+\lambda^{(k)}\,d^{(k)}\right)=\min F\left(X^{(k)}+\lambda\;d^{(k)}\right)=\min F\left(\lambda\right)$$

Dacă se face ipoteza că funcţia *F*(λ) este netedă, metodele cele mai eficiente se bazează pe aproximarea acesteia printr-o funcţie polinomială de gradul 2, al cărei minim se poate calcula uşor.

$$F=C_2\,\lambda^2+C_1\,\lambda+C_0\,,\qquad C_2>0$$

cu minimul

$$\lambda^{(k)}=-\frac{C_1}{2\,C_2}$$

Metoda 1

Pentru determinarea coeficienţilor C_0 , C_1 și C_2 sunt necesare trei valori ale funcţiei respective în punctele λ_1 , λ_2 şi λ_3 , expresia finală fiind:

$$\lambda^{(k)} = \frac{1}{2}\cdot\frac{F_1\left(\lambda_2^2-\lambda_3^2\right)+F_2\left(\lambda_3^2-\lambda_1^2\right)+F_3\left(\lambda_1^2-\lambda_2^2\right)}{F_1\left(\lambda_2-\lambda_3\right)+F_2\left(\lambda_3-\lambda_1^2\right)+F_3\left(\lambda_1-\lambda_2\right)}$$

- **Corectitudinea minimului depinde de aproximarea polinomială;**
- **Eficienţa procesului de optimizare depinde foarte mult de alegerea valorilor λ_2 şi λ_3 (pentru λ_1 se ia, de regulă, o valoare nulă).**

Metoda 2

Constă în folosirea valorii funcţiei şi a derivatei acesteia în punctul curent $X^{(k)}$ precum şi a valorii funcţiei într-un punct învecinat, $(X^{(k)} + \lambda^{(k)} \cdot d^{(k)})$.

$$C_0 = F_0 = F_{(\lambda=0)} = F\left(X^{(k)}\right)$$

$$C_1 = \left.\frac{\partial F}{\partial \lambda}\right|_{\lambda=0} = \lim_{\lambda \to 0} \frac{\sum_{j=1}^{n} \frac{\partial F(X)}{\partial X_j}}{\lambda} = \lim_{\lambda \to 0} \sum_{j=1}^{n} g_j^{(k)} \left(\frac{\lambda d_j^{(k)}}{\lambda} \right) = \left(g^{(k)}, d^{(k)}\right)$$

$$F_1 = F\left(X^{(k+1)}\right) = F\left(X^{(k)} + \lambda\ d^{(k)}\right)$$

Dacă aceste relaţii se introduc în expresia polinomială a funcţiei F(λ) se va obţine:

$$C_2 = \frac{F_1 - \lambda\left(g^{(k)}, d^{(k)}\right) - F_0}{\lambda^2}$$

Metoda 2

Constă în folosirea valorii funcţiei şi a derivatei acesteia în punctul curent $X^{(k)}$ precum şi a valorii funcţiei într-un punct învecinat, $(X^{(k)} + \lambda^{(k)} \cdot d^{(k)})$.

$$C_0 = F_0 = F_{(\lambda=0)} = F\left(X^{(k)}\right)$$

$$C_1 = \left.\frac{\partial F}{\partial \lambda}\right|_{\lambda=0} = \lim_{\lambda \to 0} \frac{\sum_{j=1}^{n} \frac{\partial F(X)}{\partial X_j}}{\lambda} = \lim_{\lambda \to 0} \sum_{j=1}^{n} g_j^{(k)} \left(\frac{\lambda d_j^{(k)}}{\lambda} \right) = \left(g^{(k)}, d^{(k)}\right)$$

$$F_1 = F\left(X^{(k+1)}\right) = F\left(X^{(k)} + \lambda \; d^{(k)}\right)$$

Dacă aceste relaţii se introduc în expresia polinomială a funcţiei F(λ) se va obţine:

$$C_2 = \frac{F_1 - \lambda\left(g^{(k)}, d^{(k)}\right) - F_0}{\lambda^2}$$

Metoda 2

Introducând expresiile corespunzătoare coeficienţilor C_1 şi C_2 în relaţia de minim se obţine:

$$\lambda^{(k)} = \frac{1}{2} \cdot \frac{\lambda^2 \left(g^{(k)}, d^{(k)}\right)}{F_1 - F_0 - \lambda\left(g^{(k)}, d^{(k)}\right)} = \frac{1}{2} \cdot \frac{\lambda}{1 + \frac{1}{\lambda} \frac{F_0 - F_1}{\left(g^{(k)}, d^{(k)}\right)}}$$

Metoda 3

Metoda se bazează pe descompunerea în serii Taylor a funcţiei F, din care se vor reţine primii doi termeni:

$$F\left(X^{(k)} + \lambda\ d^{(k)}\right) = F\left(X^{(k)}\right) + \lambda\left(g^{(k)}, d^{(k)}\right) + 0{,}5\,\lambda^2\left(d^{(k)}, H^{(k)} d^{(k)}\right)$$

Metoda 3

Se derivează relaţia în raport cu λ şi se anulează:

$$\lambda^{(k)} = -\frac{\left(g^{(k)}, d^{(k)}\right)}{\left(d^{(k)}, H^{(k)} d^{(k)}\right)}, \quad \left(d^{(k)}, H^{(k)} d^{(k)}\right) > 0$$

Pentru aplicarea acestei metode trebuie să se dispună de valorile derivatei şi a hessianului funcţiei F în punctul $X^{(k)}$.

Metode de explorare directă

Plecând de la un interval iniţial $[a_0, b_0]$ care conţine valoarea optimă λ^*, metodele de explorare directă reduc iterativ intervalul în care se găseşte λ^* până când este satisfăcută o condiţie de oprire.

Reducerea se face prin eliminarea simultană sau succesivă a unor subintervale pentru care există certitudinea că nu pot conţine valoarea optimă λ^*. Aplicarea reducerii se bazează pe unimodalitatea funcţiei (are un singur extrem în interval).

Metode pentru rezolvarea problemelor de optimizare convexă fără restricţii

Metode de ordinul 0 care folosesc valorile funcţiei în punctul curent şi vecinătăţi;

Metode de ordinul 1 care folosesc şi derivata de ordinul 1 (gradientul);

Metode de ordinul 2 care folosesc derivatele de ordinul 1 şi 2 (hessianul).

Metode de ordinul 0

Avantaje:

- nu sunt necesare expresiile analitice ale derivatei funcţiei obiectiv *F*(*X*).
- nu există condiţii referitoare la continuitatea şi derivabilitatea funcţiei *F*(*X*).
- printr-o alegere corespunzătoare a criteriilor de stop, se poate evita terminarea iteraţiilor într-un punct de inflexiune.

Dezavantaj:

- au o viteză de convergenţă scăzută.

Metodele de ordinul 0

Au la bază relaţia iterativă :

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

Stabilirea noii direcţii de explorare se face pe baza unei serii de evaluări a funcţiei obiectiv $F(X)$, într-o manieră specifică fiecărei metode, iar lungimea paşilor $\lambda^{(k)}$ este proprie fiecărei metode.

Optimizarea ciclică de-a lungul axelor de coordonate

În cadrul metodei se face o explorare unidimensională pe direcţii care coincid cu axele de coordonate.
În relaţia iterativă direcţiile $d^{(k)}$ sunt succesiv axele sistemului ortogonal de coordonate:

$$d^{(0)} = \begin{bmatrix}1 & 0 & 0 & \ldots & 0\end{bmatrix}^t, d^{(1)} = \begin{bmatrix}0 & 1 & 0 & \ldots & 0\end{bmatrix}^t, \ldots, d^{(n-1)} = \begin{bmatrix}0 & 0 & 0 & \ldots & 1\end{bmatrix}^t$$

Paşii $\lambda^{(k)}$ pe direcţiile respective sunt determinaţi cu una din metodele prezentate.
Procesul iterativ se va încheia când va fi îndeplinit criteriul de stop impus în n iteraţii.
Dacă suprafeţele respective prezintă o vale sau o creastă care nu este paralelă cu axele de coordonate, metoda nu poate fi aplicată.

Exemplu numeric

Să se determine minimul funcţiei:

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

folosind metoda de optimizare ciclică de-a lungul axelor de coordonate, indicându-se ca punct de pornire X_0 = [5, 4].

Iteraţia 1

Se face deplasarea pe direcţia $d^{(0)} = [1 \quad 0]^t$, astfel încât pentru determinarea unei noi aproximaţii se va utiliza relaţia iterativă:

$$X^{(1)} = X^{(0)} + \lambda^{(0)} d^{(0)} = \begin{bmatrix} 5 \\ 4 \end{bmatrix} + \lambda^{(0)} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 + \lambda^{(0)} \\ 4 \end{bmatrix}$$

Exemplu numeric

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

$$F(\lambda^{(0)}) = (5+\lambda^{(0)})^2 + 4\cdot 4^2 - 4 = \left(\lambda^{(0)}\right)^2 + 10\cdot\lambda^{(0)} + 85$$

În continuare se face derivata expresiei $F(\lambda^{(0)})$ în raport cu $\lambda^{(0)}$ şi se anulează.

$$\frac{\partial F\left(\lambda^{(0)}\right)}{\delta\,\lambda^{(0)}} = 2\cdot\lambda^{(0)} + 10 = 0 \quad\rightarrow\quad \lambda^{(0)} = -5$$

Valoarea pasului de deplasare se introduce în relaţia iterativă obţinându-se astfel valoarea noii aproximaţii $X^{(1)}$:

$$X^{(1)} = \left[x_1^{(1)} \quad x_2^{(1)}\right]_t = \left[0 \quad 4\right]^t$$

Exemplu numeric

Iteraţia 2

Se face deplasarea pe direcţia $d^{(1)} = [0 \quad 1]^t$, astfel încât pentru determinarea unei noi aproximaţii se va utiliza relaţia iterativă:

$$X^{(2)} = X^{(1)} + \lambda^{(1)} d^{(1)} = \begin{bmatrix} 0 \\ 4 \end{bmatrix} + \lambda^{(1)} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 + \lambda^{(1)} \end{bmatrix}$$

Pentru determinarea lungimii pasului de deplasare $\lambda^{(1)}$, se introduc valorile $x_1^{(2)}$ şi $x_2^{(2)}$, în expresia funcţiei obiectiv F:

$$F(\lambda^{(1)}) = 0 + 4 \cdot \left(4 + \lambda^{(1)}\right)^2 - 4 = \left(\lambda^{(1)}\right)^2 + 8 \cdot \lambda^{(1)} + 15$$

În continuare se face derivata expresiei $F(\lambda^{(1)})$ în raport cu $\lambda^{(1)}$ şi se anulează.

$$\frac{\partial F\left(\lambda^{(1)}\right)}{\delta \lambda^{(1)}} = 2 \cdot \lambda^{(1)} + 8 = 0 \quad \rightarrow \quad \lambda^{(0)} = -4$$

Exemplu numeric

Valoarea pasului de deplasare se introduce în relaţia iterativă obţinându-se astfel valoarea noii aproximaţii $X^{(2)}$:

$$X^{(2)} = \begin{bmatrix} x_1^{(2)} & x_2^{(2)} \end{bmatrix}^t = \begin{bmatrix} 0 & 0 \end{bmatrix}^t$$

Metoda direcţiilor conjugate

Fie o funcţie obiectiv care are o expresie pătratică de forma:

$$F(X)=\frac{1}{2}X^{t}\cdot C\cdot X+X\cdot D$$

În cadrul acestei metode se pleacă dint-un punct iniţial $X^{(0)}$ şi se face o deplasare pe direcţia $d^{(0)}$, presupusă deocamdată arbitrară,

$$X^{(1)}=X^{(0)}+\lambda^{(0)}d^{(0)}$$

unde pasul $\lambda^{(0)}$ se determină cu una din metodele indicate anterior.

Metoda direcţiilor conjugate

Următoarea direcţie de deplasare se va determina din următoarea condiţie de conjugare:

$$\left(d^{(1)}, C \cdot d^{(0)}\right) = 0$$
$$\left(d^{(i)}, C \cdot d^{(j)}\right) = 0, \quad 0 \le i \ne j \le n-1$$

Folosirea direcţiilor conjugate asigură găsirea minimului unei funcţii pătratice de *n* variabile în *n* iteraţii, deplasarea în interiorul fiecarei iteraţii făcându-se pe o direcţie conjugată cu cea din iteraţia precedentă.

Exemplu numeric

Să se determine minimul funcţiei:

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

folosind metoda direcţiilor conjugate, indicându-se ca punct de pornire X_0 = [5, 4].

Iteraţia 1

Se face deplasarea pe o direcţie arbitrară $d^{(0)} = [d^{(0)}_1 \ d^{(0)}_2]$. Pentru a evidenţia acest lucru se consideră o direcţie oarecare:

$$d^{(0)} = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}^t \qquad \left| d^{(0)} \right| = 1$$

Exemplu numeric

Mărimea pasului de deplasare se va face cu ajutorul relaţiei:

$$\lambda^{(k)} = -\frac{\left(g^{(k)}, d^{(k)}\right)}{\left(d^{(k)}, H^{(k)} d^{(k)}\right)}$$

$$g = \begin{bmatrix} \dfrac{\partial F(X)}{\partial X_1} & \dfrac{\partial F(X)}{\partial X_2} \end{bmatrix}_t = \begin{bmatrix} 2X_1 & 8X_2 \end{bmatrix}^t$$

$$H = C = \begin{bmatrix} \dfrac{\partial^2 F(X)}{\partial X_1^2} & \dfrac{\partial^2 F(X)}{\partial X_1 \partial X_2} \\ \dfrac{\partial^2 F(X)}{\partial X_2 \partial X_1} & \dfrac{\partial^2 F(X)}{\partial X_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$$

$$\lambda^{(0)} = -\frac{\begin{bmatrix} 10 & 32 \end{bmatrix} \begin{bmatrix} \dfrac{1}{2} & \dfrac{\sqrt{3}}{2} \end{bmatrix}^t}{\begin{bmatrix} \dfrac{1}{2} & \dfrac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix} \begin{bmatrix} \dfrac{1}{2} & \dfrac{\sqrt{3}}{2} \end{bmatrix}^t} = -5{,}033$$

Exemplu numeric

Noua aproximaţie se va calcula utilizând relaţia iterativă:

$$X^{(1)} = X^{(0)} + \lambda^{(0)} d^{(0)} = \begin{bmatrix} 5 \\ 4 \end{bmatrix} + (-5{,}033) \cdot \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}^t = \begin{bmatrix} 2{,}484 \\ -0{,}358 \end{bmatrix}$$

Iterația 2

Se determină direcţia de deplasare *$d^{(1)}$*. Aceasta trebuie să fie conjugată cu direcţia *$d^{(0)}$*, adică trebuie să fie satisfăcută relaţia:

$$\begin{bmatrix} d_1^{(1)} & d_2^{(1)} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} = 0$$

Exemplu numeric

La condiţia de conjugare se mai adaugă condiţia de vector de modul unitate pentru $d^{(1)}$ dată de relaţia:

$$\left[d_1^{(1)} \quad d_2^{(1)}\right]\left[d_1^{(1)} \quad d_2^{(1)}\right]^t = 1$$

Rezolvarea sistemului format din ultimele două ecuaţii ne va conduce la găsirea direcţiei de deplasare $d^{(1)} = [d^{(1)}_1 \ d^{(1)}_2]$.

$$d^{(1)} = \left[-0{,}9897 \quad 0{,}1428\right]^t$$

Pentru determinarea valorii pasului de deplasare $\lambda^{(1)}$ se va calcula gradientul funcţiei în punctul $X^{(1)}$:

$$g^{(1)} = \left[2.2{,}484 \quad -8.0{,}358\right]^t = \left[4{,}9868 \quad -2{,}864\right]^t$$

$$\lambda^{(1)} = -\frac{\left[4{,}968 \quad -2{,}864\right]\left[-0{,}9897 \quad 0{,}1428\right]^t}{\left[-0{,}9897 \quad 0{,}1428\right]\begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}\begin{bmatrix} -0{,}9897 \\ 0{,}1428 \end{bmatrix}^t} = 2{,}5096$$

Exemplu numeric

Valorile pasului de deplasare şi a direcţiei se vor introduce în relaţia iterativă obţinându-se:

$$X^{(2)} = \begin{bmatrix} 2,484 \\ -0,358 \end{bmatrix} + 2,5096 \begin{bmatrix} -0,9897 \\ 0,1428 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

METODE DE ORDINUL 1

Pentru determinarea direcţiei şi a pasului de deplasare, metodele de ordinul unu apelează la calculul derivatelor de ordinul unu.

pentru o funcţie obiectiv F(X) continuă şi derivabilă, gradientul într-un punct curent $X^{(k)}$ reprezintă vectorul derivatelor parţiale de ordinul unu în punctul respectiv.

Acest vector este ortogonal la conturul lui $F(X)$ ce trece prin punctul $X^{(k)}$.

Direcţia gradientului corespunde celei mai rapide creşteri a lui $F(X)$, ceea ce permite utilizarea ei pentru maximizare sau a direcţiei opuse, pentru minimizare.

Metoda gradientului simplu

Metoda gradientului simplu este cea mai simplă metodă de ordinul 1 pentru minimizarea unei funcții de mai multe variabile.

Metoda are la bază relaţia de recurenţă:

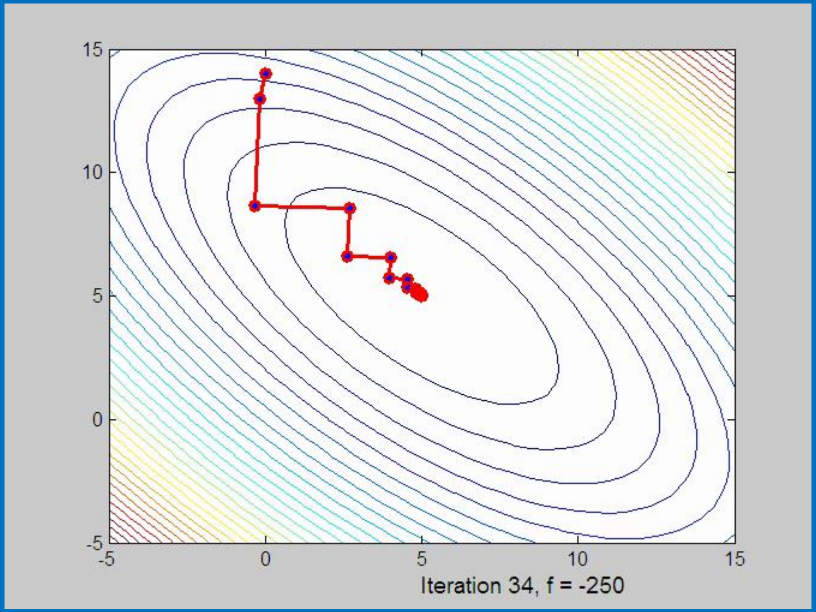
$$X^{(k+1)} = X^{(k)} + \lambda^{(k)} d^{(k)}, \quad k = 1, 2, 3, \ldots$$

unde direcţia $d^{(k)} = -\nabla F(X^{(k)})$

Avantaj: simplu de aplicat

Dezavantaj: convergenţa lentă.

Metoda gradientului simplu



- Gradientul este întotdeauna perpendicular pe normala la suprafaţa funcţiei obiectiv.
- După fiecare iteraţie noua direcţie a gradientului este întotdeauna perpendiculară pe direcția din iteraţia anterioară.
- În consecință, iteraţiile au tendința de a asigura o deplasare în zig-zag într-un mod foarte ineficient.

Metoda gradientului conjugat

În cadrul metodei gradientului conjugat noua direcţie de deplasare poate fi determinată dacă se utilizează o combinaţie între gradientul calculat în punctul *X*$^{(k)}$ şi direcţia precedentă,

$$d^{(k)} = -g^{(k)} + \beta^{(k)} d^{(k-1)}, k = 1, 2, \ldots$$

$$d^{(0)} = -g^{(0)}$$

Există mai multe variante ale metodelor de gradienţi conjugaţi, acestea deosebindu-se prin modul de determinare a parametrului scalar $\beta^{(k)}$:

- **metoda Fletcher-Reeves:**

$$\beta^{(k)} = \frac{\left(g^{(k)},\ g^{(k)}\right)}{\left(g^{(k-1)},\ g^{(k-1)}\right)}$$

- **metoda Polak-Ribière:**

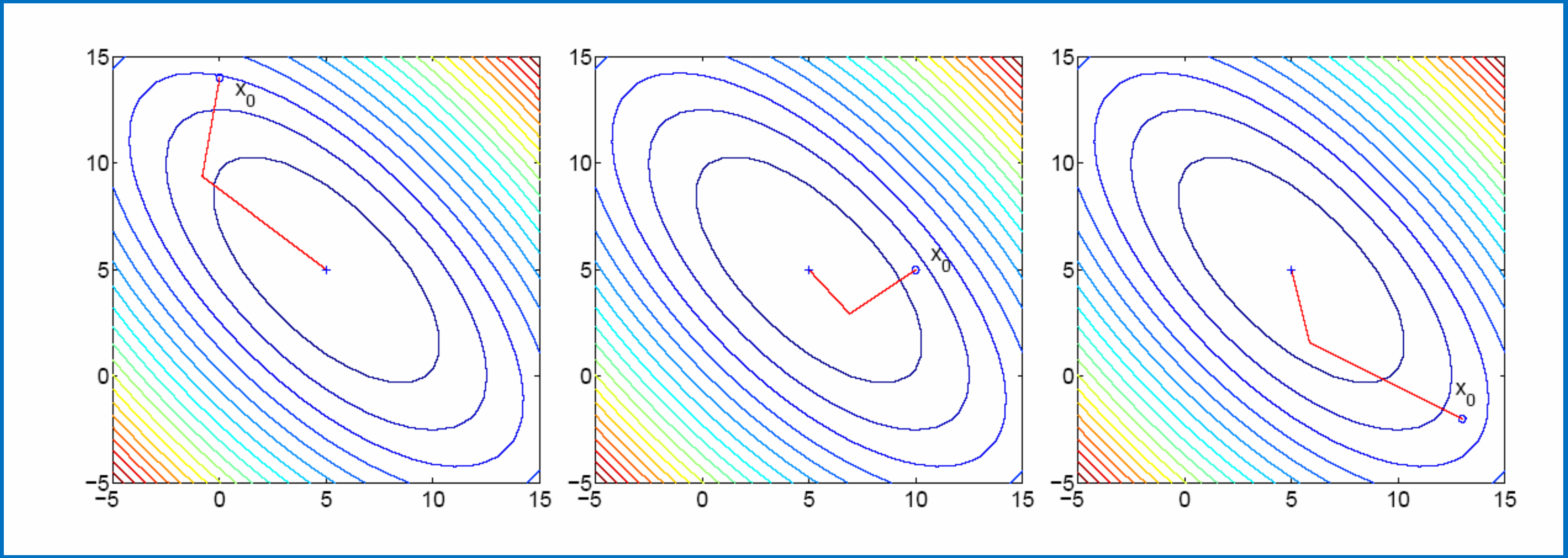
$$\beta^{(k)} = \frac{\left(g^{(k)},\ g^{(k)} - g^{(k-1)}\right)}{\left(g^{(k-1)},\ g^{(k-1)}\right)}$$

- **metoda Hestenes-Stiefel:**

$$\beta^{(k)} = \frac{\left(g^{(k)},\ g^{(k)} - g^{(k-1)}\right)}{\left(g^{(k)} - g^{(k-1)},\ d^{(k-1)}\right)}$$

Metode ale gradinetului conjugat

- Minimul unei funcţii pătratice *F*(*X*) poate fi determinat în cel mult *n* iteraţii, convergenţa lor fiind bună.



- Se pleacă din 3 puncte diferite;
- Minimul se obţine în exact 3 iteraţii

Exemplu:

Să se minimizeze funcţia:

min $\quad F(x_1, x_2, x_3) = (x_1)^2 + x_1(1 - x_2) + (x_2)^2 - x_2 x_3 + (x_3)^2 + x_3$

folosind metoda gradientului simplu.

Se consideră aproximaţia iniţială: $X^0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

Se calculează gradientul şi se determină direcţia de deplasare:

$$\nabla F(X)=\begin{bmatrix}2x_1+(1-x_2) & -x_1+2x_2-x_3 & -x_2+2x_3+1\end{bmatrix}$$

$$\mathbf{d}^0=-\nabla F(X^0)=-\begin{bmatrix}2(0)+1-0 & -0+0-0 & -0+0+1\end{bmatrix}$$

$$=-\begin{bmatrix}1 & 0 & 1\end{bmatrix}=\begin{bmatrix}-1 & 0 & -1\end{bmatrix}$$

$$X^1=\begin{bmatrix}0 & 0 & 0\end{bmatrix}+\lambda^0\cdot\begin{bmatrix}-1 & 0 & -1\end{bmatrix}$$

In continuare se detemină valoarea lui λ^0

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

$$F(X^1) = (\lambda^0)^2 + (-\lambda^0)(1) + 0 - 0 + (\lambda^0)^2 + (-\lambda^0)$$

$$= 2(\lambda^0)^2 - 2(\lambda^0)$$

În continuare se face derivata expresiei $F(\lambda^{(0)})$ în raport cu $\lambda^{(0)}$ şi se anulează.

$$\frac{df(X^1)}{d\lambda^0} = 4(\lambda^0) - 2$$

$$4(\lambda^0) = 2 \quad \Rightarrow \quad \lambda^0 = 2/4 = 1/2$$

Valoarea pasului de deplasare se introduce în relaţia iterativă obţinându-se astfel valoarea noii aproximaţii $X^{(1)}$:

$$X^1 = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} + \lambda^0 \cdot \begin{bmatrix} -1 & 0 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}$$

$$X^1 = \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}$$

Se determină noua direcţie de deplasare:

$$\mathbf{d}^1 = -\nabla F(X^1) = -\begin{bmatrix} -1+1+0 & \frac{1}{2}+0+\frac{1}{2} & 0-1+1 \end{bmatrix}$$

$$\mathbf{d}^1 = \begin{bmatrix} 0 & -1 & 0 \end{bmatrix}$$

Se determină noua aproximaţie:

$$X^2 = \left[-\frac{1}{2} \quad 0 \quad -\frac{1}{2}\right] + \lambda^1 \cdot \left[0 \quad -1 \quad 0\right]$$

$$= \left[-\frac{1}{2} \quad -\lambda^1 \quad -\frac{1}{2}\right]$$

Pentru determinarea lungimii pasului de deplasare $\lambda^{(1)}$, se introduc valorile $x_1^{(2)}$ şi $x_2^{(2)}$, în expresia funcţiei obiectiv F:

$$F(X^2) = \frac{1}{4} + \left(-\frac{1}{2}\right)\left(1 + \lambda^1\right) + (\lambda^1)^2 - (\lambda^1)\left(\frac{1}{2}\right) + \frac{1}{4} - \frac{1}{2}$$

$$= (\lambda^1)^2 - \lambda^1 - 1/2$$

În continuare se face derivata expresiei $F(\lambda^{(1)})$ în raport cu $\lambda^{(1)}$ şi se anulează.

$$\frac{dF(X^1)}{d\lambda^1} = 2(\lambda^1) - 1$$

$$2(\lambda^1) = 1 \Rightarrow \lambda^1 = 1/2$$

Valoarea pasului de deplasare se introduce în relaţia iterativă obţinându-se astfel valoarea noii aproximaţii $X^{(2)}$:

$$X^2 = \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix} + \lambda^1 \cdot \begin{bmatrix} 0 & -1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix} + \begin{bmatrix} 0 & -\frac{1}{2} & 0 \end{bmatrix}$$

$$X^2 = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

Se determină noua direcţie de deplasare:

$$\mathbf{d}^2 = -\nabla F(X^2) = -\begin{bmatrix} -1+1+\frac{1}{2} & \frac{1}{2}-1+\frac{1}{2} & \frac{1}{2}-1+1 \end{bmatrix}$$

$$\mathbf{d}^2 = \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}$$

Se determină noua aproximaţie:

$$X^3 = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} + \lambda^2 \cdot \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{1}{2}(\lambda^2+1) & -\frac{1}{2} & -\frac{1}{2}(\lambda^2+1) \end{bmatrix}$$

Pentru determinarea lungimii pasului de deplasare $\lambda^{(2)}$, se introduc valorile $x_1^{(3)}$ şi $x_2^{(3)}$, în expresia funcţiei obiectiv F:

$$F(X^3) = \frac{1}{2}(\lambda^2 + 1)^2 - \frac{3}{2}(\lambda^2 + 1) + \frac{1}{4}$$

În continuare se face derivata expresiei $F(\lambda^{(2)})$ în raport cu $\lambda^{(2)}$ şi se anulează.

$$\frac{dF(X^3)}{d\lambda^2} = (\lambda^2 + 1) - \frac{3}{2}$$

$$(\lambda^2 + 1) = \frac{3}{2} \quad \Rightarrow \quad \lambda^2 = 1/2$$

Se determină noua aproximaţie:

$$X^3 = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} + \lambda^2 \cdot \begin{bmatrix} -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} + \begin{bmatrix} -\frac{1}{4} & 0 & -\frac{1}{4} \end{bmatrix}$$

$$X^3 = \begin{bmatrix} -\frac{3}{4} & -\frac{1}{2} & -\frac{3}{4} \end{bmatrix}$$

Se determină noua direcţie de deplasare:

$$\mathbf{d}^3 = -\nabla F(X^3) = -\begin{bmatrix} 0 & \frac{1}{2} & 0 \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{2} & 0 \end{bmatrix}$$

$$X^4 = \begin{bmatrix} -\frac{3}{4} & -\frac{1}{2} & -\frac{3}{4} \end{bmatrix} + \lambda^3 \cdot \begin{bmatrix} 0 & -\frac{1}{2} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{3}{4} & -\frac{1}{2}(\lambda^3 + 1) & -\frac{3}{4} \end{bmatrix}$$

Pentru determinarea lungimii pasului de deplasare $\lambda^{(3)}$, se introduc valorile $x_1^{(4)}$ şi $x_2^{(4)}$, în expresia funcţiei obiectiv F:

$$F(X^4)=\frac{1}{4}(\lambda^3+1)^2-\frac{3}{2}(\lambda^3)-\frac{3}{2}$$

$$\frac{dF(X^4)}{d\lambda^3}=\frac{1}{2}(\lambda^3+1)-\frac{9}{8}=0$$

$$\Rightarrow \lambda^3=\frac{5}{4}$$

Se determină noua aproximaţie:

$$X^4=\begin{bmatrix}-\frac{3}{4} & -\frac{1}{2} & -\frac{3}{4}\end{bmatrix}+\begin{bmatrix}0 & -\frac{5}{8} & 0\end{bmatrix}$$

$$X^4=\begin{bmatrix}-\frac{3}{4} & -\frac{9}{8} & -\frac{3}{4}\end{bmatrix}$$

Se determină noua direcţie de deplasare:

$$\mathbf{d}^4 = -\nabla F(X^4) = -\begin{bmatrix} \frac{5}{8} & -\frac{3}{4} & \frac{5}{8} \end{bmatrix} = \begin{bmatrix} -\frac{5}{8} & \frac{3}{4} & -\frac{5}{8} \end{bmatrix}$$

Se determină noua aproximaţie:

$$X^5 = \begin{bmatrix} -\frac{3}{4} & -\frac{9}{8} & -\frac{3}{4} \end{bmatrix} + \lambda^4 \cdot \begin{bmatrix} -\frac{5}{8} & \frac{3}{4} & -\frac{5}{8} \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{1}{4}(3+\frac{5}{2}\lambda^4) & -\frac{3}{4}(\frac{3}{2}-\lambda^4) & -\frac{1}{4}(3+\frac{5}{2}\lambda^4) \end{bmatrix}$$

Pentru determinarea lungimii pasului de deplasare $\lambda^{(4)}$, se introduc valorile $x_1^{(5)}$ şi $x_2^{(5)}$, în expresia funcţiei obiectiv F:

$$F(X^5) = \frac{73}{32}(\lambda^4)^2 - \frac{43}{32}\lambda^4 - \frac{51}{64}$$

$$\frac{dF(\mathbf{x}^5)}{d\lambda^4} = \frac{73}{16}\lambda^4 - \frac{43}{32} = 0 \quad \Rightarrow \quad \lambda^4 = {}^{43}\!/_{146}$$

Se determină noua aproximaţie:

$$X^5 = \begin{bmatrix} -\frac{3}{4} & -\frac{9}{8} & -\frac{3}{4} \end{bmatrix} + \frac{43}{146} \cdot \begin{bmatrix} -\frac{5}{8} & \frac{3}{4} & -\frac{5}{8} \end{bmatrix}$$

$$X^5 = \begin{bmatrix} -\frac{1091}{1168} & -\frac{66}{73} & -\frac{1091}{1168} \end{bmatrix}$$

Se verifică dacă criteriul de convergenţă este satisfăcut:

$$||\nabla F(\mathbf{X}^5)||:$$

$$\nabla F(X^5) = \begin{bmatrix} \frac{21}{584} & \frac{35}{584} & \frac{21}{584} \end{bmatrix}$$

$$\left\|\nabla F(X^5)\right\| = \sqrt{\left(21/584\right)^2 + \left(35/584\right)^2 + \left(21/584\right)^2} = 0.0786$$

$||\nabla F(X^5)||$ = 0.0786, are o valoare foarte mică, apropiată de 0, astfel încât:

$$X = \begin{bmatrix} -\frac{1091}{1168} & -\frac{66}{73} & -\frac{1091}{1168} \end{bmatrix}$$

este foarte aproape de valoare de optim, obţinută analitic:

$$\mathbf{x}^* = \begin{bmatrix} -1 & -1 & -1 \end{bmatrix}$$

METODE DE ORDINUL 2

Metode de tip Newton

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

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

Minimul se obţine prin derivarea expresiei în raport cu componentele lui $d^{(k)}$ şi anularea derivatelor respective:

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

Relaţia care stă la baza determinării direcţiei se introduce în formula iterativă rezultând:

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

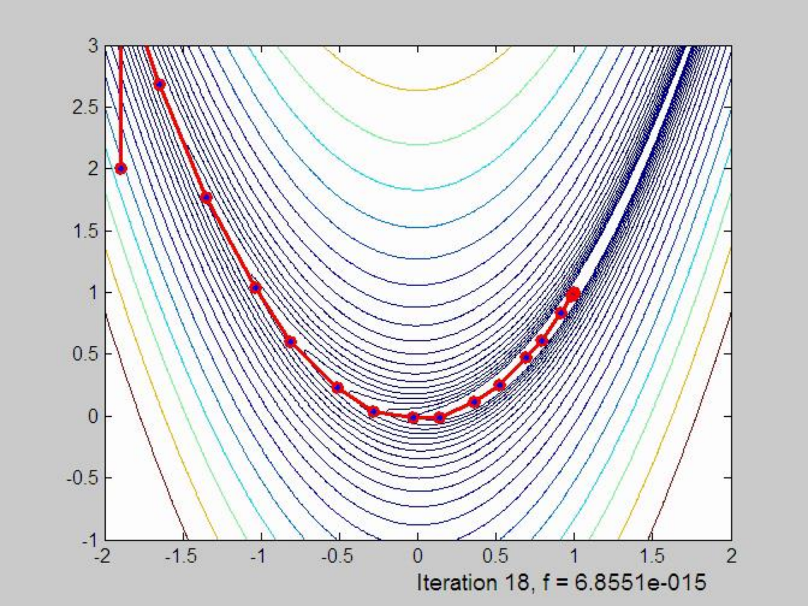
Metoda Newton

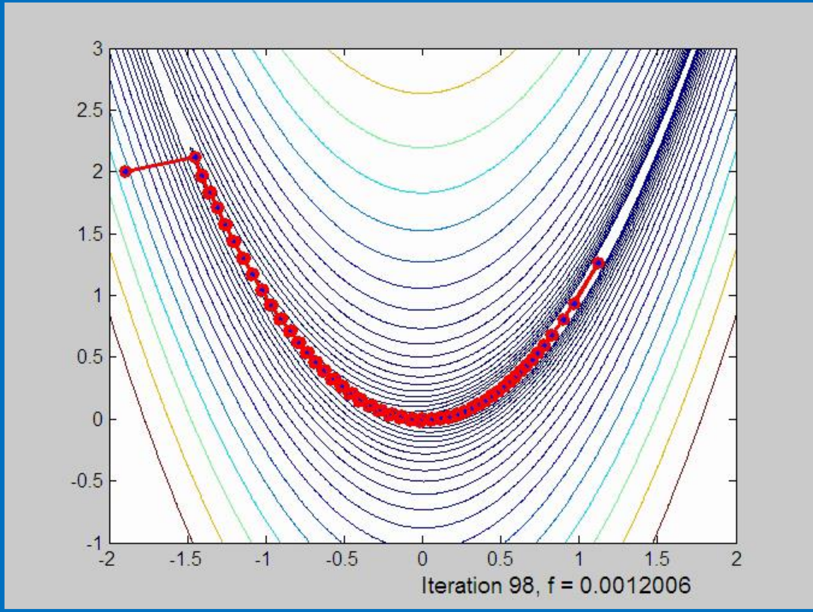
În cazul unor funcţii obiectiv pătratice, minimul va fi atins încă din prima iteraţie, pentru $\lambda^{(0)} = 1$.

Observaţii:

- **Aplicare metodei Newton cere ca matricea $H^{(k)}$ să fie pozitiv definită la fiecare iteraţie.**
- **Convergenţa metodei în apropierea optimului va fi sensibil mai bună decât a metodelor de ordinul 1.**
- **disponibilitatea derivatelor de ordinul doi permite verificarea condiţiilor de suficienţă a optimului.**

Metoda Newton





Metoda Newton – 18 iteraţii **Metoda gradientului conjugat – 98 iteraţii**

Metoda Newton

Dezavantaje:

-volumului de calcul necesar obţinerii derivatelor de ordinul doi şi pentru inversarea matricei $H^{(k)}$.

-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 $H^{(k)} = 0$).

- pot apărea unele cazuri în care $F(X^{(k+1)}) > F(X^{(k)})$, deşi $H^{(k)} > 0$, aproximaţia pătratică fiind valabilă doar într-o zonă foarte limitată.

Metoda Newton

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 $X^{(0)} = [5, 4]$, iar $\lambda^{(0)} = 1$.

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

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

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

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

Exemplul 2.

Să se determine minimul funcţiei:

$$f(x_1, x_2, x_3) = (x_1)^2 + x_1(1 - x_2) + (x_2)^2 - x_2x_3 + (x_3)^2 + x_3$$

Se calculează gradientul:

$$\nabla f(x_1, x_2, x_3) = \begin{bmatrix} 2x_1 - x_2 + 1 & -x_1 + 2x_2 - x_3 & -x_2 + 2x_3 + 1 \end{bmatrix}$$

Matricea Hessiană:

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

Matricea Hessiană inversată:

$$\left[\nabla^2 f(\mathbf{x})\right]^{-1} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 3/4 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 3/4 \end{bmatrix}$$

Se alege punctul iniţial de pornire:

$$\mathbf{x}^0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Se calculează valoarea gradientului în punctul curent:

$$\nabla f(\mathbf{x}^0) = \begin{bmatrix} 0-0+1 & -0+0-0 & -0+0+1 \end{bmatrix}$$

$$\Rightarrow \nabla f(\mathbf{x}^0) = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}$$

Se scrie relaţia iterativă pentru determinarea noii aproximaţii:

$$\mathbf{x}^1 = \mathbf{x}^0 - \left[\nabla^2 f(\mathbf{x}^0)\right]^{-1} \cdot \nabla f(\mathbf{x}^0)$$

$$= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 3/4 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 3/4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$$\therefore \mathbf{x}^1 = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$$

Se calculează gradientul funcţiei în noua aproximaţie:

$$\nabla f(\mathbf{x}^1) = \begin{bmatrix} -2+1+1 & 1-2+1 & 1-2+1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$$

Deoarece valoarea gradientului este 0, metoda a fost convergentă.

Metode cvasiNewton

.

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,

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

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

Relaţia iterativă de mai sus poate fi considerată ca fiind relaţia fundamentală care stă la baza metodelor ce utilizează derivatele.
Cel mai performant algoritm din această categorie de metode este *algoritmul Davidson-Fletcher-Powell*.

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 $\widetilde{H}^{(k)}$ se alege ca fiind matricea unitate I_n, astfel încât prima iteraţie decurge după metoda gradientului. Ulterior se face o modificare treptată de la direcţia gradientului la cea a metodei Newton, reuşindu-se astfel combinarea avantajelor aferente celor două metode.

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

$$- \frac{[\widetilde{H}^{(k)}][\Delta g^{(k)}][\Delta g^{(k)}]^t [\widetilde{H}^{(k)}]^t}{(\Delta g^{(k)}, H^{(k)} \Delta g^{(k)})}$$

Algoritmul Davidson-Fletcher-Powell

unde:

$$\Delta X^{(k)} = X^{(k+1)} - X^{(k)} \qquad \Delta g^{(k)} = g^{(k+1)} - g^{(k)}$$

Rolul matricei $A^{(k)}$ este de a asigura convergenţa lui $\tilde{H}^{(k)}$ către $H^{(k)^{-1}}$, în timp ce matricea $B^{(k)}$ are rolul anulării treptate a efectului alegerii iniţiale, adică a matricei unitate.

Pentru cazul funcţiilor obiectiv pătratice direcţiile utilizate de algoritmul Davidon-Fletcher-Powell sunt direcţii conjugate.

Algoritmul Davidson-Fletcher-Powell

Algoritmul evoluează satisfăcător dacă se îndeplinesc următoarele condiţii:

- **eroarea în evaluarea componentelor gradientului $g^{(k)}$ nu este prea mare (în cazul derivării numerice);**
- **Matricea $\widetilde{H}^{(k)}$ nu este singulară. În situaţia în care aceasta este aproape de singularitate una din soluţiile propuse constă în reiniţializarea ei cu matricea unitate (există şi alte posibilitaţi).**

Algoritmul Davidson-Fletcher-Powell

Exemplu numeric:

Să se determine minimul funcţiei:

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

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

Se calculează gradientul iar hesianul funcţiei este matricea unitate :

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

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

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:

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

$$\lambda^{(0)} = 0{,}1264$$

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

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

Se calculează matricea $\widetilde{H}^{(1)}$:

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

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

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

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 $\lambda^{(0)}$.

Deaorece:

$$g^{(2)} = \begin{bmatrix} 0 & 0 \end{bmatrix}^t$$

calculele sunt terminate.

PROGRAMAREA NELINIARĂ CU RESTRICŢII

GENERALITĂŢI

Soluţia optimă trebuie să fie aleasă în mod obligatoriu în interiorul domeniului admisibil pentru vectorul variabilelor de optimizare $X = [x_1, x_2, ..., x_n]$.

Problema care trebuie rezolvată se referă la determinarea valorii variabilelor care asigură extremul unei funcţii obiectiv:

$$\min F(X) = \min F(x_1,\ x_2,\ ..., x_n)$$

în prezenţa restricţiilor:

$$g_i(x_1,\ x_2,\ \ ...,\ x_n) \le 0, \quad i = 1, \ldots, m$$

$$x_j \ge 0, \ \ j = 1, \ldots, n$$

Condiţiile Kuhn-Tucker (K-T)

Condiţiile care trebuie să fie satisfăcute pentru un optim cu restricţii, local sau global au fost determinate de K-T având ca punct de plecare modelele clasice ale multiplicatorilor lui Lagrange.

A. Condiţiile K-T pentru cazul $g_i(X)$

Funcţia Lagrange asociată problemei PN cu restricţii (prezentată anterior), în care $F(X)$ şi $g_i(X)$ au derivate parţiale de ordinul întâi, în raport cu x_j, $j = 1, \ldots, n$, respectiv $g_i(X)$ sunt doar restricţii de egalitate, are următoarea formă:

$$\Phi(x,\lambda) = F(x) + \sum_{i=1}^{m} \lambda_i g_i(X)$$

unde λ_i, $i = 1, \ldots, m$, sunt multiplicatorii Lagrange.

Condiţiile Kuhn-Tucker (K-T)

Un punct extrem al funcţiei $\Phi(X, \lambda)$, care poate fi punctul de minim, se poate obţine prin anularea derivatelor de ordinul 1:

$$\frac{\partial \Phi(X,\lambda)}{\partial x_j} = \frac{\partial F(X)}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i(X)}{\partial x_j} = 0, \quad j = 1,\ldots,n$$

$$\frac{\partial \Phi(X,\lambda)}{\partial \lambda_i} = g_i(X) = 0, \quad i = 1,\ldots,m$$

Relaţiile reprezintă condiţiile necesare dar nu şi suficiente pentru minim. Dacă funcţia obiectiv $F(X)$ este convexă, iar restricţiile sunt liniare atunci este asigurată şi suficienţa.

Concluzie:

Minimul funcţiei $F(X)$, în prezenţa restricţiilor de egalitate $g_i(X) = 0$, este obţinut pentru acele valori ale lui X care minimizează fără restricţii funcţia auxiliară $\Phi(X, \lambda)$.

Condiţiile Kuhn-Tucker (K-T)

ALGORITM:

1. **Se aleg iniţial valori pentru λ_i, $i = 1, ..., m$;**
2. **Se rezolvă sistemul**

$$\frac{\partial \Phi(X,\lambda)}{\partial x_j} = \frac{\partial F(X)}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i(X)}{\partial x_j} = 0, \qquad j = 1, \dots, n$$

de *n* ecuaţii cu *n* necunoscute.

3. **Soluţiile determinate care depind de valorile lui λ_i sunt introduse în cele *m* restricţii.**
4. **Dacă restricţiile nu sunt satisfăcute, se aleg alte valori pentru λ_i, $i = 1, ..., m$, până când cele (n+m) ecuaţii vor fi satisfăcute de valorile alese.**

Condiţiile Kuhn-Tucker (K-T)

B. Condiţiile K-T pentru cazul $g_i(X) \leq 0$

Utilizând variabile auxiliare, w_i, $i = 1, \ldots, m$, inegalităţile se pot transforma în egalităţi:

$$g_i(X) + w_i^2 = 0, \quad i = 1, \ldots, m$$

Funcţia Lagrange asociată problemei PN cu restricţii are forma:

$$\Phi(X, w, \lambda) = F(X) + \sum_{i=1}^{m} \lambda_i \left(w_i^2 + g_i(X) \right)$$

care conduce la un punct extrem dacă sunt îndeplinite condiţiile:

Condiţiile Kuhn-Tucker (K-T)

$$\frac{\partial \Phi(X,w,\lambda)}{\partial x_j} = \frac{\partial F(X)}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i(X)}{\partial x_j} = 0, \quad j = 1,\ldots,n$$

$$\frac{\partial \Phi(X,w,\lambda)}{\partial \lambda_i} = g_i(X) + w_i^2 = 0, \quad i = 1,\ldots,m$$

$$\frac{\partial \Phi(X,w,\lambda)}{\partial w_i} = 2\lambda_i w_i = 0, \quad i = 1,\ldots,m$$

Deoarece mărimile w_i sunt variabile auxiliare, acestea pot fi eliminate.

$g_i(X) = 0$, independent de i, numai dacă $w_i = 0$.

A treia relaţie poate fi rescrisă, ţinînd cont de relaţia a doua:

$$\lambda_i g_i(X) = 0, \qquad i = 1,\ldots,m$$

Condiţiile Kuhn-Tucker (K-T)

$$\frac{\partial \Phi(X, w, \lambda)}{\partial x_j} = \frac{\partial F(X)}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i(X)}{\partial x_j} = 0, \quad j = 1, \ldots, n$$

$$g_i(X) \leq 0, \qquad i = 1, \ldots, m$$

$$\lambda_i g_i(X) = 0, \qquad i = 1, \ldots, m$$

$$\lambda_i \geq 0, \quad i = 1, \ldots, m$$

condiţiile Kuhn-Tucker

Condiţiile necesare pentru un punct de minim local şi global.

În situaţia în care *F*(*X*) şi $g_i(X)$ sunt convexe, relaţiile ne dau minimul global.

Condiţiile Kuhn-Tucker (K-T)

Observaţii

Semnul multiplicatorilor λ_i, $i = 1, \ldots, m$, se schimbă în funcţie de natura optimului căutat şi de semnul restricţiilor.

Natura restricţiilor	Natura optimului	
	max $F(X)$	min $F(X)$
$g_i(X) \le 0,\ \ i = 1, \ldots, m$	$\lambda_i \le 0$	$\lambda_i \ge 0$
$g_i(X) \ge 0,\ \ i = 1, \ldots, m$	$\lambda_i \ge 0$	$\lambda_i \le 0$

Condiţiile Kuhn-Tucker ne indică faptul că pentru o restricţie care nu este activă, de tipul:

$$\sum_{j=1}^{n} a_{i\,j} \cdot x_j = b_i$$

multiplicatorul $\lambda_i = 0$.

Metode ale programării neliniare cu restricţii

Metodele programării neliniare cu restricţii se împart în două mari categorii:

- metode ale direcţiilor admisibile;
- metode ale funcţiilor penalizare.

A.Metode ale direcţiilor admisibile

Sunt concepute astfel încât să folosească tehnici de minimizare fără restricţii, dar în acelaşi timp să considere şi restricţiile de inegalitate.

În acest sens se alege un punct iniţial $X^{(0)}$ şi o direcţie astfel încât o deplasare mică de-a lungul ei să nu modifice restricţiile, iar funcţia obiectiv $F(X)$ să se îmbunătăţească.

Metode ale programării neliniare cu restricţii

Direcţia de deplasare care satisface prima cerinţă se numeşte *direcţia admisibilă*.

Astfel, plecând din punctul iniţial $X^{(0)}$, pe direcţia $d^{(0)} = -\nabla F(X^{(0)})$, se obţine $X^{(1)}$.

Procedeul se repetă până în momentul în care este imposibil să se mai găsească o direcţie admisibilă.

Metode ale programării neliniare cu restricţii

B. Metode ale direcţiilor admisibile

Problema minimizări funcţiei $F(X)$ în prezenţa restricţiilor $g_i(X) \geq 0$, $i = 1, 2, \ldots, m$, poate fi abordată prin definirea unei funcţii de penalizare de tipul:

$$\varphi(y) = \begin{cases} 0, & y \geq 0 \\ \infty, & y < 0 \end{cases}$$

şi considerarea minimizării funcţiei fără restricţii:

$$\psi(X) = F(X) + \sum_{i=1}^{m} \varphi(g_i(X))$$

Metode ale programării neliniare cu restricţii

Observaţii

Dacă toate restricţiile $g_i(X) \geq 0$ sunt satisfăcute, al doilea termen nu contribuie cu nimic la valoarea lui $\Psi(X)$.

Dacă avem o resticție de tipul $g_i(X) < 0$, atunci $\varphi(g_i) = \infty$ şi vom fi departe de minim.

Al doilea termen din suma respectivă penalizează orice încălcare a restricţiilor.

Metode ale programării neliniare cu restricţii

Dificultăţi în abordare

Există situaţii pentru care trebuie aleasă o funcţie de penalizare mai puţin rigidă, care să prezinte proprietăţile de regularitate necesară. De exemplu, atunci când funcţia obiectiv prezintă discontinuităţi.
O alegere a unei funcţii de tipul:

$$\varphi(y)=\begin{cases}0, & y\geq 0\\ y^2 & y<0\end{cases}$$

este mai potrivită. Ea este continuă şi are derivată continuă pentru toate valorile lui *y*. Această metodă nu este foarte eficientă în toate cazurile.

Metode ale programării neliniare cu restricţii

Metoda Fiacco-McCormick

Pentru a evita dificultăţile întâlnite anterior, metoda determină punctul de minim optimul dinspre interiorul mulţimii restricţiilor. Se defineşte funcţia,

$$Z(X,r) = F(X) + r \cdot \sum_{i=1}^{m} \frac{1}{g_i(X)}, \qquad r > 0$$

Minimul acestei funcţii trebuie să se realizeze în interiorul mulţimii de restricţii deoarece pe această frontieră este îndeplinită condiția:

$$Z(X, r_0) \to +\infty \quad (g_i(X) = 0)$$

Punctul de minim depinde de r_0 şi va fi notat $X(r_0)$. Metoda prezintă interes mai ales atunci când avem de considerat restricţii puternic neliniare, eliminând complet mişcarea de-a lungul frontierelor mulţimii restricţiilor.

Metode ale programării neliniare cu restricţii

ALGORITM

1.Se alege $r_0 > 0$ şi $X^{(0)}$ în interiorul mulţimii de restricţii.
Se consideră problema minimizării fără restricţii a funcţiei Z(X, r).

2.Se repetă acest proces de minimizare pentru un şir de valori $\{r\}=\{r_0 > r_1 > r_2 > ... > r_k > 0\}$.
Prin micşorarea lui r, influenţa termenului al doilea, care penalizează apropierea de frontiera mulţimii restricţiilor, se reduce şi efortul în minimizarea a funcției $Z(X, r)$ se concentrează asupra lui $F(X)$.

Şirul $\{X(r_0), X(r_1), ...\}$ se poate apropia din ce în ce mai mult de frontiera mulţimii restricţiilor, dacă aceasta este avantajos pentru micşorare funcției $F(X)$.