Metoda gradientului proiectat (metoda Rosen)

În cazul problemelor de optimizare convexe ale căror restricții sunt liniare se poate folosi metoda gradientului proiectat. În principiu, această metodă poate fi folosită şi pentru cazul restricțiilor neliniare dar eficacitatea sa este redusă conducând la paşi de deplasare foarte mici pe frontiera domeniului.

Particularitățile acestei metode constau în modul în care se stabileşte direcția de deplasare din punctul curent:

a. în interiorul domeniului generat de restricții direcția de deplasare va fi antigradientul funcției (pentru problemele de minim);
b. atunci când aproximarea curentă a soluției optimale se găseşte pe frontiera domeniului se utilizează ca direcție de deplasare pentru îmbunătățirea soluției proiecția antigradientului pe frontieră.

În plus, această metodă are avantajul de folosi condițiile Karush-Kuhn-Tucher atât pentru a genera noi direcții de deplasare cât şi în calitate de criteriu de oprire a iterațiilor în punctul de optim.

Considerăm următoarea problemă de optimizare:

$$\begin{aligned} &\min F(x) \\ &g_i(x) = \sum_{j=1}^{n} a_{ij} x_j - b_i \le 0 \quad i = \overline{1,m} \end{aligned} \tag{9.1}$$

Sub formă matriceală sistemul de restricții poate fi scris ca:

$$A \cdot x \le b \tag{9.2}$$

unde:

$$\begin{aligned} &A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_m \end{bmatrix}^T \\ &b = \begin{bmatrix} b_1 & \cdots & b_m \end{bmatrix}^T \end{aligned} \tag{9.3}$$

a_i – vectorul linie al coeficienților restricției i.

Condiția ca domeniul soluțiilor admisibile să fie limitat în $\mathbb{R}^n$ este ca $n < m - 1$.

Într-un punct oarecare din $\mathbb{R}^n$ restricțiile pot fi active sau inactive.

Se numesc restricții active într-un punct acele restricții p care, în acest punct, sunt verificate ca egalitate:

$$g_p(x) = \sum_{j=1}^{n} a_{pj} - b_p = 0 \tag{9.4}$$

Se numesc restricții inactive într-un punct acele restricții r care, în acest punct, sunt verificate ca inegalitate:

$$g_r(x) = \sum_{j=1}^{n} a_{rj} - b_r < 0 \tag{9.5}$$

Referitor la liniile matricei A merită subliniată semnificația din punct de vedere matematic al acestora. Astfel, considerând o restricție activă în punctul x, deoarece aceasta este liniară, gradientul său este constant şi egal cu:

$$\begin{aligned}\nabla g_i(x) &= \begin{bmatrix} \dfrac{\partial g_i(x)}{\partial x_1} & \dfrac{\partial g_i(x)}{\partial x_2} & \cdots & \dfrac{\partial g_i(x)}{\partial x_n} \end{bmatrix}^T = \\ &= \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix}^T = a_i^T\end{aligned} \tag{9.6}$$

deci linia i a matricei A (sau a_i) reprezintă tocmai gradientul transpus al restricției i (sau normala la hiperplanul reprezentat de ecuația $g_i(x) = 0$).

În $\mathbb{R}^n$ pot exista mai multe poziții posibile ale punctului curent şi, în funcție de aceasta, mai multe moduri de determinare a direcției de deplasare spre punctul următor:

a. Punctul curent se găseşte într-un vârf sau pe o muchie a poliedrului soluțiilor admisibile.

b. Punctul curent se găseşte pe una dintre fețele acestui poliedru.

c. Punctul curent se găseşte în interiorul domeniului soluțiilor admisibile. În acest caz, de exemplu punctul Y^1 din figura 9.1, se foloseşte ca direcție de deplasare antigradientul $-\nabla F(x)$ (eventual normalizat), până se atinge punctul Y'^1, caz în care se regăseşte una dintre cele două situații anterioare.

Se consideră că la pasul k aproximația curentă a soluției optimale este x_k şi că acest punct se găseşte la intersecția a q hiperplane de frontieră (ecuațiile corespunzătoare acestor hiperplane reprezentând restricțiile active). Se notează cu I^k mulțimea indicilor restricțiilor active în acest punct:

$$I^k = \{j | g_j = 0\} \qquad \text{card}(I^k) = q \tag{9.7}$$

şi cu A_q matricea obținută din A prin păstrarea doar a liniilor ai căror indici apar în I^k.

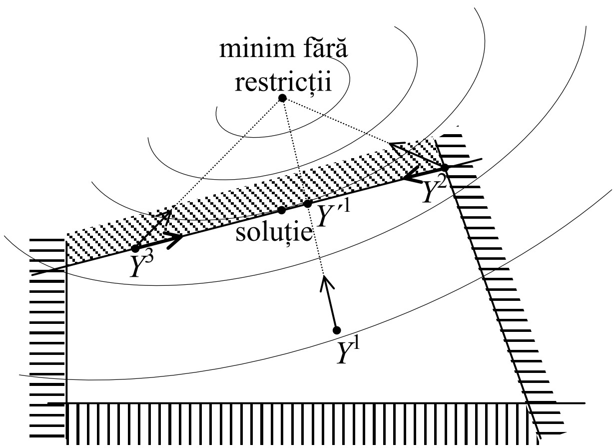

Fig.9.1 *Direcția de deplasare în funcție de poziția punctului curent*

Se pune problema de a se găsi acea direcție de deplasare din punctul $x^{(k)}$ de-a lungul căreia funcția obiectiv scade cel mai repede şi care păstrează punctul următor în interiorul sau pe frontiera domeniului soluțiilor admisibile.

În acest scop se împarte spațiul $\mathbb{R}^n$ în două subspații ortogonale.

Pentru aceasta se folosesc definițiile mulțimilor afine şi subspațiilor paralele cu mulțimi afine.

O mulțime $M \subset R^n$ se numeşte *mulțime afină* dacă:

$$\forall x', x'' \in M \Rightarrow tx' + (1-t)x'' \in M \quad \forall t \in R \tag{9.8}$$

Def. Fie $L, R \subset R^n$ două mulțimi afine. L este *paralelă* cu M dacă există $a \in R^n$ astfel încât

$$M = L + a \tag{9.9}$$

Tr. Orice mulțime afină nenulă $M \subset R^n$ este paralelă cu un singur subspațiu $L \subset R^n$ care este:

$$L = M - M = \{z \in R^n \,|\, \exists x, y \in M \text{ a.î. } z = x - y\} \tag{9.10}$$

Se va presupune în continuare că matricea A_q a coeficienților restricțiilor active este formată din vectori a_i liniar independenți.

Ținând cont de aceste definiții şi ipoteze se pot defini următoarele subspații:

a) un subspațiu D definit ca subspațiul paralel cu mulțimea afină D' a punctelor de intersecție a hiperplanelor care reprezintă condițiile active în punctul curent:

$$D' = \{x \in R^n \,|\, a_i x - b_i = 0, i \in I^k\} \tag{9.11}$$

Acest subspațiu are dimensiune n-r. Dacă n=r subspațiul D reprezintă subspațiul format doar din elementul nul al lui R^n.

b) subspațiul $D^{\perp}$ care reprezintă complementul ortogonal al lui D şi este tocmai subspațiul generat de vectorii linie din A_q.

Prin definirea acestor subspații, orice vector din R^n poate fi reprezentat în mod unic sub forma:

$$x = x_D + x_{D^{\perp}} \quad \text{cu } x_D \in D, x_{D^{\perp}} \in D^{\perp} \tag{9.12}$$

Vectorii x_D şi $x_{D^{\perp}}$ se numesc proiecțiile vectorului x pe cele două subspații.

Direcția de deplasare care permite simultan apropierea de optim şi menținerea în domeniul de admisibilitate va corespunde proiecției antigradientului pe subspațiul D şi poate fi determinată ca:

$$d^{(k)} = -P_q \nabla F\left(x^{(k)}\right) \tag{9.13}$$

în care P_q se numeşte matrice de proiecție şi este egală cu:

$$P_q = \left\lfloor I - A_q^T \left(A_q A_q^T\right)^{-1} A_q \right\rfloor \tag{9.14}$$

În condițiile în care restricțiilor active le corespund vectori a_i liniar independenți se poate arăta că expresia matricei de proiecție este corectă (matricea $A_q A_q^T$ este nesingulară).

După cum se poate observa, în cazul punctelor din interiorul domeniului de admisibilitate, deoarece nici o restricție nu este activă (A_q=0), va rezulta $P_q=I$ şi se obține ca direcție de deplasare antigradientul. Din această cauză, se poate considera că relațiile anterioare reprezintă metoda generală de determinare a direcției de deplasare, indiferent de poziția punctului curent.

Condiții de optim

Ca pentru toate metodele iterative, odată atins un punct trebuie verificat dacă acesta este un punct de optim (caz în care algoritmul se opreşte) sau nu (caz în care se determină o nouă direcție de deplasare).

Aceste condiții sunt diferite după cum punctul curent se găsește în interiorul sau pe frontiera domeniului de admisibilitate.

a. Condiții de optim în interiorul domeniului de admisibilitate.

În acest caz restricțiile nu influențează condițiile de optim acestea fiind aceleaşi ca în cazul programării fără restricții:

$$\begin{aligned} &\nabla F\left(x^{(k)}\right)=0 \\ &\nabla^2 F\left(x^{(k)}\right)>0 \end{aligned} \tag{9.15}$$

b. Condiții de optim pe frontiera domeniului de admisibilitate

Atunci când în punctul curent există restricții active condițiile necesare şi suficiente ca punctul curent să fie punct de optim rezultă prin particularizarea condițiilor Karush-Kuhn-Tucker. Pornind de la:

$$-\nabla F\left(x^{(k)}\right)=A^T \cdot \lambda \tag{9.16}$$

rezultă, sub formă vectorială:

$$\nabla F\left(x^{(k)}\right)=-\sum_{j\in I_k}\lambda_j a_j \tag{9.17}$$

Deci, gradientul funcției obiectiv poate fi scrisă ca o combinație liniară ai gradienților funcțiilor active, ceea ce este echivalent cu a spune că gradientul funcției obiectiv aparține subspațiului generat de cei q vectori linie liniar independenți din A ai căror indici apar în I_k. În aceste condiții proiecția vectorului gradient pe subspațiul ortogonal pe acesta trebuie să fie nulă:

$$P_q\nabla F\left(x^{(k)}\right)=0 \tag{9.18}$$

Condiția suficientă se obține pornind tot de la (9.16) ținând cont că acele componentele lui λ corespunzătoare restricțiilor active trebuie să fie nenegative (pentru forma problemei de optimizare considerată) iar componentele corespunzătoare restricțiilor inactive trebuie să fie nule. În aceste condiții (9.16) devine:

$$-\nabla F\left(x^{(k)}\right)=A_q^T \cdot \lambda_q \tag{9.19}$$

în care λ_q reprezintă vectorul obținut din λ păstrând doar componentele corespunzătoare restricțiilor active.

Prin înmulțire la stânga cu A_q rezultă:

$$A_q\left(-\nabla F\left(x^{(k)}\right)\right)=A_q A_q^T \cdot \lambda_q \tag{9.20}$$

sau, ținând cont că $A_q A_q^T$ este o matrice nesingulară:

$$\left(A_q A_q^T\right)^{-1} A_q\left(-\nabla F\left(x^{(k)}\right)\right)=\lambda_q \tag{9.21}$$

De aici rezultă forma condiției suficiente:

$$\left(A_q A_q^T\right)^{-1} A_q\left(-\nabla F\left(x^{(k)}\right)\right)\geq 0 \tag{9.22}$$

deoarece componentele lui λ_q nu pot fi negative.

În consecința, poate fi enunțată următoarea teoremă:

Tr. O soluție admisibilă situată la intersecția a q hiperplane frontieră liniar independente este soluție a problemei de programare (9.1) dacă şi numai dacă sunt îndeplinite condițiile:

$$\begin{aligned} &P_q \nabla F\left(x^{(k)}\right)=0 \\ &\left(A_q A_q^T\right)^{-1} A_q \nabla F\left(x^{(k)}\right)\leq 0 \end{aligned} \tag{9.23}$$

Utilizarea condițiilor Karush-Kuhn-Tucker pentru determinarea unei noi soluții admisibile

În anumite cazuri se poate ajunge în situația în care condiția necesară este îndeplinită (deci noua direcție de deplasare din punctul curent este nulă) dar condiția suficientă nu deoarece $\left(A_q A_q^T\right)^{-1} A_q\left(-\nabla F\left(x^k\right)\right)$ mai are componente negative.

Deci, pentru continuarea procesului de optimizare se impune determinarea unei noi direcții de deplasare.

Rosen şi Jacoby au propus eliminarea din matricea A_q a liniilor corespunzătoare condițiilor active cărora le corespund valori negative ale multiplicatorilor lui Lagrange, deci componente negative în $\left(A_q A_q^T\right)^{-1} A_q\left(-\nabla F\left(x^k\right)\right)$.

Folosind această matrice modificată se poate determina o nouă matrice proiecție şi, în continuare, o nouă direcție de deplasare.

Utilizarea practică a arătat că această metodă poate conduce la mărirea numărului de iterații, rezultate mai bune obținându-se prin eliminarea din matricea A_q a unei singure linii, şi anume cea căreia îi corespunde cea mai mică componentă (negativă) a lui $\left(A_q A_q^T\right)^{-1} A_q\left(-\nabla F\left(x^k\right)\right)$. În acest fel se obține matricea A_{q-1}.

Algoritmul metodei gradientului proiectat

Există două metode de a demara acest algoritm:

-pornirea dintr-un punct interior domeniului soluțiilor admisibile şi folosirea antigradientului ca direcție de deplasare până la atingerea unui punct de pe frontieră, după care se trece la folosirea metodei gradientului proiectat;

-inițializarea algoritmului cu un punct situat pe frontieră şi anume la intersecția a unui număr de hiperplane din sistemul de restricții.

În continuare se va considera că se cunoaşte deja un punct de pe frontiera domeniului de admisibilitate.

Trebuie subliniat faptul că, odată determinată o direcție de deplasare, se va verifica dacă pasul de deplasare care va fi folosit nu conduce la încălcarea unei alte restricții.

Pentru aceasta se observă că încălcarea unei restricții impune realizarea intersecției cu unul dintre hiperplanurile ale căror indice nu apare în I^k:

$$
\begin{aligned}
&x = x^{(k)} + \mu_{\max} d^{(k)} \\
&0 = a_i x - b_i = a_i \left(x^{(k)} + \mu_{\max} d^{(k)}\right) - b_i = \\
&= a_i x^{(k)} - b_i + a_i \mu_{\max} d^{(k)}
\end{aligned}
\tag{9.24}
$$

cu soluția:

$$
\mu_{\max} = \min_{j \notin I_k} \left\{ \mu_j \,\middle|\, \mu_j = \frac{g_j\left(x^{(k)}\right)}{\left\langle a_j, d^{(k)} \right\rangle}, \quad \mu_j > 0 \right\}
\tag{9.25}
$$

care reprezintă cea mai mare valoare a lui μ care nu conduce la încălcarea unei alte restricții.

Pasul de deplasare efectiv folosit trebuie să îndeplinească condiția:

$$
\mu \leq \mu_{\max}
\tag{9.26}
$$

Valoarea optimă a pasului de deplasare poate fi obținută din condiția:

$$
\frac{dF\left(x^{(k)} + \mu d^{(k)}\right)}{d\mu} = 0
\tag{9.27}
$$

Deoarece găsirea acestui minim este dificil de obținut, Rosen propune înlocuirea soluției exacte prin rădăcina unei ecuații de gradul întâi care se obține prin interpolarea liniară a derivatei din ecuația anterioară.

Astfel, se determină $\nabla F\left(x^{(k)} + \mu_{\max} d^{(k)}\right)$ și se verifică semnul acestuia:

a. dacă $\nabla F\left(x^{(k)} + \mu_{\max} d^{(k)}\right) \leq 0$ atunci funcția F este descrescătoare în raport cu μ pe intervalul $\left[0, \mu_{max}\right]$ și atunci:

$$
\mu = \mu_{\max}
\tag{9.28}
$$

b. dacă $\nabla F\left(x^{(k)} + \mu_{\max} d^{(k)}\right) > 0$ atunci minimul funcției este atinsă undeva în intervalul $\left[0, \mu_{\max}\right]$, valoarea aproximativă fiind obținută prin metoda coardei:

$$\mu \approx \frac{\left\langle d, \nabla F\left(x^{(k)}\right)\right\rangle}{\left\langle d, \nabla F\left(x^{(k)}\right) - \nabla F\left(\tilde{x}^{(k)}\right)\right\rangle} \mu_{max} \tag{9.29}$$

Probleme

1. Demonstrație matrice de proiecție

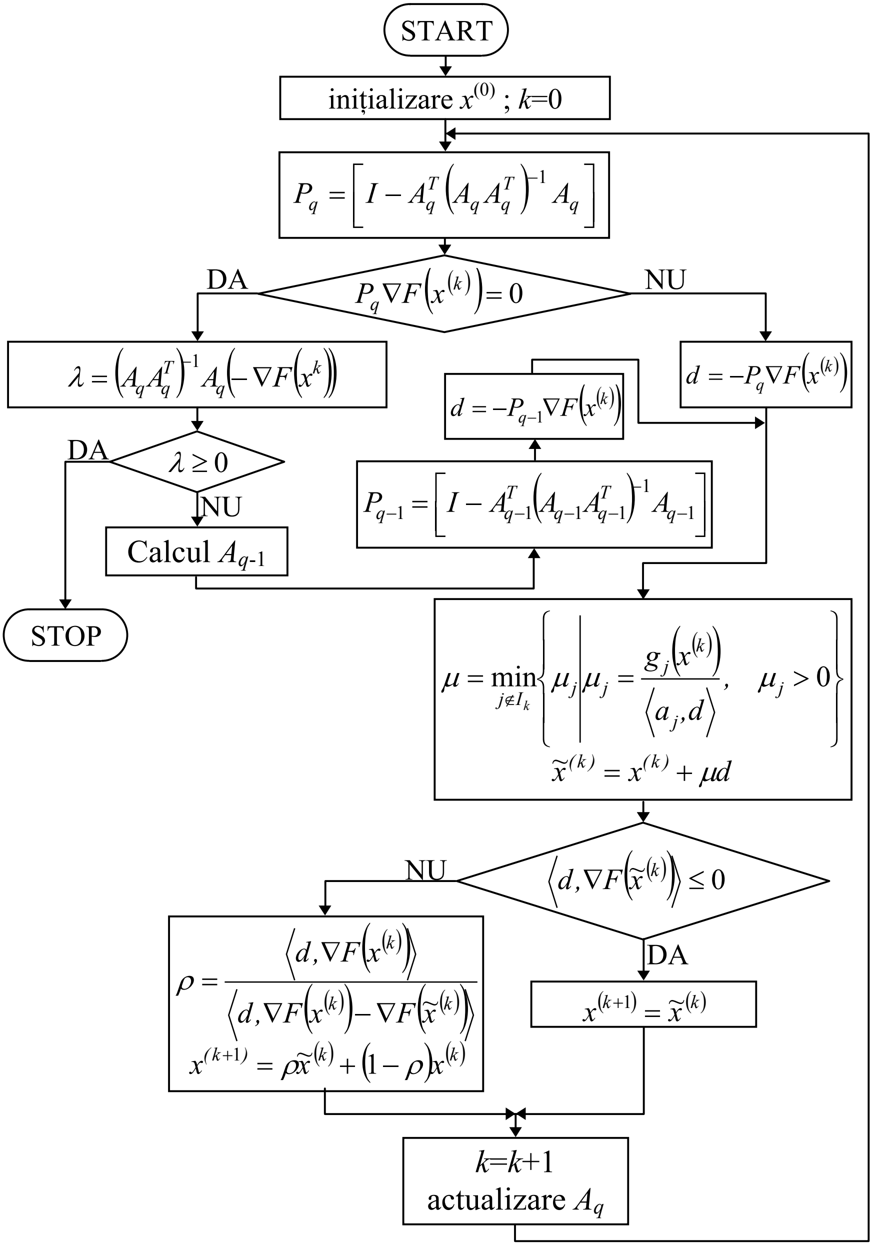

Fig.? *Algoritmul Rosen*

SPAȚII AFINE

1. Spațiu afin asociat unui spațiu vectorial real

Fie V un spațiu vectorial de dimensiune finită peste corpul R şi o mulțime nevidă de puncte $\mathrm{M} = \{A, B, C, \ldots\}$

Def: Structură afină

Se numeşte structură afină pe mulțimea M , asociată spațiului vectorial *V,* aplicația $\varphi : M \times M \to V$ care verifică următoarele condiții:

i) $\forall \quad A, B, C \in M \quad \varphi(A,B) + \varphi(B,C) = \varphi(A,C)$

ii) $\forall \quad O \in M$ fixat, aplicația $\varphi_0 : M \to V$ prin $\varphi_0(A) = \varphi(O, A)$ este bijectivă.

Mulțimea *M* înzestrată cu o structură afină se numeşte spațiu afin.

Obs:

1. Elementele lui *M* se numesc puncte iar mulțimea acestor puncte se numeşte spațiu bază a spațiului afin.
2. Dacă spațiul *V* are dimensiunea *n* atunci *M* este spațiu afin *n*-dimensional notat M^n
3. Elementele lui *V* se numesc vectori directori (cu excepția vectorului nul) iar *V* se numeşte spațiu director al spațiului afin.

Elementele fundamentale ale unui spațiu afin sunt punctele şi vectorii.

Imaginile perechilor de puncte, distincte sau nu, prin aplicația structură afină φ sunt vectori din *V*:

i) $\varphi(A,B) = \overline{v}$ unde *A* este originea iar *B* este extremitatea lui $\overline{v}$

ii) $\varphi(B, A) = -\overline{v}$

iii) i) $\varphi(A, A) = \overline{0}$