1.4. Problema UNLP (optimizare neliniara fara constrangeri). Metoda gradientului.

O problema de optimizare neliniara unidimensionala fara constrangeri ((UNLP) rezida in determinarea vectorului coloana $(x^*)^T = (x_1^*, ..., x_n^*)^T$ pentru functia obiectiv $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ cu proprietatea ca

$$\min_{x \in \mathbb{R}^{nn}} f(x) = f(x^*) = f(x_1^*, ..., x_n^*)$$

Metodele numerice de cautare a minimumului reprezinta o procedura iterativa, astfel incat, pornind de la un vector initial $x^{(0)}$, se genereaza un sir de vectori $(x^{(k)})_{k \geq 1}$, pentru care

$$f(x^{(0)}) > f(x^{(1)}) > ... > f(x^{(k)}) > ...$$

Din punct de vedere formal deplasarea intre punctele $x^{(k)}$ si $x^{(k+1)}$ este data de relatia

$$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)},$$

unde vectorul $d^{(k)}$ reprezinta directia de deplasare iar α_k determina marimea acestei deplasari.

Conform sensului geometric al gradientului $\nabla f(x)$ pentru functia $f(x)$, vectorul $\nabla f(x)$ indica **directia cresterii celei mai rapide** . Prin urmare este firesc sa alegem in calitate de directie de deplasare directia **descresterii celei mai rapide (abrupte)**, adica directia opusa gradientului, **directie determinata de vectorul** $-\nabla f(x)$ numita si **antigradient.**

Or, alegand antigradientul ca directie de deplasare, obtinem procedura iterativa in forma

$$x^{(k+1)} = x^{(k)} - \alpha_k \nabla f(x^{(k)}).$$

Valoarea parametrului α_k este determinata din conditia minimizarii functiei $f(x^{(k)} - \alpha \nabla f(x^{(k)})$ in dependenta de α in sensul ca

$$\alpha_k : f(x^{(k)} - \alpha_k \nabla f(x^{(k)})) = \min_{\alpha \in \mathbb{R}} f(x^{(k)} - \alpha \nabla f(x^{(k)})).(\mathbf{1})$$

Aceasta metoda a gradientului in care lungimea pasului se alege din conditia ($\mathbf{1}$) se numeste metoda celei mai rapide descresteri. Putem spune ca aceasta este **procedura ideala** pentru alegerea lungimii pasului, Insa din punct de vedere al aplicarii acestei metode pe calculator verificarea conditiei ($\mathbf{1}$) devine dificila. In acest scop aceata metoda esste modificata in felul urmator.

Pasul 1. Se alege o valoare arbitrara α (una si aceeasi pentru fiecare iteratie), se determina punctul $z = x^{(k)} - \alpha \nabla f(x^{(k)}$ si se calculeaza $f(z)$;

Pasul 2. Se verifica conditia

$$f(z) - f(x^{(k)}) \leq \delta \alpha \left\| \nabla f(x^{(k)}) \right\|^2, (\mathbf{2})$$

unde δ este o constanta arbitrara (de ex. $\delta = 1/2$) din intervalul $(0, 1)$;

Pasul 3. Daca este indeplinita inegalitatea (**2**), atunci valoarea α este acceptata, luand $\alpha_k = \alpha$.In caz contrar se fractioneaza α prin inmultirea lui cu un numar arbitrar γ (de ex. $\gamma = 1/2$) din intervalul $(0, 1)$,procesul acesta continuand pana cand este satisfacuta conditia (**2**).

Se recomanda ca procedura de calcul sa continue pana cand

$$\left\| \nabla f(x^{(k)}) \right\| < \varepsilon,$$

unde $\varepsilon > 0$ este considerata precizia atingerii minimumului.

Metoda gradientuluiin in care α_k este determinat conform algoritmului descris mai sus se numeste **metoda gradientului cu fractionarea pasului.**

Un exemplu de aplicare a acestui Algoritm poate fi gasit in lucrarea lui *V. Moraru si E. Tutunaru, Programare Matematica,UTM, Chisinau, 1999.*

1.5. Metoda directiilor conjugate.

Metoda directiilor conjugate poate fi privita ca pe o metoda intermediara intre metoda gradient, ce foloseste informatii de ordinul I si metodele ce folosesc informatii de ordinul II (metode de tip Newton). Aceasta metoda este motivata de dorinta de a accelera viteza de convergenta in metoda gradientului si, totodata, de a evita necesitatea de a folosi Hessiana. un caz particular al metodei directiilor conjugate este metoda gradientilor conjugati, care a fost initial dezvoltata pentru probleme patratice. Aceasta tehnica a fost extinsa asupra problemelor de optimizare generale, prin aproximare, deoarece se poate arata ca in apropierea unui punct de minim local functia obiectiv este, aproximativ, patratica.

1.5.1. Metoda directiilor conjugate pentru QP.

Metoda directiilor conjugate este, deasemenea, o metoda de ordinul I, adica foloseste informatia extrasa din valoarea functieisia gradientului acesteia, insa prezinta o rata de convergenta mai buna decat a metodei gradient, cel putin, pentru cazul patratic. Sa luam urmatoarea problema patratica (QP) strict convexa:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}(Ax, x) + b^T x + c,$$

unde $A \succ 0$ (adica este o matrice pozitiv definita). Solutia optima a acestei probleme de optimizare este echivalenta cu rezolvarea unui sistem de ecuatii liniare de forma $Ax = -b$. Intrucat A este inversabila, solutia problemei de optimizare sau solutia sistemului liniar este $x^* = -A^{-1}b$. In cele mai multe cazuri, calculul inversei este foarte costisitor iar complexitatea aritmetica a unei metode de calcul numeric matriceal pentru gasirea solutiei este de ordinul $O(n^3)$.In cele ce urmeaza vom prezenta o metoda numerica de optimizare mai simpla si , de obicei, mai putin costisitoare numeric pentru aflarea solutiei optime x^*.

Este vorba de **algoritmul lui Hestenes-Stiefel pentru minimizarea unei functii patratice** (si rezolvarea unui sistem de ecuatii lineare cu matricea pozitiv definita):

Pasul. 1. Se alege arbitrar $x^{(0)} \in \mathbb{R}^n$ si se determina $\nabla f(x^{(0)})$.Daca $\nabla f(x^{(0)}) = 0$, atunci $x^{(0)}$ este o solutie optima. STOP. In caz contrar luam

$d^{(0)} = -\nabla f(x^{(0)})$, $k = 0$ si trecem la pasul urmator.

Pasul 2. Se determina lungimea pasului α_k de-a lungul directiei $d^{(k)}$, care pleaca din $x^{(k)}$,ceea ce revine la minimizarea in raport cu parametrul scalar α al functiei $\varphi(\alpha) = f(x^{(k)} + \alpha d^{(k)})$. Determinarea lui α_k poate fi realizata conform formulei

$$\alpha_k = -\frac{(\nabla f(x^{(k)}))^T d^{(k)}}{(Ad^{(k)})^T d^{(k)}}$$

sau printr-o metoda de aproximare a punctelor de extrem pentru functii de o singura variabila.

Pasul 3. Se construieste o aproximare noua:

$$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$$

si se calculeaza $f(x^{(k+1)})$ si $\nabla f(x^{(k+1)})$. Daca $\nabla f(x^{(k+1)}) = 0$, atunci $x^* = x^{(k+1)}$ este o solutie optima. STOP. In caz contrar trecem la pasul urmator.

Pasul 4. Se ia directia

$$d^{(k+1)} = -\nabla f(x^{(k+1)}) + \frac{\left\|\nabla f(x^{(k+1)})\right\|^2}{\left\|\nabla f(x^{(k)})\right\|^2} d^{(k)},$$

dupa care se trece la **Pasul 2,** luand $k+1$ inloc de k.

Plecand de la aproximatia initiala $x^{(0)}$, acest algoritm conduce la gasirea punctului de minim x^* al functiei patratice dupa calculul a cel mult n directii conjugate.

Exemplu. Fie functia $f(x_1, x_2) = x_1^2 + x_1x_2 + \frac{1}{2}x_2^2 - 3x_1 - 2x_2$.Observam ca aceasta poate fi scrisa in forma

$$f(x) = \frac{1}{2}(Ax, x) + b^T x,$$

unde

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, b = (-3, -2)^T.$$

Luand aproximatia initiala $x^{(0)} = (-3, -2)^T$, vom obtine in 2 iteratii:

$$d^{(0)} = (1, 0)^T, \alpha_0 = \frac{1}{2}, x^{(1)} = (\frac{1}{2}, 2)^T;$$

$$d^{(1)} = (\frac{1}{4}, -\frac{1}{2})^T, \alpha_0 = 2, x^{(2)} = (1, 1)^T = x^*.$$

1.5.2. Minimizarea unei functii strict convexe.

Asa cum am mentionat anterior, metoda lui Hestenes-Stiefel poate fi folosita si pentru minimizarea functiilor nepatratice. In caz ca functia nu mai este una patratica este evident ca nu mai garantam convergenta intr-un numar finit de pasi. Algoritmul descris in p.1.5.1. se va relua, in caz nepatratic, in mod ciclic dupa iteratia cu numarul n, inlocuind $x^{(0)}$ cu $x^{(n)}$,atata timp cat $\left\|d^{(k)}\right\| \geq \varepsilon$, unde ε este eroarea de calcul admisa la aflarea punctului de minim. In felul acesta putem descrie metoda numita **metoda gradientului conjugat sau metoda lui Fletcher-Reevs:**

$$\begin{aligned}
x^{(k+1)} &= x^{(k)} + \alpha_k d^{(k)}, \\
d^{(0)} &= -\nabla f(x^{(0)}), \\
d^{(k)} &= -\nabla f(x^{(k)}) + \beta_{k-1} d^{(k-1)}, k \geq 1, \\
\alpha_k &: \quad f(x^{(k)} + \alpha_k d^{(k)}) = \min_{\alpha \geq 0} f(x^{(k)} + \alpha d^{(k)}).
\end{aligned}$$

Metodele de directii conjugate se deosebesc prin modul in care se definesc coeficientii β_k, $k \geq 0$. Astfel, in varianta sa originala, asa cum a fost ea propusa de Fletcher si Reevs acesti coeficienti se determina prin formula:

$$\beta_{k-1} = \begin{cases} \frac{\left\|\nabla f(x^{(k)})\right\|^2}{\left\|\nabla f(x^{(k-1)})\right\|^2}, k \notin I, \\ 0, k \in I, \end{cases}$$

unde I este o multime de indici de forma $I = \{0, n, 2n, 3n, ...\}$, adica algoritmul se reinoieste peste fiecare n pasi.

O alta varianta a metodei gradientului conjugat este cea propusa de **Polak si Ribiere** (1969), varianta in care

$$\beta_{k-1} = \begin{cases} \frac{(\nabla f(x^{(k)}), \nabla f(x^{(k)}) - \nabla f(x^{(k-1)}))}{\left\|\nabla f(x^{(k-1)})\right\|^2}, k \notin I, \\ 0, k \in I, \end{cases}$$

unde $I = \{0, n, 2n, 3n, ...\}$.

Remarca. In cazul functiilor patratice aceste doua variante de calcul a coeficientilor $\beta_k$ccoincid. In cazul minimizarii functiilor convexe nepatratice, algoritmul propus de Fletcher si Reevs are proprietati de convergenta mai slabe, de aceea in acest caz este recomandat algoritmul lui Polak si Ribiere.