Algebră Liniară și Optimizare

Curs 8

Conf. dr. Anca Ignat

Conf. dr. Corina Forăscu

April 21, 2026

Optimizare numerică

Probleme de minimizare

$$\min \{ f(x) \, ; \; x \in D \} , \quad D \subseteq \mathbb{R}^n , \quad f : \mathbb{R}^n \longrightarrow \mathbb{R}$$

O problemă de maximizare se poate reduce la una de minimizare astfel:

$$\max \{f(x) \, ; \; x \in D \} = - \min \{-f(x) \, ; \; x \in D\}.$$

Tipuri de probleme de optimizare

- liniară sau neliniară - funcţia f este liniară sau neliniară
- pătratică - funcţia f este pătratică
- convexă - funcţia f este convexă
- fără restricţii - $D \subseteq \mathbb{R}^n$
- cu restricţii - $D = \{x \in \mathbb{R}^n \,;\, g(x) \leq 0\}$
- continuă - x ia valori reale, $D \subseteq \mathbb{R}^n$
- discretă - x ia valori discrete, D mulţime de valori discrete
- în numere întregi - $D \subseteq \mathbb{Z}^n$
- mixtă - unele componente ale vectorului x sunt discrete şi celelalte sunt continui
- ...

Optimizare - Definiţii

Considerăm problema de minimizare fără restricţii:

$$\min \{ f(x) \, ; \; x \in \mathbb{R}^n \} \quad f : \mathbb{R}^n \longrightarrow \mathbb{R}$$

Un punct $x^* \in \mathbb{R}^n$ se numeşte **punct de minim global** pentru funcţia f dacă $f(x^*) \leq f(x)$, $\forall x \in \mathbb{R}^n$.
Un punct $x^* \in \mathbb{R}^n$ se numeşte **punct de minim local** pentru funcţia f dacă există o vecinătate V a punctului x^* astefl încât $f(x^*) \leq f(x)$, $\forall x \in V$.
Un punct $x^* \in \mathbb{R}^n$ se numeşte **punct de minim strict local** pentru funcţia f dacă există o vecinătate V a punctului x^* astefl încât $f(x^*) < f(x)$, $\forall x \in V, x \neq x^*$.

$$V = S(x^*, r) = \{ x \in \mathbb{R}^n \, ; \; \|x - x^*\| \leq r \}$$

Optimizare - Definiţii

Dacă funcţia $f \in C^1(\mathbb{R}^n)$, se numeşte **gradient** al funcţiei f în punctul x, vectorul:

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x),\ \frac{\partial f}{\partial x_2}(x),\ \ldots\ ,\ \frac{\partial f}{\partial x_n}(x)\right)^T$$

Dacă funcţia $f \in C^2(\mathbb{R}^n)$, se numeşte **matrice Hessiană** a funcţiei f în punctul x, matricea:

$$H_f(x) = \nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right)_{i,j=1,\ldots,n} \in \mathbb{R}^{n\times n}$$

Matricea Hessiană

$$H_f(x) = \nabla^2 f(x) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2}(x) & \dfrac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \dfrac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ & & \vdots & \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1}(x) & \dfrac{\partial^2 f}{\partial x_n \partial x_2}(x) & \cdots & \dfrac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}$$

Dezvoltare în serie Taylor de ordinul 1

$$f \in C^2(\mathbb{R}^n) \quad , \quad a = (a_1, a_2, \ldots a_n)^T \in \mathbb{R}^n$$

$$\begin{aligned} f(x) &= f(a) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a)(x_i - a_i) + R(\|x - a\|^2) \\ &= f(a) + \nabla f(a)^T (x - a) + R(\|x - a\|^2) \end{aligned}$$

Dezvoltare în serie Taylor de ordinul al 2-lea

$$f \in C^2(\mathbb{R}^n) \quad , \quad a = (a_1, a_2, \ldots a_n)^T \in \mathbb{R}^n$$

$$\begin{aligned} f(x) &= f(a) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a)(x_i - a_i) + \\ &\quad + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(a)(x_i - a_i)(x_j - a_j) + \bar{R}(\|x - a\|^3) \\ &= f(a) + \nabla f(a)^T (x - a) + \frac{1}{2}(x - a)^T H_f(a)(x - a) + \bar{R}(\|x - a\|^3) \end{aligned}$$

Optimizare - condiţii necesare de optim de ordinul 1

Teorema 1

Fie $f \in C^1(\mathbb{R}^n)$ şi $x^* \in \mathbb{R}^n$ un punct de minim local pentru funcţia f. Atunci $\nabla f(x^*) = 0$.

Un punct $x^* \in \mathbb{R}^n$ pentru care $\nabla f(x^*) = 0$ se numeşte **punct staţionar/critic** pentru funcţia f.
Teorema de mai sus spune că orice punct de minim local pentru funcţia f trebuie să fie punct staţionar. Punctele de minim local pot fi căutate printre soluţiile sistemului neliniar de ecuaţii $\nabla f(x) = 0$.

Optimizare - condiţii necesare de optim de ordinul al 2-lea

Teorema 2

Fie $f \in C^2(\mathbb{R}^n)$ şi $x^* \in \mathbb{R}^n$ un punct de minim local pentru funcţia f. Atunci $\nabla f(x^*) = 0$ şi matricea Hessiană $H_f(x^*)$ este pozitiv semidefinită.

O matrice $A \in \mathbb{R}^{n \times n}$ se numeşte **pozitiv definită** dacă:

$$\left(Ax,\ x\right)_{\mathbb{R}^n} > 0 \quad , \quad \forall\, x \in \mathbb{R}^n\ ,\ x \neq 0$$

O matrice $A \in \mathbb{R}^{n \times n}$ se numeşte **pozitiv semidefinită** dacă:

$$\left(Ax,\ x\right)_{\mathbb{R}^n} \geq 0 \quad , \quad \forall\, x \in \mathbb{R}^n$$

Optimizare - condiţii suficiente de optim de ordinul al 2-lea

Teorema 3

Fie $f \in C^2(\mathbb{R}^n)$, $x^* \in \mathbb{R}^n$ un punct staţionar, $\nabla f(x^*) = 0$ pentru care matricea Hessiană $H_f(x^*)$ este pozitiv definită. Atunci x^* este punct de minim strict local pentru funcţia f.

Optimizare - funcţii convexe

O funcţie $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ se numeşte **funcţie convexă** dacă:

$$f\left(ax_1+(1-a)x_2\right) \leq af(x_1)+(1-a)f(x_2)\,,\ \forall\, x_1,\, x_2 \in \mathbb{R}^n\,,\ \forall\, a \in [0,1]$$

Teorema 4

Daca funcţia $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ este convexă atunci orice punct de minim local este punct de minim global. Dacă, în plus, $f \in C^1(\mathbb{R}^n)$ atunci orice punct staţionar ($\nabla f(x^*) = 0$) este punct de minim global.

Optimizare - Metode de descreştere

Se numeşte **direcţie de descreştere** a funcţiei f în punctul x, un vector $d \in \mathbb{R}^n$ pentru care:

$$f(x + \alpha\, d) \leq f(x) \quad , \quad \forall \alpha \in [0, \bar{\alpha}).$$

Teorema 5

Fie $f \in C^1(\mathbb{R}^n$. Un vector $d \in \mathbb{R}^n$ este direcţie de descreştere pentru funcţia f în punctul x dacă şi numai dacă:

$$\nabla f(x)^T d = \big(d, \nabla f(x)\big)_{\mathbb{R}^n} < 0.$$

Algoritm de optimizare

Metodă de descreştere

1. alege $x \in \mathbb{R}^n$
2. do
   - găseşte o direcţie de descreştere d a funcţiei f în x;
   - găseşte $\tilde{\alpha} > 0$ astfel ca:

     $$f(x + \tilde{\alpha}\, d) = \min \{ f(x + \alpha\, d)\ ;\ \alpha \in [0, \bar{\alpha}) \}$$

     (ajustarea pasului -line search)
   - $x = x + \tilde{\alpha}\, d$

   while (nu am găsit soluţia);

Ajustarea pasului (line search)

Pentru găsirea lui $\tilde{\alpha}$, avem de rezolvat o problemă de minimizare unidimenională:

$$\min \{ g(a) \, ; \, a \in [b, c] \}. \tag{1}$$

Pentru rezolvarea acestei probleme, de obicei, se folosesc algoritmi care aproximează solţia ecuaţiei neliniare $g'(a) = 0$. Se construiesc şiruri de vectori (a_k), care, în anumite cazuri, converg la soluţia problemei $g'(a) = 0$. Pentru ca soluţia găsită să fie şi soluţia problemei (1), este necesar ca $g''(a^*) \geq 0$.

Metoda Newton

Fie a_k aproximarea curentă soluţiei problemei de minimizare (1). Folosim dezvoltarea în serie Taylor a funcţiei g:

$$g(a) = g(a_k) + g'(a_k)(a - a_k) + \frac{1}{2!} g''(a_k)(a - a_k)^2 + \frac{1}{3!} g'''(b_k)(a - a_k)^3.$$

Fie funcţia q definită de:

$$q(a) = g(a_k) + g'(a_k)(a - a_k) + \frac{1}{2} g''(a_k)(a - a_k)^2. \qquad (2)$$

Metoda Newton

Pentru valori ale lui a în apropiere lui a_k putem presupune că funcţia q aproximează funcţia g, $g(a) \approx q(a)$. Pentru a afla următorul element din şir, a_{k+1}, vom minimiza funcţia q (în locul funcţiei g).

Calculăm următorul element din şir, a_{k+1}, ca soluţie a problemei de minimizare:

$$q(a_{k+1}) = \min \{ q(a) \, ; \, a \in [b, c] \}.$$

Funcţia q este un polinom de grad 2 în necunoscuta a, de unde vom obţine următoarea formulă pentru a_{k+1}:

$$a_{k+1} = a_k - \frac{g'(a_k)}{g''(a_k)} \quad , \quad k = 0, 1, 2, \ldots$$

Primul element al şirului, a_0 se alege la întâmplare.

Metoda secantei

Dacă în relaţia (2), se face următoarea aproximare:

$$g''(a_k) \approx \frac{g'(a_k) - g'(a_{k-1})}{a_k - a_{k-1}}$$

obţinem următoarea metodă de aproximare, numită **metoda secantei**:

$$a_{k+1} = a_k - g'(a_k)\frac{a_k - a_{k-1}}{g'(a_k) - g'(a_{k-1})} \quad , \quad k = 0, 1, \ldots$$

Pentru a construi şirul a_k, trebuie precizate primele două elemente din şir, a_0 şi a_1.

Aproximare spline cubică

Pentru a construi elementul a_{k+1} al şirului, vom folosi elementele a_{k-1}, a_k şi valorile $g(a_{k-1})$, $g(a_k)$, $g'(a_{k-1})$, $g'(a_k)$. Aproximăm funcţia g cu un polinom de gradul al 3-lea:

$$g(a) \approx q(a) = c_0\, a^3 + c_1\, a^2 + c_2\, a + c_3.$$

Funcţia q, respectiv constantele $c_i, i = 0, \ldots, 3$, se calculează astfel ca:

$$\begin{array}{ccc} q(a_{k-1}) = g(a_{k-1}) & , & q(a_k) = g(a_k) \\ q'(a_{k-1}) = g'(a_{k-1}) & , & q'(a_k) = g'(a_k) \end{array}$$

Aproximare spline cubică

Elementul a_{k+1} este punctul de minim al funcţiei q şi se calculează folosind următoarele formule:

$$
\begin{aligned}
a_{k+1} &= a_k - (a_k - a_{k-1})\frac{g'(a_k) + u_2 - u_1}{g'(a_k) - g'(a_{k-1}) + 2u_2} \\
u_1 &= g'(a_{k-1}) + g'(a_k) - 3\frac{g(a_k) - g(a_{k-1})}{a_k - a_{k-1}} \\
u_2 &= \sqrt{u_1^2 - g'(a_{k-1})g'(a_k)}
\end{aligned}
$$

Ajustarea inexactă a pasului

$$f(x + \tilde{\alpha}\, d) \approx \min \{ f(x + \alpha\, d)\ ;\ \alpha \in [0, \bar{\alpha}).$$

- nu se obţine $\tilde{\alpha}$ optimal;
- pentru reducerea timpului de lucru, procedeul de calcul a punctului de optim se opreşte înainte de a ajunge la soluţie, în funcţie de anumite criterii/teste de oprire.

Ajustarea inexactă a pasului

Regula lui Armijo

$$g(a) = f(x_k + a\, d_k) \quad , \quad \bar{g}(a) = g(0) + \epsilon g'(0) \;,\; \epsilon \in (0, 1).$$

O valoare $\bar{a}$ este acceptabilă după regula lui Armijo dacă:

1. $g(\bar{a}) \leq \bar{g}(\bar{a})$
2. $g(\sigma\bar{a}) \geq \bar{g}(\sigma\, \bar{a})$

$\sigma = 2$ sau 10, $\epsilon = 0.2$.

Ajustarea inexactă a pasului - regula Armijo

$k = 0;$

se alege a_0;

while ($g(a_k) > \bar{g}(a_k)$)

- $a_k = \dfrac{1}{\sigma} a_k$
- $k = k + 1;$

Ajustarea inexactă a pasului

Testul Goldstein

Dat $\epsilon \in (0, 2)$, $\bar{\alpha}$ este considerat acceptabil dacă:

$$g(\bar{\alpha}) \leq \quad \text{şi } g(\bar{\alpha}) > g(0) + (1-\epsilon)g'(0)$$

Alternativ, valoarea $\bar{\alpha}$ este acceptată dacă:

$$\epsilon \leq \frac{f(x_{k+1}) - f(x_k)}{\bar{\alpha}\nabla f(x_k)^T d_k} \leq 1 - \epsilon$$

Ajustarea inexactă a pasului

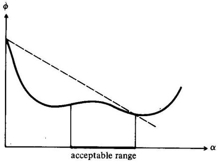
Testul Wolfe

Pentru $\epsilon \in (0, 2)$, o valoare $\bar{\alpha}$ îndeplineşte condiţia de acceptare dacă:

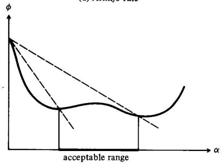
$$g(\bar{\alpha}) \leq \bar{g}(\bar{\alpha}) = g(0) + \epsilon\, g'(0)$$

$$g'(\bar{\alpha}) \geq (1 - \epsilon) g'(0)$$

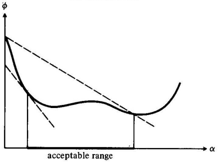
Ajustarea inexactă a pasului



(a) Armijo rule



(b) Goldstein test



(c) Wolfe test

Optimizare - forme pătratice

Fie $A \in \mathbb{R}^{n\times n}$ o matrice simetrică şi pozitiv (semi)definită, $b \in \mathbb{R}^n$. Se numeşte **formă pătratică** o funcţie $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ definită astfel:

$$
\begin{aligned}
f(x) = \quad & \frac{1}{2}\left(Ax, x\right)_{\mathbb{R}^n} - \left(b, x\right)_{\mathbb{R}^n} + c \\
= \quad & \frac{1}{2}x^T Ax - x^T b + c \quad , \quad c \in \mathbb{R}
\end{aligned}
$$

Optimizare - forme pătratice

Gradientul şi matricea Hessiană a unei forme pătratice sunt:

$$\nabla f(x) = Ax - b \quad , \quad H_f(x) = \nabla^2 f(x) = A.$$

Dacă matricea A este pozitiv definită, avem $\det A \neq 0$ şi funcţia f este strict convexă. Orice punct de minim local este punct de minim global.
Problema de minimizare:

$$f(x^*) = \min \{ f(x) \, ; \, x \in \mathbb{R}^n \}$$

are un punct de minim unic, $x^* \in \mathbb{R}^n$. Acest punct unic de minim este soluţia sistemului liniar:

$$Ax = b \quad , \quad x^* = A^{-1} b$$

Metoda pantei maxime (steepest descent)

Considerăm problema de minimizare pentru funcţia $f : \mathbb{R}^n \longrightarrow \mathbb{R}$:

$$\min\{f(x)\,;\,x \in \mathbb{R}^n\}.$$

Soluţia acestei probleme poate fi aproximată folosind un şir construit cu o metodă de descreştere:

$$x_0 \text{ dat} \quad , \quad x_{k+1} = x_k + \alpha_k d_k \quad , \; k = 0, 1, \ldots$$

unde d_k este o direcţie de descreştere a funcţiei f în x_k iar $\alpha_k > 0$ este soluţia problemei de minimizare:

$$f(x_k + \alpha_k\, d_k) = \min\{f(x_k + \alpha\, d_k)\,;\,\alpha \in [0, \bar{\alpha})\}.$$

Metoda pantei maxime consideră ca direcţia de descreştere:

$$d_k = -\nabla f(x_k).$$

Metoda pantei maxime - forme pătratice

$$g(x) = Ax - b$$

$$x_{k+1} = x_k - \alpha_k g_k \quad , \quad g_k = g(x_k) = Ax_k - b.$$

$$\alpha_k = \arg\min \{ f(x_k - \alpha g_k) \, ; \, \alpha \in [0, \bar{\alpha}) \}$$

$$\begin{aligned} f(x) &= \frac{1}{2}(x_k - \alpha g_k)^T A (x_k - \alpha g_k) - (x_k - \alpha g_k)^T b + c \\ &= \frac{1}{2}\left(g_k^T A g_k\right)\alpha^2 - \left(g_k^T A x_k - g_k^T b\right)\alpha + f(x_k) \\ &= \frac{1}{2}\left(g_k^T A g_k\right)\alpha^2 - \left(g_k^T g_k\right)\alpha + f(x_k) \end{aligned}$$

Metoda pantei maxime - forme pătratice

$f(x_k - \alpha\, g_k)$ - funcţie polinomială de gradul 2 în α, coeficientului lui α^2 este pozitiv, $g_k^T A g_k > 0$ (A este pozitiv definită). Punctul de minim al lui $f(x_k - \alpha\, g_k)$ este:

$$\alpha_{\min} = \alpha_k = \frac{g_k^T g_k}{g_k^T A g_k}$$

$$\alpha_k = \frac{\|g_k\|_2^2}{\left(A g_k, g_k\right)_{\mathbb{R}^n}}$$

Metoda pantei maxime - forme pătratice

$$x_0 \; - \; \text{dat}$$

$$g_k = Ax_k - b \quad , \quad \alpha_k = \frac{g_k^T g_k}{g_k^T A g_k}$$

$$x_{k+1} = x_k - \alpha_k g_k \quad , \quad k = 0, 1, \ldots$$

Metoda pantei maxime - f oarecare

$$x_0 \;-\; \text{dat}$$

$$g_k = \nabla f(x_k)$$

$$\alpha_k = \arg\min \{ f(x_k - \alpha\, g_k)\,;\ \alpha \in [0, \bar{\alpha}) \}$$

$$x_{k+1} = x_k - \alpha_k\, g_k \quad , \quad k = 0, 1, \ldots$$

Metoda pantei maxime

Pentru f formă pătratică:

$$E(x) = \frac{1}{2}(x - x^*)^T A(x - x^*) = f(x) + \frac{1}{2}(x^*)^T A x^*$$

$$\nabla E(x) = \nabla f(x) = g(x)$$

Pentru forme pătratice, şirul construit cu metoda pantei maxime satisface relaţia:

$$E(x_{k+1}) = \left[1 - \frac{\left(g_k^T g_k\right)^2}{\left(g_k^T A g_k\right)\left(g_k^T A^{-1} g_k\right)}\right] E(x_k)$$

Inegalitatea lui Kantorovici

Pentru $A \in \mathbb{R}^{n\times n}$, simetrică şi pozitiv definită are loc inegalitatea:

$$\frac{\left(x^T x\right)^2}{\left(x^T A x\right)\left(x^T A^{-1} g_x\right)} \geq \frac{4\, c\, C}{(c + C)^2}$$

unde c este cea mai mică valoare proprie a matricei A şi C este cea mai mare valoare proprie

Metoda pantei maxime - convergenţa

Teorema 6 (cazul pătratic)

Fie $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ o formă pătratică. Pentru orice iteraţie iniţială x_0, şirul construit cu metoda pantei maxime converge la x^* unicul punct de minim al funcţiei f, $x_k \longrightarrow x^*$, $k \longrightarrow \infty$. Are loc următoarea relaţie:

$$E(x_{k+1}) \leq \left(\frac{C - c}{C + c} \right)^2 E(x_k).$$

Metoda pantei maxime - convergenţa

Teorema 7 (cazul general)

Fie funcţia $f : \mathbb{R}^n \longrightarrow \mathbb{R}, f \in C^2(\mathbb{R}^n$. Presupunem că $x^* \in \mathbb{R}^n$ este un punct de minim local al funcţiei f. Presupunem că matricea Hessiană $H_f(x^*) = \nabla^2 f(x^*)$ are $c > 0$ cea mai mică valoare proprie şi $C > 0$ cea mai mare valoare proprie (matricea Hessiană $H_f(x^*)$ este pozitiv definită). Dacă şirul $\left(x_k\right)$ construit cu metoda pantei maxime converge la x^*, $x_k \longrightarrow x^*, k \longrightarrow \infty$ atunci are loc următoarea relaţie:

$$f(x_{k+1}) - f(x^*) \leq \left(\frac{C - c}{C + c}\right)^2 \left[f(x_k) - f(x^*) \right].$$

Metoda gradientului descendent

$$x_0 \; - \; \text{dat}$$

$$x_{k+1} = x_k - \eta \nabla f(x_k) \quad , \quad k = 0, 1, \ldots$$

η este o valoare mică, pozitivă ($\eta = 10^{-1}, 10^{-2}, \ldots$), care poartă numele de **rată de învăţare**.
Pentru probleme de maximizare metoda se numeşte a gradientului ascendent şi are forma:

$$x_0 \; - \; \text{dat}$$

$$x_{k+1} = x_k + \eta \nabla f(x_k) \quad , \quad k = 0, 1, \ldots$$

Metoda Newton

Fie x_k iteraţia curentă a şirului care aproximează soluţia problemei de minimizare:

$$\min\{f(x)\,;\, x \in \mathbb{R}^n\}.$$

Scriem dezvoltarea în serie Taylor de ordinul 2 în jurul punctului x_k:

$$f(x) = g(y) + R(\|x - x_k\|^3) \quad , \quad y = x - x_k$$

$$\begin{aligned} g(y) &= f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k) \\ &= \frac{1}{2} y^T A y - y^T b + c\,, \end{aligned}$$

$$A = \nabla^2 f(x_k) \quad , \quad b = -\nabla f(x_k) \quad , \quad c = f(x_k)$$

Metoda Newton

Problema de minimizare:

$$\min\{\, g(y)\, ;\, y \in \mathbb{R}^n \,\}$$

are un unic punct de minim y^*, dacă matricea A este simetrică şi pozitiv definită. Aceast punct de minim este soluţia sistemului liniar $Ay = b$:

$$y^* = A^{-1}b = \left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k)$$

Şirul construit cu **metoda Newton** are următoarea formulă de recurenţă:

$$x_{k+1} = x_k - \left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k)\ ,\ k = 0, 1, \ldots\ \ , x_0 - \text{dat}$$

Metoda Newton - convergenţa

Teorema 8

Fie $f \in C^3(\mathbb{R}^n)$ o funcţie care are un punct de minim local x^*, pentru care matricea Hessiană, $\nabla^2 f(x^*)$, este pozitiv definită. Dacă iteraţia iniţială x_0 este suficient de apropiată de soluţia x^*, $\|x_0 - x^*\| \leq r$, atunci şirul $\{x_k\}$ construit cu metoda lui Newton converge la x^* şi ordinul de convergenţă este cel puţin 2:

$$x_k \longrightarrow x^* \quad , \quad k \longrightarrow \infty \quad , \quad \|x_0 - x^*\| \leq r.$$

Metode cvasi-Newton - BFGS

Se folosesc aproximări, calculate iterativ, ale matricei Hessiene sau ale inversei acestei matrice.

Metoda Broyden-Fletcher-Goldfarb-Shanno

1. x_0 - dat, $H_0 = I_n$ sau $H_0 = a\, I_n$, $k = 0$;
2. Calculeaza direcţia de căutare $p_k = -H_k \nabla f(x_k)$;
3. Calculează $\alpha_k > 0$ cu ajustarea inexactă a pasului Wolfe
4. $x_{k+1} = x_k + \alpha_k p_k$
5. $s_k = x_{k+1} - x_k \quad , \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k) \quad , \quad \rho_k = \dfrac{1}{y_k^T s_k}$
6. Se actualizează matricea H_{k+1}
$$H_{k+1} = \left(I - \rho_k y_k s_k^T\right) H_k \left(I - \rho_k y_k s_k^T\right) + \rho_k s_k s_k^T$$
7. $k = k + 1$

Metode cvasi-Newton

Metoda **SR1** (symmetric-rank-1)

$$H_{k+1} = H_k + \frac{1}{(s_k - H_k y_k)^T y_k}\left[(s_k - H_k y_k)(s_k - H_k y_k)^T\right]$$

Metode **Broyden**

$$B_{k+1} = B_k + \frac{1}{u_k^T s_k}(y_k - B_k s_k)u_k^T$$

Alegerea vectorului u_k defineşte metoda de tip Broyden (este 'parametrul' metodei Broyden).
Cu H_k se notează aproximări ale inversei matricei Hessiene iar cu B_k aproximări ale matricei Hessiene.

Metoda gradienţilor conjugaţi (a direcţiilor conjugate)

Fie $A \in \mathbb{R}^{n\times n}$, o matrice simetrică, $A = A^T$ şi pozitiv definită şi $b \in \mathbb{R}^n$.

Fie problema de minimizare a formei pătratice definite cu matricea A şi vectorul b:

$$\min\left\{ f(x) = \frac{1}{2}x^T A x - x^T b\,;\ x \in \mathbb{R}^n \right\}.$$

Definiţie

Pentru o matrice $A \in \mathbb{R}^{n\times n}$, simetrică, $A = A^T$, doi vectori $d_1, d_2 \in \mathbb{R}^n$ se numesc ***A*-ortogonali** sau **conjugaţi în raport cu *A*** dacă:

$$d_1^T A d_2 = \left(Ad_2, d_1\right)_{\mathbb{R}^n} = \left(d_2, Ad_1\right)_{\mathbb{R}^n} = \left(Ad_1, d_2\right)_{\mathbb{R}^n} = 0.$$

Metoda gradienţilor conjugaţi

$$A = 0_{n\times n} \quad \Rightarrow \quad \forall\, d_1, d_2 \text{ sunt } A \text{ ortogonali}$$
$$A = I_n \quad \Rightarrow \quad \text{ortogonalitate clasică, } \left(d_1, d_2\right)_{\mathbb{R}^n} = 0.$$

Vectorii $\{d_0, d_1, \ldots, d_k\}$ se numesc ***A*-ortogonali** sau ***A*-conjugaţi** dacă:

$$d_i^T A d_j = \left(A d_j, d_i\right)_{\mathbb{R}^n} = 0 \;,\; \forall i \neq j, i, j = 0, 1, \ldots, k.$$

Propoziţie

Fie $A \in \mathbb{R}^{n\times n}$, o matrice simetrică, $A = A^T$ şi pozitiv definită şi $\left\{d_0, d_1, \ldots, d_k\right\}$ direcţii A-conjugate, $d_i \neq 0, \forall i = 0, \ldots, k$. Atunci vectorii $\left\{d_0, d_1, \ldots, d_k\right\}$ sunt liniar independenţi.

Metoda gradienţilor conjugaţi

Dacă $A \in \mathbb{R}^{n\times n}$, este matrice simetrică, $A = A^T$ şi pozitiv definită, iar $\left\{d_0, d_1, \ldots, d_{n-1}\right\}$ sunt n direcţii A-conjugate atunci $\left\{d_0, d_1, \ldots, d_{n-1}\right\}$ formează o bază în $\mathbb{R}^n$.

$$x^* = \arg\min\{f(x)\,;\, x \in \mathbb{R}^n\} \Leftrightarrow x^* = A^{-1}b \quad (Ax^* = b)$$

Putem scrie soluţia x^* în raport cu baza de n vectori A-conjugaţi, $\left\{d_0, d_1, \ldots, d_{n-1}\right\}$:

$$x^* = \alpha_0\, d_0 + \alpha_1\, d_1 + \cdots + \alpha_{n-1}\, d_{n-1}$$

Metoda gradienţilor conjugaţi

Din relaţia:

$$\left(Ax^*, d_i\right)_{\mathbb{R}^n} = \left(b, d_i\right)_{\mathbb{R}^n} = d_i^T A x^* = \alpha_i \, d_i^T A d_i$$

rezultă că:

$$\alpha_i = \frac{d_i^T b}{d_i^T A d_i} = \frac{\left(b\,, d_i\right)_{\mathbb{R}^n}}{\left(Ad_i\,, d_i\right)_{\mathbb{R}^n}}$$

Dacă ştim n direcţii A-conjugate, soluţia sistemului $Ax = b$ şi a problemei de minimizare cu forme pătratice se poate scrie astfel:

$$x^* = \sum_{i=0}^{n-1} \frac{\left(b\,, d_i\right)_{\mathbb{R}^n}}{\left(Ad_i\,, d_i\right)_{\mathbb{R}^n}} d_i$$

Metoda gradienţilor conjugaţi

Procesul iterativ:

$$
\begin{aligned}
& x_0 \in \mathbb{R}^n - \text{dat} \\
& g_k = Ax_k - b \quad , \quad \alpha_k = -\frac{g_k^T d_k}{d_k^T A d_k} \\
& x_{k+1} = x_k + \alpha_k\, d_k \ \ , \ \ k = 0, 1, 2, \ldots, n-1
\end{aligned}
$$

are proprietatea că $x_n = x^*$.
Considerăm următoarele relaţii:

$$x_k = x_0 + \alpha_0\, d_0 + \alpha_1\, d_1 + \cdots + \alpha_{k-1}\, d_{k-1}$$

$$x^* = x_0 + \beta_0\, d_0 + \beta_1\, d_1 + \cdots + \beta_{n-1}\, d_{n-1}$$

Metoda gradienţilor conjugaţi

Avem urmtoarele relaţii:

$$\alpha_i = \frac{d_i^T A(x_k - x_0)}{d_i^T A d_i} \quad , \quad \beta_i = \frac{d_i^T A(x^* - x_0)}{d_i^T A d_i}$$

$$d_k^T A(x_k - x_0) = 0.$$

$$x^* - x_k = (\beta_0 - \alpha_0)\, d_0 + \cdots + (\beta_{k-1} - \alpha_{k-1})\, d_{k-1} + \beta_k\, d_k + \ldots \beta_{n-1}\, d_{n-1}.$$

$$g_k^T d_i = 0 \quad \forall i < k.$$

Algoritmul gradienţilor conjugaţi

$$x_0 \in \mathbb{R}^n \ , \ g_0 = Ax_0 - b$$
$$d_0 = -g_0 = b - Ax_0$$
$$\alpha_k = -\frac{g_k^T d_k}{d_k^T A d_k}$$
$$x_{k+1} = x_k + \alpha_k d_k$$
$$g_{k+1} = Ax_{k+1} - b \text{ sau } g_{k+1} = g_k + \alpha_k A d_k$$
$$\beta_k = \frac{g_{k+1}^T A d_k}{d_k^T A d_k}$$
$$d_{k+1} = -g_{k+1} + \beta_k d_k$$

Metoda gradienţilor conjugaţi

Fie $y_1, y_2, \ldots, y_p \in \mathbb{R}^n$. Facem următoarea notaţie:

$$\begin{aligned}\text{span}\{y_1, y_2, \ldots, y_p\} &= \{y \in \mathbb{R}^n ; y = a_1 y_1 + \cdots + a_p y_p,\ a_i \in \mathbb{R}\} \\ &= \text{subspaţiul generat de vectorii } y_1, y_2, \ldots\end{aligned}$$

$$\mathcal{K}(g_0, k) = \text{span}\{g_0, Ag_0, \ldots, A^k g_0\}$$

Acest subspaţiu se numeşte **subspaţiu Krylov de grad k pentru vectorul g_0**.

Metoda gradienţilor conjugaţi

Teorema 9

Presupunem că $x_k \neq x^*$. Avem relaţiile:

(1) $g_k^T g_i = 0 \quad , \ \forall\, i = 0, 1, \ldots, k-1$

(2) $\mathrm{span}\{g_0, g_1, \ldots, g_k\} = \mathrm{span}\{g_0, Ag_0, \ldots, A^k g_0\}$

(3) $\mathrm{span}\{d_0, d_1, \ldots, d_k\} = \mathrm{span}\{g_0, Ag_0, \ldots, A^k g_0\}$

(4) $d_k^T A d_i = 0 \quad , \ \forall\, i = 0, 1, \ldots, k-1$

Şirul x_k converge la soluţia x^* în cel mult n paşi.

Algoritmul gradienţilor conjugaţi - forma practică

$$
\begin{aligned}
& x_0 \in \mathbb{R}^n - \text{dat} \text{ , } k = 0 \\
& g_0 = Ax_0 - b \text{ , } d_0 = -g_0 \\
& \textbf{while } (g_k \neq 0) \\
& \qquad \alpha_k = -\frac{g_k^T g_k}{d_k^T A d_k} \\
& \qquad x_{k+1} = x_k + \alpha_k d_k \\
& \qquad g_{k+1} = g_k + \alpha_k A d_k \\
& \qquad \beta_k = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \\
& \qquad d_{k+1} = -g_{k+1} + \beta_k d_k \\
& \qquad k = k + 1
\end{aligned}
$$

Metoda gradienţilor conjugaţi - convergenţa

Cu ajutorul unei matrice $A\ in\mathbb{R}^{n\times n}$ simetrică şi pozitiv definită, se defineşte următoarea normă matriceală:

$$\|x\|_A = \sqrt{(Ax, x)_{\mathbb{R}^n}}$$

Teorema 10

Dacă matricea A are doar r valori proprii distincte, atunci algoritmul gradienţilor conjugaţi calculează soluţia x^* în cel mult r iteraţii

Metoda gradienţilor conjugaţi - convergenţa

Teorema 11

Dacă $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ sunt valorile proprii ale matricei A atunci:

$$\|x_{k+1} - x^*\|_A^2 \leq \left(\frac{\lambda_{n-k} - \lambda_1}{\lambda_{n-k} + \lambda_1}\right)^2 \|x_0 - x^*\|_A^2$$

$$\|x_{k+1} - x^*\|_A^2 = 2E(x_{k+1})$$

$$E(x_{k+1}) \leq \left(\frac{\lambda_{n-k} - \lambda_1}{\lambda_{n-k} + \lambda_1}\right)^2 E(x_0)$$

Metodele gradienţilor conjugaţi neliniare

$$\min\{f(x)\,;\,x\in\mathbb{R}^n\}\quad,\quad f\ \text{funcţie oarecare, neliniară}$$

Pentru a calcula elementul x_{k+1}, se foloseşte aproximarea pătrtică a funcţiei f dată de dezvoltarea în serie Taylor în jurul punctului x_k:

$$f(x)\approx f(x_k)=f(x_k)+(x-x_k)^T\nabla f(x_k)+\frac{1}{2}(x-x_k)^T\nabla^2 f(x_k)(x-x_k)$$

În metoda gradienţilor conjugaţi se fac următoarele înlocuiri:

$$\begin{aligned} g_k &\quad\leftrightarrow\quad \nabla f(x_k)\\ A &\quad\leftrightarrow\quad \nabla^2 f(x_k)\end{aligned}$$

Algoritmul gradienţilor conjugaţi - funcţii oarecare

$$
\begin{aligned}
& x_0 \in \mathbb{R}^n - dat\ ,\ k = 0 \\
& g_0 = \nabla f(x_0); d_0 = -g_0 \\
& \textbf{while } (g_k \neq 0) \\
& \quad A = \nabla^2 f(x_k); \\
& \quad \alpha_k = -\frac{g_k^T d_k}{d_k^T A d_k} \\
& \quad x_{k+1} = x_k + \alpha_k d_k \\
& \quad g_{k+1} = \nabla f(x_{k+1}) \\
& \quad \beta_k = \frac{g_{k+1}^T A d_k}{d_k^T A d_k} \\
& \quad d_{k+1} = -g_{k+1} + \beta_k d_k \\
& \quad k = k + 1
\end{aligned}
$$

Algoritmul gradienţilor conjugaţi - metoda Fletcher-Reeves

$x_0 \in \mathbb{R}^n - dat\ ,\ k = 0$

$g_0 = \nabla f(x_0); d_0 = -g_0$

while $(g_k \neq 0)$

$\alpha_k = \min \{ f(x_k + \alpha d_k)\, ;\alpha \in [0, \bar{\alpha})\, \}$

(exact sau inexact cu testul Wolfe)

$x_{k+1} = x_k + \alpha_k\, d_k$

$g_{k+1} = \nabla f(x_{k+1})$

$\beta_k^{\mathrm{FR}} = \dfrac{g_{k+1}^T g_k}{g_k^T g_k}$

$d_{k+1} = -g_{k+1} + \beta_k^{\mathrm{FR}}\, d_k$

$k = k + 1$

Metoda Fletcher-Reeves

Sunt vectorii d_k construiţi cu metoda Fletcher-Reeves direcţii de descreştere?

$$d_k = -g_k + \beta_{k-1} d_{k-1}$$

$$g_k^T d_k = -g_k^T g_k + \beta_{k-1} g_k^T d_{k-1}$$

Dacă se foloseşte ajustarea exactă a pasului: α_{k-1} este punctul de minim local pentru funcţia f pe direcţia d_{k-1}, ceea ce implică faptul că:

$$\nabla f(x_k)^T d_{k-1} = g_k^T d_{k-1} = 0$$

Prin urmare:

$$g_k^T d_k = -g_k^T g_k = -\|g_k\|_2^2 < 0 \Rightarrow d_k \text{ direcţie de descreştere.}$$

Metoda Fletcher-Reeves

Dacă se foloseşte ajustarea inexactă a pasului am putea avea $g_k T\ d_k > 0$ (d_k direcţie de creştere!!) dar dacă se foloseşte testului Wolfe avem:

$$f(x_k + \alpha_k\ d_k) \leq f(x_k) + \epsilon\ \alpha_k\ g_k^T\ d_k$$

$$|\nabla f(x_k + \alpha_k\ d_k)^T d_k| \leq (1 - \epsilon)|g_k^T d_k|$$

$$\epsilon \in \left(0, \frac{1}{2}\right) \Rightarrow \quad g_k^T\ d_k < 0.$$

Metoda Polak-Ribière

Este o variantă a metodei Fletcher-Reeves:

$$\beta_k^{\mathrm{PR}} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{g_k^T g_k}$$

Dacă se face ajustarea inexactă a pasului cu testul lui Wolfe, nu putem deduce că d_k sunt direcţii de descreştere. Se foloseşte modificarea:

$$\beta_k^{+} = \max\{\beta_k^{\mathrm{PR}}, 0\}$$

şi un test Wolfe adaptat pentru a face direcţiile d_k, direcţii de descreştere.

Varianta Hestenes-Stiefel

$$\beta_k^{\mathrm{HS}} = \frac{g_{k+1}^T(g_{k+1} - g_k)}{(g_{k+1} - g_k)^T d_k}$$