Calcul Numeric

Cursul 11

2022

Anca Ignat

Metoda pantei maxime

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

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

d_k direcţie de descreştere a lui f în x_k

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

d_k direcţie de descreştere dacă :

$$\nabla f(x_k)^T d_k < 0.$$

Metoda pantei maxime:

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

Cazul pătratic:

$$f(x) = \frac{1}{2} x^T A x - x^T b = \frac{1}{2}\left(Ax, x\right)_{\mathbb{R}^n} - \left(b, x\right)_{\mathbb{R}^n}$$

$b \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$ simetrică şi pozitiv definită

A pozitiv definită $\rightarrow$ $\det A \neq 0$, f este strict convexă

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

$$f(x^*) = \min\{f(x);\ x \in \mathbb{R}^n\}$$ x^* punct unic de minim $\Leftrightarrow$

x^* soluţia sistemului liniar $Ax = b$, $x^* = A^{-1}b$

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

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

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

$$f(x_k - \alpha g_k) = \frac{1}{2}(x_k - \alpha g_k)^T A(x_k - \alpha g_k) - (x_k - \alpha g_k)^T b =$$

$$= \frac{1}{2}\left(g_k^T A g_k\right)\alpha^2 - \left(g_k^T A x_k - g_k^T b\right)\alpha + f(x_k) =$$

$$= \frac{1}{2}\left(g_k^T A g_k\right)\alpha^2 - \left(g_k^T g_k\right)\alpha + f(x_k)$$

$f(x_k - \alpha g_k)$ ecuaţie de gr. 2 în α, coef. lui $\alpha^2, g_k^T A g_k > 0$

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

Metoda pantei maxime pentru funcționale pătratice:

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

$$\boldsymbol{x_{k+1} = x_k - \left(\frac{g_k^T g_k}{g_k^T A g_k} \right) g_k \quad , \quad k = 0,1,\ldots}$$

$$\boldsymbol{g_k = A x_k - b}$$

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

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

Şirul construit cu metoda pantei maxime satisface:

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

Inegalitatea lui Kantorovich

$\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{A} = \boldsymbol{A}^T$, $\boldsymbol{A} > \boldsymbol{0}$ pozitiv definită

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

$\boldsymbol{c}$, $\boldsymbol{C}$ - cea mai mica şi cea mai mare valoare proprie a lui $\boldsymbol{A}$.

Teoremă

Cazul pătratic: Pentru orice iteraţie iniţială x_0, şirul construit cu metoda pantei maxime converge la x^* unicul punct de minim al funcţiei f. Avem:

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

Cazul general: $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2(\mathbb{R}^n)$, f are un punct de minim local x^*. Presupunem că $F(x^*) = \nabla^2 f(x^*)$ are $c > 0$ cea mai mică valoare proprie şi $C > 0$ cea mai mare valoare proprie. Dacă şirul $\{x_k\}$ construit cu metoda pantei maxime converge la x^* (la fiecare pas $A = \nabla^2 f(x_k)$, $b = \nabla f(x_k)$), $x_k \to x^*$

atunci $f(x_k) \to f(x^*)$ converge liniar cu o rată de convergenţă $\leq \left(\frac{C-c}{C+c}\right)^2$.

Metoda Newton

x_k - punctul curent de aproximare

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

g – funcţională pătratică, $y = x - x_k$,

$$g(y) = \frac{1}{2} y^T A y - y^T b + c$$

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

$y^* = \arg\min\{g(y);\ y \in \mathbb{R}^n\}$ este unica soluție a sistemului liniar $Ay = b$, $y^* = A^{-1}b = -\left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k)$

Metoda Newton

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

Teoremă

Fie $\boldsymbol{f} \in \boldsymbol{C}^{3}(\mathbb{R}^{\boldsymbol{n}})$ care are un punct de minim local $\boldsymbol{x}^*$ astfel ca matricea $\nabla^2 \boldsymbol{f}(\boldsymbol{x}^*) > \mathbf{0}$ este pozitiv definită. Dacă punctul de început $\boldsymbol{x}_{\mathbf{0}}$ este suficient de aproape de $\boldsymbol{x}^*$, $\|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \boldsymbol{r}$, atunci şirul $\{\boldsymbol{x}_{\boldsymbol{k}}\}$ generat cu metoda lui Newton converge la $\boldsymbol{x}^*$ şi ordinul de convergenţă este cel puţin 2.

Metoda gradienţilor conjugaţi (a direcţiilor conjugate)

$$\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{n}\times\boldsymbol{n}} \;\;,\;\; \boldsymbol{A} = \boldsymbol{A}^{\boldsymbol{T}} \;\;,\;\; \boldsymbol{A} > \boldsymbol{0} \text{ pozitiv definită}$$

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

Definiţie

Pentru o matrice $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{n}\times\boldsymbol{n}}$, $\boldsymbol{A} = \boldsymbol{A}^{\boldsymbol{T}}$ simetrică, doi vectori $\boldsymbol{d}_1, \boldsymbol{d}_2 \in \mathbb{R}^{\boldsymbol{n}}$ se numesc ***A*-ortogonali** sau **conjugaţi în raport cu *A*** dacă:

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

$$A = 0_{n\times n} \Rightarrow \forall d_1, d_2 \text{ sunt } A - \text{ortogonali}$$

$$A = I_n \Rightarrow \text{ortogonalitate clasică } (d_1, d_2)_{\mathbb{R}^n} = 0$$

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

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

Propoziţie

Fie $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{A} = \boldsymbol{A}^T$, $\boldsymbol{A} > \boldsymbol{0}$, şi $\{\boldsymbol{d}_0, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_k\}$ direcţii $\boldsymbol{A}$-conjugate, $\boldsymbol{d}_i \neq \boldsymbol{0}$, $\forall \boldsymbol{i} = \boldsymbol{0}, \ldots, \boldsymbol{k}$. Atunci vectorii $\{\boldsymbol{d}_0, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_k\}$ sunt liniar independenţi.

$$\left.\begin{array}{l} \boldsymbol{A} \in \mathbb{R}^{n \times n} \ , \ \boldsymbol{A} = \boldsymbol{A}^T \ , \ \boldsymbol{A} > \boldsymbol{0} \\ \\ \{\boldsymbol{d}_0, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_{n-1}\} \ \text{- direcţii } \boldsymbol{A}-\text{conjugate} \end{array}\right\} \Rightarrow \begin{array}{c} \{\boldsymbol{d}_0, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_{n-1}\} \\ \text{bază în } \mathbb{R}^n \end{array}$$

$$\boldsymbol{x}^* = \arg\min\{ \boldsymbol{f}(\boldsymbol{x});\ \boldsymbol{x} \in \mathbb{R}^n \} \Leftrightarrow \boldsymbol{x}^* \text{ soluţia sist. } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

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

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

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

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

Considerăm procesul iterativ

$$x_0 \in \mathbb{R}^n - \text{dat}$$

$$x_{k+1} = x_k + \alpha_k d_k \quad , k = 0,1,2,...,n-1$$

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

are proprietatea că $x_n = x^*$.

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

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

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

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

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

Corolar

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

Algoritmul gradienţilor conjugaţi

$$x_0 \in \mathbb{R}^n \;,\; g_0 = Ax_0 - b$$

$$d_0 = -g_0 = b - Ax_0$$

$$\alpha_k = -\frac{g_k^T d_k}{d_k^T A d_k}$$

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

$$g_{k+1} = Ax_{k+1} - b \text{ sau } g_{k+1} = g_k + \alpha_k A d_k$$

$$\beta_k = \frac{g_{k+1}^T A d_k}{d_k^T A d_k}$$

$$d_{k+1} = -g_{k+1} + \beta_k d_k$$

$$y_1, y_2, \dots, y_p \in \mathbb{R}^n$$

$$\text{span}\{y_1, y_2, \dots, y_p\} = \{y = a_1 y_1 + \cdots + a_p y_p \in \mathbb{R}^n;\ a_i \in \mathbb{R}, i = 1, \dots, p\}$$

= subspaţiul generat de vectorii $y_1, y_2, \dots, y_p$

Teoremă

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

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

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

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

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

Şirul $x_k \to x^*$ în cel mult n paşi.

$\mathcal{K}(\boldsymbol{g_0}, \boldsymbol{k}) = \mathrm{span}\{\boldsymbol{g_0}, \boldsymbol{A g_0}, ..., \boldsymbol{A^k g_0}\}$ - se numeşte **subspaţiu Krylov de grad *k*** pentru $\boldsymbol{g_0}$.

Forma practică a metodei gradienţilor conjugaţi

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

$$g_0 = Ax_0 - b, \quad d_0 = -g_0$$

$$\textit{while}\,(g_k \neq 0)$$

$$\begin{cases} \alpha_k = \dfrac{g_k^T g_k}{d_k^T A d_k} \\ x_{k+1} = x_k + \alpha_k d_k \\ g_{k+1} = g_k + \alpha_k A d_k \\ \beta_k = \dfrac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \\ d_{k+1} = -g_{k+1} + \beta_k d_k \\ k = k + 1; \end{cases}$$

Teoremă

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

Teoremă

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

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

$$g_k \leftrightarrow -\nabla f(x_k)$$

$$A \leftrightarrow \nabla^2 f(x_k)$$

Varianta pentru funcţii oarecare a metodei gradienţilor conjugaţi:

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

$$g_0 = \nabla f(x_0), \quad d_0 = -g_0$$

$$while\,(g_k \neq 0)$$

$$\begin{cases} A = \left[\nabla^2 f(x_k)\right] \\ \alpha_k = -\dfrac{g_k^T d_k}{d_k^T A d_k} \\ x_{k+1} = x_k + \alpha_k d_k \\ g_{k+1} = \nabla f(x_{k+1}) \\ \beta_k = \dfrac{g_{k+1}^T A d_k}{d_k^T A d_k} \\ d_{k+1} = -g_{k+1} + \beta_k d_k \\ k = k+1; \end{cases}$$

Metoda Fletcher-Reeves

α_k se calculează folosind metoda ajustării pasului

$$\beta_k = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}$$

$$x_0 \in \mathbb{R}^n - \text{dat},$$
$$g_0 = \nabla f(x_0), \quad d_0 = -g_0 \;, \; k = 0$$

$$\boldsymbol{while\,(g_k \neq 0)}$$

$$\begin{cases} \\ \alpha_k = \min\{ f(x_k + \alpha d_k);\, \alpha \in [0, \bar{\alpha})\} \\ \text{(exact sau inexact cu testul lui Wolfe)} \\ x_{k+1} = x_k + \alpha_k d_k \\ g_{k+1} = \nabla f(x_{k+1}) \\ \beta_k^{FR} = \dfrac{g_{k+1}^T g_k}{g_k^T g_k} \\ d_{k+1} = -g_{k+1} + \beta_k^{FR} d_k \\ k = k+1; \end{cases}$$

Se pune problema dacă $\boldsymbol{d}_k$ sunt direcţii de descreştere?

$$\boldsymbol{d}_k = -\boldsymbol{g}_k + \boldsymbol{\beta}_{k-1}\boldsymbol{d}_{k-1}$$

$$\boldsymbol{g}_k^T\boldsymbol{d}_k = -\boldsymbol{g}_k^T\boldsymbol{g}_k + \boldsymbol{\beta}_{k-1}\boldsymbol{g}_k^T\boldsymbol{d}_{k-1}$$

Dacă se foloseşte ajustarea pasului exactă:
$\boldsymbol{\alpha}_{k-1}$ este punct de minim local pentru $\boldsymbol{f}$ pe direcţia $\boldsymbol{d}_{k-1}$ prin urmare $\boldsymbol{g}_k^T\boldsymbol{d}_{k-1} = \mathbf{0}$ $(\boldsymbol{g}_k = \nabla\boldsymbol{f}(\boldsymbol{x}_k))$.
$\Rightarrow$ $\boldsymbol{g}_k^T\boldsymbol{d}_k = -\boldsymbol{g}_k^T\boldsymbol{g}_k = -\|\boldsymbol{g}_k\|_2^2 < \mathbf{0} \Rightarrow \boldsymbol{d}_k$ direcţie de descreştere

Dacă se foloseşte ajustarea pasului inexactă am putea avea $g_k^T d_k > 0$ (d_k direcţie de creştere!!) dar folosind testul lui Wolfe deducem:

$$f(x_k + \alpha_k d_k) \le f(x_k) + \varepsilon \alpha_k g_k^T d_k$$

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

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

Metoda Polak-Ribière

- variantă a metodei Fletcher-Reeves

$$\boldsymbol{\beta}_k^{PR} = \frac{\boldsymbol{g}_{k+1}^T(\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)}{\boldsymbol{g}_k^T \boldsymbol{g}_k}$$

Dacă se face ajustarea pasului inexactă cu testul lui Wolfe nu putem deduce că $\boldsymbol{d}_k$ sunt direcţii de descreştere.

Se foloseşte $\boldsymbol{\beta}_k^+ = \mathbf{max}\{\boldsymbol{\beta}_k^{PR}, \mathbf{0}\}$ şi un test Wolfe adaptat pentru a obţine $\boldsymbol{d}_k$ direcţii de descreştere.

Varianta Hestenes-Stiefel

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

Precondiţionare

Se consideră norma:

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

Evaluarea erorii în metoda pantei maxime:

$$\| x^{(k)} - x^* \|_A \le \left(\frac{k(A) - 1}{k(A) + 1} \right)^k \| x^{(0)} - x^* \|_A$$

$k(A) = \dfrac{\lambda_n}{\lambda_1}$ – **numărul de condiţionare spectral**

$0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ **valorile proprii ale matricii** A

Avem convergenţă rapidă dacă numărul de condiţionare spectrală al matricei ***A*** este apropiat de **1** (***k(A)≥1*** întotdeauna).

Ideea precondiţionării este de a transforma sistemul ***Ax=b***

astfel încât să îmbunătăţim proprietăţile spectrale.

$$\boldsymbol{Ax = b} \;\; \rightleftarrows \;\; \boldsymbol{\tilde{A}x = \tilde{b}} \;,\; \textbf{cu} \;\; \boldsymbol{k(\tilde{A}) << k(A)}$$

Precondiţionare

$$Ax = b \rightarrow M^{-1}Ax = M^{-1}b \text{ (la stânga)}$$

$$\rightarrow AM^{-1}y = b \;, x = M^{-1}y \text{ (la dreapta)}$$

$$\rightarrow M_1^{-1}AM_2^{-1}y = M_1^{-1}b \;, x = M_2^{-1}\, y \text{ (split)}, M = M_1 M_2$$

cu M matrice nesingulară , M "$\approx$" A. Matricea M sau M^{-1} poartă numele de ***matrice de precondiţionare***.

Cum trebuie să alegem matricea $\boldsymbol{M}$?

- sistemul precondiţionat ($\tilde{\boldsymbol{A}}\boldsymbol{x} = \tilde{\boldsymbol{b}}$) să fie uşor de rezolvat (convergenţă rapidă)

- matricea de precondiţionare să fie economic de construit şi aplicat – ietraţiile să nu fie costisitor de construit

Matricea de precondiţionare Jacobi

$$M = \mathrm{diag}[a_{11}, a_{22}, \ldots, a_{nn}]$$

Matrice de precondiţionare SSOR

$$M = (D+L)\,D^{-1}\,(D+L)^T \quad (A = L + D + L^T)$$

$$M(\omega) = \frac{1}{2-\omega}\left(\frac{1}{\omega}D + L\right)\left(\frac{1}{\omega}D\right)^{-1}\left(\frac{1}{\omega}D + L\right)^T, \ \omega \in (0,2)$$

Pentru ω - optimal, în anumite cazuri:

$$k(M(\omega_{opt})^{-1}A) = O(\sqrt{k(A)})$$

(ω_{opt} - foarte costisitor de calculat)