

VII Wykład: 10 XII 10 r.

Opracowano na podstawie: [3, ss. 381–393], [5], [6, ss. 247–260], [8, ss. 146–159], [7] i [9, ss. 402–408].

VII–1.0.1. Wstęp

Będziemy zajmować się wyznaczeniem ciągu $x^k \in \mathbb{R}^n$, $k = 0, 1, 2, \dots$ zbieżnego do rozwiązania $x^* \in \mathbb{R}^n$ układu

$$Ax = b,$$

gdzie macierz $A \in \mathbb{R}^{n \times n}$ jest symetryczna $A = A^T$ i dodatnio określona tzn. $x^T Ax = \langle x, Ax \rangle > 0$, $x \neq 0$. Podstawą rozważanych tu metod będzie funkcja

$$\Phi(x) = \frac{1}{2}x^T Ax - x^T b = \frac{1}{2}\langle x, Ax \rangle - \langle x, b \rangle - \text{energia układu } Ax = b.$$

VII–1.1 Lemat

Niech macierz A będzie symetryczna i dodatnio określona. Wtedy, dla $x \in \mathbb{R}^n$ i kierunku $v \in \mathbb{R}^n$, $v \neq 0$ minimum funkcji

$$\varphi(t) \stackrel{\text{def}}{=} \Phi(x + tv), \quad t \in \mathbb{R}$$

jest przyjęte dla

$$t_0 = \frac{\langle v, b - Ax \rangle}{\langle v, Av \rangle}. \quad (\text{VII–1.1})$$

i wynosi

$$\varphi(t_0) = \Phi(x + t_0 v) = \Phi(x) - \frac{1}{2} \frac{\langle v, b - Ax \rangle^2}{\langle v, Av \rangle}. \quad (\text{VII–1.2})$$

Dowód:

Dla danych wektorów $x, v \in \mathbb{R}^n$ i dowolnej liczby rzeczywistej $t \in \mathbb{R}$ mamy

$$\begin{aligned} \varphi(t) &= \Phi(x + tv) = \\ &= \frac{1}{2}\langle x + tv, A(x + tv) \rangle - \langle x + tv, b \rangle = \frac{1}{2}\langle x + tv, Ax + tAv \rangle - \langle x + tv, b \rangle = \\ &= \frac{1}{2}(\langle x, Ax \rangle + t\langle x, Av \rangle + t\langle v, Ax \rangle + t^2\langle v, Av \rangle) - \langle x, b \rangle - t\langle v, b \rangle = \\ &= \Phi(x) + t\langle v, Ax - b \rangle + \frac{1}{2}t^2\langle v, Av \rangle. \end{aligned}$$

W przeprowadzonych rachunkach skorzystaliśmy z własności $\langle x, Av \rangle = \langle v, Ax \rangle$ aby otrzymać

$$\varphi(t) = \Phi(x) + t\langle v, Ax - b \rangle + \frac{1}{2}t^2\langle v, Av \rangle. \quad (\text{VII–1.3})$$

Prawa strona (VII–1.3) jest trójmianem kwadratowym o współczynniku przy t^2 dodatnim w takim razie istnieje $t_0 \in \mathbb{R}$ o własności

$$\Phi(x + t_0 v) = \min_{t \in \mathbb{R}} \Phi(x + tv)$$

Warunek konieczny na ekstremum w t_0 jest następujący

$$\left. \frac{d\Phi(x + tv)}{dt} \right|_{t=t_0} = 0.$$

Ze wzoru (VII-1.3) otrzymujemy, że

$$\left. \frac{d\Phi(x + tv)}{dt} \right|_{t=t_0} = \langle v, Ax - b \rangle + t_0 \langle v, Av \rangle = 0.$$

A więc

$$t_0 = \frac{\langle v, b - Ax \rangle}{\langle v, Av \rangle}.$$

Dla tak wyznaczonego t_0 obliczamy

$$\begin{aligned} \varphi(t_0) &= \Phi(x + t_0 v) = \\ &= \Phi(x) + \frac{\langle v, b - Ax \rangle}{\langle v, Av \rangle} \langle v, Ax - b \rangle + \frac{1}{2} \frac{\langle v, b - Ax \rangle^2}{\langle v, Av \rangle^2} \langle v, Av \rangle = \\ &= \Phi(x) - \frac{1}{2} \frac{\langle v, b - Ax \rangle^2}{\langle v, Av \rangle}. \end{aligned}$$

□

VII-1.2 Lemat

Jeżeli macierz A jest symetryczna i dodatnio określona, to wektor $x^ \in \mathbb{R}^n$ jest rozwiązaniem układu*

$$Ax = b$$

wtedy i tylko wtedy, gdy

$$\Phi(x^*) = \min_{x \in \mathbb{R}^n} \Phi(x). \quad (\text{VII-1.4})$$

Dowód:

Założymy, że $x^* \in \mathbb{R}^n$ spełnia równanie $Ax^* = b$. Przyjmując w (VII-1.3), dla dowolnego $x \in \mathbb{R}^n$ za $v = x - x^*$ i $t = 1$ otrzymujemy:

$$\Phi(x) = \Phi(x^* + v) = \Phi(x^*) + \langle v, Ax^* - b \rangle + \frac{1}{2} \langle v, Av \rangle \geq \Phi(x^*).$$

Skąd wynika, iż

$$\Phi(x^*) = \min_{x \in \mathbb{R}^n} \Phi(x).$$

Niech¹ $x^* \in \mathbb{R}^n$ spełnia (VII-1.4) i założmy nie wprost, że $Ax^* - b \neq 0$. Gdy teraz przyjmimy $v = Ax^* - b$ to z lematu VII-1.1 mamy

$$\Phi(x^* + t_0 v) = \Phi(x^*) - \frac{1}{2} \frac{\langle v, b - Ax^* \rangle^2}{\langle v, Av \rangle} < \Phi(x^*).$$

A więc założenie $Ax^* - b \neq 0$ doprowadziło do sprzeczności z (VII-1.4).

□

W [10, t. II, ss. 242–246] lemat VII-1.2 uzasadnia się w następujący sposób

VII-1.3 Uwaga

Niech

$$F(x) = \frac{1}{2} (Ax - b)^T A^{-1} (Ax - b) = \frac{1}{2} \langle Ax - b, A^{-1} (Ax - b) \rangle,$$

¹ K.T. Istnienie takiego x^* wynika z ogólnej teorii form dwuliniowych dodatnio określonych.

gdzie macierz A jest symetryczna dodatnio określona. Wtedy

$$F(x) = \Phi(x) + \frac{1}{2}b^T A^{-1}b$$

oraz $x^* \in \mathbb{R}^n$ jest rozwiązaniem $Ax^* = b$ jedynie wtedy, gdy spełnia

$$0 = F(x^*) = \min_{x \in \mathbb{R}^n} F(x).$$

Dowód:

Mamy

$$(A^{-1}A)^T = (AA^{-1})^T = I$$

skąd

$$A^T(A^{-1})^T = (A^{-1})^T A^T = I.$$

Co daje

$$(A^{-1})^T = (A^T)^{-1}.$$

Jeżeli więc macierz A jest symetryczna, to A^{-1} też. Następnie mamy

$$x^T A^{-1}x = (A^{-1}x)^T A(A^{-1}x)$$

skąd wynika, że symetryczna macierz A jest dodatnio określona jedynie wtedy, gdy A^{-1} też jest dodatnio określona. Teraz na podstawie definicji dodatniej określoności wynika teza.

Poza tym mamy

$$\begin{aligned} F(x) &= \frac{1}{2}(Ax - b)^T A^{-1}(Ax - b) = \\ &= \frac{1}{2}(x^T Ax - x^T b - b^T x + b^T A^{-1}b) = \Phi(x) + \frac{1}{2}b^T A^{-1}b \end{aligned}$$

□

VII-1.4 Definicja

Jeżeli $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$, to

$$\text{grad } \Phi(x) = \nabla \Phi(x) \stackrel{\text{def}}{=} [\Phi_{x_1}(x), \Phi_{x_2}(x), \dots, \Phi_{x_n}(x)]^T$$

gdzie

$$\Phi_{x_i}(x) = \frac{\partial \Phi(x)}{\partial x_i}, \quad i = 1, 2, \dots$$

Interpretacja geometryczna $\nabla \Phi(x)$ jest następująca.

Wiemy, że jeżeli $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ jest funkcją różniczkowalną, to pochodna w $\varphi'(t_0)$ opisuje szybkość zmian wartości funkcji $y = \varphi(t)$ w t_0 . Weźmy teraz punkt $x \in \mathbb{R}^n$ i wektor $v \in \mathbb{R}^n$, $\|v\|_2 = 1$ oraz dla $t \in \mathbb{R}$ rozważmy funkcję

$$\varphi(t) = \Phi(x + tv).$$

Wtedy, jeżeli $\Phi = \Phi(x)$ jest różniczkowalna, to funkcja $\varphi = \varphi(t)$ też jest różniczkowalna i pochodna $\varphi'(0)$ opisuje szybkość zmian funkcji $\varphi = \varphi(t)$ w 0. Ze wzoru na różniczkowanie funkcji złożonej mamy

$$\varphi'(0) = \left. \frac{\partial \Phi(x + tv)}{\partial t} \right|_{t=0} = \langle \nabla \Phi(x + tv) \big|_{t=0}, v \rangle = \langle \nabla \Phi(x), v \rangle$$

i otrzymana liczba opisuje szybkość zmian funkcji $\Phi = \Phi(x)$ w punkcie x w kierunku wektora v . Z geometrycznej interpretacji iloczynu skalarnego mamy

$$\min_{\|v\|_2=1} \langle \nabla \Phi(x), v \rangle = -\langle \nabla \Phi(x), \frac{\nabla \Phi(x)}{\|\nabla \Phi(x)\|_2} \rangle.$$

Tak więc wektor

$$v = -\frac{\nabla \Phi(x)}{\|\nabla \Phi(x)\|_2} \quad (\text{VII-1.5})$$

określa kierunek najszybszego malenia – spadku, funkcji $\Phi = \Phi(x)$.

Dla funkcji $\Phi = \Phi(x)$ opisującej energię układu $Ax = b$ mamy

$$\Phi(x) = \frac{1}{2}x^T Ax - x^T b = \frac{1}{2} \sum_{i=1}^n x_i \sum_{j=1}^n a_{ij} x_j - \sum_{i=1}^n x_i b_i.$$

Teraz obliczamy

$$\begin{aligned} \Phi_{x_k}(x) &= \frac{\partial \Phi(x)}{\partial x_k} = \frac{1}{2} \sum_{i=1}^n \frac{\partial}{\partial x_k} \left(x_i \sum_{j=1}^n a_{ij} x_j \right) - b_k = \\ &= \frac{1}{2} \sum_{i=1}^n \left(\frac{\partial x_i}{\partial x_k} \sum_{j=1}^n a_{ij} x_j + x_i \sum_{j=1}^n a_{ij} \frac{\partial x_j}{\partial x_k} \right) - b_k = \\ &= \frac{1}{2} \left(\sum_{i=1}^n \frac{\partial x_i}{\partial x_k} \sum_{j=1}^n a_{ij} x_j + \sum_{i=1}^n x_i \sum_{j=1}^n a_{ij} \frac{\partial x_j}{\partial x_k} \right) - b_k = \\ &= \frac{1}{2} \left(\sum_{j=1}^n a_{kj} x_j + \sum_{i=1}^n x_i a_{ik} \right) - b_k \end{aligned}$$

gdzie skorzystaliśmy z

$$\frac{\partial x_i}{\partial x_k} = \begin{cases} 1, & i = k \\ 0, & i \neq k \end{cases}$$

W ten sposób wyliczyliśmy, że gradient funkcji energii

$$\nabla \Phi(x) = \frac{1}{2}(Ax + A^T x) - b = Ax - b \quad (\text{VII-1.6})$$

jest wektorem residualnym (resztowym)

$$r = Ax - b.$$

VII-1.0.2. Metoda najszybszego spadku

Na podstawie lematu VII-1.2 rozwiązanie układu $Ax = b$ sprowadziliśmy do znalezienia punktu $x^* \in \mathbb{R}^n$ w którym funkcja energii $\Phi = \Phi(x)$ osiąga swoje minimum. W pierwszym kroku wybieramy w sposób dowolny punkt $x^0 \in \mathbb{R}^n$. Ze wzoru (VII-1.5) i (VII-1.6) określamy kierunek $v^0 \in \mathbb{R}^n$ w którym funkcja $\Phi = \Phi(x)$ najszybciej maleje. Na wybranym kierunku przy pomocy wzoru (VII-1.1) lematu VII-1.1 wyznaczamy nowy punkt x^1 dla którego całą procedurę powtarzamy. Przy tym sposobie postępowania otrzymujemy następujący algorytm

$$x^{k+1} = x^k + t_k v^k, \quad k = 0, 1, 2, \dots$$

gdzie

$$v^k = -\nabla\Phi(x^k) = b - Ax^k, \quad t_k = \frac{\langle v^k, b - Ax^k \rangle}{\langle v^k, Av^k \rangle} = \frac{v^{kT} v^k}{v^{kT} Av^k}.$$

Zaznaczmy tutaj, że w tej metodzie wybór kierunku oraz wyznaczenie na nim minimum jakie osiąga funkcja $\Phi = \Phi(x)$ są procesami niezależnymi. Zauważmy [10, t II, ss. 243], że kolejny punkt x^{k+1} iteracji w metodzie najszybszego spadku jest wyznaczony jednoznacznie z warunku

$$\Phi(x^{k+1}) = \min_{t \in \mathbb{R}} \Phi(x^k + tr^k), \quad k = 0, 1, 2, \dots$$

gdzie kierunek v^k szukania minimum

$$v^k = r^k = Ax^k - b$$

jest wektorem resztowym. W kolejnej metodzie – sprzężonych gradientów, punkt x^{k+1} będzie wyznaczony z warunku

$$\Phi(x^{k+1}) = \min_{t_0, t_1, \dots, t_k \in \mathbb{R}} \Phi(x^k + t_0 r^0 + t_1 r^1 + \dots + t_k r^k).$$

Zacniemy od wprowadzenia A. Iserlesa przedstawionego w [5, NA, w. 2, 3, 4].

VII–1.0.3. Metoda sprzężonych gradientów

Dla macierzy A symetrycznej, dodatnio określonej wprowadzamy w $\mathbb{R}^n$ nowy iloczyn skalarny

$$\langle x, y \rangle_A = \langle x, Ay \rangle_A = x^T Ay,$$

który generuje normę w $\mathbb{R}^n$ indukowaną przez formę A

$$\|x\|_A = \sqrt{\langle x, x \rangle_A} = \sqrt{x^T Ax}.$$

VII–1.5 Definicja (kierunki sprzężone)

Kierunki $p, q \in \mathbb{R}^n$, $p \neq 0$, $q \neq 0$ nazywamy A-sprężonymi lub A-ortogonalnymi, gdy

$$\langle p, q \rangle_A = p^T Aq = 0.$$

W metodzie najszybszego spadku kolejny punkt x^{k+1} jest wyznaczony ze wzoru postaci

$$x^{k+1} = x^k + t_k v^k,$$

gdzie v^k jest kierunkiem szukania minimum. Po nałożeniu macierzy A i odjęciu wektora b otrzymujemy

$$Ax^{k+1} - b = (Ax^k - b) + t_k Av^k,$$

skąd mamy, iż

$$\nabla\Phi(x^{k+1}) = \nabla\Phi(x^k) + t_k Av^k.$$

Oznaczmy

$$g^k \stackrel{\text{ozn}}{=} \nabla\Phi(x^k) = Ax^k - b.$$

wtedy mamy, że

$$g^{k+1} = g^k + t_k Av^k. \quad (\text{VII–1.7})$$

Jeśli teraz założymy, że kierunki v i v^k są sprzężone, to z równości

$$\langle v, g^{k+1} \rangle = \langle v, g^k \rangle + t_k \langle v, v^k \rangle_A$$

wynika, że jeżeli $\langle v, g^k \rangle = 0$ to też $\langle v, g^{k+1} \rangle = 0$. W metodzie sprzężonych gradientów podstawą jest lemat VII–1.1 a kolejne kierunki szukania minimum $\Phi(x)$ energii układu będą spełniać

$$\langle v^j, g^{k+1} \rangle = \langle v^j, \nabla \Phi(x^{k+1}) \rangle = 0, \quad j = 0, 1, \dots, k.$$

Dla dowolnego x^0 przyjmujemy $v^0 = -g^0$ i określamy ciąg

$$x^{k+1} = x^k + t_k v^k, \text{ gdzie } t_k = -\frac{\langle v^k, g^k \rangle}{\langle v^k, Av^k \rangle} \text{ dla } k = 0, 1, 2, \dots \quad (\text{VII-1.8})$$

a kierunek szukania minimum poprawiamy w ten sposób

$$v^k = -g^k + \beta_k v^{k-1}, \quad k = 1, 2, \dots \quad (\text{VII-1.9})$$

aby kierunki v^k i v^{k-1} były sprzężone tzn.

$$\begin{aligned} 0 &= \langle v^k, Av^{k-1} \rangle = \langle -g^k + \beta_k v^{k-1}, Av^{k-1} \rangle = \\ &= -\langle g^k, Av^{k-1} \rangle + \beta_k \langle v^{k-1}, Av^{k-1} \rangle \end{aligned}$$

skąd otrzymujemy

$$\beta_k = \frac{\langle g^k, Av^{k-1} \rangle}{\langle v^{k-1}, Av^{k-1} \rangle}, \quad k = 1, 2, \dots \quad (\text{VII-1.10})$$

Z równości (VII–1.8) mamy jak poprzednio $g^{k+1} = g^k + t_k Av^k$ skąd otrzymujemy, iż

$$\begin{aligned} \langle v^k, g^{k+1} \rangle &= \langle v^k, g^k + t_k Av^k \rangle = \langle v^k, g^k \rangle + t_k \langle v^k, Av^k \rangle = \\ &= \langle v^k, g^k \rangle - \frac{\langle v^k, g^k \rangle}{\langle v^k, Av^k \rangle} \langle v^k, Av^k \rangle = 0. \end{aligned} \quad (\text{VII-1.11})$$

Zachodzi więc

$$\langle v^{k-1}, g^k \rangle = 0, \quad \text{dla } k = 1, 2, \dots$$

a z (VII–1.9) otrzymujemy, że

$$\begin{aligned} \langle v^k, g^k \rangle &= \langle -g^k + \beta_k v^{k-1}, g^k \rangle = -\langle g^k, g^k \rangle + \beta_k \langle v^{k-1}, g^k \rangle = \\ &= -\|g^k\|_2^2 < 0. \end{aligned}$$

skąd wynika, że występująca we wzorze (VII–1.8) liczba t_k jest dodatnia.

VII–1.6 Twierdzenie

Jeżeli A jest symetryczną dodatnio określoną macierzą a ciągi $v^k, g^k \in \mathbb{R}^n$ są określone przez (VII–1.8), (VII–1.9) i (VII–1.10), to wtedy dla $k = 1, 2, \dots$ zachodzi

- (1) $\text{lin}\{g^0, g^1, \dots, g^{k-1}\} = \text{lin}\{v^0, v^1, \dots, v^{k-1}\},$
- (2) $\langle v^{k-1}, Av^j \rangle = \langle v^{k-1}, v^j \rangle_A = 0$ dla $j = 0, 1, \dots, k-2$ gdzie $k \geq 2,$
- (3) $\langle g^j, g^k \rangle = 0$ dla $j = 0, 1, \dots, k-1.$

Dowód:

Dowód będzie indukcyjny. Dla $k = 1$ $\text{lin}\{g^0\} = \text{lin}\{v^0\}$ bo z definicji $v^0 = -g^0$. Punkt **(2)** trzeba wykazać dopiero dla $k \geq 2$ a **(3)** wynika z $v^0 = -g^0$ i (VII-1.11).

Założmy, że dla $k \geq 1$ własności **(1)** — **(3)** z tezy twierdzenia są wykazane.

Wtedy, jeżeli $\text{lin}\{g^0, g^1, \dots, g^{k-1}\} = \text{lin}\{v^0, v^1, \dots, v^{k-1}\}$, to z równości (VII-1.9) wynika że też $\text{lin}\{g^0, g^1, \dots, g^k\} = \text{lin}\{v^0, v^1, \dots, v^k\}$.

Dowód własności **(2)** zaczniemy od uwagi, że liczba β_k została określona równością (VII-1.10) w ten sposób, by v^k z (VII-1.9) spełniało

$$\langle v^k, Av^{k-1} \rangle = 0.$$

Tak więc do sprawdzenia pozostało

$$\langle v^k, Av^j \rangle = 0, \quad j = 0, 1, \dots, k-2 \quad \text{przy} \quad k \geq 2.$$

Z określenia (VII-1.9) kierunku v^k mamy

$$\begin{aligned} \langle v^k, Av^j \rangle &= \\ &= \langle -g^k + \beta_k v^{k-1}, Av^j \rangle = -\langle g^k, Av^j \rangle + \beta_k \langle v^{k-1}, Av^j \rangle = \\ &= -\langle g^k, Av^j \rangle, \quad \text{dla} \quad j = 0, 1, \dots, k-2, \end{aligned}$$

gdzie $\langle v^{k-1}, Av^j \rangle = 0$, $j = 0, 1, \dots, k-2$ na podstawie założenia indukcyjnego. Z (VII-1.8) mamy, że $x^{j+1} = x^j + t_j v^j$, $j = 0, 1, \dots, k-2$ skąd otrzymujemy (VII-1.7) a więc równość

$$g^{j+1} - g^j = t_j Av^j \quad \text{z} \quad t_j > 0 \quad \text{gdzie} \quad j = 0, 1, \dots, k-2.$$

Kontynuując poprzednie przekształcenia otrzymujemy dla $j = 0, 2, \dots, k-2$

$$\begin{aligned} \langle v^k, Av^j \rangle &= \\ &= -\langle g^k, Av^j \rangle = -\frac{1}{t_j} \langle g^k, g^{j+1} - g^j \rangle = -\frac{1}{t_j} \left(\langle g^k, g^{j+1} \rangle - \langle g^k, g^j \rangle \right) = 0 \end{aligned}$$

z przesłanki indukcyjnej punktu **(3)** tezy.

Pozostał punkt **(3)** czyli mamy sprawdzić, że $\langle g^j, g^{k+1} \rangle = 0$, $j = 0, 1, \dots, k$. Dla tego celu, na podstawie udowodnionej własności **(1)** wystarczy wykazać $\langle v^j, g^{k+1} \rangle = 0$, $j = 0, 1, \dots, k$. Przypadek $j = k$ jest zawarty w (VII-1.11). Co więcej z przesłanki indukcyjnej punktów **(1)** i **(3)** wynika, że

$$\langle v^j, g^k \rangle = 0, \quad j = 0, 1, \dots, k-1.$$

Skąd wynika, że aby udowodnić

$$\langle v^j, g^{k+1} \rangle = 0, \quad j = 0, 1, \dots, k-1$$

wystarczy wykazać, iż

$$\langle v^j, g^{k+1} - g^k \rangle = 0, \quad j = 0, 1, \dots, k-1.$$

a ponieważ

$$g^{k+1} - g^k = t_k Av^k, \quad \text{z} \quad t_k > 0$$

to mamy

$$\begin{aligned}\langle v^j, g^{k+1} \rangle &= \\ &= \langle v^j, g^{k+1} - g^k \rangle = \frac{1}{t_k} \langle v^j, Av^k \rangle = \frac{1}{t_k} \langle v^k, Av^j \rangle = 0.\end{aligned}$$

Skorzystaliśmy tu z $A = A^T$ i udowodnionego już kroku indukcyjnego dla punktu (2) tezy. $\square$

VII-1.7 Wniosek

Dla dowolnego x^0 i ciągu określonego równościami (VII-1.8), (VII-1.9) i (VII-1.10) tylko skończenie wiele g^k , $k = 0, 1, \dots$ jest różnych od zera.

Dowód:

Wynika to z punktu (3) twierdzenia VII-1.6 i skończonego wymiaru przestrzeni $\mathbb{R}^n$. $\square$

Na podstawie twierdzenia VII-1.6 przeformułujemy algorytm sprzężonych gradientów: równania (VII-1.8), (VII-1.9) i (VII-1.10). Przyjmiemy tu $x^0 = 0$ i zamiast wektora g^k – gradientu, będziemy używać

$$r^k = b - Ax^k = -g^k \quad \text{– residuum.}$$

Ponadto mamy $g^k - g^{k-1} = t_{k-1} Av^{k-1}$ skąd wyrażenie (VII-1.10) przyjmuje postać

$$\begin{aligned}\beta_k &= \frac{\langle g^k, Av^{k-1} \rangle}{\langle v^{k-1}, Av^{k-1} \rangle} = \frac{\langle g^k, g^k - g^{k-1} \rangle}{\langle v^{k-1}, g^k - g^{k-1} \rangle} = \\ &= \frac{\langle g^k, g^k \rangle - \langle g^k, g^{k-1} \rangle}{\langle v^{k-1}, g^k \rangle - \langle v^{k-1}, g^{k-1} \rangle} = \frac{\|g^k\|_2^2}{\|g^{k-1}\|_2^2} = \frac{\|r^k\|_2^2}{\|r^{k-1}\|_2^2},\end{aligned}$$

gdzie skorzystaliśmy z twierdzenia VII-1.6 oraz równości $\langle v^{k-1}, g^{k-1} \rangle = -\|g^{k-1}\|_2^2$.

Teraz **algorytm sprzężonych gradientów** przybiera postać:

1. Dla $k = 0$ kładziemy $x^0 = 0$, $r^0 = b$, $v^0 = r^0$.
2. Jeśli $k > 0$, to $v^k = r^k + \beta_k v^{k-1}$, gdzie $\beta_k = \frac{\|r^k\|_2^2}{\|r^{k-1}\|_2^2}$.
3. Obliczamy: $d^k = Av^k$ i $t_k = \frac{\|r^k\|_2^2}{\langle v^k, d^k \rangle}$.
4. Wyznaczamy: $x^{k+1} = x^k + t_k v^k$ i $r^{k+1} = r^k - t_k d^k$.

VII-1.0.4. Zakończenie

Podstawa opracowania: [5], [2, t. II, ss. 338–344] i [8, ss. 159–162] a zwłaszcza [4].

VII-1.8 Definicja (podprzestrzeń Kryłowa)

Niech $A \in \mathbb{R}^{n \times n}$ będzie macierzą, $v \in \mathbb{R}^n$, $v \neq 0$ wektorem i $m \in \mathbb{N}$. Liniowa przestrzeń²

$$\mathcal{K}_m(A, v) = \text{lin}\{v, Av, A^2v, \dots, A^{m-1}v\}$$

jest podprzestrzenią Kryłowa rzędu m przestrzeni $\mathbb{R}^n$.

² $\text{lin} \equiv \text{span}$ – najmniejsza przestrzeń liniowa zawierająca (generowana, rozpięta przez) wektory ... [1, ss. 52]

VII-1.9 Lemat

Jeżeli $\delta_m = \dim \mathcal{K}_m(A, v)$, to wtedy $\delta_m \leq \delta_{m+1}$, $m = 1, 2, \dots$ oraz istnieje takie $k \in \mathbb{N}$, że $\delta_m = m$, dla $m = 1, \dots, k$ i $\delta_m = k$ dla $m = k+1, k+2, \dots$.

Założmy dodatkowo, że $v = \sum_{i=1}^{\hat{k}} c_i w_i$, $c_1, \dots, c_{\hat{k}} \neq 0$ gdzie $w_1, \dots, w_{\hat{k}} \neq 0$ są wektorami własnymi odpowiadającym różnym wartościom własnym $\lambda_1, \dots, \lambda_{\hat{k}}$, $\lambda_i \neq \lambda_j$ dla $i \neq j$ macierzy A , wtedy $k = \hat{k}$.

Dowód:

Ponieważ $\mathcal{K}_m(A, v) \subseteq \mathcal{K}_{m+1}(A, v)$ więc $\delta_m \leq \delta_{m+1}$ a z $\mathcal{K}_m(A, v) \subseteq \mathbb{R}^n$ mamy oszacowanie $\delta_m \leq n$. Z założenia $v \neq 0$ wynika, że $\delta_1 = \dim \mathcal{K}_1(A, v) = \dim \{v\} = 1$ i $\delta_m \leq m$ bo przestrzeń $\mathcal{K}_m(A, v)$ jest generowana przez m wektorów. Stąd dobrze jest określona liczba

$$k \stackrel{\text{def}}{=} \max\{m \in \mathbb{N} : \delta_m = m\}$$

Z definicji m wynika, że $\delta_m < m$ dla $m \geq k+1$ i w szczególności dla $m = k+1$ mamy

$$\delta_{k+1} = \delta_k = k$$

skąd wynika, że

$$\mathcal{K}_{k+1}(A, v) = \mathcal{K}_k(A, v).$$

Więc

$$A^k v \in \mathcal{K}_k(A, v).$$

Istnieją takie liczby $\theta_0, \dots, \theta_{k-1} \in \mathbb{R}$, że

$$A^k v = \sum_{j=0}^{k-1} \theta_j A^j v = \theta_0 v + \theta_1 A v + \theta_2 A^2 v + \dots + \theta_{k-1} A^{k-1} v$$

dalej dla $s = 0, 1, 2, \dots$ zachodzi

$$A^{k+s} v = \sum_{j=0}^{k-1} \theta_j A^{j+s} v = \theta_0 A^s v + \theta_1 A^{s+1} v + \theta_2 A^{s+2} v + \dots + \theta_{k-1} A^{k-1+s} v.$$

Z ostatniej równości poprzez proste rozumowanie indukcyjne względem s wynika, iż

$$A^{k+s} v \in \mathcal{K}_k(A, v) \quad \text{dla} \quad s = 0, 1, 2, \dots$$

Z definicji VII-1.8 przestrzeni Kryłowa otrzymujemy teraz, że

$$\mathcal{K}_m(A, v) = \mathcal{K}_k(A, v) \quad \text{dla} \quad m \geq k$$

skąd oczywiście wynika, że

$$\delta_m = \delta_k \quad \text{dla} \quad m \geq k.$$

Niech $v = \sum_{i=1}^{\hat{k}} c_i w_i$, gdzie w_i są wektorami własnymi macierzy A a λ_i odpowiadającymi tym wektorom różnymi wartościami własnymi. Wtedy

$$A^j v = \sum_{i=1}^{\hat{k}} c_i A^j w_i = \sum_{i=1}^{\hat{k}} c_i \lambda_i^j w_i \quad \text{dla} \quad j = 0, 1, 2, \dots \quad (\text{VII-1.12})$$

więc

$$\mathcal{K}_k(A, v) \subseteq \text{lin}\{w_1, \dots, w_{\hat{k}}\}$$

Przy poczynionych założeniach o wartościach własnych mamy

$$\dim \operatorname{lin}\{w_1, \dots, w_{\hat{k}}\} = \hat{k}.$$

Tak więc wykazaliśmy, że $k \leq \hat{k}$. Załóżmy teraz, że $k < \hat{k}$. Zostało już wykazane, iż $\delta_{\hat{k}} = \delta_k$ więc układ wektorów

$$v, Av, A^2v, \dots, A^{\hat{k}-1}v \quad (\text{VII-1.13})$$

jest liniowo zależny. Istnieją więc liczby $\alpha_0, \alpha_1, \dots, \alpha_{\hat{k}-1}$ nie wszystkie równe 0, dla których zachodzi

$$\sum_{j=0}^{\hat{k}-1} \alpha_j A^j v = \alpha_0 v + \alpha_1 Av + \alpha_2 A^2 v + \dots + \alpha_{\hat{k}-1} A^{\hat{k}-1} v = 0$$

Z wzoru (VII-1.12) otrzymujemy

$$\begin{aligned} 0 &= \sum_{j=0}^{\hat{k}-1} \alpha_j A^j v = \sum_{j=0}^{\hat{k}-1} \alpha_j A^j \sum_{i=1}^{\hat{k}} c_i w_i = \sum_{j=0}^{\hat{k}-1} \alpha_j \sum_{i=1}^{\hat{k}} c_i \lambda_i^j w_i = \\ &= \sum_{i=1}^{\hat{k}} c_i \left(\sum_{j=0}^{\hat{k}-1} \alpha_j \lambda_i^j \right) w_i. \end{aligned}$$

Ponieważ wektory $w_1, \dots, w_{\hat{k}}$ są liniowo niezależne i $c_1, c_2, \dots, c_{\hat{k}} \neq 0$, to

$$\sum_{j=0}^{\hat{k}-1} \alpha_j \lambda_i^j = 0, \quad \text{dla } i = 1, 2, \dots, \hat{k}.$$

Otrzymaliśmy układ równań postaci

$$\begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \dots & \lambda_1^{\hat{k}-1} \\ 1 & \lambda_2 & \lambda_2^2 & \dots & \lambda_2^{\hat{k}-1} \\ 1 & \lambda_3 & \lambda_3^2 & \dots & \lambda_3^{\hat{k}-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_{\hat{k}} & \lambda_{\hat{k}}^2 & \dots & \lambda_{\hat{k}}^{\hat{k}-1} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{\hat{k}-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

gdzie współczynniki tworzą macierz Vandermondea. Z założenia $\lambda_i \neq \lambda_j$, $i \neq j$ wynika, iż jedyne rozwiązanie tego układu jest $\alpha_0 = \alpha_1 = \dots = \alpha_{\hat{k}-1} = 0$. Co jest sprzeczne z liniową zależnością wektorów w (VII-1.13). W ten sposób wykazaliśmy, że $k = \hat{k}$

□

Pokażemy teraz, że:

VII-1.10 Lemat

Dla algorytmu sprzężonych gradientów (str. 8) zachodzi

$$r^k, v^k \in \mathcal{K}_{k+1}(A, b).$$

Dowód:

Dla pierwszych dwóch iteracji mamy

$$r^0 = b,$$

$$r^1 = r^0 - t_0 A v^0 = (I - t_0 A) b,$$

$$r^2 = r^1 - t_1 A v^1 = r^1 - t_1 A(r^1 + \beta_1 b) = [(I - t_1 A)(I - t_0 A) - t_1 \beta_1 A] b$$

więc $r^m \in \mathcal{K}_{m+1}(A, b)$ dla $m = 0, 1, 2$. Z punktu (1) twierdzenia VII-1.6 otrzymujemy, że także $v^m \in \mathcal{K}_{m+1}(A, b)$ dla $m = 0, 1, 2$. Dowodzimy dalej indukcyjnie. Jeżeli $r^m, v^m \in \mathcal{K}_{m+1}(A, b)$, to wtedy dla $m \leq j$ zachodzi $r^{j+1} = r^j - t_j A v^j$ więc i $r^{j+1} \in \mathcal{K}_{j+2}(A, b)$ oraz na podstawie punktu (1) twierdzenia VII-1.6 mamy $v^{j+1} \in \mathcal{K}_{j+2}(A, b)$. Z punktu (3) twierdzenia VII-1.6 mamy, że residua r^j są wzajemnie ortogonalne a więc liczba różnych od zera wektorów residualnych jest ograniczona z góry przez wymiar przestrzeni $\mathcal{K}_n(A, b)$. $\square$

VII-1.0.5. Uzupełnienie**VII-1.11 Lemat (CGS)**

Jeżeli macierz A jest symetryczna i dodatnio określona w przestrzeni $\mathbb{R}^n$, to w przestrzeni liniowej $\mathbb{R}^n$ istnieje baza A -ortonormalna unormowana.

Dowód³:

Forma dwuliniowa $\langle x, y \rangle_A = \langle x, Ay \rangle$ jest iloczynem skalarnym w $\mathbb{R}^n$. Niech wektory $u^1, \dots, u^n$ tworzą bazę w tej przestrzeni. Definiujemy teraz indukcyjnie układ wektorów $v^1, \dots, v^n$

$$v^1 = \frac{u^1}{\|u^1\|_A},$$

$$v^k = \frac{w^k}{\|w^k\|_A}, \text{ gdzie } w^k = u^k - \sum_{i=1}^{k-1} \langle u^k, v^i \rangle_A v^i, \text{ dla } k = 2, \dots, n.$$

Wektory $w^k, k = 2, \dots, n$ zostały wprowadzone dla uproszczenia zapisu. Wykażemy teraz, że indukcyjnie, że dla $k = 2, \dots, n$ zachodzi

$$\langle v^k, v^i \rangle_A = 0 \quad \text{gdzie } i = 1, \dots, k-1.$$

Założmy więc, że $k = 2$ – pierwszy krok indukcji. Wtedy mamy

$$\begin{aligned} \langle w^2, v^1 \rangle_A &= \\ &= \langle u^2 - \langle u^2, v^1 \rangle_A v^1, v^1 \rangle_A = \langle u^2, v^1 \rangle_A - \langle u^2, v^1 \rangle_A \langle v^1, v^1 \rangle_A = \\ &= \langle u^2, v^1 \rangle_A - \langle u^2, v^1 \rangle_A = 0 \end{aligned}$$

bo $\langle v^1, v^1 \rangle_A = 1$. Jasne jest teraz, iż $\langle v^2, v^1 \rangle_A = 0$. Niech teraz dla $1 < k < n, i = 1, \dots, k-1$ zachodzi

$$\langle v^k, v^i \rangle_A = 0.$$

³ CGS $\equiv$ Classical Gram-Schmidt

Otrzymujemy teraz dla $i = 1, 2, \dots, k$

$$\begin{aligned} \langle w^{k+1}, v^i \rangle_A &= \\ &= \langle u^{k+1} - \sum_{j=1}^k \langle u^{k+1}, v^j \rangle_A v^j, v^i \rangle_A = \langle u^{k+1}, v^i \rangle_A - \sum_{j=1}^k \langle u^{k+1}, v^j \rangle_A \langle v^j, v^i \rangle_A = \\ &= \langle u^{k+1}, v^i \rangle_A - \langle u^{k+1}, v^i \rangle_A = 0. \end{aligned}$$

A więc też⁴

$$\langle v^{k+1}, v^i \rangle_A = 0 \quad \text{dla } i = 1, 2, \dots, k.$$

□

VII–1.12 Lemat (MGS)

Jeżeli macierz A jest symetryczna i dodatnio określona w przestrzeni $\mathbb{R}^n$, to w przestrzeni liniowej $\mathbb{R}^n$ istnieje baza A -ortonormalna unormowana.

Dowód⁵:

Podobnie jak w lemacie VII–1.11, niech wektory $u^1, \dots, u^n$ tworzą bazę w przestrzeni $\mathbb{R}^n$. Definiujemy indukcyjnie układ wektorów $v^1, \dots, v^k$, $k = 1, \dots, n$ równocześnie w każdym kroku modyfikując pozostałe wektory z $u^{k+1}, \dots, u^n$, $k = 1, \dots, n-1$ w ten sposób, aby

$$\langle v^i, u^{j,k} \rangle_A = 0, \quad i = 1, \dots, k, \quad j = k+1, \dots, n, \quad k = 1, \dots, n-1.$$

Dla $k = 1$ przyjmujemy:

$$\begin{aligned} v^1 &= \frac{u^1}{\|u^1\|_A}, \\ u^{j,2} &= u^j - \langle u^j, v^1 \rangle_A v^1, \quad j = 2, \dots, n. \end{aligned} \tag{VII–1.14}$$

Oczywiście zachodzi tu, dla $j = 2, \dots, n$

$$\begin{aligned} \langle v^1, u^{j,2} \rangle_A &= \\ &= \langle v^1, u^j - \langle u^j, v^1 \rangle_A v^1 \rangle_A = \langle u^j, v^1 \rangle_A - \langle u^j, v^1 \rangle_A = 0. \end{aligned}$$

Dla $k = 2, \dots, n$ definiujemy indukcyjnie wektory:

$$\begin{aligned} v^k &= \frac{u^{k,k}}{\|u^{k,k}\|_A}, \\ u^{j,k+1} &= u^{j,k} - \langle u^{j,k}, v^k \rangle_A v^k, \quad j = k+1, \dots, n. \end{aligned} \tag{VII–1.15}$$

Sprawdzamy warunki ortogonalności

$$\begin{aligned} \langle v^k, u^{j,k+1} \rangle_A &= \\ &= \langle v^k, u^{j,k} - \langle u^{j,k}, v^k \rangle_A v^k \rangle_A = \langle u^{j,k}, v^k \rangle_A - \langle u^{j,k}, v^k \rangle_A = 0, \end{aligned}$$

dla $j = k+1, \dots, n$ oraz dla $i = 1, \dots, k-1$, $j = k+1, \dots, n$

$$\begin{aligned} \langle v^i, u^{j,k+1} \rangle_A &= \\ &= \langle v^i, u^{j,k} - \langle u^{j,k}, v^k \rangle_A v^k \rangle_A = \langle v^i, u^{j,k} \rangle_A - \langle u^{j,k}, v^k \rangle_A \langle v^i, v^k \rangle_A = 0. \end{aligned}$$

□

⁴ Ponieważ $u^1, \dots, u^n$ jest bazą, to $u^k \neq 0$, $k = 1, 2, \dots, n$.

⁵ MGS $\equiv$ Modified Gram–Schmidt

VII-1.13 Twierdzenie

Niech ([6, T. 1, ss. 251]) układ $v^1, \dots, v^n$ będzie A -ortogonalną unormowaną bazą przestrzeni liniowej $\mathbb{R}^n$. Dla dowolnego $x^0 \in \mathbb{R}^n$ określamy rekurencyjnie ciąg

$$x^i = x^{i-1} + \langle b - Ax^{i-1}, v^i \rangle v^i, \quad i = 1, 2, \dots, n. \quad (\text{VII-1.16})$$

Wtedy

$$Ax^n = b.$$

Dowód:

Wprowadzamy następujące oznaczenie

$$t_i = \langle b - Ax^{i-1}, v^i \rangle, \quad i = 1, 2, \dots, n.$$

Wtedy wzór rekurencyjny (VII-1.16) przyjmuje następującą postać

$$x^i = x^{i-1} + t_i v^i.$$

Skąd mamy

$$Ax^i = Ax^{i-1} + t_i Av^i, \quad i = 1, \dots, n$$

$$Ax^n = Ax^0 + t_1 Av^1 + \dots + t_n Av^n.$$

Teraz od obu stron ostatniej równości odejmujemy wektor b i przemnażamy skalarnie przez v^i , $i = 1, \dots, n$ aby otrzymać

$$\begin{aligned} \langle Ax^n - b, v^i \rangle &= \\ &= \langle Ax^0 - b, v^i \rangle + t_1 \langle Av^1, v^i \rangle + \dots + t_n \langle Av^n, v^i \rangle = \\ &= \langle Ax^0 - b, v^i \rangle + t_1 \langle v^1, v^i \rangle_A + \dots + t_n \langle v^n, v^i \rangle_A = \\ &= \langle Ax^0 - b, v^i \rangle + t_i. \end{aligned}$$

Dalej mamy, że

$$\begin{aligned} t_i &= \langle b - Ax^{i-1}, v^i \rangle = \\ &= \langle b - Ax^0, v^i \rangle + \langle Ax^0 - Ax^1, v^i \rangle + \dots + \langle Ax^{i-2} - Ax^{i-1}, v^i \rangle = \\ &= \langle b - Ax^0, v^i \rangle + \langle -t_1 Av^1, v^i \rangle + \dots + \langle -t_{i-1} Av^{i-1}, v^i \rangle = \\ &= \langle b - Ax^0, v^i \rangle - t_1 \langle v^1, v^i \rangle_A + \dots - t_{i-1} \langle v^{i-1}, v^i \rangle_A = \langle b - Ax^0, v^i \rangle. \end{aligned}$$

Pokazaliśmy w ten sposób, że

$$\langle Ax^n - b, v^i \rangle = 0 \quad \text{dla} \quad i = 1, 2, \dots, n.$$

Tak więc, dla dowolnego $w \in \mathbb{R}^n$ zachodzi $\langle Ax^n - b, w \rangle = 0$ co jest możliwe jedynie wtedy, gdy

$$Ax^n - b = 0.$$

□

VII–1.0.6. Zastosowanie

Rozważania tego wykładu porównamy z wynikami jakie osiągnęliśmy na poprzednim dla układu

$$\begin{bmatrix} 4 & 3 & 0 \\ 3 & 4 & -1 \\ 0 & -1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 24 \\ 30 \\ -24 \end{bmatrix}$$

MatLabowski skrypt dla metody sprzężonych gradientów (CG):

```
% Metoda sprzezonych gradientow
function [x,e] = CG(A,b,eps,xd,It)
[n,m] = size(A);
e      = zeros(It,1);
d      = sqrt(xd'*xd);
% Ustalanie warunkow poczatkowych
x      = zeros(n,1);
r      = b;
v      = r;
Nr1    = r'*r; % kwadrat normy wektora resztowego
k      = 0;    % licznik iteracji
% while (sqrt(Nr1)>eps) & (k<It) % tu jest kryterium zatrzymania
while (k<It)
    if 0<k
        beta = Nr1/Nr0;
        v = r + beta*v;
    end;
    d = A*v;
    t = Nr1/(v'*d);
    x = x+t*v;
    r = r-t*d;
    Nr0 = Nr1;
    Nr1 = r'*r;
    k = k+1;
    e(k) = sqrt((x-xd)*(x-xd));
end;
```

Główny skrypt

```
clear all;
clc;
A      = [4 3 0; 3 4 -1; 0 -1 4];
b      = [24; 30; -24];
xd     = A\b;
a      = 1;
It     = 10;
[x1,e1] = SOR(A,b,1.25,xd,It);
[x2,e2] = CG(A,b,0.0001,xd,It);
plot([a:It],e1(a:It),'k-.',[a:It],e2(a:It),'k:');
axis([a It -0.01 0.8]);
legend('SOR','CG',1);
```

Na Rys. VII–1 widać, że w przypadku metody sprzężonych gradientów dokładne rozwiązanie zostało osiągnięte dla $k=3$ iteracji podczas gdy dla metody SOR z $\omega = 1.25$ trzeba było wykonać $k = 6$ iteracji.

Rysunek VII–1. Wykresy błędów względnych dla $k = 1, \dots, 10$.

Literatura

- [1] A. Białynicki-Birula, Algebra liniowa z geometrią, PWN, Warszawa, 1976.
- [2] A. Björck and G. Dahlquist, Numerical mathematics and scientific computation, SIAM, Philadelphia, 2003,
Wersję roboczą polecił A. Szustalewicz. Jest ona dostępna pod adresem <http://www.mai.liu.se/~akbjjo/NMbook.html>.
- [3] R.L. Burden and Faires, Numerical methods, 3, Brooks-Cole, Florence, 2003,
Dostępna na stronie WWW autorów, 7.XII.03r.
- [4] I.C.F. Ipsen and C.D. Meyer, The idea behind Krylow methods, AMM **105** (1998), no. 10, 889–899.
- [5] A. Iserles and A. Shadrin, Numerical analysis, 7.XII.03r: dostępna pod adresem <http://www.damtp.cam.ac.uk/user/na/na.html>, 2003,
Lecture notes for the course, DAMTP, University of Cambridge.
- [6] D. Kincaid and W. Cheney, Numerical analysis, Brooks, USA, 1996,
Książka jest w bibliotece Instytutu Informatyki Uniwersytetu Wrocławskiego.
- [7] K. Moszyński, Metody numeryczne dla informatyków, 27.XII.03r: dostępna pod adresem: <http://www.mimuw.edu.pl/~kmoszyns/>, 2003/2004.
- [8] A. Quarteroni, R. Sacco, and F. Saleri, Numerical mathematics, Springer, 2000.
- [9] A. Ralston, Wstęp do analizy numerycznej, PWN, Warszawa, 1983,
Tłumaczył R. Zuber, wydanie polskie opracował St. Paszkowski. Do wydania polskiego jest dedykacja autora.
- [10] J. Stoer, Wstęp do metod numerycznych, PWN, Warszawa, 1979,
Tłumaczyła E. Gurbiel.

Wrocław, dnia 10 grudnia 2010

(–) Krzysztof Tabisz
tabisz@math.uni.wroc.pl