2.5 Metoda sdružených gradientů jako Krylovovská metoda

Tato metoda byla vyvinuta v 50. letech 20. století jako přímá metoda pro řešení soustav lineárních rovnic s SPD maticí. Jako iterační metoda je používána až od 70. let. Algoritmus metody sdružených gradientů (zkráceně CG z angl. conjugate gradient) můžeme chápat jako rozvinutí algoritmu metody největšího spádu, viz další část. Efektivita metody sdružených gradientů souvisí s plným využitím tzv. Krylovova prostoru.

Definice. Prostor $\mathcal{K}_m(A,v) = \mathrm{Lin}\left(v, Av, \ldots, A^{m-1}v\right)$ je tzv. Krylovův podprostor (pro dané $m \in \mathbb{N}$, matici A a vektor v). Jedná se o podprostor $\mathbb{R}^n$, kde n je řád matice A. Dimenze $\mathcal{K}_m(A,v)$ je nejvýše rovna menšímu z čísel m a n.

Věta 2.3. *V případě Richarsonovy metody nebo metody největšího spádu platí*

$$u_{i+1} = u_i + \omega_i \cdot r_i \;\; a \;\; r_{i+1} = r_i - \omega_i \cdot A r_i.$$

Odtud plyne indukcí, že pro $i \geq 0$ je

$$r_{i+1} \in \mathcal{K}_{i+2}(r_0, A) \;\; a \;\; u_{i+1} = u_0 + \sum_{k=0}^{i} \omega_k \cdot r_k \in u_0 + \mathcal{K}_{i+1}(r_0, A).$$

Iterace Richarsonovy metody nebo metody největšího spádu tedy patří do množiny určené Krylovým podprostorem, ale nevyužívají jej optimálně, tj. neplatí

$$\|u_{i+1} - u^\star\|_{(m)} = \min\left\{\|w - u^\star\|_{(m)} : \;\; w \in u_0 + \mathcal{K}_{i+1}(r_0, A)\right\} \tag{7}$$

v některé z norem $\|\cdot\|_{(m)}$, $m = 0, 1, 2$. Plné využití Krylovova prostoru znamená vznik efektivnější metody, u které je konvergence určena konvergenčním faktorem typu (5).

Věta 2.4. *Uvažujme iterace $u_{i+i} \in u_0 + \mathcal{K}_{i+1}(r_0, A)$. Potom*

$$e_{i+1} \in e_0 + \mathcal{K}_{i+1}(r_0, A) = \mathcal{K}_{i+2}(e_0, A), \;\; e_{i+1} = P_{i+1}(A) e_0.$$

Podmínka optimality (7) a srovnání s Čebyševovým polynomem (5), (6) implikují

$$\begin{aligned} \|e_{i+1}\|_{(m)} &= \min_{P_{i+1}:\, P_{i+1}(0)=1} \|P_{i+1}(A) e_0\|_{(m)} \leq \min_{P_{i+1}:\, P_{i+1}(0)=1} \max_{\lambda \in \sigma(A)} |P_{i+1}(\lambda)| \, \|e_0\|_{(m)} \\ &\leq \min_{P_{i+1}:\, P_{i+1}(0)=1} \max_{\lambda \in \sigma(A)} \left|P^C_{i+1}(\lambda)\right| \|e_0\|_{(m)} \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{i+1}. \end{aligned}$$

Důkaz. Důkaz obsahuje zřejmé, ale důležité kroky, proto je připomeneme:
I. Chyba po $(i+1)$-ní iteraci

$$\begin{aligned} e_{i+1} &= u_{i+1} - u^* = u_i - u^* + \omega_i \cdot (-A e_i) = e_i - \omega_i A e_i = \\ &= (I - \omega_i A) \cdot e_i = \prod_{k=0}^{i} (I - \omega_k A)\, e_0 = P_{i+1}(A) \cdot e_0, \end{aligned}$$

kde $P_{i+1}(z) = \prod_{k=0}^{i} (I - \omega_k z)$ je polynom stupně $i+1$ s absolutním členem 1, tj. $P_{i+1}(0) = 1$. Pokud $e_0 = \sum_{s=1}^{n} \alpha_s v_s$ je vyjádření počáteční chyby v bázi tvořené vlastními vektory A, potom

$$e_{i+1} = P_{i+1}(A) \cdot \sum_{s=1}^{n} \alpha_s v_s = \sum_{s=1}^{n} \alpha_s P_{i+1}(A)\, v_s = \sum_{s=1}^{n} \alpha_s P_{i+1}(\lambda_s)\, v_s$$

II. $\|\cdot\|_{(m)}$ norma chyby. Nechť A je symetrická, tedy existuje ortonormální systém vlastních vektorů $\{v_s\}$, $e_0 = \sum_s \alpha_s v_s$. Potom

$$\begin{aligned} \|e_{i+1}\|_{(m)} &= \left\| \sum_s \alpha_s P_{i+1}(\lambda_s)\, v_s \right\|_{(m)} \\ \|e_{i+1}\|^2_{(m)} &= \left(\sum_s \alpha_s P_{i+1}(\lambda_s)\, v_s, \sum_s \alpha_s P_{i+1}(\lambda_s)\, v_s \right)_{(m)} \\ &= \sum_s \alpha_s^2 \lambda_s^m P^2_{i+1}(\lambda_s) \leq \left(\max_{\lambda_s \in \sigma(A)} P^2_{i+1}(\lambda_s) \right) \sum_s \alpha_s^2 \lambda_s^m \\ \|e_{i+1}\|_{(m)} &\leq \max_{\lambda_s \in \sigma(A)} |P_{i+1}(\lambda_s)| \cdot \|e_0\|_{(m)} \end{aligned}$$

III. Naše metoda je optimální, tedy chyba iterace je menší, než jakou bychom dostali při dosazení normovaného Čebyševova polynomu. A pro Čebyševův polynom umíme najít odhad konvergenčního faktoru, viz Sekce 2.3.1. □

Poznámka 2.4. Máme odhady

$$\|e_{i+1}\|_{(m)} \leq \xi \left(\frac{c-1}{c+1}\right)^{i+1} \|e_0\|_{(m)},$$

kde $\xi = 1$ a $c = cond(A)$ u Richardsonovy metody a metody největšího spádu, $\xi = 2$ a $c = \sqrt{cond(A)}$ u optimální Krylovovské metody.

Poznámka 2.5. Kolik iterací je potřeba pro dosažení dané přesnosti ε?

$$\begin{aligned} \|e_{i+1}\|_{(m)} &\leq \varepsilon \|e_0\|_{(m)} \\ \xi \left(\frac{c-1}{c+1}\right)^{i+1} &\leq \varepsilon \end{aligned}$$

$$(i+1) \ln\left(\frac{c-1}{c+1}\right) \leq \ln(\varepsilon/\xi) \Rightarrow (i+1) \geq -\ln(\varepsilon/\xi) \left[-\ln\left(\frac{c-1}{c+1}\right)\right]^{-1}.$$

Protože

$$-\ln\left(\frac{c-1}{c+1}\right) = \ln\left(\frac{c+1}{c-1}\right) = \ln\left(\frac{1+1/c}{1-1/c}\right) = 2\left(\frac{1}{c} + \frac{1}{3}\left(\frac{1}{c}\right)^3 + \cdots\right) \geq 2 \cdot \frac{1}{c},$$

platí

$$i + 1 \geq \frac{1}{2} \cdot c \cdot \ln(\xi/\varepsilon).$$

Počet iterací tedy roste s $c = cond(A)$ u Richardsonovy metody a metody největšího spádu (respektive pomaleji), s $c = \sqrt{cond(A)}$ u Krylovovské metody.

Zamysleme se dále nad realizací podmínky optimality (7).

Věta 2.5. *Pokud $\{v_i\}$ je báze prostoru řešení, potom je podmínka optimality*

$$\left\| u_0 + \sum \alpha_k v_k - u^\star \right\|^2_{(m)} = \|u_0 - u^\star + V\underline{\alpha}\|^2_{(m)} \to \min,$$

ekvivalentní řešení soustavy

$$G\underline{\alpha} = \underline{\xi}$$

kde V je matice se sloupci tvořenými vektory v_r, $G = V^T A^m V = (G_{rs}) = \big((v_r, v_s)_{(m)}\big)$ je Grammova matice a $\underline{\xi} = -V^T A^m (u_0 - u^\star) = (\xi_r) = \big((-u_0 + u^\star, v_r)_{(m)}\big)$. Pravou stranu $\underline{\xi}$ lze určit pokud $m \geq 1$ a násobení A převede neznámou chybu $e_0 = u_0 - u^\star$ na residuum.

Poznámka 2.6. Dobré je pracovat s bází, ve které je G dobře podmíněná, nejlépe bází ortogonální. Báze tvořená vektory $A^i r_0$ dobře podmíněná není, viz mocninná metoda pro výpočet vlastních čísel a vlastních vektorů.

V další části ukážeme, že metoda sdružených gradientů je Krylovovská metoda s podmínkou optimality v normě $\|\cdot\|_{(1)} = \|\cdot\|_A$. Z toho pak plyne, že pro ni bude platit konvergenční odhad $c = \sqrt{cond(A)}$. Půjdeme cestou zobecnění metody největšího spádu, využitím konstrukce A-ortogonální báze Krylovova prostoru, která nám ukáže, jak podmínku optimality realizovat.