2.6 Metoda sdružených gradientů jako zobecnění metody největšího spádu

Algoritmus metody sdružených gradientů můžeme také chápat jako rozvinutí algoritmu metody největšího spádu. Iterační předpis má opět tvar

$$u_{i+i} = u_i + \omega_i \cdot d_i,$$

ale vektor d_i již (na rozdíl od metody největšího spádu) nemusí být roven reziduu, ale je vybrán jako lineární kombinace residua a předchozího směru, přičemž tato lineární kombinace je A-ortogonální k předchozímu směru.

Volba koeficientu ω_i. Pokud uvažujeme obecný směr d_i a ne nutně $d_i = r_i$, bude optimální koeficient v A-normě obdobou koeficientu, který se vyskytuje v metodě největšího spádu.

$$\begin{aligned} \varphi(\omega) &= \|u_{i+1} - u^*\|_A^2 = \|e_i + \omega d_i\|_A^2 = (Ae_i, e_i) + 2\omega (Ae_i, d_i) + \omega^2 (Ad_i, d_i) \\ \varphi'(\omega) &= 2(Ae_i, d_i) + 2\omega (Ad_i, d_i) = 0 \Rightarrow \end{aligned}$$

$$\omega_i = -\frac{(Ae_i,\, d_i)}{(Ad_i,\, d_i)} = \frac{(r_i,\, d_i)}{(Ad_i,\, d_i)}$$

Volba směru d_i**:** Kromě volby $d_i = r_i$ vedoucí k metodě největšího spádu máme mnoho dalších možností, např. cyklicky používat základní směrové vektory $d_i = (0, \dots, 0, 1, 0, \dots, 0)$, což povede k Jacobiho metodě $\omega_i = \frac{r_i}{a_{ii}}$. V případě metody sdružených gradientů vyjdeme z rezidua, ale budeme jej ortogonalizovat k předchozím směrům. Přesněji, provedeme ortogonalizalizaci jen k jednomu předchozímu směru (a později uvidíme, že ortogonalita k dalším předchozím směrům se již objeví sama). Tedy volíme

$$\begin{aligned} d_0 &= r_0, \\ d_{i+1} &= r_{i+1} + \beta_i d_i, \;\; (d_{i+1},\, d_i)_A = 0. \end{aligned}$$

$$(d_{i+1},\, Ad_i) = (r_{i+1} + \beta_i d_i,\, Ad_i) = (r_{i+1},\, Ad_i) + \beta_i (Ad_i,\, d_i) = 0$$

$$\Rightarrow \beta_i = -\frac{(r_{i+1},\, Ad_i)}{(Ad_i,\, d_i)}.$$

U metody sdružených gradientů se standardně značí α_i místo ω_i. Metoda sdružených gradientů jako zobecnění metody největšího spádu má při tomto značení tvar Algoritmu 4.

Věta 2.6. *V metodě sdružených gradientů platí*

$$\alpha_i = \frac{(r_i,\, d_i)}{(Ad_i,\, d_i)} = \frac{(r_i,\, r_i)}{(Ad_i,\, d_i)} \tag{8}$$

$$\beta_i = -\frac{(r_{i+1},\, Ad_i)}{(Ad_i,\, d_i)} = \frac{(r_{i+1},\, r_{i+1})}{(r_i,\, r_i)} \tag{9}$$

Důkaz. Platí

$(r_i,\, d_i) = (r_i,\, r_i + \beta_{i-1} d_{i-1}) = (r_i,\, r_i) + \beta_{i-1}(r_i,\, d_{i-1}) = (r_i,\, r_i)$,
protože $(r_i,\, d_{i-1}) = (r_{i-1} - \alpha_{i-1} Ad_{i-1},\, d_{i-1}) = (r_{i-1},\, d_{i-1}) - \alpha_{i-1}(Ad_{i-1},\, d_{i-1}) = 0$

$\beta_i = -\frac{(r_{i+1},\, Ad_i)}{(Ad_i,\, d_i)} = -\frac{\left(r_{i+1},\, \frac{1}{\omega_i}(r_i - r_{i+1})\right)}{(Ad_i,\, d_i)} = -\frac{(r_{i+1},\, r_i) - (r_{i+1},\, r_{i+1})}{(r_i,\, r_i)} = \frac{(r_{i+1},\, r_{i+1})}{(r_i,\, r_i)}$,
protože $(r_{i+1},\, r_i) = (r_i,\, r_i) - \omega_i (Ad_i,\, r_i) = (r_i,\, r_i) - \omega_i (Ad_i,\, d_i - \beta_{i-1} d_{i-1}) =$
$= (r_i,\, r_i) - \frac{(r_i,\, d_i)}{(Ad_i,\, d_i)} \left((Ad_i,\, d_i) - \beta_{i-1}(Ad_i,\, d_{i-1})\right) = (r_i,\, r_i) - (r_i,\, d_i) = 0$ □

Větu využijeme pro vytvoření efektivnější varianty algoritmu metody CG.

Algoritmus 4 (Metoda sdružených gradientů, 3 skalární součiny)

$u_0 \longleftarrow$	počáteční odhad
$r_0 = b - A \cdot u_0$	
$d_0 = r_0$	
for $i = 0, 1, \dots$	
$\quad v_i = A \cdot d_i$	
$\quad \alpha_i = \frac{(r_i,\, d_i)}{(v_i,\, d_i)}$	
$\quad u_{i+i} = u_i + \alpha_i \cdot d_i$	minimalizace ve směru d_i
$\quad r_{i+1} = r_i - \alpha_i \cdot v_i$	
$\quad \beta_i = -\frac{(r_{i+1}, Ad_i)}{(v_i, d_i)}$	
$\quad d_{i+1} = r_{i+1} + \beta_i d_i$	ortogonalizace rezidua
end	

Algoritmus 5 (Metoda sdružených gradientů, pouze 2 skalární součiny s využitím Věty 2.6)

$u_0 \longleftarrow$
$r_0 = b - A \cdot u_0$
$d_0 = r_0$
$\rho_0 = (r_0,\, r_0)$
`for` $i = 0, 1, \dots$
$\quad w_i = A \cdot d_i$
$\quad \alpha_i = \frac{\rho_i}{(w_i, d_i)}$
$\quad u_{i+i} = u_i + \alpha_i \cdot d_i$
$\quad r_{i+1} = r_i - \alpha_i \cdot w_i$
$\quad \rho_{i+1} = (r_{i+1},\, r_{i+1})$
`if` $\rho_{i+1} \le \varepsilon^2 \cdot \rho_0$ `then STOP` `%možná ukončovací podmínka`
$\quad \beta_i = \frac{\rho_{i+1}}{\rho_i}$
$\quad d_{i+1} = r_{i+1} + \beta_i d_i$
`end`

Vedle A-ortogonality směru d_{i+1} vůči předchozímu směru existují v metodě CG i další (skryté) ortogonality. Samozřejmě předpokládáme dokonalou přesnost výpočtů, neuvažujeme nepřesnou aritmetiku.

Věta 2.7. *Nechť (d_i) a (r_i) jsou posloupnosti směrů a reziduí generované metodou sdružených gradientů. Potom $\forall i \geq 0,\ \forall j < i$ platí*

$$\begin{aligned}(r_i,\, d_j) &= 0,\\ (r_i,\, r_j) &= 0,\\ (Ad_i,\, d_j) &= 0.\end{aligned}$$

Důkaz (matemetickou indukcí):

1. $\underline{i = 1:}$

$$\begin{aligned}\alpha_0 = \frac{(r_0,\, d_0)}{(Ad_0,\, d_0)} \Rightarrow (r_1,\, d_0) &= (r_0,\, d_0) - \alpha_0\,(Ad_0,\, d_0) = (r_0 - \alpha_0 Ad_0,\, d_0) = 0\\ (r_1,\, r_0) &= (r_1,\, d_0) = 0\\ (Ad_1,\, d_0) &= 0\end{aligned}$$

2. $\underline{i > 1:}$

$$\begin{aligned}(r_{i+1},\, d_j) &= \begin{cases} j = i\ldots & (r_i - \alpha_i Ad_i,\, d_i) = (r_i,\, d_i) - \alpha_i\,(Ad_i,\, d_i) = 0\\ j < i\ldots & (r_i - \alpha_i Ad_i,\, d_j) = (r_i,\, d_j) - \alpha_i\,(Ad_i,\, d_j) = 0\end{cases}\\ (r_{i+1},\, r_j) &= (r_{i+1},\, d_j - \beta_{j-1} d_{j-1}) = (r_{i+1},\, d_j) - \beta_{j-1}\,(r_{j+1},\, d_{j-1}) = 0\\ (Ad_{i+1},\, d_j) &= (d_{i+1},\, Ad_j) = \left(d_{i+1},\, \frac{1}{\alpha_j}\,(r_j - r_{j+1})\right) =\\ &= \left(r_{i+1} + \beta_i d_i,\, \frac{1}{\alpha_j}\,(r_j - r_{j+1})\right) =\\ &= \left(r_{i+1},\, \frac{1}{\alpha_j}\,(r_j - r_{j+1})\right) + \beta_i \left(d_i,\, \frac{1}{\alpha_j}\,(r_j - r_{j+1})\right) =\\ &= \left(r_{i+1},\, \frac{1}{\alpha_j}\,(r_j - r_{j+1})\right) + \beta_i\,(d_i,\, Ad_j) = 0 \text{ pro j<i.}\end{aligned}$$

Pro $j = i$ zřejmé. □

Lemma 2.1. *Pro posloupnosti iterací $u_0, u_1, \ldots$, reziduí $r_0, r_1, \ldots$ a směrů $d_0, d_1, \ldots$ generované metodou sdružených gradientů platí, že residua i směry generují Krylovův prostor, přesněji*

$$Lin\,\{d_0, \ldots, d_i\} = Lin\,\{r_0, \ldots, r_i\} = Lin\,\left\{r_0, \ldots,\, A^i r_0\right\} = \mathcal{K}_{i+1}\,(r_0, A)\,. \tag{10}$$

V důsledku toho

$$u_{i+1} = u_0 + \sum_{k=0}^{i} \alpha_k d_k \in u_0 + Lin\,\{d_0, \ldots, d_i\} = u_0 + \mathcal{K}_{i+1}\,(r_0, A)\,. \tag{11}$$

Důkaz. Lze provést matematickou indukcí. Pro $i = 0$ rovnosti (10) platí.

Pokud platí $\mathrm{Lin}\,\{d_0, \ldots, d_i\} = \mathrm{Lin}\,\{r_0, \ldots, r_i\}$, potom $d_{i+1} = r_{i+1} + \beta_i d_i$, takže $d_{i+1} \in \mathrm{Lin}\,\{r_0, \ldots, r_{i+1}\}$ a $\mathrm{Lin}\,\{d_0, \ldots, d_i,\, d_{i+1}\} \subset \mathrm{Lin}\,\{r_0, \ldots, r_i,\, r_{i+1}\}$. Obdobně plyne opačná inkluze.

Pokud platí $\mathrm{Lin}\,\{r_0, \ldots, r_i\} = \mathrm{Lin}\left\{r_0,\, Ar_0, \ldots,\, A^i r_0\right\} = \mathcal{K}_{i+1}\,(r_0, A)$, potom $r_{i+1} = r_i - \alpha_i A d_i \in r_i + A\mathcal{K}_{i+1}\,(r_0, A) \in \mathcal{K}_{i+2}\,(r_0, A)$. Opačně $A^i r_0 \in A\mathcal{K}_{i+1}\,(r_0, A) = A\mathrm{Lin}\,\{r_0, \ldots, r_i\} = A\mathrm{Lin}\,\{d_0, \ldots, d_i\}$, přičemž $\alpha_i A d_i = r_i - r_{i+1}$. Tedy $\mathcal{K}_{i+2}\,(r_0, A) \subset \mathrm{Lin}\,\{r_0, \ldots, r_{i+1}\}$. □

Věta 2.8. *Metoda sdružených gradientů je Krylovovská metoda vzhledem k A-normě, tedy*

$$u_{i+i} \in u_0 + \mathcal{K}_{i+1}(r_0, A) \tag{12}$$

a

$$\|u_{i+1} - u^\star\|_A = \min\,\{\|w - u^\star\|_A :\ \ w \in u_0 + \mathcal{K}_{i+1}(r_0, A)\}\,. \tag{13}$$

Důkaz. Z Lemma (2.1) plyne (12) a z definice máme

$$u_{i+1} = u_0 + \sum_{k=0}^{i} \alpha_k d_k,\ \ \alpha_k = \frac{(r_k,\, d_k)}{(Ad_k,\, d_k)}.$$

$\mathrm{Lin}\,\{d_0, \ldots, d_i\} = \mathcal{K}_{i+1}\,(r_0, A)$ a uvažujme bázi $\{d_k\}$. Podmínka Krylovovské ortogonality je potom ekvivalentní Grammovské soustavě, viz Věta 2.5, která říká, že v případě optimality (13) musí být optimální koeficienty

$$\alpha_k^{opt} = \frac{(r_0,\, d_i)}{(Ad,\, d_i)}.$$

Ale z rekurentního vztahu pro reziduum $r_k = r_{k-1} - \alpha_k \cdot Ad_k$ plyne

$$r_k = r_0 - \sum_{j=0}^{k-1} \alpha_j A d_j,$$

proto

$$(r_k,\, d_k) = \left(r_0 - \sum_{j=0}^{k-1} \alpha_k A d_j,\, d_k\right) = (r_0,\, d_k)\,.$$

Je tedy $\alpha_k = \alpha_k^{opt}$ a platí podmínka optimality (13). □

2.7 Konvergence metody sdružených gradientů

Metoda sdružených gradientů dává v A-normě nejlepší aproximaci v Krylovově prostoru. To implikuje platnost Čebyševovského odhadu

$$\|e_{i+1}\|_A \leq \xi \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^{i+1} \|e_i\|_A\,,$$

kde $\kappa = cond(A)$.

Poznámka 2.7. Polynom generovaný CG je vzhledem k optimalitě lepší než čebyševovský, ten používáme jen jako odhad. Ve srovnání s čebyševovskou metodou, metoda CG nepotřebuje odhady spektra a dokáže reagovat na distribuci spektra, např. shluky vlastních čísel. Čebyševova metoda ale nepoužívá skalární součiny, v tom je vhodnější pro paralelizaci.

Poznámka 2.8. Pokud se Krylovův prostor přidáním další mocniny $A^i r_0$ přestane rozšiřovat, stane se invariantním vůči A a tedy i A^{-1}. Protože obsahuje r_0 musí pak obsahovat i chybu e_0 a podmínka optimality zaručí, že Krylovovská metoda dojde k přesnému řešení. Při přesné aritmetice tedy dostaneme přesné řešení po $k \leq n$ iteracích, kde n je dimenze řešené soustavy. Prakticky toto jednak nenastane, jednak chceme soustavu vyřešit po $k \ll n$ iteracích.

2.8 Reprezentace matice v bázích Krylovova prostoru

Připomeňme, že pokud $\{v_i\}$ je báze V, $\{w_i\}$ je báze W, potom operátor

$$A : V \to W$$

reprezentujeme v daných bázích maticí (a_{ij}), kde

$$Av_i = \sum_j a_{ij} w_j.$$

My uvažujeme řešení soustavy s maticí A metodou CG. Matice A reprezentuje operátor $\mathcal{K}_k(r_0, A) \to \mathcal{K}_{k+1}(r_0, A)$.

Uvažujme nyní ortogonální bázi tvořenou normovanými rezidui, $\{v_i\} = \{\bar{r}_0, \ldots, \bar{r}_{k-1}\}$ a $\{w_i\} = \{\bar{r}_0, \ldots, \bar{r}_k\}$, $\bar{r}_i = \frac{r_i}{\|r_i\|}$. Chceme vyjádřit

$$A\bar{r}_i = \sum a_{ij} \bar{r}_j$$

Platí

$$\begin{aligned} Ad_i &= \frac{1}{\alpha_i}(r_i - r_{i+1}) \\ Ad_i &= A(r_i + \beta_{i-1} d_{i-1}) = Ar_i + \beta_{i-1} Ad_{i-1} = Ar_i + \beta_{i-1} \frac{1}{\alpha_{i-1}}(r_{i-1} - r_i) \end{aligned}$$

Z rovnosti

$$\frac{1}{\alpha_i}(r_i - r_{i+1}) = Ar_i + \beta_{i-1} \frac{1}{\alpha_{i-1}}(r_{i-1} - r_i)$$

pak vyjádříme

$$Ar_i = \frac{1}{\alpha_i}(r_i - r_{i+1}) - \beta_{i-1} \frac{1}{\alpha_{i-1}}(r_{i-1} - r_i) = \left(\frac{1}{\alpha_i} + \frac{\beta_{i-1}}{\alpha_{i-1}}\right) r_i - \frac{1}{\alpha_i} r_{i+1} - \frac{\beta_{i-1}}{\alpha_{i-1}} r_{i-1}$$

a substitucí $\bar{r}_i = \frac{r_i}{\|r_i\|}$ dostaneme

$$\begin{aligned} A\bar{r}_i &= \left(\frac{1}{\alpha_i} + \frac{\beta_{i-1}}{\alpha_{i-1}}\right)\bar{r}_i - \frac{1}{\alpha_i}\frac{\|r_{i+1}\|}{\|r_i\|}\bar{r}_{i+1} - \frac{\beta_{i-1}}{\alpha_{i-1}}\frac{\|r_{i-1}\|}{\|r_i\|}\bar{r}_{i-1} = \\ &= \left(\frac{1}{\alpha_i} + \frac{\beta_{i-1}}{\alpha_{i-1}}\right)\bar{r}_i - \frac{1}{\alpha_i} \cdot \sqrt{\beta_i} \cdot \bar{r}_{i+1} - \frac{\beta_{i-1}}{\alpha_{i-1}} \frac{1}{\sqrt{\beta_{i-1}}} \bar{r}_{i-1} \end{aligned}$$

Tak dostáváme vyjádření matice A v bázi $\{\bar{r}_i\}$, které dává třídiagonální matici

$$T_{k+1,k} = \begin{pmatrix} \ddots & & & & \\ \ddots & \ddots & \ddots & & \\ & t_{i,i-1} & t_{i,i} & t_{i+1,i} & \\ & & \ddots & \ddots & \ddots \\ & & & & \ddots \end{pmatrix}, \text{ kde } \begin{aligned} t_{i,i} &= \tfrac{1}{\alpha_i} + \tfrac{\beta_{i-1}}{\alpha_{i-1}} \\ t_{i+1,i} &= \tfrac{\sqrt{\beta_i}}{\alpha_i} \\ t_{i,i+1,} &= \tfrac{\sqrt{\beta_i}}{\alpha_i}. \end{aligned}$$

Matici $T_{k+1,k}$ o rozměrech $(k+1) \times k$ lze tedy vytvořit s využitím koeficientů α_i a β_i počítaných v iteracích CG metody. Odebráním posledního řádku dostaneme symetrickou třídiagonální matici T_k. Pokud označíme $V_i = [\bar{r}_0, \ldots, \bar{r}_{i-1}]$ matici se sloupci tvořenými vektory $\bar{r}_i$, potom

$$\begin{aligned} (AV_k)_{pq} &= [A\bar{r}_0, \ldots, A\bar{r}_{i-1}]_{pq} = (A\bar{r}_q)_p = \sum_j (T_{k+1,k})_{qj} (V_{k+1})_{pj} = \\ &= (V_{k+1} T_{k+1,k})_{pq}, \end{aligned}$$

tedy

$$AV_k = V_{k+1} T_{k+1,k}$$

a

$$V_k^T A V_k = V_k^T V_{k+1} T_{k+1,k} = T_k.$$

Takový přechod k třídiagonální matici využívá i Lanzoszova metoda výpočtu vlastních čísel a vektorů matic. T_k tedy můžeme využít pro odhad vlastních čísel, zejména minimálního a maximálního vlastního čísla, čísla podmíněnosti a vztahu mezi reziduem a chybou.

Poznámka 2.9. K vztahu mezi reziduem a chybou poznamenejme, že

$$\|e_k\| = \left\|A^{-1} r_k\right\| \leq \left\|A^{-1}\right\| \|r_k\| \leq \varepsilon \left\|A^{-1}\right\| \|r_0\| \leq \varepsilon \left\|A^{-1}\right\| \|Ae_0\| \leq$$

$$\leq \varepsilon \left\|A^{-1}\right\| \|A\| \|e_0\| \leq \varepsilon\kappa \|e_0\|.$$

Poznámka 2.10. Pokud uvažujeme A-ortonormální bázi tvořenou normovanými směry $\bar{d}_i = d_i / \|d_i\|_A$ potom využitím vztahů $Ad_i = \frac{1}{\alpha_i}(r_i - r_{i+1})$ a $r_i = d_i - \beta_{i-1} d_{i-1}$ dojdeme opět k reprezentaci pomocí třídiagonální matice.