Veta 1. *Ak A je symetrická kladne definitná matica, tak Gaussova-Seidlova metóda konverguje.*

Dôkaz: Pre prípad symetrickej matice dostávame rozklad

$$A = D + L + L^T,$$

kde D je diagonálna matica obsahujúca diagonálne prvky matice A a L a L^T sú zvyšné dolné a horné trojuholníkové časti matice $A - D$. Keďže A je kladne definitná, diagonála matice D obsahuje len kladné prvky, a teda je tiež kladne definitná (ak by $a_{ii}<0$, tak $e_i^T A e_i = a_{ii} < 0$, kde $e_i = (0, \ldots, 1, \ldots, 0)$ je vektor kanonickej bázy).

Na odhad pre konvergenciu použijeme 2-normu, čím zredukujeme problém na odhad veľkosti absolútnych hodnôt vlastných čísel, keďže pre symetrickú maticu A je $\|A\|_2 = |\lambda|_{\max}$. Najskôr zjednodušíme maticu $Q_{GS} = -(D + L)^{-1} L^T$.

$$Q_{GS} = -(D+L)^{-1}L^T = -D^{-\frac{1}{2}}(I + D^{-\frac{1}{2}}LD^{-\frac{1}{2}})^{-1}D^{-\frac{1}{2}}L^T D^{-\frac{1}{2}} D^{\frac{1}{2}}.$$

Označme $L_1 = D^{-\frac{1}{2}}LD^{-\frac{1}{2}}$. Potom

$$Q_{GS} = -D^{-\frac{1}{2}}(I + L_1)^{-1}L_1^T D^{\frac{1}{2}}.$$

Matica $-(I + L_1)^{-1}L_1^T$ je teda podobná s maticou Q_{GS}, a tým pádom má rovnaké vlastné čísla.

Ak x označuje jednotkový vlastný vektor, t.j. $\|x\|_2 = \sqrt{x^* x} = 1$ a $-(I + L_1)^{-1}L_1^T x = \lambda x$ (x môže byť komplexný vektor a λ komplexná vlastná hodnota, * je znamená hermitovské združenie), máme

$$\begin{aligned} -(I + L_1)^{-1}L_1^T x &= \lambda x \\ -L_1^T x &= \lambda(I + L_1)x \end{aligned}$$

a prenásobením posledného vzťahu zľava x^* a úpravou

$$-x^* L_1^T x = \lambda x^*(I + L_1)x = \lambda(I + x^* L_1 x).$$

Označme $z = x^* L_1^T x$. Potom platí

$$\begin{aligned} -z &= \lambda(1 + \bar{z}) \\ \lambda &= -\frac{z}{1 + \bar{z}}. \end{aligned}$$

Z toho dostaneme, že

$$|\lambda|^2 = \lambda\bar{\lambda} = \frac{z}{1+\bar{z}} \cdot \frac{\bar{z}}{1+z} = \frac{|z|^2}{1+z+\bar{z}+|z|^2}$$

a navyše

$$\begin{aligned} 1+z+\bar{z} &= 1 + x^* L_1^T x + x^* L_1 x = x^*(I + L_1 + L_1^T)x \\ &= x^* D^{-\frac{1}{2}}(D + L + L^T)D^{-\frac{1}{2}} \\ &= x^* D^{-\frac{1}{2}} A D^{-\frac{1}{2}} x > 0. \end{aligned}$$

Preto

$$|\lambda|^2 = \frac{|z|^2}{1+z+\bar{z}+|z|^2} < 1,$$

a teda aj $\|Q\|_2 = |\lambda|_{\max} < 1$.

Poznámka: Ak sa na výpočet Jacobiho metódou pozrieme po súradniciach uvidíme, že j-ta súradnica riešenia v x^m sa spočíta ako

$$x_j^{(m)} = \frac{1}{a_{jj}}(b_j - \sum_{k \neq j} a_{jk} x_k^{(m-1)}). \tag{1}$$

To ale znamená, že v čase, keď počítame $x_j^{(m)}$ už poznáme $x_l^{(m)}$ pre $l < j$, čo sú súradnice zlepšeného riešenia. Ak tieto nové súradnice použijeme vo výpočte (1), t.j. rozdelíme sumu v zátvorke na $k < j$ a $k > j$ a v $k < j$ nahradíme $x_k^{(m-1)}$ novými súradnicami $x_k^{(m)}$, dostaneme

$$x_j^{(m)} = \frac{1}{a_{jj}} \left(b_j - \sum_{k<j} a_{jk} x_k^{(m)} - \sum_{k>j} a_{jk} x_k^{(m-1)} \right),$$

čo nie je ale nič iné ako Gaussova-Seidelova metóda. Ako si možno všimnúť

v prípade že niektorý prvok diagonály matice A je nulový, metódy zlyhajú (delenie prvkom a_{ii}). Tento problém možno obísť výmenou riadkov alebo stĺpcov, avšak za cenu straty istoty konvergencie.

Ak by sme chceli všeobecný prípad $Ax = b$ previesť na symetrický $A^T Ax = A^T b$, môže sa stať, že $A^T A$ je zle podmienená ($c_2(A^T A) = c_2(A)^2$).

Metódy založené na minimalizácii kvadratickej formy

Pri týchto metódach hľadáme riešenie systému $Ax = b$ nepriamo ako minimum nejakej vhodne definovanej funkcie $F(x)$. Ako príklady takých funkcií môžeme uviesť

$$F_1(x) = (b - Ax)^T(b - Ax),$$

alebo ak $r = b - Ax$ je reziduum a $\bar{x}$ presné riešenie, tak

$$F_2(x) = (\bar{x} - x)^T(\bar{x} - x) = \Delta x^T \Delta x = (A^{-1}r)^T(A^{-1}r).$$

Symetrické, kladne definitné matice

Pre prípad systému $Ax = b$ so symetrickou a kladne definitnou maticou systému A môžeme definovať funkciu F ako

$$F(x) = \Delta x^T A \Delta x = (\bar{x} - x)^T A(\bar{x} - x).$$

Vďaka kladnej definitnosti máme $F(x) \geq 0$ a $F(x) = 0$ práve vtedy keď $x = \bar{x}$.

Keďže

$$\begin{aligned} F(x) = (\bar{x} - x)^T A(\bar{x} - x) &= \bar{x}^T \underbrace{A\bar{x}}_{b} - x^T \underbrace{A\bar{x}}_{b} - \underbrace{\bar{x}^T A}_{b^T} x + x^T Ax \\ &= x^T Ax - 2b^T x + \bar{x}^T b \end{aligned}$$

a $\bar{x}^T b$ nezávisí od x, stačí minimalizovať funkciu

$$f(x) = x^T Ax - 2b^T x$$

Samotný iteračný krok potom prebieha tak, že k už získanému riešeniu pripočítame vhodný $\alpha \in \mathbb{R}$ násobok nejakého vhodne zvoleného vektora v, čiže $x + \alpha v$ bude nové zlepšené odhad riešenia. To vedie k otázke, ako voliť α a v.

Najskôr predpokladajme, že v sme zvolili. Potom α je zrejme vhodné voliť tak, aby $x + \alpha v$ minimalizovalo funkciu f. Po dosadení $x - \alpha v$ dostaneme

$$\begin{aligned} g(\alpha) = f(x + \alpha v) &= (x + \alpha v)^T A(x + \alpha v) - 2b^T(x + \alpha v) = \\ &= \alpha^2 v^T Av + 2\alpha(v^T Ax - b^T v) + x^T Ax - 2b^T x \\ &= \alpha^2 v^T Av + 2\alpha v^T(Ax - b) + f(x) \\ &= \alpha^2 v^T Av - 2\alpha v^T r + f(x) \end{aligned}$$

čo je kvadratická funkcia premennej α, pričom $v^T Av > 0$, takže minimum g získame pre α spĺňajúce

$$g'(\alpha) = 2\alpha v^T Av - 2v^T r = 0,$$

teda pre

$$\alpha = \frac{v^T r}{v^T Av}$$

Iteračný proces teda môžeme zapísať ako

$$x^{(k)} = x^{(k-1)} + \alpha_k v^{(k)} = x^{(k-1)} + \frac{v^{(k)T} r^{(k-1)}}{v^{(k)T} Av^{(k)}} v \tag{2}$$

Ostáva už len určiť vektory $v^{(k)}$. Na zlepšovanie riešenia stačí, aby $\alpha_k \neq 0$, čo znamená, že

$$v^{(k)T} r^{(k-1)} \neq 0.$$

Vektory $v^{(k)}$ treba teda voliť tak, aby neboli kolmé na reziduá. Rôznym spôsobom voľby vektorov $v^{(k)}$ potom dostávame rôzne metódy.

Metóda najväčšieho spádu

Prirodzená voľba vektora v je zrejme smer, v ktorom sa funkcia $f(x)$ mení najviac. Keďže veľkosť zmeny meriame deriváciou, vektor v hľadáme tak, aby číslo

$$\lim_{t\to\infty} \frac{f(x+tv) - f(x)}{t}$$

bolo v absolútnej hodnote čo najväčšie. Počítajme teda:

$$\lim_{t\to\infty} \frac{f(x+tv) - f(x)}{t} = \left.\frac{df(x+tv)}{dt}\right|_{t=0} = g'(0) = -2v^T r = -2\|v\|_2\|r\|_2 \cos\varphi.$$

Pri voľbe v tak, aby $\|v\|_2 = 1$ je toto číslo najväčšie pre $\varphi = 0$ (prípadne $\varphi = \pi$, ale v tom prípade α bude mať len opačné znamienko). Zvolíme teda $v = r$, takže v iteráciách $v^{(k)} = r^{(k-1)}$ a

$$x^{(k)} = x^{(k-1)} + \frac{r^{(k-1)T} r^{(k-1)}}{r^{(k-1)T} Ar^{(k-1)}} v$$

Samotný algoritmus prebieha sa dá zhrnúť nasledovne

- $x = x^0$

- $r^{(k)} = b - Ax^{(k)}$; $\alpha_{k+1} = \dfrac{r^{(k)T} r^{(k)}}{r^{(k)T} A r^{(k)}}$; $x^{(k+1)} = x^{(k)} + \alpha_{k+1} r^{(k)}$.

Pozrime sa teraz na rýchlosť konvergencie tejto metódy. Pomerne jednoducho sa dá ukázať, že

$$\frac{F(x^{(k+1)})}{F(x^{(k)})} = 1 - \frac{(r^{(k)T} r^{(k)})^2}{(r^{(k)T} A r^{(k)})(r^{(k)T} A^{-1} r^{(k)})}. \tag{3}$$

V predchádzajúcom výraze označme $r = r^{(k)}$ a zapíšme r v báze ortonormálnych vlastných vektorov $(x_1, \ldots, x_n)$, kde $Ax_i = \lambda_i x$. Potom $r = \sum_{i=1}^n \beta_i x_i$ a

$$\begin{aligned}
(r^T r)^2 &= \left[\left(\sum_{i=1}^n \beta_i x_i^T\right)\left(\sum_{i=1}^n \beta_i x_i\right)\right]^2 = \left(\sum_{i=1}^n \beta_i^2\right)^2, \\
(r^T A r)^2 &= \left[\left(\sum_{i=1}^n \beta_i x_i^T\right)\left(\sum_{i=1}^n \beta_i \lambda_i x_i\right)\right] = \sum_{i=1}^n \beta_i^2 \lambda_i, \\
(r^T A^{-1} r)^2 &= \left[\left(\sum_{i=1}^n \beta_i x_i^T\right)\left(\sum_{i=1}^n \beta_i \frac{1}{\lambda_i} x_i\right)\right] = \sum_{i=1}^n \frac{\beta_i^2}{\lambda_i}.
\end{aligned}$$

Po dosadení do (3) máme

$$\frac{F(x^{(k+1)})}{F(x^{(k)})} = 1 - \frac{\left(\sum_{i=1}^n \beta_i^2\right)^2}{\left(\sum_{i=1}^n \beta_i^2 \lambda_i\right)\left(\sum_{i=1}^n \frac{\beta_i^2}{\lambda_i}\right)} = 1 - \frac{1}{H},$$

kde

$$H = \frac{\left(\sum_{i=1}^n \beta_i^2 \lambda_i\right)\left(\sum_{i=1}^n \frac{\beta_i^2}{\lambda_i}\right)}{\left(\sum_{i=1}^n \beta_i^2\right)\left(\sum_{i=1}^n \beta_i^2\right)}$$

a pri označení $s = \sum_{i=1}^n \beta_i^2$

$$H = \left(\sum_{i=1}^n \frac{\beta_i^2 \lambda_i}{s}\right)\left(\sum_{i=1}^n \frac{\beta_i^2}{s\lambda_i}\right).$$

Toto číslo je možné odhadnúť použitím Kantorovičovej nerovnosti

Veta 2. *Ak $x_1, \ldots, x_n$ sú kladné reálne čísla a $\gamma_1, \ldots, \gamma_n \geq 0$, také, že $\sum_{j=1}^n \gamma_i = 1$, tak*

$$\left(\sum_{i=1}^n \gamma_i x_i\right)\left(\sum_{i=1}^n \gamma_i x_i^{-1}\right) \leq \frac{(x_1 + x_n)^2}{4x_1 x_2}.$$

Z toho dostávame odhad

$$H = \left(\sum_{i=1}^{n} \frac{\beta_i^2 \lambda_i}{s} \right) \left(\sum_{i=1}^{n} \frac{\beta_i^2}{s\lambda_i} \right) \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n},$$

a potom

$$\frac{F(x^{(k+1)})}{F(x^{(k)})} \leq 1 - \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2} = \left[\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n} \right].$$

Dá sa ukázať, že potom

$$\|\Delta x^{(k)}\|_2 \leq \left[\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n} \right]^k \|x^{(0)}\|_2 \sqrt{\frac{\lambda_1}{\lambda_n}}.$$

Takže pokiaľ λ_1 a λ_n sú rádovo blízke, metóda konverguje celkom rýchlo. Pokiaľ ale $\lambda_1 \gg \lambda_n$, čiže pre zle podmienené matice, môže byť konvergencia pomalá.

Poznámky: Rýchlosť algoritmu možno zlepšiť

1. $r^{(k+1)} = b - Ax^{(k+1)} = b - A(x^{(k)} + \alpha_{k+1} r^k) = (b - Ax^{(k)}) - \alpha_{k+1} A r^{(k)}$ $= r^{(k)} - \alpha_{k+1} A r^{(k)}$, pričom $Ar^{(k)}$ počítame už pri α_{k+1}, čiže nemusíme počítať $Ax^{(k+1)}$.

2. $x^{(k+1)} = x^{(k)} + \omega_{k+1}\alpha_{k+1}x^{(k)}$, kde $\omega_{k+1} = \frac{\alpha_k}{\alpha_{k+1}}$.

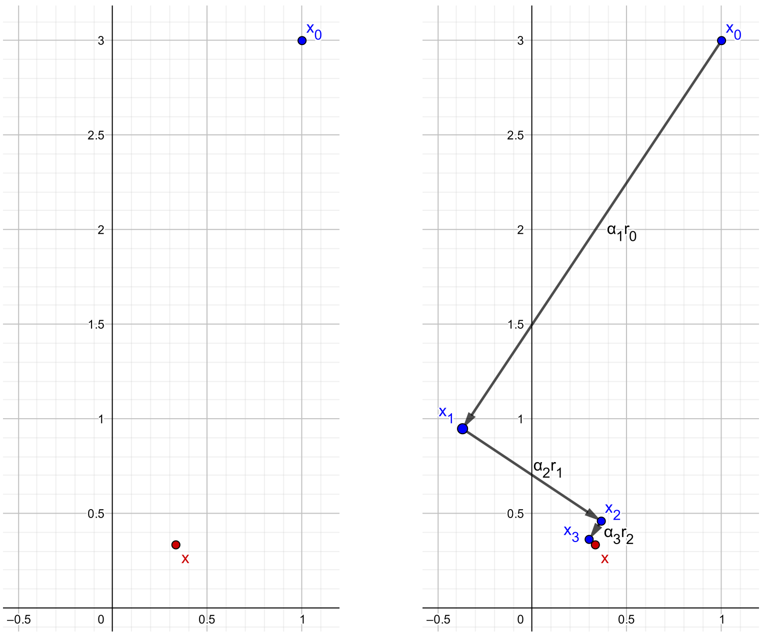
Čo si môžeme všimnúť je, že

$$\begin{aligned} r^{(k)T} r^{(k+1)} &= r^{(k)T}(r^{(k)} - \alpha_{k+1} A r^{(k)}) = r^{(k)T} r^{(k)} - r^{(k)T} \frac{r^{(k)T} r^{(k)}}{r^{(k)T} A r^{(k)}} A r^{(k)} \\ &= r^{(k)T} r^{(k)} - r^{(k)T} r^{(k)} \frac{r^{(k)T} A r^{(k)}}{r^{(k)T} A r^{(k)}} = 0. \end{aligned}$$

Takže $r^{(k)T} \perp r^{(k+1)}$.

Nižšie je animácia a obrázok zobrazujúci tri iterácie prezentovanej metódy pre systém

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$



Obr. 1: Vľavo je animácia postupu výpočtu metódou najväčšieho spádu, vpravo sú zobrazené riešenia získaných z troch iterácií.