Pokročilé numerické metody I
Systémy lineárních rovnic

Jiří Zelinka

Iterační metody pro systémy lineárních rovnic

Systém

$$A\boldsymbol{x} = \boldsymbol{b}$$

převedeme na

$$\boldsymbol{x} = T\boldsymbol{x} + \boldsymbol{g}$$

$\hat{\boldsymbol{x}}$ – řešení
$\hat{\boldsymbol{x}} = (E - T)^{-1}\boldsymbol{g}$ za předpokladu, že $E - T$ je regulární.

$\boldsymbol{x}^0 \in \mathbb{R}^n$ – libovolná počáteční aproximace. Posloupnost $\left\{\boldsymbol{x}^k\right\}_{k=0}^{\infty}$ určená rekurentně vztahem

$$\boldsymbol{x}^{k+1} = T\boldsymbol{x}^k + \boldsymbol{g}, \qquad k = 0, 1, \ldots$$

se nazývá **iterační posloupnost** a matice T se nazývá **iterační matice**.

Problémy:

1. Jak zvolit iterační matici T, tj. jakým způsobem převést systém $A\boldsymbol{x} = \boldsymbol{b}$ na systém $\boldsymbol{x} = T\boldsymbol{x} + \boldsymbol{g}$?
2. Za jakých předpokladů posloupnost $\left\{\boldsymbol{x}^k\right\}_{k=0}^{\infty}$ konverguje pro libovolnou počáteční aproximaci k přesnému řešení $\hat{\boldsymbol{x}}$?

$$\begin{aligned}
\boldsymbol{x}^1 &= T\boldsymbol{x}^0 + \boldsymbol{g}, \\
\boldsymbol{x}^2 &= T\boldsymbol{x}^1 + \boldsymbol{g} = T(T\boldsymbol{x}^0 + \boldsymbol{g}) + \boldsymbol{g} = T^2\boldsymbol{x}^0 + (T + E)\boldsymbol{g}, \\
\boldsymbol{x}^3 &= T\boldsymbol{x}^2 + \boldsymbol{g} = T^3\boldsymbol{x}^0 + (T^2 + T + E)\boldsymbol{g}, \\
&\vdots \\
\boldsymbol{x}^{k+1} &= T^{k+1}\boldsymbol{x}^0 + (T^k + T^{k-1} + \ldots + E)\boldsymbol{g}.
\end{aligned}$$

Definice

Řekneme, že matice H je **konvergentní**, jestliže

$$\lim_{k\to\infty} H^k = O,$$

kde O je nulová matice, konvergence je bodová.

Věta 1

Následující tvrzení jsou ekvivalentní:

1. H je konvergentní matice.
2. $\lim\limits_{k\to\infty} \|H^k\| = 0$ pro nějakou přidruženou maticovou normu.
3. $\rho(H) < 1$ ($\rho(H)$ je spektrální poloměr H).
4. $\lim\limits_{k\to\infty} H^k \boldsymbol{x} = \boldsymbol{o}$ pro libovolný vektor $\boldsymbol{x} \in \mathbb{R}^n$.

Lemma

Nechť $\rho(T) < 1$. Pak $E - T$ je regulární a platí

$$(E - T)^{-1} = E + T + T^2 + \dots$$

Věta 2 (Hlavní věta o konvergenci iteračního procesu)
Posloupnost $\{\boldsymbol{x}^k\}_{k=0}^{\infty}$ určená iteračním procesem $\boldsymbol{x} = T\boldsymbol{x} + \boldsymbol{g}$ konverguje pro každou počáteční aproximaci $\boldsymbol{x}^0 \in \mathbb{R}^n$ právě tehdy, když $\rho(T) < 1$, přičemž

$$\lim_{k\to\infty} \boldsymbol{x}^k = \hat{\boldsymbol{x}},\ \hat{\boldsymbol{x}} = T\boldsymbol{x}^* + \boldsymbol{g}$$

Důsledek

Nechť pro nějakou přidruženou maticovou normu platí $\|T\| < 1$. Pak posloupnost $\{\boldsymbol{x}^k\}_{k=0}^{\infty}$ generovaná iteračním procesem $\boldsymbol{x} = T\boldsymbol{x} + \boldsymbol{g}$ konverguje k řešení $\hat{\boldsymbol{x}} = (E - T)^{-1}\boldsymbol{g}$ pro každou počáteční aproximaci $\boldsymbol{x}^0 \in \mathbb{R}^n$. Dále platí

$$\begin{aligned}
\|\hat{\boldsymbol{x}} - \boldsymbol{x}^k\| &\leq \|T\|^k \|\hat{\boldsymbol{x}} - \boldsymbol{x}^0\|, \\
\|\hat{\boldsymbol{x}} - \boldsymbol{x}^k\| &\leq \frac{\|T\|^k}{1 - \|T\|} \|\boldsymbol{x}^1 - \boldsymbol{x}^0\|.
\end{aligned}$$

Jacobiova iterační metoda

$$A\boldsymbol{x} = \boldsymbol{b}, \qquad A = \begin{pmatrix} a_{11} & \cdots & \cdots & a_{1n} \\ a_{21} & & & a_{2n} \\ \vdots & & & \vdots \\ a_{n1} & \cdots & \cdots & a_{nn} \end{pmatrix}$$

Z první rovnice vypočteme x_1:

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \Rightarrow x_1 = \frac{1}{a_{11}}(b_1 - a_{12}x_2 - \ldots - a_{1n}x_n)$$

$$x_1^{k+1} = \frac{1}{a_{11}}(b_1 - a_{12}x_2^k - \ldots - a_{1n}x_n^k),$$

z druhé rovnice vypočteme x_2:

$$x_2^{k+1} = \frac{1}{a_{22}}(b_2 - a_{21}x_1^k - a_{23}x_3^k - \ldots - a_{1n}x_n^k),$$

Obecně z i-té rovnice vypočteme x_i:

$$x_i^{k+1} = -\sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{a_{ij}}{a_{ii}} x_j^k + \frac{b_i}{a_{ii}},$$

až z n-té rovnice vypočteme x_n.

Na pravé straně jsou vždy všechny prvky $A\mathbf{x}$ kromě diagonálního, kterým se rovnice vydělí.

Maticový zápis:

Matici A zapišme ve tvaru

$$A = D + L + U,$$

kde

$$D = \begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix},$$

$$
L = \begin{pmatrix} 0 & & & 0 \\ a_{21} & \ddots & & \\ \vdots & \ddots & \ddots & \\ a_{n1} & \cdots & a_{n,n-1} & 0 \end{pmatrix},
$$

$$
U = \begin{pmatrix} 0 & a_{12} & \cdots & a_{1n} \\ & \ddots & \ddots & \vdots \\ & & \ddots & a_{n-1,n} \\ 0 & & & 0 \end{pmatrix}.
$$

D je diagonální matice, L je dolní trojúhelníková matice s nulami na diagonále a U je horní trojúhelníková matice s nulami na diagonále.

$$Ax = (D + L + U)x = b$$

$$Dx = -(L + U)x + b.$$

Pokud $a_{ii} \neq 0$, $i = 1, \ldots, n$, je matice D regulární a z předchozí rovnice lze vypočítat

$$x = -D^{-1}(L + U)x + D^{-1}b.$$

$$D^{-1} = \begin{pmatrix} \frac{1}{a_{11}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{a_{nn}} \end{pmatrix},$$

Maticový tvar Jacobiovy iterační metody

Jacobiova iterační matice: $T_J = -D^{-1}(L+U)$

$$\mathbf{x}^{k+1} = T_J \mathbf{x}^k + D^{-1}\mathbf{b},$$

$T_J = (t_{ij})$, $t_{ij} = -\frac{a_{ij}}{a_{ii}}$ pro $i \neq j$, $t_{ii} = 0$ pro $i = 1, \ldots, n$.

$$T_J = \begin{pmatrix} 0 & -\dfrac{a_{12}}{a_{11}} & \cdots & -\dfrac{a_{1n}}{a_{11}} \\ -\dfrac{a_{21}}{a_{22}} & 0 & & -\dfrac{a_{2n}}{a_{22}} \\ \vdots & & \ddots & \vdots \\ -\dfrac{a_{n1}}{a_{nn}} & -\dfrac{a_{n2}}{a_{nn}} & \cdots & 0 \end{pmatrix}, \quad D^{-1}\mathbf{b} = \begin{pmatrix} \dfrac{b_1}{a_{11}} \\ \dfrac{b_2}{a_{22}} \\ \vdots \\ \dfrac{b_n}{a_{nn}} \end{pmatrix}.$$

Věta 3 (O konvergenci Jacobiovy iterační metody)
Posloupnost $\{\boldsymbol{x}^k\}_{k=0}^{\infty}$ generovaná metodou $\boldsymbol{x}^{k+1} = T_J\boldsymbol{x}^k + D^{-1}\boldsymbol{b}$ konverguje pro každou počáteční aproximaci $\boldsymbol{x}^0 \in \mathbb{R}^n$ právě tehdy, když $\varrho(T_J) < 1$.

Odhad chyby:

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}^k\|_\infty \leq \frac{\|T_J\|_\infty^k}{1 - \|T\|_\infty}\|\boldsymbol{x}^1 - \boldsymbol{x}^0\|_\infty.$$

Silné řádkové sumační kriterium:

Nechť matice A je ryze řádkově diagonálně dominantní, tj.

$$|a_{ii}| > \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|, \qquad i = 1, \dots, n.$$

Pak Jacobiova iterační metoda konverguje pro každou počáteční aproximaci $\boldsymbol{x}^0 \in \mathbb{R}^n$.

Silné sloupcové sumační kriterium:

Nechť matice A je ryze sloupcově diagonálně dominantní, tj.

$$|a_{kk}| > \sum_{\substack{i=1 \\ i\neq k}}^{n} |a_{ik}|, \qquad k = 1, \dots, n.$$

Pak Jacobiova iterační metoda konverguje pro každou počáteční aproximaci $\boldsymbol{x}^0 \in \mathbb{R}^n$.

Gaussova-Seidelova iterační metoda

Z první rovnice vypočteme x_1:

$$a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n = b_1 \Rightarrow x_1 = \frac{1}{a_{11}}(b_1-a_{12}x_2-\ldots-a_{1n}x_n)$$

$$x_1^{k+1} = \frac{1}{a_{11}}(b_1 - a_{12}x_2^k - \ldots - a_{1n}x_n^k),$$

z druhé rovnice vypočteme x_2, pro x_1 použijme novou iteraci:

$$x_2^{k+1} = \frac{1}{a_{22}}(b_2 - a_{21}x_1^{k+1} - a_{23}x_3^k - \ldots - a_{1n}x_n^k),$$

ze třetí rovnice vypočteme x_3, pro x_1 a x_2 použijme novou iteraci:

$$x_3^{k+1} = \frac{1}{a_{33}}(b_3 - a_{31}x_1^{k+1} - a_{32}x_2^{k+1} - a_{34}x_4^k \ldots - a_{1n}x_n^k),$$

Obecně

$$x_i^{k+1} = -\sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{k+1} - \sum_{j=i+1}^{n} \frac{a_{ij}}{a_{ii}} x_j^k + \frac{b_i}{a_{ii}}, \qquad i = 1, \dots, n.$$

Maticový zápis:

$$\begin{aligned} A\boldsymbol{x} = \boldsymbol{b} \quad \Rightarrow \quad (D + L + U)\boldsymbol{x} &= \boldsymbol{b} \\ (D + L)\boldsymbol{x} &= -U\boldsymbol{x} + \boldsymbol{b}. \end{aligned}$$

$a_{ii} \neq 0$, $i = 1, \dots, n$, $\quad \Rightarrow$ matice $D + L$ je regulární a

$$\boldsymbol{x} = -(D + L)^{-1} U\boldsymbol{x} + (D + L)^{-1}\boldsymbol{b}.$$

Položme $T_G = -(D + L)^{-1} U$, Gaussova-Seidelova iterační metoda je tvaru

$$\boldsymbol{x}^{k+1} = T_G \boldsymbol{x}^k + \boldsymbol{g}, \qquad \boldsymbol{g} = (D + L)^{-1}\boldsymbol{b}.$$

Věta 4

Posloupnost $\{\boldsymbol{x}^k\}_{k=0}^{\infty}$ generovaná Gaussovou-Seidelovou iterační metodou $\boldsymbol{x}^{k+1} = -(D+L)^{-1}U\boldsymbol{x}^k + (D+L)^{-1}\boldsymbol{b}$. konverguje pro každou počáteční aproximaci $\boldsymbol{x}^0 \in \mathbb{R}^n$ právě tehdy, když $\varrho(T_G) < 1$.

Silné řádkové sumační kriterium:

Nechť matice A je ryze řádkově diagonálně dominantní, tj.

$$|a_{ii}| > \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}|, \qquad i = 1, \dots, n.$$

Pak Gaussova-Seidelova iterační metoda konverguje pro každou počáteční aproximaci $\boldsymbol{x}^0 \in \mathbb{R}^n$.

Silné sloupcové sumační kriterium:

Nechť matice A je ryze sloupcově diagonálně dominantní, tj.

$$|a_{kk}| > \sum_{\substack{i=1 \\ i \neq k}}^{n} |a_{ik}|, \qquad k = 1, \dots, n.$$

Pak Gaussova-Seidelova iterační metoda konverguje pro každou počáteční aproximaci $\boldsymbol{x}^0 \in \mathbb{R}^n$.

Věta 5

Nechť A je pozitivně definitní matice. Pak Gaussova-Seidelova metoda konverguje pro každou počáteční aproximaci.

Věta 6 (Stein-Rosenberg)
Nechť pro prvky matice A platí $a_{ij} \leq 0$ pro všechna $i \neq j$ a $a_{ii} > 0$, $i = 1, \dots, n$. Pak platí právě jedno z následujících tvrzení:

- $0 < \varrho(T_G) < \varrho(T_J) < 1$
- $1 < \varrho(T_J) < \varrho(T_G)$
- $\varrho(T_J) = \varrho(T_G) = 0$
- $\varrho(T_J) = \varrho(T_G) = 1.$

To znamená, že konvergují-li obě metody, Gaussova-Seidelova metoda konverguje rychleji.

Relaxační metody

Modifikace Gaussovy–Seidelovy metody, ω – relaxační parametr

$$x_i^{k+1} = (1-\omega)x_i^k + \frac{\omega}{a_{ii}}\left[b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{k+1} - \sum_{j=i+1}^{n} a_{ij}x_j^k\right].$$

Relaxační metodu lze maticově zapsat takto

$$\boldsymbol{x}^{k+1} = (D+\omega L)^{-1}[(1-\omega)D - \omega U]\boldsymbol{x}^k + \omega(D+\omega L)^{-1}\boldsymbol{b}$$

$$T_\omega = (D+\omega L)^{-1}[(1-\omega)D - \omega U]$$

Hodnoty parametru ω:

Pro	$0 < \omega < 1$	se iterační metody nazývají **metodami dolní relaxace**. Tyto metody jsou vhodné v případě, že Gaussova-Seidelova metoda nekonverguje.
Pro	$\omega = 1$	je relaxační metoda totožná s Gaussovou-Seidelovou metodou.
Pro	$1 < \omega$	se metody nazývají **metodami horní relaxace**, nebo častěji SOR metodami (SOR = Successive Over-Relaxation). Tyto metody lze užít ke zrychlení konvergence Gaussovy-Seidelovy metody.

Věta 7 (Kahan)
Nechť $a_{ii} \neq 0$, $i = 1, \ldots, n$. Pak

$$\varrho(T_\omega) \geq |\omega - 1|.$$

Důsledek: Má smysl uvažovat jen $\omega \in (0, 2)$.

Věta 8 (Ostrowski-Reich)
Pro pozitivně definitní matici A platí $\varrho(T_\omega) < 1$ pro všechna $\omega \in (0, 2)$.

Třídiagonální matice:

$$A = (a_{ij})_{i,j=1}^{n}, \qquad a_{ij} = 0, \text{ pro } |i - j| > 1$$

Věta 9
Nechť A je třídiagonální pozitivně definitní matice. Pak $\varrho(T_G) = \varrho^2(T_J) < 1$ a optimální hodnota relaxačního parametru je dána vztahem

$$\omega = \omega_{opt} = \frac{2}{1 + \sqrt{1 - \varrho^2(T_J)}}.$$

Při této volbě je $\varrho(T_\omega) = |1 - \omega|$.

Cykly v iteračních metodách

Například pro systémy

$$\begin{aligned} x_1 + kx_2 &= b_1 \\ x_1 - kx_2 &= b_2 \end{aligned}$$

Jacobiova metoda: cyklus délky 4
Gaussova–Seidelova metoda: cyklus délky 2

Relaxační metody: cykly různých délek

$$\begin{aligned} x_1 + x_2 &= 1 \\ qx_1 + x_2 &= 1. \end{aligned}$$

Pro $\omega = 2$:

$$T_2 = \begin{pmatrix} -1 & -2 \\ 2q & 4q-1 \end{pmatrix}, \quad \lambda_{1,2} = \cos\varphi \pm i\sin\varphi, \quad q = \frac{1}{2}(1+\cos\varphi)$$

$\varphi = 2\pi l/p$, $0 < l < p/2 \Rightarrow$ existuje cyklus délky p.

Body cyklu leží na elipse se středem v hledaném řešení.

Věta 10

Jestliže A je pozitivně definitní matice, pak řešení systému $A\boldsymbol{x} = \boldsymbol{b}$ je ekvivalentní minimalizaci kvadratické funkce

$$Q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b}.$$

Tato kvadratická funkce má jediné minimum, kterého nabývá v řešení systému $A\boldsymbol{x} = \boldsymbol{b}$, tj. pro $\hat{\boldsymbol{x}} = A^{-1}\boldsymbol{b}$.

Důsledek

$\min Q(\boldsymbol{x}) = Q(\hat{\boldsymbol{x}}) = -\frac{1}{2}\boldsymbol{b}^T A^{-1}\boldsymbol{b}.$

Geometrická interpretace

Nechť $\hat{\boldsymbol{x}}$ je řešení systému $A\boldsymbol{x} = \boldsymbol{b}$ a nechť $\boldsymbol{x} = \hat{\boldsymbol{x}} + \boldsymbol{s}$. Pak

$$Q(\boldsymbol{x}) + \frac{1}{2}\hat{\boldsymbol{x}}^T\boldsymbol{b} = \frac{1}{2}(\hat{\boldsymbol{x}}+\boldsymbol{s})^T A\,(\hat{\boldsymbol{x}}+\boldsymbol{s}) - (\hat{\boldsymbol{x}}+\boldsymbol{s})^T\boldsymbol{b} + \frac{1}{2}\hat{\boldsymbol{x}}^T\boldsymbol{b} = \frac{1}{2}\boldsymbol{s}^T A\boldsymbol{s}\,.$$

Plocha o rovnici

$$\frac{1}{2}\boldsymbol{s}^T A\boldsymbol{s} = konst.$$

je hyperelipsoid v proměnných $s_1,\dots,s_n$ se středem v $\boldsymbol{s} = \boldsymbol{o}$, tj. v $\hat{\boldsymbol{x}}$, tedy i rovnice $Q(\boldsymbol{x}) = konst.$ představuje hyperelipsoid, směry jeho os jsou dány vlastními vektory matice A, délky poloos vlastními čísly.

Minimalizace funkce Q:

vybereme počáteční aproximaci $\boldsymbol{x}_1$ a pak určíme $\boldsymbol{x}_2$ tak, že zvolíme nějaký směr $\boldsymbol{v}_1$ a vzdálenost t_1 ve směru $\boldsymbol{v}_1$. Obecně

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + t_k\boldsymbol{v}_k, \quad i = 1, 2, \dots$$

Vektory $\boldsymbol{v}_k$ se nazývají *směrové vektory*.

Metoda maximálního spádu (steepest descent)

$\boldsymbol{x}_1$ je počáteční aproximace řešení $\hat{\boldsymbol{x}}$, položme $\boldsymbol{r}_1 = \boldsymbol{b} - A\boldsymbol{x}_1$: *reziduový vektor*. Hledáme takový vektor $\boldsymbol{v}_1$, $\|\boldsymbol{v}_1\| = \|\boldsymbol{b} - A\boldsymbol{x}_1\|$, pro který

$$\frac{\delta}{\delta t} Q(\boldsymbol{x}_1 + t\boldsymbol{v}_1)\bigg|_{t=0} = min.$$

Výsledek: $\boldsymbol{v}_1 = \boldsymbol{r}_1 = \boldsymbol{b} - A\boldsymbol{x}_1$.

Dále hledáme minimum Q ve směru $\boldsymbol{v}_1 = \boldsymbol{r}_1$ od bodu $\boldsymbol{x}_1$, tedy minimum kvadratické funkce $Q(\boldsymbol{x}_1 + t\boldsymbol{r}_1)$ v bodě t_1.

Výsledek: $t_1 = \frac{\boldsymbol{r}_1^T \boldsymbol{r}_1}{\boldsymbol{r}_1^T A \boldsymbol{r}_1}$, takže $\boldsymbol{x}_2 = \boldsymbol{x}_1 + \frac{\boldsymbol{r}_1^T \boldsymbol{r}_1}{\boldsymbol{r}_1^T A \boldsymbol{r}_1}\boldsymbol{r}_1$ a obecně

$$t_{k+1} = \frac{{\boldsymbol{r}_k}^T \boldsymbol{r}_k}{{\boldsymbol{r}_k}^T A \boldsymbol{r}_k}, \quad \boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \frac{{\boldsymbol{r}_k}^T \boldsymbol{r}_k}{{\boldsymbol{r}_k}^T A \boldsymbol{r}_k}\boldsymbol{r}_k\,.$$

Vlastnosti iteračního procesu:

1. $\boldsymbol{r}_{k+1} = \boldsymbol{r}_k - \frac{\boldsymbol{r}_k{}^T \boldsymbol{r}_k}{\boldsymbol{r}_k{}^T A \boldsymbol{r}_k} A\boldsymbol{r}_k = \boldsymbol{r}_k - t_k A\boldsymbol{r}_k$
2. $(\boldsymbol{r}_{k+1}, \boldsymbol{r}_k) = \boldsymbol{r}_k^T \boldsymbol{r}_{k+1} = 0$
3. $Q(\boldsymbol{x}_{k+1}) < Q(\boldsymbol{x}_k)$, $k = 1, 2, \dots$

Rychlost konvergence

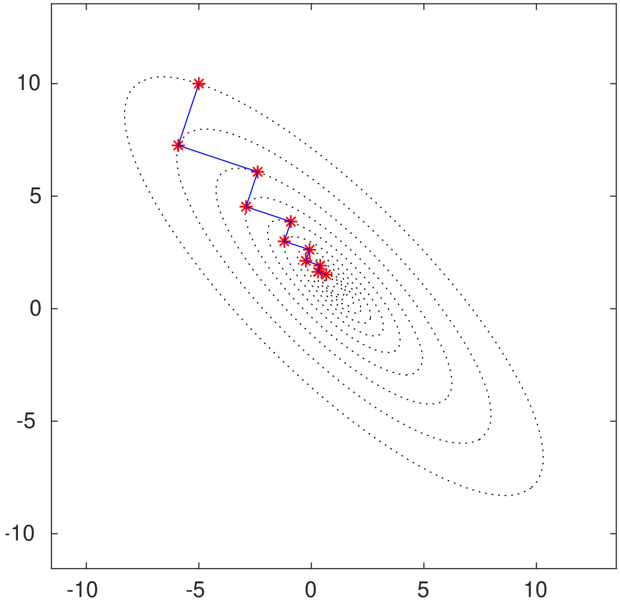
Buď $\hat{\boldsymbol{x}}$ řešení systému $A\boldsymbol{x} = \boldsymbol{b}$. Pak přímým výpočtem dostaneme $Q(\hat{\boldsymbol{x}}) = -\frac{1}{2}\boldsymbol{b}^T\hat{\boldsymbol{x}} = -\frac{1}{2}\boldsymbol{b}^T A^{-1}\boldsymbol{b}$.

Věta 11

Označme $E(\boldsymbol{x}_k) = Q(\boldsymbol{x}_k) - Q(\hat{\boldsymbol{x}}) = \frac{1}{2}(\boldsymbol{x}_k - \hat{\boldsymbol{x}})^T A(\boldsymbol{x}_k - \hat{\boldsymbol{x}})$.

Pak $E(\boldsymbol{x}_{k+1}) \leq \left(\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}}\right)^2 E(\boldsymbol{x}_k)$

Metoda maxmálního spádu



Metoda sdružených (konjugovaných) gradientů

Nechť $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$ tvoří bázi, $\hat{\boldsymbol{x}}$ je řešení $Ax = b$, $\hat{\boldsymbol{x}} = \sum_{i=1}^{n} c_i \boldsymbol{p}_i$.

Pokud jsou $\boldsymbol{p}_i$ ortogonální, pak $c_i = \frac{(\hat{\boldsymbol{x}}, \boldsymbol{p}_i)}{(\boldsymbol{p}_i, \boldsymbol{p}_i)}$.

Problém: neznáme $\hat{\boldsymbol{x}}$.

Výchozí rovnost násobíme maticí A:

$A\hat{\boldsymbol{x}} = \boldsymbol{b} = \sum_{i=1}^{n} c_i A\boldsymbol{p}_i$. Kdyby platilo

$$(A\boldsymbol{p}_j, \boldsymbol{p}_i) = \boldsymbol{p}_i^T A\boldsymbol{p}_j = 0 \text{ pro } i \neq j, \text{ pak } c_i = \frac{(\boldsymbol{b}, \boldsymbol{p}_i)}{(A\boldsymbol{p}_i, \boldsymbol{p}_i)}.$$

Proto definujme nový skalární součin $(\boldsymbol{x}, \boldsymbol{y})_A = (A\boldsymbol{x}, \boldsymbol{y})$, vektory, pro něž $(\boldsymbol{x}, \boldsymbol{y})_A = 0$, se nazývají A-ortogonální (A-sdružené, A-konjugované).

Cílem metody je sestrojit A-ortogonální vektory $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$ podobně jako u Gramova–Schmidtova ortogonalizačního procesu.

Modifikovaný Gramův–Schmidtův ortogonalizační proces:
Z dané báze $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n$ vytváříme novou, A-ortogonální bázi. Položme $\boldsymbol{p}_1 = \boldsymbol{r}_1$ a dále

$$\boldsymbol{p}_k = \boldsymbol{r}_k + \sum_{i=1}^{k-1} \beta_{k,i} \boldsymbol{p}_k, \text{ pak } \beta_{k,i} = -\frac{(\boldsymbol{r}_k, \boldsymbol{p}_i)_A}{(\boldsymbol{p}_i, \boldsymbol{p}_i)_A}.$$

Jako bázi $\boldsymbol{r}_k$ bereme postupně rezidua, v každém kroku počítáme koeficienty α i β.

Snadno se ověří (viz metoda maximálního spádu), že pro iteraci $\boldsymbol{x}_k$ a směr daný vektorem $\boldsymbol{p}$ má funkce Q minimum v

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}, \text{ pro } \alpha_k = \frac{\boldsymbol{p}^T \boldsymbol{r}_k}{\boldsymbol{p}^T A \boldsymbol{p}} = \frac{(\boldsymbol{r}_k, \boldsymbol{p})}{(\boldsymbol{p}, \boldsymbol{p})_A}.$$

Počáteční krok

$\boldsymbol{x}_1$ – počáteční iterace, $\boldsymbol{p}_1 = \boldsymbol{r}_1 = \boldsymbol{b} - A\boldsymbol{x}_1$.

Následně $\boldsymbol{x}_2 = \boldsymbol{x}_1 + \alpha_1 \boldsymbol{p}_1$ pro $\alpha_1 = \frac{(\boldsymbol{r}_1, \boldsymbol{p}_1)}{(\boldsymbol{p}_1, \boldsymbol{p}_1)_A}$, potom

$\boldsymbol{r}_2 = \boldsymbol{b} - A\boldsymbol{x}_2 = \boldsymbol{r}_1 - \alpha_1 A\boldsymbol{p}_1$ a

$\boldsymbol{p}_2 = \boldsymbol{r}_2 + \beta_{2,1}\boldsymbol{p}_1$ pro $\beta_{2,1} = -\frac{(\boldsymbol{r}_2, \boldsymbol{p}_1)_A}{(\boldsymbol{p}_1, \boldsymbol{p}_1)_A}$.

Obecný krok

Pro iteraci $\boldsymbol{x}_k$, reziduum $\boldsymbol{r}_k = b - A\boldsymbol{x}_k$ a směrový vektor $\boldsymbol{p}_k$ položíme

$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$, kde $\alpha_k = \frac{(\boldsymbol{r}_k, \boldsymbol{p}_k)}{(\boldsymbol{p}_k, \boldsymbol{p}_k)_A}$.

$\boldsymbol{r}_{k+1} = b - A\boldsymbol{x}_{k+1} = \boldsymbol{r}_k - \alpha_k A\boldsymbol{p}_k$,

$\boldsymbol{p}_{k+1} = \boldsymbol{r}_{k+1} + \sum_{i=1}^{k} \beta_{k+1,i}\boldsymbol{p}_k$ pro $\beta_{k+1,i} = -\frac{(\boldsymbol{r}_{k+1}, \boldsymbol{p}_i)_A}{(\boldsymbol{p}_i, \boldsymbol{p}_i)_A}$.

Věta 12 (Vlastnosti definovaných vektorů)

1. $\mathcal{L}(\boldsymbol{r}_1, \dots, \boldsymbol{r}_k) = \mathcal{L}(\boldsymbol{p}_1, \dots, \boldsymbol{p}_k)$ pro $k = 1, 2, \dots$.
2. Vektory $\boldsymbol{p}_1, \boldsymbol{p}_2, \dots$ jsou A-ortogonální.
3. Vektor $\boldsymbol{r}_k - \boldsymbol{r}_{k+1}$ je kolmý k $\boldsymbol{p}_i$ pro $i \neq k$.
   Důkaz: $\boldsymbol{r}_k - \boldsymbol{r}_{k+1} = \alpha_k A\boldsymbol{p}_k$, takže pro $i \neq k$ dostáváme $(\boldsymbol{p}_i, \boldsymbol{r}_k - \boldsymbol{r}_{k+1}) = \alpha_k(\boldsymbol{p}_i, A\boldsymbol{p}_k) = \alpha_k(\boldsymbol{p}_i, \boldsymbol{p}_k)_A = 0$.
4. $(\boldsymbol{r}_k, \boldsymbol{p}_k)_A = (\boldsymbol{p}_k, \boldsymbol{p}_k)_A, \ \forall k$.
   Důkaz indukcí: Pro $k = 1$ platí, nechť platí pro k, pak

$$(\boldsymbol{r}_{k+1}, \boldsymbol{p}_{k+1})_A = (\boldsymbol{p}_{k+1} - \sum_{i=1}^{k} \beta_{k+1,i}\boldsymbol{p}_i, \boldsymbol{p}_{k+1})_A = (\boldsymbol{p}_{k+1}, \boldsymbol{p}_{k+1})_A$$

5. Vektory $\boldsymbol{r}_1, \boldsymbol{r}_2, \dots$ jsou ortogonální.

Důkaz indukcí: Nejprve dokážeme $(\boldsymbol{r}_2, \boldsymbol{r}_1) = 0$:

$$(\boldsymbol{r}_2, \boldsymbol{r}_1) = (\boldsymbol{r}_1, \boldsymbol{r}_1) - \alpha_1(A\boldsymbol{p}_1, \boldsymbol{r}_1) = (\boldsymbol{r}_1, \boldsymbol{r}_1) - \frac{(\boldsymbol{r}_1, \boldsymbol{p}_1)}{(\boldsymbol{p}_1, \boldsymbol{p}_1)_A}(\boldsymbol{p}_1, \boldsymbol{r}_1)_A = 0,$$

neboť $\boldsymbol{p}_1 = \boldsymbol{r}_1$.

Dále nechť jsou kolmé vektory $\boldsymbol{r}_1, \dots, \boldsymbol{r}_k$ pro $k \geq 2$. Pak

$$(\boldsymbol{r}_{k+1}, \boldsymbol{r}_k) = (\boldsymbol{r}_k, \boldsymbol{r}_k) - \alpha_k(A\boldsymbol{p}_k, \boldsymbol{r}_k) = (\boldsymbol{r}_k, \boldsymbol{r}_k) -$$
$$- \frac{(\boldsymbol{r}_k, \boldsymbol{p}_k)}{(\boldsymbol{p}_k, \boldsymbol{p}_k)_A}(\boldsymbol{p}_k, \boldsymbol{r}_k)_{A(\text{viz 4.})} = (\boldsymbol{r}_k, \boldsymbol{r}_k) - (\boldsymbol{r}_k, \boldsymbol{r}_k + \sum_{i<k} \beta_{k,i}\boldsymbol{p}_i) =$$
$$= (\boldsymbol{r}_k, \boldsymbol{r}_k) - (\boldsymbol{r}_k, \boldsymbol{r}_k + \sum_{i<k} \gamma_i \boldsymbol{r}_i) = 0 \text{ pro nějaké koeficienty } \gamma_i$$

vzhledem k vlastnosti 1. a indukčnímu předpokladu.

Nakonec dokážeme $(\boldsymbol{r}_{k+1}, \boldsymbol{r}_i) = 0$ pro $i < k$:

$$(\boldsymbol{r}_{k+1}, \boldsymbol{r}_i) = (\boldsymbol{r}_k - (\boldsymbol{r}_k - \boldsymbol{r}_{k+1}), \boldsymbol{r}_i) = (\boldsymbol{r}_k, \boldsymbol{r}_i) -$$
$$-(\boldsymbol{r}_k - \boldsymbol{r}_{k+1}, \boldsymbol{r}_i) = 0 - (\boldsymbol{r}_k - \boldsymbol{r}_{k+1}, \boldsymbol{p}_i - \sum_{j<i} \beta_{i,j}\boldsymbol{p}_j) = 0$$

vzhledem k vlastnosti 3.

6. Pro nějaké k (nejvýše $k = n + 1$) je $\boldsymbol{r}_k = 0$.
   Plyne okamžitě z 5., jedná se o teoretický výsledek prakticky nedosažitelný kvůli zaokrouhlovacím chybám.
7. $(\boldsymbol{r}_k, \boldsymbol{p}_k) = (\boldsymbol{r}_k, \boldsymbol{r}_k),\ (\boldsymbol{r}_k, \boldsymbol{p}_i) = 0\ \forall k,\ \forall i < k.$
   Důkaz obou rovností plyne z vlastností 1. a 5. (viz též podobná část v důkazu 5.):
   $(\boldsymbol{r}_k, \boldsymbol{p}_k) = (\boldsymbol{r}_k, \boldsymbol{r}_k + \sum_{i<k} \beta_{k,i}\boldsymbol{p}_i) = (\boldsymbol{r}_k, \boldsymbol{r}_k).$
8. $\beta_{k+1,i} = 0$ pro $i < k$
   Důkaz plyne z kolmosti reziduí:

$$\begin{aligned}
\beta_{k+1,i} &= -\frac{(\boldsymbol{r}_{k+1}, \boldsymbol{p}_i)_A}{(\boldsymbol{p}_i, \boldsymbol{p}_i)_A}, \\
(\boldsymbol{r}_{k+1}, \boldsymbol{p}_i)_A &= (\boldsymbol{r}_{k+1}, A\boldsymbol{p}_i) = \frac{1}{\alpha_i}(\boldsymbol{r}_{k+1}, \boldsymbol{r}_i - \boldsymbol{r}_{i+1}) = 0
\end{aligned}$$

9. $\beta_{k+1,k} = \beta_{k+1} = \frac{(\boldsymbol{r}_{k+1},\boldsymbol{r}_{k+1})}{(\boldsymbol{r}_k,\boldsymbol{r}_k)} = \frac{\boldsymbol{r}_{k+1}^T\boldsymbol{r}_{k+1}}{\boldsymbol{r}_k^T\boldsymbol{r}_k}$

   Důkaz plyne z vlastnosti 7. a z kolmosti reziduí:

$$\beta_{k+1} = -\frac{(\boldsymbol{r}_{k+1},\boldsymbol{p}_k)_A}{(\boldsymbol{p}_k,\boldsymbol{p}_k)_A} = -\frac{\boldsymbol{r}_{k+1}^T A\boldsymbol{p}_k}{\boldsymbol{p}_k^T A\boldsymbol{p}_k} = -\frac{\frac{1}{\alpha_k}\boldsymbol{r}_{k+1}^T(\boldsymbol{r}_k-\boldsymbol{r}_{k+1})}{\frac{1}{\alpha_k}\boldsymbol{p}_k^T(\boldsymbol{r}_k-\boldsymbol{r}_{k+1})} =$$

$$= -\frac{(\boldsymbol{r}_{k+1},\boldsymbol{r}_k)-(\boldsymbol{r}_{k+1},\boldsymbol{r}_{k+1})}{(\boldsymbol{p}_k,\boldsymbol{r}_k)-(\boldsymbol{p}_k,\boldsymbol{r}_{k+1})} = \frac{(\boldsymbol{r}_{k+1},\boldsymbol{r}_{k+1})}{(\boldsymbol{r}_k,\boldsymbol{r}_k)}$$

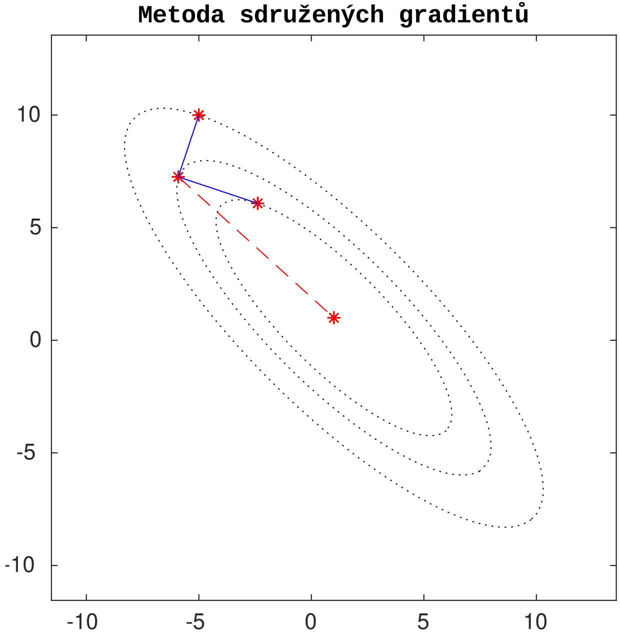
Celkový algoritmus metody sdružených gradientů:

$\boldsymbol{x}_1$ je počáteční iterace, lze použít $\boldsymbol{x}_1 = \boldsymbol{o}$.

$$
\begin{aligned}
\boldsymbol{p}_1 &= \boldsymbol{r}_1 = \boldsymbol{b} - A\boldsymbol{x}_1 \\
pro\ k &= 1, \dots n: \\
\alpha_k &= \frac{(\boldsymbol{r}_k, \boldsymbol{p}_k)}{(\boldsymbol{p}_k, \boldsymbol{p}_k)_A} = \frac{\boldsymbol{p}_k^T \boldsymbol{r}_k}{\boldsymbol{p}_k^T A \boldsymbol{p}_k} \\
\boldsymbol{x}_{k+1} &= \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k \\
\boldsymbol{r}_{k+1} &= \boldsymbol{b} - A\boldsymbol{x}_{k+1} = \boldsymbol{r}_k - \alpha_k A \boldsymbol{p}_k \\
\beta_{k+1} &= \frac{(r_{k+1}, r_{k+1})}{(r_k, r_k)} = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k} \\
\boldsymbol{p}_{k+1} &= \boldsymbol{r}_{k+1} + \beta_{k+1} \boldsymbol{p}_k
\end{aligned}
$$

Algoritmus končíme pokud $\|\boldsymbol{r}_{k+1}\|$ je dostatečně malá.

Odhad chyby: $\|\hat{\boldsymbol{x}} - \boldsymbol{x}_k\|_A \leq 2 \left(\frac{\sqrt{C_p(A)}-1}{\sqrt{C_p(A)}+1} \right)^k \|\hat{\boldsymbol{x}} - \boldsymbol{x}_1\|_A$


Metoda sdružených gradientů
10
5
0
-5
-10
-10
-5
0
5
10

Metoda bikonjugovaných gradientů

Matice soustavy A nemusí být symetrická. Vytváříme psloupnost dvojic reziduí $\boldsymbol{r}_k$, $\tilde{\boldsymbol{r}}_k$, které jsou biortogonální, tj.

$$(\boldsymbol{r}_k, \tilde{\boldsymbol{r}}_l) = 0, \text{ pro } k \neq l, \; (\boldsymbol{r}_k, \tilde{\boldsymbol{r}}_k) \neq 0,$$

a dvojic směrů $\boldsymbol{p}_k$, $\tilde{\boldsymbol{p}}_k$, které jsou bikonjugované, tj.

$$(A\,\boldsymbol{p}_k, \tilde{\boldsymbol{p}}_l) = 0, \text{ pro } k \neq l, \; (A\,\boldsymbol{p}_k, \tilde{\boldsymbol{p}}_k) \neq 0.$$

Algoritmus metody:

$\boldsymbol{v}_1 = \boldsymbol{r}_1 = \boldsymbol{b} - A\boldsymbol{x}_1$

$\tilde{\boldsymbol{r}}_1$ volíme tak, aby $(\boldsymbol{r}_1, \tilde{\boldsymbol{r}}_1) \neq 0$, $\tilde{\boldsymbol{v}}_1 = \tilde{\boldsymbol{r}}_1$.

Pro $k = 1, \ldots n$:

$\alpha_k = \frac{(\boldsymbol{r}_k, \tilde{\boldsymbol{r}}_k)}{(A, \boldsymbol{p}_k, \tilde{\boldsymbol{p}}_k)}$

$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$

$\boldsymbol{r}_{k+1} = \boldsymbol{r}_k - \alpha_k A \boldsymbol{p}_k$, $\tilde{\boldsymbol{r}}_{k+1} = \tilde{\boldsymbol{r}}_k - \alpha_k A^T \tilde{\boldsymbol{p}}_k$

$\beta_k = \frac{(\boldsymbol{r}_{k+1}, \tilde{\boldsymbol{r}}_{k+1})}{(\boldsymbol{r}_k, \tilde{\boldsymbol{r}}_k)}$

$\boldsymbol{p}_{k+1} = \boldsymbol{r}_{k+1} + \beta_k \boldsymbol{p}_k$

$\tilde{\boldsymbol{p}}_{k+1} = \tilde{\boldsymbol{r}}_{k+1} + \beta_k \tilde{\boldsymbol{p}}_k$

Pro metodu bikonjugovaných gradientů je známo jen málo teoretických výsledků. Pro pozitivně definitní matici A je tato metoda ekvivalentní metodě sdružených gradientů za cenu dvojnásobně více výpočtů.

Selhání metody:
$(\boldsymbol{r}_k, \tilde{\boldsymbol{r}}_k) \approx 0$ nebo $(A\,\boldsymbol{p}_k, \tilde{\boldsymbol{p}}_k) \approx 0$.

Věta 13
Buď M maximální počet iterací, během nichž metoda bikonjugovaných gradientů neselže. Pak pro $k = 1, \ldots, M$ platí
$\boldsymbol{r}_k = b - A\boldsymbol{x}_k$

Věta 14
Buď M maximální počet iterací, během nichž metoda bikonjugovaných gradientů neselže. Pak pro $k = 1, \ldots, M$ jsou vektory $\boldsymbol{r}_k$, $\tilde{\boldsymbol{r}}_k$ biortogonální a vektory $\boldsymbol{p}_k$, $\tilde{\boldsymbol{p}}_k$ bikonjugované.

Metoda se dá dobře paralelizovat.

Numerická minimalizace funkce více proměnných

Metoda maximálního spádu

Pro kvadratickou funkci

$$Q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b}$$

je její gradient roven $\nabla Q = Ax - b = -r$, což je směr maximálního růstu. Obecně dává gradient funkce f více proměnných směr maximální růstu, takže při hledání minima funkce můžeme využít směr maximálního spádu $-\nabla f$:

Buď x_1 počáteční iterace, položme $\boldsymbol{v}_1 = -\nabla f(x_1)$ a hledáme minimum funkce f ve směru $\boldsymbol{p}_1$, tj. $\min f(\boldsymbol{x}_1 + t\boldsymbol{p}_1)$ v bodě t_1. Bod t_1 musíme najít nějakou numerickou jednorozměrnou minimalizační metodou, pak $x_2 = \boldsymbol{x}_1 + t_1\boldsymbol{p}_1$ a podobně pokračujeme dál.

Nelineární metoda sdružených gradientů

Fletcherova-Reevesova metoda

Algoritmus:

x_1 je počáteční iterace, $\nabla f_1 = \nabla f(x_1)$, $\boldsymbol{p}_1 = -\nabla f_1$.

Pro $k = 1, 2, \ldots$

najdeme α_k po něž je $f(x_k + \alpha_k \boldsymbol{p}_k)$ minimální,

$x_{k+1} = x_k + \alpha_k \boldsymbol{p}_i$,

určíme $\nabla f_{k+1} = \nabla f(x_{k+1})$ a pak

$\beta_k = \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k}$,

$\boldsymbol{p}_{i+1} = -\nabla f_{k+1} + \beta_k \boldsymbol{p}_k$.