NUMERICKÉ METÓDY LINEÁRNEJ ALGEBRY

06. Gradientné metódy.

Ing. Marek Macák, PhD.

Konzultácie: podľa potreby/dohody

11. Marec 2024

Úvod - Gradientné metódy

- Budeme venovať iteračným metódam založeným na minimalizácii kvadratickej funkcie, ktorej jediným minimom je práve riešenie systému $Ax = b$.
- Pre ľubovoľnú maticu A uvažujme napríklad kvadratickú funkciu

$$J(x) = r^T r = (b - Ax)^T (b - Ax) \tag{1}$$

  ktorá je pozitívne definitná a svoje jediné minimum nadobúda, keď x je riešením systému $Ax = b$.
- V rámci jednoduchosti sa však zameriame len na symetrické a pozitívne definitné matice A.

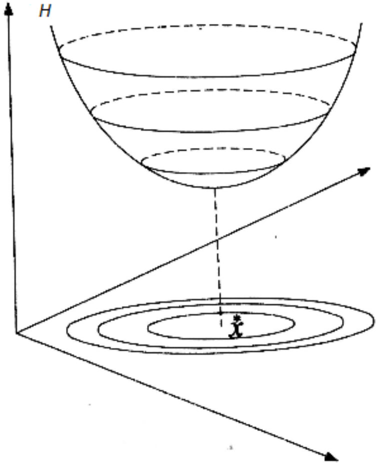
Úvod - Gradientné metódy

- Pre symetrickú a pozitívne definitnú maticu A je riešenie systému $Ax = b$ ekvivalentné minimalizácii kvadratickej funkcie

$$H(x) = \frac{1}{2}x^T Ax - x^T b \tag{2}$$

- Táto kvadratická funkcia potom nadobúda svoje jediné minimum, keď x je riešením systému $Ax = b$.

Úvod - Gradientné metódy



Minimalizácia kvadratickej funkcie H.

Úvod - Gradientné metódy

- Hlavná myšlienka gradientných metód spočíva v zvolení počiatočnej aproximácie $x^{(0)}$ a následnom urcení $x^{(1)}$ pomocou vzdialenosti $\alpha_{(0)}$ v zvolenom smere $v^{(0)}$ od bodu $x^{(0)}$.
- Všeobecný zápis týchto metód je teda tvaru

$$x^{(k)} = x^{(k-1)} + \alpha_{(k-1)} v^{(k-1)}, \quad k = 1, 2, \ldots \tag{3}$$

- Jednotlivé metódy potom dostaneme konkrétnou voľbou smerových vektorov $v^{(k-1)}$.

Metóda najväčšieho spádu

- Smerové vektory $v^{(k-1)}$ zvolíme v smere najväcšieho spádu, t.j. najväcšej zmeny kvadratickej funkcie H.
- Vzdialenosť $\alpha^{(k-1)}$ určime tak, aby funkcia H nadobúdala v bode

$$x^{(k)} = x^{(k-1)} + \alpha^{(k-1)} v^{(k-1)}, \tag{4}$$

svoje minimum v smere $v^{(k-1)}$.

Metóda najväčšieho spádu

Geometricky môžeme metódu interpretovať nasledovne:

- Nech $x^{(0)}$ je počiatočná aproximácia riešenia sústavy $Ax = b$.
- Priamka prechádzajúca bodmi $x^{(0)}$ a $v^{(0)}$ je všeobecne tvaru $x = x^{(0)} + \alpha v^{(0)}, \alpha \in \mathbb{R}$
- Body ležiace na tejto priamke predstavujú krivku

$$H\left(x^{(0)} + \alpha v^{(0)}\right) = \frac{1}{2} x^{(0)T} A x^{(0)} + \alpha v^{(0)T}\left(Ax^{(0)} - b\right) + \frac{1}{2}\alpha^2 v^{(0)T} A v^{(0)} - x^{(0)T} b, \tag{5}$$

na ploche zodpovedajúcej kvadratickej funkcii in $H \in \mathbb{R}^n$.

Metóda najväčšieho spádu

- Hladáme smerový vektor $v^{(0)}$ ktorom má plocha H najväčší spád. Hladáme maximum prvej derivácie funkcie H pre $\alpha = 0$

$$\left.\frac{\partial H(x^{(0)} + \alpha v^{(0)})}{\partial \alpha}\right|_{\alpha=0} = v^{(0)\,T}\left(Ax^{(0)} - b\right) \tag{6}$$

  Skalárny súcin z výrazu je maximálny pre $v^{(0)} = Ax^{(0)} - b = -r^{(0)}$. Teda smer najväčšieho spádu plochy H je opačný vzhľadom k smeru reziduového vektoru $r^{(0)}$.
- Ďalej hladáme na priamke x bod $x^{(1)}$ v ktorom nadobúda plocha H svoje minimum

$$\frac{\partial H(x^{(0)} + \alpha r^{(0)})}{\partial \alpha} = r^{(0)\,T} r^{(0)} + \alpha r^{(0)\,T} A r^{(0)} = 0 \quad \rightarrow \quad \alpha = -\frac{r^{(0)\,T} r^{(0)}}{r^{(0)\,T} A r^{(0)}} \tag{7}$$

Metóda najväčšieho spádu

- Jednoznačnosť minima overíme pomocou druhej derivácie

$$\frac{\partial^2 H(x^{(0)} + \alpha r^{(0)})}{\partial \alpha^2} = r^{(0)T} A r^{(0)} \tag{8}$$

Pre pozitívne definitnú maticu A a $r^{(0)} \neq \mathbf{0}$ je $r^{(0)T} A r^{(0)} > 0$. Funkcia H je teda konvexná vzhľadom k premennej α a minimum, ktoré nadobúda v bode

$$x^{(1)} = x^{(0)} + \frac{r^{(0)T} r^{(0)}}{r^{(0)T} A r^{(0)}} r^{(0)} \tag{9}$$

je jediné.

Metóda najväčšieho spádu

- Všeobecný zápis metódy najväčšieho spádu je potom tvaru

$$x^{(k)} = x^{(k-1)} + \alpha_{k-1} r^{(k-1)} \tag{10}$$

kde

$$\alpha_{k-1} = -\frac{r^{(k-1)T} r^{(k-1)}}{r^{(k-1)T} A r^{(k-1)}}. \tag{11}$$

Metóda združených gradientov

- Je založená na konštrukcii smerových vektorov $v^{(i)}$ tak, aby boli A-ortogonálne.
- Nech $A \in \mathbb{M}_n$ je symetrická a pozitívne definitná matica. Nenulové vektory $v^{(i)}$ a $v^{(j)}$ nazývame združené vzhľadom k A alebo A-ortogonálne, ak platí

$$v^{(j)T} A v^{(i)} = 0, \quad i \neq j \tag{12}$$

- Nenulové vektory $v^{(0)}, v^{(1)}, \dots, v^{(m-1)}$ sú A-ortogonálne, a tiež platí, že sú navzájom lineárne nezávislé.

Metóda združených gradientov

- Takto definované A-ortogonálne vektory $v^{(0)}, v^{(1)}, \ldots, v^{(m-1)}$ tvoria bázu n-rozmerného euklidovského priestoru a presné riešenie $x^\star$ systému $Ax = b$ môžeme vyjadriť ako ich lineárnu kombináciu

$$x^\star = \alpha_0 v^{(0)} + \alpha_1 v^{(1)} + \ldots + \alpha_{n-1} v^{(n-1)} \tag{13}$$

kde

$$\alpha_i = \frac{v^{(i)T} A x^\star}{v^{(i)T} A v^{(i)}} = \frac{v^{(i)T} b}{v^{(i)T} A v^{(i)}}. \tag{14}$$

- Je dôležité poznamenať, že koeficienty α_i teda vieme vypočítať aj bez znalosti presného riešenia $x^\star$.

Metóda združených gradientov

- Nech $v^{(0)}, v^{(1)}, ..., v^{(m-1)}$ sú nenulové A–ortogonálne vektory v $\mathbb{R}^n$. Potom pre ľubovoľnú počiatočnú aproximáciu $x^{(0)} \in \mathbb{R}^n$ má postupnosť $x^{(k)}$ daná vzťahom

$$x^{(k)} = x^{(k-1)} + \alpha_{k-1} v^{(k-1)}, \quad kde \quad \alpha_{k-1} = \frac{v^{(k-1)T} r^{(k-1)}}{v^{(k-1)T} A v^{(k-1)}} \quad a \quad k \geq 1 \qquad (15)$$

nasledujúce vlastnosti:

  - Postupnosť konverguje k presnému riešeniu $x^\star$ systému $Ax = b$ v nanajvýš n krokoch.
  - Kvadratická funkcia $H(x)$ dosahuje svoje minimum v bode $x^\star$.

Metóda združených gradientov

- Ako teda zostrojiť vhodnú postupnosť A-ortogonálnych vektorov?
- Nech $u^{(0)}, u^{(1)}, \ldots, u^{(m-1)}$ tvoria bázu v $\mathbb{R}^n$. Potom postupnosť $\{v^i\}$

$$\begin{aligned} v^{(0)} &= u^{(0)}, && (16) \\ v^{(i)} &= u^{(i)} + \sum_{j=0}^{i-1} \beta_{i,j} v^{(j)}, \quad i = 1, 2, \ldots, n-1. && (17) \end{aligned}$$

kde koeficienty $\beta_{i,j}$ sú tvaru

$$\beta_{i,j} = -\frac{v^{(j)T} A u^{(i)}}{v^{(j)T} A v^{(j)}} \tag{18}$$

Metóda združených gradientov

- Ak zvolíme za postupnosť $\{u^i\}$ postupnosť reziduových vektorov $\{r^i\}$ dostaneme metódu A–združených gradientov. Počiatočnú aproximáciu $x^{(0)} \in \mathbb{R}^n$ máme v tvare

$$v^{(0)} = u^{(0)} = r^{(0)} = b - Ax^{(0)} \tag{19}$$

a pre $i = 1, 2, ...$ platia nasledujúce vzťahy

$$\begin{aligned}
\alpha_{i-1} &= \frac{v^{(i-1)T} r^{(i-1)}}{v^{(i-1)T} A v^{(i-1)}}, & (20)\\
x^{(i)} &= x^{(i-1)} + \alpha_{i-1} v^{(i-1)}, & (21)\\
r^{(i)} &= b - Ax^{(i)} = r^{(i-1)} - \alpha_{i-1} A v^{(i-1)}, & (22)\\
\beta_{i-1} &= -\frac{v^{(i-1)T} A r^{(i)}}{v^{(i-1)T} A v^{(i-1)}}, & (23)\\
v^{(i)} &= r^{(i)} + \beta_{i-1} v^{(i-1)}, & (24)
\end{aligned}$$

Zastavovacie kritéria

- Iteračne metódy vytvára postupnosť $\{x^{(i)}\}$ vektorov konvergujúcich k vektoru $x^\star$ spĺňajúcemu $Ax^\star = b$.
- Aby bol výpočet efektívny, musí sa rozhodnúť, kedy sa má zastaviť.
- Dobré zastavenie kritérium by malo:
  - určiť, kedy je chyba $e^{(i)} = x^{(i)} - x^\star$ dostatočne malá na zastavenie,
  - zastaviť výpočet, ak chyba už neklesá alebo klesá príliš pomaly, a
  - obmedziť maximálny čas strávený výpočtom.

Zastavovacie kritéria

- Používateľ môže zadávať veličiny *maxit*, $||b||$, *stop_tol* a najlepšie aj $||A||$:
  - Celé číslo *maxit* je maximálny počet iterácií, ktoré bude môcť algoritmus vykonať.
  - Reálne číslo $||A||$ je norma A. Každá rozumná (rádovo) aproximácia absolútnej hodnoty najväčšej položky matice A.
  - Reálne číslo $||b||$ je norma b. Opäť platí, že akákoľvek rozumná aproximácia absolútnej hodnoty najväčšej hodnoty vektora b.
  - Reálne číslo *stop_tol* meria, ako malý bude zostatok $r^{(i)} = Ax^{(i)} - b$ konečného riešenia $x^{(i)}$. Jedným zo spôsobov, ako zvoliť *stop_tol*, je približná aproximácia hodnôt v A a b vzhľadom na $||A||$, resp. $||b||$. Napríklad, voľba $stop\_tol = 10^{-6}$ znamená, že používateľ berie do úvahy hodnoty v A a b s chybou v rozsahu $\pm 10^{-6}||A||$ a $\pm 10^{-6}||b||$. *stop_tol* by malo byt väčšie ako strojová presnosť ϵ.

Zastavovacie kritéria

- Algoritmus výpočtu vieme zapísať nasledovne:

$i = 1$
Do $\quad i = i + 1$
  Spočítaj aproximáciu riešenia $x^{(i)}$.
  Spočítaj rezíduum $r^{(i)} = Ax^{(i)} - b$.
  Spočítaj $||r^{(i)}||$ a $||x^{(i)}||$.
until $i \geq maxit$ or $||r^{(i)}|| \leq stop\_tol \cdot (||A|| \cdot ||x^{(i)}|| - ||b||)$.

- Ak $||A||$ nie je k dispozícii, môže byt zastavovacie kritérium nahradene

until $i \geq maxit$ or $||r^{(i)}|| \leq stop\_tol \cdot ||b||$.

- Ak poznáme $||A^{-1}||$

until $i \geq maxit$ or $||r^{(i)}|| \leq stop\_tol \cdot ||x^{(i)}||/||A^{-1}||$.

Zastavovacie kritéria

- Existujú dva prístupy k obmedzeniu nepresnosti vypočítaného riešenia $Ax = b$.
- Keďže $e^{(i)} = x^{(i)} - x^{\star}$, ktorú budeme nazývať dopredná chyba, je ťažké odhadnúť priamo, zavedieme spätnú chybu, ktorá nám umožňuje ohraničiť doprednú chybu.
- Normovaná spätná chyba je definovaná ako najmenšia možná hodnota $max\{||\delta A||/||A||, ||\delta b||/||b||\}$, kde $x^{(i)}$ je presné riešenie $(A + \delta A)x^{(i)} = (b + \delta b)$.
- Spätná chyba sa dá ľahko vypočítať z rezidua $r^{(i)} = Ax^{(i)} - b$.
- Za predpokladu, že máme nejaké ohranicenie pre inverznú hodnotu A, môžeme ohraničiť doprednú chybu v termínoch spätnej chyby prostredníctvom jednoduchej rovnice

$$e^{(i)} = x^{(i)} - x^{\star} = A^{-1}(Ax^{(i)} - b) = A^{-1}r^{(i)}, \tag{25}$$

co vedie na $||e^{(i)}|| \leq ||A^{-1}|| \cdot ||r^{(i)}||$.

Zastavovacie kritéria

- Kriterium:

$$\boxed{\textbf{until } i \geq maxit \text{ or } ||r^{(i)}|| \leq stop\_tol \cdot (||A|| \cdot ||x^{(i)}|| - ||b||).}$$

- Sa rovná požiadavke, aby chyby $||\delta A||$ a $||\delta b||$ spĺňali $||\delta A|| \leq stop\_tol \cdot ||A||$ a $||\delta b|| \leq stop\_tol \cdot ||b||$.
- Toto kritérium dáva ohraničenie doprednej chyby

$$||e^{(i)}|| \leq ||A^{-1}|| \cdot ||r^{(i)}|| \leq stop\_tol \cdot ||A^{-1}|| \cdot (||A|| \cdot ||x^{(i)}|| - ||b||). \tag{26}$$

Zastavovacie kritéria

- Kriterium:

$$\boxed{\textbf{until } i \geq maxit \text{ or } ||r^{(i)}|| \leq stop\_tol \cdot ||b||.}$$

- Sa rovná požiadavke, aby chyby $||\delta A||$ a $||\delta b||$ spĺňali $\delta A = 0$ a $||\delta b|| \leq stop\_tol \cdot ||b||$.
- Jedným z problémov je, že ak $||A|| \cdot ||x|| \geq ||b||$, čo môže nastať len vtedy, ak je A veľmi zle podmienena a x leží v nulovom priestore A. Potom môže byť pre akúkoľvek metódu ťažké splniť toto kritérium zastavenia.
- Na To, že A je zle podmienena vedie na

$$1 << \frac{||A|| \cdot ||x||}{||b||} = \frac{||A|| \cdot ||A^{-1}b||}{||b||} \leq ||A|| \cdot ||A^{-1}|| \tag{27}$$

- Toto kritérium dáva ohraničenie doprednej chyby

$$||e^{(i)}|| \leq ||A^{-1}|| \cdot ||r^{(i)}|| \leq stop\_tol \cdot ||A^{-1}|| \cdot ||b||. \tag{28}$$

Zastavovacie kritéria

- Kriterium:

$$\boxed{\textbf{until } i \geq maxit \text{ or } ||r^{(i)}|| \leq stop\_tol \cdot ||x^{(i)}||/||A^{-1}||.}$$

- Toto kritérium vedie na ohraničenie doprednej chyby

$$\frac{||e^{(i)}||}{||x^{(i)}||} \leq \frac{||A^{-1}|| \cdot ||r^{(i)}||}{||x^{(i)}||} \leq stop\_tol. \tag{29}$$

umožňuje používateľovi určiť požadovanú relatívnu hranicu presnosti *stop_tol* vo vypočítanom riešení $x^{(i)}$.