Podmienenosť problému a stabilita algoritmu

Ing. Gabriel Okša, CSc.

Matematický ústav
Slovenská akadémia vied
Bratislava

Stavebná fakulta STU

Obsah

Vektorové normy

- Norma n-rozmerného vektora je funkcia $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$, ktorá spĺňa nasledujúce tri podmienky pre všetky vektory x a y a skaláry $\alpha \in \mathbb{C}$:
  1. $\|x\| \geq 0$, a $\|x\| = 0$ práve vtedy, ak $x = 0$;
  2. $\|x + y\| \leq \|x\| + \|y\|$ (trojuholníková nerovnosť);
  3. $\|\alpha\, x\| = |\alpha|\, \|x\|$.
- Najdôležitejšie normy: $\|x\|_1 = \sum_{i=1}^n |x_i|$, $\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2\right)^{1/2} = \sqrt{x^\star x}$, $\|x\|_\infty = \max_{1\leq i\leq n} |x_i|$.
- Trieda p-noriem pre $1 \leq p < \infty$: $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$.
- Vážená p-norma: Nech $W = \mathrm{diag}(w_1, \ldots, w_n)$ je diagonálna matica reálnych váh. Potom: $\|x\|_W = \left(\sum_{i=1}^n |w_i\, x_i|^p\right)^{1/p}$.

Maticové normy - 1/3

- Indukovaná maticová norma: Každá matica $A \in \mathbb{C}^{m\times n}$ reprezentuje lineárne zobrazenie $A : \mathbb{C}^n \to \mathbb{C}^m$, t.j. zo svojho definičného oboru do oboru hodnôt. Nech sú v $\mathbb{C}^n$ a $\mathbb{C}^m$ dané vektorové normy $\|\cdot\|_{(n)}$ a $\|\cdot\|_{(m)}$. Potom indukovaná maticová norma $\|A\|_{(m,n)}$ je najmenšie číslo C také, že pre všetky $x \in \mathbb{C}^n$ platí: $\|Ax\|_{(m)} \leq C\,\|x\|_{(n)}$.
- Vďaka vlastnosti č. 3 v definícii vektorovej normy je akcia A na ľubovoľný vektor určená jej akciou na jednotkové vektory. Preto:

$$\|A\|_{(m,n)} = \sup_{x\in\mathbb{C}^n,\, x\neq 0} \frac{\|Ax\|_{(m)}}{\|x\|_{(n)}} = \max_{x\in\mathbb{C}^n,\, \|x\|_{(n)}=1} \|Ax\|_{(m)}.$$

Príklad: $\|A\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}|$ ... max. stĺpcová suma;
$\|A\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{n} |a_{ij}|$ ... max. riadková suma.

Maticové normy - 2/3

- $\|A\|$ nemusí byť indukovaná vektorovou normou, t.j. akciou matice na jednotkové vektory. Maticová norma je funkcia $\|\cdot\| : \mathbb{C}^{m\times n} \to \mathbb{R}$, ktorá spĺňa nasledovné axiómy pre všetky matice A, B a skaláry $\alpha \in \mathbb{C}$:
  1. $\|A\| \geq 0$, a $\|A\| = 0$ práve vtedy, ak $A = 0$;
  2. $\|A + B\| \leq \|A\| + \|B\|$ (trojuholníková nerovnosť);
  3. $\|\alpha A\| = |\alpha| \, \|A\|$;
  4. $\|A B\| \leq \|A\| \, \|B\|$.
- **Príklad**: Frobeniova norma matice nie je indukovaná vektorovou normou:

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}.$$

Maticové normy - 3/3

- Základné vlastnosti vektorových a maticových noriem ($A \in \mathbb{C}^{m\times n},\ x \in \mathrm{C}^n$):

$$\begin{aligned} |x^\star y| &\leq \|x\|_2 \, \|y\|_2 \quad (\mathrm{Cauchy-Schwartz}); \\ \|x\|_\infty &\leq \|x\|_2; \\ \|x\|_2 &\leq \sqrt{n}\, \|x\|_\infty; \\ \|A\|_\infty &\leq \sqrt{n}\, \|A\|_2; \\ \|A\|_2 &\leq \sqrt{m}\, \|A\|_\infty; \end{aligned}$$

- $\|A\|_2 = \sup_{\|x\|_2=1} \|Ax\|_2$ je tzv. 2-norma (tiež spektrálna norma) matice A. Nie je to Frobeniova norma, pretože je to vektorovo indukovaná norma! Pre symetrické matice je $\|A\|_2 = |\lambda_{\max}|$, kde $|\lambda_{\max}|$ je najväčšie vlastné číslo matice A v absolútnej hodnote.

Podmienenosť problému

- Problém, ktorý treba vyriešiť, možno modelovať ako vektorovú funkciu $f: X \to Y$ z normovaného vektorového priestoru vstupných dát X do normovaného vektorového priestoru riešení Y. Obvykle je f nelineárna funkcia, ale je často aspoň spojitá.
- Skúmame chovanie f v danom dátovom vektore $x \in X$ vzhľadom na malé zmeny (perturbácie) bodu x. Keď urobím malú zmenu v x, ako sa táto perturbácia prenesie na riešenie $f(x)$ v Y?
- Dobre podmienený problém má tú vlastnosť, že všetky malé perturbácie bodu x vedú iba k malým zmenám $f(x)$.

Zle podmienený problém:
niektoré malé perturbácie bodu x vedú k veľkým zmenám $f(x)$.

Absolútne a relatívne číslo podmienenosti - 1/3

- Nech δx je malá perturbácia x a $\delta f = f(x + \delta x) - f(x)$. Absolútne číslo podmienenosti $\hat{\kappa} = \hat{\kappa}(x)$ problému f v bode x je definované ako:

$$\hat{\kappa} = \sup_{\delta x} \frac{\|\delta f\|}{\|\delta x\|} \quad \text{pri} \quad \delta x \to 0.$$

- Relatívne číslo podmienenosti $\kappa = \kappa(x)$ problému f v bode x je definované ako:

$$\kappa = \sup_{\delta x} \left(\frac{\|\delta f\|}{\|f(x)\|} \Big/ \frac{\|\delta x\|}{\|x\|} \right) \quad \text{pri} \quad \delta x \to 0.$$

$\hat{\kappa}$ a κ - 2/3

- Ak je f diferencovateľná, definujme Jakobián $J(x)$ funkcie f v bode x ako maticu, ktorej prvok (i, j) je parciálna derivácia $\partial f_i / \partial x_j$ vypočítaná v bode x. Potom:

$$\hat{\kappa} = \|J(x)\|,$$

$$\kappa = \frac{\|J(x)\|}{\|f(x)\|/\|x\|},$$

  kde maticová norma $\|J(x)\|$ je indukovaná vektorovými normami na X a Y.
- κ (rel.) je vhodnejšie ako $\hat{\kappa}$ (abs.), pretože počítanie v pohyblivej rádovej čiarke vnáša do výpočtu relatívne chyby (nie absolútne).
  Dobre podmienená úloha: $\kappa = 1,\ 10,\ 10^2$.
  Zle podmienená úloha: $\kappa = 10^6,\ 10^{12}$.

$\hat{\kappa}$ a κ - 3/3

Príklad 1: Je daná funkcia $f: \mathbb{R}^2 \to \mathbb{R},\ f(x) = x_1 - x_2$, kde $x = (x_1, x_2)^T$.
Číslo podmienenosti κ s použiím ∞-normy v $\mathbb{R}^2$ i v $\mathbb{R}$:
Jakobián: $J(x) = \left[\frac{\partial f}{\partial x_1}\ \frac{\partial f}{\partial x_2}\right] = [1, -1]$, t.j. $\|J(x)\|_\infty = 2$.
$\kappa = \frac{\|J(x)\|_\infty}{\|f(x)\|_\infty / \|x\|_\infty} = \frac{2}{|x_1 - x_2| / \max\{|x_1|, |x_2|\}}$.

Takže κ je veľké, ak $|x_1 - x_2| \approx 0$, t.j. problém je zle podmienený, ak $x_1 \approx x_2$ ("katastrofické odčítanie" dvoch čísiel).

Príklad 2: Nech $x > 0$ a $f(x) = \sqrt{x}$. Zistite κ.
Jakobián: $J(x) = f'(x) = 1/(2\sqrt{x}) = \|J(x)\|_\infty$.
Ďalej: $\|f(x)\|_\infty = \sqrt{x},\ \|x\|_\infty = x$ (lebo $x > 0$).
Takže:
$\kappa = \frac{\|J(x)\|_\infty}{\|f(x)\|_\infty / \|x\|_\infty} = \frac{1/(2\sqrt{x})}{\sqrt{x}/x} = \frac{1}{2}$, t.j. problém výpočtu $\sqrt{x}$ je dobre podmienený pre všetky $x > 0$.

Problém versus algoritmus

- Zopakujme, že matematický problém je zobrazenie $f: X \to Y$ z vektorového priestoru dát X do vektorového priestoru riešení Y.
- Algoritmus na riešenie f možno chápať ako iné zobrazenie $\tilde{f}: X \to Y$ medzi tými istými vektorovými priestormi.
- Spresnenie definície algoritmu: Nech je pevne daný problém f, počítač s floating-point aritmetikou podľa IEEE FPS (1985), algoritmus pre riešenie problému f a implementácia tohto algoritmu vo forme programu na danom počítači. Zoberme vstupné dáta (vektor) $x \in X$, vložme ich do počítača ako vstup pre program (t.j. v tom momente ich zaokrúhlime: $\mathrm{fl}(x_i) = x_i(1+\epsilon_i)$, $|\epsilon_i| \leq \mu$). Potom vykonajme program v počítači. Výsledkom je množina vypočítaných floating-point čísiel z Y, ktorú označíme $\tilde{f}(x)$.

Presnosť algoritmu

- Na výstup algoritmu $\tilde{f}(x)$ vplývajú minimálne zaokrúhľovacie chyby počas behu programu, ktorý implementuje daný algoritmus. Dokonca na dvoch rôznych počítačoch a na rovnakých vstupných dátach môže ten istý algoritmus (program) dať rôzne výsledky.
- Dobrý algoritmus pre problém f by mal dať "presný" výsledok.

  Absolútna chyba výpočtu: $\|\tilde{f}(x) - f(x)\|$ by mala byť malá pre každé $x \in X$.

  Relatívna chyba výpočtu: $\frac{\|\tilde{f}(x)-f(x)\|}{\|f(x)\|}$ by mala byť malá pre každé $x \in X$, t.j. $= O(\mu)$ ("rádovo μ"), kde μ je jednotka zaokrúhľovania.
- Pre zle podmienené problémy je táto požiadavka príliš silná. Už samotné zaokrúhlenie vstupných dát môže podstatne zmeniť výsledok.

Stabilita algoritmu

Namiesto požiadavky na vysokú relatívnu presnosť výpočtu nastupuje požiadavka na stabilitu algoritmu.

Algoritmus je stabilný, ak pre každý vstup $x \in X$ platí

$$\frac{\|\tilde{f}(x) - f(\tilde{x})\|}{\|f(\tilde{x})\|} = O(\mu)$$

pre nejaké $\tilde{x}$ relatívne blízke k x:

$$\frac{\|\tilde{x} - x\|}{\|x\|} = O(\mu).$$

Slovne: Stabilný algoritmus dáva takmer správny výsledok pre takmer správne vstupné dáta.

Spätná stabilita algoritmu

Mnohé algoritmy v NLA spĺňajú podmienku, ktorá je silnejšia a jednoduchšia ako stabilita.

Algoritmus $\tilde{f}$ pre problém f je spätne stabilný ('backward stable'), ak pre každý vstup $x \in X$ platí:

$$\tilde{f}(x) = f(\tilde{x})$$

pre nejaké $\tilde{x}$ relatívne blízke k x:

$$\frac{\|\tilde{x} - x\|}{\|x\|} = O(\mu).$$

Slovne: Spätne stabilný algoritmus dáva presne správny výsledok pre takmer správne vstupné dáta.

Zmysel symbolu $O(\mu)$ - 1/2

- Označenie

$$\phi(t) = O(\psi(t))$$

znamená, že existuje konštanta $C > 0$ tak, že pre t dostatočne blízko ku stanovenej limite (napr. $t \to 0$ alebo $t \to \infty$) platí: $|\phi(t)| \leq C\,\psi(t)$.

Príklad: $\sin^2 t = O(t^2)$ pri $t \to 0$ znamená, že existuje $C > 0$ tak, že pre dostatočne malé t je $|\sin^2 t| \leq C\,t^2$.

- Iný druh označenia: $\phi(s,t) = O(\psi(t))$ rovnomerne vzhľadom na s.
Znamená to, že konštanta $C > 0$ nezávisí od s.

Príklad: $(\sin^2 t)\,(\sin^2 s) = O(t^2)$ rovnomerne vzhľadom na s pri $t \to 0$.

Zmysel symbolu $O(\mu)$ - 2/2

- Takže v NLA tvrdenie: $\|\text{computed}\| = O(\mu)$ znamená tri veci:
  1. $\|\text{computed}\|$ reprezentuje normu vypočítaného výsledku pri použití algoritmu $\tilde{f}$ na problém f, pričom výsledok závisí od vstupných dát $x \in X$ a od μ.
  2. Implicitný limitný proces je $\mu \to 0$. To znamená, že ak by algoritmus bežal na viacerých počítačoch s hodnotou μ klesajúcou k nule, potom $\|\text{computed}\|$ musí klesať proporcionálne s μ (alebo rýchlejšie).
  3. Označenie $O(\cdot)$ platí rovnomerne vzhľadom na $x \in X$.

  Príklad: Nech vypočítané riešenie $\tilde{x}$ lin. sústavy $Ax = b$ s regulárnou maticou A rádu m má relatívnu chybu: $\frac{\|\tilde{x}-x\|}{\|x\|} = O(\kappa(A)\,\mu)$, kde $\kappa(A) = \|A\|\,\|A^{-1}\|$ je tzv. číslo podmienenosti matice A. Konštanta $C > 0$ nezávisí od A a b, ale závisí od m, pretože zmena m znamená zmenu dimenzie priestoru vstupov X a priestoru riešení Y!

Presnosť versus spätná stabilita

- Je spätne stabilný algoritmus aj presný? Závisí to od čísla podmienenosti $\kappa(x)$ problému.
- **Veta**: Nech je na riešenie problému $f: X \to Y$ s číslom podmienenosti $\kappa(x)$ použitý spätne stabilný algoritmus $\tilde{f}: X \to Y$, ktorý je implementovaný na počítači spĺňajúcom IEEE Floating Point Standard (1985). Potom relatívna chyba výpočtu je:

$$\frac{\|\tilde{f}(x) - f(x)\|}{\|f(x)\|} = O(\kappa(x)\,\mu).$$

- Takže ak je $\kappa(x)$ malé, potom výsledok výpočtu bude presný v relatívnom zmysle. Ak je $\kappa(x)$ veľké, potom relatívna presnosť výsledku môže byť nízka (t.j. nemáme zaručenú vysokú relatívnu presnosť).

Príklady

1. Spätná stabilita odčítania dvoch floating-point čísiel:
$f(x_1, x_2) = x_1 - x_2 : \mathbb{C}^2 \to \mathbb{C}$, takže algoritmus pre f je
$\tilde{f}(x_1, x_2) = \mathrm{fl}[\mathrm{fl}(x_1) - \mathrm{fl}(x_2)]$;
$\mathrm{fl}(x_1) = x_1\,(1+\epsilon_1),\ \mathrm{fl}(x_2) = x_2\,(1+\epsilon_2),\ |\epsilon_i| \le \mu$;
$\mathrm{fl}[\mathrm{fl}(x_1) - \mathrm{fl}(x_2)] = [x_1\,(1+\epsilon_1) - x_2\,(1+\epsilon_2)]\,(1+\epsilon_3) =$
$x_1\,(1+\epsilon_1)\,(1+\epsilon_3) - x_2\,(1+\epsilon_2)\,(1+\epsilon_3) =$
$x_1\,(1+\epsilon_1+\epsilon_3+\epsilon_1\epsilon_3) - x_2\,(1+\epsilon_2+\epsilon_3+\epsilon_2\epsilon_3) =$
$x_1\,(1+\epsilon_4) - x_2\,(1+\epsilon_5), \quad |\epsilon_4|, |\epsilon_5| \le 2\mu + O(\mu^2)$.
Takže $\tilde{f}(x_1, x_2)$ sa presne rovná rozdielu $\tilde{x}_1 - \tilde{x}_2$, kde:

$$\frac{|\tilde{x}_1 - x_1|}{|x_1|} = O(\mu), \quad \frac{|\tilde{x}_2 - x_2|}{|x_2|} = O(\mu),$$

a akékoľvek $C > 2$ je použiteľné vnútri "$O(\cdot)$" symbolov. Potom pre euklidovskú normu na $\mathbb{C}^2$:
$\frac{\|\tilde{x} - x\|}{\|x\|} = O(\mu)$.

Príklady

2. Je pravda alebo nie?
a) $\sin x = O(1)$ pre $x \to +\infty$;
b) $\sin x = O(1)$ pre $x \to 0$;
c) $\ln x = O(x^{1/100})$ pre $x \to +\infty$;
d) $n! = O((n/e)^n)$ pre $n \to +\infty$;
e) $S = O(V^{2/3})$ pre $V \to +\infty$, kde S je plocha a V je objem gule.

3. Nech pre $x \in \mathbb{R}$ sa súčet $1 + x$ počíta pomocou algoritmu $\tilde{f}(x) = \mathrm{fl}(1 + \mathrm{fl}(x))$ (t.j. číslo 1 je v počítači reprezentované presne). Ukážte, že pre tento algoritmus je absolútna chyba $O(\mu)$, ale algoritmus nie je spätne stabilný. (Návod: uvažujte prípad $|x| \approx 0$).