Predavanje 1

Uvod v numerične metode

Prvi sklop izročkov

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

4. oktober 2021

Viri in obveznosti

Viri:

- Bojan Orel, Osnove numerične matematike, Založba FE in FRI.
- Bor Plestenjak: Razširjen uvod v numerične metode, DMFA založništvo.

Obveznosti:

- Predavanja: 3 ure na teden
- Vaje: 2 uri na teden
- 3 domače naloge: 50% ocene
- Pisni izpit iz teorije: 50% ocene

Vsebina predmeta

1. Računanje in vloga napak pri numerični matematiki
2. Reševanje sistemov linearnih enačb
   - Gausova eliminacija in LU razcep - cena in problemi
   - Pivotiranje
   - Iterativne metode - Jacobijeva in Seidlova iteracija
3. Reševanje nelinearnih enačb
   - Tangentna oz. Newtonova metoda
   - Metoda fiksne točke
4. Aproksimacija in interpolacija
   - Lagrangeov in Newtonov interpolacijski polinom
   - Aproksimacija po metodi najmanjših kvadratov
5. Numerično odvajanje in integriranje
   - Trapezna metoda
   - Simpsonova metoda
   - Rombergova metoda
6. Numerično reševanje diferencialnih enačb
   - Eulerjeva metoda
   - Runge-Kutta metode

Numerično in simbolno računanje

Numerično računanje:

- Takoj v formulo vstavljamo števila
- Pridemo do numeričnega rezultata - numerične rešitve

Simbolno računanje:

- simboli predstavljajo števila
- izraz preoblikujemo s simbolnim račuanjem do novega simbolnega izraza - analitična rešitev

Primer

- *Numerično:*

$$\frac{(17.36)^2 - 1}{17.36 + 1} = 16.36; \quad 0.25,\ 0.33333\ldots(?),\ 3.14159\ldots(?)$$

- *Simbolno:*

$$\frac{x^2 - 1}{x + 1} = x - 1; \quad \frac{1}{4},\ \frac{1}{3},\ \pi,\ \tan 83$$

Kaj zanima numerično matematiko?

Metoda. . . matematična konstrukcija, s katero rešujemo problem

Algoritem. . . koraki metode

Implementacija. . . zapis algoritma v izbranem jeziku

Kaj pomeni 'biti numerično dober'?

majhna sprememba podatkov $\Rightarrow$ majhna napaka rezultata

Tipična vprašanja numerične matematike:

- Ali je problem občutljiv?
- Ali je metoda 'dobra'?
- Ali je algoritem robusten - deluje na širokem spektru problemov?
- Ali je implementacija hitra - časovna in prostorska zahtevnost?

Občutljivih problemov NM ne more rešiti

Problem je **občutljiv** , če se ob majhni spremembi začetnih podatkov točen rezultat zelo spremeni.

Občutljivost je odvisna le od narave problema in ne od izbrane numerične metode.

Primer (presečišča premic)

Sistem in njegova perturbacija

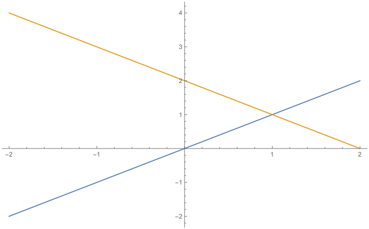
$$\begin{aligned} x + y &= 2 \quad &\to \quad x + y &= 1.9999 \\ x - y &= 0 \quad &\to \quad x - y &= 0.0002 \end{aligned}$$

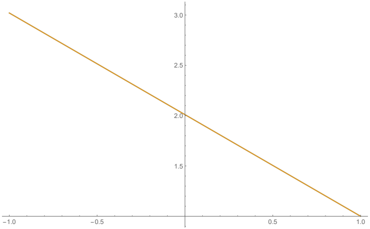
ima rešitvi $x = y = 1$ *oz.* $x = 1.00005$ *in* $y = 0.99985$. *Problem je neobčutljiv, saj je šlo za spremembo za isti velikostni razred.*

Sistem in njegova perturbacija

$$\begin{aligned} x + 0.99y &= 1.99 \quad \rightarrow \quad & x + 0.99y &= 1.9899 \\ 0.99x + 0.98y &= 1.97 \quad \rightarrow \quad & 0.99x + 0.98y &= 1.9701 \end{aligned}$$

ima rešitvi $x = y = 1$ oz. $x = 2.97$ in $y = -0.99$. Problem je občutljiv, saj je majhna sprememba začetnih podatkov povzročila veliko spremembo rezultata.





Na čem temeljijo numerične metode?

- Neskončne procese nadomestimo s končnimi (Taylorjeva vrsta) .
- Neskončno razsežne prostore nadomestimo s končno razsežnimi (funkcije nadomestimo s polinomi).
- Diferencialne enačbe nadomestimo z algebraičnimi (znebimo se vseh parcialnih odvodov iz enačb).
- Nelinearne probleme nadomestimo z linearnimi (linearna aproksimacija v točki).
- Matrike nadomestimo z enostavnejšimi (upoštevamo samo zgornjetrikotni del).

Zakaj sploh potrebujemo numerično matematiko?

Znanost, ki temelji na matematičnih izračunih, je neposredno odvisna od NM.

Nekatere katastrofe so se zgodile zaradi slabega numeričnega računanja (`http://www-users.math.umn.edu/~arnold//disasters/`):

- *Nesreča Misije Patriot, Zalivska vojna 1991, Savdska Arabija*, 28 žrtev: slaba analiza zaokrožitvenih napak.

  Čas zadetka iraške rakete, usmerjene na Savdsko Arabijo, je bil računan na vsako desetino sekunde v 24-bitnem sistemu. Ker velja

  $$\frac{1}{10} = 2^{-4} + 2^{-5} + 2^{-8} + 2^{-9} + 2^{-12} + 2^{-13} + 2^{-16} + 2^{-17} + 2^{-20} + 2^{-21} + \underbrace{+2^{-24} + 2^{-25} + 2^{-28} + \ldots}_{\text{zanemarimo}},$$

  je vsako desetinko sekunde napaka $9.5 \cdot 10^{-8}\ \mathrm{s}$. Po 100 urah računanja je bila napaka $9.5 \cdot 10^{-8}\ \mathrm{s} \cdot 100 \cdot 60 \cdot 60 \cdot 10 = 0.34\ \mathrm{s}$. Ker je hitrost rakete $1.676\ \mathrm{km/s}$, je bila pozicija rakete za več kot $500\ \mathrm{m}$ napačno predvidena in je ta ušla radarjem.

- *Eksplozija rakete Ariana 5, Francoska Gvajana, 1996*: posledica prekoračitve obsega števil.
  `https://www.youtube.com/watch?v=PK_yguLapgA`
  `https://www.youtube.com/watch?v=W3YJeoYgozw`

  Ob prenovi rakete so 'pozabili' nadgraditi uporabljen številski sistem, ki je horizontalno hitrost meril v 16-bitnem sistemu (1 bit porabimo za predznak). Največja hitrost v tem sistemu je

  $$2^0 + 2^1 + \ldots + 2^{13} + 2^{14} = \frac{2^{15} - 1}{2 - 1} = 32767.$$

  Ker je prenovljena raketa po 37 sekundah presegla to hitrost, je prišlo do zaustavitve motorjev...

- *Potop naftne ploščadi Sleipner A, Stavanger, Norveška, 1991*, miljarda dolarjev škode: nenatančna obdelava obremenitev pri reševanju PDE-jev.
  `https://www.youtube.com/watch?v=eGdiPs4THW8`

Ponovitev predstavljivih števil

Števila shranjujemo v obliki

$$\boxed{x = \pm 0.d_1 d_2 d_3 \dots d_m \times \beta^e,}$$

kjer je

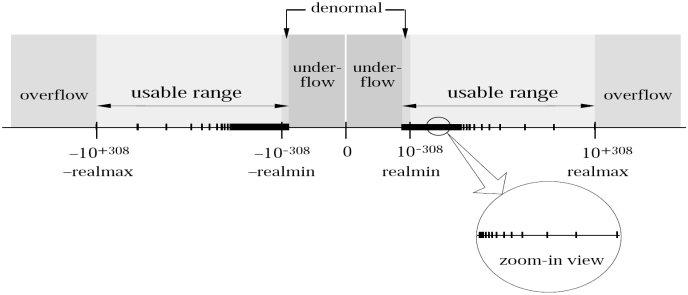
- β naravno število (v računalništvu $\beta = 2$),
- $d_1 d_2 d_3 \dots d_m$ mantisa, e eksponent.

Primer (baza 10)

- *1000.12345 zapišemo kot* $+(0.100012345)_{10} \times 10^4$.
- *0.000812345 zapišemo kot* $+(0.812345)_{10} \times 10^{-3}$.

Prekoračitev in podkoračitev

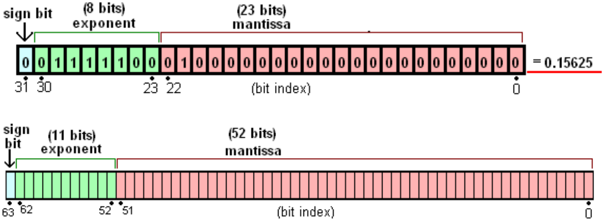
Floating Point Number Line



- izračuni preblizu 0 lahko povzročijo **podkoračitev**
- preveliki izračuni lahko povzročijo **prekoračitev**
- prekoračitev je v splošnem hujši problem

IEEE standard

- IEEE Enojna natančnost: števila so predstavljena z 32 biti.
- IEEE Dvojna natančnost: števila so predstavljena z 64 biti.



Kaj so zaokrožitvene napake?

- Večine realnih števil ne moremo predstaviti v strojni aritmetiki $\Rightarrow$ zaokrožujemo in delamo zaokrožitvene napake.

- IEEE standard... zaokroži x do najbližjega predstavljivega števila $\mathrm{fl}(x)$. Naj bosta $x_- \leqslant x \leqslant x_+$ najbližji predstavljivi števili števila x. Potem je

$$\boxed{\mathrm{fl}(x) = \begin{cases} x_-, & \text{če je } x - x_- < x_+ - x, \\ x_+, & \text{če je } x - x_- \geqslant x_+ - x. \end{cases}}$$

- Kako velika je napaka? Recimo, da je x bližje x_-:

$$\begin{aligned} x &= (0.1b_2b_3\ldots b_mb_{m+1})_2 \times 2^e, \\ x_- &= (0.1b_2b_3\ldots b_m)_2 \times 2^e, \\ x_+ &= \left((0.1b_2b_3\ldots b_m)_2 + 2^{-m}\right) \times 2^e, \\ x - x_- &\leqslant \frac{x_+ - x_-}{2} = 2^{e-m-1}, \end{aligned}$$

$$\frac{x - x_-}{x} \leqslant \frac{2^{e-m-1}}{1/2 \times 2^e} \leqslant \underbrace{2^{-m}}_{u} \ldots$$ **osnovna zaokrožitvena napaka**

Torej je

$$x_- = x_- - x + x \geqslant -ux + x = x(1-u).$$

Podobno $x_+ \leqslant x(1+u)$. Sledi

$$\boxed{\mathrm{fl}(x) = x(1+\delta)}, \quad \text{kjer je } |\delta| < u.$$

Analogna izpeljava velja v primeru, ko je x negativen.

Kako računamo s predstavljivimi števili?

Za predstavljivi števili x, y in katerokoli od osnovnih operacij $\odot \in \{+, -, \cdot, :\}$ število $x \odot y$ ni nujno predstavljivo. Po zgornjem pa velja

$$\boxed{\mathrm{fl}(x \odot y) = (x \odot y)(1 + \delta)}, \quad \text{kjer je } |\delta| \leqslant u.$$

Seštevanje numerično ni asociativna operacija, tj.

$$\boxed{(a + b) + c \neq a + (b + c)}:$$

$$\begin{aligned}
(a+b)+c &= \mathrm{fl}(\mathrm{fl}(a+b)+c) = \mathrm{fl}((a+b)(1+\delta_1)+c) \\
&= [(a+b)(1+\delta_1)+c]\,(1+\delta_2) \\
&= [(a+b+c)+(a+b)\delta_1)]\,(1+\delta_2) \\
&= (a+b+c)\left[1+\frac{a+b}{a+b+c}\delta_1(1+\delta_2)+\delta_2\right]
\end{aligned}$$

Podobno

$$a+(b+c) = (a+b+c)\left[1+\frac{b+c}{a+b+c}\delta_3(1+\delta_4)+\delta_4\right].$$

Če pozabimo na člena $\delta_1\delta_2$ in $\delta_3\delta_4$, dobimo

$$(a+b)+c = (a+b+c)(1+\epsilon_3) \quad \text{kjer je} \quad \epsilon_3 \approx \frac{a+b}{a+b+c}\delta_1 + \delta_2,$$
$$a+(b+c) = (a+b+c)(1+\epsilon_4) \quad \text{kjer je} \quad \epsilon_4 \approx \frac{b+c}{a+b+c}\delta_3 + \delta_4.$$

Sklep: Ko seštevamo števila, je za čim manjšo napako najbolje začeti z najmanjšim in prištevati večje.

Napake pri numeričnem računanju

- **Neodstranljiva napaka** D_n ... nenatančni začetni podatki.
- **Napaka metode** D_m ... npr. neskončni proces aproksimiramo s končnim.
- **Zaokrožitvena napaka** D_z ... računanje s približki in zaokroževanje.

Celotna napaka D je

$$\boxed{D = D_n + D_m + D_z}.$$

Primer ($\sin \frac{\pi}{10}$ računamo v desetiškem sistemu z $m = 4$)

- D_n: $fl(\frac{\pi}{10}) = 0.3142 \cdot 10^0$. *Ocenimo:* $|D_n| \approx \sin'(\frac{\pi}{10})|x - fl(x)| \leqslant \frac{1}{2} \cdot 10^{-4}$.
- D_m : $\sin x \approx x - x^3/6$. *Ocenimo:* $|D_m|| \leqslant x^5/120 = 2.6 \cdot 10^{-5}$.
- D_z: $fl(x - fl(fl(fl(x \cdot x) \cdot x)/6))$. *Ocenimo:* $|D_z| \leqslant 3.0 \cdot 10^{-5}$.

Stabilnost meri kakovost metode

Stabilnost metode preverimo z **analizo zaokrožitvenih napak**.

Vrste napak (x naj bo točna vrednost, $\bar{x}$ pa približek zanjo):

- Prva delitev:
  - **Absolutna napaka** : $\boxed{\bar{x} - x}$.
  - **Relativna napaka** : $\boxed{\dfrac{\bar{x} - x}{x}}$.
- Druga delitev:
  - **Direktna napaka**: Numerična napaka rezultata.
  - **Obratna napaka**: Koliko je potrebno spremeniti začetne podatke, da dobimo izračunan rezultat.

Velja

$$\boxed{|\text{direktna napaka}| \approx \text{občutljivost} \times |\text{obratna napaka}|}.$$

Izračunana vrednost je blizu pravi, če rešujemo neobčutljiv problem z obratno stabilno metode.

Odštevanje in seštevanje sta lahko 'katastrofalni'

odštevanje dveh približno enakih števil

seštevanje dveh približno nasprotnih števil

$$a = x.xxxx\,xxxx\,xxx1\ \overbrace{ssss\ldots}^{\text{izguba}}$$

$$b = x.xxxx\,xxxx\,xxx0\ \overbrace{tttt\ldots}^{\text{izguba}}$$

Potem

$$\begin{array}{rll} & \overbrace{x.xxx\,xxxx\,xxx1}^{\text{končna natančnost}} & \\ - & x.xxx\,xxxx\,xxx0 & \\ \hline = & 0.000\,0000\,0001 & ????\,???? \\ = & 1.\ \underbrace{????\,????}_{\text{izguba natančnosti}}\ \cdot \beta^{-m} & \end{array}$$

S ponavljanjem se napake seštevajo.