

Sveučilišta J. J. Strossmayera u Osijeku

Odjel za matematiku

Rudolf Scitovski

# **N U M E R I Č K A   M A T E M A T I K A**

**Drugo izdanje**

Osijek, 2004.

Dr. sc. Rudolf Scitovski, Numerička matematika

Odjel za matematiku

Sveučilište u Osijeku

Gajev trg 6

HR-31 000 Osijek

e-mail: `scitowsk@mathos.hr`

Izdavač:

Odjel za matematiku Sveučilišta u Osijeku

Recenzenti:

dr. sc. Dragan Jukić, Odjel za matematiku Sveučilišta u Osijeku

dr. sc. Miljenko Marušić, PMF – Matematički odjel, Zagreb

Lektor:

Ivanka Ferčec, Elektrotehnički fakultet Osijek

**CIP – Katalogizacija u publikaciji**  
**Gradska i sveučilišna knjižnica, Osijek**

UDK 517(075.8)

UDK 512(075.8)

**SCITOVSKI, Rudolf**

Numerička matematika/ Rudolf Scitovski, –  
Osijek : Odjel za matematiku Sveučilišta u Osijeku, 1999.

- 173 str.:

73 ilustr. ; 24 cm

Na vrhu nasl. str.: Sveučilišta J. J. Strossmayera  
u Osijeku. - Bibliografija: str. 168-170.

ISBN 953 – 6032 – 24 – 4

990225022

Ovaj udžbenik objavljuje se uz suglasnost Senata Sveučilišta J. J. Strossmayera u Osijeku  
pod brojem xx/04

© Rudolf Scitovski, 2004.

Tisak: Grafika d.o.o., Osijek

## PREDGOVOR DRUGOM IZDANJU

U posljednjih 5 godina prema ovom udžbeniku u većoj ili manjoj mjeri izvodila se nastava iz područja numeričke matematike na Odjelu za matematiku, Elektrotehničkom fakultetu, Strojarskom fakultetu i Građevinskom fakultetu Sveučilišta u Osijeku, te na Veleučilištu u Karlovcu. Novo izdanje korigirano je i dopunjeno u gotovo svim poglavljima. Posebno su provjereni i korigirani rezultati pratećih zadataka. Zahvaljujem se kolegama sa Odjela za matematiku Sveučilišta u Osijeku (posebno dr. sc. N. Truharu i mr. sc. K. Sabi i D. Marković) i Matematičkog odjela PMF-a Sveučilišta u Zagrebu koji su me upozorili na neke tiskarske i druge pogreške.

Osijek, rujna, 2004.

Rudolf Scitovski

## PREDGOVOR

Obzirom da je ovaj udžbenik kompletno pripremljen za digitalni tisak, bilo je moguće najprije tiskati manji broj primjeraka, a onda nakon godinu dana ponoviti izdanje uz korekcije naknadno uočenih pogrešaka na koje su me upozorili suradnici ili studenti: Z. Drmač, H. Glavaš, D. Ivanović, J. Job, T. Marošević, M. Pešut, K. Sabo, R. Sotirov, na čemu im se iskreno zahvaljujem.

Gotovo uvijek u primijenjenim istraživanjima umjesto stvarne fizikalne slike promatranog problema analiziramo nekakvu njegovu idealizaciju nastalu: zanemarivanjem nekih faktora, linearizacijom nelinearnog problema, uzimanja približnih vrijednosti ulaznih veličina u matematičkom modelu, itd. I pored toga, rješavanje problema može biti složen matematički i numerički problem, a vrlo često bez pomoći računala ne može se ni zamisliti njegova realizacija. U ovom tekstu opisane su numeričke metode, koje se najčešće koriste u ovakvim situacijama. Na taj način uvijek dobivamo neku aproksimaciju stvarnog rješenja. Budući da stvarno rješenje najčešće ne znamo, važno je znati barem ocjenu veličine greške, koju smo pri tome učinili.

Ovaj tekst pisan je tako da podrazumijeva poznavanje osnova matematičke analize, linearne algebre i programiranja, a namijenjen je studentima viših i visokih škola inženjerskog usmjerenja.

Tekst sadrži puno primjera koji mogu doprinijeti razumijevanju izložene materije. Na kraju svakog poglavlja nalaze se brojni zadaci: od sasvim jednostavnih, do onih koji mogu poslužiti kao teme seminarskih i sličnih radova. Gdje god je to moguće dana su rješenja zadataka, a kod nekih i odgovarajuće upute za rješavanje. Izrada većine zadataka povezana je s mogućnošću korištenja nekih gotovih programa (MATHEMATICA, MATLAB i sl.), kao i programa navedenih u ovoj knjizi. Predviđeno je da student sam prilagođava navedene programe i konstruira nove u skladu sa zadanim problemom. Većina programa urađena je *Mathematica*-programom jer je Ministarstvo znanosti i tehnologije omogućilo posjedovanje legalne licence ovog programa svakoj visokoškolskoj instituciji.

Sadržaj naveden u ovoj knjizi, kao i priloženi programi, mogu poslužiti i u nekim praktičnim istraživanjima. U posebnim primjedbama navedene su i neke druge mogućnosti, koje se mogu pojaviti u praktičnom radu. Također, dane su i odgovarajuće upute za daljnje samostalno usavršavanje. U tom smislu, na kraju svakog poglavlja u posebnoj primjedbi čitatelj se upućuje na odgovarajuću noviju literaturu iz tog područja. Poklavlje 9. *Numeričko rješavanje parcijalnih diferencijalnih jednadžbi* napisano je prema radu GALIĆ (1989).

Zahvaljujem se recenzentima dr. sc. D. Jukiću i dr. sc. M. Marušiću, te lektoru Ivanki Ferčec na korisnim sugestijama i prijedlozima u cilju pojednostavljivanja i popravljivanja ovog teksta. Također, zahvaljujem se kolegama dr. sc. M. Benšić, dr. sc. Z. Drmaču, dr. sc. B. Guljašu, dr. sc. T. Maroševiću, mr. sc. K. Seršiću, R. Sotirov na korisnim primjedbama. Posebno se zahvaljujem asistentu Elektrotehničkog fakulteta Alfonzu Baumgartneru za obradu teksta u L<sup>A</sup>T<sub>E</sub>X-u i kolegici Renati Sotirov za izradu ilustracija.

Osijek, travnja, 2000.

Rudolf Scitovski

## Sadržaj:

<b>1. Pogreške</b>	<b>1</b>
1.1 Vrste pogrešaka . . . . .	1
1.2 Apsolutna i relativna pogreška . . . . .	4
1.3 Signifikantne znamenke . . . . .	6
1.4 Pogreške kod izračunavanja vrijednosti funkcije . . . . .	8
1.5 Inverzni problem u teoriji pogrešaka . . . . .	10
1.6 Zadaci . . . . .	12
<b>2. Interpolacija. Spline interpolacija</b>	<b>16</b>
2.1 Interpolacija . . . . .	16
2.1.1 Lagrangeov oblik interpolacijskog polinoma . . . . .	18
2.1.2 Newtonov oblik interpolacijskog polinoma . . . . .	20
2.1.3 Ocjena pogreške . . . . .	23
2.2 Spline interpolacija . . . . .	25
2.2.1 Linearni interpolacijski spline . . . . .	26
2.2.2 Kubični interpolacijski spline . . . . .	28
2.3 Zadaci . . . . .	33
<b>3. Rješavanje sustava linearnih jednadžbi</b>	<b>38</b>
3.1 Norma vektora i matrice . . . . .	39
3.2 Uvjetovanost sustava linearnih jednadžbi . . . . .	41
3.3 Rješavanje trokutastih sustava . . . . .	43
3.4 Gaussova metoda eliminacije . . . . .	44
3.5 LU-dekompozicija . . . . .	47
3.6 Cholesky - dekompozicija . . . . .	51
3.7 QR-dekompozicija . . . . .	52
3.8 Iterativne metode . . . . .	57
3.9 Dekompozicija na singularne vrijednosti . . . . .	60

3.10 Zadaci . . . . .	63
<b>4. Rješavanje nelinearnih jednadžbi</b>	<b>67</b>
4.1 Metoda bisekcije . . . . .	71
4.2 Metoda jednostavnih iteracija . . . . .	72
4.3 Newtonova metoda . . . . .	77
4.3.1 Modifikacije Newtonove metode . . . . .	82
4.4 Rješavanje sustava nelinearnih jednadžbi . . . . .	85
4.5 Zadaci . . . . .	88
<b>5. Aproksimacija funkcija</b>	<b>93</b>
5.1 Najbolja $L_2$ aproksimacija . . . . .	94
5.1.1 Ortogonalni polinomi . . . . .	97
5.1.2 Čebiševljevi polinomi . . . . .	98
5.2 Najbolja $L_\infty$ aproksimacija . . . . .	100
<b>6. Problemi najmanjih kvadrata</b>	<b>103</b>
6.1 Linearni problemi najmanjih kvadrata . . . . .	108
6.1.1 Rješavanje LPNK preko sustava normalnih jednadžbi . . . . .	109
6.1.2 Rješavanje LPNK pomoću QR-dekompozicije . . . . .	110
6.1.3 Rješavanje LPNK pomoću dekompozicije na singularne vrijednosti . . . . .	111
6.2 Nelinearni problemi najmanjih kvadrata . . . . .	115
6.2.1 Gauss-Newtonova metoda . . . . .	117
6.2.2 Marquardtova metoda . . . . .	120
6.3 Zadaci . . . . .	121
<b>7. Numerička integracija</b>	<b>125</b>
7.1 Trapezno pravilo . . . . .	125
7.2 Newton-Cotesova formula . . . . .	130
7.3 Simpsonovo pravilo . . . . .	130
7.4 Zadaci . . . . .	134
<b>8. Numeričko rješavanje običnih diferencijalnih jednadžbi</b>	<b>137</b>
8.1 Eulerova metoda . . . . .	138

8.2	Metoda Runge - Kutta . . . . .	141
8.3	Metoda diskretizacije . . . . .	146
8.4	Zadaci . . . . .	147
<b>9.</b>	<b>Numeričko rješavanje parcijalnih diferencijalnih jednačbi</b>	<b>150</b>
9.1	Rješavanje Poissonove dif. jednačbe metodom konačnih diferencija	152
9.2	Zadaci . . . . .	158
	<b>Literatura</b>	<b>161</b>
	<b>Indeks</b>	<b>164</b>

## 1. Pogreške

Vrlo često praksa nameće potrebu operiranja s približnim umjesto sa stvarnim veličinama, a umjesto stvarne fizikalne situacije, vrlo često u praksi promatramo idealiziranu sliku stvarnosti. Također, umjesto stvarnog rezultata, često se zadovoljavamo njegovom aproksimacijom. U svim takvim i sličnim situacijama moramo biti svjesni s kakvom greškom ulazimo u račun i što je još važnije, moramo unaprijed znati kakve će to posljedice imati na konačni rezultat.

### 1.1 Vrste pogrešaka

Općenito, pogreške u praktičnom radu i stručnim i znanstvenim istraživanjima možemo podijeliti u nekoliko skupina:

#### a) Pogreške zaokruživanja

Često u praksi neki broj zaokružujemo na nekoliko prvih signifikantnih znamenki (vidi *Primjebu* 1.4, str. 8), čime svjesno pravimo grešku. Neki puta u račun nije moguće uzeti stvarni broj (primjerice transcendentan broj  $\pi$  ili  $e$ ), a neki puta iz praktičnih razloga nije potrebno operirati s potpunim brojem (primjerice stopa rasta proizvodnje).

**Primjer 1.1** *Zadan je kvadrat sa stranicom  $a = 100$  cm. Njegova površina je  $P = 10\,000$  cm<sup>2</sup>. Treba odrediti stranicu kvadrata koji će imati dvostruko manju površinu.*

S  $x$  označimo stranicu traženog kvadrata. Treba biti  $a^2 = 2x^2$ , odakle je  $x = a\frac{\sqrt{2}}{2}$ . Broj  $\sqrt{2}$  je iracionalan broj (beskonačni neperiodični decimalni broj), pa smo prisiljeni u račun uzeti njegovu približnu vrijednost. Ako uzmemo  $\sqrt{2} \approx 1.41$ , dobivamo stranicu kvadrata  $x = 70.5$  cm, a odgovarajuća površina je  $P = 4970.25$  cm<sup>2</sup>. Ako uzmemo  $\sqrt{2} \approx 1.41421$ , dobit ćemo stranicu kvadrata  $x = 70.7105$  i površinu  $P = 4999.97481$  cm<sup>2</sup>. Koju ćemo vrijednost aproksimacije broja  $\sqrt{2}$  upotrijebiti, ovisi o tome s kakvom točnošću želimo dobiti rezultat.

#### b) Pogreške nastale zbog nepreciznosti ulaznih podataka

Često se u praksi koriste približne vrijednosti nekih konstanti koje utječu na veličinu pogreške konačnog rezultata. Primjerice, često za veličinu akceleracije sile teže uzimamo vrijednost  $9.81$  ms<sup>-2</sup>, iako se zna da ona čak i mijenja vrijednost



ovisno o zemljopisnom položaju.

### c) Pogreška metode

Različiti iterativni procesi (primjerice, rješavanje sustava linearnih ili nelinearnih jednadžbi, traženje lokalnih ekstrema funkcije, itd.) po prirodi su beskonačni iterativni procesi, koje u praksi zaustavljamo na nekom koraku (kad postignemo željenu točnost aproksimacije).

### d) Pogreška modela

U praksi gotovo uvijek, zbog složenosti stvarnog problema, promatramo idealizirani model kao njegovu aproksimaciju.

**Primjer 1.2** *Prilikom proučavanja balističkih putanja u eksternoj balistici najvažniji problem predstavlja dobro definiranje zakona otpora zraka za leteći objekt. Da bi se postavljanje problema pojednostavilo, uvijek prvo promatramo balistički problem u vakumu (dakle, balistički problem bez otpora zraka), čime se stvara idealizirana situacija u kojoj se problem lako rješava. No, rezultati koji se dobivaju ovakvom idealizacijom stvarnosti mogu se jako razlikovati od stvarnih. Usporedimo za primjer maksimalni domet običnog puščanog metka (10 g) i teškog topovskog projektila (150 kg) iste početne brzine  $v_0 = 900 \text{ ms}^{-1}$  u vakumu i u zraku.*

maksimalni domet projektila	vakum	zrak
puščani metak	82 km	3.5 km
teški projektil	82 km	40 km

Očigledne razlike (vidi MOLITZ (1967)) nastale su u najvećoj mjeri zbog toga što je otpor zraka prema letećem objektu obrnuto proporcionalan težini projektila.

### e) Strojna pogreška

Za potrebe numeričkih izračunavanja realan broj  $a \in \mathbb{R}$  obično zapisujemo u obliku

$$a = \pm m \times b^e, \quad (1.1)$$

gdje je  $b \in \mathbb{Z}$ ,  $b \geq 2$  baza,  $\frac{1}{b} \leq m < 1$  mantisa a  $e \in \mathbb{Z}$  eksponent realnog broja  $a$ . Ljudi uglavnom za bazu koriste  $b = 10$  (*dekadski sustav*), a kod većine računala koristi se  $b = 2$  (*binarni sustav*) ili  $b = 16$  (*heksadecimalni sustav*).

U kompjuterskom zapisu koristi se tzv. zapis s pomičnim zarezom (*floating-point representation*) broja

$$a = \pm 0.m_1m_2 \cdots m_t \times b^e, \quad (1.2)$$

gdje je mantisa  $m = 0.m_1m_2 \cdots m_t$  zapisana pomoću  $t$  nenegativnih cijelih brojeva  $0 \leq m_i < b$ ,  $m_1 \neq 0$ . Kažemo da je broj  $a$  reprezentiran u  $t$ -znamenkastoj floating-point aritmetici. Broj  $t$  je fiksna za svako pojedino računalo i predstavlja običnu točnost (*single precision*) računala. Ako realni broj  $a$  ima više od  $t$  znamenki, računalo će ga postupkom zaokruživanja ("rounding") ili odbacivanja ("chopping") pretvoriti u  $t$ -znamenkasti broj. Svako računalo obično ima i mogućnost izračunavanja i u *dvostrukoj preciznosti* ( $\approx 2t$  znamenki mantise). Eksponent  $e$  također ima

svoje granice<sup>1</sup>

$$-L \leq e \leq U, \quad L, U \in \mathbb{Z}_+$$

“*Overflow*” znači da smo pokušali zapisati broj  $a > b^U$ , a “*underflow*” da smo pokušali zapisati broj  $a < b^{-L}$

**Primjedba 1.1** Svakom računalu pridružuje se najmanji broj  $\epsilon > 0$  za koji je  $1 + \epsilon > 1$ . Broj  $\epsilon$  nazivamo točnost računala, a približno ga možemo izračunati sljedećim programom:

```
10 EPS = 1
20 EPS = .5*EPS: EPS1 = EPS+1
30 IF EPS1 > 1 THEN 20
40 PRINT EPS
```

Neke brojeve je moguće, a neke nije moguće prikazati u obliku (1.2). Primjerice, brojevi  $\sqrt{2}, e, \pi$  ne mogu se prikazati u tom obliku. Primijetite također, da je primjerice broj  $\frac{1}{10}$  moguće prikazati u obliku (1.2) s bazom  $b = 10$ , ali ne i s bazom  $b = 2$ .

Zaokruživanjem ili odbacivanjem decimala dobivamo tzv. floating-point aproksimaciju  $fl(a)$  broja  $a$ , za koju vrijedi

$$fl(a) = a(1 + \delta),$$

gdje je  $\delta$  pogreška aproksimacije.

**Primjer 1.3** Potražimo manji po apsolutnoj vrijednosti korijen kvadratne jednadžbe

$$x^2 + 2px - q = 0, \quad p > 0, q > 0, \quad p \gg q.$$

Dobivamo

$$x_0 = -p + \sqrt{p^2 + q} \quad (*)$$

Ako računamo u floating-point aritmetici, lako se može dogoditi da je  $fl(p) = fl(\sqrt{p^2 + q})$  i da je rezultat  $x_0 = 0$ .

Ako rezultat napišemo u obliku

$$x_0 = \frac{q}{p + \sqrt{p^2 + q}}, \quad (**)$$

dobit ćemo znatno točniji rezultat.

Primjerice, u 4-znamenkastoj floating-point aritmetici za  $p = 0.125 \times 10^3$  i  $q = 0.125 \times 10^{-1}$  dobivamo

$$\begin{aligned} x_0 &= 0.00005 && (\text{prava vrijednost}) \\ x'_0 &= 0 && (\text{prema } (*)) \\ x''_0 &= 0.00005 && (\text{prema } (**)) \end{aligned}$$

Niže je naveden *Mathematica*-modul za pretvaranje realnog broja  $x$  u  $k$ -znamenkasti floating-point broj

---

<sup>1</sup>Računala koja su građena prema IEEE standardima za bazu koriste  $b = 2$ . Vrijednost za  $t$  u običnoj preciznosti je 23, a u dvostrukoj preciznosti  $t = 52$ . Granice eksponenta su  $L = 126$  i  $U = 127$ , što znači da ta računala “prepoznaju” brojeve veličine  $1 \times 10^{-38} \leq a \leq 2 \times 10^{38}$

```

fl[w_, k_] := Module[{sw},
  sw = MantissaExponent[w];
  sw[[1]] = Round[10^k sw[[1]] ]/10^k;
  ww = sw[[1]] 10^sw[[2]] //N
]

```

## 1.2 Apsolutna i relativna pogreška

Označimo s  $a$  stvarnu vrijednost neke poznate (primjerice  $\sqrt{2}$ ) ili nepoznate (primjerice korijen jednadžbe  $f(x) = 0$ ) veličine, a s  $a^*$  njezinu približnu vrijednost. Obično kažemo da je  $a^*$  aproksimacija od  $a$ .

**Definicija 1.1** Razliku  $(a - a^*)$  između stvarne veličine  $a$  i njene aproksimacije  $a^*$  nazivamo **pogreška aproksimacije**. Apsolutnu vrijednost pogreške aproksimacije nazivamo **apsolutna pogreška aproksimacije** i označavamo

$$\Delta a^* = |a - a^*|. \quad (1.3)$$

**Primjer 1.4** Niz  $(a_n)$  definiran rekursivnom formulom

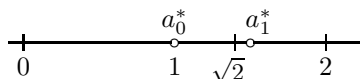
$$a_{n+1} = \frac{1}{2} \left( a_n + \frac{2}{a_n} \right), \quad a_0 = 1, \quad n = 0, 1, 2, \dots$$

konvergira prema broju  $\sqrt{2}$ . U Tablici 1.1 prikazane su redom prve četiri aproksimacije broja  $\sqrt{2}$ , kao i odgovarajuće pogreške aproksimacija.

$n$	$a_n^*$	$a - a_n^*$
0	1	+0.414214
1	1.5	-0.085786
2	1.416667	-0.002453
3	1.414216	-0.000002

Tablica 1.1.

Vidi se također, da pogreške aproksimacije mogu biti pozitivni ili negativni brojevi. Apsolutna pogreška aproksimacije na brojevnom pravcu predstavlja udaljenost točke koja odgovara aproksimaciji  $a^*$  do točke koja odgovara stvarnoj vrijednosti  $a$  (vidi Sliku 1.1).



Slika 1.1. Aproksimacije broja  $\sqrt{2}$

U praksi stvarna vrijednost  $a$  često nije poznata, ali se zna da se pogreška aproksimacije kreće u intervalu  $[-\epsilon, \epsilon]$ , za neki  $\epsilon > 0$ . To znači da za aproksimaciju

$a^*$  vrijedi

$$-\epsilon \leq a - a^* \leq \epsilon, \quad (1.4)$$

što možemo zapisati kao

$$|a - a^*| \leq \epsilon. \quad (1.5)$$

Spomenuti broj  $\epsilon > 0$  nazivamo **granica pogreške aproksimacije**. Nejednadžbu (1.4) možemo pisati u obliku

$$a^* - \epsilon \leq a \leq a^* + \epsilon. \quad (1.6)$$

Broj  $a^* - \epsilon$  je najmanja, a broj  $a^* + \epsilon$  najveća vrijednost koju može primiti aproksimacija broja  $a$ , pa često simbolički pišemo

$$a = a^* \pm \epsilon. \quad (1.7)$$

To znači da je stvarna veličina  $a$  aproksimirana brojem  $a^*$  i da pri tome apsolutna pogreška nije veća od  $\epsilon$ .

**Primjer 1.5** Obavljen je niz mjerenja neke veličine  $R$  i dobiveni su sljedeći rezultati

mjerenje	1	2	3	4	5	6
$R$	29.2	29.3	29.25	29.28	29.24	29.26

Aritmetička sredina  $\bar{R} = 29.255$  predstavlja jednu aproksimaciju veličine  $R$ , pri čemu možemo smatrati da je granica apsolutne pogreške  $\epsilon = 0.055$ . Zbog toga pišemo  $R = 29.255 \pm 0.055$ .

**Primjer 1.6** Zadani su radijus  $r$  i visina  $h$  cilindra s odgovarajućim granicama pogreške

$$r = 2 \pm 0.01 \text{ cm}, \quad h = 100 \pm 0.01 \text{ cm}.$$

Iako su granice grešaka obje veličine brojčano jednake, one nemaju isto značenje jer je veličina  $h$  relativno mnogo veća od veličine  $r$ . Točnost izmjerene veličine  $h$  cilindra relativno mnogo je veća nego točnost izmjerene veličine  $r$ . Dakle, veličinu pogreške treba promatrati u odnosu na promatranu veličinu.

**Definicija 1.2** Omjer između apsolutne pogreške  $\Delta a^*$  i apsolutne vrijednosti veličine  $a$  ( $a \neq 0$ ) nazivamo **relativna pogreška**  $\delta a^*$  i pišemo

$$\delta a^* = \frac{\Delta a^*}{|a|}. \quad (1.8)$$

Relativna pogreška je u stvari veličina apsolutne pogreške izmjerena u odnosu na promatranu veličinu  $a$ . Zbog toga se ona često izražava u postocima (množenjem relacije (1.8) sa 100) ili u promilima (množenjem relacije (1.8) s 1000).

**Primjedba 1.2** Kako je u praksi obično  $a^* \approx a$ , onda se relativna pogreška često izražava na sljedeći način

$$\delta a^* \approx \frac{\Delta a^*}{|a^*|}, \quad a^* \neq 0. \quad (1.9)$$

Mi ćemo nadalje relativnu pogrešku aproksimacije  $a^*$  također računati na ovaj način.

Spomenimo još da je često puta u praksi prikladno koristiti tzv. mješovitu pogrešku

$$e = \frac{|a - a^*|}{1 + |a^*|}$$

Primijetite da je za  $|a^*| \ll 1$  mješovita pogreška  $e$  slična apsolutnoj pogreški, a za  $|a^*| \gg 1$ , mješovita pogreška  $e$  slična je relativnoj pogreški.

**Primjer 1.7** U Primjeru 1.1, str. 1 relativna pogreška aproksimacije  $a^* = 1.41$  broja  $\sqrt{2}$  je

$$\delta a^* = \frac{|\sqrt{2} - 1.41|}{1.41} \approx 0.00299 \approx 3\text{‰}.$$

U Primjeru 1.6, str. 1 relativne pogreške veličina  $r$  i  $h$  su

$$\delta_{r^*} \approx \frac{\Delta_{r^*}}{|r^*|} = \frac{0.01}{2} = 0.005, \quad \delta_{h^*} \approx \frac{\Delta_{h^*}}{|h^*|} = \frac{0.01}{100} = 0.0001.$$

Mogli bismo reći da je veličina  $h$  zadana 50 puta pouzdanije od veličine  $r$ .

**Zadatak 1.1** Ako je  $f : [a, b] \rightarrow \mathbb{R}$  neprekidna funkcija, onda za integral

$$I = \int_a^b f(x) dx \text{ vrijedi}$$

$$m(b-a) \leq I \leq M(b-a), \quad m = \min_{x \in [a,b]} f(x), \quad M = \max_{x \in [a,b]} f(x).$$

Kao jednu od aproksimacija integrala  $I$  možemo uzeti aritmetičku sredinu

$$I^* = \frac{m(b-a) + M(b-a)}{2} = (b-a) \frac{m+M}{2}.$$

Pokažite da tada vrijedi

$$\Delta I^* \leq (b-a) \frac{M-m}{2}, \quad \delta I^* \leq \frac{M-m}{M+m}.$$

Ocijenite na takav način integral  $\int_0^2 \frac{5-x}{9-x^2} dx$ , te izračunajte apsolutnu i relativnu pogrešku.

Rješenje:  $I^* = 1.1$ ,  $\Delta I^* \leq 0.1$ ,  $\delta I^* \leq 9\text{‰}$

## 1.3 Signifikantne znamenke

Svaki pozitivni realni broj  $a$  u dekadskom sustavu možemo zapisati u obliku

$$a = b_m 10^m + b_{m-1} 10^{m-1} + \dots + b_{m-n+1} 10^{m-n+1} + b_{m-n} 10^{m-n} \dots, \quad b_m \neq 0, m \in \mathbb{Z} \quad (1.10)$$

gdje su  $b_i \in \{0, 1, \dots, 9\}$  znamenke broja  $a$ . Primjerice,

$$0.02305 = 2 \cdot 10^{-2} + 3 \cdot 10^{-3} + 0 \cdot 10^{-4} + 5 \cdot 10^{-5}.$$

Neka je

$$a^* = b_m^* 10^m + b_{m-1}^* 10^{m-1} + \dots + b_{m-n+1}^* 10^{m-n+1} + b_{m-n}^* 10^{m-n} + \dots$$

aproksimacija broja  $a$ , takva da im se podudaraju prvih  $n$  znamenki

$$b_m^* = b_m, \quad b_{m-1}^* = b_{m-1}, \quad \dots, \quad b_{m-n+1}^* = b_{m-n+1}.$$

Tada apsolutnu pogrešku  $\Delta a^*$  možemo ovako ocijeniti

$$\begin{aligned} \Delta a^* &= |a - a^*| = |(b_{m-n} - b_{m-n}^*) 10^{m-n} + \dots| \leq |b_{m-n} - b_{m-n}^*| 10^{m-n} + \dots \\ &\leq 9 \cdot 10^{m-n} + \dots = 9 \cdot 10^{m-n} (1 + 10^{-1} + \dots) \leq 9 \cdot 10^{m-n} \frac{1}{1 - \frac{1}{10}} = 10^{m-n+1}. \end{aligned}$$

Ovo je motivacija za uvođenje sljedeće definicije:

**Definicija 1.3** *Neka je*

$$a^* = b_m^* 10^m + b_{m-1}^* 10^{m-1} + \dots, \quad b_m^* \neq 0, \quad m \in \mathbb{Z}$$

*aproksimacija broja  $a$  zadanog s (1.10). Kaže se da su prvih  $n$  znamenki  $b_m^*, \dots, b_{m-n+1}^*$  broja  $a^*$  signifikantne (pouzdate) ako je  $n$  najveći pozitivni cijeli broj za koji vrijedi*

$$\Delta a^* = |a - a^*| \leq \frac{1}{2} \times 10^{m-n+1}. \quad (1.11)$$

Primjerice, ako je  $a = 0.0024357 \pm 0.3 \times 10^{-4}$ , tada je  $m = -3$  i vrijedi

$$\Delta a^* = 0.3 \times 10^{-4} < \frac{1}{2} \times 10^{-4} = \frac{1}{2} \times 10^{-3-n+1} \Rightarrow n = 2,$$

odakle zaključujemo da su samo prve dvije znamenke broja  $a^*$  signifikantne pa možemo pisati  $a^* = 0.0024$ .

Ako je  $a = 243.5731 \pm 0.7 \times 10^{-2}$ , tada je  $m = 2$  i vrijedi

$$\Delta a^* = 0.7 \times 10^{-2} = 0.07 \times 10^{-1} < \frac{1}{2} \times 10^{-1} = \frac{1}{2} \times 10^{2-n+1} \Rightarrow n = 4,$$

odakle zaključujemo da su prve 4 znamenke broja  $a^*$  signifikantne i možemo pisati  $a^* = 243.6$ .

Broj 1.416667 iz *Primjera 1.4* ima prve tri, a broj 1.414216 ima prvih šest signifikantnih znamenki.

**Primjedba 1.3** Broj signifikantnih znamenki približnog broja  $a^*$  možemo definirati i na drugi način - kao najveći pozitivni cijeli broj  $n$  za koji vrijedi

$$\delta a^* \approx \frac{|a - a^*|}{|a^*|} < \frac{1}{2} \times 10^{-n+1}. \quad (1.12)$$

Broj signifikantnih znamenki procijenjen na taj način može se od Definicije 1.3 razlikovati najviše za jedan.

**Primjedba 1.4** Vrlo često u praksi brojeve aproksimiramo zaokruživanjem. Pri tome treba se držati sljedećih pravila:

- ako se iza znamenke na koju zaokružujemo broj nalazi znamenka manja od 5, znamenka na koju zaokružujemo ne mijenja se;
- ako se iza znamenke na koju zaokružujemo broj nalazi znamenka veća ili jednaka 5, znamenka na koju zaokružujemo povećava se za 1;

**Primjer 1.8** Broj  $a = 2.351850$  treba zaokružiti na jedno, dva, tri i četiri decimalna mjesta. Redom dobivamo sljedeće aproksimacije broja  $a$ :

$$2.4, \quad 2.35, \quad 2.352, \quad 2.3519.$$

## 1.4 Pogreške kod izračunavanja vrijednosti funkcije

Zadana je realna funkcija  $n$  varijabli  $z = f(x_1, \dots, x_n)$ . Treba izračunati apsolutnu pogrešku vrijednosti funkcije u točki  $x^* = (x_1^*, \dots, x_n^*)$  ako je, u skladu s (1.7),  $x_i = x_i^* \pm \Delta x_i^*$ ,  $i = 1, \dots, n$ .

Prema teoremu o srednjoj vrijednosti je

$$\Delta z^* = |f(x_1, \dots, x_n) - f(x_1^*, \dots, x_n^*)| = \left| \sum_{i=1}^n \frac{\partial f(\tilde{x}_1, \dots, \tilde{x}_n)}{\partial x_i} (x_i - x_i^*) \right|,$$

gdje je  $\tilde{x}_i$  između  $x_i$  i  $x_i^*$ . Budući da brojevi  $\tilde{x}_i$  nisu poznati, a  $x_i^* \approx x_i$ , možemo pisati

$$\Delta z^* \leq \sum_{i=1}^n \left| \frac{\partial f(\tilde{x}_1, \dots, \tilde{x}_n)}{\partial x_i} \right| \Delta x_i^* \approx \sum_{i=1}^n \left| \frac{\partial f(x_1^*, \dots, x_n^*)}{\partial x_i} \right| \Delta x_i^*.$$

Zato je

$$\Delta z^* \approx \sum_{i=1}^n |\partial_i f^*| \Delta x_i^*, \quad \text{gdje je } \partial_i f^* := \frac{\partial f(x_1^*, \dots, x_n^*)}{\partial x_i}. \quad (1.13)$$

**Primjer 1.9** Treba odrediti apsolutnu i relativnu pogrešku pri izračunavanju volumena kugle  $V = \frac{4}{3}r^3\pi$ , ako je radijus kugle zadan s  $r = 10.2 \pm 0.01$  cm, a broj  $\pi \approx 3.14$ .

Imamo

$$r^* = 10.2, \quad \Delta r^* = 0.01, \quad \pi^* = 3.14, \quad \Delta \pi = 0.0016,$$

$$\partial_r V(r^*, \pi^*) = 4r^{*2}\pi^* = 4 \cdot 10.2^2 \cdot 3.14 = 1306.74,$$

$$\partial_\pi V(r^*, \pi^*) = \frac{4}{3}r^{*3} = \frac{4}{3} \cdot 10.2^3 = 1414.94,$$

$$\Delta V^* \approx |\partial_r V^*| \Delta r^* + |\partial_\pi V^*| \Delta \pi^* = 1306.74 \cdot 0.01 + 1414.94 \cdot 0.0016 = 15.33.$$

Dakle,

$$V = \frac{4}{3}r^{*3}\pi^* \pm \Delta V^* = 4442.92 \pm 15.33 \text{ cm}^3.$$

Kako je  $\Delta V^* \approx 0.5 \times 10^{3-2+1}$ , signifikantne su samo prve dvije znamenke. Relativna pogreška prema (1.9) je

$$\delta V^* \approx \frac{\Delta V^*}{|V^*|} = \frac{15.33}{4442.92} = 0.00345 \approx 3.5\%.$$

Prema *Primjedbi* 1.3 to bi značilo da su prve tri znamenke signifikantne.

Za navedeni račun možemo koristiti kratki *Mathematica*-program:

```
In[1]:=V[r_, pi_] := 4 r^3 pi/3;
w = {r -> 10.2, pi -> 3.14};
Vz = V[r, pi] /. w
DV = .01 Abs[D[V[r, pi], r] /. w] + .0016 Abs[D[V[r, pi], pi] /. w]
dV = DV/Vz
```

**Primjer 1.10** Treba izračunati pogrešku pri izračunavanju površine kruga radijusa 100 m, ako uzmemo  $\pi \approx 3.14$ .

Imamo

$$\pi^* = 3.14, \quad \Delta \pi^* = 0.0016, \quad r^* = 100, \quad \Delta r^* = 0,$$

$$\Delta P^* \approx |\partial_r P^*| \Delta r^* + |\partial_\pi P^*| \Delta \pi^* = 0 + 10\,000 \cdot 0.0016 = 16 \text{ m}^2,$$

$$\delta P^* \approx \frac{\Delta P^*}{|P^*|} = \frac{16}{31\,400} \approx 0.5\%.$$

Približna površina je  $P^* = r^2 \pi^* = 31\,400 \pm 16 \text{ m}^2$ , a stvarna površina je  $P \approx 31\,415.93 \text{ m}^2$ , što se podudara s gornjom procjenom.

**Primjer 1.11** Treba procijeniti pogrešku kod izračunavanja volumena valjka ( $V = r^2 \pi h$ ), ako su mu radijus  $r \approx 2 \text{ m}$  i visina  $h \approx 3 \text{ m}$  zadani s točnošću na dvije decimale (3 signifikantne znamenke), a broj  $\pi$  uzimamo s točnošću na 5 decimala (6 signifikantnih znamenki).

Imamo

$$r^* = 2, \quad h^* = 3, \quad \pi^* = 3.14159,$$

$$\Delta r^* = 0.005, \quad \Delta h^* = 0.005, \quad \Delta \pi = 0.00000265,$$

$$\partial_r V^* = 2r^* \pi^* h^* = 2 \cdot 2 \cdot 3.14159 \cdot 3 = 37.6991,$$

$$\partial_\pi V^* = r^{*2} h^* = 2^2 \cdot 3 = 12, \quad \partial_h V^* = r^{*2} \pi^* = 2^2 \cdot 3.14159 = 12.5664,$$

$$\Delta V^* \approx |\partial_r V^*| \Delta r^* + |\partial_\pi V^*| \Delta \pi^* + |\partial_h V^*| \Delta h^* \approx 0.25 \text{ m}^3,$$

$$\delta V^* \approx \frac{\Delta V^*}{|V^*|} = \frac{0.25}{37.6991} = 0.0066 \approx 7\%.$$

To znači da su u broju  $V^* = 37.6991$  samo prve dvije znamenke signifikantne.



**Primjedba 1.5** *Prilikom izvođenja računskih operacija s približnim brojevima potrebno je unaprijed znati s kakvom točnošću možemo očekivati dobiveni rezultat. Primjenom formule (1.13) uz pretpostavku da su apsolutne pogreške ulaznih veličina malene i to je moguće procijeniti:*

- *Prilikom zbrajanja ili oduzimanja dva približna broja dobivamo broj čija apsolutna pogreška ne prelazi zbroj apsolutnih pogrešaka pribrojnika. Naime, ako je  $z = x + y$ , onda je*

$$\Delta z^* = |z - z^*| = |x + y - x^* - y^*| \leq |x - x^*| + |y - y^*| = \Delta x^* + \Delta y^*.$$

- *Prilikom množenja dva približna broja dobivamo broj čija relativna pogreška ne prelazi zbroj relativnih pogrešaka članova produkta. Naime, ako je  $z = x \cdot y$ , onda prema (1.13) vrijedi*

$$\Delta z^* \approx |\partial_x z^*| \Delta x^* + |\partial_y z^*| \Delta y^* = |y^*| \Delta x^* + |x^*| \Delta y^*,$$

odakle slijedi

$$\delta z^* \approx \frac{\Delta z^*}{|z|} \approx \frac{|y^*| \Delta x^* + |x^*| \Delta y^*}{|xy|} = \frac{\Delta x^*}{|x|} + \frac{\Delta y^*}{|y|} = \delta x^* + \delta y^*.$$

- *Prilikom dijeljenja dva približna broja dobivamo broj čija relativna pogreška ne prelazi zbroj relativnih pogrešaka članova kvocijenta. Naime, ako je  $z = x/y$ , onda prema (1.13) vrijedi*

$$\Delta z^* \approx |\partial_x z^*| \Delta x^* + |\partial_y z^*| \Delta y^* = \frac{1}{|y^*|} \Delta x^* + \left| \frac{x^*}{-y^{*2}} \right| \Delta y^*,$$

odakle slijedi

$$\delta z^* \approx \frac{\Delta z^*}{|z^*|} \approx \left| \frac{y^*}{x^*} \right| \left( \frac{\Delta x^*}{|y^*|} + \frac{|x^*|}{|y^*|^2} \Delta y^* \right) = \frac{\Delta x^*}{|x^*|} + \frac{\Delta y^*}{|y^*|} \approx \delta x^* + \delta y^*.$$

**Primjedba 1.6** *U primjenama se također upotrebljava tzv. standardna pogreška funkcije (naročito ako je broj nezavisnih varijabli  $x_1, \dots, x_n$  nešto veći – vidi primjerice DAHLQUIST (1972)). Uz pretpostavku (vidi GALIĆ (1999)) da su sve pogreške  $x_i - x_i^*$ ,  $i = 1, \dots, n$  nezavisne i normalno distribuirane slučajne varijable s matematičkim očekivanjem 0 i standardnim devijacijama  $\Delta x_1^*, \dots, \Delta x_n^*$ , standardna pogreška funkcije  $z = f(x_1, \dots, x_n)$  zadana je s*

$$\Delta z^* \approx \left( \sum_{i=1}^n (\partial_i f^*)^2 (\Delta x_i^*)^2 \right)^{1/2}.$$

## 1.5 Inverzni problem u teoriji pogrešaka

U prethodnom poglavlju pokazali smo kako se može procijeniti pogreška prilikom izračunavanja vrijednosti funkcije ako su vrijednosti nezavisnih varijabli približni brojevi. U praksi se obično postavlja obrnuti problem:

s kojom točnošću moramo uzeti vrijednosti nezavisnih varijabli promatrane funkcije, tako da njezina vrijednost bude u granicama unaprijed zadane točnosti?

Za rješavanje ovog problema koristit ćemo tzv. “princip jednakih efekata”:

pretpostavljamo da svi parcijalni diferencijali iz (1.13) imaju jednaki utjecaj na veličinu apsolutne pogreške, tj.

$$|\partial_1 f^*| \Delta x_1^* = |\partial_2 f^*| \Delta x_2^* = \dots = |\partial_n f^*| \Delta x_n^* \quad (1.14)$$

Iz (1.13) i (1.14) lako slijedi

$$\Delta x_i^* \approx \frac{\Delta z^*}{n |\partial_i f^*|}, \quad i = 1, \dots, n. \quad (1.15)$$

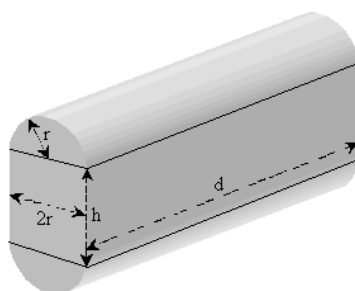
**Primjer 1.12** Kolike smiju biti apsolutne pogreške radijusa  $r$  i visine  $h$  valjka volumena  $V = 12\pi \pm 0.1 \text{ m}^3$  (dozvoljena tolerancija je 100 litara!)? Neka je pri tome  $r^* = 2 \text{ m}$ . Broj  $\pi$  uzet ćemo dovoljno točno da njegova pogreška ne utječe na rezultat.

Najprije iz  $12\pi = r^{*2}\pi h^*$  slijedi  $h^* = 3 \text{ m}$ . Kako je  $\partial_r V^* = 2\pi r^* h^* = 37.7$ ,  $\partial_h V^* = r^{*2}\pi = 12.57$ , iz (1.15) dobivamo

$$\Delta r^* \approx \frac{\Delta V^*}{2|\partial_r V^*|} = 0.001 \text{ m}, \quad \Delta h^* \approx \frac{\Delta V^*}{2|\partial_h V^*|} = 0.004 \text{ m}.$$

Dakle, radijus  $r$  moramo imati s točnošću do na 1 mm, a visinu do na 4 mm.

**Primjer 1.13** Treba načiniti spremnik za tekućinu oblika prikazanog na Slici 1.2 volumena 1000 l, širine baze  $2r^* = 80 \text{ cm}$  i visine baze  $2r^* + h^* = 100 \text{ cm}$ . Kolika mora biti dužina  $d$  spremnika? Kolike smiju biti pogreške  $\Delta r^*$ ,  $\Delta h^*$ ,  $\Delta d^*$ , ako je dozvoljena tolerancija volumena 10 litara?



Slika 1.2 Spremnik za tekućinu

Kako je  $V^* = r^* d^* (r^* \pi + 2h^*)$ , onda je

$$d^* = \frac{V^*}{r^* (r^* \pi + 2h^*)} = \frac{1\,000\,000}{40(40\pi + 40)} \approx 151 \text{ cm}.$$

Kako je

$$\partial_r V^* = 2 d^* (r^* \pi + h^*) = 43\,990,$$

$$\partial_h V^* = 2 d^* r^* = 12\,080, \quad \partial_d V^* = r^* (r^* \pi + 2h^*) = 6\,627,$$

onda iz (1.15) dobivamo

$$\begin{aligned} \Delta r^* &\approx \frac{\Delta V^*}{3|\partial_r V^*|} = \frac{10\,000}{3 \cdot 43\,990} = 0.08 \text{ cm}, \\ \Delta h^* &\approx \frac{\Delta V^*}{3|\partial_h V^*|} = 0.3 \text{ cm}, \\ \Delta d^* &\approx \frac{\Delta V^*}{3|\partial_d V^*|} = 0.5 \text{ cm}. \end{aligned}$$

Dakle,

$$r = 40 \pm 0.08 \text{ cm}, \quad h = 20 \pm 0.3 \text{ cm}, \quad d = 151 \pm 0.5 \text{ cm}.$$

**Primjedba 1.7** Više detalja o teoriji pogrešaka i primjenama može se vidjeti primjerice u BOHTE (1993), DEMIDOVICH (1981).

Posebno o problemu kompjuterske preciznosti u numeričkim metodama može se vidjeti u GILL (1981, 1991), PRESS (1992), STOER (1993), itd.

## 1.6 Zadaci

**Zadatak 1.2** Korištenjem poznatih formula izračunajte po apsolutnoj vrijednosti manji korijen kvadratne jednadžbe  $2x^2 + 984x - 0.05 = 0$  u a) 2-znamenkastoj, b) 6-znamenkastoj, c) 7-znamenkastoj, d) 8-znamenkastoj i e) 10-znamenkastoj floating-point aritmetici. Usporedite rezultat s točnom vrijednošću tog korijena:  $x_0 = 0.000050813$ . Kako treba napisati formulu da bi rješenje i u nižoj floating-point aritmetici bilo prihvatljivo? Kakva su ta rješenja?

Uputa: Koristite priloženi Mathematica-program.

Rješenje: a) 4, b) 0, c) 0.0001, d) 0.00005, e) 0.0000508.

**Zadatak 1.3** U  $k$ -znamenkastoj ( $k = 3, 4, 6, 8$ ) floating-point aritmetici izračunajte vrijednost polinoma  $P(x) = -32x^3 + 96x^2 - 90x + 27$  u točki  $a = 0.6$  direktno i koristeći Hornerovu shemu. Što primjećujete?

**Zadatak 1.4** Broj  $a^*$  je aproksimacija broja  $a$ . Za koliko znamenki broja  $a^*$  možemo reći da su signifikantne?

$$\begin{aligned} \text{a) } a &= 23.395, & a^* &= 23.40 & \text{b) } a &= 0.00275, & a^* &= 0.00266 \\ \text{c) } a &= 243317, & a^* &= 243315 & \text{d) } a &= 0.012815, & a^* &= 0.0130 \end{aligned}$$

Rješenje:

$$\begin{aligned} \text{a) } \Delta a^* &= 0.5 \cdot 10^{-2} = 0.5 \times 10^{1-4+1}, & 4 \text{ signifikantne znamenke} \\ \text{b) } \Delta a^* &= 0.09 \times 10^{-3} < 0.5 \times 10^{-3-1+1}, & 1 \text{ signifikantna znamenka} \\ \text{c) } \Delta a^* &= 2 < 0.5 \times 10^1 = 0.5 \times 10^{5-5+1}, & 5 \text{ signifikantnih znamenki} \\ \text{d) } \Delta a^* &= 0.185 \times 10^{-3} < 0.5 \times 10^{-2-2+1}, & 2 \text{ signifikantne znamenke} \end{aligned}$$

**Zadatak 1.5** Broj  $a$  zaokružite na jedno, dva, tri i četiri decimalna mjesta.

$$\text{a) } a = 2.71828 \quad \text{b) } a = 0.753550 \quad \text{c) } a = 0.97965$$

Rješenje: a) 2.7, 2.72, 2.718, 2.7183  
 b) 0.8, 0.75, 0.754, 0.7536  
 c) 1, 0.98, 0.980, 0.9797

**Zadatak 1.6** Zadana je godišnja kamatna stopa  $p$ . Ispodgodišnji konformni kamatnjak  $p_m$  (vidi primjerice SCITOVSKI (1993)) računa se po formuli

$$p_m = 100 \left( \sqrt[m]{1 + \frac{p}{100}} - 1 \right),$$

gdje je  $m$  broj jednakih podintervala na koje dijelimo godinu. Jedna aproksimacija konformnog kamatnjaka koja se koristi u praksi je tzv. relativni kamatnjak, dobiva se kao linearna aproksimacija konformnog kamatnjaka pomoću Taylorove formule

$$p_r = \frac{p}{m}.$$

Ako je  $m = 12$ , izračunajte apsolutnu i relativnu pogrešku aproksimacije, te relativnu pogrešku u % za različite vrijednosti godišnje kamatne stope

$$p = 2\%, 3.5\%, 5\%, 7.5\%, 12\%, 24\%, 50\%, 120\%, 360\%, 1000\%.$$

Rješenje:  $\Delta p_r = |p_m - p_r|$ ,  $\delta p_r = \frac{\Delta p_r}{p_m}$

$p$	2%	3.5%	7.5%	12%	24%	50%	120%	360%	1000%
$\Delta p_r$	0.0015	0.0046	0.02	0.05	0.19	0.73	3.21	16.44	61.21
$\delta p_r$	0.0091	0.0159	0.03	0.05	0.11	0.21	0.47	1.21	2.76
%	1	1.6	3	5	11	21	47	121	276

**Zadatak 1.7** Kako se ponaša apsolutna i relativna pogreška u prethodnom zadatku ako za  $m$  uzmemo 2, 4 ili 365?

**Zadatak 1.8** Neka je  $x = 2.00 \pm 0.005$ ,  $y = 3.00 \pm 0.005$ ,  $z = 4.00 \pm 0.005$ . Procijenite pogrešku prilikom izračunavanja vrijednosti niže navedenih funkcija u točki  $(x^*, y^*, z^*)$

$$a) \quad f(x, y, z) = 3x + y - z, \quad b) \quad f(x, y, z) = x \frac{y}{z}, \quad c) \quad f(x, y) = x \sin \frac{y}{40}.$$

Rješenje: a)  $\Delta f^* \approx 0.025$ , b)  $\Delta f^* \approx 0.008$ , c)  $\Delta f^* \approx 0.003$

**Zadatak 1.9** Procijenite apsolutnu i relativnu pogrešku pri izračunavanju površine  $P$  pravokutnika sa stranicama

$$a = 29.3 \pm 0.05 \text{ cm}, \quad b = 18.1 \pm 0.04 \text{ cm}.$$

Rješenje:  $\Delta P^* \approx 2.08 \text{ cm}^2$ ,  $\delta P^* \approx 0.0039 \approx 4\%$ .

**Zadatak 1.10** Procijenite apsolutnu i relativnu pogrešku pri izračunavanju volumena  $V$  paralelepipeda sa stranicama

$$a = 29.3 \pm 0.05 \text{ cm}, \quad b = 18.1 \pm 0.04 \text{ cm}, \quad c = 11.2 \pm 0.03 \text{ cm}.$$

Rješenje:  $\Delta V \approx 39.17 \text{ cm}^3$ ,  $\delta V^* \approx 0.0066 \approx 7\text{‰}$ .

**Zadatak 1.11** Procijenite apsolutnu i relativnu pogrešku pri izračunavanju volumena  $V = 2ar^2\pi^2$  torusa koji nastaje rotacijom kruga radijusa  $r$ , čije je središte za  $a$  udaljeno od centra rotacije, ako je

- a)  $r = 7.5 \pm 0.05 \text{ cm}$ ,  $a = 30 \pm 0.1 \text{ cm}$ .  
 b)  $r = 20 \pm 0.2 \text{ cm}$ ,  $a = 100 \pm 0.5 \text{ cm}$ .

Rješenje: a)  $\Delta V^* \approx 555.2 \text{ cm}$ ,  $\delta V^* \approx 0.017 \approx 2\%$ .  
 b)  $\Delta V^* \approx 19739.2 \text{ cm}$ ,  $\delta V^* \approx 0.025 \approx 2.5\%$ .

**Zadatak 1.12** Prema poučku o kosinusima, ako su za neki trokut poznate dvije stranice  $a, b$  i kut  $\gamma$  između njih, onda se treća stranica  $c$  može izračunati po formuli

$$c^2 = a^2 + b^2 - 2ab \cos \gamma.$$

Procijenite apsolutnu i relativnu pogrešku pri izračunavanju treće stranice  $c$  trokuta, ako je

- a)  $a = 10.2 \pm 0.05 \text{ cm}$ ,  $b = 7.5 \pm 0.05 \text{ cm}$ ,  $\gamma = 31.3^\circ \pm 0.05^\circ$ .  
 b)  $a = 230.7 \pm 0.05 \text{ m}$ ,  $b = 123.5 \pm 0.05 \text{ m}$ ,  $\gamma = 27.4^\circ \pm 0.05^\circ$ .

Rješenje: kut  $\gamma$  najprije treba pretvoriti u radijane. Dobivamo:

- a)  $\Delta c^* \approx 0.0524 \text{ cm}$ ,  $\delta c^* \approx 0.01 \approx 1\%$ ,  $c^* \approx 5.4367 \text{ cm}$ .  
 b)  $\Delta c^* \approx 0.16 \text{ m}$ ,  $\delta c^* \approx 0.0012 \approx 0.1\text{‰}$ ,  $c^* \approx 133.733 \text{ m}$ .

**Zadatak 1.13** Procijenite apsolutnu i relativnu pogrešku prilikom:

- a) zbrajanja  $n$  približnih brojeva  
 b) množenja  $n$  približnih brojeva  
 c) kvadriranja jednog približnog broja  
 d) potenciranja  $s$ -tom potencijom jednog približnog broja  
 e) izračunavanja drugog korijena nekog približnog broja  
 f) izračunavanja  $n$ -tog korijena nekog približnog broja  
 g) izračunavanja prirodnog logaritma nekog pozitivnog realnog broja

**Zadatak 1.14** Treba načiniti tetraedar volumena  $V = 1 \pm 0.01 \text{ l}$ . S kojom točnošću treba imati stranicu  $a$  tetraedra?

Uputa: Volumen tetraedra stranice  $a$  je  $V = \frac{a^3\sqrt{2}}{12}$ .

Rješenje:  $\Delta a^* \approx 0.0068 \text{ cm}$ ,  $\delta a^* \approx 0.0033 \approx 0.3\text{‰}$ ,  $a^* \approx 2.03965 \text{ dm}$ .

**Zadatak 1.15** Ako je  $z = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$  i  $x_i = x_i^* \pm \Delta x_i^*$ , pokažite da je  $\delta z^* \approx \sum_{i=1}^n |\alpha_i| \delta x_i^*$ .

Uputa: Najprije logaritmirajte  $\ln z = \alpha_1 \ln x_1 + \cdots + \alpha_n \ln x_n$ .

**Zadatak 1.16** Procijenite apsolutnu pogrešku pri izračunavanju vrijednosti funkcije  $f(x) = \sqrt{1+x^2} - x + 200$ , ako je  $x = 100 \pm 1$ .

Rješenje:  $f(x) = 200.0050 \pm 0.5 \times 10^{-4}$ .

**Zadatak 1.17** Balistički projektil ispaljen u vakuumu početnom brzinom  $v_0$  pod kutem  $\alpha$  u odnosu na ravni teren, past će na udaljenosti  $d = \frac{v_0^2}{g} \sin 2\alpha$ , gdje je  $g = 9.81 \text{ ms}^{-2}$  akceleracija sile teže (vidi Primjer 4.11, str. 83). Procijenite apsolutnu i relativnu pogrešku pri izračunavanju daljine gađanja ako je

$$a) v_0 = 115 \pm 1 \text{ ms}^{-1}, \quad \alpha = 15^\circ \pm 0.1^\circ.$$

$$b) v_0 = 900 \pm 10 \text{ ms}^{-1}, \quad \alpha = 15^\circ \pm 0.1^\circ.$$

Rješenje: a)  $\Delta d^* \approx 15.8 \text{ m}$ ,  $\delta d^* \approx 0.023 \approx 2\%$ ,  $d^* \approx 674.057 \text{ m}$ .

b)  $\Delta d^* \approx 1167.04 \text{ m}$ ,  $\delta d^* \approx 0.028 \approx 3\%$ ,  $d^* \approx 41\,284.4 \text{ m}$ .

**Zadatak 1.18** Maksimalni domet balističkog projektila na ravnom terenu u vakuumu računa se po formuli (vidi Primjer 4.11, str. 83):  $X = \frac{v_0^2}{g}$ . Procijenite apsolutnu i relativnu pogrešku pri izračunavanju maksimalnog dometa projektila, ako je

$$a) v_0 = 115 \pm 1 \text{ ms}^{-1}, \quad g = 9.81 \pm 0.001 \text{ ms}^{-2}.$$

$$b) v_0 = 900 \pm 10 \text{ ms}^{-1}, \quad g = 9.81 \pm 0.001 \text{ ms}^{-2}.$$

Rješenje: a)  $\Delta X^* \approx 23.58 \text{ m}$ ,  $\delta X^* \approx 0.017 \approx 1.7\%$ .

b)  $\Delta X^* \approx 1843.28 \text{ m}$ ,  $\delta X^* \approx 0.022 \approx 2.2\%$ .

**Zadatak 1.19** Vrijeme leta balističkog projektila ispaljenog pod kutem  $\alpha$  početnom brzinom  $v_0$  na ravnom terenu u vakuumu računa se po formuli (vidi Primjer 4.11, str. 83):  $T = \frac{2v_0}{g} \sin \alpha$ . Procijenite apsolutnu i relativnu pogrešku pri izračunavanju vremena leta  $T$  projektila, ako je

$$a) v_0 = 115 \pm 1 \text{ ms}^{-1}, \quad g = 9.81 \pm 0.001 \text{ ms}^{-2}, \quad \alpha = 15^\circ \pm 0.1^\circ$$

$$b) v_0 = 900 \pm 10 \text{ ms}^{-1}, \quad g = 9.81 \pm 0.001 \text{ ms}^{-2}, \quad \alpha = 15^\circ \pm 0.1^\circ$$

Rješenje: a)  $\Delta T^* \approx 0.093 \text{ s}$ ,  $\delta T^* \approx 0.015 \approx 1.6\%$ ,  $T^* \approx 6.068 \text{ s}$ .

b)  $\Delta T^* \approx 0.84 \text{ s}$ ,  $\delta T^* \approx 0.018 \approx 2\%$ ,  $T^* \approx 47.49 \text{ s}$ .

**Zadatak 1.20** Balističkim projektilom početne brzine  $v_0 = 300 \text{ ms}^{-1}$  u vakuumu na ravnom terenu gađa se točka udaljena  $d = 5000 \text{ m}$ . S kojom točnošću treba imati izlazni kut  $\alpha$  (u stupnjevima), ako se tolerira apsolutna pogreška pogotka cilja od  $\Delta d^* = 50 \text{ m}$ ?

Uputa: Vidi Zadatak 1.17. Veličine  $v_0$  i  $g$  smatraju se točnim.

Rješenje:  $\alpha \approx 0.288 \pm 0.003$  (odnosno  $\alpha \approx 16.5^\circ \pm 0.18^\circ$ ).

## 2. Interpolacija. Spline interpolacija

Često za neku funkciju  $f$  nemamo analitički izraz, ali poznajemo njenu vrijednost u nekoliko točaka:  $x_0 < x_1 < \dots < x_n$ . U području podataka, tj. u intervalu  $[x_0, x_n]$  treba aproksimirati funkciju  $f$  nekom jednostavnijom poznatom funkcijom  $g$ , tako da bude

$$g(x_i) = f(x_i), \quad i = 0, 1, \dots, n, \quad (2.1)$$

Problem određivanja funkcije  $g$  na osnovi zahtjeva (2.1) nazivamo **problem interpolacije**. Funkcija  $g$  obično se bira u klasi polinoma, trigonometrijskih, eksponencijalnih, racionalnih ili nekih drugih funkcija. Kada odredimo funkciju  $g$ , onda možemo i procijeniti vrijednosti funkcije  $f$  u nekoj točki  $x$ ,  $x \neq x_i$ , tako da stavimo  $f(x) \approx g(x)$ .

### 2.1 Interpolacija

Pretpostavimo da je funkcija  $f : [a, b] \rightarrow \mathbb{R}$  neprekidna i da je njena vrijednost poznata u  $n + 1$  točaka (zvat ćemo ih **čvorovi interpolacije**)

$$a \leq x_0 < x_1 < \dots < x_n \leq b, \quad (2.2)$$

u kojima prima vrijednosti

$$y_i = f(x_i), \quad i = 0, 1, \dots. \quad (2.3)$$

Na osnovi podataka  $(x_i, y_i)$ ,  $i = 1, \dots, n$ , treba rekonstruirati funkciju  $f$ , tj. pronaći novu funkciju  $g : [a, b] \rightarrow \mathbb{R}$ , tako da bude  $g(x_i) = y_i$ . Pri tome pogrešku aproksimacije obično izražavamo u  $L_\infty$ -normi

$$\|f - g\|_\infty = \max_{x \in [a, b]} |f(x) - g(x)|$$

ili u  $L_2$ -normi

$$\|f - g\|_2 = \sqrt{\int_a^b (f(x) - g(x))^2 dx}.$$

Prijeđimo na problem interpolacije polinomom. Treba pronaći interpolacijski polinom stupnja  $n$

$$P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0,$$





```

ts = Timing[ mat = Table[0, {i,0,n}, {j,0,n}];
Do[Do[If[j < 2, mat[[i,1]]=1,
      mat[[i,j]] = N[x[[i]]^(j-1)], {i,n+1}, {j,n+1}]];
(* Interpolacijski polinom *)
koef = LinearSolve[mat, y];
intpol[xx_] := Sum[koef[[i]] xx^(i-1), {i,n+1}];
(* Graf funkcije f, interpolacijski polinom p i podaci *)
slf = Plot[f[x], {x, x[[1]], x[[n+1]]}, DisplayFunction -> Identity];
slpod=ListPlot[pod, PlotStyle -> AbsolutePointSize[4],
      DisplayFunction -> Identity];
slp=Plot[intpol[xx],{xx,x[[1]],x[[n+1]]},DisplayFunction->Identity];
Show[slf, slp, slpod, DisplayFunction -> $DisplayFunction];
];
Print["Vrijeme racunanja = ", ts[[1]]\ ]

```

Vrijenost polinoma  $P$  u točki  $z$  možemo izračunati Hornerovom shemom korištenjem jednostavnog potprograma:

```

Horner[a_,n_,z_] := Module[{s}, p = a[[n]];
Do[p = p z + a[[n-i]], {i,n-1}]; p];

```

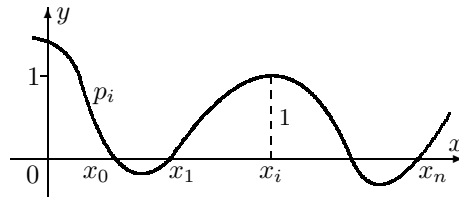
Ovdje je  $a$  vektor koeficijenata polinoma, a  $n$  broj tih koeficijenata.

### 2.1.1 Lagrangeov oblik interpolacijskog polinoma

Pogledajmo najprije jedan specijalni slučaj: *treba pronaći polinom  $p_i$  stupnja  $n$  za koji vrijedi*

$$p_i(x_j) = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases}, \quad i, j = 0, 1, \dots, n. \quad (2.6)$$

Geometrijski to znači da treba pronaći polinom čiji graf presijeca os  $x$  u točkama  $x_0, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ , a u  $x_i$  prima vrijednost 1 (Slika 2.2).



Slika 2.2. Polinom  $p_i$

Kako polinom  $p_i$  iščezava u točkama  $x_0, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ , mora biti

$$p_i(x) = C_i(x - x_0)(x - x_1) \cdots (x - x_{i-1})(x - x_{i+1}) \cdots (x - x_n) \quad (2.7)$$

gdje je  $C_i$  konstanta koju ćemo odrediti iz uvjeta  $p_i(x_i) = 1$

$$C_i = \frac{1}{(x_i - x_0)(x_i - x_1) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_n)}. \quad (2.8)$$

Uvrštavajući (2.8) u (2.7), dobivamo traženi polinom

$$p_i(x) = \frac{(x-x_0)(x-x_1)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)}{(x_i-x_0)(x_i-x_1)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_n)}.$$

Prema tome, polinom  $P_n$  za koji vrijedi (2.4) uz uvjet (2.2) glasi

$$\begin{aligned} P_n(x) &= \sum_{i=0}^n y_i p_i(x) \\ &= \sum_{i=0}^n y_i \frac{(x-x_0)(x-x_1)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)}{(x_i-x_0)(x_i-x_1)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_n)}. \end{aligned} \quad (2.9)$$

Očigledno je

$$P_n(x_j) = \sum_{i=0}^n p_i(x_j) y_i = p_j(x_j) y_j = y_j.$$

Polinom  $P_n$  obično nazivamo **Lagrangeov oblik interpolacijskog polinoma**

Primijetimo, međutim, da izračunavanje vrijednosti interpolacijskog polinoma u Lagrangeovom obliku u nekoj točki  $x \neq x_i$  zahtijeva veliki broj računskih operacija, pa time i značajno vrijeme rada računala.

**Primjer 2.1** *Odrediti interpolacijski polinom čiji graf prolazi točkama  $T_0(-1, 4)$ ,  $T_1(2, 7)$ ,  $T_2(4, 29)$ .*

Prema (2.9) imamo

$$P_2(x) = \frac{(x-2)(x-4)}{(-1-2)(-1-4)} 4 + \frac{(x+1)(x-4)}{(2+1)(2-4)} 7 + \frac{(x+1)(x-2)}{(4+1)(4-2)} 29$$

ili nakon sređivanja

$$P_2(x) = 2x^2 - x + 1.$$

**Primjer 2.2** *Treba odrediti interpolacijski polinom za funkciju  $f(x) = \ln x - 1 + \frac{1}{x}$  i sljedeće čvorove interpolacije*

$$\begin{array}{ll} a) & 1, 2, 4, 8, 10 \quad b) \quad 2, 4, 8, 10 \\ c) & 4, 8, 10 \quad d) \quad 2, 4, 8 \end{array}$$

Za svaki od tih slučajeva treba izračunati  $P(5.25)$ , te apsolutnu i relativnu pogrešku ove aproksimacije.

Prema (2.9) dobivamo (koristite priloženi program)

- a)  $P_4(x) = 0.0004455x^4 - 0.009929x^3 + 0.06339x^2 + 0.0658x - 0.11971$
- b)  $P_3(x) = 0.0007619x^3 - 0.023923x^2 + 0.3438x - 0.4048$
- c)  $P_2(x) = -0.007160x^2 + 0.227947x - 0.161$
- d)  $P_2(x) = -0.013256x^2 + 0.30111x + 0.35605$

	$f(5.25)$	$y^* = P(5.25)$	$\Delta y^*$
a)	-0.848704	-0.874536	0.0258
b)	-0.848704	-0.850896	0.0022
c)	-0.848704	-0.838456	0.0102
d)	-0.848704	-0.859408	0.0107

Tablica 2.1.

Rezultate provjerite priloženim *Mathematica*-programom, str. 17.

**Primjer 2.3** Za različite vrijednosti izlaznog kuta  $\alpha$  i konstantnu početnu brzinu  $v_0 = 115 \text{ ms}^{-1}$  primjenom formule (B.4), str. 84 izračunate su daljine gađanja  $X$  balističkog projektila u vakuumu

$\alpha^\circ$	5°	15°	25°	35°	45°
$X \text{ (m)}$	234	674	1033	1267	1348

Na osnovu ovih podataka treba odrediti polinom koji će za proizvoljnu daljinu  $X$  davati vrijednost izlaznog kuta  $\alpha$ .

Primjenom Lagrangeove interpolacijske formule (2.9) dobivamo polinom  $P_4$  s koeficijentima

$$a_0 = 57.95152, a_1 = -0.41569, a_2 = 0.00101, a_3 \approx a_4 \approx 0.$$

Pomoću ove funkcije za zadane daljine gađanja: 500 m, 750 m i 1000 m izračunati su odgovarajući izlazni kutevi  $\alpha^*$ , te apsolutne i relativne pogreške u odnosu na prave vrijednosti  $\alpha$  dobivene iz formule (B.4), str. 84.

$X \text{ (m)}$	$\alpha(^{\circ})$	$\alpha^*(^{\circ})$	$\Delta\alpha^*$
500	10.885	5.676	5.209
750	16.901	18.452	1.551
1000	23.942	24.505	0.563

Tablica 2.2.

**Primjedba 2.1** Za zadane točke  $T_i(x_i, y_i)$ ,  $i = 0, 1, \dots, n$ , možemo potražiti polinom  $P_m$  stupnja nižeg od  $n$  ( $m < n$ ), za koji više neće biti ispunjen uvjet (2.4), ali čiji će graf prolaziti blizu zadanih točaka. Tako se u prethodnom primjeru pokazalo da su posljednja dva koeficijenta u interpolacijskom polinomu zanemarivo malena, pa ima smisla postavljeno problem pokušati približno riješiti pomoću polinoma 2. stupnja. U tom slučaju aproksimacija se obično pravi tako da zahtijevamo da suma kvadrata odstupanja zadanih točaka od odgovarajućih točaka na grafu polinoma bude minimalna. Ova metoda naziva se metoda najmanjih kvadrata (vidi t.6, str. 103).

### 2.1.2 Newtonov oblik interpolacijskog polinoma

Pretpostavimo da su zadane vrijednosti  $y_i = f(x_i)$ ,  $i = 0, 1, \dots, n$  neprekidne funkcije  $f : [a, b] \rightarrow \mathbb{R}$  u čvorovima  $a \leq x_0 < x_1 < \dots < x_n \leq b$ . Interpolacijski

polinom tražit ćemo u obliku

$$P_n(x) = a_0 + a_1(x - x_0) + \cdots + a_n(x - x_0)(x - x_1) \cdots (x - x_{n-1}). \quad (2.10)$$

Nakon što odredimo koeficijente  $a_0, a_1, \dots, a_n$  vrijednost polinoma  $P_n$  u nekoj točki  $x \neq x_i$  možemo izračunati prema

$$P_n(x) = (\cdots ((a_n(x - x_{n-1}) + a_{n-1})(x - x_{n-2}) + \cdots + a_1)(x - x_0) + a_0 \quad (2.11)$$

što zahtijeva znatno manje izračunavanja nego kod Lagrangeovog oblika interpolacijskog polinoma. Primijetimo da (2.11) po formi odgovara tzv. Hornerovoj shemi za izračunavanje vrijednosti polinoma (vidi JUKIĆ (1998), str. 47).

Prijedimo na izračunavanje koeficijenata  $a_i$  u prikazu (2.10). Ako je  $n = 1$ , onda je interpolacijski polinom čiji graf prolazi točkama  $T_0(x_0, y_0)$  i  $T_1(x_1, y_1)$  linearna funkcija (pravac kroz dvije točke)

$$P_1(x) = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0),$$

što uz oznake  $f[x_0] := y_0$ ,  $f[x_0, x_1] := \frac{y_1 - y_0}{x_1 - x_0}$ , možemo pisati

$$P_1(x) = f[x_0] + f[x_0, x_1](x - x_0).$$

Uzmimo sada  $n = 2$ . Polinom  $P_2$ , čiji graf treba prolaziti točkama  $T_0, T_1, T_2$  ima oblik

$$P_2(x) = f[x_0] + f[x_0, x_1](x - x_0) + \alpha(x - x_0)(x - x_1). \quad (2.12)$$

Očigledno je  $P_2(x_0) = y_0$  i  $P_2(x_1) = y_1$ . Parametar  $\alpha$  treba odrediti tako da bude  $P_2(x_2) = y_2$ . Ako u (2.12) uvrstimo  $x = x_2$ , dobivamo

$$\alpha = \frac{y_2 - y_0}{(x_2 - x_1)(x_2 - x_0)} - \frac{y_1 - y_0}{(x_1 - x_0)(x_2 - x_1)} = \frac{y_2 - y_1}{(x_2 - x_1)(x_2 - x_0)} - \frac{y_1 - y_0}{(x_1 - x_0)(x_2 - x_0)}.$$

Ako uvedemo oznake

$$f[x_0, x_1] := \frac{y_1 - y_0}{x_1 - x_0}, \quad f[x_1, x_2] := \frac{y_2 - y_1}{x_2 - x_1}, \quad f[x_0, x_1, x_2] := \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0}$$

onda je očigledno  $\alpha = f[x_0, x_1, x_2]$ , pa polinom (2.12) glasi

$$P_2(x) = f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1).$$

Općenito, za  $(n + 1)$  točku  $T_i(x_i, y_i)$ ,  $i = 0, 1, \dots, n$  dobivamo

$$P_n(x) = f[x_0] + f[x_0, x_1](x - x_0) + \cdots + f[x_0, x_1, \dots, x_n](x - x_0) \cdots (x - x_{n-1}), \quad (2.13)$$

gdje brojeve

$$f[x_0, x_1, \dots, x_i] = \frac{f[x_1, \dots, x_i] - f[x_0, x_1, \dots, x_{i-1}]}{x_i - x_0}, \quad i = 1, \dots, n. \quad (2.14)$$

zovemo podijeljene razlike, a polinom (2.13) Newtonov oblik interpolacijskog polinoma  $n$ -tog reda.

**Primjer 2.4** Treba odrediti Newtonov oblik interpolacijskog polinoma koji će proći točkama  $T_0(-2, -5)$ ,  $T_1(2, 3)$ ,  $T_2(4, 211)$  i izračunati njegovu vrijednost za  $x = 1$ .

Koeficijente  $f[x_0, x_1]$  i  $f[x_0, x_1, x_2]$  koji su potrebni u formuli (2.13) izračunat ćemo prema sljedećoj shemi:

$$\begin{array}{llll} x_0 = -2 & f[x_0] = -5 & & \\ & & f[x_0, x_1] = 2 & \\ x_1 = 2 & f[x_1] = 3 & & f[x_0, x_1, x_2] = 17 \\ & & f[x_1, x_2] = 104 & \\ x_2 = 4 & f[x_2] = 211 & & \end{array}$$

Prema (2.13) imamo  $P_2(x) = -5 + 2(x + 2) + 17(x + 2)(x - 2)$ . Vrijednost  $P_2(1)$  treba računati prema (2.11):

$$P_2(1) = ((17(1 - 2) + 2)(1 + 2) - 5 = -50.$$

Rezultate provjerite niže navedenim *Mathematica*-programom za izračunavanje podijeljenih razlika prema (2.14) i vrijednosti polinoma u točki prema (2.11).

```
In[1]:= (* Podijeljene razlike *)
n=3; x={-2,2,4}; y={-5,3,211}; z=1; a = y; t = a;
Do[kk=k-1;
  Do[ii=n-i-kk;
    t[[ii]]=(y[[ii+1]]-y[[ii]])/(x[[ii+kk+1]]-x[[ii]]), {i,n-1}];
    a[[k+1]]=t[[1]]; y = t, {k,n-1}]; a

Out[1]:= {-5,2,17}

In[2]:= (* Vrijednost polinoma u točki z - Hornerova shema *)
p = a[[n]];
Do[ii=n-i; p=p (z - x[[ii]]) + a[[ii]], {i,n-1}]; p

Out[2]:= -50
```

**Zadatak 2.1** Izvedite Newtonov oblik interpolacijskog polinoma za jednoliko raspoređene čvorove, tj. ako je  $h := x_1 - x_0 = x_2 - x_1 = \dots = x_n - x_{n-1}$ . Pokažite da u tom slučaju vrijedi

$$f[x_0, x_1, \dots, x_k] = \frac{\Delta^k y_0}{h^k k!}, \quad k = 0, 1, \dots, n,$$

gdje je:  $\Delta^1 y_i = y_{i+1} - y_i, \dots, \Delta^k y_i = \Delta^{k-1} y_{i+1} - \Delta^{k-1} y_i, i = 0, 1, \dots, n - 1$ .

## 2.1.3 Ocjena pogreške

**Teorem 2.1** Neka je  $f \in C_{[a,b]}^{n+1}$  funkcija čije vrijednosti su poznate u  $(n+1)$  točaka  $x_i$ ,  $i = 0, 1, \dots, n$ ,

$$a = x_0 < x_1 < \dots < x_n = b, \quad y_i = f(x_i), \quad i = 0, 1, \dots, n$$

i neka je  $P_n$  odgovarajući interpolacijski polinom.

Tada za svaki  $\bar{x} \in [a, b]$  postoji  $\xi \in (a, b)$ , tako da je

$$f(\bar{x}) - P_n(\bar{x}) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \omega(\bar{x}), \quad \omega(\bar{x}) = (\bar{x} - x_0) \cdots (\bar{x} - x_n). \quad (2.15)$$

**Dokaz.** Za  $\bar{x} = x_i$  tvrdnja je očigledna. Za  $\bar{x} \neq x_i$ ,  $i = 0, 1, \dots, n$  definirajmo pomoću funkciju

$$g(x) := f(x) - P_n(x) - k\omega(x), \quad (2.16)$$

gdje ćemo konstantu  $k$  odrediti tako da bude  $g(\bar{x}) = 0$ . Na taj način funkcija  $g$  imat će barem  $(n+2)$  nultočke:  $\bar{x}, x_0, x_1, \dots, x_n$ . Prema Rolleovom teoremu funkcija  $g'$  ima barem  $(n+1)$  nultocku, funkcija  $g''$  barem  $n$  nultocka, ..., a funkcija  $g^{(n+1)}$  ima barem jednu nultocku  $\xi \in (a, b)$ .

Primijetite da je  $\frac{d^{n+1}}{dx^{n+1}} P_n(x) = 0$ . Osim toga, kako je  $\omega$  polinom  $(n+1)$ -og stupnja s vodećim koeficijentom 1, imamo  $\frac{d^{n+1}}{dx^{n+1}} \omega(x) = (n+1)!$ . Zato je

$$0 = g^{(n+1)}(\xi) = f^{(n+1)}(\xi) - k(n+1)!$$

odakle slijedi  $k = \frac{f^{(n+1)}(\xi)}{(n+1)!}$ . Na taj način odredili smo konstantu  $k$  tako da je  $g(\bar{x}) = 0$ . Zato iz (2.16) slijedi (2.15).  $\square$

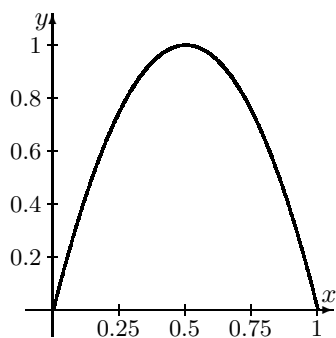
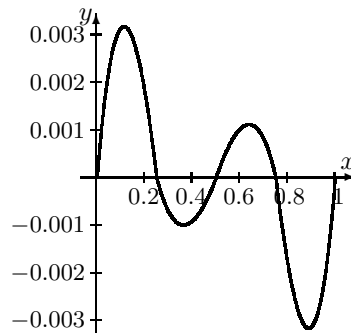
Ako označimo  $M_{n+1} := \max_{x \in [a,b]} |f^{(n+1)}(x)|$ , onda iz (2.15) slijedi

$$|f(x) - P_n(x)| \leq \frac{M_{n+1}}{(n+1)!} |\omega(x)|, \quad (2.17)$$

što znači da pogreška aproksimacije ovisi o ponašanju funkcije  $\omega$ . To ćemo analizirati u sljedeća dva primjera.

**Primjer 2.5** Funkciju  $f(x) = \sin(\pi x)$  na intervalu  $[0, 1]$  treba aproksimirati interpolacijskim polinomom koji će prolaziti sljedećim točkama

$x$	0	0.25	0.5	0.75	1
$f(x)$	0	0.707107	1	0.707107	0

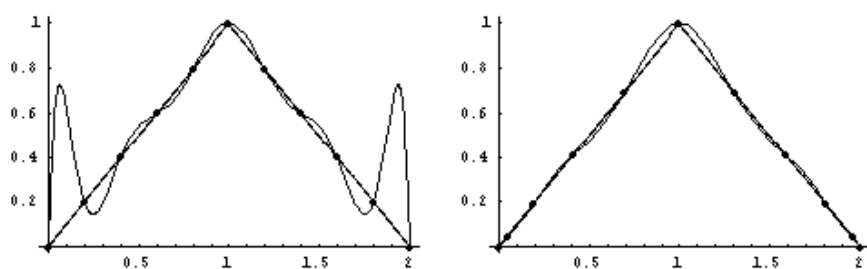
Slika 2.3.a  $f(x) = \sin \pi x$ Slika 2.3.b  $\omega(x) = \prod_{i=0}^4 (x - x_i)$ 

Prema (2.13) dobivamo Newtonov oblik interpolacijskog polinoma  $P_4$  s koeficijentima:  $f[x_0] = 0$ ,  $f[x_0, x_1] = 2.82843$ ,  $f[x_0, x_1, x_2] = -3.31371$ ,  $f[x_0, x_1, x_2, x_3] = -1.83011$ ,  $f[x_0, x_1, x_2, x_3, x_4] = 3.66022$ . Na *Slici 2.3* prikazani su grafovi funkcija  $f$  i  $\omega$ . Uočljiva su relativno značajnija odstupanja na rubovima promatranog intervala, ali su maksimalna odstupanja malena:

$$\|f - P_4\|_{\infty} = \max_{x \in [a, b]} |f(x) - P_4(x)| \leq 0.00027$$

**Primjer 2.6** Zadana je funkcija  $f(x) = 1 - |x - 1|$ ,  $x \in [0, 2]$ . Treba pronaći interpolacijski polinom koji će prolaziti točkama  $(x_i, f(x_i))$ ,  $x_i = \frac{2i}{n}$ ,  $i = 0, 1, \dots, n$ .

Na *Slici 2.4* lijevo prikazan je graf funkcije  $f$  i odgovarajućeg interpolacijskog polinoma za  $n = 11$ . Vidi se da na rubovima područja interpolacije dolazi do jakih oscilacija, zbog čega je ovakav interpolacijski polinom praktično neuporabiv.



Slika 2.4 Interpolacija funkcije  $f(x) = 1 - |x - 1|$ ,  $x \in [0, 2]$ . Lijevo je prikazana interpolacija u jednoliko raspoređenim čvorovima, a desno interpolacija u čvorovima koji su nultočke Čebiševljevog polinoma  $T_{10}$

Pogreška (2.17) kod interpolacijskog polinoma ovisi o ponašanju funkcije  $\omega$ . Za jednoliko raspoređene čvorove i nešto veći  $n$  pokazuje se (vidi također *Primjer 2.6*) da funkcija  $\omega$  na rubovima područja interpolacije ima jake oscilacije, što ukazuje na pojavu velikih grešaka kod interpolacije. Čvorove interpolacije  $x_0, x_1, \dots, x_n$

trebalo bi birati tako da bude

$$\max_{a \leq x \leq b} |\omega(x)| \longrightarrow \min. \quad (2.18)$$

Primijetite da je  $\omega$  normirani polinom stupnja  $n+1$ . Prema *Teoremu 5.1*, str. 101, minimum u (2.18) postići će se i biti jednak  $2^{-n}$  ako je  $\omega(x) = 2^{-n}T_{n+1}$ , gdje je  $T_{n+1}$   $(n+1)$ -vi Čebiševljev polinom<sup>1</sup> ekspaniran na interval  $[a, b]$ . To drugim riječima znači da optimalan izbor čvorova interpolacije predstavljaju nul-točke polinoma  $T_{n+1}$ :

$$x_k = \frac{a+b}{2} + \frac{b-a}{2} \cos\left(\frac{n-k}{n}\pi\right), \quad k = 0, 1, \dots, n \quad (2.19)$$

One su gušće raspoređene pri rubovima intervala  $[a, b]$ . U tom slučaju interpolacijski polinom  $P_n$  imat će najmanju moguću pogrešku prema (2.17)

$$|f(x) - P_n(x)| \leq \frac{M_{n+1}}{2^n(n+1)!}, \quad x \in [a, b].$$

Na *Slici 2.4* desno prikazan je interpolacijski polinom za funkciju iz *Primjera 2.6* s čvorovima u nultočkama Čebiševljevog polinoma  $T_{10}$  rastegnuto na interval  $[0, 2]$ .

Najviše zbog navedenih problema kod interpolacije prije 30-tak godina pojavile su se tzv. spline-interpolacije.

**Primjedba 2.2** Za primjene je također zanimljiva tzv. *Hermiteova interpolacija*. U čvorovima interpolacije zadaje se vrijednost funkcije i njene derivacije (“nagib tangente”). O rješavanju ovakvih problema također se može vidjeti u niže navedenoj literaturi.

**Primjedba 2.3** Postoji vrlo opsežna literatura o interpolaciji funkcija (vidi primjerice IVANŠIĆ (1998), STOER (1993), DEMIDOVICH (1981), STOER (1994), ORTEGA (1981) i SCHWARZ (1986). Odgovarajuća FORTRAN – podrška može se naći kod PRESS (1992), a posebno u THE NAG LIBRARY (PHILIPS (1986)).

## 2.2 Spline interpolacija

Ako je broj čvorova interpolacije velik, odgovarajući interpolacijski polinom je visokog stupnja i kao takav neuporabiv u primjenama (vidi *Primjer 2.6*). Umjesto interpolacijskog polinoma, možemo pokušati interpolirati funkciju  $\varphi : [a, b] \rightarrow \mathbb{R}$ , takvu da je

$$\varphi|_{[x_{i-1}, x_i]} = \varphi_i,$$

pri čemu je  $\varphi_i : [x_{i-1}, x_i] \rightarrow \mathbb{R}$  neka jednostavnija funkcija (primjerice polinom 1. stupnja) takva da je

$$\varphi_i(x_{i-1}) = y_{i-1}, \quad \varphi_i(x_i) = y_i$$

---

<sup>1</sup>Čebiševljev polinom stupnja  $n$  programom *Mathematica* dobiva se naredbom `ChebyshevT[n, x]`



### 2.2.1 Linearni interpolacijski spline

Pretpostavimo da je  $f : [a, b] \rightarrow \mathbb{R}$  neprekidna funkcija i da poznamo njene vrijednosti  $y_0, y_1, \dots, y_n$  u  $(n + 1)$  točaka-čvorova

$$a = x_0 < x_1 < \dots < x_n = b.$$

Funkciju  $f$  interpolirat ćemo neprekidnom po dijelovima linearnom funkcijom  $\varphi : [a, b] \rightarrow \mathbb{R}$ . Lako se vidi da takva funkcija postoji i da je jedinstvena. Za  $i = 1, \dots, n$  definirajmo

$$\begin{aligned} \varphi|_{[x_{i-1}, x_i]} &= \varphi_i, \\ \varphi_i(x) &= y_{i-1} + \frac{y_i - y_{i-1}}{x_i - x_{i-1}}(x - x_{i-1}), \quad x \in [x_{i-1}, x_i]. \end{aligned} \quad (2.20)$$

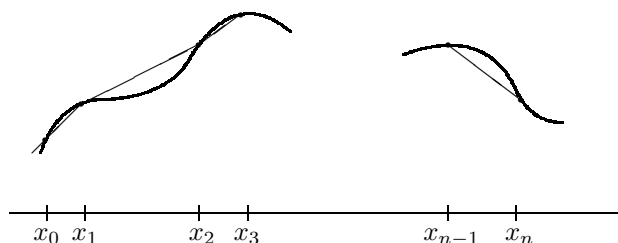
Kako je

$$\varphi_i(x_i) = y_i = \varphi_{i+1}(x_i), \quad i = 1, \dots, n-1,$$

funkcija  $\varphi$  je neprekidna na čitavom intervalu  $[a, b]$ , a linearna na svakom podintervalu  $[x_{i-1}, x_i]$ . Kako je osim toga

$$\varphi(x_0) = \varphi_1(x_0) = y_0, \quad \varphi(x_i) = \varphi_i(x_i) = y_i, \quad i = 1, \dots, n,$$

ona interpolira funkciju  $f$  u čvorovima  $x_0, x_1, \dots, x_n$ . Nazivamo je **linearni interpolacijski spline**<sup>2</sup> (vidi *Sliku 2.5*).



Slika 2.5. Linearni interpolacijski spline

Međutim, linearni interpolacijski spline nije derivabilna funkcija u čvorovima  $x_0, x_1, \dots, x_n$  i stoga nije pogodan za daljnju uporabu.

Linearni interpolacijski spline možemo definirati i na drugi način - analogno kako smo u t.2.1.1, str. 18 uveli Lagrangeov oblik interpolacijskog polinoma. U tu svrhu definirajmo tzv. **hat-funkcije**  $C_i$ ,  $i = 1, \dots, n-1$ .

$$C_i(x) = \begin{cases} (x - x_{i-1})/(x_i - x_{i-1}), & x \in [x_{i-1}, x_i] \\ (x_{i+1} - x)/(x_{i+1} - x_i), & x \in [x_i, x_{i+1}] \\ 0, & \text{inače} \end{cases}$$

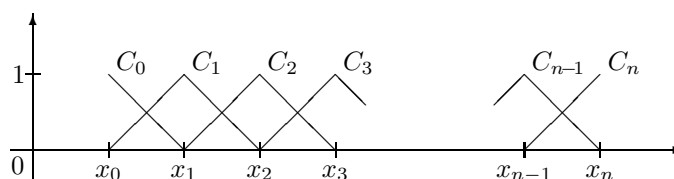
$$C_0(x) = \begin{cases} \frac{x_1 - x}{x_1 - x_0}, & x \in [x_0, x_1] \\ 0, & \text{inače} \end{cases} \quad C_n(x) = \begin{cases} \frac{x - x_{n-1}}{x_n - x_{n-1}}, & x \in [x_{n-1}, x_n] \\ 0, & \text{inače} \end{cases}$$

<sup>2</sup>Naziv dolazi od engleske riječi **spline** [splain], koja označava dug, savitljiv, uzak drveni klin

Hat-funkcije  $C_i$  imaju svojstvo

$$C_i(x_j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

(usporedi s (2.6) str. 18), a njihovi grafovi prikazani su na *Slici 2.6*.



Slika 2.6. Hat-funkcije

Sada linearni interpolacijski spline možemo definirati formulom

$$\varphi(x) = \sum_{i=0}^n y_i C_i(x) \quad (2.21)$$

(usporedi s (2.9), str. 19).

**Primjedba 2.4** Ako s  $\mathcal{L}$  označimo skup svih linearnih interpolacijskih spline-ova određenih čvorovima  $a = x_0 < x_1 < \dots < x_n = b$ , onda je lako vidjeti da je  $\mathcal{L}$  vektorski potprostor na prostoru  $C_{[a,b]}$  svih neprekidnih funkcija na intervalu  $[a, b]$ . Definirajmo linearni operator  $L : C_{[a,b]} \rightarrow \mathcal{L}$ ,  $L(f) = \varphi$ , koji neprekidnoj funkciji  $f$  pridružuje linearni interpolacijski spline  $\varphi$  u čvorovima  $x_0, \dots, x_n$ . Specijalno, za neki linearni interpolacijski spline  $\varphi \in \mathcal{L} \subset C_{[a,b]}$  vrijedi  $L(\varphi) = \varphi$ , pa onda i općenito za svaki  $f \in C_{[a,b]}$  vrijedi

$$L(L(f)) = L(f), \quad \text{tj.} \quad L^2 = L.$$

Dakle, linearni operator  $L$  je projektor.

**Zadatak 2.2** Pokažite da jednu bazu u prostoru  $\mathcal{L}$  čine hat-funkcije  $\{C_0, C_1, \dots, C_n\}$ , te da za svaki  $f \in C_{[a,b]}$  vrijedi

$$L(f) = \sum_{i=0}^n y_i C_i,$$

gdje je  $y_i = f(x_i)$ ,  $i = 0, 1, \dots, n$ .

**Zadatak 2.3** Izračunajte integral funkcije (2.21) u granicama od  $x_0$  do  $x_n$ . Pokažite da za ekvidistantno raspoređene čvorove dobivamo poznato generalizirano trapezno pravilo (vidi t.7.1, str. 125).

Uz pretpostavku da je funkcija  $f \in C_{[a,b]}^2$ , prema *Teoremu 2.1*, str. 23, za svaki  $x \in [x_{i-1}, x_i]$  postoji  $\xi_i \in (x_{i-1}, x_i)$ , takav da vrijedi

$$f(x) - C_i(x) = \frac{1}{2} f''(\xi_i)(x - x_{i-1})(x - x_i). \quad (2.22)$$

Označimo  $h_i := x_i - x_{i-1}$ . Budući da kvadratna funkcija  $x \mapsto (x - x_{i-1})(x - x_i)$  postiže minimum u točki  $x_0 = \frac{x_{i-1} + x_i}{2}$ , onda za svaki  $x \in [x_{i-1}, x_i]$  vrijedi

$$|(x - x_{i-1})(x - x_i)| \leq \left| \left( \frac{x_{i-1} + x_i}{2} - x_{i-1} \right) \left( \frac{x_{i-1} + x_i}{2} - x_i \right) \right| \leq \frac{h_i^2}{4}.$$

Ako označimo

$$h_{max} := \max_{i=0, \dots, n} h_i \quad M_2 := \max_{x \in [a, b]} |f''(x)|,$$

tada prema (2.22) vrijedi ocjena pogreške za linearni interpolacijski spline

$$|f(x) - L(f)(x)| \leq \frac{M_2}{8} h_{max}^2 \quad (2.23)$$

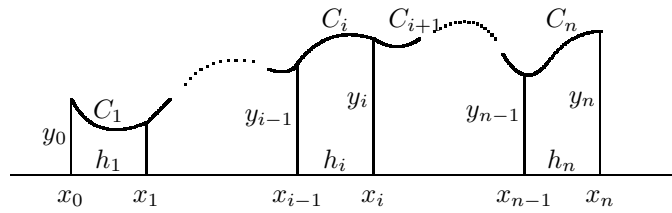
### 2.2.2 Kubični interpolacijski spline

Neprekidnu funkciju  $f : [a, b] \rightarrow \mathbb{R}$  čije vrijednosti  $y_0, y_1, \dots, y_n$  poznajemo u  $(n + 1)$  čvorova  $a = x_0 < x_1 < \dots < x_n = b$ , interpolirat ćemo funkcijom  $C : [a, b] \rightarrow \mathbb{R}$ ,

$$C(x) = C_i(x), \quad x \in [x_{i-1}, x_i], \quad i = 1, \dots, n, \quad (2.24)$$

gdje su  $C_i$ ,  $i = 1, \dots, n$ , kubični polinomi koji trebaju zadovoljavati sljedeće uvjete (usporedi *Sliku 2.7*).

- (i)  $C_i(x_{i-1}) = y_{i-1}$ ,  $i = 1, \dots, n$ ;
- (ii)  $C_i(x_i) = y_i$ ,  $i = 1, \dots, n$ ;
- (iii)  $C'_i(x_i) = C'_{i+1}(x_i)$ ,  $i = 1, \dots, n - 1$ ;
- (iv)  $C''_i(x_i) = C''_{i+1}(x_i)$ ,  $i = 1, \dots, n - 1$ .



Slika 2.7. Kubični interpolacijski spline

Uvjeti (i) – (iv) osiguravaju da će funkcija  $C$  zadovoljiti interpolacijske uvjete u čvorovima  $x_0, x_1, \dots, x_n$ , ali i da će biti klase  $C^2_{[a, b]}$ . Budući da  $n$  kubičnih polinoma  $C_1, \dots, C_n$  ima ukupno  $4n$  neodređenih koeficijenata, a zahtjevima (i) – (iv) određeno je  $4n - 2$  uvjeta, za jednoznačno određivanje funkcije  $C$  potrebno je postaviti još dva dodatna uvjeta. Najčešće zahtijevamo još da bude

$$C''_1(x_0) = C''_n(x_n) = 0, \quad (2.25)$$

iako je moguće zadati i neke druge zahtjeve. Funkciju  $C$  koja zadovoljava uvjete (i)-(iv), te dodatne uvjete (2.25) nazivamo **prirodni kubični interpolacijski spline**. Funkcija  $C$  je klase  $C_{[a,b]}^2$ , koja na intervalu  $[a, b]$  interpolira funkciju  $f$  poznatu u čvorovima  $x_0, x_1, \dots, x_n$ .

**Primjedba 2.5** Problem određivanja koeficijenata kubičnih polinoma  $C_i$  na taj način sveo se na rješavanje sustava linearnih jednačbi. Primijetimo da već za 50 čvorova treba riješiti sustav od 200 jednačbi s 200 nepoznanica. Ako je pri tome još matrica sustava loše uvjetovana (vidi t.3.2, str. 41), mogu nastupiti ozbiljni numerički problemi.

**Teorem 2.2** Zadani su podaci  $(x_i, y_i)$ ,  $i = 0, 1, \dots, n$  koji zadovoljavaju uvjet  $x_0 < x_1 < \dots < x_n$ . Tada postoji jedinstveni prirodni kubični interpolacijski spline  $C$ , pri čemu su polinomi  $(C_i, i = 1, \dots, n)$  zadani s

$$C_i(x) = \left( y_{i-1} - s_{i-1} \frac{h_i^2}{6} \right) + b_i(x - x_{i-1}) + \frac{s_{i-1}}{6h_i}(x - x_{i-1})^3 + \frac{s_i}{6h_i}(x - x_{i-1})^3 \quad (2.26)$$

gdje je

$$b_i = d_i - (s_i - s_{i-1}) \frac{h_i}{6},$$

$$d_i = \frac{y_i - y_{i-1}}{h_i}, \quad h_i = x_i - x_{i-1},$$

a brojevi  $(s_i, i = 0, \dots, n)$  zadovoljavaju sustav jednačbi

$$s_{i-1}h_i + 2s_i(h_i + h_{i+1}) + s_{i+1}h_{i+1} = 6(d_{i+1} - d_i), \quad i = 1, \dots, n-1$$

$$s_0 = s_n = 0. \quad (2.27)$$

**Dokaz.** Za  $i = 1, \dots, n$  označimo

$$s_{i-1} := C_i'''(x_{i-1}), \quad s_i := C_i'''(x_i).$$

Primijetimo da je zbog (2.25)  $s_0 = s_n = 0$ . Kako je  $x \rightarrow C_i''(x)$  linearna funkcija čiji graf treba prolaziti točkama  $(x_{i-1}, s_{i-1})$ ,  $(x_i, s_i)$ , mora biti

$$C_i''(x) = \frac{s_{i-1}}{h_i}(x_i - x) + \frac{s_i}{h_i}(x - x_{i-1}) \quad (2.28)$$

Primitivna funkcija funkcije (2.28) je

$$C_i'(x) = b_i - \frac{s_{i-1}}{2h_i}(x_i - x)^2 + \frac{s_i}{2h_i}(x - x_{i-1})^2, \quad (2.29)$$

gdje je  $b_i$  konstanta. Nadalje, primitivna funkcija funkcije (2.29) je

$$C_i(x) = a_i + b_i x + \frac{s_{i-1}}{6h_i}(x_i - x)^3 + \frac{s_i}{6h_i}(x - x_{i-1})^3, \quad (2.30)$$

gdje je  $a_i$  konstanta. Konstante  $a_i$ ,  $b_i$  odredit ćemo iz uvjeta (i)-(ii):

$$C_i(x_{i-1}) = y_{i-1}, \quad C_i(x_i) = y_i, \quad i = 1, \dots, n$$

tj.

$$a_i + b_i x_{i-1} + s_{i-1} \frac{h_i^2}{6} = y_{i-1}, \quad (2.31)$$

$$a_i + b_i x_i + s_i \frac{h_i^2}{6} = y_i. \quad (2.32)$$

Oduzimajući (2.31) od (2.32) dobivamo

$$b_i = d_i - (s_i - s_{i-1}) \frac{h_i}{6}, \quad d_i = \frac{y_i - y_{i-1}}{h_i}. \quad (2.33)$$

Iz (2.31) dobivamo

$$a_i = -b_i x_{i-1} + y_{i-1} - s_{i-1} \frac{h_i^2}{6},$$

iz čega neposredno slijedi (2.26).

Još treba odrediti brojeve  $s_1, \dots, s_{n-1}$ . Njih ćemo odrediti iz uvjeta (iii). Iz (2.29) dobivamo

$$C'_i(x_i) = b_i + \frac{s_i}{2} h_i, \quad C'_{i+1}(x_i) = b_{i+1} - \frac{s_i}{2} h_{i+1}.$$

Na taj način uvjet (iii) postaje

$$d_i - (s_i - s_{i-1}) \frac{h_i}{6} + \frac{s_i}{2} h_i = d_{i+1} - (s_{i+1} - s_i) \frac{h_{i+1}}{6} - s_i \frac{h_{i+1}}{2},$$

iz čega neposredno slijedi (2.27).

Kako je  $s_0 = s_n = 0$ , (2.27) je sustav od  $(n-1)$  jednadžbe s  $(n-1)$  nepoznanicom:  $s_1, \dots, s_{n-1}$ , koji možemo pisati

$$\mathbf{H}\mathbf{s} = \mathbf{b}, \quad (2.34)$$

$$\mathbf{H} = \begin{bmatrix} 2(h_1 + h_2) & h_2 & \cdots & 0 & 0 \\ h_2 & 2(h_2 + h_3) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 2(h_{n-2} + h_{n-1}) & h_{n-1} \\ 0 & 0 & \cdots & h_{n-1} & 2(h_{n-1} + h_n) \end{bmatrix},$$

$$\mathbf{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_{n-1} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 6(d_2 - d_1) \\ \vdots \\ 6(d_n - d_{n-1}) \end{bmatrix}.$$

Primijetimo da je matrica sustava (2.34) trodijagonalna, dijagonalno dominantna i simetrična matrica, pa kao takva pozitivno definitna. Zato postoji jedinstveno rješenje, a onda i jedinstveni prirodni kubični interpolacijski spline.  $\square$

**Primjedba 2.6** Ako je  $f \in C^4[a, b]$  i  $h_{\max} = \max_i h_i$ , onda vrijedi (vidi STEWART (1997))

$$\begin{aligned}\|f - C\|_{\infty} &\leq \frac{5}{384} h_{\max}^4 \\ \|f' - C'\|_{\infty} &\leq \frac{1}{24} h_{\max}^3 \\ \|f'' - C''\|_{\infty} &\leq \frac{3}{8} h_{\max}^2\end{aligned}$$

### Algoritam SPLINE3

Korak 1. Učitati podatke  $(x_i, y_i)$ ,  $i = 0, 1, \dots, n$ ;

Korak 2. Izračunati

$$h_i = x_i - x_{i-1}, \quad d_i = \frac{y_i - y_{i-1}}{h_i}, \quad i = 1, \dots, n-1$$

i riješiti sustav (2.34);

Korak 3. Izračunati  $b_i = d_i - (s_i - s_{i-1}) \frac{h_i}{6}$ ,  $i = 1, \dots, n$  i definirati polinome (2.26).

Prilikom rješavanja sustava (2.34) treba iskoristiti specijalnu strukturu matrice sustava (vidi *Zadatak 3.24*, str. 65). Specijalno, ako su razmaci između čvorova ekvidistantni i iznose  $h$ , onda imamo

$$\mathbf{H} = \begin{bmatrix} 4 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 4 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 4 & 1 \\ 0 & 0 & 0 & \cdots & 1 & 4 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \frac{6}{h}(d_2 - d_1) \\ \vdots \\ \frac{6}{h}(d_n - d_{n-1}) \end{bmatrix}.$$

**Zadatak 2.4** Izradi program koji će prema algoritmu SPLINE3 izračunavati prirodni kubični interpolacijski spline.

**Primjer 2.7** Tablično je zadana funkcija

$x_i$	0	0.25	0.5	0.75	1
$y_i$	1	2	1	0	1

U ovom slučaju je  $n = 4$ , a svi  $h_i$  jednaki su 0.25. Sustav (2.34) glasi

$$\begin{bmatrix} 1 & 0.25 & 0 \\ 0.25 & 1 & 0.25 \\ 0 & 0.25 & 1 \end{bmatrix} \mathbf{s} = \begin{bmatrix} -48 \\ 0 \\ 48 \end{bmatrix}.$$

a njegovo rješenje je:  $s_1 = -48$ ,  $s_2 = 0$ ,  $s_3 = 48$ . Osim toga, iz uvjeta (2.25) imamo:  $s_0 = s_4 = 0$ . Sada iz (2.26) možemo izračunati četiri kubična polinoma  $C_i$ ,  $i = 1, 2, 3, 4$ , koji grade traženi prirodni kubični interpolacijski spline

$$C(x) = \begin{cases} C_1(x) = 1 + 6x - 32x^3, & 0 \leq x \leq 0.25 \\ C_2(x) = 18x - 48x^2 + 32x^3, & 0.25 \leq x \leq 0.5 \\ C_3(x) = 18x - 48x^2 + 32x^3, & 0.5 \leq x \leq 0.75 \\ C_4(x) = 27 - 90x + 96x^2 - 32x^3, & 0.75 \leq x \leq 1 \end{cases}$$

Ako želimo primjerice izračunati vrijednost spline-a  $C$  u točki 0.7, onda prvo moramo ustanoviti da je  $0.7 \in [0.5, 0.75]$ . Tada je  $C(0.7) = C_3(0.7) = 0.056$ .

**Primjedba 2.7** Prirodni kubični interpolacijski spline  $C$ , koji interpolira funkciju  $f$  u čvorovima  $a = x_0 < x_1 < \dots < x_n = b$  ima svojstvo da za svaku funkciju  $g \in C^2[a, b]$ , koja također interpolira funkciju  $f$  u navedenim čvorovima, vrijedi

$$\int_a^b (g''(x))^2 dx \geq \int_a^b (C''(x))^2 dx. \quad (*)$$

Naime,

$$\begin{aligned} 0 &\leq \int_a^b (g''(x) - C''(x))^2 dx = \int_a^b (g''(x))^2 dx - 2 \int_a^b g''(x) C''(x) dx + \int_a^b (C''(x))^2 dx \\ &= \int_a^b (g''(x))^2 dx - \int_a^b (C''(x))^2 dx - 2 \int_a^b C''(x) (C''(x) - g''(x)) dx. \end{aligned}$$

Pri tome, zbog  $C'''(a) = C'''(b) = 0$ , parcijalnom integracijom dobivamo

$$\begin{aligned} \int_a^b C''(x) (C''(x) - g''(x)) dx &= [C''(x) (C'(x) - g'(x))]_a^b \\ &\quad - \int_a^b C'''(x) (C'(x) - g'(x)) dx \\ &= - \int_a^b C'''(x) (C'(x) - g'(x)) dx. \end{aligned}$$

Kako je  $C'''$  po dijelovima konstantna funkcija, imamo

$$\begin{aligned} \int_a^b C''(x) (C''(x) - g''(x)) dx &= - \sum_{i=1}^{n-1} C'''(x_i+) \int_{x_i}^{x_{i+1}} (C'(x) - g'(x)) dx \\ &= - \sum_{i=1}^{n-1} C'''(x_i+) (C(x_{i+1}) - g(x_{i+1}) - C(x_i) + g(x_i)) = 0. \end{aligned}$$

Svojstvo (\*) nazivamo monimalno svojstvo prirodnog kubičnog spline-a. To svojstvo vrijedi i za kubični spline s drukčijim rubnim uvjetima.

**Primjedba 2.8** Spline-interpolacije pojavile su se u matematičkoj literaturi u posljednjih tridesetak godina, a zbog mogućnosti učinkovite primjene kod rješavanja običnih i parcijalnih diferencijalnih jednadžbi i u mnogim drugim situacijama, i sada se intenzivno proučavaju. Kvalitetna i pregledna knjiga s odgovarajućom programskom podrškom (FORTRAN) iz ovog područja je DE BOOR (1978), ali se osnovni pojmovi i neke primjene mogu vidjeti i u IVANŠIĆ (1998), ORTEGA (1981), SCHWARZ (1986), STEWART (1998), FORSYTHE (1974) itd. Program Matematika omogućava izravno korištenje spline-a bilo kojeg reda razumne veličine. Odgovarajuća FORTRAN programska podrška može se naći u PRESS (1992), kao i u THE NAG LIBRARY (PHILIPS (1986)). O interpolaciji funkcija jedne i više varijabli može se vidjeti primjerice kod LANCASTER (1986).

## 2.3 Zadaci

**Zadatak 2.5** Odredite interpolacijski polinom za sljedeće tablično zadane funkcije

$$\begin{array}{ll} a) \begin{array}{c|ccc} x & 2 & 5 & 10 \\ y & 51 & 150 & 435 \end{array} & b) \begin{array}{c|cccc} x & -1 & 0 & 1 & 3 \\ y & -8 & 2 & 6 & 44 \end{array} \\ c) \begin{array}{c|cccc} x & -2 & 1 & 2 & 5 \\ y & 34 & -2 & -10 & 62 \end{array} & d) \begin{array}{c|ccccc} x & 0 & 1 & 2 & 5 & -2 \\ y & -8 & -1 & 38 & 1127 & -70 \end{array} \end{array}$$

Rješenje:

$$\begin{array}{ll} a) P_2(x) = 3x^2 + 12x + 15 & b) P_3(x) = 2x^3 - 3x^2 + 5x + 2 \\ c) P_3(x) = x^3 - 15x + 12 & d) P_4(x) = x^4 + 5x^3 - 6x^2 + 7x - 8 \end{array}$$

**Zadatak 2.6** Za tablično zadane funkcije odredite Newtonov oblik interpolacijskog polinoma te izračunajte njegovu vrijednost za  $\xi = 10$ .

$$\begin{array}{ll} a) \begin{array}{c|ccc} x & 2 & 4 & 5 \\ y & 91 & 171 & 220 \end{array} & b) \begin{array}{c|cccc} x & -3 & 0 & 2 & 4 \\ y & -28 & 8 & 52 & 224 \end{array} \\ c) \begin{array}{c|cccc} x & -2 & -1 & 0 & 1 \\ y & -4 & -6 & -2 & 2 \end{array} & d) \begin{array}{c|ccccc} x & -2 & 0 & 2 & 3 & 4 \\ y & 88 & 6 & 20 & 108 & 370 \end{array} \end{array}$$

Rješenje:

$$\begin{array}{ll} a) P_2(x) = 91 + 40(x-2) + 3(x-2)(x-4) & p_2(10) = 555 \\ b) P_3(x) = -28 + 12(x+3) + 2(x+3)x + 2(x+3)x(x-2) & P_3(10) = 2468 \\ c) P_3(x) = -4 - 2(x+2) + 3(x+2)(x+1) - 1(x+2)(x+1)x & P_3(10) = -952 \\ d) P_4(x) = 88 - 41(x+2) + 12(x+2)x + 3(x+2)x(x-2) + 2(x+2)x(x-2)(x-3) & P_4(10) = 17356 \end{array}$$

**Zadatak 2.7** Zadane su vrijednosti funkcije  $f(x) = 3x^2 + \pi^{-4} \ln(\pi - x)^2 + 1$  u sljedećim točkama

$$\begin{array}{c|cccc} x & 3.13 & 3.14 & 3.15 & 3.16 \\ \hline f(x) & 30.29918 & 30.44653 & 30.66939 & 30.87478 \end{array}$$

Odredite interpolacijski polinom, čiji će graf prolaziti ovim točkama, te izračunajte njegovu vrijednost u točki  $\xi = 3.1416$ .

Rješenje:  $P_3(3.1416) = 30.32619$ ,

$$P_3(x) = 483453.48 - 460723.76x + 146356.15x^2 - 15496.67x^3$$

**Zadatak 2.8** Zadane su vrijednosti  $\Gamma$ -funkcije u sljedećim točkama

$$\begin{array}{c|cccc} x & 0.2 & 0.4 & 0.6 & 0.8 \\ \hline \Gamma(x) & 4.5910 & 2.2182 & 1.4892 & 1.1642 \end{array}$$

Odredite interpolacijski polinom, čiji će graf prolaziti ovim točkama, te izračunajte njegovu vrijednost u točkama: 0.3, 0.5, 0.7. Izračunajte apsolutnu i relativnu pogrešku ovih aproksimacija.



Rješenje: Koeficijenti interpolacijskog polinoma su

$$a_0 = 9.8474, \quad a_1 = -35.5573, \quad a_2 = 51.5425, \quad a_3 = -25.8292$$

$x$	$\Gamma(x)$	$P(x)$	$\Delta P$
0.3	2.9917	3.122	0.1303
0.5	1.7724	1.726	0.0464
0.7	1.2980	1.354	0.0560

**Zadatak 2.9** Funkcija  $f$  zadana je u nekoliko točaka:  $x_0, x_1, \dots$ . Odredite interpolacijski polinom  $P$ , te izračunajte  $P(c_1)$  i  $P(c_2)$  ako je

- a)  $f(x) = x + 2/x$ ,  
 (1)  $x_0 = 1, x_1 = 2, x_2 = 2.5$ ,  $c_1 = 1.5, c_2 = 1.2$ ,  
 (2)  $x_0 = 0.5, x_1 = 1, x_2 = 2, x_3 = 2.5$ ,  $c_1 = 1.5, c_2 = 1.2$ ,  
 b)  $f(x) = 8x2^{-x}$ ,  
 (1)  $x_0 = 0, x_1 = 1, x_2 = 2$ ,  $c_1 = 1.5, c_2 = 1.3$ ,  
 (2)  $x_0 = 0, x_1 = 1, x_2 = 2, x_3 = 3$ ,  $c_1 = 1.5, c_2 = 1.3$ ,  
 c)  $f(x) = 2 \sin(\pi x/6)$ ,  
 (1)  $x_0 = 0, x_1 = 1, x_2 = 3$ ,  $c_1 = 2, c_2 = 2.4$ ,  
 (2)  $x_0 = 0, x_1 = 1, x_2 = 3, x_3 = 5$ ,  $c_1 = 2, c_2 = 2.4$ ,  
 (3)  $x_0 = 0, x_1 = 1, x_2 = 3$ ,  $c_1 = 4, c_2 = 3.5$ ,  
 (4)  $x_0 = 0, x_1 = 1, x_2 = 3, x_3 = 5$ ,  $c_1 = 4, c_2 = 3.5$ .

Rješenje:

- a) (1)  $P(x) = 0.4x^2 - 1.2x + 3.8$ ,  $P(1.5) = 2.9$ ,  $P(1.2) = 2.936$ ;  
 (2)  $P(x) = -0.8x^3 + 4.8x^2 - 8.8x + 7.8$ ,  $P(1.5) = 2.7$ ,  $P(1.2) = 2.7696$   
 b) (1)  $P(x) = -2x^2 + 6x$ ,  $P(1.5) = 4.5$ ,  $P(1.3) = 4.42$ ;  
 (2)  $P(x) = 0.5x^3 + 3.5x^2 + 7x$ ,  $P(1.5) = 4.125$ ,  $P(1.3) = 4.2835$   
 c) (1)  $P(x) = -\frac{1}{6}x^2 + \frac{7}{6}x$ ,  $P(2) = \frac{7}{6}$ ,  $P(2.4) = 1.84$ ;  
 (2)  $P(x) = -.01(6)x^3 - .1x^2 + 1.11(6)x$ ,  $P(2) = 1.7$ ,  $P(2.4) = 1.8736$   
 (3)  $P(x) = -.1(6)x^2 + 1.1(6)x$ ,  $P(4) = 2$ ,  $P(3.5) = 2.04167$   
 (4)  $P(x) = -.01(6)x^3 - .1x^2 + 1.11(6)x$ ,  $P(4) = 1.8$ ,  $P(3.5) = 1.96875$

**Zadatak 2.10** Za različite vrijednosti izlaznog kuta  $\alpha$  i konstantnu početnu brzinu  $v_0 = 300 \text{ ms}^{-1}$  primjenom formule (B.4), str. 84 izračunati su dometi  $X$  balističkog projektila u vakuumu

$\alpha^\circ$	5°	10°	25°	45°
$X \text{ (m)}$	1593	3138	7028	9174

Na osnovi ovih podataka odredite funkciju-polinom koja će za proizvoljnu daljinu  $X$  davati vrijednost izlaznog kuta  $\alpha$ . Pomoću ove funkcije za zadane daljine: 2000 m, 3000 m, 5000 m, 7000 m, 9000 m izračunajte izlazne kuteve  $\alpha^*$ , apsolutne i relativne pogreške.

Rješenje:  $\alpha(X) = -3.2516 + 0.00669X - 0.0000011X^2 + 1.0410^{-10}X^3$ .

$X$ (m)	$\alpha$	$\alpha^*$	$\Delta\alpha^*$
2000	6.296	6.507	0.211
3000	9.543	9.613	0.070
5000	16.512	15.407	1.105
7000	24.865	24.819	0.046
9000	39.407	42.858	3.451

**Zadatak 2.11** Olovnu kuglicu puštamo da u zraku pada s različitim visina  $h$  i mjerimo vrijeme pada  $t$

$h$ (m)	0.5	0.8	1	1.2
$t$ (s)	0.32	0.40	0.45	0.49

Odredite interpolacijski polinom

a) za funkciju  $h \mapsto t(h)$ ,      b) za funkciju  $t \mapsto h(t)$ .

Rješenje:

$$\begin{aligned} a) \quad t(h) &= 0.2257 + 0.0874h + 0.2679h^2 - 0.131h^3 \\ b) \quad h(t) &= -3.56697 + 26.7947t - 61.3122t^2 + 54.0473t^3 \end{aligned}$$

**Zadatak 2.12** Olovnu kuglicu puštamo da u ulju pada s različitim visina  $h$  i mjerimo vrijeme pada  $t$

$h$ (cm)	4	8	10	16	20
$t$ (s)	2.5	5	6.25	10	12.5

Odredite interpolacijski polinom za funkciju  $h \mapsto t(h)$ .

Rješenje: a)  $t(h) = 0.625h$ .

**Zadatak 2.13** Principom kontradikcije u dokazivanju pokažite da za dane podatke  $(x_i, y_i)$ ,  $i = 0, 1, \dots, n$ , koji zadovoljavaju uvjet  $x_0 < x_1 < \dots < x_n$ , postoji jedinstveni interpolacijski polinom stupnja  $\leq n$ .

**Zadatak 2.14** Polinom  $P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_2 x^2 + a_1 x + a_0$  može se napisati u obliku (Hornerova shema)

$$P_n(x) = (((((a_n x + a_{n-1})x + a_{n-2})x + a_{n-3}) \dots)x + a_1)x + a_0. \quad (*)$$

Izradite program kojim ćete učitati stupanj polinoma, sve njegove koeficijente, te za zadanu vrijednost  $x = a$  primjenom formule (\*) izračunati vrijednost polinoma u točki  $a$ .

**Zadatak 2.15** Izradite program kojim ćete učitati  $n$  točaka  $T_i(x_i, y_i)$ ,  $i = 0, 1, \dots, n$  i definirati linearni interpolacijski spline. Za zadani broj  $x_0 \in [x_1, x_n]$  izračunajte vrijednost linearnog interpolacijskog spline-a.

**Zadatak 2.16** Za zadane podatke  $(x_i, y_i)$ ,  $i = 0, 1, \dots, n$  odredite kvadratni interpolacijski spline i izradite odgovarajući program.

**Zadatak 2.17** Zadana je periodična funkcija  $f$  perioda  $2\pi$  i točke  $(x_i, y_i)$ ,  $i = 0, \dots, m$ ,  $x_i = -\pi + 2\pi i/m$ ,  $y_i = f(x_i)$ . Tada postoji trigonometrijski polinom

$$T_n(x) = \frac{a_0}{2} + \sum_{j=1}^n (a_j \cos(jx) + b_j \sin(jx)), \quad (2n < m),$$

$$a_j = \frac{2}{m} \sum_{k=1}^m f(x_k) \cos(jx_k), \quad b_j = \frac{2}{m} \sum_{k=1}^m f(x_k) \sin(jx_k), \quad j = 0, \dots, n,$$

takav da je

$$\sum_{k=1}^m (f(x_k) - T_n(x_k))^2 \rightarrow \min.$$

Izradite program kojim ćete na osnovi zadanih podataka definirati trigonometrijski polinom  $T_n$ .

**Zadatak 2.18** Pokažite da je dijagonalno dominantna matrica regularna.

**Zadatak 2.19** Na osnovi podataka

$$a) \begin{array}{c|cccc} x_i & 0 & 2 & 4 & 5 \\ y_i & 10 & 15 & 12 & 14 \end{array}, \quad \xi = 1 \quad b) \begin{array}{c|cccc} x_i & 1 & 3 & 5 & 7 \\ y_i & 8 & 12 & 10 & 14 \end{array}, \quad \xi = 4.5$$

odredite prirodni kubični interpolacijski spline i izračunajte  $C(\xi)$ .

Rješenje:

$$\begin{aligned} a) \quad C_1(x) &= 10 + \frac{43}{11}x - \frac{31}{88}x^3, \quad x \in [0, 2] \\ C_2(x) &= \frac{12}{11} + \frac{190}{11}x - \frac{147}{22}x^2 + \frac{67}{88}x^3, \quad x \in [2, 4] \\ C_3(x) &= \frac{1124}{11} - \frac{644}{11}x + \frac{135}{11}x^2 - \frac{9}{11}x^3, \quad x \in [4, 5], \\ C(1) &= 13.5568 \\ b) \quad C_1(x) &= 5.25 + 2.25x + 0.75x^2 - 0.25x^3, \quad x \in [1, 3] \\ C_2(x) &= -15 + 22.5x - 6x^2 + 0.5x^3, \quad x \in [3, 5] \\ C_3(x) &= 78.75 - 33.75x + 5.25x^2 - 0.25x^3, \quad x \in [5, 7] \\ C(4.5) &= 10.3125 \end{aligned}$$

**Zadatak 2.20** Na osnovi podataka

$$\begin{array}{c|ccccc} x_i & 0 & 1 & 2 & 3 & 4 & 5 \\ y_i & 1 & 2 & 1.5 & 2 & 2.5 & 2 \end{array}, \quad \xi = 2.5,$$

odredite prirodni kubični interpolacijski spline i izračunajte  $C(\xi)$ .

Uputa: Sustav (2.34) u ovom slučaju glasi:

$$\begin{bmatrix} 4 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 1 & 4 & 1 \\ 0 & 0 & 1 & 4 \end{bmatrix} \mathbf{s} = \begin{bmatrix} -9 \\ 6 \\ 0 \\ -6 \end{bmatrix},$$

*i ima rješenje*  $\mathbf{s} = (0, -2.8134, 2.25359, -0.200957, -1.44976, 0)^T$ .

Rješenje:

$$C_1(x) = 1 + 1.4689x - 0.4689x^3, \quad x \in [0, 1]$$

$$C_2(x) = -0.313397 + 5.409091x - 3.94019x^2 + 0.844498x^3, \quad x \in [1, 2]$$

$$C_3(x) = 9.71531 - 9.63397x + 3.58134x^2 - 0.409091x^3, \quad x \in [2, 3]$$

$$C_4(x) = 4.28947 - 4.20813x + 1.77273x^2 - 0.208134x^3, \quad x \in [3, 4]$$

$$C_5(x) = -24.4952 + 17.3804x - 3.6244x^2 + 0.241627x^3, \quad x \in [4, 5],$$

$$C(2.5) = 2.28439$$

**Zadatak 2.21** *Problem iz Primjera 2.3, str. 20 riješite primjenom kubičnog interpolacijskog spline-a. Koristite program u Dodatku.*

Rješenje:  $C(500) = 10.54086$ ,  $C(750) = 17.2716$ ,  $C(1000) = 24.32858$ .

**Zadatak 2.22** *Program u Dodatku za izračunavanje kubičnog interpolacijskog spline-a dopunite tako da računa prvu, drugu i treću derivaciju dobivenog interpolacijskog spline-a u točki  $\xi \in [x_0, x_n]$ .*

### 3. Rješavanje sustava linearnih jednačbi

Promatramo sustav linearnih jednačbi

$$\mathbf{Ax} = \mathbf{b}, \quad (3.1)$$

gdje je  $\mathbf{A} \in \mathbb{R}^{m \times n}$  matrica,  $\mathbf{x} \in \mathbb{R}^m$  vektor nepoznanica, a  $\mathbf{b} \in \mathbb{R}^n$  vektor slobodnih koeficijenata. Osim klasičnih metoda za rješavanje sustava (3.1), kao što je Gaussova metoda, postoje metode koje se zasnivaju na različitim dekompozicijama matrice  $\mathbf{A}$ , kao što je Cholesky dekompozicija, QR dekompozicija, dekompozicija na singularne vrijednosti, itd. Posebnu klasu metoda predstavljaju tzv. iterativne metode, kao što je Gauss-Seidelova ili Jacobijeva metoda. Također, od posebnog su praktičnog značaja tzv. veliki rijetki sustavi linearnih jednačbi za koje postoje specijalne metode za rješavanje.

**Primjer 3.1** *Za početak pogledajmo jednostavni sustav jednačbi*

$$\begin{aligned} 2x_1 + 400x_2 &= 200 \\ x_1 + x_2 &= 1 \end{aligned}$$

*čije je rješenje*

$$x_1 = 0.5025125628 \dots, \quad x_2 = 0.4974874372 \dots$$

Riješit ćemo ovaj sustav na računalu s 2-znamenkastom floating-point aritmetikom. Množeći prvu jednačbu s  $(-0.5)$  i dodavajući drugoj, dobivamo

$$\begin{aligned} 2x_1 + 400x_2 &= 200 \\ -200x_2 &= -99 \end{aligned}$$

Iz druge jednačbe dobivamo  $x_2^* = 0.50$ , a nakon toga iz prve,  $x_1^* = (200 - 400 \times 0.50)/2 = 0$ .

Ako bismo račun proveli u 3-znamenkastoj floating-point aritmetici, dobili bismo  $x_2^{**} = 0.497$ ,  $x_1^{**} = 0.600$ .

Iz ovog jednostavnog primjera vidi se da je važno znati procijeniti veličinu pogreške aproksimacije rješenja.

Ovisno o veličini i strukturi matrice  $\mathbf{A}$ , prilikom rješavanja sustava (3.1) u praktičnim primjenama mogu nastupiti različiti numerički problemi. Pokušat ćemo proanalizirati razloge ovim pojavama i uputiti na mogućnost dobivanja što korektnijeg rješenja u nekim za praksu važnim slučajevima.

### 3.1 Norma vektora i matrice

**Definicija 3.1** Neka je  $X$  realni vektorski prostor. Funkciju  $\|\cdot\| : X \rightarrow [0, \infty)$ , koja svakom vektoru  $\mathbf{x} \in X$  pridružuje nenegativan realan broj (kojeg ćemo označiti s  $\|\mathbf{x}\|$ ) zovemo **norma** ako vrijedi

- (i)  $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$  (pozitivna definitnost),
- (ii)  $\|\lambda \mathbf{x}\| = |\lambda| \|\mathbf{x}\|$  za svaki  $\lambda \in \mathbb{R}$  i za svaki  $\mathbf{x} \in X$ ,
- (iii)  $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$  za svaki  $\mathbf{x}, \mathbf{y} \in X$  (nejednakost trokuta).

Vrijednost funkcije  $\|\cdot\|$  na nekom vektoru  $\mathbf{x} \in X$  zvat ćemo norma vektora  $\mathbf{x}$ . U ovom poglavlju razmatrat ćemo vektorski prostor  $\mathbb{R}^n$ , čiji su elementi vektori stupci. Zato obično pišemo  $\mathbf{x} = (x_1, \dots, x_n)^T$ .

Najčešće korištene vektorske norme su

$$\begin{aligned} \|\mathbf{x}\|_1 &= \sum_{i=1}^n |x_i|, & (L_1 \text{ norma}) \\ \|\mathbf{x}\|_2 &= \sqrt{\mathbf{x}^T \mathbf{x}} = \left( \sum_{i=1}^n x_i^2 \right)^{1/2}, & (L_2 \text{ ili Euklidova norma}) \\ \|\mathbf{x}\|_\infty &= \max_{i=1, \dots, n} |x_i|, & (L_\infty \text{ norma ili Čebiševljeva norma}) \end{aligned}$$

**Zadatak 3.1** Pokažite da spomenute norme imaju sva svojstva navedena u Definiciji 3.1.

Može se pokazati da su sve tri navedene norme međusobno ekvivalentne (vidi primjerice MARDEŠIĆ (1988)), tj. za svaki par  $i, j \in \{1, 2, \infty\}$  postoje konstante  $\alpha, \beta > 0$  takve da vrijedi

$$\alpha \|\mathbf{x}\|_i \leq \|\mathbf{x}\|_j \leq \beta \|\mathbf{x}\|_i \quad \text{za svaki } \mathbf{x} \in \mathbb{R}^n.$$

**Zadatak 3.2** Jedinična kuglina ljuska (sfera) sa središtem u  $\mathbf{O} \in \mathbb{R}^n$  definira se s

$$\partial K_n = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = 1\}.$$

Nacrtajte jedinične kugline ljuske za  $n = 2$  i norme  $\|\cdot\|_1$ ,  $\|\cdot\|_2$ ,  $\|\cdot\|_\infty$ .

**Primjedba 3.1** Odgovarajuće norme u vektorskom prostoru  $C_{[a,b]}$ ,  $[a, b] \subset \mathbb{R}$  svih neprekidnih funkcija na intervalu  $[a, b]$  bile bi

$$\|f\|_1 = \int_a^b |f(t)| dt, \quad \|f\|_2 = \left( \int_a^b f^2(t) dt \right)^{1/2}, \quad \|f\|_\infty = \max_{t \in [a,b]} |f(t)|$$

**Definicija 3.2** Neka je  $\mathbb{R}^{n \times n}$  skup svih realnih matrica tipa  $(n \times n)$ . Funkciju  $\|\cdot\| : \mathbb{R}^{n \times n} \rightarrow [0, \infty)$ , koja svakoj matrici  $\mathbf{A} \in \mathbb{R}^{n \times n}$  pridružuje nenegativan realan broj (kojeg ćemo označiti s  $\|\mathbf{A}\|$ ) zovemo **norma matrice  $\mathbf{A}$**  ako vrijedi

- (i)  $\|\mathbf{A}\| = 0 \Leftrightarrow \mathbf{A} = \mathbf{0}$ , za svaki  $\mathbf{A} \in \mathbb{R}^{n \times n}$
- (ii)  $\|\lambda \mathbf{A}\| = |\lambda| \|\mathbf{A}\|$  za svaki  $\lambda \in \mathbb{R}$  i  $\mathbf{A} \in \mathbb{R}^{n \times n}$ ,
- (iii)  $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$  za svaki  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$
- (iv)  $\|\mathbf{A} \cdot \mathbf{B}\| \leq \|\mathbf{A}\| \cdot \|\mathbf{B}\|$  za sve  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$

Ako smo na  $\mathbb{R}^n$  definirali vektorsku normu  $\|\cdot\|$ , onda njoj možemo pridružiti matričnu normu na sljedeći način

$$\|\mathbf{A}\| = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|}, \quad (3.2)$$

koju ćemo zvati **inducirana matrična norma**. Primijetite da je u (3.2) norma matrica  $\mathbf{A}$  definirana kao najveći mogući omjer norme vektora  $\mathbf{A}\mathbf{x}$  i norme vektora  $\mathbf{x}$ ,  $\mathbf{x} \neq \mathbf{0}$ . Iz (3.2) slijedi  $\|\mathbf{A}\| \geq \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|}$ , odnosno

$$\|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\| \|\mathbf{x}\|. \quad (3.3)$$

Matrične norme inducirane ranije spomenutim vektorskim normama su

$$\begin{aligned} \|\mathbf{A}\|_1 &= \max_{j=1, \dots, n} \|\mathbf{a}_j\|_1, \quad \text{gdje je } \mathbf{a}_j \text{ } j\text{-ti stupac matrice } \mathbf{A} \\ \|\mathbf{A}\|_2 &= \sigma_1(\mathbf{A}) - \text{najveća singularna vrijednost matrice } \mathbf{A} \quad (\text{vidi t.3.9, str. 60}) \\ \|\mathbf{A}\|_\infty &= \max_{i=1, \dots, n} \|\mathbf{a}_i^T\|_1, \quad \text{gdje je } \mathbf{a}_i \text{ } i\text{-ti redak matrice } \mathbf{A}. \end{aligned}$$

Ako za neku matričnu i za neku vektorsku normu vrijedi (3.3), kažemo da su one **kompatibilne** (ili konzistentne).

Osim navedenih, važna matrična norma je tzv. **Frobeniusova norma** matrice  $\mathbf{A} \in \mathbb{R}^{m \times n}$

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

Frobeniusova norma nije inducirana matrična norma, ali je kompatibilna s Euklidovom vektorskom normom (vidi primjerice [19], tj. vrijedi

$$\|\mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{A}\|_F \|\mathbf{x}\|_2.$$

Ako je  $\tilde{\mathbf{x}}$  aproksimacija vektora  $\mathbf{x}$ , a  $\tilde{\mathbf{A}}$  aproksimacija matrice  $\mathbf{A}$ , onda analogno kao što smo to uradili u t.1.2, str. 4, možemo definirati

	Vektor	Matrica
Pogreška aproksimacije	$\Delta \mathbf{x} = \tilde{\mathbf{x}} - \mathbf{x}$	$\Delta \mathbf{A} = \tilde{\mathbf{A}} - \mathbf{A}$
Apsolutna pogreška aproksimacije	$\ \Delta \mathbf{x}\  = \ \tilde{\mathbf{x}} - \mathbf{x}\ $	$\ \Delta \mathbf{A}\  = \ \tilde{\mathbf{A}} - \mathbf{A}\ $
Relativna pogreška aproksimacije	$\frac{\ \tilde{\mathbf{x}} - \mathbf{x}\ }{\ \mathbf{x}\ }$	$\frac{\ \tilde{\mathbf{A}} - \mathbf{A}\ }{\ \mathbf{A}\ }$

### 3.2 Uvjetovanost sustava linearnih jednadžbi

Promatramo sustav jednadžbi  $\mathbf{Ax} = \mathbf{b}$  uz pretpostavku da je  $\mathbf{A}$  kvadratna regularna matrica. Tada sustav ima jedinstveno rješenje  $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ . Možemo postaviti ovakvo pitanje:

- (i) *Koliko će se promijeniti rješenje  $\mathbf{x}$  sustava, ako se promijeni vektor slobodnih koeficijenata  $\mathbf{b}$ , a matrica  $\mathbf{A}$  ostane nepromijenjena?*

Neka je  $\mathbf{x}$  rješenje sustava  $\mathbf{Ax} = \mathbf{b}$ , a  $\tilde{\mathbf{x}} = \mathbf{x} + \Delta \mathbf{x}$  rješenje “perturbiranog sustava”  $\mathbf{A}\tilde{\mathbf{x}} = \mathbf{b} + \Delta \mathbf{b}$ , tj.

$$\mathbf{A}(\mathbf{x} + \Delta \mathbf{x}) = \mathbf{b} + \Delta \mathbf{b}. \quad (3.4)$$

Iz (3.1) i (3.4) dobivamo

$$\Delta \mathbf{x} = \mathbf{A}^{-1}\Delta \mathbf{b}, \quad (3.5)$$

Korištenjem kompatibilne matrične i vektorske norme, iz (3.1) i (3.5) dobivamo

$$\|\mathbf{b}\| \leq \|\mathbf{A}\|\|\mathbf{x}\|, \quad (3.6)$$

$$\|\Delta \mathbf{x}\| \leq \|\mathbf{A}^{-1}\|\|\Delta \mathbf{b}\|. \quad (3.7)$$

Dijeljenjem (3.7) s  $\|\mathbf{x}\|$  i korištenjem (3.6) dobivamo

$$\frac{\|\tilde{\mathbf{x}} - \mathbf{x}\|}{\|\mathbf{x}\|} \leq \|\mathbf{A}^{-1}\| \cdot \|\mathbf{A}\| \frac{\|\Delta \mathbf{b}\|}{\|\mathbf{b}\|}, \quad (3.8)$$

čime smo ocijenili relativnu pogrešku vektora  $\tilde{\mathbf{x}}$  pomoću relativne pogreške vektora  $\mathbf{b}$ .

Vrijedi također (vidi primjerice GOLUB (1996)):

- (ii) *Ako se matrica sustava  $\mathbf{A}$  promijeni za  $\Delta \mathbf{A}$ , a vektor  $\mathbf{b}$  ostane nepromijenjen, onda relativnu pogrešku aproksimacije rješenja  $\tilde{\mathbf{x}}$  možemo ocijeniti s*

$$\frac{\|\tilde{\mathbf{x}} - \mathbf{x}\|}{\|\tilde{\mathbf{x}}\|} \leq \|\mathbf{A}^{-1}\| \cdot \|\mathbf{A}\| \frac{\|\Delta \mathbf{A}\|}{\|\mathbf{A}\|}. \quad (3.9)$$

- (iii) *Ako se matrica sustava  $\mathbf{A}$  promijeni za  $\Delta \mathbf{A}$ , a vektor  $\mathbf{b}$  za  $\Delta \mathbf{b}$ , onda relativnu pogrešku aproksimacije rješenja  $\tilde{\mathbf{x}}$  možemo ocijeniti s*

$$\frac{\|\tilde{\mathbf{x}} - \mathbf{x}\|}{\|\mathbf{x}\|} \leq \|(\mathbf{A} + \Delta \mathbf{A})^{-1}\| \|\mathbf{A}\| \left( \frac{\|\Delta \mathbf{b}\|}{\|\mathbf{b}\|} + \frac{\|\Delta \mathbf{A}\|}{\|\mathbf{A}\|} \right). \quad (3.10)$$



U sva tri slučaja u ocjeni relativne pogreške aproksimacije rješenja pojavio se broj  $\|\mathbf{A}^{-1}\|\|\mathbf{A}\|$ . Ovaj broj zvat ćemo broj uvjetovanosti regularne matrice  $\mathbf{A}$  (*condition number*) i označavati s

$$\text{cond}(\mathbf{A}) = \|\mathbf{A}^{-1}\|\|\mathbf{A}\|.$$

Broj uvjetovanosti  $\text{cond}(\mathbf{A})$  uvijek je veći ili jednak 1. Naime,

$$1 \leq \|\mathbf{I}\| = \|\mathbf{A}\mathbf{A}^{-1}\| \leq \|\mathbf{A}\|\|\mathbf{A}^{-1}\| = \text{cond}(\mathbf{A}).$$

Ako je  $\text{cond}(\mathbf{A})$  blizu 1, kažemo da je matrica **dobro uvjetovana**, a ako je  $\text{cond}(\mathbf{A})$  mnogo veći od 1, kažemo da je matrica **loše uvjetovana** (*ill-conditioned*).

Iz (3.8), (3.9) i (3.10) vidi se da će i male pogreške u matrici ili vektoru slobodnih koeficijenata sustava s loše uvjetovanom matricom sustava rezultirati značajnom pogreškom u rješenju sustava.

**Primjer 3.2** Rješenje sustava linearnih jednadžbi, gdje su matrica sustava  $\mathbf{A}$  i vektor  $\mathbf{b}$  zadani s

$$\mathbf{A} = \begin{bmatrix} 0.234 & 0.458 \\ 0.383 & 0.750 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0.224 \\ 0.367 \end{bmatrix},$$

je  $\mathbf{x} = (-1, 1)^T$ . Promijenimo vektor  $\mathbf{b}$  za  $\Delta\mathbf{b} = (0.00009, 0.000005)^T$ .

Egzaktno rješenje sustava  $\mathbf{A}\mathbf{x} = \mathbf{b} + \Delta\mathbf{b}$  je  $\tilde{\mathbf{x}} = (-0.241774, 0.612791)^T$ , pri čemu je  $\Delta\mathbf{x} = \tilde{\mathbf{x}} - \mathbf{x} = (0.758256, -0.387209)^T$ . Izračunajmo relativne pogreške vektora  $\mathbf{b}$  i  $\mathbf{x}$  u normi  $\|\cdot\|_\infty$ .

$$\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} = \frac{0.00009}{0.367} \approx 2.45 \times 10^{-4}, \quad \frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} = \frac{0.758256}{1} = 0.758256.$$

Dakle, relativna pogreška rješenja je više od 3 000 puta veća od relativne pogreške vektora  $\mathbf{b}$ .

Načinimo sada malu promjenu u elementu  $a_{11}$  matrice  $\mathbf{A}$ , tako da je

$$\mathbf{A} + \Delta\mathbf{A} = \begin{bmatrix} 0.233 & 0.458 \\ 0.383 & 0.750 \end{bmatrix}.$$

Egzaktno rješenje sustava  $(\mathbf{A} + \Delta\mathbf{A})\tilde{\mathbf{x}} = \mathbf{b}$  je  $\tilde{\mathbf{x}} = (0.129518, 0.423193)^T$ . Sada je

$$\frac{\|\tilde{\mathbf{x}} - \mathbf{x}\|}{\|\tilde{\mathbf{x}}\|} = \frac{1.129518}{0.423193} \approx 2.67, \quad \frac{\|\Delta\mathbf{A}\|}{\|\mathbf{A}\|} = \frac{0.001}{1.133} \approx 0.00075,$$

pa je relativna pogreška aproksimacije rješenja opet više od 3 000 puta veća od relativne pogreške u matrici  $\mathbf{A}$ .

Razloge za ove pojave treba tražiti u veličini broja uvjetovanosti matrice  $\mathbf{A}$ . Kako je

$$\mathbf{A} = \begin{bmatrix} 0.234 & 0.458 \\ 0.383 & 0.750 \end{bmatrix}, \quad \mathbf{A}^{-1} = \begin{bmatrix} 8720.93 & -5325.58 \\ -4453.49 & 2720.93 \end{bmatrix},$$

dobivamo  $\|\mathbf{A}\|_\infty = 1.133$  i  $\|\mathbf{A}^{-1}\|_\infty = 14\,046.51$ , pa je broj uvjetovanosti matrice  $\mathbf{A}$   $\text{cond}(\mathbf{A}) = 15\,914.70$ .

**Zadatak 3.3** *Provedite čitav račun iz Primjera 3.2 za norme  $\|\cdot\|_1$  i  $\|\cdot\|_2$ .*

Uputa: Broj uvjetovanosti  $\text{cond}(\mathbf{A})$  u  $\|\cdot\|_2$  može se izračunati (vidi primjerice GOLUB (1996)) po formuli

$$\text{cond}(\mathbf{A}) = \frac{\sigma_1}{\sigma_n},$$

gdje je  $\sigma_1$  najveća, a  $\sigma_n$  najmanja singularna vrijednost matrice  $\mathbf{A}$  (vidi također t. 3.9, str. 60). Singularne vrijednosti kvadratne matrice  $\mathbf{A}$  mogu se lako izračunati programom *Mathematica* naredbom: `SingularValues[A][[2]]`.

**Primjedba 3.2** *Moglo bi se pomisliti da je veličina determinante matrice u direktnoj vezi s uvjetovanosti matrice. Da te dvije veličine nisu u direktnoj vezi vidi se već iz sljedećeg primjera. Tako je primjerice dijagonalna matrica  $\mathbf{D} = \text{diag}(10^{-10}, 10^{-10})$  perfektno uvjetovana kao jedinična matrica ( $\text{cond}(\mathbf{D}) = 1$ ), a njezina determinanta je skoro nula ( $\det \mathbf{D} = 10^{-20}$ ). S druge strane, dijagonalna matrica  $\mathbf{S} = \text{diag}(10^5, 10^{-5})$  je loše uvjetovana ( $\text{cond}(\mathbf{S}) = 10^{10}$ ), iako je  $\det \mathbf{S} = 1$ .*

### 3.3 Rješavanje trokutastih sustava

Sustav linearnih jednadžbi  $\mathbf{U}\mathbf{x} = \mathbf{b}$ , gdje je  $\mathbf{U} \in \mathbb{R}^{n \times n}$  gornja<sup>1</sup> trokutasta matrica

$$\mathbf{U} = \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ 0 & u_{22} & \dots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & u_{nn} \end{bmatrix}, \quad u_{ii} \neq 0, i = 1, \dots, n$$

lako se rješava rješavanjem unazad (engl: *Back Substitution* (BS)):

$$x_i = \frac{1}{u_{ii}} \left( b_i - \sum_{k=i+1}^n u_{ik} x_k \right), \quad i = n, n-1, \dots, 1. \quad (3.11)$$

Slično, sustav linearnih jednadžbi  $\mathbf{L}\mathbf{x} = \mathbf{b}$ , gdje je  $\mathbf{L} \in \mathbb{R}^{n \times n}$  donja<sup>2</sup> trokutasta matrica

$$\mathbf{L} = \begin{bmatrix} l_{11} & 0 & \dots & 0 \\ l_{21} & l_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \dots & l_{nn} \end{bmatrix}, \quad l_{ii} \neq 0, i = 1, \dots, n$$

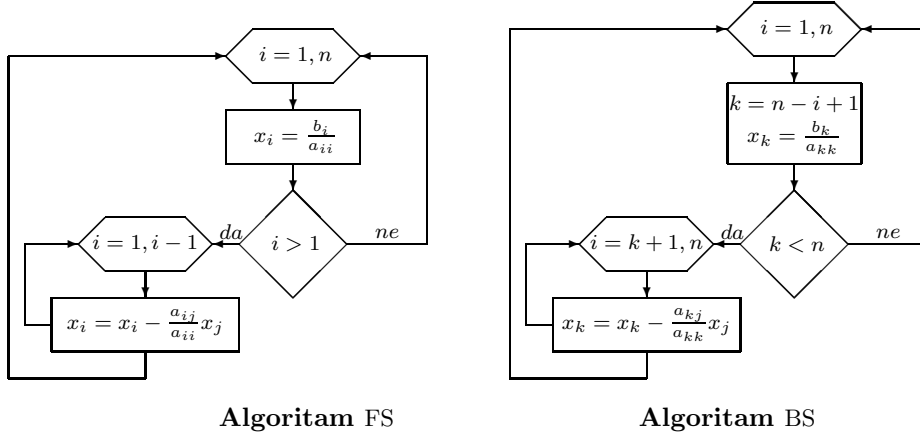
rješava se rješavanjem unaprijed (engl: *Forward Substitution* (FS)):

$$x_i = \frac{1}{l_{ii}} \left( b_i - \sum_{k=1}^{i-1} l_{ik} x_k \right), \quad i = 1, \dots, n. \quad (3.12)$$

<sup>1</sup>engl: upper = gornji

<sup>2</sup>engl: lower = donji

Niže su navedeni odgovarajući dijagrami toka, *Mathematica* – programi za rješavanje donjeg trokutastog (Algoritam FS) i gornjeg trokutastog (Algoritam BS) sustava.



```

FS[n_, a_, b_] := Module[{i, j},
  x[[1]] = b[[1]]/a[[1,1]];
  Do[
    x[[i]] = (b[[i]] - Sum[a[[i,j]] x[[j]], {j,i-1}))/a[[i,i]]
    ,{i,2,n}]
]

BS[n_, a_, b_] := Module[{i, j, k},
  x[[n]] = b[[n]]/a[[n,n]];
  Do[k = n-i;
    x[[k]] = (b[[k]] - Sum[a[[k,j]] x[[j]], {j,k+1,n}))/a[[k,k]]
    ,{i,n-1}]
]

```

**Zadatak 3.4** Specijalizirajte algoritme FS i BS za slučaj trodijagonalne matrice<sup>3</sup> (kakva se primjerice pojavljuje kod prirodnog kubičnog spline-a, str. 28).

Uputa: Od cijele matrice treba “pamtiti” samo tri vektora: glavnu dijagonalu, te donju i gornju sporednu dijagonalu matrice **A**.

### 3.4 Gaussova metoda eliminacije

Zadana je kvadratna regularna matrica **A** i vektor slobodnih koeficijenata **b**

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}, \quad (3.13)$$

<sup>3</sup>Trodijagonalna matrica je specijalni slučaj tzv. vrpčastih matrica, koje spadaju u tzv. rijetko popunjene matrice (sparse matrices).

Prema *Kronecker-Capellijevom* teoremu sustav  $\mathbf{Ax} = \mathbf{b}$  je rješiv i ima jedinstveno rješenje<sup>4</sup>.

Pretpostavimo da je tzv. pivot-element  $a_{11} \neq 0$ . Najprije ćemo primjenom Gaussovih transformacija matricu  $\mathbf{A}$  svesti na oblik

$$\mathbf{A}^{(2)} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2}^{(2)} & \cdots & a_{nn}^{(2)} \end{bmatrix}, \quad \mathbf{b}^{(2)} = \begin{bmatrix} b_1 \\ b_2^{(2)} \\ \vdots \\ b_n^{(2)} \end{bmatrix},$$

gdje je, uz oznaku  $m_{i1} = \frac{a_{i1}}{a_{11}}$ ,

$$a_{ik}^{(2)} = a_{ik} - m_{i1}a_{1k}, \quad b_i^{(2)} = b_i - m_{i1}b_1, \quad (i, k = 2, \dots, n) \quad (3.14)$$

Uz pretpostavku da je sljedeći pivot-element  $a_{22}^{(2)} \neq 0$ , na sličan način dobivamo

$$\mathbf{A}^{(3)} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2n}^{(2)} \\ 0 & 0 & a_{33}^{(3)} & \cdots & a_{3n}^{(3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & a_{n3}^{(3)} & \cdots & a_{nn}^{(3)} \end{bmatrix}, \quad \mathbf{b}^{(3)} = \begin{bmatrix} b_1 \\ b_2^{(2)} \\ b_3^{(3)} \\ \vdots \\ b_n^{(3)} \end{bmatrix},$$

gdje je, uz oznaku  $m_{i2} = \frac{a_{i2}^{(2)}}{a_{22}^{(2)}}$ ,

$$a_{ik}^{(3)} = a_{ik}^{(2)} - m_{i2}a_{2k}^{(2)}, \quad b_i^{(3)} = b_i^{(2)} - m_{i2}b_2^{(2)}, \quad (i, k = 3, \dots, n) \quad (3.15)$$

Uz pretpostavku da su svi pivot-elementi bili različiti od nule, nakon  $(n-1)$  ovakvih koraka (*iteracija*) matrica  $\mathbf{A}$  prijeći će u oblik gornje trokutaste matrice, gdje su svi dijagonalni elementi različiti od nule. Takav sustav jednadžbi lako se rješava (*Algoritam BS*, str. 44).

**Primjedba 3.3** *Primjenom Gaussovih transformacija matricu sustava mogli bismo svesti i na dijagonalnu matricu. U tom slučaju govorimo o Gauss-Jordanovoj metodi, koja igra važnu ulogu kod simplex-metode za rješavanje problema linearnog programiranja.*

Vidjeli smo da je prilikom provođenja Gaussovog postupka važno da su svi pivot-elementi:  $a_{11}, a_{22}^{(2)}, \dots$  različiti od nule. To neće biti uvijek osigurano, iako je matrica sustava regularna. Ipak, postoje dva važna slučaja, gdje će spomenuti uvjet biti ispunjen. To su slučajevi kada je matrica sustava

- dijagonalno dominantna, tj  $|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|$

<sup>4</sup>Iako se Gaussovom metodom može rješavati opći sustav linearnih jednadžbi (gdje broj jednadžbi nije jednak broju nepoznanica ili gdje matrica sustava nije regularna), mi ćemo se zadržati na ovom specijalnom slučaju, koji je i najčešći u primjenama.

- simetrična i pozitivno definitna, tj  $\mathbf{A}^T = \mathbf{A}$  &  $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$  za sve  $\mathbf{x} \neq \mathbf{0}$ .  
Prema poznatom Sylvesterovom kriteriju, simetrična matrica  $\mathbf{A}$  je pozitivno definitna onda i samo onda ako su svi njezini glavni minori<sup>5</sup> pozitivni.

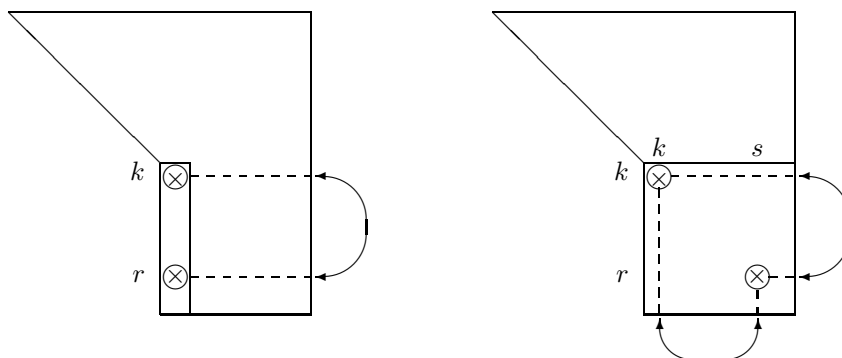
Dijagonalna dominantnost, odnosno pozitivna definitnost vrlo su jaki uvjeti na matricu  $\mathbf{A}$ . Da bi se ipak proširila klasa matrica za koje bi Gaussov postupak bio provediv, ranije opisanu Gaussovu metodu eliminacije malo ćemo modificirati.

Jedan pokušaj u tom smislu je tzv. strategija parcijalnog pivotiranja:

*u  $k$ -toj iteraciji treba zamijeniti  $k$ -ti s  $r$ -tim retkom, gdje je  $r$  najmanji indeks retka takav da je*

$$|a_{rk}^{(k)}| = \max_{k \leq i \leq n} |a_{ik}^{(k)}|$$

(vidi shemu na Slici 3.1)



Slika 3.1 Strategija parcijalnog (lijevo) i strategija potpunog (desno) pivotiranja

Drugi pokušaj je tzv. strategija potpunog pivotiranja:

*u  $k$ -toj iteraciji najprije pronađemo cijele brojeve  $r$  i  $s$ , za koje je*

$$|a_{rs}^{(k)}| = \max_{k \leq i, j \leq n} |a_{ij}^{(k)}|,$$

*a onda sukcesivno napravimo zamjenu  $k$ -tog i  $s$ -tog stupaca, te  $k$ -tog i  $r$ -tog retka (vidi shemu na Slici 3.1)*

**Zadatak 3.5** Pokažite da je Gaussov postupak bez pivotiranja provediv onda i samo onda ako su svi glavni minori matrica sustava različiti od nule.

**Primjer 3.3** Ako bismo u Primjeru 3.1, str. 38 za slučaj 2-znamenkaste floating-point aritmetike primijenili strategiju parcijalnog pivotiranja, dobili bismo isti (loši) rezultat. Primjenom strategije potpunog pivotiranja dobili bismo “točan” rezultat:  $x_1^* = x_2^* = 0.50$ .

<sup>5</sup>Glavni minori matrice  $\mathbf{A}$  su determinante  $\det \mathbf{A}_k$ ,  $k = 1, \dots, n$ , gdje je  $\mathbf{A}_k$  matrica dobivena od elemenata matrice  $\mathbf{A}$  na presjeku prvih  $k$  stupaca i prvih  $k$  redaka

U 3-znamenkastoj floating-point aritmetici već uz primjenu strategije parcijalnog pivotiranja dobili bismo “točan” rezultat  $x_1^{**} = 0.503$ ,  $x_2^{**} = 0.498$ .

Ovaj primjer pokazuje da se strategijom parcijalnog, a naročito strategijom potpunog pivotiranja također može postići i numerička stabilnost Gaussovog postupka.

**Zadatak 3.6** Izradite program koji će u  $k$ -znamenkastoj floating-point aritmetici ( $k = 2, 3, \dots, 8$ ) rješavati sustav linearnih jednadžbi  $\mathbf{Ax} = \mathbf{b}$ :

- Gaussovom metodom bez pivotiranja;
- Gaussovom metodom uz strategiju parcijalnog pivotiranja;
- Gaussovom metodom uz strategiju potpunog pivotiranja;

Program ispitajte na sljedećim primjerima:

(a)

$$\begin{bmatrix} 0.0001 & 0.5 \\ 0.4 & -0.4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.1 \end{bmatrix},$$

Egzaktno rješenje zaokruženo na 4 znamenke je  $\mathbf{x} = (0.9999, 0.9998)^T$ .

(b)

$$\begin{array}{rcl} \epsilon x_1 & & = b \\ & \epsilon x_2 & = c, \\ x_1 + x_2 + x_3 & = & a \end{array}$$

gdje je  $\epsilon, a, b, c \in \mathbb{R}$ . Egzaktno rješenje je  $x_1 = \frac{b}{\epsilon}$ ,  $x_2 = \frac{c}{\epsilon}$ ,  $x_3 = a - \frac{b}{\epsilon} - \frac{c}{\epsilon}$ .

(c) Sustav kome je matrica sustava tzv. Hilbertova matrica  $H_{ij} = \frac{1}{i+j-1}$ ,  $i = 1, \dots, n$ , s proizvoljnim vektorom slobodnih koeficijenata.

## 3.5 LU-dekompozicija

Pretpostavimo da je  $\mathbf{A} \in \mathbb{R}^{n \times n}$  regularna kvadratna matrica, kojoj su svi glavni minori različiti od nule. Tada je na jedinstven način moguće načiniti rastav (BJÖRCK (1974))

$$\mathbf{A} = \mathbf{LU}$$

gdje je  $\mathbf{L}$  donja trokutasta matrica, kojoj su na glavnoj dijagonali jedinice, a  $\mathbf{U}$  gornja trokutasta matrica, čiji dijagonalni elementi nisu nule

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ l_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & 1 \end{bmatrix} \cdot \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{bmatrix} \quad (3.16)$$

**Zadatak 3.7** Matrica  $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$  je regularna kvadratna matrica. Pokažite da za ovu matricu ipak nije moguće načiniti LU-dekompoziciju.

Najprije ćemo pokazati da Gaussova metoda eliminacije opisana u prethodnoj točki primijenjena na kvadratnu matricu  $\mathbf{A}$ , kojoj su svi glavni minori različiti od nule generira LU-dekompoziciju te matrice.

Pretpostavimo da je  $\mathbf{A}$  kvadratna regularna matrica kojoj su svi glavni minori različiti od nule. Ako uvedemo vektore

$$\mathbf{u}_1 = (0, m_{21}, \dots, m_{n1})^T, \quad \mathbf{v}_1 = (1, 0, \dots, 0)^T,$$

i elementarnu matricu  $\mathbf{M}_1$ :

$$\mathbf{M}_1 = \mathbf{I} - \mathbf{u}_1 \mathbf{v}_1^T = \begin{bmatrix} 1 & & & \\ -m_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ -m_{n1} & 0 & \dots & 1 \end{bmatrix},$$

gdje su elementi  $m_{i,j}$  zadani kao u t.3.4, str.44, onda matricu  $\mathbf{A}^{(2)}$  dobivenu u prvom koraku Gaussovog postupka eliminacije možemo pisati kao

$$\mathbf{A}^{(2)} = \mathbf{M}_1 \mathbf{A},$$

Matricu  $\mathbf{A}^{(3)}$  možemo pisati kao

$$\mathbf{A}^{(3)} = \mathbf{M}_2 \mathbf{M}_1 \mathbf{A},$$

gdje je  $\mathbf{M}_2 = \mathbf{I} - \mathbf{u}_2 \mathbf{v}_2^T$  elementarna matrica s vektorima

$$\mathbf{u}_2 = (0, 0, m_{32}, \dots, m_{n2})^T, \quad \mathbf{v}_2 = (0, 1, 0, \dots, 0)^T$$

Na kraju Gaussovog postupka matrica  $\mathbf{A}$  primit će oblik gornje trokutaste matrice, koju ćemo označiti s  $\mathbf{U}$  i zapisati kao

$$\mathbf{U} := \mathbf{M}_{n-1} \cdots \mathbf{M}_2 \mathbf{M}_1 \mathbf{A}$$

Sve matrice  $\mathbf{M}_1, \dots, \mathbf{M}_{n-1}$  su regularne donje trokutaste matrice s jedinicama na glavnoj dijagonali. Zato su takve i njihove inverzne matrice. Možemo pisati

$$\mathbf{A} = \mathbf{L} \mathbf{U} \quad \text{gdje je} \quad \mathbf{L} = \mathbf{M}_1^{-1} \mathbf{M}_2^{-1} \cdots \mathbf{M}_{n-1}^{-1}.$$

Matrica  $\mathbf{L}$  je donja trokutasta matrica s jedinicama na glavnoj dijagonali jer su sve matrice  $\mathbf{M}_k$  takvog oblika.

**Zadatak 3.8** Pokažite da je inverzna matrica matrice  $\mathbf{M}_k = \mathbf{I} - \mathbf{u}_k \mathbf{v}_k^T$  iz gornjeg izvoda zadana s

$$\mathbf{M}_k^{-1} = \mathbf{I} + \mathbf{u}_k \mathbf{v}_k^T.$$

**Primjedba 3.4** Općenitije može se pokazati da vrijedi tzv. SHERMAN-MORRISONOVA<sup>6</sup> lema:

<sup>6</sup>J. SHERMAN, W. J. MORRISON, *Adjustments of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix*, Ann. Math. Stat. **20**(1949), 621.

Neka je  $\mathbf{A} \in \mathbb{R}^{n \times n}$  regularna matrica i  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ . Tada je matrica  $\mathbf{A} + \mathbf{u}\mathbf{v}^T$  regularna onda i samo onda ako je  $\sigma = 1 + \mathbf{v}^T \mathbf{A}^{-1} \mathbf{u} \neq 0$ . U tom slučaju vrijedi formula

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{1}{\sigma} \mathbf{A}^{-1} \mathbf{u}\mathbf{v}^T \mathbf{A}^{-1}.$$

**Primjedba 3.5** Ako svi glavni minori regularne kvadratne matrice  $\mathbf{A}$  nisu različiti od nule (vidi Primjer 3.7), onda postoji matrica permutacija  $\mathbf{P}$ , takva da je  $\mathbf{PA} = \mathbf{LU}$  LU-dekompozicija matrice  $\mathbf{PA}$ . Matrica  $\mathbf{P}$  pojavljuje se prilikom provođenja Gaussovog postupka uz primjenu strategije parcijalnog pivotiranja (vidi GILL (1991)).

U nekim praktičnim situacijama (primjerice – analiza modela nacionalne privrede) treba više puta rješavati sustav  $\mathbf{Ax} = \mathbf{b}_i$ , gdje je matrica sustava (“matrica tehnologije”) uvijek ista, a vektor slobodnih koeficijenata (“outputi”) se mijenja. Također se može dogoditi da treba riješiti sustave  $\mathbf{Ax}_1 = \mathbf{b}_1$  i  $\mathbf{Ax}_2 = \mathbf{b}_2$ , gdje je  $\mathbf{b}_2$  neka funkcija od  $\mathbf{x}_1$ . U takvim situacijama korisno je poznavati LU-dekompoziciju matrice sustava  $\mathbf{A} = \mathbf{LU}$ .

Sustav  $\mathbf{Ax} = \mathbf{b}$  tada glasi

$$\mathbf{LUx} = \mathbf{b}, \quad (3.17)$$

Ako označimo

$$\mathbf{z} := \mathbf{Ux}, \quad (3.18)$$

onda (3.17) postaje sustav

$$\mathbf{Lz} = \mathbf{b}, \quad (3.19)$$

koji se lako rješava jer je matrica sustava  $\mathbf{L}$  donja trokutasta matrica s jedinicama na glavnoj dijagonali (*Algoritam FS*). Rješenje  $\mathbf{z}^*$  sustava (3.19) uvrstimo u (3.18), koji se sada lako rješava jer je matrica sustava  $\mathbf{U}$  gornja trokutasta matrica čiji dijagonalni elementi nisu nule (*Algoritam BS*). Rješenje  $\mathbf{x}^*$  sustava (3.18) je i rješenje polaznog sustava.

Primijetimo također da poznavajući LU-dekompoziciju matrice  $\mathbf{A}$ , prema *Binet-Cauchyjevom teoremu* (vidi primjerice JUKIĆ (1998), str. 244) lako možemo izračunati determinantu matrice  $\mathbf{A}$

$$\det \mathbf{A} = u_{11}u_{22} \cdots u_{nn}.$$

Matrice  $\mathbf{L}$  i  $\mathbf{U}$  u rastavu  $\mathbf{A} = \mathbf{LU}$  možemo dobiti primjerice tzv. Croutovom metodom (vidi STOER (1993)):

$$\begin{aligned} (1) \quad & a_{1j} = \sum_{k=1}^1 l_{1k}u_{kj}, \quad u_{1j} := a_{1j}, \quad j = 1, \dots, n \\ (2) \quad & a_{i1} = \sum_{k=1}^1 l_{ik}u_{k1}, \quad l_{i1} := a_{i1} := \frac{a_{i1}}{u_{11}}, \quad i = 2, \dots, n \\ (3) \quad & a_{2j} = \sum_{k=1}^2 l_{2k}u_{kj}, \quad u_{2j} := a_{2j} - l_{21}u_{1j}, \quad j = 2, \dots, n \\ & \vdots \end{aligned}$$



ili općenito:

**Algoritam** LU: za  $i = 1, \dots, n$

$$u_{ij} := a_{ij} - \sum_{k=1}^{i-1} l_{ik} u_{kj}, \quad j = i, n$$

$$l_{ji} := \frac{1}{u_{ii}} \left( a_{ji} - \sum_{k=1}^{i-1} l_{jk} u_{ki} \right), \quad j = i+1, \dots, n$$

```

LU[a_,n_]:=Module[{i,j,k},
  L=Table[0,{i,n},{j,n}]; U=Table[0,{i,n},{j,n}]; Do[L[[i,i]]=1,{i,n}];
  Do[
    Do[
      U[[i,j]]=a[[i,j]]-Sum[L[[i,k]]U[[k,j]],{k,i-1}],
      {j,i,n}];
    Do[
      L[[j,i]]=(a[[j,i]]-Sum[L[[j,k]]U[[k,i]],{k,i-1}))/U[[i,i]],
      {j,i+1,n}],{i,n}];
  Print["L=",MatrixForm[L],", U=",MatrixForm[U]]
]
```

**Primjer 3.4** Treba riješiti sustav linearnih jednadžbi gdje su matrica sustava  $\mathbf{A}$  i vektor slobodnih koeficijenata  $\mathbf{b}$ , zadani s

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 11 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Navedenim *Mathematica*-programom ili BASIC-programom u dodatku dobivamo

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & 0 & 2 \end{bmatrix}$$

Sukcesivnim rješavanjem jednadžbi (3.19) i (3.18), dobivamo rješenje sustava

$$x_1 = -1/3, \quad x_2 = 1/3, \quad x_3 = 0.$$

Rješenje možemo dobiti korištenjem FS i BS modula (nakon što smo aktivirali: LU[3,A]) kako slijedi

$$\text{FS}[3, \mathbf{L}, \mathbf{b}]; \quad \text{BS}[3, \mathbf{U}, \mathbf{x}]; \quad \mathbf{x}$$

**Zadatak 3.9** Izvedite formule navedene u Algoritmu LU koristeći definiciju jednakosti dviju matrica. Specijalizirajte algoritam za slučaj trodijagonalne matrice

**Primjer 3.5** Za kvadratnu regularnu matricu  $\mathbf{A} \in \mathbb{R}^{n \times n}$  vrijedi  $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ . Ako s  $\mathbf{x}_1, \dots, \mathbf{x}_n$  označimo vektore-stupce matrice  $\mathbf{A}^{-1}$ , a s:  $\mathbf{e}_1, \dots, \mathbf{e}_n$  vektore-stupce jedinične matrice  $\mathbf{I}$ , onda matricnu jednadžbu  $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$  možemo zapisati kao  $n$  sustava linearnih jednadžbi

$$\mathbf{A}\mathbf{x}_1 = \mathbf{e}_1, \quad \mathbf{A}\mathbf{x}_2 = \mathbf{e}_2, \quad \dots, \quad \mathbf{A}\mathbf{x}_n = \mathbf{e}_n,$$

Korištenjem LU-dekompozicije matrice  $\mathbf{A}$ , te ranije sagrađenih algoritama FS i BS možemo izračunati inverznu matricu. Primjerice za matricu iz *Primjera 3.4*, str. 50 imamo

```
(* Invertiranje matrice *)
In[1]:= n=3;
a={{1,4,7}, {2,5,8}, {3,6,11}};
LU[n,a];
inv=Table[0, {i,n}, {j,n}];
Do[FS[n,L,IdentityMatrix[n][[i]]];
  BS[n,U,x]; inv[[i]]=x, {i,n}]
Transpose[inv]/MatrixForm
Out[2]/MatrixForm=
```

$$\begin{bmatrix} -\frac{7}{6} & \frac{1}{3} & \frac{1}{2} \\ -\frac{1}{3} & \frac{5}{3} & -1 \\ \frac{1}{2} & -1 & \frac{1}{2} \end{bmatrix}$$

### 3.6 Cholesky - dekompozicija

Pretpostavimo da je  $\mathbf{A} \in \mathbb{R}^{n \times n}$  simetrična pozitivno definitna matrica. Tada je Gaussov postupak provediv bez strategije pivotiranja, pri čemu su svi pivot-elementi  $a_{11}, a_{22}^{(2)}, \dots$  pozitivni. Zbog simetričnosti matrice  $\mathbf{A}$ , množeći slijeva elementarnom matricom  $\mathbf{M}_1$  (vidi *t.3.5*, str. 48) i sdesna matricom  $\mathbf{M}_1^T$ , dobivamo

$$\tilde{\mathbf{A}}^{(2)} = \mathbf{M}_1 \mathbf{A} \mathbf{M}_1^T = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2}^{(2)} & \cdots & a_{nn}^{(2)} \end{bmatrix}.$$

Ponavljajući postupak kao u *t.3.5*, str. 48, nakon  $(n-1)$  koraka dobivamo

$$\mathbf{M}_{n-1} \cdots \mathbf{M}_2 \mathbf{M}_1 \mathbf{A} \mathbf{M}_1^T \mathbf{M}_2^T \cdots \mathbf{M}_{n-1}^T = \text{diag}(a_{11}, a_{22}^{(2)}, \dots, a_{nn}^{(n)}) =: \mathbf{D} \quad (3.20)$$

Kako su svi dijagonalni elementi matrice  $\mathbf{D}$  pozitivni, možemo definirati

$$\sqrt{\mathbf{D}} = \text{diag} \left( \sqrt{a_{11}}, \sqrt{a_{22}^{(2)}}, \dots, \sqrt{a_{nn}^{(n)}} \right),$$

a (3.20) zapisati kao

$$\mathbf{A} = \mathbf{M}_1^{-1} \cdots \mathbf{M}_{n-1}^{-1} \sqrt{\mathbf{D}} \sqrt{\mathbf{D}} (\mathbf{M}_{n-1}^T)^{-1} \cdots (\mathbf{M}_1^T)^{-1}.$$

Ako sada definiramo

$$\begin{aligned} \mathbf{L} &:= \mathbf{M}_1^{-1} \cdots \mathbf{M}_{n-1}^{-1} \sqrt{\mathbf{D}}, \\ \mathbf{L}^T &:= \sqrt{\mathbf{D}} (\mathbf{M}_{n-1}^T)^{-1} \cdots (\mathbf{M}_1^T)^{-1}, \end{aligned}$$

dobivamo Cholesky-dekompoziciju matrice  $\mathbf{A}$ :

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T, \quad (3.21)$$

gdje je  $\mathbf{L}$  donja trokutasta matrica. Dekompoziciju (3.21) neki puta je zgodnije pisati u obliku  $\mathbf{A} = \mathbf{R}^T\mathbf{R}$ , gdje je  $\mathbf{R} = \mathbf{L}^T$  gornja trokutasta matrica. Ako je Cholesky dekompozicija matrice  $\mathbf{A}$  poznata, onda sustav  $\mathbf{Ax} = \mathbf{b}$ , odnosno sustav  $\mathbf{L}\mathbf{L}^T\mathbf{x} = \mathbf{b}$  rješavamo tako da najprije *Algoritmom FS* riješimo sustav  $\mathbf{Lz} = \mathbf{b}$ . Rješenje polaznog sustava tada je  $\mathbf{x}^* = \mathbf{Lz}$ .

Elemente matrice  $\mathbf{L}$  redom, stupac po stupac, možemo dobiti na sljedeći način.

**Algoritam** Cholesky

Za  $k = 1, \dots, n$

$$l_{kk} = \left( a_{kk} - \sum_{p=1}^{k-1} l_{kp}^2 \right)^{1/2}, \quad l_{ik} = \frac{1}{l_{kk}} \left( a_{ik} - \sum_{p=1}^{k-1} l_{ip}l_{kp} \right), \quad i = k+1, \dots, n.$$

(pri čemu navedene sume nećemo izvoditi ako je gornja granica sume manja od donje granice).

```
Cholesky[a_, n_] :=
Module[{i,j,k,p}, L=Table[0,{i,n},{j,n}];
Do[L[[k,k]] =
Sqrt[a[[k,k]] - Sum[L[[k,p]]^2, {p,k-1}]];
Do[
L[[i,k]] =
(a[[i,k]] - Sum[L[[i,p]] L[[k,p]], {p,k-1}))/L[[k,k]], {i,k+1,n}];
Print["L=", MatrixForm[L]]
]
```

**Primjer 3.6** Cholesky dekompozicijom treba riješiti sustav  $\mathbf{Hx} = \mathbf{b}$ , gdje je  $\mathbf{H}$  Hilbertova matrica 3. reda (str. 47), a  $\mathbf{b} = (1, 1, 1)^T$ .

*Algoritmom Cholesky* lako dobivamo

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2\sqrt{3}} & 0 \\ \frac{1}{3} & \frac{1}{2\sqrt{3}} & \frac{1}{6\sqrt{5}} \end{bmatrix}$$

i rješenje sustava  $\mathbf{x}^* = (3, -24, 30)^T$ .

### 3.7 QR-dekompozicija

QR-dekompozicija, koju bismo željeli sada izložiti, matricu  $\mathbf{A} \in \mathbb{R}^{m \times n}$  rastavlja na produkt  $\mathbf{QR}$ , gdje je  $\mathbf{Q}$  ortogonalna, a  $\mathbf{R}$  gornja trokutasta matrica. Na taj način sustav  $\mathbf{Ax} = \mathbf{b}$  prelazi u trokutasti sustav

$$\mathbf{Rx} = \mathbf{Q}^T\mathbf{b},$$

koji se lako rješava *Algoritmom BS*. Ono što je ovdje od izuzetnog značaja je činjenica da se pri tome broj uvjetovanosti nije promijenio, tj.

$$\text{cond}(\mathbf{A}) = \text{cond}(\mathbf{R}).$$

I to nije sve! QR-dekompozicija je provediva i u slučaju ako je matrica  $\mathbf{A}$  singularna, pa čak i ako je  $\mathbf{A}$  pravokutna matrica nepunog ranga po retcima ili stupcima. Mi ćemo razmotriti dovoljno općeniti slučaj, koji se javlja kod rješavanja sustava linearnih jednadžbi i kod rješavanja tzv. linearnog problema najmanjih kvadrata (*t.6.1*, str. 108).

Pretpostavimo dakle, da je matrica  $\mathbf{A} \in \mathbb{R}^{m \times n}$  ( $m \geq n$ ) punog ranga po stupcima ( $\text{rang } \mathbf{A} = n$ ), tada postoji ortogonalna matrica<sup>7</sup>  $\mathbf{Q} \in \mathbb{R}^{m \times m}$  i gornja trokutasta matrica  $\mathbf{R}$  s elementima na glavnoj dijagonali koji su različiti od nule, tako da je  $\mathbf{A} = \mathbf{QR}$ . Ovaj rastav nazivamo QR-dekompozicija matrice  $\mathbf{A}$ .

$$\begin{array}{c} \begin{array}{|c|} \hline n \\ \hline \mathbf{A} \\ \hline m \end{array} = \begin{array}{|c|} \hline m \\ \hline \mathbf{Q} \\ \hline m \end{array} = \begin{array}{|c|} \hline n \\ \hline \begin{array}{c} \text{---} \text{ } 0 \text{ ---} \\ \text{ } 0 \end{array} \\ \hline m \end{array} \left. \vphantom{\begin{array}{|c|} \hline n \\ \hline \mathbf{A} \\ \hline m \end{array}} \right\} \mathbf{R}_n \\ \mathbf{R} \end{array}$$

QR-dekompozicija matrice  $\mathbf{A}$  može se dobiti na više načina. Spomenimo samo Gram-Schmidtov postupak ortogonalizacije, Givensove rotacije i Householderove transformacije (vidi BJÖRCK (1990), GILL (1991), GOLUB (1996), LAWSON (1996), STEWART (1998)). Mi ćemo QR-dekompoziciju graditi primjenom Householderovih transformacija.

U tu svrhu najprije ćemo uvesti pojam tzv. Householderove matrice:

$$\mathbf{H} = \mathbf{H}(\mathbf{u}) := \mathbf{I} - 2\mathbf{u}\mathbf{u}^T, \quad \mathbf{u}^T \mathbf{u} = 1, \quad \mathbf{u} \in \mathbb{R}^m, \quad \mathbf{H} \in \mathbb{R}^{m \times m}, \quad (3.22)$$

Vektor  $\mathbf{u}$  iz (3.22) nazivamo Householderov vektor. Householderova matrica je očigledno simetrična i ortogonalna. Posebno je važno sljedeće svojstvo ove matrice:

**Lema 3.1** *Za dva vektora  $\mathbf{a}$ ,  $\mathbf{b}$  jednake Euklidove norme postoji Householderov vektor  $\mathbf{u}$ , tako da je  $\mathbf{H}\mathbf{a} = \mathbf{b}$ , odnosno, zbog simetričnosti matrice  $\mathbf{H}$ ,  $\mathbf{H}\mathbf{b} = \mathbf{a}$ .*

**Dokaz.** Ako je  $\mathbf{a} = \mathbf{b}$ , onda je  $\mathbf{H} = \mathbf{I}$ . Pretpostavimo zato da su  $\mathbf{a}$ ,  $\mathbf{b}$  dva vektora takva da je

$$\mathbf{a} \neq \mathbf{b} \quad \& \quad \|\mathbf{a}\| = \|\mathbf{b}\| \quad (3.23)$$

---

<sup>7</sup>Ortogonalna matrica  $\mathbf{Q}$  je takva kvadratna matrica za koju vrijedi  $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$ , odnosno

$$\sum_k q_{ik}q_{jk} = \delta_{ij} \quad \& \quad \sum_k q_{ki}q_{kj} = \delta_{ij},$$

tj. njeni retci i njeni stupci čine ortonormirani sustav vektora. Ona je regularna matrica za koju je  $\mathbf{Q}^{-1} = \mathbf{Q}^T$ . Ima važno svojstvo da čuva  $L_2$  normu, tj.

$$\|\mathbf{Q}\mathbf{a}\|^2 = (\mathbf{Q}\mathbf{a})^T(\mathbf{Q}\mathbf{a}) = \mathbf{a}^T\mathbf{Q}^T\mathbf{Q}\mathbf{a} = \mathbf{a}^T\mathbf{a} = \|\mathbf{a}\|^2.$$

Ortogonalna matrica geometrijski predstavlja rotaciju ili simetriju prostora

Ovo također znači da vektori  $\mathbf{a}, \mathbf{b}$  ne mogu biti nulvektori. Primijetimo također da zbog  $\mathbf{a} \neq \mathbf{b}$  slijedi  $\|\mathbf{a} - \mathbf{b}\| > 0$ . Zato možemo definirati Householderov vektor

$$\mathbf{u} = \frac{1}{\|\mathbf{a} - \mathbf{b}\|}(\mathbf{a} - \mathbf{b}).$$

Kako je

$$\mathbf{a} - \mathbf{H}\mathbf{b} = (\mathbf{a} - \mathbf{b}) + 2 \frac{\mathbf{a} - \mathbf{b}}{\|\mathbf{a} - \mathbf{b}\|} \frac{(\mathbf{a} - \mathbf{b})^T}{\|\mathbf{a} - \mathbf{b}\|} \mathbf{b} = \frac{\mathbf{a} - \mathbf{b}}{\|\mathbf{a} - \mathbf{b}\|^2} (\|\mathbf{a} - \mathbf{b}\|^2 + 2(\mathbf{a} - \mathbf{b})^T \mathbf{b}),$$

onda zbog

$$\|\mathbf{a} - \mathbf{b}\|^2 + 2(\mathbf{a} - \mathbf{b})^T \mathbf{b} = \mathbf{a}^T \mathbf{a} - 2\mathbf{a}^T \mathbf{b} + \mathbf{b}^T \mathbf{b} + 2\mathbf{a}^T \mathbf{b} - 2\mathbf{b}^T \mathbf{b} = \mathbf{a}^T \mathbf{a} - \mathbf{b}^T \mathbf{b},$$

vrijedi  $\mathbf{H}\mathbf{b} = \mathbf{a}$ . □

**Primjer 3.7** Vektori  $\mathbf{a} = (\frac{9}{2}, 1)^T$  i  $\mathbf{b} = (3, \frac{7}{2})^T$  imaju svojstvo (3.23). Definirat ćemo Householderov vektor  $\mathbf{u}$ , tako da bude  $\mathbf{H}\mathbf{a} = \mathbf{b}$ .

Stavimo

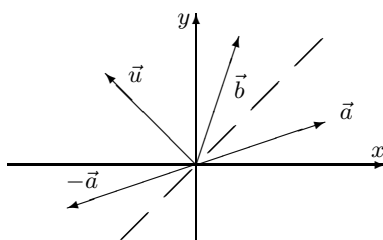
$$\mathbf{u} = \frac{(\mathbf{a} - \mathbf{b})}{\|\mathbf{a} - \mathbf{b}\|} = \frac{2}{\sqrt{34}} \left( \frac{3}{2}, -\frac{5}{2} \right)^T.$$

Householderova matrica glasi

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - 2 \frac{4}{34} \begin{bmatrix} \frac{3}{2} \\ -\frac{5}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{3}{2} & -\frac{5}{2} \end{bmatrix} = \begin{bmatrix} \frac{8}{17} & \frac{15}{17} \\ \frac{15}{17} & -\frac{8}{17} \end{bmatrix}$$

i vrijedi

$$\mathbf{H}\mathbf{a} = \begin{bmatrix} \frac{8}{17} & \frac{15}{17} \\ \frac{15}{17} & -\frac{8}{17} \end{bmatrix} \cdot \begin{bmatrix} \frac{9}{2} \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ \frac{7}{2} \end{bmatrix} = \mathbf{b}.$$



Slika 3.2

Na Slici 3.2 vidi se da su vektori  $\mathbf{a}$  i  $\mathbf{b}$  osno simetrični na pravac kroz ishodište, koji je okomit na vektor  $\mathbf{u}$ . Dakle, Householderova matrica u geometrijskom smislu predstavlja zrcaljenje ili rotaciju prostora.

**Zadatak 3.10** Zadana je točka  $T(4, 3)$ . Odredite točku  $T'$  na pozitivnom dijelu osi  $x$ , tako da bude  $\|\vec{\mathbf{r}}_T\| = \|\vec{\mathbf{r}}_{T'}\|$ , gdje su  $\vec{\mathbf{r}}_T, \vec{\mathbf{r}}_{T'}$  radij-vektori točaka  $T, T'$ . Odredite također odgovarajuću Householderovu matricu  $\mathbf{H}$ , koja će vektor  $\vec{\mathbf{r}}_T$  preslikavati u vektor  $\vec{\mathbf{r}}_{T'}$ .

Rješenje:  $T'(5, 0)$ ,  $\mathbf{H} = \frac{1}{5} \begin{bmatrix} 4 & 3 \\ 3 & -4 \end{bmatrix}$ .

**Primjedba 3.6** Množenje nekog vektora  $\mathbf{c}$  s Householderovom matricom  $\mathbf{H}$  može se jednostavno provesti. Naime, zbog

$$\mathbf{H}\mathbf{c} = \mathbf{c} - 2(\mathbf{u}^T\mathbf{c})\mathbf{u},$$

dovoljno je napraviti skalarni produkt  $\mathbf{u}^T\mathbf{c}$ , množenje vektora sa skalarom  $2(\mathbf{u}^T\mathbf{c})\mathbf{u}$  i oduzimanje vektora  $\mathbf{c} - 2(\mathbf{u}^T\mathbf{c})\mathbf{u}$ .

U cilju svođenja matrice  $\mathbf{A}$  na gornje trokutasti oblik, njene stupce promatrat ćemo kao vektore u  $\mathbb{R}^m$ . Pretpostavimo općenito da je  $\mathbf{a} \in \mathbb{R}^m$ , da je  $k \geq 1$  i da je barem jedna od komponenti  $a_k, \dots, a_m$  vektora  $\mathbf{a}$  različita od nule. Konstruirat ćemo Householderovu matricu  $\mathbf{H}$  tako da se prvih  $(k-1)$  komponenti vektora  $\mathbf{b} := \mathbf{H}\mathbf{a}$  i vektora  $\mathbf{a}$  podudaraju, a da se posljednjih  $(m-k)$  komponenti vektora  $\mathbf{b}$  poništavaju. Dakle, treba biti

$$(\mathbf{H}\mathbf{a})_i = b_i = \begin{cases} a_i, & i = 1, \dots, k-1 \\ \pm\gamma, & i = k \\ 0, & i = k+1, \dots, m \end{cases}, \quad \gamma = \sqrt{a_k^2 + \dots + a_m^2}.$$

Pri tome ćemo predznak  $(\pm\gamma)$  birati tako da bude suprotan predznaku od  $a_k$ . Primijetimo da je stoga  $b_k = -\text{sgn}(a_k)\gamma$ . Očigledno je  $\|\mathbf{a}\| = \|\mathbf{b}\|$ .

Potražimo oblik vektora  $\mathbf{u}$ . Koristeći jednakost  $|a_k| = a_k \text{sgn}(a_k)$ , dobivamo

$$\begin{aligned} \|\mathbf{a} - \mathbf{b}\|^2 &= (a_k + \text{sgn}(a_k)\gamma)^2 + a_{k+1}^2 + \dots + a_m^2 \\ &= a_k^2 + 2\gamma a_k \text{sgn}(a_k) + \gamma^2 + a_{k+1}^2 + \dots + a_m^2 = 2\gamma(\gamma + |a_k|). \end{aligned}$$

Zato je

$$\mathbf{u} = \frac{\mathbf{a} - \mathbf{b}}{\|\mathbf{a} - \mathbf{b}\|} = \frac{1}{\sqrt{2\gamma(\gamma + |a_k|)}} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ a_k + \text{sgn}(a_k)\gamma \\ a_{k+1} \\ \vdots \\ a_m \end{bmatrix}.$$

**Primjer 3.8** Treba konstruirati Householderovu matricu  $\mathbf{H}$  tako da poništi drugu komponentu vektora  $\mathbf{a} = (4, 3)^T$ .

U ovom slučaju je  $k = 1$ ,  $\gamma = \sqrt{16 + 9} = 5$ ,

$$\mathbf{u} = \frac{1}{\sqrt{10(5+4)}} \begin{bmatrix} 4+5 \\ 3 \end{bmatrix} = \frac{1}{\sqrt{90}} \begin{bmatrix} 9 \\ 3 \end{bmatrix}, \quad \mathbf{H} = -\frac{1}{5} \begin{bmatrix} 4 & 3 \\ 3 & -4 \end{bmatrix}.$$

Dobivamo:  $\mathbf{H}\mathbf{a} = (-5, 0)^T$ .

**Zadatak 3.11** Konstruirajte Householderovu matricu  $\mathbf{H}$  tako da poništi samo treću komponentu vektora  $\mathbf{a} = (7, 4, 3)^T$ .

Rješenje:

$$k = 2, \quad \gamma = 5, \quad \mathbf{u} = \frac{1}{\sqrt{90}} \begin{bmatrix} 0 \\ 9 \\ 3 \end{bmatrix}, \quad \mathbf{H} = -\frac{1}{5} \begin{bmatrix} 5 & 0 & 0 \\ 0 & -4 & -3 \\ 0 & -3 & 4 \end{bmatrix}, \quad \mathbf{H}\mathbf{a} = \begin{bmatrix} 7 \\ -5 \\ 0 \end{bmatrix}.$$

**Primjer 3.9** Primjenom Householderovih transformacija, matricu  $\mathbf{A}$  treba transformirati u gornju trokutastu

$$\mathbf{A} = \begin{bmatrix} 10 & 9 & 18 \\ 20 & -15 & -15 \\ 20 & -12 & 51 \end{bmatrix}.$$

1. Najprije stavimo  $k = 1$  i  $\mathbf{a} = (10, 20, 20)^T$ . Dobivamo

$$\gamma = 30, \quad \mathbf{u}^{(1)} = \frac{1}{\sqrt{60(30+10)}} \begin{bmatrix} 10+30 \\ 20 \\ 20 \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix},$$

$$\mathbf{H}_1 = -\frac{1}{3} \begin{bmatrix} 1 & 2 & 2 \\ 2 & -2 & 1 \\ 2 & 1 & -2 \end{bmatrix}, \quad \mathbf{A}^{(1)} := \mathbf{H}_1 \mathbf{A} = \begin{bmatrix} -30 & 15 & -30 \\ 0 & -12 & -39 \\ 0 & -9 & 27 \end{bmatrix}.$$

2. Stavimo  $k = 2$ ,  $\mathbf{a} = (15, -12, -9)^T$ . Dobivamo

$$\gamma = 15, \quad \mathbf{u}^{(2)} = \frac{1}{\sqrt{30(15+12)}} \begin{bmatrix} 0 \\ -12-15 \\ -9 \end{bmatrix} = \frac{1}{\sqrt{10}} \begin{bmatrix} 0 \\ -3 \\ -1 \end{bmatrix},$$

$$\mathbf{H}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.8 & -0.6 \\ 0 & -0.6 & 0.8 \end{bmatrix}, \quad \mathbf{A}^{(2)} := \mathbf{H}_2 \mathbf{A}^{(1)} = \begin{bmatrix} -30 & 15 & -30 \\ 0 & 15 & 15 \\ 0 & 0 & 45 \end{bmatrix}.$$

Kako je  $\mathbf{A}^{(2)} = \mathbf{H}_2 \mathbf{A}^{(1)} = \mathbf{H}_2 \mathbf{H}_1 \mathbf{A}$ , dobivamo dekompoziciju  $\mathbf{A} = \mathbf{Q}\mathbf{R}$ , gdje je

$$\mathbf{Q} = (\mathbf{H}_1 \mathbf{H}_2) = \begin{bmatrix} -\frac{1}{3} & \frac{14}{15} & -\frac{2}{15} \\ -\frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ -1 & -\frac{2}{15} & \frac{11}{15} \end{bmatrix}, \quad \mathbf{R} = \mathbf{A}^{(2)}.$$

Slično kao u prethodnom primjeru, sukcesivnom primjenom Householderovih matrica  $\mathbf{H}_k$  na matricu  $\mathbf{A} \in \mathbb{R}^{m \times n}$  možemo dobiti gornju trokutastu matricu  $\mathbf{R}$

$$\mathbf{R} = \mathbf{H}_m \mathbf{H}_{m-1} \cdots \mathbf{H}_2 \mathbf{H}_1 \mathbf{A},$$

iz čega zbog  $\mathbf{H}_k^{-1} = \mathbf{H}_k^T = \mathbf{H}_k$  slijedi

$$\mathbf{A} = \mathbf{H}_1 \mathbf{H}_2 \cdots \mathbf{H}_m \mathbf{R}.$$

Kako je produkt ortogonalnih matrica opet ortogonalna matrica, dobivamo QR-dekompoziciju matrice  $\mathbf{A}$

$$\mathbf{A} = \mathbf{Q}\mathbf{R}, \quad \text{gdje je } \mathbf{Q} = \mathbf{H}_1 \mathbf{H}_2 \cdots \mathbf{H}_{m-1} \mathbf{H}_m.$$

### 3.8 Iterativne metode

Do sada razmatrane metode za rješavanje sustava  $\mathbf{Ax} = \mathbf{b}$  (Gaussov postupak, LU-dekompozicija, Cholesky-dekompozicija, QR-dekompozicija) spadaju u tzv. *direktne metode*. Budući da je broj potrebnih računskih operacija kod ovih metoda reda veličine  $n^3$ , za velike matrice s puno elemenata koji iščezavaju (tzv. *large sparse matrices*) ove metode ne mogu se preporučiti. Osim toga, treba ekonomizirati i s brojem elemenata matrice za koje treba rezervirati mjesto u memoriji računala. Takve situacije javljaju se primjerice kod proučavanja električnih mreža, kod velikih ekonometrijskih modela nacionalne privrede, kod spline-interpolacija, kod rješavanja rubnih problema za obične i parcijalne diferencijalne jednačbe, itd. Upravo se iterativne metode, o kojima ćemo reći samo nekoliko osnovnih činjenica, koriste u takvim situacijama.

Spomenut ćemo dvije klasične iterativne metode za rješavanje sustava  $\mathbf{Ax} = \mathbf{b}$ ,  $\mathbf{A} \in \mathbb{R}^{n \times n}$ . Pretpostavljamo da je  $a_{ii} \neq 0$ ,  $i = 1, \dots, n$ .

- **Jacobijeva metoda**

$$\begin{aligned} x_1^{(k+1)} &= \frac{1}{a_{11}}(b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)} - \dots - a_{1,n-1}x_{n-1}^{(k)} - a_{1n}x_n^{(k)}) \\ x_2^{(k+1)} &= \frac{1}{a_{22}}(b_2 - a_{21}x_1^{(k)} - a_{23}x_3^{(k)} - \dots - a_{2,n-1}x_{n-1}^{(k)} - a_{2n}x_n^{(k)}) \\ x_3^{(k+1)} &= \frac{1}{a_{33}}(b_3 - a_{31}x_1^{(k)} - a_{32}x_2^{(k)} - \dots - a_{3,n-1}x_{n-1}^{(k)} - a_{3n}x_n^{(k)}) \\ &\vdots \\ x_n^{(k+1)} &= \frac{1}{a_{nn}}(b_n - a_{n1}x_1^{(k)} - a_{n2}x_2^{(k)} - a_{n3}x_3^{(k)} - \dots - a_{n,n-1}x_{n-1}^{(k)}) \end{aligned}$$

odnosno:

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left( b_i - \sum_{j \neq i}^n a_{ij}x_j^{(k)} \right), \quad i = 1, \dots, n \quad (3.24)$$

- **Gauss-Seidelova metoda**

$$\begin{aligned} x_1^{(k+1)} &= \frac{1}{a_{11}}(b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)} - \dots - a_{1,n-1}x_{n-1}^{(k)} - a_{1n}x_n^{(k)}) \\ x_2^{(k+1)} &= \frac{1}{a_{22}}(b_2 - a_{21}x_1^{(k+1)} - a_{23}x_3^{(k)} - \dots - a_{2,n-1}x_{n-1}^{(k)} - a_{2n}x_n^{(k)}) \\ x_3^{(k+1)} &= \frac{1}{a_{33}}(b_3 - a_{31}x_1^{(k+1)} - a_{32}x_2^{(k+1)} - \dots - a_{3,n-1}x_{n-1}^{(k)} - a_{3n}x_n^{(k)}) \\ &\vdots \\ x_n^{(k+1)} &= \frac{1}{a_{nn}}(b_n - a_{n1}x_1^{(k+1)} - a_{n2}x_2^{(k+1)} - a_{n3}x_3^{(k+1)} - \dots - a_{n,n-1}x_{n-1}^{(k+1)}) \end{aligned}$$

odnosno:

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left( b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij}x_j^{(k)} \right), \quad i = 1, \dots, n \quad (3.25)$$



**Primjer 3.10** *Jacobijevom odnosno Gauss-Seidelovom metodom riješit ćemo sustav iz Primjera 2.7, str. 31:*

$$\begin{bmatrix} 1 & 0.25 & 0 \\ 0.25 & 1 & 0.25 \\ 0 & 0.25 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -48 \\ 0 \\ 48 \end{bmatrix}.$$

Jacobijeva metoda daje rješenje  $\mathbf{x}^* = (-48, 0, 48)^T$  već u prvom koraku. Gauss-Seidelova metoda također konvergira rješenju. Prvih nekoliko iteracija prikazano je u *Tablici 3.1*.

$k$	$x_1^k$	$x_2^k$	$x_3^k$
0	0	0	0
1	-48.0	12.0	45.0
2	-51.0	1.5	47.625
3	-48.375	0.1875	47.9531
4	-48.0469	0.02344	47.9941
5	-48.0059	0.00293	47.9993
6	-48.0007	0.00037	47.9999
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Tablica 3.1

Izračunavanje je obavljeno na bazi niže navedenih *Mathematica*-modula.

```
Jacobi[a_,b_,n_,it_] := Module[{i,k,xs,xn}, xs=Table[0, {i,n}]; xn=xs;
Do[
  Do[
    xn[[i]]=(b[[i]]-Apply[Plus,ReplacePart[a[[i]],0,{i}] xs])/a[[i,i]],
    {i, n}];
  Print["It_", k, " = ", N[xn]]; xs=xn,
  {k, it}]

GS[a_,b_,n_,it_] := Module[{i,k,xs,xn}, xs = Table[0, {i,n}]; xn=xs;
Do[
  Do[
    xn[[i]]=(b[[i]] - Sum[a[[i,j]] xn[[j]], {j, i-1}]
      - Sum[a[[i,j]] xs[[j]], {j, i+1,n}))/a[[i,i]], {i,n}];
  Print["It_", k, " = ", N[xn]]; xs = xn,
  {k, it}]
]
```

U cilju ispitivanja konvergencije navedenih, pa i drugih iterativnih metoda, na Banachovom prostoru  $\mathbb{R}^n$  (snabdjevenom s jednom od normi navedenih u *t.3.1*) uvest ćemo metriku  $d : \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, \infty)$ ,

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|, \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

čime  $\mathbb{R}^n$  postaje i potpuni metrički prostor.

**Definicija 3.3** Preslikavanje  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  zovemo kontrakcija ako postoji takav realni broj  $q < 1$ , da bude

$$d(F(\mathbf{x}), F(\mathbf{y})) < q d(\mathbf{x}, \mathbf{y}) \quad \text{za sve } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

**Primjer 3.11** Funkcija  $F : [1, \infty) \rightarrow \mathbb{R}$ ,  $F(x) = \frac{x}{2} + \frac{1}{x}$  je kontrakcija s  $q = 1/2$ . Naime

$$\begin{aligned} d(F(x), F(y)) &= |F(x) - F(y)| = \left| \frac{x}{2} + \frac{1}{x} - \frac{y}{2} - \frac{1}{y} \right| = \left| \frac{1}{2}(x - y) - \frac{x - y}{xy} \right| \\ &= |x - y| \left| \frac{1}{2} - \frac{1}{xy} \right| \leq \frac{1}{2} |x - y| = \frac{1}{2} d(x, y) \end{aligned}$$

**Teorem 3.1** (Banachov teorem o fiksnoj točki) Neka je funkcija  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  kontrakcija. Tada postoji jedinstvena fiksna točka  $\mathbf{x}^* \in \mathbb{R}^n$  funkcije  $F$ , tj.

$$F(\mathbf{x}^*) = \mathbf{x}^*.$$

Osim toga, za proizvoljni  $\mathbf{x}_0 \in \mathbb{R}^n$ , niz definiran rekurzivnom formulom  $\mathbf{x}_{k+1} = F(\mathbf{x}^k)$  konvergira prema  $\mathbf{x}^*$  i vrijedi formula za ocjenu pogreške

$$d(\mathbf{x}^{k+1}, \mathbf{x}^*) \leq \frac{q}{1-q} d(\mathbf{x}^k, \mathbf{x}^{k+1}) \leq \frac{q^k}{1-q} d(\mathbf{x}_0, F(\mathbf{x}_0)), \quad k = 0, 1, \dots$$

Ako matricu sustava  $\mathbf{A}$  rastavimo na donji trokut, dijagonalu i gornji trokut,

$$\mathbf{A} = \mathbf{D}(\mathbf{L} + \mathbf{I} + \mathbf{U}), \quad \mathbf{D} = \text{diag}(a_{11}, \dots, a_{nn}),$$

dobivamo matični oblik

- Jacobijeve metode

$$\mathbf{x}^{(k+1)} = -(\mathbf{L} + \mathbf{U})\mathbf{x}^{(k)} + \mathbf{D}^{-1}\mathbf{b}, \quad k = 0, 1, \dots$$

- Gauss-Seidelove metode

$$\mathbf{x}^{(k+1)} = -\mathbf{L}\mathbf{x}^{(k+1)} - \mathbf{U}\mathbf{x}^{(k)} + \mathbf{D}^{-1}\mathbf{b} \quad k = 0, 1, \dots$$

Sljedeća rekurzivna formula obuhvaća obje metode

$$\mathbf{x}^{(k+1)} = \mathbf{B}\mathbf{x}^{(k)} + \mathbf{c}, \quad k = 0, 1, \dots; \quad (3.26)$$

za  $\mathbf{B} = \mathbf{B}_J := -(\mathbf{L} + \mathbf{U})$  dobivamo Jacobijevu, a za  $\mathbf{B} = \mathbf{B}_{GS} := -(\mathbf{I} + \mathbf{L})^{-1}\mathbf{U}$  dobivamo Gauss-Seidelovu metodu.

Primijetimo da je funkcija  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $F(\mathbf{x}) = \mathbf{B}\mathbf{x} + \mathbf{c}$  kontrakcija onda i samo onda ako je  $\|\mathbf{B}\| = q < 1$ . Tada prema Banachovom teoremu o fiksnoj točki niz  $(x^{(n)})$  definiran s (3.26) konvergira prema jedinstvenom rješenju sustava  $\mathbf{A}\mathbf{x} = \mathbf{b}$  i vrijedi ocjena pogreške

$$\|\mathbf{x}^{(k)} - \mathbf{x}^*\| \leq \frac{q}{1-q} \|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|.$$

Kod Jacobijeve metode matrica  $\mathbf{B}_J$  ima elemente  $b_{ij} = \frac{a_{ij}}{a_{ii}}$ ,  $i \neq j$ ,  $b_{ii} = 0$ , pa je

$$\|\mathbf{B}_J\|_\infty = \max_{i=1,\dots,n} \sum_{j \neq i}^n \frac{|a_{ij}|}{|a_{ii}|}.$$

Dakle, Jacobijeva metoda konvergiraće ako je  $\mathbf{A}$  strogo dijagonalno dominantna matrica.

Slično se može pokazati da će i Gauss-Seidelova metoda konvergirati ako je  $\mathbf{A}$  strogo dijagonalno dominantna matrica.

**Primjedba 3.7** Jednostavnom modifikacijom Gauss-Seidelove metode moguće je ubrzati konvergenciju iterativnog procesa. Proces (3.25) možemo zapisati kao

$$x_i^{(k+1)} = x_i^{(k)} + r_i^{(k)}, \quad i = 1, \dots, n, \quad (3.27)$$

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

Uvođenjem relaksacijskog parametra  $\omega$ , proces (3.27) možemo modificirati tako da iterativni proces

$$x_i^{(k+1)} = x_i^{(k)} + \omega r_i^{(k)}, \quad i = 1, \dots, n, \quad (3.28)$$

koji se u literaturi nalazi pod imenom Successive overrelaxation (SOR) method, brže konvergira. Može se pokazati da odgovarajuća matrica  $\mathbf{B}$  iz (3.26) u ovom slučaju glasi

$$\mathbf{B} := \mathbf{B}_\omega = (\mathbf{I} + \omega \mathbf{L})^{-1} ((\mathbf{I} - \omega \mathbf{I}) - \omega \mathbf{U}),$$

i da odgovarajući iterativni proces konvergira za  $\omega \in (0, 2)$ .

### 3.9 Dekompozicija na singularne vrijednosti

Svaku matricu  $\mathbf{A} \in \mathbb{R}^{m \times n}$  možemo zapisati kao (vidi primjerice BJÖRCK (1996), GILL (1991), GOLUB (1996) ili LAWSON (1996))

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \quad (3.29)$$

gdje su  $\mathbf{U} \in \mathbb{R}^{m \times m}$  i  $\mathbf{V} \in \mathbb{R}^{n \times n}$  ortogonalne matrice, a  $\mathbf{S} \in \mathbb{R}^{m \times n}$  dijagonalna matrica

$$\mathbf{S} = \text{diag}(\sigma_1, \dots, \sigma_p), \quad p = \min(m, n), \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0. \quad (3.30)$$

Nenegativne brojeve  $\sigma_1, \dots, \sigma_p$  zovemo singularne vrijednosti matrice  $\mathbf{A}$ , a rastav (3.29) rastav na singularne vrijednosti (*Singular Value Decomposition*).

Kako je  $\mathbf{A}^T \mathbf{A} = \mathbf{V} \mathbf{S}^T \mathbf{S} \mathbf{V}^T$ , kvadrati singularnih vrijednosti  $\sigma_i^2$  su svojstvene vrijednosti<sup>8</sup> simetrične pozitivno semidefinitne kvadratne matrice  $\mathbf{A}^T \mathbf{A}$ , a stupci matrice  $\mathbf{V}$  su odgovarajući ortonormirani svojstveni vektori. Analogno, kako je  $\mathbf{A} \mathbf{A}^T = \mathbf{U} \mathbf{S} \mathbf{S}^T \mathbf{U}^T$ , kvadrati singularnih vrijednosti  $\sigma_i^2$  također su i svojstvene vrijednosti matrice  $\mathbf{A} \mathbf{A}^T$ , a stupci matrice  $\mathbf{U}$  su odgovarajući ortonormirani svojstveni vektori.

Ako s  $\mathbf{u}_i \in \mathbb{R}^m$  i  $\mathbf{v}_i \in \mathbb{R}^n$  označimo stupce matrice  $\mathbf{U}$ , odnosno  $\mathbf{V}$ , i ako je  $\text{rang } \mathbf{A} = k \leq p$ , tada iz (3.29) slijedi

$$\mathbf{A} \mathbf{v}_i = \sigma_i \mathbf{u}_i \quad \& \quad \mathbf{A}^T \mathbf{u}_i = \sigma_i \mathbf{v}_i, \quad i = 1, \dots, p. \quad (3.31)$$

Vektore  $\mathbf{u}_i$  zovemo lijevi, a vektore  $\mathbf{v}_i$  desni singularni vektori matrice  $\mathbf{A}$ .

**Primjedba 3.8** Singularne vrijednosti matrice  $\mathbf{A} \in \mathbb{R}^{m \times n}$  mogu se interpretirati i kao svojstvene vrijednosti matrice  $(\mathbf{A}^T \mathbf{A})^{1/2}$

Nadalje, iz (3.29)-(3.30) lako dobivamo  $\mathbf{A}^T \mathbf{A} = \mathbf{V} \mathbf{S}^T \mathbf{S} \mathbf{V}^T$ . Zato SVD matrice  $\mathbf{A} \in \mathbb{R}^{m \times n}$  možemo izračunati tako da najprije načinimo spektralnu dekompoziciju  $\mathbf{A}^T \mathbf{A} = \mathbf{V} \mathbf{D} \mathbf{V}^T$  i definiramo dijagonalnu matricu  $\mathbf{S} \in \mathbb{R}^{m \times n}$ , čiji je gornji  $(n \times n)$  blok matrica  $+\sqrt{\mathbf{D}}$ . Nakon toga matricu  $\mathbf{U}$  možemo dobiti iz jednadžbe  $\mathbf{U} \mathbf{S} = \mathbf{A} \mathbf{V}$ . Nažalost, opisani postupak nije numerički stabilan. Numerički stabilni SVD algoritmi obično se provode u dvije faze: najprije se provede ortogonalna bidijagonalizacija matrice  $\mathbf{A}$ , a nakon toga se nad dobivenom bidijagonalnom matricom provede prilagođeni SVD postupak (vidi primjerice GOLUB (1996), TREFETHEN (1997)).

**Primjer 3.12** (Björck, 1996) Uzmimo primjerice da matrica  $\mathbf{A}$  ima dva stupca  $\mathbf{a}_1, \mathbf{a}_2$  koji su ujedno jedinični vektori, a kut između njih je  $\gamma$ , tj.  $\mathbf{a}_1^T \mathbf{a}_2 = \cos \gamma$ .

Tada je

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 1 & \cos \gamma \\ \cos \gamma & 1 \end{bmatrix},$$

sa svojstvenim vrijednostima  $\lambda_1 = 2 \cos^2 \frac{\gamma}{2}$ ,  $\lambda_2 = 2 \sin^2 \frac{\gamma}{2}$ . Singularne vrijednosti matrice  $\mathbf{A}$  su

$$\sigma_1 = \sqrt{2} \cos \frac{\gamma}{2}, \quad \sigma_2 = \sqrt{2} \sin \frac{\gamma}{2}.$$

<sup>8</sup>Kažemo da je  $\lambda \in \mathbb{C}$  svojstvena vrijednost kvadratne matrice  $\mathbf{A} \in \mathbb{R}^{n \times n}$  ako postoji  $\mathbf{x} \neq \mathbf{0}$  iz  $\mathbb{R}^n$  tako da bude  $\mathbf{A} \mathbf{x} = \lambda \mathbf{x}$ . Spomenuti vektor  $\mathbf{x}$  nazivamo svojstveni vektor matrice  $\mathbf{A}$ . Svaka kvadratna matrica  $\mathbf{A} \in \mathbb{R}^{n \times n}$  ima  $n$  (ne nužno različitih) svojstvenih vrijednosti  $\lambda_1, \dots, \lambda_n$ . Skup  $\sigma(\mathbf{A}) = \{\lambda_1, \dots, \lambda_n\}$  naziva se spektar, a broj  $\rho(\mathbf{A}) = \max\{|\lambda_i| : \lambda_i \in \sigma(\mathbf{A})\}$  spektralni radijus matrice  $\mathbf{A}$ . Ako je  $\mathbf{A}$  još i simetrična, onda su sve svojstvene vrijednosti  $\lambda_i$  realne. U tom slučaju može se definirati ortogonalna matrica  $\mathbf{V} \in \mathbb{R}^{n \times n}$  čiji su stupci ortonormirani svojstveni vektori, tako da je  $\mathbf{A} = \mathbf{V} \mathbf{D} \mathbf{V}^T$ , gdje je  $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_n)$ . Rastav  $\mathbf{A} = \mathbf{V} \mathbf{D} \mathbf{V}^T$  nazivamo spektralna dekompozicija matrice  $\mathbf{A}$ . Spomenimo još da vrijedi (vidi primjerice GILL (1991), PARLETT (1980)):

– matrica  $\mathbf{A} \in \mathbb{R}^{n \times n}$  je pozitivno definitna onda i samo onda ako je  $\sigma(\mathbf{A}) \subset \mathbb{R}_+$ ;

–  $\text{Tr}(\mathbf{A}) := \sum_{i=1}^n \mathbf{a}_{ii} = \sum_{i=1}^n \lambda_i$ ;

–  $\prod_{i=1}^n \lambda_i = \det \mathbf{A}$ .

Svojstveni vektori matrice  $\mathbf{A}^T \mathbf{A}$

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

ujedno su desni singularni vektori matrice  $\mathbf{A}$ .

S numeričkog aspekta, ako je  $\gamma^2$  manji od najmanjeg floating-point broja računala, onda bi zbog  $\cos \gamma \approx 1 - \frac{\gamma^2}{2} = 1$ , matrica  $\mathbf{A}^T \mathbf{A}$  imala samo jednu svojstvenu vrijednost  $\lambda_1 = 2$  i samo jednu singularnu vrijednost  $\sigma_1 = \sqrt{2}$  različitu od nule. Manja singularna vrijednost bi se izgubila.

**Zadatak 3.12** Izračunaj lijeve singularne vektore  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$  iz prethodnog primjera. Formiraj matrice  $\mathbf{U}$  i  $\mathbf{V}$  i provjeri jednakost (3.29).

Primjenom SVD možemo riješiti sustav linearnih jednadžbi  $\mathbf{A}\mathbf{x} = \mathbf{b}$ , gdje je  $\mathbf{A} \in \mathbb{R}^{n \times n}$  kvadratna regularna matrica, koja može biti i loše uvjetovana. Korištenjem rastava (3.29) polazni sustav možemo pisati

$$\mathbf{U}\mathbf{S}\mathbf{V}^T \mathbf{x} = \mathbf{b} \implies \mathbf{S}\mathbf{V}^T \mathbf{x} = \mathbf{U}^T \mathbf{b},$$

na osnovi čega, supstitucijom  $\mathbf{z} = \mathbf{V}^T \mathbf{x}$ , polazni sustav transformiramo u dijagonalni

$$\mathbf{S} \mathbf{z} = \mathbf{U}^T \mathbf{b}.$$

SVD-dekompoziciju matrice  $\mathbf{A}$  možemo dobiti korištenjem programa *Mathematica* naredbom `SingularValues[A]` (vidi također BASIC program u *Dodatku*).

**Primjedba 3.9** Broj uvjetovanosti proizvoljne matrice  $\mathbf{A} \in \mathbb{R}^{m \times n}$  ranga  $k < p = \min\{m, n\}$  (vidi GILL (1991), GOLUB (1996)) može se definirati na sljedeći način:

$$\text{cond}(\mathbf{A}) = \|\mathbf{A}\|_2 \|\mathbf{A}^+\|_2 = \frac{\sigma_1}{\sigma_k}, \quad (3.32)$$

gdje su  $\sigma_1 > \sigma_2 > \dots > \sigma_k > 0$  singularne vrijednosti matrice  $\mathbf{A}$ , a  $\mathbf{A}^+ = \mathbf{V}\mathbf{S}^+ \mathbf{U}^T$  pseudoinverzna matrica od  $\mathbf{A}$ , pri čemu je  $\mathbf{S}^+ = \text{diag}(s_1, \dots, s_p)$ , gdje je

$$s_i = \begin{cases} \frac{1}{\sigma_i}, & \sigma_i \neq 0 \\ 0, & \sigma_i = 0 \end{cases} \quad (3.33)$$

**Primjedba 3.10** Više o raznim numeričkim metodama za rješavanje sustava linearnih jednadžbi može se naći u BJÖRCK (1996), DAHLQUIST (1972), GILL (1991), GOLUB (1996), SCHWARZ (1986) itd, a odgovarajuća FORTRAN-programaska podrška može se naći u FORSYTHE (1974), PRESS (1992), te posebno u *The NAG-Library* (PHILIPS (1986), KÖCKLER (1990)). Mogu se također koristiti i gotovi *Mathematica*-programi (WOLFRAM (1998)).

### 3.10 Zadaci

**Zadatak 3.13** Koliko je računskih operacija potrebno da se za zadanu kvadratnu matricu  $\mathbf{A}$  reda  $n$  provede

- a) LU-dekompozicija (Gaussov postupak bez pivotiranja);  
 b) Cholesky-dekompozicija; c) QR-dekompozicija.

**Zadatak 3.14** Nekom od navedenih metoda riješite sustave linearnih jednadžbi, te izračunajte  $\text{cond}(\mathbf{A})$  za navedene sustave.

$$\begin{aligned} \text{a)} \quad & \begin{aligned} x_1 + 3x_2 + 9x_3 + x_4 &= 10 \\ x_1 + x_2 + x_3 + 5x_4 &= 6 \\ 2x_1 + 7x_2 + 3x_3 + 2x_4 &= 15 \\ 4x_1 + 2x_2 + x_3 + x_4 &= 9 \end{aligned} \end{aligned}$$

$$\begin{aligned} \text{b)} \quad & \begin{aligned} 100x_1 - 24x_2 + 48x_3 - 23x_4 &= 39 \\ 5x_1 + 100x_2 - 44x_3 - 31x_4 &= 72 \\ 10x_1 - 3x_2 + 100x_3 + 55x_4 &= 56 \\ -12x_1 + 7x_2 - 11x_3 + 100x_4 &= 47 \end{aligned} \end{aligned}$$

Rješenje: a)  $\mathbf{x} = \frac{1}{7}(9, 10, 3, 4)^T$ ,  $\text{cond}(\mathbf{A}) = 2.68$ .

b)  $\mathbf{x} = (0.615488, 0.955458, 0.249673, 0.504441)^T$ ,  $\text{cond}(\mathbf{A}) = 1.075$ .

**Zadatak 3.15** U 4-znamenkastoj, odnosno 6-znamenkastoj floating-point aritmetici uz primjenu neke od navedenih metoda riješite sustav linearnih jednadžbi  $\mathbf{K}\mathbf{x} = \mathbf{b}$ ,

$$\mathbf{K} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ -1 & 1 & 1 & \cdots & 1 \\ -1 & -1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{n \times n}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

**Zadatak 3.16** Rješenje sustava linearnih jednadžbi, gdje su matrica sustava  $\mathbf{A}$  i vektor  $\mathbf{b}$  zadani s

$$\mathbf{A} = \begin{bmatrix} 0.550 & 0.423 \\ 0.484 & 0.372 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0.127 \\ 0.112 \end{bmatrix},$$

je  $\mathbf{x} = (1, -1)^T$ . Promijenimo vektor  $\mathbf{b}$  za  $\Delta\mathbf{b} = (0.00007, 0.00028)^T$ . Odredite egzaktno rješenje sustava  $\mathbf{A}\mathbf{x} = \mathbf{b} + \Delta\mathbf{b}$  i izračunajte relativne pogreške vektora  $\mathbf{b}$  i  $\mathbf{x}$ . Koliko puta je relativna promjena u rješenju veća od relativne promjene u vektoru  $\mathbf{b}$ ?

Rješenje:  $\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} = \frac{0.00028}{0.127} \approx 2.2 \times 10^{-3}$ ,  $\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} = \frac{0.91}{1} = 0.91$ . Relativna promjena rješenja  $\mathbf{x}$  413 puta je veća od relativne promjene u vektoru  $\mathbf{b}$ .

**Zadatak 3.17** Načinimo malu promjenu u elementu  $a_{21}$  matrice  $\mathbf{A}$  iz Zadatka 3.16, tako da je

$$\mathbf{A} + \Delta\mathbf{A} = \begin{bmatrix} 0.550 & 0.423 \\ 0.483 & 0.372 \end{bmatrix}.$$

Odredite egzaktno rješenje sustava  $(\mathbf{A} + \Delta\mathbf{A})\tilde{\mathbf{x}} = \mathbf{b}$ . Koliko puta je relativna promjena u rješenju  $\mathbf{x}$  veća od relativne promjene u matrici? Koliki je  $\text{cond}(\mathbf{A})$  u  $\|\cdot\|_\infty$ ?

Rješenje: Relativna promjena u rješenju je oko 1000 puta veća od relativne promjene u matrici  $\mathbf{A}$ .  $\text{cond}(\mathbf{A}) = 1039.52$

**Zadatak 3.18** Izradite program koji će učitati red matrice  $n$ , elemente matrice  $\mathbf{A}$ , te primjenom Gaussove metode eliminacije izračunati inverznu matricu  $\mathbf{A}^{-1}$ .

**Zadatak 3.19** Izradite program koji će učitati red matrice  $n$ , elemente matrice  $\mathbf{A}$ , te primjenom LU-dekompozicije izračunati inverznu matricu  $\mathbf{A}^{-1}$ .

**Zadatak 3.20** Izradite program koji će Gaussovom metodom uz strategiju potpunog pivotiranja riješiti sustav  $\mathbf{Ax} = \mathbf{b}$ , te niže navedenim postupkom iterativno popraviti rješenje:

Neka su  $\bar{\mathbf{L}}$  i  $\bar{\mathbf{U}}$  neke aproksimacije matrica  $\mathbf{L}$  i  $\mathbf{U}$ , a  $\bar{\mathbf{x}}$  odgovarajuća aproksimacija rješenja sustava  $\mathbf{Ax} = \mathbf{b}$ . Označimo  $\mathbf{x}^{(1)} = \bar{\mathbf{x}}$  i računamo vektor-rezidual

$$\mathbf{r}^{(1)} = \mathbf{b} - \mathbf{Ax}^{(1)},$$

riješimo sustav

$$\bar{\mathbf{L}}\bar{\mathbf{U}}\delta\mathbf{x}^{(1)} = \mathbf{r}^{(1)},$$

i izračunamo sljedeću aproksimaciju rješenja sustava  $\mathbf{Ax} = \mathbf{b}$  na sljedeći način

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \delta\mathbf{x}^{(1)}.$$

Navedeni postupak možemo ponoviti nekoliko puta, a preporučljivo ga je izvesti u double-precision.

**Zadatak 3.21** Pokažite da je  $\mathbf{x}^{(2)}$  iz prethodnog zadatka točno rješenje sustava  $\mathbf{Ax} = \mathbf{b}$ , ako je  $\bar{\mathbf{L}}\bar{\mathbf{U}} = \mathbf{A}$  i ako su  $\mathbf{r}^{(1)}$  i  $\delta\mathbf{x}^{(1)}$  točno izračunati.

**Zadatak 3.22** Izradite program koji će primjenom LU-dekompozicije riješiti sustav  $\mathbf{Ax} = \mathbf{b}$ , te na osnovi iterativnog postupka navedenog u Zadatku 8 iterativno popraviti rješenje.

**Zadatak 3.23** Neka je  $\mathbf{A} = \mathbf{LU}$ , gdje su  $\mathbf{L}$  i  $\mathbf{U}$  matrice definirane Algoritmom LU. Tada je  $\mathbf{A}^{-1} = (\mathbf{LU})^{-1} = \mathbf{U}^{-1}\mathbf{L}^{-1}$ . Uz oznake:  $\mathbf{Y} = \mathbf{L}^{-1}$  i  $\mathbf{Z} = \mathbf{U}^{-1}$  vrijedi:  $\mathbf{LY} = \mathbf{UZ} = \mathbf{I}$ , odnosno

$$\mathbf{Ly}^j = \mathbf{Uz}^j = \mathbf{e}^j, \quad j = 1, \dots, n.$$

Komponente vektora  $y^j$  možemo dobiti iz Algoritma FS, a komponente vektora  $z^j$  iz Algoritma BS. Pokažite da je tada

$$y_{ij} = 0, \quad i < j, \quad z_{ij} = 0, \quad i > j,$$

tj. da je  $\mathbf{L}^{-1}$  donja trokutasta, a  $\mathbf{U}^{-1}$  gornja trokutasta matrica.





Izradite kompjuterski program koji će učitati red matrice  $n$ , vektore  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$ ,  $\mathbf{d}$ ,  $\mathbf{e}$ , te vektor slobodnih koeficijenata  $\mathbf{y}$  i riješiti sustav nekom od ranije spomenutih metoda.

**Zadatak 3.26** Hilbertova matrica  $\mathbf{H}$  s elementima  $h_{ij} = \frac{1}{i+j-1}$  je klasični primjer loše uvjetovane matrice. Ispitajte neke komercijalne, kao i priložene programe, rješavajući sustav jednažbi, tako da za matricu sustava izaberete Hilbertove matrice različitih dimenzija. Pri tome izračunavanja obavljajte u  $k$ -znamenkastoj ( $k = 2, \dots, 10$ ) floating-point aritmetici.

**Zadatak 3.27** Pokažite da je

$$\mathbf{A} = \begin{bmatrix} 4 & 3 & -1 \\ -2 & -4 & 5 \\ 1 & 2 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ -25 & -5 & 1 \end{bmatrix} \cdot \begin{bmatrix} 4 & 3 & -1 \\ 0 & -2.5 & 4.5 \\ 0 & 0 & 8.5 \end{bmatrix}.$$

Primjenom LU-dekompozicije riješite sustav  $\mathbf{Ax} = \mathbf{b}$ , gdje je  $\mathbf{b} = (-2, 20, 7)^T$ .

Rješenje:  $\mathbf{x} = (3, -4, 2)^T$ .

**Zadatak 3.28** Zadan je sustav linearnih jednažbi

$$\begin{bmatrix} 34 & 55 \\ 55 & 89 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 21 \\ 34 \end{bmatrix}$$

Koja je aproksimacija rješenja:  $\mathbf{x}_A^* = (-0.11, 0.45)^T$  ili  $\mathbf{x}_B^* = (-0.99, 1.01)^T$  bolja u  $\mathbf{L}_1$ ,  $\mathbf{L}_2$  i  $\mathbf{L}_\infty$  normi? Kako se ovo može primijeniti za komparativnu analizu kvalitete računala (vidi MATHEWS (1992)).

**Zadatak 3.29** Cholesky dekompozicija također se može provesti i po retcima sljedećim algoritmom

Za  $k = 1, \dots, n$

$$l_{kj} = \frac{1}{l_{jj}} \left( a_{kj} - \sum_{p=1}^{j-1} l_{kp} l_{jp} \right), \quad l_{kk} = \left( a_{kk} - \sum_{p=1}^{k-1} l_{kp}^2 \right)^{1/2}, \quad j = 1, \dots, k-1.$$

(pri čemu navedene sume nećemo izvoditi ako je gornja granica sume manja od donje granice). Izradite odgovarajući Mathematica-program i ispitajte ga na prethodnim primjerima.

## 4. Rješavanje nelinearnih jednačbi

Promatramo realnu neprekidnu funkciju  $f$  definiranu na zatvorenom intervalu  $[a, b]$ . Općenito, svaki kompleksni broj  $\xi$ , koji je rješenje jednačbe

$$f(x) = 0, \quad (4.1)$$

nazivamo **nultočkom** funkcije  $f$ . Mi ćemo se ograničiti na istraživanje samo realnih nultočaka funkcije  $f$  (sjecišta grafa funkcije s osi  $x$ ). Općenito se može dogoditi da neka funkcija ima više realnih nultočaka, da su neke višestruke ili da uopće nema realnih nultočaka.

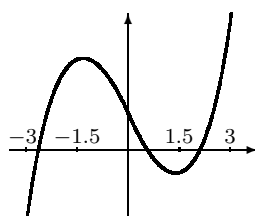
Ako je funkcija  $f$  neprekidna na intervalu  $I = [a, b]$  i ako na rubovima intervala prima suprotne vrijednosti (tj. ako je  $f(a) \cdot f(b) < 0$ ), onda (vidi JUKIĆ (1999), str. 118), postoji barem jedna točka  $\xi \in I$ , za koju vrijedi  $f(\xi) = 0$ . Ako je, osim toga, prva derivacija  $f'$  stalnog predznaka na intervalu  $I$ , onda je to i jedina nultočka funkcije  $f$  na intervalu  $I$ . Na taj se način posao oko traženja realnog rješenja jednačbe (4.1) svodi na dva koraka:

1. Separirati interval  $I$ , u kome funkcija ima nultočku,
2. Nekom iterativnom metodom odrediti aproksimaciju nultočke  $\xi$  s unaprijed zadanom točnošću.

Spomenuti interval  $I = [a, b]$ , u kome se nalazi barem jedna nultočka funkcije  $f$ , treba odrediti tako da na njegovim rubovima funkcija prima vrijednosti suprotnog predznaka, tj. da bude

$$f(a) \cdot f(b) < 0. \quad (4.2)$$

**Primjer 4.1** Treba separirati realne nultočke funkcije  $f(x) = x^3 - 6x + 2$ .



Slika 4.1  $f(x) = x^3 - 6x + 2$

Graf ove funkcije lako je nacrtati (vidi *Sliku 4.1*). Sada je dovoljno samo procijeniti predznak vrijednosti funkcije u nekim točkama.

$x$	-3	-1.5	0	1.5	3
$\operatorname{sgn} f(x)$	-1	+1	+1	-1	+1

Na taj način separirali smo tri intervala:  $I_1 = [-3, -1.5]$ ,  $I_2 = [0, 1.5]$  i  $I_3 = [1.5, 3]$ . U svakom od njih nalazi se po jedna nultočka funkcije. Budući da se u ovom slučaju radi o polinomu trećeg stupnja, ova funkcija nema drugih nultočaka.

Prilikom separiranja nultočaka funkcije možemo se koristiti i grafičkim metodama. Pri tome je naročito pogodno koristiti grafičke mogućnosti računala (programi: MATHEMATICA, DERIVE, MATLAB). Ako jednadžbu  $f(x) = 0$  preuredimo tako, da na lijevu i desnu stranu dođu funkcije čije grafove možemo lako nacrtati

$$\varphi(x) = \psi(x),$$

tada apscise točaka sjecišta grafova funkcija  $\varphi$  i  $\psi$  predstavljaju nultočke funkcije  $f$ . Na ovaj način nećemo dobiti dovoljno dobre aproksimacije nultočaka, ali tako smo u mogućnosti izolirati intervale u kojima leže nultočke.

**Primjer 4.2** Treba separirati realne nultočke funkcije

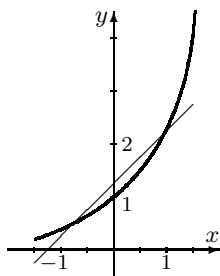
$$f(x) = e^x - x - 5/4.$$

Jednadžbu  $f(x) = 0$  napisat ćemo u obliku:  $e^x = x + 5/4$ , te nacrtati grafove funkcija  $\varphi(x) = e^x$  i  $\psi(x) = x + 5/4$  (*Slika 4.2.a*). Iz slike se vidi da funkcija  $f$  ima dvije nultočke, koje se nalaze u intervalima  $I_1 = [-1, 0]$  i  $I_2 = [0, 1]$ , što potvrđuje i sljedeća tablica

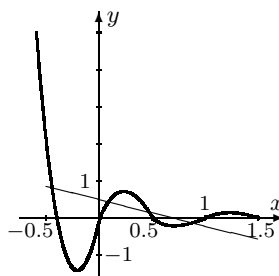
$x$	-1	0	1
$\operatorname{sgn} f(x)$	+1	-1	+1

**Primjer 4.3** Treba separirati sve pozitivne nultočke funkcije

$$f(x) = e^{-2x} \sin 6x + \frac{2}{3}x - \frac{1}{2}.$$



Slika 4.2.a  $e^x = x + 5/4$



Slika 4.2.b  $e^{-2x} \sin 6x = \frac{1}{2} - \frac{2}{3}x$

Jednadžbu  $f(x) = 0$  napisat ćemo u obliku:  $e^{-2x} \sin 6x = \frac{1}{2} - \frac{2}{3}x$ , te nacrtati grafove funkcija  $\varphi(x) = e^{-2x} \sin 6x$  i  $\psi(x) = \frac{1}{2} - \frac{2}{3}x$  (*Slika 4.2.b*). Iz slike se vidi da funkcija  $f$  ima tri pozitivne nultočke, koje treba tražiti u intervalima  $I_1 = [0, 0.25]$ ,  $I_2 = [0.25, 0.5]$  i  $I_3 = [0.5, 1]$ , što potvrđuje i sljedeća tablica

$x$	0	0.25	0.5	1
$\operatorname{sgn} f(x)$	-1	+1	-1	+1

Za rješavanje jednadžbe  $f(x) = 0$  postoji čitav niz različitih metoda, a mi ćemo se upoznati s nekim od njih, koje se najčešće koriste prilikom rješavanja praktičnih problema. Pretpostavit ćemo da je promatrana funkcija  $f$  na intervalu  $I = [a, b]$  dovoljno “glatka”, tj. da je klase  $C_{[a,b]}^2$ . Metode za rješavanje jednadžbe  $f(x) = 0$  uglavnom su iterativne metode u kojima se nekom rekursivnom formulom definira niz brojeva (aproksimacija):  $x_0, x_1, \dots, x_n, \dots$ , koji uz neke uvjete može konvergirati rješenju jednadžbe.

Ako je  $x_n$  jedna aproksimacija nultočke  $\xi$  u intervalu  $I = [a, b]$  i ako je  $f$  derivabilna funkcija, takva da je  $|f'(x)| > 0, x \in I$ , onda vrijedi ovakva ocjena apsolutne pogreške

$$|\xi - x_n| \leq \frac{|f(x_n)|}{m_1}, \quad \text{gdje je } 0 < m_1 = \min_{x \in I} |f'(x)|. \quad (4.3)$$

Naime, prema teoremu o srednjoj vrijednosti, imamo

$$f(\xi) - f(x_n) = (\xi - x_n)f'(c), \quad c \in I.$$

Oдавde je

$$|f(x_n)| = |\xi - x_n| \cdot |f'(c)| \geq m_1 \cdot |\xi - x_n|,$$

odakle zbog stroge monotonosti funkcije  $f$  slijedi ocjena (4.3).

Budući da je rješenje  $\xi$  jednadžbe  $f(x) = 0$  obično granična vrijednost jednog beskonačnog niza, postavlja se pitanje kada zaustaviti iterativni proces.

Ako želimo da apsolutna pogreška aproksimacije  $x_n$  nultočke  $\xi$  ne bude veća od  $\epsilon > 0$ , onda je prema (4.3) dovoljno ispuniti uvjet  $|f(x_n)|/m_1 < \epsilon$ . Ovo može biti jedan kriterij za zaustavljanje iterativnog procesa.

**Primjedba 4.1** *Moglo bi se pomisliti da je prirodan kriterij za zaustavljanje iterativnog procesa ispunjenje nejednakosti:  $|f(x_n)| < \epsilon$ , za neki unaprijed zadani  $\epsilon > 0$ . Međutim, lako je iskonstruirati primjer u kome će za neki  $n$  biti  $|f(x_n)| < \epsilon$ , a da ipak pogreška aproksimacije  $|\xi - x_n|$  bude značajna, tj. da  $x_n$  na brojevnom pravcu bude daleko od  $\xi$  (vidi Sliku 4.3.a). Naime, u tom slučaju je broj  $m_1$  malen, pa je  $\frac{1}{m_1}$  velik, što znači da će biti teško ispuniti zahtjev  $\frac{|f(x_n)|}{m_1} < \epsilon$ , koji će prema (4.3) osigurati da pogreška aproksimacije  $x_n$  ne premaši  $\epsilon$ .*

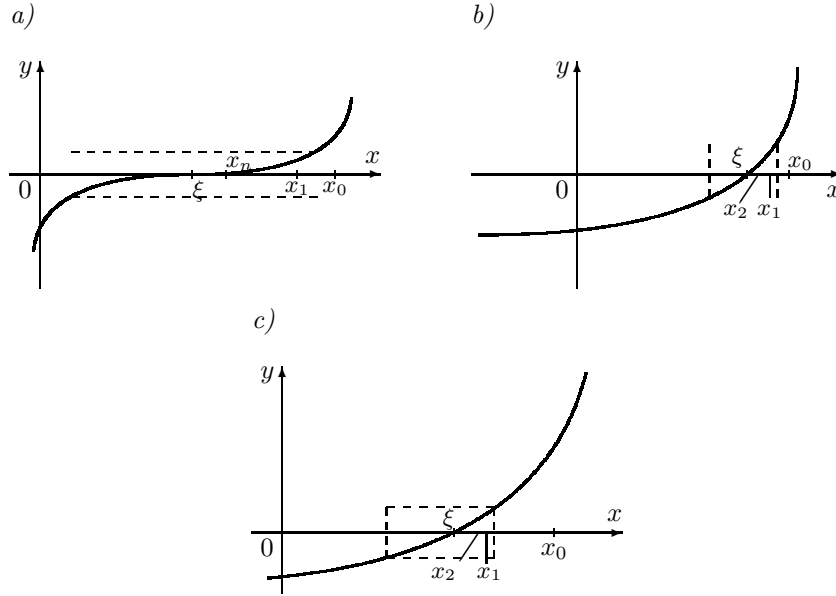
*Ako je  $m_1$  velik,  $\frac{1}{m_1}$  je malen, pa prema (4.3) pogreška aproksimacije može biti malena ( $x_n$  je blizu rješenja  $\xi$ ), a da vrijednost funkcije  $f(x_n)$  bude relativno velika (Slika 4.3.b).*

*Ako se dogodi da je  $m_1 = 0$ , metodom bisekcije (vidi t.4.1) treba suziti interval, tako da bude  $m_1 > 0$ .*

Budući da je često puta u praktičnim primjerima teško odrediti broj  $m_1$ , umjesto kriterija za zaustavljanje procesa  $\frac{|f(x_n)|}{m_1} < \epsilon$ , često se koristi (vidi primjerice MATHEWS, 1992)

$$|x_n - x_{n-1}| < \epsilon_1 \quad \& \quad |f(x_n)| < \epsilon_2, \quad (4.4)$$

gdje su  $\epsilon_1, \epsilon_2$  unaprijed zadani brojevi.



Slika 4.3. Kriteriji zaustavljanja iterativnog procesa

Za neku iterativnu metodu važno je poznavati brzinu (red) konvergencije metode. U tu svrhu uvodimo sljedeću definiciju

**Definicija 4.1** (Red konvergencije metode) *Neka niz  $(x_n)$ , dobiven nekom iterativnom metodom, konvergira prema  $\xi \in \mathbb{R}$  i neka je  $e_n = \xi - x_n$  pogreška  $n$ -te aproksimacije. Tada, ako postoje dvije pozitivne konstante  $A, r \in \mathbb{R}_+$ , takve da vrijedi*

$$\lim_{n \rightarrow \infty} \frac{|e_{n+1}|}{|e_n|^r} = A, \quad (4.5)$$

*kažemo da metoda ima red konvergencije  $r$ .*

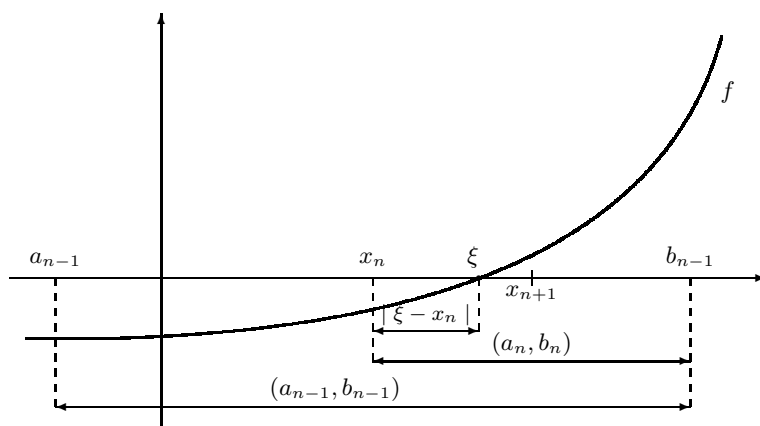
Specijalno, ako je  $r = 1$ , kažemo da metoda ima linearnu, a ako je  $r = 2$ , da ima kvadratnu brzinu konvergencije. Iz navedene definicije jasno je da će za veći  $r$  niz  $(x_n)$  brže konvergirati prema  $\xi$ .

Primijetimo također da iz Definicije 4.1 slijedi da metoda ima red konvergencije  $r$  ako postoji  $A \in \mathbb{R}_+$  i  $n_0 \in \mathbb{N}$ , tako da bude

$$|\xi - x_{n+1}| \leq A|\xi - x_n|^r, \quad n \geq n_0.$$

## 4.1 Metoda bisekcije

Pretpostavimo da je funkcija  $f$  neprekidna na intervalu  $I = [a, b]$  i da je  $f(a) \cdot f(b) < 0$ . Ako interval  $I$  raspolovimo i ako polovište nije nultočka funkcije  $f$ , dalje promatramo onu polovinu intervala  $I$  na čijim rubovima funkcija  $f$  prima vrijednosti suprotnog predznaka. Navedeni postupak dalje ponavljamo.



Slika 4.4. Metoda bisekcije

Formalno, označimo  $a_0 := a$ ,  $b_0 := b$  i navedenim postupkom definirajmo silazni niz segmenata:  $[a_1, b_1] \supset [a_2, b_2] \supset [a_3, b_3] \supset \dots$ , koji sadrži korijen jednadžbe  $f(x) = 0$ . Na taj način smo u mogućnosti definirati niz aproksimacija  $x_0, x_1, \dots, x_n, \dots$  (vidi Sliku 4.4)

$$x_n = \frac{1}{2}(a_{n-1} + b_{n-1}),$$

koji konvergira prema rješenju jednadžbe  $f(x) = 0$  linearnom brzinom jer je

$$|\xi - x_{n+1}| \leq \frac{1}{2}|\xi - x_n|,$$

pri čemu pogrešku aproksimacije  $x_n$  možemo ocijeniti (vidi Sliku 4.4) s

$$|\xi - x_n| \leq \frac{1}{2}(b_{n-1} - a_{n-1}).$$

### Algoritam BISEKCIJE

Korak 0. Učitati  $a$ ,  $b$ ,  $\epsilon$ . Ako je  $f(a) \cdot f(b) > 0$ , zaustaviti proces; u protivnom prijeći na Korak 1.

Korak 1. Staviti  $a_0 := a$ ;  $b_0 := b$  i izračunati  $x_1 = \frac{a_0 + b_0}{2}$ .

Korak 2. Ako je  $f(a_0) \cdot f(x_1) < 0$ , staviti  $b_0 := x_1$ ; u protivnom staviti  $a_0 = x_1$ .

Korak 3. Ako je  $b_0 - a_0 > 2\epsilon$ , izračunati  $x_1 = \frac{a_0 + b_0}{2}$  i prijeći na Korak 2.; u protivnom zaustaviti proces.

**Zadatak 4.1** Izradi program za računalo koji će izvoditi algoritam BISEKCIJE.

**Primjer 4.4** S točnošću  $\epsilon = 0.05$  korištenjem algoritma bisekcije izračunat ćemo nultočku funkcije  $f(x) = x^3 - 6x + 2$  na intervalu  $[0, 1.5]$  (vidi Primjer 4.1 str. 67).

Nakon 6 iteracija dobivamo  $x^* = 0.3$ . Rezultat izvođenja iterativnog postupka vidljiv je u Tablici 4.1.

$n$	$a_n$	$b_n$	$x_{n+1}$	$\text{sgn}f(x_{n+1})$	$(b_{n-1} + a_{n-1})/2$
0	0	1.5	0.75	-1	0.75
1	0	0.75	0.375	-1	0.375
2	0	0.375	0.1875	1	0.1875
3	0.1875	0.375	0.28125	1	0.09375
4	0.28125	0.375	0.328125	1	0.046875
5	0.328125	0.375	0.351563	-1	0.02344
6	0.328125	0.351563	0.339844	1	0.0117

Tablica 4.1

**Primjer 4.5**<sup>1</sup> Kao ilustraciju Primjedbe 4.1, str. 69 promatrajmo problem rješavanja sljedećih jednadžbi metodom bisekcije:

$$(a) \quad f(x) = 0.001x + \frac{1}{2} + \frac{\pi}{2} + \arctg x = 0, \quad x \in [-600, -400]$$

$$(b) \quad g(x) = 1000(x - 4) - e^x = 0, \quad x \in [4, 5].$$

a) Graf funkcije  $f$  na zadanom intervalu približno odgovara grafu prikazanom na Slici 4.3.a i ima nultočku  $\xi = -501.992061$  točnu na 6 decimala. Uz  $\epsilon = .005$  (odnosno  $\epsilon = 0.00005$ ) i kriterij zaustavljanja:  $|f(x_n)| < \epsilon$  nakon 5 (odnosno 9) iteracija dobivamo aproksimaciju  $x_5 = -506.25$  (odnosno  $x_9 = -501.9531$ ). Za aproksimaciju točnu na 2 (odnosno 4) decimala bilo bi potrebno 12 (odnosno 18) iteracija.

b) Graf funkcije  $g$  na zadanom intervalu približno odgovara grafu prikazanom na Slici 4.3.b i ima nultočku  $\xi = 4.057850$  točnu na 6 decimala. Uz  $\epsilon = .005$  (odnosno  $\epsilon = 0.00005$ ) i kriterij zaustavljanja:  $|f(x_n)| < \epsilon$  nakon 16 (odnosno 18) iteracija dobivamo puno točniju aproksimaciju nego što je tražena:  $x_{16} = 4.057846$  (odnosno  $x_{18} = 4.057850$ ). Za aproksimaciju točnu na 2 (odnosno 4) decimala potrebno je samo 4 (odnosno 12) iteracija.

## 4.2 Metoda jednostavnih iteracija

Jednadžbu  $f(x) = 0$  možemo napisati u obliku

$$x = \varphi(x), \quad (4.6)$$

gdje za funkciju  $\varphi$  obično postoji više izbora. Primjerice, pripadnu jednadžbu iz Primjera 4.2, str. 68 za  $x > -\frac{5}{4}$  možemo pisati u obliku

$$x = e^x - 5/4 \quad \text{ili} \quad x = \ln(x + 5/4).$$

<sup>1</sup>Ovaj primjer konstruirao je J. Muharemović, student III godine ETF-a u Osijeku, 1998.





konvergira prema  $\xi$  i vrijede ovakve ocjene pogreške aproksimacije

$$|\xi - x_n| \leq \frac{q^n}{1-q} |x_1 - x_0|, \quad (4.9)$$

$$|\xi - x_n| \leq \frac{q}{1-q} |x_n - x_{n-1}|. \quad (4.10)$$

Metoda ima linearnu brzinu konvergencije, tj. vrijedi

$$|\xi - x_{n+1}| \leq q |\xi - x_n|. \quad (4.11)$$

Dokaz. Kako je  $x_n = \varphi(x_{n-1})$  i  $x_{n+1} = \varphi(x_n)$ , koristeći Lagrangeov teorem o srednjoj vrijednosti, imamo

$$x_{n+1} - x_n = \varphi(x_n) - \varphi(x_{n-1}) = (x_n - x_{n-1})\varphi'(c), \quad c \in (a, b).$$

Zato je

$$|x_{n+1} - x_n| \leq q |x_n - x_{n-1}| \quad \text{za svaki } n = 1, 2, \dots \quad (4.12)$$

Sukcesivnom primjenom nejednakosti (4.12) dobivamo

$$|x_{n+1} - x_n| \leq q |x_n - x_{n-1}| \leq \dots \leq q^n |x_1 - x_0|. \quad (4.13)$$

Promatrajmo sada ovakav red

$$x_0 + (x_1 - x_0) + (x_2 - x_1) + \dots + (x_n - x_{n-1}) + \dots \quad (4.14)$$

Njegova  $n$ -ta parcijalna suma je  $s_n = x_n$ . Zbog (4.13) jedna majoranta ovog reda je red

$$|x_0| + |x_1 - x_0| + q |x_1 - x_0| + \dots + q^{n-1} |x_1 - x_0| + \dots \quad (4.15)$$

kojeg možemo zapisati ovako

$$|x_0| + |x_1 - x_0| (1 + q + \dots + q^{n-1} + \dots).$$

Kako je  $q < 1$ , geometrijski red u zagradi je konvergentan, pa je i red (4.15) konvergentan. Na taj način pronašli smo jednu konvergentnu majorantu reda (4.14), što znači da je i taj red konvergentan. Po definiciji konvergencije reda, to znači da je i odgovarajući niz  $n$ -tih parcijalnih suma ( $s_n$ ) reda konvergentan. Kako je  $s_n = x_n$ , to znači da je i niz ( $x_n$ ) konvergentan, tj. postoji realan broj  $\xi$ , takav da je  $\xi = \lim x_n$ . Zbog uvjeta (i) čitav niz ( $x_n$ ) je sadržan u  $I$ , a kako je  $I$  zatvoren skup, također je i  $\xi \in I$ .

Ako sada u (4.8) pustimo  $n \rightarrow \infty$ , onda zbog neprekidnosti funkcije  $\varphi$  imamo  $\xi = \varphi(\xi)$ , što znači da je  $\xi$  korijen jednadžbe (4.6).

Dokažimo da je  $\xi$  jedini korijen jednadžbe (4.6) u intervalu  $I$ . Pretpostavimo da je  $\bar{\xi}$  također korijen jednadžbe (4.6), tj. pretpostavimo da vrijedi  $\bar{\xi} = \varphi(\bar{\xi})$ . Tada prema teoremu o srednjoj vrijednosti mora biti

$$\bar{\xi} - \xi = \varphi(\bar{\xi}) - \varphi(\xi) = (\bar{\xi} - \xi)\varphi'(c), \quad c \in (a, b),$$

odakle je

$$(\bar{\xi} - \xi)(1 - \varphi'(c)) = 0.$$

Budući da je  $1 - \varphi'(c) \neq 0$  zbog uvjeta (ii), mora biti  $\bar{\xi} = \xi$ . Dakle,  $\xi$  je jedino rješenje jednadžbe (4.6).

U svrhu dokaza ocjene (4.9) promatramo

$$\begin{aligned} |x_{n+p} - x_n| &= |(x_{n+p} - x_{n+p-1}) + (x_{n+p-1} - x_{n+p-2}) + \cdots + (x_{n+1} - x_n)| \leq \\ &\leq |x_{n+p} - x_{n+p-1}| + |x_{n+p-1} - x_{n+p-2}| + \cdots + |x_{n+1} - x_n| \leq \\ (\text{prema (4.12)}) &\leq q^{n+p-1}|x_1 - x_0| + q^{n+p-2}|x_1 - x_0| + \cdots + q^n|x_1 - x_0| = \\ &= q^n|x_1 - x_0|(1 + q + \cdots + q^{p-1}) = q^n|x_1 - x_0|\frac{1 - q^p}{1 - q} < \frac{q^n}{1 - q}|x_1 - x_0|. \end{aligned}$$

Dakle vrijedi

$$|x_{n+p} - x_n| < \frac{q^n}{1 - q}|x_1 - x_0|.$$

Ako sada pustimo  $p \rightarrow \infty$ , dobivamo nejednakost (4.9).

U svrhu dokaza ocjene (4.10) definirajmo funkciju  $f(x) = x - \varphi(x)$ . Zbog uvjeta (ii) je  $f'(x) = 1 - \varphi'(x) \geq 1 - q > 0$ . Zato uz korištenje teorema o srednjoj vrijednosti dobivamo

$$|x_n - \varphi(x_n)| = |f(x_n)| = |f(x_n) - f(\xi)| = |(x_n - \xi)f'(c)| \geq (1 - q)|x_n - \xi|,$$

odakle, ponovno koristeći teorem o srednjoj vrijednosti dobivamo traženu ocjenu

$$|x_n - \xi| \leq \frac{|x_n - \varphi(x_n)|}{1 - q} = \frac{|\varphi(x_{n-1}) - \varphi(x_n)|}{1 - q} \leq \frac{q|x_n - x_{n-1}|}{1 - q}.$$

Koristeći  $x_{n+1} = \varphi(x_n)$  i  $\xi = \varphi(\xi)$  i teorem o srednjoj vrijednosti, dobivamo

$$|\xi - x_{n+1}| = |\varphi(\xi) - \varphi(x_n)| = |(\xi - x_n)\varphi'(c)| \leq q|\xi - x_n|.$$

odakle neposredno slijedi (4.11). □

**Primjedba 4.2** Ako je unaprijed zadana točnost  $\epsilon$  s kojom želimo dobiti aproksimaciju rješenja jednadžbe (4.6), dovoljno je desnu stranu nejednakosti (4.9) ili desnu stranu nejednakosti (4.10) učiniti manjom od  $\epsilon$ . Na taj način osigurali smo da apsolutna pogreška aproksimacije ne bude veća od  $\epsilon$ .

**Primjedba 4.3** Uvjet (ii) znači da je apsolutna vrijednost koeficijenta smjera tangente na krivulju  $\varphi$  u svakoj točki intervala  $I$  manja od 1, što i osigurava konvergenciju procesa.

**Primjer 4.6** S točnošću  $\epsilon = 0.005$  (na dvije decimale) treba pronaći pozitivnu nultočku funkcije iz Primjera 4.2. Pokazano je da se nultočka nalazi u intervalu  $I = [0, 1]$ .

Ako odgovarajuću jednadžbu napišemo u obliku

$$x = e^x - 5/4,$$

funkcija  $\varphi$  bila bi  $\varphi(x) = e^x - 5/4$ . Kako je  $\varphi'(x) = e^x > 1$  za  $x \in I$ , uvjeti *Teorema 4.1* nisu ispunjeni. Zato ćemo jednadžbu napisati u obliku

$$x = \ln(x + 5/4), \quad x > -\frac{5}{4}.$$

Sada je  $\varphi(x) = \ln(x + 5/4)$ , a  $\varphi'(x) = 4/(4x + 5)$ . Lako se vidi (nacrtaj graf funkcije  $\varphi'$ ) da je

$$\varphi'(x) \leq 4/5 < 1 \quad \text{za svaki } x \in I \quad \text{i} \quad \varphi([0, 1]) = \left(\ln \frac{5}{4}, \ln \frac{9}{4}\right) \subset [0, 1].$$

To znači da su uvjeti *Teorema 4.1* ispunjeni. Cijeli iterativni proces prikazan je u *Tablici 4.2*. U trećem, odnosno četvrtom stupcu prikazane su vrijednosti desne strane ocjene (4.9); odnosno ocjene (4.10).

$n$	$x_n$	(4.9)	(4.10)	$f(x_n)$
0	1	—	—	0.468282
1	0.810930	0.756280	0.756280	0.189070
2	0.723157	0.605024	0.351091	0.087773
3	0.679635	0.484019	0.174090	0.043522
4	0.657331	0.387215	0.089217	0.022304
5	0.645705	0.309772	0.046504	0.011626
6	0.639591	0.247818	0.024457	0.006114
7	0.636360	0.198254	0.012922	0.003230
8	0.634649	0.158603	0.006844	0.001711
9	0.633742	0.126883	0.003630	0.000907

Tablica 4.2

Vidi se da je ocjena (4.9) mnogo grublja od ocjene (4.10). Dakle, pouzdano možemo reći da apsolutna pogreška aproksimacije  $x^* = 0.63$  tražene nultočke nije veća od 0.005.

**Primjer 4.7** *S točnošću  $\epsilon = 0.005$  (na dvije decimale) treba odrediti nultočku funkcije*

$$f(x) = x - 2/x.$$

Kako je  $f(1) = -1$  i  $f(2) = 1$ , interval u kome leži nultočka je  $I = [1, 2]$ . Ako pripadnu jednadžbu napišemo u obliku  $x = 2/x := \varphi(x)$ , proces neće konvergirati jer je  $\varphi'(x) = -2/x^2$  i  $|\varphi'(x)| > 1$  na intervalu  $I$  (primjerice  $\varphi'(1) = -2$ ). Zbog toga ćemo funkciju  $\varphi$  morati drugačije definirati. Jednadžbu  $x = 2/x$  zapisat ćemo u obliku

$$x = \frac{1}{2}\left(x + \frac{2}{x}\right).$$

Označimo  $\varphi(x) := \frac{1}{2}\left(x + \frac{2}{x}\right)$ . Lako se vidi da je derivacija funkcije  $\varphi$  na intervalu  $I$  monotono rastuća funkcija. Zato je  $q = \max\{|\varphi'(1)|, |\varphi'(2)|\} = \max\{1/2, 1/4\} = 1/2$ . Osim toga je  $\varphi([1, 2]) = [\sqrt{2}, 1.5] \subset [1, 2]$ . Time su uvjeti *Teorema 4.1* ispunjeni. Tijek odgovarajućeg iterativnog procesa prikazan je u *Tablici 4.3*. Inače, lako se vidi da je jedina nultočka promatrane funkcije broj  $\sqrt{2}$ . Zato ćemo u tablici pratiti i stvarnu apsolutnu pogrešku aproksimacija  $|\sqrt{2} - x_n|$ .

$n$	$x_n$	$ \sqrt{2} - x_n $	(4.9)	(4.10)	$f(x_n)/m_1$
0	1	0.414214	—	—	—1
1	1.500000	0.085786	0.500000	0.500000	0.166667
2	1.416667	0.002453	0.255000	0.083333	0.004902
3	1.414216	0.000002	0.125000	0.002451	0.000004

Tablica 4.3.

Na ovom primjeru vidi se da iako ocjena (4.10) u drugoj iteraciji ne osigurava ni jednu točnu decimalu, ipak već imamo jednu signifikantnu decimalu. U trećoj iteraciji ocjena (4.10) osigurava dvije signifikantne decimale, iako stvarna apsolutna pogreška pokazuje da smo postigli čak pet signifikantnih znamenki, od čega četiri decimalna mjesta.

Najbolju ocjenu pogreške u ovom primjeru dobili bismo korištenjem ocjene (4.3), gdje je  $m_1 = \min |f'(x)| > 1.5$ . Provjerite!

U praksi tražena nultočka obično nije poznata i zbog toga je važno da ocjena pogreške ne bude manja od stvarne pogreške, ali da ne bude previše gruba.

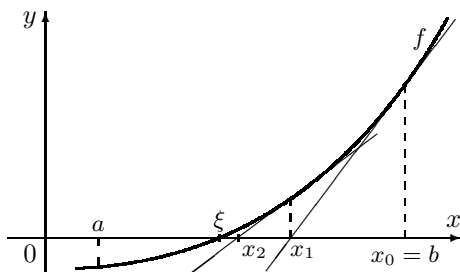
**Primjedba 4.4** *Ideja metode iteracija može se koristiti i u drugim situacijama. Tako primjerice na sličan način možemo pristupiti i rješavanju sustava linearnih (t.3.8, str. 57) ili nelinearnih (t.4.4, str. 85) jednadžbi.*

## 4.3 Newtonova metoda

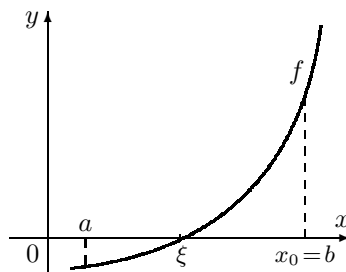
Pretpostavimo da smo na neki način odredili interval  $I = [a, b]$ , u kome se nalazi jedinstvena nultočka neprekidne i dovoljno “glatke” funkcije  $f$  za koju je  $f'(x) \neq 0 \forall x \in I$ . Izaberimo početnu aproksimaciju  $x_0 \in I$ , razvijmo funkciju  $f$  u Taylorov red u okolini točke  $x_0$  i zadržimo se na linearnom članu. Tako smo funkciju  $f$  u okolini točke  $x_0$  aproksimirali linearnom funkcijom

$$f_1(x) = f(x_0) + (x - x_0)f'(x_0) \quad (4.16)$$

čiji graf je tangenta na graf funkcije  $f$  u točki  $x_0$  (Slika 4.6.a).



Slika 4.6.a Newtonova metoda

Slika 4.6.b Izbor početne aproksimacije  $x_0$

Sada ćemo umjesto rješavanja jednadžbe (4.1), rješavati jednadžbu  $f_1(x) = 0$  (geometrijski gledano, umjesto traženja sjecišta grafa funkcije  $f$  s osi  $x$ , tražimo sjecište tangente s osi  $x$ ). Rješenje jednadžbe  $f_1(x) = 0$  označimo s  $x_1$

$$x_1 := x_0 - \frac{f(x_0)}{f'(x_0)}. \quad (4.17)$$

Ponavljajući postupak, dobivamo niz  $x_0, x_1, \dots$  zadan rekursivnom formulom<sup>3</sup>

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, \dots \quad (4.18)$$

Važno je primijetiti da u svakoj iteraciji konstruiramo lokalni aproksimant funkcije  $f$  i onda tražimo nultočku lokalnog aproksimanta.

**Teorem 4.2** *Neka funkcija  $f : I \rightarrow \mathbb{R}$  ima neprekidnu drugu derivaciju na intervalu  $I = [a, b]$ . Neka je nadalje,  $f(a) \cdot f(b) < 0$ , a prva ( $f'$ ) i druga ( $f''$ ) derivacija funkcije  $f$  na intervalu  $I$  imaju stalan predznak.*

*Tada, ako je  $x_0 \in I$  izabran tako da bude*

$$f(x_0) \cdot f''(x_0) > 0, \quad (4.19)$$

*niz definiran s (4.18) konvergira prema jedinstvenom rješenju  $\xi$  jednadžbe  $f(x) = 0$ .*

*Pri tome vrijedi ocjena pogreške aproksimacije*

$$|\xi - x_n| \leq \frac{M_2}{2m_1}(x_n - x_{n-1})^2, \quad (4.20)$$

*gdje je*

$$m_1 = \min_{x \in I} |f'(x)|, \quad M_2 = \max_{x \in I} |f''(x)|, \quad (4.21)$$

*a metoda ima kvadratnu brzinu konvergencije, tj. vrijedi*

$$|\xi - x_{n+1}| \leq \frac{M_2}{2m_1}(\xi - x_n)^2. \quad (4.22)$$

Dokaz. Pogledajmo slučaj<sup>4</sup>

$$f(a) < 0, \quad f(b) > 0, \quad f'(x) > 0, \quad f''(x) > 0 \quad \text{za svaki } x \in I.$$

Izaberimo  $x_0 \in I$  tako da bude ispunjeno (4.19) (primjerice možemo uzeti  $x_0 = b$ ). Kako je po pretpostavci  $f''(x_0) > 0$ , onda je i  $f(x_0) > 0$ , pa je  $x_0 > \xi$  (vidi Sliku 4.6.b) jer je  $f$  rastuća funkcija.

Primjenom matematičke indukcije pokazat ćemo da je niz definiran s (4.18) monotono padajući i ograničen odozdo s  $\xi$ .

<sup>3</sup>U literaturi često ovu metodu možemo naći pod imenom *Newton-Raphsonova metoda*.

<sup>4</sup>Ostali slučajevi dokazuju se analogno

- Pretpostavimo da je  $x_n > \xi$ . Prema Taylorovoj formuli je

$$f(x) = f(x_n) + f'(x_n) \cdot (x - x_n) + \frac{1}{2} \cdot f''(c) \cdot (x - x_n)^2, \quad c \in (a, b).$$

Specijalno, za  $x = \xi$  imamo

$$0 = f(\xi) = f(x_n) + f'(x_n) \cdot (\xi - x_n) + \frac{1}{2} \cdot f''(c) \cdot (\xi - x_n)^2, \quad c \in (a, b).$$

Kako je  $f''(c) > 0$  i  $(\xi - x_n) \neq 0$  (zbog  $x_n > \xi$ ), onda je

$$f(x_n) + f'(x_n) \cdot (\xi - x_n) < 0.$$

Oдавде je

$$\xi < x_n - \frac{f(x_n)}{f'(x_n)} = x_{n+1},$$

pa je i  $x_{n+1} > \xi$ . Dakle niz  $(x_n)$  je odozdo ograničen sa  $\xi$ .

- Kako je  $x_n > \xi$  za svaki  $n \in \mathbb{N}$ , onda zbog strogog monotonog rasta funkcije  $f$  mora biti i  $f(x_n) > f(\xi) = 0$  za svaki  $n \in \mathbb{N}$ . Osim toga po pretpostavci je i  $f'(x_n) > 0$  za svaki  $n \in \mathbb{N}$ . Zato iz (4.18) slijedi

$$x_n - x_{n+1} = \frac{f(x_n)}{f'(x_n)} > 0,$$

što znači da je niz  $(x_n)$  monotonno padajući.

Budući da je niz  $(x_n)$  monotonno padajući i ograničen odozdo, on je konvergentan i čitav se nalazi u  $I$ , pa zbog toga postoji realni broj  $\bar{\xi} \in I$ , takav da bude  $\bar{\xi} = \lim x_n$ . Ako sada u (4.18) pustimo  $n \rightarrow \infty$ , dobivamo

$$\bar{\xi} = \bar{\xi} - \frac{f(\bar{\xi})}{f'(\bar{\xi})},$$

odakle (zbog  $f'(\bar{\xi}) > 0$ ) slijedi  $f(\bar{\xi}) = 0$ . Kako je  $\bar{\xi}$  jedinstvena nultočka funkcije  $f$  na intervalu  $I$ , mora biti  $\bar{\xi} = \xi$ .

Prijeđimo na dokaz ocjene (4.20). Prema Taylorovoj formuli vrijedi

$$f(x_n) = f(x_{n-1}) + f'(x_{n-1}) \cdot (x_n - x_{n-1}) + \frac{1}{2} \cdot f''(c) \cdot (x_n - x_{n-1})^2, \quad c \in (a, b).$$

Zbog (4.18), ostaje samo

$$f(x_n) = \frac{1}{2} f''(c) (x_n - x_{n-1})^2,$$

odakle je

$$|f(x_n)| \leq \frac{1}{2} M_2 (x_n - x_{n-1})^2. \quad (4.23)$$

Iz (4.3) koristeći (4.23), dobivamo ocjenu (4.20).

U svrhu dokaza relacije (4.22), ponovo koristimo Taylorovu formulu

$$0 = f(\xi) = f(x_n) + f'(x_n)(\xi - x_n) + \frac{1}{2} \cdot f''(c)(\xi - x_n)^2, \quad c \in (a, b),$$

odakle je

$$0 = \frac{f(x_n)}{f'(x_n)} + \xi - x_n + \frac{1}{2} \cdot \frac{f''(c)}{f'(x_n)}(\xi - x_n)^2.$$

Koristeći ovdje (4.18), dobivamo

$$\xi - x_{n+1} = -\frac{1}{2} \cdot \frac{f''(c)}{f'(x_n)}(\xi - x_n)^2,$$

odakle slijedi (4.22).  $\square$

**Primjedba 4.5** Ako je početna aproksimacija  $x_0$  izabrana u intervalu za koji vrijedi

$$\frac{M_2}{2m_1} \leq 2,$$

onda će se u svakom koraku broj signifikantnih decimala aproksimacije podvostručavati. Primjerice ako je  $n$ -ta aproksimacija imala  $k$  signifikantnih decimala, tj.  $|\xi - x_n| < \frac{1}{2}10^{-k}$ , tada je prema (4.22)

$$|\xi - x_{n+1}| \leq \frac{M_2}{2m_1}(\xi - x_n)^2 \leq 2 \frac{1}{4}10^{-2k} = \frac{1}{2}10^{-2k},$$

tj.  $(n+1)$ -va aproksimacija ima  $2k$  signifikantnih decimala.

**Primjedba 4.6** Za Newtonovu metodu kažemo da je lokalno konvergentna. To znači da postoji neka okolina oko nultočke  $\xi$  u kojoj proizvoljno možemo birati početnu aproksimaciju  $x_0$ , pa da iterativni proces konvergira. Ako  $x_0$  izaberemo izvan te okoline, iterativni proces neće konvergirati.

Primjerice, za funkciju  $f(x) = \arctg x$  ako je  $|x_0| < 1.39$ , Newtonova metoda konvergirat će prema  $\xi = 0$ , a ako je  $|x_0| > 1.40$ , Newtonova metoda će divergirati.

**Zadatak 4.2** Za koju će vrijednost početne aproksimacije  $x_0$  Newtonova metoda za funkciju  $f(x) = \arctg x$  proizvoditi “kruženje” ( $x_1 = -x_0$ ,  $x_2 = x_0, \dots$ )?

**Primjer 4.8** S točnošću  $\epsilon = 0.00005$  (4 signifikantne decimala) treba odrediti pozitivnu nultočku funkcije

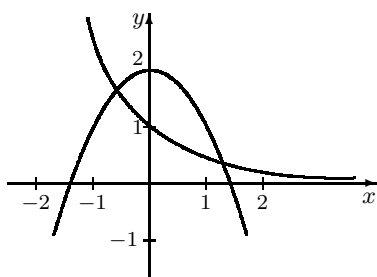
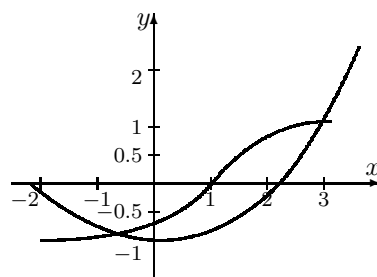
$$f(x) = e^{-x} + x^2 - 2.$$

Ova funkcija ima dvije nultočke. Jedna se nalazi u intervalu  $I_1 = [-1, 0]$  a druga u intervalu  $I_2 = [1, 2]$  (Slika 4.7). Mi ćemo potražiti nultočku iz intervala  $[1, 2]$ . Kako je  $f'(x) = -e^{-x} + 2x$  rastuća funkcija na  $I_2$ , bit će  $m_1 = f'(1) \geq 1.6$ , a kako je  $f''(x) = e^{-x} + 2$  padajuća funkcija na  $I_2$ , uzet ćemo  $M_2 = f''(1) \leq 2.4$ . Zbog  $f(2) \cdot f''(2) > 0$  za početnu aproksimaciju izabrat ćemo  $x_0 = 2$ .

Nakon 3 iteracije dobivamo  $x^* = 1.3160$ . Tijek iterativnog procesa prikazan je u Tablici 4.4 ( $g_0$  je desna strana ocjene (4.3), a  $g_2$  desna strana ocjene (4.20)).

$n$	$x_n$	$g_0$	$g_2$	$f(x_n)$	$f'(x_n)$
0	2	1.334585	—	2.135336	3.864665
1	1.4475	0.206462	0.228965	0.330339	2.659780
2	1.3233	0.010823	0.011569	0.017316	2.380286
3	1.3160	0.000037	0.000040	0.000060	2.363792

Tablica 4.4

Slika 4.7  $e^{-x} = 2 - x^2$ Slika 4.8  $\arctg(x - 1) = \frac{1}{5}x^2 + 1$ 

Za izvođenje iterativnog postupka može se koristiti sljedeći MATHEMATICA-program

```
f[x_] := Exp[-x] + x^2 - 2
n=0; x0=2; m1=1.6; m2=2.4; eps=.00005;
x=x0;
While[N[Abs[f[x]]/m1]>eps,
  Print["Iteracija ",n,":  xn=",x," greska=",f[x]/m1];
  x=N[x-f[x]/f'[x]]; n=n+1
]
Print["Iteracija ",n,":  xn=",x," greska=",f[x]/m1];
```

**Primjer 4.9** S točnošću  $\epsilon = 0.00005$  treba odrediti pozitivnu nultočku funkcije

$$f(x) = \arctg(x - 1) - \frac{1}{5}x^2 + 1.$$

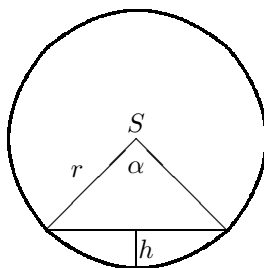
Ova funkcija ima dvije nultočke. Jedna se nalazi u intervalu  $I_1 = [-1, 0]$ , a druga u intervalu  $I_2 = [3, 4]$  (Slika 4.8). Mi ćemo potražiti nultočku iz intervala  $[3, 4]$ . Kako je  $f'$  negativna i padajuća funkcija na  $I_2$ , uzet ćemo  $m_1 = |f'(3)| \approx 1$ , a kako je  $f''$  negativna i rastuća funkcija na  $I_2$ , uzet ćemo  $M_2 = |f''(3)| \approx 0.56$ . Zbog  $f(4) \cdot f''(4) > 0$  za početnu aproksimaciju izabrat ćemo  $x_0 = 4$ . Nakon 3 iteracije dobivamo  $x^* = 3.2850$ . Tijek iterativnog procesa prikazan je u Tablici 4.5.

$n$	$x_n$	(4.3)	(4.20)	$f(x_n)$	$f'(x_n)$
0	4	0.950954	—	-0.950954	-1.500000
1	3.366031	0.095113	0.112537	-0.095113	-1.194853
2	3.286428	0.001621	0.001774	-0.001621	-1.154000
3	3.285023	0.000001	0.000001	-0.000001	-1.153272

Tablica 4.5



**Primjer 4.10** Spremnik za naftu ima oblik položenog cilindra radijusa  $r = 1m$ . Kolika je visina  $h$  razine nafte ako je spremnik napunjen s  $q = 1/4$  svog volumena?



Slika 4.9 Razina spremnika

Označimo s  $d$  duljinu spremnika, a s  $\alpha$  kut kružnog isječka (Slika 4.9). Količina (volumen) nafte zadana je formulom  $V_h = \frac{1}{2}r^2\alpha d - \frac{1}{2}r^2\sin\alpha \cdot d$ , odnosno

$$V_h = \frac{1}{2}r^2d(\alpha - \sin\alpha). \quad (*)$$

gdje je  $d$  duljina spremnika.

S druge strane kako je volumen punog spremnika  $V = r^2\pi \cdot d$ , onda je  $q$ -ti dio njegovog volumena jednak  $V_q = r^2\pi \cdot d \cdot q$ . Kut  $\alpha$ , za koji je  $V_h = V_q$ , dobivamo iz jednadžbe

$$\alpha - \sin\alpha - 2\pi q = 0. \quad (**)$$

Kako je  $\cos \frac{\alpha}{2} = (r-h)/r$ , nakon što smo iz (\*\*) odredili  $\alpha$ , visinu  $h$  razine nafte dobivamo iz formule

$$h = r \left(1 - \cos \frac{\alpha}{2}\right). \quad (***)$$

Primjenom Newtonove metode tangenti uz  $\alpha_0 = 3$  i  $\epsilon = 0.00005$ , dobivamo  $\alpha = 2.3099$  ( $\approx 132^\circ$ ). Nakon toga iz (\*\*\*) dobivamo  $h = 59.6cm$ .

### 4.3.1 Modifikacije Newtonove metode

Činjenica da u svakom koraku Newtonove metode moramo računati i vrijednost funkcije i njene derivacije, u nekim slučajevima otežava primjenu ove metode. Navest ćemo nekoliko modifikacija Newtonove metode u kojima se izbjegava izračunavanje derivacije funkcije u svakoj iteraciji.

Najvažnija modifikacija je tzv. **metoda sekanti**. U intervalu  $I = [a, b]$  izabrat ćemo dvije početne aproksimacije  $x_0$  i  $x_1$  te povući sekantu na krivulju kroz točke  $(x_0, f(x_0))$ ,  $(x_1, f(x_1))$ . Sljedeću aproksimaciju  $x_2$  dobit ćemo kao sjecište sekante s osi  $x$  (Slika 4.10)

$$x_2 = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)}, \text{ ako je } f(x_1) \neq f(x_0).$$

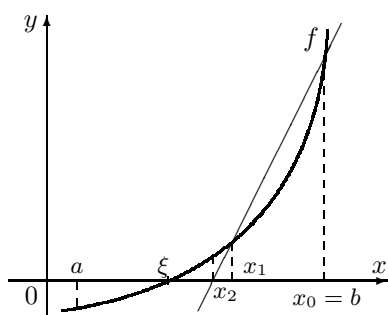
Ponavljajući postupak, dobivamo niz definiran rekurzivnom formulom

$$x_{n+1} = \frac{x_{n-1} f(x_n) - x_n f(x_{n-1})}{f(x_n) - f(x_{n-1})}, \quad f(x_n) \neq f(x_{n-1}), \quad n = 1, 2, \dots \quad (4.24)$$

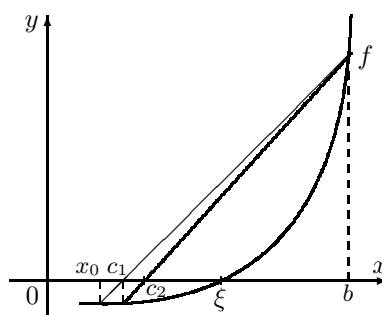
Ako su početne aproksimacije  $x_0, x_1$  birane dovoljno blizu rješenja  $\xi$ , onda uz uvjete  $f'(\xi) \neq 0$  i  $f''(\xi) \neq 0$ , metoda sekanti ima brzinu konvergencije  $\frac{1+\sqrt{5}}{2} \approx 1.618$ ,  $n \gg 1$  (vidi primjerice DAHLQUIST (1972), SCHWARZ (1986)).

**Zadatak 4.3** Pokažite da se rekursivna formula (4.24) može dobiti izravno iz Newtonove metode (4.18), str. 78, tako da  $f'(x_n)$  aproksimiramo podjeljenom razlikom

$$\frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}$$



Slika 4.10 Metoda sekanti



Slika 4.11 Regula falsi

**Regula falsi** (*metoda krivih položaja*) je metoda, koju također možemo smatrati jednom modifikacijom Newtonove metode, odnosno jednom varijantom metode sekanti.<sup>5</sup>

Pretpostavimo da je funkcija  $f$  neprekidna na zatvorenom intervalu  $[a, b]$  i da je  $f(a) \cdot f(b) < 0$ . Označimo:  $x_0 := a$ ,  $b_0 := b$  i povucimo pravac točkama  $(x_0, f(x_0))$ ,  $(b_0, f(b_0))$ . On siječe os  $x$  u točki

$$c_1 = \frac{x_0 f(b_0) - b_0 f(x_0)}{f(b_0) - f(x_0)}, \quad x_0 < c_1 < b_0$$

Ako je  $f(c_1) = 0$ , nultočka je pronađena; u protivnom postupimo na sljedeći način

$$\text{ako je } f(c_1)f(x_0) > 0, \quad \begin{matrix} x_1 = c_1; \\ b_1 = b_0 \end{matrix} \quad \text{inače} \quad \begin{matrix} x_1 = c_1 \\ b_1 = x_0 \end{matrix}$$

Ponavljajući postupak, dobivamo niz  $(x_n)$  koji linearnom brzinom konvergira prema jednom korijenu jednadžbe  $f(x) = 0$  na intervalu  $[a, b]$ . Prema izloženom vidi se da ovu metodu možemo smatrati i jednom varijantom metode bisekcije.

Za ocjenu pogreške kod svih ovih modifikacija možemo koristiti (4.3), str. 69.

**Primjer 4.11** Gibanje balističkog projektila u vakuumu određeno je sustavom diferencijalnih jednadžbi (vidi primjerice MOLITZ (1967))

$$\frac{d^2 x(t)}{dt^2} = 0, \quad \frac{d^2 y(t)}{dt^2} = -g(y)$$

<sup>5</sup>Također se može pokazati (vidi STOER (1993)) da je regula falsi specijalni slučaj metode jednostavnih iteracija

s početnim uvjetima

$$x(0) = y(0) = 0, \quad \frac{dx(0)}{dt} = v_0 \cos \alpha, \quad \frac{dy(0)}{dt} = v_0 \sin \alpha,$$

gdje su  $(x, y)$  koordinate (u metrima) proizvoljne točke  $P$  na putanji,  $t$  vrijeme (u sekundama) potrebno da projektil dođe u točku  $P$ ,  $\alpha$  izlazni kut,  $v_0$  početna brzina projektila (u  $\text{ms}^{-1}$ ) i  $g$  akceleracija sile teže (vidi Sliku 4.12). Za male visine  $y$  može se pretpostaviti da je funkcija  $g$  konstantna (približno  $9.81\text{ms}^{-2}$ ). Rješenje sustava diferencijalnih jednačbi možemo zapisati u parametarskom obliku

$$\begin{aligned} (a) \quad x(t) &= v_0 t \cos \alpha \\ (b) \quad y(t) &= v_0 t \sin \alpha - \frac{1}{2} g t^2 \end{aligned} \quad (B.1)$$

ili eksplicitno (isključujući vrijeme  $t$  iz jednačbi (B.1))

$$y = x \tan \alpha - \frac{x^2}{4h \cos^2 \alpha}, \quad h = \frac{v_0^2}{2g}. \quad (B.2)$$

Lako se vidi da je vrijeme leta projektila (dok ne padne u točku  $X$ )

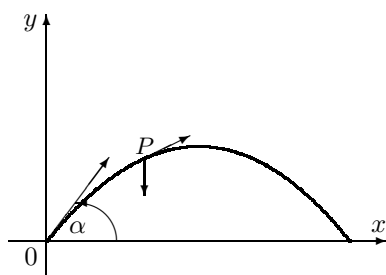
$$T = \frac{2v_0}{g} \sin \alpha, \quad (B.3)$$

uz domet

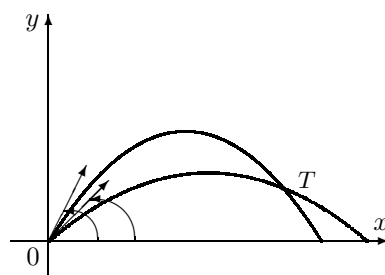
$$X = 2h \sin 2\alpha. \quad (B.4)$$

Maksimalni domet projektila postiže se za  $\alpha = \frac{\pi}{4} \text{ rad}$  ( $45^\circ$ ) i iznosi  $X_M = 2h$ .

Zadana je točka  $T(x_0, y_0)$ . Treba odrediti veličinu izlaznog kuta  $\alpha$  tako da projektil pogodi točku  $T$  (vidi Sliku 4.13).



Slika 4.12 Gibanje balističkog projektila



Slika 4.13 Gađanje točke  $T$

Da bi projektil mogao pogoditi točku  $T$ , njene koordinate moraju zadovoljavati jednačbu

$$y_0 = x \tan \alpha - \frac{x_0^2}{4h \cos^2 \alpha},$$

što možemo pisati

$$y_0 \cos^2 \alpha - \frac{1}{2} x_0 \sin 2\alpha + \frac{x_0^2}{4h} = 0. \quad (*)$$

Jednadžbu (\*) možemo riješiti nekom od metoda navedenih u ovom poglavlju. Jednadžba (\*) može imati jedno ili dva rješenja (tzv. donji i gornji dijapazon kuteva). U *Tablici 4.6* prikazani su dobiveni rezultati za neke karakteristične situacije (kutevi su dani u stupnjima).

$x_n$	$y_n$	$v_0$	$\alpha_1$	$\alpha_2$	max. doimet( $m$ )
1 000	10	115	24.66	65.92	1 348
50 000	500	900	19.28	71.29	82 569
300	1 000	1 200	73.33	—	146 789
500	3 000	500	—	—	25 484

Tablica 4.6

## 4.4 Rješavanje sustava nelinearnih jednadžbi

Spomenimo na kraju mogućnost generalizacije prethodno navedenih metoda na rješavanje sustava nelinearnih jednadžbi

$$f_i(x_1, x_2, \dots, x_n) = 0, \quad i = 1, \dots, n, \quad (4.25)$$

odnosno  $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ , gdje je  $\mathbf{f}(\mathbf{x}) = (f_1(x), \dots, f_n(x))^T$ ,  $\mathbf{x} = (x_1, \dots, x_n)^T$ . Analogno metodi jednostavnih iteracija (t.4.2 str. 72) sustav (4.25) možemo zapisati u obliku

$$x_i = \varphi_i(x_1, x_2, \dots, x_n), \quad i = 1, \dots, n. \quad (4.26)$$

što upućuje na iterativni postupak

$$x_i^{(k+1)} = \varphi_i(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}), \quad i = 1, \dots, n, \quad k = 0, 1, \dots \quad (4.27)$$

Newtonova metoda (t.4.3, str. 77) također se može generalizirati na rješavanje sustava nelinearnih jednadžbi (4.25). Najprije izaberimo početnu aproksimaciju  $\mathbf{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})^T$  i svaku od funkcija  $f_i$  razvijemo u Taylorov red u okolini  $\mathbf{x}^{(0)}$ , te linearnu aproksimaciju označimo s  $\tilde{f}_i$ :

$$\tilde{f}_i(\mathbf{x}) = f_i(\mathbf{x}^{(0)}) + \sum_{j=1}^n \frac{\partial f_i(\mathbf{x}^{(0)})}{\partial x_j} (x_j - x_j^{(0)}), \quad i = 1, \dots, n. \quad (4.28)$$

Sada umjesto sustava (4.25) rješavamo sustav

$$\tilde{f}_i(\mathbf{x}) = 0, \quad i = 1, \dots, n,$$

što možemo pisati u matičnom obliku na sljedeći način

$$\mathbf{J}^{(0)} \mathbf{s}^{(0)} = -\mathbf{f}(\mathbf{x}^{(0)}), \quad (4.29)$$

$$\mathbf{J}^{(0)} = \begin{bmatrix} \frac{\partial f_1(\mathbf{x}^{(0)})}{\partial x_1} & \dots & \frac{\partial f_1(\mathbf{x}^{(0)})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n(\mathbf{x}^{(0)})}{\partial x_1} & \dots & \frac{\partial f_n(\mathbf{x}^{(0)})}{\partial x_n} \end{bmatrix}, \quad \mathbf{s}^{(0)} = \begin{bmatrix} x_1 - x_1^{(0)} \\ \vdots \\ x_n - x_n^{(0)} \end{bmatrix}, \quad \mathbf{f}(\mathbf{x}^{(0)}) = \begin{bmatrix} f_1(\mathbf{x}^{(0)}) \\ \vdots \\ f_n(\mathbf{x}^{(0)}) \end{bmatrix}.$$

Matricu  $\mathbf{J}$  nazivamo Jacobijeva matrica ili Jacobijan sustava. Nova aproksimacija rješenja tada je

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \mathbf{s}^{(0)}.$$

Općenito, dobivamo iterativni postupak

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{s}^{(k)}, \quad k = 0, 1, \dots \quad (4.30)$$

gdje je  $\mathbf{s}^{(k)}$  rješenje sustava

$$\mathbf{J}^{(k)} \mathbf{s} = -\mathbf{f}(\mathbf{x}^{(k)}). \quad (4.31)$$

Primijetite da se iterativni postupak (4.30)–(4.31) uz pretpostavku regularnosti Jacobijana  $\mathbf{J}^{(k)}$  može zapisati u obliku

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \left(\mathbf{J}^{(k)}\right)^{-1} \mathbf{f}(\mathbf{x}^{(k)}), \quad k = 0, 1, \dots, \quad (4.32)$$

što podsjeća na Newtonov iterativni postupak (4.18), str. 78. Inače, iterativni proces (4.32) rijetko se koristi jer je postupak (4.30)–(4.31) numerički stabilniji.

Pitanje konvergencije spomenutih metoda nećemo razmatrati. Zainteresirani čitatelj to može pogledati u više knjiga iz područja numeričke analize spomenutih u *Primjedbi 4.8*, str. 87 ili u popisu referenci na kraju teksta.

**Primjedba 4.7** *Vrlo važno mjesto među metodama za rješavanje sustava nelinearnih jednadžbi zauzimaju tzv. kvazi-Newtonove metode, koje je uveo BROYDEN (1961), a koje su nastale kao generalizacija metode sekanti (vidi t.4.3.1, str. 82):*

(a) Broydenova metoda

$$\mathbf{B}_k(\mathbf{x}^{k+1} - \mathbf{x}^k) = -\mathbf{f}(\mathbf{x}^k), \quad k = 0, 1, \dots,$$

gdje su  $\mathbf{B}_k$  matrice koje uz dobar izbor početne aproksimacije  $\mathbf{B}_0$  sve više nalikuju Jacobijanu, a izračunavaju se iz rekurzivne formule:

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \frac{(\mathbf{y}_k - \mathbf{B}_k \mathbf{s}_k) \mathbf{s}_k^T}{(\mathbf{s}_k, \mathbf{s}_k)}, \quad k = 0, 1, \dots,$$

gdje je

$$\mathbf{y}_k = \mathbf{f}(\mathbf{x}^{k+1}) - \mathbf{f}(\mathbf{x}^k), \quad \mathbf{s}_k = \mathbf{x}^{k+1} - \mathbf{x}^k$$

(b) Davidon-Fletcher-Powell (DFP) metoda

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \mathbf{H}_k \mathbf{f}(\mathbf{x}^k), \quad k = 0, 1, \dots,$$

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{\mathbf{s}_k \mathbf{s}_k^T}{(\mathbf{y}_k, \mathbf{s}_k)} - \frac{\mathbf{H}_k \mathbf{y}_k \mathbf{y}_k^T \mathbf{H}_k}{(\mathbf{y}_k, \mathbf{H}_k \mathbf{y}_k)}, \quad k = 0, 1, \dots,$$

(c) Broyden-Fletcher-Goldfarb-Schano (BFGS) metoda

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \mathbf{H}_k \mathbf{f}(\mathbf{x}^k), \quad k = 0, 1, \dots,$$

$$\mathbf{H}_{k+1} = \left( \mathbf{I} - \frac{\mathbf{s}_k \mathbf{y}_k^T}{(\mathbf{y}_k, \mathbf{s}_k)} \right) \mathbf{H}_k \left( \mathbf{I} - \frac{\mathbf{y}_k \mathbf{s}_k^T}{(\mathbf{y}_k, \mathbf{s}_k)} \right) + \frac{\mathbf{s}_k \mathbf{s}_k^T}{(\mathbf{y}_k, \mathbf{s}_k)}, \quad k = 0, 1, \dots,$$

Primijetimo da je  $\text{rang}(\mathbf{B}_{k+1} - \mathbf{B}_k) = 1$ . Zbog toga se kaže da je Broydenova metoda metoda ranga 1. Ako je Jacobijan simetrična matrica, bolje je koristiti DFP ili BFGS metodu, u kojima su korekcije Jacobijana  $\mathbf{H}_k$  ranga 2. Može se pokazati da rekurzivne formule za generiranje korekcija  $\mathbf{H}_k$  kod DFP i BFGS metode čuvaju simetričnost i pozitivnu definitnost korekcije. Više detalja o ovim metodama može se naći kod DENNIS (1977).

**Zadatak 4.4** Izradite program za rješavanje sustava nelinearnih jednadžbi navedenim metodama i ispitajte ga na sljedećim primjerima

(a) (Dennis, Schnabel (1996))

$$\begin{aligned} 2(x_1 + x_2)^2 + (x_1 - x_2)^2 - 8 &= 0 \\ 5x_1^2 + (x_2 - 3)^2 - 9 &= 0 \end{aligned}$$

$$\text{Rješenje: } \mathbf{x}_1^* = (1, 1)^T, \quad \mathbf{x}_2^* = (-1.18347, 1.58684)^T.$$

(b) (Dennis, Schnabel (1996))

$$\mathbf{f}(x_1, x_2) = (x_1^2 + x_2^2 - 2, e^{x_1-1} + x_2^3 - 2)^T = 0$$

$$\text{Rješenje: } \mathbf{x}^* = (1, 1)^T.$$

(c) (Dennis, Schnabel (1996))

$$\mathbf{f}(x_1, x_2, x_3) = (x_1, x_2^2 + x_2, e^{x_3} - 1)^T = 0$$

(d) *Rosenbrock parabolic valey*

$$\text{gardi } f(x_1, x_2) = 0, \quad f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

$$\text{Rješenje: } \mathbf{x}^* = (1, 1)^T.$$

(e)  $\text{grad } f(x_1, x_2, x_3) = 0, \quad f(x_1, x_2, x_3) = \sum_{i=1}^{11} \left( x_3 \exp\left(-\frac{(t_i - x_1)^2}{x_2}\right) - y_i \right)^2$ ,  
gdje su brojevi  $(t_i, y_i)$ ,  $i = 1, \dots, 11$  zadani tablicom

$t_i$	0	1	2	3	4	5	6	7	8	9	10
$y_i$	.001	.01	.04	.12	.21	.25	.21	.12	.04	.01	.001

$$\text{Rješenje: } \mathbf{x}^* = (4.93, 2.62, 0.28)^T.$$

**Primjedba 4.8** Više detalja o numeričkim metodama za rješavanje nelinearnih jednadžbi može se naći u brojnoj literaturi (primjerice DEMIDOVICH (1981), IVANŠIĆ (1998), KUREPA (1990), MATHEWS (1992), SCHWARZ (1986), STOER (1993) itd.) FORTRAN kodovi mogu se naći u FORSYTHE (1974) ili PRESS (1992). O rješavanju sustava nelinearnih jednadžbi i problemima ekstreme funkcija više varijabli može se naći u DENNIS (1996), DENNIS (1977), FLETCHER (1987), ORTEGA (1970) itd. Kvalitetna FORTRAN programska podrška nalazi se u The NAG-Library (vidi PHILLIPS (1986), KÖCKLER (1990)). Naravno, kvalitetan inženjerski "alat" za rješavanje ovakvih problema je program Mathematica (WOLFRAM (1998)).

## 4.5 Zadaci

**Zadatak 4.5** Nekom iterativnom metodom riješite jednadžbu  $x^2 - 2x - 2 = 0$  s točnošću  $\epsilon = 0.5 \cdot 10^{-4}$ , tj. na 4 signifikantne decimale.

Rješenje:  $x_1 \approx 2.7321$ ,  $x_2 \approx -0.7321$ .

**Zadatak 4.6** Uz točnost  $\epsilon = 0.5 \times 10^{-6}$  odredite  $\sqrt[3]{2}$  rješavajući jednadžbu  $x^3 - 2 = 0$ .

Rješenje:  $\sqrt[3]{2} \approx 1.259921$ .

**Zadatak 4.7** S točnošću  $\epsilon = 0.5 \cdot 10^{-4}$  riješite jednadžbu

a)  $x^3 - 2x - 2 = 0$ ,      b)  $\sin x - 5x + 1/2 = 0$ .

Rješenje: a)  $\xi \approx 1.7693$ ,      b)  $\xi \approx 0.12492$

**Zadatak 4.8** S točnošću  $\epsilon = 0.5 \cdot 10^{-4}$  odredite sjecište grafova krivulja zadanih jednadžbama  $y = 1/(x-1)$  i  $x^2 - y^2 = 1$ .

Rješenje:  $x_0 \approx 1.7167$ ,  $y_0 \approx 1.3953$ .

**Zadatak 4.9** S točnošću  $\epsilon = 0.5 \cdot 10^{-4}$  odredite realne nultočke polinoma

$$P(x) = x^5 - 5x^4 - 4x^3 + 20x^2 - 5x + 25.$$

Rješenje:  $x_1 \approx -2.2361$ ,  $x_2 \approx 2.2361$ ,  $x_3 = 5$ .

**Zadatak 4.10** S točnošću  $\epsilon = 0.5 \cdot 10^{-4}$  odredite najmanji pozitivni korijen jednadžbe  $x \sin x + 1 = 0$ .

Rješenje:  $x \approx 3.4368$ .

**Zadatak 4.11** Jednadžba  $x^3 - x = 0$  ima tri korijena:  $-1$ ,  $0$ ,  $1$ . Ispitajte, prema kojem korijenu jednadžbe će konvergirati Newtonov iterativni proces (4.18, str. 78) ako se izabere proizvoljna početna aproksimacija  $x_0 \in \mathbb{R}$ ?

**Zadatak 4.12** Grafičkom metodom odredite intervale u kojima funkcija  $f$  ima nultočke (DAHLQUIST (1972))

a)  $f(x) = 4 \sin x + 1 - x$ ,      b)  $f(x) = 1 - x - e^{-2x}$ ,      c)  $f(x) = (x+1)e^{x-1}$ ,  
d)  $f(x) = x^4 - 4x^3 + 2x^2 - 8$ ,      e)  $f(x) = e^x + x^2 + x$ ,      f)  $f(x) = e^x - x^2 - 2x - 2$

Rješenje

a)  $I_1 = [-3, -1]$ ,  $I_2 = [-1, 1]$ ,  $I_3 = [2, 3]$ ,      b)  $I_1 = [-.5, .5]$ ,  $I_2 = [.5, 1]$ ,  
c)  $I = [0, 1]$ ,      d)  $I_1 = [-2, 0]$ ,  $I_2 = [3, 4]$ ,  
e) nema nultočaka      f)  $I = [2, 4]$ .

**Zadatak 4.13** Zadana je funkcija  $f$ . Separirajte intervale u kojima se nalaze njezine nultočke, za svaki od tih intervala odredite broj  $m_1$  iz ocjene (4.3), te metodom bisekcije ili metodom iteracija s točnošću  $\epsilon = 0.5 \times 10^{-4}$  odredite nultočke. Provjeri red konvergencije metode računajući  $|e_{n+1}|/|e_n|$  za svaki  $n$ .

a)  $f(x) = e^{-x} \sin(3x+2) + x - 1/2$       b)  $f(x) = x \operatorname{th}(x) - x^2 + 1$   
c)  $f(x) = x^2 - 1 + \ln(x+1)$   
d)  $f(x) = e^{-x} - 0.5 \cdot (x-1)^2 + 1$       e)  $f(x) = \frac{\sin x}{x} - (x-2)^3$ .

Rješenje:

- a)  $I_1 = [-1, -0.4]$ ,  $m_1 \approx 3$ ,  $x_1 \approx -0.4498$ ,  $I_2 = [0, \frac{1}{2}]$ ,  $m_1 \approx 0.49$ ,  $x_2 \approx 0.2832$ ,  
 $I_3 = [0.75, 1]$ ,  $m_1 \approx 0.79$ ,  $x_3 \approx 0.9047$ .  
 b)  $I_1 = [-2, -1]$ ,  $m_1 \approx 0.8$ ,  $x_1 \approx -1.5572$ ,  $I_2 = [1, 2]$ ,  $m_1 \approx 0.8$ ,  $x_2 \approx 1.5572$ .  
 c)  $I = [0, 1]$ ,  $m_1 \approx 1$ ,  $\xi \approx 0.6896$ .  
 d)  $I = [2, 3]$ ,  $m_1 \approx 1.14$ ,  $\xi \approx 2.4727$ .  
 e)  $I = [2, 3]$ ,  $m_1 \approx 0.5$ ,  $x_1 \approx 2.5879$ .

**Zadatak 4.14** Funkcija  $f(x) = x + \ln x$  ima jednostruku realnu nultočku  $\xi$  na intervalu  $[.1, 1]$ . Koji od niže navedenih iterativnih procesa konvergira prema  $\xi$ ? Zašto?

- a)  $x_{n+1} = -\ln x_n$     b)  $x_{n+1} = e^{-x_n}$     c)  $x_{n+1} = \frac{1}{2}(x_n + e^{-x_n})$

**Zadatak 4.15** Neka je  $\varphi : I = [a, b] \rightarrow \mathbb{R}$  neprekidno derivabilna funkcija, a jednadžba  $x = \varphi(x)$  neka ima jedinstveno rješenje na  $I$ . Za proizvoljno izabrani  $x_0 \in I$  grafički prikazite dobivanje prve tri aproksimacije pomoću metode iteracija, ako je

- a)  $0 < \varphi'(x) < 1$ ,  $x \in I$     b)  $0 > \varphi'(x) > -1$ ,  $x \in I$   
 c)  $\varphi'(x) > 1$ ,  $x \in I$     d)  $\varphi'(x) < -1$ ,  $x \in I$ .

**Zadatak 4.16** Neka funkcija  $f : I \rightarrow \mathbb{R}$  ima neprekidnu drugu derivaciju na intervalu  $I = [a, b]$  i neka u tom intervalu jednadžba  $f(x) = 0$  ima jedinstveno rješenje  $\xi$ . Neka je nadalje,  $f(a) \cdot f(b) < 0$ , a prva ( $f'$ ) i druga ( $f''$ ) derivacija funkcije  $f$  na intervalu  $I$  imaju stalan predznak. Za izabrani  $x_0 \in I$  grafički prikazite dobivanje prve tri aproksimacije za sve četiri kombinacije predznaka prve ( $f'$ ) i druge ( $f''$ ) derivacija funkcije  $f$ :

- a) kod Newtonove metode tangenti,    b) kod metode sekanti,  
 c) kod metode krivih položaja (regula falsi).

**Zadatak 4.17** Rješenje jednadžbe  $x = a/x$ ,  $a > 0$  je  $\xi = \pm\sqrt{a}$ . Pokažite da niz  $x_{n+1} = a/x_n$ ,  $n = 0, 1, \dots$  (formula (4.8) za funkciju  $\varphi(x) = a/x$ ) nije konvergentan ni za jedan izbor  $x_0 \in \mathbb{R}$ .

**Zadatak 4.18** Rekursivnom formulom zadan je niz (formula (4.18) za funkciju  $f(x) = x^2 - a$

$$x_{n+1} = \frac{1}{2} \left( x_n + \frac{a}{x_n} \right), \quad a > 0, \quad x_0 > a, \quad n = 0, 1, \dots$$

Bez pozivanja na Teorem 4.2 dokažite konvergenciju ovog niza te odredite graničnu vrijednost.

Uputa: Treba pokazati da je niz monotono padajući i ograničen odozdo.

**Zadatak 4.19** Slično kao u Primjeru 4.7, str. 76 definirajte i analizirajte iterativne postupke za rješavanje jednadžbe:

- a)  $f(x) = x^3 - a$ ,  $a \in \mathbb{R}$  (treći korijen realnog broja  $a$ ),  
 b)  $f(x) = x^k - a$ ,  $a > 0$ ,  $k \in \mathbb{N}$  ( $k$ -ti korijen pozitivnog broja  $a$ ).

Uputa: Koristite formulu (4.18), str. 78.

Rješenje: a)  $x_{n+1} = \frac{1}{3}(2x_n + \frac{a}{x_n^2})$ ; b)  $x_{n+1} = \frac{1}{k}((k-1)x_n + \frac{a}{x_n^{k-1}})$ .



**Zadatak 4.20** Realni korijen jednadžbe  $x^3 - x - 4 = 0$  je  $x_0 = \sqrt[3]{2 + \frac{1}{9}\sqrt{321}} + \sqrt[3]{2 - \frac{1}{9}\sqrt{321}}$ . Izračunajte  $x_0$  s točnošću na 4 decimale iz ove formule, te također korištenjem Newtonove metode za  $x_0 = 2$ .

Rješenje:  $x_0 = 1.7963$ .

**Zadatak 4.21** Za Primjer 4.10, str. 82 odredite visine svih nivoa nafte koji odgovaraju količinama 2%, 4%, 6%, ..., 100%. Izradite odgovarajuću tablicu.

**Zadatak 4.22** Ako su u Primjeru 4.10, str. 82 poznate visine  $h$  razine nafte, izrazite odgovarajući volumen nafte kao funkciju od  $r$ ,  $l$ ,  $h$ . Za zadane visine razine nafte  $h : 2, 4, \dots, 100$  cm izračunajte odgovarajuće količine nafte. Izradite odgovarajuću tablicu.

Uputa: Uvrštavajući (\*\*) u (\*) dobivamo  $V$  kao funkciju od  $h$ .

**Zadatak 4.23** Slično kao u Primjeru 4.11, str. 83 izvedite jednadžbe gibanja balističkog projektila uz dodatnu pretpostavku da je otpor zraka proporcionalan brzini projektila (konstantu proporcionalnosti označimo s  $c$ ). Izračunajte vrijeme leta  $T$  i domet  $X$  projektila ako je:  $\alpha = 45^\circ$ ,  $v_0 = 226.3 \text{ ms}^{-1}$ ,  $c = 10$ . Nacrtajte graf odgovarajuće balističke krivulje.

Rješenje:  $T \approx 8.7 \text{ s}$ ,  $X \approx 929 \text{ m}$ ,

$$x = cv_0 \cos \alpha \cdot (1 - e^{-t/c}), \quad y = (cv_0 \sin \alpha + gc^2) \cdot (1 - e^{-t/c}) - gct.$$

**Zadatak 4.24** Ako je zadana početna brzina  $v_0$  balističkog projektila iz Primjera 4.11, str. 83, pronađite uvjet kojega moraju ispunjavati koordinate  $x_0$  i  $y_0$  točke  $T$  da bi se balističkim projektilom mogla pogoditi točka  $T$ .

Rješenje:  $4h^2 > 4hy_0 + x_0^2$ .

**Zadatak 4.25** Kada će jednadžba (\*) iz Primjera 4.11, str. 83 imati jedno, a kada dva rješenja?

Rješenje: Jednadžba (\*) imat će jedno rješenje ako je  $y_0 \leq x_0$ .

**Zadatak 4.26** Izradite program koji će lokalizirati nultočku neke funkcije  $f$ , a zatim primjenom metode bisekcije (str. 71) utvrditi dovoljno mali (zadane duljine  $\epsilon$ ) interval u kome se nalazi nultočka.

**Zadatak 4.27** Provjerite da jednadžba:  $x - 2 \sin x = 0$  ima tri realne nultočke:  $\xi_1 = 0$ ,  $\xi_2 \in [\frac{\pi}{2}, 2]$ ,  $\xi_3 \in [-2, -\frac{\pi}{2}]$ . Pokažite da za proizvoljni  $x_0 \in [\frac{\pi}{2}, 2]$  niz definiran rekurzivnom formulom

$$x_{k+1} = 2 \sin x_k, \quad k = 0, 1, \dots$$

konvergira prema  $\xi_2$ . Za koje  $x_0$  će Newtonova metoda konvergirati prema  $\xi_2$ ? Usporedite red konvergencije navedenih metoda. Izradite odgovarajuće programe.

**Zadatak 4.28** Može li se Newtonova metoda primijeniti za rješavanje jednadžbe  $f(x) = 0$ , ako je:

$$a) f(x) = x^2 - 14x + 50; \quad b) f(x) = x^{1/3}; \quad c) f(x) = (x - 3)^{1/2}, \quad x_0 = 4.$$

**Zadatak 4.29** Pokažite da Newtonov iterativni postupak konvergira prema jedinstvenom rješenju jednadžbe  $e^{2x} + 3x + 2 = 0$  za proizvoljni  $x_0$ .

**Zadatak 4.30** Može se pokazati da Newtonova metoda ima kvadratnu brzinu konvergencije u jednostrukoj, a linearnu u dvostrukoj nultočki (vidi primjerice MATH-NEWS (1992)). Ustanovite to na primjeru funkcije  $f(x) = x^3 - 3x + 2$  ( $\xi_1 = -2$  je jednostruka, a  $\xi_2 = 1$  dvostruka nultočka).

Uputa: U iterativnom procesu za svaki  $n$  računajte  $|e_{n+1}|/|e_n|^2$  za  $\xi_1 = -2$ , odnosno  $|e_{n+1}|/|e_n|$  za  $\xi_2 = 1$ .

**Zadatak 4.31** Newtonov iterativni postupak (4.18) ima linearnu brzinu konvergencije ako je  $\xi$  dvostruka nultočka funkcije  $f$ . Pokaži da tada niz definiran s

$$x_{n+1} = x_n - 2 \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, \dots$$

konvergira prema  $\xi$  kvadratnom brzinom konvergencije. Provjerite to na primjeru funkcije  $f(x) = x^3 - 3x + 2$ , kojoj je  $\xi = 1$  dvostruka nultočka.

**Zadatak 4.32** Pokažite da niz definiran Newtonovim iterativnim postupkom:

- a) konvergira za  $f(x) = e^{2x} + 3x + 2$  i za proizvoljni  $x_0$ ;
- b) divergira za  $f(x) = x \cdot e^{-x}$  i  $x_0 = 2$ ;
- c) ne konvergira prema rješenju, a članovi niza se ponavljaju skoro ciklički ( $x_{k+4} \approx x_k$ ,  $k = 0, 1, \dots$ ) za  $f(x) = x^3 - x - 3$  i  $x_0 = -3$ ; ako izaberemo  $x_0 = 2$ , niz konvergira rješenju  $\xi = 1.671699881$ ;
- d) divergira u širem smislu za  $f(x) = \arctg x$  i  $x_0 = 1.45$ ; ako izaberemo  $x_0 = 0.5$ , niz konvergira prema rješenju  $\xi = 0$ .

Objasnite nastale fenomene i nacrtajte odgovarajuće slike.

**Zadatak 4.33** Lokalni minimum neprekidne dovoljno "glatke" funkcije  $f$  može se tražiti Newtonovim iterativnim postupkom

$$x_{k+1} = x_k - [f''(x_k)]^{-1} f'(x_k), \quad k = 0, 1, \dots$$

za povoljno odabrani  $x_0$ . Izradite odgovarajući program i ispitajte ga na primjeru funkcije  $f(x) = x^3 - 2x - 5$ , koja postiže minimum u točki 0.8165.

**Zadatak 4.34** U Dodatku je priložen program za traženje lokalnog minimuma funkcije  $F$ . Na početku programa unosi se analitički izraz minimizirajuće funkcije, učitavaju se granice intervala  $[XA, XB]$  u kome je lokaliziran minimum i željena točnost aproksimacije EPS. Program je napisan prema ideji izloženoj u knjizi (vidi također FORSYTHE (1974)):

R. P. BRENT, *Algorithms for minimization without derivatives*, Englewood Cliffs, N.J., Prentice-Hall, 1973.

Proces minimizacije obavlja se kombinacijom metode zlatnog reza i sukcesivne paraboličke aproksimacije minimizirajuće funkcije. Na početku programa ispituje se relativna preciznost računala (veličina PREC). Ispitajte navedeni program na nekoliko

primjera, te rezultate komparirajte s metodom navedenom u Zadatku 4.33. Primijetite da navedena metoda ne zahtijeva ni prvu ni drugu derivaciju minimizirajuće funkcije.

**Zadatak 4.35** Nultočku funkcije  $g$  na separiranom intervalu  $[a, b]$  može se tražiti i tako da na tom intervalu minimiziramo funkciju  $f(x) = g^2(x)$ . Primjenom programa za jednodimenzionalnu minimizaciju iz Dodatka riješite neke od prethodno navedenih jednažbi.

## 5. Aproksimacija funkcija

U t.2, str.16 opisan je problem *interpolacije* funkcije. Zadanu funkciju  $f$  zamijenili smo polinomom  $P$  (kao funkcijom čije se vrijednosti lako izračunavaju), tako da se funkcija  $f$  i polinom  $P$  podudaraju u zadanim točkama (čvorovima interpolacije). Sada bismo željeli istu funkciju  $f$  **aproksimirati** nekom jednostavnom – lako izračunljivom – funkcijom  $f^*$ , koja će danu funkciju  $f$  “*dobro predstavljati*” u čitavom području definicije. Pri tome ćemo kvalitetu aproksimacije  $f^*$  mjeriti “*udaljenošću*” funkcije  $f$  od njene aproksimacije  $f^*$

$$d(f, f^*) = \|f - f^*\|,$$

gdje je norma  $\|\cdot\|$  najčešće zadana s

$$\begin{aligned} \|f\|_p &= \left( \int_a^b \omega(x) |f(x)|^p dx \right)^{1/p} & (L_p \text{ norma}, p \geq 1), \\ \|f\|_\infty &= \max_{a \leq x \leq b} |f(x)| \omega(x) & (\text{Čebiševljeva } L_\infty \text{ norma}), \end{aligned}$$

gdje je  $\omega(x) > 0$  tzv. težinska funkcija.

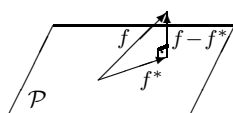
**Primjedba 5.1** *Primijetite da od svih navedenih normi jedino za  $L_2$ -normu možemo definirati skalarni produkt*

$$(f, g) = \int_a^b \omega(x) f(x) g(x) dx, \quad (5.1)$$

tako da je  $\|f\|_2^2 = (f, f)$ . Ako je  $\omega(x) = 1$ , onda  $d_2(f, f^*) = \|f - f^*\|_2$  predstavlja približnu površinu između grafova funkcija  $f$  i  $f^*$ .

Ako je  $\mathcal{P} \subset C_{[a,b]}$  normirani potprostor nekih jednostavnih funkcija (primjerice polinoma), onda kažemo da je  $f^* \in \mathcal{P}$  najbolja aproksimacija funkcije  $f$  na potprostoru  $\mathcal{P}$  ako vrijedi

$$\|f - f^*\| \leq \|f - u\| \text{ za sve } u \in \mathcal{P}$$



Slika 5.1. Najbolja aproksimacija



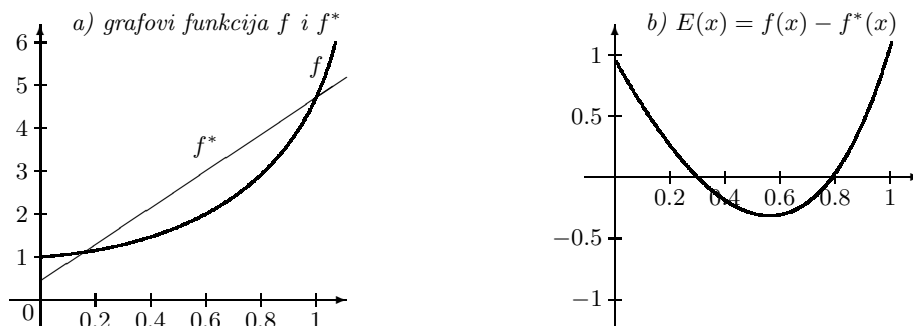
Lako se vidi da su u ovom slučaju koeficijenti matrice sustava (5.4) zadani s

$$(\varphi_i, \varphi_j) = \int_0^1 \varphi_i(x) \varphi_j(x) dx = \int_0^1 x^{i+j} dx = \frac{1}{i+j+1}, \quad i, j = 0, 1,$$

a vektor slobodnih koeficijenata

$$\left( \frac{e^2 - 1}{2}, \frac{e^2 + 1}{4} \right)^T = (3.19453, 2.09726)^T.$$

Rješenje sustava (5.4) je  $a_0^* = 0.194528$ ,  $a_1^* = 6.0$ .



Slika 5.2 Aproksimacija funkcije  $f(x) = e^{2x}$

Na Slici 5.2 prikazani su grafovi funkcija  $f$  i  $f^*$ , te graf funkcije pogreške  $E(x) = f(x) - f^*(x)$ .

Primijetite da je matrica sustava u ovom primjeru poznata Hilbertova matrica. Ona se u literaturi obično uzima kao primjer loše uvjetovane matrice. Primjerice, za  $n = 10$ , uvjetovanost Hilbertove matrice je  $2 \times 10^7$ . Zato je traženje najbolje  $L_2$  aproksimacije rješavanjem sustava (5.4) često puta nepouzđano.

Ako bi funkcije  $\varphi_0, \varphi_1, \dots, \varphi_n$  činile ortogonalan sustav funkcija, tj. ako bi bilo

$$(\varphi_i, \varphi_j) = 0, \quad i \neq j \quad \text{ i } \quad \|\varphi_i\| \neq 0 \quad \text{ za sve } i = 0, 1, \dots, n,$$

onda bi matrica sustava (5.4) bila dijagonalna, a rješenje bi mogli eksplicitno zapisati

$$a_i^* = \frac{(f, \varphi_i)}{(\varphi_i, \varphi_i)}, \quad i = 0, 1, \dots, n. \quad (5.5)$$

**Primjer 5.2** Lako se može pokazati (vidi primjerice KUREPA (1977)) da sustav funkcija

$$1, \sin x, \cos x, \dots, \sin nx, \cos nx,$$

čini ortogonalan sustav funkcija na  $C_{[-\pi, \pi]}$  uz  $\omega(x) = 1$  i da je

$$\|1\| = \sqrt{2\pi}, \quad \|\sin kx\| = \|\cos kx\| = \sqrt{\pi}, \quad k = 1, \dots, n.$$

**Zadatak 5.1** Pokažite da je sustav funkcija  $\varphi_i(x) = \cos ix$ ,  $i = 0, 1, \dots, n$ , ortogonalan na  $[0, \pi]$  uz  $\omega(x) = 1$  i da je

$$\|\varphi_0\| = \sqrt{\pi}, \quad \|\varphi_i\| = \sqrt{\frac{\pi}{2}}, \quad i = 1, \dots, n.$$

**Zadatak 5.2** Pokažite da je sustav funkcija  $\varphi_i(x) = \cos ix$ ,  $i = 0, 1, \dots, n$ , definiran na  $D = \left\{x_i = \frac{2i+1}{n+1} \frac{\pi}{2} : i = 0, 1, \dots, n\right\} \subset [0, \pi]$  čini ortogonalan sustav sa skalarnim produktom

$$(f, g) = \sum_{i=0}^n f(x_i)g(x_i),$$

i da je  $\|\varphi_0\| = \sqrt{n+1}$ ,  $\|\varphi_i\| = \sqrt{\frac{n+1}{2}}$ ,  $i = 1, \dots, n$ .

**Primjedba 5.2** Ako su  $f, g$  ortogonalne funkcije, onda za njih vrijedi Pitagorin teorem

$$\|f + g\|^2 = \|f\|^2 + \|g\|^2. \quad (5.6)$$

Općenitije, za ortogonalan sustav  $\varphi_0, \varphi_1, \dots, \varphi_n$  vrijedi

$$\left\| \sum_{i=0}^n a_i \varphi_i \right\|^2 = \sum_{i=0}^n a_i^2 \|\varphi_i\|^2.$$

Nadalje, kako je  $f - f^*$  okomit na  $L(\varphi_0, \varphi_1, \dots, \varphi_n)$  i kako je  $f^* = \sum_{i=0}^n a_i^* \varphi_i$ , onda je  $(f - f^*, f^*) = 0$  i

$$\|f - f^*\| = \|f\|^2 - 2(f, f^*) + \|f^*\|^2 = \|f\|^2 - \|f^*\|^2 = \|f\|^2 - \sum_{i=0}^n (a_i^*)^2 \|\varphi_i\|^2. \quad (5.7)$$

Iz (5.7) izravno slijedi poznata Besselova nejednakost

$$\sum_{i=0}^n (a_i^*)^2 \|\varphi_i\|^2 \leq \|f\|^2. \quad (5.8)$$

**Primjedba 5.3** Budući da je sustav funkcija iz Primjera 5.2, str. 95, ortogonalan za svaki  $n \in \mathbb{N}$ , možemo promatrati i beskonačni ortogonalni sustav funkcija

$$1, \sin x, \cos x, \dots, \sin nx, \cos nx, \dots,$$

i neku funkciju  $f$  pokušati prikazati pomoću sume reda

$$\mathcal{F}(x) := \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos kx + b_k \sin kx, \quad (5.9)$$

gdje je prema (5.5)

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) dx, \quad a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx dx, \quad b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx dx, \quad (5.10)$$

Red (5.9) nazivamo Fourierov red, a brojeve  $a_0, a_1, b_1, \dots$  Fourierovi koeficijenti funkcije  $f$ . Kada će red (5.9) za neki  $x_0$  biti konvergentan i da li će u tom slučaju biti  $\mathcal{F}(x_0) = f(x_0)$  određeno je Dirichletovim teoremom (vidi primjerice KUREPA (1977)).

### 5.1.1 Ortogonalni polinomi

Pretpostavimo da je na  $C_{[a,b]}$  definiran skalarni produkt (5.1) s težinskom funkcijom  $\omega$ . Neka je  $L(\varphi_0, \varphi_1, \dots, \varphi_n)$  potprostor razapet linearno nezavisnim funkcijama  $\varphi_0, \varphi_1, \dots, \varphi_n$ . U tom potprostoru definirat ćemo novu ortogonalnu bazu  $\psi_0, \psi_1, \dots, \psi_n$  (Gram-Schmidtov postupak ortogonalizacije):

$$\begin{aligned}\psi_0 &= \varphi_0, \\ \psi_1 &= \varphi_1 - \frac{(\varphi_1, \psi_0)}{(\psi_0, \psi_0)} \psi_0, \\ &\vdots \\ \psi_n &= \varphi_n - \sum_{j=0}^{n-1} \frac{(\varphi_n, \psi_j)}{(\psi_j, \psi_j)} \psi_j.\end{aligned}\tag{5.11}$$

Budući da je  $\psi_k$  linearna kombinacija vektora  $\varphi_0, \varphi_1, \dots, \varphi_k$ , za svaki  $k = 0, 1, \dots, n$ , vrijedi  $L(\varphi_0, \varphi_1, \dots, \varphi_n) = L(\psi_0, \psi_1, \dots, \psi_n)$ . Zato najbolju aproksimaciju funkcije  $f$  možemo tražiti kao linearnu kombinaciju ortogonalnih funkcija  $\psi_0, \psi_1, \dots, \psi_n$ .

**Primjer 5.3** Neka je  $(\cdot, \cdot)$  skalarni produkt definiran na  $C_{[-1,1]}$  uz  $\omega(x) = 1$ . Sustav funkcija  $\varphi_i(x) = x^i$ ,  $i = 0, 1, \dots, 5$  ortogonalizirat ćemo prema Gram-Schmidtovom postupku (5.11).

Niže navedenim *Mathematica*-programom dobivamo

$$\begin{aligned}\psi_0(x) &= 1, & \psi_1(x) &= x, & \psi_2(x) &= x^2 - \frac{1}{3}, \\ \psi_3(x) &= x^3 - \frac{3}{5}x, & \psi_4(x) &= x^4 - \frac{6}{7}x^2 + \frac{3}{35}, & \psi_5(x) &= x^5 - \frac{10}{9}x^3 + \frac{5}{21}x.\end{aligned}\tag{5.12}$$

(\* Gram - Schmidtov postupak ortogonalizacije \*)

```
fi[i_, x_] := x^(i-1);
om[t_] := 1; a=-1; b=1; n=6;
Do[
  psi[i,t] = fi[i,t] -
    Sum[Integrate[om[t] fi[i,t] psi[j,t], {t,a,b}] psi[j,t]/
      Integrate[om[t] psi[j,t]^2, {t,a,b}], {j,i-1}];
  ,{i,n}]
ortpol = Table[{i-1, Simplify[psi[i,t]]}, {i,n}];
Print[TableForm[ortpol] ]
```

Polinomi (5.12) se do na konstantu podudaraju s poznatim Laguerreovim polinomima, koji se obično definiraju kao

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} [(x^2 - 1)^n], \quad n \in \mathbb{N}.\tag{5.13}$$



**Zadatak 5.3** Pokažite da za Legendreove polinome (5.13) vrijedi

a) Svojstvo ortogonalnosti

$$\int_{-1}^1 P_m(x)P_n(x) dx = \begin{cases} 0, & m \neq n \\ \frac{2}{2n+1}, & m = n, \end{cases} \quad m, n = 0, 1, \dots$$

b) Legendreov polinom  $P_n$ ,  $n \geq 1$ , u intervalu  $(-1, 1)$  ima  $n$  jednostrukih nultočaka

Uputa: primijetite da je  $(x^2 - 1)/_{x=-1,1} = 0$  i koristite Rolleov teorem.

c)  $P_0(x) = 1$ ,  $P_1(x) = x$ , a za  $n = 1, 2, \dots$  vrijedi rekurzivna formula:

$$P_{n+1}(x) = \frac{2n+1}{n+1}xP_n(x) - \frac{n}{n+1}P_{n-1}(x).$$

**Primjedba 5.4** Uz pretpostavku da je poznat ortogonalan sustav  $\psi_0, \psi_1, \dots, \psi_n$  na intervalu  $[-1, 1]$ , jednostavnom transformacijom

$$\gamma : [a, b] \longrightarrow [-1, 1], \quad \gamma(x) = \frac{2x}{b-a} - \frac{a+b}{b-a}, \quad (5.14)$$

možemo dobiti sustav funkcija

$$\tilde{\psi}_i(x) = \psi_i(\gamma(x)), \quad i = 0, 1, \dots, n,$$

koji je ortogonalan na  $[a, b]$ .

**Zadatak 5.4** Koristeći Legendreove polinome odrediti najbolju  $L_2$  aproksimaciju funkcije  $f(x) = e^{2x}$ ,  $x \in [0, 1]$  na potprostoru polinoma stupnja  $\leq 1$ . Rezultate usporedite s Primjerom 5.1.

Uputa: Koristeći transformaciju (5.14) na Legendreovim polinomima  $P_0$  i  $P_1$  dobivamo:  $\tilde{\psi}_0(x) = 1$ ,  $\tilde{\psi}_1(x) = 2x - 1$ , pa je  $f^* = a_0^* \cdot 1 + a_1^*(2x - 1)$ , gdje se  $a_0^*$  i  $a_1^*$  računaju prema (5.5).

### 5.1.2 Čebiševljevi polinomi

Sustav ortogonalnih polinoma na  $[-1, 1]$  s težinskom funkcijom  $\omega(x) = \frac{1}{\sqrt{1-x^2}}$  nazivaju se Čebiševljevi polinomi<sup>1</sup>. Mogu se definirati pomoću eksplicitne formule

$$T_n(x) = \cos(n \arccos x), \quad n = 0, 1, \dots \quad (5.15)$$

Korištenjem jednostavne relacije

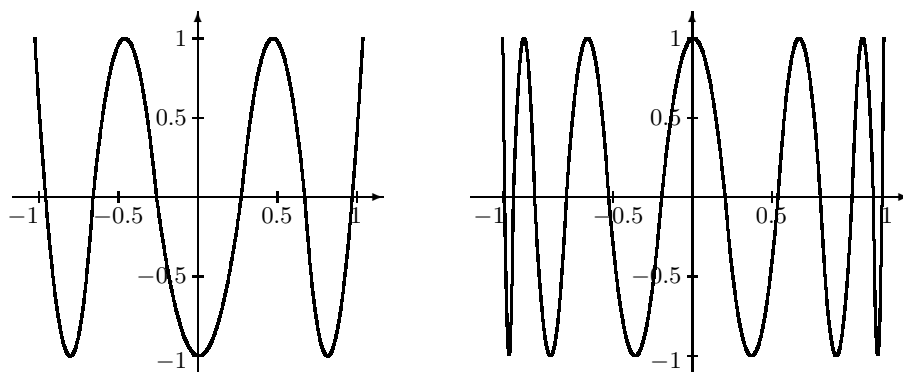
$$\cos(n+1)\varphi + \cos(n-1)\varphi = 2\cos\varphi \cos n\varphi, \quad n \in \mathbb{N},$$

induktivno zaključujemo da su  $T_n$  polinomi stupnja  $n$ . Nadalje, koristeći supstituciju  $\varphi = \arccos x$ , dobivamo

$$\int_{-1}^1 T_i(x)T_j(x) \frac{dx}{\sqrt{1-x^2}} = \int_{\pi}^0 \cos i\varphi \cos j\varphi d\varphi = \frac{1}{2} \int_{-\pi}^{\pi} \cos i\varphi \cos j\varphi d\varphi = \frac{\pi}{2} \delta_{ij}.$$

<sup>1</sup>Oznaka polinoma "T" dolazi od engleske transkripcije prezimena Čebišev: Tchebycheff

odakle neposredno slijedi ortogonalnost Čebiševljevih polinoma (vidi *Primjer 5.2*, str. 95)



Slika 5.3 Čebiševljevi polinomi  $T_6$  i  $T_{12}$

Čebiševljevi polinomi imaju sljedeća svojstva:

- a)  $|T_n(x)| \leq 1$ ,  $x \in [-1, 1]$ ,  $n = 0, 1, \dots$ ;
- b) Čebiševljev polinom  $T_n$  ima na intervalu  $[-1, 1]$   $n$  različitih nultočaka

$$x_k = \cos\left(\frac{2k-1}{n} \frac{\pi}{2}\right), \quad k = 1, \dots, n;$$

- c) Čebiševljev polinom  $T_n$  na intervalu  $[-1, 1]$  ima  $n+1$  različitih točaka

$$\xi_k = \cos \frac{k\pi}{n}, \quad k = 0, 1, \dots, n,$$

u kojima naizmjenično postiže globalne minimume i maksimume;

- d)  $T_0(x) = 1$ ,  $T_1(x) = x$ , a za  $n = 1, 2, \dots$  vrijedi rekursivna formula:

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \quad n = 1, 2, \dots$$

**Zadatak 5.5** Izračunajte  $\|T_n\|$ ,  $n = 0, 1, \dots$ , te odgovarajuće koeficijente  $a_k^*$ ,  $k = 0, 1, \dots$ , najbolje  $L_2$  aproksimacije funkcije.

Rezultat:  $\|T_0\| = \sqrt{\pi}$ ,  $\|T_k\| = \sqrt{\frac{\pi}{2}}$ ,  $k = 0, 1, \dots$ ,  $a_k^* = \frac{2}{\pi} \int_0^\pi f(\varphi) \cos k\varphi d\varphi$ ,  $k = 0, 1, \dots$

**Zadatak 5.6** Nultočke, kao i točke u kojima Čebiševljev polinom naizmjenično postiže globalne minimume i maksimume nisu jednoliko razmještene na intervalu  $[-1, 1]$ . One su gušće raspoređene pri rubovima intervala, a rjeđe oko središta. Označite točke  $x_k, \xi_k$ ,  $k = 0, 1, \dots, n$  za  $n = 10$ , tako da na trigonometrijskoj kružnici najprije označite točke koje odgovaraju realnim brojevima  $\frac{2k-1}{n} \frac{\pi}{2}$  i  $\frac{k\pi}{n}$ ,  $k = 0, 1, \dots, n$ .

## 5.2 Najbolja $L_\infty$ aproksimacija

Pretpostavimo da je  $f \in C_{[a,b]}$  i da je  $\varphi_0, \varphi_1, \dots, \varphi_n$  dani sustav funkcija. Tražimo vektor parametara  $a^* = (a_0^*, \dots, a_n^*)^T \in \mathbb{R}^{n+1}$  koji će minimizirati funkciju

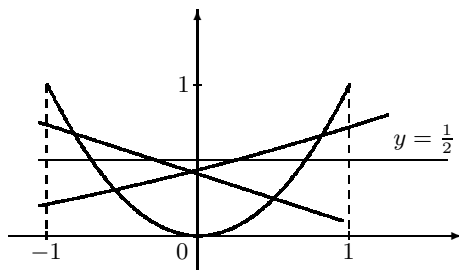
$$F(a_0, \dots, a_n) = \left\| \sum_{i=0}^n a_i \varphi_i - f \right\|_\infty = \max_{a \leq x \leq b} \left| \sum_{i=0}^n a_i \varphi_i(x) - f(x) \right|. \quad (5.16)$$

Funkcija  $f^* = \sum_{i=0}^n a_i^* \varphi_i$  je najbolja  $L_\infty$  aproksimacija funkcije  $f$  na  $L(\varphi_0, \varphi_1, \dots, \varphi_n)$ .

Mi ćemo se zadržati u okvirima jednog specijalnog slučaja: tražit ćemo najbolju  $L_\infty$  aproksimaciju funkcije  $f$  na prostoru polinoma stupnja  $\leq n$ , tj. na prostoru razapetom funkcijama  $\varphi_i(x) = x^i$ ,  $i = 0, 1, \dots, n$ .

**Primjer 5.4** Za danu funkciju  $f(x) = x^2$ ,  $x \in [-1, 1]$  treba pronaći najbolju  $L_\infty$  aproksimaciju na prostoru polinoma stupnja  $\leq 1$ .

Treba pronaći parametre  $a_0^*, a_1^* \in \mathbb{R}$ , tako da bude  $F(a_0, a_1) = \max_{-1 \leq x \leq 1} |x^2 - (a_0 + a_1 x)|$  minimalno. Budući da funkcija  $F$  nije derivabilna na  $[-1, 1]$ , ne možemo upotrijebiti svoje predznanje o istraživanju ekstrema derivabilnih funkcija.



Slika 5.4 Linearna  $L_\infty$  aproksimacija kvadratne funkcije  $f(x) = x^2$

Pokazat će se (vidi *Teorem 5.1*) da se najbolja aproksimacija dobije za  $a_0^* = \frac{1}{2}$ ,  $a_1^* = 0$ , tj. da je najbolja  $L_\infty$  aproksimacija kvadratne funkcije  $f(x) = x^2$  u klasi linearnih funkcija, funkcija  $f^*(x) = \frac{1}{2}$ . Primijetite (vidi *Sliku 5.4*) da su u tom slučaju maksimalna odstupanja u točkama  $-1, 0, 1$  po apsolutnoj vrijednosti jednaka  $\frac{1}{2}$ , a da po predznaku alterniraju.

Pokušajmo riješiti općenitiji problem:

za funkciju  $f(x) = x^n$ ,  $x \in [-1, 1]$  treba pronaći najbolju  $L_\infty$  aproksimaciju na prostoru  $\mathcal{P}_{n-1}$  polinoma stupnja  $\leq n-1$ .

Odgovor će dati sljedeći teorem.

**Teorem 5.1** *Između svih normiranih polinoma<sup>2</sup> stupnja  $\leq n$  polinom  $2^{1-n}T_n$ , gdje je  $T_n$   $n$ -ti Čebiševljev polinom, ima najmanju  $L_\infty$  normu na intervalu  $[-1, 1]$ , koja iznosi  $2^{1-n}$ .*

**Dokaz.** Uočimo najprije da je koeficijent uz najvišu potenciju Čebiševljevog polinoma  $T_n$  jednak  $2^{n-1}$  (uspredi svojstvo d) Čebiševljevog polinoma, str. 99). Zato je polinom  $2^{1-n}T_n$  normiran i vrijedi:  $|2^{1-n}T_n(x)| \leq 2^{1-n}$  (svojstvo a), str. 99). Pretpostavimo sada da postoji neki drugi normirani polinom  $p_n$ , takav da je  $|p_n(x)| < 2^{1-n}$  za sve  $x \in [-1, 1]$ . Neka su nadalje,  $\xi_0, \xi_1, \dots, \xi_n$  točke u kojima Čebiševljev polinom  $T_n$  prima ekstremne vrijednosti  $-1$  ili  $1$  (vidi t.5.1.2, str. 98). Tada vrijedi

$$\begin{aligned} p_n(\xi_0) &< 2^{1-n}T_n(\xi_0) = 2^{1-n} \\ p_n(\xi_1) &> 2^{1-n}T_n(\xi_1) = -2^{1-n} \\ p_n(\xi_2) &< 2^{1-n}T_n(\xi_2) = 2^{1-n} \quad \text{itd.} \end{aligned}$$

To bi značilo da “polinom geške”  $E(x) := p_n(x) - 2^{1-n}T_n(x)$ ,  $n$  puta mijenja predznak na intervalu  $[-1, 1]$ . To međutim nije moguće jer je  $E$  polinom stupnja  $(n-1)$  (polinomi  $p_n$  i  $2^{1-n}T_n$  su normirani!)  $\square$

Ovo važno svojstvo Čebiševljevih polinoma već smo iskoristili u t.2.1.3 prilikom optimalnog izbora čvorova interpolacije funkcije (str. 25). Sada se također vidi i zašto je rješenje zadatka u *Primjeru 5.4* bila baš funkcija  $x \mapsto \frac{1}{2}$ . Naime, prema upravo dokazanom teoremu, vrijedi

$$\min_{a_0, a_1} \|x^2 - (a_0 + a_1x)\|_\infty = \left\| \frac{1}{2}T_2 \right\|_\infty,$$

gdje je  $T_2(x) = 2x^2 - 1$ .

**Zadatak 5.7** *Odredite najbolju  $L_\infty$  aproksimaciju funkcije  $f(x) = x^n$ ,  $x \in [-1, 1]$  na prostoru polinoma stupnja  $\leq n-1$ .*

**Odgovor:** Prema *Teoremu 5.1* najbolja aproksimacija je funkcija  $f^*(x) = x^n - 2^{1-n}T_n(x)$ .

---

<sup>2</sup>Kažemo da je polinom normiran ako mu je koeficijent uz najvišu potenciju jednak 1

Prethodne tvrdnje mogu se generalizirati kroz sljedeći važni teorem.

**Teorem 5.2** (de la Vallée Poussin) *Neka je  $f \in C_{[a,b]}$ . Polinom  $p^*$  stupnja  $\leq n$  je najbolja  $L_\infty$  aproksimacija funkcije  $f$  na  $[a, b]$  onda i samo onda ako postoje barem  $n + 2$  točke,*

$$a \leq x_0 < x_1 < \cdots < x_{n+1} \leq b$$

*u kojima funkcija pogreške  $E = p_n^* - f$  prima maksimalne vrijednosti s alternirajućim predznakom, tj.*

$$E(x_{i+1}) = -E(x_i), \quad i = 0, 1, \dots, n, \quad |E(x_i)| = \|E\|_\infty.$$

**Zadatak 5.8** *Primjenom de la Vallée Poussin teorema odredite najbolju  $L_\infty$  aproksimaciju funkcije  $f(x) = \sin x$  na  $[0, \frac{\pi}{2}]$  u klasi polinoma prvog stupnja.*

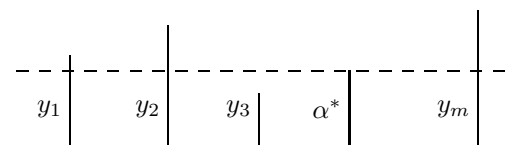
Rješenje:  $p_1^*(x) = 0.105 + \frac{2t}{\pi}$ .

## 6. Problemi najmanjih kvadrata

Problem najbolje  $L_2$ -aproksimacije funkcije koja je zadana na konačnom skupu točaka obično se u literaturi naziva **problem najmanjih kvadrata**<sup>1</sup>.

Najprije ćemo na nekoliko tipičnih primjera pokazati neke probleme iz primjena, koji se rješavaju metodom najmanjih kvadrata. Još neke druge primjene mogu se vidjeti u SCITOVSKI (1993).

**Primjer 6.1** *Pretpostavimo da su  $y_1, y_2, \dots, y_m$  rezultati mjerenja neke veličine  $A$ . Treba odrediti aproksimaciju  $\alpha^*$  veličine  $A$  tako da sve izmjerene vrijednosti  $y_1, y_2, \dots, y_m$  budu "što bliže" aproksimaciji  $\alpha^*$*



Slika 6.1. Aritmetička sredina podataka

Pri tome pojam "što bliže" shvaćamo u smislu zahtjeva da suma kvadrata odstupanja svih mjerenja od aproksimacije  $\alpha$  bude minimalna, tj. da bude

$$F_2(\alpha) = \sum_{i=1}^m (\alpha - y_i)^2 \longrightarrow \min. \quad (*)$$

Ovaj princip nazivamo princip najmanjih kvadrata ili češće **metoda najmanjih kvadrata**, a problem određivanja aproksimacije  $\alpha^*$  na osnovi principa najmanjih kvadrata, nazivamo **problem najmanjih kvadrata**. Problem najmanjih kvadrata je specijalni problem ekstrema bez ograničenja. Lako se vidi da je rješenje problema (\*) aritmetička sredina podataka

$$\alpha^* = \frac{1}{m} \sum_{i=1}^m y_i$$

---

<sup>1</sup>Princip najmanjih kvadrata postavio je Carl Friedrich Gauss 1795. godine prilikom izučavanja kretanja nebeskih tijela, što je objavio 1809 godine u radu "*Teoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*", (Perthes and Besser, Hamburg). Treba također reći da je francuski matematičar Legendre 1805. prvi publicirao algebarsku proceduru metode najmanjih kvadrata

Dakle, aritmetička sredina mjerenja  $y_1, \dots, y_m$  je takva veličina koja ima svojstvo da je suma kvadrata odstupanja te veličine od mjerenja  $y_1, \dots, y_m$  najmanja.

**Primjer 6.2** *Zadani su podaci*

$x_i$	-4	-2	2	4
$y_i$	8	6	3	5

Treba pronaći linearnu funkciju  $f(x) = a_1 + a_2x$  tako da njezin graf prolazi što bliže danim točkama  $T_i(x_i, y_i)$ ,  $i = 1, 2, 3, 4$  (vidi Sliku 6.2.a).

Jasno je da bi nam već prve dvije točke  $T_1$  i  $T_2$  bile dovoljne za određivanje tražene funkcije. Ali što je onda s preostalim točkama – podacima? Zato ćemo pokušati parametre  $a_1, a_2$ , odrediti tako da bude

$$y_i \approx a_1 + a_2x_i, \quad i = 1, \dots, 4, \quad (*)$$

pri čemu će nam “ $\approx$ ” u (\*) značiti minimalno kvadratno odstupanje. Drugim riječima, definirat ćemo odstupanja

$$r_i = a_1 + a_2x_i - y_i, \quad i = 1, \dots, 4$$

i pokušati riješiti problem ekstrema

$$F(a_1, a_2) = \sum_{i=1}^n (a_1 + a_2x_i - y_i)^2 = \sum_{i=1}^n r_i^2 = \|\mathbf{r}\|_2^2 \rightarrow \min, \quad (**)$$

gdje je  $\mathbf{r}$  vektor s komponentama  $r_i$ . Odgovarajući problem (\*\*) također je jedan problem najmanjih kvadrata.

**Primjedba 6.1** Problem (\*\*) često zovemo i problem diskretne  $L_2$  aproksimacije. Problem bismo mogli postaviti i tako da umjesto (\*\*) promatramo problem

$$F_1(a_1, a_2) = \sum_{i=1}^n |a_1 + a_2x_i - y_i| = \|\mathbf{r}\|_1 \rightarrow \min, \quad (***)$$

ili

$$F_\infty(a_1, a_2) = \max_{i=1, \dots, m} |a_1 + a_2x_i - y_i| = \|\mathbf{r}\|_\infty \rightarrow \min. \quad (****)$$

Problem (\*\*\* ) naziva se problem najbolje  $L_1$  aproksimacije, a problem (\*\*\*\*) problem najbolje  $L_\infty$  aproksimacije. Ocem  $L_1$  aproksimacija smatra se hrvatski znanstvenik R. Bošković, a ocem  $L_\infty$  aproksimacija ruski matematičar P. L. Čebišev. Inače funkcije  $F_1$  i  $F_\infty$  nisu derivabilne, pa su zato problemi (\*\*\*\*) i (\*\*\*\*\*) znatno složeniji i zbog toga ih nećemo dalje promatrati. Više o  $L_1$  i  $L_\infty$  aproksimacijama može se vidjeti kod RICE (1964, 1969).

**Primjer 6.3** Promatramo sustav linearnih jednadžbi  $\mathbf{J}\mathbf{a} = \mathbf{y}$ , gdje je

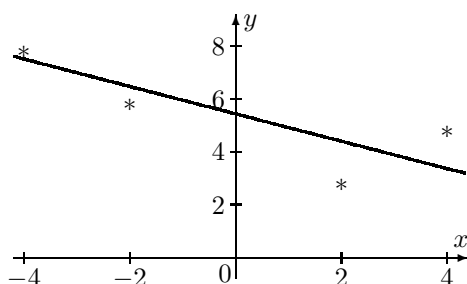
$$\mathbf{J} = \begin{bmatrix} 1 & -4 \\ 1 & -2 \\ 1 & 2 \\ 1 & 4 \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 8 \\ 6 \\ 3 \\ 5 \end{bmatrix}.$$

Ovaj sustav nije rješiv jer vektor  $\mathbf{y}$  ne leži u potprostoru  $L(\mathbf{j}_1, \mathbf{j}_2)$ , gdje su  $\mathbf{j}_1, \mathbf{j}_2$  stupci matrice  $\mathbf{J}$ . Kažemo da se radi o **preodređenom sustavu**. Potražiti ćemo najbolju  $L_2$  aproksimaciju vektora  $\mathbf{y}$  u potprostoru  $L(\mathbf{j}_1, \mathbf{j}_2)$ , tj. potražiti ćemo parametre  $a_1, a_2$ , tako da bude

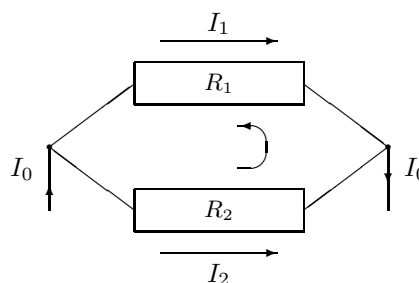
$$F(a_1, a_2) = \|\mathbf{J}\mathbf{a} - \mathbf{y}\|_2^2 = \sum_{i=1}^n (a_1 + a_2 x_i - y_i)^2 \rightarrow \min.$$

Primijetite da će za optimalne vrijednosti parametara  $a_1 = 5.5$ ,  $a_2 = -0.45$  suma kvadrata odstupanja lijevih od desnih strana jednadžbi biti minimalna. I u ovom slučaju radi se o jednom problemu najmanjih kvadrata.

Iz navedenog primjera jasno je da i problemu rješavanja sustava linearnih jednadžbi također možemo pristupiti kao jednom problemu najmanjih kvadrata.



Slika 6.2.a) Linearna regresija



Slika 6.2.b) Električna mreža

**Primjer 6.4** U električnoj mreži bez izvora, sastavljenoj od otpora  $R_k$ , struje  $I_k$  izračunavaju se pomoću Kirchhoffovih zakona:

“Pravilo čvora”: Za svaki čvor mreže vrijedi  $\sum_k I_k = 0$ . Pri tome, dolazeće struje smatraju se pozitivnim, a odlazeće negativnim;

“Pravilo petlje”: Za svaku zatvorenu petlju (konturu) vrijedi  $\sum_k U_k = \sum_k I_k R_k = 0$ .

Pri tome, ako konturu obilazimo u nekom smjeru, onda se struje u tom smjeru smatraju pozitivnim, a struje u suprotnom smjeru, negativnim.

Za primjer na Slici 6.2.b i zadane otpore  $R_1, R_2$  i ulaznu struju  $I_0$  treba izračunati struje  $I_1$  i  $I_2$ . Imamo

$$\begin{array}{rrrr} I_0 & -I_1 & -I_2 & = 0 \\ -I_0 & +I_1 & +I_2 & = 0 \\ & -I_1 R_1 & +I_2 R_2 & = 0 \end{array}$$

Na taj način obično dobivamo više jednadžbi nego nepoznanica, a zbog prisutnosti pogrešaka u mjerenjima, takav sustav je često preodređen i nema rješenje, pa se obično zadovoljavamo najboljom  $L_2$  aproksimacijom rješenja.

**Primjer 6.5** SOS - poziv u pomoć s nekog broda primiljen je na  $m$  ( $m \geq 2$ ) radio stanica. Poznate su koordinate položaja tih radiostanica:  $S_i(x_i, y_i), i = 1, \dots, m$ , kao i kutevi:  $\alpha_i, i = 1, \dots, m$ , pod kojima su stanice primile poziv s broda (vidi



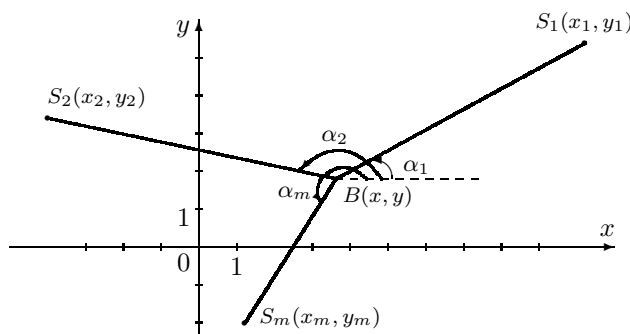
Sliku 6.3.) Treba odrediti koordinate  $x, y$  položaja broda koji šalje poziv u pomoć. Zbog jednostavnosti nećemo uzimati u obzir zakrivljenost Zemljine površine.

Očigledno vrijedi

$$\alpha_i = \arctg \left( \frac{y - y_i}{x - x_i} \right), \quad i = 1, \dots, m. \quad (*)$$

Kada bi svi izmjereni kutevi bili potpuno točni, sustav (\*) imao bi jedinstveno rješenje  $(x, y)$ . Budući da se to u praksi ne može očekivati, onda ćemo rješenje sustava (\*) potražiti u smislu najmanjih kvadrata

$$F(x, y) = \sum_{i=1}^m \left[ \arctg \left( \frac{y - y_i}{x - x_i} \right) - \alpha_i \right]^2 \rightarrow \min.$$



Slika 6.3. Traženje pozicije broda

Pretpostavimo općenito da zavisna varijabla  $y$  ovisi o nezavisnoj varijabli  $x$  po funkcionalnom zakonu

$$y = f(x; \mathbf{a}), \quad \mathbf{a} \in \mathbb{R}^n,$$

gdje je  $\mathbf{a} = (a_1, \dots, a_n)^T$  vektor parametar. Osim toga, neka su zadani podaci

$$(x_i, y_i), \quad i = 1, \dots, m, \quad m \geq n.$$

Nadalje, neka je  $\mathbf{r}$  vektor odstupanja (reziduala) s komponentama

$$r_i(\mathbf{a}) = f(x_i; \mathbf{a}) - y_i. \quad (6.1)$$

Treba pronaći točku  $\mathbf{a}^* = (a_1^*, \dots, a_n^*) \in \mathbb{R}^n$ , u kojoj funkcija  $F : \mathbb{R}^n \rightarrow \mathbb{R}$ , definirana s

$$F(\mathbf{a}) = \frac{1}{2} \sum_{i=1}^m r_i^2(\mathbf{a}) = \frac{1}{2} \|\mathbf{r}(\mathbf{a})\|_2^2, \quad \mathbf{a} \in \mathbb{R}^n, \quad (6.2)$$

postiže globalni minimum.

Funkciju  $f$  nazivamo funkcija-model, vektor  $\mathbf{r}$  vektor odstupanja ili reziduala, a brojeve  $a_1^*, \dots, a_n^*$ , optimalni parametri. Problem određivanja optimalnih parametara  $a_1^*, \dots, a_n^*$  nazivamo problem najmanjih kvadrata (u daljnjem tekstu: PNK).

$$(\mathbf{H}_k)_{ij} = \frac{\partial^2 \mathbf{r}_k}{\partial a_i \partial a_j}, \quad (6.6)$$

Ako želimo pronaći kritične točke minimizirajuće funkcije (6.1), odnosno (6.3), potrebno je gradijent funkcije  $F$  izjednačiti s nulom. Tako dobivamo sustav od  $n$  jednadžbi s  $n$  nepoznanica  $a_1, \dots, a_n$

$$\mathbf{J}^T \mathbf{r} = \mathbf{0} \quad \text{ili} \quad \begin{array}{c} \sum_{i=1}^m r_i \frac{\partial r_i}{\partial a_1} = 0 \\ \vdots \\ \sum_{i=1}^m r_i \frac{\partial r_i}{\partial a_n} = 0 \end{array} \quad (6.7)$$

Ovaj sustav jednadžbi bit će linearan ako su sve funkcije  $r_i = r_i(a_1, \dots, a_n)$  linearne u svakom od argumenata  $a_1, \dots, a_n$ . To će se dogoditi onda ako je funkcija-model  $f$  linearna u svim parametrima  $a_1, \dots, a_n$ . Tada govorimo o linearnom problemu najmanjih kvadrata (LPNK).

U suprotnom, ako je funkcija-model  $f$  nelinearna u barem jednom od parametara:  $a_1, \dots, a_n$ , govorimo o nelinearnom problemu najmanjih kvadrata (NPNK). U tom slučaju je (6.7) sustav nelinearnih jednadžbi.

Primijetimo da su *Primjeri 6.1, 6.2. i 6.3.* linearni problemi, a *Primjer 6.4.* nelinearni problem najmanjih kvadrata.

## 6.1 Linearni problemi najmanjih kvadrata

Pretpostavimo da je funkcija-model  $f$  linearna u svim parametrima  $a_1, \dots, a_n$  i da je općenito oblika

$$f(x; a_1, \dots, a_n) = a_1 \varphi_1(x) + \dots + a_n \varphi_n(x), \quad x \in \mathbb{R} \quad (6.8)$$

gdje su  $\varphi_i$  neprekidne funkcije. Primjerice za polinom-model, funkcije  $\varphi_i$  su potencije:  $\varphi_i(x) = x^{i-1}$ ,  $i = 1, \dots, n$ .

Jacobijeva matrica za funkciju-model (6.8) ne ovisi o  $\mathbf{a}$ . Naime, kako u ovom slučaju (6.1) glasi

$$r_i(\mathbf{a}) = a_1 \varphi_1(x) + \dots + a_n \varphi_n(x) - y_i$$

imamo

$$\mathbf{J}_{ij} = \frac{\partial r_i}{\partial a_j} = \varphi_j(x_i), \quad i = 1, \dots, m, \quad j = 1, \dots, n.$$

Hessijan funkcije  $F$  također je bitno jednostavniji i također ne ovisi o  $\mathbf{a}$

$$\mathbf{H}_F = \mathbf{J}^T \mathbf{J}. \quad (6.9)$$

Vektor odstupanja  $\mathbf{r}$  s komponentama  $r_i = a_1 \varphi_1(x_i) + \dots + a_n \varphi_n(x_i) - y_i$  možemo pisati

$$\mathbf{r} = \mathbf{J}\mathbf{a} - \mathbf{y}, \quad (6.10)$$

gdje je  $\mathbf{y} = (y_1, \dots, y_m)^T$ . Traži se  $\mathbf{a} \in \mathbb{R}^n$  tako da bude

$$F(\mathbf{a}^*) = \min_{\mathbf{a} \in \mathbb{R}^n} F(\mathbf{a}), \quad \text{gdje je } F(\mathbf{a}) = \frac{1}{2} \|\mathbf{r}(\mathbf{a})\|^2 = \frac{1}{2} \|\mathbf{J}\mathbf{a} - \mathbf{y}\|^2. \quad (6.11)$$

Često puta se u literaturi LPNK (6.11) označava kao

$$\mathbf{J}\mathbf{a} \simeq \mathbf{y}. \quad (6.12)$$

Primjerice, za podatke i funkciju-model iz *Primjera 6.2* imamo:  $\varphi_1(x) = 1$ ,  $\varphi_2(x) = x$ ,

$$\mathbf{J} = \begin{bmatrix} 1 & -4 \\ 1 & -2 \\ 1 & 2 \\ 1 & 4 \end{bmatrix}, \quad \mathbf{H}_F = \mathbf{J}^T \mathbf{J} = \begin{bmatrix} 4 & 0 \\ 0 & 40 \end{bmatrix}.$$

### 6.1.1 Rješavanje LPNK preko sustava normalnih jednadžbi

Kritične točke funkcije  $F$  dobit ćemo rješavanjem jednadžbe (6.4)

$$\text{grad } F = \mathbf{J}^T \mathbf{r} = 0. \quad (6.13)$$

Hesijan  $\mathbf{H}_F$  ne ovisi o vektoru  $\mathbf{a}$  i zadan je formulom (6.9). Sljedeća lema pokazuje kada će Hesijan  $\mathbf{H}_F$  biti pozitivno definitan

**Lema 6.1** *Neka je  $\mathbf{J} \in \mathbb{R}^{m \times n}$ ,  $m > n$ . Matrica  $\mathbf{J}^T \mathbf{J}$  je pozitivno definitna onda i samo onda ako je  $\mathbf{J}$  punog ranga po stupcima ( $\text{rang } \mathbf{J} = n$ ).*

**Dokaz.** (*Nužnost*) Neka je  $\mathbf{J}^T \mathbf{J}$  pozitivno definitna. Pretpostavimo da je  $\text{rang } \mathbf{J} < n$ , tj. da su stupci od  $\mathbf{J}$  linearno zavisni. Tada bi za neki  $\mathbf{a}_0 \neq \mathbf{0}$  imali  $\mathbf{J}\mathbf{a}_0 = \mathbf{0}$ , odakle bi slijedilo  $\mathbf{a}_0^T \mathbf{J}^T \mathbf{J} \mathbf{a}_0 = 0$ , što bi značilo da  $\mathbf{J}^T \mathbf{J}$  nije pozitivno definitna.

(*Dovoljnost*) Pretpostavimo da je  $\text{rang } \mathbf{J} = n$ , tj. da su stupci od  $\mathbf{J}$  linearno nezavisni. Tada za  $\mathbf{a} \neq \mathbf{0} \Rightarrow \mathbf{J}\mathbf{a} \neq \mathbf{0}$ , pa je

$$(\mathbf{H}_F \mathbf{a}, \mathbf{a}) = (\mathbf{J}^T \mathbf{J} \mathbf{a}, \mathbf{a}) = \mathbf{a}^T (\mathbf{J}^T \mathbf{J} \mathbf{a}) = (\mathbf{J}\mathbf{a})^T (\mathbf{J}\mathbf{a}) = \|\mathbf{J}\mathbf{a}\|^2 > 0.$$

□

Dakle, LPNK, za koji je pripadna Jacobijeva matrica punog ranga (što je u praksi gotovo uvijek ispunjeno), rješiv je i postoji jedinstveno rješenje, koje možemo dobiti tako da pronađemo kritične točke funkcije  $F$ . Zato ćemo (6.10) uvrstiti (6.13). Dobivamo jednadžbu

$$\mathbf{J}^T \mathbf{J} \mathbf{a} - \mathbf{J}^T \mathbf{y} = 0 \quad (6.14)$$

Lako se vidi da je rješenje ove jednadžbe

$$\mathbf{a}^* = \mathbf{J}^+ \mathbf{y}, \quad \text{gdje je } \mathbf{J}^+ = (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T.$$

Jednadžbu (6.14) nazivamo sustav normalnih jednadžbi, a matricu  $\mathbf{J}^+$  pseudoinverzna matrica ili Moore-Penroseov generalizirani inverz<sup>2</sup> matrice  $\mathbf{J}$ .

**Primjedba 6.3** U slučaju ako je funkcija-model (6.8) polinom  $n$ -tog stupnja, a broj podataka  $m$  jednak  $(n+1)$ , rješavanjem odgovarajućeg LPNK dobivamo interpolacijski polinom.

**Primjedba 6.4** Kako je  $\text{cond}(\mathbf{J}^T \mathbf{J}) = \text{cond}^2(\mathbf{J})$  (vidi primjerice GILL (1991), matrica  $\mathbf{J}^T \mathbf{J}$  bit će vrlo loše uvjetovana ako je matrica  $\mathbf{J}$  loše uvjetovana, a rješenje LPNK, dobiveno rješavanjem sustava normalnih jednadžbi (6.14), bit će vrlo nepouzdan.

**Primjer 6.6** Zadano je

$$\mathbf{J} = \begin{bmatrix} 1 & 1 \\ 10^{-3} & 0 \\ 0 & 10^{-3} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 2 \\ 10^{-3} \\ 10^{-3} \end{bmatrix}$$

Može se pokazati (vidi Primjedbu 3.9, str. 62) da je  $\text{cond}(\mathbf{J}) \approx 1.4 \times 10^3$  i da je rješenje odgovarajućeg LPNK  $\mathbf{a}^* = (1, 1)^T$  uz  $\|\mathbf{r}(\mathbf{a}^*)\| = 0$ .

Ako bismo ovaj LPNK rješavali preko sustava normalnih jednadžbi na računalu s 6-znamenkastom floating-point aritmetikom, matrica  $fl(\mathbf{J}^T \mathbf{J}) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$  bila bi singularna. U 7-znamenkastoj floating-point aritmetici rješenje bi bilo  $\mathbf{a}^* = (2.000001, 0)^T$ .

### 6.1.2 Rješavanje LPNK pomoću QR-dekompozicije

Minimum funkcije  $F(\mathbf{a}) = \frac{1}{2} \|\mathbf{J}\mathbf{a} - \mathbf{y}\|^2$  iz (6.11) potražiti ćemo na sljedeći način. Kako je  $\mathbf{J} \in \mathbb{R}^{m \times n}$ , najprije ćemo algoritmom opisanim u t.3.7, str. 52, načiniti dekompoziciju  $\mathbf{J} = \mathbf{Q}\mathbf{R}$ , gdje je  $\mathbf{Q} \in \mathbb{R}^{m \times m}$  ortogonalna, a  $\mathbf{R} \in \mathbb{R}^{m \times n}$  gornja trokutasta matrica. Vektor odstupanja  $\mathbf{r} = \mathbf{J}\mathbf{a} - \mathbf{y}$ , sada možemo ovako pisati

$$\mathbf{r} = \mathbf{Q}\mathbf{R}\mathbf{a} - \mathbf{y} \quad (6.15)$$

Budući da ortogonalna matrica čuva normu, vrijedi

$$\|\mathbf{r}\| = \|\mathbf{Q}\mathbf{R}\mathbf{a} - \mathbf{y}\| = \|\mathbf{Q}(\mathbf{R}\mathbf{a} - \mathbf{Q}^T \mathbf{y})\| = \|\mathbf{R}\mathbf{a} - \mathbf{Q}^T \mathbf{y}\|, \quad (6.16)$$

pa je

$$\begin{aligned} 2F(\mathbf{a}) = \|\mathbf{r}\|^2 = \|\mathbf{R}\mathbf{a} - \mathbf{Q}^T \mathbf{y}\|^2 &= \|\mathbf{R}_n \mathbf{a} - (\mathbf{Q}^T \mathbf{y})_n\|_{(n)}^2 + \|(\mathbf{Q}^T \mathbf{y})_{m-n}\|_{(m-n)}^2 \\ &\geq \|(\mathbf{Q}^T \mathbf{y})_{m-n}\|_{(m-n)}^2, \end{aligned}$$

gdje je  $\mathbf{R}_n$  matrica sastavljena od prvih  $n$  redaka matrice  $\mathbf{R}$ , a  $(\mathbf{Q}^T \mathbf{y})_n$ , odnosno  $(\mathbf{Q}^T \mathbf{y})_{m-n}$  vektori sastavljeni od prvih  $n$ , odnosno posljednjih  $m-n$  komponenti

<sup>2</sup>Više o svojstvima i značenju Moore-Penroseovog generaliziranog inverza može se vidjeti primjerice u GILL (1991), GOLUB (1989) ili LAWSON (1974)

originalnih vektora. Indeksi  $(n)$  i  $(m - n)$  označavaju odgovarajuće vektorske  $L_2$  norme u  $\mathbb{R}^n$ , odnosno  $\mathbb{R}^{m-n}$ . Zato se minimum funkcije  $F$  postiže na vektoru  $\mathbf{a}^* \in \mathbb{R}^n$ , za koji je

$$\mathbf{R}_n \mathbf{a}^* = (\mathbf{Q}^T \mathbf{y})_n. \quad (6.17)$$

pri čemu je

$$F(\mathbf{a}^*) = \frac{1}{2} \|\mathbf{r}(\mathbf{a}^*)\|^2 = \frac{1}{2} \|(\mathbf{Q}^T \mathbf{y})_{m-n}\|_{(m-n)}^2. \quad (6.18)$$

**Primjer 6.7** LPNK iz Primjera 6.2

$$\mathbf{J} \mathbf{a} = \begin{bmatrix} 1 & 4 \\ 1 & 2 \\ 1 & -2 \\ 1 & -4 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \simeq \mathbf{y} = \begin{bmatrix} 8 \\ 6 \\ 3 \\ 5 \end{bmatrix}.$$

riješit ćemo primjenom QR-dekompozicije.

Primjenom *Mathematica*-naredbe `QRDecomposition[J]` (ili odgovarajućeg BASIC-programa u *Dodatku*) dobivamo sustav (6.17)

$$\begin{bmatrix} -2. & 0. \\ 0 & 6.32456 \end{bmatrix} \begin{bmatrix} a_1^* \\ a_2^* \end{bmatrix} = \begin{bmatrix} -11. \\ -2.84605 \end{bmatrix},$$

odakle dobivamo  $\mathbf{a}^* = (5.5, -0.45)^T$  i  $F(\mathbf{a}^*) = 1.1068$ .

**Primjer 6.8** U Primjeru 2.3, str. 20, pokazalo se da su posljednja dva koeficijenta u interpolacijskom polinomu zanemarivo malena, pa ima smisla potražiti polinom stupnja 2, koji će u smislu najmanjih kvadrata najbolje aproksimirati zadane podatke i dati funkcionalnu zavisnost  $\alpha = P_2(X)$ .

Primjenom QR-dekompozicije za rješavanje odgovarajućeg LPNK dobivamo

$$a_0 = 5.0467, a_1 = -0.0033, a_2 = 0.000023.$$

Usporedite rezultate prikazane u *Tablici 6.1* s *Tablicom 3.2*, str. 20.

$X(m)$	$\alpha$	$\alpha^*$	$\Delta\alpha^*$
500	10.885	9.14	1.745
750	16.901	15.49	1.411
1000	23.942	24.71	0.768

Tablica 6.1.

### 6.1.3 Rješavanje LPNK pomoću dekompozicije na singularne vrijednosti

Ako Jacobijeva matrica  $\mathbf{J}$  nije punog ranga, prema *Lemi 6.1*, str. 109, Hesijan  $H_F$  odgovarajućeg LPNK  $\mathbf{J} \mathbf{a} \simeq \mathbf{y}$  nije pozitivno definitan. LPNK u tom slučaju ima beskonačno mnogo rješenja. Od svih tih rješenja potražiti ćemo ono koje ima najmanju normu. U tom slučaju i u slučaju kada je Jacobijan  $\mathbf{J}$  loše uvjetovana matrica koristit ćemo rastav matrice na singularne vrijednosti (SVD).

Promatrajmo ponovo problem minimizacije (6.11) funkcije

$$F(\mathbf{a}) = \frac{1}{2} \|\mathbf{r}(\mathbf{a})\|^2, \quad \text{gdje je } \mathbf{r} = \mathbf{J}\mathbf{a} - \mathbf{y},$$

s time da je  $\text{rang}(\mathbf{J}) = k \leq n < m$ . Neka je  $\mathbf{J} = \mathbf{U}\mathbf{S}\mathbf{V}^T$  dekompozicije matrica  $\mathbf{J}$  na singularne vrijednosti, gdje su  $\mathbf{U}$ ,  $\mathbf{S}$  i  $\mathbf{V}$  matrice definirane u t.3.9, str.60. Budući da ortogonalna matrica čuva normu, vrijedi

$$\|\mathbf{r}\| = \|\mathbf{J}\mathbf{a} - \mathbf{y}\| = \|\mathbf{U}\mathbf{S}\mathbf{V}^T\mathbf{a} - \mathbf{y}\| = \|\mathbf{S}\mathbf{V}^T\mathbf{a} - \mathbf{U}^T\mathbf{y}\|.$$

Zato, uz oznaku  $\mathbf{z} = \mathbf{V}^T\mathbf{a}$ , imamo

$$\begin{aligned} 2F(\mathbf{a}) = \|\mathbf{r}\|^2 = \|\mathbf{S}\mathbf{z} - \mathbf{U}^T\mathbf{y}\|^2 &= \|\mathbf{S}_k\mathbf{z} - (\mathbf{U}^T\mathbf{y})_k\|_{(k)}^2 + \|(\mathbf{U}^T\mathbf{y})_{m-k}\|_{(m-k)}^2 \\ &\geq \|(\mathbf{U}^T\mathbf{y})_{m-k}\|_{m-k}^2, \end{aligned}$$

gdje je  $\mathbf{S}_k$  matrica sastavljena od prvih  $k$  redaka dijagonalne matrice  $\mathbf{S}$ , a  $(\mathbf{U}^T\mathbf{y})_k$ , odnosno  $(\mathbf{U}^T\mathbf{y})_{m-k}$  vektori sastavljeni od prvih  $k$ , odnosno posljednjih  $m-k$  komponenti originalnih vektora. Indeksi  $(k)$  i  $(m-k)$  označavaju odgovarajuće vektorske  $L_2$  norme u  $\mathbb{R}^k$ , odnosno  $\mathbb{R}^{m-k}$ . Zato se minimum funkcije  $F$  postiže na vektoru  $\mathbf{z}^* \in \mathbb{R}^n$ , za koji je

$$\mathbf{S}_k\mathbf{z}^* = (\mathbf{U}^T\mathbf{y})_k. \quad (6.19)$$

Na taj način određeno je prvih  $k$  komponenti vektora  $\mathbf{z}^*$ ,

$$z_i^* = \frac{u_i^T \mathbf{y}}{\sigma_i}, \quad i = 1, \dots, k,$$

gdje su  $\mathbf{u}_1, \dots, \mathbf{u}_k$  prvih  $k$  stupaca matrice  $\mathbf{U}$ . Preostalih  $(n-k)$  komponenti  $z_{k+1}^*, \dots, z_n^*$  vektora  $\mathbf{z}^*$  biramo tako da rješenje  $\mathbf{a}^*$  ima minimalnu Euklidovu normu. Dakle,

$$\mathbf{a}^* = \mathbf{V}\mathbf{z}^* = \sum_{i=1}^k \frac{u_i^T \mathbf{y}}{\sigma_i} \mathbf{v}_i, \quad \|\mathbf{a}^*\| = \min_{\mathbf{J}\mathbf{a} - \mathbf{y} = \mathbf{r}} \|\mathbf{a}\|, \quad (6.20)$$

gdje su  $\mathbf{v}_1, \dots, \mathbf{v}_k$  prvih  $k$  stupaca matrice  $\mathbf{V}$ . Minimalna vrijednost funkcije  $F$  je

$$F(\mathbf{a}^*) = \frac{1}{2} \|(\mathbf{U}^T\mathbf{y})_{m-k}\|_{(m-k)}^2 = \sum_{i=k+1}^m (\mathbf{u}_i^T \mathbf{y})^2. \quad (6.21)$$

**Primjer 6.9** Promatramo LPNK  $\mathbf{J}\mathbf{a} \simeq \mathbf{y}$ , gdje je

$$\mathbf{J} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Budući da je  $\text{rang} \mathbf{J} = k = 1$ , LPNK ima beskonačno mnogo rješenja. SVD ove matrice glasi

$$\mathbf{J} = \begin{bmatrix} -.707107 & -.707107 \\ -.707107 & .707107 \end{bmatrix} \cdot \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} -.707107 & -.707107 \\ -.707107 & .707107 \end{bmatrix}.$$

Primjenom formule (6.20) dobivamo rješenje (koje ima minimalnu  $L_2$  normu!)

$$\mathbf{a}^* = -\frac{1}{2} \left( (-.707107, -.707107) \cdot (1, 2)^T \right) \cdot (-.707107, -.707107)^T = (.75, .75)^T.$$

Isti rezultat možemo dobiti i izravno, tražeći minimum funkcije

$$F(a_1, a_2) = \frac{1}{2}(a_1 + a_2 - 1)^2 + \frac{1}{2}(a_1 + a_2 - 2)^2$$

Lako se vidi da ova funkcija postiže globalni minimum  $F^* = 0.25$  na beskonačno mnogo vektora  $\mathbf{a}^* = (a_1^*, a_2^*)^T$ , čije komponente  $a_1, a_2$  zadovoljavaju jednadžbu pravca  $\eta = \frac{3}{2} - \xi$ . Među njima vektor najmanje  $L_2$  norme dobit ćemo tako da potražimo minimum funkcije

$$\xi \mapsto \xi^2 + \left( \frac{3}{2} - \xi \right)^2.$$

On se postiže za  $\xi = \frac{3}{4}$ . Dakle, vektor najmanje  $L_2$  norme je  $\mathbf{a}^* = (0.75, 0.75)^T$  i to je rješenje postavljenog LPNK.

**Primjedba 6.5** Ako je rang  $\mathbf{A} = k$ , onda dijagonalnu matricu  $\mathbf{S}$  iz SVD rastava (3.29), str. 60 možemo pisati kao  $\mathbf{S} = \sum_{i=1}^k \mathbf{S}_i$ , gdje je  $\mathbf{S}_i$  dijagonalna matrica koja na mjestu  $(i, i)$  ima element  $\sigma_i$ , a sve ostale nule. Zato SVD matrice  $\mathbf{A}$  možemo pisati

$$\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^T = \mathbf{U} \left( \sum_{i=1}^k \mathbf{S}_i \right) \mathbf{V}^T = \sum_{i=1}^k \mathbf{U} \mathbf{S}_i \mathbf{V}^T = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T. \quad (6.22)$$

Najbolja  $L_2$  aproksimacija ranga  $\rho < k$  matrice  $\mathbf{A}$ , tada je (vidi BJÖRCK (1996), KALMAN (1996))

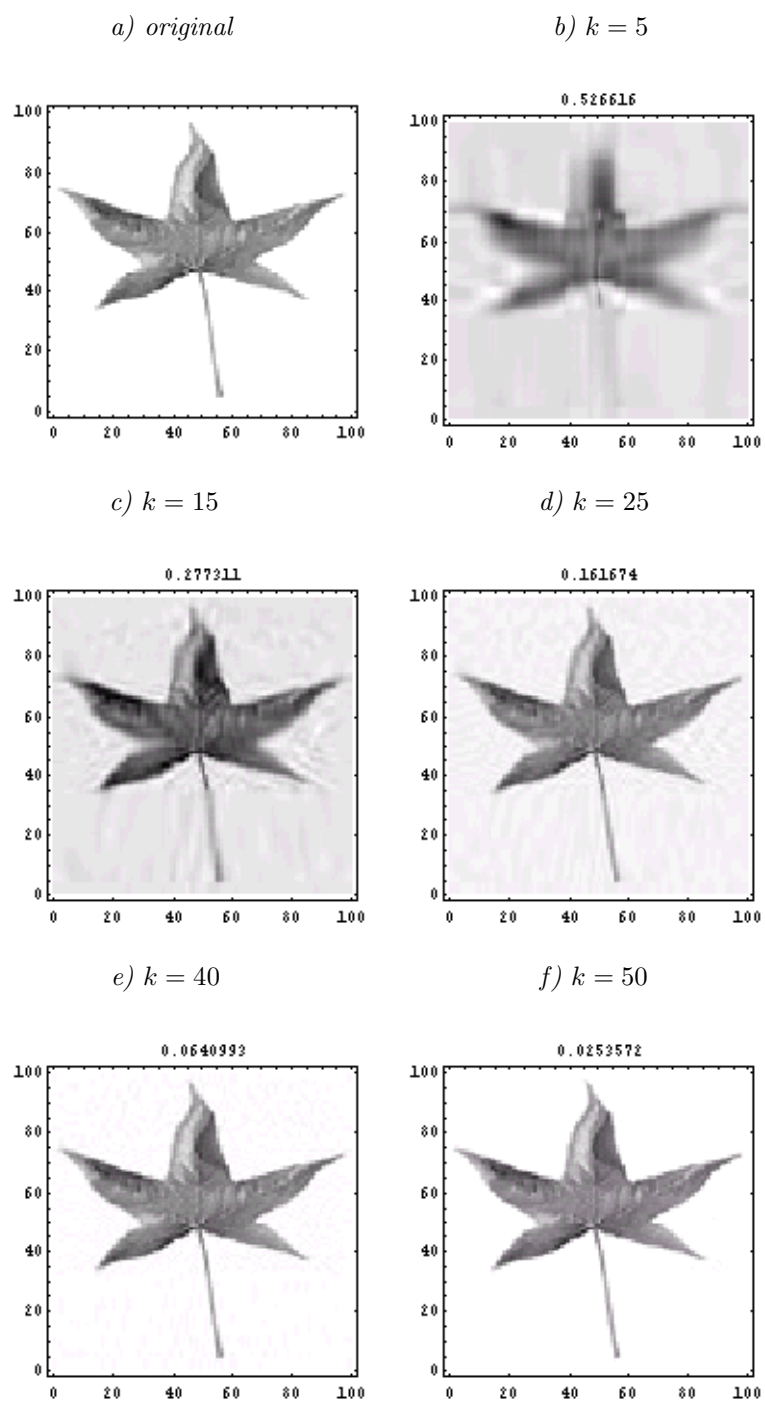
$$\tilde{\mathbf{A}} = \sum_{i=1}^{\rho} \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \text{ pri čemu je } \|\mathbf{A} - \tilde{\mathbf{A}}\|_2 = \sigma_{\rho+1}. \quad (6.23)$$

**Zadatak 6.1** Kako izgleda rastav (6.22) za matricu  $\mathbf{A}$  iz Primjera 3.12, str. 61?

Ova činjenica može se upotrijebiti u svrhu kompresije slike (vidi primjerice DEMEL (1997), KALMAN (1996), TREFETHEN (1997)). Crno-bijela slika lista (vidi Sliku 6.4.a) zapisana je u obliku matrice veličine  $100 \times 100$  s elementima iz intervala  $[0, 1]$ , pri čemu 0 znači potpuno crno, a 1 potpuno bijelo. Aproksimirajući ovu matricu koristeći redom prvih  $\rho = 5, 15, 25, 40, 50$  singularnih vrijednosti u formuli (6.22) dobivamo slike aproksimacije originala<sup>3</sup>

<sup>3</sup>Navedeni primjer izradili su K. Petrić i K. Sabo, studenti Pedagoškog fakulteta u Osijeku, smjer *Matematika-informatika*.





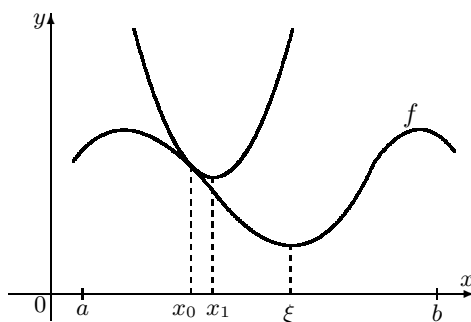
Slika 6.4 Kompresija slike

## 6.2 Nelinearni problemi najmanjih kvadrata

Ako funkcija-model  $f$  nije linearna u barem jednom od svojih parametara  $a_1, \dots, a_n$ , govorimo o nelinearnom problemu najmanjih kvadrata (NPNK). Ovaj problem rješavat ćemo kao nelinearni problem minimuma bez ograničenja. Najpoznatija metoda za rješavanje ovakvih problema je tzv. **Newtonova metoda** (vidi DENNIS (1996), FLETCHER (1987), ORTEGA (1970)).

Radi ilustracije razmotrimo jednodimenzionalni problem minimizacije. Pretpostavimo da je  $f \in C_{[a,b]}^2$  i da se u točki  $\xi$  postiže jedinstveni lokalni minimum funkcije  $f$ . Izaberimo  $x_0 \in (a, b)$  i na osnovi Taylorove formule funkciju  $f$  u okolini točke  $x_0$  aproksimiramo kvadratnom funkcijom

$$k_1(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2$$



Slika 6.5. Newtonova metoda

Sljedeću aproksimaciju  $x_1$  tražene točke  $\xi$  birat ćemo tako da odredimo minimum kvadratne funkcije  $k_1$

$$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)}.$$

Ponavljajući postupak dobivamo niz  $x_0, x_1, \dots, x_n, \dots$  zadan rekursivnom formulom

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}, \quad n = 0, 1, \dots$$

koji uz neke uvjete (vidi primjerice DENNIS (1996)) konvergira prema  $\xi$ . Ovo je tzv. **obična Newtonova metoda minimizacije**. Primijetite da se ona podudara s Newtonovom metodom (t.4.3, str. 77) za rješavanje jednadžbe  $f'(x) = 0$ .

Analogno prethodnom, lokalni minimum dva puta neprekidno derivabilne funkcije  $F : \mathbb{R}^n \rightarrow \mathbb{R}$  mogli bismo potražiti iterativnim postupkom (vidi primjerice DENNIS (1996), FLETCHER (1987), ORTEGA (1970))

$$\mathbf{a}^{k+1} = \mathbf{a}^k - (\mathbf{H}_F(\mathbf{a}^k))^{-1} \text{grad } F(\mathbf{a}^k), \quad \mathbf{a}^0 \in \mathbb{R}^n,$$

gdje je  $\mathbf{H}_F(\mathbf{a}^k)$  Hessijan funkcije  $F$  u točki  $\mathbf{a}^k$ . U cilju povećanja numeričke stabilnosti i osiguranja konvergencije, navedeni iterativni postupak modificira se na sljedeći način:

$$\mathbf{a}^{k+1} = \mathbf{a}^k - \lambda_k \mathbf{s}^k, \quad \mathbf{a}^0 \in \mathbb{R}^n, \quad (6.24)$$

gdje je  $\mathbf{s}^k$  rješenje jednadžbe

$$\mathbf{H}_F(\mathbf{a}^k) \mathbf{s} = \text{grad } F(\mathbf{a}^k). \quad (6.25)$$

Vektor  $\mathbf{s}^k$  nazivamo vektor smjera silaska. Broj  $\lambda_k \in (0, 1]$  je duljina koraka u smjeru vektora silaska koji je određen sljedećim algoritmom

**Algoritam** (određivanje duljine koraka)

Korak 0. izabрати brojeve  $\tau \in (0, \frac{1}{2})$ ,  $\nu \in (0, 1)$ ;

Korak 1. staviti  $\lambda = 1$ ;

Korak 2. provjeriti nejednakost

$$F(\mathbf{a}^k - \lambda \mathbf{s}^k) - F(\mathbf{a}^k) < \tau \lambda (\mathbf{g}^k, \mathbf{s}^k), \quad (6.26)$$

gdje je  $\mathbf{g}^k = \text{grad } F(\mathbf{a}^k)$ , a  $(\mathbf{g}^k, \mathbf{s}^k)$  uobičajeni skalarni produkt;

Korak 3. ako je nejednakost (6.26) ispunjena, staviti  $\lambda_k = \lambda$ ; inače smanjiti duljinu koraka stavljajući  $\lambda := \lambda \nu$  i prijeći na Korak 2.

Algoritam se ponavlja tako dugo dok se ne ispuni uvjet iz Koraka 2. jer na taj način osiguravamo smanjenje vrijednosti minimizirajuće funkcije

$$F(\mathbf{a}^{k+1}) < F(\mathbf{a}^k),$$

i konvergenciju niza aproksimacija  $(\mathbf{a}^k)$  prema lokalnom minimumu funkcije  $F$ . Iterativni postupak (6.24)-(6.25) nazivamo Newtonova metoda minimizacije

Uz neke uvjete na funkciju  $F$  i odgovarajući izbor početne aproksimacije  $\mathbf{a}^0$  Newtonova metoda konvergira prema lokalnom minimumu  $\mathbf{a}^*$  funkcije  $F$  (DENNIS (1996)). Ako je Hessijan funkcije  $F$  u točki  $\mathbf{a}^k$  pozitivno definitna matrica, može se pokazati da je za  $\mathbf{g}^k \neq \mathbf{0}$  uvjet (6.26) uvijek moguće ispuniti. Ako je  $\mathbf{g}^k = \mathbf{0}$ , onda je  $\mathbf{a}^k$  kritična točka (pa i mogući kandidat za lokalni minimum) funkcije  $F$ .

**Zadatak 6.2** Pomoću Newtonove metode pronađite lokalni minimum funkcije

$$f(x_1, x_2) = e^{x_1}(4x_1^2 + 2x_2^2 + 4x_1x_2 + 2x_2 + 1).$$

Rješenje:  $x_1^* = 0.5$ ,  $x_2^* = -1$ ,  $f(x_1^*, x_2^*) = 0$ .

### 6.2.1 Gauss-Newtonova metoda

Za dane podatke  $(x_i, y_i)$ ,  $i = 1, \dots, m$ ,  $m \geq n$  i model-funkciju  $x \mapsto f(x; \mathbf{a})$  nelinearnu u vektoru parametara  $\mathbf{a} \in \mathbb{R}^n$  treba pronaći globalni minimum funkcije (6.2)

$$F(\mathbf{a}) = \frac{1}{2} \sum_{i=1}^m (f(x_i; \mathbf{a}) - y_i)^2 = \frac{1}{2} \sum_{i=1}^m r_i^2(\mathbf{a}) = \frac{1}{2} \|r(\mathbf{a})\|_2^2, \quad \mathbf{a} \in \mathbb{R}^n.$$

Prilikom korištenja Newtonove metode potrebno je poznavati sve prve i druge parcijalne derivacije minimizirajuće funkcije. Budući da je u našem slučaju minimizirajuća funkcija (6.2), odnosno (6.3) specijalnog oblika, Newtonovu metodu moguće je prilagoditi i pojednostaviti.

Uz pretpostavku da je u blizini rješenja  $\mathbf{a}^*$  član  $\mathbf{J}^T(\mathbf{a})\mathbf{J}(\mathbf{a})$  dominantan u izrazu (6.5), Hessijan  $\mathbf{H}_F$  funkcije  $F$  možemo aproksimirati s

$$\mathbf{H}_F(\mathbf{a}) \approx \mathbf{J}^T(\mathbf{a})\mathbf{J}(\mathbf{a}). \quad (6.27)$$

Gradijent minimizirajuće funkcije zadan je s (6.4). Na taj način u Newtonovom iterativnom postupku (6.24)-(6.25) vektor smjera silaska  $\mathbf{s}^k$  dobivamo rješavajući sustav

$$\mathbf{J}^T(\mathbf{a}^k)\mathbf{J}(\mathbf{a}^k)\mathbf{s}^k = \mathbf{J}^T(\mathbf{a}^k)\mathbf{r}(\mathbf{a}^k). \quad (6.28)$$

Uz pretpostavku da je  $\text{rang } \mathbf{J}(\mathbf{a}^k) = n$ , matrica  $\mathbf{J}^T(\mathbf{a}^k)\mathbf{J}(\mathbf{a}^k)$  je pozitivno definitna, sustav (6.28) ima jedinstveno rješenje, a uvjet (6.26) iz kojeg biramo duljinu koraka  $\lambda_k$  uvijek se može ispuniti.

Sustav (6.28) možemo shvatiti kao sustav normalnih jednadžbi za LPNK

$$\mathbf{J}(\mathbf{a}^k)\mathbf{s}^k \simeq \mathbf{r}(\mathbf{a}^k). \quad (6.29)$$

Na taj način rješavanje NPNK Gauss-Newtonovom metodom svodi se na rješavanje niza linearnih problema najmanjih kvadrata, koje je najbolje rješavati primjenom QR-dekompozicije (vidi t.6.1.2, str. 110).

**Primjer 6.10** Zadani su podaci  $(x_i, y_i)$ ,  $i = 1, \dots, m$ , gdje je

$$x_i = i, \quad y_i = 0.1e^{0.25x_i} + \epsilon_i, \quad \text{gdje je } \epsilon_i \sim N[0, \sigma^2].$$

Treba pronaći optimalne parametre  $b^*$ ,  $c^*$  funkcije-modela  $f(x; b, c) = be^{cx}$ .

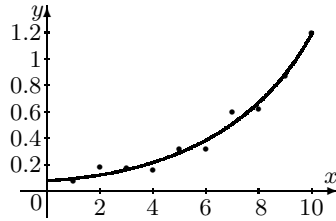
Niže je naveden *Mathematica*-program i dobiveni rezultati.

```
In[1]:= Needs["Statistics`Master`"];
(* Podaci *)
In[2]:= m=10; b=.1; c=.25; SeedRandom[17];
f[x_,b_,c_]:=b Exp[c x];
slf=Plot[f[x,b,c], {x,0,m}, DisplayFunction->Identity];
F[x_,y_,b_,c_]:=Apply[Plus,(y-f[x,b,c])^2];
```

```

x=Table[i, {i,m}];
y=Table[f[i,b,c]+Random[NormalDistribution[0,.05]],{i,m}];
pod=Table[{x[[i]],y[[i]]}, {i,m}];
slpod=ListPlot[pod, PlotStyle->{AbsolutePointSize[5]},
               PlotRange->{0,1.5}, DisplayFunction->Identity];
Show[slf,slpod,DisplayFunction->$DisplayFunction];

```



```

In[5]:= b0=.5; c0=.1; F0=F[x,y,b0,c0]; tau=.5; eps=.005; k=0;
Print["It ",0,":  b = ",b0,"  c = ",c0,"  F = ",F0];
While[b=b0; c=c0; lam=1; k=k+1;
      jac=Table[{f[x[[i]],1,c], x[[i]] f[x[[i]],b,c]}, {i,m}];
      While[q=-Transpose[jac].(y-f[x,b,c]);
            p=-Inverse[Transpose[jac].jac].q;
            {b0,c0} = {b,c} + lam p; F1=F[x,y,b0,c0];
            F1-F0>-tau lam Plus@@(p q), lam=lam/2.
      ];
      norm=Sqrt[Abs[Plus@@q^2]]; norm>eps,
      Print["It ",k,":  b = ",b0,"  c = ",c0,"  F = ",F1,
            "  lam=",lam,"  grad=",norm]
];

```

It 0: b = 0.5 c = 0.1 F = 1.74005  
 It 1: b = 0.263802 c = 0.139479 F = 0.268953 lam=0.5 grad=20.5445  
 It 2: b = 0.0625254 c = 0.226529 F = 0.887862 lam=1 grad=2.18007  
 It 3: b = 0.0973459 c = 0.271372 F = 0.149347 lam=1 grad=16.6483  
 It 4: b = 0.0979424 c = 0.253021 F = 0.0255789 lam=1 grad=10.6933  
 It 5: b = 0.0987964 c = 0.250451 F = 0.0248364 lam=1 grad=0.687365

**Primjer 6.11** Promatramo problem iz Primjera 6.5, str. 105. Podaci o koordinatama točaka  $\mathbf{S}_i(x_i, y_i)$ ,  $i = 1, 4$ , kao i kutevima  $\alpha_i$ ,  $i = 1, 4$  (u stupnjevima i radijanima prenešenim u interval  $[-\frac{\pi}{2}, \frac{\pi}{2}]$ ) vide se u Tablici 6.2.

Za ovaj problem najmanjih kvadrata vektor parametara je  $\mathbf{a} = (x, y)^T$ , komponente  $r_i$  ( $i = 1, 4$ ) vektora reziduala su

$$r_i = f(x_i, y_i; x, y) - \alpha_i, \quad \text{gdje je } f(x_i, y_i; x, y) = \arctg \frac{y - y_i}{x - x_i}, \quad (a)$$

a elementi Jacobijeve matrice su

$$\mathbf{J}_{i1} = \frac{y_i - y}{(x - x_i)^2 + (y - y_i)^2}, \quad \mathbf{J}_{i2} = \frac{x - x_i}{(x - x_i)^2 + (y - y_i)^2}, \quad i = 1, 4.$$

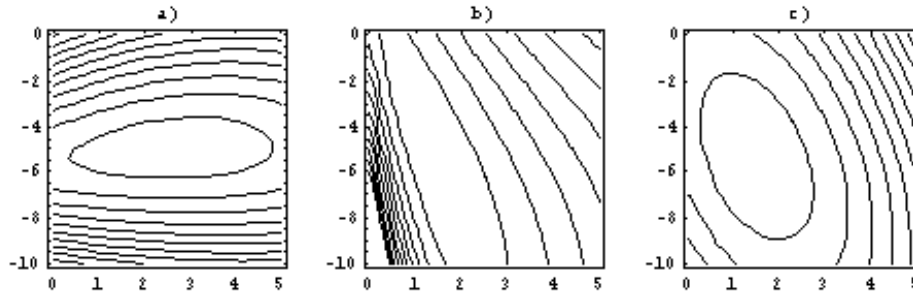
$x_i$	$y_i$	$\alpha_i(^{\circ})$	$\alpha_i(rad)$	$\alpha_i(rad)$	$\alpha_i(^{\circ})$	$r_i$	ODST u %
13	7	50	0.8727	0.8606	49.3	0.0121	0.2%
-2	14	100	-1.396	-1.328	103.9	-0.0685	0.5%
-7	-7	190	0.1745	0.2059	191.8	-0.0314	0.4%
15	-11	330	-0.5236	-0.4558	333.9	-0.0678	0.6%

Tablica 6.2

Problem ćemo riješiti Gauss-Newtonovom metodom primjenom ranije spomenutih programa. Uz početnu aproksimaciju  $\mathbf{a}^0 = (2, -3)^T$  i točnost  $\epsilon = 0.00001$ , nakon 3 iteracije dobivene su koordinate tražene točke  $\mathbf{B}_1(2.705558, -4.972401)$ . Tijek iterativnog procesa prikazan je u *Tablici 6.3*. Proces se zaustavlja kada se postigne  $\|\mathbf{g}^k\| < \epsilon$ .

$k$	$x^k$	$y^k$	$\ \mathbf{g}^k\ $	$F(x_k, y_k)$
0	2	-3	0.033842	0.040778
1	2.62135	-5.052535	0.001222	0.005276
2	2.696991	-4.973894	0.000049	0.005214
3	2.705558	-4.972401	0.000004	0.005214

Tablica 6.3.



Slika 6.6. Izo-krivulje minimizirajućih funkcija

U *Tablici 6.2*. mogu se također vidjeti veličine "teoretskih kuteva"  $\bar{\alpha}_i$  (u radijanima i stupnjevima), pojedini reziduali (odstupanja) i relativna odstupanja u %. Na *Slici 6.6.a* prikazane su tzv. izo-krivulje (*ContourPlot*) odgovarajuće minimizirajuće funkcije  $F$ .

Isti problem mogli bismo riješiti i tako da komponente  $(r_i, i = 1, \dots, 4)$  vektora reziduala zadamo s

$$r_i = f(x_i, y_i; x, y) - \operatorname{tg} \alpha_i, \quad \text{gdje je} \quad f(x_i, y_i; x, y) = \frac{y - y_i}{x - x_i}. \quad (b)$$

Tada se minimum postiže u točki  $\mathbf{B}_2(1.44266, -5.49173)$ , pri čemu dobivamo vrijednost funkcije  $F^* = 0.0208275$ , a izo-krivulje odgovarajuće minimizirajuće funkcije vide se na *Slici 6.6.b*.

Ako komponente  $r_i (i = 1, 4)$  vektora reziduala zadamo s

$$r_i = (x - x_i) \operatorname{tg} \alpha_i - (y - y_i), \quad (c)$$

tada se radi o LPNK. Minimum se postiže u točki  $\mathbf{B}_3(1.49565, -5.3101)$ , a vrijednost funkcije  $F$  je  $F^* = 3.35062$ . Izo-krivulje odgovarajuće minimizirajuće funkcije (*Slika 6.6.c*) sada su koncentrične elipse.

Koji pristup (a), (b) ili (c) ćemo primijeniti ovisi o drugim zahtjevima i dodatnoj analizi (vidi DRAPER (1981), ROOS (1990)).

### 6.2.2 Marquardtova metoda

U Gauss-Newtonovom iterativnom postupku mogu nastupiti problemi onda ako je  $\text{rang } \mathbf{J}(\mathbf{a}^k) < n$  za neki  $k$ . Naime, tada je matrica  $\mathbf{J}^T(\mathbf{a}^k)\mathbf{J}(\mathbf{a}^k)$  singularna. U tom slučaju, Hessijan  $\mathbf{H}(\mathbf{a}^k)$  umjesto s (6.27), aproksimirat ćemo s

$$\mathbf{H}_F(\mathbf{a}) = \mathbf{J}(\mathbf{a})^T \mathbf{J}(\mathbf{a}) + \mu \mathbf{I}, \quad \mu > 0, \quad (6.30)$$

gdje je  $\mathbf{I}$  jedinična matrica. Vektor smjera silaska bit će rješenje jednadžbe

$$(\mathbf{J}^T(\mathbf{a}^k)\mathbf{J}(\mathbf{a}^k) + \mu_k \mathbf{I})\mathbf{s}^k = \mathbf{J}^T(\mathbf{a}^k)\mathbf{r}(\mathbf{a}^k). \quad (6.31)$$

Sustav (6.31) možemo shvatiti kao sustav normalnih jednadžbi za LPNK

$$\begin{bmatrix} \mathbf{J}(\mathbf{a}^k) \\ \sqrt{\mu_k} \mathbf{I} \end{bmatrix} \mathbf{s}^k \simeq \begin{bmatrix} \mathbf{r}(\mathbf{a}^k) \\ 0 \end{bmatrix}, \quad (6.32)$$

Strategija biranja parametra  $\mu$  sastoji se u tome da uvijek bude ispunjen uvjet (6.26), a da se pri tome nastojimo zadržati na što manjoj vrijednosti parametra  $\mu$ . Naime, u tom slučaju vektor smjera silaska blizak je vektoru smjera silaska kod Gauss-Newtonove metode, što osigurava bržu konvergenciju procesa. Kada  $\mu \rightarrow +\infty$ , vektor smjera silaska približava se smjeru antigradijenta, što također osigurava konvergenciju procesa, ali znatno sporiju (DENNIS (1983), FLETCHER (1987), SCHWARZ (1986)).

**Primjedba 6.6** *Kod gotovo svih iterativnih metoda, pa tako i kod navedenih metoda za rješavanje NPNK, važan je izbor kvalitetne početne aproksimacije. Pri tome, mogu se koristiti neki uvjeti iz teorema o egzistenciji rješenja, mogu se koristiti rezultati dobiveni logaritmiranjem ili nekim drugim postupkom (vidi primjere SCITOVSKI (1993,1994)), a u slučaju  $n = 2$  (funkcija-model ima dva nezavisna parametra) mogu se iskoristiti grafičke mogućnosti računala (programski paketi: MATHEMATICA, MATLAB, DERIVE i sl.)*

**Primjedba 6.7** *Više o problemima najmanjih kvadrata, njihovom numeričkom rješavanju i praktičnim primjenama može se naći kod BJÖRCK (1996), GILL (1981, 1991), GOLUB (1996), SCITOVSKI (1993), itd. Spomenimo da se programska podrška navedena u PRESS (1992) nažalost, ne može uspješno koristiti za sve probleme najmanjih kvadrata, koji se mogu javiti u primjenama. Programi za rješavanje nelinearnih problema najmanjih kvadrata ovdje su napisani tako da se u svakom koraku rješava sustav normalnih jednadžbi, što ozbiljno umanjuje stabilnost numeričkog procesa. Za praktičnu primjenu bolje je preporučiti THE NAG-LIBRARY (PHILIPS (1986)) ili Mathematica Package: Statistics'NonlinearFit'.*

## 6.3 Zadaci

**Zadatak 6.3** U 4-znamenkastoj floating-point aritmetici preko sustava normalnih jednadžbi riješite

$$\begin{array}{rcl} \epsilon x_1 & & = 20 \\ & \epsilon x_2 & = 30 \\ & & \epsilon x_3 = 40 \\ x_1 + x_2 + x_3 & = & 10 \end{array}$$

a)  $\epsilon = 0.1$       b)  $\epsilon = 0.01$       c)  $\epsilon = 0.001$

Egzaktno rješenje je

$$\mathbf{x} = \left( \frac{10(-3 + \epsilon + 2\epsilon^2)}{3\epsilon + \epsilon^3}, \frac{10(1 + 3\epsilon)}{3 + \epsilon^2}, \frac{10(3 + \epsilon + 4\epsilon^2)}{3\epsilon + \epsilon^3} \right)^T$$

**Zadatak 6.4** Riješite  $LPNK \mathbf{J}\mathbf{a} \simeq \mathbf{y}$ , gdje je

$$\mathbf{J}^T = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \quad \mathbf{y}^T = [1, 2, 3, 1, 2, 1].$$

Rješenje:  $\mathbf{a} = (3, \frac{7}{4}, \frac{5}{4})^T$ .

**Zadatak 6.5** Pokušajte riješiti Zadatak 6.3 i Zadatak 6.4 rješavanjem sustava normalnih jednadžbi pomoću Cholesky dekompozicije.

**Zadatak 6.6** Olovnu kuglicu puštamo da pada u vakuumu sa različitih visina  $h$  i mjerimo vrijeme dok kuglica ne padne na tlo

$h(m)$	7.5	10	15	20
$t(s)$	1.23	1.44	1.75	2.01

Poznat je zakon gibanja:  $h = \frac{1}{2}at^2$ . Primjenom metode najmanjih kvadrata procijenite veličinu parametra  $a$ .

Rješenje:  $a = 9.8$

**Zadatak 6.7** Umjesto interpolacijskog polinoma u Zadatku 2.10, str. 34, primjenom metode najmanjih kvadrata odredite odgovarajući polinom 2. stupnja.

Rješenje:  $\alpha(X) = 6.810254 - 0.00454X + 610^{-7}X^2$ .

$X(m)$	$\alpha$	$\alpha^*$	$\Delta\alpha^*$
2000	6.296	6.507	0.211
3000	9.543	9.613	0.070
5000	16.512	15.407	1.105
7000	24.865	24.819	0.046
9000	39.407	42.858	3.451



**Zadatak 6.8** Umjesto interpolacijskog polinoma u Zadatku 2.11, str. 35, za funkciju  $t \mapsto h(t)$  primjenom metode najmanjih kvadrata odredite polinom 2. stupnja.

Rješenje:  $h(t) = -0.060599 + 0.23098t + 4.7645t^2$

**Zadatak 6.9** Mjerenjem je ustanovljeno da električni otpor  $R$  bakrene žice ovisi o temperaturi  $t$  na sljedeći način

$t^\circ\text{C}$	19.1	25	30.1	36	40	45.1	50
$R(\Omega)$	76.3	77.8	79.75	80.8	82.35	83.9	85.1

Uz pretpostavku da se radi o linearnoj ovisnosti, odredite parametre  $a$  i  $b$  linearne funkcije  $R(t) = a + bt$ .

Rješenje:  $R(t) = 70.76228 + 0.288072t$

**Zadatak 6.10** Zadana je funkcija model  $f(x; a_1, a_2) = a_1 + a_2x$  i podaci  $(x_i, y_i)$ ,  $i = 1, \dots, n$ . Pokažite da odgovarajući PNK ima jedinstveno rješenje ako je  $x_i \neq x_j$  za barem jedan par indeksa  $i, j$ .

**Zadatak 6.11** Ako je  $\mathbf{A}$  kvadratna regularna matrica, pokažite da se tada pseudo-inverzna matrica  $\mathbf{A}^+$  podudara s  $\mathbf{A}^{-1}$ .

**Zadatak 6.12** Za vektor  $\mathbf{x} = (-4, 1, 4, 2, 5, 1, 1)^T$  odredite Householderov vektor  $\mathbf{u}_k$ ,  $k = 1, \dots, 7$ , tako da je samo  $k$ -ta komponenta transformiranog vektora  $\mathbf{H}\mathbf{x}$  različita od nule.

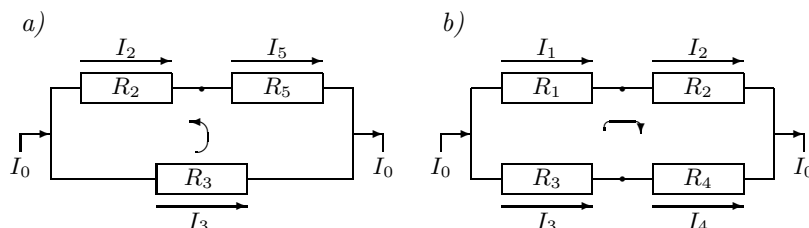
**Zadatak 6.13** Neka je  $\mathbf{H}$  Householderova matricu s Householderovim vektorom  $\mathbf{u}$ . Pokažite da vrijedi

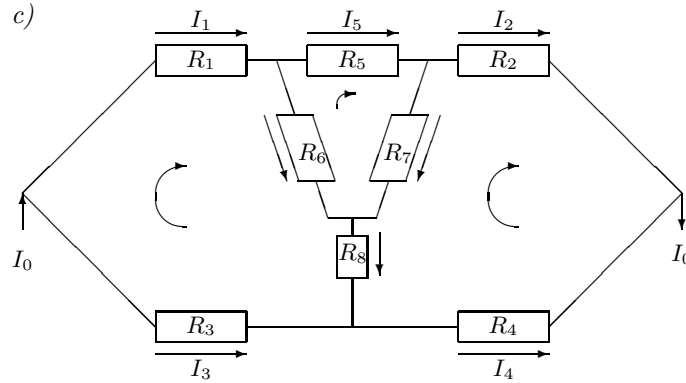
- ako je  $\mathbf{s}$ , vektor okomit na vektor  $\mathbf{u}$ , tada je  $\mathbf{H}\mathbf{s} = \mathbf{s}$ ;
- ako je  $\mathbf{z}$ , vektor paralelan s vektorom  $\mathbf{u}$ , tada je  $\mathbf{H}\mathbf{z} = -\mathbf{z}$ ;
- ako je  $\mathbf{x} = \mathbf{s} + \mathbf{z}$ , gdje je  $\mathbf{s}$  vektor okomit na vektor  $\mathbf{u}$ , a  $\mathbf{z}$  vektor paralelan s vektorom  $\mathbf{u}$ , tada je  $\mathbf{H}\mathbf{x} = \mathbf{s} - \mathbf{z}$ .

Kako se zove geometrijska transformacija koja vektor  $\mathbf{a}$  prevodi u  $\mathbf{H}\mathbf{a}$ ?

**Zadatak 6.14** Za mreže prikazane na slikama a), b) i c) izračunajte veličine struja ako je

- $I_0 = 5\text{A}$ ,  $R_1 = 50\Omega$ ,  $R_2 = 10\Omega$ ,  $R_3 = 5\Omega$
- $I_0 = 10\text{A}$ ,  $R_1 = 100\Omega$ ,  $R_2 = 10\Omega$ ,  $R_3 = 20\Omega$ ,  $R_4 = 30\Omega$
- $I_0 = 10\text{A}$ ,  $R_1 = 5\Omega$ ,  $R_2 = 50\Omega$ ,  $R_3 = 5\Omega$ ,  $R_4 = 50\Omega$ ,  
 $R_5 = 10\Omega$ ,  $R_6 = 12\Omega$ ,  $R_7 = 20\Omega$ ,  $R_8 = 30\Omega$





Rješenje:

- a)  $I_1 = 0.38462A$ ,  $I_2 = 0.38462A$ ,  $I_3 = 4.61539A$   
b)  $I_1 = 3.125A$ ,  $I_2 = 3.125A$ ,  $I_3 = 6.875A$ ,  $I_4 = 6.875A$   
c)  $I_1 = 4.485A$ ,  $I_2 = 4.6996A$ ,  $I_3 = 5.515A$ ,  $I_4 = 5.3A$   
 $I_5 = 3.519A$ ,  $I_6 = 0.9657A$ ,  $I_7 = 1.1803A$ ,  $I_8 = 0.2146A$

**Zadatak 6.15** Pokažite da je

$$\mathbf{Q} = \begin{bmatrix} \cos(-60^\circ) & \sin(-60^\circ) \\ -\sin(-60^\circ) & \cos(-60^\circ) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$$

ortogonalna matrica. Također, pokažite da je  $\mathbf{Q}^T(1, \sqrt{3})^T = (2, 0)^T$ . Kako se zove geometrijska transformacija koja vektor  $\mathbf{a}$  prevodi u vektor  $\mathbf{Q}^T \mathbf{a}$ ?

**Zadatak 6.16** Pokažite da je

$$\mathbf{Q} = \begin{bmatrix} \cos 30^\circ & \sin 30^\circ \\ \sin 30^\circ & -\cos 30^\circ \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix}$$

ortogonalna simetrična matrica. Također, pokažite da je  $\mathbf{Q}^T(1, \sqrt{3})^T = (\sqrt{3}, -1)^T$ . Kako zovemo geometrijsku transformaciju koja vektor  $\mathbf{a}$  prevodi u  $\mathbf{Q}^T \mathbf{a}$ ?

**Zadatak 6.17** Zadane su točke:  $S_1(8, 6)$ ,  $S_2(-4, 5)$ ,  $S_3(1, -3)$  i kutovi  $\alpha_1 = 42^\circ$ ,  $\alpha_2 = 158^\circ$ ,  $\alpha_3 = 248^\circ$  pod kojima se “vidi” tražena točka  $B$  kao u Primjeru 6.5, str. 105, odnosno Primjeru 6.11, str. 118. Izračunajte koordinate točke  $B$  prema sve tri varijante navedene u Primjeru 6.11.

Rješenje:

**Zadatak 6.18** Zadani su podaci  $(x_i, y_i)$ ,  $i = 1, \dots, m$ , na osnovi kojih su metodom najmanjih kvadrata određeni parametri  $k$  i  $l$  linearne regresije  $f(x) = ax + b$ .

a) Pokažite da vrijedi

$$a = \frac{1}{D} (m \sum x_k y_k - \sum x_k \sum y_k),$$

$$b = \frac{1}{D} (\sum x_k^2 \sum y_k - \sum x_k \sum x_k y_k), \quad D = m \sum x_k - (\sum x_k)^2,$$

- b) Ako je:  $\bar{x} = \frac{1}{m} \sum x_k$ ,  $\bar{y} = \frac{1}{m} \sum y_k$ , pokažite da točka  $(\bar{x}, \bar{y})$  leži na pravcu  $y = ax + b$ . Također, pokaži da vrijedi

$$a = \frac{1}{D} \sum (x_k - \bar{x})(y_k - \bar{y}), \quad b = \bar{y} - a\bar{x}, \quad D = \sum (x_k - \bar{x})^2.$$

- c) Kako glase formule za  $a$  i  $b$  ako središte koordinatnog sustava postavimo u točku  $(\bar{x}, \bar{y})$  ?

**Zadatak 6.19** Kako se problem procjene parametara u niže navedenim funkcijama-modelima može linearizirati (vidi SCITOVSKI (1993))

$$\begin{array}{lll} a) y = \frac{a}{x} + b, & b) y = \frac{a}{x+c}, & c) y = \frac{1}{ax+b}, \\ d) y = \frac{x}{a+bx}, & e) y = a \ln x + b, & f) y = a \exp(bx), \\ g) y = \frac{1}{(ax+b)^2}, & h) y = ax \exp(-bx), & i) y = \frac{A}{(1+b \exp(-cx))} \end{array}$$

Za proizvoljno izabrane podatke, parametre u modelima procijenite kao NPNK i kao LPNK dobiven linearizacijom. Usporedite dobivene parametre te nacrtajte odgovarajuće izo-krivulje. Također, grafički prikažite podatke i dobiveni regresijski model.

**Zadatak 6.20** Newtonovom metode minimizacije odredite lokalne minimume funkcije iz Zadatka 4.9, str. 88.

**Zadatak 6.21** Pokažite da funkcija  $F(x, y) = (x - y)^2(x + y)^2$ ,  $x, y \in \mathbb{R}$  postiže svoj globalni minimum  $F^* = 0$  u beskonačno mnogo točaka. Odredite geometrijska mjesta tih točaka. Graf funkcije  $F$  prikazan je na naslovnoj stranici ovog udžbenika.

**Zadatak 6.22** Analizirajte problem minimuma funkcije  $F$ . Uz koje uvjete je njezin Hessijan pozitivno definitan, a uz koje pozitivno semidefinitan? Kakve to posljedice ima na egzistenciju lokalnog i globalnog minimuma? Nacrtajte graf funkcije  $F$  i njene izo-krivulje.

$$\begin{array}{ll} a) F(x, y) = (ax + by + c)^2, & a, b, c \in \mathbb{R}. \\ b) F(x, y) = (a_1x + b_1y + c_1)^2 + (a_2x + b_2y + c_2)^2, & a_i, b_i \in \mathbb{R} \end{array}$$

Kako se mogu generalizirati navedeni problemi?

Uputa: Pokažite da su glavni minori Hesseove matrice funkcije  $F$

$$\begin{array}{ll} a) \Delta_1 = 2a^2, & \Delta_2 = 0; \\ b) \Delta_1 = 2(a_1^2 + a_2^2), & \Delta_2 = 2(a_1b_2 + a_2b_1)^2. \end{array}$$

**Zadatak 6.23** Riješite sljedeće LPNK

$$\begin{array}{ll} a) ax + by + c \simeq 0 & b) \begin{array}{l} a_1x + b_1y + c_1 \simeq 0 \\ a_2x + b_2y + c_2 \simeq 0 \end{array} \end{array}$$

Rješenje: a)  $x = \frac{ac}{a^2+b^2}$ ,  $y = \frac{bc}{a^2+b^2}$ ,  $ab \neq 0$ .

## 7. Numerička integracija

Ako je  $f : [a, b] \rightarrow \mathbb{R}$  neprekidna funkcija, a  $G$  njena primitivna funkcija, onda se Riemannov integral na intervalu  $[a, b]$  može izračunati primjenom Newton-Leibnizove formule:

$$I = \int_a^b f(x) dx = G(b) - G(a).$$

Međutim, u praksi se najčešće pojavljuju takve situacije, gdje nije moguće primijeniti ovu formulu. Može se dogoditi:

- *da primitivnu funkciju  $G$  nije moguće dobiti elementarnim metodama (vidi Primjer 7.5 i zadatke 7.3, 7.4 i 7.5)*
- *da je podintegralna funkcija poznata samo u nekoliko (konačno) točaka (kao u Primjeru 7.2 i Zadatku 7.6).*

Da bismo ipak aproksimativno (približno) izračunali vrijednost integrala  $I$ , podintegralnu funkciju  $f$  interpolirat ćemo (vidi t.2) nekom jednostavnijom funkcijom  $\varphi$  i na taj način dobiti aproksimaciju integrala  $I$ :

$$I^* = \int_a^b \varphi(x) dx.$$

Pri tome, aproksimirajuća funkcija  $\varphi$  treba biti takva, da za zadanu točnost  $\epsilon > 0$ , bude:

$$\Delta I^* = |I - I^*| < \epsilon$$

Uz pretpostavku poznavanja funkcije  $f$  u  $n+1$  točaka  $x_0, \dots, x_n \in [a, b]$ , za funkciju  $\varphi$  možemo primjerice uzeti *Lagrangeov interpolacijski polinom* (t.2.1.1), str. 18.

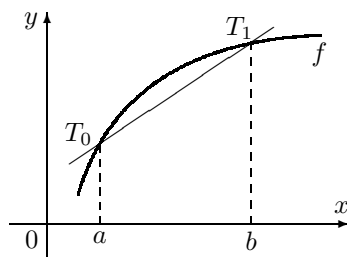
### 7.1 Trapezno pravilo

Funkciju  $f : [a, b] \rightarrow \mathbb{R}$  interpolirat ćemo linearnom funkcijom  $P_1$  (interpolacijskim polinomom stupnja 1) u čvorovima interpolacije

$$x_0 = a, \quad x_1 = b.$$

Graf funkcije  $P_1$  je pravac koji prolazi točkama  $T_0(a, f(a))$ ,  $T_1(b, f(b))$ ,

$$P_1(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a).$$



Slika 7.1 Trapezno pravilo

Sada lako dobivamo

$$I^* = \int_a^b P_1(x) dx = \frac{b-a}{2} (f(a) + f(b)). \quad (7.1)$$

Geometrijski,  $I^*$  predstavlja površinu trapeza sa stranicama  $f(a)$  i  $f(b)$  i visinom  $h = b - a$ . Apsolutna pogreška  $\Delta I^*$  predstavlja veličinu površine između pravca  $L_1$  i grafa funkcije  $f$  (vidi Sliku 7.1). Vrijedi:

**Teorem 7.1** Neka je  $f \in C_{[a,b]}$ . Tada postoji  $c \in (a, b)$ , tako da bude:

$$I = \int_a^b f(x) dx = \frac{b-a}{2} (f(a) + f(b)) - \frac{(b-a)^3}{12} f''(c). \quad (7.2)$$

**Dokaz.** Prema Teoremu 2.1, str. 23, postoji  $\xi \in (a, b)$  (ovisan o  $x$ ) tako da je

$$E = I - I^* = \int_a^b (f(x) - P_1(x)) dx = \int_a^b \frac{f''(\xi)}{2} (x-a)(x-b) dx.$$

Kako je  $(x-a)(x-b) \leq 0$  za sve  $x \in [a, b]$ , koristeći poopćeni teorem o srednjoj vrijednosti integralnog računa<sup>1</sup>, dobivamo

$$E = I - I^* = \frac{1}{2} f''(c) \int_a^b (x-a)(x-b) dx, \quad c \in (a, b).$$

Uvođenjem supstitucije  $x = a + (b-a)t$  dobivamo

$$E = I - I^* = \frac{1}{2} f''(c) (b-a)^3 \int_0^1 t(t-1) dt = -\frac{(b-a)^3}{12} f''(c). \quad \square$$

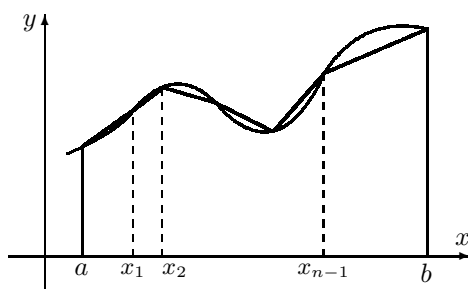
<sup>1</sup>Teorem. Ako je  $f \in C_{[a,b]}$ , a  $g$  integrabilna funkcija stalnog predznaka na  $[a, b]$  (ili je  $g(x) \geq 0$  ili je  $g(x) \leq 0$ ), onda postoji  $c \in [a, b]$ , takav da vrijedi (vidi primjerice [3], II. dio)

$$\int_a^b f(x)g(x) dx = f(c) \int_a^b g(x) dx$$

Ako je interval integracije  $[a, b]$  velik, i pogreška  $E$  bit će velika. U cilju postizanja bolje aproksimacije  $I^*$  integrala  $I$ , interval  $[a, b]$  podijelit ćemo na više podintervala i onda na svakom od njih primijeniti trapezno pravilo (7.2).

Pretpostavimo da funkciju  $f$  poznajemo u  $n + 1$  točaka  $x_0, \dots, x_n \in [a, b]$ , ali da je pri tome ispunjeno:

$$x_1 - x_0 = \dots = x_n - x_{n-1} = h \quad \& \quad x_0 = a, \quad x_n = b.$$



Slika 7.2 Generalizirano trapezno pravilo

Očigledno je  $h = \frac{b-a}{n}$ , a točke  $x_0, \dots, x_n$  dijele interval  $[a, b]$  na  $n$  jednakih podintervala duljine  $h$ .

Označimo:

$$y_i = f(x_i), \quad i = 0, \dots, n.$$

Na svakom podintervalu primijenit ćemo trapezno pravilo (7.2). Tako za interval  $[x_{i-1}, x_i]$  dobivamo:

$$\int_{x_{i-1}}^{x_i} f(x) dx = \frac{h}{2} (y_{i-1} + y_i) - \frac{h^3}{12} f''(c_i), \quad c_i \in (x_{i-1}, x_i).$$

Cijeli integral  $I$  bit će:

$$I = \int_a^b f(x) dx = \sum_{i=1}^n \int_{x_{i-1}}^{x_i} f(x) dx = \frac{h}{2} (y_0 + 2y_1 + \dots + 2y_{n-1} + y_n) - \frac{h^3}{12} \sum_{i=1}^n f''(c_i). \quad (7.3)$$

Zbog neprekidnosti funkcije  $f''$  na intervalu  $(a, b)$ , postoji  $c \in (a, b)$ , takav da je

$$\frac{1}{n} \sum_{i=1}^n f''(c_i) = f''(c). \quad (7.4)$$

Uvrštavajući (7.4) u (7.3), dobivamo generalizirano trapezno pravilo:

$$I = I^* + E_n,$$

gdje je:

$$I^* = \frac{h}{2} (y_0 + 2y_1 + \dots + 2y_{n-1} + y_n), \quad (7.5)$$

$$E_n = -\frac{b-a}{12}h^2f''(c). \quad (7.6)$$

Niže je naveden *Mathematica*-modul, koji primjenom generalizirane trapezne formule izračunava približnu vrijednost integrala funkcije  $f$ .

```
Trapez[n_, a_, b_, f_] := Module[{i, h}, h=(b-a)\n;
  int = (h/2) (f[a] + f[b] + 2 Sum[f[a+i h], {i, n-1}])//N;
  Print["n=", n, " I=", int]
]
```

**Primjedba 7.1** Ako je zadana točnost  $\epsilon > 0$  s kojom treba izračunati integral  $I$  i ako označimo:

$$M_2 = \max_{x \in [a,b]} |f''(x)|, \quad (7.7)$$

onda iz (7.6) možemo ocijeniti apsolutnu pogrešku  $\Delta I^*$  aproksimacije  $I^*$ :

$$\Delta I^* = |R_n| \leq \frac{b-a}{12}h^2M_2 < \epsilon. \quad (7.8)$$

Iz (7.8) tada lako dobivamo broj podintervala na koji treba podijeliti interval  $[a, b]$ , da primjenom generaliziranog trapeznog pravila postignemo točnost  $\epsilon$ :

$$n > (b-a)\sqrt{\frac{M_2}{\epsilon} \cdot \frac{b-a}{12}}. \quad (7.9)$$

**Primjer 7.1** Znamo da je:

$$I = \int_0^1 \frac{4}{1+x^2} dx = \pi$$

Izračunat ćemo približnu vrijednost ovog integrala primjenom generaliziranog trapeznog pravila i usporediti rezultate.

Kako je:

$$f''(x) = \frac{8(3x^2 - 1)}{(1+x^2)^3},$$

lako dobivamo  $M_2 = |f''(0)| = 8$  (vidi *Sliku 7.6*). U *Tablici 7.1* prikazani su rezultati izračunavanja dobiveni za različite vrijednosti točnosti  $\epsilon$  primjenom sljedećeg *Mathematica* – programa:

```
f[x_] := 4/(1 + x^2);          a=0;  b=1;  M2=8;  eps0=5;
Do[
  eps = eps0/10.^i;
  n = Floor[(b-a) Sqrt[M2 (b-a)/(12 eps)]] + 1;
  Trapez[n, a, b, f]; Print["E=", Abs[int - Pi] ],
  {i, 6}]
```

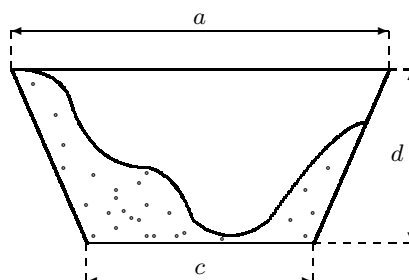
$\epsilon$	br. točnih znamenki	$n$	$I^*$	$\Delta I^*$
0.5	1	2	3	0.041593
0.05	2	4	3.1	0.010416
0.005	3	12	3.14	0.001157
0.0005	4	37	3.141	0.000122
0.00005	5	116	3.1415	0.000012
0.000005	6	366	3.14159	0.000001

Tablica 7.1

**Primjer 7.2** Mjerenjem dubina poprečnog presjeka kanala širine  $a = 10$  m na svakih metar, dobiveni su ovakvi podaci

$x_i$ (m)	0	1	2	3	4	5	6	7	8	9	10
$y_i$ (m)	0	0.1	0.5	1.2	1.8	2.3	2.1	2.5	2.1	1.5	0.9

Kanal treba tako prokopati da dobije oblik pravilnog trapeza sa stranicama  $a = 10$  m,  $c = 6$  m i dubine  $d = 3$  m. Treba procijeniti količinu otkopa koji će u tu svrhu biti potrebno uraditi na jednom dužnom metru kanala.



Slika 7.3 Presjek kanala

Površinu poprečnog presjeka  $\Delta P$  zemlje u kanalu, koju treba iskopati dobit ćemo primjenom trapezne formule (7.5)

$$\Delta P = 24 - 14.55 = 9.45 \text{ m}^2,$$

a količina iskopa na jednom dužnom metru kanala je  $\Delta V = 9.45 \text{ m}^3$ .

Zašto u ovom primjeru ne bi bilo dobro najprije odrediti interpolacijski polinom  $P_{10}$  i primijeniti formulu  $\Delta P = \frac{a+c}{2}d - \int_0^{10} P_{10}(x) dx$ ? Provjerite da biste tako dobili rezultat  $\Delta P = 6.48 \text{ m}^2$ .



## 7.2 Newton-Cotesova formula

U cilju bolje aproksimacije funkcije  $f$ , pa time i bolje aproksimacije integrala  $I$ , interval  $[a, b]$  mogli bismo podijeliti na  $n$  jednakih podintervala jednoliko raspoređenim čvorovima  $x_i = a + i\frac{b-a}{n}$ , a funkciju  $f$  tada aproksimirati interpolacijskim polinomom  $n$ -tog reda u Lagrangeovom obliku

$$P_n(x) = \sum_{k=0}^n f(x_k) p_k(x), \quad p_k(x) = \prod_{\substack{i=0 \\ i \neq k}}^n \frac{x - x_i}{x_k - x_i}.$$

Približnu vrijednost  $I^*$  integrala  $I$  mogli bismo dobiti kao

$$I_{(n)}^* = \int_a^b P_n(x) dx = \sum_{k=0}^n f(x_k) \int_a^b p_k(x) dx = (b-a) \sum_{k=0}^n \omega_k f(x_k),$$

gdje su  $\omega_k = \frac{1}{b-a} \int_a^b p_k(x) dx$  težine. Uz supstituciju  $x = a + (b-a)t$  one postaju jednostavnije

$$\omega_k = \frac{1}{b-a} \int_a^b p_k(x) dx = \frac{1}{b-a} \int_a^b \prod_{\substack{i=0 \\ i \neq k}}^n \frac{x - x_i}{x_k - x_i} dx = \int_0^1 \prod_{\substack{i=0 \\ i \neq k}}^n \frac{nt - i}{k - i} dt.$$

Tako dobivamo poznatu kvadraturu ili Newton-Cotesovu formulu  $n$ -tog reda za aproksimaciju integrala  $I$ :

$$\int_a^b P_n(x) dx = (b-a) \sum_{k=0}^n \omega_k f(x_k), \quad \omega_k = \int_0^1 \prod_{\substack{i=0 \\ i \neq k}}^n \frac{nt - i}{k - i} dt \quad (7.10)$$

**Zadatak 7.1** Pokažite da su  $\omega_k$ ,  $k = 0, 1, \dots, n$  iz (7.10) racionalni brojevi sa svojom svojstvom  $\sum_{k=0}^n \omega_k = 1$ .

**Zadatak 7.2** Primjenom Newton-Cotesove formule izvedite trapeznu formulu (7.2).

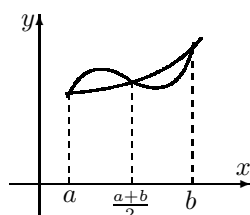
## 7.3 Simpsonovo pravilo

Specijalno, iz Newton-Cotesove formule (7.10) za  $n = 2$  dobit ćemo poznato Simpsonovo pravilo. Primijetite da je u ovom slučaju:  $x_0 = a$ ,  $x_1 = \frac{a+b}{2}$ ,  $x_2 = b$ . Najprije ćemo izračunati težine

$$\begin{aligned} \omega_0 &= \frac{1}{2} \int_0^1 (2t-1)(2t-2) dt = \frac{1}{6} \\ \omega_1 &= - \int_0^1 2t(2t-2) dt = \frac{2}{3} \\ \omega_2 &= \frac{1}{2} \int_0^1 2t(2t-1) dt = \frac{1}{6}. \end{aligned}$$

Nakon toga lako dobivamo Simpsonovo pravilo (vidi također *Sliku 7.4*).

$$I^* = \int_a^b P_2(x) dx = \frac{b-a}{6} \left( f(a) + 4f\left(\frac{a+b}{2}\right) + f(b) \right). \quad (7.11)$$

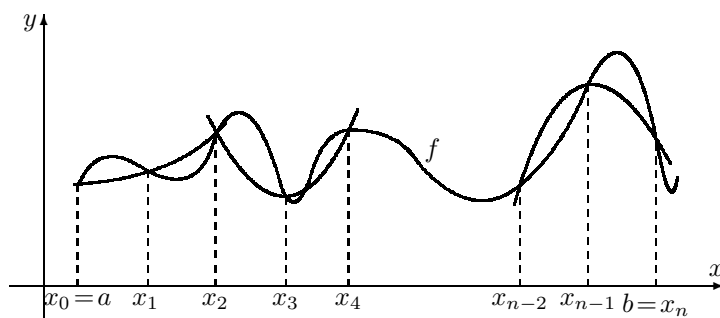


Slika 7.4 Simpsonovo pravilo

Može se također pokazati (vidi primjerice STOER (1993) ili STEWART (1996)) da je pogreška aproksimacije zadana s

$$E = I - I^* = -\frac{(b-a)^5}{90} f^{(4)}(c), \quad c \in (a, b). \quad (7.12)$$

Ako je interval integracije  $[a, b]$  velik, i pogreška  $E$  bit će velika. U cilju postizanja bolje aproksimacije  $I^*$  integrala  $I$  interval  $[a, b]$  podijelit ćemo na parni broj ( $n = 2m$ ) podintervala duljine  $h = \frac{b-a}{n}$  čvorovima  $x_i = a + ih$ ,  $i = 0, 1, \dots, n$ . Uz oznaku  $y_i = f(x_i)$ ,  $i = 0, 1, \dots, n$ , redom na po dva podintervala primijenit ćemo Simpsonovo pravilo (7.11) (vidi *Sliku 7.5*). Lako se može pokazati da vrijedi generalizirano Simpsonovo pravilo:



Slika 7.5 Generalizirano Simpsonovo pravilo

$$I = I^* + E_n, \quad (7.13)$$

$$I^* = \frac{h}{3} ((y_0 + y_{2m}) + 4(y_1 + \dots + y_{2m-1}) + 2(y_2 + \dots + y_{2m-2})), \quad (7.14)$$

$$E_n = -\frac{b-a}{180} h^4 f^{(4)}(c), \quad c \in (a, b). \quad (7.15)$$

Niže je naveden *Mathematica*-modul, koji primjenom generalizirane Simpsonove formule izračunava približnu vrijednost integrala funkcije  $f$ .

```
Simpson[n_, a_, b_, f_] := Module[{i,h,n1}, h = N[(b-a)/n]; n1=n/2;
  int = (h/3) (f[a] + f[b] + 4 Sum[f[a+(2 i-1) h], {i,n1}]
    + 2 Sum[f[a+2 i h], {i,n1-1}]);
  Print["n=", n, " I=", int]
]
```

**Primjedba 7.2** Ako je zadana točnost  $\epsilon > 0$  s kojom treba izračunati integral  $I$  i ako označimo:

$$M_4 = \max_{x \in [a,b]} |f^{(4)}(x)|, \quad (7.16)$$

onda iz (7.15) možemo ocijeniti apsolutnu pogrešku  $\Delta I^*$  aproksimacije  $I^*$

$$\Delta I^* = |E_n| \leq \frac{b-a}{180} h^4 M_4 < \epsilon. \quad (7.17)$$

Iz (7.17) tada lako dobivamo broj podintervala na koji treba podijeliti interval  $[a, b]$ , da primjenom generaliziranog trapeznog pravila postignemo točnost  $\epsilon$ :

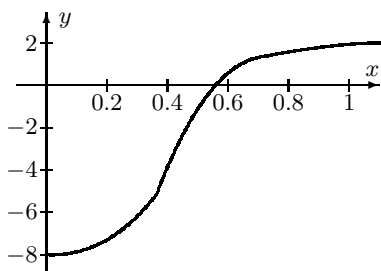
$$n > (b-a) \sqrt[4]{\frac{M_4}{\epsilon} \cdot \frac{b-a}{180}}. \quad (7.18)$$

**Primjer 7.3** Integral iz Primjera 7.1 približno ćemo izračunati primjenom generaliziranog Simpsonovog pravila.

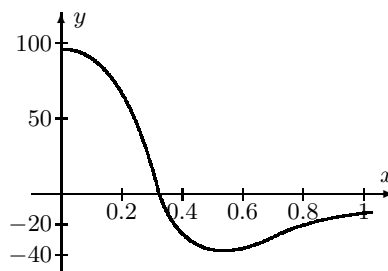
Kako je:

$$f^{(4)}(x) = \frac{96(5x^4 - 10x^2 + 1)}{(x^2 + 1)^5}$$

onda imamo (vidi Sliku 7.7):  $M_4 = |f^{(4)}(0)| = 96$ .



Slika 7.6  $\frac{d^2}{dx^2} \left( \frac{4}{1+x^2} \right)$



Slika 7.7  $\frac{d^4}{dx^4} \left( \frac{4}{1+x^2} \right)$

U Tablici 7.2 prikazani su rezultati izračunavanja (primjenom niže navedenog *Mathematica*-programa) za različite vrijednosti točnosti  $\epsilon$ .

```
f[x_] := 4/(1 + x^2);      a=0;  b=1;  M4=96;  eps0=.5;
Do[
  eps=eps0/10.^i; n=Floor[(b-a) ((M4 (b-a)/(180 eps))^(1/4)) + 1;
  If[OddQ[n], n=n+1, ];
  Simpson[n, a, b, f];      Print["E=", Abs[int - Pi] ],
  {i, 6}]
```

$\epsilon$	br. točnih znamenki	$n$	$I^*$	$\Delta I^*$
0.05	2	2	3.1	0.008259
0.005	3	4	3.14	0.000024
0.0005	4	6	3.142	0
0.00005	5	8	3.1416	0
0.000005	6	12	3.14159	0

Tablica 7.2

**Zadatak 7.3** Izvedite Newton-Cotesove formule reda  $n = 3$  i  $n = 4$ , te izradite odgovarajuće Mathematica – programe.

**Primjer 7.4** Riješit ćemo problem postavljen u Primjeru 7.2 primjenom generaliziranog Simpsonovog pravila. Dobivamo:  $I^* = 14.77$ , odnosno:  $\Delta P = 24 - 14.77 = 9.234 \text{ m}^3$ .

**Primjer 7.5** Prilikom određivanja zakrivljenosti zavoja ceste pojavljuju se tzv. Fresnelovi integrali:

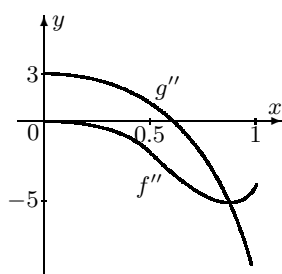
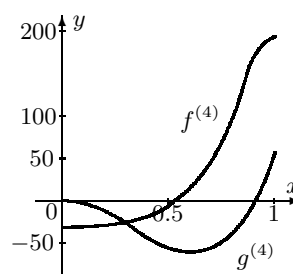
$$I = \int_0^s \cos\left(\frac{\pi}{2}t^2\right) dt, \quad J = \int_0^s \sin\left(\frac{\pi}{2}t^2\right) dt \quad s > 0.$$

Budući da primitivne funkcije ovih podintegralnih funkcija ne možemo naći, njihove vrijednosti odredit ćemo numeričkom integracijom. Za  $s = 0.5$  i  $s = 1$  izračunat ćemo ove integrale primjenom generaliziranog trapeznog i Simpsonovog pravila uz točnost  $\epsilon = 0.00005$  (4 korektne decimale). Da bismo odredili konstante  $M_2$  i  $M_4$  za navedne podintegralne funkcije nacrtat ćemo grafove njihovih drugih (Slika 7.8) i četvrtih (Slika 7.9) derivacija na intervalu  $[0, 1]$ .

$f(t) = \cos(\frac{\pi}{2}t^2)$	$M_2$	$M_4$
$s = 0.5$	3.5	29.6
$s = 1$	6	186

$g(t) = \sin(\frac{\pi}{2}t^2)$	$M_2$	$M_4$
$s = 0.5$	3	52
$s = 1$	10	70

Za  $s = 0.5$  dobivamo:  $I = 0.4923$ ,  $J = 0.0647$ , a za  $s = 1$ :  $I = 0.7799$ ,  $J = 0.4382$

Slika 7.8 Procjena  $M_2$ Slika 7.9 Procjena  $M_4$

Pri tome se broj podintervala  $n$ , na koji smo morali podijeliti interval integracije, kretao kao u niže navedenoj tablici:

$n$ ( $s = 0.5$ )	Trapezno pravilo	Simpsonovo pravilo	$n$ ( $s = 1$ )	Trapezno pravilo	Simpsonovo pravilo
$I$	28	4	$I$	101	12
$J$	26	4	$J$	130	10

**Primjedba 7.3** Više o numeričkim metodama integracije i njenim teorijskim i praktičnim aspektima može se naći u IVANŠIĆ (1998), SCHWARZ (1986), STOER (1993) itd. Programaska podrška (FORTRAN) naći će se u PRESS (1989), PHILIPS (1986), a može se koristiti i gotovi korisnički software: MATHEMATICA, MATLAB.

## 7.4 Zadaci

**Zadatak 7.4** Primjenom generaliziranog trapeznog pravila za zadanu podintegralnu funkciju  $f$  i interval integracije  $[a, b]$  uz točnost  $\epsilon = 0.00005$  odredite konstantu  $M_2$ , broj podintervala ( $n$ ) na koji treba podijeliti interval integracije  $[a, b]$ , te vrijednost aproksimacije integrala  $I$ .

- a)  $f(x) = \sin^2 x$ ,  $[0, \pi/4]$ ;
- b)  $f(x) = \cos^2 x$ ,  $[0, \pi/4]$ ;
- c)  $f(x) = \sin x - \ln x + e^x$ ,  $[0.2, 1]$ ;
- d)  $f(x) = \sqrt{1 + 2x}$ ,  $[0, 1]$ ;
- e)  $f(x) = x^2(1 - x^2)^{-1/2}$ ,  $[0, 0.5]$ ;
- f)  $f(x) = \frac{\sin x}{x}$ ,  $[\pi/4, \pi/2]$ ;
- g)  $g(x) = \frac{5-x}{9-x^2}$ ,  $[0, 2]$ .

Uputa: prilikom određivanja konstante  $M_2$  može se koristiti kompjuterska grafika, a za izračunavanje integrala navedeni *Mathematica*-modul.

Rješenje:

- a)  $M_2 = 2$ ,  $n = 41$ ,  $I^* = 0.1427$ .
- b)  $M_2 = 2$ ,  $n = 41$ ,  $I^* = 0.6427$ .
- c)  $M_2 = f''(0.2) = 26$ ,  $n = 149$ ,  $I^* = 2.4148$ .
- d)  $M_2 = |f''(0)| = 1$ ,  $n = 41$ ,  $I^* = 1.3987$ .
- e)  $M_2 = |f''(0.5)| = 5$ ,  $n = 33$ ,  $I^* = 0.04532$ .
- f)  $M_2 = |f''(\pi/4)| \simeq 0.3$ ,  $n = 16$ ,  $I^* = 0.6118$ .
- g)  $M_2 = |f''(2)| \simeq 0.6$ ,  $n = 90$ ,  $I^* = 1.0473$ .

**Zadatak 7.5** Primjere iz Zadatka 7.4 riješite primjenom generaliziranog Simpsonovog pravila.

Rješenje:

- a)  $M_4 = |f^{(4)}(0)| = 8, \quad n = 6, \quad I^* = 0.1427.$   
 b)  $M_4 = |f^{(4)}(0)| = 8, \quad n = 6, \quad I^* = 0.6427.$   
 c)  $M_4 = f^{(4)}(0.2) = 3752, \quad n = 20, \quad I^* = 2.4148.$   
 d)  $M_4 = |f^{(4)}(0)| = 15, \quad n = 8, \quad I^* = 1.3987.$   
 e)  $M_4 = |f^{(4)}(0.5)| \simeq 120, \quad n = 6, \quad I^* = 0.0453.$   
 f)  $M_4 = |f^{(4)}(\pi/4)| \simeq 0.2, \quad n = 2, \quad I^* = 0.6118.$   
 g)  $M_4 = |f^{(4)}(2)| \simeq 8, \quad n = 14, \quad I^* = 1.0473.$

**Zadatak 7.6** Duljina elipse  $\frac{x^2}{a^2} + y = 1$  ( $a > 0$ ) jednaka je:

$$I = 4 \int_0^{\pi/2} \sqrt{1 - (1 - a^2) \sin^2 t} dt.$$

Za  $a \neq 1$  ne može se elementarno naći primitivna funkcija ove podintegralne funkcije. Za  $a = 0.5$  i  $a = 1.5$  izračunajte vrijednost ovog integrala primjenom generaliziranog trapeznog i Simpsonovog pravila uz  $\epsilon = 0.00005$ .

Rješenje: Za  $a = 0.5$  dobivamo:

- a) Trapezno pravilo:  $M_2 = f''(\pi/2) = 6, \quad n = 197, \quad I^* = 4.8442;$   
 b) Simpsonovo pravilo:  $M_4 = |f^{(4)}(\pi/2)| = 78, \quad n = 18, \quad I^* = 0.8442;$

**Zadatak 7.7** Za podintegralnu funkciju u nepravom integralu:

$$I(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt$$

ne postoji elementarna primitivna funkcija. Za  $x = 0.5, 1, 5$  primjenom trapeznog ili Simpsonovog pravila približno izračunajte vrijednost ovog integrala s točnošću  $\epsilon = 0.00005$ .

Uputa: koristite činjenicu da je  $I(x) = 0.5 + \frac{1}{\sqrt{2\pi}} \int_0^x \exp\left(-\frac{t^2}{2}\right) dt$

Rješenje:  $I^*(0.5) = 0.6914, \quad I^*(1) = 0.8414, \quad I^*(5) = 1.$  Pri tome za sva tri intervala integracije imamo:

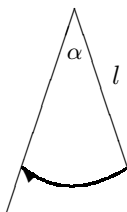
$$M_2 = |f''(0)| \simeq 0.4, \quad M_4 = |f^{(4)}(0)| \simeq 1.2.$$

Potreban broj podintervala ( $n$ ) vidi se u sljedećoj tablici:

$n$	$x = 0.5$	$x = 1$	$x = 5$
Trapezno pravilo	10	26	289
Simpsonovo pravilo	2	4	26

**Zadatak 7.8** Matematičko njihalo s kutem odklona  $\alpha$  ima period  $T = 4K\sqrt{\frac{l}{g}}$  gdje je:

$$K = \int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - \sin^2 \frac{\alpha}{2} \sin^2 \phi}}.$$



Elementarnim metodama ne može se odrediti primitivna funkcija za ovu podintegralnu funkciju. Za  $\alpha = 5^\circ, 10^\circ, 45^\circ$  primjenom trapeznog ili Simpsonovog pravila približno izračunajte vrijednost ovog integrala s točnošću  $\epsilon = 0.00005$ .

Rješenje:

- a)  $\alpha = 5^\circ$ ,  $M_2 = |f''(0)| \simeq 0.0075$ ,  $M_4 = |f^{(4)}(0)| \simeq 0.03$ ,  $K^* = 1.56783$ .  
 b)  $\alpha = 10^\circ$ ,  $M_2 = |f''(0)| \simeq 0.03$ ,  $M_4 = |f^{(4)}(0)| \simeq 0.13$ ,  $K^* = 1.55915$ .  
 c)  $\alpha = 45^\circ$ ,  $M_2 \simeq 0.5$ ,  $M_4 \simeq 4.25$ ,  $K^* = 1.41574$ .

Potreban broj podintervala ( $n$ ) vidi se u tablici:

$n$	$\alpha = 5^\circ$	$\alpha = 10^\circ$	$\alpha = 45^\circ$
Trapezno pravilo	7	14	57
Simpsonovo pravilo	4	4	10

**Zadatak 7.9** Širina rijeke je 20 m. Mjerenjem dubina na svakih 2 m njenog poprečnog presjeka dobiveni su sljedeći podaci:

$x_i(\text{m})$	0	2	4	6	8	10	12	14	16	18	20
dubina(m)	0.2	0.5	0.9	1.1	1.3	1.7	2.1	1.5	1.1	0.6	0.2

Primjenom trapeznog i Simpsonovog pravila izračunajte površinu poprečnog presjeka rijeke.

Uputa: Postupite slično kao u *Primjeru 7.2*, str. 129, te iskoristite navedene *Mathematica*-module.

Rješenje:

- a) Trapezno pravilo :  $S^* = 22 \text{ m}^2$   
 b) Simpsonovo pravilo :  $S^* = 21.9 \text{ m}^2$

**Zadatak 7.10** Izvedite Newton-Cottesovu formulu za  $n = 3$ . Primijenite dobiveni rezultat kod numeričke integracije.

Rješenje:  $I^* = \frac{b-a}{8} \left( f(a) + 3f\left(\frac{2a+b}{3}\right) + 3f\left(\frac{a+2b}{3}\right) + f(b) \right)$

## 8. Numeričko rješavanje običnih diferencijalnih jednačbi

Prilikom rješavanja diferencijalnih jednačbi u praktičnim primjenama vrlo često nailazimo na slučajeve, koje ne možemo elementarno riješiti ili bi elementarno rješenje bilo previše složeno. Tada problem možemo pokušati riješiti numeričkim metodama. Primjenom numeričkih metoda za rješavanje diferencijalnih jednačbi s početnim ili rubnim uvjetom, dobivamo tablično zadanu funkciju, pri čemu korak u vrijednosti nezavisne varijable možemo dovoljno fino izabrati.

Promatramo sljedeći problem:

– za zadanu funkciju  $f(x, y)$  treba naći funkciju  $y(x)$ ,  $x \in [x_0, b]$ , koja zadovoljava diferencijalnu jednačbu 1. reda:

$$\frac{dy}{dx} = f(x, y), \quad (8.1)$$

uz početni uvjet:

$$y(x_0) = y_0. \quad (8.2)$$

Ovaj problem nazivamo Cauchyjev problem ili inicijalni problem<sup>1</sup>. U nekim slučajevima (primjerice, ako je funkcija  $f$  linearna) ovaj problem može se riješiti elementarno (kao u *Primjeru 8.1*). Međutim, u većini praktičnih problema prisiljeni smo tražiti aproksimativno rješenje. Štoviše, u praktičnim istraživanjima obično se javlja problem rješavanja sustava diferencijalnih jednačbi:

$$\frac{dy_i}{dx} = f_i(x, y_1, \dots, y_k), \quad i = 1, \dots, k,$$

s početnim uvjetima:  $y_i(x_0) = c_i$ ,  $i = 1, \dots, k$ , ili u vektorskom obliku:

$$\frac{d\mathbf{y}}{dx} = \mathbf{f}(x, \mathbf{y}), \quad \mathbf{y}(x_0) = \mathbf{c}, \quad (8.3)$$

gdje je:  $\mathbf{y} = (y_1, \dots, y_k)^T$ ,  $\mathbf{c} = (c_1, \dots, c_k)^T$ ,  $\mathbf{f} = (f_1, \dots, f_k)^T$ .

---

<sup>1</sup>Engl. = initial value problem, njem. = Anfangswertaufgabe



**Primjedba 8.1** Diferencijalna jednačba višeg reda može se svesti na sustav diferencijalnih jednačbi 1. reda. Primjerice diferencijalna jednačba 3. reda:

$$y''' = g(x, y, y', y''),$$

s početnim uvjetima:  $y(0) = c_1$ ,  $y'(0) = c_2$ ,  $y''(0) = c_3$ , uz supstitucije:

$$z_1 = y, \quad z_2 = y', \quad z_3 = y'',$$

prelazi u sustav diferencijalnih jednačbi 1. reda:

$$\begin{aligned} z_1' &= z_2, & z_1(0) &= c_1 \\ z_2' &= z_3, & z_2(0) &= c_2 \\ z_3' &= g(x, z_1, z_2, z_3), & z_3(0) &= c_3. \end{aligned}$$

Zato ima smisla posebnu pozornost posvetiti rješavanju Cauchyjevog problema (8.1)-(8.2), odnosno (8.3).

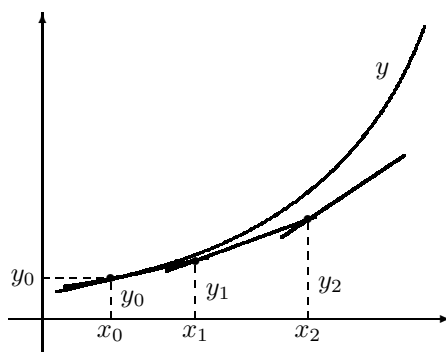
## 8.1 Eulerova metoda

Promatramo problem (8.1)-(8.2). Kako je zbog (8.1) u točki  $x_0$  poznat koeficijent smjera tangente na traženu funkciju  $x \mapsto y(x)$ :

$$y'(x_0) = f(x_0, y_0),$$

onda možemo funkciju  $y$  u okolini točke  $x_0$  aproksimirati linearnim aproksimantom, čiji je graf tangenta (*vidi Sliku 8.1*):

$$y = y_0 + f(x_0, y_0)(x - x_0). \quad (8.4)$$



Slika 8.1. Eulerova metoda

Za maleni  $h$ , točka  $x_1 = x_0 + h$  je bliska točki  $x_0$ , pa ćemo vrijednost tražene funkcije  $y$  u točki  $x_1$  aproksimirati vrijednošću linearnog aproksimanta (8.4)

$$y_1 \equiv y(x_1) = y_0 + f(x_0, y_0)h.$$

Ponavljajući postupak, dobivamo Eulerov iterativni proces:

$$y_{k+1} = y_k + f(x_k, y_k)h, \quad k = 0, 1, \dots \quad (8.5)$$

koji daje aproksimaciju funkcije  $y$  u točkama:

$$x_i = x_0 + i h, \quad i = 0, 1, \dots$$

Eulerova metoda<sup>2</sup> ne upotrebljava se za rješavanje praktičnih problema. Ako duljina koraka  $h$  nije jako malena, pogreške se mogu brzo akumulirati, a ako odaberemo potrebno maleni korak  $h$ , broj iteracija postaje nerazumno velik (*vidi Primjer 8.2*). Ipak, ova metoda ima važno teorijsko značenje, jer se druge, mnogo efikasnije metode, zasnivaju na ovoj ideji.

**Primjer 8.1** Treba riješiti Cauchyjev problem:

$$y' = x + y, \quad y(0) = 1,$$

na intervalu  $[0, 1]$ .

Rješenje problema je funkcija  $y = 2e^x - x - 1$ . Potražiti ćemo rješenje i pomoću Eulerove metode uz korake  $h = .2$  i  $.1$  uz korištenje *Mathematica*-modula:

```
Euler[a_, b_, h_, f_, y0_] := Module[{i, x, y}, n = Floor[(b - a) / h];
  If[IntegerQ[n], n = n, n = n + 1]; tab = Table[{a + (i - 1) h, 0}, {i, n + 1}];
  x = a; y = y0; tab[[1, 2]] = y0;
  Do[
    x = x + h; y = y + f[x, y] h; tab[[i, 2]] = y
    , {i, 2, n + 1}]
```

Dobiveni rezultati, kao i usporedba sa stvarnim vrijednostima prikazani su u *Tablici 8.1*.

$x_k$	$y(x_k)$	$h = 0.2$		$h = 0.1$	
		$y_k$	$y(x_k) - y_k$	$y_k$	$y(x_k) - y_k$
0	1	1	0	1	0
0.1	1.11034			1.11	-.00034
0.2	1.24281	1.24	-.00281	1.241	-.00181
0.3	1.39972			1.3951	-.00462
0.4	1.58365	1.568	-.01565	1.57461	-.00904
0.5	1.79744			1.78207	-.01537
0.6	2.04424	2.0016	-.04264	2.02028	-.02396
0.7	2.32751			2.29231	-.03520
0.8	2.65108	2.56192	-.08916	2.60154	-.04954
0.9	3.01921			2.95169	-.06752
1	3.43656	3.2743	-.16226	3.34686	-.08970

Tablica 8.1. Rješavanje inicijalnog problema  $y' = x + y$ ,  $y(0) = 1$

<sup>2</sup>Ovu metodu neki puta naći ćemo u literaturi pod imenom: **metoda tangenti** ili Cauchyjeva metoda jer je Cauchy otprilike 100 godina kasnije (1828.) dokazao konvergenciju ove metode

**Primjer 8.2** *Promatrajmo jednostavni Cauchyjev problem:*

$$y' = y, \quad y(0) = 1.$$

U ovom slučaju Eulerov iterativni postupak (8.5) postaje:

$$y_{k+1} = y_k(1 + h), \quad k = 0, 1, \dots$$

ili općenito, za neki  $n \in \mathbb{N}$  imamo

$$y_n = y_{n-1}(1 + h) = y_{n-1}(1 + h)^2 = \dots = y_0(1 + h)^n = (1 + h)^n.$$

Da bismo izračunali vrijednost tražene funkcije  $y$  u nekoj fiksnoj točki  $x > 0$ , u intervalu  $[0, x]$  odredit ćemo  $n$  jednoliko raspoređenih točaka  $x_i$  s razmakom  $h = \frac{x}{n}$ . Primjenom Eulerove metode dobivamo

$$y(x) \approx y_n = \left(1 + \frac{x}{n}\right)^n. \quad (*)$$

Kada pustimo  $n \rightarrow \infty$  (tj.  $h \rightarrow 0$ ), dobivamo egzaktno rješenje:  $y(x) = \exp(x)$ , što znači da Eulerova metoda konvergira.

U svrhu ispitivanja brzine konvergencije Eulerove metode u ovom slučaju raspišimo (\*) pomoću Taylorove formule

$$y(x) \approx \exp\left(n \ln\left(1 + \frac{x}{n}\right)\right) = \exp\left(x - \frac{x^2}{2n} + \dots\right) \approx e^x e^{-\frac{x^2}{2n}}.$$

Tako dobivamo približnu pogrešku aproksimacije u fiksnoj točki  $x$ :

$$\Delta y(x) = |y(x) - e^x| \approx e^x \left(1 - e^{-\frac{x^2}{2n}}\right).$$

Tako primjerice, ako želimo vrijednost funkcije  $y$  u točki  $x = 1$  dobiti s jednom točnom decimalom ( $\epsilon = 0.05$ ), potrebno je staviti  $n = 27$  (odnosno  $h \approx 0.04$ ). Za 5 točnih decimala ( $\epsilon = 0.5 \times 10^{-5}$ ) treba staviti  $n = 271\,828$  ( $h \approx 0.000004$ ).

**Primjedba 8.2** *Kod Eulerove metode sljedeću aproksimaciju funkcije uvijek izračunavamo samo na bazi jedne prethodne aproksimacije. Zato kažemo da je to jednokoračna metoda. Eulerova metoda može se poboljšati pomoću tzv. simetrične formule*

$$y_{k+1} = y_{k-1} + 2hf(x_k, y_k), \quad k = 0, 1, \dots$$

*Simetrična formula je dvokoračna metoda.*

*Osim navedenih, postoje i brojne druge metode za numeričko rješavanje Cauchyjevog problema, kao primjerice: Richardson-ove ekstrapolacije (specijalno Bulirsch-Stoerova metoda), višekoračne metode: predictor-corrector metoda, Milneova metoda, Adams-ova metoda itd. (vidi primjerice GEAR (1971), ORTEGA (1981), STOER (1993)).*

## 8.2 Metoda Runge - Kutta

Promatramo Cauchyjev problem (8.1)-(8.2):

$$y' = f(x, y), \quad y(x_0) = y_0.$$

Posebno važno mjesto za rješavanje ovog problema u primjenama zauzima dobro poznata Runge–Kutta (RK) metoda<sup>3</sup>. Pretpostavimo da poznajemo aproksimaciju  $y_n$  tražene funkcije  $x \mapsto y(x)$  u točki  $x_n$ . Želimo odrediti  $(n+1)$ -vu aproksimaciju  $y_{n+1}$  u točki  $x_n + h$ . U tu svrhu na intervalu  $(x_n, x_n + h)$  u nekoliko strateških točaka aproksimirat ćemo vrijednost funkcije  $x \mapsto f(x, y(x))$ , te pomoću njih što bolje aproksimirati razliku  $y_{n+1} - y_n$ . Egzaktni izvod i analiza ovih metoda izlazi iz okvira ove skripte, a može se naći primjerice kod SCHWARZ (1986)) STOER (1993).

Najjednostavniji primjer iz familije RK metoda je tzv. Heunova metoda

$$\begin{aligned} y_{n+1} &= y_n + \frac{1}{2}(k_1 + k_2), \\ k_1 &= hf(x_n, y_n), \\ k_2 &= hf(x_n + h, y_n + k_1). \end{aligned} \quad (8.6)$$

**Primjedba 8.3** *Kako je*

$$y(x_{n+1}) - y(x_n) = \int_{x_n}^{x_n+h} \frac{dy}{dx} dx = \int_{x_n}^{x_n+h} f(x, y(x)) dx, \quad (8.7)$$

ako pretpostavimo da je  $f$  samo funkcija od  $x$ , Heunova metoda (8.6) odgovara trapeznom pravilu (7.5), str. 127, koja ima pogrešku reda veličine  $O(h^2)$ . Primjetimo također da je za svaku aproksimaciju  $y_n$  potrebno dva puta izračunavati vrijednost funkcije  $f$ .

Klasična RK metoda definirana je s

$$\begin{aligned} y_{n+1} &= y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4), \\ k_1 &= hf(x_n, y_n), \\ k_2 &= hf(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}), \\ k_3 &= hf(x_n + \frac{h}{2}, y_n + \frac{k_2}{2}), \\ k_4 &= hf(x_n + h, y_n + k_3) \end{aligned} \quad (8.8)$$

**Primjedba 8.4** *Ako pretpostavimo da je  $f$  samo funkcija od  $x$ , onda iz (8.7) možemo pokazati da RK-metoda (8.8) odgovara Simpsonovoj formuli (7.14), str. 131, uz zamjenu  $h \mapsto \frac{h}{2}$ . Sjetimo se da Simpsonova formula ima pogrešku reda veličine  $O(h^2)$ , što se prenosi i na RK metodu (8.8) i u općem slučaju – kada je  $f$  funkcija od  $x$  i od  $y$ . Primijetimo također da je kod RK metode (8.8) za svaku aproksimaciju  $y_n$  potrebno četiri puta izračunavati vrijednost funkcije  $f$ .*

<sup>3</sup>Ideju je prvi izložio C. Runge u radu *Über die numerische Auflösung von Differentialgleichungen*, Mathematische Annalen **46**(1895), 167–178, a kasnije razvio W. Kutta u radu *Beitrag zur näherungsweise Integration von Differentialgleichungen*, Zeitschrift für Mathematik und Physik **46**(1901), 435–453

**Primjer 8.3** Na primjeru Cauchyjevog problema

$$y' = 30(\sin x - y), \quad y(0) = 0,$$

čije je egzaktno rješenje lako dobiti (linearna diferencijalna jednačba !),

$$y(x) = \frac{30}{901} (30 \sin x - \cos x + e^{-30x}),$$

na intervalu  $[0, 1]$  usporedit ćemo efikasnost Eulerove i RK metode (8.8).

Rezultati prikazani u Tablici 8.2 dobiveni su korištenjem niže navedenog *Mathematica*-modula

```

RK[a_, b_, h_, f_, y0_] := Module[{i, x, y, k1, k2, k3, k4},
  n = Floor[(b-a)/h];
  If[IntegerQ[n], n=n, n=n+1]; tab=Table[{a+(i-1) h, 0}, {i, n+1}];
  x=a; y=y0; tab[[1,2]] = y0;
  Do[
    k1 = h f[x, y];          k2 = h f[x+h/2, y+k1/2];
    k3 = h f[x+h/2, y+k2/2]; k4 = h f[x+h, y+k3];
    y = y + (1/6) (k1 + 2 k2 + 2 k3 + k4);
    tab[[i, 2]] = y;      x = x+h, {i, 2, n+1}
  ]

```

		Eulerova metoda		RK metoda	
$x_k$	$y(x_k)$	$y_k$	$y(x_k) - y_k$	$y_k$	$y(x_k) - y_k$
0	0	0	0	0	0
0.1	0.068250	0.29950	-0.23125	0.112391	-.04414
0.2	0.165899	-0.00299	0.16889	0.228523	-.06262
0.3	0.263387	0.89255	-0.62915	0.349063	-.08567
0.4	0.358318	-0.61684	0.97515	0.475312	-.11699
0.5	0.449673	2.67195	-2.22228	0.60946	-.15979
0.6	0.536535	-3.64997	4.18651	0.754908	-.21837
0.7	0.618036	9.23259	-8.61456	0.916725	-.29869
0.8	0.693362	-16.31310	17.00650	1.10226	-.40890
0.9	0.76176	34.97620	-34.21450	1.32199	-.56023
1	0.822547	-67.42800	68.25060	1.59067	-.76813

Tablica 8.2  $y' = 30(\sin x - y)$ ,  $y(0) = 0$ ,  $x \in [0, 1]$

**Primjer 8.4** Treba riješiti Cauchyjev problem:

$$y' = xy^2 + 1, \quad y(0) = 0,$$

na intervalu  $[0, 1]$ .

Primjenom RK metode (8.8) dobivamo rezultate vidljive u *Tablici 8.3*.

$x_k$	$y_k$
0	0
0.1	0.1
0.2	0.2
0.3	0.302
0.4	0.4065
0.5	0.5162
0.6	0.6345
0.7	0.7666
0.8	0.92
0.9	1.1073
1	1.3503

Tablica 8.3  $y' = xy^2 + 1$ ,  $y(0) = 0$

$t_k$	$v_k$	$v(t_k)$
0	0	0
2	18.568	18.573
4	32.193	32.211
6	40.093	40.123
8	44.070	44.1
10	45.933	45.955
12	46.766	46.79
14	47.152	47.16
16	47.319	47.323
18	47.391	47.394
20	47.424	47.425

Tablica 8.4  $\dot{v} = g - \frac{\gamma}{m}v^2$ ,  $v(0) = 0$

**Primjer 8.5** *Promatramo slobodni pad tijela mase  $m$  u zraku.*

Uz pretpostavku da je otpor zraka proporcionalan kvadratu brzine, vrijedi jednadžba gibanja

$$m\ddot{s} = mg - \gamma\dot{s}^2, \quad (*)$$

gdje je  $g = 9.81 \text{ ms}^{-2}$ ,  $\gamma = C\delta S$ , gdje je  $S$  površina poprečnog presjeka tijela ( $\text{m}^2$ ),  $\delta$  specifična gustoća zraka ( $\text{kg m}^{-3}$ ), a  $C$  konstanta bez dimenzije koja ovisi o obliku tijela - tzv. form-faktor (vidi MOLITZ (1967)). Veličina  $\delta$  u jednadžbi (\*) je funkcija temperature (u K) i visine ( $s$ ), ali uz pretpostavku da se radi o malim visinama (nekoliko stotina metara), možemo pretpostaviti da je ona konstantna i jednaka svojoj vrijednosti na razini mora  $\delta_0 = 1.22 \text{ kgm}^{-3}$ .

Uz oznake:  $v = \dot{s}$  i  $\frac{dv}{dt} = \ddot{s}$ , jednadžba (\*) glasi:

$$\frac{dv}{dt} = g - \frac{\gamma}{m}v^2,$$

čije rješenje je:

$$v = b \tanh(abt), \quad (**)$$

gdje je:  $a = \gamma/m$ ,  $b^2 = mg/\gamma$ . Primijetimo da za  $t \rightarrow \infty$  brzina  $v$  teži konačnoj veličini  $b$ .

Za  $m = 70 \text{ kp}$ ,  $S = 0.25 \text{ m}^2$  i  $C = 1$  primjenom RK metode riješit ćemo jednadžbu (\*), na intervalu  $[0, 20]$  (pad tijela pratimo prvih 20 sekundi) uz korak u metodi  $h = 2$ . Dobiveni rezultati u *Tablici 8.4* uspoređeni su s egzaktnim vrijednostima iz jednadžbe (\*\*).

**Primjedba 8.5** *Runge-Kuttovom metodom možemo rješavati i sustave diferencijalnih jednadžbi. Postupak je analogan iterativnom procesu (8.8). Primjerice za sustav:*

$$y' = f(x, y, z), \quad z' = g(x, y, z), \quad y(x_0) = y_0, \quad z(x_0) = z_0,$$

imamo:

$$\begin{aligned} y_1 &= y_0 + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4), \\ z_1 &= z_0 + \frac{1}{6}(m_1 + 2m_2 + 2m_3 + m_4), \end{aligned} \quad (8.9)$$

gdje je:

$$\begin{aligned} k_1 &= hf(x_0, y_0, z_0) & m_1 &= hg(x_0, y_0, z_0) \\ k_2 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{k_1}{2}, z_0 + \frac{m_1}{2}\right), & m_2 &= hg\left(x_0 + \frac{h}{2}, y_0 + \frac{k_1}{2}, z_0 + \frac{m_1}{2}\right), \\ k_3 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{k_2}{2}, z_0 + \frac{m_2}{2}\right), & m_3 &= hg\left(x_0 + \frac{h}{2}, y_0 + \frac{k_2}{2}, z_0 + \frac{m_2}{2}\right), \\ k_4 &= hf(x_0 + h, y_0 + k_3, z_0 + m_3), & m_4 &= hg(x_0 + h, y_0 + k_3, z_0 + m_3). \end{aligned}$$

Primjerice, za funkcije  $f(x, y, z) = x + z$ ,  $g(x, y, z) = -x$  i početni uvjet  $y(0) = z(0) = 1$  na intervalu  $[0, 1]$  dobivamo vrijednosti funkcija  $x \mapsto y(x)$  i  $x \mapsto z(x)$  prikazane u Tablici 8.5.

$x_k$	$y_k$	$z_k$
0	1	1
0.1	1.1048	0.995
0.2	1.2187	0.98
0.3	1.3405	0.955
0.4	1.4693	0.92
0.5	1.6042	0.875
0.6	1.744	0.82
0.7	1.8878	0.755
0.8	2.0347	0.68
0.9	2.1835	0.595
1	2.333	0.5

Tablica 8.5

$x_k$	$y_k$	$y'_k$
0	1	0
0.2	1.0202	0.2040
0.4	1.0833	0.4333
0.6	1.1972	0.7183
0.8	1.3771	1.1017
1.0	1.6487	1.6487
1.2	2.0543	2.4652
1.4	2.6642	3.7298
1.6	3.5960	5.7536
1.8	5.0518	9.0932
2	7.3862	14.7724

Tablica 8.6

**Primjedba 8.6** Diferencijalna jednadžba višeg reda rješava se svodenjem na sustav diferencijalnih jednadžbi 1. reda. Tako primjerice Cauchyjev problem:

$$y'' = g(x, y, y'), \quad y(x_0) = \alpha, \quad y'(x_0) = \beta,$$

svodimo na sustav:

$$\begin{aligned} y' &= z, & y(x_0) &= y_0, \\ z' &= g(x, y, z), & z(x_0) &= z_0, \end{aligned}$$

koji možemo riješiti RK metodom:

$$\begin{aligned} y_1 &= y_0 + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4), \\ z_1 &= z_0 + \frac{1}{6}(m_1 + 2m_2 + 2m_3 + m_4), \end{aligned}$$

$$\begin{aligned} k_1 &= h z_0 & m_1 &= hg(x_0, y_0, z_0) \\ k_2 &= h\left(z_0 + \frac{m_1}{2}\right), & m_2 &= hg\left(x_0 + \frac{h}{2}, y_0 + \frac{k_1}{2}, z_0 + \frac{m_1}{2}\right), \\ k_3 &= h\left(z_0 + \frac{m_2}{2}\right), & m_3 &= hg\left(x_0 + \frac{h}{2}, y_0 + \frac{k_2}{2}, z_0 + \frac{m_2}{2}\right), \\ k_4 &= h(z_0 + m_3), & m_4 &= hg(x_0 + h, y_0 + k_3, z_0 + m_3). \end{aligned}$$

**Primjer 8.6** Treba riješiti Cauchyjev problem:

$$y'' - xy' - y = 0, \quad y(0) = 1, \quad y'(0) = 0.$$

Problem ćemo pretransformirati na sustav diferencijalnih jednadžbi 1. reda

$$\begin{aligned} y' &= z, & y(0) &= 1, \\ z' &= xz + y, & z(0) &= 0, \end{aligned}$$

a rješenje potražiti na intervalu  $[0, 2]$  primjenom RK metode s korakom  $h = 0.2$  (vidi Tablicu 8.6).

**Primjer 8.7** Tzv. Lotka–Volterrin biološki model grabežljivac - plijen opisuje se sustavom diferencijalnih jednadžbi (vidi primjerice GEAR (1971)):

$$\frac{dx}{dt} = x(p_1 - p_2y), \quad \frac{dy}{dt} = y(p_3x - p_4).$$

Uz početne uvjete  $x(0) = 1$ ,  $y(0) = 0.3$  i vrijednosti parametara:

$$p_1 = 0.77069, \quad p_2 = p_3 = 2.1268, \quad p_4 = 1.984,$$

primjenom RK metode dobivamo funkcije  $x$  i  $y$  tablično prikazane u Tablici 8.7 i na Slici 8.2.a.

$t$	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5
$x$	1.0	1.052	1.060	1.013	0.934	0.860	0.819	0.8198	0.854	0.915
$y$	0.3	0.332	0.381	0.427	0.446	0.428	0.386	0.341	0.307	0.291
$t$	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5
$x$	0.986	1.044	1.062	1.026	0.950	0.873	0.824	0.816	0.844	0.901
$y$	0.296	0.324	0.370	0.419	0.445	0.434	0.396	0.350	0.312	0.293

Tablica 8.7 Podaci za Lotka–Volterrin model: grabežljivac - plijen

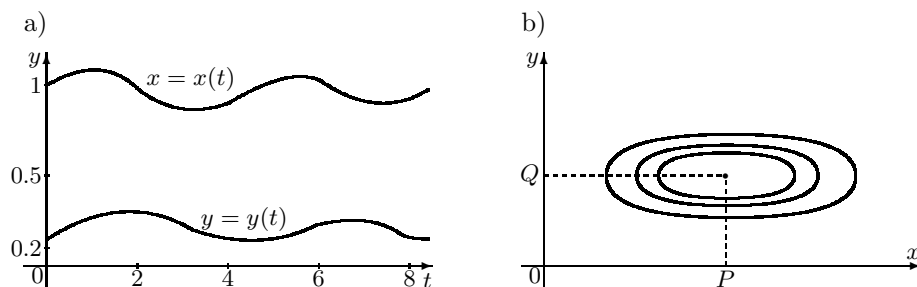
Možemo potražiti i tzv. fazno rješenje, odnosno opće rješenje  $y = y(x; c)$  rješavajući jednadžbu:

$$\frac{dy}{dx} = \frac{y(p_3x - p_4)}{x(p_1 - p_2y)}.$$

Uz oznake:  $P = p_4/p_3$ ,  $Q = p_1/p_2$ , dobivamo familiju elipsa (vidi također Sliku 8.2.b):

$$\frac{(x - P)^2}{p_4} + \frac{(y - Q)^2}{p_1} = c.$$

Konstanta  $c$  određuje se iz početnog uvjeta:  $x(0) = x_0$ ,  $y(0) = y_0$ , čime je izabrana jedna elipsa. Funkcije  $x(t)$  i  $y(t)$  su vremenske periodične funkcije.



Slika 8.2 Lotka–Volterrin model: grabežljivac - plijen



### 8.3 Metoda diskretizacije

Osim inicijalnih problema koje smo do sada razmatrali, rješavanje praktičnih problema često vodi na rješavanje tzv. **rubnih problema**<sup>4</sup>.

Za danu funkciju  $f(x, y, y')$  treba odrediti funkciju  $x \mapsto y(x)$ ,  $x \in [a, b]$ , koja zadovoljava diferencijalnu jednadžbu 2. reda:

$$y'' = f(x, y, y'), \quad (8.10)$$

uz tzv. rubne uvjete:

$$y(a) = \alpha, \quad y(b) = \beta, \quad (a < b). \quad (8.11)$$

Postoji također više metoda za rješavanje rubnih problema (vidi primjerice STOER, BULIRSCH (1993)). Mi ćemo ovdje samo kratko opisati tzv. **metodu diskretizacije**.

Interval  $[a, b]$  podijelit ćemo na  $n$  jednakih dijelova, stavljajući:

$$h = \frac{b-a}{n}, \quad x_i = a + ih.$$

Prvu ( $y'$ ) i drugu ( $y''$ ) derivaciju funkcije  $y$  možemo aproksimirati s:

$$y'(x_k) \approx \frac{y_{k+1} - y_{k-1}}{2h}, \quad y''(x_k) \approx \frac{y_{k+1} - 2y_k + y_{k-1}}{h^2}.$$

Na taj način rubni problem (8.10)-(8.11) svodi se na rješavanje sustava tzv. **diferencijskih jednadžbi**:

$$y_{k+1} - 2y_k + y_{k-1} = h^2 f_k, \quad k = 1, 2, \dots, n-1, .$$

$$y_0 = \alpha, \quad y_n = \beta,$$

gdje je:

$$f_k = f\left(x_k, y_k, \frac{y_{k+1} - y_{k-1}}{2h}\right).$$

ili u matričnom obliku:

$$\mathbf{A} \mathbf{y} = h^2 \mathbf{F}(\mathbf{y}) - \mathbf{r}, \quad (8.12)$$

gdje je:

$$\mathbf{A} = \begin{bmatrix} -2 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \cdots & 0 \\ 0 & 1 & -2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \end{bmatrix}, \quad \mathbf{F}(\mathbf{y}) = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{n-1} \end{bmatrix}, \quad \mathbf{r} = \begin{bmatrix} \alpha \\ 0 \\ \vdots \\ \beta \end{bmatrix}.$$

<sup>4</sup>Engl. = two point boundary value problem, njem. = Randwertaufgabe

Ako je funkcija  $f$  linearna, sustav (8.12) je sustav linearnih jednačbi i možemo ga pokušati riješiti nekom metodom navedenom u t.3, str.38 (posebno vidi *Zadatak 3.24*, str. 65). Ako je funkcija  $f$  nelinearna, onda treba koristiti neku iterativnu metodu za rješavanje sustava nelinearnih jednačbi (t.4.4, str. 85).

**Primjedba 8.7** Postoji opsežna literatura o numeričkom rješavanju običnih diferencijalnih jednačbi (vidi npr. COLLATZ (1966), ORTEGA (1981), STOER (1993)). FORTRAN-programaska podrška iz ovog područja može se naći u FORSYTHE (1974), PRESS (1989) i u *The Nag-Library* (Philips (1986)). U programu Mathematica mogu se naći gotovi moduli za rješavanje ovakvih problema, te svi potrebni prateći “alati”.

## 8.4 Zadaci

**Zadatak 8.1** Numerički riješite diferencijalnu jednačbu s početnim uvjetom na intervalu  $[a, b]$ . Numerički dobiveno rješenje usporedite s egzaktnim, te nacrtajte pripadne grafove.

- a)  $y' + \frac{1}{x}y = \frac{\sin x}{x}$ ,  $y(\frac{\pi}{2}) = 1$ ,  $x \in [\frac{\pi}{2}, \pi]$ ,
- b)  $y' = 3x^2y$ ,  $y(0) = 2$ ,  $x \in [0, 1]$ ,
- c)  $x + yy' = 0$ ,  $y(0) = 2$ ,  $x \in [0, 2]$ ,
- d)  $2xy' = y$ ,  $y(1) = 2$ ,  $x \in [1, 3]$ ,
- e)  $y' + y = 0$ ,  $y(0) = 1$ ,  $x \in [0, 0.6]$ ,
- f)  $y' = y_2$ ,  $y(0) = 0.25$ ,  $x \in [0, 3]$ ,
- g)  $y' = 1 + xy + y^2$ ,  $y(0) = 0$ ,  $x \in [0, 2]$

Rješenje:

- a)  $y = -\frac{\cos x}{x} + \frac{\pi}{2x}$ , b)  $y = 2 \exp(x^3)$ , c)  $x^2 + y^2 = 4$ , d)  $y^2 = 4x$ ,
- e)  $y = e^{-x}$ , f)  $y = 1/(4 - x)$ , g)

**Zadatak 8.2** Primjenom Runge-Kutta metode riješite Cauchyjev problem:

$$y' = 100(\sin x - y), \quad y(0) = 0,$$

na intervalu  $[0, \pi/2]$  za različite veličine koraka  $h = 0.1, 0.01, 0.001$ . Dobivene rezultate usporedite s egzaktnim rješenjem:

$$y = \frac{1}{1.0001}(\sin x - 0.01 \cos x + 0.01e^{-100x}).$$

Nacrtajte odgovarajuće grafove funkcije-rješenja i funkcije-aproksimacije rješenja.

**Zadatak 8.3** Numerički riješite Cauchyjev problem na intervalu  $I = [0, 1]$ :

- a)  $y'' = \sqrt{1 + y'^2}$ ,  $y(0) = 1$ ,  $y'(0) = 0$ ,
- b)  $y'' - xy' + y = 0$ ,  $y(0) = 0$ ,  $y'(0) = 1$ .

Rješenje: a)  $y = \operatorname{ch} x$ , b)

**Zadatak 8.4** Numerički riješite sustav diferencijalnih jednadžbi:

$$\begin{array}{ll} \text{a)} & \begin{array}{l} y' = -7y + z, \quad y(0) = 0, \\ z' = -2y - 5z, \quad z(0) = 1, \\ x \in [0, \pi] \end{array} \\ \text{b)} & \begin{array}{l} y' = z, \quad y(1) = 1 \\ z' = -y, \quad z(1) = 1 \\ x \in [1, 2] \end{array} \end{array}$$

**Zadatak 8.5** Primjenom Runge-Kutta metode riješite Cauchyjev problem:

a)  $y'' + 1001y' + 1000y = 0, \quad y(0) = 0, \quad y'(0) = 1, \quad x \in [0, 0.2].$

b)  $y'' + y = x, \quad y(0) = 1, \quad y'(0) = 2, \quad x \in [0, 1].$

za različite veličine koraka  $h = 0.01, 0.005, 0.001$ . Dobivene rezultate usporedite s egzaktnim rješenjem, te nacrtajte odgovarajuće grafove funkcije rješenja i funkcije-aproksimacije rješenja.

Rješenje: a)  $y = (e^{-x} - e^{-1000x})/999$ , b)  $y = \cos x - \sin x + x$ .

**Zadatak 8.6** Rješenje rubnog problema:

$$y'' + y = x, \quad y(0) = 1, \quad y(\pi/2) = \pi/2 - 1,$$

je funkcija:

$$y = \cos x - \sin x + x.$$

Riješite ovaj problem metodom diskretizacije koja je spomenuta u t.8.3, str. 146 uzevši  $m = 5$  ili  $m = 10$ , te usporedite rješenja s egzaktnim rješenjem. Nacrtajte grafove funkcije rješenja i funkcije-aproksimacije rješenja.

**Zadatak 8.7** Za koju vrijednost parametra  $\lambda$  rubni problem:

$$y'' + \lambda y = 0, \quad y(0) = 0, \quad y(1) = 0,$$

ima netrivialno rješenje ( $y \neq 0$ ).

Rješenje:  $\lambda = n^2\pi^2, n = 1, 2, \dots$

**Zadatak 8.8** Zadatak 8.7 riješite metodom diskretizacije uz  $n = 3$ .

Rješenje:

$$\lambda = 9 \text{ (prava vrijednost: } \pi^2 = 9.8696),$$

$$\lambda = 27 \text{ (prava vrijednost: } 4\pi^2 = 39.48).$$

**Zadatak 8.9** S podacima iz Primjera 8.5, str. 143 odredite put  $s$  kao funkciju vremena  $t$  na intervalu vremena  $[0, 20]$ .

Uputa: riješite diferencijalnu jednadžbu  $\frac{ds}{dt} = b \operatorname{th}(abt)$ .

**Zadatak 8.10** S podacima iz Primjera 8.5, str. 143 odredite brzinu  $v$  kao funkciju puta  $s$  na intervalu  $[0, 800]$ .

Rješenje:  $v = \sqrt{\frac{mg}{\gamma}} \sqrt{1 - \exp(-\frac{2\gamma}{m})}$ .

**Zadatak 8.11** Pokažite da se Runge-Kuttova metoda podudara sa Simpsonovom formulom ako funkcija  $f$  ne ovisi o  $y$ .

**Zadatak 8.12** Izradite kompjuterski program za metodu diskretizacije (t.8.3, str. 146) koristeći LU-dekompoziciju matrice.

**Zadatak 8.13** Slično kao u (8.8), str. 141 definirajte Runge-Kutta metodu za rješavanje diferencijalne jednačbe 3. reda iz Primjedbe 8.1, str. 138 Izradite odgovarajući kompjuterski program.

**Zadatak 8.14** Riješite Cauchyjev problem zadan s:

- a)  $x' = x - xy, \quad x(0) = 4,$   
 $y' = -y + xy, \quad y(0) = 1, \quad t \in [0, 8], h = 0.1$
- b)  $x' = -3x - 2y - 2xy^2, \quad x(0) = 0.8,$   
 $y' = 2x - y + 2y^3, \quad y(0) = 0.6, \quad t \in [0, 4], h = 0.1$
- c)  $x' = y^2 - x^2, \quad x(0) = 2,$   
 $y' = 2xy, \quad y(0) = 0.1, \quad t \in [0, 1.5], h = 0.05$
- d)  $x' = x^2 - y^2, \quad x(0) = 2,$   
 $y' = 2xy, \quad y(0) = 0.6, \quad t \in [0, 1.6], h = 0.02$
- e)  $x' = 1 - y, \quad x(0) = -1.2,$   
 $y' = x^2 - y^2, \quad y(0) = 0, \quad t \in [0, 5], h = 0.1$
- f)  $x' = x^3 - 2xy^2, \quad x(0) = 1,$   
 $y' = 2x^2y - y^3, \quad y(0) = 0.2, \quad t \in [0, 2], h = 0.025$

**Zadatak 8.15** Za sustave diferencijalnih jednačbi iz Zadatka 8.14, nađite fazno rješenje  $y = y(x; c)$ , te nacrtajte odgovarajući pramen (familiju) krivulja.

**Zadatak 8.16** Riješite sljedeće rubne probleme za  $h = 0.2, 0.1, 0.05$ :

- a)  $x'' = \frac{2t}{1+t^2}x'(t) - \frac{2}{1+t^2}x(t) + 1, \quad x(0) = 1.25, \quad x(4) = -0.95,$
- b)  $x'' = -\frac{2}{t}x'(t) + \frac{2}{t^2}x(t) + \frac{10 \cos(\ln t)}{t^2}, \quad x(1) = 1, \quad x(3) = -1,$
- c)  $x'' + \frac{1}{t}x'(t) + \frac{16}{t^2}x(t) = \frac{1}{t^2}, \quad x(1) = 0.75, \quad x(7) = 0.3,$
- d)  $x'' + \frac{2}{t}x'(t) - \frac{2}{t^2}x(t) = \frac{\sin t}{t^2}, \quad x(1) = -0.02, \quad x(6) = 0.02.$

## 9. Numeričko rješavanje parcijalnih diferencijalnih jednačbi

Problem rješavanja parcijalnih diferencijalnih jednačbi pojavljuje se u mnogim područjima primjena (elektrotehnika, fizika, kemija, tehnologija, itd). Općenito, parcijalna diferencijalna jednačba može sadržavati četiri nezavisne varijable (tri prostorne  $x, y, z$  i jedna vremenska  $t$ ), ali se najčešće broj varijabli može reducirati na dvije nezavisne varijable. To je razlog što ćemo razmatrati<sup>1</sup> jednačbu oblika

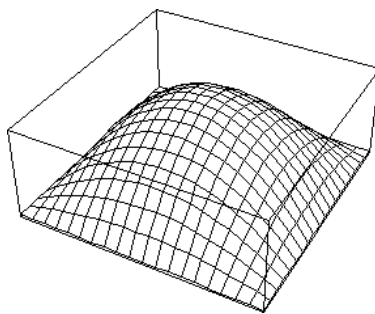
$$F(x, y, u, u_x, u_y, u_{xx}, u_{xy}, u_{yy}) = 0, \quad (9.1)$$

gdje su  $x, y$  nezavisne varijable,  $u = u(x, y)$  tražena funkcija, a

$$u_x = \frac{\partial u}{\partial x}, \quad u_y = \frac{\partial u}{\partial y}, \quad u_{xx} = \frac{\partial^2 u}{\partial x^2}, \quad u_{xy} = \frac{\partial^2 u}{\partial x \partial y} = \frac{\partial^2 u}{\partial y \partial x}, \quad u_{yy} = \frac{\partial^2 u}{\partial y^2},$$

njezine prve i druge parcijalne derivacije. Zato kažemo da je (9.1) parcijalna diferencijalna jednačba drugog reda.

Rješenje jednačbe (9.1) je funkcija  $u$ , koja uvrštena u tu jednačbu daje identitet. Grafički, rješenje jednačbe (9.1) predstavlja plohu u prostoru (*Slika 9.1*).



Slika 9.1. Graf funkcije  $u(x, y)$

Kažemo da je jednačba (9.1) linearna ako je funkcija  $F$  linearna u svim argumentima koji se navode u (9.1) osim eventualno u  $x$  ili  $y$ . Tada se (9.1) može

<sup>1</sup>Pri tome u cijelosti pratimo tijek izlaganja u radu GALIĆ (1989)

pisati u obliku

$$A \frac{\partial^2 u}{\partial x^2} + 2B \frac{\partial^2 u}{\partial x \partial y} + C \frac{\partial^2 u}{\partial y^2} + a \frac{\partial u}{\partial x} + b \frac{\partial u}{\partial y} + cu = f, \quad (9.2)$$

pri čemu koeficijenti  $A, B, C, a, b, c$  i funkcija  $f$  može ovisiti samo o varijablama  $x, y$ . Specijalno, ako navedeni koeficijenti ne ovise o varijablama  $x, y$ , onda kažemo da je (9.2) linearna diferencijalna jednačba drugog reda s konstantnim koeficijentima. Tada broj  $D = AC - B^2$  nazivamo **diskriminantom jednačbe** (9.2). Ovisno o predznaku diskriminante razlikujemo sljedeće tipove diferencijalnih jednačbi

$$\begin{aligned} D > 0 & \quad - \text{ eliptički tip,} \\ D < 0 & \quad - \text{ hiperbolički tip,} \\ D = 0 & \quad - \text{ parabolčki tip.} \end{aligned}$$

Još ćemo posebno spomenuti tzv. osnovne parcijalne diferencijalne jednačbe matematičke fizike

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2} \quad (\text{valna jednačba - oscilacija žice}) \quad (9.3)$$

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2} \quad (\text{jednačba provođenja topline}) \quad (9.4)$$

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \quad \text{ili} \quad \Delta u = 0 \quad (\text{Laplaceova jednačba}) \quad (9.5)$$

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x, y) \quad \text{ili} \quad \Delta u = f \quad (\text{Poissonova jednačba}). \quad (9.6)$$

Ovdje su  $x, y$  prostorne varijable, a  $t$  varijabla koja predstavlja vrijeme. Kako u jednačbama (9.5) i (9.6) nema vremenske varijable, kažemo da one opisuju **stacionarne probleme**. Lako se vidi da je (9.3) hiperbolička, (9.4) parabolčka diferencijalna jednačba, a (9.5) i (9.6) da su eliptičke diferencijalne jednačbe.

Parcijalne diferencijalne jednačbe u općem slučaju imaju beskonačno mnogo rješenja. Zato, ako je fizikalni proces opisan parcijalnim diferencijalnim jednačbama, za jednoznačno definiranje procesa potrebno je jednačbama pridružiti dopunske uvjete. U jednostavnim slučajevima ti dopunski uvjeti sastoje se od **početnih uvjeta** (zadana funkcija ili njena derivacija za  $t = t_0$ ) i **rubnih uvjeta** (zadana funkcija ili gradijent na rubu područja od interesa). Diferencijalnu jednačbu zajedno s početnim i rubnim uvjetima nazivamo **rubni problem**.

Problem rješavanja parcijalnih diferencijalnih jednačbi izučava se u numeričkoj analizi kao sastavnom dijelu primijenjene matematike. Postoji niz metoda pomoću kojih se dolazi do približnog rješenja rubnog problema. Najpoznatija numerička metoda za rješavanje ovakvih problema je **metoda konačnih diferencija** (ili *metoda mreže*) s idejom da se derivacije funkcija jedne ili više varijabli aproksimiraju **podijeljenim razlikama**. Druge važne metode za rješavanje ovakvih problema su: *metoda kolokacije*, *Galerkinova metoda*, *Ritzova metoda* i *metoda konačnih elemenata*. Mi ćemo na primjeru Poisson-ove diferencijalne jednačbe pokazati **metodu konačnih diferencija** za rješavanje parcijalnih diferencijalnih jednačbi eliptičkog tipa.

## 9.1 Rješavanje Poissonove dif. jednačbe metodom konačnih diferencija

Važna jednačba matematičke fizike je Poissonova diferencijalna jednačba

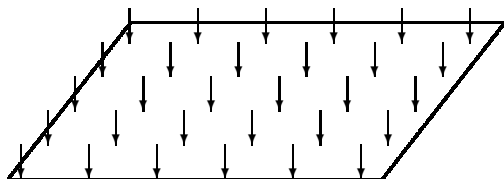
$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x, y), \quad (x, y) \in \Omega \subset \mathbb{R}^2, \quad (9.7)$$

odnosno

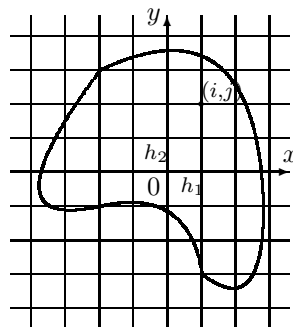
$$\Delta u = f, \quad (x, y) \in \Omega \subset \mathbb{R}^2,$$

koja fizikalno predstavlja problem ravnoteže membrane na području  $\Omega$ . Zanima nas progib (pomak točaka) membrane, koji nastaje djelovanjem neke distributivne sile  $f$  na membranu (Slika 9.2).

Budući da u ovom slučaju pomak točaka membrane  $u(x, y)$ , nije funkcija vremena  $t$ , već samo funkcija položaja točke, uz Poissonovu diferencijalnu jednačbu nisu potrebni početni, već samo rubni uvjeti.



Slika 9.2 Djelovanje sile  $f$  na membranu



Slika 9.3 Mreža

Na rubu ( $\partial\Omega$ ) područja  $\Omega$  zadaje se vrijednost funkcije  $u$  u obliku

$$u(x, y) = \varphi(x, y) \quad \text{za } (x, y) \in \partial\Omega. \quad (9.8)$$

Prema tome, naš rubni problem glasi

$$\begin{aligned} \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} &= f(x, y), \quad (x, y) \in \Omega \subset \mathbb{R}^2, \\ u(x, y) &= \varphi(x, y) \quad \text{za } (x, y) \in \partial\Omega. \end{aligned} \quad (9.9)$$

Rubni problem (9.9) u literaturi možemo naći pod nazivom Dirichletov unutarnji problem. Ovaj problem uvijek ima jedinstveno rješenje.

Prijeđimo na opis metode konačnih diferencija za rješavanje rubnog problema (9.9). Najprije ćemo područje  $\Omega$  pokriti s dvije familije pravaca paralelnih s koordinatnim osima  $x$  i  $y$

$$x_i = ih_1, \quad y_j = jh_2, \quad i, j \in \mathbb{Z},$$

gdje su  $h_1$  i  $h_2$  zadani koraci mreže (Slika 9.3). Ako je  $h_1 = h_2 = h$ , onda govorimo o kvadratnoj mreži.

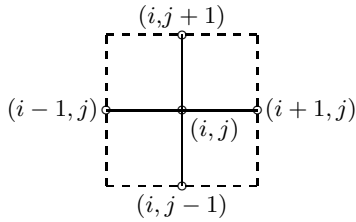
Točke  $(x_i, y_i) = (ih_1, jh_2)$  u kojima se sijeku pravci zovemo **čvorovi** i označavamo samo s  $(i, j)$ . S  $\Omega_h$  označit ćemo skup onih čvorova mreže koji leže unutar ili na rubu područja  $\Omega$ . Skup od pet točaka (vidi Sliku 9.4)

$(i, j), (i-1, j), (i+1, j), (i, j-1), (i, j+1)$  – prva osnovna shema,

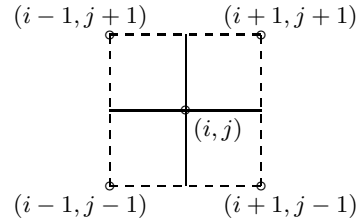
ili

$(i, j), (i+1, j+1), (i-1, j+1), (i-1, j-1), (i+1, j-1)$  – druga osnovna shema

nazivamo peterotočkasta zvijezda sa središtem u točki  $(i, j)$ .



Slika 9.4.a Prva osnovna shema



Slika 9.4.b Druga osnovna shema

Kaže se da je čvor  $(i, j)$  **unutarnji čvor** ako svi čvorovi zvijezde sa središtem u čvoru  $(i, j)$  pripadaju skupu  $\Omega_h$ , a odresci pravaca, koji spajaju središte s ostalim čvorovima zvijezde, pripadaju području  $\Omega$ . Kaže se da je čvor  $(k, l)$  **rubni (granični) čvor** ako barem jedan čvor zvijezde sa središtem u  $(k, l)$  ne pripada skupu  $\Omega_h$  ili ako barem jedan odrezak, koji spaja središte s nekim čvorom zvijezde presijeca rub  $\partial\Omega$ . Vrijednost funkcije  $u = u(x, y)$  u čvoru  $(i, j)$  označit ćemo s  $u_{i,j}$ . S  $u_h$  označit ćemo tablično zadanu funkciju približnih vrijednosti funkcije  $u$  u svim čvorovima koji pripadaju skupu  $\Omega_h$ . Tabličnu funkciju  $u_h$  zvat ćemo **mrežna funkcija**.

Osnovna ideja *metode mreže* sastoji se u tome da se parcijalne derivacije promatrane diferencijalne jednačbe zamijene **konačnim diferencijama**.

Prema Taylorovoj formuli vrijedi

$$u(x + h_1, y) = u(x, y) + h_1 \frac{\partial u}{\partial x}(x, y) + \frac{h_1^2}{2!} \frac{\partial^2 u}{\partial x^2}(x, y) + \frac{h_1^3}{3!} \frac{\partial^3 u}{\partial x^3}(x, y) + \dots \quad (9.10)$$

$$u(x - h_1, y) = u(x, y) - h_1 \frac{\partial u}{\partial x}(x, y) + \frac{h_1^2}{2!} \frac{\partial^2 u}{\partial x^2}(x, y) - \frac{h_1^3}{3!} \frac{\partial^3 u}{\partial x^3}(x, y) + \dots \quad (9.11)$$

Oduzimajući (9.11) od (9.10), zanemarujući sve članove počevši od drugog i dijeljenjem s  $2h_1$ , dobivamo aproksimaciju prve parcijalne derivacije funkcije  $u$  po varijabli  $x$  u točki  $(x, y)$

$$\frac{\partial u}{\partial x}(x, y) \approx \frac{u(x + h_1, y) - u(x - h_1, y)}{2h_1},$$



što koristimo za izračunavanje te parcijalne derivacije u čvoru  $(i, j)$

$$\left(\frac{\partial u}{\partial x}\right)_{i,j} \approx \frac{u_{i+1,j} - u_{i-1,j}}{2h_1}. \quad (9.12)$$

Analogno dobivamo

$$\left(\frac{\partial u}{\partial y}\right)_{i,j} \approx \frac{u_{i,j+1} - u_{i,j-1}}{2h_2}. \quad (9.13)$$

Zbrajajući (9.10) i (9.11), zanemarujući sve članove počevši od trećeg i dijeljenjem s  $h_1$ , dobivamo aproksimaciju druge parcijalne derivacije funkcije  $u$  po varijabli  $x$  u točki  $(x, y)$

$$\frac{\partial^2 u}{\partial x^2}(x, y) \approx \frac{[u(x + h_1, y) - 2u(x, y) + u(x - h_1, y)]}{h_1^2},$$

što možemo koristiti za približno izračunavanje te parcijalne derivacije u čvoru  $(i, j)$

$$\left(\frac{\partial^2 u}{\partial x^2}\right)_{i,j} \approx \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h_1^2}. \quad (9.14)$$

Analogno dobivamo

$$\left(\frac{\partial^2 u}{\partial y^2}\right)_{i,j} \approx \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{h_2^2}. \quad (9.15)$$

Uvrštavanjem diferencija (9.14) i (9.15) u diferencijalnu jednačbu (9.9) dobivamo

$$\frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h_1^2} + \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{h_2^2} = f_{i,j}, \quad (9.16)$$

gdje je  $f_{i,j}$  vrijednost funkcije  $f$  u čvoru  $(i, j)$ . Ako stavimo  $h_1 = h_2 = h$ , onda se (9.16) može pisati

$$u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{i,j} = h^2 f_{i,j}. \quad (9.17)$$

Algebarske jednačbe (9.16), odnosno (9.17) nazivamo diferencijskim ili mrežnim jednačbama u čvoru  $(i, j)$ .

**Zadatak 9.1** Napišite diferencijsku (mrežnu) jednačbu u čvoru  $(i, j)$  prema drugoj osnovnoj shemi (Slika 9.4.b).

Svakom čvoru iz područja  $\Omega_h$  pridružiti ćemo pripadnu mrežnu jednačbu oblika (9.16) ili (9.17). Na taj način dobit ćemo sustav od toliko linearnih algebarskih jednačbi koliko ima čvorova u  $\Omega_h$ . Dakle, rješavanje rubnog problema (9.9) metodom mreže sveli smo na rješavanje sustava linearnih algebarskih jednačbi oblika (9.16), koji možemo pokušati riješiti nekom od metoda navedenih u t.3, str. 38.

**Primjedba 9.1** U praktičnim primjenama red matrice sustava (9.16) može biti tako velik da raspoloživa memorija računala nije dostatna. Zato se za rješavanje ovakvih sustava obično koriste specijalne metode za velike rijetke matrice (vidi primjerice Press 1992).

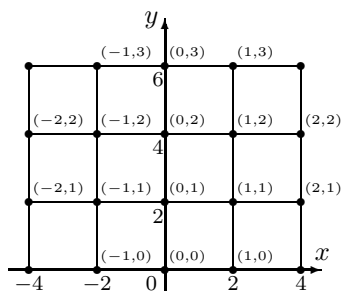
**Primjedba 9.2** Postoji više načina za ocjenu pogreške metode mreže (vidi primjerice AGANOVIĆ (1985) ili DAHLQUIST (1972)). Pokazuje se da pogreška ovisi o veličini koraka mreže. Što je korak mreže manji (mreža - gušća), pogreška je manja.

**Primjer 9.1** Metodom mreže treba riješiti rubni problem

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = -2, \quad (x, y) \in \Omega \subset \mathbb{R}^2,$$

$$u(x, y) = \varphi(x, y) \quad \text{za } (x, y) \in \partial\Omega \quad \Omega = (-4, 4) \times (0, 6).$$

U području  $\Omega$  nacrtat ćemo kvadratnu mrežu s korakom  $h = 2$  (Slika 9.5).



Slika 9.5. Mreža na području  $\Omega$

Područje  $\Omega$  simetrično je s obzirom na os  $y$ , pa se može očekivati da će i rješenje  $u$  rubnog problema biti simetrično s obzirom na os  $y$ . Zbog toga možemo promatrati samo jednu, primjerice desnu, polovinu područja  $\Omega$ . Ako numeraciju čvorova uredimo prema Slici 9.5, onda treba odrediti rješenje samo za četiri čvora  $(0, 1)$ ,  $(1, 1)$ ,  $(0, 2)$ ,  $(1, 2)$  jer se rješenje  $u$  na rubnim čvorovima, zbog rubnog uvjeta  $u/\partial\Omega = 0$ , poništava. Za svaki od navedenih čvorova napisat ćemo odgovarajuće jednačbe konačnih diferencija koristeći prvu osnovnu shemu. Prema (9.17) imamo

$$\begin{aligned} (0, 1): \quad & u_{1,1} + u_{-1,1} + u_{0,2} + u_{0,0} - 4u_{0,1} = 2^2(-2) \\ (0, 2): \quad & u_{1,2} + u_{-1,2} + u_{0,3} + u_{0,1} - 4u_{0,2} = 2^2(-2) \\ (1, 1): \quad & u_{2,1} + u_{0,1} + u_{1,2} + u_{1,0} - 4u_{1,1} = 2^2(-2) \\ (1, 2): \quad & u_{2,2} + u_{0,2} + u_{1,3} + u_{1,1} - 4u_{1,2} = 2^2(-2) \end{aligned}$$

Zbog simetričnosti rješenja vrijedi

$$u_{1,1} = u_{-1,1}, \quad u_{1,2} = u_{-1,2},$$

a zbog  $u/\partial\Omega = 0$  vrijedi

$$u_{0,0} = u_{1,0} = u_{0,3} = u_{1,3} = u_{2,1} = u_{2,2} = 0,$$

pa prethodni sustav diferencnih jednačbi možemo zapisati

$$\begin{array}{lclclcl} (0,1): & -4u_{0,1} & +u_{0,2} & +2u_{1,1} & & = -8 \\ (0,2): & u_{0,1} & -4u_{0,2} & & +2u_{1,2} & = -8 \\ (1,1): & u_{0,1} & & -4u_{1,1} & +u_{1,2} & = -8 \\ (1,2): & & u_{0,2} & +u_{1,1} & -4u_{1,2} & = -8 \end{array}$$

Tako smo dobili sustav od četiri jednačbe s četiri nepoznanice, koji možemo riješiti nekom od metoda navedenih u t.3, str. 38. Dobivamo

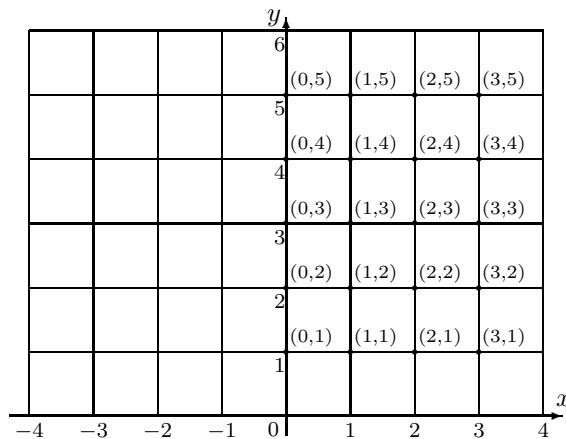
$$u_{0,1} = 5.71429, \quad u_{0,2} = 5.71429, \quad u_{1,1} = 4.57143, \quad u_{1,2} = 4.57143.$$

Kompletna mrežna funkcija  $u_h$  prikazana je u *Tablici 9.1*.

$y \setminus x$	-4	-2	0	2	4
0	0	0	0	0	0
2	0	4.57143	5.71429	4.57143	0
4	0	4.57143	5.71429	4.57143	0
6	0	0	0	0	0

Tablica 9.1. Mrežna funkcija  $u_h$

**Primjer 9.2** Metodom mreže riješit ćemo rubni problem naveden u *Primjeru 9.1*, ali ćemo za korak u mreži uzeti  $h = 1$ . Čvorove ćemo numerirati kao na *Slici 9.6*.



Slika 9.6. Mreža na području  $\Omega$  s korakom  $h = 1$

Zbog simetričnosti promatramo samo čvorove u desnoj poluravnini. Za te čvorove napisat ćemo odgovarajuće jednačbe konačnih diferencija koristeći također prvu osnovnu shemu. Prema (9.17) dobivamo sustav jednačbi

$$\begin{aligned}
u_{1,1} + u_{-1,1} + u_{0,2} + u_{0,0} - 4u_{0,1} &= -2 \\
u_{1,2} + u_{-1,2} + u_{0,3} + u_{0,1} - 4u_{0,2} &= -2 \\
u_{1,3} + u_{-1,3} + u_{0,4} + u_{0,2} - 4u_{0,3} &= -2 \\
u_{1,4} + u_{-1,4} + u_{0,5} + u_{0,3} - 4u_{0,4} &= -2 \\
u_{1,5} + u_{-1,5} + u_{0,6} + u_{0,4} - 4u_{0,5} &= -2 \\
u_{2,1} + u_{0,1} + u_{1,2} + u_{1,0} - 4u_{1,1} &= -2 \\
u_{2,2} + u_{0,2} + u_{1,3} + u_{1,1} - 4u_{1,2} &= -2 \\
u_{2,3} + u_{0,3} + u_{1,4} + u_{1,2} - 4u_{1,3} &= -2 \\
u_{2,4} + u_{0,4} + u_{1,5} + u_{1,3} - 4u_{1,4} &= -2 \\
u_{2,5} + u_{0,5} + u_{1,6} + u_{1,4} - 4u_{1,5} &= -2 \\
u_{3,1} + u_{1,1} + u_{2,2} + u_{2,0} - 4u_{2,1} &= -2 \\
u_{3,2} + u_{1,2} + u_{2,3} + u_{2,1} - 4u_{2,2} &= -2 \\
u_{3,3} + u_{1,3} + u_{2,4} + u_{2,2} - 4u_{2,3} &= -2 \\
u_{3,4} + u_{1,4} + u_{2,5} + u_{2,3} - 4u_{2,4} &= -2 \\
u_{3,5} + u_{1,5} + u_{2,6} + u_{2,4} - 4u_{2,5} &= -2 \\
u_{4,1} + u_{2,1} + u_{3,2} + u_{3,0} - 4u_{3,1} &= -2 \\
u_{4,2} + u_{2,2} + u_{3,3} + u_{3,1} - 4u_{3,2} &= -2 \\
u_{4,3} + u_{2,3} + u_{3,4} + u_{3,2} - 4u_{3,3} &= -2 \\
u_{4,4} + u_{2,4} + u_{3,5} + u_{3,3} - 4u_{3,4} &= -2 \\
u_{4,5} + u_{2,5} + u_{3,6} + u_{3,4} - 4u_{3,5} &= -2
\end{aligned} \quad (*)$$

Zbog simetričnosti rješenja vrijedi

$$u_{1,1} = u_{-1,1}, \quad u_{1,2} = u_{-1,2}, \quad u_{1,3} = u_{-1,3}, \quad u_{1,4} = u_{-1,4}, \quad u_{1,5} = u_{-1,5}$$

a zbog  $u/\partial\Omega = 0$ , svi članovi kojima je prvi indeks 4 ili kojima je drugi indeks 0 ili 6 jednaki su nuli. Na taj način prethodni sustav se reducira na sustav od 20 jednačbi s 20 nepoznanica, koje ćemo poredati ovim redom

$$\begin{aligned}
&u_{0,1} \quad u_{0,2} \quad u_{0,3} \quad u_{0,4} \quad u_{0,5} \quad u_{1,1} \quad u_{1,2} \quad u_{1,3} \quad u_{1,4} \quad u_{1,5} \\
&u_{2,1} \quad u_{2,2} \quad u_{2,3} \quad u_{2,4} \quad u_{2,5} \quad u_{3,1} \quad u_{3,2} \quad u_{3,3} \quad u_{3,4} \quad u_{3,5}
\end{aligned}$$

Matrica novog sustava je vrlo "rijetka" tj. relativno je malo elemenata koji su različiti od nule i oni su smješteni na glavnoj dijagonali, te u prvoj i petoj sporednoj dijagonali u gornjem i donjem trokutu<sup>2</sup>. Matrica sustava prema tome glasi:

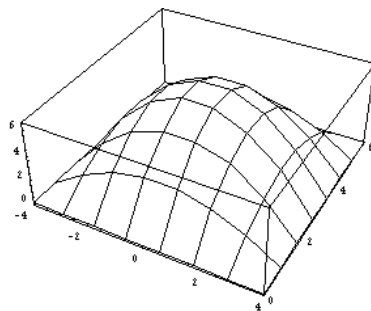
$$\begin{bmatrix}
-4 & 1 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & -4 & 1 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -4 & 1 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -4 & 1 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -4 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & -4 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & -4 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & -4 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & -4 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & -4
\end{bmatrix}$$

<sup>2</sup>Primijetimo da uz navedene redukcije, prvi stupac u sustavu (\*) predstavlja petu gornju dijagonalu novog sustava, drugi stupac - petu donju dijagonalu, treći stupac - prvu gornju dijagonalu, četvrti stupac - prvu donju dijagonalu, a peti stupac - glavnu dijagonalu novog sustava

Uz točnost  $\epsilon = 0.00005$  (4 decimale) SOR metodom (vidi str. 60) dobivamo rješenje prikazano u *Tablici 9.2*.

$y \setminus x$	-4	-3	-2	-1	0	1	2	3	4
0	0	0	0	0	0	0	0	0	0
1	0	2.04280	3.12354	3.65712	3.81848	3.65712	3.12354	2.04280	0
2	0	3.04767	4.79423	5.68647	5.95967	5.68647	4.79423	3.04764	0
3	0	3.35365	5.31924	6.33486	6.64727	6.33486	5.31924	3.35365	0
4	0	3.04767	4.79423	5.68647	5.95967	5.68647	4.79423	3.04764	0
5	0	2.04280	3.12354	3.65712	3.81848	3.65712	3.12354	2.04280	0
6	0	0	0	0	0	0	0	0	0

Tablica 9.2. Mrežna funkcija  $u_h$



Slika 9.7. Deformirana membrana

Relativno dugo vrijeme rada računala prilikom rješavanja ovakvih problema upućuje na korištenje specijalnih metoda za rješavanje "velikih rijetkih sustava" (vidi primjerice GOLUB (1989), PRESS (1989), SCHWARZ (1986)). Vidi također *Zadatak 3.24* i *Zadatak 3.25*, str. 65.

Graf mrežne funkcije  $u_h$  prikazan je na *Slici 9.7*. Kao što se vidi na slici, promatrana membrana je uslijed djelovanja konstantne sile  $f$  deformirana tako da je ostala simetrična u odnosu na os  $y$  i na pravac  $y = 3$ , ali je na rubovima ostala učvršćena.

**Primjedba 9.3** Postoji opsežna literatura o egzistenciji i jedinstvenosti rješenja, kao i o numeričkim metodama za rješavanje parcijalnih diferencijalnih jednadžbi (vidi AGANOVIĆ (1985), MATHEWS (1992), ORTEGA (1981), SCHWARZ (1986), itd.) Odgovarajuću programsku podšku (FORTRAN) može se naći u PRESS (1992), kao i u *The NAG-Library* (PHILIPS (1986), KÖCKLER (1990)).

## 9.2 Zadaci

**Zadatak 9.2** Riješite rubni problem iz *Primjera 9.1* koristeći metodu mreže s korakom  $h = 2$ , tako da u jednadžbama konačnih diferencija koristiš drugu osnovnu shemu (vidi *Zadatak 9.1*, str. 154).

Rješenje:

$y \setminus x$	-4	-2	0	2	4
0	0	0	0	0	0
2	0	2.857143	3.428572	2.857143	0
4	0	2.857143	3.428572	2.857143	0
6	0	0	0	0	0

**Zadatak 9.3** Riješite rubni problem iz Primjera 9.1 koristeći metodu mreže s korakom  $h = 1$ , tako da u jednadžbama konačnih diferencija koristite drugu osnovnu shemu (vidi Zadatak 9.1, str. 154). Pripadni sustav jednadžbi riješite nekom od navedenih iterativnih metoda u t.3.8, str. 57

Rješenje:

$y \setminus x$	-4	-3	-2	-1	0	1	2	3	4
0	0	0	0	0	0	0	0	0	0
1	0	1.13415	1.64634	1.91463	1.98780	1.91463	1.64634	1.13415	0
2	0	1.60975	2.53658	2.97561	3.12195	2.97561	2.53658	1.60975	0
3	0	1.76829	2.79268	3.32927	3.47561	3.32927	2.79268	1.76829	0
4	0	1.60975	2.53658	2.97561	3.12195	2.97561	2.53658	1.60975	0
5	0	1.13415	1.64634	1.91463	1.98780	1.91463	1.64634	1.13415	0
6	0	0	0	0	0	0	0	0	0

**Zadatak 9.4** Metodom mreže ( $h = 1$ ) nađite aproksimativno rješenje Laplaceovog rubnog problema

$$\Delta u = 0, \quad (x, y) \in \Omega = [0, 4] \times [0, 4],$$

$$\begin{aligned} u(x, 0) = 20 \quad \& \quad u(x, 4) = 180, \quad \text{za } x \in (0, 4) \\ u(0, y) = 80 \quad \& \quad u(4, y) = 0, \quad \text{za } y \in (0, 4) \end{aligned}$$

Rješenje:

$y \setminus x$	0	1	2	3	4
0		20	20	20	
1	80	55.7143	43.2143	27.1429	0
2	80	79.6429	70.0000	84.2857	0
3	80	112.857	111.786	84.2857	0
4		180	180	180	

**Zadatak 9.5** Metodom mreže ( $h = 1$ ) nađite aproksimativno rješenje Laplaceovog rubnog problema

$$\Delta u = 0, \quad (x, y) \in \Omega = [0, 4] \times [0, 4],$$

$$\begin{aligned} u(x, 4) = 180 \quad \& \quad u_y(x, 0) = 0, \quad \text{za } x \in (0, 4) \\ u(0, y) = 80 \quad \& \quad u(4, y) = 0, \quad \text{za } y \in [0, 4) \end{aligned}$$

Rješenje:

$y \setminus x$	0	1	2	3	4
0	80	71.8218	56.8543	32.2342	0
1	80	75.2165	61.6806	36.0412	0
2	80	87.3636	78.6103	50.2502	0
3	80	115.628	115.147	86.3492	0
4		180	180	180	

**Zadatak 9.6** Metodom mreže ( $h = 0.5$ ) nađite aproksimativno rješenje Laplaceovog rubnog problema

$$\Delta u = 0, \quad (x, y) \in \Omega = [0, 1.5] \times [0, 1.5],$$

$$\begin{aligned} u(x, 0) = x^4 \quad &\& \quad u(x, 1.5) = x^4 - 13.5x^2 + 5.0625, \quad \text{za } x \in [0, 1.5], \\ u(0, y) = y^4 \quad &\& \quad u(1.5, y) = 5.0625 - 13.5y^2 + y^4, \quad \text{za } y \in [0, 1.5]. \end{aligned}$$

Dobiveno rješenje usporedi s egzaktnim  $u(x, y) = x^4 - 6x^2y^2 + y^4$ .

**Zadatak 9.7** Metodom mreže ( $h = 0.2$ ) nađite aproksimativno rješenje Poissonovog rubnog problema

$$\Delta u = 2, \quad (x, y) \in \Omega = [0, 1] \times [0, 1],$$

$$\begin{aligned} u(x, 0) = x^2 \quad &\& \quad u(x, 1) = (x - 1)^2, \quad \text{za } x \in [0, 1], \\ u(0, y) = y^2 \quad &\& \quad u(1, y) = (y - 1)^2, \quad \text{za } y \in [0, 1]. \end{aligned}$$

**Zadatak 9.8** Metodom mreže ( $h = 0.2$ ) nađite aproksimativno rješenje Poissonovog rubnog problema

$$\Delta u = y, \quad (x, y) \in \Omega = [0, 1] \times [0, 1],$$

$$\begin{aligned} u(x, 0) = x^3 \quad &\& \quad u(x, 1) = x^3, \quad \text{za } x \in [0, 1], \\ u(0, y) = 0 \quad &\& \quad u(1, y) = 1, \quad \text{za } y \in [0, 1]. \end{aligned}$$

**Zadatak 9.9** Riješite  $\Delta u = -4u$ , na pravokutniku  $\Omega = [0, 1] \times [0, 1]$  s rubnim uvjetima

$$u(x, y) = \cos 2x + \sin 2y.$$

## Literatura

- [1] I. AGANOVIĆ, K. VESELIĆ *Jednadžbe matematičke fizike*, Školska knjiga, Zagreb, 1985.
- [2] A. BJÖRCK, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.
- [3] D. BLANUŠA, *Viša matematika*, Tehnička knjiga, Zagreb, 1973.
- [4] Z. BOHTE, *Numerično reševanje nelinearnih enačb*, Društvo matematikov, fizikov in astronomov Slovenije, Ljubljana, 1993.
- [5] L. COLLATZ, *The Numerical Treatment of Differential Equations*, Springer-Verlag, Berlin, Heidelberg, New York, 1966.
- [6] G. DAHLQUIST, A. BJÖRCK *Numerische Methoden*, R. Oldenbourg Verlag, München, Wien, 1972. (postoji i engleski prijevod)
- [7] C. DE BOOR, *A Practical Guide to Splines*, Springer-Verlag, New York, 1978.
- [8] J. W. DEMEL, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.
- [9] B. P. DEMIDOVICH, I. A. MARON *Computational Mathematics*, MIR, Moscow, 1981.
- [10] J. E. DENNIS, JR, R. B. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM, Philadelphia, 1996.
- [11] J. E. DENNIS JR., J. J. MORE, *Quasi-Newton methods, motivation and theory*, SIAM Review, **19**(1977), 46-89.
- [12] N. R. DRAPER, H. SMITH, *Applied Regression Analysis*, J.Wiley & Sons, New York, 1996.
- [13] J. D. FAIRES, R. BURDEN, *Numerical Methods*, 2<sup>nd</sup> Ed., Brooks/Cole Publ. Comp., New York, 1998.
- [14] R. FLETCHER, *Practical Methods of Optimization*, Wiley, New York, 1987.
- [15] G. E. FORSYTHE, M. A. MALCOLM, C. B. MOLER, *Computer Methods for Mathematical Computations*, Prentice-Hall, Inc. Englewood Cliffs, N.J., 1974.
- [16] R. GALIĆ, *Rješavanje Poissonove diferencijalne jednadžbe metodom mreže*, Elektrotehnika '88, Elektrotehnički fakultet, Osijek, 1989.
- [17] R. GALIĆ, *Vjerojatnost i statistika*, Elektrotehnički fakultet, Osijek, 1999.



- [18] P. E. GILL, W. MURRAY AND M. H. WRIGHT, *Practical Optimization*, Academic Press, 1981.
- [19] P. E. GILL, W. MURRAY AND M. H. WRIGHT, *Numerical Linear Algebra and Optimization*, Vol.1., Addison-Wesley Publishing Company, New York, 1991.
- [20] G. H. GOLUB, C. F. VAN LOAN *Matrix Computations*, The J.Hopkins University Press, Baltimore and London, 1996.
- [21] I. IVANŠIĆ, *Numerička matematika*, Element, Zagreb, 1998.
- [22] D. JUKIĆ, R. SCITOVSKI, *Matematika I*, Elektrotehnički fakultet, Prehrambeno-tehnološki fakultet, Osijek, 1999.
- [23] D. KALMAN, *A singularity valuable decomposition: the SVD of a matrix*, The College of Mathematics Journal **27**(1996), 1-23.
- [24] N. KÖCKLER, *Numerische Algorithmen in Softwaresystemen*, B. G. Teubner, Stuttgart, 1990.
- [25] S. KUREPA, *Konačno dimenzionalni vektorski prostori i primjene*, Tehnička knjiga, Zagreb, 1967.
- [26] S. KUREPA, *Matematička analiza II*, Tehnička knjiga, Zagreb, 1990.
- [27] P. LANCASTER, K. SALKASKAS, *Curve and Surface Fitting*, Academic Press, London, 1986.
- [28] C. L. LAWSON, R. J. HANSON, *Solving Least Squares Problems*, SIAM, Philadelphia, 1996.
- [29] S. MARDEŠIĆ, *Matematička analiza I*, Školska knjiga, Zagreb, 1988.
- [30] J. H. MATHEWS, *Numerical Methods for Mathematics, Science, and Engineering*, Prentice-Hall Int. Ed., London, 1992.
- [31] MOLITZ, *Äussere Balistik*, 1967.
- [32] J. M. ORTEGA, W. G. POOLE, JR., *An Introduction to Numerical Methods for Differential Equations*, Pitman Publishing inc., Massashusetts, 1981.
- [33] J. M. ORTEGA, W. C. RHEINOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970. (postoji i ruski prijevod)
- [34] J. PHILIPS, *The NAG Library: A Beginner's Guide*, Clarendon Press, Oxford, 1986.
- [35] W. H. PRESS, B. P. FLANNERY, S. A. TEUKOLSKY, W. T. VETTERLING, *Numerical Recipes*, Cambridge University Press, Cambridge, 1989.
- [36] G. J. S. ROSS, *Nonlinear Estimation*, Springer Verlag, New York, 1990.
- [37] H. R. SCHWARZ, *Numerische Mathematik*, Teubner, Stuttgart, 1986.
- [38] R. SCITOVSKI, *Problemi najmanjih kvadrata. Financijska matematika*, Ekonomski fakultet, Elektrotehnički fakultet, Osijek, 1993.

- [39] R. SCITOVSKI, R. GALIĆ, M. BENŠIĆ *Numerička analiza. Vjerojatnost i statistika*, (interna skripta), Elektrotehnički fakultet, Osijek, 1993.
- [40] G. W. STEWART, *Afternotes on Numerical Analysis*, SIAM, Philadelphia, 1996.
- [41] G. W. STEWART, *Afternotes goes to Graduate School*, SIAM, Philadelphia, 1998.
- [42] J. STOER, *Numerische Mathematik I*, 7. izdanje Springer-Verlag, Berlin, 1994.
- [43] J. STOER, R. BULIRSCH, *Introduction to Numerical Analysis*, Springer-Verlag, New York, 1993.
- [44] L. N. TREFETHEN, D. BAU, *Numerical Linear Algebra*, SIAM, Philadelphia, 1997.
- [45] S. WOLFRAM, *The Mathematica Book*, Wolfram Media, Champaign, 1998.

## Indeks

Čebiševljevi polinomi, 25, 98

Čvorovi interpolacije, 16

### Algoritam

Back Substitution, 44, 49, 52, 53

bisekcije, 71

Cholesky, 52

duljina koraka, 116

Eulerov, 139

Forward Substitution, 44, 49, 52

Gauss-Newtonov, 117

Gauss-Seidelov, 58

Gram - Schmidtov, 97

Hornerov, 22

invertiranje matrice, 51

Jacobijev, 58

LU, 50

Newtonova metoda, 81

Runge-Kutta, 142

Simpsonovo pravilo, 132

spline, 31

trapezno pravilo, 128

### Aproksimacija

funkcije, 93

matrice, 40, 113

najbolja, 93

najbolja  $L_2$  aproksimacija, 94, 113

najbolja  $L_\infty$  aproksimacija, 100

vektora, 40

Aritmetička sredina, 103

Balistika, 83

Besselova nejednakost, 96

Brzina konvergencije, 70

Cauchyjev problem, 137

Dekompozicija

Cholesky, 51, 66

LU-dekompozicija, 47

na singularne vrijednosti, 60, 111

QR-dekompozicija, 52, 110

Diferencijalna jednadžba

fazno rješenje, 145

obična, 137

parcijalna, 150

Poissonova, 152

Floating-point aritmetika, 2, 38

Gaussova metoda eliminacije, 44, 48

Givensove rotacije, 53

Gram-Schmidtov postupak ortogonalizacije, 97

Hat-funkcija, 26

Hesijan, 107

Hornerova shema, 21, 22, 35

Householderove transformacije, 53

Interpolacija, 16

Interpolacijski polinom, 16

Lagrangeov oblik, 18

Newtonov oblik, 20

ocjena pogreške, 23

Jacobijan, 107

Kontrakcija, 59

### Matrica

determinanta matrice, 49

dijagonalno dominantna, 30, 45, 60

elementarna, 48

glavni minori, 46

Gramova, 94

- Hilbertova, 47, 52, 95
- Householderova, 53
- inverzna, 50
- permutacija, 49
- pozitivno definitna, 46
- pseudoinverzna, 110
- rijetko popunjena, 44, 57, 157
- trodiagonalna, 30, 44
- vrpčasta, 65, 157
- Metoda
  - bisekcije, 71
  - Brentova, 91
  - Croutova, 49
  - diskretizacije, 146
  - Eulerova, 138
  - Gauss–Jordanova, 45
  - Gauss–Newtonova, 117
  - Gauss–Seidelova, 57
  - Gaussova metoda eliminacije, 44, 48
  - Heunova, 141
  - Jacobijeva, 57
  - jednostavnih iteracija, 72
  - konačnih diferencija, 151
  - Marquardtova, 120
  - Newtonova, 77, 115
  - regula falsi, 83
  - Runge - Kutta, 141
  - sekanti, 82
  - simetrična formula, 140
  - SOR, 60
- Minimizacija
  - jednodimenzionalna, 91, 115
  - višedimenzionalna, 115
- Model
  - Lotka–Volterra, 145
- Mrežna funkcija, 153
- Nelinearna jednadžba, 67
- Newton–Cotesova formula, 130
- Norma, 39
  - $L_1$  norma, 39
  - Čebiševljeva  $L_\infty$  norma, 39, 93
  - Euklidova  $L_2$  norma, 39
  - Frobeniusova, 40
  - matrična, 40
  - matrična inducirana, 40
  - vektorska, 39
- Nultočka funkcije, 67
- Ocjena pogreške, 69, 75, 78
- Ortogonalan sustav, 95
  - Fourierov sustav, 96
- Ortogonalni polinomi, 97
  - Čebiševljevi polinomi, 98
  - Legendreovi polinomimi, 97
- Pivot-element, 45
- Podijeljene razlike, 21
- Pogreška
  - apsolutna, 4
  - inverzni problem, 10
  - metode, 2
  - modela, 2
  - relativna, 5
  - standardna, 10
  - strojna, 2
  - zaokruživanja, 1, 8
- Pouzdanje znamenke, 7
- Problem najmanjih kvadrata, 103
  - linearni, 104, 108
  - nelinearni, 105, 115
- Separacija nultočki, 67
- Sherman–Morrisonova lema, 48
- Signifikantne znamenke, 6
- Simpsonovo pravilo, 130
  - generalizirano, 131
- Singularne vrijednosti matrice, 43, 61
- Singularni vektori matrice, 61
- Slobodni pad, 143
- Spline, 25
  - kubični interpolacijski, 28
  - linearni interpolacijski, 26
  - ocjena pogreške, 28
- Strategija
  - parcijalnog pivotiranja, 46
  - potpunog pivotiranja, 46
- Sustav linearnih jednadžbi, 38
  - diferencijskih, 146, 154
  - iterativne metode, 57
  - mrežnih, 154

- perturbirani, 41
- rješavanje unaprijed, 43
- rješavanje unazad, 43
- trokutasti sustav, 43
- Sustav nelinearnih jednadžbi, 85
  - BFGS metoda, 86
  - Broydenova metoda, 86
  - DFP metoda, 86
  - kvazi-Newtonova metoda, 86
  - metoda jednostavnih iteracija, 85
  - Newtonova metoda, 85
- Sustav normalnih jednadžbi, 94, 109
- Svojstvene vrijednosti matrice, 61
- Svojstveni vektori matrice, 61
- Sylvesterov kriterij, 46
- Težine podataka, 107
- Teorem
  - Banachov o fiksnoj točki, 59, 73
  - de la Vallée Poussin, 102
  - Pitagorin, 96
- Trapezno pravilo, 27, 125, 141
  - generalizirano, 127
- Uvjetovanost, 41
  - broj uvjetovanosti, 42, 62
- Vandermondova determinanta, 17