

PRI MIJENJEN

MATEMATIKA ■

6

Ivan Ivanšić

ELEMENT
ELEMENT
NUMERIČKA
MATEMATIKA

ISBN 953-197-526-4



9 789531 975261

Prof. dr. sc. Ivan Ivanišić

NUMERIČKA MATEMATIKA



ISBN 953-197-526-4

Ivan Ivanšić

Redoviti profesor Fakulteta elektrotehnike i računarstva
Zavod za primijenjenu matematiku

NUMERIČKA MATEMATIKA

Zagreb, 2002.

© Prof. dr. sc. Ivan Ivanšić, 1998.

Urednik
Prof. dr. sc. Neven Elezović

Slog, crteži i prijelom
Nataša Jocić, dipl. inž

Design ovitka
Palete, Zagreb

Nakladnik
ELEMENT, Zagreb, Republike Austrije 11
tel. 01/37-777-37, 01/37-777-44, 01/37-777-52
faks 01/37-736-41
<http://www.element.hr/>
e-mail: element@element.hr

Tisak
Nova promocija, Zagreb

PREDGOVOR

Numerička analiza je vrlo opsežna grana matematike koja je u punom razvoju iako su prvi numerički postupci stari koliko i cijela matematika. Ne samo da se rješavaju novi problemi, već se i za klasične probleme, koje su nekada rješavali npr. Newton, Gauss, Euler i drugi, nalaze pod stanovitim (za primjenu važnim) uvjetima novi računski efikasni postupci (algoritmi). Zbog te opsežnosti ova knjiga prvenstveno prati program predmeta "Numerička matematika" koji se predaje studentima Fakulteta elektrotehnike i računarstva (FER) u Zagrebu i time nosi naziv tog predmeta. Time prvenstveno služi studentima FER-a, ali može poslužiti i drugima u dijelu njihovog programa koji se preklapa sa gore spomenutim. Pretpostavlja se da je student prošao kroz uobičajeni tečaj matematičke analize i linearne algebre i taj se aparat koristi uz mali podsjetnik na licu mesta.

Kod proučavanja numeričkih metoda nije dovoljno poznavati samo "receipt", potrebno ga je razumijeti barem u svojem elementarnom dijelu. Stoga su gotovo svi postupci izvedeni.

Za primjenu je važno poznavati i ocjenu pogreške upotrebljenog postupka, pa su iste izvedene. U nekoliko slučajeva su navedeni gotovi rezultati ili se poziva na neki detaljno obrađeni dokaz jer je promatrani potpuno analogan. Spomenimo da je obrađeno i nekoliko "mladih postupaka" izvan gornjeg programa. Ti postupci imaju značajnu primjenu u novije vrijeme.

Svi prezentirani postupci su računski upotrebljivi i široko se koriste. Da pojasnimo misao o upotrebljivosti postupka ilustrirajmo ju na primjeru tzv. Hornerove šeme pomoći koje se računa vrijednost polinoma

$$P(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \dots + a_{n-1}x + a_n = \sum_{i=0}^n a_i x^{n-i}.$$

Prepostavimo da treba izračunati $P(x_0)$, tj. vrijednost polinoma za $x = x_0$. Čitanjem zapisa polinoma nudi se postupak da izračunamo redom sve potencije x_0^i , $i = 0, 1, 2, \dots, n$; zatim svaku pomnožimo sa a_{n-i} i onda to sve zbrojimo. Da ne duljimo oko toga da je taki postupak nespretan napišimo odmah polinom u ovom obliku (zapis izvodimo postupno):

$$\begin{aligned} P(x) &= x(a_0x^{n-1} + a_1x^{n-2} + \dots + a_{n-2}x + a_{n-1}) + a_n \\ &= x(x(a_0x^{n-2} + a_1x^{n-3} + \dots + a_{n-2}) + a_{n-1}) + a_n \\ &= x(x(x(a_0x^{n-3} + a_1x^{n-4} + \dots + a_{n-3}) + a_{n-2}) + a_{n-1}) + a_n \\ &\quad \dots \dots \dots \\ &= x(x(\dots(x(xa_0 + a_1) + a_2) + \dots + a_{n-2}) + a_{n-1}) + a_n. \end{aligned}$$

Pogledajmo u čemu je prednost ovakvog zapisa.

Zagrade nam kažu u kojem redoslijedu se provode operacije. Polazimo odnutarnje zagrade u kojoj za $x = x_0$ treba izračunati vrijednost polinoma prvog stupnja $a_0x + a_1$. Taj rezultat glasi

$$a_0x_0 + a_1 = y_1.$$

Izračunavši y_1 prelazimo na drugu zgradu (glezano naravno iz unutra prema van) i vidimo da za dobiveni y_1 treba opet izračunati vrijednost polinoma prvog stupnja, koji rezultat glasi

$$y_1 x_0 + a_2 = y_2.$$

I sljedeći korak je isti s time da ulogu od y_1 preuzima y_2 i računamo

$$y_2 x_0 + a_3 = y_3,$$

i tako redom

$$y_3 x_0 + a_4 = y_4$$

.....

$$y_{i-1} x_0 + a_i = y_i$$

.....

$$y_{n-1} x_0 + a_n = y_n = P(x_0)$$

da u zadnjem koraku izračunamo vrijednost danog polinoma u točki x_0 . Vidimo da je a_0 bio početni y_0 i da u svakom koraku imamo isti račun s promijenjenim podacima, naime mijenjaju se koeficijenti polinoma prvog stupnja dok je argument x_0 stalno isti. Nadalje, ono što smo u jednom koraku izračunali koristimo u sljedećem pa kažemo da se radi o iteracijskom postupku. Takav se postupak lako programira na računskom stroju. Spomenimo da se može i tabelirati što obično susrećemo u ovom obliku

	a_0	a_1	a_2	a_3	...	a_n	$ x_0$
+	$y_0 x_0$	$y_1 x_0$	$y_2 x_0$...		$y_{n-1} x_0$	
$a_0 = y_0$	y_1	y_2	y_3			$y_n = P(x_0)$	

Tablica se izgradi ovako. U prvi red popišemo koeficijente polinoma P , zatim prema gornjim formulama formiramo drugi i treći red. To znači da prvo u treći red upišemo $a_0 = y_0$, koji zatim pomnožimo sa x_0 (označen na desno na tablice) i potpišemo ispod a_1 . Zbrojimo li a_1 i potpisani $y_0 x_0$ (što ističe znak "+" na lijevo od tablice) dobivamo y_1 . Zatim y_1 pomnožimo sa x_0 i produkt potpišemo ispod a_2 , to zbrojeno daje y_2 , i tako redom dok ne dođemo do kraja dobivši $y_n = P(x_n)$.

Teme obrađene u ovoj knjizi daju u stanovitom smislu temeljna znanja. Zbog opsežnosti numeričke analize imamo specijalizirane knjige, koje obrađuju pojedina područja numeričke analize, kakvih prije pedesetak godina gotovo da nije bilo. Ilustrirajmo to na sustavu linearnih jednadžbi. U poglavljju "Sustavi linearnih jednadžbi" obrađene su direktnе i iteracijske metode koje su u principu primjenjive na opći slučaj. Međutim, danas se u istraživanjima primjenjuju sustavi s velikim brojem nepoznanica. Pri tome često matrica takvog sustava ima mnogo elemenata koji su nula (još se kaže da je matrica rijetka) (engl. sparse). Da bi se efikasno riješio takav sustav potrebna je metoda koja tu pojavu uzima u obzir, pa je time efikasnija od neke opće metode. Stoga će oni koji se budu bavili istraživanjima u kojima se pojavljuju posebni problemi morati posegnuti za specijaliziranom literaturom, a stečeno znanje će im pomoći da to razumiju.

I u ovoj knjizi ima pogrešaka. Ako mi na njih ukažete, onda postoji mogućnost da tih ne bude u eventualno ponovljenom izdanju.

Zagreb, 20. ožujka 1998.

Autor

SADRŽAJ

1. Interpolacija i aproksimacija funkcija	1
1.1. Interpolacijski polinom	1
1.2. Interpolacijski splajn	18
1.3. Trigonometrijska interpolacija	23
1.4. Čebiševljevi polinomi, minimaks polinom i teleskopiranje redova potencija	31
1.5. Polinom najmanjih kvadrata	36
2. Numeričko integriranje	40
2.1. Newton–Cotesove formule	41
2.2. Ponavljanje raspolažljanja i Rombergov algoritam	50
2.3. Gaussove formule	54
2.4. Integriranje brzo oscilirajućih funkcija. Filonova formula	58
3. Sustavi linearnih jednadžbi	65
3.1. Direktne metode	66
3.2. Iteracijske metode	85
4. Svojstvene vrijednosti i svojstveni vektori matrice	110
4.1. Metoda neodređenih koeficijenata	115
4.2. Danilevskijeva metoda	117
4.3. Krilovljeva metoda	124
4.4. Nalaženje karakterističnog polinoma Leverrierovom metodom	127
4.5. Misesova metoda potencija — spektralni radijus matrice	129
5. Nelinearne jednadžbe	134
5.1. Jednadžbe s jednom nepoznanicom	134
5.2. Algebarske jednadžbe	148
5.3. Sustavi nelinearnih jednadžbi	164
6. Aproksimacija rješenja običnih diferencijalnih jednadžbi	180
6.1. Metoda sukcesivne aproksimacije	181
6.2. Eulerova metoda	185
6.3. Runge–Kuttini postupci	189
Kazalo	194
Literatura	196

1.

Interpolacija i aproksimacija funkcija

1.1. Interpolacijski polinom	1
1.2. Interpolacijski splajn	18
1.3. Trigonometrijska interpolacija	23
1.4. Čebiševljevi polinomi, minimaks polinom i teleskopiranje redova potencija	31
1.5. Polinom najmanjih kvadrata	36

Jedan od osnovnih problema numeričkih metoda je kako aproksimirati danu funkciju f pomoću funkcije g koja je prikladnija za izračunavanje, te ocijeniti pogrešku koja je učinjena pri takvoj aproksimaciji. Uobičajeni oblik aproksimacije g je linearna kombinacija

$$g(x) = a_0 g_0(x) + a_1 g_1(x) + \dots + a_n g_n(x)$$

nekih “jednostavnih” funkcija g_i , $i = 0, 1, 2, \dots, n$. Najčešći izbor funkcija g_k su potencije x^k , trigonometrijske funkcije $\sin kx$, $\cos kx$, te eksponencijalne funkcije $e^{b_k x}$. Nadalje, same g_k mogu biti linearne kombinacije kao što su npr. ortogonalni polinomi itd. U novije vrijeme sve više se koriste racionalne funkcije, tj. funkcije oblika

$$g(x) = \frac{a_0 + a_1 x + \dots + a_n x^n}{b_0 + b_1 x + \dots + b_m x^m} = \frac{P_n(x)}{P_m(x)}.$$

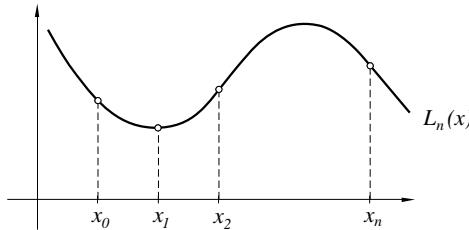
Ovdje ćemo se najviše ograničiti na aproksimaciju polinomima.

1.1. Interpolacijski polinom

Da bismo motivirali problem zamislimo da eksperimentalnim mjeranjem istražujemo nepoznatu funkciju $x \mapsto f(x)$. Time dobivamo seriju podataka

$$(x_0, f(x_0)), \quad (x_1, f(x_1)), \quad \dots, \quad (x_n, f(x_n)) \tag{1}$$

koji u ravnini predstavljaju točke grafa Γ_f (ne zaboravimo da su u (1) svi podaci aproksimativne vrijednosti).



Sl. 1.1.

Postavlja se pitanje izračunavanja aproksimativnih vrijednosti funkcije f izvan točaka x_i , tj. za vrijednosti argumenta x , $x \neq x_i$. Oblik funkcije nam je potpuno nepoznat (znamo samo njezinu vrijednosti u točkama x_0, x_1, \dots, x_n). Stoga nepoznatu funkciju f zamjenjujemo s drugom, nama poznatom funkcijom, koja ima iste vrijednosti u zadanim točkama. Svakako je najjednostavnija takva funkcija polinom.

To nas vodi na problem iznalaženja tzv. **interpolacijskog polinoma**

$$P(x) = a_0 + a_1x + \dots + a_nx^n \quad (2)$$

stupnja $\leq n$ čije se vrijednosti u točkama x_i podudaraju s vrijednostima $f(x_i)$, tj. vrijedi

$$P(x_i) = f(x_i), \quad i = 0, 1, \dots, n. \quad (3)$$

Točke x_i zovu se **čvorovi** (bazne točke ili interpolacijske točke), dok se P zove **interpolacijski polinom**.

Lako se uvjerimo da je interpolacijski polinom jedinstven, tj. da kroz $n+1$ čvoriste (1) prolazi samo jedan polinom P stupnja $\leq n$ takav da vrijedi $P(x_i) = f(x_i)$, $i = 0, \dots, n$. Naime, uvrstimo li redom točke x_i dobivamo sljedeći sustav linearnih jednadžbi po nepoznatim koeficijentima a_0, a_1, \dots, a_n .

$$\begin{aligned} a_0 + a_1x_0 + a_2x_0^2 + \dots + a_nx_0^n &= f(x_0) \\ a_0 + a_1x_1 + a_2x_1^2 + \dots + a_nx_1^n &= f(x_1) \\ &\vdots \\ a_0 + a_1x_n + a_2x_n^2 + \dots + a_nx_n^n &= f(x_n). \end{aligned} \quad (4)$$

Sustav (4) ima jedinstveno rješenje zbog toga što je determinanta sustava

$$\begin{vmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ & \vdots & & & \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{vmatrix} = \prod_{i>j} (x_i - x_j) \neq 0. \quad (5)$$

Determinanta (5) poznata je pod imenom Vandermondeova determinanta, čija je vrijednost jednaka umnošku svih razlika $x_i - x_j$, $i > j$. Rješenjem sustava (4) dobivamo koeficijente a_0, a_1, \dots, a_n , odnosno traženi interpolacijski polinom.

Lagrangeov oblik interpolacijskog polinoma

Rješavanje sustava jednadžbi (4) iz uvoda nije jednostavno, međutim, pokazuje se da se P može pronaći u zapisu koji je različit od zapisa (2) iz uvoda, pa ovdje izvodimo zapis od P u formi u kojoj se naziva Lagrangeov oblik interpolacijskog polinoma i označava sa L_n . Ograničimo li se na restrikciju od f na skup $\{x_0, x_1, \dots, x_n\}$, vidimo da tu restrikciju možemo prikazati kao linearu kombinaciju ovih $n + 1$ funkcija

$$\varphi_i(x_j) = \delta_{ij} = \begin{cases} 1, & \text{za } j = i, \\ 0, & \text{za } j \neq i, \end{cases} \quad i, j = 0, 1, \dots, n,$$

jer očito vrijedi

$$f(x_j) = \sum_{i=0}^n f(x_i) \varphi_i(x_j). \quad (6)$$

Pronađemo li interpolacijski polinom p_i za svaku od pomoćnih funkcija φ_i , dobit ćemo traženi interpolacijski polinom L_n kao

$$L_n(x) = \sum_{i=0}^n f(x_i) p_i(x). \quad (7)$$

Interpolacijski polinom p_i lako nalazimo, zato što su sva čvorišta, osim x_i , njegove nultočke. Imamo dakle da je

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

gdje je C_i konstanta, koju lako određujemo uvrštavanjem $x = x_i$. Kako je $p_i(x_i) = 1$, imamo

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

odnosno

$$C_i = \frac{1}{(x_i - x_0)(x_i - x_1) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_n)}. \quad (8)$$

Time je

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

što prema (7) daje

$$\begin{aligned} L_n(x) &= \sum_{i=0}^n f(x_i) \frac{(x - x_0)(x - x_1) \cdots (x - x_{i-1})(x - x_{i+1}) \cdots (x - x_n)}{(x_i - x_0)(x_i - x_1) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_n)} \\ &= \sum_{i=0}^n f(x_i) \prod_{\substack{i \neq j \\ 0 \leq j \leq n}} \frac{x - x_j}{x_i - x_j}. \end{aligned} \quad (9)$$

Zapis (9) zove se **Lagrangeov oblik interpolacijskog polinoma**. Vidimo da L_n nismo dobili u obliku razvijenom po potencijama od x već kao linearu kombinaciju polinoma p_i , što je posljedica postupka nalaženja L_n . Želimo li izračunati koeficijent

a_k uz potenciju x^k , moramo izvršiti pripadna množenja i zbrajanja. Za $n = 1$ imamo jednadžbu pravca kroz točke $(x_0, f(x_0))$, $(x_1, f(x_1))$, napisanu u ovom obliku:

$$L_1(x) = \frac{x - x_1}{x_0 - x_1} f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1), \quad (10)$$

dok za $n = 2$ dobivamo jednadžbu parabole kroz zadane tri točke, napisanu u ovom obliku

$$L_2(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} f(x_0) + \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} f(x_1) + \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} f(x_2). \quad (11)$$

Primjer 1.1. Odredite interpolacijski polinom za funkciju $y = \sin \pi x$ pri sljedećem izboru čvorova: $x_0 = 0$, $x_1 = \frac{1}{6}$, $x_2 = \frac{1}{2}$.

Lako sada izračunamo $f(x_0) = 0$, $f(x_1) = \sin \frac{\pi}{6} = \frac{1}{2}$, $f(x_2) = \sin \frac{\pi}{2} = 1$ pa uvrštavanjem u (11) i sređivanjem dobivamo

$$L_2(x) = \frac{7}{2}x - 3x^2.$$

Sada npr. za $x = \frac{1}{4}$ dobivamo $\sin \frac{\pi}{4} \approx 0.6875$. To je aproksimacija od $\sin \frac{\pi}{4} = \frac{\sqrt{2}}{2} \approx 0.7021$ s greškom $< 2 \cdot 10^{-2}$ što može zadovoljavati neke račune.

Lagrangeov oblik interpolacijskog polinoma za ekvidistantne čvorove

Izvedimo i Lagrangeov oblik interpolacijskog polinoma za ekvidistantne čvorove, tj. $x_{i+1} - x_i = h$, $i = 0, 1, 2, \dots, n - 1$. Veličinu h zovemo **korak interpolacije**. Označimo li sa $t = \frac{x - x_0}{h}$, imamo: $x - x_0 = th$, $x = x_0 + th$; $x - x_1 = (x - x_0) - (x_1 - x_0) = th - h$; $x_i - x_0 = ih$, $x_i - x_1 = (i - 1)h$, itd. pa možemo pisati

$$\begin{aligned} & \frac{(x - x_0)(x - x_1) \cdots (x - x_{i-1})(x - x_{i+1}) \cdots (x - x_n)}{(x_i - x_0)(x_i - x_1) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_n)} \\ &= \frac{th(th - h) \cdots [th - (i - 1)h][th - (i + 1)h] \cdots (th - nh)}{ih(i - 1)h \cdots h(-h) \cdots [-(n - i)h]} \\ &= \frac{t(t - 1) \cdots (t - n)}{t - i} \cdot \frac{(-1)^{n-i}}{i!(n - i)!} \\ &= \frac{t(t - 1) \cdots (t - n)}{t - i} \cdot \frac{(-1)^{n-i}}{n!} \cdot \frac{n!}{i!(n - i)!} \\ &= (-1)^{n-i} \binom{n}{i} \cdot \frac{1}{t - i} \cdot \frac{t(t - 1) \cdots (t - n)}{n!} \end{aligned} \quad (12)$$

Time Lagrangeov oblik interpolacijskog polinoma (9) poprima oblik

$$L_n(x) = L_n(x_0 + th) = (-1)^n \frac{t(t - 1) \cdots (t - n)}{n!} \sum_{i=0}^n (-1)^i \binom{n}{i} \frac{f(x_i)}{t - i}. \quad (13)$$

Iz (12) vidimo da koeficijenti uz $f(x_i)$ ne ovise niti od f niti od koraka h . Stoga se koeficijenti (12) mogu tabelirati. Takve tablice postoje i zovu se **tablice Lagrangeovih koeficijenata**.

Sada $L_2(x)$ izgleda ovako:

$$L_2(x) = L_2(x_0 + th) = \frac{t(t-1)(t-2)}{2} \left[\frac{f(x_0)}{t} - 2\frac{f(x_0+h)}{t-1} + \frac{f(x_0+2h)}{t-2} \right],$$

što je naravno polinom drugog stupnja u varijabli t .

Izračunavanje interpolacijskog polinoma, Aitkenova interpolacijska shema

Kao što znamo, polinomi su funkcije koje su definirane na svim realnim brojevima, pa prema tome možemo izračunati $L_n(x)$ za bilo koji realan broj $x \in \mathbf{R}$. Ipak, u tome razlikujemo dva slučaja obzirom na polaznu funkciju f . Ako izračunamo $L_n(x)$ za neki x između x_0 i x_n , tj. $x_0 < x < x_n$, onda govorimo da smo izvršili **interpolaciju**, dok za ostale x , tj. x izvan segmenta $[x_0, x_n]$, kažemo da smo izvršili **ekstrapolaciju**.

U slučaju da nam ne treba opći izraz za L_n nego samo njegove vrijednosti za neke x spretno je služiti se tzv. **Aitkenovom interpolacijskom šemom (Aitkenov algoritam)** koja se još zove i **iterirana linearna interpolacija**. Radi kraćeg zapisanja označimo $f(x_i) = y_i$. Napišimo ponovo jednadžbu linearog interpolacijskog polinoma koji prolazi točkama (x_0, y_0) i (x_1, y_1) . Označimo taj polinom sa L_{01} . Prema (10) imamo slijedeći zapis

$$\begin{aligned} L_{01}(x) &= \frac{x - x_1}{x_0 - x_1} y_0 + \frac{x - x_0}{x_1 - x_0} y_1 \\ &= \frac{1}{x_1 - x_0} [y_0(x_1 - x) - y_1(x_0 - x)] \\ &= \frac{1}{x_1 - x_0} \begin{vmatrix} y_0 & x_0 - x \\ y_1 & x_1 - x \end{vmatrix}. \end{aligned} \tag{14}$$

Analogno možemo odrediti L_{12}, L_{23}, \dots , odnosno L_{jk} , $j < k$. Promotrimo interpolacijski polinom drugog stupnja kroz točke $(x_0, y_0), (x_1, y_1), (x_2, y_2)$ koji označavamo sa L_{012} . Prema (10) imamo

$$L_{012}(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} y_0 + \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} y_1 + \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} y_2.$$

Uočimo da rastavljanjem na parcijalne razlomke vrijedi

$$\frac{1}{(x-a)(x-b)} = \frac{1}{a-b} \frac{1}{x-a} + \frac{1}{b-a} \frac{1}{x-b}$$

što primjenjeno na srednji pribrojnik od L_{012} daje

$$\frac{1}{(x_1 - x_0)(x_1 - x_2)} = \frac{1}{x_0 - x_2} \frac{1}{x_1 - x_0} + \frac{1}{x_2 - x_0} \frac{1}{x_1 - x_2}$$

tako da L_{012} možemo ovako zapisati

$$\begin{aligned} L_{012}(x) &= \frac{1}{x_2 - x_0} \left\{ \frac{(x - x_1)(x_2 - x)}{x_0 - x_1} y_0 + \frac{(x - x_0)(x_2 - x)}{x_1 - x_0} y_1 \right. \\ &\quad \left. - \frac{(x_0 - x)(x - x_2)}{x_1 - x_2} y_1 + \frac{(x - x_0)(x - x_1)}{x_2 - x_1} y_2 \right\} \\ &= \frac{1}{x_2 - x_0} \left\{ \left[\frac{x_1 - x}{x_1 - x_0} y_0 - \frac{x_0 - x}{x_1 - x_0} y_1 \right] (x_2 - x) \right. \\ &\quad \left. - \left[\frac{x_2 - x}{x_2 - x_1} y_1 - \frac{x_1 - x}{x_2 - x_1} y_2 \right] (x_0 - x) \right\} \\ &= \frac{1}{x_2 - x_0} \left\{ L_{01}(x)(x_2 - x) - L_{12}(x)(x_0 - x) \right\}. \end{aligned}$$

Izraz u zagradi je ponovo razvoj determinante drugog reda što u takvom zapisu izgleda ovako

$$L_{012}(x) = \frac{1}{x_2 - x_0} \begin{vmatrix} L_{01}(x) & x_0 - x \\ L_{12}(x) & x_2 - x \end{vmatrix}. \quad (15)$$

Izrazi (14) i (15) su potpuno analogno građeni pa naslućujemo da općenito vrijedi

$$L_n(x) = L_{01\dots n}(x) = \frac{1}{x_n - x_0} \begin{vmatrix} L_{01\dots(n-1)}(x) & x_0 - x \\ L_{12\dots n}(x) & x_n - x \end{vmatrix}. \quad (16)$$

Formulu (16) lako dokazujemo indukcijom. Primijetimo da $L_{01\dots(n-1)}$ nije ništa drugo nego interpolacijski polinom za čvorove x_0, x_1, \dots, x_{n-1} , dok je $L_{12\dots n}$ interpolacijski polinom za čvorove x_1, x_2, \dots, x_n . Zapišemo li ih u Lagrangeovom obliku (9) imamo

$$L_{01\dots(n-1)}(x) = \sum_{i=0}^{n-1} y_i \frac{(x - x_0)(x - x_1) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_{n-1})}{(x_i - x_0)(x_i - x_1) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_{n-1})}, \quad (17)$$

odnosno

$$L_{12\dots n}(x) = \sum_{i=1}^n y_i \frac{(x - x_1)(x - x_2) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_i - x_1)(x_i - x_2) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)}. \quad (18)$$

Izračunamo li determinantu (16) dobivamo

$$L_{01\dots n}(x) = \frac{(x_n - x)L_{01\dots(n-1)} - (x_0 - x)L_{12\dots n}}{x_n - x_0}.$$

Očigledno je da prvi pribrojnik sume (17) pomnožen sa $\frac{x_n - x}{x_n - x_0} = \frac{x - x_n}{x_0 - x_n}$ predstavlja prvi pribrojnik sume (9) i da zadnji pribrojnik sume (18) pomnožen sa $-\frac{x_0 - x}{x_n - x_0} = \frac{x - x_0}{x_n - x_0}$ predstavlja zadnji pribrojnik sume (9). Ostaje još da pogledamo koeficijente od y_i za $1 \leq i \leq n-1$. Tada je koeficijent od y_i jednak

$$\begin{aligned} \frac{1}{x_n - x_0} \left\{ -\frac{(x - x_0) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_{n-1})}{(x_i - x_0) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_{n-1})} (x - x_n) \right. \\ \left. + (x - x_0) \frac{(x - x_1) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_i - x_1) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)} \right\} \end{aligned}$$

$$= \frac{1}{x_n - x_0} \frac{(x-x_0)(x-x_1) \dots (x-x_{i-1})(x-x_{i+1}) \dots (x-x_{n-1})(x-x_n)[-x_i+x_n-x_i-x_0]}{(x_i-x_0)(x_i-x_1) \dots (x_i-x_{i-1})(x_i-x_{i+1}) \dots (x_i-x_{n-1})(x_i-x_n)}$$

što je točno koeficijent od y_i u prikazu (9).

Formula (16) daje drugi zapis interpolacijskog polinoma. Taj zapis je veoma podesan za izračunavanje vrijednosti interpolacijskog polinoma, posebno na računskim strojevima jer se radi o višestrukoj uzastopnoj primjeni istog postupka. Šema računanja interpolacijskog polinoma izgleda sada ovako:

x_i	y_i	$x_i - x$	$L_{i-1,i}$	$L_{i-2,i-1,i}$	$L_{i-3,i-2,i-1,i}$	$L_{i-4,i-3,i-2,i-1,i}$
x_0	y_0	$x_0 - x$				
x_1	y_1	$x_1 - x$	$L_{01}(x)$			
x_2	y_2	$x_2 - x$	$L_{12}(x)$	$L_{012}(x)$		
x_3	y_3	$x_3 - x$	$L_{23}(x)$	$L_{123}(x)$	$L_{0123}(x)$	
x_4	y_4	$x_4 - x$	$L_{34}(x)$	$L_{234}(x)$	$L_{1234}(x)$	$L_{01234}(x)$
x_5	y_5	$x_5 - x$	$L_{45}(x)$	$L_{345}(x)$	$L_{2345}(x)$	$L_{12345}(x)$

Kod iznalaženja interpolacijskog polinoma nismo ništa govorili o poređaju čvorova. Automatski smo mislili da ih uzimamo po veličini, međutim, to nismo nigdje koristili. Možemo dakle proizvoljno permutirati čvorove i zatim ispisivati Lagrangeov interpolacijski polinom, dobivamo isti rezultat. Sada se može izvršiti jedno opće kombinatoričko razmatranje i uz nešto računanja dobiti jedan opći zapis interpolacijskog polinoma iz kojega se specijaliziranjem indeksa i permutacija dobiva nekoliko drugih zapisa (oblika) interpolacijskog polinoma. To ovdje izostavljamo.

Vratimo se primjeru aproksimacije $y = \sin \pi x$ za $x = 0.25 = \frac{1}{4}$ iz 1.1. Sada bez nalaženja interpolacijskog polinoma imamo postupnim računanjem sljedeću tablicu čija je zadnja vrijednost jednakna $L_{012}(\frac{1}{4})$.

x_i	y_i	$x_i - 0.25$		
0	0	-0.25		
0.1666	0.5	-0.0834	0.75	
0.5	1	0.25	0.6252	$0.6876 \approx \sin \frac{\pi}{4}$

Newtonovi oblici interpolacijskog polinoma

Prvo ćemo izvesti tzv. Newtonov oblik s kvocijentima diferencija. Već na samom početku diferencijalnog računa sreli smo pojam kvocijenta diferencija i na njemu proveli granični proces da dobijemo derivaciju u točki. Sada ne provodimo granični proces nego kratko uvodimo odgovarajuće pojmove.

Za točke $(x_0, f(x_0))$ i $(x_1, f(x_1))$ definiramo i označavamo kvocijent diferencija funkcije f kao

$$f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{f(x_0)}{x_0 - x_1} + \frac{f(x_1)}{x_1 - x_0} \quad (19)$$

i zovemo **kvocijent diferencija**. Za tri točke $(x_0, f(x_0)), (x_1, f(x_1)), (x_2, f(x_2))$ uvodimo **kvocijent diferencija drugog reda** funkcije f kao

$$\begin{aligned} f[x_0, x_1, x_2] &= \frac{f[x_2, x_1] - f[x_1, x_0]}{x_2 - x_0} \\ &= \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0} \\ &= \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)} + \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2)} + \frac{f(x_2)}{(x_2 - x_0)(x_2 - x_1)} \end{aligned} \quad (20)$$

tj. kao kvocijent diferencija od kvocijenata diferencija prvog reda. Induktivno definiramo kvocijent diferencija n -tog reda **kvocijent diferencija n -tog reda** od f kao

$$f[x_0, x_1, \dots, x_n] = \frac{f[x_1, x_2, \dots, x_n] - f[x_0, x_1, \dots, x_{n-1}]}{x_n - x_0}. \quad (21)$$

Za jednu točku $(x_0, f(x_0))$ definiramo kvocijent diferencija 0-tog reda kao funkciju vrijednost u toj točki i označavamo $f[x_0] = f(x_0)$.

Dokažimo da za kvocijent diferencija vrijedi formula

$$\begin{aligned} f[x_0, x_1, \dots, x_n] &= \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2) \dots (x_0 - x_n)} + \dots \\ &\quad + \frac{f(x_n)}{(x_n - x_0)(x_n - x_1) \dots (x_n - x_{n-1})} \\ &= \sum_{i=0}^n \frac{f(x_i)}{\prod_{\substack{i \neq j \\ 0 \leq j \leq n}} (x_i - x_j)}. \end{aligned} \quad (22)$$

Jednakost (22) dokazuje se indukcijom. Za $n = 1$ formula (19) daje takav rastav. Pretpostavimo induktivno da rastav vrijedi za prirodne brojeve $< n$. Sada je po definiciji

$$f[x_0, x_1, \dots, x_n] = \frac{1}{x_n - x_0} \{f[x_1, x_2, \dots, x_n] - f[x_0, x_1, \dots, x_{n-1}]\}.$$

Prema induktivnoj prepostavci možemo pisati

$$\begin{aligned} f[x_1, x_2, \dots, x_n] &= \frac{f(x_1)}{(x_1 - x_2)(x_1 - x_3) \dots (x_1 - x_n)} + \dots \\ &\quad + \frac{f(x_n)}{(x_n - x_1)(x_n - x_2) \dots (x_n - x_{(n-1)})} \\ &= \sum_{i=1}^n \frac{f(x_i)}{\prod_{\substack{i \neq j \\ 1 \leq j \leq n}} (x_i - x_j)}, \end{aligned}$$

i analogno

$$f[x_0, x_1, \dots, x_{n-1}] = \sum_{i=0}^{n-1} \frac{f(x_i)}{\prod_{\substack{i \neq j \\ 0 \leq j \leq n-1}} (x_i - x_j)}.$$

Sada je

$$f[x_0, x_1, \dots, x_n] = \frac{1}{x_n - x_0} \left\{ \sum_{i=1}^n \frac{f(x_i)}{\prod_{\substack{i \neq j \\ 1 \leq j \leq n}} (x_i - x_j)} - \sum_{i=0}^{n-1} \frac{f(x_i)}{\prod_{\substack{i \neq j \\ 0 \leq j \leq n-1}} (x_i - x_j)} \right\}. \quad (23)$$

Iz (23) direktno čitamo da uz $f(x_0)$ i $f(x_n)$ imamo koeficijent koji propisuje (22). Ostaje da to provjerimo za indekse $1 \leq j \leq n-1$. Nadalje čitamo da je u (23) koeficijent od $f(x_j)$ jednak

$$\begin{aligned} \frac{1}{x_n - x_0} \left\{ \frac{1}{\prod_{\substack{i \neq j \\ 1 \leq j \leq n}} (x_i - x_j)} - \frac{1}{\prod_{\substack{i \neq j \\ 0 \leq j \leq n-1}} (x_i - x_j)} \right\} &= \frac{(x_i - x_0) - (x_i - x_n)}{(x_n - x_0) \prod_{\substack{i \neq j \\ 0 \leq j \leq n}} (x_i - x_j)} \\ &= \frac{x_n - x_0}{(x_n - x_0) \prod_{\substack{i \neq j \\ 0 \leq j \leq n}} (x_i - x_j)} = \frac{1}{\prod_{\substack{i \neq j \\ 0 \leq j \leq n}} (x_i - x_j)}, \end{aligned}$$

čime je dokaz proveden.

Iz definicije (21) i jednakosti (22) lako se dokazuju ova osnovna svojstva kvocijenta diferencija n -tog reda:

a) kvocijent diferencija n -tog reda je linearan operator, tj. vrijedi

$$(\alpha f + \beta g)[x_0, x_1, \dots, x_n] = \alpha f[x_0, x_1, \dots, x_n] + \beta g[x_0, x_1, \dots, x_n];$$

b) kvocijent diferencija je simetrična funkcija svojih argumenata x_0, x_1, \dots, x_n , tj. ne mijenja vrijednost pri proizvoljnoj permutaciji argumenata.

Ako je funkcija f zadana u točkama x_0, x_1, \dots, x_n onda se tablica

$$\begin{array}{ccccccc} f[x_0] & & f[x_0, x_1] & & & & \\ f[x_1] & & f[x_0, x_1, x_2] & & & & \\ f[x_2] & & f[x_1, x_2] & & \ddots & & \\ f[x_3] & & f[x_2, x_3] & & & & f[x_0, x_1, \dots, x_n] \\ & & \vdots & & \ddots & & \\ & & f[x_{n-2}, x_{n-1}, x_n] & & & & \\ & \vdots & & & & & \\ f[x_n] & & & & & & \end{array}$$

zove **tablica kvocijenata diferencija od f** .

Da iz Lagrangeovog oblika interpolacijskog polinoma izvedemo Newtonov postupimo ovako. Promatrajmo u prvom redu razliku

$$\begin{aligned} f(x) - L_n(x) &= f(x) - \sum_{i=0}^n f(x_i) \prod_{\substack{i \neq j \\ 0 \leq j \leq n}} \frac{x - x_j}{x_i - x_j} \\ &= \prod_{0 \leq j \leq n} (x - x_j) \left\{ \frac{f(x)}{\prod_{0 \leq j \leq n} (x - x_i)} + \sum_{i=0}^n \frac{f(x_i)}{(x_i - x) \prod_{\substack{j \neq i \\ 0 \leq j \leq n}} (x_i - x_j)} \right\}, \end{aligned}$$

pri čemu smo drugu jednakost dobili izlučivanjem produkta ispred vitičaste zagrade. Promatramo li izraz u vitičastoj zagradi uviđamo (prema (4)) da iznosi $f[x, x_0, x_1, \dots, x_n]$, tj. da je to kvocijent diferencija $(n+1)$ -og reda (x je shvaćen kao dodatno čvorište), pa možemo pisati

$$f(x) - L_n(x) = f[x, x_0, x_1, \dots, x_n] \prod_{0 \leq j \leq n} (x - x_j) \quad (24)$$

što će nam trenutak kasnije korisno poslužiti. Nadalje, zapišimo interpolacijski polinom L_n u obliku

$$L_n(x) = L_0(x) + (L_1(x) - L_0(x)) + \dots + (L_n(x) - L_{n-1}(x)) \quad (25)$$

gdje L_k označava interpolacijski polinom u Lagrangeovom obliku određen čvorovima x_0, x_1, \dots, x_k i $L_0(x) = f(x_0)$. Tada je razlika $L_k(x) - L_{k-1}(x)$ polinom k -tog stupnja kojemu su x_0, x_1, \dots, x_{k-1} nultočke jer vrijedi $f(x_j) = L_{k-1}(x_j) = L_k(x_j)$ za $j = 0, 1, 2, \dots, k-1$. Prema tome možemo pisati

$$\begin{aligned} L_k(x) - L_{k-1}(x) &= A_{k-1}(x - x_0)(x - x_1) \dots (x - x_{k-1}) \\ &= A_{k-1} \prod_{0 \leq j \leq k-1} (x - x_j), \end{aligned} \quad (26)$$

gdje koeficijent A_{k-1} moramo odrediti. Za $x = x_k$ dobivamo iz (26)

$$L_k(x_k) - L_{k-1}(x_k) = A_{k-1} \prod_{0 \leq j \leq k-1} (x_k - x_j),$$

dok iz (24) za $n = k-1$ i $x = x_k$ dobivamo

$$f(x_k) - L_{k-1}(x_k) = f[x_k, x_0, \dots, x_{k-1}] \prod_{0 \leq j \leq k-1} (x_k - x_j).$$

Kako je $L_k(x_k) = f(x_k)$ to iz zadnje dvije jednakosti i svojstva b) slijedi

$$A_{k-1} = f[x_0, x_1, \dots, x_{k-1}, x_k].$$

Uvrstimo li dobiveni rezultat u (25) imamo

$$\begin{aligned} L_n(x) &= f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \dots \\ &\quad + f[x_0, x_1, \dots, x_n](x - x_0)(x - x_1) \dots (x - x_{n-1}) \end{aligned} \quad (27)$$

koji zapis interpolacijskog polinoma zovemo **Newtonov oblik s kvocijentima diferencija**.

Stupanj pribrojnika u (27) raste, na osnovi čega direktno čitamo da je $L_n(x)$ polinom n -tog stupnja onda i samo onda ako je za dana čvorista pripadni kvocijent diferencija n -tog reda različit od nule.

Kada imamo ekvidistantne čvorove tada se interpolacijski polinom može zapisati pomoću viših diferencija unaprijed, odnosno unatrag.

Promatrajmo funkciju f u točki x . **Prva diferencija unaprijed** od f u x označava se sa $\Delta f(x)$ i definira kao

$$\Delta f(x) = f(x + h) - f(x). \quad (28)$$

Ako nam je poznata prva diferencija unaprijed od f u točki $x + h$, tj. ako znamo $\Delta f(x + h) = f(x + h + h) - f(x + h) = f(x + 2h) - f(x + h)$, onda drugu diferenciju od f u točki x definiramo kao

$$\begin{aligned} \Delta^2 f(x) &= \Delta f(x + h) - \Delta f(x) \\ &= f(x + 2h) - f(x + h) - (f(x + h) - f(x)) \\ &= f(x + 2h) - 2f(x + h) + f(x) \end{aligned} \quad (29)$$

tj. kao prvu diferenciju unaprijed od prve diferencije unaprijed. Induktivno definiramo r -tu diferenciju unaprijed od f u točki x kao

$$\Delta^r f(x) = \Delta^{r-1} f(x + h) - \Delta^{r-1} f(x). \quad (30)$$

Izvedimo vezu između diferencija unaprijed i kvocijenata diferencija. Sada je $x_{k+1} = x_k + h$, jer radimo s ekvidistantnim čvorovima, pa imamo

$$\frac{\Delta f(x_0)}{h} = \frac{f(x_0 + h) - f(x_0)}{h} = \frac{f(x_1) - f(x_0)}{h} = f[x_0, x_1]$$

iz čega slijedi

$$\Delta f(x_0) = 1!h f[x_0, x_1].$$

Za kvocijent diferencija drugog reda imamo

$$\begin{aligned} f[x_0, x_1, x_2] &= \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0} \\ &= \frac{f(x_0 + 2h) - 2f(x_0 + h) + f(x_0)}{2h^2} \\ &= \frac{\Delta^2 f(x_0)}{2h^2} \end{aligned}$$

iz čega slijedi

$$\Delta^2 f(x_0) = 2!h^2 f[x_0, x_1, x_2].$$

Općenito vrijedi formula

$$f[x_0, x_1, \dots, x_r] = \frac{\Delta^r f(x_0)}{r!h^r}, \quad x_{k+1} - x_k = h, \quad k = 0, 1, \dots, r-1, \quad (31)$$

ili

$$\Delta^r f(x_0) = r!h^r f[x_0, x_1, \dots, x_r], \quad x_{k+1} - x_k = h, \quad k = 0, 1, \dots, r-1. \quad (32)$$

Uvrstimo li redom (31) u (27) za $r = 1, 2, \dots, n$ dobivamo

$$\begin{aligned} L_n(x) &= f(x_0) + \frac{\Delta f(x_0)}{1!h}(x-x_0) + \frac{\Delta^2 f(x_0)}{2!h^2}(x-x_0)(x-x_1) + \\ &\quad \dots + \frac{\Delta^n f(x_0)}{n!h^n}(x-x_0)\dots(x-x_{n-1}). \end{aligned} \quad (33)$$

Uvedemo li u (33) $t = \frac{x-x_0}{h}$, ili $x = x_0 + th$, kao što smo to učinili za Lagrangeov oblik, formula (33) poprima ovaj oblik

$$\begin{aligned} L_n(x) &= L_n(x_0 + th) \\ &= f(x_0) + \Delta f(x_0)t + \Delta^2 f(x_0) \frac{t(t-1)}{2}! + \dots + \Delta^n f(x_0) \frac{t(t-1)\dots(t-n+1)}{n!} \end{aligned} \quad (34)$$

koji se zove **Newtonov oblik interpolacijskog polinoma s diferencijama unaprijed**. Usporedite formulu (34) sa formulom (13).

Ponekad je spretno služiti se s tzv. diferencijama unatrag. Tako se **diferencija unatrag od f u točki x** označava sa $\nabla f(x)$ i definira kao

$$\nabla f(x) = f(x) - f(x-h). \quad (35)$$

Induktivno se definira r -ta diferencija unatrag od f u točki x kao

$$\nabla^r f(x) = \nabla^{r-1} f(x) - \nabla^{(r-1)} f(x-h). \quad (36)$$

Pomičući se unatrag moramo sada krenuti od zadnjeg čvorišta x_n , tako da analogno formuli (31) sada dobivamo

$$f[x_n, x_{n-1}, \dots, x_{n-r}] = \frac{\nabla^r f(x_n)}{r!h^r} \quad (37)$$

ili

$$\nabla^r f(x_n) = r' h^r f[x_n, x_{n-1}, \dots, x_{n-r}]. \quad (38)$$

Uz (37) i (38) spomenimo samo da je zbog svojstva b) $f[x_n, x_{n-1}, \dots, x_{n-r}] = f[x_{n-r}, x_{n-r+1}, \dots, x_n]$.

Ako u formuli (27) pođemo od x_n prema x_0 ona poprima ovaj zapis

$$\begin{aligned} L_n(x) &= f(x_n) + f[x_n, x_{n-1}](x-x_n) + f[x_n, x_{n-1}, x_{n-2}](x-x_n)(x-x_{n-1}) + \dots \\ &\quad + f[x_n, x_{n-1}, \dots, x_0](x-x_n)(x-x_{n-1})\dots(x-x_1) \end{aligned} \quad (39)$$

u što redom uvrštavamo (37) za $r = 1, 2, \dots, n$ i dobivamo

$$\begin{aligned} L_n(x) &= f(x_n) + \frac{\nabla f(x_n)}{1!h}(x-x_n) + \frac{\nabla^2 f(x_n)}{2!h^2}(x-x_n)(x-x_{n-1}) + \dots \\ &\quad + \frac{\nabla^n f(x_n)}{n!h^n}(x-x_n)(x-x_{n-1})\dots(x-x_1). \end{aligned} \quad (40)$$

Ako i u (40) provedemo supstituciju $t = \frac{x-x_n}{h}$, $x = x_n + ht$ prelazi ona u

$$\begin{aligned} L_n(x) &= L_n(x_n + th) = f(x_n) + \nabla f(x_n) \frac{t}{1!} + \nabla^2 f(x_n) \frac{t(t+1)}{2!} + \dots \\ &\quad + \nabla^n f(x_n) \frac{t(t+1)\dots(t+n-1)}{n!} \end{aligned} \quad (41)$$

što se zove **Newtonov oblik interpolacijskog polinoma s diferencijama unatrag**.

Spomenimo na kraju da postoje i drugi oblici interpolacijskog polinoma u što se ovdje ne upuštamo.

Ocjena pogreške, minimiziranje ocjene pogreške

Ocijenimo pogrešku $|f(x) - L_n(x)|$ uz pretpostavku da funkcija f ima na segmentu $[a, b]$, $a \leq x_0 < x_1 < \dots < x_n \leq b$, neprekinute derivacije do uključivo $n+1$ -og reda. U tom cilju promatrajmo pomoćnu funkciju

$$\varphi(t) = f(t) - L_n(t) - \frac{f(x) - L_n(x)}{(x - x_0)(x - x_1) \dots (x - x_n)}(t - x_0)(t - x_1) \dots (t - x_n) \quad (42)$$

gdje je $x \neq x_i$ ($i = 0, 1, 2, \dots, n$) proizvoljna ali fiksirana točka segmenta $[a, b]$. Ta funkcija ima na $[a, b]$ derivacije do uključivo $(n+1)$ -og reda. Osim toga φ ima barem $n+2$ nultočke na $[a, b]$ jer vrijedi

$$\varphi(x_0) = \varphi(x_1) = \dots = \varphi(x_n) = \varphi(x) = 0.$$

Prema Rolleovom teoremu derivacija φ' se poništava barem jednom između susjednih nultočaka, pa se dakle poništava u barem $n+1$ točaka segmenta $[a, b]$. Neka su

$$\xi_0^{(1)}, \xi_1^{(1)}, \dots, \xi_n^{(1)} \in (a, b)$$

točke u kojima se φ' poništava, tj. vrijedi

$$\varphi'(\xi_i^{(1)}) = 0, \quad i = 0, 1, 2, \dots, n.$$

Primjenimo ponovo Rolleov teorem na funkciju φ' . Time analogno prethodnom dobivamo barem n točaka

$$\xi_0^{(2)}, \xi_1^{(2)}, \xi_2^{(2)}, \dots, \xi_{n-1}^{(2)}$$

u kojima se poništava druga derivacija φ'' , tj.

$$\varphi''(\xi_i^{(2)}) = 0, \quad i = 0, 1, 2, \dots, n-1.$$

Primijenimo li redom gornji postupak na funkcije $\varphi''', \dots, \varphi^{(n+1)}$ dolazimo do zaključka da postoji barem jedna točka

$$\xi_0^{(n+1)} = \xi \in (a, b)$$

za koju vrijedi

$$\varphi^{(n+1)}(\xi) = 0.$$

(Istaknimo da točka ξ ovisi o x .) Izračunamo li $(n+1)$ -vu derivaciju od φ dobivamo

$$\varphi^{(n+1)}(t) = f^{(n+1)}(t) - \frac{f(x) - L_n(x)}{(x - x_0)(x - x_1) \dots (x - x_n)}(n+1)!.$$

Uvrstimo li $t = \xi$ imamo

$$f^{(n+1)}(\xi) - \frac{f(x) - L_n(x)}{(x - x_0)(x - x_1) \dots (x - x_n)}(n+1)! = 0. \quad (43)$$

Iz (43) dobivamo pogrešku (ili ostatak) u ovom obliku

$$R_n(x) = f(x) - L_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)(x - x_1) \dots (x - x_n). \quad (44)$$

Neka je

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

Tada dobivamo sljedeću ocjenu absolutne pogreške

$$|R_n(x)| \leq \frac{M_{n+1}}{(n+1)!} |(x-x_0)(x-x_1)\dots(x-x_n)|. \quad (45)$$

Istaknimo na kraju da se (45) odnosi samo na točke segmenta $[x_0, x_n]$, tj. $x \in [x_0, x_n]$.

Ako se služimo ekvidistantnim čvorištima, onda (44) i (45) poprimaju ove oblike:

$$R_n(x) = \frac{h^{n+1}}{(n+1)!} t(t-1)(t-2)\dots(t-n) f^{(n+1)}(\xi), \quad (46)$$

$$|R_n(x)| \leq \frac{M_{n+1}h^{n+1}}{(n+1)!} |t(t-1)(t-2)\dots(t-n)|. \quad (47)$$

Analiziramo li formulu (44) vidimo da je $(n+1)!$ konstanta, nadalje je $f^{(n+1)}(\xi)$ određeno promatranom funkcijom f pa na te faktore ne možemo utjecati. Postavimo li pitanje da li možemo utjecati na produkt $(x-x_0)(x-x_1)\dots(x-x_n)$, onda u stvari pitamo da li izborom čvorišta možemo utjecati na pogreške. Naravno, želja je da ju minimiziramo. Uočimo dvije činjenice vezane za produkt $(x-x_0)(x-x_1)\dots(x-x_n)$. To je polinom $(n+1)$ -og stupnja u normiranom obliku (vodeći koeficijent je jedinica) i nadalje x_0, x_1, \dots, x_n su njegove jednostrukе nultočke. Proučavanjem tzv. Čebiševljevih polinoma u normiranom obliku dolazi se do ovakvog rezultata: *Među svim normiranim polinomima n -tog stupnja promatranih na segmentu $[-1, 1]$, normirani Čebiševljev polinom n -tog stupnja ima najmanju gornju ogragu za svoju absolutnu vrijednost.*

Taj nam rezultat kaže da za čvorove x_0, x_1, \dots, x_n treba uzeti nultočke Čebiševljevog polinoma $(n+1)$ -og stupnja. Gornja tvrdnja je izrečena za segment $[-1, 1]$. To ne predstavlja ograničenje, jer transformacija

$$x = \frac{2\xi - b - a}{b - a}$$

preslikava segment $[a, b]$ na $[-1, 1]$, dok njena inverzna

$$\xi = \frac{x(b-a) + (b+a)}{2}$$

preslikava segment $[-1, 1]$ na $[a, b]$. Čebiševljev polinom n -tog stupnja je definiran ovako

$$T_n(x) = \cos n\varphi, \quad \varphi = \arccos x. \quad (48)$$

Prema tome nama trebaju nultočke od T_{n+1} , tj. rješenja jednadžbe

$$\cos(n+1)\varphi = 0, \quad \varphi = \arccos x.$$

Riješimo li jednadžbu u φ dobivamo

$$(n+1)\varphi = \frac{\pi}{2} + k\pi, \quad k = 0, 1, 2, \dots, n,$$

odnosno,

$$\varphi_k = \frac{(2k+1)\pi}{2(n+1)}, \quad k = 0, 1, \dots, n,$$

što uvršteno u $\varphi = \arccos x$ i riješeno u varijabli x daje

$$x_k = \cos \frac{(2k+1)\pi}{2(n+1)}, \quad k = 0, 1, 2, \dots, n. \quad (49)$$

To su dakle nultočke od T_{n+1} i one se sve nalaze unutar segmenta $[-1, 1]$. Trebamo li čvorove unutar segmenta $[a, b]$ treba primijeniti gornju transformaciju po kojoj su

$$\xi_k = \frac{x_k(b-a) + (b+a)}{2}, \quad k = 0, 1, \dots, n \quad (50)$$

traženi čvorovi unutar $[a, b]$. Odredimo li interpolacijski polinom L_n uz tako odabrane čvorove (a često se čvorovi i inače mogu birati) dobit ćemo minimalnu pogrešku prema ocjeni pogreške (45). Istaknimo još jednom, ovo razmatranje je minimiziralo desnu stranu formule (45), dakle uz takvu ocjenu pogreške, možemo pogrešku minimizirati izabravši za čvorove interpolacijskog polinoma L_n nultočke Čebiševljevog polinoma T_{n+1} .

Ocjena pogreške od netočnih funkcijskih vrijednosti $f(x_i)$

Formula (45) daje ocjenu pogreške koju činimo ako funkcijsku vrijednost $f(x)$ zamijenimo vrijednošću interpolacijskog polinoma $L_n(x)$. U toj su formuli $f(x_i)$ točne vrijednosti od f u čvorovima x_i . Međutim, te vrijednosti nisu točne. Ocijenimo time uvedenu pogrešku. Opći oblik interpolacijskih formula je linearna kombinacija polinoma s koeficijentima koji su funkcijске vrijednosti u čvorovima tj. za dani x računamo sumu oblika

$$\sum_{i=0}^n f(x_i)P_i(x) \approx L_n(x) \quad (51)$$

koja aproksimira $L_n(x)$. Kako u realnom slučaju nemamo točne $f(x_i)$ to u lijevu stranu uvrštavamo približne vrijednosti $f^*(x_i) = f(x_i) + \eta_i$, tako da smo izračunali

$$\sum_{i=0}^n f^*(x_i)P_i(x) = \sum_{i=0}^n f(x_i)P_i(x) + \sum_{i=0}^n \eta_i P_i(x) = \sum_{i=0}^n f(x_i)P_i(x) + \varepsilon. \quad (52)$$

Dobili smo da je time uzrokovana pogreška jednaka

$$\varepsilon = \sum_{i=0}^n \eta_i P_i(x). \quad (53)$$

Ako znamo granice od η_i (npr. znamo točnost mjerjenja $f(x_i)$), onda iz (53) možemo ocijeniti gornju ogragu pogreške. Napose, ako je $|\eta_i| \leq \eta$, dobivamo ocjenu

$$|\varepsilon| \leq \eta \sum_{i=0}^n |P_i(x)|. \quad (54)$$

Uočimo da (54) daje ocjenu uz pretpostavku da točno računamo vrijednosti polinoma P_i .

U slučaju linearne interpolacije

$$\begin{aligned} L_1(x) &= f(x_0) \frac{x - x_1}{x_0 - x_1} + f(x_1) \frac{x - x_0}{x_1 - x_0} \\ &= f(x_0) \frac{x_1 - x}{x_1 - x_0} + f(x_1) \frac{x - x_0}{x_1 - x_0} \end{aligned}$$

imamo $P_0(x) = \frac{x_1 - x}{x_1 - x_0}$ i $P_1(x) = \frac{x - x_0}{x_1 - x_0}$, pa prema (54) za x između x_0 i x_1 , tj. $x_0 < x < x_1$ imamo

$$\begin{aligned} \eta(|P_0(x)| + |P_1(x)|) &= \eta\left(\left|\frac{x_1 - x}{x_1 - x_0}\right| + \left|\frac{x - x_0}{x_1 - x_0}\right|\right) \\ &= \eta \frac{|x_1 - x| + |x - x_0|}{|x_1 - x_0|} = \eta \left|\frac{x_1 - x_0}{x_1 - x_0}\right| = \eta. \end{aligned}$$

Dobiveni rezultat kaže da kod linearne interpolacije pogreška izazvana netočnošću $f(x_0)$ i $f(x_1)$ ne prelazi tu netočnost, tj. pogreška u računanju $L_1(x)$ uzrokovana pogrešnim funkcijskim vrijednostima $f(x_0)$ i $f(x_1)$ ne prelazi pogrešku tih funkcijskih vrijednosti.

Napomena.

- 1) Dobiveni rezultat ne možemo primijeniti na korake Aitkenove šeme, jer npr. ako x nije između x_0 i x_1 , onda to ne vrijedi za $L_{01}(x)$.
- 2) U gornjem razmatranju nisu uzete u obzir pogreške izračunavanja $P_i(x)$.

Neke daljnje primjene interpolacije. Inverzna interpolacija

Početni cilj je bio aproksimacija funkcije u točkama izvan čvorova. Međutim, aparat interpolacije može se primijeniti i na neke druge zadaće. Opisujemo ideju nekoliko takovih postupaka.

Ekstrem funkcije jedne varijable. Prepostavimo da treba odrediti ekstrem funkcije jedne varijable i točku u kojoj funkcija poprima ekstrem. U prvom koraku sastavimo tablicu funkcijskih vrijednosti s velikim korakom i pogledajmo da li možemo locirati položaj ekstrema.

Ako x_0 ne možemo locirati, usitnimo korak i razmotrimo finiju tablicu, itd. Naučno, to ne možemo provoditi bez kraja i konca. U nekom koraku možemo kroz tri ili četiri čvora oko lociranog položaja ekstrema naći kvadratni ili kubični interpolacijski polinom. Zatim deriviramo taj polinom i odredimo položaj njegovog ekstrema koji aproksimira položaj ekstrema funkcije, dok vrijednost polinoma u toj točki aproksimira ekstrem funkcije.

Jednadžba s jednom nepoznanicom. Pregledavanjem tablica kao gore lociramo položaj rješenja jednadžbe $f(x) = 0$, tj. položaj nultočke od f . Kroz okolna dva ili tri čvora odredimo interpolacijski polinom prvog, odnosno drugog stupnja. Za dovoljno bliža čvorišta nultočka interpolacijskog polinoma aproksimira nultočku funkcije pa time rješavamo linearu odnosno kvadratnu jednadžbu. To se može i iterirati tako da se dobivena nultočka uzme za novi čvor pri čemu se jedan stari čvor odbaci. To nas u slučaju linearne interpolacije vodi na metodu sekante, a u slučaju kvadratne interpolacije na tzv. Mullerovu metodu.

U slučaju invertibilne funkcije $y = f(x)$, tj. funkcije koja dopušta inverznu funkciju $x = g(y) = f^{-1}(y)$, rješavanje jednadžbe $f(x) = d$ svodi se na nalaženje

funkcijske vrijednosti inverzne funkcije, tj. rješenje je jednako $f^{-1}(d)$. Taj slučaj ćemo komentirati nakon opisa inverzne interpolacije.

Inverzna interpolacija. Neka funkcija $y = f(x)$ dopušta inverznu u okolini od $y = d$ (f ne mora imati inverznu na cijeloj domeni, sjetite se trigonometrijskih funkcija), tj. na nekoj okolini od d imamo funkciju $x = g(y) = f^{-1}(y)$ tako da za y iz te okoline vrijedi $f(g(y)) = y$. Pođemo li od čvorova $(x_i, y_i = f(x_i))$, možemo ih shvatiti kao čvorove od g , tj. $(y_i, g(y_i))$ i naći interpolacijski polinom od g . Time dobivamo polinom $L_n(y)$, za koji naravno vrijedi

$$L_n(y_i) = x_i, \quad i = 0, 1, \dots, n.$$

Opisani postupak nalaženja $L_n(y)$ zovemo **inverzna interpolacija**.

Vratimo li se na jednadžbu $f(x) = d$, onda je njeno rješenje jednako $x = c = g(d)$, čiju aproksimaciju dobivamo ako izračunamo $L_n(d)$, tj. dobili smo

$$c = g(d) = f^{-1}(d) \approx L_n(d).$$

Ovo napose može biti korisno ako u jednadžbi $f(x) = d$ variramo d .

Racionalna interpolacija

Pod racionalnom interpolacijom razumijevamo nalaženje racionalne funkcije

$$R(x) = \frac{P_p(x)}{Q_q(x)} \quad (55)$$

koja prolazi zadanim čvorovima $(x_i, f(x_i))$, $i = 0, 1, 2, \dots, n$, pri čemu su P_p i Q_q polinomi stupnja p , odnosno q . Neka je

$$\begin{aligned} P_p(x) &= a_p x^p + a_{p-1} x^{p-1} + \dots + a_1 x + a_0 \\ Q_q(x) &= b_q x^q + b_{q-1} x^{q-1} + \dots + b_1 x + b_0. \end{aligned} \quad (56)$$

Da nađemo $R(x)$ treba iz uvjeta $R(x_i) = f(x_i)$ odrediti koeficijente a_k i b_k kojih ima ukupno $(p+1) + (q+1) = p+q+2$. Kako se bronik i nazivnik mogu pomnožiti istom konstantom to jedan koeficijent možemo izabrati, (obično se uzima $b_0 = 1$ ili $\sum b_k^2 = 1$), pa smo time ostali sa $p+q+1$ nepoznatih koeficijenata. Prema tome da odredimo R trebamo toliko različitih čvorova koliko je i nepoznanica tj. $n+1 = p+q+1$, odnosno

$$n = p + q. \quad (57)$$

Opišimo praktični postupak rješavanja problema racionalne interpolacije. Pretpostavimo oblike (56) od P_p i Q_q i uvrstimo redom čvorove u (55)

$$R(x_i) = \frac{P_p(x_i)}{Q_q(x_i)} = f(x_i), \quad i = 0, 1, 2, \dots, n = p+q.$$

Nakon množenja sa $Q_q(x_i)$ dobivamo homogen sustav linearnih jednadžbi

$$P_p(x_i) - f(x_i)Q_q(x_i) = 0, \quad i = 0, 1, 2, \dots, n. \quad (58)$$

Taj sustav ima uvijek netrivijalnih rješenja u nepoznanice a_k , b_k jer ima jednu nepoznanicu više. Nadalje, Q_q ne može biti identički nula u rješenju problema racionalne interpolacije.

Sljedeći korak je da provjerimo da li je Q_q nula u nekom čvoru. Ako nije, onda smo našli rješenje problema (55), tj. našli smo racionalnu funkciju koja prolazi zadanim čvorovima. To rješenje je jedinstveno do na množenje brojnika i nazivnika istim faktorom.

Ako je $Q_q(x_i) = 0$ za svaki čvor, onda formula (58) povlači da je i $P_p(x_i) = 0$. Iz toga slijedi da polinomi P_p i Q_q imaju $(x - x_i)$ kao zajednički faktor, koji možemo skratiti. Gotovo je izvjesno da u tom čvoru ne vrijedi $R(x_i) = f(x_i)$, pa ako se to dogodi onda nemamo rješenje problema.

Ipak se možemo pitati što smo to dobili u drugom slučaju, jer nakon skraćivanja imamo i dalje racionalnu funkciju. Radi kraćeg pisanja neka je x_0 jedini čvor u kome je $Q_q(x_0) = 0$. Tada dobivena racionalna funkcija prolazi kroz čvorove x_1, x_2, \dots, x_n i predstavlja rješenje problema

$$R(x_i) = f(x_i), \quad i = 1, 2, \dots, n,$$

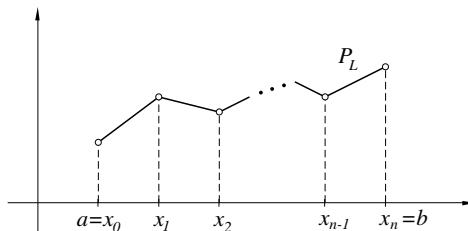
a ne polaznog. Naravno da se može dogoditi da je Q_p nula u više čvorova.

Na kraju spomenimo da za $q = 0$, tj. $Q_q = b_0 \neq 0$ dobivamo interpolacijski polinom kao poseban slučaj racionalne interpolacije.

1.2. Interpolacijski splajn

Promatrajmo funkciju $f : [a, b] \rightarrow \mathbf{R}$ i čvorove $a = x_0, x_1, \dots, x_n = b$. Na svakom segmentu $[x_{i-1}, x_i]$ možemo promatrati interpolacijski polinom 1. stupnja

$$L_{i1}(x) = \frac{x - x_i}{x_{i-1} - x_i} f(x_{i-1}) + \frac{x - x_{i-1}}{x_i - x_{i-1}} f(x_i), \quad i = 1, \dots, n-1, n. \quad (1)$$



Sl. 1.2.

Na slici 1.2 je skiciran graf Γ_L tako dobivene funkcije L , koja se još zove po dijelovima linearne funkcije. Istaknimo da je to neprekinuta funkcija, ali nije diferencijabilna. Time smo iz interpolacijskih polinoma 1. stupnja sastavili neprekinutu funkciju $L : [a, b] \rightarrow \mathbf{R}$, koja predstavlja izvjesnu aproksimaciju funkcije f . Uzmemimo li više čvorova dobit ćemo bolju aproksimaciju. Ta se ideja proširuje i na polinome višeg stupnja. Promatrajmo segmente $[x_0, x_1], \dots, [x_{n-1}, x_n]$ i označimo taj izbor čvorova sa Δ . Zamislimo da smo nad pojedinim segmentom $[x_{i-1}, x_i]$ našli polinom m -toga stupnja

$$P_{im}(x) = a_{i0} + a_{i1}x + \dots + a_{im}x^m, \quad x_{i-1} \leq x \leq x_i. \quad (2)$$

Od tih n polinoma zahtijevamo sljedeće: U početnom i krajnjem čvoru vrijedi $P_{1,m}(x_0) = f(x_0)$, odnosno $P_{n,m}(x_n) = f(x_n)$. U svakom nutarnjem čvoru x_i , $i = 1, 2, \dots, n - 1$, polinomi P_{im} i $P_{i+1,m}$, tj. polinomi nad susjednim segmentima imaju redom: jednake vrijednosti i to jednake funkcijskoj vrijednosti u čvoru, jednake vrijednosti prvih derivacija, jednake vrijednosti drugih derivacija, i tako dalje sve do derivacija $(m - 1)$ -og reda. Zapisano formulama to izgleda ovako:

$$\begin{aligned} P_{im}(x_{i-1}) &= f(x_{i-1}) & i = 1, 2, \dots, n; \\ P_{im}(x_i) &= f(x_i) & i = 1, 2, \dots, n; \\ P'_{im}(x_i) &= P'_{i+1,m}(x_i) & i = 1, 2, \dots, n - 1; \\ P''_{im}(x_i) &= P''_{i+1,m}(x_i) & i = 1, 2, \dots, n - 1; \\ &\vdots \\ P_{im}^{(m-1)}(x_i) &= P_{i+1,m}^{(m-1)}(x_i) & i = 1, 2, \dots, n - 1. \end{aligned} \quad (3)$$

Tako dobivena funkcija, označimo ju sa $S_{f,\Delta}^m$, zove se **interpolacijski splajn m -toga stupnja** (engl. *spline*). U oznaci $S_{f,\Delta}^m$ je istaknut red m , funkcija f i izbor čvorova Δ . Dakle, $S_{f,\Delta}^m$ je po dijelovima polinom m -toga stupnja čiji se dijelovi u unutarnjim čvorištima glatko nadovezuju. Jednadžbe (3) predstavljaju linearan sustav jednadžbi uz nepoznate koeficijente a_{ij} , $i = 1, 2, \dots, n$, $j = 0, 1, 2, \dots, m$. Sustav (3) ima ukupno $n(m + 1)$ nepoznanica i $2n + (m - 1)(n - 1) = n(m + 1) - (m - 1)$ jednadžbi.

Sada vidimo da je za $m = 1$ funkcija L sa slike ništa drugo nego interpolacijski splajn prvog stupnja, tj. vrijedi $L = S_{f,\Delta}^1$. Sustav jednadžbi (3) poprima u tom slučaju oblik

$$\begin{aligned} P_{i1}(x_{i-1}) &= a_{i0} + a_{i1}x_{i-1} = f(x_{i-1}) \\ P_{i1}(x_i) &= a_{i0} + a_{i1}x_i = f(x_i). \end{aligned} \quad (4)$$

To je sustav od $2n$ jednadžbi s $2n$ nepoznanica. Riješimo li taj sustav dobivamo

$$\begin{aligned} a_{i1} &= \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}, & i = 1, 2, \dots, n \\ a_{i0} &= f(x_{i-1}) - a_{i1}x_{i-1}. \end{aligned} \quad (5)$$

Sredimo li (1) po potencijama od x , dobit ćemo formule (5), s time da je u drugoj formuli provedena supstitucija već poznatog koeficijenta a_{i1} . Spomenimo još da se interpolacijski splajn prvog stupnja koristi u trapeznoj formuli za aproksimaciju integrala, a koristi se i u drugim prilikama.

Ovdje ćemo još izvesti formule za koeficijente interpolacijskog splajna trećeg stupnja ili kraće **kubnog splajna** ne ulazeći u razlog koji posebno ističe važnost trećeg stupnja. Tražimo dakle $S_{f,\Delta}^3$. Polazimo od

$$P_{i3}(x) = a_{i0} + a_{i1}x + a_{i2}x^2 + a_{i3}x^3, \quad x_{i-1} \leq x \leq x_i, \quad i = 1, 2, \dots, n. \quad (6)$$

Kako tražimo n polinoma trećeg stupnja vidimo da treba odrediti $4n$ nepoznatih koeficijenata a_{ij} , $i = 1, 2, \dots, n$, $j = 0, 1, 2, 3$. Jednadžbe (3) sada glase

$$\begin{aligned} P_{i3}(x_{i-1}) &= f(x_{i-1}), & i = 1, 2, \dots, n; \\ P_{i3}(x_i) &= f(x_i), & i = 1, 2, \dots, n; \\ P'_{i3}(x_i) &= P'_{i+1,3}(x_i), & i = 1, 2, \dots, n - 1; \\ P''_{i3}(x_i) &= P''_{i+1,3}(x_i), & i = 1, 2, \dots, n - 1. \end{aligned} \quad (7)$$

To su ukupno $4n - 2$ jednadžbe s $4n$ nepoznanica a_{ij} .

Prema tome, linearnom sustavu (7) treba dodati još dvije jednadžbe. Njih dobivamo iz uvjeta koje zahtijevamo da vrijede na rubu. Uobičajeno ti uvjeti glase

$$\begin{aligned} P''_{13}(x_0) &= 0 \\ P''_{n3}(x_n) &= 0 \end{aligned} \quad (8)$$

(i vode na tridiagonalni sustav) tj. na krajevima segmenta zahtijevamo da je druga derivacija splajna $S_{f,\Delta}^3$ jednaka nuli. Da bi riješili sustav (7)–(8) uočimo sljedeće. Druga derivacija polinoma trećeg stupnja je polinom prvog stupnja. Iz $P''_{i3}(x_i) = P''_{i+1,3}(x_i)$ slijedi da je druga derivacija od $S_{f,\Delta}^3$ neprekidna po dijelovima linearna funkcija koju možemo shvatiti kao splajn prvog stupnja. Primijenimo (1) na P''_{i3} . To daje

$$P''_{i3}(x) = \frac{x - x_i}{x_{i-1} - x_i} P''_{i3}(x_{i-1}) + \frac{x - x_{i-1}}{x_i - x_{i-1}} P''_{i3}(x_i) \quad \text{na } [x_{i-1}, x_i]. \quad (9)$$

Označimo li $h_i = x_i - x_{i-1}$, (9) prelazi u

$$P''_{i3}(x) = \frac{x_i - x}{h_i} P''_{i3}(x_{i-1}) + \frac{x - x_{i-1}}{h_i} P''_{i3}(x_i) \quad \text{na } [x_{i-1}, x_i]. \quad (10)$$

Uvedimo kraće oznake

$$P''_{i3}(x_i) = M_i, \quad i = 1, 2, \dots, n$$

tako da (10) poprima oblik

$$P''_{i3}(x) = \frac{x_i - x}{h_i} M_{i-1} + \frac{x - x_{i-1}}{h_i} M_i \quad \text{na } [x_{i-1}, x_i]. \quad (11)$$

Dvostrukom integracijom dobivamo

$$P_{i3}(x) = \frac{(x_i - x)^3}{6h_i} M_{i-1} + \frac{(x - x_{i-1})^3}{6h_i} M_i + C_i x + D_i \quad (12)$$

gdje su C_i i D_i konstante integracije koje treba odrediti. Određujemo ih iz zadanih vrijednosti na rubovima, tj. iz $P_{i3}(x_{i-1}) = f(x_{i-1})$ i $P_{i3}(x_i) = f(x_i)$. Imamo dakle

$$f(x_{i-1}) = \frac{(x_i - x_{i-1})^3}{6h_i} M_{i-1} + C_i x_{i-1} + D_i$$

$$f(x_i) = \frac{(x_i - x_{i-1})^3}{6h_i} M_i + C_i x_i + D_i$$

odnosno

$$\begin{aligned} C_i x_{i-1} + D_i &= f(x_{i-1}) - \frac{h_i^2}{6} M_{i-1} \\ C_i x_i + D_i &= f(x_i) - \frac{h_i^2}{6} M_i. \end{aligned} \quad (13)$$

Oduzmemmo li prvu jednadžbu od druge dobivamo

$$C_i(x_i - x_{i-1}) = f(x_i) - f(x_{i-1}) + \frac{h_i^2}{6}(M_{i-1} - M_i)$$

ili

$$C_i = \frac{1}{h_i} \left\{ f(x_i) - f(x_{i-1}) + \frac{h_i^2}{6}(M_{i-1} - M_i) \right\}.$$

Sada iz druge jednadžbe (13) imamo

$$D_i = f(x_i) - \frac{h_i^2}{6}M_i - \frac{x_i}{h_i} \left\{ f(x_i) - f(x_{i-1}) + \frac{h_i^2}{6}(M_{i-1} - M_i) \right\}.$$

Izračunate vrijednosti od C_i i D_i uvrstimo sada u (12) tako da dobivamo

$$\begin{aligned} P_{i3}(x) &= \frac{(x_i - x)^3}{6h_i}M_{i-1} + \frac{(x - x_{i-1})^3}{6h_i}M_i + \frac{x}{h_i} \left\{ f(x_i) - f(x_{i-1}) + \frac{h_i^2}{6}(M_{i-1} - M_i) \right\} \\ &\quad + f(x_i) - \frac{h_i^2}{6}M_i - \frac{x_i}{h_i} \left\{ f(x_i) - f(x_{i-1}) + \frac{h_i^2}{6}(M_{i-1} - M_i) \right\} \\ &= \frac{(x_i - x)^3}{6h_i}M_{i-1} + \frac{(x - x_{i-1})^3}{6h_i}M_i + \frac{x_i - x}{h_i} \left\{ f(x_{i-1}) - \frac{M_{i-1}h_i^2}{6} \right\} \\ &\quad + \frac{x}{h_i} \left[f(x_i) - \frac{h_i^2}{6}M_i \right] + f(x_i) - \frac{h_i^2}{6}M_i - \frac{x_{i-1} + h_i}{h_i} \left[f(x_i) - \frac{h_i^2}{6}M_i \right] \\ &= \frac{(x_i - x)^3}{6h_i}M_{i-1} + \frac{(x - x_{i-1})^3}{6h_i}M_i + \frac{x_i - x}{h_i} \left\{ f(x_{i-1}) - \frac{M_{i-1}h_i^2}{6} \right\} \\ &\quad + \frac{x}{h_i} \left[f(x_i) - \frac{h_i^2}{6}M_i \right] + f(x_i) - \frac{h_i^2}{6}M_i - \frac{x_{i-1}}{h_i} \left[f(x_i) - \frac{h_i^2}{6}M_i \right] - f(x_i) + \frac{h_i^2}{6}M_i, \end{aligned}$$

što nakon sređivanja glasi

$$\begin{aligned} P_{i3}(x) &= \frac{(x_i - x)^3}{6h_i}M_{i-1} + \frac{(x - x_{i-1})^3}{6h_i}M_i \\ &\quad + \frac{x_i - x}{h_i} \left\{ f(x_{i-1}) - \frac{M_{i-1}h_i^2}{6} \right\} + \frac{x - x_{i-1}}{h_i} \left\{ f(x_i) - \frac{M_ih_i^2}{6} \right\} \end{aligned} \tag{14}$$

na $[x_{i-1}, x_i]$, $i = 1, 2, 3, \dots, n$. (14) predstavlja traženi splajn 3. stupnja $S_{f,\Delta}^3$ u kome su još ostali nepoznati parametri M_i , $i = 0, 1, 2, \dots, n$. Prisjetimo se da M_i predstavlja vrijednost druge derivacije splajna u čvoru x_i . Kako do sada nismo koristili uvjet na prve derivacije to ćemo ga sada upotrijebiti. Ti uvjeti glase

$$P'_{i3}(x_i) = P'_{i+1,3}(x_i) \quad i = 1, 2, \dots, n-1.$$

To predstavlja $n-1$ jednadžbu s $n+1$ nepoznanica M_i , pa su nam potrebna dva dodatna uvjeta. To su uvjeti (8) koji određuju M_0 i M_n , tj.

$$P''_{13}(x_0) = M_0 = 0, \quad P''_{n3}(x_n) = M_n = 0.$$

Deriviranjem (14) dobivamo

$$\begin{aligned} P'_{i,3}(x) &= -\frac{(x_i - x)^2}{2h_i}M_{i-1} + \frac{(x - x_{i-1})^2}{2h_i}M_i + \frac{1}{h_i} \left[f(x_i) - f(x_{i-1}) + \frac{h_i^2}{6}(M_{i-1} - M_i) \right], \\ P'_{i+1,3}(x) &= -\frac{(x_{i+1} - x)^2}{2h_{i+1}}M_i + \frac{(x - x_i)^2}{2h_{i+1}}M_{i+1} + \frac{1}{h_{i+1}} \left[f(x_{i+1}) - f(x_i) + \frac{h_{i+1}^2}{6}(M_i - M_{i+1}) \right], \end{aligned}$$

što za $x = x_i$ daje jednakost

$$\begin{aligned} \frac{h_i^2}{2h_i}M_i + \frac{1}{h_i} \left[f(x_i) - f(x_{i-1}) + \frac{h_i^2}{6}(M_{i-1} - M_i) \right] \\ = -\frac{h_{i+1}^2}{2h_{i+1}}M_i + \frac{1}{h_{i+1}} \left[f(x_{i+1}) - f(x_i) + \frac{h_{i+1}^2}{6}(M_i - M_{i+1}) \right]. \end{aligned}$$

Sredimo li tu jednakost po nepoznanicama M_i imamo

$$\frac{h_i}{6}M_{i-1} + \frac{h_i + h_{i+1}}{3}M_i + \frac{h_{i+1}}{6}M_{i+1} = \frac{f(x_{i+1}) - f(x_i)}{h_{i+1}} - \frac{f(x_i) - f(x_{i-1})}{h_i}.$$

Time smo rješavanje sustava (7), koji je imao $4n$ nepoznanica, sveli na sustav od $n-1$ jednadžbe s $n-1$ nepoznanicom M_1, M_2, \dots, M_{n-1} koji glasi

$$\frac{h_i}{6}M_{i-1} + \frac{h_i + h_{i+1}}{3}M_i + \frac{h_{i+1}}{6}M_{i+1} = \frac{f(x_{i+1}) - f(x_i)}{h_{i+1}} - \frac{f(x_i) - f(x_{i-1})}{h_i}, \quad (15)$$

$$i = 1, 2, 3, \dots, n-1.$$

U matričnom zapisu

$$CM = d$$

je

$$M = \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_{n-1} \end{bmatrix}, \quad d = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_{n-1} \end{bmatrix}$$

pri čemu je $d_i = \frac{f(x_{i+1}) - f(x_i)}{h_{i+1}} - \frac{f(x_i) - f(x_{i-1})}{h_i}$ i

$$C = [c_{ij}] = \left[\begin{array}{ccccccc} \frac{h_1+h_2}{3} & \frac{h_2}{6} & 0 & \dots & \dots & \dots & 0 \\ \frac{h_2}{6} & \frac{h_2+h_3}{3} & \frac{h_3}{6} & 0 & \dots & \dots & 0 \\ 0 & \frac{h_3}{6} & \frac{h_3+h_4}{3} & \frac{h_4}{6} & 0 & \dots & 0 \\ & & & \ddots & \ddots & \ddots & \\ 0 & \dots & \dots & 0 & \frac{h_{n-2}}{6} & \frac{h_{n-2}+h_{n-1}}{3} & \frac{h_{n-1}}{6} \\ 0 & \dots & \dots & \dots & 0 & \frac{h_{n-1}}{6} & \frac{h_{n-1}+h_n}{3} \end{array} \right]$$

je tridiagonalna matrica (izgled matrice C sugerira naziv).

Elemente C možemo ovako zapisati

$$c_{ij} = \begin{cases} \frac{h_i}{6} & \text{za } j = i-1, \\ \frac{h_i + h_{i+1}}{3} & \text{za } j = i, \\ \frac{h_{i+1}}{6} & \text{za } j = i+1, \\ 0 & \text{za } |i-j| > 1. \end{cases}$$

Rješavanjem sustava (15) dobivamo tražene parametre M_i koje zatim uvrstimo u (14). Izraz u (14) predstavlja traženi splajn 3. stupnja na segmentu $[x_{i-1}, x_i]$, $i = 1, 2, 3, \dots, n$. Postoje i sažeti zapisi kubnog splajna što nije od posebne koristi.

1.3. Trigonometrijska interpolacija

Trigonometrijski i fazni interpolacijski polinom

Trigonometrijska interpolacija u svom najvažnijem obliku koristi linearne kombinacije funkcija $\cos mx$ i $\sin mx$ gdje je m prirodan broj, tj. izraz koji ima jedan od ovih oblika

$$\psi(x) = \frac{A_0}{2} + \sum_{m=1}^M (A_m \cos mx + B_m \sin mx), \quad (1)$$

$$\psi(x) = \frac{A_0}{2} + \sum_{m=1}^{M-1} (A_m \cos mx + B_m \sin mx) + \frac{A_M}{2} \cos Mx, \quad (2)$$

pri čemu se (1) koristi u slučaju kada imamo **neparan** broj čvorova $N = 2M + 1$, dok se (2) koristi kada imamo **paran** broj čvorova $N = 2M$. Označimo li sa (x_k, y_k) , $y_k = f(x_k)$ čvorove, onda se koeficijenti A_m , B_m određuju iz uvjeta

$$\psi(x_k) = y_k, \quad k = 0, 1, 2, \dots, N - 1 \quad (3)$$

a ψ se zove **trigonometrijski interpolacijski polinom**. Interpolacija trigonometrijskim polinomom prikladna je za periodičke funkcije poznatog perioda. Istaknimo da je ψ periodička funkcija od x s periodom 2π . Transformacijom $\xi = \frac{b-a}{2\pi}x + a$ preslikavamo $[0, 2\pi]$ na $[a, b]$, stoga se opća razmatranja vrše na segmentu $[0, 2\pi]$. Naravno da nas formule (1) i (2) podsjećaju na parcijalne sume Fourierovog reda.

Sada se postavlja problem iznalaženja koeficijenata A_m i B_m . Očigledno je sa (3) dan linearan sustav jednadžbi u nepoznate koeficijente A_m i B_m čije rješenje tražimo. Da dođemo do formula za koeficijente A_m , B_m činimo izvjesna pojednostavnjenja. Osnovno pojednostavnenje je da razmatramo jednoliku raspodjelu čvorova i da pređemo na eksponencijalni zapis trigonometrijskih funkcija služeći se De Moivreovom formulom $e^{ikx} = \cos kx + i \sin kx$, otkuda kao što znamo, imamo

$$\begin{aligned} \cos kx &= \frac{e^{ikx} + e^{-ikx}}{2}, \\ \sin kx &= \frac{e^{ikx} - e^{-ikx}}{2i}. \end{aligned}$$

Kod jednolike (ekvidistantne) raspodjele čvorova dijelimo segment $[0, 2\pi]$ na N jednakih dijelova pomoću kojih određujemo nepoznate koeficijente. Sada su nam čvorovi dani sa

$$x_k = \frac{2\pi k}{N}, \quad k = 0, 1, 2, \dots, N - 1. \quad (4)$$

Zbog periodičnosti, točka $x = 2\pi$, koja predstavlja $(N+1)$ -vi čvor, ne predstavlja novi uvjet, pa je obično ne spominjemo. Sada nam jednolika raspodjela čvorova omogućava da problem nalaženja koeficijenata svedemo na traženje koeficijenata tzv. **faznog interpolacijskog polinoma** koji ima ovaj oblik

$$P(x) = \beta_0 + \beta_1 e^{ix} + \dots + \beta_{N-1} e^{(N-1)ix} \quad (5)$$

pri čemu su β_j kompleksni koeficijenti, koji se određuju iz uobičajenog uvjeta interpolacije

$$P(x_k) = y_k, \quad k = 0, 1, 2, \dots, N - 1. \quad (6)$$

U tu svrhu provedimo slijedeći račun. Svakako vrijedi

$$\begin{aligned} e^{-mix_k} &= e^{-\frac{2\pi imk}{N}} = e^{-\frac{2\pi imk}{N}} \cdot e^{\frac{2\pi iNk}{N}} \\ &= e^{\frac{2\pi i(N-m)k}{N}} = e^{(N-m)ix_k}. \end{aligned}$$

To nam omogućava da $\cos mx_k$ i $\sin mx_k$ zapišemo u obliku

$$\cos mx_k = \frac{e^{imx_k} + e^{i(N-m)x_k}}{2}, \quad \sin mx_k = \frac{e^{imx_k} - e^{i(N-m)x_k}}{2i}.$$

Uvrstimo li to u ψ i dobiveni izraz sredimo po potencijama od e^{ix_k} , dobivamo fazni polinom P s koeficijentima β_j , $j = 1, 2, \dots, N - 1$. Time su u slučaju ekvidistantnog izbora čvorova koeficijenti faznog polinoma povezani s koeficijentima trigonometrijskog polinoma na sljedeći način:

a) Ako je N neparan, tj. $N = 2M + 1$, onda vrijedi:

$$\beta_0 = \frac{A_0}{2}, \quad \beta_j = \frac{1}{2}(A_j - iB_j), \quad \beta_{N-j} = \frac{1}{2}(A_j + iB_j), \quad j = 1, 2, \dots, M; \quad (7)$$

odnosno,

$$A_0 = 2\beta_0, \quad A_m = \beta_m + \beta_{N-m}, \quad B_m = i(\beta_m - \beta_{N-m}), \quad m = 1, 2, \dots, M.$$

b) Ako je N paran, tj. $N = 2M$, onda vrijedi:

$$\begin{aligned} \beta_0 &= \frac{A_0}{2}, \\ \beta_j &= \frac{1}{2}(A_j - iB_j), \\ \beta_{N-j} &= \frac{1}{2}(A_j + iB_j), \quad j = 1, 2, \dots, M - 1, \\ \beta_M &= \frac{A_M}{2}; \end{aligned} \quad (8)$$

odnosno,

$$\begin{aligned} A_0 &= 2\beta_0, \\ A_m &= \beta_m + \beta_{N-m}, \\ B_m &= i(\beta_m - \beta_{N-m}), \quad m = 1, 2, \dots, M - 1, \\ A_M &= 2\beta_m. \end{aligned}$$

Izvedimo formule (7). Uvrštenjem u ψ imamo

$$\psi(x) = \frac{A_0}{2} + \sum_{m=1}^M \left(A_m \frac{e^{imx_k} + e^{i(N-m)x_k}}{2} + B_m \frac{e^{imx_k} - e^{i(N-m)x_k}}{2i} \right).$$

Kako m varira između 1 i M to $N - m$ varira između $M + 1$ i $2M = N - 1$, tako da imamo zastupljene sve potencije od e^{ix_k} za $1 \leq j \leq N - 1$. Vidimo da prvi pribrojnici razlomaka daju potencije od e^{ix_k} između 1 i M , pa ako ih zbrojimo imamo

$$A_m \frac{e^{imx_k}}{2} + B_m \frac{e^{imx_k}}{2i} = \frac{1}{2}(A_m - iB_m)e^{imx_k},$$

dakle za $1 \leq j \leq M$ imamo

$$\beta_j = \frac{1}{2}(A_j - iB_j), \quad j = 1, 2, \dots, M.$$

Drugi pribrojnici razlomaka imaju potencije između $M + 1$ i $N - 1$, ako ih zbrojimo imamo

$$A_m \frac{e^{i(N-m)x_k}}{2} - B_m \frac{e^{i(N-m)x_k}}{2} = \frac{1}{2}(A_m + iB_m)e^{i(N-m)x_k},$$

dakle za $1 \leq j \leq M$ imamo $M + 1 \leq N - j \leq 2M = N - 1$, odnosno

$$\beta_{N-j} = \frac{1}{2}(A_j + iB_j), \quad j = 1, 2, \dots, M.$$

Očigledno je $\beta_0 = \frac{A_0}{2}$. Formule koje daju A_m i B_m pomoću koeficijenata β_j slijede direktno iz gornjih, zbrajanjem, odnosno oduzimanjem odgovarajućih jednakosti.

Istaknimo odmah da i uz jednoliku raspodjelu čvorova, trigonometrijski interpolacijski polinom ψ i njemu odgovarajući fazni polinom P poprimaju u čvorovima iste vrijednosti tj. $\psi(x_k) = P(x_k) = y_k$, međutim, općenito je $\psi(x) \neq P(x)$ u točkama $x \neq x_k$, tj. izvan čvorova.

Fazni polinom je po svojoj građi jednostavniji od trigonometrijskog, pa se mnoga zaključivanja lakše provode na nivou faznog polinoma. Da to potkrijepimo provedimo npr. sljedeću supstituciju $\omega = e^{ix}$, $\omega_k = e^{ix_k}$.

Tada P poprima sljedeći oblik.

$$P(x) = p(\omega) = \beta_0 + \beta_1\omega + \dots + \beta_{N-1}\omega^{N-1},$$

što je običan polinom u ω . Primjetimo sljedeće: kod ekvidistantne podjele $j \neq k$ povlači da je $\omega_j \neq \omega_k$, $0 \leq j, k \leq N - 1$, što opet povlači da imamo N različitih čvorova u varijabli ω i nadalje, u svakom čvoru vrijedi

$$P(x_k) = p(\omega_k) = y_k, \quad k = 0, 1, 2, \dots, N - 1.$$

Prema tome (sada je varijabla kompleksna), p je običajeni interpolacijski polinom. Iz toga odmah zaključujemo da je kod ekvidistantnog izbora čvorova fazni interpolacijski polinom P jedinstven. Preko rezultata (7) taj se rezultat prenosi i na trigonometrijski interpolacijski polinom ψ , dakle ψ je jedinstven za ekvidistantne čvorove.

Teorem 1.1. Za jednoliku podjelu segmenta $[0, 2\pi]$ na N dijelova, koeficijenti faznog interpolacijskog polinoma (5) dani su formulom

$$\beta_j = \frac{1}{N} \sum_{k=0}^{N-1} y_k e^{-\frac{2\pi k}{N} i}, \quad (9)$$

gdje je $x_k = \frac{2\pi k}{N}$, $k = 0, 1, 2, \dots, N - 1$.

Da dokazemo formule (9) postupimo ovako. Za $\omega_k = e^{ix_k} = e^{\frac{2\pi k}{N} i}$ očigledno vrijedi:

$$(\omega_k)^j = (\omega_j)^k, \quad (\omega_k)^{-j} = \overline{(\omega_k)^j}, \quad 0 \leq j, \quad k \leq N - 1,$$

gdje $\bar{\omega}$ označava konjugirano kompleksnu vrijednost. Pokažimo prvo da vrijedi sljedeća jednakost

$$\sum_{k=0}^{N-1} (\omega_k)^j (\bar{\omega}_k)^{-m} = N \delta_{mj} = \begin{cases} N & \text{za } j = m \\ 0 & \text{za } j \neq m \end{cases} \quad (10)$$

kojom ćemo se kasnije poslužiti.

Da provjerimo (10) primijetimo da su $\omega_k = e^{\frac{2\pi i}{N} k}$ ništa drugo nego N -ti korijeni jedinice, tj. rješenja jednadžbe $z^N - 1 = 0$. Uočimo li ω_{j-m} i shvatimo ga kao nultočku polinoma $z^N - 1$, onda iz faktorizacija

$$z^N - 1 = (z - 1) \sum_{k=0}^{N-1} z^k$$

slijedi da je

$$\omega_{j-m} - 1 = 0,$$

ako je ω_{j-m} nultočka prvog faktora, odnosno da je

$$\sum_{k=0}^{N-1} (\omega_{j-m})^k = 0$$

ako je ω_{j-m} nultočka drugog faktora. U prvom slučaju imamo $\omega_{j-m} = 1$, ili $e^{\frac{2\pi i(j-m)}{N}} = 1$, što daje $j - m = 0$, odnosno, $m = j$. U tom se slučaju suma (10) sastoji od N pribrojnika jednakih 1, pa je time slučaj $m = j$ formule (10) provjerен. U drugom slučaju imamo

$$\sum_{k=0}^{N-1} (\omega_k)^j \cdot (\bar{\omega}_k)^{-m} = \sum_{k=0}^{N-1} (\omega_k)^{j-m} = \sum_{k=0}^{N-1} (\omega_{j-m})^k = 0$$

čime je (10) u potpunosti provjereno.

Vratimo se sada dokazu formule (9). U tom cilju promatrajmo N -komponentne vektore s kompleksnim brojevima kao skalarima.

Vektori su ortogonalni ako je njihov skalarni produkt nula. Stoga se prisjetimo definicije skalarmog produkta vektora s kompleksnim komponentama. Ako su

$$\omega_1 = \begin{bmatrix} z_{11} \\ \vdots \\ z_{1n} \end{bmatrix} \quad \text{i} \quad \omega_2 = \begin{bmatrix} z_{21} \\ \vdots \\ z_{2n} \end{bmatrix}$$

dva vektora, onda je skalarni produkt (ω_1, ω_2) definiran kao

$$(\omega_1, \omega_2) = \sum_{k=1}^n z_{1k} \cdot \overline{z_{2k}}.$$

Promatrajmo sada sljedećih N vektora s kompleksnim komponentama

$$\omega_m = \begin{bmatrix} 1 \\ (\omega_1)^m \\ (\omega_2)^m \\ \vdots \\ (\omega_{N-1})^m \end{bmatrix}, \quad m = 0, 1, 2, \dots, N-1.$$

Vidimo sada da suma u formuli (10) predstavlja skalarni produkt (ω_j, ω_m) i kaže se da su vektori ω_j i ω_m međusobno ortogonalni (tu smo zatrebali uočenu činjenicu $(\omega_k)^{-j} = \overline{(\omega_k)^j}$). Pri tome spomenimo da ti vektori nisu jedinični jer prema (10) vrijedi $(\omega_m, \omega_m) = N$. Ortogonalnost koristimo sada ovako. Uvrstimo li čvor $x_k = \frac{2\pi k}{N}$, $k = 0, 1, \dots, N-1$ u fazni interpolacijski polinom dobivamo linearни sustav jednadžbi

$$\begin{aligned} P(x_k) &= y_k = \beta_0 + \beta_1 e^{ix_k} + \beta_2 e^{2ix_k} + \dots + \beta_{N-1} e^{(N-1)ix_k} \\ &= \beta_0 + \beta_1 \omega_k + \beta_2 (\omega_k)^2 + \dots + \beta_{N-1} (\omega_k)^{N-1}, \quad k = 0, 1, 2, \dots, N-1, \end{aligned}$$

u nepoznate koeficijente β_j . Taj sustav u vektorskom zapisu izgleda ovako

$$y = \beta_0 \omega_0 + \beta_1 \omega_1 + \beta_2 \omega_2 + \dots + \beta_{N-1} \omega_{N-1} \quad (11)$$

gdje je s y označen vektor

$$y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix}.$$

Pomnožimo li (11) skalarno s ω_j dobit ćemo na osnovu gore rečenog

$$(y, \omega_j) = \beta_j (\omega_j, \omega_j) = \beta_j N$$

ili

$$\beta_j N = \sum_{k=0}^{N-1} y_k \overline{(\omega_k)^j} = \sum_{k=0}^{N-1} y_k (\omega_k)^{-j} = \sum_{k=0}^{N-1} y_k e^{-\frac{-2\pi j k}{N}}$$

pa dijeljenje s N daje formulu (9). Time smo dokazali formule (9) koje daju koeficijente faznog interpolacijskog polinoma za ekvidistantnu podjelu.

Vratimo se sada polaznom trigonometrijskom polinomu (1) i formulama (7) i (8), iz kojih dobivamo formule za A_m i B_m . Uočimo li da je $\beta_0 = \beta_N = \frac{1}{N} \sum_{k=0}^{N-1} y_k$ možemo prema (7) i (8) pisati

$$A_0 = 2\beta_0 = \beta_0 + \beta_N.$$

Sada za $m = 0, 1, 2, \dots, M-1$ imamo

$$\begin{aligned} A_m &= \beta_m + \beta_{N-m} = \frac{1}{N} \sum_{k=0}^{N-1} y_k \left(e^{-mix_k} + e^{-(N-m)ix_k} \right) \\ &= \frac{2}{N} \sum_{k=0}^{N-1} y_k \cos mx_k = \frac{2}{N} \sum_{k=0}^{N-1} y_k \cos \frac{2\pi m k}{N}, \end{aligned} \quad (12)$$

dok za paran N , $N = 2M$, vrijedi $M = N - M$, pa možemo pisati

$$A_M = 2\beta_M = \beta_M + \beta_{N-M}. \quad (13)$$

Za koeficijente B_m , neovisno o paritetu od N , imamo

$$\begin{aligned} B_m &= i(\beta_m - \beta_{N-m}) = \frac{i}{N} \sum_{k=0}^{N-1} y_k (e^{-mik} - e^{(N-m)ik}) \\ &= \frac{2}{N} \sum_{k=0}^{N-1} y_k \sin mx_k = \frac{2}{N} \sum_{k=0}^{N-1} y_k \sin \frac{2\pi mk}{N}. \end{aligned} \quad (14)$$

Time smo u stvari dokazali ovaj teorem.

Teorem 1.2. Za jednoliku podjelu segmenta $[0, 2\pi]$ na N dijelova, koeficijenti trigonometrijskog interpolacijskog polinoma (1) dani su formulama (12) i (14)

Skrenimo pažnju na to da su sva razmatranja vršena na segmentu $[0, 2\pi]$. Za opći segment $[a, b]$ potrebno je izvršiti transformaciju

$$x = \frac{2\pi}{b-a}(\xi - a),$$

koja preslikava segment $[a, b]$ na segment $[0, 2\pi]$, dok inverzna transformacija

$$\xi = \frac{b-a}{2\pi}x + a$$

preslikava segment $[0, 2\pi]$ na $[a, b]$. To naravno daje jednoliku podjelu na $[a, b]$ i vrijednosti čvorova su

$$\xi_k = a + \frac{b-a}{N}k, \quad k = 0, 1, 2, \dots, N-1.$$

Brzi Fourierovi transformati

U prethodnom smo paragrafu promatrali fazni interpolacijski polinom $P(x) = \sum_{j=0}^{N-1} \beta_j e^{i j x}$, te za ekvidistantna čvorista (x_k, y_k) , $x_k = \frac{2\pi k}{N}$, $k = 0, 1, 2, \dots, N-1$ dobili formulu za njegove koeficijente β_j koja glasi

$$\beta_j = \frac{1}{N} \sum_{k=0}^{N-1} y_k e^{-\frac{2\pi j k}{N}}, \quad j = 0, 1, \dots, N-1. \quad (15)$$

Izračunavanje takvih izraza pojavljuje se u Fourierovoj analizi, pa je zbog toga od prvenstvene važnosti. Takvi se izrazi pojavljuju i kod diskretnе aproksimacije slike po Fourierovoj transformaciji (Fourierovog transformata) koja se definira kao

$$F(s) = \int_{-\infty}^{\infty} f(t) e^{-2\pi i s t} dt \quad (16)$$

s time, da F računamo u N ekvidistantnih točaka varijable s . Međutim, kod numeričkog izračunavanja izraza oblika (15) poznavali smo do nedavno postupke koji zahtjevaju mnogo množenja. Točnije, red množenja je bio N^2 , što za velike N , kakve zahtjeva diskretna aproksimacija integrala (16), predstavlja poteškoću. J.W. Cooley i

J.W. Tukey otkrili su 1965.g. postupak za brzo izračunavanje izraza (15), ali za posebne vrijednosti od N . Kod njihovog postupka je red množenja za takve N jednak $N \log N$. Ta metoda kao i njene varijante dobila je ime **brzi Fourierovi transformati** (engl. *Fast Fourier Transforms*), jer je njome ubrzano računanje.

Zadržat ćemo se na **Cooley-Tukeyevoj metodi**. Metoda se bazira na cjelobrojnoj faktorizaciji broja N čime se postiže raščlanjivanje polaznog problema na potprobleme nižeg stupnja. To se raščlanjivanje provodi rekursivno i najlakše funkcioniра kada je N oblika $N = 2^n$ i $n > 0$ prirodan broj. To je ujedno i najvažniji i najjednostavniji slučaj. Ovdje se ograničavamo samo na taj slučaj, iako analogne tehnike funkcioniраju i u općenitijem slučaju, tj. kada je $N = N_1 \cdot N_2 \cdot \dots \cdot N_n$, gdje su N_k , $k = 1, 2, \dots, n$ prirodni brojevi.

Prikažimo Cooley-Tukeyev pristup pomoću faznih interpolacijskih polinoma. Neka je $N = 2M$. Promatrajmo dva fazna interpolacijska polinoma $(M - 1)$ -og stupnja $Q(x)$ i $R(x)$ određena sa

$$Q(x_{2m}) = y_{2m}, \quad R(x_{2m}) = y_{2m+1}, \quad m = 0, 1, \dots, M - 1.$$

Vidimo da i Q i R prolaze kroz čvorove parnog indeksa, međutim, translacijom za $\frac{2\pi}{N}$ iz R dobivamo fazni interpolacijski polinom $\tilde{R}(x) = R\left(x - \frac{2\pi}{N}\right) = R\left(x - \frac{\pi}{M}\right)$, koji prolazi kroz čvorove neparnog indeksa. Zbog

$$e^{Mx_k i} = e^{\frac{2\pi M k}{N} i} = e^{\pi k i} = \begin{cases} +1, & \text{za } k \text{ paran} \\ -1, & \text{za } k \text{ neparan} \end{cases}$$

možemo fazni interpolacijski polinom P koji prolazi kroz sve čvorove izraziti pomoću Q i \tilde{R} na ovaj način

$$P(x) = Q(x) \frac{1 + e^{Mxi}}{2} + R\left(x - \frac{\pi}{M}\right) \frac{1 - e^{Mxi}}{2}, \quad (17)$$

ili

$$2P(x) = Q(x)(1 + e^{Mxi}) + R\left(x - \frac{\pi}{M}\right)(1 - e^{Mxi}). \quad (18)$$

Uočimo da je svaki pribrojnik u (17) polinom u e^{xi} stupnja $\leq N - 1$, pa je takva i njihova suma. Nadalje,

$$\begin{aligned} P(x_k) &= Q(x_k) \frac{1 + e^{Mx_k i}}{2} + R\left(x_k - \frac{\pi}{M}\right) \frac{1 - e^{Mx_k i}}{2} \\ &= \begin{cases} y_k \frac{1 + 1}{2} + R\left(x_k - \frac{\pi}{M}\right) \frac{1 - 1}{2} = y_k, & k \text{ paran} \\ Q(x_k) \frac{1 - 1}{2} + y_k \frac{1 + 1}{2} = y_k, & k \text{ neparan} \end{cases} \end{aligned}$$

tj. za svaki k vrijedi $P(x_k) = y_k$, pa time (17) predstavlja rastav faznog interpolacijskog polinoma određenog svim čvorištima (ne smijemo zaboraviti da je $N = 2M$).

Ovakav rastav faznog interpolacijskog polinoma sugerira sljedeće razmatranje. Pođemo li od $N = 2^n = 2 \cdot 2^{n-1} = 2M$ onda prema (17) traženi fazni interpolacijski polinom P stupnja $N - 1 = 2^n - 1 = 2M - 1$ dobivamo pomoću dva fazna interpolacijska polinoma $(N - 1) = (2 \cdot 2^{n-1} - 1)$ -og stupnja. Zbog $2^{n-1} = 2 \cdot 2^{n-2}$ možemo prema (17) svaki od njih dobiti pomoću dva fazna interpolacijska polinoma $(2 \cdot 2^{n-1} - 1)$ -og stupnja što ukupno daje $2^2 = 4$ fazna interpolacijska polinoma

$(2 \cdot 2^{n-2} - 1)$ -og stupnja. U sljedećem koraku bi imali $2^3 = 8$ faznih interpolacijskih polinoma $(2 \cdot 2^{n-3} - 1)$ -og stupnja i tako dalje redom dok ne dođemo do zadnjeg koraka tj. do $2^n = N$ faznih interpolacijskih polinoma 0-og stupnja. To nam sugerira sljedeću rekurzivnu šemu od n koraka.

Opišimo sada opći korak te šeme. Neka je $m \leq n$, $N = 2^n$. Zapišimo N u obliku

$$N = 2 \cdot 2^{m-1} \cdot 2^{n-m} = 2MS, \quad M = 2^{m-1}, \quad S = 2^{n-m}.$$

Po gore opisanom u tom koraku treba odrediti $2^{n-m} = S$ faznih interpolacijskih polinoma $(2 \cdot 2^{m-1} - 1) = (2M - 1)$ -og stupnja. Označimo te polinome sa

$$P_s^{(m)}(x) = \beta_{s0}^{(m)} + \beta_{s1}^{(m)} e^{ix} + \dots + \beta_{s,2M-1}^{(m)} e^{(2M-1)ix}, \quad s = 0, 1, 2, \dots, S-1.$$

Prema rekurziji (17) oni se određuju pomoću $2S$ faznih interpolacijskih polinoma $P_s^{(m-1)}$, $s = 0, 1, \dots, 2S-1$, pri čemu (17) daje

$$2P_s^{(m)}(x) = P_s^{(m-1)}(x) \left(1 + e^{Mix} \right) + P_{S+s}^{(m-1)} \left(x - \frac{\pi}{M} \right) \left(1 - e^{Mix} \right).$$

Ta relacija povlači jednakost među koeficijentima istog stupnja, pa dobivamo sljedeće rekurzivne relacije između koeficijenata gornjih faznih interpolacijskih polinoma

$$2\beta_{sj}^{(m)} = \beta_{sj}^{(m-1)} + \beta_{S+s,j}^{(m-1)} \left(e^{-\frac{2\pi i}{M}} \right)^j \quad s = 0, 1, \dots, S-1$$

$$2\beta_{s,M+j}^{(m)} = \beta_{s,j}^{(m-1)} - \beta_{S+s,j}^{(m-1)} \left(e^{-\frac{2\pi i}{M}} \right)^j \quad j = 0, 1, \dots, M-1.$$

Uvedemo li kraću oznaku $\varepsilon_m = e^{-\frac{2\pi i}{M}}$, $m = 0, 1, 2, \dots, n$, možemo dobiveni rezultat pisati ovako

$$2\beta_{sj}^{(m)} = \beta_{sj}^{(m-1)} + \beta_{S+s,j}^{(m-1)} (\varepsilon_m)^j \quad s = 0, 1, \dots, S-1$$

$$2\beta_{s,M+j}^{(m)} = \beta_{s,j}^{(m-1)} - \beta_{S+s,j}^{(m-1)} (\varepsilon_m)^j \quad j = 0, 1, \dots, M-1$$

$$\varepsilon_m = e^{-\frac{2\pi i}{M}}, \quad m = 0, 1, 2, \dots, n.$$

Rekurzija počinje od 0 prema n , tj. da stavimo

$$\beta_{s0}^{(0)} = y_s, \quad s = 0, 1, 2, \dots, N-1$$

što su fazni interpolacijski polinomi 0-og stupnja i završava s

$$\beta_j = \beta_{0j}^{(n)}, \quad j = 0, 1, 2, \dots, N-1$$

što su koeficijenti traženog faznog interpolacijskog polinoma i predstavljaju sume oblika (15). Opisana rekurzija je tipična za Cooley-Tukeyevu metodu.

1.4. Čebiševljevi polinomi, minimaks polinom i teleskopiranje redova potencija

Čebiševljevi polinomi

Promatramo li potencije $1 = x^0, x, x^2, \dots, x^n$ na segmentu $[-1, 1]$ onda znamo da apsolutna vrijednost svakog od njih poprima maksimalnu vrijednost 1 u točki $x = \pm 1$, a minimalna 0 u točki $x = 0$. Ako f aproksimiramo s

$$P_n(x) = a_0 + a_1 x + \dots + a_n x^n \quad (1)$$

na $[-1, 1]$, onda zanemarivanje članova s višom potencijom, ili mala izmjena koeficijenata a_1, a_2, \dots, a_n , izaziva malu pogrešku za male x (oko nule), a vjerojatno znatnu pogrešku oko krajeva segmenta, tj. oko ± 1 . Ograničimo razmatranje na segment $[-1, 1]$. To ne predstavlja bitno ograničenje jer supstitucijom

$$x = \frac{2\xi - b - a}{b - a} \quad (2)$$

transformiramo segment $[a, b]$ na segment $[-1, 1]$, dok

$$\xi = \frac{x(b - a) + (b + a)}{2} \quad (3)$$

transformiramo segment $[-1, 1]$ na $[a, b]$.

Naša je opća želja da polinomom P_n aproksimiramo funkciju tako da pogreška aproksimacije $|f(x) - P_n(x)|$ bude "jednoličnije" raspoređena duž segmenta, zatim da se P_n lako nalazi i da maksimalnu pogrešku svede na minimum. To ne možemo općenito postići ali, promatramo li kosinus funkcije $\cos \varphi, \cos 2\varphi, \dots, \cos n\varphi$ vidimo da svaka ima (po apsolutnoj vrijednosti) jednake maksimalne vrijednosti koje su jednoliko raspoređene nad segmentom, recimo $[0, \pi]$. Nadalje, ekstremne vrijednosti dviju takvih funkcija $\cos j\varphi$ i $\cos k\varphi$, $k \neq j$ ne događaju se općenito na istom mjestu. Međutim, koristimo li kosinus funkcije onda ih trebamo aproksimirati da bi izračunali njihove vrijednosti. Jednostavan i koristan rezultat dobivamo ako transformiramo $\cos n\varphi$ na segmentu $[0, \pi]$ u polinom n -tog stupnja na segmentu $[-1, 1]$. Polinomi

$$T_n(x) = \cos n\varphi, \quad \varphi = \arccos x, \quad n = 0, 1, 2, \dots \quad (4)$$

zovu se **Čebiševljevi polinomi**. Da je T_n polinom n -tog stupnja vidimo tako da $\cos n\varphi$ prikažemo kao polinom n -tog stupnja od $\cos \varphi$ i zatim uvrstimo $\varphi = \arccos x$. Odmah vidimo da je $T_0(x) = \cos 0 = 1$, dok je Čebiševljev polinom prvog stupnja

$$T_1(x) = \cos \varphi = \cos(\arccos x) = x.$$

Odredimo još T_2 iz čega će nam biti jasna rekurzivna formula koja vrijedi za Čebiševljeve polinome. Znamo da vrijedi $\cos 2\varphi = 2\cos^2 \varphi - 1$. To nam daje

$$\cos 2\varphi = 2\cos^2(\arccos x) - 1$$

ili

$$T_2(x) = 2x^2 - 1.$$

Općenito, primjenom trigonometrijskog identiteta

$$\cos n\varphi = 2 \cos \varphi \cos(n-1)\varphi - \cos(n-2)\varphi \quad (5)$$

izračunavamo Čebiševljeve polinome višeg stupnja, pomoću onih nižeg stupnja. Formula (5) daje ovu rekurzivnu formulu

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x). \quad (6)$$

Da nađemo formulu (5) podimo od $\cos \alpha + \cos \beta = 2 \cos \frac{\alpha + \beta}{2} \cos \frac{\alpha - \beta}{2}$ i uvrstimo $\alpha = n\varphi$, $\beta = (n-2)\varphi$. Prvih deset Čebiševljevih polinoma glasi:

$$\begin{aligned} T_0(x) &= 1, \\ T_1(x) &= x, \\ T_2(x) &= 2x^2 - 1, \\ T_3(x) &= 4x^3 - 3x, \\ T_4(x) &= 8x^4 - 8x^2 + 1, \\ T_5(x) &= 16x^5 - 20x^3 + 5x, \\ T_6(x) &= 32x^6 - 18x^4 + 18x^2 - 1, \\ T_7(x) &= 64x^7 - 112x^5 + 56x^3 - 7x, \\ T_8(x) &= 128x^8 - 256x^6 + 160x^4 - 32x^2 + 1, \\ T_9(x) &= 256x^9 - 576x^7 + 432x^5 - 120x^3 + 9x. \end{aligned} \quad (7)$$

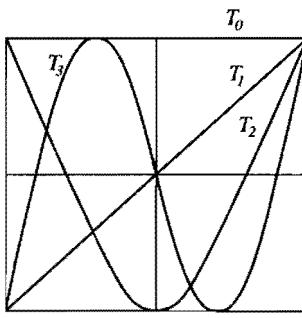
Želimo li potencije 1 , x , x^2 , \dots , x^9 izraziti pomoću Čebiševljevih polinoma onda lako dobivamo:

$$\begin{aligned} 1 &= T_0(x), \\ x &= T_1(x), \\ x^2 &= \frac{1}{2}\{T_0(x) + T_2(x)\}, \\ x^3 &= \frac{1}{4}\{3T_1(x) + T_3(x)\}, \\ x^4 &= \frac{1}{8}\{3T_0(x) + 4T_2(x) + T_4(x)\}, \\ x^5 &= \frac{1}{16}\{10T_1(x) + 5T_3(x) + T_5(x)\}, \\ x^6 &= \frac{1}{32}\{10T_0(x) + 15T_2(x) + 6T_4(x) + T_6(x)\}, \\ x^7 &= \frac{1}{64}\{35T_1(x) + 21T_3(x) + 7T_5(x) + T_7(x)\}, \\ x^8 &= \frac{1}{128}\{35T_0(x) + 56T_2(x) + 28T_4(x) + 8T_6(x) + T_8(x)\}, \\ x^9 &= \frac{1}{256}\{126T_1(x) + 84T_3(x) + 36T_5(x) + 9T_7(x) + T_9(x)\}. \end{aligned} \quad (8)$$

Na slici 1.3. vidimo grafove prva četiri Čebiševljeva polinoma. Primijetimo da Čebiševljevi polinomi na segmentu $[-1, 1]$ (zbog svog porijekla $T_n(x) = \cos n\varphi$) imaju $n + 1$ ekstremnu vrijednost (sve imaju istu absolutnu vrijednost 1) i to naizmjence pozitivnu i negativnu. Prema tome T_n ima n različitih nultočaka, sve su realne i nalaze se u intervalu $(-1, 1)$. Nultočke od T_n su dane sa

$$x_k = \cos \frac{(2k+1)\pi}{2n}, \quad k = 0, 1, 2, \dots, n-1. \quad (9)$$

Naime, iz $T_n(x) = 0$ dobivamo zahtjev $\cos n\varphi = 0$, što je ispunjeno za $n\varphi = (2k+1)\frac{\pi}{2}$, tj. $\varphi = \frac{(2k+1)\pi}{2n}$, $k = 0, 1, 2, \dots, n-1$. Supstitucija (4) daje odmah rezultat (9).



Sl. 1.3.

Promatrajmo sada normirani oblik Čebiševljevog polinoma

$$Q_n(x) = \frac{T_n(x)}{2^{n-1}}, \quad n \geq 2 \quad (10)$$

tj. podijelimo T_n s vodećim koeficijentom koji iznosi 2^{n-1} (što je očigledno iz rekurzivne formule (6)). Označimo sa \mathcal{P}_n skup svih normiranih polinoma n -tog stupnja, tj. skup svih polinoma n -tog stupnja s vodećim koeficijentom jednakim 1. Vrlo korisno svojstvo Čebiševljevih polinoma dano je sljedećim teoremom.

Teorem 1.3. *Među svim normiranim polinomima \mathcal{P}_n , promatrаниm na segmentu $[-1, 1]$, Q_n ima najmanju gornju ogragu za svoju absolutnu vrijednost.*

Dokaz. Kako je $\max |T_n(x)| = 1$ u svim ekstremnim točkama to imamo da je $\max |Q_n(x)| = \frac{1}{2^{n-1}}$ u svim ekstremnim točkama, kojih ima točno $n+1$. Isto tako znamo da se maksimum i minimum pojavljuju naizmjence. Prepostavimo da u \mathcal{P}_n postoji polinom S_n koji na $[-1, 1]$ ima manje maksimalne vrijednosti od Q_n . Promatrajmo razliku

$$D(x) = Q_n(x) - S_n(x).$$

Neka su x_0, x_1, \dots, x_n apscise ekstremnih vrijednosti od Q_n . Kako je po pretpostavci $|S_n(x_k)| \leq |Q_n(x_k)|$, $k = 0, 1, \dots, n$ to slijedi da razlika

$$D(x_k) = Q_n(x_k) - S_n(x_k)$$

mijenja predznak za svaki sljedeći k , tj. ako je $D(x_0) < 0$, onda je $D(x_1) > 0$, $D(x_2) < 0$, itd., i ako je $D(x_0) > 0$, onda je $D(x_1) < 0$, $D(x_2) > 0$, itd. (Naime, ako je x_k minimum od Q_n , onda je $D(x_k) < 0$, dok je $D(x_k) > 0$, ako je x_k točka maksimuma od Q_n .) Na taj način zaključujemo da D mijenja predznak točno n puta što ima za posljedicu da D ima barem n nultočaka na $[-1, 1]$. To je nemoguće zato što je D polinom stupnja $\leq n - 1$ (Q_n i S_n imaju vodeće koeficijente jednake 1) i može imati najviše $n - 1$ nultočku. Time je teorem dokazan.

Minimaks princip

Dobiveni rezultat možemo iskoristiti na sljedeći način. Prisjetimo se ocjene pogreške prilikom aproksimacije interpolacijskim polinomom u Lagrangeovom obliku. Ista je dovedena u oblik

$$f(x) - L_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} (x - x_0)(x - x_1) \dots (x - x_n). \quad (11)$$

Osnovni cilj aproksimacije je postići minimalnu pogrešku. Iz građe formule (11) vidimo da na faktor $f^{(n+1)}(\xi)$ možemo jako malo utjecati. Naime, ako znamo njegovu ogradu onda smo zadovoljni, prema tome, u ocjenama ga smatramo konstantom. Kako je $(n+1)!$ konstanta, jedini dio koji eventualno možemo minimizirati je polinom $(x - x_0)(x - x_1) \dots (x - x_n)$. To je polinom $(n+1)$ -og stupnja, a x_0, x_1, \dots, x_n su njegove nultočke. Da ga minimiziramo moramo prema gornjem teoremu zahtjevati da taj polinom bude jednak Q_{n+1} . To drugim riječima znači, da minimiziramo pogrešku (11) moramo čvorove izabrati tako da to budu nultočke od Q_{n+1} , što su zapravo nultočke Čebiševljevog polinoma T_{n+1} , te su prema (9) dane sa

$$x_k = \cos \frac{(2k+1)\pi}{2n+2}, \quad k = 0, 1, 2, \dots, n. \quad (12)$$

Za opći segment $[a, b]$ treba prema (3) iz 4.1. uzeti

$$\xi_k = \frac{x_k(b-a) + (b+a)}{2} \quad (13)$$

za čvorove interpolacije. Odredimo li na $[a, b]$ interpolacijski polinom, označimo ga sa M_n , sa čvorovima danim prema (13) onda dobivamo interpolacijski polinom koji minimizira maksimalnu pogrešku. Primjena polinoma M_n je moguća samo onda kada imamo slobodu u odabiranju čvorova interpolacije. Srećom to je čest slučaj kod eksperimentalnih istraživanja. Takvo minimiziranje maksimalne pogreške interpolacijskog polinoma ponekad se zove **minimaks princip**.

Minimaks ili optimalni polinom

Opći problem može se ovako formulirati: za danu funkciju f naći polinom P_n koji minimizira maksimalnu pogrešku, tj. odrediti polinom iz zahtjeva

$$\max_{a \leq x \leq b} |f(x) - P_n^*(x)| = \text{minimalan}. \quad (14)$$

Polinom P_n^* određen zahtjevom (14) zove se **minimaks polinom** ili **optimalni aproksimacijski polinom**. Za polinom P_n^* vrijedi sljedeće (dokaz izostavljam):

- a) P_n^* je jedinstven.

b) Devijacija

$$D(x) = f(x) - P_n^*(x) \quad (15)$$

postiže svoju ekstremnu vrijednost u barem $n + 2$ točke na $[-1, 1]$ (ili $[a, b]$) koja je naizmjence pozitivna i negativna.

Pretpostavka o neprekidnosti od f povlači neprekidnost devijacije D , pa zbog promjene predznaka D ima barem $n + 1$ različitu nultočku. To znači da f i P_n^* imaju jednakе vrijednosti u barem $(n + 1)$ -oj točki, odnosno, P_n^* je stanoviti interpolacijski polinom stupnja $\leq n$. Kada je f $(n + 1)$ -puta diferencijabilna i $f^{(n+1)} = \text{const.}$, tada je interpolacijski polinom M_n , tj. s čvorovima danim prema (13), jednak minimaks polinomu P_n^* . Ako je $f^{(n+1)} \neq \text{const.}$, onda je M_n samo aproksimacija od P_n^* . Nalaženje P_n^* , kada je $f^{(n+1)}$ nepoznata ili $f^{(n+1)} \neq \text{const.}$ dosta je složeno i nećemo ovdje opisivati.

Teleskopiranje redova potencija

Istaknimo da su Čebiševljevi polinomi ortogonalni na $[-1, 1]$, pa nam se nameće pitanje razvoja u Fourierov red po polinomima T_n . Ako se f razvije u Fourierov red po Čebiševljevim polinomima, tj.

$$f(x) = \sum_{k=0}^{\infty} c_k T_k(x), \quad (16)$$

onda se može pokazati da će Fourierov polinom (tj. parcijalna suma reda (16))

$$S_n(x) = \sum_{k=0}^n c_k T_k(x) \quad (17)$$

obično biti vrlo dobra aproksimacija minimaks polinoma. Nažalost, određivanje koeficijenata c_k je mukotrpan posao. Postoje tablice tih koeficijenata za neke funkcije kao što su: trigonometrijske, arkus, eksponencijalna, hiperbolička, area i neke druge. Kako su koeficijenti Fourierovog reda (16) dani pomoću integrala, njihove se aproksimacije mogu dobiti pomoću numeričke integracije.

Ako je f polinom n -tog stupnja onda se koeficijenti c_i lako odrede pomoću formula (8). Jednostavno zamijenimo x^i sa svojim razvojem po Čebiševljevim polinomima i sredimo izraz po polinomima T_j . Na takvu situaciju nailazimo npr. kada f aproksimiramo Taylorovim polinomom. U takvim situacijama obično možemo zanemariti neke Čebiševljeve polinome višeg reda znajući da je time učinjena pogreška malena (gornja ograda za $T_n(x)$ na $[-1, 1]$ jednaka je 1). Tada dobiveni polinom možemo natrag srediti po potencijama od x (to općenito više nije Taylorov polinom od f) koji je sada nižeg stupnja od polaznog Taylorovog polinoma. Takav se postupak zove **Čebiševsko ekonomiziranje ili teleskopiranje reda potencija**. To ilustriramo na primjeru kosinusne funkcije. Pođemo li od Taylorovog reda

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \frac{x^{10}}{10!} + \dots + (-1)^k \frac{x^{2k}}{2k!} + \dots$$

i promatramo Taylorov polinom $n = 2k$ -tog stupnja, onda je pogreška (red alternira) manja od apsolutne vrijednosti prvog zanemarenog člana. Ograničimo li se na segment

$[-1, 1]$ i $n = 8$, dobivamo da je pogreška manja od $\frac{1}{10!} = 2.7610^{-7}$ za $x \in [-1, 1]$. Imamo dakle

$$\begin{aligned}\cos x &\approx 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} \\&= T_0 - \frac{1}{2!} \cdot \frac{1}{2}(T_0 + T_2) + \frac{1}{4!} \cdot \frac{1}{8}(3T_0 + 4T_2 + T_4) \\&\quad - \frac{1}{6!} \cdot \frac{1}{32}(10T_0 + 15T_2 + 6T_4 + T_6) \\&\quad + \frac{1}{8!} \cdot \frac{1}{128}(35T_0 + 56T_2 + 28T_4 + 8T_6 + T_8).\end{aligned}$$

Sredimo li dobiveni izraz po Čebiševljevim polinomima dobivamo

$$\begin{aligned}\cos x &\approx 0.76519775T_0 - 0.22980686T_2 + 0.0049533419T_4 \\&\quad - 4.185265 \cdot 10^{-5}T_6 + 1.937624 \cdot 10^{-7}T_8.\end{aligned}$$

Zanemarimo li članove T_6 i T_8 uvodimo dodatnu pogrešku koja nije veća od $4.186 \cdot 10^{-5} + 1.938 \cdot 10^{-7} = 4.206 \cdot 10^{-5}$. Dodamo li to pogrešci Taylorovog reda $2.76 \cdot 10^{-7}$ vidimo da je ukupna pogreška manja od $4.234 \cdot 10^{-5}$. Dakle, pogreška te aproksimacije je manja od $5 \cdot 10^{-5}$ na segmentu $[-1, 1]$. Istaknimo da je dobivena aproksimacija polinom četvrtog stupnja.

Ocjena pogreške aproksimacije Taylorovim polinomom četvrtog stupnja ($\cos x \approx 1 - \frac{x^2}{2!} + \frac{x^4}{4!}$) na $[-1, 1]$ iznosi $\frac{1}{6!} = \frac{1}{720} = 0.00179 = 139 \cdot 10^{-5}$ što je oko 28 puta veći iznos. Poslužimo li se sa (7) možemo dobivenu aproksimaciju srediti po potencijama od x , pa imamo

$$\cos x = 0.99995795 - 0.99924045x^2 + 0.03962674x^4.$$

Prepostavimo sada da je $5 \cdot 10^{-5}$ bila zahtjevana točnost. Uvjerimo se da tu točnost ne možemo postići teleskopiranjem Taylorovog polinoma šestog stupnja što bi mogli pomicati nakon gornjeg računa. Sam Taylorov polinom šestog stupnja zadovoljava tom zahtjevu jer je $\frac{1}{8!} = 2.5 \cdot 10^{-5} < 5 \cdot 10^{-5}$. Međutim, kada se $1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!}$ prikaže pomoću Čebiševljevih polinoma dobiva se uz T_8 koeficijent $4.34 \cdot 10^{-5}$ što zajedno sa $2.5 \cdot 10^{-5}$ daje pogrešku $6.84 \cdot 10^{-5} > 5 \cdot 10^{-5}$.

1.5. Polinom najmanjih kvadrata

U čvorovima zamišljamo da interpolacijski polinom ima vrijednosti funkcije, tj. u čvorovima je greška nula, međutim, izvan čvorova greška može biti značajna. Nadalje, vrijednosti $f(x_i)$, $i = 0, 1, 2, \dots, n$ redovito su netočne, često je taj polinom nižeg stupnja od n i što je najvažnije redovito imamo i više od $n+1$ točaka u kojima poznamo približnu vrijednost od f . Kako je sa $(n+1)$ -om točkom jedinstveno određen interpolacijski polinom stupnja $\leq n$, to je općenito nemoguće naći interpolacijski polinom stupnja $m < n$ koristeći svih $n+1$ zadanih točaka (uočite u zahtjevu striktnu

nejednakost). Stoga se prirodno nameće pitanje aproksimacije funkcije f polinomom P (koji možda nigdje ne poprima vrijednost od f) takvim da su njegove vrijednosti "što je moguće bliže" vrijednostima od f . Jedan od postupaka sastoji se u tome da zahtijevamo da u čvorovima bude minimalna suma kvadrata razlika između funkcije i polinoma. Neka je dakle dano $n + 1$ točaka (čvorova)

$$(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n) \quad (1)$$

gdje zbog kraćeg zapisivanja uvodimo $y_i = f(x_i)$. Traži se polinom

$$P_m(x) = a_0 + a_1 x + \dots + a_m x^m \quad (2)$$

tako da suma kvadrata odstupanja

$$E(a_0, a_1, \dots, a_m) = \sum_{i=0}^n \{P_m(x_i) - y_i\}^2 \quad (3)$$

bude minimalna. Dobiveni polinom zovemo **polinom najmanjih kvadrata**. Uočimo da je funkcija E polinom drugog stupnja s nepoznanicama a_0, a_1, \dots, a_m . Dakle, da odredimo koeficijente od P_m treba (3) parcijalno derivirati po nepoznatim koeficijentima a_k i parcijalne derivacije izjednačiti s nulom te riješiti dobiveni sustav linearnih jednadžbi. Te parcijalne derivacije glase

$$\frac{\partial E}{\partial a_k} = 2 \sum_{i=0}^n \{P_m(x_i) - y_i\} x_i^k = 0, \quad k = 0, 1, \dots, m$$

ili nakon dijeljenja sa 2 i prebacivanja konstantnog člana na desnu stranu imamo

$$\sum_{i=0}^n x_i^k P_m(x_i) = \sum_{i=0}^n x_i^k y_i, \quad k = 0, 1, \dots, m. \quad (4)$$

Ponovimo, (4) je linearan sustav jednadžbi koji se sastoji od $m + 1$ jednadžbe sa $m + 1$ nepoznanicom a_0, a_1, \dots, a_m . Zbog toga što je kvadratni polinom (3) ograničen odozdo s nulom to ima minimum iz čega zaključujemo da sustav (4) ima rješenje.

Primjer 1.2. Za $m = 2$ polazimo od polinoma 2. stupnja $P_2(x) = c + bx + ax^2$, tako da je

$$E(a, b, c) = \sum_{i=0}^n [c + bx_i + ax_i^2 - y_i]^2.$$

Parcijalnim deriviranjem po c, b i a dobivamo linearni sustav od tri jednadžbe s tri nepoznanice koji sređen glasi

$$\begin{aligned} (n+1)c + \left(\sum_{i=0}^n x_i\right)b + \left(\sum_{i=0}^n x_i^2\right)a &= \sum_{i=0}^n y_i \\ \left(\sum_{i=0}^n x_i\right)c + \left(\sum_{i=0}^n x_i^2\right)b + \left(\sum_{i=0}^n x_i^3\right)a &= \sum_{i=0}^n x_i y_i \\ \left(\sum_{i=0}^n x_i^2\right)c + \left(\sum_{i=0}^n x_i^3\right)b + \left(\sum_{i=0}^n x_i^4\right)a &= \sum_{i=0}^n x_i^2 y_i. \end{aligned}$$

Napišimo sada sustav (4) u razvijenom obliku tako da imamo (sve sume idu od $i = 0$ do $i = n$, pa je indeks sumacije izostavljen)

$$\begin{aligned} (n+1)a_0 + \left(\sum x_i\right)a_1 + \left(\sum x_i^2\right)a_2 + \dots + \left(\sum x_i^m\right)a_m &= \sum y_i \\ \left(\sum x_i\right)a_0 + \left(\sum x_i^2\right)a_1 + \left(\sum x_i^3\right)a_2 + \dots + \left(\sum x_i^{m+1}\right)a_m &= \sum x_i y_i \\ \left(\sum x_i^2\right)a_0 + \left(\sum x_i^3\right)a_1 + \left(\sum x_i^4\right)a_2 + \dots + \left(\sum x_i^{m+2}\right)a_m &= \sum x_i^2 y_i \quad (5) \\ &\vdots \\ \left(\sum x_i^m\right)a_0 + \left(\sum x_i^{m+1}\right)a_1 + \left(\sum x_i^{m+2}\right)a_2 + \dots + \left(\sum x_i^{2m}\right)a_m &= \sum x_i^m y_i \end{aligned}$$

Rješenje sustava (5) daje koeficijente a_0, a_1, \dots, a_m traženog polinoma najmanjih kvadrata koji u traženim točkama možemo računati pomoću Hornerove sheme. U primjenama je to gotovo redovito polinom niskog stupnja.

Do sustava (5) možemo i ovako doći, što će nam trenutak kasnije poslužiti da izvedemo stanoviti zaključak. Napišemo li niz jednakosti (koje bi imali u slučaju interpolacijskog polinoma):

$$\begin{aligned} a_0 + a_1 x_0 + a_2 x_0^2 + \dots + a_m x_0^m &= y_0 \\ a_0 + a_1 x_1 + a_2 x_1^2 + \dots + a_m x_1^m &= y_1 \\ &\vdots \\ a_0 + a_1 x_n + a_2 x_n^2 + \dots + a_m x_n^m &= y_n \end{aligned} \quad (6)$$

Matrica tog sustava je formata $(n+1) \times (m+1)$, dakle općenito više jednadžbi nego nepoznanica i izgleda ovako:

$$A = \begin{bmatrix} 1 & x_0 & x_0^2 & \dots & x_0^m \\ 1 & x_1 & x_1^2 & \dots & x_1^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^m \end{bmatrix}, \quad (7)$$

što podsjeća na Vandermondeovu determinantu (vidi 1.1).

Sam sustav (6) možemo u matričnom obliku ovako zapisati

$$Aa = y, \quad (8)$$

gdje su vektori a i y jednaki

$$a = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix}, \quad y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}. \quad (9)$$

Zamislimo sada da na sustavu (6) provedemo sljedeći račun. Zbrojimo li sve jednadžbe sustava (6) dobivamo prvu jednadžbu sustava (5). Zatim redom pomnožimo jednadžbu sustava (6) i to prvu s x_0 , drugu s x_1 , itd. i sve zbrojimo; dobivamo drugu jednadžbu sustava (5). Sada je već jasno kako se dobije r -ta jednadžba sustava (5), treba u sustavu (6) prvu pomnožiti s x_0^r , drugu s x_1^r , itd. redom i izračunato zbrojiti.

Ako se zapitamo kako je ovim postupkom nastao sustav (5) iz sustava (6), onda vidi-mo da u matričnom zapisu treba (8) pomnožiti slijeva s transponiranim matricom A^T matrice A . Dakle, u matričnom zapisu sustav (5) glasi

$$A^T A a = A^T y. \quad (10)$$

Gornji izvod nam služi da uvidimo sljedeće: *Ako je $m = n$, onda je polinom najmanjih kvadrata jednak interpolacijskom polinomu.* Naime, za $m = n$ matrica A je kvadratna. Tada je A regularna, jer je njena determinanta upravo Vandermondeova determinanta. Time je regularna i njena transponirana matrica. Kako se u tom slučaju (5) dobiva iz (6) množenjem s regularnom matricom A^T slijedi da su tada sustavi (5) i (6) ekvivalentni. Kako u tom slučaju (6) određuje interpolacijski polinom zaključujemo da je tada polinom najmanjih kvadrata jednak interpolacijskom.

Primjer 1.3. Odredimo "pravac" najmanjih kvadrata za sljedeći niz podataka

x_i	0	1	2	3	4
$y_i = f(x_i)$	1.00	3.85	6.50	9.35	12.05

Pripadni sustav (5) glasi

$$5a_0 + 10a_1 = 32.75$$

$$10a_0 + 30a_1 = 93.10$$

čije je rješenje $a_0 = 1.03$ i $a_1 = 2.76$. Dakle, jednadžba "pravca" najmanjih kvadrata glasi

$$y = 1.03 + 2.76x.$$

U točkama $x = 0, 1, 2, 3, 4$ "pravac" najmanjih kvadrata ima vrijednosti $y = 1.03; 3.79; 6.55; 9.31; 12.07$, što više nisu vrijednosti dane gornjom tablicom.

Uočimo da je polinom najmanjih kvadrata određen iz uvjeta (3), tj. iz zahtjeva da je suma kvadrata odstupanja u promatranih točkama minimalna. Kako je $P_m(x) = a_0 + a_1x + \dots + a_mx^m$ linearna kombinacija linearne nezavisnih funkcija $x^0, x^1, x^2, \dots, x^m$, to se postupak može proširiti i na linearnu kombinaciju drugih linearne nezavisnih funkcija $\varphi_0, \varphi_1, \dots, \varphi_m$, tj. možemo promatrati

$$\varphi(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \dots + a_m\varphi_m(x) \quad (11)$$

i odrediti koeficijente a_0, a_1, \dots, a_m iz uvjeta da suma kvadrata odstupanja

$$F(a_0, a_1, \dots, a_m) = \sum_{i=0}^n \{\varphi(x_i) - y_i\}^2 \quad (12)$$

bude minimalna. Općenito se takav tip aproksimacije naziva **L_2 -aproksimacija**. Daljnji korak u poopćenju je da se pojedinim čvorovima pripiše određena težina w_i , tako da se koeficijenti a_0, a_1, \dots, a_n određuju iz uvjeta da je

$$G(a_0, a_1, \dots, a_m) = \sum_{i=0}^n w_i \{\varphi(x_i) - y_i\}^2 \quad (13)$$

minimalno. Ovo zadnje se primjenjuje i na polinom najmanjih kvadrata.

2.

Numeričko integriranje

2.1. Newton–Cotesove formule	41
2.2. Ponavljanje raspolažljanja i	
Rombergov algoritam	50
2.3. Gaussove formule	54
2.4. Integriranje brzo oscilirajućih funkcija	
Filonova formula	58

U matematičkoj analizi smo problem nalaženja određenog integrala $\int_a^b f(x)dx$ rješavali pomoću Leibniz–Newtonove formule

$$\int_a^b f(x)dx = F(b) - F(a)$$

gdje je $F'(x) = f(x)$. Kako u mnogim primjenama ne znamo $F(x)$, to je potrebno aproksimirati vrijednost određenog integrala. U primjenama je često i $f(x)$ nepoznata, tj. poznajemo njene približne vrijednosti samo u nekim točkama. Zamislimo da nam za takvu funkciju trebaju Fourierovi koeficijenti po nekom sustavu ortogonalnih funkcija. Upravo ovih par rečenica kažu da su nam potrebne numeričke metode približnog izračunavanja određenih integrala.

Opća ideja je da funkciju $f(x)$ aproksimiramo jednostavnijom funkcijom $g(x)$ i da iz $f(x) \approx g(x)$ dobijemo

$$\int_a^b f(x)dx = \int_a^b g(x)dx + R(f)$$

te da izračunavanje prvog člana desne strane bude jednostavno i da po mogućnosti možemo ocijeniti pogrešku $R(f)$.

2.1. Newton–Cotesove formule

Opće Newton–Cotesove formule

Neka je $g(x)$ interpolacijski polinom, tj.

$$\begin{aligned} g(x) &= L_n(x) = \sum_{i=0}^n y_i \prod_{\substack{j \neq i \\ j=0}}^n \frac{x - x_j}{x_i - x_j} \\ &= \sum_{i=0}^n y_i \frac{\prod_{j=0}^n (x - x_j)}{(x - x_i) \prod_{\substack{j \neq i \\ j=0}}^n (x_i - x_j)} \end{aligned} \quad (1)$$

za koji znamo da vrijedi $L_n(x_i) = f(x_i) = y_i$, $i = 0, 1, 2, \dots, n$.

Imamo dakle

$$\int_a^b f(x) dx = \int_a^b L_n(x) dx + R_n(f) \quad (2)$$

gdje je $R_n(f)$ pogreška aproksimacije. Uvrštavanjem (1) u (2) dobivamo formulu za aproksimaciju

$$\begin{aligned} \int_a^b f(x) dx &\approx \int_a^b \sum_{i=0}^n y_i \frac{\prod_{j=0}^n (x - x_j)}{(x - x_i) \prod_{\substack{j \neq i \\ j=0}}^n (x_i - x_j)} dx \\ &= \sum_{i=0}^n y_i \int_a^b \frac{\prod_{j=0}^n (x - x_j)}{(x - x_i) \prod_{\substack{j \neq i \\ j=0}}^n (x_i - x_j)} dx \\ &= \sum_{i=0}^n y_i A_i \end{aligned} \quad (3)$$

gdje su

$$A_i = \int_a^b \frac{\prod_{j=0}^n (x - x_j)}{(x - x_i) \prod_{\substack{j \neq i \\ j=0}}^n (x_i - x_j)} dx, \quad i = 0, 1, 2, \dots, n. \quad (4)$$

Da odredimo A_i primijetimo da isti ne ovise o f . Ako je dakle f polinom $\leq n$ -tog stupnja, onda je $L_n(x) = f(x)$. Uzmemو li napose $f(x) = x^k$, $k = 0, 1, \dots, n$,

onda imamo da je $R_n(x^k) = 0$. Uvrstimo li to redom u (3) dobivamo sljedeći sustav linearnih jednadžbi

$$\begin{aligned} I_0 &= \int_a^b x^0 dx = \sum_{i=0}^n A_i x_i^0 = \sum_{i=0}^n A_i \\ I_1 &= \int_a^b x^1 dx = \sum_{i=0}^n A_i x_i^1 = \sum_{i=0}^n x_i A_i \\ I_2 &= \int_a^b x^2 dx = \sum_{i=0}^n A_i x_i^2 = \sum_{i=0}^n x_i^2 A_i \\ &\vdots \\ I_n &= \int_a^b x^n dx = \sum_{i=0}^n A_i x_i^n = \sum_{i=0}^n x_i^n A_i \end{aligned} \quad (5)$$

gdje je

$$I_k = \int_a^b x^k dx = \frac{b^{k+1} - a^{k+1}}{k+1}, \quad k = 0, 1, 2, \dots, n.$$

Determinanta sustava jednažbi (5) je različita od nule jer je to Vandermondeova determinanta čija je vrijednost jednaka

$$D = \prod_{i>j} (x_i - x_j) \neq 0,$$

tako da sustav (5) ima jedinstveno rješenje. Rješivši sustav (5), možemo dobivene vrijednosti A_i uvrstiti u (3) otkuda dobivamo

$$\int_a^b f(x) dx \approx \sum_{i=0}^n y_i A_i. \quad (6)$$

Uočimo da smo zaobišli efektivno nalaženje interpolacijskog polinoma i njegovo integriranje.

Ako je $a = x_0$ i $b = x_n$ kaže se da je (3) **formula zatvorenog tipa**, inače **otvorenog tipa**.

Zadržat ćemo se na zatvorenom tipu formule (3), tj. kada je

$$x_0 = a \quad \text{i} \quad x_n = b$$

uz ekvidistantnu (jednoliku) podjelu

$$h = \frac{b-a}{n} = \frac{x_n - x_0}{n}, \quad x_i = x_0 + ih, \quad y_i = f(x_i), \quad i = 0, 1, \dots, n.$$

Izvedimo sada eksplicitne formule za koeficijente A_i . Uvedimo parametar

$$t = \frac{x - x_0}{h}, \quad \text{odnosno} \quad x = x_0 + th.$$

Tada imamo

$$\begin{aligned}x - x_0 &= th \\x - x_1 &= (x - x_0) - (x_1 - x_0) = th - h \\x - x_2 &= (x - x_1) - (x_2 - x_1) = th - 2h \\\vdots \\x - x_n &= th - nh.\end{aligned}$$

Nadalje je

$$\begin{aligned}x_i - x_0 &= ih \\x_i - x_1 &= (x_i - x_0) - (x_1 - x_0) = ih - h = (i - 1)h \\\vdots \\x_i - x_{i-1} &= h \\x_i - x_{i+1} &= -h \\\vdots \\x_i - x_n &= -(n - i)h.\end{aligned}$$

Uvrštenjem u (4) imamo

$$\begin{aligned}A_i &= \int_0^n \frac{th(th-h)\dots(th-ih)\dots(th-nh)}{(th-ih)ih(i-1)h\dots h(-h)\dots(-(n-i)h)} h dt \\A_i &= h \frac{(-1)^{n-i}}{i!(n-i)!} \int_0^n \frac{t(t-1)\dots(t-n)}{t-i} dt, \\&= \frac{x_n - x_0}{n} \frac{(-1)^{n-i}}{i!(n-i)!} \int_0^n \frac{t(t-1)\dots(t-n)}{t-i} dt, \quad i = 0, 1, 2, \dots, n.\end{aligned}$$

Uvedemo li

$$H_i = \frac{A_i}{x_n - x_0} = \frac{1}{n} \frac{(-1)^{n-i}}{i!(n-i)!} \int_0^n \frac{t(t-1)\dots(t-n)}{t-i} dt, \quad i = 0, 1, 2, \dots, n \quad (7)$$

preveli smo (4) u sljedeći oblik

$$\int_{x_0}^{x_n} f(x) dx \approx (x_n - x_0) \sum_{i=0}^n H_i y_i = (b - a) \sum_{i=0}^n H_i y_i \quad (8)$$

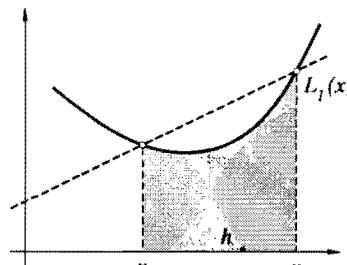
gdje su H_i dani sa (7) jednom zauvijek i zovu se **Cotesovi koeficijenti**, dok se formula (8) zove **Newton-Cotesova formula**.

Cotesovi koeficijenti imaju ova dva svojstva:

1. $\sum_{i=0}^n H_i = 1$;
2. $H_i = H_{n-i}$.

Primjeri Newton – Cotesovih formula

1. Osnovna trapezna formula.



Sl. 2.1.

To je slučaj $n = 1$

$$H_0 = - \int_0^1 \frac{t(t-1)}{t} dt = - \int_0^1 (t-1) dt = \frac{1}{2}$$

$$H_1 = \int_0^1 t dt = \frac{1}{2}.$$

Imamo dakle sljedeću aproksimaciju

$$\int_{x_0}^{x_1} f(x) dx \approx \frac{x_1 - x_0}{2} (y_0 + y_1) = \frac{h}{2} (y_0 + y_1). \quad (9)$$

Sada je prema (2) pogreška aproksimacije jednaka

$$R_1 = \int_{x_0}^{x_1} f(x) dx - \frac{h}{2} (y_0 + y_1) = \int_{x_0}^{x_0+h} f(x) dx - \frac{h}{2} [f(x_0) + f(x_0 + h)].$$

Pretpostavimo da je f klase C^2 , tj. da na promatranom segmentu ima neprekidnu drugu derivaciju. U dobivenoj formuli možemo varirati h , pa postupimo ovako. Prvo dvaput deriviramo R_1 po h (uocimo da se h nalazi na gornjoj granici integrala). Imamo dakle

$$\begin{aligned} R'_1 &= f(x_0 + h) - \frac{1}{2} [f(x_0) + f(x_0 + h)] - \frac{h}{2} f'(x_0 + h) \\ &= \frac{1}{2} [f(x_0 + h) - f(x_0)] - \frac{h}{2} f'(x_0 + h) \\ R''_1 &= \frac{1}{2} f'(x_0 + h) - \frac{1}{2} f'(x_0 + h) - \frac{h}{2} f''(x_0 + h) \\ &= -\frac{h}{2} f''(x_0 + h), \end{aligned}$$

pri čemu je $R_1(0) = R'_1(0) = 0$.

Sada obratno, integriramo dobiveni rezultat pa imamo

$$R'_1 = R'_1(0) + \int_0^h R'' dh = -\frac{1}{2} \int_0^h h f''(x_0 + h) dh$$

tako da primjenom teorema o srednjoj vrijednosti integralnog računa, koji ponovimo glasi

(Ako su funkcije $\varphi(x)$ i $\psi(x)$ neprekidne na $[a, b]$ i $\psi(x)$ ne mijenja predznak na $[a, b]$, onda je

$$\int_a^b \varphi(x)\psi(x)dx = \varphi(\xi) \int_a^b \psi(x)dx$$

gdje je $a < \xi < b$.),
dobivamo

$$R'_1 = -\frac{1}{2}f''(\xi_1) \int_0^h h dh = -\frac{h^2}{4}f''(\xi_1), \quad x_0 < \xi_1 < x_0 + h = x_1.$$

Ponovnom integracijom i primjenom teorema o srednjoj vrijednosti imamo

$$\begin{aligned} R_1 &= R_1(0) + \int_0^h R'_1 dh = -\frac{1}{4} \int_0^h h^2 f''(\xi_1) dh = -\frac{1}{4} f''(\xi) \int_0^h h^2 dh \\ &= -\frac{h^3}{12} f''(\xi), \quad x_0 < \xi < x_1. \end{aligned}$$

Dakle pogreška iznosi

$$R_1 = -\frac{h^3}{12} f''(\xi), \tag{10}$$

pa ako je $|f''(x)| \leq M_2$ na $[x_0, x_1]$ dobivamo ovu ocjenu apsolutne vrijednosti pogreške

$$|R_1| \leq \frac{h^3}{12} M_2. \tag{11}$$

2. Trapezna formula.

To je po dijelovima primijenjena osnovna trapezna formula. Iz

$$\int_{x_0}^{x_n} f(x)dx = \int_{x_0}^{x_1} f(x)dx + \int_{x_1}^{x_2} f(x)dx + \dots + \int_{x_{n-1}}^{x_n} f(x)dx$$

imamo

$$\begin{aligned} \int_a^b f(x)dx &= \int_{x_0}^{x_n} f(x)dx \approx \frac{h}{2}[y_0 + 2y_1 + 2y_2 + \dots + 2y_{n-1} + y_n] \\ a = x_0, \quad b = x_n, \quad h &= \frac{b-a}{n}, \quad y_i = f(x_i). \end{aligned} \tag{12}$$

Pogreške po segmentima moramo zbrajati, pa imamo

$$\begin{aligned} R_T &= -\frac{h^3}{12} f''(\xi_1) - \frac{h^3}{12} f''(\xi_2) - \dots - \frac{h^3}{12} f''(\xi_n), \\ &= -\frac{h^3}{12} \sum_{i=1}^n f''(\xi_i), \quad x_{i-1} < \xi_i < x_i. \end{aligned}$$

Neprekidnost od f'' osigurava postojanje η , $x_0 < \eta < x_n$, za koji vrijedi

$$\sum_{i=1}^n f''(\xi_i) = n f''(\eta),$$

$f''(\eta)$ je srednja vrijednost od $f''(\xi_i)$ tako da je

$$R_T = -\frac{h^3}{12}nf''(\eta), \quad x_0 < \eta < x_n,$$

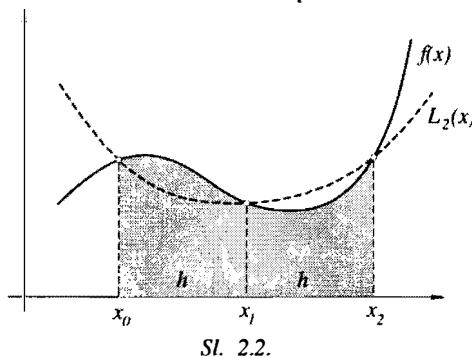
odnosno,

$$R_T = -\frac{(b-a)^3}{12n^2}f''(\eta), \quad x_0 < \eta < x_n. \quad (13)$$

Analogno je

$$|R_T| \leq \frac{(b-a)^3}{12n^2}M_2. \quad (14)$$

3. Osnovna Simpsonova formula.



Sl. 2.2.

To je slučaj $n = 2$, pa imamo

$$H_0 = \frac{1}{2} \cdot \frac{1}{2} \int_0^2 (t-1)(t-2)dt = \frac{1}{6}$$

$$H_1 = -\frac{1}{2} \cdot \frac{1}{1} \int_0^2 t(t-2)dt = \frac{2}{3},$$

$$H_2 = \frac{1}{2} \cdot \frac{1}{2} \int_0^2 t(t-1)dt = \frac{1}{6}.$$

Imamo dakle sljedeću aproksimaciju

$$\int_{x_0}^{x_2} f(x)dx \approx (x_2 - x_0) \left[\frac{1}{6}y_0 + \frac{2}{3}y_1 + \frac{1}{6}y_2 \right] = \frac{2h}{6}[y_0 + 4y_1 + y_2]$$

ili

$$\int_{x_0}^{x_2} f(x)dx \approx \frac{h}{3}[y_0 + 4y_1 + y_2]. \quad (15)$$

Provodeći analogan postupak kao u 1. primjeru, s time da bi tražili treću derivaciju i potom natrag integrirali dobiva se uz pretpostavku da f ima neprekidnu četvrtu derivaciju ($f \in C^4$) ova formula

$$R_2 = -\frac{h^5}{90}f^{(IV)}(\xi) \quad x_0 < \xi < x_2. \quad (16)$$

Treba uočiti da je eksponent od h porastao za 2. Stoga je osnovna Simpsonova formula točnija, a kako račun nije složen isto je razlog da je jedna od najčešće korištenih.

Ako je f bila polinom ≤ 3 (uočimo $\leq 3!$) onda je $R_2 = 0$.

4. Simpsonova formula.

To je po dijelovima primijenjena osnovna Simpsonova formula. Istaknimo da za njenu primjenu broj n mora biti paran.

$$a = x_0, \quad b = x_n, \quad h = \frac{b-a}{n} \quad \text{i} \quad n = 2m \text{ (paran).}$$

Tada iz

$$\int_{x_0}^{x_n} f(x)dx = \int_{x_0}^{x_2} f(x)dx + \int_{x_2}^{x_4} f(x)dx + \dots + \int_{x_{n-2}}^{x_n} f(x)dx$$

imamo

$$\int_a^b f(x)dx = \int_{x_0}^{x_n} f(x)dx \approx \frac{h}{3}[(y_0+y_n)+4(y_1+y_3+\dots+y_{n-1})+2(y_2+y_4+\dots+y_{n-2})]. \quad (17)$$

Pogreška je jednaka

$$R_S = -\frac{h^5}{90} \sum_{k=1}^m f^{(IV)}(\xi_k) \quad x_{2k-2} < \xi_k < x_{2k}, \quad (18)$$

tako da služeći se neprekidnošću od $f^{(IV)}$ nalazimo $x_0 < \eta < x_n$ za koji je $f(\eta)$ srednja vrijednost od $f(\xi_k)$ tj.

$$f^{(IV)}(\eta) = \frac{1}{m} \sum_{k=1}^m f^{(IV)}(\xi_k),$$

pa imamo

$$R_s = -\frac{mh^5}{90} f^{(IV)}(\eta) = -\frac{m}{\frac{2m}{90}} \cdot h^4 f^{(IV)}(\eta),$$

$$R_s = -\frac{(b-a)h^4}{180} f^{(IV)}(\eta), \quad (19)$$

odnosno

$$|R_s| \leq \frac{(b-a)h^4}{180} M_4$$

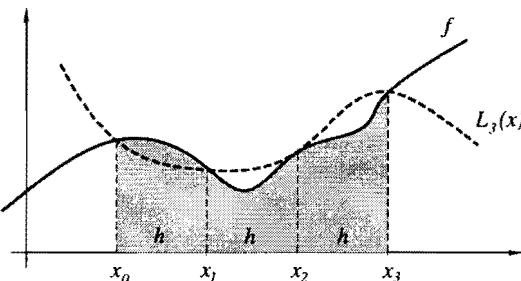
gdje je $|f^{(IV)}(x)| \leq M_4$ na $[x_0, x_n]$.

5. Druga osnovna Simpsonova formula.

Ta se formula dobiva za $n = 3$.

$$\int_{x_0}^{x_3} f(x)dx \approx \frac{3h}{8}[y_0 + 3y_1 + 3y_2 + y_3] \quad (20)$$

$$R_3 = -\frac{3h^5}{80} f^{(IV)}(\xi) \quad (21)$$



Sl. 2.3.

Osnovna formula za $n = 4$.

$$\int_{x_0}^{x_4} f(x)dx \approx \frac{2h}{45}[7y_0 + 32y_1 + 12y_2 + 32y_3 + 7y_4] \quad (22)$$

$$R_4 = -\frac{8h^7}{945}f^{(VI)}(\xi) \quad (23)$$

Ako se primjeni po dijelovima, onda n mora biti oblika $n = 3m$.

Osnovna formula za $n = 5$.

$$\int_{x_0}^{x_5} f(x)dx \approx \frac{5h}{288}[19y_0 + 75y_1 + 50y_2 + 50y_3 + 75y_4 + 19y_5] \quad (24)$$

$$R_5 = -\frac{275h^7}{12096}f^{(VI)}(\xi) \quad (25)$$

Osnovna formula za $n = 6$.

$$\int_{x_0}^{x_6} f(x)dx \approx \frac{h}{140}[41y_0 + 216y_1 + 27y_2 + 272y_3 + 27y_4 + 216y_5 + 41y_6] \quad (26)$$

$$R_6 = -\frac{9h^9}{1400}f^{(VIII)}(\xi) \quad (27)$$

Formule višeg reda imaju teorijski bolju točnost, računski nisu spretne jer su Cotesovi koeficijenti veliki, pa se računske pogreške akumuliraju. Stoga se najčešće koristi Simpsonova formula (17).

Promatrane formule za aproksimaciju integrala imaju sljedeći oblik

$$\int_a^b f(x)dx = \sum_{i=1}^n A_i f(x_i) + R$$

gdje su x_i dani čvorovi, a koeficijente A_i moramo izračunati i R je pogreška.

Različite formule imaju različitu točnost. Uzmimo da promatramo ekvidistantne čvorove, s istim brojem čvorova n , tj. $h = \frac{b-a}{n}$. Usپoredimo pogreške trapezne i Simpsonove formule. Imamo

$$R_T = -\frac{b-a}{12} h^2 f''(\eta)$$

i

$$R_S = -\frac{b-a}{180} h^4 f^{(IV)}(\eta).$$

Usپoredimo li potencije od h imamo h^2 odnosno h^4 , pa formulu smatramo točnijom što je viša potencija od h , prema čemu je Simpsonova formula točnija, što dolazi do izražaja pri dovoljno malom h .

Istaknimo da pri fiksiranom broju čvorova točnost bitno ovisi o rasporedu čvorova pa i manje točna formula može dati točniji rezultat.

Primjer 2.1. $\int_{-1}^1 (-8 + 45x^2 - 25x^4) dx = 4.$

Za $h = 1$ trapezna formula daje

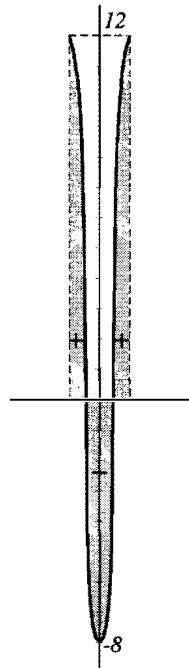
$$I_T = \frac{1}{2}f(-1) + f(0) + \frac{1}{2}f(1) = 6 - 8 + 6 = 4$$

(točan rezultat),

dok Simpsonova daje

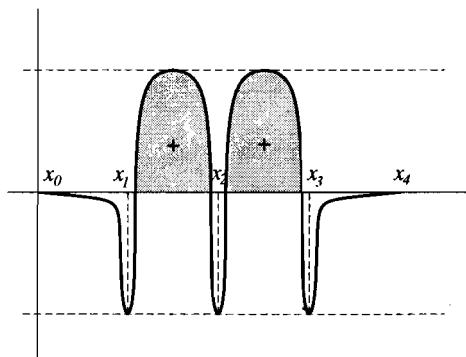
$$\begin{aligned} I_S &= \frac{1}{3}[f(-1) + 4f(0) + f(1)] \\ &= \frac{1}{3}(12 - 32 + 12) \\ &= -\frac{8}{3}, \end{aligned}$$

pa niti ne osigurava predznak integrala.



Sl. 2.4.

Uočimo da pri lošem rasporedu čvorova možemo dobiti potpuno krive rezultate što ilustriramo slikom.



Sl. 2.5.

Za funkciju na slici je $\int_{x_0}^{x_1} f(x)dx > 0$, no i trapezna i Simpsonova formula će dati negativne vrijednosti jer su $f(x_i) \leq 0$, a koeficijenti A_i u tim formulama su ≥ 0 . Pojavu s gornje slike imamo kada podintegralna funkcija ima puno nultočaka, ili kada ima puno ekstrema (tada prva derivacija ima puno nula) što povećava vrijednosti viših derivacija čime se smanjuje točnost aproksimacije. Stoga h treba birati tako da je znatno manji od razmaka susjednih nultočaka funkcije i razmaka susjednih nultočaka njene derivacije f' .

2.2. Ponavljanje raspolažljivanja i Rombergov algoritam

Vratimo se trapeznoj formuli

$$\int_a^b f(x)dx = \int_{x_0}^{x_n} f(x)dx = \frac{h}{2} [y_0 + 2y_1 + \dots + 2y_{n-1} + y_n] - \frac{(b-a)^3}{12n^2} f''(\eta).$$

Kako je pogreška proporcionalna s $\frac{1}{n^2}$, to raspolažljanjem očekujemo njeni smanjenje ($f''(\eta)$ je općenito različito za različite η). Prepostavimo da smo proveli račun za $n = n_1$ i $n = n_2$, $n_1 \neq n_2$. Tada je

$$I = I_{n_1} + R_{n_1} = I_{n_2} + R_{n_2} \quad (1)$$

gdje su I_{n_i} aproksimacije trapeznom formulom, a R_{n_i} pogreške, $i = 1, 2$. Sada je

$$\frac{R_{n_2}}{R_{n_1}} = \frac{-\frac{(b-a)^3}{12n_2^2} f''(\eta_2)}{-\frac{(b-a)^3}{12n_1^2} f''(\eta_1)}, \quad \eta_1, \eta_2 \in (a, b).$$

Pretpostavivši da su $f''(\eta_1)$ i $f''(\eta_2)$ jednake imamo

$$R_{n_2} = \left(\frac{n_1}{n_2} \right)^2 R_{n_1},$$

pa supstitucijom u drugu jednakost od (1) imamo

$$\begin{aligned} I_{n_1} + R_{n_1} &= I_{n_2} + \left(\frac{n_1}{n_2}\right)^2 R_{n_1} \\ R_{n_1} \left(1 - \left(\frac{n_1}{n_2}\right)^2\right) &= I_{n_2} - I_{n_1}, \end{aligned}$$

ili

$$R_{n_1} = \frac{I_{n_2} - I_{n_1}}{1 - \left(\frac{n_1}{n_2}\right)^2}.$$

Uvrstimo li dobiveno u prvu jednakost od (1) dobivamo

$$I \approx I_{n_1} + \frac{I_{n_2} - I_{n_1}}{1 - \left(\frac{n_1}{n_2}\right)^2}. \quad (2)$$

Ako je nadalje $n_2 = 2n_1$ dobivamo

$$I \approx \frac{4}{3}I_{n_2} - \frac{1}{3}I_{n_1}. \quad (3)$$

Prisjetimo se da smo u računu pretpostavili $f''(\eta_1) = f''(\eta_2)$ pa stoga očekujemo da je

$$\frac{4}{3}I_{n_2} - \frac{1}{3}I_{n_1} \approx I$$

poboljšana aproksimacija integrala I . To zovemo **Richardsonova ekstrapolacija**.

Kod raspolavljanja segmenta $[a, b]$ korak h ima oblik $\frac{b-a}{2^N}$, a broj čvorova iznosi $2^N + 1$. Krenuvši od $[a, b]$ provedimo rekurzivno raspolavljanje. Uvedemo li prikladne označke kao

$$\begin{aligned} T_{0,1} &= \frac{b-a}{1} \left\{ \frac{1}{2} [f(a) + f(b)] \right\} \\ T_{1,1} &= \frac{b-a}{2} \left\{ \frac{1}{2} [f(a) + f(b)] + f\left(a + \frac{b-a}{2}\right) \right\} \\ &= \frac{1}{2} \left\{ T_{0,1} + (b-a) f\left(a + \frac{b-a}{2}\right) \right\} \\ T_{2,1} &= \frac{b-a}{4} \left\{ \frac{1}{2} [f(a) + f(b)] + \sum_{i=1}^3 f\left(a + \frac{b-a}{4}i\right) \right\} \\ &= \frac{1}{2} \left\{ T_{1,1} + \frac{(b-a)}{2} \sum_{i=1}^3 f\left(a + \frac{b-a}{4}i\right) \right\} \end{aligned} \quad (4)$$

$$\begin{aligned}
 T_{3,1} &= \frac{b-a}{8} \left\{ \frac{1}{2}[f(a) + f(b)] + \sum_{i=1}^7 f\left(a + \frac{b-a}{8}i\right) \right\} \\
 &= \frac{1}{2} \left\{ T_{2,1} + \frac{(b-a)}{4} \sum_{\substack{i=1 \\ \Delta i=2}}^7 f\left(a + \frac{b-a}{8}i\right) \right\} \\
 T_{4,1} &= \frac{1}{2} \left\{ T_{3,1} + \frac{(b-a)}{8} \sum_{\substack{i=1 \\ \Delta i=2}}^{15} f\left(a + \frac{b-a}{16}i\right) \right\} \\
 &\vdots
 \end{aligned}$$

dobivamo rekurzivni zapis pripadnih trapeznih formula. Tako je $T_{0,1}$ trapezna formula za segment $[a, b]$, $T_{1,1}$ je trapezna formula kada $[a, b]$ raspolovimo itd..

$\Delta i = 2$ ispod znaka sume znači da indeks sumacije skače za 2, npr. u $T_{4,1}$ poprima vrijednosti 1, 3, 5, 7, 9, 11, 13, 15 jer su pribrojnici s indeksima 2, 4, 6, 8, 10, 12, 14 ujedno pribrojnici od $T_{3,1}$. Izvedite sami formule za $T_{1,1}$ i $T_{2,1}$.

Induktivno vrijedi

$$T_{N,1} = \frac{1}{2} \left\{ T_{N-1,1} + \frac{b-a}{2^{N-1}} \sum_{\substack{i=1 \\ \Delta i=2}}^{2^N-1} f\left(a + \frac{b-a}{2^N}i\right) \right\} \quad (5)$$

s pogreškom

$$R_{T_{N,1}} = -\frac{(b-a)^3}{12 \cdot 2^{2N}} f''(\xi).$$

Primjenimo li Richardsonovu ekstrapolaciju na par $T_{N,1}$ i $T_{N+1,1}$ nadamo se poboljšanoj aproksimaciji. Označimo tu aproksimaciju s $T_{N,2}$, tako da imamo

$$T_{N,2} = \frac{4T_{N+1,1} - T_{N,1}}{3}.$$

Npr.

$$T_{0,2} = \frac{4T_{1,1} - T_{0,1}}{3} = \frac{b-a}{6} \left\{ f(a) + 4f\left(\frac{a+b}{2}\right) + f(b) \right\}$$

što je elementarna Simpsonova formula. Greška te formule je proporcionalna s $\frac{1}{n^4}$ o čemu treba voditi računa kod ponovljene Richardsonove ekstrapolacije. Nastavimo li niz

$$T_{0,2}, T_{1,2}, T_{2,2}, T_{3,2}, \dots, T_{N,2}$$

dobivamo Simpsonovu formulu za $n = 2^N$ nakon što smo N puta raspolovili segment $[a, b]$. Time je $h = \frac{b-a}{2^{N+1}}$ pa je pogreška aproksimacije jednaka

$$R_{T_{N,2}} = -\frac{(b-a)^5}{2880 \cdot 2^{4N}} f^{(IV)}(\eta).$$

Sada ponovo možemo primijeniti Richardsonovu ekstrapolaciju na par susjednih članova niza $T_{0,2}, T_{1,2}, T_{2,2}, \dots$ dobivši novi niz aproksimacija $T_{0,3}, T_{1,3}, T_{2,3}, T_{3,3}, \dots, T_{N,3}$. Zbog toga što je sadašnja pogreška proporcionalna s $\frac{1}{n^4}$ imamo sljedeću formulu

$$T_{N,3} = \frac{16T_{N+1,2} - T_{N,2}}{15}.$$

Sagradićmo li niz $T_{0,3}, T_{1,3}, \dots$ možemo pokazati da je $T_{N,3}$ po dijelovima primijenjena elementarna Newton – Cotesova osnovna formula za $n = 4$, pa je pogreška takve aproksimacije oblika

$$-\frac{(b-a)^7}{1935360 \cdot 2^{4N}} f^{(VI)}(\eta).$$

Ponovimo li opet Richardsonovu ekstrapolaciju na nizu $T_{0,3}, T_{1,3}, \dots$ dobivamo

$$T_{N,4} = \frac{64T_{N+1,3} - T_{N,3}}{63}.$$

Induktivno imamo opću ekstrapolacijsku formulu oblika

$$T_{N,j} = \frac{4^{j-1}T_{N+1,j-1} - T_{N,j-1}}{4^{j-1} - 1} \quad (6)$$

koja se pripisuje Rombergu. Vidimo da $T_{N,j}$ ima dva indeksa pa navedimo što se događa kada jedan od njih teži u beskonačno. Jasno nam je da bi povećanjem N trebali kao limes dobiti pravu vrijednost integrala. To se može dokazati, tako da vrijedi

$$\lim_{N \rightarrow \infty} T_{N,j} = \int_a^b f(x) dx.$$

Nadalje, niz

$$T_{0,1}, T_{0,2}, T_{0,3}, \dots$$

također konvergira k pravoj vrijednosti integrala, tj.

$$\lim_{j \rightarrow \infty} T_{0,j} = \int_a^b f(x) dx.$$

Početno razmatranje u konstrukciji Rombergovih nizova može dati krivu sugestiju da su nizovi $T_{N,j}$ po dijelovima primijenjene Newton – Cotesove aproksimacije. To je istina za $j \leq 3$, dok za $j > 3$ to nije slučaj.

Rombergove nizove možemo složiti u jednostavnu tablicu u kojoj je N indeks retka, a j indeks stupca.

$T_{0,1}$	$T_{0,2}$	$T_{0,3}$	$T_{0,4}$	$T_{0,5}$	\dots	$T_{0,j}$
$T_{1,1}$	$T_{1,2}$	$T_{1,3}$	$T_{1,4}$	$T_{1,5}$		
$T_{2,1}$	$T_{2,2}$	$T_{2,3}$				
\vdots	\vdots	\vdots				
$T_{N-2,1}$	$T_{N-2,2}$	$T_{N-2,3}$				
$T_{N-1,1}$	$T_{N-1,2}$					
$T_{N,1}$						

Rombergova tablica

U praktičnoj primjeni ove tehnike prvo izračunamo elemente prvog stupca pomoću formula (4) i (5). Zatim popunimo trokutnu tablicu služeći se formulom (6). Kako svaki stupac i redak konvergira prema vrijednosti integrala to na dijagonali trokutne tablice dobivamo seriju aproksimacija.

Što se tiče pogreške koja odgovara aproksimaciji

$$\int_a^b f(x)dx \approx T_{N,j}$$

može se pokazati (u što se ne upuštamo) da je jednaka

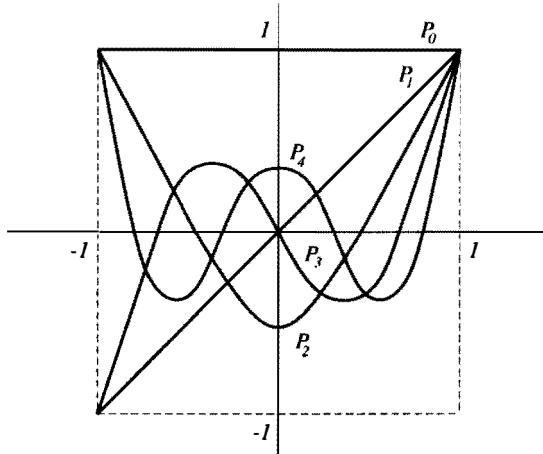
$$R_{N,j} = \frac{k(j)}{2^{2jN}} f^{(ij)}(\eta), \quad a < \eta < b$$

gdje je $k(j)$ konstanta koja ovisi o a , b i j , a ne ovisi o N .

2.3. Gaussove formule

Ekvidistantnost čvorova (jednolika podjela segmenta $[a, b]$) bitan su uvjet Newton – Cotesovih formula. Dopustimo sada nejednoliku podjelu i upoznajmo tzv. **Gaussove formule** za numeričku integraciju. Da to provedemo moramo se prisjetiti tzv. **Legendreovih polinoma**. To su polinomi oblika

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n, \quad n = 0, 1, 2, \dots$$



Sl. 2.6.

Prvih pet polinoma glasi

$$P_0(x) = 1$$

$$P_1(x) = x$$

$$P_2(x) = \frac{1}{2}(3x^2 - 1)$$

$$P_3(x) = \frac{1}{2}(5x^3 - 3x)$$

$$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3)$$

Te smo polinome sreli kod Fourierovih redova kao primjer niza ortogonalnih funkcija na $[-1, 1]$, tj. vrijedi

$$\int_{-1}^1 P_n(x)P_m(x)dx = 0, \quad n \neq m,$$

ali vrijedi i više, naime vrijedi da je

$$\int_{-1}^1 P_n(x)Q_k(x)dx = 0 \quad k < n, \quad (1)$$

gdje je Q_k proizvoljan polinom stupnja nižeg od n , tj. $k < n$. Daljnja svojstva Legendreovih polinoma:

$$P_n(1) = 1, \quad P_n(-1) = (-1)^n \quad n = 0, 1, \dots, \\ P_n \text{ ima na } (-1, 1) n \text{ jednostrukih korijena.}$$

Označimo nezavisnu varijablu s t i provedimo razmatranje na $[-1, 1]$. (Kasnije ćemo sve transformirati na $[a, b]$.) Postavimo si zadatak da nađemo čvorove t_1, t_2, \dots, t_n i koeficijente A_1, A_2, \dots, A_n tako da formula za aproksimaciju integrala izgleda ovako

$$\int_{-1}^1 f(t)dt \approx \sum_{i=1}^n A_i y_i, \quad \text{gdje je } y_i = f(t_i) \quad (2)$$

i da daje točne vrijednosti za sve polinome do odgovarajućeg mogućeg stupnja N , koji ćemo odmah odrediti.

Kako u zadatku treba odrediti n vrijednosti t_i i n vrijednosti A_i , što je ukupno $2n$ vrijednosti, a polinom $(2n-1)$ -og stupnja je određen s $2n$ koeficijenata to zaključujemo da je u općem slučaju $N = 2n-1$. Dovoljno je ograničiti se na polinome

$$f(t) = 1, t, t^2, \dots, t^{2n-1} \quad (3)$$

jer su ostali polinomi stupnja $\leq 2n-1$ njihove linearne kombinacije. Time za prave vrijednosti polinoma u (3) imamo da vrijedi

$$\int_{-1}^1 t^k dt = \sum_{i=1}^n A_i t_i^k, \quad k = 0, 1, 2, \dots, 2n-1. \quad (4)$$

Kako je

$$\int_{-1}^1 t^k dt = \frac{1 - (-1)^{k+1}}{k+1} = \begin{cases} \frac{2}{k+1} & \text{ako je } k \text{ paran} \\ 0 & \text{ako je } k \text{ neparan} \end{cases} \quad (5)$$

to dobivamo ovakav sustav jednadžbi u t_i i A_i

$$\begin{aligned} \sum_{i=1}^n A_i &= 2 \\ \sum_{i=1}^n A_i t_i &= 0 \\ &\dots \\ \sum_{i=1}^n A_i t_i^{2n-2} &= \frac{2}{2n-1} \\ \sum_{i=1}^n A_i t_i^{2n-1} &= 0 \end{aligned} \tag{6}$$

što je nelinearan sustav jednadžbi. Svedimo problem rješavanja sustava (6) na rješavanje linearog sustava u nepoznanice A_i posluživši se Legendreovim polinomom $P_n(x)$.

Promatrajmo polinome

$$Q_{n+k}(t) = t^k P_n(t) \quad k = 0, 1, 2, \dots, n-1 \tag{7}$$

gdje je $P_n(t)$ Legendreov polinom n -tog stupnja. Stupanj polinoma u (7) nije veći od $2n-1$, pa za njih u (2) vrijedi jednakost, tj. imamo

$$\int_{-1}^1 Q_{n+k}(t) dt = \int_{-1}^1 t^k P_n(t) dt = \sum_{i=1}^n A_i t_i^k P_n(t_i), \quad k = 0, 1, \dots, n-1.$$

Sada se poslužimo svojstvom (1) koje kaže da je

$$\int_{-1}^1 t^k P_n(t) dt = 0 \quad \text{za } k < n, \quad \text{tj. za } k = 0, 1, \dots, n-1$$

tako da uvrštavanjem u (2) imamo

$$\sum_{i=1}^n A_i t_i^k P_n(t_i) = 0, \quad k = 0, 1, \dots, n-1. \tag{8}$$

Iz $P_n(t_i) = 0$ slijedi da su jednadžbe (8) zadovoljene za proizvoljne vrijednosti A_i . Ali

$$P_n(t_i) = 0, \quad i = 1, 2, \dots, n \tag{9}$$

kaže da su t_i nultočke (korijeni) Legendreovog polinoma n -tog stupnja koje su različite, realne i nalaze se unutar $(-1, 1)$. Uvrstivši tako određene t_i u prvih n jednadžbi sustava (6), dobivamo ovaj linearan sustav jednadžbi s nepoznanicama A_i

$$\begin{aligned} \sum_{i=1}^n A_i &= 2 \\ \sum_{i=1}^n t_i A_i &= 0 \end{aligned}$$

$$\begin{aligned}
 \sum_{i=1}^n t_i^2 A_i &= \frac{2}{3} \\
 \sum_{i=1}^n t_i^3 A_i &= 0 \\
 &\dots \\
 \sum_{i=1}^n t_i^k A_i &= \begin{cases} \frac{2}{k+1} & k \text{ paran} \\ 0 & k \text{ neparan} \end{cases} \\
 &\dots \\
 \sum_{i=1}^n t_i^{n-1} A_i &= \begin{cases} \frac{2}{n} & k \text{ neparan} \\ 0 & k \text{ paran.} \end{cases}
 \end{aligned} \tag{10}$$

Determinanta podsustava od prvih n jednadžbi je Vandermondeova determinanta pa rješenjem sustava (10) dobivamo **Gaussove formule** za numeričku integraciju koje glase

$$\int_{-1}^1 f(x)dx \approx \sum_{i=1}^n A_i f(t_i) \tag{11}$$

gdje su A_i rješenja sustava (10) u kome su čvorovi t_i korjeni Lagrangeovog polinoma n -tog stupnja.

Linearna transformacija

$$x = \frac{b+a}{2} + \frac{b-a}{2}t \tag{12}$$

prevodi $[-1, 1]$ na $[a, b]$, pa imamo

$$\int_a^b f(x)dx = \frac{b-a}{2} \int_{-1}^1 f\left(\frac{b+a}{2} + \frac{b-a}{2}t\right)dt$$

tako da dobivamo

$$\int_a^b f(x)dx \approx \frac{b-a}{2} \sum_{i=1}^n f(x_i)A_i \tag{13}$$

gdje je $x_i = \frac{b+a}{2} + \frac{b-a}{2}t_i$, $i = 1, 2, \dots, n$, a t_i su korjeni (nule) Legendreovog polinoma n -tog stupnja.

Pogreška aproksimacije se može dovesti u ovaj oblik

$$R_n = \frac{(b-a)^{2n+1}(n!)^4 f^{(2n)}(\xi)}{[(2n)!]^3(2n+1)}. \tag{14}$$

Nedostatak Gaussovih formula je iracionalnost korijena t_i , dok se s druge strane ističu visokom točnošću.

Primjer 2.2. $n = 3$.

Tada je

$$P_3(t) = \frac{1}{2}(5t^3 - 3t)$$

čije su nultočke

$$t_1 = -\sqrt{\frac{3}{5}}$$

$$t_2 = 0$$

$$t_3 = \sqrt{\frac{3}{5}}$$

pa za $n = 3$ sustav (10) ima ovaj oblik

$$A_1 + A_2 + A_3 = 2$$

$$-\sqrt{\frac{3}{5}}A_1 + \sqrt{\frac{3}{5}}A_3 = 0$$

$$\frac{3}{5}A_1 + \frac{3}{5}A_3 = \frac{2}{3}$$

čije rješenje glasi $A_1 = A_3 = \frac{5}{9}$, $A_2 = \frac{8}{9}$. Imamo dakle formulu

$$\int_{-1}^1 f(t)dt \approx \frac{1}{9} \left[5f\left(-\sqrt{\frac{3}{5}}\right) + 8f(0) + 5f\left(\sqrt{\frac{3}{5}}\right) \right].$$

2.4. Integriranje brzo oscilirajućih funkcija. Filonova formula

Provđimo prvo jedno opće razmatranje. Neka je potrebno izračunati

$$\int_a^b f(x)p(x)dx. \quad (1)$$

Mislimo li na $\int_a^b f(x)dx$ imamo da je u (1) $p(x) = 1$. No, u produktu možemo aproksimirati jedan faktor i time očekivati da smo aproksimirali produkt. U takvom postupku funkcija $p(x)$ se zove **težinska funkcija**.

Kao što smo kod Gaussovih formula proveli razmatranje na $[-1, 1]$ i zatim sve transformirali na $[a, b]$ sa

$$x = \frac{b+a}{2} + \frac{b-a}{2}t \quad (2)$$

tako to možemo i sada učiniti. Pokazat će se na kraju da smo kod Gaussovih formula imali Legendreov polinom kao težinsku funkciju.

Neka su $t_0, t_1, t_2, \dots, t_n \in [-1, 1]$ točke koje odgovaraju točkama $x_0, x_1, x_2, \dots, x_n \in [a, b]$ po transformaciji (2). Aproksimirajmo f na $[a, b]$ s L_n koji prolazi čvorovima x_i , $i = 0, 1, \dots, n$. Dakle, na $[a, b]$ je

$$L_n(x) \approx f(x),$$

pa imamo

$$\int_a^b f(x)p(x)dx \approx \int_a^b L_n(x)p(x)dx. \quad (3)$$

Pogreška iznosi

$$\begin{aligned} R_n &= \int_a^b f(x)p(x)dx - \int_a^b L_n(x)p(x)dx \\ &= \int_a^b p(x)[f(x) - L_n(x)]dx. \end{aligned}$$

Kako znamo ocijeniti razliku $f(x) - L_n(x)$,

$$f(x) - L_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)(x-x_1)\dots(x-x_n),$$

to dobivamo

$$R_n = \int_a^b p(x) \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)(x-x_1)\dots(x-x_n)dx, \quad (4)$$

iz čega imamo ovakvu ocjenu absolutne vrijednosti pogreške

$$|R_n| \leq \max_{x \in [a,b]} |f^{(n+1)}(\xi)| \int_a^b \frac{|p(x)|(x-x_0)\dots(x-x_n)}{(n+1)!} dx. \quad (5)$$

Provđemo li supstituciju (2) u formuli za pogrešku dobivamo

$$\begin{aligned} \frac{1}{(n+1)!} \int_a^b p(x)(x-x_0)\dots(x-x_n)dx &= \frac{1}{(n+1)!} \int_{-1}^1 p\left(\frac{b+a}{2} + \frac{b-a}{2}t\right) \\ &\cdot (t-t_0)\dots(t-t_n) \left(\frac{b-a}{2}\right)^{n+2} dt, \end{aligned}$$

jer je $x - x_i = \frac{b+a}{2} + \frac{b-a}{2}t - \left(\frac{b+a}{2} + \frac{b-a}{2}t_i\right) = \frac{b-a}{2}(t-t_i)$ i $dx = \frac{b-a}{2}dt$.

Označimo li sa

$$D(t_0, t_1, \dots, t_n) = \int_{-1}^1 \frac{\left|p\left(\frac{b+a}{2} + \frac{b-a}{2}t\right)(t-t_0)\dots(t-t_n)\right|}{(n+1)!} dt$$

imamo ovaj oblik absolutne vrijednosti pogreške

$$|R_n| \leq D(t_0, t_1, \dots, t_n) \left|f^{(n+1)}(\xi)\right| \left(\frac{b-a}{2}\right)^{n+2}. \quad (6)$$

U Lagrangeovom obliku L_n glasi

$$L_n(x) = \sum_{i=0}^n f(x_i) \prod_{\substack{j=0 \\ j \neq i}}^n \frac{x - x_j}{x_i - x_j}.$$

Supstitucijom (2) u integral (3) imamo

$$\begin{aligned} \int_a^b p(x)L_n(x)dx &= \int_{-1}^1 p\left(\frac{b+a}{2} + \frac{b-a}{2}t\right) \sum_{i=0}^n f(x_i) \prod_{\substack{j=0 \\ j \neq i}}^n \frac{\frac{b-a}{2}(t-t_j)}{\frac{b-a}{2}(t_i-t_j)} \cdot \frac{b-a}{2} dt \\ &= \frac{b-a}{2} \sum_{i=0}^n f(x_i) \int_{-1}^1 p\left(\frac{b+a}{2} + \frac{b-a}{2}t\right) \prod_{\substack{j=0 \\ j \neq i}}^n \frac{t-t_j}{t_i-t_j} dt. \end{aligned}$$

Dakle

$$\int_a^b p(x)L_n(x)dx = \frac{b-a}{2} \sum_{i=0}^n f(x_i) D_i \quad (7)$$

gdje je

$$D_i = \int_{-1}^1 p\left(\frac{b+a}{2} + \frac{b-a}{2}t\right) \prod_{\substack{j=0 \\ j \neq i}}^n \frac{t-t_j}{t_i-t_j} dt, \quad i = 0, 1, 2, \dots, n. \quad (8)$$

Time dobivamo aproksimaciju integrala u obliku

$$\int_a^b f(x)p(x)dx \approx \frac{b-a}{2} \sum_{i=0}^n f(x_i) D_i. \quad (9)$$

Spomenimo bez dokaza sljedeće:

Ako je p parna obzirom na polovište od $[a, b]$ i ako su čvorovi x_i simetrično raspoređeni oko polovišta, odnosno $t_i = -t_{n-i}$, onda vrijedi

$$D_i = D_{n-i}.$$

Neka je potrebno izračunati integral (Fourierova transformacija)

$$\int_a^b f(x)e^{i\omega x} dx, \quad e^{i\omega x} = \cos \omega x + i \sin \omega x \quad (10)$$

pri čemu je $\omega(b-a) \gg 1$ (mnogo veći od) i f glatka funkcija. Sada funkcije

$$\begin{aligned} \operatorname{Re}\left(f(x)e^{i\omega x}\right) &= f(x) \cos \omega x \\ \operatorname{Im}\left(f(x)e^{i\omega x}\right) &= f(x) \sin \omega x \end{aligned}$$

imaju na $[a, b]$ puno nula, naime $\frac{\omega(b-a)}{\pi}$ je broj nula na $[a, b]$, pa ako je $\omega(b-a) \gg 1$, dijeljenje s π neće to bitno smanjiti. Za $f(x)e^{i\omega x}$, uz uvjet $\omega(b-a) \gg 1$, kažemo da je **brzo oscilirajuća** na $[a, b]$. Kako polinom n -tog stupnja ima najviše n različitih nula to se takva funkcija može dobro aproksimirati polinomom n -tog stupnja samo ako je

$$n \gg \frac{\omega(b-a)}{\pi}.$$

Da to izbjegnemo možemo postupiti tako da funkciju $e^{i\omega x}$ shvatimo kao težinsku funkciju. Prema (7) imamo

$$\begin{aligned} \int_a^b L_n(x) e^{i\omega x} dx &= \frac{b-a}{2} \sum_{i=0}^n f(x_i) D_i \\ D_i &= \int_{-1}^1 e^{i\omega \left(\frac{b+a}{2} + \frac{b-a}{2} t\right)} \prod_{\substack{j=0 \\ j \neq i}}^n \frac{t - t_j}{t_i - t_j} dt \\ &= e^{i\omega \frac{a+b}{2}} \int_{-1}^1 e^{i\left(\omega \frac{b-a}{2}\right)t} \prod_{\substack{j=0 \\ j \neq i}}^n \frac{t - t_j}{t_i - t_j} dt, \end{aligned}$$

pa gornja aproksimacija poprima oblik

$$\int_a^b f(x) e^{i\omega x} dx \approx \frac{b-a}{2} e^{i\omega \frac{a+b}{2}} \sum_{i=0}^n f(x_i) A_i(p), \quad (11)$$

gdje je

$$A_i(p) = \int_{-1}^1 e^{ip t} \prod_{\substack{j=0 \\ j \neq i}}^n \frac{t - t_j}{t_i - t_j} dt, \quad p = \omega \frac{b-a}{2}. \quad (12)$$

Pogreška je dana s

$$|R_n| \leq D(t_0, t_1, \dots, t_n) \left| f^{(n+1)}(\xi) \right| \left(\frac{b-a}{2} \right)^{n+2} \quad (13)$$

gdje zbog $|e^{i\omega x}| = 1$, $D(t_0, t_1, \dots, t_n)$ poprima oblik

$$D(t_0, t_1, \dots, t_n) = \int_{-1}^1 \frac{|(t-t_0)(t-t_1)\dots(t-t_n)|}{(n+1)!} dt. \quad (14)$$

U primjenama se susreće algoritam za $n = 2$, $t_0 = -1$, $t_1 = 0$, $t_2 = 1$. Pripadna formula se zove **Filonova formula**. U tu svrhu možemo izračunati

$$\begin{aligned} A_0(p) &= \int_{-1}^1 e^{ip t} \frac{(t-0)(t-1)}{(-1-0)(-1-1)} dt = \frac{1}{2} \int_{-1}^1 e^{ip t} (t^2 - t) dt \\ A_1(p) &= \int_{-1}^1 e^{ip t} \frac{(t+1)(t-1)}{(0+1)(0-1)} dt = - \int_{-1}^1 e^{ip t} (t^2 - 1) dt \\ A_2(p) &= \int_{-1}^1 e^{ip t} \frac{(t+1)(t-0)}{(1+1)(1-0)} dt = \frac{1}{2} \int_{-1}^1 e^{ip t} (t^2 + t) dt. \end{aligned}$$

Da to izračunamo trebamo redom:

$$\begin{aligned} \int_{-1}^1 e^{ip t} dt &= \frac{1}{ip} [e^{ip} - e^{-ip}] \\ \int_{-1}^1 e^{ip t} t dt &= \frac{1}{ip} [e^{ip} + e^{-ip}] + \frac{1}{p^2} [e^{ip} - e^{-ip}] \end{aligned}$$

$$\int_{-1}^1 e^{ip t} t^2 dt = \frac{1}{ip} \left(1 - \frac{2}{p^2} \right) [e^{ip} - e^{-ip}] + \frac{2}{p^2} [e^{ip} + e^{-ip}].$$

Uvrstimo li te rezultate imamo

$$\begin{aligned} A_0 &= \frac{1}{2} \left\{ \int_{-1}^1 e^{ip t} t^2 dt - \int_{-1}^1 e^{ip t} t dt \right\} \\ &= \frac{1}{2} \left\{ \frac{1}{ip} \left(1 - \frac{2}{p^2} \right) [e^{ip} - e^{-ip}] + \frac{2}{p^2} [e^{ip} + e^{-ip}] - \frac{1}{ip} [e^{ip} + e^{-ip}] - \frac{1}{p^2} [e^{ip} - e^{-ip}] \right\} \\ &= \frac{1}{2} \left\{ \frac{1}{ip} \left(1 - \frac{2}{p^2} \right) 2i \sin p + \frac{2}{p^2} 2 \cos p - \frac{1}{ip} 2 \cos p - \frac{1}{p^2} 2i \sin p \right\} \\ &= \frac{1}{p} \left[2 \frac{\cos p}{p} + \left(1 - \frac{2}{p^2} \right) \sin p \right] + i \frac{1}{p} \left[\cos p - \frac{\sin p}{p} \right] \\ &= E + iF, \end{aligned}$$

$$\begin{aligned} A_1 &= - \left\{ \int_{-1}^1 e^{ip t} t^2 dt - \int_{-1}^1 e^{ip t} dt \right\} \\ &= - \left\{ \frac{1}{ip} \left(1 - \frac{2}{p^2} \right) [e^{ip} - e^{-ip}] + \frac{2}{p^2} [e^{ip} + e^{-ip}] - \frac{1}{ip} [e^{ip} - e^{-ip}] \right\} \\ &= - \frac{2}{p^2} \left\{ - \frac{1}{ip} [e^{ip} - e^{-ip}] + [e^{ip} + e^{-ip}] \right\} \\ &= - \frac{2}{p^2} \left\{ - \frac{2}{p} \sin p + 2 \cos p \right\} = - \frac{4}{p^2} \left\{ \cos p - \frac{\sin p}{p} \right\} \\ &= - \frac{4}{p} F, \end{aligned}$$

$$\begin{aligned} A_2 &= \frac{1}{2} \left\{ \int_{-1}^1 e^{ip t} t^2 dt + \int_{-1}^1 e^{ip t} t dt \right\} \\ &= \frac{1}{2} \left\{ \frac{1}{ip} \left(1 - \frac{2}{p^2} \right) [e^{ip} - e^{-ip}] + \frac{2}{p^2} [e^{ip} + e^{-ip}] + \frac{1}{ip} [e^{ip} + e^{-ip}] + \frac{1}{p^2} [e^{ip} - e^{-ip}] \right\} \\ &= \frac{1}{2} \left\{ \frac{1}{ip} \left(1 - \frac{2}{p^2} \right) 2i \sin p + \frac{2}{p^2} 2 \cos p + \frac{1}{ip} 2 \cos p + \frac{1}{p^2} 2i \sin p \right\} \\ &= \frac{1}{p} \left[2 \frac{\cos p}{p} + \left(1 - \frac{2}{p^2} \right) \sin p \right] - i \frac{1}{p} \left[\cos p - \frac{\sin p}{p} \right] \\ &= E - iF, \end{aligned}$$

Uvršteno u (11) daje

$$\int_a^b f(x) e^{i\omega x} dx \approx \frac{b-a}{2} e^{i\omega \frac{b+a}{2}} \left[f(a)A_0 + f\left(\frac{a+b}{2}\right)A_1 + f(b)A_2 \right].$$

Uvrstimo li $\alpha = \omega \frac{b+a}{2}$, imamo $e^{i\alpha} = \cos \alpha + i \sin \alpha$ tako da je

$$\begin{aligned}
\int_a^b f(x)e^{i\omega x} dx &\approx \frac{b-a}{2} [\cos \alpha + i \sin \alpha] \left[f(a)(E + iF) \right. \\
&\quad \left. - f\left(\frac{a+b}{2}\right) \frac{4}{p} F + f(b)(E - iF) \right] \\
&= \frac{b-a}{2} \left\{ \left(f(a)E \cos \alpha + f(b)E \cos \alpha - f\left(\frac{a+b}{2}\right) \frac{4}{p} F \cos \alpha \right. \right. \\
&\quad \left. \left. - f(a)F \sin \alpha + f(b)F \sin \alpha \right) + i \left(f(a)E \sin \alpha + f(b)E \sin \alpha \right. \right. \\
&\quad \left. \left. - f\left(\frac{a+b}{2}\right) \frac{4}{p} F \sin \alpha + f(a)F \cos \alpha - f(b)F \cos \alpha \right) \right\} \\
&= \frac{b-a}{2} \left\{ f(a)[E \cos \alpha - F \sin \alpha] + f(b)[E \cos \alpha + F \sin \alpha] \right. \\
&\quad \left. - f\left(\frac{a+b}{2}\right) \frac{4}{p} F \cos \alpha \right\} + i \frac{b-a}{2} \left\{ f(a)[E \sin \alpha \right. \\
&\quad \left. + F \cos \alpha] + f(b)[E \sin \alpha - F \cos \alpha] \right. \left. - f\left(\frac{a+b}{2}\right) \frac{4}{p} F \sin \alpha \right\}
\end{aligned}$$

gdje je

$$\begin{aligned}
E &= \frac{1}{p} \left[2 \frac{\cos p}{p} + \left(1 - \frac{2}{p^2} \right) \sin p \right], \\
F &= \frac{1}{p} \left[\cos p - \frac{\sin p}{p} \right], \\
\alpha &= \omega \frac{a+b}{2}, \\
p &= \omega \frac{b-a}{2}.
\end{aligned}$$

Rastavimo li na realni i imaginarni dio imamo

$$\begin{aligned}
\int_a^b f(x) \cos \omega x dx &\approx \frac{b-a}{2} \left\{ f(a)[E \cos \alpha - F \sin \alpha] + f(b)[E \cos \alpha + F \sin \alpha] \right. \\
&\quad \left. - f\left(\frac{a+b}{2}\right) \frac{4}{p} F \cos \alpha \right\} \\
\int_a^b f(x) \sin \omega x dx &\approx \frac{b-a}{2} \left\{ f(a)[E \sin \alpha + F \cos \alpha] + f(b)[E \sin \alpha - F \cos \alpha] \right. \\
&\quad \left. - f\left(\frac{a+b}{2}\right) \frac{4}{p} F \sin \alpha \right\}
\end{aligned}$$

ili

$$\int_a^b f(x) \cos \omega x dx \approx \frac{b-a}{2} \left\{ [f(a) + f(b)]E \cos \alpha + [f(b) - f(a)]F \sin \alpha - f\left(\frac{a+b}{2}\right) \frac{4}{p} F \cos \alpha \right\}$$

$$\int_a^b f(x) \sin \omega x dx \approx \frac{b-a}{2} \left\{ [f(a) + f(b)]E \sin \alpha + [f(a) - f(b)]F \cos \alpha - f\left(\frac{a+b}{2}\right) \frac{4}{p} F \sin \alpha \right\}.$$

U primjenama se susrećemo i s formulama za $n = 4$, $t_0 = -1$, $t_1 = -0.5$, $t_2 = 0$, $t_3 = 0.5$, $t_4 = 1$.

3.

Sustavi linearnih jednadžbi

3.1. Direktne metode	66
3.2. Iteracijske metode	85

Promatramo opći sustav linearnih jednadžbi

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ &\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m, \end{aligned} \tag{1}$$

tj. m jednadžbi sa n nepoznanica. Matrični zapis istog glasi

$$Ax = b \tag{2}$$

gdje je $A = [a_{ij}]$ ($m \times n$)-matrica, x je n -komponentni vektor i b je m -komponentni vektor. Opća teorija egzistencije rješenja sustava linearnih jednadžbi kaže da pored matrice A treba promatrati proširenu matricu sustava, u oznaci B , koja se dobiva iz matrice A dodavanjem vektora b kao $(n+1)$ -vog stupca, dakle B izgleda ovako:

$$B = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{bmatrix} \tag{3}$$

te da treba usporediti njihove rangove. Krajnji rezultat glasi:

- 1) Ako je $\text{rang } A \neq \text{rang } B$, onda sustav nema rješenje;
- 2) Ako je $\text{rang } A = \text{rang } B$, onda sustav ima rješenje i to:
 - 2a) ako je $\text{rang } A = \text{rang } B = n = \text{broj nepoznanica}$, onda sustav ima jedinstveno rješenje;
 - 2b) ako je $\text{rang } A = \text{rang } B = r < n$, onda sustav ima beskonačno rješenja, naime, stanovitih $n - r$ nepoznanica možemo zadati nakon čega je ostalih r određeno. Tada se obično govori o $(n - r)$ -parametarskom rješenju sustava.

Nama su važna oba ova slučaja. Na prvi slučaj redovito nailazimo kada imamo sustav od n jednadžbi sa n nepoznanica za koji je determinanta matrice A različita od nule, $\det A \neq 0$, dok na drugi slučaj nailazimo kod linearog optimiziranja.

U ovom poglavlju se zadržavamo na sustavu od n jednadžbi sa n nepoznanica

$$Ax = b$$

za koji je

$$\det A \neq 0.$$

Tada je

$$x = A^{-1}b \quad (4)$$

rješenje sustava koje zahtijeva računanje inverzne matrice. Direktno računanje A^{-1} zahtijeva mnogo vremena pa se formula (4) rijetko koristi.

Nalaženje rješenja sustava (1) pomoću Cramerovog pravila

$$x_i = \frac{\det A_i}{\det A}, \quad i = 1, 2, \dots, n \quad (5)$$

gdje je

$$A_i = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1i-1} & b_1 & a_{1i+1} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2i-1} & b_2 & a_{2i+1} & \dots & a_{2n} \\ \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{ni-1} & b_n & a_{ni+1} & \dots & a_{nn} \end{bmatrix} \quad (6)$$

ima istu manu kao i nalaženje inverzne matrice jer treba računati vrijednosti determinata visokog reda. Stoga su razvijene različite metode rješavanja sustava linearnih jednadžbi koje su računski jednostavnije i zahtijevaju manje vremena. Neke od tih metoda su **direktne** (nekada govorimo točne), a druge su **iteracijske** koje daju aproksimaciju rješenju.

Ponovimo ovdje još sljedeće. Ako sustav pomnožimo s regularnom matricom C ($\det C \neq 0$), onda je novodobiveni sustav $CAx = Cb$ ekvivalentan polaznom. Ako usporedimo novi sustav s polaznim, onda je novi sustav dobiven iz polaznog linearnim kombiniranjem jednadžbi polaznog. Ako npr. želimo zamijeniti r -tu jednadžbu sa s -tom, onda to možemo postići množenjem sa matricom I_{rs} koja se dobiva iz jedinične matrice I zamjenom r -tog i s -tog retka.

Prvo ćemo proučiti neke direktne metode, a zatim neke iteracijske.

3.1. Direktne metode

Prvo ćemo opisati tzv. Gaussovnu metodu eliminacije čiju smo varijantu koristili još u srednjoj školi pod imenom metoda eliminacije. Obično je to bila tzv. Gauss-Jordanova metoda koju također opisuјemo.

Gaussova metoda eliminacije

Najraširenija direktna metoda ili algoritam je postupak postupnog eliminiranja nepoznanica. Radi lakšeg praćenja opisat ćemo to na sustavu 4 jednadžbe sa 4 nepoznanicama.

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 &= a_{15} \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 &= a_{25} \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 &= a_{35} \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 &= a_{45} \end{aligned} \quad (1)$$

Neka je $a_{11} \neq 0$. Podijelimo li prvu jednadžbu s a_{11} dobivamo

$$x_1 + b_{12}x_2 + b_{13}x_3 + b_{14}x_4 = b_{15}, \quad (2)$$

gdje je

$$b_{1j} = \frac{a_{1j}}{a_{11}}, \quad j > 1.$$

Koristeći jednadžbu (2) lako eliminiramo x_1 iz ostalih jednadžbi sustava (1). Treba jednadžbu (2) pomnožiti s a_{21} i odbiti od druge jednadžbe sustava (1), zatim jednadžbu (2) množimo s a_{31} i odbijemo od treće jednadžbe sustava (1), i tako redom. Time dobivamo sustav s jednom jednadžbom i jednom nepoznanicom manje:

$$\begin{aligned} a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 + a_{24}^{(1)}x_4 &= a_{25}^{(1)} \\ a_{32}^{(1)}x_2 + a_{33}^{(1)}x_3 + a_{34}^{(1)}x_4 &= a_{35}^{(1)} \\ a_{42}^{(1)}x_2 + a_{43}^{(1)}x_3 + a_{44}^{(1)}x_4 &= a_{45}^{(1)} \end{aligned} \quad (3)$$

gdje se koeficijenti $a_{ij}^{(1)}$ ($i, j \geq 2$) računaju prema formuli

$$a_{ij}^{(1)} = a_{ij} - a_{i1}b_{1j} \quad (i, j \geq 2).$$

Sada ponovimo postupak na sustavu (3) i eliminiramo nepoznanicu x_2 , tj. podijelimo prvu jednadžbu sustava (3) s $a_{22}^{(1)}$ (koji naravno treba biti $\neq 0$) i dobivamo

$$x_2 + b_{23}^{(1)}x_3 + b_{24}^{(1)}x_4 = b_{25}^{(1)}, \quad (4)$$

gdje je

$$b_{2j}^{(1)} = \frac{a_{2j}^{(1)}}{a_{22}^{(1)}}, \quad j > 2.$$

Eliminirajući x_2 iz druge i treće jednadžbe sustava (4) dobivamo sljedeći sustav

$$\begin{aligned} a_{33}^{(2)}x_3 + a_{34}^{(2)}x_4 &= a_{35}^{(2)} \\ a_{43}^{(2)}x_3 + a_{44}^{(2)}x_4 &= a_{45}^{(2)} \end{aligned} \quad (5)$$

gdje je

$$a_{ij}^{(2)} = a_{ij}^{(1)} - a_{i2}^{(1)}b_{2j}^{(1)}, \quad (i, j \geq 3).$$

Ako sada prvu jednadžbu sustava (5) podijelimo sa $a_{33}^{(2)}$ (koji naravno treba biti $\neq 0$), dobivamo

$$x_3 + b_{34}^{(2)}x_4 = b_{35}^{(2)} \quad (6)$$

gdje je

$$b_{3j}^{(2)} = \frac{a_{3j}^{(2)}}{a_{33}^{(2)}}, \quad j > 3.$$

Eliminirajući x_3 iz druge jednadžbe sustava (5) dobivamo

$$a_{44}^{(3)}x_4 = a_{45}^{(3)}, \quad (7)$$

gdje je

$$a_{ij}^{(3)} = a_{ij}^{(2)} - a_{i3}^{(2)}b_{3j}^{(2)}, \quad (i, j \geq 4).$$

Iz (7) slijedi

$$x_4 = \frac{a_{45}^{(3)}}{a_{44}^{(3)}} = \dot{b}_{45}^{(3)}.$$

Ostale nepoznanice sada nalazimo postupno iz jednadžbi (6), (4) i (2), pa imamo

$$\begin{aligned} x_3 &= b_{35}^{(2)} - b_{34}^{(2)}x_4 \\ x_2 &= b_{25}^{(1)} - b_{24}^{(1)}x_4 - b_{23}^{(1)}x_3 \\ x_1 &= b_{15} - b_{14}x_4 - b_{13}x_3 - b_{12}x_2. \end{aligned}$$

Promatramo li apstraktno opisani postupak, vidimo da isti prevodi polazni sustav $Ax = b$ u njemu ekvivalentan $A_n x = b_n$ pri čemu je A_n gornje trokutasta matrica oblika

$$A_n = \begin{bmatrix} 1 & b_{12} & b_{13} & \dots & b_{1n-1} & b_{1n} \\ 0 & 1 & b_{23}^{(1)} & \dots & b_{2n-1}^{(1)} & b_{2n}^{(1)} \\ 0 & 0 & 1 & \dots & b_{3n-1}^{(2)} & b_{3n}^{(2)} \\ & & & \vdots & & \\ 0 & 0 & 0 & \dots & 1 & b_{n-1n}^{(n-2)} \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}, \quad b_n = \begin{bmatrix} b_{1n+1} \\ b_{2n+1}^{(1)} \\ b_{3n+1}^{(2)} \\ \vdots \\ b_{n-1n+1}^{(n-2)} \\ b_{nn+1}^{(n-1)} \end{bmatrix}.$$

Provedemo li sljedeću supstituciju u sustavu jednadžbi $\xi = x + e$, gdje je

$$e = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \text{ dobivamo}$$

$$A(\xi - e) = b,$$

ili

$$A\xi = b + Ae = b',$$

što je novi sustav jednadžbi. Komponente rješenja $\{\xi_i\}$ novog sustava razlikuju se za 1 od komponenata rješenja $\{x_i\}$ polaznog sustava, a koeficijenti desne strane su sume svih koeficijenata odgovarajuće polazne jednadžbe. Taj sustav obično rješavamo paralelno radi kontrole.

Ocjena pogreške zbog netočnosti A i b .

Ponekad treba ocijeniti pogrešku Δx rješenja x linearog sustava ako su poznate pogreške ΔA i Δb matrice A i vektora b . Imamo:

$$(A + \Delta A)(x + \Delta x) = b + \Delta b$$

$$Ax + A\Delta x + (\Delta A)x + (\Delta A)(\Delta x) = b + \Delta b.$$

Zanemarivši član $\Delta A \Delta x$ dobivamo linearni sustav u Δx

$$\Delta A x = \Delta b - (\Delta A)x, \quad (8)$$

koji onda treba riješiti u Δx . Usporedimo li sustav (8) s polaznim, vidimo da ima promijenjenu samo desnu stranu.

Izračunavanje determinante Gaussovom metodom.

Izračunavanje vrijednosti determinante prema definiciji njene vrijednosti računski gledano je posao koji traži puno vremena. Stoga se redovito koriste postupci koji smanjuju broj računskih operacija.

Gaussova se metoda može primijeniti i na izračunavanje vrijednosti determinante. Neka je $\det(A)$ determinanta matrice A . Gledamo li homogen sustav pridružen matrici A , imamo

$$Ax = 0,$$

koji po Gaussovoj metodi zamjenjujemo sa ekvivalentnim

$$A_n x = 0$$

$$A_n = \begin{bmatrix} 1 & b_{12} & \dots & b_{1n} \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & 1 \end{bmatrix}.$$

Pogledamo li koje su operacije vršene, onda vidimo da smo vrijednost $\det(A)$ redom dijelili sa a_{11} , $a_{22}^{(1)}$, ..., $a_{nn}^{(n-1)}$, pa dakle vrijedi

$$\det A_n = 1 = \frac{\det A}{a_{11} a_{22}^{(1)} \dots a_{nn}^{(n-1)}}$$

odnosno,

$$\det A = a_{11} a_{22}^{(1)} \dots a_{nn}^{(n-1)}. \quad (9)$$

(Prisjetimo se dviju činjenica upotrebljenih pri gornjem računu. Te su da se determinanta množi brojem da joj se jedan stupac ili jedan redak pomnoži tim brojem i da je vrijednost trokutne determinante jednaka produktu dijagonalnih elemenata.)

Nalaženje inverzne matrice Gaussovom metodom

Gaussova se metoda može primijeniti i na izračunavanje inverzne matrice. Neka je $A = [a_{ij}]$. Elemente inverzne matrice od A označimo sa x_{ij} , tj.

$$A^{-1} = [x_{ij}].$$

Tada imamo $AA^{-1} = I$. Množenjem dobivamo sustav linearnih jednadžbi sa n^2 nepoznanica, koji se raspada na n sustava sa po n nepoznanica. To su ovi sustavi

$$\sum_{k=1}^n a_{ik}x_{kj} = \delta_{ij}, \quad (i, j = 1, 2, 3, \dots, n), \quad (10)$$

gdje je

$$\delta_{ij} = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

Rješenjem tih sustava dobivamo elemente inverzne matrice.

Uskoro ćemo vidjeti da je Gauss – Jordanova metoda prikladnija za nalaženje inverzne matrice.

Primjer 3.1. Da na konkretnom primjeru ponovimo postupak riješimo sustav

$$\begin{aligned} 2x_1 - 7x_2 + 4x_3 &= 9 \\ x_1 + 9x_2 - 6x_3 &= 1 \\ -3x_1 + 8x_2 + 5x_3 &= 6 \end{aligned}$$

čije rješenje lako nađemo: $x_1 = 4$, $x_2 = 1$, $x_3 = 2$. Postupak ćemo pratiti na pripadnim proširenim matricama. Imamo

$$B = [A, b] = \left[\begin{array}{ccc|c} 2 & -7 & 4 & 9 \\ 1 & 9 & -6 & 1 \\ -3 & 8 & 5 & 6 \end{array} \right]$$

Eliminacija nepoznanice x_1 prevodi B u

$$B_1 = \left[\begin{array}{ccc|c} 1 & -\frac{7}{2} & 2 & \frac{9}{2} \\ 0 & \frac{25}{2} & -8 & -\frac{7}{2} \\ 0 & -\frac{5}{2} & 11 & \frac{39}{2} \end{array} \right]$$

Eliminiramo li x_2 dobivamo

$$B_2 = \left[\begin{array}{ccc|c} 1 & -\frac{7}{2} & 2 & \frac{9}{2} \\ 0 & 1 & -\frac{16}{25} & -\frac{7}{25} \\ 0 & 0 & \frac{47}{5} & \frac{94}{5} \end{array} \right]$$

Konačno dijeljenjem zadnjeg retka sa $\frac{47}{5}$ završavamo postupak eliminacije i dobivamo

$$B_3 = \left[\begin{array}{ccc|c} 1 & -\frac{7}{2} & 2 & \frac{9}{2} \\ 0 & 1 & -\frac{16}{25} & -\frac{7}{25} \\ 0 & 0 & 1 & 2 \end{array} \right].$$

Iz B_3 čitamo odmah $x_3 = 2$. Da nađemo x_2 treba u

$$x_2 - \frac{16}{25}x_3 = -\frac{7}{25}$$

uvrstiti $x_3 = 2$ tako da imamo $x_2 = -\frac{7}{25} + \frac{16}{25} \cdot 2 = \frac{25}{25} = 1$.

Konačno iz $x_1 - \frac{7}{2}x_2 + 2x_3 = \frac{9}{2}$ izračunavamo x_1 kao $x_1 = \frac{7}{2} \cdot 1 - 2 \cdot 2 + \frac{9}{2} = 4$.

Vrijednost $\det A$ iznosi prema (9)

$$\det A = a_{11}a_{22}^{(1)}a_{33}^{(2)} = 2 \cdot \frac{25}{2} \cdot \frac{47}{5} = 5 \cdot 47 = 235.$$

Da Gaussovom metodom odredimo A^{-1} potrebno je riješiti ova tri linearna sustava

$$\begin{aligned} & \left. \begin{aligned} 2x_{11} - 7x_{21} + 4x_{31} &= 1 \\ x_{11} + 9x_{21} - 6x_{31} &= 0 \\ -3x_{11} + 8x_{21} + 5x_{31} &= 0 \end{aligned} \right\} \quad i = 1, 2, 3; \quad j = 1, \\ & \left. \begin{aligned} 2x_{12} - 7x_{22} + 4x_{32} &= 0 \\ x_{12} + 9x_{22} - 6x_{32} &= 1 \\ -3x_{12} + 8x_{22} + 5x_{32} &= 0 \end{aligned} \right\} \quad i = 1, 2, 3; \quad j = 2, \\ & \left. \begin{aligned} 2x_{13} - 7x_{23} + 4x_{33} &= 0 \\ x_{13} + 9x_{23} - 6x_{33} &= 0 \\ -3x_{13} + 8x_{23} + 5x_{33} &= 1 \end{aligned} \right\} \quad i = 1, 2, 3; \quad j = 3. \end{aligned}$$

Njihovim rješenjem dobivamo

$$A^{-1} = \frac{1}{235} \begin{bmatrix} 93 & 67 & 6 \\ 13 & 22 & 16 \\ 35 & 5 & 25 \end{bmatrix}.$$

Gauss-Jordanova metoda eliminacije

Pod tim nazivom se obično razumijeva modificirana Gaussova metoda, u smislu da se kod ove metode ne moramo vraćati natrag da nađemo $x_{n-1}, x_{n-2}, \dots, x_1$ (prati opis Gaussove metode). Prvi korak je isti kao kod Gaussove metode. Drugi korak provodimo gotovo kao i prije s time da x_2 eliminiramo i iz prve, te ovih daljnijih jednadžbi. Zatim se analogno x_3 eliminira iz prve, druge, četvrte i svih dalnjih jednadžbi. Nastavljanjem tog postupka dobivamo na kraju rješenje. Ako algebarski mislimo, onda je efekt primjene Gauss-Jordanove metode množenje s A^{-1} , tj. dobili smo

$$x = A^{-1}b$$

bez iznalaženja A^{-1} .

Nalaženje inverzne matrice Gauss-Jordanovom metodom.

Nakon uočenog algebarskog efekta Gauss-Jordanove metode nameće se ovakav postupak nalaženja inverzne matrice. Zamislimo da je

$$M = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} & 1 & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2n} & 0 & 1 & \dots & 0 \\ \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} & 0 & 0 & \dots & 1 \end{bmatrix} = [A \quad I]$$

matrica sustava jednadžbi (nepoznanice nisu zapisane kao ni konstantni članovi desne strane). To je matrica formata $n \times 2n$ koja se u blok zapisu sastoji od matrice A i jedinične matrice I . Provedemo li na M Gauss-Jordanov postupak, koji provodimo isključivo nad prvih n stupaca, bit će M pomnožena s A^{-1} . No, pomnoženo po blokovima to daje

$$A^{-1}M = [A^{-1}A \quad A^{-1}I] = [I \quad A^{-1}]$$

dakle u drugih n stupaca dobivamo elemente od A^{-1} .

Gauss-Jordanova i Gaussova metoda s izborom vodećeg elementa

Prepostavka pod kojom smo proveli Gaussovou i Gauss-Jordanovu metodu je da su svi elementi $a_{kk}^{(k-1)} \neq 0$, jer smo s njima dijelili pripadnu jednadžbu i provodili eliminaciju nepoznanice x_k . U slučaju da u nekom koraku naiđemo na nulu onda možemo ispremještati daljnje jednadžbe ili nepoznanice tako da u taj položaj dovedemo koeficijent različit od nule. Pri tome naravno treba voditi računa o nastaloj permutaciji. Daljnji ozbiljni nedostatak vođenja postupka na opisani način leži u tome da, iako je $a_{kk}^{(k-1)} \neq 0$ isti može po absolutnoj vrijednosti biti malen (blizu nule). Kako se s $a_{kk}^{(k-1)}$ dijeli, to u dalnjem računu može izazvati znatno nagomilavanje pogreške. Da se to izbjegne uvodi se algoritam kod kojega u k -tom koraku ne eliminiramo x_k nego neku drugu nepoznanicu x_j , $j \neq k$. To se najčešće ostvaruje izborom vodećeg (glavnog) elementa, pa se taj algoritam naziva **Gauss-Jordanova, odnosno Gaussova metoda s izborom vodećeg (glavnog) elementa**. Osnovna ideja metode sastoji se u tome da u sljedećem koraku ne eliminiramo nepoznanicu sa sljedećim indeksom nego onu nepoznanicu koja ima po absolutnoj vrijednosti najveći koeficijent. Takav se element naziva **vodeći, glavni ili pivot element**. Na taj način izbjegavamo dijeljenje s brojem koji je u usporedbi s ostalima malen i smanjujemo pogrešku. To se može raditi na različite načine. Prva ideja koja se nameće je da u matrici A potražimo element s najvećom absolutnom vrijednošću i eliminiramo njegovu nepoznanicu iz preostalih jednadžbi, te da na novodobivenom sustavu ponovimo postupak s nekom dalnjom nepoznanicom. Na Gauss-Jordanovoj metodi to izgleda ovako. Ako je a_{kj} vodeći element od A , tj. $|a_{kj}|$ ima najveću absolutnu vrijednost, onda pomoću k -te jednadžbe eliminiramo nepoznanicu x_j iz svih ostalih jednadžbi. Da nastavimo postupak, moramo odrediti novi vodeći (glavni) element u novonastaloj matrici sustava. Taj tražimo izvan k -toga retka i j -toga stupca, tj. treba precrtatи k -ti redak i j -ti stupac novonastale matrice i među preostalim elementima naći onaj koji ima najveću absolutnu vrijednost. Jasno je da u trećem izboru treba precrtatи dva retka i dva stupca u novonastaloj matrici, tj. one koji su korišteni u prvom i drugom koraku, i među preostalim elementima naći vodeći element. Time korak po korak dolazimo do dvoredne podmatrice u predzadnjem koraku i jednoredne podmatrice u zadnjem koraku. Krajnje dobivena matrica će biti ispermutirana jedinična matrica (stupci su ispermutirani), o čemu treba voditi računa kod očitavanja komponenata nađenog rješenja.

Druge dvije jednostavne varijante izbora vodećeg elementa su po recima, odnosno stupcima, koje su razumljive nakon gornjeg opisa izbora vodećeg elementa.

Ilustrirajmo Gauss-Jordanovu metodu s vodećim elementom (po gore opisanom izboru) na jednostavnom primjeru.

Primjer 3.2. Riješimo ovaj sustav jednadžbi Gauss-Jordanovom metodom izbora vodećeg elementa.

$$-3x_1 + 8x_2 + 5x_3 = 6$$

$$2x_1 - 7x_2 + 4x_3 = 9$$

$$x_1 + 9x_2 - 6x_3 = 1.$$

Račun provedimo tablično na proširenoj matrici B . Imamo

$$B = \left[\begin{array}{ccc|c} -3 & 8 & 5 & 6 \\ 2 & -7 & 4 & 9 \\ 1 & 9 & -6 & 1 \end{array} \right]$$

Vidimo da je $a_{32} = 9$ vodeći element matrice A . Time eliminiramo x_2 iz sustava, pa nova proširena matrica sustava glasi

$$B_1 = \left[\begin{array}{ccc|c} -\frac{35}{9} & 0 & \frac{31}{3} & \frac{46}{9} \\ \frac{25}{9} & 0 & -\frac{2}{3} & \frac{88}{9} \\ \frac{1}{9} & 1 & -\frac{2}{3} & \frac{1}{9} \end{array} \right]$$

Vodeći element nove matrice B_1 tražimo izvan drugog stupca i trećeg retka. Nalazimo da je novi vodeći element $\frac{31}{3}$ koji se nalazi u poziciji prvog retka i trećeg stupca. Time eliminiramo nepoznaciju x_3 , pa nova proširena matrica glasi

$$B_2 = \left[\begin{array}{ccc|c} -\frac{35}{93} & 0 & 1 & \frac{46}{93} \\ \frac{235}{93} & 0 & 0 & \frac{940}{93} \\ -\frac{13}{93} & 1 & 0 & \frac{41}{93} \end{array} \right]$$

Kod izbora trećeg vodećeg elementa smo upućeni na prvi stupac i drugi redak, dakle $\frac{235}{93}$. Pomoću njega eliminiramo x_1 iz prve i treće jednadžbe. Nova proširena matrica glasi

$$B_3 = \left[\begin{array}{ccc|c} 0 & 0 & 1 & 2 \\ 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 1 \end{array} \right].$$

Ponovimo da prva tri stupca predstavljaju permutirane stupce jedinične matrice, dok četvrti stupac sadrži rješenje koje nije u prirodnom poretku, jer nismo u tom poretku provodili eliminaciju varijabli. Citamo li četvrti stupac odozgo prema dolje nailazimo prvo na element 2 što predstavlja vrijednost od x_3 traženog rješenja, dakle $x_3 = 2$. Zatim nailazimo na 4 što predstavlja vrijednost od x_1 traženog rješenja, $x_1 = 4$, i konačno nailazimo na 1 što je vrijednost od x_2 rješenja polaznog sustava. Konačno zapisano imamo $x_1 = 4$, $x_2 = 1$, $x_3 = 2$ ili

$$x = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}.$$

Opišimo ukratko Gaussov metodu s izborom vodećeg elementa, s time da je vodeći element onaj koji ima najveću apsolutnu vrijednost. Prvi korak je isti kao kod Gauss-Jordanove metode. Naime, nađemo element matrice A s najvećom apsolutnom vrijednošću. Ako je to bio a_{kj} , onda pomoću k -te jednadžbe eliminiramo x_j iz svih ostalih jednadžbi. U drugom koraku nađemo vodeći element novodobivene matrice izvan k -tog retka i j -tog stupca. Neka je to $a_{rs}^{(1)}$. Sada r -tu jednadžbu koristimo za eliminaciju s -te nepoznacije iz svih jednadžbi osim k -te koja je korištena u prvom koraku. U novonastaloj matrici nađemo opet vodeći element izvan k -tog i r -tog retka i izvan j -tog i s -tog stupca i eliminiramo pripadnu nepoznaciju iz svih jednadžbi koje prethodno nisu korištene u ovom postupku, tj. iz svih jednadžbi osim k -te i r -te. Uočimo da u nekom koraku eliminiramo nepoznaciju dobivenu izborom vodećeg elementa iz svih jednadžbi koje nisu korištene u prethodnim koracima eliminacije. Korak po korak dolazimo do jednoredne podmatrice, tj. do jednog elementa, pa dijeljenjem s njime njegove jednadžbe završavamo postupak. Uočimo da iz dobivenog treba postupno naći rješenje, što sve zajedno ilustriramo jednostavnim primjerom. Naime, zadnju nepoznaciju očitavamo direktno dok prethodne treba izračunati.

Primjer 3.3. Riješimo ovaj sustav jednadžbi Gaussovom metodom izbora vodećeg elementa:

$$\begin{aligned}x_1 + 2x_2 - x_3 + 2x_4 &= 4 \\2x_1 + 3x_2 - x_3 + 4x_4 &= 6 \\4x_1 + 5x_2 - 3x_3 + 8x_4 &= 12 \\2x_1 + 3x_2 - 2x_3 + 3x_4 &= 6.\end{aligned}$$

Kao i u prethodnom primjeru račun ćemo provesti tablično nad proširenom matricom sustava B .

$$B = [A, b] = \left[\begin{array}{cccc|c} 1 & 2 & -1 & 2 & 4 \\ 2 & 3 & -1 & 4 & 6 \\ 4 & 5 & -3 & 8 & 12 \\ 2 & 3 & -2 & 3 & 6 \end{array} \right]$$

Element $a_{34} = 8$ ima najveću apsolutnu vrijednost, pa pomoću treće jednadžbe eliminiramo x_4 iz prve, druge i četvrte. Time dobivamo novu proširenu matricu

$$B_1 = \left[\begin{array}{cccc|c} 0 & \frac{3}{4} & -\frac{1}{4} & 0 & 1 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{5}{8} & -\frac{3}{8} & 1 & \frac{3}{2} \\ \frac{1}{2} & \frac{9}{8} & -\frac{7}{8} & 0 & \frac{3}{2} \end{array} \right].$$

Novi vodeći element tražimo sada izvan trećeg retka i četvrtog stupca. Vidimo da $\frac{9}{8}$ ima najveću apsolutnu vrijednost i da je to element četvrtog retka i drugog stupca. To znači da ćemo se poslužiti četvrtom jednadžbom i eliminirati x_2 iz prve i druge jednadžbe, treću preskačemo jer je korištena u prethodnom koraku. Time dobivamo

$$B_2 = \left[\begin{array}{cccc|c} -\frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\ -\frac{2}{9} & 0 & \frac{8}{9} & 0 & -\frac{2}{3} \\ \frac{1}{2} & \frac{5}{8} & \frac{3}{8} & 1 & \frac{3}{2} \\ \frac{4}{9} & 1 & -\frac{7}{9} & 0 & \frac{4}{3} \end{array} \right].$$

U sljedećem koraku treba precrtati 2. i 4. redak te 3. i 4. stupac. Vidimo da je $\frac{8}{9}$ sljedeći vodeći element pomoću koga eliminiramo x_3 iz prve jednadžbe. To daje

$$B_3 = \left[\begin{array}{cccc|c} -\frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} \\ -\frac{1}{4} & 0 & 1 & 0 & -\frac{3}{4} \\ \frac{1}{2} & \frac{5}{8} & -\frac{3}{8} & 1 & \frac{3}{2} \\ \frac{4}{9} & 1 & -\frac{7}{9} & 0 & \frac{4}{3} \end{array} \right].$$

Zadnji korak je da prvu jednadžbu podijelimo s $-\frac{1}{4}$ što daje

$$B_4 = \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & -1 \\ -\frac{1}{4} & 0 & 1 & 0 & -\frac{3}{4} \\ \frac{1}{2} & \frac{5}{8} & -\frac{3}{8} & 1 & \frac{3}{2} \\ \frac{4}{9} & 1 & -\frac{7}{9} & 0 & \frac{4}{3} \end{array} \right].$$

Koraci natrag su sada sljedeći. Zadnje eliminirana varijabla se nalazi prva, predzadnje eliminirana druga i tako redom. U našem primjeru čitamo da je $x_1 = -1$. Iz druge jednadžbe dobivamo x_3 kao

$$x_3 = -\frac{3}{4} + \frac{1}{4}x_1, \quad x_3 = -\frac{3}{4} - \frac{1}{4} = -1.$$

Sljedeća nepoznanica koju računamo je x_2 , a računamo ju iz četvrte jednadžbe kao

$$x_2 = \frac{4}{3} - \frac{4}{9}x_1 + \frac{7}{9}x_3, \quad x_2 = 1.$$

Konačno, iz treće jednadžbe nalazimo x_4 ,

$$x_4 = \frac{3}{2} - \frac{1}{2}x_1 - \frac{5}{8}x_2 + \frac{3}{8}x_3, \quad x_4 = 1.$$

Prema tome rješenje glasi $x_1 = -1$, $x_2 = 1$, $x_3 = -1$, $x_4 = 1$ ili

$$x = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}.$$

LU (LR) faktorizacija, Choleskyjeva šema

Polazimo od nezavisnog sustava ($\det A \neq 0$) s n nepoznanica

$$Ax = b. \quad (11)$$

Za potrebe ovog razmatranja označimo komponente vektora b ovako

$$b = \begin{bmatrix} a_{1,n+1} \\ a_{2,n+1} \\ \vdots \\ a_{n,n+1} \end{bmatrix}.$$

Osnovna misao je da matricu A faktoriziramo $A = BC$ i da pri tome izaberemo prikladno građene matrice B i C . Gaussov algoritam sugerira trokutaste matrice. Stoga pretpostavimo da su B i C oblika

$$B = \begin{bmatrix} b_{11} & 0 & \dots & 0 \\ b_{21} & b_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix}, \quad (12)$$

$$C = \begin{bmatrix} 1 & c_{12} & c_{13} & \dots & c_{1n} \\ 0 & 1 & c_{23} & \dots & c_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}. \quad (13)$$

Dakle B je donje (engl. *lower*) ili lijevo (engl. *left*) trokutasta matrica, a C je gornje (engl. *upper*) ili desno (engl. *right*) trokutasta matrica, otkuda i potječu pokrate LU (LR) u oznaci faktorizacije. Korist takve faktorizacije je sljedeća. Uvrstimo li $A = BC$ u (11) dobivamo

$$B(Cx) = b$$

pri čemu smo Cx stavili u zagradu što smijemo zbog asocijativnosti matričnog množenja. Supstitucijom

$$y = Cx$$

dobivamo sustav

$$By = b$$

no kako je B donje trokutasta, to lako izračunamo y , koji zatim uvrstimo u

$$Cx = y.$$

Kako je C gornje trokutasta lako riješimo i taj sustav što je onda rješenje polaznog sustava. Preostaje dakle odrediti elemente matrica B i C . Radi se o ukupno n^2 nepoznatih elemenata. Sustav jednadžbi u elemente od B i C dobivamo iz jednakosti $A = BC$, što po elementima daje

$$a_{ij} = \sum_{k=1}^n b_{ik}c_{kj}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, n. \quad (14)$$

Vidimo da se radi o kvadratnom sustavu jednadžbi, međutim, sistematskim ispisivanjem tih jednadžbi nameće se šema rješavanja tog sustava koja se naziva **Choleskyjeva šema**. Prirodno pitanje u ovom trenutku je postojanje rješenja sustava (14). Na kraju ćemo navesti teorem koji osigurava postojanje i jedinstvenost rješenja. Napomenimo da postoje i neke druge faktorizacije u što se ovdje ne upuštamo.

Uvažimo li nule iz B i C dobivamo ovaj eksplisitni zapis sustava (14):

$$\begin{aligned} b_{i1} &= a_{i1}, \\ b_{ij} &= a_{ij} - \sum_{k=1}^{j-1} b_{ik}c_{kj}, \quad i \geq j > 1, \\ c_{1j} &= \frac{a_{1j}}{b_{11}}, \\ c_{ij} &= \frac{1}{b_{ii}} \left(a_{ij} - \sum_{k=1}^{i-1} b_{ik}c_{kj} \right), \quad 1 < i < j. \end{aligned} \quad (15)$$

Vidimo da se prvi stupac od B podudara s prvim stupcem od A . Nakon toga se lako nađe prvi redak od C . Kada imamo prvi redak od C možemo naći drugi stupac od B

$$b_{i2} = a_{i2} - b_{i1}c_{12} \quad i = 2, 3, \dots$$

Našavši drugi stupac od B možemo izračunati elemente drugog retka od C , tj.

$$c_{2j} = \frac{1}{b_{22}} (a_{2j} - b_{21}c_{1j}) \quad j = 3, 4, \dots$$

Zatim nalazimo treći stupac od B , a nakon toga treći redak od C i tako redom.

Kako su B i C trokutaste matrice, to kao u Gaussovoj metodi nalazimo

$$\begin{aligned} y_1 &= \frac{a_{1,n+1}}{b_{11}} \\ y_i &= \frac{1}{b_{ii}} \left(a_{i,n+1} - \sum_{k=1}^{i-1} b_{ik}y_k \right) \quad i > 1, \end{aligned} \quad (16)$$

odnosno

$$x_n = y_n$$

$$x_i = y_i - \sum_{k=i+1}^n c_{ik}x_k \quad i < n. \quad (17)$$

Da iskažemo dovoljan uvjet koji osigurava egzistenciju rješenja sustava (14) neka nam D_k označava k rednu determinantu

$$D_k = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{vmatrix} \quad k = 1, 2, \dots, n$$

koju dobivamo iz matrice A tako da uzmemo elemente na križanju prvih k redaka i prvih k stupaca. Vrijedi:

Teorem 3.1. Neka je A kvadratna matrica reda n i neka je

$$D_k \neq 0 \quad \text{za } k = 1, 2, \dots, n.$$

Tada matrica A ima jedinstvenu LU faktorizaciju.

Ilustrirajmo postupak sljedećim jednostavnim primjerom.

Primjer 3.4.

$$\begin{aligned} 3x_1 + x_2 - x_3 + 2x_4 &= 6 \\ -5x_1 + x_2 + 3x_3 - 4x_4 &= -12 \\ 2x_1 &+ x_3 - x_4 = 1 \\ x_1 - 5x_2 + 3x_3 - 3x_4 &= 3. \end{aligned}$$

Nadimo elemente matrica B i C koje na kraju glase:

$$B = \begin{bmatrix} 3 & 0 & 0 & 0 \\ -5 & \frac{8}{3} & 0 & 0 \\ 2 & -\frac{2}{3} & 2 & 0 \\ 1 & -\frac{16}{3} & 6 & \frac{5}{2} \end{bmatrix}, \quad C = \begin{bmatrix} 1 & \frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \\ 0 & 1 & \frac{1}{2} & -\frac{1}{4} \\ 0 & 0 & 1 & -\frac{5}{4} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Prvi stupac matrice B podudara se s prvim stupcem matrice A . Zatim redom računamo elemente prvog retka matrice C :

$$c_{12} = \frac{1}{3}, \quad c_{13} = -\frac{1}{3}, \quad c_{14} = \frac{2}{3}.$$

Sljedeći korak daje drugi stupac od B , tako da imamo

$$b_{22} = a_{22} - b_{21}c_{12} = 1 - \left(-5 \cdot \frac{1}{3}\right) = \frac{8}{3},$$

$$b_{32} = a_{32} - b_{31}c_{12} = 0 - 2 \cdot \frac{1}{3} = -\frac{2}{3},$$

$$b_{42} = a_{42} - b_{41}c_{12} = -5 - \frac{1}{3} = -\frac{16}{3}.$$

Sada možemo izračunati drugi redak od C , pa i mamo

$$c_{23} = \frac{1}{b_{22}}(a_{23} - b_{21}c_{13}) = \frac{3}{8}\left(3 - (-5)\left(-\frac{1}{3}\right)\right) = \frac{1}{2},$$

$$c_{24} = \frac{1}{b_{22}}(a_{24} - b_{21}c_{14}) = \frac{3}{8}\left((-4) - (-5)\frac{2}{3}\right) = -\frac{1}{4}.$$

Nakon ovoga prelazimo na treći stupac od B .

$$b_{33} = a_{33} - \sum_{k=1}^2 b_{3k}c_{k3} = 2,$$

$$b_{43} = a_{43} - \sum_{k=1}^2 b_{4k}c_{k3} = 6.$$

U C je još preostao element c_{34} koji iznosi

$$c_{34} = \frac{1}{b_{33}}\left(a_{34} - \sum_{k=1}^2 b_{3k}c_{k4}\right) = -\frac{5}{4}.$$

Konačno

$$b_{44} = a_{44} - \sum_{k=1}^3 b_{4k}c_{k4} = \frac{5}{2}.$$

Rješenje sustava $By = b$ glasi $y_1 = 2$, $y_2 = -\frac{3}{4}$, $y_3 = -\frac{7}{4}$, $y_4 = 3$. Konačno, rješenje sustava $Cx = y$ je rješenje polaznog sustava i iznosi $x_1 = 1$, $x_2 = -1$, $x_3 = 2$, $x_4 = 3$.

Bunemanova metoda

Pri rješavanju nekih važnih problema nailazimo na "velike sustave" (s velikim brojem nepoznatih, na stotine) linearnih jednadžbi koji su posebno građeni. Iako se ne radi o općim sustavima isti u primjenama predstavljaju važne slučajeve. U novije vrijeme nađene su izvjesne direktnе metode za posebno građene sustave i što je najvažnije u tim su slučajevima te metode superiorne iteracijskim metodama. Jedna od prvih takovih metoda je **Bunemanov algoritam** iz 1969. godine. Istaknimo da je zadnjih 20-tak godina razvijeno više algoritama (metoda) rješavanja tzv. "velikih sustava" kod kojih je matrica sustava "rijetka" (engl. *sparse*), tj. ima relativno malo elemenata različitih od nula. Ovdje ukratko opisujemo početni dio Bunemanovog algoritma.

Matricu A nazivamo tridiagonalna ako ima oblik

$$A = \begin{bmatrix} a_{11} & a_{12} & 0 & 0 & \dots & 0 \\ a_{21} & a_{22} & a_{23} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & 0 \\ 0 & \dots & 0 & a_{n-1,n-2} & a_{n-1,n-1} & a_{n-1,n} \\ 0 & 0 & \dots & 0 & a_{n,n-1} & a_{nn} \end{bmatrix}$$

Isto tako zovemo matricu tridiagonalnom ako takav oblik ima u nekom bloku zapisu, što ćemo uskoro vidjeti u opisu Bunemanove metode. Takve se matrice pojavljuju kod rješavanja mnogih važnih tehničkih problema.

Radi opće informacije i navikavanja na nešto općenitija razmatranja zamislimo da u ovom slučaju promatramo linearan sustav jednadžbi s kompleksnim nepoznanicama. Prema tome, elementi matrice sustava su kompleksni brojevi, komponente vektora su kompleksni brojevi i nepoznanice su kompleksne varijable. Bunemanov algoritam je primjenjiv na linearan sustav jednadžbi

$$Mz = b \quad (18)$$

koji u blok zapisu izgleda ovako

$$M = \begin{bmatrix} A & I & 0 & \dots & \dots & \dots & 0 \\ I & A & I & 0 & \dots & \dots & 0 \\ 0 & I & A & I & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & I & A & I \\ 0 & \dots & \dots & 0 & I & A & \end{bmatrix}, \quad z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_q \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_q \end{bmatrix}, \quad (19)$$

gdje je I ($p \times p$) jedinična matrica, dok je A ($p \times p$) hermitska (u slučaju realnih elemenata simetrična) tridiagonalna matrica. Time se M sastoji od q blokova po recima i po stupcima (format matrice M je $qp \times qp$), z_i i b_i su p -komponentni vektori (podvektori od z i b koji su qp -komponentni). Građu od M koristimo tako da rekurzivno reduciramo broj nepoznanica, pri čemu dobivamo slično građen sustav s manje nepoznanica. Raspišemo li sustav (19) po blokovima imamo

$$\begin{aligned} Az_1 + z_2 &= b_1 \\ z_{j-1} + Az_j + z_{j+1} &= b_j \quad j = 2, 3, \dots, q-1 \\ z_{q-1} + Az_q &= b_q. \end{aligned} \quad (20)$$

Iz triju uzastopnih jednadžbi

$$\begin{aligned} z_{j-2} + Az_{j-1} + z_j &= b_{j-1}, \\ z_{j-1} + Az_j + z_{j+1} &= b_j, \\ z_j + Az_{j+1} + z_{j+2} &= b_{j+1}, \end{aligned} \quad (21)$$

mogemo za sve parne $j = 2, 4, \dots$, eliminirati varijable z_{j-1} i z_{j+1} . To provodimo tako da u (21) zbrojimo prvu i treću jednadžbu i od njihovog zbroja odbijemo s A pomnoženu drugu jednadžbu. Imamo dakle odbijanje sljedećih jednadžbi

$$\begin{aligned} z_{j-2} + Az_{j-1} + 2Iz_j + Az_{j+1} + z_{j+2} &= b_{j-1} + b_{j+1} \\ Az_{j-1} + A^2z_j + Az_{j+1} &= Ab_j, \end{aligned}$$

što daje

$$z_{j-2} + (2I - A^2)z_j + z_{j+2} = b_{j-1} - Ab_j + b_{j+1}.$$

Prema tome za **neparan** q sustav (19) se reducira na

$$\begin{bmatrix} 2I - A^2 & I & 0 & \dots & \dots & 0 \\ I & 2I - A^2 & I & 0 & \dots & 0 \\ 0 & I & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \ddots & \ddots & I \\ 0 & \dots & \dots & 0 & I & 2I - A^2 \end{bmatrix} \begin{bmatrix} z_2 \\ z_4 \\ \vdots \\ z_{q-1} \end{bmatrix} = \begin{bmatrix} b_1 + b_3 - Ab_2 \\ b_3 + b_5 - Ab_4 \\ \vdots \\ b_{q-2} + b_q - Ab_{q-1} \end{bmatrix} \quad (22)$$

u kome se nalaze samo nepoznanice (podvektori) s parnim indeksom kojih ukupno ima $\frac{q-1}{2}$. Prebacimo li u svakoj drugoj jednadžbi sustava (20) sve parne nepoznanice

na desnu stranu dobivamo sustav jednadžbi u neparne nepoznanice (kojih ukupno ima $\frac{q+1}{2}$):

$$\begin{aligned} Az_1 &= b_1 - z_2 \\ Az_3 &= b_3 - z_2 - z_4 \\ Az_5 &= b_5 - z_4 - z_6 \\ &\dots \\ Az_q &= b_q - z_{q-1} \end{aligned}$$

ili u blok zapisu

$$\left[\begin{array}{cccccc} A & 0 & \dots & \dots & 0 \\ 0 & A & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & A \end{array} \right] \begin{bmatrix} z_1 \\ z_3 \\ \vdots \\ z_q \end{bmatrix} = \begin{bmatrix} b_1 - z_2 \\ b_1 - z_2 - z_4 \\ \vdots \\ b_q - z_{q-1} \end{bmatrix} \quad (23)$$

koji se opet raspada na $\frac{(q+1)}{2}$ podsustava oblika

$$Az_{2k+1} = b_{2k+1} - z_{2k} - z_{2k+2}, \quad k = 0, 1, \dots, \frac{1}{2}(q-1). \quad (24)$$

Uočimo da podsustavi (24) imaju različite desne strane, dok su im lijeve strane jednakе. Prema tome, rješavanje sustava (19) je svedeno na rješavanje sustava (22) s "polovicom" nepoznanica, a potom sustava (23), odnosno podsustava (24). Nadalje, sustav (22) je građen kao i sustav (19), umjesto matrice A nalazi se matrica $2I - A^2$, tako da sustav (22) možemo ovako zapisati

$$M^{(1)}z^{(1)} = b^{(1)} \quad (25)$$

gdje je

$$M^{(1)} = \begin{bmatrix} A^{(1)} & I \\ I & A^{(1)} \\ \ddots & \ddots & \ddots \\ I & A^{(1)} \end{bmatrix}, \quad A^{(1)} = 2I - A^2$$

$$z^{(1)} = \begin{bmatrix} z_1^{(1)} \\ z_2^{(1)} \\ \vdots \\ z_{q_1}^{(1)} \end{bmatrix} = \begin{bmatrix} z_2 \\ z_4 \\ \vdots \\ z_{q-1} \end{bmatrix}, \quad b^{(1)} = \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \\ \vdots \\ b_{q_1}^{(1)} \end{bmatrix} = \begin{bmatrix} b_1 + b_3 - Ab_2 \\ b_3 + b_5 - Ab_4 \\ \vdots \\ b_{q-2} + b_q - Ab_{q-1} \end{bmatrix},$$

pa ako je i $\frac{q-1}{2}$ neparan, onda se redukcija nepoznanica može primjeniti i na (25). Ako je polazni q , označimo ga sada s q_0 , oblika

$$q_0 = 2^{k+1} - 1 = 2 \cdot 2^k - 1,$$

onda postupak redukcije možemo ponoviti, jer je broj blokova u (25) jednak $\frac{q_0-1}{2} = \frac{2^{k+1}-1-1}{2} = 2^k - 1$, dakle neparan. U tom slučaju se redukcija može nastaviti dok u (23) ne dođemo do jednog bloka. Polazeći od $q_0 = 2^{k+1} - 1$ to se ostvaruje u $(k-1)$ -oj redukciji. Naime, nakon prve redukcije (23) ima $2^k = \frac{2^{k+1}-1+1}{2}$ blokova, nakon druge

redukcije ima 2^{k-1} blokova, tako da ćemo nakon n -te redukcije u (23) imati 2^{k-n-1} blokova, otkuda i $k-n-1=0$ slijedi da je zadnja redukcija bila $(k-1)$ -va. Prema tome nastavljajući opisanu redukciju dobivamo niz matrica $A^{(r)}$ i vektora $b^{(r)}$ prema sljedećem postupku.

Označimo $A^{(0)} = A$; $b_j^{(0)} = b_j$, $j = 1, 2, \dots, q_0$, $q_0 = 2^{k+1} - 1$. Sada za $r = 0, 1, 2, \dots, k-1$ označimo

$$(i) \quad A^{(r+1)} = 2I - (A^{(r)})^2,$$

$$(ii) \quad b_j^{(r+1)} = b_{2j-1}^{(r)} + b_{2j+1}^{(r)} - A^{(r)}b_{2j}^{(r)}, \quad j = 1, 2, \dots, 2^{k-r} - 1 = q_{r+1}.$$

Time za svaki $r+1$, $r = 0, 1, 2, \dots, k-1$ dobivamo sustav linearnih jednadžbi

$$M^{(r+1)}z^{(r+1)} = b^{(r+1)}$$

koji u blok raspisanom obliku glasi

$$\begin{bmatrix} A^{(r+1)} & I & 0 & \dots & 0 \\ I & A^{(r+1)} & I & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & I & A^{(r+1)} & I \\ 0 & \dots & 0 & I & A^{(r+1)} \end{bmatrix} \begin{bmatrix} z_1^{(r+1)} \\ z_2^{(r+1)} \\ \vdots \\ z_{q_{r+1}}^{(r+1)} \end{bmatrix} = \begin{bmatrix} b_1^{(r+1)} \\ b_2^{(r+1)} \\ \vdots \\ b_{q_{r+1}}^{(r+1)} \end{bmatrix}. \quad (26)$$

Rješenje $z^{(r+1)}$ tog sustava odmah daje podvektore s parnim indeksima rješenja $z^{(r)}$ odgovarajućeg sustava jednadžbi $M^{(r)}z^{(r)} = b^{(r)}$ u r -tom koraku, tj. imamo

$$\begin{bmatrix} z_2^{(r)} \\ z_4^{(r)} \\ \vdots \\ z_{q_{r-1}}^{(r)} \end{bmatrix} = \begin{bmatrix} z_1^{(r+1)} \\ z_2^{(r+1)} \\ \vdots \\ z_{q_{r+1}}^{(r+1)} \end{bmatrix}, \quad (27)$$

dok podvektore s neparnim indeksima od $z^{(r)}$ možemo dobiti rješavanjem sustava

$$\begin{bmatrix} A^{(r)} & 0 & \dots & 0 \\ 0 & A^{(r)} & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & \dots & 0 & A^{(r)} \end{bmatrix} \begin{bmatrix} z_1^{(r)} \\ z_3^{(r)} \\ \vdots \\ z_{q_r}^{(r)} \end{bmatrix} = \begin{bmatrix} b_1^{(r)} - z_2^{(r)} \\ b_3^{(r)} - z_2^{(r)} - z_4^{(r)} \\ \vdots \\ b_{q_r}^{(r)} - z_{q_r-1}^{(r)} \end{bmatrix}. \quad (28)$$

Našavši niz matrica $A^{(r)}$ i vektora $b_j^{(r)}$, $r = 0, 1, \dots, k-1$ možemo natrašće naći konačno rješenje $z = z^{(0)}$ polaznog sustava (19). Taj postupak izgleda ovako. U početnom koraku određujemo $z^{(k)} = z_k^{(1)}$ rješenjem sustava jednadžbi (to je sustav (28) redukcijom sveden na jedan blok)

$$A^{(k)}z^{(k)} = b^{(k)} = b_1^{(k)}.$$

Sada za $r = k-1, k-2, \dots, 0$:

a) uvrštavamo $z_{2j}^{(r)} = z_j^{(r+1)}$, $j = 1, 2, \dots, q_{r+1} = 2^{k-r} - 1$;

b) za $j = 1, 3, 5, \dots, q_r$ izračunamo vektor $z_j^{(r)}$ rješavanjem sustava

$$A^{(r)}z_j^{(r)} = b_j^{(r)} - z_{j-1}^{(r)} - z_{j+1}^{(r)}, \quad (z_0^{(r)} = z_{q_r+1}^{(r)} = 0).$$

Na kraju imamo rješenje polaznog sustava kao $z = z^{(0)}$.

To je taj početni dio Bunemanove metode, koji nam daje ideju kako izgleda neki moderan algoritam rješavanja specijalnog sustava jednadžbi. Iako opisani algoritam lijepo izgleda on ima računske mane. U prvom redu eksplizitno računanje $A^{(r+1)} = 2I - (A^{(r)})^2$ je vrlo skupo, jer povećanjem r od tridiagonalne matrice $A^{(0)} = A$ dobivamo ubrzo "gustu" matricu $A^{(r)}$ koja je "trakasta", a širina njene trake iznosi $2^r + 1$. Trakasta matrica ima oblik

$$A = \begin{bmatrix} * & \dots & * & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & 0 & \dots & 0 \\ * & \ddots & \ddots & \ddots & \dots & 0 \\ 0 & \dots & \ddots & \ddots & \ddots & * \\ 0 & \dots & 0 & * & \dots & * \end{bmatrix}, \quad a_{ij} = 0 \quad \text{za} \quad |i - j| \geq m$$

i m se zove širina trake. Nadalje, povećanjem r postaje skupo i rješavanje sustava b), tj. sustava

$$A^{(r)} z_j^{(r)} = b_j^{(r)} - z_{j-1}^{(r)} - z_{j+1}^{(r)}.$$

Sve se te mane mogu otkloniti na način da se izbjegne eksplizitno računanje od $A^{(r)}$ iskoristivši stanoviti prikaz od $A^{(r)}$ kao produkta tridiagonalnih matrica. Taj dio metode izostavljamo, jer smo početni i elementarnim trikovima htjeli pokazati kako u principu izgleda neki moderan algoritam. Onaj tko to zatreba neka se posluži specijaliziranom literaturom.

O upotrebljivosti izračunatog rješenja

Poslužimo li se bilo kojom od prikazanih metoda da nađemo rješenje sustava linearnih jednadžbi $Ax = b$, onda je u većini slučajeva ono što izračunamo stanovita aproksimacija \tilde{x} točnog rješenja x . Postavlja se pitanje kako ocijeniti točnost od \tilde{x} , tj. kako ocijeniti pogrešku $x - \tilde{x}$. Da bi mjerili pogrešku moramo imati način mjeranja "veličine" vektora. Za to nam služi norma vektora, u oznaci $\|x\|$, koja predstavlja poopćenje poznatog pojma apsolutne vrijednosti vektora (drugi korijen iz sume kvadrata komponenata). Drugim riječima formule koje daju ograde za veličinu $\|x - \tilde{x}\|$ su formule za ocjenu pogreške. Te formule izostavljamo jer zahtijevaju određeno znanje o inverznoj matrici A^{-1} matrice A . Sada ćemo prikazati jednu korisnu ocjenu, novijeg datuma, koja ne zahtijeva nikakvo znanje o A^{-1} .

Neka je $A = (a_{ik})$ matrica. S $|A| = (|a_{ik}|)$ označavamo matricu čiji su elementi apsolutne vrijednosti elemenata matrice A . Analogno, ako je $b = (b_1, b_2, \dots, b_n)^T$ vektor, onda $|b|$ označava vektor čije su komponente apsolutne vrijednosti komponenta od b , tj. $|b| = (|b_1|, |b_2|, \dots, |b_n|)^T$. Znak $A \leq B$ među matricama (ili vektorima) znači da po elementima (komponentama) vrijedi $a_{ij} \leq b_{ij}$. Naravno da ima neuporedivih matrica i vektora, dovoljno je da je nejednakost narušena na jednom paru elemenata.

Promatrajmo sustav linearnih jednadžbi

$$A_0 x = b_0. \tag{29}$$

U primjenama su, iz različitih razloga, A_0 i b_0 netočni, npr. zbog pogrešaka mjeranja. Prema tome, razumno je prihvati približno rješenje \tilde{x} sustava $A_0 x = b_0$ kao "ispravno" ako je \tilde{x} točno rješenje nekog "okolinski bliskog" sustava jednadžbi

$$A \tilde{x} = b, \tag{30}$$

pri čemu to znači da A i b pripadaju skupovima \mathcal{A} , odnosno \mathcal{B} , definiranim ovako:

$$\begin{aligned} A \in \mathcal{A} &= \{A : |A - A_0| \leq \Delta A\}, \\ b \in \mathcal{B} &= \{b : |b - b_0| \leq \Delta b\}. \end{aligned} \quad (31)$$

(31) znači da A (odnosno b), pripada maloj okolini od A_0 (odnosno b_0) u prostoru matrica (odnosno vektora). U tom smislu dokažimo da je istinit sljedeći teorem, nakon čega ćemo komentirati njegovu korist.

Teorem 3.2. Neka je $\Delta A \geq 0$ i $\Delta b \geq 0$ te $A \in \mathcal{A}$ i $b \in \mathcal{B}$ neka su definirani prema (31). Podimo od sustava $A_0x = b_0$. Tada su svakom približnom rješenju \tilde{x} sustava $A_0x = b_0$ pridruženi matrica $A \in \mathcal{A}$ i vektor $b \in \mathcal{B}$ za koje vrijedi

$$A\tilde{x} = b \quad (32)$$

ako i samo ako je

$$|r(\tilde{x})| \leq \Delta A |\tilde{x}| + \Delta b, \quad (33)$$

gdje je $r(\tilde{x}) = b_0 - A_0\tilde{x}$, ostatak za \tilde{x} .

Dokaz. Prepostavimo prvo (32), tj. da za neki $A \in \mathcal{A}$ i $b \in \mathcal{B}$ vrijedi

$$A\tilde{x} = b.$$

Tada je

$$A = A_0 + \delta A, \quad \text{pri čemu je} \quad |\delta A| \leq \Delta A$$

i

$$b = b_0 + \delta b, \quad \text{pri čemu je} \quad |\delta b| \leq \Delta b,$$

otkuda slijedi da je

$$\begin{aligned} |r(\tilde{x})| &= |b_0 - A_0\tilde{x}| = |b - \delta b - (A - \delta A)\tilde{x}| \\ &= |- \delta b + (\delta A)\tilde{x}| \leq |\delta b| + |\delta A||\tilde{x}| \\ &\leq \Delta b |\tilde{x}| + \Delta A |\tilde{x}|. \end{aligned}$$

Obratno, neka vrijedi (33), tj.

$$|r(\tilde{x})| \leq \Delta A |\tilde{x}| + \Delta b,$$

treba odrediti A i b . Označimo ovako komponente promatranih objekata:

$$r(\tilde{x}) = (\rho_1, \rho_2, \dots, \rho_n)^T$$

$$\Delta b + \Delta A |\tilde{x}| = (\sigma_1, \sigma_2, \dots, \sigma_n)^T$$

i primijetimo da su svi $\sigma_i \geq 0$. Nadalje, označimo s $\delta A = (\delta a_{ij})$, $\delta b = (\delta b_1, \delta b_2, \dots, \delta b_n)^T$, $\tilde{x} = (\xi_1, \xi_2, \dots, \xi_n)^T$ i definirajmo

$$\begin{aligned} \delta a_{ij} &= (\text{sign} \xi_j) \frac{\rho_i}{\sigma_i} \Delta a_{ij}, \\ \delta b_i &= -\frac{\rho_i}{\sigma_i} \Delta b_i, \end{aligned} \quad (34)$$

pri čemu, ako je $\sigma_i = 0$, onda definiramo da je $\frac{\rho_i}{\sigma_i} = 0$. Iz (33) slijedi da je $\rho_i = 0$ ako je $\sigma_i = 0$ i da je

$$|\rho_i| \leq \sigma_i, \quad \text{odnosno} \quad \left| \frac{\rho_i}{\sigma_i} \right| \leq 1, \quad i = 1, 2, \dots, n. \quad (35)$$

Dokažimo da vrijedi

$$A = A_0 + \Delta A \in \mathcal{A} \quad \text{i} \quad b = b_0 + \delta b \in \mathcal{B}.$$

Naime, pogledamo li komponente od

$$|A - A_0| = |A_0 + \Delta A - A_0| = |\Delta A| \quad \text{i} \quad |b - b_0| = |\delta b|$$

vidimo da zbog (35) vrijedi

$$|\delta a_{ij}| = \left| \frac{\rho_i}{\sigma_i} \right| \Delta a_{ij} \leq \Delta a_{ij} \quad \text{i} \quad |\delta b_i| = \left| \frac{\rho_i}{\sigma_i} \right| \Delta b_i \leq \Delta b_i.$$

Još je ostalo da provjerimo jednakost $A\tilde{x} = b$, tj. da je \tilde{x} točno rješenje sustava određenog s $A = A_0 + \Delta A$ i $b = b_0 + \delta b$. Pođimo od $\rho_i = \sigma_i \frac{\rho_i}{\sigma_i}$, $i = 1, 2, \dots, n$.

Ispišimo ovako lijevu i desnu stranu tih jednakosti

$$\begin{aligned} \rho_i &= b_i - \sum_{j=1}^n a_{ij} \xi_j = \left(\Delta b_i + \sum_{j=1}^n \Delta a_{ij} |\xi_j| \right) \frac{\rho_i}{\sigma_i} \\ &= \frac{\rho_i}{\sigma_i} \Delta b_i + \sum_{j=1}^n \frac{\rho_i}{\sigma_i} \Delta a_{ij} (\operatorname{sign} \xi_j) \xi_j \\ &= -\delta b_i + \sum_{j=1}^n \delta a_{ij} \xi_j. \end{aligned}$$

Prema tome za $i = 1, 2, \dots, n$ vrijedi

$$b_i - \sum_{j=1}^n a_{ij} \xi_j = -\delta b_i + \sum_{j=1}^n \delta a_{ij} \xi_j,$$

otkuda dobivamo

$$\sum_{j=1}^n (a_{ij} + \delta a_{ij}) \xi_j = b_i + \delta b_i,$$

tj. vrijedi (32)

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

što je i trebalo dokazati.

Istaknimo korisnost dokazanog teorema. On omogućava da dođemo do zaključka o prihvatljivosti rješenja iz malešnosti njegovog ostatka. Na primjer, ako sve komponente od A_0 i b_0 imaju istu relativnu točnost ε , tj. ako je

$$\Delta A = \varepsilon |A_0| \quad \text{i} \quad \Delta b = \varepsilon |b_0|,$$

onda je uvjet teorema ispunjen ako je

$$|r(\tilde{x})| = |b_0 - A_0 \tilde{x}| \leq \varepsilon (|b_0| + |A_0| |\tilde{x}|), \tag{36}$$

pa iz (36) možemo izračunati najmanji ε za koji se \tilde{x} može prihvatiti kao upotrebljivo rješenje. Uočimo da je za poznate ΔA i Δb , te nađeni \tilde{x} , računanje desne strane nejednakosti (33) jednostavan posao.

3.2. Iteracijske metode

Opća linearna iteracija

Kod iteracijskih metoda polazimo od linearog sustava jednadžbi

$$Ax = b, \quad \det(A) \neq 0 \quad (1)$$

i nekog početnog vektora $x^{(0)}$ tako da iteracijom oblika

$$x^{(k+1)} = \phi(x^{(k)}), \quad k = 0, 1, 2, \dots \quad (2)$$

generiramo niz vektora

$$x^{(0)} \rightarrow x^{(1)} \rightarrow x^{(2)} \rightarrow \dots$$

od koga tražimo da konvergira rješenju sustava

$$x = A^{-1}b = \lim_{k \rightarrow \infty} x^{(k)}. \quad (3)$$

Naravno da proizvoljna funkcija ϕ ne dolazi u obzir. Ovdje ćemo se ograničiti na slučaj kada je ϕ linearna funkcija. Prevedemo li sustav (1) u oblik

$$x = \alpha x + \beta \quad (4)$$

od koga zahtijevamo da je **ekvivalentan** polaznom sustavu (1), onda, za odabrani početni $x^{(0)}$, generiramo niz

$$x^{(k+1)} = \alpha x^{(k)} + \beta, \quad k = 0, 1, 2, \dots \quad (5)$$

Svaki izbor prelaza sa (1) na (4) potencijalno daje **linearno iteracijsku metodu**. Nadalje kažemo da je **iteracijska metoda konvergentna** ako za svaki početni vektor $x^{(0)}$ pripadni niz vektora $\{x^{(k)}\}$ konvergira prema točnom rješenju sustava $x = A^{-1}b$.

Opišimo jedan dosta općenit prelaz s (1) na (4) kojim ćemo se služiti u više dalnjih iteracijskih metoda. Neka je B regularna matrica. Dodavanjem

$$Bx - Bx = 0 \quad (6)$$

jednadžbi (1) imamo ekvivalentan sustav

$$Bx + (A - B)x = b. \quad (7)$$

Pomnožimo sada (7) s B^{-1} . To daje ekvivalentan sustav oblika

$$x = (I - B^{-1}A)x + B^{-1}b, \quad (8)$$

pa uvrstivši

$$\alpha = I - B^{-1}A, \quad \beta = B^{-1}b \quad (9)$$

dobivamo (4), odnosno, prema (5) niz vektora

$$x^{(k+1)} = (I - B^{-1}A)x + B^{-1}b, \quad k = 0, 1, 2, \dots \quad (10)$$

Svaki izbor regularne matrice B potencijalno daje, naravno pod određenim uvjetima, konvergentnu iteracijsku metodu.

Primijetimo da općenito vrijedi:

Teorem 3.3. Ako niz $\{x^{(k)}\}$ dobiven iteracijom (5) konvergira, onda je njegov limes rješenje polaznog sustava (1).

Dokaz. Iz

$$\lim_{k \rightarrow \infty} x^{(k+1)} = \lim_{k \rightarrow \infty} (\alpha x^{(k)} + \beta) = \alpha \lim_{k \rightarrow \infty} x^{(k)} + \beta$$

i $\lim_{k \rightarrow \infty} x^{(k)} = \lim_{k \rightarrow \infty} x^{(k+1)} = x$, slijedi $x = \alpha x + \beta$, tj. $x = \lim_{k \rightarrow \infty} x^{(k)}$ je rješenje sustava (4), a kako je isti ekvivalentan sustavu (1) to je $x = \lim_{k \rightarrow \infty} x^{(k)}$ rješenje polaznog sustava (1), tj. $x = A^{-1}b$. Prema tome, članovi konvergentnog iteracijskog niza $x^{(0)}, x^{(1)}, \dots, x^{(k)}, \dots$ su aproksimacije točnog rješenja sustava (1).

Ponovimo što znači da niz vektora $\{x^{(k)}\}$ konvergira. Neka su $x_i^{(k)}$, $i = 1, 2, \dots, n$ komponente vektora $x^{(k)}$,

$$x^{(k)} = [x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}]^T.$$

Konvergencija niza vektora definira se pomoću konvergencije njihovih komponenata. Dakle vektor $x = [x_1 x_2 \dots x_n]^T$ je limes niza vektora $\{x^{(k)}\}$ ako i samo ako je

$$x_i = \lim_{k \rightarrow \infty} x_i^{(k)}, \quad i = 1, 2, \dots, n.$$

Analogno se definira i konvergencija niza matrica $\{A^{(k)}\}$, $A^{(k)} = (a_{ij}^{(k)})$, $k = 1, 2, \dots$. Tada je matrica $A = (a_{ij})$ limes niza matrica $\{A^{(k)}\}$ ako i samo ako je

$$a_{ij} = \lim_{k \rightarrow \infty} a_{ij}^{(k)}, \quad i, j = 1, 2, \dots, n.$$

U ovom ponavljanju smo se ograničili na kvadratne matrice jer u ovim paragrafima isključivo radimo s kvadratnim matricama.

Da naslutimo kada je neka linearna iteracijska metoda (5) konvergentna provjerimo ovako kratko razmatranje. Označimo ovako pogrešku aproksimacije u k -tom koraku

$$p_k = x^{(k)} - x$$

gdje je s x označeno točno rješenje i za njega vrijedi (4). Imamo dakle ove jednakosti:

$$\begin{aligned} x^{(k)} &= \alpha x^{(k-1)} + \beta, \\ x &= \alpha x + \beta. \end{aligned}$$

Oduzimanjem tih dviju jednakosti dobivamo rekurziju među susjednim pogreškama

$$p_k = x^{(k)} - x = \alpha[x^{(k-1)} - x] = \alpha p_{k-1},$$

iz koje korak po korak nalazimo

$$p_k = \alpha^k p_0, \quad k = 0, 1, 2, \dots \tag{11}$$

Naravno, $x^{(k)} \rightarrow x$ ako i samo ako $p_k \rightarrow 0$. Na desnoj strani formule (11) imamo produkt $\alpha^k p_0$, pri čemu je $p_0 = x^{(0)} - x$ zapravo proizvoljan vektor jer je točno rješenje x danog sustava konstantan vektor x dok $x^{(0)}$ kod konvergentne metode možemo uzeti po volji. Naslućujemo da vrijedi:

Teorem 3.4. Linearno iteracijska metoda je konvergentna ako i samo ako je $\lim_{k \rightarrow \infty} \alpha^k = 0$.

Dokaz. Uvjet je dovoljan, jer $\lim_{k \rightarrow \infty} \alpha^k = 0$ povlači prema (11) da je $\lim_{k \rightarrow \infty} p_k = 0$, dakle $x^{(k)} \rightarrow x = A^{-1}b$. Prepostavimo sada obratno da je $\lim_{k \rightarrow \infty} p_k = 0$ za svaki izbor $x^{(0)}$. Iz (11) dobivamo da je $0 = (\lim_{k \rightarrow \infty} \alpha^k)p_0$ i to za svaki p_0 . Dakle, $\lim_{k \rightarrow \infty} \alpha^k$ je operator koji se poništava na svakom vektoru p_0 kakav je jedino 0.

Teorem 3.4 nam nije od posebne direktnе koristi, no taj teorem kaže, ako znamo kriterij za koje matrice kod te metode vrijedi $\alpha^k \rightarrow 0$ dobili smo kriterij konvergencije metode. Taj ćemo rezultat koristiti u sljedećem paragrafu.

Konvergentnost metode i ocjena pogreške

Da izrečemo stanovite teoreme o konvergentnim metodama potrebno je ponoviti nekoliko pojmove. Podimo od n -redne kvadratne matrice $A = (a_{ij})$. Vektor $v \neq 0$ zovemo **svojstvenim vektorom** matrice (operatora) A ako postoji skalar λ takav da vrijedi

$$Av = \lambda v. \quad (12)$$

Skalar λ zovemo **svojstvena vrijednost** matrice A koja odgovara svojstvenom vektoru v . Znamo da su svojstvene vrijednosti korjeni algebarske jednadžbe

$$\det(\lambda I - A) = 0 \quad (13)$$

koja se zove **karakteristična jednadžba** matrice A . U razvijenom zapisu determinante (13) izgleda ovako

$$\begin{vmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \dots & \lambda - a_{nn} \end{vmatrix} = 0. \quad (14)$$

Neka su $\lambda_1, \dots, \lambda_m$ (najviše n različitih) svojstvene vrijednosti od A , tj. korjeni jednadžbe (13). **Spektralni radius** matrice A se definira kao

$$\rho(A) = \max\{|\lambda_1|, |\lambda_2|, \dots, |\lambda_m|\}, \quad (15)$$

tj. jednak je najvećoj absolutnoj vrijednosti svojstvenih vrijednosti matrice A .

Da dokažemo sljedeći teorem potrebno je računati potencije matrice, dakle matičnu funkciju $f(A) = A^k$, gdje je k prirodan broj. Kod takvog računanja se služimo **Jordanovom formom** matrice. To je dijagonalna blok-matrica kojoj se na dijagonalu nalaze **Jordanovi blokovi** koji odgovaraju pojediniм svojstvenim vrijednostima, dakle blok-matrica oblika

$$J = \begin{bmatrix} J(\lambda_1) & 0 & \dots & 0 \\ 0 & J(\lambda_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_m(\lambda_m) \end{bmatrix}, \quad (16)$$

a svaki Jordanov blok se sastoji od **elementarnih Jordanovih klijetki**. Naime, svaki Jordanov blok gledan odvojeno je opet dijagonalna blok-matrica, čiji se dijagonalni

blokovi zovu elementarne Jordanove klijetke. Dakle, pojedini Jordanov blok je ovako građen

$$J(\lambda_i) = \begin{bmatrix} J_1(\lambda_i) & 0 & \dots & 0 \\ 0 & J_2(\lambda_i) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_p(\lambda_i) \end{bmatrix}, \quad (17)$$

pri čemu je $J_r(\lambda_i)$ ili matrica oblika

$$J_r(\lambda_i) = \begin{bmatrix} \lambda_i & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda_i & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \lambda_i & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda_i \end{bmatrix},$$

tj. na dijagonali ima svojstvenu vrijednost λ_i , a na sporednoj dijagonali jedinice, ili je to dijagonalna matrica kojoj se λ_i nalazi na dijagonali, tj. oblika

$$J_r(\lambda_i) = \begin{bmatrix} \lambda_i & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_i \end{bmatrix}. \quad (18)$$

Osnovni rezultat matrične algebre kaže da je matrica slična svojoj Jordanovoj formi, tj. postoji regularna matrica T takva da vrijedi

$$A = T^{-1}JT, \quad (19)$$

otkuda slijedi da su i cijelobrojne potencije od A slične cijelobrojnim potencijama od J , tj. vrijedi

$$A^k = (T^{-1}JT)(T^{-1}JT) \dots (T^{-1}JT) = T^{-1}J^kT, \quad (20)$$

pa se računanje potencije A^k svodi na računanje J^k i naknadno množenje s T^{-1} i T . Kako je J dijagonalna blok-matrica njene cijelobrojne potencije su opet blok matrice s cijelobrojnim potencijama njenih blokova, tj. vrijedi

$$J^k = \begin{bmatrix} J_1(\lambda_1)^k & 0 & \dots & 0 \\ 0 & J_2(\lambda_2)^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_m(\lambda_m)^k \end{bmatrix}, \quad (21)$$

tako da trebamo samo znati kako se računaju cijelobrojne potencije elementarnih Jordanovih klijetki. Ako je klijetka dijagonalan blok, onda znamo da je potencija opet dijagonalna čija se dijagonala sastoјi od potencija dijagonalnih elemenata. Kako je s potencijom klijetke koja na gornjoj sporednoj dijagonali ima jedinice? Ovdje ćemo napisati gotov rezultat koji ovako izgleda. Neka je $J_r(\lambda_i)$ n -redna i želimo ispisati

$J_r(\lambda_i)^k$ za $k \geq n$. Tada je

$$J_r(\lambda_i)^k = \begin{bmatrix} \lambda_i^k & k\lambda_i^{k-1} & \frac{k(k-1)}{2!}\lambda_i^{k-2} & \dots & \dots & \frac{k(k-1)\dots(k-n+2)}{(n-1)!}\lambda_i^{k-(n-1)} \\ 0 & \lambda_i^k & k\lambda_i^{k-1} & \dots & \dots & \frac{k(k-1)\dots(k-n+3)}{(n-2)!}\lambda_i^{k-(n-2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_i^k & k\lambda_i^{k-1} \\ 0 & 0 & 0 & \dots & 0 & \lambda_i^k \end{bmatrix} \quad (22)$$

drugim riječima, duž dijagonale imamo k -te potencije, a na sporednim gornjim dijagonalama imamo jednake članove sa sve nižim potencijama, dok ne dođemo do $\lambda_i^{k-(n-1)}$ u poziciji $(1, n)$. Kako će nama trebati samo limes takvih potencija kada $k \rightarrow \infty$ to smo ispustili potencije $< n$. Čitatelj može pogledati i neku drugu knjigu ako mu treba detaljnija informacija. Isto tako čitatelj može pokušati potencirati 3-rednu Jordanovu klijetku, pa će ubrzo osjetiti kako se računaju elementi iznad dijagonale. Nas zanima kada će limes pojedinih elemenata biti nula, tj. kada je

$$\lim_{k \rightarrow \infty} J_r(\lambda_i)^k = 0.$$

Za ilustraciju pogledajmo kada je

$$\lim_{k \rightarrow \infty} \frac{k(k-1)\dots(k-n+2)}{(n-1)!} \lambda_i^{k-(n-1)} = 0$$

tj. limes člana u poziciji $(1, n)$. Faktor $k(k-1)\dots(k-n+2)$ raste neograničeno kada $k \rightarrow \infty$. Da bi limes bio nula nužno je da $\lambda_i^k \rightarrow 0$ za $k \rightarrow \infty$. No, to je onda i samo onda ako je $|\lambda_i| < 1$. Taj je uvjet i dovoljan da bude

$$\lim_{k \rightarrow \infty} \frac{k(k-1)\dots(k-n+2)}{(n-1)!} \lambda_i^{k-(n-1)} = 0,$$

jer se prvi faktor ponaša kao $(n-1)$ -va potencija od k dok je drugi faktor eksponencijalna funkcija.

Tako smo ovim razmatranjem dokazali ovaj teorem.

Teorem 3.5. *Niz matrica $\{A^k\}$ konvergira k nul matrici ako i samo ako je $\rho(A) < 1$ ($\rho(A)$ je spektralni radijus od A).*

U teoremu 3.4 smo vidjeli da je linearno iteracijska metoda

$$x^{(k+1)} = \alpha x^{(k)} + \beta$$

konvergentna ako i samo ako je $\lim_{k \rightarrow \infty} \alpha^k = 0$. Povežemo li to s prethodnim teoremom dobivamo nuždan i dovoljan uvjet konvergencije linearno iteracijske metode koji onda glasi:

Teorem 3.6. *Linearno iteracijska metoda*

$$x^{(k+1)} = \alpha x^{(k)} + \beta$$

je konvergentna ako i samo ako je $\rho(\alpha) < 1$.

Do sada smo se sreli s pojmom **duljine** ili **norme** vektora koja je bila drugi kriterij iz sume kvadrata njegovih komponenata. Taj ćemo pojam poopćiti i dalje kratko zvati **norma vektora**, a poslužit će nam u formulaciji dovoljnog uvjeta konvergencije linearno iteracijske metode.

Definicija 3.1. Norma vektora $\|v\|$ je nenegativna realna funkcija na vektorskom prostoru koja zadovoljava sljedeće uvjete:

- $\|v\| > 0$ za svaki $v \neq 0$, (pozitivnost)
- $\|\lambda v\| = |\lambda| \|v\|$, gdje je $|\lambda|$ apsolutna vrijednost skalara λ , (homogenost)
- $\|u + v\| \leq \|u\| + \|v\|$, (nejednakost trokuta).

Evo nekoliko uobičajenih normi vektora. Neka je $v = [v_1 v_2 \dots v_n]^T$.

$$\text{Maksimum norma: } \|v\|_\infty = \max_i |v_i| \quad (23)$$

$$\text{Apsolutna norma: } \|v\|_1 = \sum_{i=1}^n |v_i| \quad (24)$$

$$\text{Euklidska norma: } \|v\|_2 = \left(\sum_{i=1}^n |v_i|^2 \right)^{1/2} \quad (25)$$

$$l_p \text{ norma: } \|v\|_p = \left(\sum_{i=1}^n |v_i|^p \right)^{1/p} \quad (26)$$

Iz (26) dobivamo (24) i (25) za $p = 1$, odnosno $p = 2$. Ako $p \rightarrow \infty$, onda (26) daje (23) što pojašnjava oznaku maksimum norme.

Za svaku normu vrijedi

$$\|u - v\| \geq \| \|u\| - \|v\| \| \quad (27)$$

gdje vanjske vertikalne crte znače uobičajenu apsolutnu vrijednost (modul) realnog broja. Iz zapisa u uvjetu b) i nejednakosti (27) je jasno zašto se za oznaku norme uvodi znak s dvije vertikalne crte $\| \cdot \|$. Lako se uvjerimo u točnost nejednakosti (27). Naime iz c) imamo

$$\|u\| = \|(u - v) + v\| \leq \|u - v\| + \|v\|,$$

pa je prema tome

$$\|u - v\| \geq \|u\| - \|v\|.$$

Zamjenimo li uloge od u i v i upotrijebimo b) imamo

$$\|u - v\| = \|v - u\| \geq \|v\| - \|u\|,$$

pa iz dobivenih dviju nejednakosti slijedi (27).

Među normama vektora $\| \cdot \|_1$, $\| \cdot \|_2$ i $\| \cdot \|_\infty$ vrijede ove nejednakosti

$$\|v\|_\infty \leq \|v\|_1 \leq n \|v\|_\infty \quad (28)$$

$$\|v\|_\infty \leq \|v\|_2 \leq \sqrt{n} \|v\|_\infty \quad (29)$$

$$\frac{1}{\sqrt{n}} \|v\|_1 \leq \|v\|_2 \leq \|v\|_1. \quad (30)$$

Označimo sa $M(n, n)$ skup n -rednih kvadratnih matrica. Kako svaku $n \times n$ matricu možemo shvatiti kao n^2 -komponentni vektor to vidimo da se pojam norme vektora može proširiti i na matrice, no uzet ćemo u obzir i množenje matrica pa imamo sljedeću definiciju.

Definicija 3.2. Norma matrice je nenegativna realna funkcija na $M(n, n)$ koja zadovoljava sljedeće uvjete:

- $\|A\| > 0$ za svaku $A \neq 0$ (pozitivnost)
- $\|\lambda A\| = |\lambda| \|A\|$, gdje je $|\lambda|$ absolutna vrijednost skalara λ (homogenost)
- $\|A + B\| \leq \|A\| + \|B\|$ (nejednakost trokuta)
- $\|AB\| \leq \|A\| \|B\|$ (submultiplikativnost)

U općim razmatranjima se $\|\cdot\|$ koja zadovoljava samo a), b) i c) zove opća norma matrice, a norma koja zadovoljava i d) zove submultiplikativna norma matrice. (Ako u d) vrijedi jednakost, onda se zove multiplikativna norma matrice.) Mi trebamo submultiplikativne norme matrica.

Navodimo nekoliko uobičajenih normi matrica

$$\text{Redna norma: } \|A\|_r = \max_i \sum_{j=1}^n |a_{ij}| \quad (31)$$

$$\text{Stupčana norma: } \|A\|_s = \max_j \sum_{i=1}^n |a_{ij}| \quad (32)$$

$$\text{Euklidska norma: } \|A\|_2 = \left(\sum_{j=1}^n \sum_{i=1}^n |a_{ij}|^2 \right)^{1/2} \quad (33)$$

$$\text{Maksimum norma: } \|A\|_\infty = \max_{i,j} |a_{ij}| \quad (34)$$

Istaknimo da maksimum norma (34) nije submultiplikativna, ne zadovoljava uvjet d), pa je time primjer opće norme matrice. Nadalje, norma (33) se u nekim knjigama zove Schurova norma, a u nekim Schmidtova norma.

Primjer 3.5. Za $A = \begin{bmatrix} 1 & 0 \\ 2 & 2 \end{bmatrix}$ imamo: $\|A\|_r = \max\{1, 4\} = 4$; $\|A\|_s = \max\{3, 2\} = 3$; $\|A\|_2 = \sqrt{1+4+4} = 3$; $\|A\|_\infty = \max\{0, 1, 2\} = 2$.

Sada kada smo ponovili pojmove norme vektora i norme matrica, možemo ekvivalentno pomoću norme opisati konvergenciju niza vektora, odnosno, matrica. (U 2.1 smo opisali konvergencije pomoću komponenata vektora, odnosno, elemenata matrica.) Kažemo da niz vektora $x^{(k)}$ konvergira vektoru a ako i samo ako $\|x^{(k)} - a\| \rightarrow 0$ kada $k \rightarrow \infty$. Slično, niz matrica $A^{(k)}$ konvergira matrici A ako i samo ako $\|A^{(k)} - A\| \rightarrow 0$ kada $k \rightarrow \infty$.

Dokažimo sada sljedeći jednostavan i za nas koristan teorem.

Teorem 3.7. Ako je bilo koja submultiplikativna norma matrice A manja od jedinice, onda niz potencija $\{A^k\}$ konvergira k nul matrici ($\lim_{k \rightarrow \infty} A^k = 0$).

Dokaz. Kako vrijedi

$$\|A^k - 0\| = \|A^k\| = \|AA^{k-1}\| \leq \|A\| \|A^{k-1}\| \leq \dots \leq \|A\|^k$$

i kako je po pretpostavci $\|A\| < 1$ to imamo da $\|A^k\| \rightarrow 0$ za $k \rightarrow \infty$. Iz konvergencije niza normi $\|A^k\|$ zaključujemo da niz A^k konvergira, a iz toga da je limes niza normi nula zaključujemo da je limes nul matrica (uvjet a) definicije norme matrice), dakle $\lim_{k \rightarrow \infty} A^k = 0$.

Teorem 3.7 zajedno s teoremom 3.4 daje nam dovoljan uvjet konvergencije linearno iteracijske metode što izričemo kao sljedeći teorem.

Teorem 3.8. Linearno iteracijska metoda

$$x^{(k+1)} = \alpha x^{(k)} + \beta$$

je konvergentna ako je bilo koja submultiplikativna norma matrice α manja od 1.

Teoremi 3.5 i 3.7 nam omogućavaju da povežemo submultiplikativne norme matrica s njenim spektralnim radijusom. Vrijedi sljedeći teorem, koji u kombinaciji s teoremom 3.6 daje teorem 3.8.

Teorem 3.9. Spektralni radius $\rho(A)$ matrice A nije veći od bilo koje njene submultiplikativne norme.

Dokaz. Neka je $\varepsilon > 0$, i $\|\cdot\|$ proizvoljna submultiplikativna norma matrice, i A neka matrica. Pomoću A definirajmo matricu B ovako:

$$B = \frac{1}{\|A\| + \varepsilon} A.$$

Sada je prema uvjetu b) definicije 3.2

$$\|B\| = \left\| \frac{1}{\|A\| + \varepsilon} \right\| \|A\| = \frac{1}{\|A\| + \varepsilon} \|A\|$$

iz čega slijedi da je $\|B\| < 1$. Teorem 3.7 stoga povlači da je $\lim_{k \rightarrow \infty} B^k = 0$. Iz teorema 3.5 zaključujemo da je $\rho(B) < 1$, tj. absolutna vrijednost svake svojstvene vrijednosti od B je manja od 1. Ako je λ_i svojstvena vrijednost matrice A , onda je $\frac{\lambda_i}{\|A\| + \varepsilon}$ svojstvena vrijednost matrice B , jer iz

$$\det(\lambda_i I - A) = 0$$

slijedi da je (služeći se Binet-Cauchyjevim teoremom)

$$\det\left(\frac{1}{\|A\| + \varepsilon} I(\lambda_i I - A)\right) = 0,$$

jer je $\det\left(\frac{1}{\|A\| + \varepsilon} I\right) = \left(\frac{1}{\|A\| + \varepsilon}\right)^n \neq 0$. Nakon množenja imamo

$$\det\left(\frac{\lambda_i}{\|A\| + \varepsilon} I - \frac{1}{\|A\| + \varepsilon} A\right) = \det\left(\frac{\lambda_i}{\|A\| + \varepsilon} I - B\right) = 0.$$

Prema rečenom vrijedi $\left| \frac{\lambda_i}{\|A\| + \varepsilon} \right| < 1$, odnosno $|\lambda_i| < \|A\| + \varepsilon$ i to za svaki $\varepsilon > 0$, pa je dakle $\rho(A) \leq \|A\|$.

Da ocijenimo pogrešku linearno iteracijske metode uvest ćemo dodatni uvjet na normu matrice. Kažemo da je submultiplikativna norma matrice $\|\cdot\|$ **konzistentna** (suglasna) s normom vektora $\|\cdot\|_v$ ako vrijedi

$$\|Ax\|_v \leq \|A\| \|x\|_v \quad \text{za sve } x \in A.$$

Npr. redna norma matrice je konzistentna s maksimum normom vektora, stupčana norma matrice s apsolutnom normom vektora i euklidska norma matrice s euklidskom normom vektora.

Provđimo sada ovakvo opće razmatranje. Podimo od dviju susjednih aproksimacija $x^{(k-1)}$ i $x^{(k)}$, $k \geq 1$, dobivenih linearom iteracijom

$$x^{(k)} = \alpha x^{(k-1)} + \beta.$$

Tada za $p \geq 1$ možemo pisati

$$x^{(k+p)} - x^{(k)} = (x^{(k+1)} - x^{(k)}) + (x^{(k+2)} - x^{(k+1)}) + \dots + (x^{(k+p)} - x^{(k+p-1)})$$

tako da višestrukom primjenom nejednakosti trokuta vrijedi

$$\|x^{(k+p)} - x^{(k)}\| \leq \|x^{(k+1)} - x^{(k)}\| + \|x^{(k+2)} - x^{(k+1)}\| + \dots + \|x^{(k+p)} - x^{(k+p-1)}\|. \quad (35)$$

Nadalje vrijedi

$$x^{(m+1)} - x^{(m)} = (\alpha x^{(m)} + \beta) - (\alpha x^{(m-1)} + \beta) = \alpha(x^{(m)} - x^{(m-1)}),$$

tako da za konzistentne norme matrica i vektora dobivamo sljedeću nejednakost (koristimo istu oznaku za obje norme)

$$\|x^{(m+1)} - x^{(m)}\| = \|\alpha(x^{(m)} - x^{(m-1)})\| \leq \|\alpha\| \|x^{(m)} - x^{(m-1)}\|. \quad (36)$$

Iteriramo li (36) dobivamo za $m > k \geq 1$

$$\|x^{(m+1)} - x^{(m)}\| = \|\alpha\|^{m-k} \|x^{(k+1)} - x^{(k)}\|. \quad (37)$$

Time nejednakost (35) poprima ovaj oblik

$$\begin{aligned} \|x^{(p+k)} - x^{(k)}\| &\leq \|x^{(k+1)} - x^{(k)}\| + \|\alpha\| \|x^{(k+2)} - x^{(k)}\| + \dots + \|\alpha\|^{p-1} \|x^{(k+1)} - x^{(k)}\| \\ &= (1 + \|\alpha\| + \dots + \|\alpha\|^{p-1}) \|x^{(k+1)} - x^{(k)}\| \\ &= \frac{1 - \|\alpha\|^p}{1 - \|\alpha\|} \|x^{(k+1)} - x^{(k)}\|. \end{aligned} \quad (38)$$

Prepostavimo li $\|\alpha\| < 1$ dobivamo sljedeću nejednakost

$$\|x^{(p+k)} - x^{(k)}\| \leq \frac{1}{1 - \|\alpha\|} \|x^{(k+1)} - x^{(k)}\|. \quad (39)$$

Ako $p \rightarrow \infty$, onda $x^{(p+k)} \rightarrow x$ točnom rješenju, tako da smo dobili ovu ocjenu pogreške:

$$\|x - x^{(k)}\| \leq \frac{1}{1 - \|\alpha\|} \|x^{(k+1)} - x^{(k)}\| \leq \frac{1}{1 - \|\alpha\|} \|\alpha\| \|x^{(k)} - x^{(k-1)}\|,$$

ili konačno

$$\|x - x^{(k)}\| \leq \frac{\|\alpha\|}{1 - \|\alpha\|} \|x^{(k)} - x^{(k-1)}\|. \quad (40)$$

Ako je

$$\|\alpha\| < \frac{1}{2},$$

onda je $\|\alpha\| < 1 - \|\alpha\|$, pa (40) poprima jednostavan oblik

$$\|x - x^{(k)}\| \leq \|x^{(k)} - x^{(k-1)}\|$$

što je jednostavno građena formula, jer ocjenu pogreške računamo iz posljednjih dviju aproksimacija. To znači, ako je (uz $\|\alpha\| \leq \frac{1}{2}$)

$$\|x^{(k)} - x^{(k-1)}\| < \varepsilon,$$

onda je i pogreška manja od ε , tj. vrijedi

$$\|x - x^{(k)}\| < \varepsilon.$$

Iz (40) čitamo sljedeće. Ako ustanovimo da vrijedi

$$\|x^{(k)} - x^{(k-1)}\| \leq \frac{1 - \|\alpha\|}{\|\alpha\|} \varepsilon, \quad \|\alpha\| < 1,$$

onda je

$$\|x - x^{(k)}\| = \varepsilon.$$

Iz formule (40) dobivamo postupno sljedeću nejednakost

$$\|x - x^{(k)}\| \leq \frac{\|\alpha\|^k}{1 - \|\alpha\|} \|x^{(1)} - x^{(0)}\|. \quad (41)$$

Napose, ako smo uzeli $x^{(0)} = \beta$, onda je $x^{(1)} = \alpha\beta + \beta$, tako da je

$$\|x^{(1)} - x^{(0)}\| = \|\alpha\beta + \beta - \beta\| = \|\alpha\beta\| \leq \|\alpha\| \|\beta\|.$$

U tom slučaju ocjena pogreške (40) poprima ovaj oblik

$$\|x - x^{(k)}\| \leq \frac{\|\alpha\|^{k+1}}{1 - \|\alpha\|} \|\beta\|. \quad (42)$$

Iz (42) možemo izračunati broj koraka potrebnih da postignemo zadalu točnost. Nai-me iz

$$\frac{\|\alpha\|^{k+1}}{1 - \|\alpha\|} \|\beta\| \leq \varepsilon$$

slijedi

$$\|\alpha\|^{k+1} \leq \frac{\varepsilon}{\|\beta\|} (1 - \|\alpha\|)$$

otkuda možemo ocijeniti najmanji k za koji vrijedi (42).

Vratimo se sada linearnej iteraciji uvedenoj formulom (10) i pitajmo se koji izbori matrice B daju korisniju metodu. Općenito možemo reći da korisniju metodu dobivamo ako B bolje udovoljava ovim uvjetima:

- (i) da se sustav $Bx^{(k+1)} + (A - B)x^{(k)} = b$ lako riješi u $x^{(k+1)}$;
- (ii) da su svojstvene vrijednosti od $\alpha = I - B^{-1}A$ po absolutnoj vrijednosti (modulu) što je moguće manje.

Na kraju ovog paragrafa navedimo sljedeći standardni sustav matrice A , sustava jednadžbi $Ax = b$ kojim ćemo se poslužiti pri izboru matrice B u sljedećih nekoliko paragrafa. Napišimo A kao

$$A = D - E - F \quad (43)$$

gdje su

$$D = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a_{nn} \end{bmatrix} \quad (44)$$

$$E = - \begin{bmatrix} 0 & 0 & \dots & 0 \\ a_{21} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn-1} & 0 \end{bmatrix} \quad (45)$$

$$F = - \begin{bmatrix} 0 & a_{12} & \dots & a_{1n} \\ 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{nn} \\ 0 & \dots & 0 & 0 \end{bmatrix} \quad (46)$$

i pretpostavimo da je

$$a_{ii} \neq 0, \quad i = 1, 2, \dots, n.$$

Jacobijeva metoda iteracije

Polazimo od sustava

$$Ax = b, \quad \det A \neq 0, \quad a_{ii} \neq 0, \quad i = 1, 2, \dots, n. \quad (47)$$

Ako u rastavu (43) izaberemo $B = D$ dobivamo **Jacobijevu metodu iteracije**. Sada je

$$\alpha = I - D^{-1}A \quad \text{i} \quad \beta = D^{-1}b. \quad (48)$$

Kako je D dijagonalna i $a_{ii} \neq 0$, $i = 1, 2, \dots, n$, to je D regularna i njena inverzna matrica je također dijagonalna čiji su elementi $\frac{1}{a_{ii}}$, tj.

$$D^{-1} = \begin{bmatrix} \frac{1}{a_{11}} & 0 & \dots & 0 \\ 0 & \frac{1}{a_{22}} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \frac{1}{a_{nn}} \end{bmatrix}, \quad (49)$$

tako da lako izračunamo elemente matrice α i komponente vektora β . Uzmimo da je $i \neq j$ i pomnožimo u mislima i -ti redak od D^{-1} s j -tim stupcem od A . Odbivši to od (i, j) -toga elementa od I dobivamo

$$\alpha_{ij} = 0 - \frac{1}{a_{ii}} a_{ij} = -\frac{a_{ij}}{a_{ii}}, \quad i \neq j. \quad (50)$$

Ako je $i = j$, onda rezultat množenja i -tog retka od D^{-1} s i -tim stupcem od A i odbijanja od jedinice iz I daje

$$\alpha_{ii} = 1 - \frac{1}{a_{ii}} a_{ii} = 0. \quad (51)$$

Komponente vektora β su očigledno jednake

$$\beta_i = \frac{b_i}{a_{ii}}. \quad (52)$$

Prema tome, Jacobijeva iteracija izgleda ovako

$$x^{(k+1)} = \alpha x^{(k)} + \beta, \quad k = 0, 1, 2, \dots \quad (53)$$

gdje su elementi od α i β dani sa (50), (51) i (52).

U razvijenom obliku to izgleda ovako

$$x_i^{(k+1)} = \beta_i + \sum_{j=1}^n \alpha_{ij} x_j^{(k)}, \quad i = 1, 2, \dots, n. \quad (54)$$

Drugi put izvoda formula (54) bi bio da iz i -te jednadžbe sustava (47) direktno izračunamo nepoznanicu x_i što bi dalo

$$x_i = \frac{b_i}{a_{ii}} - \sum_{\substack{j=1 \\ j \neq i}}^n \frac{a_{ij}}{a_{ii}} x_j, \quad i = 1, 2, \dots, n \quad (55)$$

iz čega direktno dobivamo (54).

Što se tiče početnog vektora $x^{(0)}$ uobičajeno je uzeti $x^{(0)} = \beta$, jer ga jednostavno već znamo.

Da bi niz (53) konvergirao nužno je i dovoljno prema teoremu 3.6 iz 2.2 da je spektralni radius matrice α manji od 1. Dovoljan uvjet konvergencije je prema teoremu 3.8 da je bilo koja submultiplikativna norma matrice α manja od 1.

Primjeri sustava za koje konvergira Jacobijeva metoda iteracije su tzv. dijagonalno dominantni sustavi. Sustav (47)

$$Ax = b$$

je **dijagonalno dominantan** ako vrijedi

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|, \quad i = 1, 2, \dots, n \quad (56)$$

Da u tim slučajevima dokažemo konvergenciju Jacobijeve metode pokažimo da je redna norma pripadne matrice α manja od 1. Prema (50) i (51) imamo

$$\alpha_{ij} = \begin{cases} -\frac{a_{ij}}{a_{ii}} & \text{za } i \neq j \\ 0 & \text{za } i = j \end{cases}$$

pa je pripadna redna norma jednaka

$$\begin{aligned} \|\alpha\|_r &= \max_i \sum_{j=1}^n |\alpha_{ij}| = \max_i \sum_{\substack{j=1 \\ j \neq i}}^n \left| \frac{a_{ij}}{a_{ii}} \right| \\ &= \max_i \frac{1}{|a_{ii}|} \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| < 1 \end{aligned}$$

pri čemu zadnja nejednakost slijedi iz (56). Teorem 3.8 povlači konvergentnost Jacobićeve metode u tom slučaju.

Primjetimo sljedeće. Ako u sustavu (47) provedemo supstituciju

$$x_j = c_j z_j, \quad c_j \neq 0, \quad j = 1, \dots, n \quad (57)$$

dobit ćemo sustav

$$A'z = b, \quad (58)$$

gdje je $a'_{ij} = a_{ij}c_j$. Sada vidimo da vrijedi sljedeće. Promatrana metoda za sustav (47) konvergira ako i samo ako konvergira za sustav (58). Ta tvrdnja slijedi direktno iz (57), jer $z_j^{(k)} \rightarrow z_j$ ako i samo ako $x_j^{(k)} \rightarrow c_j z_j$, $j = 1, 2, \dots, n$.

Ovu primjedbu ćemo iskoristiti ovako. Za gore uvedeni pojam dijagonalno dominantnog sustava mogli bi reći da je dijagonalno dominantan po recima. Što ako vrijedi

$$|a_{jj}| > \sum_{\substack{i=1 \\ i \neq j}}^n |a_{ij}|, \quad j = 1, 2, \dots, n, \quad (59)$$

da li i tada konvergira Jacobijeva metoda? Za sustav koji zadovoljava (59) mogli bi reći da je **dijagonalno dominantan po stupcima**. Pokažimo sljedeće: Jacobijeva metoda konvergira za sustav dijagonalno dominantan po stupcima.

Poslužit ćemo se supstitucijom (57) u kojoj odabiremo

$$c_j = \frac{1}{a_{jj}}, \quad j = 1, 2, \dots, n. \quad (60)$$

Elementi matrice A' pripadnog sustava (58) su sada jednaki

$$a'_{ij} = \frac{a_{ij}}{a_{jj}}, \quad i, j = 1, 2, \dots, n. \quad (61)$$

Predemo li na Jacobijevu iteraciju za taj novi sustav dobivamo

$$z = \alpha'z + \beta' \quad (62)$$

gdje je

$$\alpha'_{ij} = \begin{cases} -\frac{a_{ij}}{a_{jj}} & \text{za } i \neq j \\ 1 & \\ 0 & \text{za } i = j \end{cases} \quad (63)$$

Izračunamo li stupčanu normu od α' dobivamo zbog (59)

$$\|\alpha'\|_s = \sum_{\substack{j=1 \\ j \neq i}}^n |\alpha'_{ij}| = \sum_{\substack{j=1 \\ j \neq i}}^n \left| \frac{a_{ij}}{a_{jj}} \right| < 1,$$

čime smo dokaz proveli.

Dakle, Jacobijeva metoda iteracije je konvergentna za sustave dijagonalno dominantne bilo po retcima bilo po stupcima.

Gauss – Seidelova metoda iteracije

Polazimo ponovo od sustava jednadžbi

$$Ax = b, \quad \det A \neq 0 \quad (64)$$

i u standardnom rastavu $A = D - E - F$ izabiremo

$$B = D - E. \quad (65)$$

Sada je prema (9)

$$\begin{aligned} \alpha &= I - B^{-1}A = I - (D - E)^{-1}(D - E - F) = (D - E)^{-1}F, \\ \beta &= (D - E)^{-1}b, \end{aligned} \quad (66)$$

a pripadni se iteracijski postupak

$$x^{(k+1)} = (D - E)^{-1}Fx^{(k)} + (D - E)^{-1}b \quad (67)$$

zove **Gauss – Seidelova metoda iteracije**.

Da nađemo zapis Gauss – Seidelove iteracije u razvijenom obliku postupimo ovako. Nećemo računati $(D - E)^{-1}$ nego ćemo se poslužiti sa (7) tj. polazimo od

$$Bx^{(k+1)} + (A - B)x^{(k)} = b \quad (68)$$

što uvrštavanjem $B = D - E$ poprima oblik

$$(D - E)x^{(k+1)} - Fx^{(k)} = b$$

ili

$$(D - E)x^{(k+1)} = Fx^{(k)} + b. \quad (69)$$

Matrica $D - E$ je donja trokutasta i jednaka je

$$D - E = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1} & \dots & \dots & a_{nn} \end{bmatrix},$$

pa (69) lako raspišemo i taj zapis glasi

$$\sum_{j=1}^i a_{ij}x_j^{(k+1)} = - \sum_{j=i+1}^n a_{ij}x_j^{(k)} + b_i, \quad i = 1, 2, \dots, n. \quad (70)$$

Zapišemo li prvo zadnji pribrojnik lijeve strane imamo

$$a_{ii}x_i^{(k+1)} + \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} = - \sum_{j=i+1}^n a_{ij}x_j^{(k)} + b_i, \quad i = 1, 2, \dots, n$$

što nakon prenošenja drugog pribrojnika na desnu stranu i dijeljenja s a_{ii} daje ovaj krajnji zapis

$$x_i^{(k+1)} = \frac{b_i}{a_{ii}} - \sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{(k+1)} - \sum_{j=i+1}^n \frac{a_{ij}}{a_{ii}} x_j^{(k)}, \quad i = 1, 2, \dots, n. \quad (71)$$

Ako (71) raspišemo po indeksu i dobivamo

$$\begin{aligned} x_1^{(k+1)} &= \frac{b_1}{a_{11}} - \sum_{j=2}^n \frac{a_{1j}}{a_{11}} x_j^{(k)} \\ x_2^{(k+1)} &= \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}} x_1^{(k+1)} - \sum_{j=3}^n \frac{a_{2j}}{a_{22}} x_j^{(k)} \\ x_3^{(k+1)} &= \frac{b_3}{a_{33}} - \frac{a_{31}}{a_{33}} x_1^{(k+1)} - \frac{a_{32}}{a_{33}} x_2^{(k+1)} - \sum_{j=4}^n \frac{a_{3j}}{a_{33}} x_j^{(k)} \\ &\vdots \\ x_n^{(k+1)} &= \frac{b_n}{a_{nn}} - \sum_{j=1}^{n-1} \frac{a_{nj}}{a_{nn}} x_j^{(k+1)}. \end{aligned} \quad (72)$$

Usporedimo li Gauss – Seidelovu iteraciju s Jacobijskom možemo reći da je nastala iz Jacobijske tako da prilikom računanja nove aproksimacije nepoznanice x_i u iteracijski korak uvrstimo nove aproksimacije prethodnih nepoznanica x_1, x_2, \dots, x_{i-1} , dok za daljnje nepoznanice $x_{i+1}, x_{i+2}, \dots, x_n$ uvrštavamo prethodne aproksimacije. Još se ponekad kaže da je Jacobijska metoda **totalno koračna** dok se za Gauss – Seidelovu kaže da je **jedno koračna**.

Prirodno je pitanje usporedba Jacobijske i Gauss – Seidelove metode. Obično Gauss – Seidelova metoda ubrzava konvergenciju iako vodi na opsežnije računanje. Isto tako, ako jedna metoda konvergira ne znači da i druga i obratno, čak su mogući slučajevi da Gauss – Seidelova iteracija konvergira sporije od Jacobijske.

Dok smo kod Jacobijske metode veoma lako utvrdili da konvergira za po recima dijagonalno dominantne sustave to je za Gauss – Seidelovu metodu nešto složenije no ipak vrijedi:

Ako je sustav $Ax = b$ dijagonalno dominantan po recima, tj. ako vrijedi

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \quad i = 1, 2, \dots, n,$$

onda je Gauss – Seidelova metoda konvergentna.

Drugi važan slučaj sustava za koje konvergira Gauss – Seidelova iteracija su tzv. pozitivno definitni sustavi. Za sustav

$$Ax = b$$

kažemo da je **pozitivno definitan** ako je A simetrična matrica, $A = A^T$ (transponirana matrica), i ako je matrici A pridružena kvadratna forma (polinom 2. stupnja)

$$u(x_1, \dots, x_n) = x^T A x = a_{11}x_1^2 + a_{22}x_2^2 + \dots + a_{nn}x_n^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + \dots + 2a_{n-1,n}x_{n-1}x_n$$

pozitivno definitna. Da je $u(x_1, \dots, x_n)$ pozitivno definitna znači da je $u(x_1, \dots, x_n) > 0$ osim za $x_1 = x_2 = \dots = x_n = 0$ kada poprima vrijednost nula.

Navedimo neke činjenice o pozitivno definitnim matricama. Za kvadratnu matricu A imamo uvjek definirane glavne podmatrice reda $1, 2, \dots, n-1, n$ kao

$$A_i = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1i} \\ a_{21} & a_{22} & \dots & a_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ii} \end{bmatrix}, \quad i = 1, 2, \dots, n.$$

Glavne minore su subdeterminante kod kojih je dijagonalna sastavljena od elemenata dijagonale od A . Npr. jednoredne glavne minore su dijagonalni elementi matrice A . Kriterij pozitivne definitnosti glasi ovako:

Teorem 3.10. *A je pozitivno definitna ako i samo ako je*

$$\det(A_i) > 0 \quad \text{za } i = 1, 2, \dots, n$$

Navedimo i ova svojstva pozitivno definitnih matrica.

- (i) Ako je A pozitivno definitna onda postoji A^{-1} koja je također pozitivno definitna.
- (ii) Sve glavne podmatrice su pozitivno definitne.
- (iii) Sve glavne minore su pozitivne.

Iz (iii) vidimo da su svi dijagonalni elementi pozitivno definitne matrice pozitivni.

Dakle, za pozitivno definitne sustave vrijedi: *Ako je sustav $Ax = b$ pozitivno definitan onda za njega konvergira Gauss–Seidelova metoda iteracije.*

Ta tvrdnja je poseban slučaj Teorema 3.11.

Na prvi pogled može izgledati da je gornja tvrdnja o konvergenciji Gauss–Seidelove metode za pozitivno definitne sustave ograničenog dometa. Međutim, svaki sustav

$$Ax = b, \quad \det A \neq 0$$

ekvivalentan je sustavu

$$(A^T A)x = A^T b \tag{73}$$

koji je pozitivno definitan.

Da je sustav (73) pozitivno definitan dosta se lako dokazuje. Iz

$$(A^T A)^T = A^T A$$

čitamo odmah da je matrica sustava (73) simetrična. Ispišemo li pripadnu kvadratnu formu za $A^T A$ imamo

$$\begin{aligned} u(x_1, x_2, \dots, x_n) &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_{ki} a_{kj} x_i x_j \\ &= \sum_{k=1}^n \sum_{i=1}^n \sum_{j=1}^n a_{ki} x_i a_{kj} x_j \\ &= \sum_{k=1}^n \left(\sum_{i=1}^n a_{ki} x_i \sum_{j=1}^n a_{kj} x_j \right). \end{aligned}$$

Kako vrijednost sume ne ovisi o oznaci indeksa sumacije to imamo

$$u(x_1, x_2, \dots, x_n) = \sum_{k=1}^n \left(\sum_{i=1}^n a_{ki} x_i \right)^2 \geq 0$$

jer je suma kvadrata nenegativna. No, suma kvadrata je nula ako i samo ako su svi pribrojnici nula, tj.

$$\sum_{i=1}^n a_{ki} x_i = 0, \quad k = 1, 2, \dots, n$$

što u matričnom zapisu daje

$$Ax = 0.$$

Vidimo da smo dobili homogen sustav pridružen sustavu $Ax = b$. Zbog $\det A \neq 0$ homogen sustav ima samo trivijalno rješenje iz čega slijedi da je kvadratna forma sustava (73) pozitivno definitna, odnosno, da je sustav (73) pozitivno definitan.

Spomenimo da će za pozitivno definitne sustave konvergirati SOR metoda čiji je poseban slučaj Gauss–Seidelova metoda. S druge strane opet vidimo da se, teorijski gledano, svaki sustav može rješavati Gauss–Seidelovom metodom prelaskom na sustav (73) koji ima isto rješenje. Taj prelaz svakako uključuje množenje s A^T što je dodatno računanje.

Rješavanje sustava linearnih jednadžbi iteracijskim metodama kao što su Jacobijska i Gauss–Seidelova metoda ponekad se zove **relaksacija**. Smisao tog naziva dolazi odatle što se pogreška početne aproksimacije postupno smanjuje ili relaksira nastavkom računanja. Gauss–Seidelova metoda i SOR metoda uvelike se koriste prilikom aproksimacije rješenja parcijalnih diferencijalnih jednadžbi metodom konačnih diferencija i u drugim problemima. U sljedećem paragrafu je prikazana jedna vrlo jednostavna metoda relaksacije.

Metoda relaksacije

Promatramo ponovo sustav linearnih jednadžbi

$$Ax = b, \quad \det A \neq 0 \tag{74}$$

i prevedimo ga u oblik

$$0 = -Ax + b. \tag{75}$$

Ako u (75) uvrstimo neku aproksimaciju $x^{(k)}$ točnog rješenja $x = A^{-1}b$, onda na lijevoj strani nećemo dobiti 0 nego neki ostatak $r^{(k)}$, tj. imamo

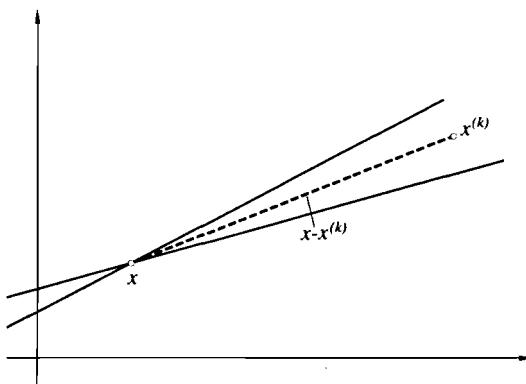
$$r^{(k)} = -Ax^{(k)} + b, \quad (76)$$

koji ostatak možemo smatrati mjerom koliko dobro $x^{(k)}$ zadovoljava sustav jednadžbi (74). Ako su komponente od $r^{(k)}$ malene, onda $x^{(k)}$ gotovo da zadovoljava sustav jednadžbi (74). Metoda relaksacije se sastoji u tome da postupno smanjujemo komponente od $r^{(k)}$. Primijetimo sljedeće. Pomnožimo li (76) s A^{-1} slijeva dobivamo

$$A^{-1}r^{(k)} = A^{-1}b - x^{(k)}. \quad (77)$$

Kako je $A^{-1}b$ točno rješenje sustava (74), to desna strana jednakosti (77) predstavlja pogrešku aproksimacije, jer je to razlika između točnog rješenja $x = A^{-1}b$ i aproksimacije $x^{(k)}$.

Promatramo li lijevu stranu od (77) vidimo da je ona produkt $A^{-1}r^{(k)}$ koji može imati velike komponente iako $r^{(k)}$ ima male komponente. Pokušajmo to ilustrirati grafički. Zamislimo da rješavamo sustav dvije jednadžbe s dvije nepoznanice. Grafički gledano tražimo koordinate sjecišta dvaju pravaca. Zamislimo situaciju ilustriranu sljedećom slikom:



Sl. 3.1.

Na slici je x njihovo sjecište, dakle točno rješenje sustava. Ako je $x^{(k)}$ aproksimacija kao na slici, onda će $x^{(k)}$ gotovo zadovoljavati obje jednadžbe, jer je $x^{(k)}$ blizu i jednog i drugog pravca, tj. komponente od $r^{(k)}$ su male. Međutim, pogreška je velika jer je predstavljena vektorom (crtkano) koji spaja x i $x^{(k)}$. Kada bi kut između tih pravaca bio veći uz isto sjecište x , za istu aproksimaciju $x^{(k)}$ bi $r^{(k)}$ imao veće komponente i smatrali bi ga slabom aproksimacijom. Vidimo da to ovisi o promatranoj sustavu jednadžbi $Ax = b$ ili kratko rečeno o matrici A . Pojava koju smo ilustrirali gornjom slikom događa se kod tzv. **nestabilnih ili slabo uvjetovanih** sustava (engl. *ill-conditioned systems*), a to su oni kod kojih je $\det A$ malena u usporedbi s elementima a_{ij} matrice A . Tada male promjene elemenata (npr. izazvane zaokruživanjem) mogu izazvati velike promjene rješenja. Iako je tako, relaksacijske metode se koriste zbog jednostavnog računanja. Često se koriste da se poboljša $x^{(0)}$, pa se tako poboljšana početna aproksimacija iskoristi u drugoj metodi.

Ovdje opisujemo jedan jednostavan postupak relaksacije. Prevedimo sustav (74) u oblik $x = \alpha x + \beta$, odnosno

$$-x + \alpha x + \beta = 0 \quad (78)$$

gdje su kao i kod Jacobijeve metode

$$\begin{aligned} \alpha_{ij} &= \begin{cases} -\frac{a_{ij}}{a_{ii}} & \text{za } i \neq j, \\ 0 & \text{za } i = j, \quad i, j = 1, 2, \dots, n \end{cases} \\ \beta_i &= \frac{b_i}{a_{ii}} \quad i = 1, 2, \dots, n. \end{aligned}$$

U razvijenom obliku sustav (78) izgleda ovako:

$$\begin{aligned} -x_1 &+ \alpha_{12}x_2 + \alpha_{13}x_3 + \dots + \alpha_{1n}x_n + \beta_1 = 0 \\ -x_2 &+ \alpha_{21}x_1 + \alpha_{23}x_3 + \dots + \alpha_{2n}x_n + \beta_2 = 0 \\ &\vdots \\ -x_n &+ \alpha_{n1}x_1 + \alpha_{n2}x_2 + \alpha_{n3}x_3 + \dots + \beta_n = 0. \end{aligned} \quad (79)$$

Neka je $x^{(0)} = [x_1^{(0)} \ x_2^{(0)} \ \dots \ x_n^{(0)}]^T$ početna aproksimacija. Uvrstimo li $x^{(0)}$ u (78) dobivamo

$$r^{(0)} = -x^{(0)} + \alpha x^{(0)} + \beta$$

ili u razvijenom obliku

$$\begin{aligned} r_1^{(0)} &= -x_1^{(0)} + \sum_{j=1}^n \alpha_{1j}x_j^{(0)} + \beta_1 \\ &\vdots \\ r_s^{(0)} &= -x_s^{(0)} + \sum_{j=1}^n \alpha_{sj}x_j^{(0)} + \beta_s \\ &\vdots \\ r_n^{(0)} &= -x_n^{(0)} + \sum_{j=1}^n \alpha_{nj}x_j^{(0)} + \beta_n \end{aligned} \quad (80)$$

Ako sada aproksimaciji s -te komponente $x_s^{(0)}$ dodamo prirast $\delta x_s^{(0)}$, onda iz (80) vidimo da će se odgovarajući $r_s^{(0)}$ smanjiti za $\delta x_s^{(0)}$ dok će se svi ostali $r_i^{(0)}$, $i \neq s$, povećati za $\alpha_{is}\delta x_s^{(0)}$. Prema tome, da postignemo $r_s^{(1)} = 0$, dovoljno je izabrati prirast

$$\delta x_s^{(0)} = r_s^{(0)}, \quad (81)$$

jer iz (80) slijedi

$$0 = -(x_s^{(0)} + r_s^{(0)}) + \sum_{j=1}^n \alpha_{sj}x_j^{(0)} + \beta_s,$$

dok će se svi ostali $r_i^{(1)}$ promijeniti i iznositi

$$r_i^{(1)} = r_i^{(0)} + \alpha_{is} r_s^{(0)}, \quad i \neq s. \quad (82)$$

Jasno je sada da smo poništavanjem $r_s^{(0)}$ zadovoljili s -tu jednadžbu i da poništavanjem svih r_i dobivamo rješenje sustava jer su tada sve jednadžbe zadovoljene. Ako se pitamo kako sada izgleda nova aproksimacija $x^{(1)}$ onda po komponentama imamo

$$x_j^{(1)} = \begin{cases} x_j^{(0)} & \text{za } j \neq s \\ x_s^{(0)} + r_s^{(0)} & \text{za } j = s. \end{cases}$$

tj. nova aproksimacija $x^{(1)}$ ima promijenjenu samo s -tu komponentu. (Geometrijski gledano $x^{(1)}$ je točka koja se nalazi na s -toj hiperravnini i u sljedećem će koraku preći na neku drugu hiperravninu, itd.)

U najjednostavnijem obliku metoda relaksacije se sastoji u tome da u k -tom koraku poništimo ovaj $r_i^{(k-1)}$ koji ima **najveću apsolutnu vrijednost**. Postupak zauzavljamo kada su svi r_i po apsolutnoj vrijednosti manji od zadanih brojeva. Istaknimo na kraju da je ta metoda konvergentna.

SOR metoda ili metoda sukcesivne nadrelaksacije

Promatrajmo sustav linearnih jednadžbi

$$Ax = b, \quad \det A \neq 0 \quad (83)$$

i prisjetimo se kako je na početku uvedena linearna iteracija pomoću regularne matrice B . Sustavu (83) smo dodali $Bx - b = 0$ i dobili

$$Bx + (A - B)x = b$$

te množenjem s B^{-1} izračunali x kao

$$x = B^{-1}b + B^{-1}(B - A)x \quad (84)$$

otkuda smo dobili iteracijski postupak

$$x^{(k+1)} = B^{-1}b + (I - B^{-1}A)x^{(k)}. \quad (85)$$

Uz pokrate $\alpha = I - B^{-1}A$ i $\beta = B^{-1}b$ imali smo zapis

$$x^{(k+1)} = \alpha x^{(k)} + \beta. \quad (86)$$

U formulama (35)–(38) smo uveli rastav matrice $A = D - E - F$, pa smo zatim uzeli da je $B = D$ i dobili Jacobijevu metodu, dok smo uvezši $B = D - E$ dobili Gauss–Seidelovu metodu. Uvedimo sada **SOR metodu ili metodu sukcesivne nadrelaksacije** (SOR dolazi od engleskog *successive over-relaxation*). Kod te metode koristimo sljedeću familiju matrica $B(\omega)$:

$$B(\omega) = \frac{1}{\omega}(D - \omega E) \quad (87)$$

pri čemu se ω zove **relaksacijski parametar**. Odmah vidimo da izbor $\omega = 1$ daje Gauss – Seidelovu metodu, tako da se $\omega < 1$ zove **podrelaksacijski parametar**, a $\omega > 1$ **nadrelaksacijski parametar**.

Da osiguramo konvergenciju metode nužno je i dovoljno da spektralni radius matrice

$$\alpha(\omega) = I - B^{-1}(\omega)A \quad (88)$$

bude manji od 1, tj. da vrijedi

$$\rho(\alpha(\omega)) < 1. \quad (89)$$

Pokazat ćemo da za $\rho(\alpha(\omega))$ vrijedi

$$\rho(\alpha(\omega)) \geq |\omega - 1|. \quad (90)$$

Kombiniramo li (89) i (90) dobivamo ove granice za ω

$$0 < \omega < 2. \quad (91)$$

Dakle, $\omega \in (0, 2)$ može dati konvergentnu metodu.

Dokažimo nejednakost (90). U prvom redu zapišimo $B(\omega)$ na ovaj način

$$B(\omega) = \frac{1}{\omega}(D - \omega E) = \frac{1}{\omega}D(I - \omega D^{-1}E) \quad (92)$$

i izračunajmo

$$\begin{aligned} B(\omega) - A &= \frac{1}{\omega}D(I - \omega D^{-1}E) - D + E + F \\ &= \frac{1}{\omega}D[I - \omega D^{-1}E - \omega I + \omega D^{-1}E + \omega D^{-1}F] \\ &= \frac{1}{\omega}D[(1 - \omega)I + \omega D^{-1}F]. \end{aligned} \quad (93)$$

Pomnožimo li (93) s $B^{-1}(\omega)$ dobivamo ovakav zapis od $\alpha(\omega)$

$$\begin{aligned} \alpha(\omega) &= I - B^{-1}(\omega)A = \frac{1}{\omega}B^{-1}(\omega)D[(1 - \omega)I + \omega D^{-1}F] \\ &= \frac{1}{\omega}\left[\frac{1}{\omega}D(I - \omega D^{-1}E)\right]^{-1}D[(1 - \omega)I + \omega D^{-1}F] \\ &= \frac{1}{\omega}\omega(I - \omega D^{-1}E)^{-1}D^{-1}D[(1 - \omega)I + \omega D^{-1}F] \\ &= (I - \omega D^{-1}E)^{-1}[(1 - \omega)I + \omega D^{-1}F]. \end{aligned} \quad (94)$$

Prisjetimo se Binet – Cauchyjeva teorema koji kaže da je determinanta produkta dviju matrica jednaka produktu determinanti. Uočimo nadalje da je matrica

$$I - \omega D^{-1}E \quad (95)$$

donje trokutasta i da su njeni dijagonalni elementi jedinice koje dolaze od pribrojnika I jer je ωDE donje trokutasta s dijagonalnim elementima jednakim nuli. Dakle, neovisno o ω vrijedi

$$\det(I - \omega D^{-1}E) = 1. \quad (96)$$

Da ocijenimo $\rho(\alpha(\omega))$. Promatrajmo karakteristični polinom od $\alpha(\omega)$, tj.

$$\varphi(\lambda) = \det(\lambda I - \alpha(\omega)).$$

Zbog (96) i Binet – Cauchyjeva teorema vrijedi

$$\begin{aligned} \varphi(\lambda) &= \det(\lambda I - \alpha(\omega)) = \det((I - \omega D^{-1}E)(\lambda I - \alpha(\omega))) \\ &= \det((I - \omega D^{-1}E)[\lambda I - (I - \omega D^{-1}E)^{-1}[(1 - \omega)I + \omega D^{-1}F]]) \\ &= \det(\lambda(I - \omega D^{-1}E) - (1 - \omega)I - \omega D^{-1}F) \\ &= \det((\lambda + \omega - 1)I - \lambda \omega D^{-1}E - \omega D^{-1}F). \end{aligned} \quad (97)$$

Kako je vodeći koeficijent karakterističnog polinoma 1, to je konstantni član $\varphi(0)$ jednak produktu njegovih korijena tj. produktu svojstvenih vrijednosti $\lambda_i(\alpha(\omega))$. Imamo dakle da je

$$\prod_{i=1}^n \lambda_i(\alpha(\omega)) = \varphi(0) = \det((\omega - 1)I - \omega D^{-1}F) = (\omega - 1)^n, \quad (98)$$

jer je $(\omega - 1)I - \omega D^{-1}F$ gornje trokutasta matrica čiji su svi dijagonalni elementi jednaki $\omega - 1$. Iz (98) zaključujemo da je

$$\rho(\alpha(\omega)) = \max_i |\lambda_i(\alpha(\omega))| \geq |\omega - 1|,$$

čime smo dokazali nejednakost (90).

Dokazavši (90) izvršimo jedno kratko razmatranje koje nas navodi na izbor matrice $B(\omega)$ dan formulom (87). Kako smo već rekli, za $\omega = 1$ dobivamo Gauss–Seidelovu iteraciju, pa je za očekivati da je ista izvor formule (87). Pretpostavimo sada da su nam poznate komponente

$$x_j^{(k+1)} \quad \text{za } j = 1, 2, \dots, i-1,$$

$(k+1)$ -ve aproksimacije $x^{(k+1)}$. Uvedimo pomoćnu veličinu $\tilde{x}_i^{(k+1)}$ prema Gauss–Seidelovoj iteraciji (formula (71)) tj. tada za $\tilde{x}_i^{(k+1)}$ vrijedi

$$a_{ii}\tilde{x}_i^{(k+1)} = b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij}x_j^{(k)}. \quad (99)$$

Možemo očekivati da ćemo poboljšati $(k+1)$ -vu aproksimaciju od x_i ako za nju uzmemos težinsku aritmetičku sredinu $x_i^{(k)}$ (prethodne aproksimacije od x_i) i $\tilde{x}_i^{(k+1)}$ koju daje (99). Prema toj zamisli bi poboljšana aproksimacija od x_i bila dana sa

$$x_i^{(k+1)} = (1 - \omega)x_i^{(k)} + \omega x_i^{(k+1)} = x_i^{(k)} + \omega(\tilde{x}_i^{(k+1)} - x_i^{(k)}). \quad (100)$$

Da nađemo poboljšanu aproksimaciju $x_i^{(k+1)}$ treba iz (100) eliminirati $\tilde{x}_i^{(k+1)}$ i smisleno odabrati težinu ω . Pomnožimo (100) s a_{ii} i u dobiveni rezultat uvrstimo $a_{ii}\tilde{x}_i^{(k+1)}$:

$$a_{ii}x_i^{(k+1)} = a_{ii}x_i^{(k)} + \omega \left[b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij}x_j^{(k)} - a_{ii}x_i^{(k)} \right] \quad i = 1, 2, \dots, n \quad k \geq 0, \quad (101)$$

otkuda možemo prepoznati (87). Dijeljenjem (101) s ω i prebacivanjem na lijevu stranu prve sume imamo

$$\frac{1}{\omega}a_{ii}x_i^{(k+1)} - \sum_{j=1}^{i-1} (-a_{ij})x_j^{(k+1)} = \frac{1}{\omega}a_{ii}x_i^{(k)} - a_{ii}x_i^{(k)} + \sum_{j=i+1}^n (-a_{ij})x_j^{(k)} + b_i \\ i = 1, 2, \dots, n \quad k \geq 0, \quad (102)$$

Iz (102) čitamo ovaj matrični zapis

$$\frac{1}{\omega}Dx^{(k+1)} - Ex^{(k+1)} = \frac{1}{\omega}Dx^{(k)} - Dx^{(k)} + Fx^{(k)} + b. \quad (103)$$

Iz (103) lako izračunamo

$$\begin{aligned} \left(\frac{1}{\omega}D - E\right)x^{(k+1)} &= \left(\frac{1}{\omega}D - D + F\right)x^{(k)} + b \\ &= \left[\left(\frac{1}{\omega}D - E\right) - (D - E - F)\right]x^{(k)} + b \\ &= \left[\left(\frac{1}{\omega}D - E\right) - A\right]x^{(k)} + b \end{aligned}$$

ili

$$\left(\frac{1}{\omega}D - E\right)x^{(k+1)} + \left[A - \left(\frac{1}{\omega}D - E\right)\right]x^{(k)} = b. \quad (104)$$

Označimo li sa

$$B(\omega) = \frac{1}{\omega}D - E = \frac{1}{\omega}(D - \omega E)$$

dobivamo (85), pa možemo reći da su SOR iteracije nastale u nastojanju da se poboljša Gauss-Seidelova iteracija.

Dokažimo sljedeći teorem.

Teorem 3.11. Za pozitivno definitne sustave i svaki $\omega \in (0, 2)$ konvergira SOR metoda.

Dokaz. Da dokažemo konvergenciju dovoljno je (a i nužno) da dokažemo da je spektralni radius (vidi teorem 3.6)

$$\rho(\alpha(\omega)) = \rho(I - B(\omega)^{-1}A) < 1 \quad \text{za } \omega \in (0, 2).$$

Kao i u dokazu nejednakosti (90) služit ćemo se osnovnim činjenicama i spretnim računanjem. Kako je A pozitivno definitna, to je A simetrična s pozitivnim dijagonalnim elementima, pa zbog simetrije vrijedi $F = E^T$. Promatrajmo matrice

$$B = \frac{1}{\omega}D - E \quad (105)$$

i

$$\begin{aligned} B + B^T - A &= \frac{1}{\omega}D - E + \frac{1}{\omega}D^T - E^T - (D - E - F) \\ &= \frac{1}{\omega}D - E + \frac{1}{\omega}D - F - D + E + F \\ &= \left(\frac{2}{\omega} - 1\right)D, \end{aligned} \quad (106)$$

gdje pišemo kraće B umjesto $B(\omega)$. Sada $\omega \in (0, 2)$ ima za posljedicu da je $\frac{2}{\omega} - 1 > 0$, pa su obje te matrice pozitivno definitne, jer je B donje trokutasta s pozitivnim elementima na dijagonalni, a druga je dijagonalna s isto tako pozitivnim elementima na dijagonalni.

Kao prvo dokažimo da su svojstvene vrijednosti λ matrice

$$A^{-1}(2B - A) \quad (107)$$

pozitivne. Naime, ako je x svojstveni vektor za λ , onda vrijedi

$$A^{-1}(2B - A)x = \lambda x$$

ili množenjem s A

$$(2B - A)x = \lambda Ax. \quad (108)$$

Pomnožimo li (108) s x^T slijeva dobivamo

$$x^T(2B - A)x = \lambda x^T Ax, \quad (109)$$

otkuda čitamo da je faktor od λ upravo pridružena kvadratna forma matrice A . Transponiranjem (109) imamo

$$x^T(2B^T - A)x = \lambda x^T Ax, \quad (110)$$

pa zbrajanjem (109) i (110) dobivamo

$$x^T(B + B^T - A)x = \lambda x^T Ax. \quad (111)$$

Kako smo vidjeli $B + B^T - A$ je pozitivno definitna, pa je lijeva strana od (111) pridružena kvadratna forma matrice $B + B^T - A$, dakle za $x \neq 0$ je pozitivna. Kako je i $x^T Ax$ pozitivna za $x \neq 0$ slijedi da je $\lambda > 0$. Time smo dokazali da matrica (107) ima pozitivne svojstvene vrijednosti.

Sljedeći trik kojim ćemo se poslužiti je spretan zapis matrice $\alpha(\omega)$. Označimo s Q matricu

$$Q = A^{-1}(2B - A) = 2A^{-1}B - I, \quad (112)$$

za koju vrijede ove jednakosti

$$Q - I = 2A^{-1}B - 2I \quad (113)$$

$$Q + I = 2A^{-1}B \quad (114)$$

$$(Q + I)^{-1} = \frac{1}{2}B^{-1}A. \quad (115)$$

Pomnožimo li međusobno prvu i treću jednakost dobivamo

$$(Q - I)(Q + I)^{-1} = (2A^{-1}B - 2I)\frac{1}{2}B^{-1}A = I - B^{-1}A = \alpha(\omega). \quad (116)$$

Kako je B regularna trokutasta matrica to postoji B^{-1} , pa iz (115) slijedi da postoji i $(Q + I)^{-1} = \frac{1}{2}B^{-1}A$.

Označimo sa μ svojstvenu vrijednost od $\alpha(\omega)$ i s x pripadni svojstveni vektor. Tada iz

$$(Q - I)(Q + I)^{-1}x = \alpha(\omega)x = \mu x \quad (117)$$

slijedi da za vektor

$$y = (Q + I)^{-1}x \neq 0, \quad x = (Q + I)y \quad (118)$$

vrijedi

$$(Q - I)y = \mu(Q + I)y$$

ili

$$(1 - \mu)Qy = (1 + \mu)y. \quad (119)$$

Kako je $y \neq 0$, slijedi da mora biti $\mu \neq 1$, tako da dobivamo

$$Qy = \frac{1 + \mu}{1 - \mu}y,$$

tj. $\lambda = \frac{1+\mu}{1-\mu}$ je svojstvena vrijednost od $Q = A^{-1}(2B - A)$ što smo tražili. Izračunamo li μ dobivamo

$$\mu = \frac{\lambda - 1}{\lambda + 1},$$

pri čemu smo već prije vidjeli da je $\lambda > 0$ za svaki $0 < \omega < 2$, pa smo zapravo dobili da je $|\mu| < 1$, odnosno da je $\rho(\alpha(\omega)) < 1$. Time je dokaz kompletiran. Kako se Gauss–Seidelova iteracija dobiva za $\omega = 1$, to smo dobili spomenuti rezultat o konvergenciji Gauss–Seidelove iteracije za pozitivno definitne matrice.

4.

Svojstvene vrijednosti i svojstveni vektori matrice

4.1. Metoda neodređenih koeficijenata	115
4.2. Danilevskijeva metoda	117
4.3. Krilovljeva metoda	124
4.4. Nalaženje karakterističnog polinoma Leverrierovom metodom	127
4.5. Misesova metoda potencija — spektralni radijus matrice	129

Neka je $A = [a_{ij}]$ kvadratna matrica i $I = [\delta_{ij}]$ jedinična matrica istog reda. Uz svaku kvadratnu matricu (linearni operator) A vezan je poseban skup vektora koje zovemo **svojstveni vektori matrice** A i poseban skup skalarova koje zovemo **svojstvene vrijednosti matrice** A . Skalar λ je svojstvena vrijednost od A ako postoji **ne-nul** vektor $x \neq 0$ za koji vrijedi

$$Ax = \lambda x. \quad (1)$$

Sam vektor x se zove **svojstveni vektor matrice** A **koji pripada svojstvenoj vrijednosti** λ . Jednakost (1) možemo pomnožiti skalarom $\mu \neq 0$ što daje

$$A(\mu x) = \lambda(\mu x)$$

iz čega čitamo da je svaki vektor kolinearan sa svojstvenim vektorom i sam svojstveni vektor, odnosno, svojstveni vektor je određen do na skalar $\mu \neq 0$. Jednakost (1) možemo zapisati u obliku

$$(\lambda I - A)x = 0 \quad (2)$$

ili

$$(A - \lambda I)x = 0 \quad (3)$$

što je homogen linearan sustav jednadžbi u x . Kako homogen sustav ima netrivijalnih rješenja $x \neq 0$ ako i samo ako je determinanta sustava nula, to dobivamo da su svojstvene vrijednosti od A korijeni algebarske jednadžbe

$$D(\lambda) = \det(\lambda I - A) = 0 \quad (4)$$

ili

$$D_1(\lambda) = \det(A - \lambda I) = 0. \quad (5)$$

Kako je $D(\lambda) = (-1)^n D_1(\lambda)$ to algebarske jednadžbe (4) i (5) imaju iste korijene. Ovo ističemo zbog toga što se u pojedinim knjigama (4) zove karakteristična jednadžba matrice A , dok se u drugim tako zove (5). $D(\lambda) = 0$ i $D_1(\lambda) = 0$ su iste jednadžbe, međutim, $D(\lambda)$ i $D_1(\lambda)$ nisu isti polinomi. Nekada je spretno raditi s $D(\lambda)$, a nekada s $D_1(\lambda)$. Mi ćemo zvati

$$D(\lambda) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \dots & \lambda - a_{nn} \end{vmatrix} \quad (6)$$

$$= \lambda^n + p_1 \lambda^{n-1} + p_2 \lambda^{n-2} + \dots + p_n$$

karakteristični polinom matrice A . Očigledno je da je to polinom n -tog stupnja, jer se u razvoju determinante pojavljuje pribrojnik $(\lambda - a_{11})(\lambda - a_{22}) \dots (\lambda - a_{nn})$. Razvije li se determinanta (6) i rezultat sredi po potencijama od λ , dobit će se koeficijenti p_i . Taj rezultat glasi da je do na predznak koeficijent p_i jednak sumi svih dijagonalnih minora i -tog reda determinante

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

Točno vrijedi

$$p_1 = - \sum_{i=1}^n a_{ii}$$

$$p_2 = \sum_{i < j} \begin{vmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{vmatrix}$$

$$p_3 = - \sum_{i < j < k} \begin{vmatrix} a_{ii} & a_{ij} & a_{ik} \\ a_{ji} & a_{jj} & a_{jk} \\ a_{ki} & a_{kj} & a_{kk} \end{vmatrix}$$

$$\vdots$$

$$p_n = (-1)^n \det(A). \quad (7)$$

S druge strane Vièteov teorem daje

$$D(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n)$$

gdje su λ_i rješenja karakteristične jednadžbe $D(\lambda) = 0$, tj. svojstvene vrijednosti matrice A , pa imamo

$$D(0) = (-1)^n \lambda_1 \lambda_2 \dots \lambda_n = p_n = (-1)^n \det(A)$$

odnosno

$$\det(A) = \lambda_1 \lambda_2 \dots \lambda_n. \quad (8)$$

Dakle, p_n je do na predznak jednak $\det(A)$ koja je s druge strane jednaka produktu svojstvenih vrijednosti. Iz (8) slijedi da nula nije svojstvena vrijednost regularne

matrice i da je nula barem jedna svojstvena (kratnost!) vrijednost svake singularne matrice.

Istaknimo da je broj dijagonalnih minora k -tog reda jednak

$$\binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{k!},$$

pa da prema (7) izračunamo koeficijente karakterističnog polinoma potrebno je ukupno izračunati

$$\binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n} = 2^n - 1$$

determinanta različitog reda. To je računski pozamašan posao, pa su nam potrebne druge računski spretneje metode.

Ponovimo još neke činjenice iz linearne algebri vezane za svojstvene vrijednosti i svojstvene vektore:

1. Ako je A regularna i λ njena svojstvena vrijednost, onda je $1/\lambda$ svojstvena vrijednost njene inverzne matrice A^{-1} .
2. Ako je $f(A)$ polinom od A ($f(A)$ je opet jedna matrica), onda su $f(\lambda_i)$ svojstvene vrijednosti od $f(A)$, gdje su λ_i svojstvene vrijednosti matrice A .
3. Ako A ima n različitih svojstvenih vrijednosti, onda A ima n linearno nezavisnih svojstvenih vektora.
4. Ako A ima višestruke svojstvene vrijednosti, onda A može imati manje od n svojstvenih vektora. Međutim, ako je A hermitska ili realna simetrična, onda A ima n linearno nezavisnih svojstvenih vektora. Nadalje, svojstvene vrijednosti hermitских ili realnih simetričnih su realni brojevi.
5. Ako je λ_i svojstvena vrijednost od A , onda je $\overline{\lambda_i}$ (konjugirano kompleksna vrijednost) svojstvena vrijednost njoj adjungirane matrice A^* .
6. Svaka matrica zadovoljava svoju karakterističnu jednadžbu (Hamilton Caylejev teorem), tj. vrijedi

$$A^n + p_1 A^{n-1} + \dots + p_n I = 0.$$

7. Slične matrice imaju jednake karakteristične polinome, pa prema tome i jednake svojstvene vrijednosti. Prisjetimo se da su A i B **slične**, ako postoji regularna matrica T za koju vrijedi

$$B = T^{-1}AT.$$

Matrica T je općenito kompleksna i kada polazimo od realne matrice A .

Prisjetimo se i elementarnih transformacija matrica:

I. Zamjena dvaju redaka ili dvaju stupaca.

II. Množenje retka (stupca) brojem različitim od nule.

III. Pribrojanje elementa jednog retka (stupca) odgovarajućim elementima nekog drugog retka (stupca) pomnoženih brojem različitim od nule.

Za matrice kažemo da su ekvivalentne ako se jedna dobiva iz druge pomoću konično elementarnih transformacija. Osnovno svojstvo ekvivalentnih matrica je da imaju isti rang.

Lako se uvjerimo da se svaka elementarna transformacija može realizirati pomoću množenja matrica i to tako da matricu A pomnožimo s regularnom matricom koja se dobiva iz jedinične matrice I na ovaj način:

Ako se elementarna transformacija provodi na retcima (stupcima) od A , onda matricu A treba pomnožiti slijeva (zdesna) matricom koja se dobiva iz jedinične matrice I tako da se na I provede ista elementarna transformacija.

Tako npr. da u matrici

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

zamijenimo drugi i treći redak treba A pomnožiti slijeva s

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

što daje

$$EA = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}.$$

Da drugom stupcu od A pribrojimo prvi pomnožen s 2 treba A pomnožiti zdesna s

$$E = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

što daje

$$AE = \begin{bmatrix} a_{11} & 2a_{11}+a_{12} & a_{13} \\ a_{21} & 2a_{21}+a_{22} & a_{23} \\ a_{31} & 2a_{31}+a_{32} & a_{33} \end{bmatrix}.$$

Nama će odmah zatrebati i inverzna matrica ovako transformirane jedinične matrice. Tako, ako je E nastala iz I zamjenom dvaju redaka (stupaca), onda se E^{-1} dobiva iz I zamjenom istoimenih redaka (stupaca), tj. $E^{-1} = E$. Ako je E dobivena iz I množenjem retka (stupca) s $c \neq 0$, onda se E^{-1} dobiva iz I množenjem istoimenog retka (stupca) s $\frac{1}{c}$. Ako je E dobivena iz I zbrajanjem i -toga retka (stupca) pomnoženog s $\lambda \neq 0$ k -tom retku (stupcu) onda se E^{-1} dobiva tako da se i -ti redak (stupac) pomnoži s $-\lambda$ i probroji k -tom retku (stupcu). Tako npr. ako je redom:

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \text{onda je} \quad E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix};$$

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad \text{onda je} \quad E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix};$$

$$E = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{onda je} \quad E^{-1} = \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Da nađemo svojstvene vrijednosti matrice treba riješiti algebarsku jednadžbu (4). Za matrice reda ≤ 4 to možemo riješiti i direktno. Ako želimo računati svojstvene vrijednosti za $n \geq 5$, onda su nam na vidiku dva načina napada tog problema. Prvi bi bio da nađemo $D(\lambda)$ te nađemo rješenja algebarske jednadžbe nekom iteracijskom

metodom. Drugi bi bio da se poslužimo nekom metodom koja iteracijskim postupkom transformira matricu tako da na limesu dobijemo oblik iz kojeg možemo pročitati njene svojstvene vrijednosti.

Samo nalaženje svojstvenih vektora svodi se općenito na nalaženje netrivialnih rješenja homogenog sustava linearnih jednadžbi

$$(\lambda I - A)x = 0.$$

Ilustrirajmo važnost svojstvenih vrijednosti i svojstvenih vektora na primjeru sustava linearnih diferencijalnih jednadžbi.

$$\begin{aligned} \frac{dy_1}{dt} &= a_{11}y_1 + a_{12}y_2 + \dots + a_{1n}y_n \\ \frac{dy_2}{dt} &= a_{21}y_1 + a_{22}y_2 + \dots + a_{2n}y_n \\ &\vdots \\ \frac{dy_n}{dt} &= a_{n1}y_1 + a_{n2}y_2 + \dots + a_{nn}y_n. \end{aligned} \tag{9}$$

Predimo prvo na matrični zapis. Označimo li s

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \dot{y} = \begin{bmatrix} \frac{dy_1}{dt} \\ \frac{dy_2}{dt} \\ \vdots \\ \frac{dy_n}{dt} \end{bmatrix},$$

(9) poprima ovaj matrični zapis

$$\dot{y} = Ay. \tag{10}$$

Ako se pitamo da li rješenje ima oblik

$$y = e^{\lambda t}x, \quad x \neq 0 \tag{11}$$

gdje je x neki ne-nul vektor, onda uvrštavanje u (10) daje

$$\lambda e^{\lambda t}x = Ae^{\lambda t}x = e^{\lambda t}(Ax)$$

pa dijeljenjem s $e^{\lambda t}$ dobivamo da su λ i x podvrgnuti uvjetu

$$\lambda x = Ax. \tag{12}$$

Pročitamo li uvjet (12) vidimo da je λ svojstvena vrijednost matrice A i x svojstveni vektor koji pripada svojstvenoj vrijednosti λ . Ako pročitamo taj izvod natraške, onda vidimo da ako pođemo od svojstvene vrijednosti λ i svojstvenog vektora x koji pripada toj svojstvenoj vrijednosti, da je vektor $y = e^{\lambda t}x$ rješenje sustava (10). Dobili smo dakle ovaj rezultat:

$y = e^{\lambda t}x$ je rješenje sustava $\dot{y} = Ay$ ako i samo ako je λ svojstvena vrijednost od A i x svojstveni vektor koji njoj pripada.

Sljedeći korak je naći opće rješenje sustava (10). Navedimo slučaj da su $\lambda_1, \lambda_2, \dots, \lambda_n$ različite svojstvene vrijednosti od A i da su x_1, x_2, \dots, x_n njima pripadni

svojstveni vektori. Tada je

$$y = \sum_{j=1}^n C_j x_j e^{\lambda_j t} \quad (13)$$

opće rješenje sustava (9), gdje su C_1, C_2, \dots, C_n proizvoljne konstante. Ako je

$$x_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{bmatrix}$$

onda je i -ta komponenta od y dana s

$$y_i(t) = \sum_{j=1}^n C_j x_{ij} e^{\lambda_j t}.$$

Ako svojstvene vrijednosti nisu jednostrukе, onda je opće rješenje složenije građeno.

U idućih pet paragrafa obrađene su opće metode predviđene programom. Na njima se najlakše stiču osnovna znanja iz te problematike. Navedimo nazive nekih metoda, koje ovdje nisu obrađene, od kojih su neke primjenjive na određene klase matrica: Jacobijska metoda za simetrične matrice, Rutishauserova metoda (QR algoritam), metoda tridiagonalizacije, Givenova metoda, Housholderova metoda, Lanczosov algoritam.

4.1. Metoda neodređenih koeficijenata

Pođimo od, moglo bi se reći za učenje najjednostavnije, metode nalaženja karakterističnog polinoma

$$D(\lambda) = \det(\lambda I - A) = \lambda^n + p_1 \lambda^{n-1} + p_2 \lambda^{n-2} + \dots + p_n. \quad (1)$$

Da nađemo svojstvene vrijednosti i vektore treba naći korijene algebarske jednadžbe $D(\lambda) = 0$, što je posebna problematika, a potom naći svojstvene vektore koji pripadaju tim svojstvenim vrijednostima. **Metoda neodređenih koeficijenata** se sastoji u tome da u (1), u lijevu i desnu stranu uvrstimo n posebnih vrijednosti od λ i sastavimo sustav linearnih jednadžbi s nepoznatim koeficijentima p_1, p_2, \dots, p_n . Dakle, uvrstimo u (1) redom $\lambda = 0, 1, 2, \dots, n-1$. Time dobivamo ovaj linearni sustav jednadžbi

$$\begin{aligned} p_n &= D(0) \\ 1^n + p_1 1^{n-1} + p_2 1^{n-2} + \dots + p_n &= D(1) \\ 2^n + p_1 2^{n-1} + p_2 2^{n-2} + \dots + p_n &= D(2) \\ &\vdots \\ (n-1)^n + p_1(n-1)^{n-1} + p_2(n-1)^{n-2} + \dots + p_n &= D(n-1). \end{aligned} \quad (2)$$

Kako je $p_n = D(0) = \det(-A) = (-1)^n \det(A)$ taj koeficijent znamo izračunati, pa prebacivanjem konstantnih članova na desnu stranu sustav (2) prelazi u

$$\begin{aligned} p_1 + p_2 + \dots + p_{n-1} &= D(1) - D(0) - 1 \\ 2^{n-1}p_1 + 2^{n-2}p_2 + \dots + 2p_{n-1} &= D(2) - D(0) - 2^n \\ &\vdots \\ (n-1)^{n-1}p_1 + (n-1)^{n-2}p_2 + \dots + (n-1)p_{n-1} &= D(n-1) - D(0) - (n-1)^n. \end{aligned} \quad (3)$$

U matričnom zapisu sustav (3) izgleda ovako

$$C_n p = d \quad (4)$$

gdje su

$$C_n = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 2^{n-1} & 2^{n-2} & \dots & 2 \\ 3^{n-1} & 3^{n-2} & \dots & 3 \\ \vdots & \vdots & & \vdots \\ (n-1)^{n-1} & (n-1)^{n-2} & \dots & n-1 \end{bmatrix}, \quad p = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_{n-1} \end{bmatrix} \quad \text{i}$$

$$d = \begin{bmatrix} D(1) - D(0) - 1^n \\ D(2) - D(0) - 2^n \\ D(3) - D(0) - 3^n \\ \vdots \\ D(n-1) - D(0) - (n-1)^n \end{bmatrix}.$$

Iz (4) slijedi da je

$$p = C_n^{-1}d. \quad (5)$$

Istaknimo da C_n ovisi samo o n , pa se za fiksni n može C_n^{-1} odrediti. Dakle, za fiksni n po ovoj metodi treba izračunati n n -rednih determinanata

$$D(k) = \det(kI - A), \quad k = 0, 1, 2, \dots, n-1$$

i riješiti linearni sustav jednadžbi (3).

Spomenimo i ovo. Uvrštavanje posebnih vrijednosti u (1) i izračunavanje vrijednosti od D u tim točkama podsjeća na interpolaciju. Kako je $D(\lambda)$ polinom n -tog stupnja, to zadavši $\lambda = 0, 1, \dots, n-1, n$ i izračunavši $D(0), D(1), \dots, D(n-1), D(n)$, formule za interpolacijski polinom daju ispis karakterističnog polinoma $D(\lambda)$, jer je polinom n -tog stupnja određen sa svojih $n+1$ točaka. Takve formule postoje, u što se nećemo upuštati. Ne zaboravimo da one uključuju izračunavanje determinanata $D(0), D(1), \dots, D(n)$, a zatim računanje koeficijenata prema formuli za interpolacijski polinom.

Nalaženje pripadnih svojstvenih vektora svodi se na nalaženje netrivijalnih rješenja homogenog sustava linearnih jednadžbi

$$Ax = \lambda x,$$

za izračunatu svojstvenu vrijednost λ .

4.2. Danilevskijeva metoda

Nalaženje karakterističnog polinoma Danilevskijevom metodom

Bit Danilevskijeve metode je u tome da polaznoj matrici A nađemo njoj sličnu matricu P čiji je karakteristični polinom u tzv. **normalnom Frobeniusovom obliku** koji izgleda ovako

$$D_1(\lambda) = \begin{vmatrix} \sigma_1 - \lambda & \sigma_2 & \sigma_3 & \sigma_4 & \dots & \sigma_{n-1} & \sigma_n \\ 1 & -\lambda & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & -\lambda & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & -\lambda & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -\lambda \end{vmatrix}. \quad (1)$$

Jasno da je $D(\lambda) = (-1)^n D_1(\lambda)$ karakteristični polinom ove matrice

$$P = \begin{bmatrix} \sigma_1 & \sigma_2 & \sigma_3 & \dots & \sigma_{n-1} & \sigma_n \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} \quad (2)$$

samo treba vidjeti da su elementi prvog retka od P sa -1 pomnoženi koeficijenti karakterističnog polinoma matrice P . Da se u to uvjerimo razvijemo determinantu (1) po prvom retku. Taj razvoj izgleda ovako:

$$\begin{aligned} D_1(\lambda) &= (\sigma_1 - \lambda) \begin{vmatrix} -\lambda & 0 & 0 & \dots & 0 & 0 \\ 1 & -\lambda & 0 & \dots & 0 & 0 \\ 0 & 1 & -\lambda & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -\lambda \end{vmatrix} - \sigma_2 \begin{vmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -\lambda & 0 & \dots & 0 & 0 \\ 0 & 1 & -\lambda & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -\lambda \end{vmatrix} \\ &\quad + \sigma_3 \begin{vmatrix} 1 & -\lambda & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & -\lambda & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -\lambda \end{vmatrix} - \sigma_4 \begin{vmatrix} 1 & -\lambda & 0 & \dots & 0 & 0 \\ 0 & 1 & -\lambda & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -\lambda \end{vmatrix} \\ &\quad + \dots + (-1)^{n-1} \sigma_n \begin{vmatrix} 1 & -\lambda & 0 & \dots & 0 & 0 \\ 0 & 1 & -\lambda & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \end{vmatrix} \end{aligned}$$

pri čemu su crtano determinante razbijene u blokove, da lakše uočimo donje ili gornje trokutaste subdeterminante, tako da čitamo

$$\begin{aligned} D_1(\lambda) &= (\sigma_1 - \lambda)(-\lambda)^{n-1} - \sigma_2(-\lambda)^{n-2} + \sigma_3(-\lambda)^{n-3} - \sigma_4(-\lambda)^{n-4} + \dots + (-1)^{n-1}\sigma_n \\ &= (-\lambda)^n + \sigma_1(-\lambda)^{n-1} - \sigma_2(-\lambda)^{n-2} + \sigma_3(-\lambda)^{n-3} - \sigma_4(-\lambda)^{n-4} + \dots + (-1)^{n-1}\sigma_n \\ &= (-1)^n(\lambda^n - \sigma_1\lambda^{n-1} - \sigma_2\lambda^{n-2} - \sigma_3\lambda^{n-3} - \sigma_4\lambda^{n-4} - \dots - \sigma_n). \end{aligned} \quad (3)$$

Kako vrijedi $D(\lambda) = (-1)^n D_1(\lambda)$ (vidi uvodni tekst), to karakteristični polinom $D(\lambda)$ matrice P glasi

$$D(\lambda) = \lambda^n - \sigma_1\lambda^{n-1} - \sigma_2\lambda^{n-2} - \sigma_3\lambda^{n-3} - \sigma_4\lambda^{n-4} - \dots - \sigma_n.$$

Iz zapisa (6) uvoda

$$D(\lambda) = \det(\lambda I - A) = \lambda^n + p_1\lambda^{n-1} + \dots + p_n$$

čitamo da su $p_i = -\sigma_i$, $i = 1, 2, \dots, n$.

Ako dakle pođemo od neke matrice A i uspijemo naći regularnu matricu S takvu da vrijedi

$$P = S^{-1}AS \quad (\text{sličnost}) \quad (4)$$

i da je P oblika (2), onda smo našli karakteristični polinom matrice A jer slične matrice imaju jednake karakteristične polinome. Nađemo li nadalje korijene karakterističnog polinoma našli smo svojstvene vrijednosti matrice A . Nalaženje korijena polinoma je problem za sebe i spada u poglavlje o nelinearnim jednadžbama.

Opišimo sada Danilevskijevu metodu, tj. postupak kako se iz A izgradi P . To postižemo pomoću $(n-1)$ -dne transformacije sličnosti i to postepenim transformiranjem redaka matrice A . Počinjemo sa zadnjim redkom matrice A

$$a_{n1} \ a_{n2} \ \dots \ a_{nn-1} \ a_{nn}$$

koji prevodimo u oblik

$$0 \ 0 \ \dots \ 1 \ 0.$$

Prepostavimo da je $a_{nn-1} \neq 0$. Tada treba elemente $(n-1)$ -og stupca podijeliti s a_{nn-1} nakon čega n -ti redak poprima oblik

$$a_{n1} \ a_{n2} \ \dots \ 1 \ a_{nn}.$$

Da dobijemo nule na ostalim mjestima n -og retka treba transformirani $(n-1)$ -vi stupac redom množiti s a_{n1} , a_{n2} , ..., a_{nn-2} , a_{nn} i redom oduzeti od 1-og, 2-og, ..., $(n-2)$ -og, n -og stupca. Time dobivamo matricu čiji je n -ti redak jednak n -tom retku matrice P . Spomenute transformacije na stupcima su elementarne transformacije provedene na stupcima od A . Ako te transformacije provedemo na jediničnoj matrici, onda dobivamo matricu

$$M_{n-1} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ m_{n-1,1} & m_{n-1,2} & \dots & m_{n-1,n-1} & m_{n-1,n} \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

u kojoj su

$$\begin{aligned} m_{n-1j} &= -\frac{a_{nj}}{a_{nn-1}} \quad \text{za } j \neq n-1, \\ m_{n-1n-1} &= \frac{1}{a_{nn-1}}. \end{aligned} \quad (5)$$

Opisane elementarne transformacije realiziramo tako da A pomnožimo zdesna s M_{n-1} , što daje

$$AM_{n-1} = B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n-1} & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n-1} & b_{2n} \\ \vdots & \vdots & & \vdots & \vdots \\ b_{n-11} & b_{n-12} & \dots & b_{n-1n-1} & b_{n-1n} \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

Iz opisa postupka i množenja matrica čitamo da su

$$\begin{aligned} b_{ij} &= a_{ij} + a_{in-1}m_{n-1j} \quad \text{za } 1 \leq i \leq n; \quad j \neq n-1 \\ b_{in-1} &= a_{in-1}m_{n-1n-1} \quad \text{za } 1 \leq i \leq n. \end{aligned} \quad (6)$$

Međutim, matrica $B = AM_{n-1}$ nije nužno slična matrici A . Da bi osigurali sličnost moramo B pomnožiti slijeva s M_{n-1}^{-1} što daje

$$C = M_{n-1}^{-1}B = M_{n-1}^{-1}AM_{n-1}. \quad (7)$$

Prema rečenom u uvodu pripadna inverzna matrica glasi

$$M_{n-1}^{-1} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn-1} & a_{nn} \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}.$$

Uočimo da se M_{n-1}^{-1} razlikuje od jedinične samo u $(n-1)$ -om retku, zbog čega su elementi matrice C jednaki elementima matrice B u svim recima osim u $(n-1)$ -om, tj. C ima oblik

$$C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n-1} & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n-1} & c_{2n} \\ \vdots & \vdots & & \vdots & \vdots \\ c_{n-11} & c_{n-12} & \dots & c_{n-1n-1} & c_{n-1n} \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

gdje je prema rečenom

$$\begin{aligned} c_{ij} &= b_{ij} \quad \text{za } 1 \leq i \leq n-2, \quad i = n \\ c_{n-1j} &= \sum_{k=1}^n a_{nk}b_{kj} \quad \text{za } 1 \leq j \leq n. \end{aligned} \quad (8)$$

Na taj način, množenje matrice B slijeva s M_{n-1}^{-1} mijenja samo $(n-1)$ -vi redak od B , čiji se elementi računaju po formuli (8). To je prvi korak **Danilevskijeve metode**.

Da nastavimo pretpostavimo da je

$$c_{n-1, n-2} \neq 0$$

tako da na C ponovimo analogne transformacije nad $(n-1)$ -im retkom pomoću $(n-2)$ -og stupca. Uočimo da pri tome ne narušavamo već postignuto u n -tom retku. Time dobivamo matricu

$$D = M_{n-2}^{-1} C M_{n-2} \quad (9)$$

koja je slična matrici A i ima zadnja dva retka u željenom obliku. Nastavivši postupak dalje dobivamo Frobeniusovu matricu

$$P = M_1^{-1} M_2^{-1} \dots M_{n-2}^{-1} M_{n-1}^{-1} A M_{n-1} M_{n-2} \dots M_2 M_1 \quad (10)$$

iz koje prema (3) dobivamo

$$D_1(\lambda) = (-1)^n (\lambda^n - \sigma_1 \lambda^{n-1} - \sigma_2 \lambda^{n-2} - \dots - \sigma_n).$$

Ovakav postupak je provodiv uz pretpostavku da je u k -tom koraku

$$d_{n-k+1, n-k} \neq 0,$$

što ne mora biti ispunjeno u svakom koraku. Pretpostavimo da smo došli do matrice ovog oblika

$$D = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1k-1} & d_{1k} & \dots & d_{1n-1} & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2k-1} & d_{2k} & \dots & d_{2n-1} & d_{2n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ d_{k1} & d_{k2} & \dots & 0 & d_{kk} & \dots & d_{kn-1} & d_{kn} \\ 0 & 0 & \dots & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

u kojoj je $d_{k,k-1} = 0$. Tada mogu nastupiti 2 slučaja.

1. Na lijevo od $d_{k,k-1}$ imamo element $d_{kl} \neq 0$, $l < k-1$. Tada zamijenimo $(k-1)$ -vi i l -ti stupac i istovremeno zamijenimo $(k-1)$ -vi i l -ti redak da bi zadržali sličnost sa A i potom nastavimo postupak.

2. Na lijevo od $d_{k,k-1}$ su same nule, tj. D ima oblik

$$D = \left[\begin{array}{cc|cc|cc} d_{11} & \dots & d_{1k-1} & | & d_{1k} & \dots & d_{1n-1} & d_{1n} \\ \vdots & & \vdots & | & \vdots & & \vdots & \vdots \\ d_{k-1,1} & \dots & d_{k-1, k-1} & | & d_{k-1, k} & \dots & d_{k-1, n-1} & d_{k-1, n} \\ 0 & \dots & 0 & | & d_{kk} & \dots & d_{kn-1} & d_{kn} \\ 0 & \dots & 0 & | & 1 & \dots & 0 & 0 \\ \vdots & & \vdots & | & \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & | & 0 & \dots & 1 & 0 \end{array} \right] = \begin{bmatrix} D_1 & L \\ 0 & D_2 \end{bmatrix}.$$

Tada je

$$\det(D - \lambda I) = \det(D_1 - \lambda I) \cdot \det(D_2 - \lambda I)$$

pri čemu je D_2 već u normalnom Frobeniusovom obliku pa se postupak nastavi na D_1 .

Nalaženje svojstvenih vektora Danilevskijevom metodom

Danilevskijeva metoda se može iskoristiti za nalaženje svojstvenih vektora matrice ako su nam poznate njene svojstvene vrijednosti. Tako, neka je λ svojstvena vrijednost matrice A . Time je λ svojstvena vrijednost i njene Frobeniusove matrice P (sličnost). Imamo dakle

$$(P - \lambda I)y = 0$$

ili

$$\begin{bmatrix} -p_1 - \lambda & -p_2 & -p_3 & \dots & -p_{n-1} & -p_n \\ 1 & -\lambda & 0 & \dots & 0 & 0 \\ 0 & 1 & -\lambda & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -\lambda \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \vec{0}. \quad (11)$$

(11) predstavlja matrični zapis ovog homogenog sustava linearnih jednadžbi s nepozanicama y_1, y_2, \dots, y_n ,

$$\begin{aligned} (-p_1 - \lambda)y_1 - p_2y_2 - p_3y_3 - \dots - p_{n-1}y_{n-1} - p_ny_n &= 0 \\ y_1 - \lambda y_2 &= 0 \\ y_2 - \lambda y_3 &= 0 \\ &\vdots \\ y_{n-1} - \lambda y_n &= 0. \end{aligned} \quad (12)$$

Zbog $\det(P - \lambda I) = 0$ sustav (12) ima netrivialnih rješenja. Sustav možemo do na koeficijent proporcionalnosti ovako riješiti. Izaberimo $y_n = 1$. Tada korak po korak imamo

$$\begin{aligned} y_n &= 1 \\ y_{n-1} &= \lambda \\ y_{n-2} &= \lambda^2 \\ &\vdots \\ y_1 &= \lambda^{n-1}, \end{aligned}$$

pa je na taj način vektor

$$y = \begin{bmatrix} \lambda^{n-1} \\ \lambda^{n-2} \\ \vdots \\ \lambda \\ 1 \end{bmatrix} \quad (13)$$

svojstveni vektor matrice P koji pripada svojstvenoj vrijednosti λ . Da nađemo svojstveni vektor matrice A koji pripada svojstvenoj vrijednosti λ , prisjetimo se kako smo dobili matricu P iz matrice A . Imali smo formulu (10) u prethodnom paragrafu koja je glasila

$$P = M_1^{-1} \dots M_{n-1}^{-1} A M_{n-1} \dots M_1. \quad (14)$$

Kako je $Py = \lambda y$ to uvrštavanjem dobivamo

$$M_1^{-1} \dots M_{n-1}^{-1} A M_{n-1} \dots M_1 y = \lambda y$$

što množenjem slijeva redom s M_1, M_2, \dots, M_{n-1} daje

$$A(M_{n-1} \dots M_1 y) = \lambda(M_{n-1} \dots M_1 y). \quad (15)$$

Iz (15) čitamo da je

$$x = M_{n-1} \dots M_1 y \quad (16)$$

svojstveni vektor matrice A koji pripada svojstvenoj vrijednosti λ . Dakle, da nađemo x moramo zapamtiti sve matrice M_i , $i = 1, 2, \dots, n-1$.

Primijetimo i sljedeće

$$M_1 y = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1n} \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^n m_{1k} y_k \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^n m_{1k} \lambda^{n-k} \\ \lambda^{n-2} \\ \lambda^{n-3} \\ \vdots \\ \lambda \\ 1 \end{bmatrix}$$

iz čega vidimo da M_1 mijenja samo prvu komponentu vektora y . Analogno M_2 mijenja samo drugu komponentu vektora $M_1 y$, itd. Time korak po korak nalazimo svojstveni vektor x matrice A , pri čemu je čitav račun, osim rješavanja algebarske jednadžbe $D_1(\lambda) = 0$, vrlo jednostavan.

Primjer 4.1. Neka je

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 2 & 3 \\ 3 & 2 & 1 & 2 \\ 4 & 3 & 2 & 1 \end{bmatrix}.$$

Vidimo da je $a_{43} = 2$. Da u tom položaju dobijemo jedinicu treba treći stupac podijeliti s 2, što povlači da je element matrice M_3 u položaju (3, 3) jednak $\frac{1}{2}$; tj. $m_{33} = \frac{1}{2}$. Da u A poništimo ostale elemente četvrtog retka treba s 2 podijeljeni treći stupac redom pomnožiti s 4, 3 i 1 te redom oduzeti od 1-og, 2-og i 4-og stupca. To sve treba provesti na jediničnoj matrici tako da dobivamo

$$M_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -2 & -\frac{3}{2} & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

pa matrica $B = AM_3$ glasi

$$B = \begin{bmatrix} -5 & -\frac{5}{2} & \frac{3}{2} & \frac{5}{2} \\ -2 & -\frac{3}{2} & \frac{1}{2} & \frac{3}{2} \\ 1 & \frac{1}{2} & \frac{1}{2} & \frac{3}{2} \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Pripadna inverzna matrica M_3^{-1} glasi

$$M_3^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 4 & 3 & 2 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

tako da je $C = M_3^{-1}B$ jednaka

$$C = M_3^{-1}B = M_3^{-1}AM_3 = \begin{bmatrix} -5 & -\frac{5}{2} & \frac{3}{2} & \frac{5}{2} \\ -2 & -2 & 1 & 2 \\ -24 & -15 & 11 & 19 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Time je proveden prvi korak Danilevskijeve metode. Za drugi korak trebamo $c_{32} \neq 0$. Kako je $c_{32} = -15$, to M_2 glasi

$$M_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\frac{24}{15} & -\frac{1}{15} & \frac{11}{15} & \frac{19}{15} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

i pripadna inverzna

$$M_2^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -24 & -15 & 11 & 19 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Matrica

$$\begin{aligned} D = M_2^{-1}CM_2 &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ -24 & -15 & 11 & 19 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -5 & -\frac{5}{2} & \frac{3}{2} & \frac{5}{2} \\ -2 & -2 & 1 & 2 \\ -24 & -15 & 11 & 19 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\frac{24}{15} & -\frac{1}{15} & \frac{11}{15} & \frac{19}{15} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ -24 & -15 & 11 & 19 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & \frac{1}{6} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{18}{15} & \frac{2}{15} & -\frac{7}{15} & -\frac{8}{15} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & \frac{1}{6} & -\frac{1}{3} & -\frac{2}{3} \\ ⑥ & 5 & 34 & 24 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{aligned}$$

je slična matrici A . Time je obavljen drugi korak Danilevskijeve metode. Kako je A četvrtog reda, to teorijski preostaje još jedan korak. Kako je $d_{21} = 6 \neq 0$ to nalazimo

$$M_1 = \begin{bmatrix} \frac{1}{6} & -\frac{5}{6} & -\frac{34}{6} & -4 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad M_1^{-1} = \begin{bmatrix} 6 & 5 & 34 & 24 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

odnosno

$$\begin{aligned}
 P = M_1^{-1} D M_1 &= \begin{bmatrix} 6 & 5 & 34 & 24 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & \frac{1}{6} & -\frac{1}{3} & -\frac{2}{3} \\ 6 & 5 & 34 & 24 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{6} & -\frac{5}{6} & -\frac{34}{6} & -4 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
 &= \begin{bmatrix} 6 & 5 & 34 & 24 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -\frac{1}{6} & 1 & \frac{32}{6} & \frac{10}{3} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 40 & 56 & 20 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.
 \end{aligned}$$

Otkuda čitamo

$$D(\lambda) = (-1)^4 D_1(\lambda) = \lambda^4 - 4\lambda^3 - 40\lambda^2 - 56\lambda - 20.$$

4.3. Krilovljeva metoda

Nalaženje karakterističnog polinoma Krilovljevom metodom

U ovoj metodi ćemo problem nalaženja koeficijenata karakterističnog polinoma svesti na rješavanje sustava linearnih jednadžbi. Neka je

$$D(\lambda) = \det(\lambda I - A) = \lambda^n + p_1\lambda^{n-1} + \dots + p_n \quad (1)$$

karakteristični polinom. Prema **Hamilton-Cayleyevom teoremu** matrica A poništava (zadovoljava) svoj karakteristični polinom, tj. vrijedi

$$A^n + p_1 A^{n-1} + \dots + p_n I = 0. \quad (2)$$

Neka je $0 \neq y^{(0)} = \begin{bmatrix} y_1^{(0)} \\ \vdots \\ y_n^{(0)} \end{bmatrix}$ vektor različit od nulvektora. Pomnožimo li (2) zdesna s $y^{(0)}$ dobivamo

$$A^n y^{(0)} + p_1 A^{n-1} y^{(0)} + \dots + p_n y^{(0)} = 0. \quad (3)$$

Označimo li s

$$y^{(k)} = A^k y^{(0)}, \quad k = 1, 2, \dots, n \quad (4)$$

jednakost (3) će poprimiti oblik

$$y^{(n)} + p_1 y^{(n-1)} + \dots + p_n y^{(0)} = 0 \quad (5)$$

što je linearan sustav jednadžbi s nepoznanicama p_1, p_2, \dots, p_n . Jednakost (5) možemo i ovako zapisati

$$\begin{bmatrix} y_1^{(n-1)} & y_1^{(n-2)} & \dots & y_1^{(0)} \\ y_2^{(n-1)} & y_2^{(n-2)} & \dots & y_2^{(0)} \\ \vdots & \vdots & \ddots & \vdots \\ y_n^{(n-1)} & y_n^{(n-2)} & \dots & y_n^{(0)} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} = - \begin{bmatrix} y_1^{(n)} \\ y_2^{(n)} \\ \vdots \\ y_n^{(n)} \end{bmatrix}, \quad (6)$$

pri čemu su komponente od $y^{(k)}$ označene ovako

$$y^{(k)} = \begin{bmatrix} y_1^{(k)} \\ y_2^{(k)} \\ \vdots \\ y_n^{(k)} \end{bmatrix}, \quad k = 0, 1, 2, \dots, n.$$

Pojedina jednadžba sustava (5) glasi

$$y_j^{(n-1)} p_1 + y_j^{(n-2)} p_2 + \dots + y_j^{(0)} p_n = -y_j^{(n)}, \quad j = 1, 2, \dots, n. \quad (7)$$

Rješavanjem tog sustava nalazimo nepoznanice p_1, p_2, \dots, p_n , tj. koeficijente karakterističnog polinoma (1). Time je problem nalaženja karakterističnog polinoma sведен na rješavanje sustava linearnih jednadžbi. Vektore $y^{(k)}$ nalazimo postupno (iterativno) prema (4), jer je

$$y^{(k)} = A^k y^{(0)} = A(A^{k-1} y^{(0)}) = A y^{(k-1)},$$

tako da za komponente vektora $y^{(k)}$, $k = 1, 2, \dots, n$ nalazimo ove formule

$$\begin{aligned} y_i^{(1)} &= \sum_{j=1}^n a_{ij} y_j^{(0)} \\ y_i^{(2)} &= \sum_{j=1}^n a_{ij} y_j^{(1)} \\ &\vdots \\ y_i^{(n)} &= \sum_{j=1}^n a_{ij} y_j^{(n-1)}. \end{aligned} \quad (8)$$

Uočimo da su komponente početnog vektora $y^{(0)}$ bile proizvoljne. Spomenimo da možemo naići na sustav (5) koji nema jedinstveno rješenje. Tada se preporuča promijeniti početni vektor $y^{(0)}$.

Nalaženje svojstvenih vektora Krilovljevom metodom

Krilovljeva se metoda može iskoristiti i za nalaženje svojstvenih vektora matrice. Opisimo taj postupak uz pretpostavku da znamo $D(\lambda)$ i njegove korijene

$$D(\lambda) = \det(\lambda I - A) = \lambda^n + p_1 \lambda^{n-1} + \dots + p_n. \quad (9)$$

Neka su ti korijeni $\lambda_1, \lambda_2, \dots, \lambda_n$ različiti.

Označimo s $x^{(1)}, x^{(2)}, \dots, x^{(n)}$ pripadne svojstvene vektore. Poslužimo se s prethodno korištenim vektorima iz 2.1

$$y^{(0)}, y^{(1)} = Ay^{(0)}, \dots, y^{(n-1)} = A^{n-1} y^{(0)} \quad (10)$$

na sljedeći način. Rastavimo $y^{(0)}$ po svojstvenim vektorima $x^{(i)}$, za koje iz opće teorije znamo da su linearno nezavisni. Imamo dakle

$$y^{(0)} = c_1 x^{(1)} + c_2 x^{(2)} + \dots + c_n x^{(n)} \quad (11)$$

gdje su c_i koeficijenti tog rastava. Kako je

$$Ax^{(i)} = \lambda_i x^{(i)}$$

to množenjem slijeva s A dobivamo redom

$$\begin{aligned} A^2 x^{(i)} &= A \lambda_i x^{(i)} = \lambda_i A x^{(i)} = \lambda_i \lambda_i x^{(i)} = \lambda_i^2 x^{(i)} \\ A^3 x^{(i)} &= \lambda_i^3 x^{(i)} \\ &\vdots \\ A^{n-1} x^{(i)} &= \lambda_i^{n-1} x^{(i)}. \end{aligned} \tag{12}$$

Uvrštavanjem (12) u $y^{(1)}, y^{(2)}, \dots, y^{(n-1)}$ dobivamo

$$\begin{aligned} y^{(0)} &= c_1 x^{(1)} + c_2 x^{(2)} + \dots + c_n x^{(n)} \\ y^{(1)} &= c_1 \lambda_1 x^{(1)} + c_2 \lambda_2 x^{(2)} + \dots + c_n \lambda_n x^{(n)} \\ y^{(2)} &= c_1 \lambda_1^2 x^{(1)} + c_2 \lambda_2^2 x^{(2)} + \dots + c_n \lambda_n^2 x^{(n)} \\ &\vdots \\ y^{(n-1)} &= c_1 \lambda_1^{n-1} x^{(1)} + c_2 \lambda_2^{n-1} x^{(2)} + \dots + c_n \lambda_n^{n-1} x^{(n)}. \end{aligned} \tag{13}$$

U dobivenom sustavu jednadžbi (13) su nam nepoznati svojstveni vektori $x^{(i)}$ i koeficijenti c_i , tako da isti nije linearan. Usredotočimo li se na svojstvene vrijednosti λ_i vidimo da je svaka jednakost, formalno gledano, "linearna" kombinacija istih potencija svojstvenih vrijednosti λ_i , pa možemo provesti ovakvo razmatranje (dosjetka).

Neka je

$$\varphi_i(\lambda) = \lambda^{n-1} + q_{1i}\lambda^{n-2} + \dots + q_{n-1i}, \quad i = 1, 2, \dots, n \tag{14}$$

proizvoljnih n polinoma $(n-1)$ -og stupnja. Iskoristimo koeficijente polinoma

$$1, q_{1,i}, q_{2,i}, \dots, q_{n-1,i}, \quad i = 1, 2, \dots, n$$

i pomoću njih načinimo linearnu kombinaciju vektora $y^{(n-1)}, y^{(n-2)}, \dots, y^{(1)}, y^{(0)}$. Dobivamo

$$\begin{aligned} y^{(n-1)} + q_{1i}y^{(n-2)} + \dots + q_{n-1i}y^{(0)} \\ &= 1(c_1 \lambda_1^{n-1} x^{(1)} + \dots + c_n \lambda_n^{n-1} x^{(n)}) \\ &\quad + q_{1i}(c_1 \lambda_1^{n-2} x^{(1)} + \dots + c_n \lambda_n^{n-2} x^{(n)}) \\ &\quad \vdots \\ &\quad + q_{n-1,i}(c_1 x^{(1)} + \dots + c_n x^{(n)}) \\ &= c_1 \varphi_i(\lambda_1) x^{(1)} + c_2 \varphi_i(\lambda_2) x^{(2)} + \dots + c_n \varphi_i(\lambda_n) x^{(n)}, \quad i = 1, 2, \dots, n, \end{aligned} \tag{15}$$

pri čemu smo u zadnjem koraku sumu sredili po vektorima $x^{(i)}$. U ovom razmatranju su polinomi φ_i bili proizvoljni, pa sada možemo izvršiti njihov izbor. Uzmimo da je

$$\varphi_i = \frac{D(\lambda)}{\lambda - \lambda_i} = \frac{(\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n)}{\lambda - \lambda_i}, \quad i = 1, 2, \dots, n. \tag{16}$$

Uz taj izbor vidimo da vrijedi

$$\begin{aligned}\varphi_i(\lambda_j) &= 0 \quad \text{ako je } i \neq j \\ \varphi_i(\lambda_i) &= D'(\lambda_i),\end{aligned}$$

derivacija od $D(\lambda)$ u točki $\lambda = \lambda_i$. Da je to zaista $D'(\lambda_i)$ vidimo iz oblika $D(\lambda) = \prod_{j=1}^n (\lambda - \lambda_j)$ čija je derivacija $D'(\lambda) = \sum_{i=1}^n \prod_{\substack{j=1 \\ j \neq i}}^n (\lambda - \lambda_j)$, pa kada uvrstimo

$\lambda = \lambda_i$ ostaje samo onaj pribrojnik u kojemu je izostavljen faktor $\lambda - \lambda_i$. Za izbor prema (16) linearne kombinacije (15) poprimaju oblik

$$c_i \varphi_i(\lambda_i) x^{(i)} = y^{(n-1)} + q_{1i} y^{(n-2)} + \dots + q_{n-1i} y^{(0)}, \quad i = 1, 2, \dots, n. \quad (17)$$

Kako je $\varphi_i(\lambda_i) \neq 0$, pa ako je i $c_i \neq 0$, ondā možemo izračunati $x^{(i)}$, tj. svojstveni vektor matrice A .

Preostalo je još da odredimo koeficijente q_{ji} . Kako je

$$D(\lambda) = \varphi_i(\lambda)(\lambda - \lambda_i)$$

to imamo

$$\begin{aligned}\lambda^n + p_1 \lambda^{n-1} + p_2 \lambda^{n-2} + \dots + p_n \\ &= (\lambda^{n-1} + q_{1i} \lambda^{n-2} + q_{2i} \lambda^{n-3} + \dots + q_{n-1i})(\lambda - \lambda_i) \\ &= \lambda^n + (q_{1i} - \lambda_i) \lambda^{n-1} + (q_{2i} - \lambda_i q_{1i}) \lambda^{n-2} + (q_{3i} - \lambda_i q_{2i}) \lambda^{n-3} + \dots \\ &\quad + (q_{n-1i} - \lambda_i q_{n-2i}) \lambda - \lambda_i q_{n-1i}.\end{aligned}$$

Koeficijenti $q_{0i} = 1$ su zadani, pa izjednačavanjem koeficijenata uz iste potencije dobivamo formule za postupno izračunavanje koeficijenata q_{ji} :

$$\begin{aligned}q_{0i} &= 1 \quad i = 1, 2, \dots, n \\ q_{ji} &= \lambda_i q_{j-1i} + p_j \quad j = 1, 2, \dots, n-1.\end{aligned} \quad (18)$$

4.4. Nalaženje karakterističnog polinoma Leverrierovom metodom

Ovdje prikazujemo još jedan postupak nalaženja karakterističnog polinoma pri čemu ćemo se poslužiti sumom potencija korijena karakteristične jednadžbe. Pođimo od

$$D(\lambda) = \det(\lambda I - A) = \lambda^n + p_1 \lambda^{n-1} + \dots + p_n \quad (1)$$

i neka su $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$ svи korijeni karakteristične jednadžbe, tj. u tom popisu se svaki korijen pojavljuje onoliko puta kolika mu je kratnost. Označimo s

$$s_k = \lambda_1^k + \lambda_2^k + \dots + \lambda_n^k, \quad k = 0, 1, 2, \dots, n$$

sumu k -tih potencija korijena. Za takve sume i koeficijente polinoma vrijede tzv. **Newtonove formule** koje glase

$$s_k + p_1 s_{k-1} + p_2 s_{k-2} + \dots + p_{k-1} s_1 = -kp_k, \quad k = 1, 2, 3, \dots, n. \quad (2)$$

Iz (2) lako redom izračunamo

$$\begin{aligned} p_1 &= -s_1 \\ p_2 &= -\frac{1}{2}(s_2 + p_1 s_1) \\ p_3 &= -\frac{1}{3}(s_3 + p_1 s_2 + p_2 s_1) \\ &\vdots \\ p_n &= -\frac{1}{n}(s_n + p_1 s_{n-1} + \dots + p_{n-1} s_1) \end{aligned} \tag{3}$$

samo nam trebaju s_1, s_2, \dots, s_n . Međutim, te se veličine mogu izračunati služeći se činjenicom da su $\lambda_1^k, \lambda_2^k, \dots, \lambda_n^k$ svojstvene vrijednosti matrice A^k i što znamo da je

$$s_1 = \lambda_1 + \lambda_2 + \dots + \lambda_n = \sum_{i=1}^n a_{ii} = S_p(A)$$

($S_p(A) = \text{trag matrice } A$). Kako su λ_i^k svojstvene vrijednosti od A^k , to onda vrijedi

$$s_k = \lambda_1^k + \lambda_2^k + \dots + \lambda_n^k = S_p(A^k), \tag{4}$$

pa smo problem reducirali na računanje traga od A^k , $k = 2, \dots, n$. Za $k = 2$ dijagonalni elementi od A^2 su dani s

$$a_{ii}^{(2)} = \sum_{k=1}^n a_{ik} a_{ki}, \quad i = 1, 2, \dots, n$$

tako da je

$$s_2 = \lambda_1^2 + \lambda_2^2 + \dots + \lambda_n^2 = \sum_{i=1}^n a_{ii}^{(2)} = \sum_{i=1}^n \sum_{k=1}^n a_{ik} a_{ki}.$$

Prema tome, označimo li s $a_{ij}^{(k)}$ elemente od A^k imamo da je

$$s_k = \lambda_1^k + \lambda_2^k + \dots + \lambda_n^k = \sum_{i=1}^n a_{ii}^{(k)} = \sum_{i=1}^n \sum_{k=1}^n a_{ik} a_{ki}^{(n-1)}.$$

Iz izведенog vidimo da Leverrierova metoda zahtjeva računanje potencija matrice A što može biti mukotrplno ako je red matrice A visok.

Nalaženje svojstvenih vektora svodi se i sada na nalaženje netrivijalnih rješenja homogenog sustava linearnih jednadžbi

$$(\lambda I - A)x = 0$$

za danu svojstvenu vrijednost λ .

4.5. Misesova metoda potencija — spektralni radijus matrice

Neka su $\lambda_1, \lambda_2, \dots, \lambda_n$ svojstvene vrijednosti matrice A , tj. korijeni algebarske jednadžbe

$$D(\lambda) = \det(\lambda I - A) = 0. \quad (1)$$

Svojstvene vrijednosti su općenito kompleksni brojevi i skup $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ se zove **spekter matrice A** dok se

$$\rho = \max_i |\lambda_i| \quad (2)$$

zove **spektralni radijus** matrice A . Indeksirajmo svojstvene vrijednosti tako da vrijedi

$$|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|,$$

pa kada je $|\lambda_1| > |\lambda_2|$ ili $|\lambda_1| = |\lambda_2| = \dots = |\lambda_i|$ i $|\lambda_1| > |\lambda_{i+1}|$, uobičajeno je λ_1 zvati **dominantna svojstvena vrijednost**. Tada je $\rho = |\lambda_1|$. U različitim problemima je potrebno ocijeniti vrijednost spektralnog radijusa ρ . **Misesova metoda potencija** je jedna iteracijska metoda izračunavanja dominantne svojstvene vrijednosti, ili po apsolutnoj vrijednosti najveće svojstvene vrijednosti, bez nalaženja karakterističnog polinoma.

Neka svojstvenim vrijednostima $\lambda_1, \lambda_2, \dots, \lambda_n$ pripadaju linearne nezavisne svojstveni vektori $x^{(1)}, x^{(2)}, \dots, x^{(n)}$.

1. Slučaj. Neka je

$$|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_n|. \quad (3)$$

Ako su elementi od A realni, onda je λ_1 realan. Naime, tada je $D(\lambda)$ polinom s realnim koeficijentima, pa se kompleksni korijeni javljaju u parovima konjugirano kompleksnih vrijednosti. Kada bi λ_1 bio kompleksan, onda bi $\lambda_2 = \bar{\lambda}_1$ i $|\lambda_1| = |\lambda_2|$, što se protivi pretpostavci 1. slučaja.

Podimo od proizvoljnog vektora y i rastavimo ga po svojstvenim vektorima matrice A , tj. napišimo

$$y = c_1 x^{(1)} + c_2 x^{(2)} + \dots + c_n x^{(n)} \quad (4)$$

gdje su c_j koeficijenti tog rastava. Pomnožimo (4) s A . Time dobivamo

$$Ay = \sum_{j=1}^n c_j Ax^{(j)} = \sum_{j=1}^n c_j \lambda_j x^{(j)} \quad (5)$$

jer je $x^{(j)}$ svojstveni vektor koji pripada svojstvenoj vrijednosti λ_j . Dobiveni vektor Ay zovemo **iteracija** vektora y . No to možemo nastaviti tako da imamo drugu iteraciju vektora y jednaku

$$A(Ay) = A^2y = \sum_{j=1}^n c_j \lambda_j Ax^{(j)} = \sum_{j=1}^n c_j \lambda_j^2 x^{(j)}$$

i tako redom, m -ta iteracija glasi

$$y^{(m)} = A^m y = \sum_{j=1}^n c_j \lambda_j^m x^{(j)}. \quad (6)$$

Neka je e_1, e_2, \dots, e_n baza vektorskog prostora. Neka je

$$y^{(m)} = \begin{bmatrix} y_1^{(m)} \\ y_2^{(m)} \\ \vdots \\ y_n^{(m)} \end{bmatrix}$$

gdje su $y_i^{(m)}$ ($i = 1, 2, \dots, n$) komponente od $y^{(m)}$ u bazi e_1, e_2, \dots, e_n . S druge strane neka su komponente svojstvenih vektora u toj bazi x_{ij} , tj. neka je

$$x^{(j)} = \sum_{i=1}^n x_{ij} e_i. \quad (7)$$

Uvrstimo li (7) u (6) dobivamo

$$y^{(m)} = \sum_{j=1}^n c_j \lambda_j^m \sum_{i=1}^n x_{ij} e_i = \sum_{i=1}^n e_i \left(\sum_{j=1}^n c_j x_{ij} \lambda_j^m \right). \quad (8)$$

Dakle, i -ta komponenta od $y^{(m)}$ je faktor od e_i , tj.

$$y_i^{(m)} = \sum_{j=1}^n c_j x_{ij} \lambda_j^m. \quad (9)$$

Analogno je i -ta komponenta od $y^{(m+1)}$ jednaka

$$y_i^{(m+1)} = \sum_{j=1}^n c_j x_{ij} \lambda_j^{m+1}. \quad (10)$$

Podijelimo li (10) s (9) dobivamo

$$\frac{y_i^{(m+1)}}{y_i^{(m)}} = \frac{c_1 x_{i1} \lambda_1^{m+1} + c_2 x_{i2} \lambda_2^{m+1} + \dots + c_n x_{in} \lambda_n^{m+1}}{c_1 x_{i1} \lambda_1^m + c_2 x_{i2} \lambda_2^m + \dots + c_n x_{in} \lambda_n^m}. \quad (11)$$

Prepostavimo da je $c_1 \neq 0$ i $x_{i1} \neq 0$. To nije ograničenje jer to možemo postići promjenom početnog vektora y i baze e_1, e_2, \dots, e_n . Transformirajmo (11) u ovaj oblik

$$\frac{y_i^{(m+1)}}{y_i^{(m)}} = \lambda_1 \frac{1 + \frac{c_2 x_{i2}}{c_1 x_{i1}} \left(\frac{\lambda_2}{\lambda_1} \right)^{m+1} + \dots + \frac{c_n x_{in}}{c_1 x_{i1}} \left(\frac{\lambda_n}{\lambda_1} \right)^{m+1}}{1 + \frac{c_2 x_{i2}}{c_1 x_{i1}} \left(\frac{\lambda_2}{\lambda_1} \right)^m + \dots + \frac{c_n x_{in}}{c_1 x_{i1}} \left(\frac{\lambda_n}{\lambda_1} \right)^m}. \quad (12)$$

Iz (12) zaključujemo da prelaskom na limes za $m \rightarrow \infty$ dobivamo (zbog $\frac{\lambda_i}{\lambda_1} < 1$ za $i \geq 2$) sljedeći rezultat

$$\lambda_1 = \lim_{m \rightarrow \infty} \frac{y_i^{(m+1)}}{y_i^{(m)}}. \quad (13)$$

Iz (13) dobivamo u 1. slučaju sljedeću aproksimaciju dominantne svojstvene vrijednosti

$$\lambda_1 \approx \frac{y_i^{(m+1)}}{y_i^{(m)}}, \quad i = 1, 2, \dots, n. \quad (14)$$

Ako smo uzeli dovoljno veliki korak iteracije, tj. dovoljno visoku potenciju m možemo prema (14) postići traženu točnost. Da ubrzamo konvergenciju iteracije spretno je promatrati iteracije koje su potencije broja 2, tj. $m = 2^k$. U tom slučaju imamo kvadriranje matrica, tj. gledamo potencije

$$A, A^2 = A \cdot A, A^4 = A^2 \cdot A^2, A^8 = A^4 \cdot A^4, \dots, A^{2^k} = A^{2^{k-1}} \cdot A^{2^{k-1}} = A^m.$$

Izračunamo li još $y^{(m)} = A^m y$ i $y^{(m+1)} = Ay^{(m)}$ dobivamo

$$\lambda_1 \approx \frac{y_i^{(m+1)}}{y_i^{(m)}}, \quad i = 1, 2, \dots, n.$$

Dobivši aproksimaciju od λ_1 možemo naći i aproksimaciju svojstvenog vektora koji pripada svojstvenoj vrijednosti λ_1 . Naime, *vektor*

$$y^{(m)} = A^m y \quad (15)$$

je aproksimacija svojstvenog vektora od A koji pripada svojstvenoj vrijednosti λ_1 .

Evo kako to vidimo. Iz formule (6) imamo

$$A^m y = c_1 \lambda_1^m x^{(1)} + \sum_{j=2}^n c_j \lambda_j^m x^{(j)}$$

gdje je $x^{(j)}$ svojstveni vektor koji pripada λ_j . Izlučimo li $c_1 \lambda_1^m$ dobivamo

$$A^m y = c_1 \lambda_1^m \left\{ x^{(1)} + \sum_{j=2}^n \frac{c_j}{c_1} \left(\frac{\lambda_j}{\lambda_1} \right)^m x^{(j)} \right\}. \quad (16)$$

Kako $\left(\frac{\lambda_j}{\lambda_1} \right)^m \rightarrow 0$ za $m \rightarrow \infty$ i $j \geq 2$, to za dovoljno velike m dobivamo

$$A^m y \approx c_1 \lambda_1^m x^{(1)} \quad (17)$$

tj. $A^m y$ aproksimira vektor $c_1 \lambda_1^m x^{(1)}$ koji se do na koeficijent razlikuje od $x^{(1)}$ pa je dakle i sam svojstveni vektor od A koji pripada svojstvenoj vrijednosti λ_1 .

Napomena. Kako obično ne znamo da li smo u slučaju 1, to nas formalna primjena opisanog postupka može kod neprikladnog izbora vektora y dovesti u situaciju da formula (13) ne daje korijen polinoma $D(\lambda)$ ili da (13) nema limes. Tada treba primijeniti početni vektor y .

2. Slučaj je komplementaran prvom, a taj je kada apsolutna vrijednost dominantne svojstvene vrijednosti ima kratnost, tj. vrijedi

$$|\lambda_1| = |\lambda_2| = \dots = |\lambda_s| > |\lambda_{s+1}|. \quad (18)$$

Tada do formule (11) možemo ponoviti izvod. Sljedeći korak izvoda sada glasi

$$\frac{y_i^{(m+1)}}{y_i^{(m)}} = \lambda_1 \frac{c_1 x_{i1} + \dots + c_s x_{is} + c_{s+1} x_{i,s+1} \left(\frac{\lambda_{s+1}}{\lambda_1} \right)^{m+1} + \dots + c_n x_{in} \left(\frac{\lambda_n}{\lambda_1} \right)^{m+1}}{c_1 x_{i1} + \dots + c_s x_{is} + c_{s+1} x_{i,s+1} \left(\frac{\lambda_{s+1}}{\lambda_1} \right)^m + \dots + c_n x_{in} \left(\frac{\lambda_n}{\lambda_1} \right)^m}.$$

I sada imamo da za $m \rightarrow \infty$ vrijedi $\left(\frac{\lambda_k}{\lambda_1} \right)^m \rightarrow 0$ za $k > s$, pa ako je još i $c_1 x_{i1} + \dots + c_s x_{is} \neq 0$ (time izbjegavamo pojavu neodređenog oblika $\frac{0}{0}$), onda

ponovo zaključujemo da je

$$\lambda_i = \lim_{m \rightarrow \infty} \frac{y_i^{(m+1)}}{y_i^{(m)}}, \quad i = 1, 2, \dots, n$$

i analogno je

$$y^{(m)} = A^m y$$

aproksimacija svojstvenog vektora koji pripada svojstvenoj vrijednosti λ_i .

Kao i u prvom slučaju i ovdje možemo ubrzati iteracijski postupak da računamo A^m za $m = 2^k$, tj. da gledamo potencije

$$A, A^2, A^4, A^8, \dots, A^{2^k}.$$

Izvedimo još jednu formulu za aproksimaciju apsolutne vrijednosti dominantne svojstvene vrijednosti (ili spektralnog radiusa) matrice A , i to zbog jednostavnosti samo za prvi slučaj.

Znamo da je trag matrice $Sp(A)$ jednak sumi njenih svojstvenih vrijednosti

$$Sp(A) = \sum_{i=1}^n \lambda_i.$$

Kako su λ_i^m svojstvene vrijednosti od A^m , to imamo

$$Sp(A^m) = \sum_{i=1}^n \lambda_i^m.$$

Računanje traga se pojednostavljuje ako računamo samo potencije A^m , $m = 2^k$.

Naime, označimo li s $a_{ij}^{(m)}$ elemente od A^m , onda za potencije broja 2 imamo

$$\begin{aligned} Sp(A) &= \sum_{i=1}^n a_{ii} \\ Sp(A^2) &= \sum_{i=1}^n \sum_{k=1}^n a_{ik} a_{ki} \\ Sp(A^4) &= \sum_{i=1}^n \sum_{k=1}^n a_{ik}^{(2)} a_{ki}^{(2)} \\ &\vdots \\ Sp(A^{2^{k+1}}) &= \sum_{i=1}^n \sum_{k=1}^n a_{ik}^{(2^k)} a_{ki}^{(2^k)}. \end{aligned} \tag{19}$$

Sada u 1. slučaju imamo

$$\lambda_1^m + \lambda_2^m + \dots + \lambda_n^m = \lambda_1^m \left[1 + \left(\frac{\lambda_2}{\lambda_1} \right)^m + \dots + \left(\frac{\lambda_n}{\lambda_1} \right)^m \right] = Sp(A^m). \tag{20}$$

Vađenje m -tog korijena iz (20) daje

$$|\lambda_1| \left\{ 1 + \left(\frac{\lambda_2}{\lambda_1} \right)^m + \dots + \left(\frac{\lambda_n}{\lambda_1} \right)^m \right\}^{\frac{1}{m}} = \sqrt[m]{Sp(A^m)},$$

tako da za $m \rightarrow \infty$ postižemo

$$|\lambda_1| = \lim_{m \rightarrow \infty} \sqrt[m]{Sp(A^m)}, \quad (21)$$

i za dovoljno velik m pripadnu aproksimaciju

$$|\lambda_1| \approx \sqrt[m]{Sp(A^m)}. \quad (22)$$

Sljedeća dosjetka eliminira vađenje m -tog korijena $A^{m+1} = A^m A$, što je još jedno množenje matrica. Tada je

$$\frac{\lambda_1^{m+1} + \lambda_2^{m+1} + \dots + \lambda_n^{m+1}}{\lambda_1^m + \lambda_2^m + \dots + \lambda_n^m} = \frac{Sp(A^{m+1})}{Sp(A^m)}$$

ili

$$\lambda_1 \frac{1 + \left(\frac{\lambda_2}{\lambda_1}\right)^{m+1} + \dots + \left(\frac{\lambda_n}{\lambda_1}\right)^{m+1}}{1 + \left(\frac{\lambda_2}{\lambda_1}\right)^m + \dots + \left(\frac{\lambda_n}{\lambda_1}\right)^m} = \frac{Sp(A^{m+1})}{Sp(A^m)}.$$

Kako je $\frac{\lambda_i}{\lambda_1} < 1$, to za dovoljno velike m dobivamo aproksimaciju

$$\lambda_1 \approx \frac{Sp(A^{m+1})}{Sp(A^m)}. \quad (23)$$

5.

Nelinearne jednadžbe

5.1. Jednadžbe s jednom nepoznanim	134
5.2. Algebarske jednadžbe	148
5.3. Sustavi nelinearnih jednadžbi	164

5.1. Jednadžbe s jednom nepoznanim

Prisjetimo se osnovnih činjenica. Neka je dana jednadžba

$$f(x) = 0. \quad (1)$$

Riješiti jednadžbu (1) znači naći one x za koje vrijedi jednakost (1). Svi takvi x čine skup rješenja jednadžbe (1). Uobičajena podjela jednadžbi je na:

1. Algebarske koje su oblika

$$a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n = 0, \quad a_0 \neq 0. \quad (2)$$

2. Transcendentne, tj. one koje nisu algebarske.

Prisjetimo se da je rješenje algebarskih jednadžbi uobičajeno zvati **korijeni** jednadžbe. Ponovimo i neke osnovne činjenice o algebarskim jednadžbama s realnim koeficijentima $a_k \in \mathbf{R}$.

a1) Kompleksna rješenja (korijeni) se javljaju u parovima konjugirano kompleksnih brojeva. Ako $a + ib$ ima kratnost s , onda i $a - ib$ ima kratnost s .

a2) Ako je stupanj algebarske jednadžbe neparan, onda ona ima barem jedno realno rješenje (korijen).

a3) Algebarska jednadžba ima n rješenja uračunavši njihove kratnosti.

a4) Iz poznavanja svih rješenja algebarske jednadžbe možemo naći algebarsku jednadžbu u normiranom obliku koji glasi

$$(x - x_1)(x - x_2) \dots (x - x_n) = 0$$

gdje su x_k općenito kompleksni brojevi.

a5) $x = 0$ je rješenje algebarske jednadžbe onda i samo onda ako je $a_n = 0$.

a6) Ako je $a_n \neq 0$, onda su rješenja x_k algebarske jednadžbe u kompleksnoj ravnini unutar kružnog vijenca

$$r < |x_k| < R. \quad (3)$$

Jedna jednostavna, ali gruba ocjena veličina r i R izgleda ovako.

Označimo sa

$$A = \max\{|a_1|, |a_2|, \dots, |a_n|\},$$

gdje su a_k koeficijenti jednadžbe (2). Tada vrijedi

$$|x_k| < 1 + \frac{A}{|a_0|} = R. \quad (4)$$

Označimo sa

$$B = \max\{|a_0|, |a_1|, \dots, |a_{n-1}|\} \quad \text{i} \quad a_n \neq 0.$$

Tada vrijedi

$$|x_k| > \frac{1}{1 + \frac{B}{|a_n|}} = r. \quad (5)$$

a7) Algebarske jednadžbe do uključivo četvrtog stupnja, $n \leq 4$, možemo riješiti direktno, kao npr. kvadratne jednadžbe. Ta rješenja su dana formulama, koje se zbog složenosti rijetko koriste, a rješenja se izračunavaju pomoću koeficijenata jednadžbe i realnih brojeva. Te formule sadrže u sebi zbrajanje, odbijanje, množenje, dijeljenje, potenciranje i korijenovanje. Za opće jednadžbe stupnja $n \geq 5$ to je neizvedivo, što znači da ne možemo napisati formula s konačno spomenutih operacija nad koeficijentima jednadžbe i realnim brojevima. To ne znači da se neke specijalne algebarske jednadžbe stupnja ≥ 5 ne mogu točno riješiti. Tako npr. znamo naći rješenja algebarske jednadžbe

$$x^7 - x = 0,$$

a to su 0 i svi šesti korijeni jedinice. Stoga i kod rješavanja algebarskih jednadžbi trebamo metode kojima aproksimiramo rješenja.

Navedimo i nekoliko činjenica potrebnih kod aproksimativnog rješavanja bilo algebarskih bilo transcendentnih jednadžbi koje su nam manje ili više poznate iz diferencijalnog računa:

b1) Osnovna pretpostavka je da jednadžba ima samo izolirana rješenja. (To je kod algebarskih jednadžbi uvek ispunjeno.)

b2) Postupak aproksimativnog rješavanja sastoji se iz dva koraka.

b2.1) Odjeljivanje rješenja, tj. iznalaženje što je moguće kraćih intervala $[a, b]$ unutar kojih imamo samo jedno rješenje jednadžbe. Za odjeljivanje rješenja često je praktički dovoljno provoditi proces raspolažanja koji se sastoji u tome da dani segment $[a, b]$ raspolovimo, zatim podijelimo na 4, 8, ... dijelova i odredimo predznak funkcija vrijednosti na krajevima pojedinih segmenata.

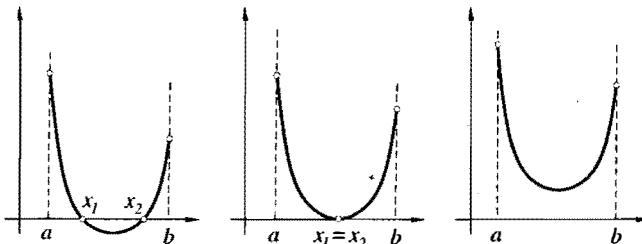
b2.2) Iznalaženje rješenja do tražene točnosti.

b3) Ako je f neprekidna i na krajevima segmenta $[a, b]$ prima vrijednosti s protivnim predznacima, tj. $f(a)f(b) < 0$, onda unutar segmenta $[a, b]$ postoji barem jedno rješenje jednadžbe $f(x) = 0$.

b4) Ako derivacija f' na $[a, b]$ ima isti predznak i vrijedi $f(a)f(b) < 0$, onda je rješenje jedinstveno (f je strogo monotona na $[a, b]$).

b5) Ako je $f(a)f(b) > 0$ i f ima neprekidnu drugu derivaciju f'' na $[a, b]$, koja na $[a, b]$ ne mijenja predznak (f je konveksna ili konkavna), onda unutar $[a, b]$ nastupa jedna od tri mogućnosti (vidi sliku):

- b5.1) $f(x) = 0$ ima dva rješenja;
- b5.2) $f(x) = 0$ ima jedno dvostruko rješenje;
- b5.3) $f(x) = 0$ nema rješenja.



Sl. 5.1.

Spomenimo bez dokaza ovaj dovoljan uvjet kada smo u slučaju b5.3). Neka dodatno f'' nema nultočku na $[a, b]$. Ako vrijedi

$$\frac{f(b)}{f'(b)} - \frac{f(a)}{f'(a)} \geq b - a$$

onda $f(x) = 0$ nema rješenja.

Za ocjenu pogreške aproksimacije lako se pomoću teorema o srednjoj vrijednosti dokaže sljedeća činjenica:

b6) Neka je ξ točna, a \bar{x} aproksimativna vrijednost rješenja jednadžbe $f(x) = 0$ koja pripada segmentu $[a, b]$ na kojem je $|f'(x)| \geq m > 0$. Tada vrijedi sljedeće ocjena

$$|\bar{x} - \xi| \leq \frac{|f(\bar{x})|}{m}. \quad (6)$$

Metoda raspolavljanja

Promatrajmo jednadžbu

$$f(x) = 0$$

gdje je f neprekidna funkcija na $[a, b]$ i $f(a)f(b) < 0$. Raspolovimo $[a, b]$ na dva podsegmenta $\left[a, \frac{a+b}{2}\right]$ i $\left[\frac{a+b}{2}, b\right]$ i provjerimo na kojemu od njih f mijenja predznak. Taj podsegment ponovo raspolovimo i potražimo na kojem dijelu f mijenja predznak. Nastavimo li postupak raspolavljanja dobivamo niz segmenata

$$[a_1, b_1], [a_2, b_2], \dots, [a_n, b_n], \dots$$

jedan unutar drugog za koje vrijedi

$$f(a_n)f(b_n) < 0, \quad n = 1, 2, \dots$$

$$b_n - a_n = \frac{1}{2^n}(b - a),$$

tako da je rješenje ξ jednako

$$\xi = \lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n,$$

dok su a_n i b_n aproksimacije od ξ .

Metoda iteracije

Ovo je jedan od najvažnijih postupaka za numeričko rješavanje jednadžbi. Neka je dana jednadžba

$$f(x) = 0, \quad (7)$$

i neka je za početak f neprekidna funkcija. Dodavanjem x na lijevu i desnu stranu jednadžbe (7) dobivamo

$$x + f(x) = x,$$

što prelazi u

$$\varphi(x) = x, \quad (8)$$

gdje je $\varphi(x) = x + f(x)$. Jasno je da su jednadžbe (7) i (8) ekvivalentne. Na jednadžbu (8) možemo primijeniti sljedeći postupak. Izaberimo na bilo koji način aproksimativnu vrijednost x_0 rješenja jednadžbe (8). Uvrstimo li x_0 u lijevu stranu od (8) dobivamo

$$x_1 = \varphi(x_0).$$

Uvrstimo li sada x_1 u lijevu stranu od (8) dobivamo

$$x_2 = \varphi(x_1).$$

Nastavimo li postupak dobivamo

$$x_n = \varphi(x_{n-1}), \quad (n = 1, 2, \dots). \quad (9)$$

Ako dobiveni niz $\{x_n\}$ konvergira, onda graničnim prelazom dobivamo

$$\xi = \lim_{n \rightarrow \infty} x_n = \varphi(\lim_{n \rightarrow \infty} x_{n-1}) = \varphi(\xi), \quad (10)$$

odnosno

$$f(\xi) = 0,$$

pa smo time našli rješenje polazne jednadžbe. Sljedeći teorem kaže pod kojim uvjetima postupak konvergira.

Teorem 5.1. *Neka je $\varphi : [a, b] \rightarrow [a, b]$ diferencijabilna na $[a, b]$. (Uočite da je kodomena od φ jednaka domeni.) Ako je*

$$|\varphi'(x)| \leq q < 1 \quad \text{za } a < x < b,$$

onda postupak iteracije

$$x_n = \varphi(x_{n-1}), \quad n = 1, 2, \dots$$

konvergira i to neovisno o početnoj vrijednosti $x_0 \in [a, b]$, a

$$\xi = \lim_{n \rightarrow \infty} x_n$$

je jedinstveno rješenje jednadžbe (8) na segmentu $[a, b]$.

Dokaz. Promatrajmo aproksimacije x_n i x_{n+1} koje zbog prepostavki teorema imaju smisla. Sada je

$$x_{n+1} - x_n = \varphi(x_n) - \varphi(x_{n-1}).$$

Primjenom teorema o srednjoj vrijednosti dobivamo

$$x_{n+1} - x_n = (x_n - x_{n-1})\varphi'(\bar{x}_n), \quad \bar{x}_n \in (x_{n-1}, x_n).$$

Zbog $|\varphi'(x)| \leq q$

$$|x_{n+1} - x_n| \leq q|x_n - x_{n-1}| \quad (11)$$

što redom daje

$$|x_2 - x_1| \leq q|x_1 - x_0|$$

$$\vdots$$

$$|x_{n+1} - x_n| \leq q^n|x_1 - x_0|.$$

Promatrajmo red

$$x_0 + (x_1 - x_0) + (x_2 - x_1) + \dots + (x_n - x_{n-1}) + \dots,$$

čija je $(n+1)$ -va parcijalna suma jednaka

$$s_{n+1} = x_n.$$

Promatrani red je konvergentan, jer mu je, zbog $q < 1$, majoranta konvergentan geometrijski red, pa niz $s_{n+1} = x_n$ konvergira i

$$\lim x_n = \xi \in [a, b]$$

je rješenje, tj. vrijedi

$$\xi = \varphi(\xi).$$

Treba još dokazati jedinstvenost rješenja. Neka je $\bar{\xi} \in [a, b]$ neko drugo rješenje, onda je

$$\bar{\xi} - \xi = \varphi(\bar{\xi}) - \varphi(\xi)$$

i prema tome

$$\bar{\xi} - \xi = (\bar{\xi} - \xi)\varphi'(c), \quad c \in (\xi, \bar{\xi}),$$

ili

$$(\bar{\xi} - \xi)[1 - \varphi'(c)] = 0.$$

$\bar{\xi} - \xi \neq 0$ povlači $1 = \varphi'(c)$, što je nemoguće. Dakle, ne postoji drugo rješenje. Time je teorem u potpunosti dokazan, a ujedno vidimo da postupak konvergira pri proizvoljnom izboru početne točke x_0 . Teorem vrijedi i kad je φ diferencijabilna na beskonačnom intervalu $(-\infty, \infty)$ uz isti uvjet $|\varphi'(x)| \leq q < 1$ na $(-\infty, \infty)$.

Sada treba ocijeniti pogrešku aproksimacije. Promatrajmo

$$\begin{aligned} |x_{n+p} - x_n| &= |x_{n+p} - x_{n+p-1} + x_{n+p-1} - \dots + x_{n+1} - x_n| \\ &\leq |x_{n+p} - x_{n+p-1}| + |x_{n+p-1} - x_{n+p-2}| + \dots + |x_{n+1} - x_n| \\ &\leq q^{n+p-1}|x_1 - x_0| + q^{n+p-2}|x_1 - x_0| + \dots + q^n|x_1 - x_0| \\ &= q^n|x_1 - x_0|(1 + q + q^2 + \dots + q^{p-1}). \end{aligned}$$

Dakle

$$|x_{n+p} - x_n| \leq q^n |x_1 - x_0| \frac{1 - q^p}{1 - q} < \frac{q^n}{1 - q} |x_1 - x_0|. \quad (12)$$

Iz $\lim_{p \rightarrow \infty} x_{n+p} = \xi$ dobivamo

$$|\xi - x_n| \leq \frac{q^n}{1 - q} |x_1 - x_0|, \quad (13)$$

pa je postupak iteracije tim bolji što je q manji.

Za ocjenu pogreške aproksimacije imamo i drugu formulu. Pođemo li od

$$g(x) = x - \varphi(x)$$

imamo

$$g'(x) = 1 - \varphi'(x) \geq 1 - q.$$

Kako je $g(\xi) = 0$ to imamo

$$|x_n - \varphi(x_n)| = |g(x_n) - g(\xi)| = |x_n - \xi| |g'(\bar{x}_n)| \geq (1 - q) |x_n - \xi|,$$

gdje je $\bar{x}_n \in (x_n, \xi)$, i prema tome

$$|x_n - \xi| \leq \frac{|x_n - \varphi(x_n)|}{1 - q}$$

odnosno

$$|\xi - x_n| \leq \frac{|x_{n+1} - x_n|}{1 - q},$$

što kombinirajući s (11) daje

$$|\xi - x_n| \leq \frac{q}{1 - q} |x_n - x_{n-1}|. \quad (14)$$

Ta formula omogućava ocijeniti pogrešku aproksimacije iz razlike uzastopnih aproksimativnih rješenja. Ako je zadana točnost $\varepsilon > 0$, postupak iteracije treba voditi dok ne bude

$$\frac{q}{1 - q} |x_n - x_{n-1}| < \varepsilon,$$

iz čega slijedi da postupak treba voditi dok ne bude

$$|x_n - x_{n-1}| < \frac{1 - q}{q} \varepsilon. \quad (15)$$

Najjednostavniji i najefikasniji način je da se iterativni postupak provodi dok ne postignemo

$$|x_n - x_{n-1}| < \varepsilon \quad (16)$$

gdje je ε zadana točnost.

Istaknimo odmah sljedeće. Prelaz s jednadžbe

$$f(x) = 0$$

na ekvivalentnu jednadžbu

$$x = \varphi(x)$$

nije jedinstven. Stoga treba uvijek paziti da uvođimo teorema

$$|\varphi'(x)| \leq q < 1 \quad \text{za } a < x < b$$

bude osiguran. To ilustriramo primjerom jedne kubne jednadžbe.

Primjer 5.1. Promatrajmo kubnu jednadžbu

$$x^3 - 4x^2 + x - 10 = 0.$$

Lako provjeravamo da na segmentu $[4, 6]$ postoji rješenje jednadžbe. Jasno je da tu jednadžbu možemo prevesti u oblik $x = \varphi(x)$ na više načina. Odmah se nameće oblik $x = 10 + 4x^2 - x^3$, no taj izbor nije dobar jer nemamo potrebnu ogradiju za $|\varphi'(x)| = |8x - 3x^2|$, pa nemamo garanciju da u tom slučaju postupak iteracije konvergira. Međutim, izračunamo li x iz prvog člana, tj. napišemo li

$$x = \frac{4x^2 - x + 10}{x^2}$$

imamo

$$\varphi'(x) = \frac{1}{x^2} - \frac{20}{x^3} = \frac{x - 20}{x^3}$$

pa vidimo da na $[4, 6]$ vrijedi $|\varphi'(x)| < 1$. Odabравши $x_0 = 4$ dobivamo

n	x_n	n	x_n
0	4	4	4.3063
1	4.4	5	4.3070
2	4.29	6	4.3069
3	4.31	7	4.3069

Newtonova metoda (metoda tangente)

Neka je ξ rješenje jednadžbe $f(x) = 0$ na segmentu $[a, b]$ i neka su f' i f'' neprekidne funkcije koje ne mijenjaju predznak na $[a, b]$. Ako je x_n aproksimacija rješenja, onda možemo naći bolju aproksimaciju tako da uočimo da je

$$\xi = x_n + h_n$$

i da pokušamo ocijeniti h_n . Prema Taylorovoj formuli imamo

$$0 = f(x_n + h_n) \approx f(x_n) + h_n f'(x_n)$$

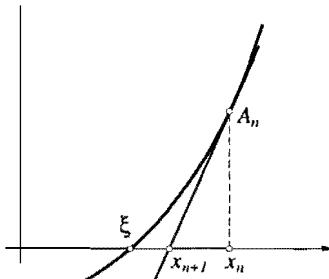
odnosno

$$h_n \approx -\frac{f(x_n)}{f'(x_n)}.$$

Odatle nalazimo novu aproksimaciju

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, 2, \dots \tag{17}$$

Geometrijski gledano, gornji postupak se sastoji u sljedećem (v. sl. 5.2).



Sl. 5.2.

Jednadžba tangente u točki $A_n(x_n, f(x_n))$ glasi

$$y - f(x_n) = f'(x_n)(x - x_n),$$

a njeno sjecište s x -osi je dano s

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)},$$

što nam predstavlja novu aproksimaciju od ξ . Iz ovog opisa je jasno otkuda dolazi ime metoda tangente. Taj se postupak još zove **Newtonov iteracijski postupak** ili **Newtonova metoda**, a ponekad i **Newton – Raphsonova metoda**. U sljedećem teoremu su dani dovoljni uvjeti pod kojima postupak konvergira.

Teorem 5.2. *Neka je $f(a)f(b) < 0$, neka su f' i f'' različite od nule i ne mijenjaju predznak na $[a, b]$. Ako podemo od neke točke $x_0 \in [a, b]$ za koju vrijedi*

$$f(x_0)f''(x_0) > 0$$

i računamo

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, 2, \dots,$$

onda niz $\{x_n\}$ konvergira k jedinstvenom rješenju ξ jednadžbe $f(x) = 0$.

Dokaz. Neka je npr. $f(a) < 0$, $f(b) > 0$, $f'(x) > 0$, $f''(x) > 0$ za $a \leq x \leq b$ (ostali slučajevi se razmatraju analogno). Za x_0 možemo odabratи b jer je $f(b)f''(b) > 0$. Pokažimo prvo da sada za sve aproksimacije vrijedi $x_n > \xi$ ($n = 0, 1, \dots$) i prema tome $f(x_n) > 0$. U prvom redu je $x_0 > \xi$. Indukcijom, neka je $x_n > \xi$. Stavimo $\xi = x_n + (\xi - x_n)$. Prema Taylorovoj formuli imamo

$$0 = f(\xi) = f(x_n) + f'(x_n)(\xi - x_n) + \frac{1}{2}f''(c_n)(\xi - x_n)^2$$

gdje je $\xi < c_n < x_n$. Kako je $f''(x) > 0$, to odbacivanjem pozitivnog pribrojnika imamo

$$f(x_n) + f'(x_n)(\xi - x_n) < 0 \tag{18}$$

odnosno

$$\xi < x_n - \frac{f(x_n)}{f'(x_n)}$$

i prema tome

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} > \xi.$$

Dakle, niz $\{x_n\}$ je ograđen, a iz (17) i činjenice da je $f(x_n) > 0$ slijedi

$$x_{n+1} < x_n \quad (19)$$

pa zaključujemo da je niz $\{x_n\}$ konvergentan. Postoji dakle

$$\bar{\xi} = \lim_{n \rightarrow \infty} x_n. \quad (20)$$

Sada još treba dokazati da je $\xi = \bar{\xi}$. Međutim, zbog neprekidnosti od f , f' i $f'(x) \neq 0$ dobivamo iz (17) graničnim prijelazom

$$\bar{\xi} = \bar{\xi} - \frac{f(\bar{\xi})}{f'(\bar{\xi})}$$

ili

$$f(\bar{\xi}) = 0.$$

Monotonost od f povlači da je $\bar{\xi} = \xi$, pa je time teorem dokazan.

Za ocjenu pogreške n -te aproksimacije možemo se poslužiti formulom (6)

$$|\xi - x_n| \leq \frac{|f(x_n)|}{m}, \quad (21)$$

gdje je $m \leq |f'(x)|$ za $x \in [a, b]$.

Izvedimo jednu drugu formulu za ocjenu pogreške. Iz Taylorove formule imamo

$$\begin{aligned} f(x_n) &= f[x_{n-1} + (x_n - x_{n-1})] \\ &= f(x_{n-1}) + f'(x_{n-1})(x_n - x_{n-1}) + \frac{1}{2}f''(\xi_{n-1})(x_n - x_{n-1})^2, \end{aligned} \quad (22)$$

gdje je $\xi_{n-1} \in (x_{n-1}, x_n)$. Prema (17) imamo

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})},$$

pa dobivamo

$$f(x_{n-1}) + f'(x_{n-1})(x_n - x_{n-1}) = 0,$$

što uvršteno u (22) daje

$$f(x_n) = \frac{1}{2}f''(\xi_{n-1})(x_n - x_{n-1})^2.$$

Dakle,

$$|f(x_n)| \leq \frac{1}{2}M(x_n - x_{n-1})^2$$

gdje je $M \geq |f''(x)|$ na $[a, b]$. Uvrstimo li dobiveno u (21) postižemo ovu formulu za ocjenu pogreške

$$|\xi - x_n| \leq \frac{M}{2m}(x_n - x_{n-1})^2. \quad (23)$$

Ocjena (23) omogućava sljedeće razmatranje. Kada provodimo konvergentan postupak, tada $x_n - x_{n-1} \rightarrow 0$, pa za dovoljno veliki N i svaki $n > N$ imamo $M|x_n - x_{n-1}| < 2m$, tako da vrijedi

$$|\xi - x_n| \leq |x_n - x_{n-1}|. \quad (24)$$

Istaknimo da $|x_n - x_{n-1}| < \varepsilon$ ne garantira pogrešku aproksimacije manju od ε , pa u tom poslu treba biti oprezan. (Naime, možda još nismo premašili N da vrijedi (24).)

Izvedimo još i formulu koja povezuje absolutne pogreške uzastopnih aproksimacija. Iz Taylorove formule

$$0 = f(\xi) = f(x_n) + f'(x_n)(\xi - x_n) + \frac{1}{2}f''(c_n)(\xi - x_n)^2, \quad c_n \in (x_n, \xi)$$

imamo

$$\xi = x_n - \frac{f(x_n)}{f'(x_n)} - \frac{1}{2} \frac{f''(c_n)}{f'(x_n)} (\xi - x_n)^2,$$

odnosno

$$\xi - x_{n+1} = -\frac{1}{2} \frac{f''(c_n)}{f'(x_n)} (\xi - x_n)^2$$

iz čega dobivamo ovu ocjenu

$$|\xi - x_{n+1}| \leq \frac{M}{2m} (\xi - x_n)^2. \quad (25)$$

Ta formula uspoređuje absolutne pogreške dviju uzastopnih aproksimacija. Ona ujedno govori o brzini konvergencije Newtonovog postupka ako se početni x_0 odabere takav da vrijedi

$$\frac{M}{2m} |\xi - x_0| \leq q < 1.$$

Posebno, ako je $\mu = \frac{M}{2m} \leq 1$ i $|\xi - x_n| < 10^{-k}$ onda iz te formule slijedi $|\xi - x_{n+1}| < 10^{-2k}$, pa ako je aproksimacija bila točna na prvih k decimalnih mjestima, onda sljedeća aproksimacija ima točno prvih $2k$ decimalnih mjestima. Kraće rečeno, za $\mu \leq 1$ Newtonova metoda udvostručava u svakom koraku broj točnih decimalnih mesta. Takav tip konvergencije iteracijskog postupka zove se **kvadratična** konvergencija. U mnogim slučajevima dovoljno je samo pet ili šest iteracija da dobijemo ξ točno na osam i više decimalnih mesta. Kod konkretnog računa obično vodimo postupak dok ne postignemo

$$|x_{n+1} - x_n| = \left| \frac{f(x_n)}{f'(x_n)} \right| = h < \varepsilon, \quad (26)$$

gdje je ε zadana točnost. Međutim, bolje je ako se možemo poslužiti npr. ocjenom pogreške (21).

Primjer 5.2. Riješite jednadžbu $e^x = 10 - x$ na 4 decimalna mesta. Sada imamo $f(x) = e^x + x - 10$, $f'(x) = e^x + 1$ i

$$x_{n+1} = x_n - \frac{e^{x_n} + x_n - 10}{e^{x_n} + 1}.$$

Lako ocjenimo da je $f(0) < 0$, $f(1) < 0$, $f(2) < 0$, ali $f(3) > 0$. Time znamo da je rješenje između 2 i 3. Izborom $x_0 = 2$ dobivamo sljedeće aproksimacije:

n	x_n	n	x_n
0	2	2	2.071
1	2.073	3	2.071

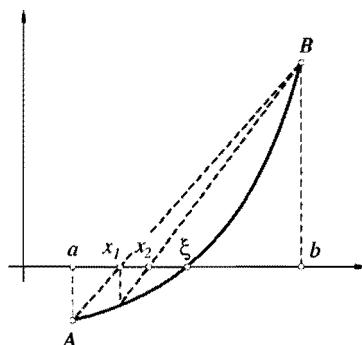
Potpunosti radi spomenimo sljedeće. Zamislimo da smo na početku proveli aproksimaciju sa uključivo kvadratnim članom Taylorove formule, tj. da smo gledali

$$0 = f(x_n + h_n) \approx f(x_n) + h_n f'(x_n) + \frac{1}{2} f''(x_n) h_n^2. \quad (27)$$

Tada se iznalaženje nove aproksimacije x_{n+1} svodi na rješavanje kvadratne jednadžbe što se može provesti, ali izostavljamo.

Metoda sekante (regula falsi)

Promatrajmo ponovo isti problem i neka je npr. $f(a) < 0$ i $f(b) > 0$ (v. sl. 5.3).



Sl. 5.3.

Jednadžba sekante kroz krajnje točke luka \widehat{AB} glasi

$$\frac{x - a}{b - a} = \frac{y - f(a)}{f(b) - f(a)}.$$

Za $y = 0$ dobivamo sjecište sekante s x -osi

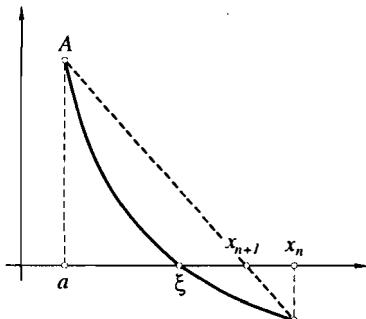
$$x_1 = a - \frac{f(a)}{f(b) - f(a)}(b - a).$$

Ponovimo sada postupak na segmentu $[x_1, b] \subset [a, b]$ kao što je ilustrirano na sl. 5.3. Taj postupak možemo nastaviti pri čemu općenito niz $\{x_i\}$ može divergirati. Da osiguramo konvergenciju postupka pretpostavimo da f'' ne mijenja predznak na $[a, b]$. Time je f konveksna funkcija na $[a, b]$ pa sekanta siječe graf od f nad $[a, b]$ samo u krajnjim točkama A, B . Neka je npr. $f''(x) > 0$ na $[a, b]$. Slučaj $f''(x) < 0$ svodi se na rješavanje jednadžbe $-f(x) = 0$. Uz $f''(x) > 0$ nastupaju dva pod slučaja.

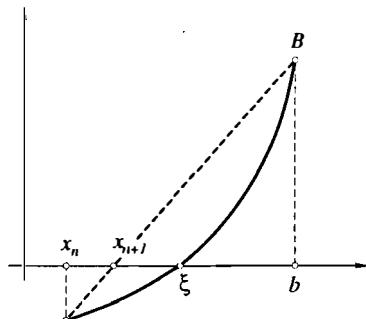
1) Za $f(a) > 0$ (v. sl. 5.4) uzimamo $x_0 = b$ za prvu aproksimaciju. Tada imamo

$$x_{n+1} = x_n - \frac{f(x_n)}{f(x_n) - f(a)}(x_n - a), \quad n = 0, 1, 2, \dots \quad (28)$$

što zbog konveksnosti od f daje jedan ograđen monotono padajući niz aproksimacija $\{x_n\}$ pa prema tome niz $\{x_n\}$ konvergira k nekom $\xi = \lim_{n \rightarrow \infty} x_n$.



Sl. 5.4.



Sl. 5.5.

2) Za $f(a) < 0$ (v. sl. 5.5) uzimamo $x_0 = a$. Tada imamo

$$x_{n+1} = x_n - \frac{f(x_n)}{f(b) - f(x_n)}(b - x_n), \quad n = 0, 1, 2, \dots \quad (29)$$

Zbog konveksnosti od f dobivamo ograđen monotono rastući niz aproksimacija $\{x_n\}$ koji konvergira k nekom $\xi = \lim_{n \rightarrow \infty} x_n$.

Kako je $\xi = \lim_{n \rightarrow \infty} x_{n+1} = \lim_{n \rightarrow \infty} x_n$, to iz (28) dobivamo

$$\xi = \xi - \frac{f(\xi)}{f(\xi) - f(a)}(\xi - a)$$

što daje $f(\xi) = 0$, tj. ξ je rješenje jednadžbe. Analogno razmatranje imamo u slučaju 2).

Usporedimo li metodu sekante s Newtonovom metodom, onda je njena prednost u tome što ne zahtijeva izračunavanje derivacije f' , međutim sporije konvergira od Newtonove metode.

Za ocjenu pogreške aproksimacije može nam poslužiti iz uvida poznata formula

$$|x_n - \xi| \leq \frac{|f(x_n)|}{m} \quad (30)$$

gdje je

$$m \leq |f'(x)|, \quad a \leq x \leq b.$$

Međutim, računski je zanimljiva formula koja daje ocjenu pogreške pomoću izračunatih susjednih aproksimacija x_{n-1} i x_n . Neka je f' ograđena na $[a, b]$, tj.

$$m \leq |f'(x)| \leq M, \quad x \in [a, b]. \quad (31)$$

Razmotrimo samo slučaj 1), tj. kada je u postupku $x_0 = b$ (analogno u slučaju 2)). Iz (28) lako dobivamo

$$-f(x_{n-1}) = \frac{f(x_{n-1}) - f(a)}{x_{n-1} - a}(x_n - x_{n-1})$$

pa dodavanjem $f(\xi) = 0$ imamo

$$f(\xi) - f(x_{n-1}) = \frac{f(x_{n-1}) - f(a)}{x_{n-1} - a}(x_n - x_{n-1}). \quad (32)$$

Po teoremu o srednjoj vrijednosti primijenjenom na obje strane jednakosti (32) dobivamo

$$(\xi - x_{n-1})f'(\xi_{n-1}) = (x_n - x_{n-1})f'(\bar{x}_{n-1}), \quad (33)$$

$\xi_{n-1} \in (x_{n-1}, \xi)$, $\bar{x}_{n-1} \in (a, x_{n-1})$. Umetanjem $0 = x_n - x_n$ u prvi faktor lijeve strane od (33) lako dobivamo

$$\begin{aligned} |\xi - x_n| &= \frac{|f'(\bar{x}_{n-1}) - f'(\xi_{n-1})|}{|f'(\xi_{n-1})|} |x_n - x_{n-1}| \\ &\leq \frac{M - m}{|f'(\xi_{n-1})|} |x_n - x_{n-1}| \\ &\leq \frac{M - m}{m} |x_n - x_{n-1}|. \end{aligned} \quad (34)$$

Ako je nadalje duljina segmenta $[a, b]$ dovoljno malena da smo postigli $M \leq 2m$, onda iz (34) dobivamo

$$|\xi - x_n| \leq |x_n - x_{n-1}|, \quad (35)$$

pa u takvoj situaciji vodimo postupak dok ne bude $|x_n - x_{n-1}| < \varepsilon$, gdje je ε dana točnost.

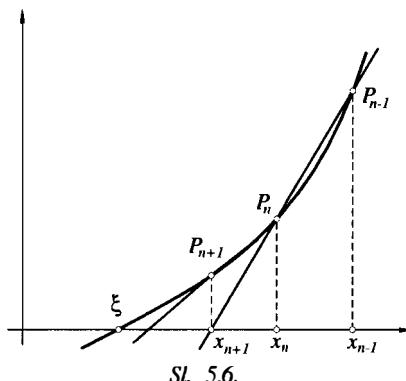
Primjer 5.3. Riješimo jednadžbu $x^r = 10$ na 3 decimalna mjesta. Logaritmiranjem imamo

$$x \ln x - \ln 10 = 0.$$

Uzmemmo li $f(x) = x \ln x - \ln 10$, onda vidimo da je $f'(x) = \ln x + 1$ iz čega zaključujemo da f pada na $(0, \frac{1}{e})$ i raste na $(\frac{1}{e}, \infty)$. U točki $x = \frac{1}{e}$ ima očito minimum. Kako za $x \rightarrow +0$, $f(x) \rightarrow -\ln 10$, to vidimo da postoji samo jedno rješenje ξ promatrane jednadžbe za koje vrijedi $\xi > \frac{1}{e}$. Bolju ocjenu intervala unutar kojega se nalazi ξ nalazimo uočavanjem da je $f(2) < 0$ i $f(3) > 0$. Odabiremo $a = 2$, $b = 3$. Iz $f''(x) = \frac{1}{x}$ vidimo da je f konveksna na $(0, \infty)$. Prema (29) imamo

n	x_n	n	x_n
0	2	3	2.5061
1	2.4798	4	2.5061
2	2.5049	5	2.5061

U dosadašnjem opisu metode sekante držali smo jedan kraj segmenta $[a, b]$ čvrstim. Postoji i modifikacija metode sekante kod koje mijenjamo oba kraja sekante. To je ilustrirano na slici 5.6.



Sada se polazi od dviju aproksimacija x_{n-1} i x_n (pretpostavljamo konveksnost od f na promatranoj okolini od ξ) koje su npr. obje veće od ξ . Iz jednadžbe sekante kroz P_{n-1} i P_n

$$y - f(x_n) = \frac{f(x_{n-1}) - f(x_n)}{x_{n-1} - x_n}(x - x_n)$$

dobivamo

$$x_{n+1} = \frac{x_{n-1}f(x_n) - x_nf(x_{n-1})}{f(x_n) - f(x_{n-1})}. \quad (36)$$

Izvedimo još i ocjenu pogreške tipa (34). Iz (36) imamo

$$-f(x_n) = (x_{n+1} - x_n) \frac{f(x_{n-1}) - f(x_n)}{x_{n-1} - x_n}$$

tako da dodavanjem $f(\xi) = 0$ imamo

$$f(\xi) - f(x_n) = \frac{f(x_{n-1}) - f(x_n)}{x_{n-1} - x_n}(x_{n+1} - x_n).$$

Primjenom teorema srednje vrijednosti na obje strane dobivamo

$$(\xi - x_n)f'(\xi_n) = \frac{f'(\bar{x}_{n-1})(x_{n-1} - x_n)}{x_{n-1} - x_n}(x_{n+1} - x_n)$$

gdje je $\xi_n \in (\xi, x_n)$, $\bar{x}_{n-1} \in (x_n, x_{n-1})$, tako da vrijedi

$$\xi - x_n = \frac{f'(\bar{x}_{n-1})}{f'(\xi_n)}(x_{n+1} - x_n). \quad (37)$$

Iz (37) dobivamo ocjenu

$$|\xi - x_n| \leq \frac{M}{m}|x_{n+1} - x_n|. \quad (38)$$

Sa stanovišta interpolacijskog polinoma mi smo kod metode sekante iterativno aproksimirali f s linearim interpolacijskim polinomom čije nultočke konvergiraju rješenju jednadžbe. To se može provesti i s kvadratnim interpolacijskim polinomom (Mullerova metoda). Isto se tako metode mogu kombinirati iz koraka u korak što ovdje izostavljam.

5.2. Algebarske jednadžbe

Neka daljnja svojstva polinoma

U ovom paragrafu ćemo se ukratko baviti problematikom algebarskih jednadžbi koje se vrlo često pojavljuju u primjenama. Tako je npr. karakteristična jednadžba linearne diferencijalne jednadžbe algebarska jednadžba čija rješenja (realna i kompleksna) treba znati da bi ispisali opće rješenje linearne diferencijalne jednadžbe. Promatrat ćemo algebarske jednadžbe s **realnim** koeficijentima. Neka je

$$P(x) = a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n = \sum_{i=0}^n a_i x^{n-i} \quad (1)$$

polinom n -og stupnja s realnim koeficijentima, tj. $a_0 \neq 0$, a_1, \dots, a_n su realni brojevi. Pripadna **algebarska jednadžba** glasi

$$a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n = 0 \quad (2)$$

koja dijeljenjem s a_0 prelazi u **normirani oblik**

$$x^n + p_1x^{n-1} + \dots + p_n = 0, \quad p_i = \frac{a_i}{a_0}, \quad i = 0, 1, 2, \dots, n. \quad (3)$$

Kao što znamo, svaki se polinom (1) može analitički proširiti na cijelu kompleksnu ravninu. To proširenje dobivamo tako da u (1) umjesto x pišemo kompleksnu varijablu $z = x + yi$, pa imamo

$$P(z) = a_0z^n + a_1z^{n-1} + \dots + a_{n-1}z + a_n \quad (4)$$

čije nultočke zovemo rješenja algebarske jednadžbe (2).

Na (4) možemo primijeniti teoriju analitičkih funkcija unutar koje lako dokazujemo niz rezultata o polinomima kao npr. osnovni teorem algebre itd. Prisjetimo se da nam teorem o argumentu primijenjen na polinome daje broj nultočaka N koje se nalaze unutar konture Γ . Ta formula glasi

$$N = \frac{1}{2\pi} \Delta_\Gamma \arg P$$

gdje $\Delta_\Gamma \arg P$ označava promjenu argumenta od P kada jednom obidiemo konturu Γ u pozitivnom smjeru. Prisjetimo se također Rouchéovog teorema koji se lako dokazuje pomoću teorema o argumentu i kaže:

Neka su f i g analitičke funkcije na području G , a $\Gamma \subset G$ kontura čije je nutarnje područje sadržano u G . Ako za točke konture Γ vrijedi nejednakost

$$|f(z)| > |g(z)|, \quad z \in \Gamma$$

onda funkcije f i $f + g$ imaju unutar Γ isti broj nultočaka.

Prema osnovnom teoremu algebre znamo da polinom (4) ima točno n nultočaka, što za nas znači da jednadžba (2) ima točno n rješenja (korijena), realnih i kompleksnih, pod uvjetom da svako rješenje uzmemos sa svojom kratnošću. Znamo da je ξ **s-kratno rješenje** od (2) (ili s -struka nultočka od (4)) ako vrijedi

$$0 = P(\xi) = P'(\xi) = \dots = P^{(s-1)}(\xi) \quad \text{i} \quad P^{(s)}(\xi) \neq 0.$$

Prisjetimo se nadalje da se kompleksna rješenja algebarske jednadžbe s **realnim koeficijentima** javljuju u parovima konjugirano kompleksnih brojeva, te ako je $\alpha + \beta i$ s -kratno, onda je i $\alpha - \beta i$ s -kratno rješenje. (Ako u jednadžbi (4) dopustimo kompleksne koeficijente, onda vrijedi sljedeće: Konjugirano kompleksne vrijednosti rješenja jednadžbe (4) su rješenja jednadžbe koja se iz (4) dobiva tako da joj se koeficijenti zamijene s konjugirano kompleksnim vrijednostima. Naravno, kratnosti se čuvaju.)

Označimo s

$$x_1, x_2, \dots, x_n \quad (5)$$

rješenja jednadžbe (2) (x_i realan ili kompleksan). Tada P možemo pisati u obliku

$$P(x) = a_0(x - x_1) \dots (x - x_n) = a_0 \prod_{i=1}^n (x - x_i) \quad (6)$$

iz čega iz jednačavanjem koeficijenata uz iste potencije dobivamo sljedeće formule

$$\begin{aligned} \sum_{i=1}^n x_i &= -\frac{a_1}{a_0}, & \sum_{i \neq j} x_i x_j &= \frac{a_2}{a_0}, \\ \sum_{i \neq j \neq k} x_i x_j x_k &= -\frac{a_3}{a_0}, \dots, & \prod_{i=1}^n x_i &= (-1)^n \frac{a_n}{a_0} \end{aligned} \quad (7)$$

koje se zovu **Vièta-ove formule**. Nadalje, ako su x_i nultočke od (2), onda su $\frac{1}{x_i}$ nultočke polinoma

$$P^*(x) = a_0 + a_1 x + \dots + a_n x^n = \sum_{i=0}^n a_i x^i. \quad (8)$$

Uzmemo li u obzir kratnost s_i , nultočke x_i , onda (6) prelazi u

$$P(x) = a_0(x - x_1)^{s_1}(x - x_2)^{s_2} \dots (x - x_m)^{s_m} \quad (9)$$

gdje su x_1, \dots, x_m različite nultočke od P i vrijedi

$$s_1 + s_2 + \dots + s_m = n.$$

Granice rješenja algebarskih jednadžbi

Ocijenimo u prvom redu absolutne vrijednosti rješenja jednadžbe (2). Vrijedi sljedeći teorem.

Teorem 5.3. *Neka je*

$$A = \max\{|a_1|, |a_2|, \dots, |a_n|\}.$$

Tada za absolutne vrijednosti $|x_i|$ rješenja jednadžbe (2) vrijedi sljedeća nejednakost

$$|x_i| < 1 + \frac{A}{|a_0|} = R \quad (10)$$

tj. sva se rješenja (realna i kompleksna) nalaze u kompleksnoj ravnini unutar kruga radijusa $R = 1 + \frac{A}{|a_0|}$.

Dokaz. Promatrajmo sljedeće analitičke funkcije

$$f(z) = a_0 z^n$$

$$g(z) = a_1 z^{n-1} + a_2 z^{n-2} + \dots + a_{n-1} z + a_n$$

koje zbrojene daju polinom (4). Sada za $|z| > 1$ vrijedi

$$\begin{aligned} |g(z)| &\leq |a_1||z|^{n-1} + \dots + |a_{n-1}||z| + |a_n| \\ &\leq A(|z|^{n-1} + \dots + |z| + 1) \\ &= A \frac{|z|^n - 1}{|z| - 1} \\ &< A \frac{|z|^n}{|z| - 1}. \end{aligned}$$

Da primijenimo gore spomenuti Rouchéov teorem moramo naći konturu na kojoj vrijedi $|f(z)| > |g(z)|$. Lako određujemo kružnicu za koju to vrijedi. Naime, iz zahtjeva

$$|a_0||z|^n > A \frac{|z|^n}{|z| - 1}$$

dobivamo da mora biti

$$|z| > 1 + \frac{A}{|a_0|}.$$

Prema Rouchéovom teoremu funkcije f i $f + g = P$ imaju isti broj nultočaka unutar kruga radijusa $R = 1 + \frac{A}{|a_0|}$. Kako f ima u $z = 0$ nultočku kratnosti n , to znači da smo tom kružnicom obuhvatili sve nultočke od P , pa je time teorem dokazan.

Odmah imamo sljedeću posljedicu.

Korolar 5.4. *Neka je $a_n \neq 0$ i*

$$B = \max\{|a_0|, |a_1|, \dots, |a_{n-1}|\}.$$

Tada sva rješenja jednadžbe (2) zadovoljavaju nejednakost

$$|x_i| > \frac{1}{1 + \frac{B}{|a_n|}} = r. \quad (11)$$

Dokaz. Uvjet $a_n \neq 0$ povlači da ishodište nije rješenje jednadžbe (2). Sada su prema (8) $\frac{1}{x_i}$ rješenja jednadžbe

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_0 = 0 \quad (12)$$

tako da primjenom teorema 5.3 na (12) dobivamo

$$\left| \frac{1}{x_i} \right| < 1 + \frac{B}{|a_n|}$$

iz čega slijedi (11).

Imamo li dakle $a_n \neq 0$, onda se sva rješenja jednadžbe (2) nalaze unutar kružnog vijenca

$$r = \frac{1}{1 + \frac{B}{|a_n|}} < |x_i| < 1 + \frac{A}{|a_0|} = R \quad (13)$$

gdje je $A = \max\{|a_1|, |a_2|, \dots, |a_n|\}$, $B = \max\{|a_0|, |a_1|, \dots, |a_{n-1}|\}$. Ograničimo li se na realna rješenja, onda su r i R donja i gornja ograda za pozitivna rješenja, a $-R$ i $-r$ za negativna rješenja.

Što se tiče nalaženja **granica realnih rješenja** algebarske jednadžbe dovoljno je naći granice za pozitivna rješenja, jer supstitucijom $-x$ umjesto x dobivamo jednadžbu $P(-x) = 0$ čija se rješenja razlikuju samo u predznaku od rješenja jednadžbe $P(x) = 0$. Sljedeći teorem daje jednostavnu ocjenu gornje ograde pozitivnih rješenja algebarske jednadžbe.

Teorem 5.5. (Lagrange) Neka je $a_0 > 0$ i a_m ($m \geq 1$) prvi u nizu a_0, a_1, \dots, a_n negativni koeficijent polinoma P . Tada su sva pozitivna rješenja jednadžbe (2) omeđena sa

$$R = 1 + \sqrt[m]{\frac{a}{a_0}}$$

gdje je a najveća absolutna vrijednost negativnih koeficijenata od P .

Dokaz. Neka je $x > 1$. Ako pozitivne koeficijente a_1, \dots, a_{m-1} u P zamijenimo nulom, a sve preostale a_m, a_{m+1}, \dots, a_n s $-a$, onda za $x > 1$ vrijedi

$$\begin{aligned} P(x) &= a_0 x^n + a_1 x^{n-1} + \dots + a_n \\ &= a_0 x^n - a(x^{n-m} + x^{n-m-1} + \dots + x + 1) \\ &= a_0 x^n - a \frac{x^{n-m+1} - 1}{x - 1} \\ &> a_0 x^n - a \frac{x^{n-m+1}}{x - 1} \\ &= \frac{x^{n-m+1}}{x - 1} \left\{ a_0 x^{m-1} (x - 1) - a \right\}. \end{aligned}$$

Prema tome, ako je $x \geq 1 + \sqrt[m]{\frac{a}{a_0}}$, onda je prvi faktor pozitivan, a za drugi vrijedi

$$\begin{aligned} a_0 x^{m-1} (x - 1) - a &\geq a_0 \left(1 + \sqrt[m]{\frac{a}{a_0}} \right)^{m-1} \sqrt[m]{\frac{a}{a_0}} - a \\ &= a_0 \left(\sqrt[m]{\frac{a}{a_0}} + \dots + \frac{a}{a_0} \right) - a > 0, \end{aligned}$$

što povlači $P(x) > 0$, čime je dokaz proveden.

Spomenimo da donju ogradu možemo naći kao recipročnu vrijednost gornje gradi jednadžbe (8) kao što smo to učunili u dokazu korolara 5.4.

Ocjena gornje gradi koju daje teorem 5.5 može biti dosta gruba.

Primjer 5.4. Za jednadžbu

$$x^4 - 35x^3 + 380x^2 - 1350x + 1000 = 0$$

imamo $a = 1350$, $a_0 = 1$, $m = 1$, pa iz teorema 5.5 slijedi da su sva pozitivna rješenja manja od 1351. Za iznalaženje donje ograde treba gledati jednadžbu

$$1 - 35x + 380x^2 - 1350x^3 + 1000x^4 = 0$$

za koju je $a_0 = 1000$, $a = 1350$ i $m = 1$, tako da teorem 5.5 daje $\frac{2350}{1000}$ za gornju ogradu njenih pozitivnih rješenja, pa je $\frac{1000}{2350} = 0.425$ donja ograda pozitivnih rješenja polazne jednadžbe.

Nalaženje ograde probanjem. Kako se u nalaženju ograde često služimo probanjem, to navodimo ovaj jednostavan postupak. Pretpostavimo $a_0 > 0$ (u protivnom pomnožimo jednadžbu s -1) i rastavimo promatrani polinom P na zbroj

$$P(x) = Q(x) + R(x) \quad (14)$$

gdje je Q polinom koji se sastoјi od svih vodećih članova s pozitivnim koeficijentima i svih negativnih članova polinoma P (to znači da Q nalazimo tako da u nizu a_0, a_1, \dots, a_n potražimo prvi negativni član te zbrojimo sve članove do tog mjesta i sve negativne članove), a R je polinom sastavljen od svih preostalih članova. Za svaki $x \geq 0$ vrijedi $R(x) \geq 0$. Polinom Q je građen ovako

$$\begin{aligned} Q(x) &= \left(a_0x^n + \dots + a_{m-1}x^{n-m+1} \right) - \left(|a_m|x^{n-m} + \dots \right) \\ &= x^{n-m+1} \left\{ \left(a_0x^{m-1} + \dots + a_{m-1} \right) - \frac{|a_m|x^{n-m} + \dots}{x^{n-m+1}} \right\} \end{aligned}$$

iz čega čitamo da su za $x > 0$ obje okrugle zgrade pozitivne i da prva raste, a druga pada kada x raste. Vrijedi dakle:

Ako nađemo $\alpha > 0$ takav da je u rastavu (14) $Q(\alpha) \geq 0$, onda je za svaki $x > \alpha$ $P(x) > 0$, tj. nađeni α je gornja ograda pozitivnih rješenja jednadžbe $P(x) = 0$.

U gore promatranom primjeru

$$Q(x) = x^4 - 35x^3 - 1350x, \quad R(x) = 380x^2 + 1000.$$

Očito je da 35 nije gornja ograda. Pokušamo li 37 dobivamo $Q(37) > 0$, pa je dakle 37 gornja ograda pozitivnih rješenja. Da nađemo donju ogradu promatramo

$$P^*(x) = 1000x^4 - 1350x^3 + 380x^2 - 35x + 1$$

koji zapisujemo u obliku $P^*(x) = Q^*(x) + R^*(x)$, gdje je

$$Q^*(x) = 1000x^4 - 1350x^3 - 35x, \quad R^*(x) = 380x^2 + 1.$$

Iz zapisa $Q^*(x) = x\{x^2(1000x - 1350) - 35\}$ lako uočavamo da je $Q^*(1, 3) < 0$. Međutim, $Q^*(1, 4) > 0$ pa je dakle 1.4 gornja ograda rješenja jednadžbe $P^*(x) = 0$, a $\frac{1}{1.4} \approx 0.714$ donja ograda za rješenja polazne algebarske jednadžbe. Time smo utvrdili da se pozitivna rješenja jednadžbe $x^4 - 35x^3 + 380x^2 - 1350x + 1000 = 0$ nalaze unutar segmenta $[0.714, 37]$.

Newtonov test. Pogledajmo kako nam Taylorova formula može korisno poslužiti. Taj postupak se zove **Newtonov test**. Razvijemo P po potencijama od $x - \alpha$. Imamo

$$P(x) = P(\alpha) + P'(\alpha)(x - \alpha) + \frac{P''(\alpha)}{2!}(x - \alpha)^2 + \dots + \frac{P^{(n)}(\alpha)}{n!}(x - \alpha)^n. \quad (15)$$

Iz (15) čitamo sljedeće:

Nađemo li $\alpha > 0$ takav da vrijedi

$$P(\alpha) \geq 0, P'(\alpha) \geq 0, \dots, P^{(n)}(\alpha) \geq 0 \quad (16)$$

onda je $P(x) > 0$ za svaki $x > \alpha$, tj. svaki α koji zadovoljava (16) je gornja ograda pozitivnih rješenja jednadžbe (2).

Ocjena broja realnih rješenja

Nakon utvrđivanja ograda za kompleksna i realna rješenja algebarske jednadžbe postavlja se znatno teže pitanje koliko realnih rješenja ima dana algebarska jednadžba. Promatramo li polinom P na segmentu $[a, b]$ lako uočavamo ova dva svojstva:

1) Ako je $P(a)P(b) < 0$, onda P na (a, b) ima neparan broj rješenja uzetih s njihovim kratnostima.

2) Ako je $P(a)P(b) > 0$, onda P na (a, b) ili nema rješenja ili ih ima paran broj.

Upoznajmo se prvo s tzv. **Descartesovim pravilom** koje daje ocjenu broja pozitivnih rješenja, a dokazat ćemo ga nakon teorema 5.7.

Teorem 5.6. (*Descartesovo pravilo*) Broj pozitivnih rješenja N_p jednadžbe

$$a_0x^n + a_1x^{n-1} + \dots + a_n = 0, \quad a_0 > 0$$

računati sa svojim kratnostima, manji je ili jednak broju promjena predznaka N u nizu (koeficijente koji su nula ne uračunavamo)

$$a_0, a_1, a_2, \dots, a_{n-1}, a_n \quad (17)$$

tj. $N_p \leq N$. Nadalje, ako je $N_p < N$, onda je razlika $N - N_p$ paran broj.

Primijenimo li Descartesovo pravilo na primjer 5.4 imamo niz

$$1, -35, 380, -1350, 1000$$

za koji je $N = 4$. Dakle u intervalu $(0.714, 37)$ ima ≤ 4 rješenja. Iz $P(0) > 0$ i $P(37) > 0$ slijedi da unutar $[0.714, 37]$ imamo 0, 2 ili 4 rješenja. Kako je očigledno $P(1) < 0$ to unutar $[1, 37]$ imamo neparan broj rješenja, pa vidimo da i unutar $[0.714, 1]$ postoji neparan broj rješenja jer ih ukupno na $[0.714, 37]$ ima paran broj, koji zbog uočenog ne može biti 0. Dakle P ima 2 ili 4 pozitivna rješenja. Lako provjeravamo da je $P(10) > 0$ i $P(15) < 0$. Iz ta dva podatka zaključujemo da unutar $[0.714, 10]$ i $[10, 37]$ imamo po dva rješenja. Kako unutar $[15, 37]$ imamo neparan broj rješenja zaključujemo da unutar $[10, 37]$ imamo dva jednostruka rješenja. Analogno, zbog $P(1) < 0$ zaključujemo da unutar $[0.714, 10]$ imamo dva jednostruka rješenja. Time smo zaključili da su sva rješenja jednadžbe

$$x^4 - 35x^3 + 380x^2 - 1350x + 1000 = 0$$

realna i to pozitivna i jednostruka.

Dokažimo sada općenitiji teorem čija je posljedica Descartesovo pravilo. U tu se svrhu promatra niz polinoma sastavljen od promatranoj polinoma i njegovih derivacija, tj. niz

$$P, P', P'', \dots, P^{(n)}. \quad (18)$$

Prisjetimo se da je n -ta derivacija polinoma (1) jednaka

$$P^{(n)}(x) = a_0 n!$$

dakle konstanta.

Teorem 5.7. (Budan-Fourier) Neka je $P(a)P(b) \neq 0$. Za broj nultočaka $N(a, b)$ od P unutar $[a, b]$, računatih sa svojim kratnostima vrijedi sljedeća ocjena

$$N(a, b) \leq V(a) - V(b) \quad (19)$$

gdje je $V(a)$, odnosno $V(b)$ broj varijacija (promjena) predznaka u nizu (18) u točki a , odnosno b . Ako je $N(a, b) < V(a) - V(b)$, onda je i razlika $V(a) - V(b) - N(a, b)$ paran broj.

Dokaz. Ako u (18) uvrstimo c , $a \leq c \leq b$ i promatramo niz vrijednosti

$$P(c), P'(c), \dots, P^{(n)}(c) \quad (20)$$

i zamislimo da c raste od a do b , onda će u (20) doći do eventualne varijacije predznaka samo kada prođemo kroz nultočku nekog od članova niza (18). Kako za svaki c između a i b vrijedi

$$V(a) - V(b) = [V(a) - V(c)] + [V(c) - V(b)]$$

vidimo da razmatranje možemo vršiti po podsegmentima od $[a, b]$. Nadalje, na $[a, b]$ se nalazi konačno nultočaka od $P, P', \dots, P^{(n-1)}$, pa je dovoljno promatrati male podsegmente koji sadrže jednu takvu nultočku, jer na podsegmentu na kojem su $P, P', \dots, P^{(n-1)}, P^{(n)}$ ne mijenjaju predznak, broj varijacija predznaka na početku i na kraju segmenta je isti. Istaknimo odmah da $c \in [a, b]$ može biti zajednička nultočka više nesusjednih članova niza (18). Npr. za polinom $P(x) = x^3 - 3x^2 + 4x - 2$ vrijedi: $P(1) = 0; P'(1) = 1; P''(1) = 0; P'''(1) = 6$.

Neka je α r -struka nultočka od P . Tada je po Taylorovoj formuli (15)

$$P(x) = \frac{P^{(r)}(\alpha)}{r!}(x - \alpha)^r + \dots + \frac{P^{(n)}(\alpha)}{n!}(x - \alpha)^n.$$

Neka je $\varepsilon > 0$ takav da ε -okolina od α ne sadrži drugih nultočaka niza (18) i neka je ε odabran dovoljno malen tako da za svaki x iz ε -okoline od α , tj. za koji vrijedi $|x - \alpha| < \varepsilon$, imamo

$$P(x) \approx \frac{P^{(r)}(\alpha)}{r!}(x - \alpha)^r, \quad P'(x) \approx \frac{P^{(r)}(\alpha)}{(r-1)!}(x - \alpha)^{r-1}, \dots, \quad P^{(r)}(x) \approx P^{(r)}(\alpha).$$

(Takov ε je moguće odabrati jer se za male argumente polinom ponaša kao član najnižeg stupnja, a za velike argumente kao član najvišeg stupnja.)

Tada za $x = \alpha - \varepsilon$ imamo

$$P(\alpha - \varepsilon) \approx \frac{P^{(r)}(\alpha)}{r!}(-\varepsilon)^r, \quad P'(\alpha - \varepsilon) \approx \frac{P^{(r)}(\alpha)}{(r-1)!}(-\varepsilon)^{r-1}, \dots, \quad P^{(r)}(\alpha - \varepsilon) \approx P^{(r)}(\alpha),$$

pa u podnizu

$$P(\alpha - \varepsilon), P'(\alpha - \varepsilon), \dots, P^{(r)}(\alpha - \varepsilon)$$

niza (18) promatranog u točki $\alpha - \varepsilon$ predznak alternira, tako da imamo ukupno r varijacija (promjena) predznaka. S druge strane, za $x = \alpha + \varepsilon$ imamo

$$P(\alpha + \varepsilon) \approx \frac{P^{(r)}(\alpha)}{r!} \varepsilon^r, P'(\alpha + \varepsilon) \approx \frac{P^{(r)}(\alpha)}{(r-1)!} \varepsilon^{r-1}, \dots, P^{(r)}(\alpha + \varepsilon) \approx P^{(r)}(\alpha)$$

otkuda vidimo da se u podnizu

$$P(\alpha + \varepsilon), P'(\alpha + \varepsilon), \dots, P^{(r)}(\alpha + \varepsilon)$$

predznak ne mijenja. Vidimo dakle, da prolaskom kroz r -struku nultočku od P samo u podnizu $P, P', \dots, P^{(r)}$ niza (18) gubimo r varijacija predznaka. Vrijedi dakle

$$V(\alpha - \varepsilon) - V(\alpha + \varepsilon) \geq r.$$

Promatramo li redom sve nultočke od P unutar $[a, b]$ dobivamo da je

$$N(a, b) \leq V(a) - V(b)$$

pri čemu smo nultočke od P računali s njihovim kratnostima. Iz gornjeg razmatranja vidimo da ako na $[a, b]$ imamo samo nultočke od P , onda u (19) vrijedi jednakost. Nejednakost u (19) može nastupiti kada na $[a, b]$ imamo nultočku nekog dalnjeg člana iz (19) ili nultočku od P koja je ujedno i nultočka neke derivacije višeg reda od kratnosti nultočke od P . Taj dio dokaza izostavljamo, iako je analogan prethodnom.

Vratimo se Descartesovom pravilu. Ako u Budan-Fourierovom teoremu uzmemos $a = 0$ i b dovoljno velik tako da P i sve njegove derivacije budu u b pozitivni (to je moguće jer je za velike argumente ponašanje polinoma određeno najvišom potencijom), onda imamo $V(b) = 0$, pa je $N \leq V(0)$. Kako koeficijenti od P imaju predznake od $P(0), P'(0), \dots, P^{(n)}(0)$, to je $V(0)$ jednako broju varijacija predznaka koeficijenata od P . Time smo dokazali Descartesovo pravilo.

Primjer 5.5. Promatrajmo polinom $P(x) = x^4 - 1$ na $[-2, 2]$ i sastavimo sljedeću tablicu predznaka

	P	P'	P''	P'''	$P^{(4)}$
$x = -2$	+	-	+	-	+
$x = 0$	-	0	0	0	+
$x = 2$	+	+	+	+	+

Vidimo da vrijedi: $V(-2) - V(0) = 3$; $V(0) - V(2) = 1$; $V(-2) - V(2) = 4$. Iz dobivenog je sigurno da P na $[0, 2]$ ima jednu nultočku, da izvan segmenta nema nultočaka, te da na segmentu $[-2, 0]$ ima 1 ili 3 nultočke. Uočimo da je 0 trostruka nultočka od P' .

Sturmov niz. Točan broj različitih nultočaka polinoma P na segmentu $[a, b]$ može se naći pomoću tzv. **Sturmova niza**. Dajemo samo postupak kako se izgrađuje Sturmov niz za dani polinom P . To je niz polinoma

$$P_0, P_1, P_2, \dots, P_m \tag{21}$$

koji dobivamo ovako. Početni član niza P_0 jednak je danom polinomu P , tj. $P_0(x) = P(x)$. Član P_1 jednak je derivaciji P' danog polinoma, tj. vrijedi $P_1(x) = P'(x)$. Polinom P_2 je ostatak dijeljenja P_0 s P_1 uzet s protivnim predznakom, tj. iz

$$P_0(x) = P_1(x)Q_1(x) + R_1(x) \quad (22)$$

nalazimo

$$P_2(x) = -R_1(x).$$

Dalje postupamo kao u traženju polinoma P_2 , tj. polinom P_{i+1} je ostatak dijeljenja polinoma P_{i-1} s P_i uzet s protivnim predznakom. Ispišemo li opisani postupak imamo redom

$$\begin{aligned} P(x) &= P_0(x) = P_1(x)Q_1(x) - P_2(x) \\ P'(x) &= P_1(x) = P_2(x)Q_2(x) - P_3(x) \\ P_2(x) &= P_3(x)Q_3(x) - P_4(x) \\ &\vdots \\ P_{m-2}(x) &= P_{m-1}(x)Q_{m-1}(x) - P_m(x), \end{aligned} \quad (23)$$

štonije ništa drugo nego, do na promjenu predznaka ostatka, Euklidov algoritam traženja najveće zajedničke mjere od P i P' . Dakle, P_i je ili nula ili polinom nižeg stupnja od P_{i-1} . Time u konačno koraka dolazimo do P_m koji je konstanta, pa nastupaju ova dva slučaja:

1) Ako je $P_m \neq 0$, onda P i P' nemaju zajedničkog djelitelja iz čega zaključujemo da P ima samo jednostrukе nultočke.

2) Ako je $P_m = 0$, onda je P_{m-1} najveća zajednička mjera od P i P' , iz čega zaključujemo da P ima višestrukih nultočaka. U tom slučaju polinom $Q = \frac{P}{P_{m-1}}$ ima jednostrukе nultočke koje se podudaraju s nultočkama od P .

Prema tome, slučaj kada imamo višestruke nultočke svodi se na slučaj jednostrukih, pa se zbog toga sljedeći teorem i izriče samo za polinome s jednostrukim nultočkama jer prema 2) umjesto P možemo promatrati $Q = \frac{P}{P_{m-1}}$. Jasno je da se u bilo kojem koraku P_i može pomnožiti s pozitivnim brojem čime možemo pojednostaviti računsko razmatranje. Sljedeći teorem navodimo bez dokaza.

Teorem 5.8. (Sturm) Ako polinom P ima jednostrukе nultočke i $P(a)P(b) \neq 0$, ondaje broj realnih nultočaka $N(a, b)$ na intervalu (a, b) jednak

$$N(a, b) = V(a) - V(b), \quad (24)$$

gdje je $V(a)$, odnosno $V(b)$, broj promjena (varijacija) predznaka u Sturmovom nizu (21) uzetom u točki a , odnosno b .

Promatramo li polinom i z primjera 5.4 imamo

$$P_0(x) = x^4 - 35x^3 + 380x^2 - 1350x + 1000$$

$$P_1(x) = 4x^3 - 105x^2 + 760x - 1350.$$

Dijeljenjem P_0 s P_1 dobivamo ostatak $-\frac{635}{16}x^2 + \frac{10\,400}{16}x - \frac{31\,250}{16}$ tako da množenjem sa $\frac{16}{5}$ možemo za P_2 uzeti

$$P_2(x) = 127x^2 - 2080x + 6250.$$

Dijeljenjem P_1 s P_2 dobivamo ostatak $-\frac{1\,348\,160}{127^2}x + \frac{9\,569\,600}{127^2}$ pa množenjem sa $\frac{127^2}{320}$ dobivamo

$$P_3(x) = 4213x - 29\,905.$$

Daljnjim računom nalazimo

$$P_4 = 1,$$

iz čega odmah zaključujemo da P ima jednostrukе nultočke. Kako su vodeći koeficijenti od P_0, P_1, P_2, P_3 i P_4 pozitivni, to za dovoljno veliki b imamo $V(b) = 0$. Uzmemo li $a = 0$ vidimo odmah da je $V(0) = 4$. Kombiniramo li to sa prethodno dobivenim ogradama za pozitivna rješenja vidimo da P ima četiri nultočke unutar $[0.714, 37]$.

Dakle rješenja jednadžbe

$$x^4 - 35x^3 + 380x^2 - 1350x + 1000 = 0$$

su sva realna i jednostruka i nalaze se unutar segmenta $[0.714, 37]$.

Ako na pitanje određivanja broja rješenja algebarske jednadžbe na $[a, b]$ gledamo čisto teorijski, onda nam Sturmov teorem daje potpun odgovor. Međutim, u praksi on zahtijeva mnogo računanja, pa ga stoga nismo dokazali, već smo dokazali Budan-Fourierov koji je u primjenama korisniji.

Dokažimo još sljedeću posljedicu teorema 5.7 koja govori o postojanju kompleksnih rješenja algebarske jednadžbe.

Korolar 5.9. *Neka je $P(a)P(b) \neq 0$. Ako P ima unutar $[a, b]$ k nultočaka i ako u (19) vrijedi*

$$V(a) - V(b) = k + 2l,$$

onda P ima barem $2l$ kompleksnih nultočaka.

Dokaz. Neka P ima k_1 realnih nultočaka unutar $(-\infty, a]$ i k_2 realnih nultočaka unutar $[b, \infty)$. Po teoremu 5.7 imamo

$$\begin{aligned} V(-\infty) - V(a) &= k_1 + 2l_1, \\ V(b) - V(\infty) &= k_2 + 2l_2, \end{aligned}$$

pa je time

$$k + k_1 + k_2 + 2l + 2l_1 + 2l_2 \leq n.$$

Napišemo li dobivenu nejednakost ovako

$$n - (k + k_1 + k_2) \geq 2l + 2l_1 + 2l_2 \geq 2l$$

vidimo da P ima barem $2l$ kompleksnih nultočaka, jer je $n - (k + k_1 + k_2)$ točan broj kompleksnih nultočaka.

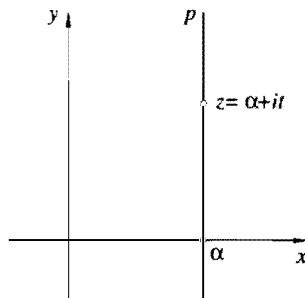
Izoliranje kompleksnih rješenja algebarske jednadžbe

Izoliranje pojedinih realnih rješenja algebarske jednadžbe obično se napada tako da se nađu donja i gornja ograda i zatim se taj segment dijeli na manje podsegmente, redovito jednakе duljine, dok u pojedinim podsegmentima ne izoliramo po jedno rješenje. Možemo se pitati da li unaprijed znamo odrediti duljinu tih podsegmenata. Spomenimo da je poznata formula koja daje ocjenu te duljine, međutim, ta formula uključuje mnogo računanja i osim toga redovito daje nisku vrijednost pa to izostavljamo. Isto tako postoji npr. **Fourierova metoda** za odjeljivanje realnih rješenja algebarske jednadžbe koja se služi Budan-Fourierovim teoremom u prethodnom potparagrafu.

U ovom potparagrafu ćemo se baviti izoliranjem kompleksnih rješenja koje pokriva i izoliranje realnih rješenja. Ono se svodi na iznalaženje pravokutnika u kompleksnoj ravnini unutar kojega se nalazi jedna nultočka polinoma

$$\begin{aligned} P(z) &= a_0 z^n + a_1 z^{n-1} + \dots + a_n \\ &= a_0(z - z_1)(z - z_2) \dots (z - z_n). \end{aligned} \quad (25)$$

Evo jednog postupka koji vodi cilju. Promatrajmo u kompleksnoj ravnini \mathbf{C} vertikalni pravac $\operatorname{Re} z = \alpha$ na kome P nema nultočku.



Sl. 5.7.

α je apscisa u kojoj pravac siječe x -os. Točke pravca možemo parametarski ovako opisati

$$z = \alpha + it, \quad -\infty < t < \infty. \quad (26)$$

Uvrstimo li (26) u P imamo

$$P(\alpha + it) = Q(t) + iR(t) = \sqrt{Q^2(t) + R^2(t)} e^{i\varphi(t)} \quad (27)$$

gdje je

$$\operatorname{tg} \varphi(t) = \frac{R(t)}{Q(t)}. \quad (28)$$

Kako je po pretpostavci $P \neq 0$ u točkama pravca $\operatorname{Re} z = \alpha$, to je za svaki t

$$Q^2(t) + R^2(t) \neq 0. \quad (29)$$

Nadalje, $\{Q^2(t) + R^2(t)\}^{1/2}$ i $\varphi(t)$ su neprekidne funkcije za $t \in (-\infty, \infty)$. Uvedimo sada pojam **indeksa polinoma**, u oznaci $I_P(\operatorname{Re} z = \alpha) = I_P$, s obzirom na vertikalni

pravac $\operatorname{Re} z = \alpha$. I_P definiramo kao promjenu argumenta od P duž pravca $\operatorname{Re} z = \alpha$ podijeljenu s π , tj.

$$I_P = \frac{1}{\pi} \{ \varphi(+\infty) - \varphi(-\infty) \}. \quad (30)$$

Veličine $\varphi(+\infty)$ i $\varphi(-\infty)$ određujemo pomoću (28).

Ako mislimo na "potpunu kompleksnu ravninu" $\overline{\mathbb{C}}$, tj. na kompleksnu ravninu s dodanom "beskonačno dalekom točkom ∞ ", onda (30) predstavlja promjenu argumenta od P , podijeljenu s π , duž konture koju pravac $\operatorname{Re} z = \alpha$ predstavlja u $\overline{\mathbb{C}}$.

Iz eksponencijalnog zapisa u (27) očitavamo odmah sljedeće svojstvo indeksa polinoma s obzirom na pravac.

Ako je $P(z) = P_1(z)P_2(z)$ onda je

$$I_P(\operatorname{Re} z = \alpha) = I_{P_1}(\operatorname{Re} z = \alpha) + I_{P_2}(\operatorname{Re} z = \alpha). \quad (31)$$

Taj ćemo rezultat primijeniti na faktorizaciju

$$P(z) = a_0(z - z_1)(z - z_2) \dots (z - z_n)$$

iz koje vidimo da je dovoljno izračunati indeks konstantnog polinoma (faktor a_0) i indeks linearног polinoma (faktori $z - z_i$) i onda te indekse zbrojiti.

Indeks konstantnog polinoma s obzirom na bilo koji vertikalni pravac je nula. (Za $P_0 = a_0$ je $P_0(\alpha + it) = a_0 e^{i\alpha}$, pa je $I_{P_0} = \frac{1}{\pi}(0 - 0) = 0$.)

Indeks linearног polinoma $P_1(z) = z - z_1$ možemo ovako izračunati. Neka je $z_1 = x_1 + iy_1$. Uvrštanjem imamo

$$P_1(\alpha + it) = (\alpha - x_1) + i(t - y_1)$$

otkuda je

$$\operatorname{tg} \varphi(t) = \frac{t - y_1}{\alpha - x_1}. \quad (32)$$

Sada nastupaju dvije mogućnosti: $\alpha - x_1 > 0$ ili $\alpha - x_1 < 0$.

Ako je $\alpha - x_1 > 0$, onda se (promjenom t od $-\infty$ do $+\infty$) $\operatorname{tg} \varphi(t)$ neprekidno mijenja od $-\infty$ do $+\infty$, pa se $\varphi(t)$ neprekidno mijenja od $-\frac{\pi}{2}$ do $+\frac{\pi}{2}$, tako da u tom slučaju dobivamo

$$I_{P_1}(\operatorname{Re} z = \alpha > x_1) = \frac{1}{\pi} \left\{ \frac{\pi}{2} - \left(-\frac{\pi}{2} \right) \right\} = 1.$$

Uočimo da se u tom slučaju nultočka z_1 nalazi **lijevo** ($x_1 < \alpha$) od pravca $\operatorname{Re} z = \alpha$.

Ako je $\alpha - x_1 < 0$, onda se nultočka z_1 nalazi desno od pravca $\operatorname{Re} z = \alpha$ i $\operatorname{tg} \varphi(t)$ se neprekidno mijenja od $+\infty$ do $-\infty$, pa se $\varphi(t)$ neprekidno mijenja od $\frac{\pi}{2}$ do $-\frac{\pi}{2}$ što uvršteno u (30) daje

$$I_{P_1}(\operatorname{Re} z = \alpha < x_1) = \frac{1}{\pi} \left\{ -\frac{\pi}{2} - \frac{\pi}{2} \right\} = -1.$$

Spojimo li to zajedno izračunali smo

$$I_{P_1}(\operatorname{Re} z = \alpha) = \begin{cases} +1 & \text{ako je } x_1 < \alpha \\ -1 & \text{ako je } x_1 > \alpha. \end{cases} \quad (33)$$

Na taj način $I_{P_1}(\operatorname{Re} z = \alpha)$ kaže da li je z_1 (a to je jedina nultočka polinoma $P_1(z) = z - z_1$) lijevo (rezultat +1) ili desno (rezultat -1) od pravca $\operatorname{Re} z = \alpha$. Primijenimo li (31) na

$$P(z) = a_0(z - z_1)(z - z_2) \dots (z - z_n)$$

dobivamo ovu formulu za I_P :

$$I_P(\operatorname{Re} z = \alpha) = k - l \quad (34)$$

gdje je k broj nultočaka polinoma P lijevo od pravca $\operatorname{Re} z = \alpha$, a l broj nultočaka od P desno od tog istog pravca. Time smo počazali da indeks I_P daje razliku između broja rješenja jednadžbe $P(z) = 0$ lijevo od pravca i broja rješenja desno od pravca na kome nema rješenja. Sada jednakosti (34) možemo priključiti jednakosti koja slijedi iz osnovnog teorema algebre

$$n = k + l \quad (35)$$

tako da iz (34) i (35) nalazimo ove formule

$$k = \frac{n + I_P(\operatorname{Re} z = \alpha)}{2} \quad \text{i} \quad l = \frac{n - I_P(\operatorname{Re} z = \alpha)}{2}. \quad (36)$$

Istaknimo da su u (36) uračunate kratnosti pojedinih rješenja jer smo formulu (31) primjenili na svaki faktor rastava $P(z) = a_0(z - z_1)(z - z_2) \dots (z - z_n)$. Uzmemo li sada dva vertikalna pravca $\operatorname{Re} z = \alpha_1$ i $\operatorname{Re} z = \alpha_2$ i neka je $\alpha_1 < \alpha_2$, onda pomoću (36) možemo naći broj korijena jednadžbe $P(z) = 0$, uračunavši njihove kratnosti unutar vertikalne trake $\alpha_1 < \operatorname{Re} z < \alpha_2$. Da dođemo do pravokutnika moramo provesti analogno razmatranje za horizontalan pravac.

Promatrajmo sada horizontalan pravac na kojem polinom P nema nultočaka. Točke takvog pravca su parametarski dane sa

$$z = t + i\beta. \quad (37)$$

Analogno imamo

$$P(t + i\beta) = Q_1(t) + iR_1(t) = \sqrt{Q_1^2(t) + R_1^2(t)} e^{i\psi(t)} \quad (38)$$

i

$$\operatorname{tg} \psi(t) = \frac{R_1(t)}{Q_1(t)}. \quad (39)$$

Indeks polinoma P s obzirom na horizontalni pravac definiramo kao

$$I_P(\operatorname{Im} z = \beta) = \frac{1}{\pi} \{\psi(+\infty) - \psi(-\infty)\}, \quad (40)$$

tako da razmatranje slično gornjem za vertikalni pravac daje

$$I_P(\operatorname{Im} z = \beta) = p - q \quad (41)$$

gdje p označava broj nultočaka od P iznad horizontalnog pravca, a q broj nultočaka od P ispod horizontalnog pravca uzetih s njihovim kratnostima. Kako je $p + q = n$ to slijede pripadne formule za p i q :

$$p = \frac{n + I_P(\operatorname{Im} z = \beta)}{2}, \quad q = \frac{n - I_P(\operatorname{Im} z = \beta)}{2}. \quad (42)$$

Nađemo li dakle vertikalni $\operatorname{Re} z = \alpha$ i horizontalni $\operatorname{Re} z = \beta$ na kojima P nema nultočaka onda oni dijele kompleksnu ravninu na 4 kvadranta. Iz pripadnih dvaju indeksa možemo izračunati broj nultočaka u svakom od tih kvadrantata. Ako je n_i broj nultočaka u i -tom kvadrantu, $i = 1, 2, 3, 4$, onda treba riješiti ovaj sustav linearnih jednadžbi

$$\begin{aligned} n_1 + n_2 &= p \\ n_1 &+ n_4 = q \\ n_2 + n_3 &= k \\ n_3 + n_4 &= l. \end{aligned} \tag{43}$$

Ako znamo granice područja koje sadrži sva rješenja jednadžbe $P(z) = 0$, onda postupnom primjenom gornje metode možemo izolirati realna i kompleksna rješenja algebarske jednadžbe. Da to provedemo treba izračunati odgovarajuće indekse što nije uvi je jednostavno.

Računanje indeksa polinoma s obzirom na pravac

Izvedimo jedan postupak računanja indeksa. Zadržimo se samo na vertikalnom pravcu $\operatorname{Re} z = \alpha$, s parametrizacijom

$$z = \alpha + it \tag{44}$$

$$P(\alpha + it) = Q(t) + iR(t) \tag{45}$$

$$\operatorname{tg} \varphi(t) = \frac{R(t)}{Q(t)}. \tag{46}$$

Kako su Q i R polinomi, to je $\operatorname{tg} \varphi(t)$ racionalna funkcija. Neka je n stupanj od P . Ako je n paran onda i Q ima stupanj n dok R ima stupanj $\leq n - 1$. Ako je n neparan, onda R ima stupanj n , a Q stupanj $\leq n - 1$. Kako P po pretpostavci nema na pravcu nultočaka, to Q i R nemaju na pravcu zajedničkih nultočaka. Funkcija $\varphi(t) = \arg P(\alpha + it)$ je neprekidna pri promjeni t od $-\infty$ do $+\infty$, međutim, $\operatorname{tg} \varphi(t) = R(t)/Q(t)$ ima diskontinuitete u realnim nultočkama t_i od $Q(t)$. Diskontinuiteti od $R(t)/Q(t)$ mogu imati jedan od sljedećih oblika:

- 1) $\lim_{t \rightarrow t_i \mp 0} \frac{R(t)}{Q(t)} = +\infty;$
- 2) $\lim_{t \rightarrow t_i \mp 0} \frac{R(t)}{Q(t)} = -\infty;$
- 3) $\lim_{t \rightarrow t_i \mp 0} \frac{R(t)}{Q(t)} = \pm\infty;$
- 4) $\lim_{t \rightarrow t_i \mp 0} \frac{R(t)}{Q(t)} = \mp\infty.$

Da pojasnimo oznake, pogledajmo slučaj 3), koji znači sljedeće: ako se točki t_i približavamo slijeva, onda $R(t)/Q(t) \rightarrow +\infty$; a ako joj se približavamo zdesna, onda $R(t)/Q(t) \rightarrow -\infty$. Funkcija tangens ima diskontinuitete u točkama $\pm \frac{\pi}{2}, \pm \frac{3\pi}{2}, \pm \frac{5\pi}{2}, \dots$ i svi su diskontinuiteti tipa 3).

Dakle, da bi našli prirast od $\varphi(t)$ pri promjeni t od $-\infty$ do $+\infty$ moramo naći sve realne nultočke polinoma Q : $-\infty < t_1 < t_2 < \dots < t_k < \infty$ i promatrati pripadne intervale

$$(-\infty, t_1), (t_1, t_2), (t_2, t_3), \dots, (t_{k-1}, t_k), (t_k, +\infty). \quad (47)$$

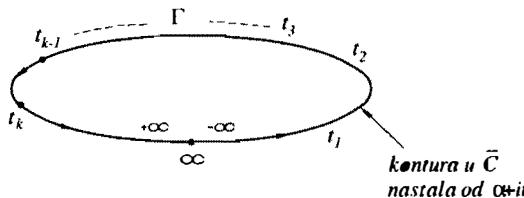
Na svakom od intervala (t_i, t_{i+1}) funkcija $\operatorname{tg} \varphi(t)$ je neprekidna, a vrijednosti argumenta $\varphi(t)$ se nalaze u intervalu

$$\left(\frac{2k_i - 1}{2}\pi, \frac{2k_i + 1}{2}\pi \right) \text{ ili } \left(-\frac{2k_i + 1}{2}\pi, -\frac{2k_i - 1}{2}\pi \right) \quad (48)$$

gdje je k_i nenegativan cijeli broj. Promatrajmo stoga prolaz parametra t kroz t_{i+1} .

Ovisno o tipu singulariteta od $\frac{R}{Q}$ u $t = t_{i+1}$, vrijednost od $\varphi(t)$ će ili ostati u intervalu

(48) ili će iz njega izići. Ako se radi o singularitetu tipa 1) ili 2), onda vrijednost od $\varphi(t)$ ostaje u tom intervalu. Ako se radi o singularitetu tipa 3), onda vrijednost od $\varphi(t)$ izlazi iz intervala (48) i raste, dok u slučaju singulariteta tipa 4) izlazi iz (48) i pada. Da to zorno predočimo mislimo na graf funkcije $\operatorname{arc} \operatorname{tg}$ i njegove grane. Istaknimo da se prelazom s jedne grane na drugu $\varphi(t)$ mijenja za π (što je period funkcije tg). Vidimo da prolazom kroz singularne točke t_1, \dots, t_k , vrijednost $\varphi(t)$ poraste za π ili padne za π ovisno o tome da li smo prošli kroz singularnu točku t_i tipa 3) ili tipa 4) pa je promjena jednak razlici između brojeva takvih točaka puta π . Ostali su još krajevi $\mp\infty$. Promatramo li konturu Γ u $\bar{\mathbb{C}}$ u koju prelazi pravac (44), onda je ∞ jedna singularna točka na toj konturi i ima neki od gornjih tipova 1) – 4).



Sl. 5.8.

Na slici smo nacrtali jednu takvu konturu s točkom ∞ . Na desno od te točke nalaze se točke intervala $(-\infty, t_1)$, a na lijevo točke intervala $(t_k, +\infty)$. Prema tome $\lim_{t \rightarrow -\infty} \varphi(t) = \lim_{t \rightarrow -\infty+0} \varphi(t)$, tj. limes u $-\infty$ je desni limes u točki $\infty \in \Gamma$ dok je $\lim_{t \rightarrow +\infty} \varphi(t) = \lim_{t \rightarrow \infty-0} \varphi(t)$ lijevi limes u točki $\infty \in \Gamma$. Prolazom kroz točku ∞ imamo istu pojavu kao i kroz točku t_{i+1} , tj. ako $\frac{R}{Q}$ ima u ∞ singularitet tipa 3) onda $\varphi(t)$ poraste za π , dok se smanji za π ako je u ∞ singularitet tipa 4). Tako smo došli do sljedećeg pravila za izračunavanje indeksa $I_P(\operatorname{Re} z = \alpha)$.

Pravilo za izračunavanje indeksa. Indeks I_P polinoma P , s obzirom na pravac $\operatorname{Re} z = \alpha$ (ili $\operatorname{Im} z = \beta$), na kome P nema nultočaka, jednak je razlici broja singulariteta tipa 3) i broja singulariteta tipa 4) od $\frac{R}{Q}$ (odnosno $\frac{R_1}{Q_1}$) u što treba ubrojati i singularitet u točki ∞ , ako je isti tipa 3) ili tipa 4).

Primjer 5.6. Promatrajmo jednadžbu

$$x^5 - 4x^4 + 6x^3 - 3x^2 + 2x + 1 = 0$$

za koju lako vidimo da ima samo jedno realno rješenje i to u intervalu $(-0.5; 0)$. Izolirat ćemo grubo kompleksna rješenja i vidjeti da se u svakom kvadrantu nalazi po jedno. Sva se rješenja nalaze unutar prstena

$$\frac{1}{7} < |z_i| < 7$$

što lako dobivamo primjenom teorema 5.3 i korolara 5.4. Pogotovo su sva rješenja unutar kvadrata $-7 \leq x \leq 7$, $-7 \leq y \leq 7$. Stoga razmotrimo imaginarnu os tj. pravac

$$z = o + it$$

i ustanovimo koliko rješenja imamo u lijevoj i desnoj poluravnini. Kako se rješenja pojavljuju u parovima konjugirano kompleksnih brojeva to su u kompleksnoj ravni simetrično raspoređena s obzirom na x -os. Izračunajmo

$$P(t) = (-4t^4 + 3t^2 + 1) + (t^5 - 6t^3 + 2t)i$$

dakle

$$R(t) = t(t^4 - 6t^2 + 2)$$

$$Q(t) = -4t^4 + 3t^2 + 1$$

i

$$\operatorname{tg} \varphi(t) = \frac{R(t)}{Q(t)} = -\frac{t}{4} \frac{t^4 - 6t^2 + 2}{t^4 - \frac{3}{4}t^2 - \frac{1}{4}}.$$

Lako nalazimo da su realna rješenja jednadžbe $Q(t) = 0$ (bikvadratna jednadžba)

$$t_1 = -1, \quad t_2 = 1.$$

Iz zapisa

$$\frac{R(t)}{Q(t)} = -\frac{t}{4} \frac{t^4 - 6t^2 + 2}{t^4 - \frac{3}{4}t^2 - \frac{1}{4}}$$

lako ustanovimo da u točkama $t_1 = -1$ i $t_2 = 1$ funkcija $\frac{R}{Q}$ ima singularitet tipa 3), dok u točki ∞ ima singularitet tipa 4) jer je

$$\lim_{t \rightarrow +\infty} \frac{R(t)}{Q(t)} = -\infty \quad \text{i} \quad \lim_{t \rightarrow -\infty} \frac{R(t)}{Q(t)} = +\infty.$$

Dakle, indeks od P , s obzirom na y -os je jednak

$$I_P = +1$$

otkuda imamo

$$k - l = 1$$

$$k + l = 5$$

tako da dobivamo $k = 3$ i $l = 2$. Znači na lijevo od y -osi imamo 3 rješenja (jedno je realno), a na desno dva rješenja. Nadalje je jasno da se u svakom kvadrantu nalazi

po jedno kompleksno rješenje i to unutar gore spomenutog prstena, odnosno kvadrata. Želimo li bolje izolirati rješenja moramo gledati indekse s obzirom na pravce koji će promatrani kvadrat subdivirati.

5.3. Sustavi nelinearnih jednadžbi

U ovom poglavlju prelazimo odmah na iteracijske metode za razliku od sustava linearnih jednadžbi kada smo prvo obradili direktne metode. Razlog je jednostavan, nemamo direktne metode za opći sustav nelinearnih jednadžbi. To ne znači da pojedine tipove sustava ne znamo direktno rješiti. Prisjetimo se Choleskijeve sheme koja predstavlja direktnu metodu nalaženja elemenata matrice kod LU faktorizacije. Kako je lakše pratiti metode na sustavima dviju jednadžbi, to ćemo prvo razviti Newtonovu metodu za takve sustave nakon čega je lakše sve pratiti na općim sustavima. S druge strane sustavi dviju jednadžbi su od interesa i kod traženja kompleksnih korijena algebarskih jednadžbi, a i inače.

Newtonova metoda za sustav dviju jednadžbi

Neka je $x = \xi$ i $y = \eta$ rješenje sustava dviju jednadžbi

$$\begin{aligned} f(x, y) &= 0 \\ g(x, y) &= 0. \end{aligned} \tag{1}$$

Prepostavimo da f i g imaju neprekidne parcijalne derivacije. Stavljujući

$$\xi = x_n + h_n, \quad \eta = y_n + k_n$$

dobivamo

$$\begin{aligned} f(x_n + h_n, y_n + k_n) &= 0 \\ g(x_n + h_n, y_n + k_n) &= 0. \end{aligned} \tag{2}$$

Primjenom Taylorove formule (za funkcije dviju varijabli), pri čemu se ograničavamo samo na linearne članove i iz jednačavanjem istih s nulom dobivamo

$$\begin{aligned} f(x_n, y_n) + h_n f'_x(x_n, y_n) + k_n f'_y(x_n, y_n) &= 0 \\ g(x_n, y_n) + h_n g'_x(x_n, y_n) + k_n g'_y(x_n, y_n) &= 0, \end{aligned} \tag{3}$$

što je linearan sustav jednadžbi s nepoznanicama h_n i k_n . Ako je Jakobijan

$$J = \begin{vmatrix} f'_x(x_n, y_n) & f'_y(x_n, y_n) \\ g'_x(x_n, y_n) & g'_y(x_n, y_n) \end{vmatrix} \neq 0,$$

onda linearni sustav (3) ima jedinstveno rješenje koje glasi

$$\begin{aligned} h_n &= -\frac{1}{J} \begin{vmatrix} f(x_n, y_n) & f'_y(x_n, y_n) \\ g(x_n, y_n) & g'_y(x_n, y_n) \end{vmatrix} \\ k_n &= -\frac{1}{J} \begin{vmatrix} f'_x(x_n, y_n) & f(x_n, y_n) \\ g'_x(x_n, y_n) & g(x_n, y_n) \end{vmatrix}. \end{aligned} \tag{4}$$

tako da za sljedeću aproksimaciju od ξ i η možemo uzeti

$$\begin{aligned}x_{n+1} &= x_n + h_n \\y_{n+1} &= y_n + k_n\end{aligned}\tag{5}$$

gdje su h_n i k_n izračunati prema (4). Početnu aproksimaciju (x_0, y_0) treba odrediti grubo promatranjem sustava (1).

Ako je $\xi = \lim x_n$, $\eta = \lim y_n$ i neka su sve točke (x_n, y_n) unutar pravokutnika oko (ξ, η) na kojemu su ispunjene pretpostavke o neprekidnosti, onda je (ξ, η) rješenje sustava (1). Naime, iz $x_n \rightarrow \xi$ i $y_n \rightarrow \eta$ i (5) slijedi da $h_n \rightarrow 0$ i $k_n \rightarrow 0$. Iz pretpostavke o neprekidnosti parcijalnih derivacija imamo da su iste ograničene na pravokutniku, pa iz (3) slijedi

$$\begin{aligned}\lim_{n \rightarrow \infty} f(x_n, y_n) &= f(\xi, \eta) = 0 \\ \lim_{n \rightarrow \infty} g(x_n, y_n) &= g(\xi, \eta) = 0,\end{aligned}$$

tj. (ξ, η) je rješenje sustava (1). Dakle, konvergentan postupak vodi ka rješenju sustava.

Primjer 5.7. Promatrajmo sustav jednadžbi

$$\begin{aligned}2x &= \sin \frac{1}{2}(x - y) \\2y &= \cos \frac{1}{2}(x + y).\end{aligned}$$

Sada možemo pisati

$$\begin{aligned}f(x, y) &= 2x - \sin \frac{1}{2}(x - y) \\g(x, y) &= 2y - \cos \frac{1}{2}(x + y),\end{aligned}$$

tako da imamo

$$\begin{aligned}f'_x(x, y) &\doteq 2 - \frac{1}{2} \cos \frac{1}{2}(x - y), \\f'_y(x, y) &= \frac{1}{2} \cos \frac{1}{2}(x - y), \\g'_x(x, y) &= \frac{1}{2} \sin \frac{1}{2}(x + y), \\g'_y(x, y) &= 2 + \frac{1}{2} \sin \frac{1}{2}(x + y).\end{aligned}$$

Postupak ćemo voditi tako da ćemo aproksimativne vrijednosti uvrštavati u (3) i računati redom (h_n, k_n) . Podjemo li od početne vrijednosti $(x_0, y_0) = (-0.16, 0.49)$ dobivamo sustav linearnih jednadžbi

$$1.5262h_0 + 0.4738k_0 = 0.0007$$

$$0.0821h_0 + 2.0821k_0 = 0.0064$$

otkuda slijedi $(h_0, k_0) = (-0.0005, 0.0031)$. Sljedeća aproksimacija iznosi $(x_1, y_1) = (-0.1605, 0.4931)$. Želimo li poboljšati aproksimaciju moramo točnije računati trigonometrijske funkcije, pa zatim odrediti točnije vrijednosti od h_1 i k_1 . Tako uz

$$\begin{aligned}\sin 0.3268 &= 0.321014, & \cos 0.3268 &= 0.947074 \\ \sin 0.1663 &= 0.165535, & \cos 0.1663 &= 0.986204\end{aligned}$$

možemo izračunati h_1 , k_1 na 6 decimalnih mesta što da je

$$(x_2, y_2) = (-0.160510, 0.493102).$$

Newtonova metoda za kompleksna rješenja (nultočke)

Promatrajmo jednadžbu

$$f(z) = 0 \quad (6)$$

pri čemu neka je f analitička funkcija na krugu U oko jednostrukih nultočaka $\zeta = \xi + i\eta$, tj. vrijedi $f(\zeta) = 0$ i $f'(\zeta) \neq 0$. Neka je $z_n \in U$ aproksimacija od ζ . Da poboljšamo aproksimaciju postupimo analogno kao u realnom slučaju. Pišemo li

$$\zeta = z_n + \Delta z_n$$

imamo prema Taylorovoj formuli za funkcije kompleksne varijable

$$0 = f(\zeta) \approx f(z_n) + \Delta z_n f'(z_n),$$

pri čemu smo članove drugog i višeg stupnja zanemarili. Izjednačavanjem desne strane s nulom dobivamo

$$\Delta z_n = -\frac{f(z_n)}{f'(z_n)}.$$

Na taj način smo dobili **Newtonov iteracijski postupak** za kompleksnu varijablu

$$z_{n+1} = z_n - \frac{f(z_n)}{f'(z_n)}, \quad n = 0, 1, 2, \dots \quad (7)$$

Odmah je vidljivo što takav postupak daje ako niz $\{z_n\}$ konvergira. Prepostavimo li $z_n \in U$, ($n = 1, 2, \dots$) i $\zeta = \lim_{n \rightarrow \infty} z_n$, imamo zbog neprekidnosti od f i f'

$$\lim_{n \rightarrow \infty} z_{n+1} = \lim_{n \rightarrow \infty} z_n - \frac{\lim_{n \rightarrow \infty} f(z_n)}{\lim_{n \rightarrow \infty} f'(z_n)}$$

što daje

$$\zeta = \zeta - \frac{f(\zeta)}{f'(\zeta)}$$

odnosno

$$f(\zeta) = 0,$$

tj. ζ je rješenje jednadžbe (6). Odredimo li u (7) realni i imaginarni dio od f , tj. napišemo $f(z) = u(x, y) + iv(x, y)$, onda (7) poprima oblik

$$\begin{aligned}x_{n+1} &= x_n + \left[\frac{v \frac{\partial u}{\partial y} - u \frac{\partial u}{\partial x}}{\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2} \right]_{\substack{x=x_n \\ y=y_n}} \\y_{n+1} &= y_n + \left[\frac{-v \frac{\partial u}{\partial y} - u \frac{\partial v}{\partial y}}{\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial v}{\partial y} \right)^2} \right]_{\substack{x=x_n \\ y=y_n}}\end{aligned}\quad (8)$$

gdje je u derivaciji $f'(z) = \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x}$ imaginarni dio zamijenjen s $-\frac{\partial u}{\partial y}$ prema Cauchy-Riemannovim jednakostima.

U posebnom slučaju ako je (7) algebarska jednadžba

$$a_0 z^n + a_1 z^{n-1} + \dots + a_{n-1} z + a_n = 0$$

s realnim koeficijentima, onda nam opisana metoda služi za aproksimaciju kompleksnih rješenja koja se javljaju u parovima konjugiranih vrijednosti. Prema tome aproksimiramo li $\xi + i\eta$ ujedno smo aproksimirali i $\xi - i\eta$. U tom slučaju su nam u i v polinomi u dvije varijable.

Ocijenimo pogrešku aproksimacije. Promatrajmo $w = f(z)$ i prepostavimo

$$|f'(z)| \geq m > 0 \quad \text{za } z \in U. \quad (9)$$

Tada je f na nekom krugu, radijusa R , oko ζ bijekcija (ta se činjenica dokazuje u teoriji funkcija kompleksne varijable), tako da je na nekoj okolini ishodišta $|w| < \rho$ definirana inverzna

$$z = f^{-1}(w)$$

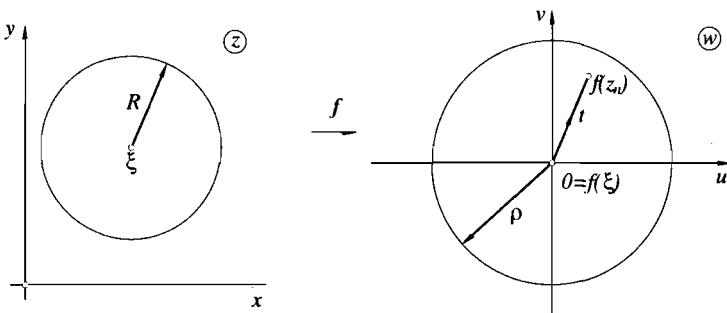
koja je analitička i vrijedi

$$z' = \frac{1}{f'(z)} = \frac{1}{f'(f^{-1}(w))}. \quad (10)$$

Prepostavimo li da je $|f(z_n)| < \rho$ imamo

$$\begin{aligned}z_n - \zeta &= f^{-1}(f(z_n)) - f^{-1}(f(\zeta)) \\&= \int_{f(\zeta)}^{f(z_n)} \frac{d}{dt} [f^{-1}(t)] dt \\&= \int_0^{f(z_n)} \frac{dt}{f' [f^{-1}(t)]}\end{aligned}\quad (11)$$

gdje je za put integracije uzeta spojnica točaka $f(\zeta) = 0$ i $f(z_n)$, v. sl.



Sl. 5.9.

Zbog $|t| < \rho$ vrijedi $|f^{-1}(t)| < R$ i prema (9)

$$|f'(f^{-1}(t))| \geq m.$$

Sada iz (11) dobivamo sljedeću ocjenu greške

$$|z_n - \zeta| \leq \int_0^{f(z_n)} \frac{|dt|}{|f'(f^{-1}(t))|} \leq \int_0^{f(z_n)} \frac{|dt|}{m} = \frac{|f(z_n)|}{m}, \quad (12)$$

što je analogon formule (6) iz uvodnog dijela ovog poglavlja.

Spomenimo da postoje teoremi koji daju dovoljne uvjete pod kojima je Newtonov postupak za kompleksnu varijablu konvergentan, no u to se ovdje nećemo upuštati.

Primjer 5.8. Promatrajmo jednadžbu

$$e^z - 0.2z + 1 = 0$$

i odredimo aproksimativnu vrijednost rješenja koje ima najmanji modul.

Istaknimo da promatrana jednadžba nema realnih rješenja. Naime, ako uvrstimo $z = x$ imamo $e^x - 0.2x + 1 = 0$, iz čega dobivamo jednakost $e^x + 1 = 0.2x$, koja očigledno nije ispunjena ni za jedno x . (Također se lako provjeri da nema imaginarnih rješenja.)

Promatramo dakle

$$f(z) = e^z - 0.2z + 1, \quad f'(z) = e^z - 0.2.$$

Izjednačimo li $f'(z) = 0$ dobivamo $e^z = 0.2$, odnosno, $z_k = \ln 0.2 + 2k\pi i \approx -1.79 + 2k\pi i$, $k = 0, \pm 1, \pm 2, \dots$. Kako te točke nisu nultočke od f to su sva rješenja gornje jednadžbe jednostruka. Da lociramo rješenja poslužimo se ovim uspoređivanjem. Promatramo li $\varphi(z) = e^z + 1$ (dakle izostavimo član $-0.2z$) vidimo da je $f(z) - \varphi(z) = 0.2z$, tj. f i φ se malo razlikuju u okolini ishodišta. Nultočke od φ su $\zeta_k = (2k+1)\pi i$, $k = 0, \pm 1, \pm 2, \dots$ pri tome je $\zeta_0 = \pi i$ najbliža ishodištu u čijoj okolini možemo očekivati nultočku od f . To nam daje razlog da za početnu

aproksimaciju traženog rješenja uzmememo $z_0 = \zeta_0 = \pi i$. Tada je

$$\begin{aligned} z_1 &= \pi i - \frac{0.2i}{1.2} = \pi i - \frac{1}{6}\pi i = \frac{5}{6}\pi i \\ z_2 &= \frac{5\pi i}{6} - \frac{0.132 - 0.024i}{-1.868 + 0.5i} = 0.069 + 2.624i \\ z_3 &= 0.069 + 2.624i - \frac{-0.061 + 0.009i}{-1.230 + 0.541i} = 0.109 + 2.646i \\ z_4 &= 0.109 + 2.646i - \frac{0 + 0.006i}{-1.178 + 0.535i} = 0.107 + 2.650i \\ z_5 &= 0.107 + 2.650i - \frac{-0.002 - 0.005i}{-1.181 + 0.525i} = 0.107 + 2.646i. \end{aligned}$$

Izračunamo li $f(z_5) = 0.002 + 0.004i$ vidimo da je

$$|f(z_5)| = \sqrt{4 \cdot 10^{-6} + 16 \cdot 10^{-6}} = \sqrt{20 \cdot 10^{-6}} \approx 0.0044$$

pa očekujemo da je z_5 već dostatno dobra aproksimacija traženog rješenja. Pokušajmo ocijeniti grešku. Promatrani niz aproksimacija nalazi se unutar pravokutnika $0 \leq x \leq 0.2$, $\frac{\pi}{2} \leq y \leq \pi$. Izračunamo li $|f'(z)|$ imamo zbog $e^{x+iy} = e^x(\cos y + i \sin y)$

$$\begin{aligned} |f'(z)|^2 &= e^{2x} \cos^2 y - 0.4e^x \cos y + 0.04 + e^{2x} \sin^2 y \\ &= e^{2x} - 0.4e^x \cos y + 0.04 \\ &= e^x(e^x - 0.4 \cos y) + 0.04 \\ &> e^x(e^x - 0.4 \cos y) \end{aligned}$$

iz čega čitamo da na promatranom pravokutniku $|f'(z)|$ ima veću vrijednost od vrijednosti u levom donjem uglu, tj. u točki $\frac{\pi}{2}i$ gdje imamo

$$\left| f'\left(\frac{\pi}{2}i\right) \right| = \sqrt{1 + 0.04} \approx 1.02 = m.$$

Prema (12) imamo

$$|z_5 - \zeta| \leq \frac{|f(z_5)|}{m} \approx \frac{0.0044}{1.02} \approx 0.0043$$

iz čega zaključujemo da se z_5 nalazi unutar kruga oko ζ radijusa < 0.043 . U svakom slučaju i realni i imaginarni dio od ζ znamo točno na tri decimalna mjesta.

Lako provjeravamo da vrijedi $f(\bar{z}) = \overline{f(z)}$, pa je prema tome i konjugirano kompleksna vrijednost $\bar{\zeta}$ rješenje promatrane jednadžbe, čija aproksimacija je onda $\bar{z}_5 = 0.107 - 2.646i$.

Newtonova metoda za opći sustav nelinearnih jednadžbi

Promatrajmo sada opći sustav nelinearnih jednadžbi. Polazimo od n jednadžbi s

n nepoznаница

$$\begin{aligned} f_1(x_1, x_2, \dots, x_n) &= 0 \\ f_2(x_1, x_2, \dots, x_n) &= 0 \\ &\vdots \\ f_n(x_1, x_2, \dots, x_n) &= 0 \end{aligned} \tag{13}$$

gdje su f_i realne funkcije od n varijabli koje imaju neprekidne parcijalne derivecije. Kao i uvijek promatramo izolirana rješenja sustava (13). To znači sljedeće: $(\xi_1, \xi_2, \dots, \xi_n)$ je izolirano rješenje ako ima okolinu u kojoj nema drugih rješenja. Okolina od $(\xi_1, \xi_2, \dots, \xi_n)$ može biti otvorena kugla u \mathbf{R}^n , tj. skup točaka $(x_1, x_2, \dots, x_n) \in \mathbf{R}^n$ za koje vrijedi

$$(x_1 - \xi_1)^2 + (x_2 - \xi_2)^2 + \dots + (x_n - \xi_n)^2 < r^2. \tag{14}$$

Sustav (13) možemo vektorski kraće zapisati što nam je često pregledniji zapis. Dakle, varijable x_1, x_2, \dots, x_n i funkcije f_1, f_2, \dots, f_n shvatimo kao komponente n -dimenzionalnih vektora

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{i} \quad f = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix}$$

tako da uz oznaku $f_i(x_1, x_2, \dots, x_n) = f_i(x)$ imamo ovaj vektorski zapis

$$f(x) = 0. \tag{15}$$

Do same metode dolazimo na isti način kao u slučaju sustava dviju jednadžbi. Pođemo li od neke aproksimacije

$$x^{(k)} = \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ \vdots \\ x_n^{(k)} \end{bmatrix}$$

i rješenja

$$\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix},$$

onda je

$$\xi = x^{(k)} + \varepsilon^{(k)}, \tag{16}$$

gdje je

$$\varepsilon^{(k)} = \begin{bmatrix} \varepsilon_1^{(k)} \\ \varepsilon_2^{(k)} \\ \vdots \\ \varepsilon_n^{(k)} \end{bmatrix}$$

prirast do točnog rješenja kojemu ne znamo komponente, ali znamo da vrijedi

$$0 = f(\xi) = f(x^{(k)} + \varepsilon^{(k)}). \quad (17)$$

I sada primijenimo Taylorovu formulu uvezši u račun samo prva dva člana i to iz jednacimo s nulom (za dovoljno malenu okolinu to je približno nula). U vektorskem zapisu Taylorova formula s prva dva člana izgleda ovako:

$$f(x^{(k)} + \varepsilon^{(k)}) = f(x^{(k)}) + f'(x^{(k)})\varepsilon^{(k)}, \quad (18)$$

gdje je $f'(x^{(k)})$ Jacobijeva matrica i zapis $f'(x^{(k)})\varepsilon^{(k)}$ je množenje matrice i vektora stupca $\varepsilon^{(k)}$. Dakle u drugoj oznaci i u razvijenom obliku je

$$f'(x^{(k)}) = J(x^{(k)}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}_{x=x^{(k)}}$$

što znači da elemente matrice treba izračunati za $x = x^{(k)}$ i zatim izvršiti naznačeno množenje u (18). Uvrstimo li taj rezultat u (17) dobivamo

$$f(x^{(k)}) + J(x^{(k)})\varepsilon^{(k)} = 0 \quad (19)$$

što je linearan sustav jednadžbi s nepozanicama $\varepsilon_1^{(k)}, \varepsilon_2^{(k)}, \dots, \varepsilon_n^{(k)}$. U razvijenom obliku sustav (19) izgleda ovako:

$$\begin{aligned} f_1(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}) + \left(\frac{\partial f_1}{\partial x_1}\right)_{x^{(k)}} \varepsilon_1 + \left(\frac{\partial f_1}{\partial x_2}\right)_{x^{(k)}} \varepsilon_2 + \dots + \left(\frac{\partial f_1}{\partial x_n}\right)_{x^{(k)}} \varepsilon_n &= 0 \\ f_2(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}) + \left(\frac{\partial f_2}{\partial x_1}\right)_{x^{(k)}} \varepsilon_1 + \left(\frac{\partial f_2}{\partial x_2}\right)_{x^{(k)}} \varepsilon_2 + \dots + \left(\frac{\partial f_2}{\partial x_n}\right)_{x^{(k)}} \varepsilon_n &= 0 \\ &\vdots \\ f_n(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}) + \left(\frac{\partial f_n}{\partial x_1}\right)_{x^{(k)}} \varepsilon_1 + \left(\frac{\partial f_n}{\partial x_2}\right)_{x^{(k)}} \varepsilon_2 + \dots + \left(\frac{\partial f_n}{\partial x_n}\right)_{x^{(k)}} \varepsilon_n &= 0 \end{aligned} \quad (20)$$

koji kada riješimo u $\varepsilon_1^{(k)}, \varepsilon_2^{(k)}, \dots, \varepsilon_n^{(k)}$ daje približnu vrijednost prirasta tako da se nadamo da je

$$x^{(k+1)} = x^{(k)} + \varepsilon^{(k)} \quad (21)$$

nova i to bolja aproksimacija rješenja $(\xi_1, \xi_2, \dots, \xi_n)$. Rješenje sustava (19) u matričnom zapisu izgleda ovako

$$\varepsilon^{(k)} = -J^{-1}(x^{(k)})f(x^{(k)}) \quad (22)$$

što uvršteno u (21) daje ovaj krajnji matrični zapis iteracijskog koraka **Newtonove metode (postupka)**

$$x^{(k+1)} = x^{(k)} - J^{-1}(x^{(k)})f(x^{(k)}), \quad k = 0, 1, 2, \dots \quad (23)$$

gdje naravno početnu aproksimaciju $x^{(0)}$ moramo odabrat. Jasno je sljedeće. Ako je niz aproksimacija konvergentan, tj.

$$\xi = \lim_{k \rightarrow \infty} x^{(k)},$$

onda je pod gornjim pretpostavkama na f_i , taj limes rješenje polaznog sustava (13). Zaključivanje je analogno onom za slučaj dviju jednadžbi.

Ako pokušamo primijeniti Newtonovu metodu na sustav linearnih jednadžbi, onda ne dobivamo ništa novo. Naime, za

$$f(x) = Ax - b$$

imamo $f'(x) = A$, pa (23) prelazi u

$$x^{(k+1)} = x^{(k)} - A^{-1}(Ax^{(k)} - b) = x^{(k)} - x^{(k)} + A^{-1}b = A^{-1}b,$$

što je direktno rješenje sustava već u prvom koraku. To direktno rješenje je dobiveno izračunavanjem inverzne matrice. Međutim, formula (23) to traži u svakom koraku ako sustav nije linearan, što je osnovna poteškoća tog iteracijskog postupka.

U svezi s Newtonovom iteracijskom metodom postavlja se više pitanja.

- a) Pod kojim uvjetima konvergira?
- b) Koja je brzina konvergencije?
- c) Jedinstvenost rješenja sustava?
- d) Stabilnost postupka obzirom na promjenu početne aproksimacije?

Za takva razmatranja potrebno je opširnije proučavati funkcije iz \mathbf{R}^n u \mathbf{R}^n

$$f : \mathbf{R}^n \rightarrow \mathbf{R}^n,$$

tj. između euklidskih prostora. Takva funkcija je dana s n realnih funkcija od n varijabli. Za tu svrhu je vektorski zapis kraći. Ponovimo oznake. Neka je

$$f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{bmatrix}.$$

Ako sve funkcije (komponente) f_i imaju neprekidne parcijalne derivacije, onda se pod derivacijom $f'(x)$ razumijeva Jacobijeva matrica

$$f'(x) = \left[\frac{\partial f_i(x)}{\partial x_j} \right].$$

Unatoč sve složenosti na gornja pitanja se mogu dati odgovori. Izreći ćemo jedan teorem koji odgovara na a), b), c) i d), ali ga nećemo dokazivati. Teorem je formuliran za rednu normu koju smo upoznali u §3.2 i bila je za matrice definirano ovako

$$||A|| = ||A||_r = \max_i \sum_j |a_{ij}|,$$

a za vektore

$$||x|| = ||x||_r = \max_i |x_i|.$$

Teorem 5.10. *Neka je*

$$f(x) = 0,$$

realan nelinearan sustav jednadžbi pri čemu je

$$f(x) = \begin{bmatrix} f_1(x_1, x_2, \dots, x_n) \\ f_2(x_1, x_2, \dots, x_n) \\ \vdots \\ f_n(x_1, x_2, \dots, x_n) \end{bmatrix} = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{bmatrix}$$

neprekidna zajedno s neprekidnim prvim i drugim parcijalnim derivacijama na području Ω . Neka je $x^{(0)} \in \Omega$ točka čija je zatvorena R -kugla (u r -normi) sadržana u Ω , tj.

$$\overline{U}_R(x^{(0)}) = \{x \in \mathbb{R}^n \mid \|x - x^{(0)}\| \leq R\} \subset \Omega \quad (24)$$

i neka su ispunjeni sljedeći uvjeti:

I) Jacobijeva matrica $J(x^{(0)}) = \left[\frac{\partial f_i}{\partial x_j} \right]_{x=x^{(0)}}$ je regularna, tj. ima inverznu za koju vrijedi

$$\|J(x^{(0)})^{-1}\| \leq A; \quad (25)$$

2)

$$\|J(x^{(0)})^{-1} \cdot f(x^{(0)})\| \leq B \leq \frac{R}{2}; \quad (26)$$

3)

$$\sum_{k=1}^n \left| \frac{\partial^2 f_i(x)}{\partial x_j \partial x_k} \right| \leq C \quad \text{za } i, j = 1, 2, \dots, n \quad i \quad x \in \overline{U}_R(x^{(0)}); \quad (27)$$

4) Konstante A , B , C zadovoljavaju sljedeću nejednakost

$$\mu = 2nABC \leq 1, \quad (28)$$

gdje je n broj nepoznanica.

a) Tada Newtonov iteracijski postupak

$$x^{(k+1)} = x^{(k)} - J(x^{(k)})^{-1} f(x^{(k)}), \quad k = 0, 1, 2, \dots$$

konvergira za gornju početnu vrijednost $x^{(0)}$ i

$$\xi = \lim_{k \rightarrow \infty} x^{(k)}$$

je rješenje polaznog sustava za koje vrijedi

$$\|\xi - x^{(0)}\| \leq 2B \leq R; \quad (29)$$

b) Vrijedi sljedeća ocjena

$$\|\xi - x^{(k)}\| \leq \left(\frac{1}{2}\right)^{k-1} \mu^{2^k-1} B; \quad (30)$$

c) Unutar kugle

$$\|x - x^{(0)}\| \leq 2B \quad (31)$$

nalazi se jedno jedino rješenje sustava;

d) Ako je

$$\frac{2}{\mu}B \leq R, \quad (32)$$

onda Newtonov iteracijski postupak konvergira unutar kugle (31) pri proizvoljnoj promjeni početne aproksimacije $x^{(0)}$ unutar kugle

$$\|x'^{(0)} - x^{(0)}\| \leq \frac{1-\mu}{2\mu}B. \quad (33)$$

Pojednostavljena Newtonova metoda.

Već smo istaknuli da se prema (23) u svakom koraku računa inverzna matrica Jacobijeve matrice. To je bitna računska poteškoća te metode. Isto smo vidjeli na što se svodi u slučaju linearnih sustava jednadžbi. Sljedeće razmišljanje ne možemo općenito osigurati, ali je dobro da ga uočimo.

Ako je $J^{-1}(x)$ neprekidna u okolini rješenja sustava ξ i ako je početna aproksimacija $x^{(0)}$ dovoljno bliza rješenju sustava ξ , onda će se $J^{-1}(x^{(0)})$ malo mijenjati prelaskom u $x^{(k)}$, pa približno vrijedi

$$J^{-1}(x^{(k)}) \approx J^{-1}(x^{(0)})$$

što vodi na pojednostavljen iteracijski postupak

$$x^{(k+1)} = x^{(k)} - J^{-1}(x^{(0)})f(x^{(k)}). \quad (34)$$

Jasno da je $x^{(1)}$ isti onaj koji dobivamo iz (23) dok se ostale aproksimacije mijenjaju. Ono što posebno želimo istaći je da i pojednostavljena iteracija konvergira k rješenju pod uvjetima 1) – 4) gornjeg teorema za danu početnu aproksimaciju $x^{(0)}$.

Metoda iteracije za sustav dviju jednadžbi

Promatrajmo sustav dviju jednadžbi

$$\begin{aligned} f(x, y) &= 0 \\ g(x, y) &= 0 \end{aligned} \quad (35)$$

za koji prepostavljamo kao i do sada da dopušta samo izolirana realna rješenja. Istaknimo ponovo da je rješenje sustava (35) par brojeva (ξ, η) koji zadovoljava sustav.

Neka je (x_0, y_0) aproksimativno rješenje sustava (35), koje smo našli nekom grupom ocjenom ili grafički. Da dođemo do iteracijskog postupka prevedimo sustav (35) u oblik

$$\begin{aligned} x &= \varphi(x, y) \\ y &= \psi(x, y) \end{aligned} \quad (36)$$

ili vektorski

$$x = \phi(x). \quad (37)$$

Iteracijski postupak provodimo kao i u slučaju jedne jednadžbe, tj. formiramo niz aproksimacija

$$\begin{aligned} x_1 &= \varphi(x_0, y_0); & y_1 &= \psi(x_0, y_0), \\ x_2 &= \varphi(x_1, y_1); & y_2 &= \psi(x_1, y_1), \\ &\dots && \\ x_{n+1} &= \varphi(x_n, y_n); & y_{n+1} &= \psi(x_n, y_n), \\ &\dots \end{aligned} \quad (38)$$

Ako iteracijski postupak konvergira, tj. ako postoje limesi

$$\xi = \lim_{n \rightarrow \infty} x_n, \quad \eta = \lim_{n \rightarrow \infty} y_n$$

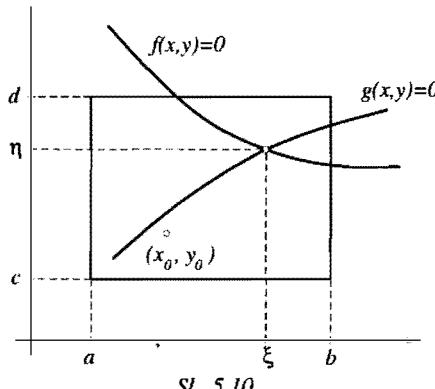
i ako su φ i ψ neprekidne funkcije dviju varijabli onda je (ξ, η) rješenje sustava (35) što uviđamo na isti način kao i u slučaju jednadžbe s jednom nepoznanicom. Prema tome, za dovoljno veliki n , (x_n, y_n) predstavlja aproksimaciju rješenja (ξ, η) . Sada navodimo teorem (bez dokaza) koji daje dovoljne uvjete pod kojima iteracijski postupak (38) konvergira.

Teorem 5.11. *Neka sustav (35) ima na pravokutniku P (v. sl. 5.10), $a \leq x \leq b$, $c \leq y \leq d$ samo jedno rješenje (ξ, η) . Ako*

- a) funkcije φ i ψ imaju neprekidne parcijalne derivacije;
- b) početna aproksimacija (x_0, y_0) i sve daljnje (x_n, y_n) , $n = 1, 2, \dots$ pripadaju pravokutniku P ;
- c) u P vrijede nejednakosti

$$\begin{aligned} \left| \frac{\partial \varphi}{\partial x} \right| + \left| \frac{\partial \psi}{\partial x} \right| &\leq q_1 < 1, \\ \left| \frac{\partial \varphi}{\partial y} \right| + \left| \frac{\partial \psi}{\partial y} \right| &\leq q_2 < 1, \end{aligned}$$

onda iteracijski postupak (38) konvergira.



Sl. 5.10.

Istaknimo da uvjet c) kaže da je na pravokutniku P stupčana norma Jacobijevе matrice preslikavanja danog s (36) manja od 1, tj. na P se traži da je norma derivacije od ϕ manja od 1, ili u oznaci

$$\|\phi'(x)\|_s = \|J(x)\|_s \leq q < 1.$$

Postupak iteracije može se ovako modificirati

$$\begin{aligned} x_{n+1} &= (x_n, y_n) \\ y_{n+1} &= (x_{n+1}, y_n) \end{aligned} \tag{39}$$

u kojem $(n+1)$ -vu aproksimaciju prve varijable koristimo za izračunavanje $(n+1)$ -ve aproksimacije druge varijable. Tako modificirana iteracija zove se **Seidelova** iteracija.

Primjer 5.9. Riješimo sustav jednadžbi

$$2x = \sin \frac{1}{2}(x - y)$$

$$2y = \cos \frac{1}{2}(x + y).$$

Provjerimo da postupak iteracije konvergira za ovaj izbor od φ i ψ :

$$x = \frac{1}{2} \sin \frac{1}{2}(x - y) = \varphi(x, y)$$

$$y = \frac{1}{2} \cos \frac{1}{2}(x + y) = \psi(x, y).$$

Imamo

$$\frac{\partial \varphi}{\partial x} = \frac{1}{4} \cos \frac{1}{2}(x - y), \quad \frac{\partial \varphi}{\partial y} = -\frac{1}{4} \cos \frac{1}{2}(x - y),$$

$$\frac{\partial \psi}{\partial x} = -\frac{1}{4} \sin \frac{1}{2}(x + y), \quad \frac{\partial \psi}{\partial y} = -\frac{1}{4} \sin \frac{1}{2}(x + y),$$

tako da vrijedi

$$\left| \frac{\partial \varphi}{\partial x} \right| + \left| \frac{\partial \psi}{\partial x} \right| \leq \frac{1}{4} + \frac{1}{4} = \frac{1}{2} < 1,$$

$$\left| \frac{\partial \varphi}{\partial y} \right| + \left| \frac{\partial \psi}{\partial y} \right| \leq \frac{1}{4} + \frac{1}{4} = \frac{1}{2} < 1$$

za svaki (x, y) , pa prema tome postupak iteracije konvergira na po volji velikom pravokutniku. Sada treba locirati rješenje. Najjednostavniji put je da nacrtamo krivulje definirane jednadžbama, ako je to moguće, i da očitano sjecište uzmememo za početnu aproksimaciju (x_0, y_0) . To je u ovom slučaju lako provedivo. Iz prve jednadžbe možemo izračunati y , dok iz druge možemo izračunati x tako da jednadžbe poprimaju oblik

$$y = x - 2 \arcsin 2x$$

$$x = 2 \arccos 2y - y.$$

Načinimo li tablice imamo za prvu jednadžbu ovo interesantno područje

x	-0.3	-0.2	-0.1	0	0.1	0.2
y	0.98	0.62	0.3	0	-0.3	-0.62

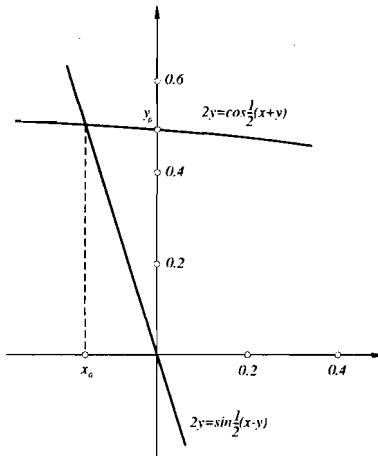
dok za drugu dobivamo

y	0.4	0.42	0.44	0.46	0.48	0.5
x	0.88	0.72	0.54	0.34	0.08	-0.5

Nacrtamo li to na milimetarskom papiru dobivamo sljedeću sliku (sl. 5.11) iz koje očitavamo sjecište u točki $(-0.16, 0.49)$. Time za početnu aproksimaciju uzimamo

$x_0 = -0.16$, $y_0 = 0.49$. Služeći se Seidelovom iteracijom dobivamo

$$\begin{aligned}x_1 &= -0.1597, & y_1 &= 0.4932 \\x_2 &= -0.1603, & y_2 &= 0.4931 \\x_3 &= -0.1605, & y_3 &= 0.4931 \\x_4 &= -0.1605.\end{aligned}$$



Sl. 5.11.

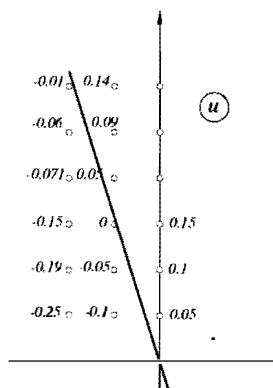
Kako nam se daljnje iteracije ponavljaju to ćemo postupak zaustaviti. To ponavljanje ujedno indicira da smo pronašli rješenje na 4 decimalna mjesta.

Početna aproksimacija. U promatranom primjeru smo mogli nacrtati krivulje $f(x, y) = 0$ i $g(x, y) = 0$ izračunavši iz prve jednadžbe y kao funkciju od x , a iz druge x kao funkciju od y i time naći početnu aproksimaciju. To nije u vijek moguće. Općenito se možemo poslužiti ovakvim grafičkim postupkom prema kojemu grubo skiciramo krivulje $f(x, y) = 0$, $g(x, y) = 0$. Pišemo

$$\begin{aligned}u &= f(x, y) \\v &= g(x, y)\end{aligned}$$

i na koordinatnoj mreži (x, y) označimo pripadne vrijednosti od u , odnosno v . (Najbolje je nacrtati dvije posebne mreže.) Prva krivulja je ona koja spaja točke $u \approx 0$, dok druga spaja točke $v \approx 0$, tako da dobivamo približno grafove. U našem primjeru bi imali

$$\begin{aligned}u &= 2x - \sin \frac{1}{2}(x - y) \\v &= 2y - \cos \frac{1}{2}(x + y).\end{aligned}$$



Sl. 5.12.

Na sl. 5.12 su u (x, y) mrežu upisane vrijednosti od u i približno skicirana krivulja dana prvom jednadžbom. Nacrtali se analogna slika za v i te dvije poklope jedna na drugu dobiva se slika poput sl. 5.11 s koje očitavamo početnu aproksimaciju (x_0, y_0) .

Neposrednu korist takvog postupka nalaženja početne aproksimacije imamo kod traženja kompleksnih rješenja jednadžbe

$$f(z) = 0. \quad (40)$$

Naime, rastavimo li $f(z)$ na realni i imaginarni dio dobivamo sustav dviju jednadžbi

$$\begin{aligned} \operatorname{Re} f(z) &= u(x, y) = 0 \\ \operatorname{Im} f(z) &= v(x, y) = 0 \end{aligned} \quad (41)$$

koji je ekvivalentan jednadžbi (40).

Metoda iteracije za opći nelinearni sustav

Nakon opisa metode iteracije za sustav dviju jednadžbi jasno je da sustav jednadžbi

$$f(x) = 0 \quad (42)$$

prevodimo u ekvivalentan oblik

$$x = \varphi(x) \quad (43)$$

koji u razvijenom obliku glasi

$$\begin{aligned} x_1 &= \varphi_1(x_1, x_2, \dots, x_n) \\ x_2 &= \varphi_2(x_1, x_2, \dots, x_n) \\ &\vdots \\ x_n &= \varphi_n(x_1, x_2, \dots, x_n) \end{aligned} \quad (44)$$

te da na njemu provodimo analogan postupak s n varijabli. Polazimo dakle od početne aproksimacije $x^{(0)}$ i računamo

$$x^{(k+1)} = \varphi(x^{(k)}), \quad k = 0, 1, 2, \dots \quad (45)$$

Također je jasno da možemo provoditi Seidelov iteracijski postupak i u ovom slučaju. Isto je tako jasno, ako je dobiveni niz $x^{(k)}$ konvergentan i ako su funkcije φ_i neprekidne da je onda

$$\xi = \lim_{k \rightarrow \infty} x^{(k)}$$

rješenje sustava (43), odnosno (42).

Uvjet koji osigurava konvergenciju iteracijskog postupka je onaj koji smo imali za dvije jednadžbe prenešen na n dimenzionalni slučaj. Prije nego navedemo takav teorem razmotrimo koristan poseban slučaj kada je φ kontrakcija. Evo kako možemo interpretirati (43). Označimo sa

$$\begin{aligned} y_1 &= \varphi_1(x_1, x_2, \dots, x_n) \\ y_2 &= \varphi_2(x_1, x_2, \dots, x_n) \\ &\vdots \\ y_n &= \varphi_n(x_1, x_2, \dots, x_n) \end{aligned} \tag{46}$$

ili kraće

$$y = \varphi(x). \tag{47}$$

Dakle, φ je preslikavanje iz \mathbf{R}^n u \mathbf{R}^n , a rješiti sustav (43) znači naći točku $x \in \mathbf{R}^n$ koju φ preslikava u samu sebe. Takve točke se zovu **fiksne** točke. Među preslikavanjima $\varphi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ posebno su istaknuta preslikavanja koja recimo "podjednako" smanjuju normu (udaljenost). Evo što to znači. Neka je $0 < q < 1$, realan broj. Kažemo da je φ **kontrakcija na području Ω** ako preslikava Ω u Ω i ako za svake dvije točke $x_1, x_2 \in \Omega$ njihove slike $y_1 = \varphi(x_1)$, $y_2 = \varphi(x_2)$ zadovoljavaju

$$\|\varphi(x_1) - \varphi(x_2)\| \leq q\|x_1 - x_2\|, \quad 0 < q < 1. \tag{48}$$

Ako je φ kontrakcija, onda iteracijski niz $\{x^{(k)}\}$ konvergira k fiksnoj točki, tj. rješenju sustava (43), što je posljedica jednog znatno općenitijeg teorema.

Vratimo se općem slučaju tj. kad φ ne mora biti kontrakcija i navedimo bez dokaza ovaj teorem.

Teorem 5.12. *Neka su φ i φ' neprekidne na ograničenom području Ω na kojemu vrijedi*

$$\|\varphi'(x)\|_r \leq q < 1, \quad \text{za svaki } x \in \Omega. \tag{49}$$

Ako članovi niza

$$x^{(k+1)} = \varphi(x^{(k)}) \quad k = 0, 1, 2, \dots$$

pripadaju u Ω , onda iteracijski postupak konvergira i

$$\xi = \lim_{k \rightarrow \infty} x^{(k)}$$

je jedinstveno rješenje sustava (42) u Ω . Nadalje, za ocjenu pogreške vrijedi

$$\|\xi - x^{(k)}\|_r \leq \frac{q^k}{1-q} \|x^{(1)} - x^{(0)}\|_r, \quad k = 0, 1, \dots \tag{50}$$

Teorem smo iskazali za rednu normu, no vrijedi i za bilo koju drugu submultiplikativnu normu koja je konzistentna s promatranom normom vektora.

6.

Aproksimacija rješenja običnih diferencijalnih jednadžbi

6.1. Metoda sukcesivne aproksimacije	181
6.2. Eulerova metoda	185
6.3. Runge–Kuttini postupci	189

Promatrajmo za početak diferencijalnu jednadžbu

$$y' = f(x, y). \quad (1)$$

Osnovni problem vezan za (1) je rješenje Cauchyjeva problema tj. treba odrediti rješenje diferencijalne jednadžbe (1) koje zadovoljava dani početni uvjet $y = y_0$ za $x = x_0$, drugim riječima treba naći integralnu krivulju koja prolazi zadanom točkom $M_0(x_0, y_0)$. Kao što znamo, prema Picardovom teoremu, takvo rješenje postoji i nadalje ono je jedinstveno, ako je f neprekidna na pravokutniku $|x - x_0| \leq a$, $|y - y_0| \leq b$ i ako f zadovoljava Lipschitzov uvjet s obzirom na varijablu y (tj. postoji $L > 0$ takav da za svaki x , $|x - x_0| < a$ i svake dvije vrijednosti y_1 i y_2 , $|y_1 - y_0| < b$, $|y_2 - y_0| < b$, $|y_2 - y_1| < b$ vrijedi $|f(x, y_2) - f(x, y_1)| \leq L|y_2 - y_1|$). Ako se prisjetimo dokaza Picardova teorema onda vidimo da nam dokaz daje i postupak aproksimacije rješenja Cauchyjevog problema. Taj postupak zovemo **metoda sukcesivne aproksimacije**.

6.1. Metoda sukcesivne aproksimacije

Pođimo od gore opisanog Cauchyevog problema

$$y' = f(x, y), \quad y(x_0) = y_0. \quad (1)$$

Potražiti ćemo rješenje za $x \geq x_0$, dok je za $x \leq x_0$ situacija potpuno analogna. Integracijom lijeve i desne strane od (1) u granicama od x_0 do x dobivamo

$$y(x) - y(x_0) = \int_{x_0}^x f[x, y(x)]dx$$

ili zbog početnog uvjeta

$$y(x) = y_0 + \int_{x_0}^x f[x, y(x)]dx. \quad (2)$$

Kako se tražena funkcija $x \rightarrow y(x)$ nalazi pod znakom integrala to je (2) integralna jednadžba čije rješenje očigledno zadovoljava promatrani Cauchyev problem (1).

Metoda sukcesivne aproksimacije sastoji se sada u sljedećem. U (2) zamjenimo nepoznanicu y sa zadatom vrijednošću y_0 na koji način postižemo prvu aproksimaciju koja glasi

$$y_1(x) = y_0 + \int_{x_0}^x f(x, y_0)dx. \quad (3)$$

Ako postupak ponovimo, tj. ako u (2) uvrstimo nađenu aproksimaciju (3) dobivamo drugu aproksimaciju

$$y_2(x) = y_0 + \int_{x_0}^x f[x, y_1(x)]dx. \quad (4)$$

Nastavivši postupak sa novodobivenim aproksimacijama imamo induktivno sljedeću formulu za n -tu aproksimaciju

$$y_n(x) = y_0 + \int_{x_0}^x f[x, y_{n-1}(x)]dx, \quad n = 1, 2, \dots \quad (5)$$

Geometrijski interpretirano, niz (5) predstavlja niz krivulja u ravnini koje prolaze kroz zajedničku točku $M_0(x_0, y_0)$.

U dokazu Picardovog teorema se pokazuje da na nekom segmentu $[x_0, x_0 + h]$ niz (5) uniformno konvergira funkciji

$$y(x) = \lim_{n \rightarrow \infty} y_n(x) \quad (6)$$

koja predstavlja jedinstveno rješenje Cauchyevog problema (1). Taj dokaz izostavljamo i prihvaćamo rezultat kao poznat. Ako je M ograda za $|f|$ na promatranom

pravokutniku $|x - x_0| < a$, $|y - y_0| < b$, onda vrijedi

$$h = \min \left\{ a, \frac{b}{M} \right\}, \quad (7)$$

što ima za posljedicu da se integralna krivulja rješenja $y(x)$ za $x_0 \leq x \leq x_0 + h$ nalazi unutar sektora omeđenog pravcima

$$y - y_0 = M(x - x_0)$$

i

$$y - y_0 = -M(x - x_0).$$

Ocijenimo sada pogrešku n -te aproksimacije $\varepsilon_n(x) = |y(x) - y_n(x)|$. Iz (2) i (5) dobivamo

$$y(x) - y_n(x) = \int_{x_0}^x [f(x, y(x)) - f(x, y_{n-1}(x))] dx$$

odnosno,

$$\varepsilon_n(x) = |y(x) - y_n(x)| \leq \int_{x_0}^x |f(x, y(x)) - f(x, y_{n-1}(x))| dx$$

a zbog Lipschitzovog uvjeta imamo

$$|f(x, y(x)) - f(x, y_{n-1}(x))| \leq L|y(x) - y_{n-1}(x)| \leq L\varepsilon_{n-1}(x).$$

Prema tome nalazimo

$$\varepsilon_n(x) \leq \int_{x_0}^x L\varepsilon_{n-1}(x) dx, \quad n = 1, 2, \dots \quad (8)$$

pri čemu za ε_0 uzimamo

$$\varepsilon_0(x) = |y(x) - y_0|. \quad (9)$$

Primjenom teorema o srednjoj vrijednosti na ε_0 dobivamo za $x_0 \leq x \leq x_0 + h$

$$\varepsilon_0(x) = |y(x) - y(x_0)| = (x - x_0)|y'(ξ)|, \quad x_0 < ξ < x + h.$$

Kako za y' vrijedi

$$|y'(\xi)| = |f(\xi, y(\xi))| \leq M$$

to postižemo

$$\varepsilon_0(x) \leq M(x - x_0).$$

Koristeći (8) imamo sada

$$\varepsilon_1(x) \leq L \int_{x_0}^x \varepsilon_0(x) dx \leq ML \int_{x_0}^x (x - x_0) dx = ML \frac{(x - x_0)^2}{2!};$$

$$\varepsilon_2(x) \leq L \int_{x_0}^x \varepsilon_1(x) dx \leq ML^2 \int_{x_0}^x \frac{(x - x_0)^2}{2!} dx = ML^2 \frac{(x - x_0)^3}{3!};$$

itd., tako da za ocjenu pogreške n -te aproksimacije vrijedi

$$\varepsilon_n(x) \leq ML^n \frac{(x - x_0)^{n+1}}{(n+1)!}, \quad n = 0, 1, 2, \dots \quad (10)$$

Iz (10) čitamo da $\varepsilon_n(x) \rightarrow 0$ za $n \rightarrow \infty$ i to uniformno na segmentu $[x_0, x_0 + h]$.

Primjetimo da za početnu aproksimaciju ne moramo uzeti konstantu y_0 već bilo koju funkciju koja je dovoljno blizu traženom rješenju i zadovoljava početni uvjet. Tako se često za početnu aproksimaciju uzima odsječak Taylorovog reda traženog rješenja.

Kako desna strana od (1) ne mora biti analitička funkcija, što ima za posljedicu da ni rješenje y nije analitička funkcija (tj. ne može se razviti u red potencija) to vidimo da se metoda sukcesivne aproksimacije može primijeniti i u toj situaciji.

Primjer 6.1. Metodom sukcesivne aproksimacije riješite diferencijalnu jednadžbu

$$y' = x - y$$

uz početni uvjet

$$y(0) = 1.$$

Prema gore izloženom uzimamo $y_0(x) = 1$, što daje

$$y_1(x) = 1 + \int_0^x (x - 1) dx = 1 - x + \frac{x^2}{2}.$$

Analogno imamo

$$y_2(x) = 1 + \int_0^x \left(x - 1 + x - \frac{x^2}{2} \right) dx$$

$$= 1 - x + x^2 - \frac{x^3}{6};$$

$$y_3(x) = 1 - x + x^2 - \frac{x^3}{3} + \frac{x^4}{24};$$

$$y_4(x) = 1 - x + x^2 - \frac{x^3}{3} + \frac{x^4}{12} - \frac{x^5}{120}.$$

itd.

Pokušamo li ocijeniti pogrešku četvrte aproksimacije y_4 , onda nas to vodi na sljedeće razmatranje. Promatramo li f na pravokutniku $|x| \leq a$, $|y - 1| \leq b$, vidimo da je $f(x, y) = x - y$ na njemu diferencijabilna a time i neprekidna. Kako je f diferencijabilna na cijeloj ravnini to a i b mogu biti proizvoljni pozitivni brojevi. Ogradu za f lako dobivamo iz

$$|f(x, y)| = |x - y| \leq |x| + |y| = |x| + |y - 1 + 1| \leq |x| + |y - 1| + 1 \leq a + b + 1 = M.$$

Ako je $a \geq 1$ onda $h = \min\left\{a, \frac{b}{M}\right\} = \frac{b}{b+a+1}$, tako da za odabранe $a = 1$ i $b = 1$ imamo $h = \frac{1}{3}$. Iz $|(x - y_2) - (x - y_1)| = |y_2 - y_1|$ vidimo da možemo uzeti

$L = 1$ (Lipschitzova konstanta). Na taj način imamo

$$\varepsilon_4(x) = |y(x) - y_4(x)| \leq 3 \cdot 1^4 \frac{x^5}{5!} = \frac{x^5}{40},$$

odnosno

$$\varepsilon_4 = \max |\varepsilon_4(x)| \leq \frac{1}{40 \cdot 243} = \frac{1}{9720} \quad \text{na } [0, \frac{1}{3}].$$

Spomenimo odmah da se metoda sukcesivne aproksimacije može primijeniti i na sustav diferencijalnih jednadžbi

$$y' = f(x, y), \quad y(x_0) = y^{(0)}, \quad (11)$$

$$\text{gdje je } y(x) = \begin{bmatrix} y_1(x) \\ y_2(x) \\ \vdots \\ y_n(x) \end{bmatrix}, \quad y'(x) = \begin{bmatrix} y'_1(x) \\ y'_2(x) \\ \vdots \\ y'_n(x) \end{bmatrix}, \quad y^{(0)} = y(x_0) = \begin{bmatrix} y_1(x_0) \\ y_2(x_0) \\ \vdots \\ y_n(x_0) \end{bmatrix} \text{ i}$$

$$f(x, y) = \begin{bmatrix} f_1(x, y_1, \dots, y_n) \\ f_2(x, y_1, \dots, y_n) \\ \vdots \\ f_n(x, y_1, \dots, y_n) \end{bmatrix}.$$

Naime, postupak se primjenjuje po komponentama, tj. ako su f_i komponente od f , onda imamo

$$\int_{x_0}^x f dx = \begin{bmatrix} \int_{x_0}^x f_1(x, y(x)) dx \\ \vdots \\ \int_{x_0}^x f_n(x, y(x)) dx \end{bmatrix}$$

dok je niz aproksimacija $y^{(p)}$ ($p = 1, 2, \dots$) određen sa

$$y^{(p)} = y^{(0)} + \int_{x_0}^x f(x, y^{(p-1)}) dx, \quad y^{(0)} = y(x_0). \quad (12)$$

Ta je metoda primjenjiva i na diferencijalne jednadžbe n -tog reda

$$\begin{aligned} y^{(n)} &= f(x, y, y', \dots, y^{n-1}) \\ y(x_0) &= y_0, \quad y'(x_0) = y'_0, \dots, y^{(n-1)}(x_0) = y_0^{(n-1)}. \end{aligned} \quad (13)$$

Ako (13) prevedemo u **normalni sustav** pridružen diferencijalnoj jednadžbi (13), koji supstitucijom $y_1 = y$ i $y_{k+1} = y^{(k)}$ glasi

$$\begin{aligned} y'_1 &= y_2 \\ y'_2 &= y_3 \\ &\dots \\ y'_{n-1} &= y_n \\ y'_n &= f(x, y_1, y_2, \dots, y_n), \end{aligned} \quad (14)$$

onda primjenom gore opisanog postupka na sustav (14) dobivamo aproksimaciju rješenja diferencijalne jednadžbe (13). Naime, prva komponenta y_1 rješenja sustava (14) predstavlja traženo rješenje, dok ostale komponente predstavljaju redom prvu i više derivacije tog rješenja.

Primjer 6.2. Odredite prve dvije sukcesivne aproksimacije rješenja sustava

$$\begin{aligned}y'_1 &= x + y_1 y_2 \\y'_2 &= x^2 - y_1^2\end{aligned}$$

koje zadovoljava ove početne uvjete

$$y_1(0) = 1, \quad y_2(0) = 0.$$

Prema (12) imamo

$$\begin{bmatrix} y_1^{(p)} \\ y_2^{(p)} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} \int_0^x (x + y_1^{(p-1)} y_2^{(p-1)}) dx \\ \int_0^x (x^2 - [y_1^{(p-1)}]^2) dx \end{bmatrix}.$$

Time dobivamo redom prvu i drugu sukcesivnu aproksimaciju

$$\begin{bmatrix} y_1^{(1)} \\ y_2^{(1)} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} \int_0^x (x + 0) dx \\ \int_0^x (x^2 - 1) dx \end{bmatrix} = \begin{bmatrix} 1 + \frac{x^2}{2} \\ -x + \frac{x^3}{3} \end{bmatrix};$$

$$\begin{bmatrix} y_1^{(2)} \\ y_2^{(2)} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} \int_0^x \left[x + \left(1 + \frac{x^2}{2}\right) \left(-x + \frac{x^3}{3}\right) \right] dx \\ \int_0^x \left[x^2 - \left(1 + x^2 + \frac{x^4}{4}\right) \right] dx \end{bmatrix} = \begin{bmatrix} 1 + \frac{x^4}{24} + \frac{x^6}{36} \\ -x + \frac{x^5}{20} \end{bmatrix}.$$

6.2. Eulerova metoda

Promatrajmo ponovo Cauchyjev problem

$$y' = f(x, y), \quad y(x_0) = y_0 \tag{1}$$

i pokušajmo izračunati aproksimativne vrijednosti y_1, y_2, \dots, y_n u točkama x_1, x_2, \dots, x_n . Neka je \hat{y} po dijelovima linearna funkcija čiji je graf nad svakim segmentom $[x_k, x_{k+1}]$ spojnica točaka (x_k, y_k) i (x_{k+1}, y_{k+1}) . To da je \hat{y} na $[x_k, x_{k+1}]$ dano sa

$$\hat{y}(x) = y_k + \frac{1}{h}(x - x_k)(y_{k+1} - y_k)$$

gdje je $h = x_{k+1} - x_k$. Ako $\hat{y}(x)$ dobro aproksimira $y(x)$ u točkama x_k , $k = 0, 1, 2, \dots, n$ i ako su diobene točke odabrane dovoljno gusto, onda će zbog neprekidnosti \hat{y} aproksimirati y na svakom podsegmentu $[x_k, x_{k+1}]$. Prema tome treba nam

postupak kojim dobivamo po dijelovima linearnu aproksimaciju. Zbog jednostavnosti redovito promatramo jednoliku podjelu, $x_k = x_0 + k\frac{a}{n}$, pri čemu smatramo da je f definirana na pravokutniku $|x - x_0| \leq a$, $|y - y_0| \leq b$ kao i u prethodnom paragrafu. Jedino što znamo o funkciji y je da zadovoljava diferencijalnu jednadžbu (1) te da je $y(x_0) = y_0$, što moramo iskoristiti kada računamo aproksimativnu vrijednost y_1 u točki $x_1 = x_0 + h$. Zatim ćemo upotrijebiti y_1 da izračunamo aproksimativnu vrijednost y_2 u točki $x_2 = x_1 + h$. Da to dobijemo poslužit ćemo se Taylorovim redom koji daje

$$y(x_k + h) = y(x_k) + hy'(x_k) + \frac{h^2}{2!}y''(x_k) + \dots \quad (2)$$

Prema tome poznavajući y i njene derivacije u točki $x = x_k$ možemo izračunati vrijednost u točki $x = x_k + h$. Kako y zadovoljava Cauchyjev problem (1) to je derivacija u točki x_k jednaka $f(x_k, y(x_k))$. Izračunat ćemo drugu i više derivacije (pomoću pravila za deriviranje složenih funkcija) dobivamo

$$y''(x_k) = \left[\frac{\partial f}{\partial x} + f \frac{\partial f}{\partial y} \right](x_k, y(x_k)),$$

tako da (2) poprima oblik

$$y(x_{k+1}) = y(x_k) + hf(x_k, y(x_k)) + \frac{h^2}{2!} \left[\frac{\partial f}{\partial x} + f \frac{\partial f}{\partial y} \right](x_k, y(x_k)) + \dots \quad (3)$$

Najjednostavniju aproksimaciju od $y(x_{k+1})$ dobivamo ako upotrijebimo samo prva dva člana Taylorovog reda (3) i u njima zamijenimo $y(x_k)$ sa aproksimacijom y_k . To redom daje

$$\begin{aligned} y_0 &= y(x_0) \\ y_1 &= y_0 + hf(x_0, y_0) \\ y_2 &= y_1 + hf(x_1, y_1) \\ &\dots \\ y_{k+1} &= y_k + hf(x_k, y_k). \end{aligned} \quad (4)$$

Postupak aproksimacije rješenja Cauchyjeva problema (1) opisan formulom (4) zove se **Eulerova metoda**. To je vrlo jednostavan postupak. Rezultati nemaju naročitu točnost jer smo y aproksimirali samo sa prva dva člana Taylorovog reda. Međutim, može dobro poslužiti kao uvod u razumijevanje kompliziranih postupaka.

Primjer 6.3. Izračunajte pomoću Eulerove metode aproksimativne vrijednosti rješenja sljedećeg Cauchyjevog problema

$$y'(x) = 1 + (y - x)^2, \quad y(0) = \frac{1}{2}$$

promatranog na segmentu $[0, 1]$. Formula (4) daje

$$y_{k+1} = y_k + h[1 + (y_k - x_k)^2], \quad k = 0, 1, \dots, n - 1,$$

i $h = \frac{1}{n}$, uz $x_0 = 0$ i $y_0 = \frac{1}{2}$. Za $n = 10$ dobivamo sljedeću tablicu

x	y	x	y
0	0.5	0.6	1.29810
0.1	0.625	0.7	1.44684
0.2	0.75256	0.8	1.60261
0.3	0.88309	0.9	1.76703
0.4	1.01709	1	1.94220
0.5	1.15517		

Druga dobra strana Eulerove metode je da se dosta lako ocjenjuje pogreška. Prema Taylorovoj formuli (uz ostatak u Lagrangeovom obliku) imamo

$$y(x_{k+1}) = y(x_k) + hf(x_k, y(x_k)) + \frac{h^2}{2!} \left[\frac{\partial f}{\partial x} + f \frac{\partial f}{\partial y} \right](\xi_k, y(\xi_k)) \quad (5)$$

gdje je $\xi_k \in (x_k, x_{k+1})$. Odbijemo li (4) od (5), dobivamo

$$\begin{aligned} y(x_{k+1}) - y_{k+1} &= y(x_k) - y_k + h[f(x_k, y(x_k)) - f(x_k, y_k)] \\ &\quad + \frac{h^2}{2!} \left[\frac{\partial f}{\partial x} + f \frac{\partial f}{\partial y} \right](\xi_k, y(\xi_k)). \end{aligned}$$

Primjenom teorema o srednjoj vrijednosti s obzirom na varijablu y dobivamo za razliku u uglatoj zagradi .

$$f(x_k, y(x_k)) - f(x_k, y_k) = \frac{\partial f(x_k, \eta_k)}{\partial y} (y(x_k) - y_k)$$

gdje je η_k između $y(x_k)$ i y_k . Prema tome imamo

$$\begin{aligned} |y(x_{k+1}) - y_{k+1}| &\leq |y(x_k) - y_k| + h \left| \frac{\partial f(x_k, \eta_k)}{\partial y} \right| |y(x_k) - y_k| \\ &\quad + \frac{h^2}{2!} \left| \left[\frac{\partial f}{\partial x} + f \frac{\partial f}{\partial y} \right](\xi_k, y(\xi_k)) \right|. \end{aligned} \quad (6)$$

Kao što znamo iz dokaza Picardovog teorema (usporedi prethodni paragraf i sl.2.) točke $(\xi_k, y(\xi_k))$ leže u pravokutniku P određenom sa

$$|x - x_0| \leq \min \left\{ a, \frac{b}{M} \right\}, \quad |y - y_0| < b,$$

gdje je M ograda za $|f|$. Međutim, i točke (x_k, y_k) leže u tom pravokutniku što se može dokazati indukcijom, tako da i točke (x_k, η_k) leže u pravokutniku P . Ako su L i D ograde za $\left| \frac{\partial f}{\partial y} \right|$ i $\left| \frac{\partial f}{\partial x} + f \frac{\partial f}{\partial y} \right|$ onda iz (6) dobivamo

$$|y(x_{k+1}) - y_{k+1}| \leq |y(x_k) - y_k| + hL|y(x_k) - y_k| + \frac{Dh^2}{2}. \quad (7)$$

Označimo li sa

$$\varepsilon_k = |y(x_k) - y_k|, \quad k = 0, 1, \dots, n \quad (8)$$

pogrešku koju činimo aproksimirajući $y(x_k)$ na y_k , onda iz (7) dobivamo

$$\varepsilon_{k+1} \leq (1 + hL)\varepsilon_k + \frac{Dh^2}{2}, \quad k = 0, 1, \dots, n - 1. \quad (9)$$

Nadalje, $\varepsilon_0 = 0$ jer je $y(x_0) = y_0$. na taj način pogreške $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_n$ zadovoljavaju nejednakosti

$$\varepsilon_{k+1} \leq A\varepsilon_k + B, \quad \varepsilon_0 = 0 \quad (10)$$

gdje je $A = 1 + hL$, $B = D\frac{h^2}{2}$.

Sada bi trebali ocijeniti pogrešku ε_k ako znamo da pogreške zadovoljavaju nejednakost (10), gdje su A i B pozitivni brojevi.

Imamo dakle

$$\begin{aligned} \varepsilon_k &\leq A\varepsilon_{k-1} + B \leq A(A\varepsilon_{k-2} + B) + B = A^2\varepsilon_{k-2} + AB + B \\ &\leq A^2(A\varepsilon_{k-3} + B) + AB + B = A^3\varepsilon_{k-3} + A^2B + AB + B \\ &\leq A^3(A\varepsilon_{k-4} + B) + A^2B + AB + B = A^4\varepsilon_{k-4} + A^3B + A^2B + AB + B \\ &\leq \dots \end{aligned}$$

Sukcesivnim uvrštavanjem dobivamo očigledno

$$\varepsilon_k \leq A^k\varepsilon_0 + A^{k-1}B + \dots + A^2B + AB + B. \quad (11)$$

Zbog $\varepsilon_0 = 0$ preostali pribrojnici desne strane od (11) čine n -tu parcijalnu sumu geometrijskog reda, tako da imamo

$$\varepsilon_k \leq \frac{B}{A-1}(A^k - 1) = \frac{Dh}{2L}[(1+hL)^k - 1]. \quad (12)$$

Kao daljnji korak naći ćemo ocjenu pogreške koja ne ovisi o k iz čega će ujedno slijediti konvergencija. Uoči li se da vrijedi

$$1 + hL \leq e^{hL}$$

(što lako uvidimo iz reda za eksponencijalnu funkciju), onda zamjenom u (12) imamo

$$\varepsilon_k \leq \frac{Dh}{2L}[(e^{hL})^k - 1] = \frac{Dh}{2L}[e^{khL} - 1].$$

Kako je $kh < \frac{b}{M}$ dobivamo

$$\varepsilon_k \leq \frac{Dh}{2L}\left[e^{\frac{hL}{M}} - 1\right] \leq \frac{Dh}{2L}e^{\frac{hL}{M}}, \quad k = 1, \dots, n. \quad (13)$$

Nejednakost (13) kaže da je pogreška proporcionalna za h te nadalje da $\varepsilon_k \rightarrow 0$ kada $h \rightarrow 0$. Jasno je kako se Eulerova metoda prenosi na sustav diferencijalnih jednadžbi

$$y' = f(x, y), \quad y(x_0) = y_0. \quad (14)$$

Treba postupak primijeniti po komponentama.

Eulerovu metodu smo dobili iz (2) tako da smo kao aproksimaciju uzeli prva dva člana Taylorova reda. Izvrši li se aproksimacija sa prva tri člana Taylorova reda dobivamo

$$y_{k+1} = y_k + hf(x_k, y_k) + \frac{h^2}{2}\left[\frac{\partial f}{\partial x} + f\frac{\partial f}{\partial y}\right](x_k, y_k), \quad k = 0, 1, \dots, n-1. \quad (15)$$

To se ponekad zove **metoda sa tri člana Taylorova reda**, koja je preciznija od Eulerove.

6.3. Runge-Kuttini postupci

Već smo istakli da smo Eulerovu metodu dobili tako da smo $y(x_{k+1})$ aproksimirali s prva dva člana Taylorovog reda. Bolja aproksimacija pomoću Taylorovog reda zahtijeva izračunavanje derivacija višeg reda što je u većini slučajeva komplikirano. Zato izvršimo slijedeće razmatranje. Podemo li od

$$y_{k+1} = y_k + h f(x_k, y_k)$$

onda smo, gledano geometrijski, u točki x_k uzeli aproksimaciju y_k i gradijent smjera $f(x_k, y_k)$ koji diferencijalna jednadžba $y' = f(x, y)$ propisuje u točki (x_k, y_k) . Time linearom interpolacijom dobivamo graf aproksimacije \hat{y} . Drugim riječima, u točki x_k smo uzeli $f(x_k, y_k)$ za aproksimaciju derivacije rješenja. Točna vrijednost derivacije u toj točki iznosi $f(x_k, y(x_k))$, pa se nameće pitanje izbora neke druge aproksimacije od $f(x_k, y(x_k))$, što u krajnjoj liniji vodi na prikidan izbor aproksimacije od $f(x, y)$ na segmentu $x_k \leq x \leq x_{k+1}$. Takva interpretacija nas vodi na opći algoritam

$$y_{k+1} = y_k + h \Phi(x_k, y_k, h) \quad (1)$$

gdje je $\Phi(x_k, y_k, h)$ neka odabrana aproksimacija od $f(x_k, y(x_k))$.

To otvara široke mogućnosti izbora Φ , a ovdje ćemo prikazati vrlo korisnu metodu koju su razvili matematičari Runge i Kutta oko 1900. godine. Kako možemo vršiti različite izvore aproksimacije Φ to imamo čitavu seriju Runge-Kuttinih postupaka (algoritama). Najlakše dolazimo do Runge-Kuttinog postupka (algoritma) drugog reda, čiji izvod nam daje put kako se dobivaju ostali. U tom slučaju promatramo težinsku aritmetičku sredinu dviju aproksimacija t_1, t_2 derivacije y' na segmentu $x_k \leq x \leq x_{k+1}$, tj. uzimamo

$$\Phi = at_1 + bt_2. \quad (2)$$

(Težinska aritmetička sredina brojeva t_1 i t_2 s težinama r i s jednaka je $\frac{rt_1 + st_2}{r + s}$).

Time prema (1) imamo

$$y_{k+1} = y_k + h(at_1 + bt_2). \quad (3)$$

Neka je

$$t_1 = f(x_k, y_k) \quad (4)$$

tj. aproksimacija od y' kao kod Eulerove metode. Tada za aproksimaciju t_2 možemo uzeti

$$\begin{aligned} t_2 &= f(x_k + ph, y_k + qhf(x_k, y_k)) \\ &= f(x_k + ph, y_k + qht_1) \end{aligned} \quad (5)$$

tj. analogno računatu aproksimaciju u nekoj međutočki na $[x_k, x_{k+1}]$, gdje su p i q konstante koje ćemo naknadno odrediti. Razvijmo prvo t_2 kao funkciju dviju varijabli p i q u Taylorov red oko točke $p = 0, q = 0$. Time imamo

$$\begin{aligned} t_2 &= f(x_k + ph, y_k + qhf(x_k, y_k)) \\ &= f(x_k, y_k) + phf_x(x_k, y_k) + qhf(x_k, y_k)f_y(x_k, y_k) + \dots \end{aligned} \quad (6)$$

Iz (6), (5), (4) i (3) dobivamo

$$\begin{aligned} y_{k+1} &= y_k + h[af(x_k, y_k) + bf(x_k, y_k)] \\ &\quad + h^2[bpf_x(x_k, y_k) + bqf(x_k, y_k)f_y(x_k, y_k)] + \dots \end{aligned} \quad (7)$$

Nadalje, razvijmo traženo rješenje y u Taylorov red oko x_k . Imamo

$$\begin{aligned} y(x_{k+1}) &= y(x_k + h) = y(x_k) + hf(x_k, y(x_k)) + \frac{h^2}{2}f'(x_k, y(x_k)) \\ &\quad + \frac{h^3}{3!}f''(\xi, y(\xi)), \quad \xi \in (x_k, x_{k+1}). \end{aligned} \quad (8)$$

Po teoremu za deriviranje složenih funkcija imamo

$$f'(x_k, y(x_k)) = f_x(x_k, y(x_k)) + f_y(x_k, y(x_k))f(x_k, y(x_k)),$$

pa uz iste potencije od h imamo u (7) i (8) ove koeficijente:

Potencije od h	Razvoj od y	Runge-Kuttin algoritam
0	$y(x_k)$	y_k
1	$f(x_k, y(x_k))$	$(a + b)f(x_k, y_k)$
2	$\frac{1}{2}[f_x(x_k, y(x_k)) + f_y(x_k, y(x_k))f(x_k, y(x_k))]$	$bpf_x(x_k, y_k) + bqf_y(x_k, y_k)f(x_k, y_k)$

Pretpostavimo li da je $y_k = y(x_k)$ i izjednačimo li koeficijente uz h i h^2 imamo

$$a + b = 1$$

i

$$bp = \frac{1}{2}, \quad bq = \frac{1}{2}$$

tako da iz

$$a = 1 - b$$

dobivamo

$$p = q = \frac{1}{2b}.$$

Sada još ostaje izbor od b . Dva uobičajena izbora su

$$b = \frac{1}{2} \quad i \quad b = 1.$$

Slučaj $b = \frac{1}{2}$. Tada je $a = \frac{1}{2}$, $p = q = 1$, pa dobivamo algoritam

$$y_{k+1} = y_k + \frac{h}{2}[f(x_k, y_k) + f(x_k + h, y_k + hf(x_k, y_k))]$$

što se može pisati i ovako

$$y_{k+1} = y_k + \frac{h}{2}[f(x_k, y_k) + f(x_{k+1}, \bar{y}_{k+1})] \quad (9)$$

gdje je

$$\bar{y}_{k+1} = y_k + hf(x_k, y_k). \quad (10)$$

Postupak koji daju (9) i (10) zove se još **poboljšana Eulerova metoda** ili **Heunova metoda**. Čitamo li (9) i (10) detaljno, onda vidimo da smo u biti dvaput upotrebljavali Eulerovu metodu. Prvo imamo \bar{y}_{k+1} kao aproksimaciju od $y(x_{k+1})$. U drugom koraku poboljšavamo aproksimaciju koju računamo iz (9).

Slučaj $b = 1$. Tada je $a = 0$, $p = q = \frac{1}{2}$, tako da dobivamo algoritam

$$y_{k+1} = y_k + h f\left(x_k + \frac{h}{2}, \bar{y}_{k+1/2}\right) \quad (11)$$

gdje je

$$\bar{y}_{k+1/2} = y_k + \frac{h}{2} f(x_k, y_k), \quad (12)$$

što se još zove i **modificirana Eulerova metoda**.

Runge-Kuttini algoritmi višeg reda dobivaju se analogno. U slučaju trećeg reda imamo

$$\Phi = at_1 + bt_2 + ct_3$$

i

$$y_{k+1} = y_k + h\Phi(x_k, y_k, h) \quad (13)$$

gdje su t_1, t_2, t_3 aproksimacije derivacije, odnosno f , u različitim točkama segmenta $[x_k, x_{k+1}]$. U tom slučaju uzimamo

$$\begin{aligned} t_1 &= f(x_k, y_k) \\ t_2 &= f(x_k + ph, y_k + pht_1), \\ t_3 &= f(x_k + rh, y_k + sht_2 + (r - s)ht_1), \end{aligned}$$

tako da Runge-Kuttin algoritam trećeg reda glasi

$$y_{k+1} = y_k + h(at_1 + bt_2 + ct_3). \quad (14)$$

Da odredimo konstante a, b, c, p, r, s , razvijemo t_2 i t_3 kao funkcije dviju varijabli u Taylorov red oko (x_k, y_k) . Traženu funkciju y razvijemo u Taylorov red (8) kao i prije. Zatim se izjednače koeficijenti uz iste potencije od h do potencije 3, što je u biti isti izvod kao i prije. Opet dobivamo manje jednadžbi nego nepoznanica koje glase:

$$\begin{aligned} a + b + c &= 1, \\ bp + cr &= \frac{1}{2}, \\ bp^2 + cr^2 &= \frac{1}{3}, \\ cps &= \frac{1}{6}. \end{aligned}$$

Odabiranjem dviju konstanti ostale su određene. Jedan izbor konstanti, koji je izvršio Kutta, daje sljedeći algoritam trećeg reda

$$\begin{aligned} y_{k+1} &= y_k + \frac{h}{6}(t_1 + 4t_2 + t_3) \\ t_1 &= f(x_k, y_k), \\ t_2 &= f\left(x_k + \frac{1}{2}h, y_k + \frac{1}{2}ht_1\right), \\ t_3 &= f(x_k + h, y_k + 2ht_2 - ht_1). \end{aligned} \quad (15)$$

(Ako f ovisi samo o x onda dobivamo Simpsonovu formulu.)

Svi algoritmi četvrtog reda polaze od

$$y_{k+1} = y_k + h(at_1 + bt_2 + ct_3 + dt_4) \quad (16)$$

gdje su t_1, t_2, t_3, t_4 aproksimacije derivacije u nekim točkama segmenta $[x_k, x_{k+1}]$. U primjenama se koristi nekoliko algoritama četvrtog reda. Sljedeći se pripisuje Kuttii:

$$\begin{aligned} y_{k+1} &= y_k + \frac{h}{6}(t_1 + 2t_2 + 2t_3 + t_4), \\ t_1 &= f(x_k, y_k), \\ t_2 &= f(x_k + \frac{1}{2}h, y_k + \frac{1}{2}ht_1), \\ t_3 &= f(x_k + \frac{1}{2}h, y_k + \frac{1}{2}ht_2) \\ t_4 &= f(x_k + h, y_k + ht_3). \end{aligned} \quad (17)$$

(17) se također svodi na Simpsonovu formulu ako je f funkcija samo od x . Jedan drugi algoritam četvrtog reda koji se također pripisuje Kuttii glasi

$$\begin{aligned} y_{k+1} &= y_k + \frac{h}{8}(t_1 + 3t_2 + 3t_3 + t_4), \\ t_1 &= f(x_k, y_k), \\ t_2 &= f(x_k + \frac{1}{3}h, y_k + \frac{1}{3}ht_1), \\ t_3 &= f(x_k + \frac{2}{3}h, y_k - \frac{1}{3}ht_1 + ht_2), \\ t_4 &= f(x_k + h, y_k + ht_1 - ht_2 + ht_3). \end{aligned} \quad (18)$$

Često se upotrebljava i slijedeći algoritam četvrtog reda a pripisuje se matematičaru Gillu.

$$\begin{aligned} y_{k+1} &= y_k + \frac{h}{6} \left[t_1 + 2\left(1 - \frac{1}{\sqrt{2}}\right)t_2 + 2\left(1 + \frac{1}{\sqrt{2}}\right)t_3 + t_4 \right], \\ t_1 &= f(x_k, y_k), \\ t_2 &= f\left(x_k + \frac{1}{2}h, y_k + \frac{1}{2}ht_1\right), \\ t_3 &= f\left(x_k + \frac{1}{2}h, y_k + \left(-\frac{1}{2} + \frac{1}{\sqrt{2}}\right)ht_1 + \left(1 - \frac{1}{\sqrt{2}}\right)ht_2\right), \\ t_4 &= f\left(x_k + h, y_k - \frac{1}{\sqrt{2}}ht_2 + \left(1 + \frac{1}{\sqrt{2}}\right)ht_3\right). \end{aligned} \quad (19)$$

Spomenimo da se u literaturi može naći više algoritama petog reda u što se ovdje ne upuštamo.

Istaknimo da se za sve Runge-Kuttine algoritme može pokazati da konvergiraju, tj. da vrijedi

$$\lim_{h \rightarrow 0} (y_k - y(x_k)) = 0.$$

To nam kaže da uz manji h dobivamo bolju aproksimaciju. Što se tiče ocjene pogreške, može se pokazati (u što se ne upuštamo) da je za algoritam i -tog reda pogreška $\leq C_i h^i$, međutim, praktički je gotovo nemoguće naći konstante C_i . Na taj način obično ne

znamo kako malen h treba uzeti da postignemo željenu točnost. Jedan način kako si pomažemo je sljedeći. Izaberemo jedan h i izračunamo y_1, y_2, \dots, y_n . Tada ponovimo račun za $\frac{h}{2}$ i usporedimo rezultate. Ako su razlike malene onda smo obično zadovoljni rezultatom, a ako su velike, onda moramo ponoviti postupak sa manjim h , dok ne postignemo zadovoljavajući rezultat.

Spomenimo još jedan kriterij za Runge-Kuttin algoritam (17) koji potječe od Collatza, a sastoji se u sljedećem: Poslije svakog koraka izračuna se

$$\left| \frac{t_3 - t_2}{t_2 - t_1} \right|.$$

Ako taj omjer postane velik (veći od nekoliko stotina), onda se h mora smanjiti. To je sasvim kvalitativan indikator, međutim, vrlo je spretan jer je dodatno računanje zanemarivo.

KAZALO

- Aitkenova šema, 5
 Aitkenov algoritam, 5
 algebarska jednadžba, 148
 apsolutna norma, 90
 brzi Fourierovi transformati, 28
 brzo oscilirajuća, 60
 Budan-Fourierov teorem, 154
 Bunemanova metoda, 78
 Choleskyjeva šema, 76
 Cooley-Tukeyeva metoda, 29
 Cotesovi koeficijenti, 43
 Čebiševljevi polinomi, 31
 Čebiševsko ekonomiziranje, 35
 čvor, 2
 Danilevskijeva metoda, 116
 Descartesovo pravilo, 153
 determinanta, 69
 diferencija unaprijed, 11
 diferencija unatrag, 12
 dijagonalno dominantan sustav, 96
 — po recima, 96
 — po stupcima, 96
 direktnе metode, 66
 dominantna svojstvena vrijednost, 129
 duljina, 89
 ekstrapolacija, 5
 euklidска norma, 90, 91
 Eulerova metoda, 185
 fazni interpolacijski polinom, 23
 fiksna točka, 179
 Filonova formula, 61
 Frobeniusov oblik, 116
 Gauss – Jordanova metoda, 70
 Gaussova metoda, 66
 Gaussove formule, 57
 Gauss – Seidelova metoda, 98
 glavni element, 71
 granice rješenja algebarskih jednadžbi, 149
 Heunova metoda, 190
 indeks polinoma s obzirom na pravac, 158
 interpolacija, 5
 interpolacijski polinom, 1
 — Lagrangeov oblik, 2
 — Newtonov oblik s diferencijama unaprijed, 12
 — Newtonov oblik s diferencijama unatrag, 12
 — Newtonov oblik s kvocijentima diferencija, 10
 interpolacijski splajn, 18
 inverzna interpolacija, 16
 inverzna matrična, 69, 71
 iteracijske metode, 84
 iterirana linearna interpolacija, 5
 Jacobijeva metoda, 95
 Jordanova forma, 87
 Jordanove elementarne klijetke, 87
 Jordanovi blokovi, 87
 karakteristična jednadžba, 87
 karakteristični polinom, 116, 124, 127
 kompleksna rješenja, 157, 166
 kontrakcija, 179
 konzistentnost normi, 93
 korak i interpolacije, 4
 Krilovljeva metoda, 124
 kubni splajn, 19
 kvadratična konvergencija, 143
 kvocijent diferencija prvog reda, 7
 kvocijent diferencija n -toga reda, 8
 Lagrangeovi koeficijenti, 4
 L_2 -aproksimacija, 39
 Legendreovi polinomi, 54
 Leverrierova metoda, 127
 linearna iteracija, 84
 $\| \cdot \|_p$ norma, 90
 LU (LR) faktorizacija, 75
 maksimum norma, 91
 metoda iteracije, 137, 174, 178
 metoda neodređenih koeficijenata, 115
 metoda raspolažanja, 136
 metoda relaksacije, 101
 metoda sekante, 144
 metoda suksesivne aproksimacije, 180
 metoda suksesivne nadrelaksacije, 104

-
- metoda tangente, 140
 - minimaks polinom, 34
 - minimaks princip, 34
 - Misesova metoda, 128
 - modificirana Eulerova metoda, 191
 - nelinearne jednadžbe, 134
 - nestabilan sustav, 102
 - Newton-Cotesova formula, 43
 - Newtonova metoda, 140, 164, 169
 - Newtonove formule, 127
 - Newtonov iteracijski postupak, 141, 166
 - Newtonov test, 152
 - Newton – Raphsonova metoda, 141
 - norma, 89
 - normalni Frobeniusov oblik, 116
 - norma matrice, 90
 - normalni sustav, 184
 - norma vektora, 89
 - normirani oblik polinoma, 148
 - numeričko integriranje, 40
 - optimalni aproksimacijski polinom, 34
 - pivot element, 71
 - poboljšana Eulerova metoda, 190
 - pojednostavljena Newtonova metoda, 174
 - polinom najmanjih kvadrata, 37
 - pozitivno definitan sustav, 99
 - racionalna interpolacija, 17
 - redna norma, 91
 - regula falsi, 144
 - relaksacija, 101
 - relaksacijski parametar, 104
 - Richardsonova ekstrapolacija, 51
 - Rombergov algoritam, 50
 - Rombergova tablica, 53
 - r -ta diferencija unaprijed, 11
 - r -ta diferencija unatrag, 12
 - Runge – Kuttini postupci, 188
 - Simpsonova formula, 47
 - s -kratno rješenje, 148
 - slabo uvjetovan sustav, 102
 - sličnost matrica, 112
 - SOR metoda, 104
 - spektar matrice, 128
 - spektralni radijus, 87, 129
 - splajn, 19
 - stupčana norma, 91
 - Sturmov niz, 155
 - sustavi linearnih jednadžbi, 65
 - sustavi nelinearnih jednadžbi, 164
 - svojstveni vektor, 87, 110, 120, 125
 - svojstvena vrijednost, 87, 110
 - teleskopiranje reda potencija, 35
 - težinska funkcija, 58
 - trapezna formula, 45
 - trigonometrijski interpolacijski polinom, 23
 - upotrebljivost rješenja, 82
 - Viéta-ove formule, 149
 - vodeći element, 71

LITERATURA

- [1] N. S. Bahvalov, *Čislenije metodi I*, Nauka, Moskva, 1975.
- [2] I. S. Berezin, N. P. Zhidkov, *Computing Methods I, 2*, Pergamon Press, Oxford, 1965.
- [3] B. Carnahan, H. A. Luther, J. O. Wilkes, *Applied Numerical Methods*, J. Wiley, New York, 1969.
- [4] B. P. Demidovič, I. A. Maron, *Computational Mathematics*, Mir Publishers, Moscow, 1987.
- [5] B. P. Demidovich, I. A. Maron, *Osnovi vičisliteljnoj matematiki*, Nauka, Moskva, 1966.
- [6] N. Elezović, *Linearna algebra*, Element, Zagreb, 1995.
- [7] F. B. Hildebrand, *Introduction to Numerical Analysis*, Mc Graw – Hill, New York, 1956.
- [8] P. Javor, *Matematička analiza I*, Element, Zagreb, 1995.
- [9] S. Kurepa, *Uvod u linearnu algebru*, Školska knjiga, Zagreb, 1975.
- [10] S. Kurepa, *Matematička analiza I, II, III*, Tehnička knjiga, Zagreb, 1975., 1971., 1975.
- [11] G. V. Milovanović, *Numerička analiza I*, Naučna knjiga, Beograd, 1988.
- [12] C. E. Pearson Ed., *Handbook of Applied Mathematics, Selected Methods and Results*, Van Nostrand, New York, 1974.
- [13] J. Stoer, R. Burlish, *Introduction to Numerical Analysis*, Springer–Verlag, Berlin, 1980.