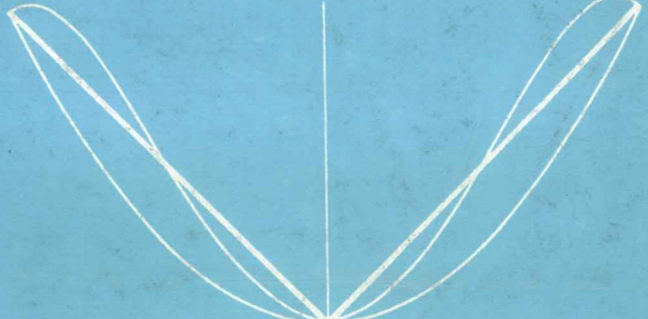
УНИВЕРЗИТЕТ У НИШУ

gradimir v. milovanović

NUMERIČKA ANALIZA

I deo



NIŠ, 1979.

UNIVERZITET U NIŠU

Dr Gradimir V. Milovanović

docent Elektronskog fakulteta

NUMERIČKA ANALIZA

I deo

NIŠ, 1979.

Recenzenti:
Dr PETAR M. VASIĆ, redovni profesor Elektrotehničkog fakulteta u Beogradu
Dr RADOSAV Ž. ĐORĐEVIĆ, redovni profesor Elektronskog fakulteta u Nišu
Dr DOBRILO Đ. TOŠIĆ, vanredni profesor Elektrotehničkog fakulteta u Beogradu

Savet Univerziteta u Nišu na svojoj sednici od 22. 06. 1979. godine doneo je odluku o štampanju ovog udžbenika

Izdavač: Univerzitet u Nišu
Tiraž: 1.000 primeraka

Štampa: Štamparija „Bakar" – Bor

P R E D G O V O R

Ovaj udžbenik numeričke analize je proistekao iz predavanja koja je autor poslednjih godina držao studentima IV godine Elektronskog fakulteta u Nišu, u okviru predmeta Numerička analiza I. Uvidom u noviju literaturu iz ove oblasti učinjen je pokušaj da se obrade numerički metodi koji nalaze široku primenu kod rešavanja različitih naučno-tehničkih problema na elektronskim računskim sistemima.

Knjiga je podeljena u sedam glava, od kojih su prve dve uvodne. U preostalih pet glava obradjuje se opšta teorija iterativnih procesa, metodi u linearnoj algebri, metodi za rešavanje nelinearnih jednačina i sistema nelinearnih jednačina, interpolacija i aproksimacija funkcija i na kraju, numeričko diferenciranje i numerička integracija. Sve glave su podeljene na poglavlja, a poglavlja na odeljke, izuzimajući prvo poglavlje u prvoj glavi. Numeracija objekata (formula, teorema, definicija i sl.) u okviru jednog odeljka izvršena je pomoću tri broja od kojih prvi ukazuje na poglavlje, drugi na odeljak, a treći na redni broj tog objekta u posmatranom odeljku. Na ovaj način je uspostavljena jednoznačna numeracija objekata u okviru jedne glave. U poslednjem poglavlju svake glave dat je spisak citirane i korišćene literature.

Namera je autora da u najskorije vreme da u štampu i II deo ove knjige, koji će obuhvatati numeričke metode za rešavanje običnih diferencijalnih jednačina (Cauchyev problem i konturni problemi), kao i metode za rešavanje parcijalnih diferencijalnih jednačina i integralnih jednačina. Posebna pažnja biće posvećena projekciono-varijacionim metodima, koji se u poslednje vreme naglo razvijaju.

Knjiga je prvenstveno namenjena studentima Elektronskog fakulteta u Nišu, medjutim, ona može biti od koristi i studentima drugih tehničkih i prirodno-matematičkih fakulteta, kao i svima onima koji u svojoj praksi koriste numeričke metode. Autor se nada da će publikovanje ove knjige ublažiti nedostatak udžbeničke literature na našem jeziku iz ove, danas veoma aktuelne matematičke discipline.

Autor se zahvaljuje profesorima dr Petru M. Vasiću, dr Radosavu Ž. Djordjeviću i dr Dobrilu Dj. Tošiću, koji su u svojstvu recenzenata, pročitali rukopis ove knjige i dali korisne sugestije, koje su uzete pri konačnoj verziji rukopisa.

U završnoj fazi pripreme rukopisa za štampu, autoru su pružili nesebičnu pomoć docent dr Ljubomir Stanković i saradnici mr Igor Milovanović, Ljubiša Kocić i Milan Kovačević, na čemu im autor duguje zahvalnost.

Takodje, autor se zahvaljuje svojoj supruzi Dobrili na razumevanju i podstreku da se ovaj, ne mali posao privede kraju.

Veći deo rukopisa je otkucao i tehnički uredio za štampu sam autor, tako da eventualni propusti i greške padaju na njegov račun. Jedan deo rukopisa otkucala je Andjelka Milinčić, daktilograf-velemajstor, na čemu joj autor duguje zahvalnost.

U Nišu, 15.12.1978. Gradimir V. Milovanović

S A D R Ž A J

1. UVOD U NUMERIČKU MATEMATIKU

1.1. ZADATAK NUMERIČKE MATEMATIKE

Savremena nauka i tehnika postavljaju niz matematičkih problema koji se klasičnim matematičkim metodima ne mogu uvek tačno rešiti ili bi njihovo rešavanje bilo suviše glomazno, s obzirom da se najčešće zahteva numerički rezultat. Na primer, poznato je da se algebarske jednačine u opštem slučaju mogu rešiti pomoću radikala, samo kada su stepena ne višeg od četvrtog. Medjutim, u praksi se vrlo često sreće problem rešavanja numeričkih algebarskih jednačina* visokog stepena. Takodje, često se javlja i problem rešavanja sistema linearnih algebarskih jednačina sa stotinu, pa i hiljadu nepoznatih ili, pak, rešavanje diferencijalnih jednačina koje se ne mogu integraliti pomoću elementarnih funkcija. Posebno važnu ulogu imaju problemi koji se svode na rešavanje parcijalnih diferencijalnih jednačina.

U eksperimentalnim naukama skoro uvek se zahteva statistička obrada rezultata merenja koja uključuje i utvrdjivanje (a nekad i otkrivanje) zavisnosti izmedju pojedinih parametara u eksperimentu.

S druge strane, čak i u najelementarnijim problemima, egzaktno rešenje koje je izraženo simbolički ne zadovoljava potrebe, s obzirom da se traži numerički rezultat. Ilustrujmo ovo prostim primerom.

*) Algebarska jednačina sa numeričkim koeficijentima.

Neka je potrebno odrediti pozitivan koren jednačine

$$x^2 - a = 0 \qquad (a > 0). \tag{1.1.1}$$

Traženo egzaktno rešenje je $x = \sqrt{a}$. Medjutim, simbol $\sqrt{\ }$ ne rešava problem, jer ne daje postupak izračunavanja broja x.

Formirajmo sada niz $\{x^{(k)}\}$, pomoću

$$x^{(0)} = a, \quad x^{(k)} = \frac{1}{2}\Big(x^{(k-1)} + \frac{a}{x^{(k-1)}}\Big) \quad (k=1,2,\ldots). \tag{1.1.2}$$

Može se pokazati da niz $\{x^{(k)}\}$ teži vrednosti $\sqrt{a}$ kada $k \to +\infty$. Pri nekom konačnom n imamo $x \simeq x^{(n)}$.

Postupak (1.1.2), kao što vidimo, omogućava nalaženje rešenja problema (1.1.1) sa tačnošću, koja zavisi od vrednosti a i n. Kako se u praktičnim izračunavanjima zahteva odredjena tačnost ε, to se za x usvaja vrednost $x^{(n)}$ koja zadovoljava uslov $|x - x^{(n)}| \leqq \varepsilon$.

Navedimo još jedan primer. Neka je potrebno odrediti vrednost sin x za x = 0.5 rad. Simbol sin, i u ovom slučaju, ne daje postupak za rešavanje problema. Da bismo izračunali traženu vrednost, funkciju $x \mapsto \sin x$ možemo da razvijemo, na primer, u Taylorov red

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots, \tag{1.1.3}$$

odakle dobijamo

$$\sin x \simeq x - \frac{x^3}{3!} = 0.479167 \tag{1.1.4}$$

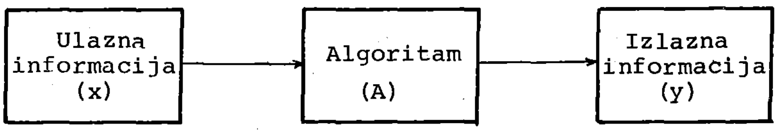
i

$$\sin x \simeq x - \frac{x^3}{3!} + \frac{x^5}{5!} = 0.479427. \tag{1.1.5}$$

Tačna vrednost za sin 0.5 sa šest decimala je 0.479425. Greške koje nastaju u (1.1.4) i (1.1.5) potiču usled uzimanja konačnog broja članova Taylorovog reda i nazivamo ih greškama odsecanja. Drugim rečima, greška nastaje iz razloga što rešavamo problem, različit od postavljenog, koji je znatno prostiji sa stanovišta računanja, a čije je rešenje u nekom smislu blisko rešenju postavljenog problema.

U opštem slučaju, problem koji treba rešavati zvaćemo ulaznom informacijom, a odgovarajući rezultat izlaznom informacijom.

Postupak transformacije ulazne informacije(x) u izlaznu informaciju(y) zvaćemo algoritmom(A). Navedena transformacija može se predstaviti blok dijagramom



ili simbolički $x \xrightarrow{A} y$.

Pri rešavanju nekog problema potrebno je izabrati pogodan algoritam koji najbrže dovodi do željenog rezultata. Ovo je naročito važno kod primene savremenih elektronskih računskih mašina, s obzirom na cenu mašinskog vremena. Izbor najracionalnijeg algoritma u prethodnom smislu, predstavlja vrlo složen problem, koji teorijski u opštem slučaju još uvek nije rešen.

Razradom i realizacijom algoritama i analizom greške u izlaznoj informaciji bavi se posebna oblast matematike, tzv. numerička matematika.

Centralni deo numeričke matematike čine numerički metodi. Oni moraju biti takvi da su pogodni sa stanovišta primene savremenih elektronskih računskih mašina. Kako računske mašine najčešće izvode samo četiri osnovne računske operacije (sabiranje, oduzimanje, množenje, deljenje), to numerički metodi moraju biti takvi, da se svode na konačan niz takvih operacija.

Na osnovu prethodnog, pogrešno je izvesti zaključak da se numerička matematika počela razvijati sa pojavom računskih mašina. Poznato je, naime, da se u delima matematičara prošlih stoleća, nalazi niz približnih metoda koji danas čine osnov numeričkoj matematici. Medjutim, nagli razvoj računske tehnike posle II svetskog rata, uslovio je brzi i sistematski razvoj numeričke matematike. Tako na primer, razradjen je niz približnih metoda u teoriji jednačina svih vrsta i za njih su dobijeni novi fundamentalni rezultati, kako teorijski, tako i praktični. Medju njima, posebno se ističu metodi za rešavanje graničnih problema u teoriji diferencijalnih jednačina (običnih i parcijalnih), metodi u teoriji integralnih i integro-diferencijalnih jednačina, kao i metodi za rešavanje apstraktnih operatorskih jednačina. Poslednjih godina dosta se radi na metodima za rešavanje graničnih problema u jednoj neiscrpnoj oblasti-teoriji nelinearnih parcijalnih jednačina. Takodje, vidno mesto zauzimaju i metodi numeričke integracije.

Medju numeričkim metodima vidno mesto zauzimaju iterativni metodi, s obzirom da su oni pogodni za primenu na računskim mašinama. Izučavanje numeričkih metoda uključuje analizu greške, stabilnost, konvergenciju (kod iterativnih metoda), kao i niz drugih svojstava. Zato se vrlo često ovaj deo numeričke matematike naziva numerička analiza.

U svom razvoju, numerička analiza se oslanja na više matematičkih disciplina od kojih u poslednje vreme vidno mesto zauzima funkcionalna analiza.

Danas u svetu postoji više specijalizovanih časopisa za numeričku matematiku. Navešćemo nekoliko od njih*:

1. Numer. Math. Numerische Mathematik. (Berlin-Göttingen-Heidelberg),
2. SIAM J. Numer. Anal. A Publication of the Society for Industrial and Applied Mathematics. (Philadelphia, Pa.),
3. Math. Comp. Mathematics of Computation. (Washington,D.C.),
4. Ž. Vyčisl. Mat. i Mat. Fiz. Akademija Nauk SSSR. Žurnal Vyčislitel'noi Matematiki i Matematičeskoi Fiziki. (Moscow) (Translated as: U.S.S.R.Comput.Math. and Math. Phys.),
5. Comput. J. The Computer Journal (London),
6. Comput. Bull. The Computer Bulletin. (London),
7. Internat. J. Comput. Math. International Journal of Computer Mathematics. (New York).

Na kraju napomenimo da se problemima praktične realizacije algoritama bavi oblast teorija programiranja koja se u poslednje vreme uspešno razvija. Jedan od njenih osnovnih zadataka je priprema i sastavljanje programa za računsku mašinu prema izabranom algoritmu.

---

* Naslovi časopisa su dati onako kako se citiraju u referativnom žurnalu Mathematical Reviews.

1.2. ELEMENTI TEORIJE GREŠAKA

1.2.1. Klasifikacija i analiza grešaka u numeričkom rešenju

U skoro svim numeričkim problemima, izlazna informacija, ili kako se češće kaže numeričko rešenje, praćeno je greškama čiji izvori mogu biti različiti. Pre nego što predjemo na klasifikaciju i analizu grešaka, uvešćemo neke pojmove i definicije.

Približan broj $\bar{x}$ je broj koji zamenjuje tačan broj x u izračunavanjima i neznatno se razlikuje od njega. Odgovarajuća greška* je

$$e = \bar{x} - x .$$

Definicija 2.1.1. Svaka cifra broja, izuzimajući nule koje služe za fiksiranje decimalne tačke, naziva se značajnom cifrom tog broja.

Primer 2.1.1. Svaki od navedenih brojeva

2.563, 15.32, 0.2687, 0.002649

ima četiri značajne cifre.

Svaki broj x može se predstaviti u normalizovanom obliku

$$x = (\pm 0.a_1 a_2 \ldots a_n a_{n+1} \ldots) b^k \qquad (a_1 \neq 0), \tag{2.1.1}$$

gde je b osnova brojnog sistema, a $a_i (i=1,2,\ldots)$ cifra brojnog sistema $(0 \leq a_i < b)$.

Broj $\pm 0.a_1 a_2 \ldots a_n a_{n+1} \ldots$ zvaćemo mantisom i označavati sa x^*. Broj k zvaćemo karakteristikom. Dakle, možemo pisati

$$x = x^* b^k .$$

Najčešće je u upotrebi binarni ili decimalni brojni sistem, tj. $b = 2$ ili $b = 10$. Jednostavnosti radi, u daljem tekstu, često ćemo uzimati $b = 10$.

*) Često se ova greška naziva apsolutnom greškom. Medjutim, ovo je pogrešno jer je apsolutna greška $|e| = |\bar{x} - x|$.

Definicija 2.1.2. Kaže se da $\bar{x}$ aproksimira broj x sa ℓ značajnih cifara ako je ℓ najveći broj cifara mantise, za koji $|\bar{x}^* - x^*|$ ne prelazi jedinicu ℓ-tog mesta.

Primer 2.1.2. Za brojeve $x = 25.245\ldots$ i $\bar{x} = 25.255$ važi

$$|\bar{x}^* - x^*| < 10^{-4},$$

pa $\bar{x}$ aproksimira x sa četiri značajne cifre.

U slučajevima kada se tačna vrednost x ne zna, uvodi se pojam granice apsolutne greške približnog broja.

Definicija 2.1.3. Pod granicom apsolutne greške Δ_x približnog broja x podrazumeva se svaki broj ne manji od apsolutne greške tog broja.

Kako je $|e| = |\bar{x} - x| \leqq \Delta_x$, imamo

$$\bar{x} - \Delta_x \leqq x \leqq \bar{x} + \Delta_x. \tag{2.1.2}$$

S obzirom da greška e nedovoljno karakteriše tačnost uvodi se i pojam relativne greške kao

$$r = \frac{e}{x} = \frac{\bar{x} - x}{x} \qquad (x \neq 0).$$

Slično se uvodi i pojam granice relativne greške

$$\varepsilon_x = \frac{\Delta_x}{|x|} \cong \frac{\Delta_x}{|\bar{x}|}.$$

Nejednakosti (2.1.2) tada postaju

$$\bar{x}(1-\varepsilon_x \operatorname{sgn}(\bar{x})) \leqq x \leqq \bar{x}(1+\varepsilon_x \operatorname{sgn}(\bar{x})).$$

U praktičnim izračunavanjima, primenom elektronskih računara, prinudjeni smo da radimo sa jednim vrlo uskim skupom brojeva. Naime, svaki realni broj oblika (2.1.1) koji se dobija kao rezultat odredjenih računskih operacija, zamenjuje se približnim brojem oblika

$$\bar{x} = (\pm 0.a_1a_2\ldots a_n)b^k \qquad (a_1 \neq 0).$$

U ovom slučaju kažemo da imamo mantisu sa n razreda.

Proces odbacivanja cifara mantise u broju x, počev od cifre a_{n+1} naziva se prosto odsecanje. Apsolutna greška pri ovome je

$$|e| \leqq b^{k-n}$$

Apsolutna greška, pri zameni broja x brojem $\bar{x}$, može se smanjiti ako se koristi tzv. postupak zakrugljivanja brojeva. Taj postupak se sastoji u sledećem:

1^o Ako je

$$a_{n+1} + a_{n+2}b^{-1} + \ldots < \frac{1}{2}b$$

koristi se prosto odsecanje;

2^o Ako je

$$a_{n+1} + a_{n+2}b^{-1} + \ldots > \frac{1}{2}b$$

cifra a_n se povećava za jedinicu, a cifre $a_{n+1}, a_{n+2},\ldots$ se odbacuju.

3^o Ako je

$$a_{n+1} + a_{n+2}b^{-1} + \ldots = \frac{1}{2}b$$

ravnopravno se mogu koristiti pravila 1^o i 2^o.

Na računskim mašinama zaokrugljivanje se najčešće izvodi dodavanjem broja $\frac{1}{2}b^{k-n}$ rezultatu x, a zatim se vrši prosto odsecanje Ovo znači da se u nerešenom slučaju 3^o uvek a_n zamenjuje sa a_n+1 (pravilo 2^o).

Napomenimo da se kod ručnog zaokrugljivanja brojeva u dekadnom sistemu ($b=10$) u nerešenom slučaju 3^o preporučuje sledeće pravilo:

Ako je cifra a_n paran broj koristiti pravilo 1^o, a ako je neparan broj koristiti pravilo 2^o.

Primer 2.1.3. Sukcesivno zaokrugljivanje broja $\pi=3.1415926535\ldots$ daje redom brojeve:

3.141592654,	3.1416,
3.14159265,	3.142,
3.1415926,	3.14,
3.141593,	3.1,
3.14159,	3.

Apsolutna greška kod zaokrugljenog broja je

$$|e| \leqq \frac{1}{2} b^{-n+k}$$

Pretpostavimo da je $b = 10$. S obzirom da je $x = x^* 10^k$ i $0.1 \leqq x^* < 1$, za apsolutnu relativnu grešku zaokrugljivanja, u tom slučaju, dobijamo ocenu

$$|r| = \frac{|e|}{|x|} \leqq \frac{\frac{1}{2}\cdot 10^{-n+k}}{|x|} = \frac{\frac{1}{2}\cdot 10^{-n+k}}{|x^*|\cdot 10^k} = \frac{10^{-n}}{2|x^*|},$$

tj.

$$|r| \leqq \frac{1}{2}\cdot 10^{-n+1} .$$

Po jednoj od mogućnih klasifikacija, greške se mogu podeliti na:

1. neotklonjive greške;
2. greške zaokrugljivanja;
3. greške metoda.

U grupu neotklonjivih grešaka ulaze greške koje potiču od netačnosti ulaznih podataka. Greške ulaznih podataka (ulazne informacije) mogu se u nekim slučajevima drastično manifestovati u izlaznoj informaciji i pri uslovu da algoritam ne unosi grešku.

<u>Primer 2.1.4.</u> Sistem jednačina (ulazna informacija)

$$2x + 6y = 8,$$
$$2x + 6.0001y = 8.0001$$

ima rešenja (izlazna informacija) $x=1$, $y=1$. Ako se koeficijenti druge jednačine neznatno promene, tj. uzme se jednačina

$$2x + 5.99999y = 8.00002,$$

rešenja su $x = 10$, $y = -2$.

Usled zaokrugljivanja medjurezultata u procesu računanja nastaju greške zaokrugljivanja.

Greške metoda se javljaju usled toga što se u numeričkoj matematici obično dati problem zamenjuje drugim koji je lakši za računanje, a čije je rešenje u izvesnom smislu blisko rešenju datog problema.

<u>Primer 2.1.5.</u> Integral $\int_a^b f(x)dx$ može se približno izračunati, na primer, zamenom funkcije f nekim polinomom P na segmentu $[a,b]$, koji je u nekom smislu blizak datoj funkciji. Medjutim, moguće je za približno izračunavanje datog integrala koristiti konačnu sumu

$$\sum_{i=1}^{n} f(x_i)\Delta x_i .$$

U oba slučaja čine se greške metoda.

Greške odsecanja su takodje greške metoda.

Zbir svih grešaka čini **totalnu** grešku.

1.2.2. Prostiranje grešaka

Kod izvodjenja operacija sa približnim brojevima postavljaju se sledeći zadaci:

1. Oceniti tačnost rezultata kada je poznata tačnost početnih podataka i obrnuto. Naime, na osnovu zadate tačnosti rezultata treba odrediti kakvu tačnost treba da imaju početni podaci.

2. Uskladiti tačnost početnih podataka ako su neki od njih zadati suviše grubo, da bi se izbegao izlišan numerički rad.

3. U procesu računanja održati tačnost medjurezultata, kako bi se došlo do konačnog rezultata sa željenom tačnošću.

Izvršimo najpre analizu greške u rezultatu $x*y$, pri zameni brojeva x i y približnim brojevima $\bar{x}$ i $\bar{y}$ respektivno, a pri uslovu da se operacija $* \in \{+,-,\times,\div\}$ izvršava tačno.

Dakle, pretpostavimo da su greške približnih brojeva $\bar{x}$ i $\bar{y}$ redom e_x i e_y, tj. $e_x=\bar{x}-x$ i $e_y=\bar{y}-y$, i odredimo grešku e_{x*y}.

1. Kod sabiranja imamo

$$\bar{x}+\bar{y}=(x+e_x)+(y+e_y)=(x+y)+(e_x+e_y),$$

tj. greška zbira e_{x+y} je

$$e_{x+y}=e_x+e_y.$$

2. Kod oduzmanja, slično dobijamo

$$e_{x-y}=e_x-e_y.$$

3. Za množenje važi

$$\bar{x}\bar{y}=(x+e_x)(y+e_y)=xy+ye_x+xe_y+e_xe_y,$$

tj.

$$e_{xy}=ye_x+xe_y,$$

s obzirom da su greške e_x i e_y obično mnogo manje od samih veličina x i y.

4. Kod deljenja imamo

$$\frac{\bar{x}}{\bar{y}} = \frac{x+e_x}{y+e_y} = \frac{x+e_x}{y} \; \frac{1}{1+\frac{e_y}{y}} \; .$$

Kako je $\left|\frac{e_y}{y}\right| << 1$, iz poslednje jednakosti sleduje

$$\frac{\bar{x}}{\bar{y}} = \frac{x+e_x}{y}\left(1 - \frac{e_y}{y} + \left(\frac{e_y}{y}\right)^2 - \ldots\right)$$

$$\simeq \frac{x}{y} + \frac{e_x}{y} - \frac{x}{y^2}\, e_y ,$$

tj.

$$e_{\frac{x}{y}} \simeq \frac{1}{y}\, e_x - \frac{x}{y^2}\, e_y ,$$

pri čemu su uzete pretpostavke kao kod množenja.

Odgovarajuće relativne greške su redom

$$r_{x+y} = \frac{e_{x+y}}{x+y} = \frac{x}{x+y}\, r_x + \frac{y}{x+y}\, r_y , \tag{2.2.1}$$

$$r_{x-y} = \frac{e_{x-y}}{x-y} = \frac{x}{x-y}\, r_x - \frac{y}{x-y}\, r_y , \tag{2.2.2}$$

$$r_{xy} = \frac{e_{xy}}{xy} \simeq r_x + r_y , \tag{2.2.3}$$

$$r_{\frac{x}{y}} = \frac{e_{\frac{x}{y}}}{\frac{x}{y}} \simeq r_x - r_y , \tag{2.2.4}$$

gde su $r_x = \frac{e_x}{x}$ i $r_y = \frac{e_y}{y}$.

<u>Primedba 2.2.1.</u> S obzirom da se tačne vrednosti x i y najčešće ne znaju, to se u izvedenim izrazima za greške koriste odgovarajuće približne vrednosti $\bar{x}$ i $\bar{y}$.

U dosadašnjem izlaganju videli smo kako se greške brojeva x i y manifestuju u rezultatu $x*y$, pod pretpostavkom da se operacija $*$ izvodi tačno. Medjutim, kod primene računskih mašina ova pretpostavka nije uvek tačna i iz tog razloga u rezultatu se javlja greška koju nazivamo mašinskom greškom. Drugim rečima, mašinske greške su posledica pseudoaritmetičkih operacija koje se u mašini izvode.

U daljem izlaganju pretpostavićemo da mašina radi sa pokretnom tačkom i n-razrednom mantisom. U tom slučaju, u memoriji mašine, kao što je rečeno u prethodnom odeljku, svaki broj x aproksimira se postupkom zaokrugljivanja, standardnim brojem

$$\bar{x} = \bar{x}^* b^k , \tag{2.2.5}$$

gde je $\bar{x}^* = \pm 0.a_1 a_2 \ldots a_n$ $(a_1 \neq 0)$ mantisa, a k karakteristika (eksponent) broja $\bar{x}$. Broj nula ima nestandardno predstavljanje u kome je $\bar{x}^*=0$ i $k=0$. Jednostavnosti radi ovde nismo uveli pretpostavku o ograničenju broja cifarskih mesta karakteristike k, a koja je realno uvek prisutna*. Nadalje, pretpostavimo da je $b=10$.

Jednakost**

$$\bar{x} = fl(\bar{x}_1 * \bar{x}_2) = \bar{x}_1 \circledast \bar{x}_2$$

označava da su $\bar{x}_1, \bar{x}_2, \bar{x}$ standardni brojevi sa pokretnom tačkom i da je $\bar{x}$ dobijeno iz $\bar{x}_1$ i $\bar{x}_2$ operacijom $\dot{*} \in \{+,-,\times,\div\}$. Drugim rečima, važi

---

*) Neka je m broj cifarskih mesta predvidjenih za predstavljanje karakteristike, uključujući i mesto za sgn k. Tada je moguće, u standardnom obliku predstavljati samo brojeve kod kojih je $-(k_m+1) \leq k \leq k_m$, gde k_m zavisi od m i b. Brojevi kod kojih je $k > k_m$ ne mogu biti predstavljeni u memoriji računara i u tom slučaju kažemo da postoji prekoračenje kapaciteta memorijskog registra. Za brojeve kod kojih je $k < -(k_m+1)$, može se dozvoliti nestandardno predstavljanje u kome je $k=-(k_m+1)$, a pretpostavka $a_1 \neq 0$ se odbacuje. Na osnovu poslednjeg, brojevi kod kojih je $k < -(k_m+n)$ tretiraju se kao nula (videti [1]).

**) fl je skraćenica od reči floating point (pokretna tačka). Kod mašina sa fiksnom tačkom (fixed point) odgovarajuća oznaka je fi.

$$fl(\bar{x}_1 * \bar{x}_2) = (\bar{x}_1 * \bar{x}_2)(1+r),$$

gde je r relativna greška zaokrugljivanja rezultata $\bar{x}_1 * \bar{x}_2$ na standardni oblik sa n-razrednom mantisom, pri čemu je

$$|r| \leqq \frac{1}{2}\cdot 10^{-n+1}.$$

Veličinu

$$e_M = fl(\bar{x}_1 * \bar{x}_2) - (\bar{x}_1 * \bar{x}_2) = (\bar{x}_1 * \bar{x}_2) r$$

nazivamo mašinskom greškom, a odgovarajuću relativnu grešku zaokrugljivanja r relativnom mašinskom greškom.

<u>Primer 2.2.1</u>. Neka je $n=4$ i $\bar{x}_1 = 0.3947\cdot 10^4$ i $\bar{x}_2 = 0.1372\cdot 10^2$. Ako se operacija sabiranja izvodi u akumulatoru, u kome je za predstavljanje mantise brojeva obezbedjeno 2n razreda (double-precision accumulator) imamo

$$\begin{array}{r} 0.39470000\cdot 10^4 \\ +\ 0.00137200\cdot 10^4 \\ \hline 0.39607200\cdot 10^4, \end{array}$$

tj. $fl(\bar{x}_1 + \bar{x}_2) = 0.3961\cdot 10^4$, pri čemu je mašinska greška -0.28. Odgovarajuća relativna greška je $-0.71\cdot 10^{-4}$.

Za pseudoaritmetičke operacije, u opštem slučaju, ne važi asocijativni zakon, tj.

$$fl(fl(\bar{x} * \bar{y}) * \bar{z}) \neq fl(\bar{x} * fl(\bar{y} * \bar{z})).$$

Na primer, ako je $*$ operacija sabiranja, imamo

$$\begin{aligned} fl(fl(\bar{x}+\bar{y})+\bar{z}) &= fl((\bar{x}+\bar{y})(1+r_1)+\bar{z}) \\ &= ((\bar{x}+\bar{y})(1+r_1)+\bar{z})(1+r_2) \end{aligned}$$

i

$$\begin{aligned} fl(\bar{x}+fl(\bar{y}+\bar{z})) &= fl(\bar{x}+(\bar{y}+\bar{z})(1+r_3)) \\ &= (\bar{x}+(\bar{y}+\bar{z})(1+r_3))(1+r_4) \end{aligned}$$

Razmotrićemo sada opšti slučaj. Neka $* \in \{+,-,\times,\div\}$ i neka je

(2.2.6) $$u = x*y \ .$$

Kao što smo ranije videli (jednakosti (2.2.1)-(2.2.4)) relativna greška rezultata operacije $*$ može se predstaviti u obliku

$$r_{x*y} = a_x r_x + a_y r_y \ ,$$

gde koeficijenti a_x i a_y zavise od x i y i operacije $*$ (videti tabelu 2.2.1).

Tabela 2.2.1

$*$	a_x	a_y
$+$	$\frac{x}{x+y}$	$\frac{y}{x+y}$
$-$	$\frac{x}{x-y}$	$\frac{-y}{x-y}$
$\times$	1	1
$\div$	1	-1

Kako je

$$\bar{x}*\bar{y} = (x*y)(1+r_{x*y}) \ ,$$

zbog prisustva mašinske greške, imamo

$$\bar{u} = fl(\bar{x}*\bar{y}) = (x*y)(1+r_{x*y})(1+r) \ ,$$

gde je r relativna mašinska greška kod izvodjenja operacije $*$.

Na osnovu poslednje jednakosti za totalnu grešku e_u^t, dobijamo

$$e_u^t = \bar{u} - u = u(r_{x*y} + r + r \cdot r_{x*y}) \ .$$

Kako je, najčešće, $r \cdot r_{x*y}$ mnogo manje od r_{x*y} i r, u daljem razmatranju koristićemo približnost

$$e_u^t \cong e_u^T = u(r_{x*y} + r) \ .$$

Odgovarajuća relativna greška je

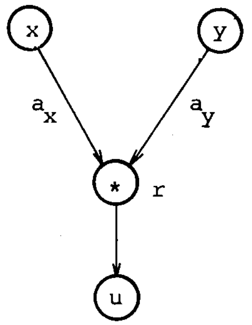
$$r_u^T = \frac{e_u^T}{u} = r_{x*y} + r \ ,$$

tj.

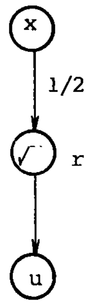
$$r_u^T = a_x r_x + a_y r_y + r. \tag{2.2.7}$$

Radi nalaženja totalne greške nekog numeričkog postupka i njene analize, predstavljaćemo numeričke postupke pomoću grafova. Kako se numerički postupak sastoji iz konačnog broja elementarnih operacija, dovoljno je znati kako se elementarna računska operacija predstavlja pomoću grafa.

Graf računske operacije (2.2.6)(sl.2.2.1) u simboličkom obliku sadrži jednakost (2.2.7). Relativna greška zaokrugljivanja r_1 rezultata operacije $*$, upisuje se pored temena grafa koje označava operaciju $*$. Smisao koeficijenata potega u grafu je u tome da relativna greška operanda ulazi u rezultat operacije pomnožena koeficijentom potega.



Sl. 2.2.1



Sl. 2.2.2

Primedba 2.2.2. Unarne operacije mogu se takodje prikazati pomoću grafa. Pokazaćemo to na primeru $u=\sqrt{x}$, mada, kao što smo videli u poglavlju 1.1, simbol $\sqrt{\ }$ ne daje numerički postupak za izračunavanje vrednosti $\sqrt{x}$. Medjutim, ova činjenica ne utiče na mogućnost analize greške. Naime, koristeći standardna označavanja, imamo

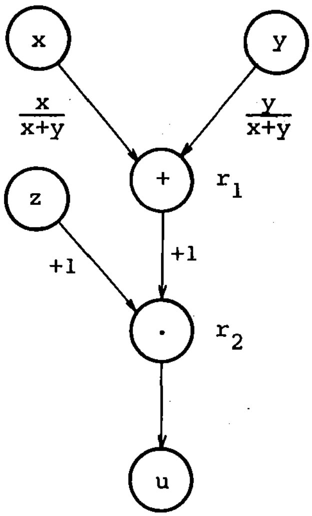
$$\bar{u} = \sqrt{\bar{x}}\,(1+r) = \sqrt{x(1+r_x)}\,(1+r) \cong \sqrt{x}(1+ \tfrac{1}{2}r_x)(1+r),$$

tj. $r_u^T = \frac{1}{2}r_x + r$. Na osnovu ove jednakosti dobija se graf unarne operacije $u = \sqrt{x}$ (sl.2.2.2). Relativna greška r, najčešće, nije veća od 10^{-n+1}, gde je n broj cifara mantise(videti [2]).

Primer 2.2.2. Ako su poznate približne vrednosti brojeva x, y,z sa relativnim greškama zaokrugljivanja r_x, r_y, r_z respektivno, odredićemo relativnu grešku u rezultatu

$$u = (x+y)z.$$

Neka su relativne mašinske greške operacija sabiranja i množenja redom r_1 i r_2. Graf odgovarajućeg računskog postupka dat je na sl. 2.2.3. Na osnovu grafa imamo redom



Sl. 2.2.3

$$r^T_{x+y} = \frac{x}{x+y} r_x + \frac{y}{x+y} r_y + r_1,$$

$$r^T_u = 1\cdot r^T_{x+y} + 1\cdot r_z + r_2$$

$$= \frac{x}{x+y} r_x + \frac{y}{x+y} r_y + r_z + r_1 + r_2.$$

Kako su sve relativne greške zaokrugljivanja po aposlutnoj vrednosti manje od $\frac{1}{2}\cdot 10^{-n+1}$ (n - broj cifara mantise), dobijamo ocenu

(2.2.8) $$|r^T_u| \leqq \left(\left|\frac{x}{x+y}\right| + \left|\frac{y}{x+y}\right| + 3\right)\frac{1}{2}10^{-n+1}.$$

Ako su brojevi x i y istog znaka, (2.2.8) se svodi na

$$|r^T_u| \leqq 2\cdot 10^{-n+1}.$$

Primer 2.2.3. Izvršimo sada analizu greške kod izračunavanja zbira

(2.2.9) $$y = x_1 + x_2 + x_3 + x_4,$$

pri čemu je $0<x_1<x_2<x_3<x_4$. Jednostavnosti radi, pretpostavimo da su brojevi x_i $(i = 1, 2, 3, 4)$ zadati tačno. Neka su relativne mašinske greške posle svake operacije sabiranja redom r_1, r_2, r_3. Graf računskog postupka (2.2.9) dat je na sl. 2.2.4. Kako je $r_{x_i} = 0$ $(i = 1, 2, 3, 4)$, na osnovu grafa dobijamo

redom

$$r^T_{x_1+x_2} = r_1,$$

$$r^T_{x_1+x_2+x_3} = \frac{x_1+x_2}{x_1+x_2+x_3} r_1+r_2,$$

$$r^T_y = \frac{x_1+x_2+x_3}{x_1+x_2+x_3+x_4} \left(\frac{x_1+x_2}{x_1+x_2+x_3} r_1 + r_2\right) + r_3,$$

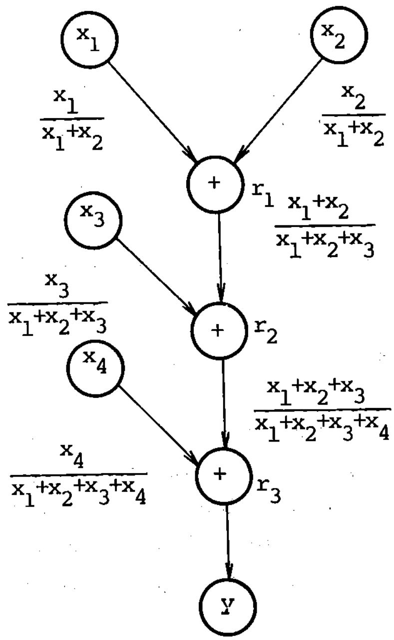
odakle je

$$e^T_y = (x_1+x_2)r_1+(x_1+x_2+x_3)r_2+(x_1+x_2+x_3+x_4)r_3. \tag{2.2.10}$$

S obzirom na $|r_i| \leq \frac{1}{2}\cdot 10^{-n+1}$ (n-broj cifara mantise), iz (2.2.10) sleduje

$$|e^T_y| \leq (3x_1+3x_2+2x_3+x_4)\ \frac{1}{2}\cdot 10^{-n+1},$$

odakle zaključujemo da je granica apsolutne greške rezultata y minimalna ukoliko se sabiranje izvodi polazeći od najmanjih brojeva.



Sl. 2.2.4

Slično se može pokazati da kod sabiranja m pozitivnih brojeva $x_1, \dots, x_m$ važi ocena

$$e_y^T = \left[(m-1)x_1 + (m-1)x_2 + (m-2)x_3 + \dots + 2x_{m-1} + x_m\right]\frac{1}{2}\cdot 10^{-n+1}.$$

Primer 2.2.4. Nadjimo zbir brojeva

$$x_1 = 0.1376\cdot 10^0, \quad x_2 = 0.4737\cdot 10^0, \quad x_3 = 0.7428\cdot 10^1,$$
$$x_4 = 0.6439\cdot 10^2, \quad x_5 = 0.5763\cdot 10^3, \quad x_6 = 0.2034\cdot 10^4,$$

zaokrugljujući sve medjurezultate na četiri značajne cifre.

Sabiranjem brojeva u datom redosledu, imamo redom

$$u_1 = x_1 = 0.1376\cdot 10^0$$
$$u_2 = fl(u_1+x_2) = 0.6113\cdot 10^0$$
$$u_3 = fl(u_2+x_3) = 0.8039\cdot 10^1$$
$$u_4 = fl(u_3+x_4) = 0.7243\cdot 10^2$$
$$u_5 = fl(u_4+x_5) = 0.6487\cdot 10^3$$
$$u_6 = fl(u_5+x_6) = 0.2683\cdot 10^4$$

Dakle, zbir je $u = u_6 = 0.2683\cdot 10^4$.

Ako sabiranje izvodimo u obrnutom redosledu, imamo

$$v_1 = x_6 = 0.2034\ 10^4$$
$$v_2 = fl(v_1+x_5) = 0.2610\cdot 10^4,$$
$$v_3 = fl(v_2+x_4) = 0.2674\cdot 10^4,$$
$$v_4 = fl(v_3+x_3) = 0.2681\cdot 10^4,$$
$$v_5 = fl(v_4+x_2) = 0.2681\cdot 10^4,$$
$$v_6 = fl(v_5+x_1) = 0.2681\cdot 10^4,$$

tj. zbir je $v = v_6 = 0.2681\cdot 10^4$.

Medjutim, tačan zbir je $s = 0.26827293\cdot 10^4$. Odgovarajuće greške dobijenih zbirova u i v su

$$e_1 = u-s \simeq 0.27 \qquad \text{i} \qquad e_2 = v-s \simeq -1.73.$$

Primer 2.2.5. Neka su brojevi iz primera 2.2.3 pozitivni i bliski po vrednostima, tj. $x_i = x_o + \delta_i$, $|\delta_i| << x_o$ $(i = 1,2,3,4)$. Korišćenjem rezultata iz pomenutog primera, zaključujemo da je

$$|e_y^T| \leqq (9x_o + 3|\delta_1| + 3|\delta_2| + 2|\delta_3| + |\delta_4|)\frac{1}{2}\cdot 10^{-n+1},$$

tj.

$$|e_y^T| \leqq 4.5x_o\cdot 10^{-n+1},$$

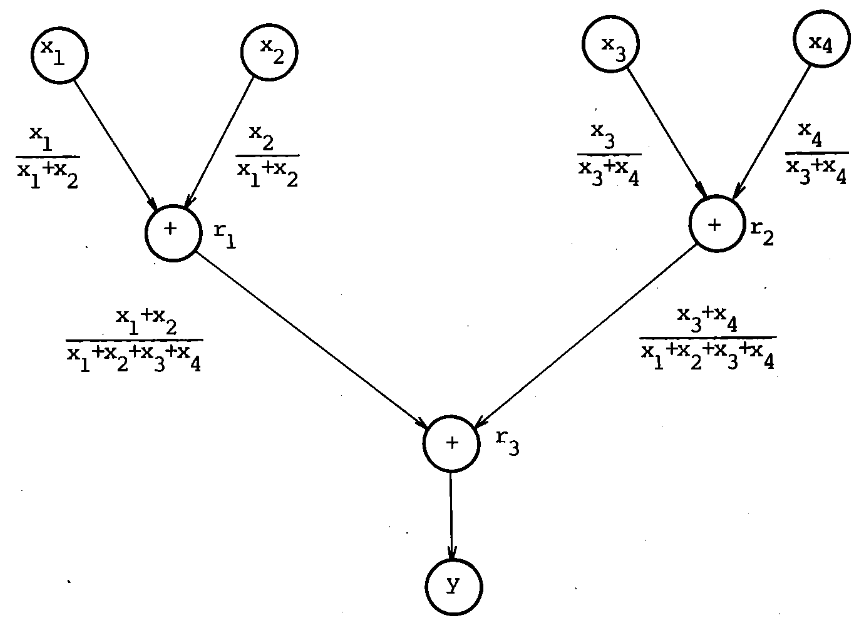
s obzirom na pretpostavku $|\delta_i| << x_o$ $(i = 1,2,3,4)$.

Izmenimo sada redosled izračunavanja. Naime, neka je

$$y' = (x_1 + x_2) + (x_3 + x_4).$$

Na osnovu grafa sa sl. 2.2.5. imamo

$$r_{y'}^T = \frac{x_1+x_2}{x_1+x_2+x_3+x_4}r_1 + \frac{x_3+x_4}{x_1+x_2+x_3+x_4}r_2 + r_3 ,$$



Sl. 2.2.5

odakle je

$$|e_{y^*}^T| \leq (2x_1+2x_2+2x_3+2x_4)\cdot\frac{1}{2}\cdot 10^{-n+1},$$

tj.

$$|e_{y^*}^T| \leq 4x_0\cdot 10^{-n+1}.$$

Dakle, na ovaj način se smanjuje granica apsolutne greške zbira četiri bliska pozitivna broja.

U opštem slučaju, ako imamo m^2 pozitivnih brojeva, približno jednakih po veličini, koje treba sabrati, granica apsolutne greške biće manja ukoliko brojeve grupišemo u m grupa po m brojeva i sabiramo brojeve u okviru svake grupe, a zatim sabiramo dobijene zbirove.

Na kraju, zadržaćemo se na problemu broja značajnih cifara približnog broja $\bar{x}$. Pretpostavićemo, kao i ranije, da računska mašina radi sa pokretnom tačkom i n-razrednom mantisom, kao i da je $b=10$.

Ako je $\bar{x}$, približan broj broja x, i ako $[\![p]\!]$ označava najbliži ceo broj* realnom broju p, broj značajnih cifara broja $\bar{x}$ može se oceniti pomoću

$$l = \left[\!\!\left[\log_{10}\frac{|x|}{|\bar{x}-x|}\right]\!\!\right]. \tag{2.2.11}$$

Drugim rečima, relativna greška r_x može se približno izraziti pomoću 10^{-l}. Naravno, uvek je $l \leq n$.

Posmatrajmo sada izračunavanje broja

$$x = f(x_1,\ldots,x_m) \quad (f : R^m \to R),$$

pomoću brojeva $x_i\,(i=1,\ldots,m)$. Pretpostavićemo da su brojevi x_i dati svojim približnim vrednostima $\bar{x}_i$ sa brojem značajnih cifara l_i.

Ako je $l_{min} = \min(l_1,\ldots,l_m)$, za grešku rezultata, može se dati približna ocena

$$e \approx \sum_{i=1}^{m} x_i\,\frac{\partial f}{\partial x_i}10^{-l_i} \leq 10^{-l_{min}}\sum_{i=1}^{m}\left|x_i\,\frac{\partial f}{\partial x_i}\right|. \tag{2.2.12}$$

---

*) Uzećemo da je $\left[\!\!\left[k+\frac{1}{2}\right]\!\!\right] = k+1$ (k ceo broj).

Na osnovu (2.2.11) i (2.2.12), broj značajnih cifara rezultata može se oceniti pomoću

(2.2.13) $$l_x = l_{min} - \delta,$$

gde je

(2.2.14) $$\delta = \left[\!\left[\log_{10} \frac{\sum_{i=1}^{m} \left| x_i \frac{\partial f}{\partial x_i} \right|}{|f(x_1,\ldots,x_m)|}\right]\!\right].$$

Veličina δ ukazuje na gubitak značajnih cifara u broju x.

Kako l_x u (2.2.13) ne može biti veće od n, to se za gubitak značajnih cifara u broju x, konačno, dobija

$$\delta_x = \min\left\{ \left[\!\left[\log_{10} \frac{\sum_{i=1}^{m} \left| x_i \frac{\partial f}{\partial x_i} \right|}{|f(x_1,\ldots,x_m)|}\right]\!\right],\ -(n - l_{min}) \right\}$$

Ispitajmo veličinu δ u nekim standardnim izračunavanjima.

1° Množenje $x=f(x_1,\ldots,x_m) = \prod_{i=1}^{m} x_i$. Kako je

$$\sum_{i=1}^{m} \left| x_i \frac{\partial f}{\partial x_i} \right| = m \left| \prod_{i=1}^{m} x_i \right|,$$

imamo

$$\delta = \left[\!\left[\log_{10} m\right]\!\right].$$

S obzirom da je, najčešće, $m \ll 100$, iz poslednje jednakosti sleduje, u najgorem slučaju, $\delta=1$. Ovo pokazuje da se gubitak značajnih cifara u ovom slučaju može zanemariti.

2° Deljenje $x = f(x_1,x_2) = \frac{x_1}{x_2}$. Iz (2.2.14) i

$$\sum_{i=1}^{2} \left| x_i \frac{\partial f}{\partial x_i} \right| = 2\left|\frac{x_1}{x_2}\right|,$$

sleduje

$$\delta = \left[\!\left[\log_{10} 2\right]\!\right] = 0,$$

što znači da nema gubitka značajnih cifara.

3° Sabiranje $x = f(x_1,\ldots,x_m) = \sum_{i=1}^{m} x_i$. Kako je

$$\sum_{i=1}^{m} \left|x_i \frac{\partial f}{\partial x_i}\right| = \sum_{i=1}^{m} |x_i|,$$

imamo

$$\delta = \left[\!\left[\log_{10} \frac{\sum_{i=1}^{m} |x_i|}{\left|\sum_{i=1}^{m} x_i\right|}\right]\!\right]. \tag{2.2.13}$$

Iz poslednje jednakosti sleduje da je $\delta=0$ samo ako su svi brojevi x_i istog znaka. Inače, u opštem slučaju, $\delta > 0$.

Dakle, kod sabiranja (oduzimanja) postoji gubitak značajnih cifara.

Kako se sva izračunavanja sastoje od konačnog broja kombinacija množenja, deljenja, sabiranja i oduzimanja, zaključujemo da je gubitak značajnih cifara prouzrokovan najčešće sabiranjima i oduzimanjima.

Efekat gubitka značajnih cifara kod oduzimanja bliskih brojeva ilustrovaćemo na sledećem primeru:

Primer 2.2.6. Izračunajmo razliku $x = x_1 - x_2$, uzimajući

$$x_1 = \sqrt{2.01} = 0.141774469\cdot 10^1, \quad x_2 = \sqrt{2} = 0.141421356\cdot 10^1.$$

Navedeni brojevi su dati sa devet tačnih značajnih cifara, medjutim rezultat se dobija sa šest značajnih cifara

$$x = 0.353113\cdot 10^{-2}.$$

Dakle, došlo je do gubitka tri značajne cifre, što se može odrediti i na osnovu (2.2.13) . Naime,

$$\delta = \left[\!\left[\log_{10} \frac{|x_1|+|x_2|}{|x_1 - x_2|}\right]\!\right] = 3 .$$

Medjutim, ako vrednost izraza za x izrazimo po formuli

$$x = \frac{0.01}{\sqrt{2.01}+\sqrt{2}},$$

kod koje izbegavamo oduzimanje bliskih brojeva, dobijamo rezultat sa svih devet značajnih cifara $x = 0.353112550\cdot 10^{-2}$.

1.2.3. Statistički prilaz u oceni grešaka

U dosadašnjoj analizi grešaka uvek smo razmatrali tzv. "najgori slučaj", tj. slučaj u kome su greške ulaznih podataka i greške zaokrugljivanja uzete tako da obezbedjuju maksimalnu grešku izlazne informacije, mada taj slučaj vrlo retko nastupa.

U ovom odeljku ukazaćemo na statistički prilaz u oceni greške na jednom jednostavnom primeru. Naime, neka je

$$u = \sum_{i=1}^{n} x_i$$

i neka su brojevi x_i dati svojim približnim vrednostima $\bar{x}_i$, pri čemu su odgovarajuće granice apsolutnih grešaka Δ_i, tj.

$$|e_i| = |\bar{x}_i - x_i| \leqq \Delta_i \qquad (i=1,\dots,n).$$

Pod pretpostavkom da ne postoje greške zaokrugljivanja u procesu računanja, greška zbira će biti

$$e_u = \sum_{i=1}^{n} e_i ,$$

odakle sleduje ocena

$$|e_u| \leqq \sum_{i=1}^{n} |e_i| \leqq \sum_{i=1}^{n} \Delta_i . \tag{2.3.1}$$

Dakle, ova ocena daje maksimalnu apsolutnu grešku, koja nastupa samo u dva slučaja, tj. kada je $e_i=\Delta_i$ $(i=1,\dots,n)$ ili kada je $e_i = -\Delta_i$ $(i=1,\dots,n)$. Naravno, sa stanovišta teorije verovatnoće i statistike, ovi slučajevi su malo verovatni. Da bismo mogli dobiti statističku ocenu za grešku e_u potrebno je uvesti pretpostavku o funkcijama raspodele grešaka e_i $(i=1,\dots,n)$. Pretpostavimo da su sve graniceapsolutnih grešaka medjusobno jednake, tj. $\Delta_i=\Delta$ $(i=1,\dots,n)$ i da greške podležu ravnomernom zakonu raspodele u intervalu $(-\Delta,\Delta)$, tj. da je gustina verovatnoće svake od grešaka e_i $(i=1,\dots,n)$ jednaka $\frac{1}{2\Delta}$. U ovom slučaju, matematičko očekivanje $m(e_i)$ i disperzija $D(e_i)$ su dati sa

$$m(e_i) = \int_{-\Delta}^{+\Delta} \frac{t}{2\Delta}\,dt = 0 \quad \text{i} \quad D(e_i) = \int_{-\Delta}^{+\Delta} \frac{t^2}{2\Delta}\,dt = \frac{1}{3}\Delta^2.$$

Pretpostavljajući da su greške e_i nezavisne veličine, za disperziju sume dobijamo

$$D(e_u) = \sum_{i=1}^{n} D(e_i) = \frac{1}{3}\,n\Delta^2.$$

Da bismo našli statističku ocenu za e_u potrebno je naći funkciju raspodele sume grešaka. S obzirom na glomaznost postupka za konstrukciju ovakve funkcije, pristupićemo izvesnoj aproksimaciji. Naime, pretpostavićemo da se suma ponaša po normalnom zakonu raspodele sa matematičkim očekivanjem 0 i standardnom devijacijom $\sigma = \Delta\sqrt{\frac{n}{3}}$.

Tada je

$$P(|e_u| < 3\sigma) = 0.9973,$$

tj. sa verovatnoćom 0.9973 možemo očekivati da je

$$|e_u| < 3\,\sigma = \Delta\sqrt{3n}. \tag{2.3.2}$$

Dakle, za statističku granicu apsolutne sume možemo uzeti veličinu $\Delta\sqrt{3n}$, koja je pri $n > 3$ manja od granice $n\Delta$, dobijene na osnovu (2.3.1).

Primetimo da sa verovatnoćom 0.68 možemo očekivati da je $|e_u| < \sigma = \Delta\sqrt{\frac{n}{3}}$.

Primer 2.3.1. Kod sabiranja 75 brojeva, koji su svi dati sa tačnošću $5\cdot 10^{-4}$, granica apsolutne greške je $n\cdot\Delta = 75.5\cdot 10^{-4} = 3.75\cdot 10^{-2}$. Medjutim, statistička granica (na osnovu (2.3.2)) je

$$|e_u| < 5\cdot 10^{-4}\sqrt{3.75} = 0{,}75\cdot 10^{-2},$$

tj. pet puta je manja.

1.3. REKURZIVNA IZRAČUNAVANJA I SUMIRANJE

1.3.1. Diferencne jednačine

Kako se mnogi problemi u numeričkoj analizi svode na rešavanje diferencnih jednačina, u ovom odeljku, daćemo ukratko neke pojmove i rezultate iz teorije ovih jednačina. Opširnije o diferencnim jednačinama može se naći, na primer, u [3], [4], [5].

Definicija 3.1.1. Pod konačnom razlikom prvog reda funkcije $x \mapsto f(x)$ podrazumevamo izraz

$$\Delta f(x) = f(x+h) - f(x) \qquad (h=\text{const} > 0).$$

Definicija 3.1.2. Pod konačnom razlikom reda n funkcije f podrazumevamo izraz $\Delta^n f(x)$ dat rekurzivno pomoću

$$\Delta^n f(x) = \Delta^{n-1} f(x+h) - \Delta^{n-1} f(x) \qquad (\Delta^o f(x)=f(x)).$$

Lako se mogu dokazati sledeće formule

$$\Delta^n f(x) = \sum_{i=o}^{n} (-1)^i \binom{n}{i} f(x+(n-i)h) \tag{3.1.1}$$

i

$$f(x+nh) = \sum_{i=o}^{n} \binom{n}{i} \Delta^i f(x).$$

Primedba 3.1.1. Slično prethodnim definicijama mogu se definisati konačne razlike niza $\{y_k\}_{k\in N}$. Naime, imamo

$$\Delta^o y_k = y_k, \quad \Delta^n y_k = \Delta^{n-1} y_{k+1} - \Delta^{n-1} y_k \quad (n=1,2,\ldots).$$

Primer 3.1.1. Konačne razlike funkcija f i g, koje su definisane pomoću $f(x)=ax^2+bx+c$ i $g(x) = e^{ax}$, su redom

$$\Delta f(x) = h(2ax+ah+b), \quad \Delta^2 f(x) = 2ah^2, \quad \Delta^n f(x) = 0 \quad (n=3,4,\ldots)$$

i

$$\Delta^n g(x) = e^{ax}(e^{ah}-1)^n \qquad (n=1,2,\ldots).$$

Konstanta h naziva se korak. Često se u prethodnim definicijama uzima h=1.

Jednačina oblika

$$F(x;f(x),\Delta f(x),\ldots,\Delta^n f(x)) = 0,$$

gde je f nepoznata funkcija, predstavlja diferencnu jednačinu reda n, ako posle transformacije pomoću (3.1.1) sadrži $f(x+nh)$ i $f(x)$; u protivnom jednačina je nižeg reda od n. Dakle, diferencna jednačina reda n ima oblik

$$G(x;f(x),f(x+h),\ldots,f(x+nh)) = 0. \tag{3.1.2}$$

<u>Primer 3.1.2</u>. Diferencna jednačina (h=1)

$$\Delta^3 f(x) - 3\Delta f(x) - 2f(x) = x+2$$

je prvog reda, jer se pomoću (3.1.1) svodi na

$$f(x+3) - 3f(x+2) = x+2,$$

tj.

$$f(z+1) - 3f(z) = z,$$

pri čemu smo uveli smenu $z=x+2$.

Ako uvedemo smenu $x=x_o+kh$ i stavimo $y_k=f(x_o+kh)$, tada (3.1.2) postaje

$$H(k;y_k,y_{k+1},\ldots,y_{k+n}) = 0.$$

U našim razmatranjima od interesa su linearne diferencne jednačine, čiji je opšti oblik

$$y_{k+n} + b_1(k)y_{k+n-1} + \ldots + b_n(k)y_k = Q(k).$$

Ako je $Q(k) \equiv 0$ jednačina je homogena; u protivnom jednačina je nehomogena.

Posmatrajmo, najpre, linearnu diferencnu jednačinu prvog reda

(3.1.3) $$y_{k+1}+b(k)y_k=Q(k).$$

Opšte rešenje homogene jednačine $y_{k+1}+b(k)y_k=0$ je dato sa

(3.1.4) $$y_k=(-1)^k C\prod_{i=o}^{k-1} b(i) \qquad \text{(C proizvoljna konstanta).}$$

Ako pretpostavimo da C zavisi od k, tj. $C=C_k$, mogućno je polazeći od rešenja (3.1.4) naći rešenje nehomogene jednačine (3.1.3). Naime, u tom slučaju dobijamo da je rešenje nehomogene jednačine data sa

$$y_k=(-1)^k\prod_{i=o}^{k-1}\left(C+\sum_{j=1}^{k}\frac{(-1)^j Q(j-1)}{\prod\limits_{i=o}^{j-1} b(i)}\right),$$

gde je C proizvoljna konstanta.

Opšte rešenje linearne diferencne jednačine n-tog reda je dato sa

$$y_k = C_1\varphi_1(k)+C_2\varphi_2(k)+\ldots+C_n\varphi_n(k)+f_p(k),$$

gde su $\varphi_1,\varphi_2,\ldots,\varphi_n$ linearno nezavisna partikularna rešenja odgovarajuće homogene jednačine, a f_p jedno partikularno rešenje nehomogene jednačine. Konstante C_i ($i=1,\ldots,n$) su proizvoljne i u konkretnim slučajevima odredjuju se iz početnih uslova.

Daćemo sada neke napomene koje se odnose na odredjivanje opšteg rešenja linearne homogene diferencne jednačine sa konstantnim koeficijentima. Pretpostavimo da je rešenje ove jednačine oblika $y_k=\lambda^k$. Tada zamenom u odgovarajućoj jednačini dobijamo

$$\lambda^{k+n}+b_1\lambda^{k+n-1}+\ldots+b_n\lambda^k=0.$$

Dakle, karakteristična jednačina je

$$\lambda^n+b_1\lambda^{n-1}+\ldots+b_n=0.$$

U zavisnosti od korena karakteristične jednačine razlikovaćemo slučajeve:

a) Ako su svi koreni $\lambda_1,\ldots,\lambda_n$ prosti, opšte rešenje je

$$y_k = C_1\lambda_1^k+\ldots+ C_n\lambda_n^k;$$

b) Ako su λ_p i λ_q konjugovano-kompleksni koreni, tj.

$$\lambda_p = \rho(\cos\theta + i\sin\theta) \quad i \quad \lambda_q = \bar{\lambda}_p,$$

tada njima u opštem rešenju odgovara izraz

$$C_p\lambda_p^k + C_q\lambda_q^k = \rho^k(A_p\cos k\theta + A_q\sin k\theta);$$

c) Ako je λ_i višestruki koren reda m, tada njemu u opštem rešenju odgovara izraz

$$(C_1 + C_2k + \ldots + C_mk^{m-1})\lambda_i^k.$$

Kod linearnih nehomogenih jednačina sa konstantnim koeficijentima moguće je u izvesnim slučajevima naći partikularno rešenje f_p.

Ako je $Q(k) = P_s(k)$ (P_s polinom stepena s) i $\lambda = 1$ koren karakteristične jednačine reda m, tada se partikularno rešenje može tražiti u obliku

$$f_p(k) = A_ok^m + A_1k^{m+1} + \ldots + A_sk^{m+s}.$$

Ukoliko je $Q(k) = P_s(k)a^k$ i $\lambda = a$ koren karakteristične jednačine reda m, f_p treba tražiti u obliku

$$f_p(k) = a^k(A_ok^m + \ldots + A_sk^{m+s}).$$

Primer 3.1.3. Rešimo diferencnu jednačinu

$$y_{k+4} + 2y_{k+3} + 3y_{k+2} + 2y_{k+1} + y_k = 0,$$

pod uslovima $y_o=y_1=y_3=0$ i $y_2=-1$.

Kako je karakteristična jednačina

$$\lambda^4 + 2\lambda^3 + 3\lambda^2 + 2\lambda + 1 = 0,$$

tj. $(\lambda^2+\lambda+1)^2=0$, imamo $\lambda_1=\lambda_2 = \frac{1}{2}(-1+ i\sqrt{3}),\lambda_3=\lambda_4= \frac{1}{2}(-1-i\sqrt{3})$.

Dakle, $\rho = 1$, $\theta = \frac{2\pi}{3}$ pa je opšte rešenje dato pomoću

$$y_k = (C_1 + C_2 k)\cos\frac{2\pi k}{3} + (C_3 + C_4 k)\sin\frac{2\pi k}{3},$$

gde su $C_i\,(i = 1,\ldots,4)$ proizvoljne konstante. Na osnovu datih uslova nalazimo $C_1 = C_2 = 0$, $C_3 = -C_4 = -2/\sqrt{3}$, tj.

$$y_k = \frac{2(k-1)}{\sqrt{3}}\sin\frac{2\pi k}{3}$$

Primetimo da se ovo rešenje može predstaviti i u obliku

$$y_k = \begin{cases} 0 & k = 0 \pmod 3, \\ k-1 & k = 1 \pmod 3, \\ 1-k & k = -1 \pmod 3. \end{cases}$$

Primer 3.1.4. Nadjimo opšti član Fibonacievog niza: 0,1,1,2,3, 5,8,13,21,... Ovde je $y_{k+2} = y_k + y_{k+1}$. Rešenje ove jednačine, pod uslovima $y_0 = 0$ i $y_1 = 1$, je

$$y_k = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right].$$

1.3.2. Napomene o izračunavanju vrednosti funkcija i i sumiranju redova

Pri izračunavanju vrednosti funkcija, na računskim mašinama, vrlo je važno u kom su obliku date odgovarajuće formule. Naime, često vrednost ekvivalentnih matematičkih izraza u numeričkom smislu nije ista(zbog grešaka o kojima je bilo reči u prethodnom poglavlju).

U ovom odeljku ukazaćemo na neke standardne načine za izračunavanje vrednosti elementarnih funkcija. Kako su ovi načini pretežno zasnovani na primeni redova, to ćemo izneti i neke probleme vezane za sumiranje redova.

Primer 3.2.1. Za izračunavanje vrednosti funkcije $x \mapsto e^x$ može se koristiti Taylorov razvoj

$$e^x = \sum_{k=0}^{n}\frac{x^k}{k!} + R_n(x), \tag{3.2.1}$$

gde je $R_n(x) = \frac{x^{n+1}}{(n+1)!} e^{\xi x}$ $(0 < \xi < 1)$.

Ako je $0 \leqq x \leqq 1$ moguće je ostatak $R_n(x)$ oceniti na sledeći način. Naime, kako je

$$R_n(x) = \sum_{i=n+1}^{+\infty} \frac{x^k}{k!} = \frac{x^{n+1}}{(n+1)!}\left(1 + \frac{x}{n+2} + \frac{x^2}{(n+2)(n+3)} + \dots\right)$$

$$< \frac{x^{n+1}}{(n+1)!}\left(1 + \frac{x}{n+2} + \left(\frac{x}{n+2}\right)^2 + \dots\right),$$

za $0 \leqq x < n+2$ imamo

$$R_n(x) < \frac{x^{n+1}}{(n+1)!}\cdot\frac{n+2}{n+2-x}.$$

Ako je $0 \leqq x \leqq 1$, iz prethodne nejednakosti sleduje

$$R_n(x) < \frac{x^{n+1}}{(n+1)!}\cdot\frac{n+2}{n+1} \leqq \frac{n+2}{(n+1)(n+1)!}. \tag{3.2.2}$$

Broj članova n u razvoju (3.2.1), koji treba uzeti, zavisi od tačnosti koja se zahteva za rezultat. Ukoliko je ovaj broj članova veliki, to će konačan rezultat, ipak, biti pogrešan usled grešaka zaokrugljivanja tokom računskog procesa.

U konkretnom slučaju, kada je $0 \leqq x \leqq 1$, uslov $R_n(x) < 10^{-7}$, na osnovu (3.2.2), ispunjen je za najmanje $n=10$. Kada se svi brojevi koji učestvuju u izračunavanjima zaokrugljuju i predstavljaju u normalizovanom obliku sa sedmorazrednom mantisom, ovaj broj članova razvoja nije veliki, s obzirom da je relativna greška zaokrugljivanja takodje reda 10^{-7}.

<u>Primer 3.2.2</u>. Izračunajmo broj e sa relativnom greškom koja je manja od 10^{-7}. Na osnovu prethodnog primera, u razvoju (3.2.1) treba uzeti $n=10$ i $x=1$. Dakle,

$$e = \sum_{k=0}^{10} u_k + R_{10}(1),$$

gde je

$$u_0 = 1, \quad u_k = \frac{1}{k} u_{k-1} \quad (k = 1, \dots, 10).$$

Zaokrugljeni brojevi u_k $(k=0,1,\ldots,10)$ predstavljeni u normalizovanom obliku sa sedmorazrednom mantisom su redom

$$\begin{array}{ll} u_0=u_1=0.1000000\cdot 10^1, & u_2=0.5000000\cdot 10^0, \\ u_3=0.1666667\cdot 10^0, & u_4=0.4166667\cdot 10^{-1}, \\ u_5=0.8333333\cdot 10^{-2}, & u_6=0.1388889\cdot 10^{-2}, \\ u_7=0.1984127\cdot 10^{-3}, & u_8=0.2480158\cdot 10^{-4}, \\ u_9=0.2755731\cdot 10^{-5}, & u_{10}=0.2755731\cdot 10^{-6}. \end{array}$$

Iz odeljka 1.2.2 poznato je da će greška biti najmanja ako sabiramo brojeve polazeći od najmanjeg, tj. od u_{10}. Tada za zbir dobijamo

$$S_0=0.2718282\cdot 10^1.$$

Broj S_0 aproksimira broj $e\ (=0.27182818284\ldots\cdot 10^1)$ sa sedam značajnih cifara. Relativna greška je manja od 10^{-7} (tačnije, manja od $0.67\cdot 10^{-7}$).

Primedba 3.2.1. U opštem slučaju za izračunavanje vrednosti funkcije $x\mapsto e^x$ poželjno je koristiti se formulom

$$e^x=e^{[x]}e^z,$$

gde su $[x]$ celi deo od x i $z=x-[x]$. Pri ovome vrednost $e^{[x]}$ izračunava se prostim množenjem, tj.

$$e^{[x]}=\begin{cases} 1 & ([x]=0), \\ \underbrace{e\ldots e}_{[x]\text{ puta}} & ([x]>0), \\ \underbrace{\frac{1}{e}\ldots\frac{1}{e}}_{-[x]\text{ puta}} & ([x]<0), \end{cases}$$

gde su

$$e=2.71828182845904\ldots \quad \text{i} \quad \frac{1}{e}=0.36787944117144\ldots\ .$$

Izračunavanje vrednosti $e^z\ (0\leq z<1)$, kao što smo ranije videli, ne predstavlja posebnu teškoću.

<u>Primer 3.2.3.</u> Za izračunavanje vrednosti funkcije $x \mapsto \log x$ $(0 < x < +\infty)$ koristimo transformaciju $x = 2^m z$, gde su z i m takvi da je $\frac{1}{2} \leq z < 1$ i $m \in Z$. Tada je

$$\log x = m \log 2 + \log z \qquad (\log 2 = 0.69314718\ldots).$$

Ako stavimo $z = \frac{1-y}{1+y}$, imamo da je

$$\log z = \log\frac{1-y}{1+y} = -2\left(y + \frac{y^3}{3} + \frac{y^5}{5} + \ldots + \frac{y^{2n-1}}{2n-1}\right) - R_n(y),$$

gde su

$$y = \frac{1-z}{1+z}$$

i

$$R_n(y) = 2\left(\frac{y^{2n+1}}{2n+1} + \frac{y^{2n+3}}{2n+3} + \ldots\right) < \frac{2}{1-y^2}\cdot\frac{y^{2n+1}}{2n+1}.$$

Kako je $0 < y \leq \frac{1}{3}$ $(\Leftarrow \frac{1}{2} \leq z < 1)$ za ostatak važe nejednakosti

$$0 < R_n(y) < \frac{1}{4\cdot 3^{2n-1}(2n+1)}.$$

Uvodjenjem oznake $u_k = \frac{y^{2k-1}}{2k-1}$ $(k = 1,\ldots,n)$, najzad dobijamo

$$\log x = m \log 2 - 2(u_1 + \ldots + u_n) - R_n,$$

gde se n odredjuje iz uslova $u_n < 4\varepsilon$ (ε zadata tačnost), jer je tada $R_n = R_n(y) \leq \frac{1}{4}u_n < \varepsilon$.

<u>Primer 3.2.4.</u> Kod sumiranja reda $\displaystyle\sum_{k=0}^{+\infty} (-1)^k \frac{4}{2k+1}$, da bi se postigla tačnost od 10^{-4}, potrebno je uzeti oko 20000 članova reda, što sleduje iz nejednakosti $4/(2n+1) < 10^{-4}$. Praktično ovakav način za sumiranje je neizvodljiv. Napomenimo da je suma ovog reda π $(= 3.1415926535\ldots)$. Dati red pripada klasi tzv. sporokonvergentnih redova.

Na osnovu poslednjeg primera može se zaključiti da kod sporokonvergentnih redova, direktno sumiranje članova reda ne dovodi do željenog rezultata(zbog velikog broja članova reda koje treba sabrati i grešaka zaokrugljivanja koje pritom nastaju). U ovakvim slučajevima treba naći neke načine za tzv. brzo sumiranje ili kako se često kaže ubrzavanje konvergencije redova. U literaturi postoji veliki broj metoda koji ovaj problem rešavaju. Navešćemo nekoliko od njih.

Posmatrajmo konvergentni alternativni red

$$(3.2.3) \qquad \sum_{k=0}^{+\infty}(-1)^k a_k = a_0 - a_1 + a_2 - a_3 + \dots \qquad (a_k \geq 0),$$

čija je suma A. Pokazaćemo da je red

$$(3.2.4) \qquad \frac{1}{2}a_0 - \frac{1}{2}(a_1-a_0) + \frac{1}{2}(a_2-a_1) - \dots$$

takodje konvergentan i da ima istu sumu kao i red (3.2.3).

Pretpostavimo da su $\{A_n\}$ i $\{B_n\}$ nizovi parcijalnih suma redova (3.2.3) i (3.2.4) respektivno. Tada je

$$(3.2.5) \qquad A_n - B_n = \frac{1}{2}(-1)^{n-1}a_{n-1} \qquad (n=1,2,\dots).$$

Kako je prvi red konvergentan, imamo da je $\lim a_n = 0$ pa iz (3.2.5) sleduje

$$\lim B_n = \lim A_n = A.$$

Primenimo izloženu transformaciju na red iz primera 3.2.4, čiju sumu označimo sa S. Imamo

$$S = 2 + \sum_{k=1}^{+\infty}(-1)^k \frac{1}{2}\left(\frac{4}{2k+1} - \frac{4}{2k-1}\right),$$

tj.

$$(3.2.6) \qquad S = 2 + \sum_{k=1}^{+\infty}\frac{4(-1)^{k-1}}{4k^2-1}.$$

Ocenimo sada koliko je članova reda (3.2.6) potrebno sabrati da bi se dobila suma sa istom tačnošću kao u primeru 3.2.4. Iz $4/(4n^2-1) < 10^{-4}$ sleduje $n > 100$. Dakle, sada se ista tačnost može postići ako se uzmu prvih sto članova reda.

Izložena transformacija za sumiranje alternativnih numeričkih redova predstavlja specijalan slučaj Euler-Abelove transformacije koja se primenjuje kod stepenih redova.

Razmotrimo stepeni red

$$f(x) = \sum_{k=0}^{+\infty} a_k x^k , \tag{3.2.7}$$

čiji je poluprečnik konvergencije R konačan. Ne umanjujući opštost razmatranja možemo uzeti $R = 1$.

Iz $f(x) = a_o + xg(x)$, gde je

$$g(x) = \sum_{k=0}^{+\infty} a_{k+1} x^k ,$$

sleduje

$$\begin{aligned} (1-x)g(x) &= (1-x) \sum_{k=0}^{+\infty} a_{k+1} x^k \\ &= \sum_{k=0}^{+\infty} a_{k+1} x^k - \sum_{k=0}^{+\infty} a_{k+1} x^{k+1} \\ &= a_o + \sum_{k=0}^{+\infty} (a_{k+1} - a_k) x^k \end{aligned}$$

odakle je

$$f(x) = a_o + \frac{x}{1-x}\Big(a_o + \sum_{k=0}^{+\infty} (a_{k+1} - a_k) x^k \Big),$$

tj.

$$f(x) = \frac{a_o}{1-x} + \frac{x}{1-x} \sum_{k=0}^{+\infty} \Delta a_k \, x^k .$$

Iz poslednje formule za $x = -1$ i $a_k \geq 0$, sleduje

$$\sum_{k=0}^{+\infty} (-1)^k a_k = \frac{1}{2} a_o - \frac{1}{2} \sum_{k=0}^{+\infty} (-1)^k \Delta a_k ,$$

što je ekvivalentno sa (3.2.4).

Sukcesivnom primenom izložene Euler-Abelove transformacije m puta na red (3.2.7) dobijamo

$$f(x)=\sum_{k=o}^{m-1}\frac{\Delta^k a_o x^k}{(1-x)^{k+1}}+\left(\frac{x}{1-x}\right)^m\sum_{k=o}^{+\infty}\Delta^m a_k x^k .$$

<u>Primer 3.2.5</u>. Sukcesivnom primenom Euler-Abelove transformacije dva puta na red $f(x)=\sum_{k=o}^{+\infty}\frac{x^k}{k+1}$, dobijamo

$$f(x)=\frac{1}{1-x}+\frac{(\frac{1}{2}-1)x}{(1-x)^2}+\left(\frac{x}{1-x}\right)^2\sum_{k=o}^{+\infty}\left(\frac{1}{k+3}-\frac{2}{k+2}+\frac{1}{k+1}\right)x^k$$

tj.

$$f(x)=\frac{1}{1-x}-\frac{x}{2(1-x)^2}+\left(\frac{x}{1-x}\right)^2\sum_{k=o}^{+\infty}\frac{2x^k}{(k+1)(k+2)(k+3)} .$$

Primenimo dobijenu formulu na izračunavanje vrednosti $\log 2$. S obzirom da je $\log 2=f(-1)$ imamo

$$\log 2\cong\frac{5}{8}+\frac{1}{2}\sum_{k=o}^{n}\frac{(-1)^k}{(k+1)(k+2)(k+3)} ,$$

odakle, za $n=2$, dobijamo $\log 2\cong 0.6958333$. Dobijeni rezultat je tačan na dve decimale.

Za ubrzavanje konvergencije redova postoje i transformacije koje se zasnivaju na ubrzavanju konvergencije nizova.

Neka je $\{S_n\}$ niz parcijalnih suma reda (3.2.3), tj. $S_n=\sum_{k=o}^{n-1}(-1)^k a_k$ $(n\in N)$. Može se pokazati da niz $\{S'_n\}$, gde je

$$S'_n=\frac{1}{2}(S_n+S_{n+1})\qquad(n=1,2,\ldots)$$

brže konvergira nego niz $\{S_n\}$ i da je $\lim S'_n=\lim S_n$. Daljom primenom iste transformacije na niz $\{S'_n\}$ dobija se niz $\{S''_n\}$, gde je

$$S''_n=\frac{1}{2}(S'_n+S'_{n+1})=\frac{1}{4}(S_n+2S_{n+1}+S_{n+2})\qquad(n=1,2,\ldots).$$

Na kraju napomenimo da postoji čitava klasa nelinearnih transformacija koje se primenjuju kod sumiranja redova ([6], [7], [8]). Jedna od najprostijih je Δ^2-transformacija

$$S_n' = S_n - \frac{(\Delta S_n)^2}{\Delta^2 S_n} \qquad (n = 0,1,\ldots)$$

o kojoj će biti reči u opštoj teoriji iterativnih procesa(odeljak 3.2.2).

1.3.3. Neke primene verižnih razlomaka

Izraz oblika

$$a_0 + \cfrac{b_1}{a_1 + \cfrac{b_2}{a_2 + \cfrac{b_3}{a_3 + \cdots}}}$$

naziva se verižni razlomak i predstavlja se u jednom od sledećih oblika

$$\left[a_0; \frac{b_1}{a_1}, \frac{b_2}{a_2}, \frac{b_3}{a_3}, \ldots\right], \quad a_0 + \frac{b_1|}{|a_1} + \frac{b_2|}{|a_2} + \frac{b_3|}{|a_3} + \cdots,$$

$$a_0 + \frac{b_1}{a_1+}\,\frac{b_2}{a_2+}\,\frac{b_3}{a_3+}\cdots .$$

Verižni razlomak može biti konačan, na primer

$$1 + \cfrac{2}{2 + \cfrac{5}{3}} = \left[1; \frac{2}{2}, \frac{5}{3}\right] = 1 + \frac{2|}{|2} + \frac{5|}{|3} = 1 + \frac{2}{2+}\,\frac{5}{3} .$$

Od interesa je razlaganje racionalne funkcije u verižni razlomak.

Neka je

$$f(x) = \frac{c_{10} + c_{11}x + c_{12}x^2 + \cdots}{c_{00} + c_{01}x + c_{02}x^2 + \cdots} \qquad (c_{10} \neq 0).$$

Tada je

$$f(x) = \cfrac{1}{\cfrac{c_{oo}}{c_{1o}} + \cfrac{c_{oo}+c_{o1}x+c_{o2}x^2+\dots}{c_{1o}+c_{11}x+c_{12}x^2+\dots} - \cfrac{c_{oo}}{c_{1o}}} = \frac{c_{1o}}{c_{oo}+xf_1(x)},$$

gde je

$$f_1(x) = \frac{c_{2o}+c_{21}x+c_{22}x^2+\dots}{c_{1o}+c_{11}x+c_{12}x^2+\dots}$$

i

$$c_{2j} = c_{1o}c_{o,j+1} - c_{oo}c_{1,j+1} \quad (j=0,1,\dots).$$

Nastavljajući ovaj proces dobija se razvoj u verižni razlomak

$$f(x) = \left[0;\ \frac{c_{1o}}{c_{oo}},\ \frac{c_{2o}x}{c_{1o}},\ \frac{c_{3o}x}{c_{2o}},\ \dots\right],$$

gde su

$$c_{ij} = -\begin{vmatrix} c_{i-2,o} & c_{i-2,j+1} \\ c_{i-1,o} & c_{i-1,j+1} \end{vmatrix} \quad (i=2,3,\dots).$$

<u>Primer 3.3.1</u>. Za funkciju $x \mapsto f(x) = \dfrac{1-x}{1-5x+6x^2}$ imamo

$$f(x) = \left[0;\ \frac{1}{1},\ \frac{-4x}{1},\ \frac{-2x}{-4},\ \frac{-12x}{-2}\right]$$

$$= \cfrac{1}{1-\cfrac{4x}{1-\cfrac{2x}{-4+6x}}}.$$

Za izračunavanje vrednosti nekih funkcija umesto Taylorovog razvoja, može se koristiti razvoj u verižni razlomak. Zato ćemo sada navesti, bez dokaza, razvoje za funkcije $x \mapsto e^x$ i $x \mapsto \operatorname{tg} x$.

Euler je dokazao razvoj

$$e^x = \left[0;\ \frac{1}{1},\ \frac{-2x}{2+x},\ \frac{x^2}{6},\ \frac{x^2}{10},\ \ldots,\ \frac{x^2}{4n+2},\ \ldots\right],$$

koji konvergira za svako realno ili kompleksno x, dok je razvoj

$$\operatorname{tg} x = \left[0;\ \frac{x}{1},\ \frac{-x^2}{3},\ \frac{-x^2}{5},\ \ldots,\ \frac{-x^2}{2n+1},\ \ldots\right]$$

dokazao Lambert. Poslednji razvoj konvergira za svako x, za koje je funkcija $x \mapsto \operatorname{tg} x$ neprekidna.

Primedba 3.3.1. Za izračunavanje vrednosti $\operatorname{tg} x$ pogodno je uvesti smenu $\operatorname{tg} x = \frac{x}{y}$. Tada se najpre izračunava vrednost

$$y = \left[1;\ \frac{-x^2}{3},\ \frac{-x^2}{5},\ \ldots\right],$$

a zatim $\operatorname{tg} x = \frac{x}{y}$.

1.3.4. Izračunavanje vrednosti polinoma

Jedan elementaran, ali važan problem je izračunavanje vrednosti polinoma.

$$(3.4.1)\qquad P(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n.$$

Ako bismo izračunavali vrednost polinoma P(x), na osnovu (3.4.1), bilo bi potrebno 2n-1 množenja i n sabiranja. Medjutim, ukoliko P(x) predstavimo u obliku

$$(3.4.2)\qquad P(x) = a_n + x(a_{n-1} + \ldots + x(a_3 + x(a_2 + x(a_1 + xa_0)))\ldots)$$

potrebno je samo n množenja i n sabiranja.

Na osnovu (3.4.2) može se formirati rekurzivni postupak

$$u_0 = a_0,\quad u_k = a_k + xu_{k-1}\quad (k = 1,\ldots,n)$$

koji posle n koraka daje vrednost polinoma, tj. $P(x) = u_n$.

Izloženi postupak je poznat kao Hornerova šema.

1.4. LITERATURA

1. J.STOER: Einführung in die Numerische Mathematik I. Berlin-Heidelberg--New York, 1972.
2. D.D.McCRACKEN, W.S.DORN: Numerical methods and FORTRAN programing. New York, 1965.
3. H.LEVY and F.LESSMAN: Finite difference equations. London, 1959.
4. K.S.MILER: An introduction to the calculus of finite differences and difference equations. New York, 1960.
5. N.E.NORLUND: Vorlesungen Uber Differenzenrechung. Berlin, 1924.
6. D.SHANKS: Non-Linear Transformations of Divergent and Slowly Convergent Sequences. Amer.Math.Monthly 57(1949), 439.
7. S.BREZINSKI: Méthode d´accélération de la convergence en analyse numérique (These). Grenoble, 1971.
8. S.BREZINSKI: Accéleration de suites a convergence logarithmique. C.R.Acad. Sc.Paris. 273(1971), (Série A), 727-730.
9. J.H.WILKINSON: Error analysis of floating point computation. Numer.Math. 2(1960), 319-340.
10. F.L.BAUER: Computational graphs and rounding error. SIAM J.Numer.Anal. 11(1974), 87-96.
11. C.T.FIKE: Computer Evaluation of Mathematical Functions. New Jersy, 1968.

2. ELEMENTI REALNE I FUNKCIONALNE ANALIZE

2.1. OSNOVNE DEFINICIJE I STAVOVI O LINEARNIM PROSTORIMA

2.1.1. Linearni prostor

Definicija 1.1.1. Skup X naziva se linearni(vektorski) prostor nad poljem K ako je:

1° U skupu X definisana jedna binarna operacija + u odnosu na koju skup X čini Abelovu grupu;

2° Ako je svakom paru (u,c) ($u \in X; c \in K$) dodeljen po jedan element cu skupa X tako da su ispunjeni uslovi:

$$
\begin{aligned}
&c_1(c_2u) = (c_1c_2)u\,,\\
&c(u_1 + u_2) = cu_1 + cu_2\,,\\
&(c_1 + c_2)u = c_1u + c_2u\,,\\
&1u = u
\end{aligned}
\tag{1.1.1}
$$

za sve elemente $u, u_i \in X$ i $c, c_i \in K (i=1,2)$.

Jedinični element polja K označen je sa 1. Elementi skupa X zovu se vektori(tačke), elementi polja K skalari, operacija + skupa X vektorsko sabiranje(unutrašnja kompozicija) i operacija $(u,c) \to cu$ množenje vektora skalarom (spoljašnja kompozicija).

Treća jednakost iz (1.1.1) pri $c_1 = 1$ i $c_2 = -1$ daje

$$ou = u + (-u) = \Theta \qquad \text{za svako } u \in X,$$

gde je Θ neutralni element skupa X za operaciju vektorskog sabiranja i ovaj element se naziva nula-vektor.

Osim toga, stavljanjem $u_2 = -u_1$ u drugoj jednakosti iz (1.1.1) dobijamo

$$c\Theta = \Theta \qquad \text{za svako } c \in K.$$

Definicija 1.1.2. Vektori $u_i (i=1,\dots,n)$ linearnog vektorskog prostora X nazivaju se linearno zavisnim ako u polju K postoje brojevi $c_i (i=1,\dots,n)$ koji istovremeno nisu jednaki nuli, tako da je

$$\sum_{i=1}^{n} c_i u_i = \theta . \tag{1.1.2}$$

Vektori $u_i (i=1,\dots,n)$ su linearno nezavisni, ako je jednakost (1.1.2) tačna samo za $c_i = o(i=1,\dots,n)$.

Definicija 1.1.3. Za beskonačno mnogo vektora kažemo da su linearno nezavisni, ako je svaki sistem konačnog broja tih vektora linearno nezavisan.

Definicija 1.1.4. Ako u vektorskom prostoru postoji n linearno nezavisnih vektora, a svaki sistem od n+1 vektora je linearno zavisan, kažemo da je prostor n-dimenzionalan. U protivnom slučaju kada u vektorskom prostoru postoji beskonačno mnogo linearno nezavisnih vektora, kažemo da je prostor beskonačno-dimenzionalan.

Definicija 1.1.5. Neka je $A = \{u_1,\dots,u_m\}$, gde su $u_k (k=1,\dots,m)$ vektori prostora X. Skup svih linearnih kombinacija ovih vektora naziva se lineal nad A i označava se sa L(A).

Definicija 1.1.6. Skup B linearno nezavisnih vektora prostora X obrazuje algebarsku ili Hamelovu bazu prostora X, ako je $L(B) = X$.

Teorema 1.1.1. Svaki vektor vektorskog prostora X može se na jedinstven način izraziti kao linearna kombinacija vektora algebarske baze tog prostora.

Napomenimo da vektorski prostor ima beskonačno mnogo različitih baza, medjutim, sve one imaju isti broj elemenata, tj. iste su kardinalnosti. Kod n-dimenzionalnog prostora baza sadrži n vektora.

Definicija 1.1.7. Svaka baza prostora naziva se koordinatni sistem tog prostora.

Neka je $B = \{u_1,\dots,u_n\}$ jedna baza prostora X. Tada se, na osnovu teoreme 1.1.1 , vektor $u \in X$ može predstaviti u obliku $u = x_1 u_1 + \dots + x_n u_n$. Dakle, ako je zadata baza B, vektor u je potpuno odredjen skalarima $x_1, \dots, x_n$ i može se korišćenjem matrične notacije predstaviti u obliku

$$u = [x_1 \ \dots \ x_n]^T .$$

Skalari $x_1, \dots, x_n$ nazivaju se koordinatama vektora.

Primer 1.1.1. Neka je $R^n = \{(x_1, \dots, x_n) \mid x_i \in R (i=1, \dots, n)\}$. Ako u ovaj skup uvedemo unutrašnju i spoljašnju kompoziciju pomoću

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n),$$

$$c(x_1, \dots, x_n) = (cx_1, \dots, cx_n),$$

on postaje vektorski prostor. Jedna baza ovog prostora je skup $\{e_1, e_2, \dots, e_n\}$, gde su

$$e_1 = (1,0,\dots,0), \ e_2 = (0,1,\dots,0), \dots, \ e_n = (0,0,\dots,1).$$

Ukoliko drugačije nije rečeno, uvek ćemo u daljem tekstu podrazumevati da je u prostoru R^n zadata pomenuta baza, koju nazivamo i prirodnom bazom. Za tačke ovog prostora koristićemo ravnopravno oznake

$$\vec{x} = (x_1, \dots, x_n) = [x_1 \ \dots \ x_n]^T .$$

Primer 1.1.2. Skup $\ell^r (r \geq 1)$ realnih(ili kompleksnih) nizova $u = \{x_k\}_{k \in N}$ za koje važi $\sum_{k=1}^{+\infty} |x_k|^r < +\infty$, obrazuje vektorski prostor ako su unutrašnja i spoljašnja kompozicija uvedene sa

$$u + v = \{x_k\}_{k \in N} + \{y_k\}_{k \in N} = \{x_k + y_k\}_{k \in N}, \quad (cu) = c\{x_k\}_{k \in N}.$$

Primer 1.1.3. Ako u skup neprekidnih funkcija $C[a,b]$ uvedemo unutrašnju i spoljašnju kompoziciju sa

$$(u+v)(t) = u(t) + v(t) \quad \text{i} \quad (cu)(t) = cu(t),$$

ovaj postaje vektorski prostor. Nula-vektor ovog prostora je funkcija koja je identički jednaka nuli na $[a,b]$. Isti je slučaj i sa skupom n puta neprekidno-diferencijabilnih funkcija $C^n[a,b]$.

Primer 1.1.4. Skup funkcija $L^r(a,b)$ $(r \geq 1)$ za koje je $\int_a^b |u(t)|^r dt < +\infty$ obrazuje vektorski prostor ako je unutrašnja i spoljašnja kompozicija uvedena kao u prethodnom primeru. Nula-vektor ovog prostora je funkcija koja je skoro svuda jednaka nuli na $[a,b]$.

2.1.2. Banachov prostor

Definicija 1.2.1. Linearni prostor X (nad poljem K) je normiran ako postoji nenegativna funkcija $u \mapsto \|u\|$ definisana za svako $u \in X$, koju nazivamo norma od u, takva da je

1° $\|u\| = 0 \Leftrightarrow u = \Theta$ (definisanost);

2° $\|cu\| = |c| \cdot \|u\|$ (homogenost);

3° $\|u+v\| \leqq \|u\| + \|v\|$ (relacija trougla),

gde su $u,v \in X$ i $c \in K$.

U normirani prostor uvodi se metrika pomoću

$$\rho(u,v) = \|u-v\| .$$

Primedba 1.2.1. Metrika na skupu X je funkcija $\rho : X^2 \mapsto [0,+\infty)$ sa sledećim svojstvima

1° $\rho(u,v) = 0 \Leftrightarrow u = v$;

2° $\rho(u,v) = \rho(v,u)$;

3° $\rho(u,v) + \rho(v,w) \geq \rho(u,w)$.

Primer 1.2.1. Vektorski prostor R^n se može normirati uvodjenjem

$$\|\vec{x}\|_p = \left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p} \qquad (1 \leqq p < +\infty) \tag{1.2.1}$$

ili

$$\|\vec{x}\|_\infty = \max_k |x_k| .$$

Od normi (1.2.1) najčešće se koriste norme za $p=1$ i $p=2$, tj.

$$\|\vec{x}\|_1 = \sum_{k=1}^{n} |x_k| \quad \text{i} \quad \|\vec{x}\|_2 = \left(\sum_{k=1}^{n} |x_k|^2\right)^{1/2} .$$

Poslednja norma poznata je kao Euklidska norma i često se označava sa $\|\vec{x}\|_E$.

Primer 1.2.2. Prostor $C[a,b]$ se može normirati, na primer, uvodjenjem norme pomoću

$$(1.2.2) \qquad \|u\| = \max_{a \leqq t \leqq b} |u(t)|$$

ili pomoću

$$(1.2.3) \qquad \|u\| = \int_a^b |u(t)|\,dt\,.$$

Primer 1.2.3. Prostor $L^r(a,b)$ se normira uvodjenjem norme pomoću

$$(1.2.4) \qquad \|u\| = \left(\int_a^b |u(t)|^r dt\right)^{1/r}.$$

Primer 1.2.4. U prostor ℓ^r norma se uvodi pomoću

$$(1.2.5) \qquad \|u\| = \left(\sum_{k=1}^{+\infty} |x_k|^r\right)^{1/r},$$

gde je $u = \{x_k\}_{k \in N}$.

Definicija 1.2.2. Neka je $\{u_n\}_{n\in N}$ niz tačaka u normiranom prostoru X i neka je $u \in X$ takvo da je $\lim\limits_{n\to+\infty} \|u_n - u\| = 0$. Tada kažemo da ovaj niz konvergira po normi ka tački u .

Definicija 1.2.3. Niz $\{u_n\}_{n\in N}$ za koji je $\lim\limits_{n,m\to+\infty} \|u_n - u_m\| = 0$ naziva se Cauchy ev niz.

Primetimo da je svaki konvergentan niz Cauchy ev niz, s obzirom da

$$\|u_n - u_m\| = \|(u_n - u) + (u - u_m)\| \leqq \|u_n - u\| + \|u_m - u\| \to 0,$$

kada $m,n \to +\infty$. Obrnuto, svaki Cauchy ev niz nije konvergentan. Na primer, granična vrednost niza neprekidnih funkcija ne mora biti neprekidna funkcija.

Definicija 1.2.4. Normiran vektorski prostor je kompletan ako u njemu svaki Cauchy ev niz konvergira.

Definicija 1.2.5. Kompletan normirani prostor naziva se Banach ov prostor.

Da li je vektorski prostor kompletan ili nije, zavisi od uvedene norme. Tako na primer, prostori ℓ^r i $L^r(a,b)$ su kompletni u odnosu na norme (1.2.4) i (1.2.5) respektivno, dok je prostor $C[a,b]$ kompletan u odnosu na normu (1.2.2), a nije kompletan u odnosu na (1.2.3).

2.1.3. Hilbertov prostor

Definicija 1.3.1. Vektorski prostor nad poljem kompleksnih brojeva ($K = C$) naziva se prostor sa skalarnim proizvodom ili unitaran prostor, ako postoji funkcija $(u,v):X^2 \to C$ koja zadovoljava sledeće uslove

1° $(u,u) \geqq 0$;

2° $(u,u) = 0 \Leftrightarrow u = \theta$;

3° $(u+v,w) = (u,w) + (v,w)$;

4° $(cu,v) = c(u,v)$;

5° $(u,v) = \overline{(v,u)}$

za svako $u,v,w \in X$ i $c \in C$.

Funkcija (u,v) se naziva skalarni proizvod.

Teorema 1.3.1. Za skalarni proizvod važi

$$(u,cv) = \bar{c}(u,v),$$

$$(u,v_1 + v_2) = (u,v_1) + (u,v_2),$$

$$|(u,v)|^2 \leqq (u,u)(v,v) \quad \text{(Schwarz ova nejednakost).}$$

Unitaran vektorski prostor može se normirati uvodjenjem norme pomoću

$$\|u\| = \sqrt{(u,u)}, \tag{1.3.1}$$

s obzirom da funkcija $u \mapsto \sqrt{(u,u)}$ ispunjava sve uslove definicije 1.2.1.

Definicija 1.3.2. Unitaran vektorski prostor sa normom (1.3.1) naziva se pred-Hilbertov prostor. Ukoliko je ovaj prostor kompletan naziva se Hilbertov.

Primer 1.3.1. Vektorski prostor R^n postaje Hilbertov prostor ako se skalarni proizvod uvede pomoću

$$(\vec{x},\vec{y}) = \sum_{k=1}^{n} x_k y_k = \vec{y}^T \vec{x} .$$

Slično, kompleksan vektorski prostor C^n postaje Hilbertov sa skalarnim proizvodom

$$(\vec{x},\vec{y}) = \sum_{k=1}^{n} x_k \bar{y}_k = \vec{y}^* \vec{x} ,$$

pri čemu $\vec{y}^*$ označava vektor koji se dobija transponovanjem vektora $\vec{y}$ i konjugovanjem njegovih komponenata. Norma koja izvire iz skalarnog proizvoda, u ovom slučaju je data sa

$$\|\vec{x}\| = \sqrt{(\vec{x},\vec{x})} = \left(\sum_{k=1}^{n} |x_k|^2 \right)^{1/2} ,$$

što ustvari predstavlja Euklidsku normu $\|\vec{x}\|_E$ (videti primer 1.2.1.).

Primer 1.3.2. Ako se u prostor ℓ^2 uvede skalarni proizvod pomoću

$$(u,v) = \sum_{k=1}^{+\infty} x_k \bar{y}_k \qquad (u = \{x_k\}_{k \in N} , \ v = \{y_k\}_{k \in N}),$$

dobija se Hilbertov prostor.

Primer 1.3.3. Sa skalarnim proizvodom

$$(1.3.2) \qquad (u,v) = \int_a^b u(t)\overline{v(t)}dt ,$$

prostor $L^2(a,b)$ postaje Hilbertov. U slučaju kada su elementi iz $L^2(a,b)$ samo realne funkcije, (1.3.2) se svodi na

$$(u,v) = \int_a^b u(t)v(t)dt .$$

Umesto skalarnog proizvoda (1.3.2) može se uzeti opštiji skalarni proizvod definisan sa

$$(u,v) = \int_a^b p(t)u(t)\overline{v(t)}dt ,$$

gde je $p: [a,b] \to R^+$ data neprekidna funkcija (težinska funkcija). Norma koja izvire iz skalarnog proizvoda uvedenog na ovaj način data je sa

$$\|u\| = \left(\int_a^b p(t)|u(t)|^2 dt \right)^{1/2} .$$

2.1.4. Ortogonalni sistemi u Hilbertovom prostoru

Definicija 1.4.1. Skup vektora $\{u_k\}_{k\in I}$ u Hilbertovom prostoru X obrazuje ortogonalan sistem ako je

$$(u_n,u_k) = \delta_{nk}\|u_k\|^2 \qquad (\forall n,k\in I),$$

gde je δ_{nk} - Kroneckerova delta i $\|u_k\| = \sqrt{(u_k,u_k)}$.

Skup indeksa može biti konačan, prebrojiv ili neprebrojiv.

Ukoliko je $\|u_k\| = 1$ $(\forall k\in I)$ kažemo da skup vektora $\{u_k\}_{k\in I}$ u X obrazuje ortonormiran sistem i takav sistem označavaćemo sa $\{u_k^*\}_{k\in I}$.

Definicija 1.4.2. Ortonormiran sistem $\{u_k^*\}_{k\in I}$ je potpun u X ako nije pravi deo nekog ortonormiranog sistema.

Teorema 1.4.1. U svakom Hilbertovom prostoru X $(\neq \{0\})$ postoji ortonormirani sistem.

<u>Teorema 1.4.2.</u> Neka su a_k Fourier ovi koeficijenti vektora $u \in X$ u odnosu na ortonormirani sistem $\{u_k^*\}_{k \in I}$. Tada su iskazi

1° $\{u_k^*\}_{k \in I}$ je potpun sistem u X

i

2° $u = \sum_{k \in I} a_k u_k^*$ za svako $u \in X$,

medjusobno ekvivalentni.

Zbog ekvivalentnosti iskaza 1° i 2° u prethodnoj teoremi, potpun ortonormiran sistem nazivamo i ortonormirana baza prostora. Slično, potpun ortogonalni sistem nazivamo ortogonalna baza prostora.

Kako, zbog jednostavnosti u primenama, ortogonalna baza ima prednosti nad algebarskom bazom u Hilbert ovom prostoru, od interesa je proučiti postupak za konstrukciju ortogonalne baze.

Neka je dat najviše prebrojiv skup linearno nezavisnih vektora $\{v_o, v_1, \dots\}$. Postupak kojim se ovom skupu vektora može pridružiti ortogonalan sistem vektora $\{u_o, u_1, \dots\}$, tako da se lineali nad ovim skupovima poklapaju, poznat je kao Gram-Schmidt ov postupak ortogonalizacije i on se može iskazati na sledeći način.

Uzmimo najpre $u_o = v_o$, a zatim u_1 predstavimo u obliku

$$u_1 = v_1 + \lambda_{1o} u_o ,$$

gde je λ_{1o} nepoznati parametar koji odredjujemo iz uslova da je vektor u_1 ortogonalan na u_o. Tada iz

$$(u_1, u_o) = (v_1, u_o) + \lambda_{1o}(u_o, u_o) = 0$$

sleduje

$$\lambda_{1o} = - \frac{(v_1, u_o)}{(u_o, u_o)} .$$

Pretpostavimo sada da smo već konstruisali vektore $u_o, u_1, \dots, u_{k-1}$. Vektor u_n predstavimo u obliku

$$u_k = v_k + \lambda_{ko}u_o + \lambda_{k1}u_1 + \cdots + \lambda_{k,k-1}u_{k-1} .$$

Nepoznate parametre λ_{ki} $(i = o,1,\ldots,k-1)$ odredjujemo iz uslova

$$(u_k,u_i) = (v_k,u_i) + \sum_{j=o}^{k-1} \lambda_{kj}(u_j,u_i) = 0 \quad (i=0,1,\ldots,k-1).$$

Dakle, imamo

$$\lambda_{ki} = -\frac{(v_k,u_i)}{(u_i,u_i)} \qquad (i=0,1,\ldots,k-1),$$

tj.

$$u_k = v_k - \sum_{i=o}^{k-1} \frac{(v_k,u_i)}{(u_i,u_i)} u_i \qquad (k=1,2,\ldots).$$

Odgovarajući ortonormirani sistem vektora je $\{u_o^*, u_1^*, \ldots\}$, gde je $u_k^* = \dfrac{u_k}{\|u_k\|}$ $(k=0,1,\ldots)$.

U specijalnom slučaju, navedenim postupkom, polazeći od prirodnog bazisa $\{1,x,x^2,\ldots\}$, moguće je u prostoru $X = L^2(a,b)$, u kome je skalarni proizvod uveden pomoću

$$(f,g) = \int_a^b p(x)f(x)g(x)dx \qquad (f,g\in X),$$

gde je $x \mapsto p(x)$ nenegativna težinska funkcija, konstruisati ortogonalni bazis $\{Q_k\}_{k\in N_o}$. Lineal nad ovim bazisom je skup svih algebarskih polinoma, koji je svuda gust* u X(videti [1]). U zavisnosti od težinske funkcije p dobijaju se različite klase ortogonalnih polinoma.

Daćemo dva primera.

Primer 1.4.1. U prostoru $L^2(-1,1)$ sa $p(x) = 1$ ortonormirani bazis obrazuje niz polinoma

---

*) Skup A je svuda gust u B ako se u svakoj okolini bilo koje tačke iz B nalazi bar jedna tačka skupa A.

$$P_k^*(x) = \sqrt{\frac{2k+1}{2}}\, P_k(x) \qquad (k = 0, 1, \ldots),$$

gde je $P_k(x) = \dfrac{1}{2^k k!} \cdot \dfrac{d^k}{dx^k}(x^2 - 1)^k$ (Polinom P_k naziva se Legendre ov polinom).

<u>Primer 1.4.2</u>. U prostoru $X = L^2(-1,1)$ sa $p(x) = (1 - x^2)^{3/2}$, odredićemo prvih pet članova ortogonalnog sistema $\{Q_k\}_{k \in N_o}$.

Izračunaćemo najpre integral

$$N_k = \int_{-1}^{+1} x^{2k}(1 - x^2)^{3/2} dx \qquad (k \in N_o).$$

Primenom parcijalne integracije na integral

$$N_{k-1} - N_k = \int_{-1}^{+1} x^{2(k-1)}(1 - x^2)^{5/2} dx \qquad (k \in N),$$

dobijamo $N_{k-1} - N_k = \dfrac{5}{2k-1} N_k$, tj. $N_k = \dfrac{2k-1}{2k+4} N_{k-1}$ $(k \in N)$.

Kako je $N_o = \dfrac{3\pi}{8}$, imamo $N_k = 3\pi \dfrac{(2k-1)!!}{(2k+4)!!}$ $(k \in N)$.

Polazeći od prirodnog bazisa $\{1, x, x^2, \ldots\}$, Gram-Schmidt-ovim postupkom ortogonalizacije nalazimo redom

$$Q_o(x) = 1,$$

$$Q_1(x) = x - \frac{(x, Q_o)}{(Q_o, Q_o)} Q_o(x) = x,$$

$$Q_2(x) = x^2 - N_1 N_o^{-1} = x^2 - \frac{1}{6},$$

$$Q_3(x) = x^3 - N_2 N_1^{-1} x = x^3 - \frac{3}{8} x,$$

$$Q_4(x) = x^4 - N_2 N_o^{-1} - (N_3 - \tfrac{1}{6} N_2)(N_2 - \tfrac{1}{3} N_1 + \tfrac{1}{36} N_o)^{-1}(x^2 - \tfrac{1}{6})$$

$$= x^4 - \frac{3}{5} x^2 + \frac{3}{80}.$$

Prva četiri člana odgovarajućeg ortonormiranog sistema $\{Q_k^*\}_{k\in N_o}$ su redom

$$Q_o^*(x) = \sqrt{\frac{8}{3\pi}},$$

$$Q_1^*(x) = \frac{4x}{\sqrt{\pi}},$$

$$Q_2^*(x) = 8\sqrt{\frac{6}{5\pi}}\,(x^2 - \frac{1}{6}),$$

$$Q_3^*(x) = \frac{32}{\sqrt{3\pi}}(x^3 - \frac{3}{8}x).$$

2.1.5. Neke opšte osobine ortogonalnih polinoma

U ovom odeljku ukazaćemo na neke osobine ortogonalnih polinoma, koje ćemo koristiti u teoriji aproksimacija i numeričkoj integraciji (opširnije o ortogonalnim polinomima može se naći, na primer u [2],[3],[4],[5]).

Neka je $\{Q_k\}_{k\in N_o}$ ortogonalan niz polinoma na (a,b) sa težinskom funkcijom $x\mapsto p(x)$.

Primedba 1.5.1. Niz $\{c_k Q_k\}_{k\in N_o}$, gde su $c_k(\neq o)$ konstante, takodje je ortogonalan na (a,b) sa težinskom funkcijom $x\mapsto p(x)$.

Teorema 1.5.1. Lineal nad bazisom $\{Q_o, Q_1, \ldots, Q_n\}$ je prostor algebarskih polinoma stepena ne višeg od n.

Dokaz. S obzirom da je

$$\int_a^b p(x)Q_k(x)\left(\sum_{i=0}^{n} \alpha_i Q_i(x)\right) dx = \alpha_k \|Q_k\|^2 \quad (0 \leqq k \leqq n),$$

uslov $\sum_{i=0}^{n} \alpha_i Q_i(x) = 0$ implicira $\alpha_i = 0$ $(i=0,1,\ldots,n)$, odakle neposredno sleduje tvrdjenje teoreme.

Teorema 1.5.2. Sve nule polinoma Q_k $(k=1,2,\ldots)$ su realne, proste i leže u intervalu (a,b).

<u>Dokaz</u>. Kako je, zbog ortogonalnosti,

$$\int_a^b p(x)Q_k(x)dx = \frac{1}{Q_o(x)}(Q_k, Q_o) = 0 \qquad (k=1,2,\ldots),$$

zaključujemo da polinom Q_k menja znak bar u jednoj tački iz (a,b). Pretpostavimo da su $x_1,\ldots,x_m (m \leqq k)$ realne nule neparnog reda polinoma Q_k, koje leže u (a,b). Definišimo polinom P pomoću

$$P(x) = Q_k(x)\omega(x),$$

gde je $\omega(x) = (x-x_1)\ldots(x-x_m)$.

Kako se polinom $x \mapsto \omega(x)$ može predstaviti u obliku

$$\omega(x) = \sum_{i=o}^{m} \alpha_i Q_i(x),$$

imamo

$$\int_a^b p(x)P(x)dx = (Q_k,\omega) = \sum_{i=o}^{m} \alpha_i (Q_k, Q_i),$$

tj.

$$\begin{aligned} (Q_k,\omega) &= 0 && (m < k), \\ &= \alpha_k \|Q_k\|^2 && (m = k). \end{aligned}$$

S druge strane, primetimo da polinom P ne menja znak na (a,b) što implicira $(Q_k,\omega) \neq 0$.

Na osnovu prethodnog zaključujemo da mora biti $m = k$, čime je dokaz završen.

<u>Teorema 1.5.3</u>. Izmedju tri uzastopna polinoma u nizu $\{Q_k\}_{k\in N_o}$ postoji rekurentna relacija

$$(1.5.1) \qquad Q_{k+1}(x) - (\alpha_k x + \beta_k)Q_k(x) + \gamma_k Q_{k-1}(x) = 0$$

gde su $\alpha_k, \beta_k, \gamma_k$ konstante.

Dokaz. Izaberimo α_k takvo da je polinom R definisan pomoću $R(x) = Q_{k+1}(x) - \alpha_k x Q_k(x)$ stepena k. Na osnovu teoreme 1.5.1, ovaj polinom možemo predstaviti u obliku

$$Q_{k+1}(x) - \alpha_k x Q_k(x) = \beta_k Q_k(x) + \sum_{j=1}^{k} \gamma_j Q_{j-1}(x), \tag{1.5.2}$$

odakle je

$$(Q_i, Q_{k+1}) - \alpha_k (xQ_i, Q_k) = \beta_k (Q_i, Q_k) + \sum_{j=1}^{k} \gamma_j (Q_i, Q_{j-1}).$$

Ako je $0 \leqq i \leqq k-2$, korišćenjem ortogonalnosti niza $\{Q_k\}_{k\in N_o}$, zaključujemo da je $\gamma_j = 0 (j=1,\ldots,k-1)$. Dakle, jednakost (1.5.2) se svodi na (1.5.1).

Ovim je dokaz završen.

Primedba 1.5.2. Rekurentna relacija (1.5.1) je tačna i za k=0, ako stavimo $Q_{-1}(x) = 0$.

Pokazaćemo sada kako se koeficijenti $\alpha_k, \beta_k, \gamma_k$, koji se pojavljuju u rekurentnoj relaciji (1.5.1), mogu izraziti pomoću koeficijenata polinoma iz ortogonalnog niza $\{Q_k\}_{k\in N_o}$.

Neka je a_k koeficijent uz x^k i b_k koeficijent uz x^{k-1} u polinomu Q_k, tj.

$$Q_k(x) = a_k x^k + b_k x^{k-1} + \cdots. \tag{1.5.3}$$

Tada, na osnovu (1.5.1) i (1.5.3), imamo

$$(a_{k+1}x^{k+1} + b_{k+1}x^k + \cdots) - (\alpha_k x + \beta_k)(a_k x^k + b_k x^{k-1} + \cdots)$$
$$+ \gamma_k (a_{k-1}x^{k-1} + b_{k-1}x^{k-2} + \cdots) = 0,$$

odakle sleduje

$$\alpha_k = \frac{a_{k+1}}{a_k} \quad \text{i} \quad \beta_k = \alpha_k \left(\frac{b_{k+1}}{a_{k+1}} - \frac{b_k}{a_k} \right) \qquad (k\in N). \tag{1.5.4}$$

Da bismo odredili γ_k podjimo od rekurentne relacije

$$Q_k(x) - (\alpha_{k-1}x + \beta_{k-1})\, Q_{k-1}(x) + \gamma_{k-1}Q_{k-2}(x) = 0.$$

Množenjem ove jednakosti sa $p(x)Q_k(x)$ i integracijom od a do b, uz korišćenje ortogonalnosti niza $\{Q_k\}_{k\in N_o}$, dobijamo

$$(1.5.5) \qquad \|Q_k\|^2 = \alpha_{k-1} \int_a^b p(x)\, x\, Q_{k-1}(x)Q_k(x)dx .$$

Slično, množenjem jednakosti (1.5.1) sa $p(x)Q_{k-1}(x)$ i odgovarajućom integracijom, dobijamo

$$0 = \alpha_k \int_a^b p(x)xQ_{k-1}(x)\,Q_k(x)\,dx - \gamma_k\|Q_{k-1}\|^2 .$$

Najzad, iz poslednje dve jednakosti sleduje

$$(1.5.6) \qquad \gamma_k = \frac{\alpha_k}{\alpha_{k-1}}\left(\frac{\|Q_k\|}{\|Q_{k-1}\|}\right)^2 = \frac{a_{k+1}a_{k-1}}{a_k^2}\left(\frac{\|Q_k\|}{\|Q_{k-1}\|}\right)^2 \qquad (k\in N).$$

Sledeći rezultat poznat je kao Christoffel-Darbouxov identitet.

<u>Teorema 1.5.4</u>. Za ortogonalan niz polinoma $\{Q_k\}_{k\in N_o}$ važi identitet

$$(1.5.7) \qquad \sum_{k=o}^{n} \frac{Q_k(x)Q_k(t)}{\|Q_k\|^2} = \frac{1}{\alpha_n\|Q_n\|^2}\cdot\frac{Q_{n+1}(x)Q_n(t) - Q_n(x)Q_{n+1}(t)}{x - t},$$

gde je α_n konstanta koja se javlja u rekurentnoj relaciji (1.5.1).

<u>Dokaz</u>. Množenjem relacije (1.5.1) sa $Q_k(t)$, dobijamo

$$(1.5.8) \qquad Q_{k+1}(t)Q_k(t) = (\alpha_k x+\beta_k)Q_k(x)Q_k(t) - \gamma_k Q_{k-1}(x)Q_k(t).$$

Ako u poslednjoj jednakosti zamenimo x sa t i t sa x i tako dobijenu jednakost oduzmemo od (1.5.8), dobijamo

$$Q_{k+1}(x)Q_k(t) - Q_{k+1}(t)Q_k(x) = \alpha_k(x-t)Q_k(x)Q_k(t)$$

$$- \gamma_k(Q_{k-1}(x)Q_k(t) - Q_{k-1}(t)Q_k(x)),$$

tj.

$$(x-t)\frac{Q_k(x)Q_k(t)}{\|Q_k\|^2} = \frac{1}{\alpha_k\|Q_k\|^2}(Q_{k+1}(x)Q_k(t) - Q_{k+1}(t)Q_k(x))$$

$$- \frac{\gamma_k}{\alpha_k\|Q_k\|^2}(Q_k(x)Q_{k-1}(t) - Q_k(t)Q_{k-1}(x)).$$

Ako stavimo $R_k = \dfrac{1}{\alpha_k\|Q_k\|^2}(Q_{k+1}(x)Q_k(t) - Q_{k+1}(t)Q_k(x))$,

i iskoristimo (1.5.6), poslednja jednakost postaje

$$(x-t)\frac{1}{\|Q_k\|^2}Q_k(x)Q_k(t) = R_k - R_{k-1},$$

odakle sumiranjem po k, dobijamo (1.5.7).

2.1.6. Klasični ortogonalni polinomi

Definicija 1.6.1. Neka je $\{Q_k\}_{k\in N_0}$ niz ortogonalnih polinoma na (a,b) sa težinskom funkcijom $x \mapsto p(x)$. Polinomi $Q_k (k\in N_0)$ nazivaju se klasičnim, ako težinska funkcija zadovoljava diferencijalnu jednačinu

$$\frac{d}{dx}(A(x)p(x)) = B(x)p(x), \tag{1.6.1}$$

gde je $x \mapsto B(x)$ polinom prvog stepena, a funkcija $x \mapsto A(x)$ u zavisnosti od a i b ima oblik

$$
\begin{aligned}
A(x) &= (x-a)(b-x) && (a \text{ i } b \text{ konačni}),\\
&= x-a && (a \text{ konačno}, b=+\infty),\\
&= b-x && (a=-\infty, b \text{ konačno}),\\
&= 1 && (a=-\infty, b=+\infty).
\end{aligned}
$$

Teorema 1.6.1. Težinska funkcija klasičnih ortogonalnih polinoma je neprekidno-diferencijabilna funkcija na (a,b) i zadovoljava uslove

$$\text{(1.6.2)} \qquad \lim_{x \to a+} x^m A(x)p(x) = 0 \quad \text{i} \quad \lim_{x \to b-} x^m A(x)p(x) = 0$$

za svako $m=0,1,\ldots$

Kako je rešenje diferencijalne jednačine (1.6.1) dato sa

$$p(x) = \frac{C}{A(x)} \exp\left(\int \frac{B(x)}{A(x)}\,dx\right),$$

gde je C proizvoljna konstanta, u zavisnosti od a i b (sa tačnošću do multiplikativne konstante) imamo

$$
\text{(1.6.3)} \qquad
\begin{aligned}
p(x) &= (x-a)^{\alpha}(b-x)^{\beta} && (a \text{ i } b \text{ konačni}),\\
&= (x-a)^{s}\exp(rx) && (a \text{ konačno}, b=+\infty),\\
&= (b-x)^{t}\exp(-rx) && (a=-\infty, b \text{ konačno}),\\
&= \exp\left(\int B(x)\,dx\right) && (a=-\infty, b=+\infty),
\end{aligned}
$$

gde su

$$\alpha = \frac{B(a)}{b-a} - 1, \quad \beta = -\frac{B(b)}{b-a} - 1, \quad s = B(a) - 1, \quad t = -B(b) - 1$$

i $B(x) = rx + q$.

Granični uslovi (1.6.2) se tada svode na

$$B(a) > 0 \qquad \text{(ako je a konačno)},$$

$$(1.6.4) \qquad B(b) < 0 \qquad \text{(ako je b konačno)},$$

$$r = B'(x) < 0.$$

Na osnovu (1.6.3) zaključujemo kakav oblik mora da ima težinska funkcija kod klasičnih ortogonalnih polinoma. Kako se primenom linearne transformacije svaki interval (a,b) može transformisati u jedan od sledeća tri intervala

$$(-1, 1), \quad (0, +\infty), \quad (-\infty, +\infty),$$

ne umanjujući opštost, za težinsku funkciju klasičnih ortogonalnih polinoma se može uzeti

$$x \mapsto (1-x)^{\alpha}(1+x)^{\beta}, \quad x \mapsto x^{s}e^{-x}, \quad x \mapsto e^{-x^2}$$

respektivno. Tada, na osnovu (1.6.2), tj. (1.6.4), parametri α, β, s moraju ispunjavati uslove

$$\alpha > -1, \quad \beta > -1, \quad s > -1.$$

Razmotrimo posebno ova tri slučaja.

1. Neka je $p(x) = (1-x)^{\alpha}(1+x)^{\beta}$ i $(a,b) = (-1,1)$. Funkcija A je u ovom slučaju data sa $A(x) = 1 - x^2$. Korišćenjem diferencijalne jednačine za težinsku funkciju odredićemo funkciju B. Dakle,

$$B(x) = \frac{1}{p(x)} \cdot \frac{d}{dx}(A(x)p(x)) = \beta - \alpha - (\alpha + \beta + 2)x.$$

Odgovarajući ortogonalni polinomi nazivaju se Jacobievi polinomi i označavaju sa $P_k^{(\alpha,\beta)}(x)$. Specijalni slučajevi Jacobievih polinoma su

1° Legendreovi polinomi $P_k(x)$ $(\alpha = \beta = 0)$,

2° Čebiševljevi polinomi prve vrste $T_k(x)$ $(\alpha = \beta = -\frac{1}{2})$,

3° Čebiševljevi polinomi druge vrste $S_k(x)$ $(\alpha = \beta = \frac{1}{2})$,

4^o Gegenbauerovi ili ultrasferni polinomi $C_k^{\lambda}(x)$ ($\alpha = \beta = \lambda - \frac{1}{2}$).

Primetimo da klasa Gegenbauerovih polinoma obuhvata polinome navedene pod $1^o, 2^o, 3^o$.

2. Neka je sada $p(x) = x^s e^{-x}$ i $(a,b) = (o,+\infty)$. Tada je $A(x) = x$ i $B(x) = -x + s + 1$. Odgovarajući polinomi se nazivaju generalisani Laguerreovi polinomi i označavaju sa $L_k^s(x)$. Polinomi koji se dobijaju za $s=o$ nazivaju se Laguerreovi polinomi i označavaju se sa $L_k(x)$ ($\equiv L_k^o(x)$).

3. Najzad, neka je $p(x) = e^{-x^2}$ i $(a,b) = (-\infty, +\infty)$. S obzirom da je tada $A(x) = 1$, dobijamo

$$B(x) = e^{x^2} \frac{d}{dx}(e^{-x^2}) = -2x.$$

Polinomi koji se dobijaju u ovom slučaju nazivaju se Hermiteovi i označavaju sa $H_k(x)$.

<u>Teorema 1.6.2</u>. Za članove niza klasičnih ortogonalnih polinoma $\{Q_k\}_{k \in N_o}$ važi formula

$$Q_k(x) = \frac{C_k}{p(x)} \cdot \frac{d^k}{dx^k}(A(x)^k p(x)) \qquad (k=o,1,\ldots), \tag{1.6.5}$$

gde su C_k konstante različite od nule.

U formuli (1.6.5), koja je poznata kao Rodriguezova formula, konstanta C_k može se odabrati na razne načine (na primer, iz uslova da je $\|Q_k\| = 1$). Jedan način izbora konstante C_k, koji ima istorijski značaj, je

$$C_k = \begin{cases} \dfrac{(-1)^k}{2^k k!} & \text{za } P_k^{(\alpha,\beta)}(x), \\ 1 & \text{za } L_k^s(x), \\ (-1)^k & \text{za } H_k(x). \end{cases}$$

Za polinome Gegenbauera i Čebiševa prve i druge vrste koristi se normalizacija(izbor konstante C_k) koja se razlikuje od opšteg slučaja Jacobievih polinoma. Naime, uzima se

$$C_k^{\lambda}(x) = \frac{(2\lambda)_k}{(\lambda+\frac{1}{2})_k} P_k^{(\lambda-\frac{1}{2},\lambda-\frac{1}{2})}(x),$$

$$T_k(x) = \frac{k!}{\left(\frac{1}{2}\right)_k} P_k^{(-1/2,-1/2)}(x),$$

$$S_k(x) = \frac{(k+1)!}{\left(\frac{3}{2}\right)_k} P_k^{(1/2,1/2)}(x),$$

gde je uvedena oznaka

$$(s)_k = s(s+1)\ldots(s+k-1) = \frac{\Gamma(s+k)}{\Gamma(s)}.$$

Ako sa a_k označimo koeficijent uz najviši stepen u polinomu Q_k, tj. $Q_k(x) = a_k x^k + \cdots$, korišćenjem Rodriguezove formule (1.6.5), iz

$$\|Q_k\|^2 = \int_a^b p(x)Q_k(x)^2dx = \int_a^b p(x)Q_k(x)(a_kx^k+\cdots)^2dx$$

sleduje

$$\|Q_k\|^2 = a_kC_k(-1)^k k! \int_a^b A(x)^k p(x)dx.$$

U specijalnim slučajevima imamo:

1° Kod Jacobievih polinoma

$$a_k = \frac{1}{2^k k!}\cdot\frac{\Gamma(2k+\alpha+\beta+1)}{\Gamma(k+\alpha+\beta+1)}$$

i

$$\|P_k^{(\alpha,\beta)}\|^2 = \frac{2^{\alpha+\beta+1}\,\Gamma(k+\alpha+1)\Gamma(k+\beta+1)}{k!\,(2k+\alpha+\beta+1)\Gamma(k+\alpha+\beta+1)};$$

2^o Kod Laguerreovih polinoma

$$a_k = (-1)^k \qquad \text{i} \qquad \|L_k^s\|^2 = k!\,\Gamma(k+s+1);$$

3^o Kod Hermiteovih polinoma

$$a_k = 2^k \qquad \text{i} \qquad \|H_k\|^2 = 2^k k!\sqrt{\pi}\,.$$

<u>Teorema 1.6.3</u>. Polinom $x \mapsto Q_k(x)$, koji je član niza klasičnih ortogonalnih polinoma $\{Q_k\}_{k\in N_0}$, je jedno partikularno rešenje linearne diferencijalne jednačine drugog reda

$$A(x)y'' + B(x)y' + \lambda_k y = 0, \tag{1.6.6}$$

gde je

$$\lambda_k = -k\left[\frac{1}{2}(k-1)A''(x) + B'(x)\right].$$

Korišćenjem diferencijalne jednačine za težinsku funkciju (1.6.1), diferencijalna jednačina (1.6.6) se može predstaviti u Sturm-Liouvilleovom obliku

$$\frac{d}{dx}(A(x)p(x)y') + \lambda_k p(x)y = 0.$$

Na kraju, u vezi sa klasičnim ortogonalnim polinomima, navedimo sledeću teoremu:

<u>Teorema 1.6.4</u>. Niz $\{Q_k^{(m)}\}_{k\in N_0}$ predstavlja niz klasičnih ortogonalnih polinoma na (a,b) sa težinskom funkcijom $p_m(x) = A(x)^m p(x)$. Odgovarajuća diferencijalna jednačina za težinsku funkciju je

$$\frac{d}{dx}(A(x)p_m(x)) = B_m(x)p_m(x),$$

gde je $B_m(x) = mA'(x) + B(x)$.

2.2. OSNOVNE DEFINICIJE I STAVOVI O OPERATORIMA

2.2.1. Linearni operatori

Neka su X i Y Banach ovi prostori. Pod operatorom* $T:X\to Y$ podrazumeva se preslikavanje

$$u\mapsto g=Tu \qquad (u\in X,\ g\in Y).$$

Definicija 2.1.1. Operator $T:X\to Y$ je homogen ako je

$$T(cu)=cTu \qquad (\forall u\in X \text{ i } \forall c\in K).$$

Definicija 2.1.2. Operator $T:X\to Y$ je aditivan ako je

$$T(u_1+u_2)=Tu_1+Tu_2 \qquad (\forall u_1,u_2\in X).$$

Definicija 2.1.3. Operator $T:X\to Y$ je linearan ako je istovremeno homogen i aditivan, tj.

$$T(c_1u_1+c_2u_2)=c_1Tu_1+c_2Tu_2 \qquad (\forall u_1,u_2\in X;\forall c_1,c_2\in Y).$$

Definicija 2.1.4. Operator $T:X\to Y$ je neprekidan ako za svaki niz $\{u_n\}_{n\in N}$ iz X važi

$$u_n\to u \quad\Rightarrow\quad Tu_n\to Tu \qquad (u\in X).$$

Definicija 2.1.5. Linearni operator $T:X\to Y$ je ograničen ako postoji nenegativan broj M takav da je

$$\|Tu\|\leqq M\|u\| \qquad (\forall u\in X). \tag{2.1.1}$$

Infimum brojeva M za koje važi (2.1.1) označava se sa $\|T\|$ i naziva norma operatora T.

Teorema 2.1.1. Linearni operator $T:X\to Y$ je ograničen ako i samo ako je neprekidan.

*) Operator $T:X\to R$ nazivamo funkcionelom.

Dokaz. Ako je operator T ograničen važi

$$\|Tu_n - Tu\| = \|T(u_n - u)\| \leq \|T\| \|u_n - u\| ,$$

odakle pri $u_n \to u$ sleduje njegova neprekidnost u proizvoljnoj tački u.

Pretpostavimo sada, da operator T nije ograničen. Tada za proizvoljno $n > 0$ postoji par elemenata $u_n, v_n \in X$, takav da je

$$u_n \neq v_n \quad \text{i} \quad \|Tu_n - Tv_n\| > 2n\|u_n - v_n\| .$$

Neka je, dalje, $\{c_n\}$ niz racionalnih brojeva sa osobinom

$$\|u_n - v_n\| < c_n < 2\|u_n - v_n\| .$$

Tada, za $k_n = \frac{1}{nc_n}(u_n - v_n)$, imamo

$$\|Tk_n\| = \frac{1}{nc_n}\|Tu_n - Tv_n\| > \frac{2n}{nc_n}\|u_n - v_n\| = \frac{2}{c_n}\|u_n - v_n\| > 1$$

i

$$\|k_n\| = \frac{1}{nc_n}\|u_n - v_n\| < \frac{1}{n} .$$

Kako $\|k_n\| \to 0$ i $\|Tk_n\| > 1$ $(n \to +\infty)$ protivureči neprekidnosti operatora T, sleduje da operator T mora biti ograničen.

Neka su T_1 i T_2 dva operatora koji preslikavaju prostor X u prostor Y.

Definicija 2.1.6. Zbir operatora $T = T_1 + T_2$ definiše se kao

$$Tu = (T_1 + T_2)u = T_1u + T_2u .$$

Ako su T_1 i T_2 linearni ograničeni operatori važi

$$\|Tu\| = \|T_1u + T_2u\| \leq \|T_1u\| + \|T_2u\| \leq (\|T_1\| + \|T_2\|)\|u\| ,$$

tj.

$$\|T_1 + T_2\| \leq \|T_1\| + \|T_2\| .$$

Neka je T_2u tačka prostora X.

Definicija 2.1.7. Proizvod operatora $T = T_1T_2$ definiše se pomoću

$$Tu = T_1T_2u = T_1(T_2u).$$

Ako su T_1 i T_2 ograničeni operatori važi

$$\|Tu\| = \|T_1(T_2u)\| \leq \|T_1\| \|T_2u\| \leq \|T_1\| \|T_2\| \|u\|,$$

tj.

$$\|T_1T_2\| \leq \|T_1\| \cdot \|T_2\|.$$

Ako za svako $u \in X \Rightarrow Tu \in X$, tada se može definisati iterirani operator T^n(n-ti stepen operatora T) kao

$$T^n = T(T^{n-1}) \qquad (\forall n \in N),$$

pri čemu je $T^0 = I$ identički operator ($Iu = u$ za svako $u \in X$).

Za operatore T^n i T^m($n,m \in N_0$) važi jednakost

$$T^nT^m = T^{n+m}.$$

Ako je T ograničen operator tada je

$$\|T^n\| \leq \|T\|^n \qquad (n \in N_0).$$

Definicija 2.1.8. Neka je $T:X \to Y$ linearan operator. Ako postoji preslikavanje T^{-1} takvo da je

$$T^{-1}(Tu) = u \qquad (\forall u \in X),$$

onda je T^{-1} inverzan operator od T.

Dakle, inverzan operator postoji ako je T biunivoko preslikavanje prostora X u Y, tj. ako važi implikacija

$$u \neq v \quad \Rightarrow \quad Tu \neq Tv,$$

što se (kod linearnih operatora) svodi na

$$Tu = \Theta \quad \Rightarrow \quad u = \Theta.$$

Teorema 2.1.2. Ako linearan operator $T:X\to Y$ ima inverzan operator T^{-1}, onda je T^{-1} takodje linearan.

Dokaz. Neka je $Tu_1 = g_1$ i $Tu_2 = g_2$, tj. $T^{-1}g_1 = u_1$ i $T^{-1}g_2 = u_2$. S obzirom da je T linearan operator imamo

$$\begin{aligned} T^{-1}(c_1g_1 + c_2g_2) &= T^{-1}(c_1Tu_1 + c_2Tu_2) \\ &= T^{-1}T(c_1u_1 + c_2u_2) \\ &= c_1u_1 + c_2u_2 \\ &= c_1T^{-1}g_1 + c_2T^{-1}g_2, \end{aligned}$$

čime je dokaz završen.

Primer 2.1.1. Odredimo operator, inverzan linearnom integralnom operatoru

$$Tf(x) = \int_a^b e^{-|\alpha(x)-\alpha(t)|}\, f(t)dt,$$

gde je α data monotono neopadajuća funkcija na $[a,b]$ i dvaput neprekidno-diferencijabilna.

Neka je $\alpha(x) - \alpha(t) = u$. Tada je

$$g(x) = Tf(x) = \int_a^x e^{-u} f(t)dt + \int_x^b e^{u} f(t)dt,$$

odakle sleduje

$$g'(x) = f(x) - \int_a^x e^{-u}\alpha'(x)f(t)dt - f(x) + \int_x^b e^{u}\alpha'(x)f(t)dt,$$

tj.

$$g'(x) = \alpha'(x)\left(-\int_a^x e^{-u}f(t)dt + \int_x^b e^{u}f(t)dt\right). \tag{2.1.2}$$

Na dalje, imamo

$$\frac{d}{dx}\left(\frac{g'(x)}{\alpha'(x)}\right) = -2f(x) + \alpha'(x)g(x),$$

odakle je

$$f(x) = T^{-1}g(x) = \frac{1}{2}\alpha'(x)g(x) - \frac{1}{2}\frac{d}{dx}\left(\frac{g'(x)}{\alpha'(x)}\right).$$

Kako iz (2.1.2) sleduje

$$g'(a) = \alpha'(a)g(a) \quad \text{i} \quad g'(b) = -\alpha'(b)g(b), \tag{2.1.3}$$

zaključujemo da je inverzan operator T^{-1} u prostoru $C^2[a,b]$ definisan za funkcije koje zadovoljavaju granične uslove (2.1.3).

Teorema 2.1.3. Neka je X Banachov prostor, I identički operator u X i $T:X \to X$ ograničen linearni operator kod koga je $\|T\| \leqq q < 1$. Tada postoji operator $(I-T)^{-1}$ za koji važi:

1° $\quad (I-T)^{-1} = \sum_{k=0}^{+\infty} T^k \quad$ (Neumannov razvoj);

2° $\quad \|(I-T)^{-1}\| \leqq \frac{1}{1-q}$.

Dokaz. S obzirom da je $\|T\| \leqq q < 1$ imamo

$$\sum_{k=0}^{+\infty} \|T^k\| \leqq \sum_{k=0}^{+\infty} \|T\|^k \leqq \frac{1}{1-q} < +\infty. \tag{2.1.4}$$

Kako je, dalje, prostor X kompletan, to iz konvergencije reda $\sum_{k=0}^{+\infty} \|T^k\|$ sleduje da je $\sum_{k=0}^{+\infty} T^k$ ograničen linearan operator.

Iz jednakosti

$$(I-T)\sum_{k=0}^{n} T^k = \sum_{k=0}^{n} T^k(I-T) = I - T^{n+1},$$

koja važi za svako $n \in N$, a s obzirom na

$$\|T^{n+1}\| \leqq \|T\|^{n+1} \to 0 \quad \Rightarrow T^{n+1} \to 0 \quad (n \to +\infty),$$

sleduje

$$(I - T)\sum_{k=0}^{+\infty} T^k = \sum_{k=0}^{+\infty} T^k(I - T) = I,$$

tj.

$$(I - T)^{-1} = \sum_{k=0}^{+\infty} T^k.$$

Korišćenjem (2.1.4) neposredno dobijamo

$$\|(I - T)^{-1}\| \leqq \sum_{k=0}^{+\infty} \|T^k\| \leqq \frac{1}{1-q}.$$

Teorema 2.1.4. Neka su $A:X \to X$ i $B:X \to X$ linearni operatori za koje postoje neprekidni inverzni operatori A^{-1} i B^{-1}. Tada je

$$\|A^{-1} - B^{-1}\| \leqq \frac{\|B^{-1}\| \cdot \|I - B^{-1}A\|}{1 - \|I - B^{-1}A\|},$$

gde je I identički operator u X.

Definicija 2.1.9. Neka je $T:X \to X$ linearan operator. Kompleksan broj λ je sopstvena ili karakteristična vrednost operatora T ako postoji vektor u različit od nula-vektora takav da je

$$Tu = \lambda u.$$

Vektor u koji odgovara sopstvenoj vrednosti λ naziva se sopstveni ili karakteristični vektor operatora T.

Teorema 2.1.5 (Banach-Steinhaus). Neka su X i Y Banachovi prostori i neka je $\{A_n\}_{n\in N}$ niz ograničenih linearnih operatora koji preslikavaju X u Y. Potrebni i dovoljni uslovi za konvergenciju niza $\{A_n u\}$ ka Au za svako $u \in X$, gde je $A:X \to Y$ ograničen linearan operator, su

1^o niz $\{\|A_n\|\}_{n\in N}$ ograničen;

2^o postoji $\lim\limits_{n\to+\infty} A_n u$ na nekom u X svuda gustom skupu.

Dokaz ove teoreme može se naći, na primer, u [1].

2.2.2. Matrica linearnog operatora na konačno-dimenzionalnim prostorima

Neka su X i Y konačno-dimenzionalni prostori sa bazama $B_u = \{u_1,\ldots,u_n\}$ i $B_v = \{v_1,\ldots,v_m\}$ respektivno i neka je $A:X\to Y$ linearan operator. Nije teško pokazati da je operator A potpuno odredjen ako su poznate slike vektora baze B_u, tj. ako su poznati vektori Au_i $(i=1,\ldots,n)$. Razložimo ove vektore po vektorima baze B_v. Tada imamo

$$\begin{aligned} Au_1 &= a_{11}v_1 + a_{21}v_2 + \cdots + a_{m1}v_m,\\ Au_2 &= a_{12}v_1 + a_{22}v_2 + \cdots + a_{m2}v_m,\\ &\vdots\\ Au_n &= a_{1n}v_1 + a_{2n}v_2 + \cdots + a_{mn}v_m. \end{aligned} \tag{2.2.1}$$

Na osnovu (2.2.1) formirajmo matricu

$$A_{vu} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & & a_{mn} \end{bmatrix}.$$

<u>Definicija 2.2.1.</u> Za A_{vu} kažemo da je matrica operatora $A:X\to Y$ u odnosu na baze B_u i B_v.

Posmatrajmo proizvoljne vektore $u\in X$ i $v\in Y$, čije su koordinatne reprezentacije, u bazama B_u i B_v, date sa

$$u = \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{i} \quad v = \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

respektivno. Veza izmedju koordinata vektora u i vektora v=Au može se iskazati pomoću jednakosti

$$\vec{y} = A_{vu}\vec{x} .$$

Formalno, operator A možemo zameniti njegovom matricom A_{vu}.

Da bismo opisali operator $A:X \to X$ dovoljno je fiksirati jednu bazu $B_u = \{u_1,\dots,u_n\}$. Naime, prethodno razmatranje ostaje u važnosti ako stavimo $Y = X$ i $B_v = B_u$. Matricu A_{uu}, u tom slučaju, prosto označavamo sa A_u.

Na kraju napomenimo da će o matričnom računu biti posebno reči u četvrtoj glavi ove knjige.

2.2.3. Bilinearni i n-linearni operatori

Skup ograničenih linearnih operatora L(X,Y) koji preslikavaju Banachov prostor X u Banachov prostor Y ima strukturu Banachovog prostora ako su u njega uvedene unutrašnja i spoljašnja kompozicija pomoću

$$(T+S)u = Tu+Su, \quad (cT)u = c(Tu) \quad (T,S \in L(X,Y);\forall c \in K),$$

dok se pod normom elementa $T \in L(X,Y)$ podrazumeva norma ograničenog linearnog operatora T u smislu definicije 2.1.5.

Neka su X i Y Banachovi prostori i operator $B:X^2 \to Y$.

Definicija 2.3.1. Operator B je bilinearan ako svakom uredjenom paru elemenata $(u,u') \in X^2$ odgovara element $g=B(u,u') \in Y$, pri čemu za svako $u_1,u_2,u_1',u_2' \in X$ i svako $c_1,c_2 \in K$ važe jednakosti:

$$B(c_1u_1+c_2u_2,u') = c_1B(u_1,u') + c_2B(u_2,u') ,$$

$$B(u,c_1u_1'+c_2u_2') = c_1B(u,u_1') + c_2B(u,u_2') .$$

Ako postoji pozitivan broj M takav da je

$$(2.3.1) \qquad \|B(u,u')\| \le M\|u\|\cdot\|u'\|$$

za svako $u,u' \in X$, operator B je ograničen.

Infimum brojeva M za koje važi (2.3.1) naziva se normom bilinearnog operatora B i označava se sa $\|B\|$.

Shodno prethodnoj definiciji bilinearna preslikavanja prostora X u prostor Y, obrazuju linearni normirani prostor, koji označavamo sa $B(X^2,Y)$. Može se pokazati da su prostori $L(X,L(X,Y))$ i $B(X^2,Y)$ izometrični.

Definicija 2.3.2. Operator $N:X^n \to Y$ je n-linearan ako svakom uredjenom sistemu elemenata $(u_1,u_2,\ldots,u_n)$ iz X odgovara element $g=N(u_1,u_2,\ldots,u_n) \in Y$, pri čemu je on linearan po svakom u_i, pri fiksiranim ostalim elementima $u_1,\ldots,u_{i-1},u_{i+1},\ldots,u_n$.

Ako postoji pozitivan broj M takav da je

$$\| N(u_1,u_2,\ldots,u_n) \| \leq M \| u_1 \| \cdot \| u_2 \| \cdots \| u_n \| ,$$

operator N je ograničen.

Sva n-linearna preslikavanja prostora X u prostor Y obrazuju linearni normiran prostor, koji označavamo sa $N(X^n,Y)$.

2.2.4. Fréchetova diferenciranja

Neka su X i Y Banachovi prostori i neka je nelinearni operator $T:X \to Y$ definisan na skupu $D \subset X$.

Definicija 2.4.1. Operator T je Fréchet-diferencijabilan u tački $u \in D$, ako postoji takav ograničeni linearan operator $A \in L(X,Y)$ da je

$$\lim_{\| h \| \to 0} \frac{\| T(u+h)-Tu-Ah \|}{\| h \|} = 0 .$$

Operator A naziva se izvod operatora T u tački u i označava sa $T'_{(u)}$.

Teorema 2.4.1. Ako je operator T Fréchet-diferencijabilan u tački u, onda je njegov izvod jedinstven.

Dokaz. Neka su $A_i (i=1,2)$ dva linearna ograničena operatora, takva da je

$$L_i = \lim_{\| h \| \to 0} \frac{\| T(u+h)-Tu-A_i h \|}{\| h \|} = 0 \qquad (i=1,2)$$

i neka je $A=A_1-A_2$.

Korišćenjem

$$\|Ah\| = \|A_1h-A_2h\| \leqq \|T(u+h)-Tu-A_1h\| + \|T(u+h)-Tu-A_2h\|$$

nalazimo

$$\lim_{\|h\|\to o} \frac{\|Ah\|}{\|h\|} = L_1 + L_2 = 0 . \tag{2.4.1}$$

Medjutim, ako je za neko h

$$\frac{\|Ah\|}{\|h\|} = \lambda \neq 0 ,$$

tada će i za svako $\varepsilon \neq 0$ biti

$$\frac{\|A(\varepsilon h)\|}{\|\varepsilon h\|} = \lambda ,$$

odakle zaključujemo da je jednakost (2.4.1) nemoguća kada je A nenula-operator.

Dakle, (2.4.1) važi ako je A nula-operator, tj. $A_1=A_2$, čime je dokaz završen.

Teorema 2.4.2. Svaki linearan ograničeni operator je Fréchet-diferencijabilan, pri čemu je $T_{(u)}\cdot = T\cdot$ za svako $u \in D$.

Dokaz. Kako je $T \in L(X,Y)$, na osnovu definiciju 2.4.1 i prethodne teoreme neposredno sleduje

$$T'_{(u)}h = Th \qquad \text{za svako } u \in D.$$

Sada ćemo dati neka pravila diferencijalnog računa:

1^o Neka su operatori $T:X\to Y$ i $H:X\to Y$ neprekidni i diferencijabilni u tački u. Tada je operator $\alpha T+\beta H (\alpha,\beta \in K)$ takodje diferencijabilan u toj tački, pri čemu je

$$(\alpha T + \beta H)'_{(u)} = \alpha T'_{(u)} + \beta H'_{(u)} ;$$

2^o Neka su X,Y,Z Banachovi prostori, a $T:X\to Y$ i $S:Y\to Z$ Fréchet-diferencijabilni operatori. Ako je $Tu=g$, $Sg=h$, izvod složenog operatora $ST(h=S(Tu)=(ST)u)$ u tački u je

$$(ST)'_{(u)} = S'_{(Tu)}T'_{(u)} ;$$

3^o Ako je A proizvoljan linearan ograničeni operator, a T Fréchet-diferencijabilan operator važi

$$(AT)'_{(u)} = AT'_{(u)} \quad \text{i} \quad (TA)'_{(u)} = T'_{(Au)}A.$$

Ako je operator T dva puta Fréchet-diferencijabilan, tada je $T''_{(u)} \varepsilon L(X,L(X,Y))$, tj. $T''_{(u)} \varepsilon B(X^2,Y)$. Slično, ako je T n puta Fréchet-diferencijabilan operator, tada je $T^{(n)}_{(u)} \varepsilon N(X^n,Y)$.

Primer 2.4.1. Neka je $X=R^n$, $Y=R$, $T\vec{x}=f(\vec{x})=f(x_1,\ldots,x_n)$, $\vec{x}=[x_1\ldots x_n]^T$, $\vec{h}=[h_1\ldots h_n]^T$, $\|\vec{x}\|_\infty=\max\limits_k |x_k|$. Pod uslovom da f ima neprekidne parcijalne izvode drugog reda, naći ćemo $T'_{(\vec{x})}$.

Neka je $\vec{c}=\vec{x}+\xi\vec{h}\,(0<\xi<1)$. Na osnovu Taylorove formule imamo

$$f(\vec{x}+\vec{h}) = f(\vec{x}) + \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}\bigg|_{\vec{x}} h_k + \frac{1}{2}\sum_{k,j} \frac{\partial^2 f}{\partial x_k \partial x_j}\bigg|_{\vec{c}} h_k h_j,$$

tj.

$$|f(\vec{x}+\vec{h})-f(\vec{x})-(\operatorname{grad} f)^T\vec{h}| \le M(f,\vec{c})\,(\max_k |h_k|)^2, \tag{2.4.2}$$

gde smo stavili

$$M(f,\vec{c}) = \frac{1}{2}\left|\sum_{k,j} \frac{\partial^2 f}{\partial x_k \partial x_j}\bigg|_{\vec{c}}\right| \quad \text{i} \quad \operatorname{grad} f = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{bmatrix}.$$

S obzirom na neprekidnost parcijalnih izvoda drugog reda funkcije f, veličina $M(f,\vec{c})$ je konačna, pa na osnovu (2.4.2) imamo

$$\frac{|f(\vec{x}+\vec{h}) - f(\vec{x}) - (\operatorname{grad} f)^T\vec{h}|}{\|\vec{h}\|} \to 0$$

kada $\|\vec{h}\| \to 0$. Dakle,

$$T'_{(\vec{x})} = (\operatorname{grad} f)^T.$$

Primer 2.4.2. Neka je $X=R^n$, $Y=R^m$, $\vec{x}\varepsilon X$, $\vec{y}\varepsilon Y$, $\|\vec{x}\|_\infty = \max\limits_{1\le k\le n} |x_k|$, $\|y\|_\infty = \max\limits_{1\le k\le m} |y_k|$, $\vec{h} = [h_1\ldots h_n]^T$ (εX) i neka je operator $T:X\to Y$ definisan pomoću

$$T\vec{x} = \vec{f}(\vec{x}) = \begin{bmatrix} f_1(x_1,\ldots,x_n) \\ \vdots \\ f_m(x_1,\ldots,x_n) \end{bmatrix}.$$

Ako pretpostavimo da funkcije f_k $(k=1,\ldots,m)$ imaju neprekidne parcijalne izrade drugog reda, primenjujući postupak iz prethodnog primera na svaku od funkcija f_k, dobijamo

$$T'_{(\vec{x})} = W(\vec{f}),$$

gde je W Jacobieva matrica za $\vec{f}$, tj.

$$W(\vec{f}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \\ \frac{\partial f_m}{\partial x_1} & & \frac{\partial f_m}{\partial x_n} \end{bmatrix}.$$

2.2.5. Taylorova formula

Najpre uvodimo pojam L-metričkog prostora.

Definicija 2.5.1. Metrički prostor X naziva se L-metrički, ako za svako u X postoji linearna ograničena funkcionela $L:X\to R$, takva da je

$$\|L\| = 1 \quad \text{i} \quad Lu = \rho(u,\Theta). \tag{2.5.1}$$

Može se pokazati da Banachov prostor poseduje svojstvo (2.5.1) (videti [1, str.224]).

Neka je D konveksan* podskup Banachovog prostora X i T proizvoljan n+1 puta Fréchet-diferencijabilan operator u oblasti D. Tada važi sledeća teorema.

Teorema 2.5.1. Ako su u i u+h zadate tačke iz D, tada je

$$T(u+h) = \sum_{k=0}^{n} \frac{1}{k!} T^{(k)}_{(u)} \underbrace{(h,h,\ldots,h)}_{k\ \text{puta}} + W(u,h), \tag{2.5.2}$$

gde je

*) Skup D je konveksan ako važi implikacija
$u, u+h \in D \Rightarrow u+th \in D$ $(0 \le t \le 1)$.

(2.5.3) $$\| W(u,h)\| \leqq \frac{1}{(n+1)!} \sup_{t\in[0,1]} \| T^{(n+1)}_{(u+th)}\| \cdot \| h\|^{n+1} .$$

Dokaz. S obzirom da je X L-metrički prostor to se za element W(u,h) nalazi linearna ograničena funkcionela L, takva da je

$$\| L\| = 1 \quad \text{i} \quad LW = \rho\,(W,\Theta) = \| W(u,h)\| .$$

Uvedimo sada pomoćnu funkciju

(2.5.4) $$t \mapsto F(t) = LT(u+th),$$

čiji su izvodi redom jednaki

$$F'(t) = LT'_{(u+th)}h ,$$

$$\vdots$$

$$F^{(n+1)}(t) = LT^{(n+1)}_{(u+th)} \underbrace{(h,h,\ldots,h)}_{n+1 \text{ puta}} ,$$

s obzirom da je $(LT)'_{(a)} = LT'_{(a)}$ i $\frac{d}{dt}(u+th) = h$.

Sada imamo

$$\| W(u,h) \| = LW = LT(u+h)-LTu - \frac{1}{1!} LT'_{(u)}h - \ldots - \frac{1}{n!}LT^{(n)}_{(u)} \underbrace{(h,\ldots,h)}_{n \text{ puta}},$$

tj.

$$\|W(u,h)\| = F(1) - F(0) - \frac{1}{1!} F'(0) - \ldots - \frac{1}{n!}F^{(n)}(0) .$$

Kako za funkciju (2.5.4) važi klasična ocena ostatka u Taylorovoj formuli

$$|F(1)-F(0) - \frac{1}{1!}F'(0)-\ldots - \frac{1}{1!}F^{(n)}(0)| \leqq \frac{1}{(n+1)!} \max_{t\in[0,1]} |F^{(n+1)}(t)| ,$$

dobijamo

$$\|W(u,h)\| \leqq \frac{1}{(n+1)!} \sup_{t\in[0,1]} (\|L\|\cdot\|T^{(n+1)}_{(u+th)}\|) \| h\|^{n+1},$$

odakle, na osnovu (2.5.1), zaključujemo da važi (2.5.3), čime je dokaz završen.

Formula (2.5.2) naziva se Taylorova formula za operatore.

Stavljanjem n=0 u (2.5.3) dobija se sledeća važna nejednakost

$$\| T(u+h) - Tu\| \leq \sup_{t\in[0,1]} \| T'_{(u+th)}\| \cdot \| h\| ,$$

koja predstavlja teoremu u srednjoj vrednosti.

2.3. LITERATURA

1. S.ALJANČIĆ: Uvod u realnu i funkcionalnu analizu. Beograd, 1968.
2. G.SZEGO: Orthogonal polynomials. New York, 1959.
3. А.Ф.НИКИФОРОВ, В.Б.УВАРОВ: Основы теорий специальных функций. Москва, 1974.
4. D.S.MITRINOVIĆ: Uvod u specijalne funkcije. Beograd, 1975.
5. L.FOX and I.B.PARKER: Chebyshev Polynomials in Numerical Analysis. Oxford, 1972.
6. L.COLLATZ: Functionalanalysis und Numerische Mathematik. Berlin-Heidelberg-New York, 1964.
7. S.KUREPA: Konačno-dimenzionalni vektorski prostori i primjene. Zagreb, 1967.

3. OPŠTA TEORIJA ITERATIVNIH PROCESA

3.1. ITERATIVNI PROCESI ZA REŠAVANJE OPERATORSKIH JEDNAČINA I EGZISTENCIJA REŠENJA

3.1.1. Osnovne napomene o rešavanju operatorskih jednačina

Neka su X i Y Banachovi prostori, D konveksan podskup prostora X i $F:D \to Y$. Posmatrajmo operatorsku jednačinu

$$Fu = \theta, \tag{1.1.1}$$

gde je θ nula-vektor prostora Y.

Veliki broj problema u nauci i tehnici svodi se na rešavanje jednačine oblika (1.1.1). Navešćemo nekoliko tipičnih primera.

<u>Primer 1.1.1</u>. Ako je $X = Y = R$, $u = x$, $F = f$, jednačine

$$f(x) = x - \cos x = 0$$

i

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n = 0$$

su oblika (1.1.1).

<u>Primer 1.1.2</u>. Ako je $X = Y = R^n$, $u = \vec{x} = [x_1 \ \dots \ x_n]^T$ i

$$Fu = F(\vec{x}) = \begin{bmatrix} f_1(x_1,\dots,x_n) \\ \cdot \\ \cdot \\ \cdot \\ f_n(x_1,\dots,x_n) \end{bmatrix},$$

gde su $f_i : R^n \to R$ date funkcije, jednačina (1.1.1) predstavlja sistem jednačina

$$f_i(x_1,\dots,x_n) = 0 \qquad (i=1,\dots,n).$$

Ako je F linearan operator, na primer, $F(\vec{x}) = A\vec{x} - b$, gde su matrica A i vektor $\vec{b}$ dati sa

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & & a_{nn} \end{bmatrix} \quad \text{i} \quad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix},$$

jednačina (1.1.1) predstavlja sistem linearnih algebarskih jednačina

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i \qquad (i=1,\ldots,n).$$

Primer 1.1.3. Neka su $X = C^2[a,b]$, $Y = C[a,b] \times R$, $u \equiv u(t)$,

$$Fu = \begin{bmatrix} f_1(u) \\ f_2(u) \end{bmatrix},$$

$$f_1(u)(t) = u''(t) - f(t,u(t),u'(t)) \qquad (t \in [a,b]),$$

$$f_2(u) = g(u(a),u(b)),$$

gde su $f: R^3 \to R$ i $g: R^2 \to R$ date funkcije.

Tada rešenje operatorske jednačine (1.1.1) predstavlja rešenje konturnog problema

$$u''(t) = f(t,u(t),u'(t)) \qquad (t \in [a,b]),$$

$$g(u(a),u(b)) = 0.$$

Predmet našeg razmatranja je rešavanje jednačine (1.1.1), tj. nalaženje takve tačke $u \in D$, koja zadovoljava (1.1.1). U tom cilju ovu jednačinu predstavimo u ekvivalentnom obliku

$$(1.1.2) \qquad u = Tu$$

tako da operator T preslikava D u D, tj. da je $Tu = H(u,Fu)$, gde operator H preslikava $D \times Y$ u D. Za jednačinu (1.1.1), oblik (1.1.2) očigledno nije jedinstven, što pokazuje sledeći primer.

Primer 1.1.4. Jednačina $f(x)=0$ se može predstaviti u ekvivalentnom obliku

$$x = x + \lambda f(x)$$

za svako λ različito od nule.

Jedan od načina za rešavanje jednačine (1.1.2), tj. jednačine (1.1.1), zasniva se na konstrukciji niza $\{u_k\}_{k\in N_0}$ pomoću

(1.1.3) $$u_{k+1} = Tu_k \qquad (k=0,1,\ldots)$$

polazeći od $u_0 \in D$. Pod izvesnim uslovima za T, niz $\{u_k\}_{k\in N_0}$ može konvergirati ka traženom rešenju jednačine (1.1.2), o čemu će biti reči u sledećem odeljku.

Formulu (1.1.3), pomoću koje se generiše niz $\{u_k\}_{k\in N}$ zvaćemo iterativnim procesom.

Pored iterativnih procesa oblika (1.1.3) postoje i opštiji iterativni procesi oblika

$$u_{k+1} = S(u_k, u_{k-1}, \ldots, u_{k-m+1}) \qquad (k=m-1, m, \ldots),$$

pri čemu $S: X^m \to X$ i kod kojih se startuje sa m početnih vrednosti $u_0, u_1, \ldots, u_{m-1} \in D$.

3.1.2. Banachov stav o nepokretnoj tački

U ovom odeljku dokazaćemo jedan vrlo važan rezultat, o egzistenciji i jedinstvenosti rešenja operatorskih jednačina.

Neka je T operator u Banachovom prostoru X. Tačke u za koje je

(1.2.1) $$u = Tu$$

zovu se nepokretne tačke operatora T. Dovoljne uslove za egzistenciju jedinstvene nepokretne tačke operatora T dao je Banach.

Definicija 1.2.1. Operator $T:X \to X$ naziva se kontrakcijom ako postoji pozitivan broj $q<1$ takav da je za bilo koje dve tačke $u,v \in X$

$$\|Tu - Tv\| \leqq q\|u - v\| .$$

Sledeća teorema je tzv. Banachov stav o nepokretnoj tački.

Teorema 1.2.1. Kontrakcija T Banachovog prostora X u samog sebe ima jednu i samo jednu nepokretnu tačku.

Dokaz. Polazeći od proizvoljne tačke $u_o \in X$ konstruišimo niz $\{u_k\}_{k\in N}$ pomoću

$$u_{k+1} = Tu_k \qquad (k=0,1,\ldots).$$

Kako je za $k \geqq 1$

$$\|u_{k+1} - u_k\| = \|Tu_k - Tu_{k-1}\| \leqq q\|u_k - u_{k-1}\| ,$$

imamo

$$\|u_{k+1} - u_k\| \leqq q^k\|u_1 - u_o\| .$$

Ako je $m>k$, tada je

$$\begin{aligned}
(1.2.2) \qquad \|u_m - u_k\| &= \|(u_{k+1} - u_k) + (u_{k+2} - u_{k+1}) + \cdots + (u_m - u_{m-1})\| \\
&\leqq \|u_{k+1} - u_k\| + \|u_{k+2} - u_{k+1}\| + \cdots + \|u_m - u_{m-1}\| \\
&\leqq (q^k + q^{k+1} + \cdots + q^{m-1})\,\|u_1 - u_o\| \\
&< \frac{q^k}{1-q}\,\|u_1 - u_o\| \qquad (q<1),
\end{aligned}$$

odakle je $\lim\limits_{k\to+\infty} \|u_m - u_k\| = 0$, što znači da je $\{u_k\}_{k\in N_o}$ Cauchyev niz.

S obzirom da je X Banachov prostor, to je on kompletan pa postoji

$$(1.2.3) \qquad \lim_{k\to+\infty} u_k = a \qquad (a \in X).$$

S druge strane, kako je

$$\|u_{k+1} - Ta\| = \|Tu_k - Ta\| \leqq q\|u_k - a\| ,$$

imamo

$$\lim_{k\to+\infty} \|u_{k+1} - Ta\| = 0,$$

tj. $\lim\limits_{k\to+\infty} u_{k+1} = Ta$.

Iz poslednje jednakosti i jednakosti (1.2.3) sleduje $a = Ta$, tj. a je nepokretna tačka operatora T. Ovim je dokazana egzistencija nepokretne tačke.

Za dokaz jedinstvenosti pretpostavimo da postoje dve nepokretne tačke a i b, tj. neka je

$$a = Ta, \quad b = Tb, \quad \|a - b\| \neq 0.$$

Kako je T kontrakcija imamo

$$\|Ta - Tb\| = \|a - b\| \leqq q\|a - b\| ,$$

odakle, s obzirom da je $0 < q < 1$, neposredno sleduje

$$\|a - b\| = 0, \quad \text{tj.} \quad a = b.$$

Ovim je dokaz teoreme završen.

Dakle, ako je T kontrakcija X u X, jednačina

$$Tu = u$$

ima jedno i samo jedno rešenje i ono može biti dobijeno kao granična vrednost niza, koji se generiše pomoću

(1.2.4) $$u_{k+1} = Tu_k \qquad (k = 0,1,\ldots),$$

gde je u_0 proizvoljna tačka iz X.

Ako se zadržimo na k-toj aproksimaciji u_k, koja je odredjena pomoću (1.2.4), iz (1.2.2), pri $m \to +\infty$, sleduje

$$\|a - u_k\| \leq \frac{q^k}{1-q}\|u_1 - u_0\| ,$$

3.1.3. Iterativni procesi za rešavanje običnih jednačina

Neka je data jednačina

$$(1.3.1) \qquad f(x) = 0,$$

gde $f:[\alpha,\beta] \to R$, i neka je

$$(1.3.2) \qquad x = \phi(x)$$

njen ekvivalentni oblik.

Teorema 1.3.1. Pretpostavimo da neprekidna funkcija ϕ zadovoljava uslove

1° $\phi:[\alpha,\beta] \to [\alpha,\beta]$;

2° ϕ ima izvod u svakoj tački $x \in [\alpha,\beta]$, takav da je $|\phi'(x)| \leqq q < 1$.

Tada jednačina (1.3.2), tj. jednačina (1.3.1), ima jedinstveno rešenje $a \in [\alpha,\beta]$ i ono se može odrediti iterativnim procesom

$$(1.3.3) \qquad x_{k+1} = \phi(x_k) \qquad (k = 0,1,\ldots)$$

sa proizvoljnim $x_o \in [\alpha,\beta]$.

Dokaz. Za dokaz ove teoreme dovoljno je pokazati da je ϕ kontrakcija na $[\alpha,\beta]$. U tom slučaju ova teorema je posledica teoreme 1.2.1.

Kako iz Lagrange ove formule

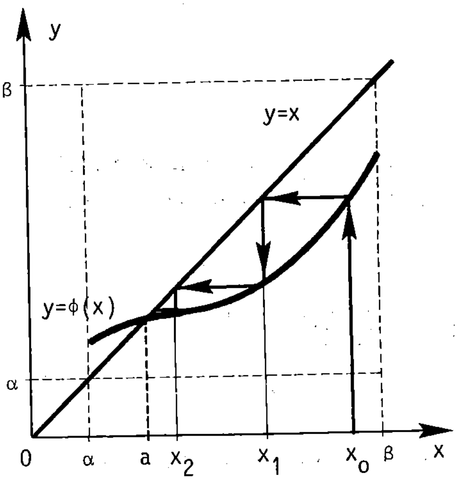
$$\phi(x) - \phi(y) = \phi'(\xi)(x - y),$$

gde su x i y proizvoljne tačke iz $[\alpha,\beta]$ i $\xi = y + \theta(x - y)$ $(0<\theta<1)$, sleduje

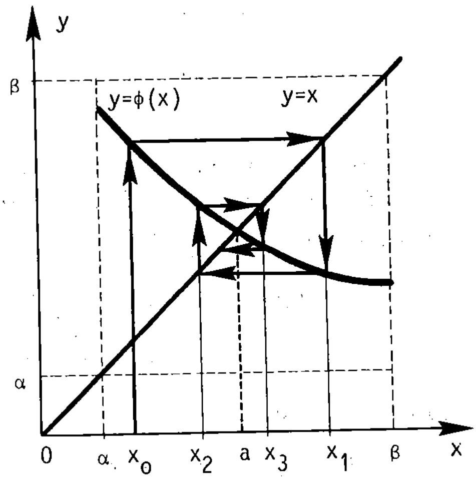
$$|\phi(x) - \phi(y)| = |\phi'(\xi)|\cdot|x - y| \leqq q|x - y|,$$

zaključujemo da je ϕ kontrakcija na $[\alpha,\beta]$, čime je dokaz teoreme završen.

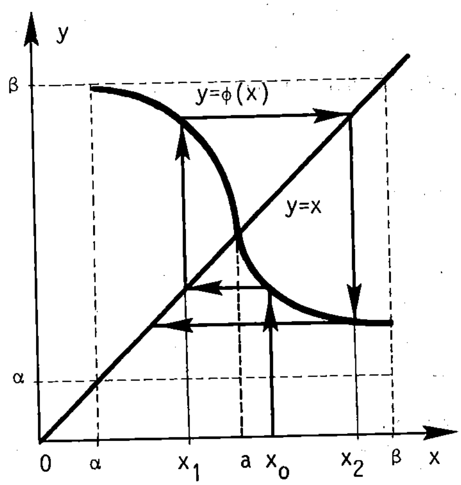
Dakle, da bi iterativni proces (1.3.3) konvergirao ka tački a, funkcija ϕ(često se naziva iterativna funkcija) mora ispunjavati odredjene uslove. Na slikama 1.3.1 do 1.3.4 data je geometrijska interpretacija iterativnih procesa oblika (1.3.3). Prva dva procesa(videti sl.1.3.1 i sl.1.3.2) su konvergentna. Interesantno je primetiti da je kod drugog od njih $\phi'(x)<0$ i da u tom slučaju niz $\{x_k\}$ konvergira oscilatorno, tj. greška $e_k=x_k-a$ alternativno menja znak. Druga dva procesa(videti sl.1.3.3 i sl. 1.3.4) su divergentna. Kod prvog od njih nije ispunjen uslov 2^o, a kod drugog uslov 1^o iz teoreme 1.3.1.



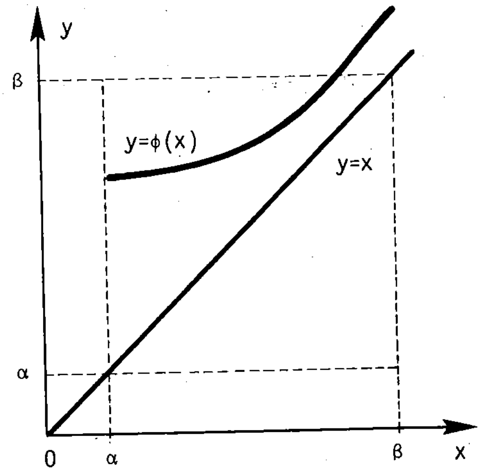
Sl. 1.3.1



Sl. 1.3.2



Sl. 1.3.3



Sl. 1.3.4

<u>Primer 1.3.1.</u> Posmatrajmo jednačinu

$$x^3 + x - 60 = 0. \tag{1.3.4}$$

Nije teško utvrditi (na primer, grafički) da ova jednačina ima koren $x = a$ koji leži u intervalu $(3,4)$. Da bismo rešili datu jednačinu, treba je prethodno svesti na oblik (1.3.2). Na primer, neki od tih oblika su

$$x = \phi_1(x) = 60 - x^3,$$
$$x = \phi_2(x) = \sqrt[3]{60 - x},$$
$$x = \phi_3(x) = \frac{60}{x^2} - \frac{1}{x}.$$

Neposrednim proveravanjem zaključujemo da od navedenih iterativnih funkcija samo ϕ_2 zadovoljava uslove teoreme 1.3.1, pri čemu je

$$|\phi_2'(x)| = \left|\frac{-1}{3\sqrt[3]{(60-x)^2}}\right| \leq \frac{1}{3(56)^{2/3}} \cong 0.022$$

kada $x \in [3,4]$.

Dakle, koren jednačine (1.3.4) može se odrediti iterativnim procesom

$$x_{k+1} = \sqrt[3]{60 - x_k} \qquad (k = 0,1,\ldots). \tag{1.3.5}$$

Polazeći od $x_0 = 4$, pomoću (1.3.5) dobijamo

k	x_k
0	4.0000000
1	3.8258623
2	3.8298239
3	3.8297338
4	3.8297359
5	3.8297358
6	3.8297358

tj. $a \cong 3.8297358$.

Pretpostavimo sada da jednačina (1.3.1) ima u $[\alpha,\beta]$ jedinstveno rešenje $x = a$, gde je segment $[\alpha,\beta]$ tako odabran da f ne menja znak na njemu. Postavlja se pitanje kako preći na ekvivalentni oblik (1.3.2), a da pri tome funkcija ϕ zadovoljava uslove teoreme 1.3.1. Izložićemo sada jedan od načina za rešavanje ovog problema.

Neka je

$$0 < m \leqq f'(x) \leqq M \qquad (\alpha \leqq x \leqq \beta). \tag{1.3.6}$$

Ukoliko je $f'(x) < 0$, umesto jednačine (1.3.1), može se uzeti jednačina $-f(x) = 0$.

Na osnovu primera 1.1.4, uzmimo

$$\phi(x) = x - \lambda f(x).$$

Izaberimo pozitivan parametar λ takav da je

$$0 \leqq \phi'(x) = 1 - \lambda f'(x) \leqq q < 1 \tag{1.3.7}$$

za $x \in [\alpha,\beta]$. Tada na osnovu (1.3.6), imamo

$$0 \leqq 1 - \lambda M \leqq 1 - \lambda m \leqq q,$$

odakle zaključujemo da će (1.3.7) biti ispunjeno ako je $\lambda = \frac{1}{M}$. U tom slučaju je

$$q = \max_{\alpha \leqq x \leqq \beta} (1 - \lambda f'(x)) = 1 - \frac{m}{M} < 1.$$

Dakle, može se uzeti

$$\phi(x) = x - \frac{1}{M} f(x).$$

<u>Primedba 1.3.1</u>. Ako je $|\phi'(x)| \geqq p > 1$ kada $x \in [\alpha,\beta]$, iterativni proces (1.3.3) divergira. Medjutim, ako se jednačina (1.3.2) predstavi u obliku

$$x = \psi(x),$$

gde je ψ inverzna funkcija od ϕ, iterativni proces

$$x_{k+1} = \Psi(x_k) \qquad (k = 0,1,\ldots)$$

biće konvergentan, jer je

$$|\Psi'(x)| = \left|\frac{1}{\phi'(\Psi(x))}\right| \leqq \frac{1}{p} = q < 1.$$

O konstrukciji iterativnih procesa oblika (1.3.3), kao i iterativnih procesa za rešavanje sistema nelinearnih jednačina biće reči u petoj glavi. S obzirom na poseban značaj iterativnih procesa u linearnoj algebri, problemi vezani za ovu oblast biće tretirani posebno u četvrtoj glavi. Napomenimo sada samo ukratko o rešavanju sistema linearnih jednačina

(1.3.8) $$A\vec{x} = \vec{b},$$

gde su matrica A i vektori $\vec{x}$ i $\vec{b}$ kao u primeru 1.1.2. Kako se sistem jednačina (1.3.8) može predstaviti(na beskonačno mnogo načina) u obliku

(1.3.9) $$\vec{x} = B\vec{x} + \vec{\beta},$$

gde su $B = [b_{ij}]_{n\times n}$ i $\vec{\beta} = [\beta_1 \ \ldots \ \beta_n]^T$, može se konstruisati iterativni proces sa operatorom $\vec{x} \mapsto T(\vec{x}) = B\vec{x} + \vec{\beta}$.

S obzirom da je T linearan operator, na osnovu

$$\|Tx - Ty\| = \|B(x - y)\| \leqq \|B\|\cdot\|x - y\|$$

zaključujemo da je $\|B\| \leqq q < 1$ dovoljan uslov da je T kontrakcija.

Dakle, ako je $\|B\| \leqq q < 1$, sistem jednačina (1.3.8), tj. (1.3.9), ima jedinstveno rešenje i odgovarajući iterativni proces

$$\vec{x}_{k+1} = T(\vec{x}_k) \qquad (k = 0,1,\ldots)$$

konvergira ka $\vec{a} = A^{-1}\vec{b}$. Napomenimo da vrednost $\|B\|$ zavisi od izbora metrike u R^n.

O egzistenciji i jedinstvenosti rešenja diferencijalnih i integralnih jednačina i odgovarajućim iterativnim procesima za njihovo rešavanje biće reči u posebnim glavama.

3.2. KARAKTERISTIKE ITERATIVNIH PROCESA I METODI ZA UBRZAVANJE KONVERGENCIJE

3.2.1. Red konvergencije iterativnih procesa

Neka su X Banachov prostor i operator $T:X \to X$ i neka niz

$$(2.1.1) \qquad u_{k+1} = Tu_k \qquad (k=0,1,\ldots)$$

konvergira ka tački $a \in X$.

Niz (2.1.1) definiše iterativni proces za rešavanje operatorske jednačine

$$u = Tu.$$

Osim iterativnih procesa oblika (2.1.1), mogu se posmatrati i opštiji iterativni procesi oblika

$$(2.1.2) \qquad u_{k+1} = S(u_k, u_{k-1}, \ldots, u_{k-m+1}) \qquad (k=m-1,m,\ldots),$$

pri čemu $S:X^m \to X$.

Definicija 2.1.1. Za iterativni proces koji konvergira ka a, kaže se da ima red konvergencije r ako je

$$(2.1.3) \qquad \|u_{k+1} - a\| = O(\|u_k - a\|^r),$$

tj. ako postoji konstanta A takva da je

$$\|u_{k+1} - a\| \leqq A\|u_k - a\|^r$$

za dovoljno veliko k.

Označimo sa $U(a)$ konveksnu okolinu tačke a. Tada, za iterativni proces (2.1.1), umesto (2.1.3) u prethodnoj definiciji, možemo uzeti

$$\|Tu - a\| = O(\|u - a\|^r) \qquad (u \in U(a)).$$

Teorema 2.1.1. Ako je operator $T:X \to X$ r-puta Fréchet-diferencijabilan u konveksnoj okolini $U(a)$, iterativni proces (2.1.1) je reda r ako su ispunjeni sledeći uslovi:

1^o $Ta = a$;

2^o $T'_{(a)}, T''_{(a)}, \dots, T^{(r-1)}_{(a)}$ su nula-operatori;

3^o $T^{(r)}_{(u)}$ je nenula-operator sa normom koja je ograničena na $U(a)$.

Dokaz. Neka $u \in U(a)$, $\|T^{(r)}_{(u)}\| \leqq M_r$ i

$$q = Ta + \frac{1}{1!}T'_{(a)}(u - a) + \dots + \frac{1}{(r-1)!}T^{(r-1)}_{(a)}(\underbrace{u-a,\dots,u-a}_{k-1 \text{ puta}}).$$

Tada na osnovu Taylorove formule, imamo

$$\|Tu - q\| \leqq \frac{1}{r!} \sup_{t \in [0,1]} \|T^{(r)}_{(a+t(u-a))}\| \cdot \|u - a\|^r ,$$

tj.

$$\|Tu - q\| \leqq \frac{M_r}{r!} \|u - a\|^r .$$

Kako je, na osnovu pretpostavki teoreme, $q = Ta = a$, imamo

$$\|Tu - a\| \leqq A\|u - a\|^r \qquad (A = \frac{M_r}{r!}),$$

čime je dokaz završen.

U daljem tekstu, kada govorimo o iterativnom procesu reda r, pretpostavljaćemo uvek da su ispunjeni uslovi teoreme 2.1.1.

Napomenimo da je red konvergencije iterativnih procesa tipa (2.1.1) uvek ceo broj, ada je kod procesa tipa (2.1.2) realan broj.

U slučaju da je $r=1$, dovoljan uslov za konvergenciju iterativnog procesa je $0 \leqq A < 1$.

Posmatrajmo sada slučaj kada je $X = R$.

Teorema 2.1.2. Neka je

$$x_{k+1} = \Phi(x_k) \qquad (k=0,1,\dots)$$

iterativni proces reda r, gde je $x \mapsto \Phi(x)$ r puta neprekidno-diferencijabilna funkcija u okolini tačke $a = \lim\limits_{k\to+\infty} x_k$. Tada je

$$(2.1.4) \qquad \lim_{k\to+\infty} \frac{x_{k+1} - a}{(x_k - a)^r} = \frac{\Phi^{(r)}(a)}{r!}.$$

Dokaz. Kako je $\Phi(a) = a$ i $\Phi^{(i)}(a) = 0 (i = 1,\dots,r-1)$, na osnovu Taylorove formule dobijamo

$$(2.1.5) \quad x_{k+1} = \Phi(x_k) = a + \frac{1}{r!}\Phi^{(r)}(a + \Theta(x_k - a))(x_k - a)^r,$$

gde je $0 < \Theta < 1$. Kako je $a = \lim\limits_{k\to+\infty} x_k$ i $\Phi^{(r)}$ neprekidna funkcija, deobom jednakosti sa $(x_k - a)^r$ i prelazom na graničnu vrednost dobijamo (2.1.4).

Definicija 2.1.2. Veličina

$$C_r = \lim_{k\to+\infty} \frac{|x_{k+1} - a|}{|x_k - a|^r}$$

naziva se faktor konvergencije ili asimptotska konstanta greške.

Primedba 2.1.1. Iterativni proces kod koga je $r = 1$ naziva se proces sa linearnom konvergencijom. Negov faktor konvergencije C_1 mora biti manji od jedinice.

Primer 2.1.1. Neka se za izračunavanje vrednosti $a = \sqrt[m]{N}$ $(N>0)$ koristi iterativni proces

$$x_{k+1} = \Phi(x_k) \qquad (k=0,1,\dots),$$

gde je

$$\Phi(x) = x\,\frac{(m-1)x^m + (m+1)N}{(m+1)x^m + (m-1)N}\,.$$

Sukcesivnim diferenciranjem jednakosti

(2.1.6) $$g(x)\Phi(x) = (m-1)x^{m+1} + (m+1)Nx,$$

gde je

$$g(x) = (m+1)x^m + (m-1)N,$$

nalazimo redom

(2.1.7) $$g(x)\Phi'(x) = -m(m+1)x^{m-1}\Phi(x) + (m-1)(m+1)x^m + (m+1)N,$$

(2.1.8) $$g(x)\Phi''(x) = -2m(m+1)x^{m-1}\Phi'(x) - m(m-1)(m+1)x^{m-2}\Phi(x) + m(m-1)(m+1)x^{m-1}\,,$$

(2.1.9) $$g(x)\Phi'''(x) = -3m(m+1)x^{m-1}\Phi''(x) - 3m(m-1)(m+1)x^{m-2}\Phi'(x) - m(m-2)(m-1)(m+1)x^{m-3}\Phi(x) + m(m-1)^2(m+1)x^{m-2}.$$

Kako je $a^m = N$, iz (2.1.6) do (2.1.9) sleduje

(2.1.10) $$\Phi(a) = a,\quad \Phi'(a) = 0,\quad \Phi''(a) = 0,\quad \Phi'''(a) = \frac{m^2-1}{2a^2}\,.$$

Ako je $|m| \neq 1$, na osnovu (2.1.10) zaključujemo da je dati iterativni proces trećeg reda. U ovom slučaju (2.1.4) se svodi na

$$\lim_{k\to+\infty} \frac{x_{k+1} - a}{(x_k - a)^3} = \frac{\Phi'''(a)}{3!} = \frac{m^2-1}{12a^2}\,.$$

Primetimo da se za $m = \pm 1$ funkcija Φ svodi na konstantu $\Phi(x) = N$ ili $\Phi(x) = 1/N$.

3.2.2. Aitkenov Δ^2 metod

U ovom odeljku razmatraćemo problem ubrzavanja konvergencije realnog niza $\{x_k\}_{k\in N_0}$, koji se generiše pomoću

$$x_{k+1} = \Phi(x_k) \qquad (k=0,1,\ldots), \tag{2.2.1}$$

pri čemu je $\lim\limits_{k\to+\infty} x_k = a$.

Pretpostavimo da je iterativni proces (2.2.1) sa linearnom konvergencijom, tj. da je

$$x_{k+1} - a = C_k(x_k - a) \qquad (k=0,1,\ldots), \tag{2.2.2}$$

gde su $C_k = C + \delta_k$ $(k=0,1,\ldots)$, C konstanta takva da je $|C|<1$ i $\delta_k \to 0$, kada $k\to+\infty$.

Iz asimptotskih relacija

$$x_{k+2} - a \sim C(x_{k+1} - a) \quad \text{i} \quad x_{k+1} - a \sim C(x_k - a) \qquad (k\to+\infty),$$

koje se dobijaju na osnovu (2.2.2), eliminacijom konstante C, sleduje

$$a \sim \frac{x_{k+2}x_k - x_{k+1}^2}{x_{k+2} - 2x_{k+1} + x_k} \qquad (k\to+\infty).$$

Dobijena asimptotska relacija sugeriše konstrukciju niza $\{x_k^*\}_{k\in N_0}$ pomoću

$$x_k^* = \frac{x_{k+2}x_k - x_{k+1}^2}{x_{k+2} - 2x_{k+1} + x_k} \qquad (k=0,1,\ldots). \tag{2.2.3}$$

Formula (2.2.3), medjutim, nije pogodna za primenu sa numeričkog stanovišta, pa se zato koriste njeni ekvivalentni matematički oblici

(2.2.4) $$x_k^* = x_k - \frac{(x_{k+1} - x_k)^2}{x_{k+2} - 2x_{k+1} + x_k},$$

(2.2.5) $$x_k^* = x_{k+1} - \frac{(x_{k+1} - x_k)(x_{k+2} - x_{k+1})}{x_{k+2} - 2x_{k+1} + x_k},$$

(2.2.6) $$x_k^* = x_{k+2} - \frac{(x_{k+2} - x_{k+1})^2}{x_{k+2} - 2x_{k+1} + x_k},$$

gde je $k = 0,1,\ldots$.

Poslednja formula se najčešće koristi u primenama, s obzirom na izvesne prednosti nad formulama (2.2.4) i (2.2.5). Naime, glavni deo u vrednosti x_k^*, po formuli (2.2.4), (2.2.5) i (2.2.6) je x_k, x_{k+1} i x_{k+2} respektivno. S druge strane, kako je glavni deo u formuli (2.2.6) najpribližniji vrednosti a (u poredjenju sa x_k i x_{k+1}), to izlazi da ova formula, sa numeričkog stanovišta, ima prednosti nad formulama (2.2.4) i (2.2.5).

Navedene formule mogu se predstaviti pomoću operatora Δ. Na primer, formula (2.2.4) dobija oblik

(2.2.7) $$x_k^* = x_k - \frac{(\Delta x_k)^2}{\Delta^2 x_k} \qquad (k = 0,1,\ldots).$$

Sledeća teorema ukazuje na to da niz $\{x_k^*\}_{k\in N_0}$ brže konvergira ka a, nego niz $\{x_k\}_{k\in N_0}$.

<u>Teorema 2.2.1</u>. Neka je za niz $\{x_k\}_{k\in N_0}$, $x_k \neq a$ i

$$x_{k+1} - a = (C + \delta_k)(x_k - a) \qquad (|C| < 1, \lim_{k\to+\infty} \delta_k = 0),$$

tada je

(2.2.8) $$\lim_{k\to+\infty} \frac{x_k^* - a}{x_k - a} = 0.$$

Dokaz. Neka je $e_k = x_k - a$. Tada je $e_{k+1} = (C + \delta_k)e_k$ i

$$\Delta x_k = \Delta e_k = e_k((C-1) + \delta_k)$$

i

$$\Delta^2 x_k = \Delta^2 e_k = e_k((C + \delta_{k+1})(C + \delta_k) - 2(C + \delta_k) + 1)$$
$$= e_k((C-1)^2 + \gamma_k),$$

gde je $\gamma_k = C(\delta_{k+1} + \delta_k) - 2\delta_k + \delta_k \delta_{k+1}$. Primetimo da uslov $\lim\limits_{k\to+\infty} \delta_k = 0$ implicira uslov $\lim\limits_{k\to+\infty} \gamma_k = 0$.

Da bismo dokazali (2.2.8) podjimo od (2.2.7).

Kako je $e_k \neq 0$, $C \neq 1$ i $\gamma_k \to 0$ $(k \to +\infty)$, imamo $\Delta^2 x_k \neq 0$, što znači da je niz $\{x_k^*\}_{k\in N_0}$ definisan pomoću (2.2.7) za svako k.

Tada je, na osnovu prethodnog,

$$x_k^* - a = e_k - \frac{((C-1) + \delta_k)^2}{(C-1)^2 + \gamma_k} e_k,$$

tj.

$$\lim_{k\to+\infty} \frac{x_k^* - a}{x_k - a} = \lim_{k\to+\infty} \frac{\gamma_k - 2(C-1)\delta_k - \delta_k^2}{(C-1)^2 + \gamma_k} = 0,$$

čime je dokaz završen.

Navedeni postupak ubrzavanja konvergencije niza $\{x_k\}_{k\in N_0}$ poznat je kao Aitkenov Δ^2 metod.

U daljem razmatranju ovog metoda uvešćemo dodatnu pretpostavku za iterativni proces (2.2.1). Naime, neka je

$$\delta_k = \omega_k e_k,$$

pri čemu $\omega_k \to \omega$, kada $k \to +\infty$. Tada (2.2.2) postaje

$$e_{k+1} = Ce_k + \omega_k e_k^2 ,$$

gde smo stavili $e_k = x_k - a$. Jasno je, medjutim, da svi iterativni procesi sa linearnom konvergencijom ne poseduju ovu osobinu.

Teorema 2.2.2. Neka su ispunjeni uslovi teoreme 2.2.1 sa $\delta_k = \omega_k e_k$ ($\omega_k \to \omega$, kada $k \to +\infty$). Tada je

$$\lim_{k \to +\infty} \frac{x_k^* - a}{(x_k - a)^2} = \frac{C\omega}{C-1} . \tag{2.2.8}$$

Dokaz. S obzirom da je

$$\begin{aligned} \gamma_k - 2(C-1)\delta_k - \delta_k^2 &= (C + \delta_k)(\delta_{k+1} - \delta_k) \\ &= e_k(C + \delta_k)(\omega_{k+1}(C + \delta_k) - \omega_k), \end{aligned}$$

imamo $x_k^* - a = \alpha_k e_k^2$, gde smo stavili

$$\alpha_k = \frac{(C + \delta_k)(\omega_{k+1}(C + \delta_k) - \omega_k)}{(C-1)^2 + \gamma_k} .$$

Kako je, dalje, $\lim\limits_{k \to +\infty} \alpha_k = \dfrac{C\omega}{C-1}$, iz prethodnog neposredno sleduje (2.2.8).

Teorema 2.2.3. Ako su ispunjeni uslovi prethodne teoreme važi

$$\lim_{k \to +\infty} \frac{x_{k+1}^* - a}{x_k^* - a} = C^2. \tag{2.2.9}$$

Dokaz. Na osnovu dokaza prethodne teoreme i jednakosti (2.2.2), imamo

$$\frac{x_{k+1}^* - a}{x_k^* - a} = \frac{\alpha_{k+1} e_{k+1}^2}{\alpha_k e_k^2} = \frac{\alpha_{k+1}}{\alpha_k}(C + \delta_k)^2 ,$$

odakle sleduje (2.2.9).

Primedba 2.2.1. Iz teoreme 2.2.3 sleduje da je i transformisani niz $\{x_k^*\}_{k\in N_0}$ sa linearnom konvergencijom, ali je njegov faktor konvergencije manji nego kod niza $\{x_k\}_{k\in N_0}$ $(C^2<|C|<1)$.

Primedba 2.2.2. Posmatrajmo iterativni proces $x_{k+1}=\Phi(x_k)$ $(k=0,1,\ldots)$, gde $x_0\in[\alpha,\beta]$. Ako za iterativnu funkciju $\Phi:[\alpha,\beta]\to[\alpha,\beta]$ pretpostavimo uslove:

1° $\Phi\in C^2[\alpha,\beta]$,

2° $|\Phi'(x)|<1$ za svako $x\in[\alpha,\beta]$,

na osnovu Taylorove formule, imamo

$$x_{k+1}=\Phi(x_k)=a+\Phi'(a)e_k+\frac{1}{2}\Phi''(\xi_k)e_k^2,$$

gde su $a=\lim\limits_{k\to+\infty}x_k$, $e_k=x_k-a$ i $\xi_k=a+\theta_k(x_k-a)$ $(0\leqq\theta\leqq1)$.

Kako iterativni proces, definisan na ovaj način, zadovoljava uslove teoreme 2.2.2 sa

$$C=\Phi'(a)\quad \text{i}\quad \omega_k=\frac{1}{2}\Phi''(\xi_k),$$

to na osnovu (2.2.8) imamo

$$\lim_{k\to+\infty}\frac{x_k^*-a}{(x_k-a)^2}=\frac{1}{2}\cdot\frac{\Phi''(a)\Phi'(a)}{\Phi'(a)-1}.$$

Primer 2.2.1. Primenom formule (2.2.6) na niz koji se generiše pomoću $x_{k+1}=\cos x_k$ $(k=0,1,\ldots)$ sa $x_0=1$, dobijamo

k	x_k	x_k^*
0	1.000000	0.728010
1	0.540302	0.733665
2	0.857553	0.736906
3	0.654290	0.738050
4	0.793480	0.738636
5	0.701369	0.738876
6	0.763960	0.738992
7	0.722102	0.739042
8	0.750418	0.739066
9	0.731404	
10	0.744237	

Primetimo da x_7^* aproksimira koren jednačine $\cos x - x = 0$ sa četiri tačne decimale.

3.2.3. O metodima bliskim Aitkenovom metodu

Posmatrajmo iterativni proces

(2.3.1) $$x_{k+1} = \Phi(x_k) \qquad (k = 0,1,\ldots).$$

Kako je $x_{k+1} = \Phi(x_k)$ i $x_{k+2} = \Phi(x_{k+1}) = \Phi(\Phi(x_k)) = \Phi^2(x_k)$, formula (2.2.3) iz prethodnog odeljka može se predstaviti u obliku

(2.3.2) $$x_k^* = \Psi(x_k) \qquad (k = 0,1,\ldots),$$

gde je

(2.3.3) $$\Psi(x) = \frac{x\,\Phi^2(x) - \Phi(x)^2}{\Phi^2(x) - 2\Phi(x) + x}\,.$$

Funkciju Ψ, definisanu sa (2.3.3), Steffensen(videti na primer [1 ,str. 241 - 246]) je dobio primenom metoda sečice (videti odeljak 5.1.4) na rešavanje jednačine

$$F(x) \equiv x - \Phi(x) = 0,$$

ne uvodeći pretpostavku o redu konvergencije procesa (2.3.1).

Dakle, za razliku od Aitkenovog Δ^2 metoda, definisanog sa (2.3.2), Steffensen je došao do procesa

$$x_{k+1} = \Psi(x_k) \qquad (k = 0,1,\ldots).$$

Opštiji metod za ubrzavanje konvergencije dao je Hauseholder (videti [2]), posmatrajući dva iterativna procesa reda r

$$x_{k+1}^{(j)} = \Phi_j(x_k^{(j)}) \qquad (j = 1,2\,;\ k = o,1,\ldots),$$

koji konvergiraju ka $x = a$ i definisanjem funkcije Ψ_1 pomoću

(2.3.4) $$\Psi_1(x) = \frac{x\phi_1(\phi_2(x)) - \phi_1(x)\phi_2(x)}{x - \phi_1(x) - \phi_2(x) + \phi_1(\phi_2(x))}.$$

Sledeći rezultat odnosi se na iterativni proces

(2.3.5) $$x_{k+1} = \Psi_1(x_k) \qquad (k = 0,1,\ldots).$$

Teorema 2.3.1(Hauseholder). Ako je

$$(\phi'_1(a) - 1)(\phi'_2(a) - 1) \neq 0$$

iterativni proces (2.3.5) ima red konvergencije $2r-1$ ako je $r>1$, ili dva ako je $r=1$.

Primedba 2.3.1. Za $\phi_1 = \phi_2 = \phi$ (2.3.4) se svodi na (2.3.3).

Pod pretpostavkom da je dato m iterativnih procesa

$$x_{k+1}^{(j)} = \phi_j(x_k^{(j)}) \qquad (j = 1,\ldots,m;\ k = 0,1,\ldots),$$

čiji je red konvergencije r, u radu [3] dato je sledeće uopštenje prethodnih rezultata.

Teorema 2.3.2. Ako je

$$\phi_{p,q,\ldots,s,t}(x) = \phi_p(\phi_q(\ldots\phi_s(\phi_t(x)))\ldots),$$

$$D_n(x) = \begin{vmatrix} x & \phi_{i_1,i_2,\ldots,i_n}(x) \\ \phi_j(x) & \phi_{l_1,l_2,\ldots,l_{n+1}}(x) \end{vmatrix},$$

$$\Delta_n(x) = \begin{vmatrix} x - \phi_{i_1,i_2,\ldots,i_n}(x) & 1 \\ \phi_j(x) - \phi_{l_1,l_2,\ldots,l_{n+1}}(x) & 1 \end{vmatrix}$$

$$(j,\ i_\alpha,\ i_\beta \in \{1,2,\ldots,m\}\ ;\quad \alpha=1,2,\ldots,n\ ;\quad \beta=1,2,\ldots,n+1),$$

i

$$1-\phi_j'(a)-\phi_{i_1}'(a)\phi_{i_2}'(a)\ldots\phi_{i_n}'(a)+\phi_{l_1}'(a)\phi_{l_2}'(a)\ldots\phi_{l_{n+1}}'(a)\neq 0,$$

iterativni proces definisan pomoću

$$x_{k+1} = D_n(x_k)\Delta_n^{-1}(x_k) \qquad (k=0,1,\ldots), \tag{3.2.6}$$

ima red konvergencije r^n+r-1 ako je $r>1$, ili dva ako je $r=1$.

3.2.4. Dva opšta metoda za ubrzavanje konvergencije iterativnih procesa

U ovom odeljku daćemo jedan opšti prilaz ubrzavanju konvergencije iterativnih procesa. Naime, metodi koji će biti izloženi omogućavaju dobijanje iterativnih procesa višeg reda, ako se podje od nekog poznatog iterativnog procesa.

Neka se niz $\{x_k\}_{k\in N_o}$ koji konvergira ka tački a generiše pomoću

$$x_{k+1} = \phi(x_k) \qquad (k=0,1,\ldots). \tag{2.4.1}$$

U radu [4] dokazana je sledeća teorema.

Teorema 2.4.1. Neka je (2.4.1) iterativni proces sa konvergencijom reda r, funkcija ϕ (r+1)-puta diferencijabilna u okolini granične tačke a i neka je $\phi'(a)\neq r$. Tada je

$$x_{k+1} = \phi(x_k) + \frac{1}{r}\phi'(x_k)(x_{k+1}-x_k),$$

tj.

$$x_{k+1} = \frac{\phi(x_k)-\frac{1}{r}\phi'(x_k)x_k}{1-\frac{1}{r}\phi'(x_k)} = x_k - \frac{x_k-\phi(x_k)}{1-\frac{1}{r}\phi'(x_k)}$$

iterativni proces najmanje reda r+1.

Na dalje posmatraćemo iterativne procese u Banachovom prostoru.

Neka je X Banachov prostor, operator $T:X\to X$ i neka niz koji se generiše pomoću

(2.4.2) $$u_{k+1} = Tu_k \qquad (k=0,1,\dots)$$

konvergira ka $a\in X$. Jednakost (2.4.2) definiše iterativni proces za rešavanje operatorske jednačine $u = Tu$.

Označimo sa $U(a)$ konveksnu okolinu granične tačke a.

U pomenutom radu [4] navedena je sledeća teorema, koja predstavlja generalizaciju teoreme 2.4.1.

Teorema 2.4.2. Neka je (2.4.2) iterativni proces reda r, operator $T:X\to X$ $(r+1)$- puta Fréchet-diferencijabilan u okolini $U(a)$ i neka postoji inverzan operator $\left[I-\frac{1}{r}T'_{(u)}\right]^{-1}$ kada $u\in U(a)$. Tada je

(2.4.3) $$u_{k+1} = Gu_k = u_k - \left[I - \frac{1}{r}T'_{(u_k)}\right]^{-1}(u_k - Tu_k)$$

iterativni proces najmanje reda $r+1$.

Osnovni nedostatak metoda, definisanog poslednjom teoremom, je što zahteva nalaženje inverznog operatora $\left[I-\frac{1}{r}T'_{(u)}\right]^{-1}$, što je u većini slučajeva vrlo složeno.

Sada ćemo izložiti jedan metod za ubrzavanje konvergencije iterativnih procesa, koji ne zahteva nalaženje inverznog operatora(videti[5]). Nažalost, ovaj metod se ne može primeniti za ubrzavanje procesa sa redom konvergencije jedan.

Teorema 2.4.3. Neka je (2.4.2) iterativni proces sa konvergencijom reda $r(\geq 2)$ i neka operator $T:X\to X$ $(r+1)$- puta Fréchet-diferencijabilan u okolini $U(a)$. Tada je

(2.4.4) $$u_{k+1} = Fu_k = Tu_k - \frac{1}{r}\,T'_{(u_k)}(u_k - Tu_k) \qquad (k=0,1,\dots)$$

iterativni proces najmanje reda $r+1$.

Dokaz. Neka $u \in U(a)$. S obzirom da je iterativni proces (2.4.2) sa konvergencijom reda r i operator T (r+1)- puta Fréchet-diferencijabilan u U(a), to su ispunjeni uslovi teoreme 2.1.1.

Na osnovu Taylorove formule imamo

$$T'_{(u)} = \frac{1}{(r-1)!} T^{(r)}_{(a)} \underbrace{(u-a,\dots,u-a)}_{r-1 \text{ puta}} + W(a,u-a) \tag{2.4.5}$$

i

$$Tu = a + \frac{1}{r!} T^{(r)}_{(a)} \underbrace{(u-a,\dots,u-a)}_{r \text{ puta}} + w(a,u-a), \tag{2.4.6}$$

gde je

$$\|W(a,u-a)\| = O(\|u-a\|^r) \quad \text{i} \quad \|w(a,u-a)\| = O(\|u-a\|^{r+1}).$$

S obzirom da je $T'_{(u)}$ linearan operator, to se (2.4.4) može predstaviti u obliku

$$Fu - a = Tu - a - \frac{1}{r}T'_{(u)}(u-a) + \frac{1}{r}T'_{(u)}(Tu-a).$$

Korišćenjem (2.4.5) i (2.4.6), dobijamo

$$Fu - a = w(a,u-a) - \frac{1}{r}W(a,u-a)(u-a) + \frac{1}{r}T'_{(u)}(Tu-a),$$

odakle sleduje

$$\|Fu-a\| \leqq \|w(a,u-a)\| + \frac{1}{r}\|W(a,u-a)\|\cdot\|u-a\| + \frac{1}{r}\|T'_{(u)}\|\cdot\|Tu-a\| .$$

Kako je $\|T'_{(u)}\| = O(\|u-a\|^{r-1})$ i $\|Tu-a\| = O(\|u-a\|^r)$, pri $r \geqq 2$, najzad dobijamo

$$\|Fu - a\| = O(\|u-a\|^{r+1}),$$

čime je dokaz teoreme završen.

Sada ćemo dati dokaz teoreme 2.4.2.

Dokaz teoreme 2.4.2. Neka $u \in U(a)$. Formuli (2.4.3) može se dati sledeći oblik

$$\left[I - \frac{1}{r}T'_{(u)}\right](u - Gu) = u - Tu,$$

tj.

$$Gu = Tu - \frac{1}{r}T'_{(u)}(u - Gu).$$

S obzirom da je $T'_{(u)}$ linearan operator, (2.4.3) se može predstaviti u ekvivalentnom obliku

$$Gu - a = Tu - a - \frac{1}{r}T'_{(u)}(u-a) + \frac{1}{r}T'_{(u)}(Gu-a)$$

odakle, korišćenjem (2.4.5) i (2.4.6), dobijamo

$$Gu - a = w(a,u-a) - \frac{1}{r}W(a,u-a)(u-a) + \frac{1}{r}T'_{(u)}(Gu-a).$$

Iz poslednje jednakosti sleduje

$$(1 - \frac{1}{r}\|T'_{(u)}\|)\,\|Gu-a\| \leqq \|w(a,u-a)\| + \frac{1}{r}\|W(a,u-a)\|\cdot\|u-a\|,$$

tj.

$$\|Gu - a\| = O(\|u-a\|^{r+1}),$$

čime je dokaz završen.

Teorema 2.4.4. Neka je (2.4.1) iterativni proces sa konvergencijom reda $r(\geq 2)$ i funkcija ϕ $(r+1)$-puta diferencijabilna u okolini granične tačke a. Tada je

$$x_{k+1} = \phi(x_k) - \frac{1}{r}\phi'(x_k)(x_k - \phi(x_k)) \qquad (k=0,1,\ldots)$$

iterativni proces najmanje reda $r+1$.

Ova teorema je posledica teoreme 2.4.3.

3.3. LITERATURA

1. A.OSTROWSKI: Solution of Equations and Systems of Equations. New York, 1966.
2. A.S.HAUSEHOLDER: Principles of Numerical Analysis. New York, 1953.
3. А.В.ПРОКОПЧЕНКО: Об итерационных процессах высших порядков. Ж.Вычисл.мат. и мат.физ. 14(1974), 230-233.
4. B.JOVANOVIĆ: A method for obtaining iterative formulas of higher order. Mat.Vesnik 9(24)(1972), 365-369.
5. G.V.MILOVANOVIĆ: A method to accelerate iterative processes in Banach space. Univ.Beograd.Publ.Elektrotehn.Fak.Ser. Mat.Fiz. №461 - №497(1974), 67-71.
6. L.COLLATZ: Funktionalanalysis und numerische Mathematik. Berlin-Heidelberg-New York, 1968.
7. J.F.TRAUB: Iterative methods for the solution of equations. New Jersey, 1964.
8. Л.В.КАНТОРОВИЧ, Г.П.АКИЛОВ: Функциональный анализ в нормированных пространствах. Москва 1959.
9. S.K. BERBERIAN: Introduction to Hilbert Space. New York, 1976.

4. NUMERIČKI METODI U LINEARNOJ ALGEBRI

4.1. ELEMENTI MATRIČNOG RAČUNA

4.1.1. Operacije sa matricama razbijenim na blokove

Definicija 1.1.1. Ako se matrica A tipa $m \times n$ mrežom horizontalnih i vertikalnih pravih razloži na više matrica, kaže se da je matrica A razbijena na blokove.

Blokovi su matrice A_{ij} tipa $m_i \times n_j$ $\left(\sum_{i=1}^{p} m_i = m \,;\, \sum_{j=1}^{q} n_j = n\right)$.

Operacije sa matricama razbijenim na blokove su formalno iste sa operacijama kod običnih matrica. Naime, važe sledeći rezultati:

Teorema 1.1.1. Neka su matrice A i B razbijene na blokove, tj.

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1q} \\ \vdots & & \\ A_{p1} & & A_{pq} \end{bmatrix} \quad \text{i} \quad B = \begin{bmatrix} B_{11} & \cdots & B_{1q} \\ \vdots & & \\ B_{p1} & & B_{pq} \end{bmatrix},$$

gde su A_{ij} i B_{ij} matrice istog tipa. Tada je

$$cA = \begin{bmatrix} cA_{11} & \cdots & cA_{1q} \\ \vdots & & \\ cA_{p1} & & cA_{pq} \end{bmatrix} \quad \text{i} \quad A + B = \begin{bmatrix} A_{11} + B_{11} & \cdots & A_{1q} + B_{1q} \\ \vdots & & \\ A_{p1} + B_{p1} & & A_{pq} + B_{pq} \end{bmatrix},$$

gde je c skalar.

Teorema 1.1.2. Neka je

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1q} \\ \vdots & & \\ A_{p1} & & A_{pq} \end{bmatrix} \quad \text{i} \quad B = \begin{bmatrix} B_{11} & \cdots & B_{1s} \\ \vdots & & \\ B_{q1} & & B_{qs} \end{bmatrix},$$

gde je broj kolona bloka A_{ij} jednak broju vrsta bloka B_{jk} $(i = 1,\ldots,p;\ j = 1,\ldots,q\,;\ k = 1,\ldots,s)$. Tada je

$$AB = \begin{bmatrix} C_{11} & \cdots & C_{1s} \\ \vdots & & \\ C_{p1} & & C_{ps} \end{bmatrix},$$

gde je

$$C_{ik} = \sum_{j=1}^{q} A_{ij}B_{jk} \qquad (i=1,\ldots,p\,;\ k=1,\ldots,s).$$

Primer 1.1.1. Neka su

$$A = \begin{bmatrix} 5 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 8 & 3 \\ 0 & 0 & 5 & 2 \end{bmatrix} \quad \text{i} \quad B = \begin{bmatrix} 1 & -2 & -1 & 0 & 0 \\ -2 & 5 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & -3 \\ 0 & 0 & 0 & -5 & 8 \end{bmatrix}.$$

Ako stavimo

$$A = \begin{bmatrix} A_{11} & \mathbf{O} \\ \mathbf{O} & A_{22} \end{bmatrix} \quad \text{i} \quad B = \begin{bmatrix} B_{11} & \mathbf{O} \\ \mathbf{O} & B_{22} \end{bmatrix},$$

gde su

$$A_{11} = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix},\quad A_{22} = \begin{bmatrix} 8 & 3 \\ 5 & 2 \end{bmatrix},\quad B_{11} = \begin{bmatrix} 1 & -2 & -1 \\ -2 & 5 & 2 \end{bmatrix},\quad B_{22} = \begin{bmatrix} 2 & -3 \\ -5 & 8 \end{bmatrix},$$

imamo

$$AB = \begin{bmatrix} A_{11}B_{11} & \mathbf{O} \\ \mathbf{O} & A_{22}B_{22} \end{bmatrix}.$$

Kako je

$$A_{11}B_{11} = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -2 & -1 \\ -2 & 5 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

i

$$A_{22}B_{22} = \begin{bmatrix} 8 & 3 \\ 5 & 2 \end{bmatrix} \cdot \begin{bmatrix} 2 & -3 \\ -5 & 8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

dobijamo

$$AB = \left[\begin{array}{ccc|cc} 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

Teorema 1.1.3. Neka je

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

regularna matrica, gde su A_{11} i A_{22} kvadratne matrice. Ako je matrica A_{22} regularna, tada je

$$A^{-1} = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix},$$

gde su

$$X_{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}, \quad X_{12} = -X_{11}A_{12}A_{22}^{-1},$$

$$X_{21} = -A_{22}^{-1}A_{21}X_{11}, \quad X_{22} = A_{22}^{-1}(I - A_{21}X_{12}).$$

Primer 1.1.2. Za matricu

$$A = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 0 & 3 & 1 & 4 \\ 2 & 7 & 6 & -1 \\ 1 & 2 & 2 & -1 \end{bmatrix},$$

odredićemo A^{-1}. Neka su

$$A_{11} = \begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix}, \quad A_{12} = \begin{bmatrix} 1 & -1 \\ 1 & 4 \end{bmatrix}, \quad A_{21} = \begin{bmatrix} 2 & 7 \\ 1 & 2 \end{bmatrix}, \quad A_{22} = \begin{bmatrix} 6 & -1 \\ 2 & -1 \end{bmatrix}.$$

Na osnovu teoreme 1.1.3, imamo

$$X_{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} = \frac{1}{6}\begin{bmatrix} -1 & 3 \\ -7 & -3 \end{bmatrix},$$

$$X_{12} = -X_{11}A_{12}A_{22}^{-1} = \frac{1}{6}\begin{bmatrix} -7 & 20 \\ 5 & -10 \end{bmatrix},$$

$$X_{21} = -A_{22}^{-1}A_{21}X_{11} = \frac{1}{6}\begin{bmatrix} 9 & 3 \\ 3 & 3 \end{bmatrix},$$

$$X_{22} = A_{22}^{-1}(I - A_{21}X_{12}) = \frac{1}{6}\begin{bmatrix} -3 & 6 \\ -3 & 6 \end{bmatrix}.$$

Dakle,

$$A^{-1} = \left[\begin{array}{cc:cc} -1 & 3 & -7 & 20 \\ -7 & -3 & 5 & -10 \\ \hdashline 9 & 3 & -3 & 6 \\ 3 & 3 & -3 & 6 \end{array}\right].$$

<u>Definicija 1.1.2.</u> Kvadratna matrica A ima svojstvo (A) ako se permutacijom vrsta i kolona može svesti na oblik

$$\begin{bmatrix} D_1 & F_1 & O & O & \ldots & O & O \\ E_1 & D_2 & F_2 & O & & O & O \\ O & E_2 & D_3 & F_3 & & O & O \\ \vdots & & & & & & \\ O & O & O & O & & E_{m-1} & D_m \end{bmatrix},$$

gde su $D_j (j = 1,\ldots,m)$ dijagonalne matrice, a E_j i F_j $(j = 1,\ldots, m-1)$ pravougaone.

Poslednju definiciju uveo je D.Young([1]). Matrice sa svojstvom (A) pojavljuju se kod iterativnih metoda za rešavanje parcijalnih jednačina eliptičkog tipa.

4.1.2. LR faktorizacija kvadratne matrice

Često se kod rešavanja sistema linearnih jednačina javlja problem predstavljanja kvadratne matrice kao proizvod dve trougaone matrice. Ovaj odeljak posvećen je ovom problemu.

Teorema 1.2.1. Ako su sve determinante

$$\Delta_k = \begin{vmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \\ a_{k1} & & a_{kk} \end{vmatrix} \qquad (k = 1,\ldots,n-1)$$

različite od nule, matrica $A = [a_{ij}]_{n\times n}$ može se predstaviti u obliku

(1.2.1) $$A = LR,$$

gde je L donja i R gornja trougaona matrica.

Trougaone matrice L i R reda n, imaju oblike:

(1.2.2) $$L = [\ell_{ij}]_{n\times n} \qquad (\ell_{ij} = 0 \Leftarrow i<j),$$

(1.2.3) $$R = [r_{ij}]_{n\times n} \qquad (r_{ij} = 0 \Leftarrow i>j).$$

Razlaganje (1.2.1), poznato kao LR faktorizacija(dekompozicija), nije jedinstveno, s obzirom na jednakost

$$LR = (cL)(\tfrac{1}{c}R) \qquad (\forall c \neq 0).$$

Medjutim, ako se dijagonalnim elementima matrice R(ili L) fiksiraju vrednosti od kojih nijedna nije jednaka nuli, razlaganje je jedinstveno.

S obzirom na (1.2.2) i (1.2.3) i imajući u vidu da je

$$a_{ij} = \sum_{k=1}^{\min(i,j)} \ell_{ik} r_{kj} \qquad (i,j = 1,\ldots,n),$$

elementi matrica L i R mogu se lako odrediti rekurzivnim postupkom, ukoliko se unapred zadaju elementi $r_{ii}(\neq 0)$ ili ℓ_{ii} $(\neq 0)(i = 1,\ldots,n)$.

Tako, na primer, neka su dati brojevi $r_{ii}(\neq 0)(i = 1,\ldots,n)$.

Tada važi

$$(1)\qquad \ell_{11}=\frac{a_{11}}{r_{11}},\qquad \left.\begin{aligned} r_{1i}&=\frac{a_{1i}}{\ell_{11}}\\ \ell_{i1}&=\frac{a_{i1}}{r_{11}}\end{aligned}\right\}\ (i=2,\dots,n);$$

$$(i)\qquad \left.\begin{aligned} \ell_{ii}&=\frac{1}{r_{ii}}\Big(a_{ii}-\sum_{k=1}^{i-1}\ell_{ik}r_{ki}\Big),\\ &\left.\begin{aligned} r_{ij}&=\frac{1}{\ell_{ii}}\Big(a_{ij}-\sum_{k=1}^{i-1}\ell_{ik}r_{kj}\Big)\\ \ell_{ji}&=\frac{1}{r_{ii}}\Big(a_{ji}-\sum_{k=1}^{i-1}\ell_{jk}r_{ki}\Big)\end{aligned}\right\}(j=i+1,\dots,n);\end{aligned}\right\}(i=2,\dots,n).$$

Slično bismo mogli iskazati i rekurzivni postupak za odredjivanje elemenata matrica L i R ako su unapred dati brojevi $\ell_{ii}(\neq 0)(i=1,\dots,n)$.

U primenama, najčešće se uzima $r_{ii}=1(i=1,\dots,n)$ ili $\ell_{ii}=1(i=1,\dots,n)$.

Primer 1.2.1. Razložimo matricu

$$A=\begin{bmatrix}1 & 4 & 1 & 3\\ 0 & -1 & 2 & -1\\ 3 & 14 & 4 & 1\\ 1 & 2 & 2 & 9\end{bmatrix}$$

u obliku (1.2.1), tako da jediničnu dijagonalu ima

a) matrica R ; b) matrica L.

a) Kako je $r_{ii}=1$ $(i=1,\dots,4)$, na osnovu izloženog rekurzivnog postupka, imamo redom:

(1) $\ell_{11}=1$,

$r_{12}=4,\ \ell_{21}=0,\ r_{13}=1,\ \ell_{31}=3,\ r_{14}=3,\ \ell_{41}=1$;

(2) $\ell_{22} = -1$,

$r_{23} = -2$, $\ell_{32} = 2$, $r_{24} = 1$, $\ell_{42} = -2$;

(3) $\ell_{33} = 5$,

$r_{34} = -2$, $\ell_{43} = -3$;

(4) $\ell_{44} = 2$.

Dakle, dobili smo

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 3 & 2 & 5 & 0 \\ 1 & -2 & -3 & 2 \end{bmatrix} \quad \text{i} \quad R = \begin{bmatrix} 1 & 4 & 1 & 3 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

b) Polazeći od $\ell_{ii} = 1 (i=1,\ldots,4)$, imamo redom

(1) $r_{11} = 1$,

$r_{12} = 4$, $\ell_{21} = 0$, $r_{13} = 1$, $\ell_{31} = 3$, $r_{14} = 3$, $\ell_{41} = 1$;

(2) $r_{22} = -1$,

$r_{23} = 2$, $\ell_{32} = -2$, $r_{24} = -1$, $\ell_{42} = 2$;

(3) $r_{33} = 5$,

$r_{34} = -10$, $\ell_{43} = -3/5$;

(4) $r_{44} = 2$,

tj.

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 3 & -2 & 1 & 0 \\ 1 & 2 & -3/5 & 1 \end{bmatrix} \quad \text{i} \quad R = \begin{bmatrix} 1 & 4 & 1 & 3 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & 5 & -10 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

U primenama vrlo često se javljaju višedijagonalne matrice, tj. matrice čiji su elementi različiti od nule samo na glavnoj dijagonali i oko glavne dijagonale. Na primer, ako je $a_{ij} \neq 0$ za $|i-j| \leqq 1$ i $a_{ij} = 0$ za $|i-j| > 1$, matrica je trodijagonalna. Obično elemente ovakve matrice predstavljamo vektorima $(a_2,\ldots,a_n)$, $(b_1,\ldots,b_n)$, $(c_1,\ldots,c_{n-1})$, tj.

$$(1.2.4)\qquad A = \begin{bmatrix} b_1 & c_1 & 0 & \dots & 0 & 0 \\ a_2 & b_2 & c_2 & & 0 & 0 \\ 0 & a_3 & b_3 & & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & & a_n & b_n \end{bmatrix}.$$

Ako je $a_{ij} \neq 0$ ($|i-j| \leq 2$) i $a_{ij} = 0$ ($|i-j| > 2$), imamo slučaj pentodijagonalne matrice.

Pretpostavimo sada da trodijagonalna matrica (1.2.4) ispunjava uslove teoreme 1.2.1. Za dekompoziciju ovakve matrice dovoljno je pretpostaviti da su

$$L = \begin{bmatrix} \beta_1 & 0 & 0 & \dots & 0 & 0 \\ \alpha_2 & \beta_2 & 0 & & 0 & 0 \\ 0 & \alpha_3 & \beta_3 & & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & & \alpha_n & \beta_n \end{bmatrix} \qquad (\beta_1\beta_2\dots\beta_n \neq 0)$$

i

$$R = \begin{bmatrix} 1 & \gamma_1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \gamma_2 & & 0 & 0 \\ 0 & 0 & 1 & & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & & 0 & 1 \end{bmatrix}.$$

Uporedjivanjem odgovarajućih elemenata matrice A i matrice

$$LR = \begin{bmatrix} \beta_1 & \beta_1\gamma_1 & 0 & \dots & 0 & 0 \\ \alpha_2 & \alpha_2\gamma_1+\beta_2 & \beta_2\gamma_2 & & 0 & 0 \\ 0 & \alpha_3 & \alpha_3\gamma_2+\beta_3 & & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & & \alpha_n & \alpha_n\gamma_{n-1}+\beta_n \end{bmatrix}$$

dobijamo sledeće rekurzivne formule za odredjivanje elemenata α_i, β_i, γ_i:

$$\beta_1 = b_1, \qquad \gamma_1 = \frac{c_1}{\beta_1},$$

$$(1.2.5)\qquad \alpha_i = a_i, \quad \beta_i = b_i - \alpha_i\gamma_{i-1}, \quad \gamma_i = \frac{c_i}{\beta_i} \quad (i=2,\dots,n-1),$$

$$\alpha_n = a_n, \quad \beta_n = b_n - \alpha_n\gamma_{n-1}.$$

4.1.3. Sopstveni vektori i sopstvene vrednosti matrica

Definicija 1.3.1. Neka je A kompleksna kvadratna matrica reda n. Svaki vektor $\vec{x} \in C^n$, koji je različit od nula-vektora, naziva se sopstveni vektor matrice A ako postoji skalar $\lambda \in C$ takav da je

$$A\vec{x} = \lambda \vec{x} . \tag{1.3.1}$$

Skalar λ naziva se odgovarajuća sopstvena vrednost.

S obzirom da se (1.3.1) može predstaviti u obliku

$$(A - \lambda I)\ \vec{x} = \vec{0},$$

zaključujemo da jednačina (1.3.1) ima netrivijalna rešenja (po $\vec{x}$) ako i samo ako je

$$\det(A - \lambda I) = 0.$$

Definicija 1.3.2. Ako je A kvadratna matrica, polinom $\lambda \mapsto P(\lambda) = \det(A - \lambda I)$ naziva se njen karakteristični polinom, a odgovarajuća jednačina $P(\lambda) = 0$ njena karakteristična jednačina.

Neka je $A = [a_{ij}]_{nxn}$. Karakteristični polinom može se predstaviti u obliku

$$P(\lambda) = \begin{vmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-\lambda & & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdot & a_{nn}-\lambda \end{vmatrix}$$

ili

$$P(\lambda) = (-1)^n(\lambda^n - p_1\lambda^{n-1} + p_2\lambda^{n-2} - \dots + (-1)^{n-1}p_{n-1}\lambda + (-1)^n p_n),$$

gde je p_k zbir svih glavnih minora reda k determinante matrice A, tj.

$$p_k = \sum_{1 \le i_1 < i_2 < \dots < i_k \le n} \det(A_{i_1 i_2 \dots i_k}^{i_1 i_2 \dots i_k}).$$

Primetimo da je

$$p_1 = \sum_{i=1}^{n} a_{ii} = \operatorname{tr} A \quad i \quad p_n = \det(A).$$

Često se umesto karakterističnog polinoma P koristi tzv. normiran karakteristični polinom H, definisan pomoću

$$H(\lambda) = (-1)^n P(\lambda) = \lambda^n - p_1\lambda^{n-1} + p_2\lambda^{n-2} - \ldots + (-1)^n p_n.$$

Sopstvene vrednosti matrice A (tj. nule polinoma P) $\lambda_i (i=1,\ldots,n)$ označavaćemo sa $\lambda_i(A)$.

Definicija 1.3.3. Skup svih sopstvenih vrednosti kvadratne matrice A naziva se spektrom te matrice.

Definicija 1.3.4. Spektralni radijus $\rho(A)$ kvadratne matrice A je broj

$$\rho(A) = \max_i |\lambda_i(A)|.$$

Teorema 1.3.1. Svaka matrica je, u matričnom smislu, nula svog karakterističnog polinoma.

Ova teorema je poznata kao Cayley-Hamiltonova teorema.

Teorema 1.3.2. Neka su $\lambda_1,\ldots,\lambda_n$ sopstvene vrednosti matrice A reda n i neka je $x \mapsto Q(x)$ skalarni polinom stepena m. Tada su

$$Q(\lambda_1),\ldots, Q(\lambda_n)$$

sopstvene vrednosti matrice $Q(A)$.

Teorema 1.3.3. Neka su $\lambda_1,\ldots,\lambda_n$ sopstvene vrednosti regularne matrice A reda n. Tada su

$$\lambda_1^{-1},\ldots,\lambda_n^{-1}$$

sopstvene vrednosti matrice A^{-1}.

Teorema 1.3.4. Sopstvene vrednosti trougaone matrice jednake su dijagonalnim elementima.

Sledeća teorema daje rekurzivni postupak za nalaženje karakterističnog polinoma trodijagonalne matrice (1.2.4).

Teorema 1.3.5. Neka je

$$A_k = \begin{bmatrix} b_1 & c_1 & 0 & \dots & 0 \\ a_2 & b_2 & c_2 & & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & & b_k \end{bmatrix} \quad \text{i} \quad H_k(\lambda) = (-1)^k \det(A_k - \lambda I).$$

Normiran karakteristični polinom $\lambda \mapsto H(\lambda) (=H_n(\lambda))$ matrice $A(=A_n)$ dobija se rekurzivnim postupkom

$$H_k(\lambda) = (\lambda - b_k) H_{k-1}(\lambda) - a_{k-1} c_{k-1} H_{k-2}(\lambda) \quad (k=2,\dots,n),$$

gde su $H_o(\lambda)=1$ i $H_1(\lambda)=\lambda-b_1$.

Definicija 1.3.5. Za matricu B kaže se da je slična matrici A ako postoji bar jedna regularna matrica C, takva da je

$$B = C^{-1} AC.$$

Teorema 1.3.6. Slične matrice imaju identične karakteristične polinome, a samim tim i iste sopstvene vrednosti.

4.1.4. Specijalne matrice i njihove osobine

Neka je $A=[a_{ij}]_{mxn}$ $(a_{ij} \varepsilon C)$. Sa $\bar{A}$ označićemo konjugovanu matricu matrici A, tj. $\bar{A}=[\bar{a}_{ij}]_{mxn}$, dok ćemo sa A^T označavati transponovanu matricu od A. Nadalje, matricu $\bar{A}^T$ označavaćemo sa A*.

Lako se uočavaju sledeće osobine

$$(A+B)^* = A^*+B^*, \quad (AB)^*=B^*A^*, \quad (A^*)^*=A,$$

$$(A^*A)^*=A^*A, \quad \det A^*=\det A, \quad (A^{-1})^*=(A^*)^{-1}.$$

Definicija 1.4.1. Ako za kvadratnu matricu A važi jednakost $A=A^T$, matrica A se naziva simetrična matrica.

Definicija 1.4.2. Ako za kvadratnu matricu A važi A=A*, matrica A se naziva hermitska matrica.

Definicija 1.4.3. Ako za kvadratnu matricu A važi A=-A*, matrica A se naziva kosohermitska matrica.

Definicija 1.4.4. Ako za kvadratnu matricu A važi A*A=I (I jedinična matrica), matrica A se naziva unitarna matrica.

Definicija 1.4.5. Ako za regularnu matricu važi $A^T = A^{-1}$, matrica A se naziva ortogonalna matrica.

Korišćenjem skalarnog proizvoda dva vektora $\vec{x} = [x_1 \dots x_n]^T$ i $\vec{y} = [y_1 \dots y_n]^T$, definisanog pomoću

$$(\vec{x},\vec{y}) = \vec{y}^*\vec{x} = \sum_{i=1}^{n} x_i \bar{y}_i ,$$

uslov za hermitsku matricu ($A = A^*$) može se predstaviti u ekvivalentnoj formi

$$(A\vec{x},\vec{y}) = (\vec{x},A\vec{y}) \qquad (\forall \vec{x},\vec{y} \in C^n).$$

Slično se uslov za kosohermitsku matricu ($A = -A^*$) može predstaviti u obliku

$$(A\vec{x},\vec{y}) + (\vec{x},A\vec{y}) = 0 \qquad (\forall \vec{x},\vec{y} \in C^n).$$

Definicija 1.4.6. Hermitska matrica A naziva se pozitivno definitna ako je za svako $\vec{x} \neq \vec{o}$ ispunjen uslov

$$(1.4.1) \qquad (A\vec{x},\vec{x}) = \vec{x}^*A\vec{x} > 0.$$

Ako je $A = [a_{ij}]_{n\times n}$ i $\vec{x} = [x_1 \dots x_n]^T$ uslov (1.4.1) može se predstaviti u obliku

$$(A\vec{x},\vec{x}) = \sum_{i,j=1}^{n} a_{ij}\bar{x}_i x_j > 0,$$

odakle zaključujemo da je $a_{ii} > 0 \, (i = 1,\dots,n)$.

Definicija 1.4.7. Simetrična pozitivno definitna matrica naziva se normalna matrica.

Teorema 1.4.1. Matrica $A = [a_{ij}]_{n\times n}$ je pozitivno definitna ako i samo ako su ispunjeni uslovi

$$(1.4.2) \qquad \Delta_k = \begin{vmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \\ a_{k1} & & a_{kk} \end{vmatrix} > 0 \quad (k = 1,\dots,n).$$

Uslovi (1.4.2) poznati su kao Sylvesterovi uslovi.

Teorema 1.4.2. Ako matrica $C = [c_{ij}]_{m\times n}$ $(m \geq n)$ ima rang n, tada je matrica C^*C pozitivno definitna.

Teorema 1.4.3. Sve sopstvene vrednosti hermitske matrice su realni brojevi.

Teorema 1.4.4. Sve sopstvene vrednosti pozitivno definitne matrice su pozitivni brojevi.

Teorema 1.4.5. Sve sopstvene vrednosti koso-hermitske matrice su čisto imaginarni brojevi.

Teorema 1.4.6. Ako je A hermitska matrica, tada je

1° $\lambda(A^* A) = \lambda(A)^2$;

2° $\lambda_{min}(A)(\vec{x},\vec{x}) \leqq (A\vec{x},\vec{x}) \leqq \lambda_{max}(A)(\vec{x},\vec{x})$.

Znaci jednakosti u 2° se postižu za sopstvene vektore koji odgovaraju sopstvenim vrednostima $\lambda_{min}(A)$ i $\lambda_{max}(A)$.

4.1.5. Jordanov kanonički oblik

Definicija 1.5.1. Kvadratna matrica reda r

$$J_1(\lambda) = [\lambda] \qquad (r = 1),$$

$$(1.5.1) \qquad J_r(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & 0 & & 0 & 0 \\ 0 & 0 & \lambda & 1 & & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & 0 & & \lambda & 1 \\ 0 & 0 & 0 & 0 & & 0 & \lambda \end{bmatrix} \qquad (r \geq 2),$$

naziva se Jordanov blok.

Definicija 1.5.2. Kvazidijagonalna matrica oblika

$$(1.5.2) \qquad J = \begin{bmatrix} J_{r_1}(\lambda) & 0 & \dots & 0 \\ 0 & J_{r_2}(\lambda) & & 0 \\ \vdots & & & \\ 0 & 0 & & J_{r_m}(\lambda) \end{bmatrix}$$

naziva se Jordanova matrica.

Jordanova matrica često se predstavlja i u obliku

(1.5.3) $$J = J_{r_1}(\lambda_1) \dotplus J_{r_2}(\lambda_2) \dotplus \dots \dotplus J_{r_m}(\lambda_m).$$

Pod Jordanovim kanoničkim oblikom matrice A podrazumeva se ona Jordanova matrica koja je slična matrici A i njena egzistencija sleduje iz sledeće teoreme(videti, na primer, [2]).

Teorema 1.5.1. Svaka klasa sličnih matrica sadrži bar jednu Jordanovu matricu.

Napomenimo, da je broj Jordanovih blokova u (1.5.2), tj. (1.5.3), jednak broju linearno nezavisnih vektora matrice A.

Primetimo, takodje, da je λ_i sopstvena vrednost matrica $J_{r_i}(\lambda_i)$, J, A.

Primedba 1.5.1. Neka je $x \mapsto Q(x) = a_0x^k + a_1x^{k-1} + \dots + a_k$ skalarni polinom i $J_r(\lambda)$ Jordanov blok (1.5.1). Tada je

$$Q(J_r(\lambda)) = \begin{bmatrix} Q(\lambda) & Q'(\lambda) & \frac{1}{2!}Q''(\lambda) & \dots & \frac{1}{(r-1)!}Q^{(r-1)}(\lambda) \\ 0 & Q(\lambda) & Q'(\lambda) & & \frac{1}{(r-2)!}Q^{(r-2)}(\lambda) \\ \vdots & & & & \\ 0 & 0 & 0 & & Q(\lambda) \end{bmatrix}.$$

Štaviše, ako je J Jordanova matrica, imamo

$$Q(J) = Q(J_{r_1}(\lambda_1)) \dotplus Q(J_{r_2}(\lambda_2)) \dotplus \dots \dotplus Q(J_{r_m}(\lambda_m)).$$

Teorema 1.5.2. Za proizvoljnu hermitsku matricu A, postoji ortogonalna matrica S, takva da je S^*AS dijagonalna matrica, tj.

$$S^*AS = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}.$$

Dve važne posledice ove teoreme su:

Posledica 1.5.1. Svaka hermitska matrica A reda n ima n linearno nezavisnih sopstvenih vektora, koji obrazuju ortogonalni sistem.

Posledica 1.5.2. Potrebni i dovoljni uslovi, da je hermitska matrica A reda n pozitivno definitna su $\lambda_i(A) > 0 (i = 1,\dots,n)$.

4.1.6. Norme vektora i matrica

U teoriji iterativnih procesa veoma važnu ulogu, pri ispitivanju konvergencije, imaju norme vektora i norme matrica.

Neka je X realan prostor R^n(kompleksan prostor C^n) sa nula-vektorom $\vec{o}(=[o \ \dots \ o]^T)$. Za vektor $\vec{x}(=[x_1 \dots x_n]^T)$ definiše se norma saglasno definiciji 1.2.1(odeljak 2.1.2). Od svih mogućnih normi vektora od interesa su norme oblike

(1.6.1) $$\|\vec{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \qquad (1 \leq p < +\infty).$$

U graničnom slučaju, kada $p \to +\infty$, iz (1.6.1) sleduje

(1.6.2) $$\|\vec{x}\|_\infty = \max_i |x_i| .$$

U praktičnim računanjima, pored norme (1.6.2), koriste se i norme koje se iz (1.6.1) dobijaju za $p=1$ i $p=2$, tj.

(1.6.3) $$\|\vec{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

i

(1.6.4) $$\|\vec{x}\|_2 = \|\vec{x}\|_E = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} .$$

Norma (1.6.4) poznata je kao Euklidska norma. Primetimo da je $\|\vec{x}\|_E = (\vec{x}^*\vec{x})^{1/2} = \sqrt{(\vec{x},\vec{x})}$.

Nije teško primetiti da za proizvoljno $\vec{x} \in X$ važe nejednakosti

$$\|\vec{x}\|_\infty \leq \|\vec{x}\|_1 \leq n\|\vec{x}\|_\infty ,$$

$$\|\vec{x}\|_\infty \leq \|\vec{x}\|_2 \leq \sqrt{n}\,\|\vec{x}\|_\infty ,$$

$$\frac{1}{\sqrt{n}}\|\vec{x}\|_1 \leq \|\vec{x}\|_2 \leq \|\vec{x}\|_1 .$$

Neka je X_M linearan realan ili kompleksan prostor kvadratnih matrica reda n, sa nula-matricom $O(\in X_M)$.

<u>Definicija 1.6.1</u>. Pod normom matrice $A \in X_M$ podrazumeva se nenegativan broj $\|A\|$, takav da je

1° $\|A\| = 0 \iff A = 0$ (definisanost);

2° $\|cA\| = |c| \cdot \|A\|$ (homogenost);

3° $\|A+B\| \leqq \|A\| + \|B\|$ (relacija trougla),

gde su $A, B \in X_M$ i $c \in C$.

Napomenimo da se norma $\|A\|$ može razmatrati i kao norma operatora A koji se primenjuje na vektore iz prostora X.

Sada dajemo definicije saglasnosti i potčinjenosti norme matrice sa normom vektora.

<u>Definicija 1.6.2</u>. Norma $\|A\|$ saglasna je sa normom $\|\vec{x}\|$, ako je

$$\|A\vec{x}\| \leqq \|A\| \cdot \|\vec{x}\| \qquad (\forall A \in X_M \wedge \forall \vec{x} \in X)$$

i

$$\|AB\| \leqq \|A\| \cdot \|B\| \qquad (\forall A, B \in X_M).$$

<u>Definicija 1.6.3</u>. Neka je norma $\|A\|$ saglasna sa normom $\|\vec{x}\|$. Ako se za svako $A \in X_M$ može naći $\vec{x} (\neq \vec{o})$, takvo da je $\|A\vec{x}\| = \|A\| \|\vec{x}\|$, za normu $\|A\|$ kaže se da je potčinjena normi $\|\vec{x}\|$.

Za svaku normu matrica, potčinjenu normi vektora, važi $\|I\| = 1$, gde je I jedinična matrica.

Može se dokazati (videti, na primer, [3]) da za svaku normu vektora $\|\vec{x}\|$ postoji bar jedna potčinjena norma $\|A\|$. Na primer, norma matrice A, definisana pomoću

$$(1.6.5) \qquad \|A\| = \max_{\|\vec{x}\| = 1} \|A\vec{x}\| = \max_{\vec{x} \neq \vec{o}} \frac{\|A\vec{x}\|}{\|\vec{x}\|},$$

potčinjena je upotrebljenoj normi vektora.

Na osnovu (1.6.5) možemo, na primer, naći potčinjenu normu matrice, za Euklidsku normu vektora. Tako imamo

$$\|A\| = \max_{\vec{x} \neq \vec{o}} \frac{\|A\vec{x}\|_E}{\|\vec{x}\|_E}.$$

Kako je $\|A\vec{x}\|_E^2 = (A\vec{x})^* (A\vec{x}) = \vec{x}^* A^* A\vec{x}$ i matrica $A^* A$ hermitska, imamo

$$\|A\vec{x}\|_E^2 \leqq \lambda_{max} \|\vec{x}\|_E^2 , \tag{1.6.6}$$

gde je λ_{max} najveća sopstvena vrednost matrice A^*A(videti teoremu 1.4.6).

Nejednakost (1.6.6) sugeriše sledeću definiciju:

Definicija 1.6.4. Pod spektralnom normom kvadratne matrice A podrazumeva se broj $\sigma(A)$, dat pomoću

$$\sigma(A) = \|A\|_{sp} = +\sqrt{\max \lambda(A^*A)} .$$

Za spektralnu normu važe sledeći rezultati:

Teorema 1.6.1. Ako je matrica A regularna, tada je

$$\sigma(A^{-1}) = +\sqrt{\frac{1}{\max \lambda(A^*A)}} .$$

Teorema 1.6.2. Spektralni radijus $\rho(A)$ matrice A nije veći od njene spektralne norme, tj. važi $\rho(A) \leqq \sigma(A)$. Ako je matrica hermitska, tada je $\rho(A) = \sigma(A)$.

Teorema 1.6.3. Spektralna norma matrice je potčinjena Euklidskoj normi vektora.

Teorema 1.6.4. Neka je A hermitska matrica. Tada spektralna norma $\sigma(A)$ ima najmanju vrednost od svih mogućnih normi $\|A\|$, saglasnih sa nekom normom vektora.

Pored spektralne norme u upotrebi su i sledeće norme matrice $A = [a_{ij}]_{n\times n}$:

1° $\|A\|_1 = \max\limits_{j}\left(\sum_{i=1}^{n} |a_{ij}|\right)$;

2° $\|A\|_2 = \varepsilon(A) = \left(\sum_{i,j=1}^{n} |a_{ij}|^2\right)^{1/2}$ (norma Schmidta) ;

3° $\|A\|_\infty = \max\limits_{i}\left(\sum_{j=1}^{n} |a_{ij}|\right)$.

Za norme 1^o, 2^o, 3^o važe sledeće teoreme:

Teorema 1.6.5. Norma $\|A\|_1$ potčinjena je normi vektora $\|\vec{x}\|_1$.

Teorema 1.6.6. Norma Schmidta $\mathcal{E}(A)$ za matricu reda n saglasna je sa Euklidskom normom (1.6.4), ali joj nije potčinjena za $n>1$.

Teorema 1.6.7. Norma $\|A\|_\infty$ potčinjena je normi vektora $\|\vec{x}\|_\infty$.

Lako se može dokazati da za proizvoljnu kvadratnu matricu reda n važe sledeće nejednakosti

$$\frac{1}{n}\, m(A) \leqq \|A\|_p \leqq m(A) \qquad (p=1,2,\infty),$$

$$\frac{1}{\sqrt{n}}\, \mathcal{E}(A) \leqq \|A\|_p \leqq \sqrt{n}\, \mathcal{E}(A) \qquad (p=1,\infty),$$

$$\frac{1}{\sqrt{n}}\, \sigma(A) \leqq \|A\|_p \leqq \sqrt{n}\, \sigma(A) \qquad (p=1,\infty),$$

$$m(A) \leqq n\,\sigma(A)\,, \quad \mathcal{E}(A) \leqq \sqrt{n}\,\sigma(A),$$

$$\frac{1}{n}\|A\|_\infty \leqq \|A\|_1 \leqq n\,\|A\|_\infty\,,$$

gde je $m(A) = n \max\limits_{i,j} |a_{ij}|$.

Pomoću regularne matrice H, mogućno je date norme $\|\vec{x}\|$ i $\|A\|$ transformisati u $\|\vec{x}\|^H$ i $\|A\|^H$, tako da ove poslednje zadovoljavaju uslove za normu. Lako je proveriti da $\|\vec{x}\|^H$ i $\|A\|^H$ predstavljaju norme, ako se definišu pomoću

$$\|\vec{x}\|^H = \|H^{-1}\vec{x}\| \quad \text{i} \quad \|A\|^H = \|H^{-1}AH\|\,.$$

Štaviše, svojstva saglasnosti i potčinjenosti ostaju u važnosti i za transformisane norme $\|\vec{x}\|^H$ i $\|A\|^H$.

Kao transformaciona matrica H, najčešće se koristi dijagonalna matrica sa pozitivnim elementima, tj.

$$H = \begin{bmatrix} h_1 & & 0 \\ & \ddots & \\ 0 & & h_n \end{bmatrix} = \operatorname{diag}(h_1,\ldots,h_n) \quad (h_i > 0\,(i=1,\ldots,n)).$$

Na ovaj način, norme $\|\vec{x}\|_1$, $\|A\|_1$ i norme $\|\vec{x}\|_\infty$, $\|A\|_\infty$ transformišu se u

$$\|\vec{x}\|_1^H = \sum_{i=1}^{n} h_i^{-1}|x_i| \,, \quad \|A\|_1^H = \max_j h_j\Big(\sum_{i=1}^{n} h_i^{-1}|a_{ij}|\Big)$$

i

$$\|\vec{x}\|_\infty^H = \max_i h_i^{-1}|x_i| \,, \quad \|A\|_\infty^H = \max_i h_i^{-1}\Big(\sum_{j=1}^{n} h_j|a_{ij}|\Big)$$

respektivno.

4.1.7. Konvergencija matričnih nizova i redova

Posmatrajmo niz vektora $\{\vec{x}^{(k)}\}_{k\in N}$, gde je

$$\vec{x}^{(k)} = \big[x_1^{(k)} \;\dots\; x_n^{(k)}\big]^T .$$

Definicija 1.7.1. Ako postoje konačne granične vrednosti

$$a_i = \lim_{k\to+\infty} x_i^{(k)} \qquad (i=1,\dots,n),$$

kažemo da niz $\{\vec{x}^{(k)}\}_{k\in N}$ konvergira ka vektoru $\vec{a} = [a_1 \dots a_n]^T$, a vektor $\vec{a}$ nazivamo graničnom vrednošću niza $\{\vec{x}^{(k)}\}_{k\in N}$.

Slično definišemo i konvergenciju niza matrica $\{A^{(k)}\}_{k\in N}$ gde je $A^{(k)} = [a_{ij}^{(k)}]_{n\times n}$.

Definicija 1.7.2. Ako postoje konačne granične vrednosti

$$a_{ij} = \lim_{k\to+\infty} a_{ij}^{(k)} \qquad (i,j=1,\dots,n),$$

kažemo da niz $\{A^{(k)}\}_{k\in N}$ konvergira ka matrici $A = [a_{ij}]_{n\times n}$. Matricu A nazivamo graničnom vrednošću niza $\{A^{(k)}\}_{k\in N}$.

Mogu se dati i druge definicije konvergencije niza vektora i niza matrica zasnovane na ranije uvedenom pojmu norme (konvergencija po normi). Naime, mi kažemo da $\vec{x}^{(k)} \to \vec{a}$ $(k\to+\infty)$, ako $\|\vec{x}^{(k)} - \vec{a}\| \to 0$, kada $k\to+\infty$. Slično, kažemo da $A^{(k)} \to A$ $(k\to+\infty)$, ako $\|A^{(k)} - A\| \to 0$, kada $k\to+\infty$. Može se, medjutim, pokazati da je definicija konvergencije po normi ekvivalentna prethodno datoj definiciji po koordinatama(videti, na primer, [4]).

Primedba 1.7.1. Na osnovu nejednakosti

$$\left| \|\vec{x}^{(k)}\| - \|\vec{a}\| \right| \leqq \| \vec{x}^{(k)} - \vec{a} \|$$

zaključujemo da iz $\vec{x}^{(k)} \to \vec{a}\ (k \to +\infty)$ sleduje $\|\vec{x}^{(k)}\| \to \|\vec{a}\|\ (k \to +\infty)$. Takodje važi

$$A^{(k)} \to A\ (k \to +\infty) \quad \Rightarrow \quad \|A^{(k)}\| \to \|A\| \quad (k \to +\infty).$$

Teorema 1.7.1. Niz matrica $\{A^k\}_{k \in N}$ konvergira ka nula matrici O, ako i samo ako su sve sopstvene vrednosti matrice A po modulu manje od jedinice.

Dokaz. S obzirom da se matrica A može transformisati na Jordanov oblik, tj. da postoji regularna matrica C, takva da je

$$J = C^{-1}AC,$$

gde je J Jordanova matrica(videti odeljak 4.1.5), imamo, za svako $k \in N$,

$$A^k = CJ^kC^{-1},$$

odakle zaključujemo da niz $\{A^k\}_{k \in N}$ konvergira(ne konvergira) ka nula matrici ako i samo ako niz $\{J^k\}_{k \in N}$ konvergira(ne konvergira) ka nula matrici.

Neka su $\lambda_1, \dots, \lambda_m$ $(m \leqq n)$ medjusobno različite sopstvene vrednosti matrice A(matrice J). Kako na osnovu primedbe 1.5.1 (sa $Q(x) = x^k$) važe ekvivalencije

$$\lim_{k \to +\infty} J_r(\lambda)^k = 0 \quad \Leftrightarrow \quad |\lambda| < 1$$

i

$$\lim_{k \to +\infty} J^k = 0 \quad \Leftrightarrow \quad \lim_{k \to +\infty} J_{r_i}(\lambda_i)^k = 0 \quad (\forall i),$$

dokaz je završen.

Sledeća teorema daje dovoljan uslov za konvergenciju niza $\{A^k\}_{k \in N}$ ka nula matrici.

Teorema 1.7.2. Ako je bilo koja norma matrice A manja od jedinice važi $A^k \to O(k \to +\infty)$.

Dokaz. Kako je

$$\|A^k - O\| = \|A^k\| = \|AA^{k-1}\| \leq \|A\| \cdot \|A^{k-1}\| \leq \ \dots \ \leq \|A\|^k$$

i kako je po pretpostavci $\|A\| < 1$, imamo $\|A^k\| \to o\ (k \to +\infty)$, tj. $A^k \to O\ (k \to +\infty)$.

Ovim je dokaz završen.

Sada ćemo dokazati jedan važan rezultat koji je u vezi sa teoremama 1.6.2.i 1.6.4.

Teorema 1.7.3. Spektralni radijus $\rho(A)$ matrice A nije veći od bilo koje njene norme.

Dokaz. Za proizvoljan pozitivan broj ε definišimo matricu B pomoću

$$B = \frac{1}{\|A\| + \varepsilon} A .$$

Kako je $\|B\| = \dfrac{1}{\|A\| + \varepsilon} \|A\| < 1$, iz teoreme 1.7.2 sleduje $B^k \to O$ $(k \to +\infty)$. S druge strane, na osnovu teoreme 1.7.1 zaključujemo da su sve sopstvene vrednosti matrice B po modulu manje od jedinice, tj. $|\lambda_i(B)| < 1$. Ako sa $\lambda_i(A)$ označimo proizvoljnu sopstvenu vrednost matrice A, tada je

$$|\lambda_i(B)| = \frac{1}{\|A\| + \varepsilon} |\lambda_i(A)| < 1,$$

tj. $|\lambda_i(A)| < \|A\| + \varepsilon$. S obzirom da se ε može uzeti dovoljno malo zaključujemo da je $|\lambda_i(A)| \leq \|A\|$, tj. $\rho(A) \leq \|A\|$.

Oslanjajući se na koncept konvergencije matričnog niza, moguće je definisati matrični red pomoću

$$\sum_{m=0}^{+\infty} B^{(m)} = \lim_{k \to +\infty} \sum_{m=0}^{k} B^{(m)} , \tag{1.7.1}$$

gde su $B^{(m)}$ $(m = 0,1,\dots)$ matrice istog reda.

Teorema 1.7.4. Ako matrični red (1.7.1) konvergira tada je

$$\lim_{k \to +\infty} B^{(k)} = O.$$

Dokaz. Neka je $S^{(k)} = \sum_{m=0}^{k} B^{(m)}$ $(k = 0,1,\ldots)$ i neka je suma reda (1.7.1) S, tj. $\lim_{k\to+\infty} S^{(k)} = S$. Tada je

$$S^{(k)} - S^{(k-1)} = B^{(k)},$$

tj.

$$\lim_{k\to+\infty} S^{(k)} - \lim_{k\to+\infty} S^{(k-1)} = \lim_{k\to+\infty} B^{(k)},$$

odakle neposredno sleduje tvrdjenje teoreme 1.7.4.

U daljem tekstu daćemo potrebne i dovoljne uslove za konvergenciju matrične geometrijske progresije

$$\sum_{m=0}^{+\infty} A^m = I + A + A^2 + \ldots, \tag{1.7.2}$$

gde je A kvadratna matrica reda n. Ovde je $B^{(m)} = A^m (m = 0,1,\ldots)$.

Teorema 1.7.5. Matrični red (1.7.2) konvergira ako i samo ako je

$$\lim_{k\to+\infty} A^k = O. \tag{1.7.3}$$

Štaviše, tada je

$$I + A + A^2 + \ldots = (I - A)^{-1}. \tag{1.7.4}$$

Dokaz. Pretpostavimo da matrični red (1.7.2) konvergira. Tada, na osnovu teoreme 1.7.4, važi (1.7.3).

Obrnuto, pretpostavimo da je ispunjen uslov (1.7.3). Tada su, na osnovu teoreme 1.7.1, sve sopstvene vrednosti matrice A po modulu manje od jedinice, tj. $|\lambda_i(A)| < 1$ $(i = 1,\ldots,n)$. Kako je

$$\det(I - A) = \prod_{i=1}^{n} (1 - \lambda_i(A)) \neq 0,$$

zaključujemo da postoji matrica $(I - A)^{-1}$. Množenjem identiteta

$$(I + A + A^2 + \ldots + A^k)(I - A) = I - A^{k+1},$$

matricom $(I - A)^{-1}$ s desne strane, dobijamo

$$I + A + A^2 + \dots + A^k = (I - A)^{-1} - A^{k+1}(I - A)^{-1},$$

odakle sleduje

$$\lim_{k\to+\infty} (I + A + A^2 + \dots + A^k) = (I - A)^{-1} - \lim_{k\to+\infty} A^{k+1}(I - A)^{-1}$$
$$= (I - A)^{-1},$$

tj. (1.7.4), čime je dokaz teoreme završen.

S obzirom na tvrdjenje teoreme 1.7.1, prvi deo teoreme 1.7.5 se može formulisati i na sledeći način:

Teorema 1.7.6. Matrični red (1.7.2) konvergira ako i samo ako su sve sopstvene vrednosti matrice A po modulu manje od jedinice.

4.2. DIREKTNI METODI U LINEARNOJ ALGEBRI

4.2.1. Uvodne napomene

Numerički problemi u linearnoj algebri mogu se klasifikovati u nekoliko grupa:

1° Rešavanje sistema linearnih algebarskih jednačina

$$A\vec{x} = \vec{b}$$

sa regularnom matricom A, izračunavanje determinante od A i inverzija matrice A;

2° Rešavanje proizvoljnog sistema linearnih jednačina metodom najmanjih kvadrata;

3° Odredjivanje sopstvenih vrednosti i sopstvenih vektora date kvadratne matrice;

4° Rešavanje zadatka linearnog programiranja.

Za rešavanje ovih problema razvijen je čitav niz metoda, koji se mogu podeliti u dve klase.

Prvu klasu ovih metoda čine tzv. direktni metodi ili kako se ponekad nazivaju tačni metodi. Osnovna karakteristika ovih metoda je ta da se posle konačnog broja transformacija(koraka) dolazi do rezultata. Ukoliko bi se sve računske operacije izvodile tačno, dobijeni rezultat bi bio apsolutno tačan. Naravno, kako se proces računanja izvodi sa zaokrugljivanjem medjurezultata, konačan rezultat je ograničene tačnosti. Ovo poglavlje je posvećeno direktnim metodima.

Drugu klasu metoda čine iterativni metodi, kod kojih se rezultat dobija posle beskonačnog broja koraka. Kao početne vrednosti rešenja, kod primene iterativnih metoda, najčešće se koriste rezultati dobijeni nekim od direktnih metoda. O opštoj teoriji iterativnih procesa bilo je reči u trećoj glavi. U poglavlju 4.3 izložićemo glavne osobine iterativnih metoda koji se koriste u linearnoj algebri. Napomenimo da se kod rešavanja sistema sa velikim brojem jednačina, kakvi se javljaju pri rešavanju parcijalnih diferencijalnih jednačina, koriste uglavnom iterativni metodi.

4.2.2. Gaussov metod eliminacije sa izborom glavnog elementa

Posmatrajmo sistem linearnih algebarskih jednačina

$$
\begin{aligned}
&a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1, \\
&a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2, \\
&\qquad\vdots \\
&a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n,
\end{aligned}
\tag{2.2.1}
$$

ili u matričnom obliku

$$A\vec{x} = \vec{b}, \tag{2.2.2}$$

gde su

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & & a_{nn} \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Za sistem jednačina (2.2.2) pretpostavljamo da ima jedinstveno rešenje.

Poznato je da se rešenja sistema (2.2.1), tj. (2.2.2), mogu izraziti pomoću Cramerovih formula

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

gde je A_i matrica dobijena iz matrice A zamenom i-te kolone vektorom $\vec{b}$. Medjutim, ove formule su nepogodne za praktična izračunavanja, s obzirom da je za izračunavanje n+1 determinanata potreban veliki broj računskih operacija. Naime, ako bismo vrednost determinante n-tog reda izračunavali razvijanjem determinante po vrstama ili kolonama, potrebno je izvršiti $S_n = n!-1$ sabiranja. Broj množenja M_n odredjen je rekurentnom relacijom

$$M_n = n + nM_{n-1} .$$

Kako je $M_1 = 0$, lako se dobija da je

$$M_n = n!\left(1 + \frac{1}{2!} + \frac{1}{3!} + \dots + \frac{1}{(n-1)!}\right) \cong n!(e-1) \qquad (n>4),$$

što znači da je ukupan broj računskih operacija $P_n = M_n + S_n \cong n!e$. Pod pretpostavkom da je za obavljanje jedne računske operacije potrebno $10\mu s$, što je slučaj kod brzih računara*, to bi za izračunavanje vrednosti determinante tridesetog reda($n=30$) bilo potrebno oko $\frac{30!\cdot e\cdot 10\cdot 10^{-6}}{3600\cdot 24\cdot 365} \cong 2.3\cdot 10^{20}$ godina. Uopšteno govoreći ovakav postupak je praktično neprimenljiv, već za determinante reda $n>5$.

Jedan od najpogodnijih direktnih metoda za rešavanje sistema linearnih jednačina je Gaussov metod eliminacije. Ovaj metod se zasniva na redukciji sistema (2.2.2), primenom ekvivalentnih transformacija, na trougaoni sistem

$$(2.2.3) \qquad R\vec{x} = \vec{c},$$

gde su

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & r_{2n} \\ & & \ddots & \\ & & & r_{nn} \end{bmatrix} \quad \text{i} \quad \vec{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} .$$

*) Na sistemu IBM 1130 vreme izvršenja pojedinih operacija u μs, u običnoj(A) i proširenoj tačnosti(B) je dato u tabeli

	A	B
sabiranje	460	440
oduzimanje	560	490
množenje	560	790
deljenje	760	2060

Sistem (2.2.3) se rešava sukcesivno polazeći od poslednje jednačine. Naime,

$$x_n = \frac{c_n}{r_{nn}},$$

$$x_i = \frac{1}{r_{ii}}\left(c_i - \sum_{k=i+1}^{n} r_{ik}x_k\right) \qquad (i = n-1,\ldots,1).$$

Napomenimo da su koeficijenti $r_{ii} \neq 0$, jer po pretpostavci sistem (2.2.2), tj. (2.2.3) ima jedinstveno rešenje.

Pokazaćemo sada kako se sistem (2.2.1) može redukovati na ekvivalentan sistem sa trougaonom matricom.

Pod pretpostavkom da je $a_{11} \neq 0$, izračunajmo najpre faktore

$$m_{i1} = \frac{a_{i1}}{a_{11}} \qquad (i = 2,\ldots,n),$$

a zatim množenjem prve jednačine u sistemu (2.2.1) sa m_{i1} i oduzimanjem od i-te jednačine, dobijamo sistem od n-1 jednačina

$$\begin{aligned} a_{22}^{(2)}x_2 + \ldots + a_{2n}^{(2)}x_n &= b_2^{(2)}, \\ &\vdots \\ a_{n2}^{(2)}x_2 + \ldots + a_{nn}^{(2)}x_n &= b_n^{(2)}, \end{aligned} \tag{2.2.4}$$

gde su

$$a_{ij}^{(2)} = a_{ij} - m_{i1}a_{1j}, \quad b_i^{(2)} = b_i - m_{i1}b_1 \quad (i,j = 2,\ldots,n).$$

Pod pretpostavkom da je $a_{22}^{(2)} \neq 0$, primenjujući isti postupak na (2.2.4) sa $m_{i2} = a_{i2}^{(2)}/a_{22}^{(2)}$ $(i = 3,\ldots,n)$ dobijamo sistem od n-2 jednačine

$$\begin{aligned} a_{33}^{(3)}x_3 + \ldots + a_{3n}^{(3)}x_n &= b_3^{(3)}, \\ &\vdots \\ a_{n3}^{(3)}x_3 + \ldots + a_{nn}^{(3)}x_n &= b_n^{(3)}, \end{aligned}$$

gde su

$$a_{ij}^{(3)} = a_{ij}^{(2)} - m_{i2}a_{2j}^{(2)}, \quad b_i^{(3)} = b_i^{(2)} - m_{i2}b_2^{(2)} \quad (i,j = 3,\ldots,n).$$

Nastavljajući ovaj postupak, posle n-1 koraka dolazimo do jednačine

$$a_{nn}^{(n)}x_n = b_n^{(n)}.$$

Iz dobijenih sistema, uzimanjem prvih jednačina, dolazimo do sistema jednačina

$$\begin{aligned}
a_{11}^{(1)}x_1 + a_{12}^{(1)}x_2 + a_{13}^{(1)}x_3 + \ldots + a_{1n}^{(1)}x_n &= b_1^{(1)},\\
a_{22}^{(2)}x_2 + a_{23}^{(2)}x_3 + \ldots + a_{2n}^{(2)}x_n &= b_2^{(2)},\\
a_{33}^{(3)}x_3 + \ldots + a_{3n}^{(3)}x_n &= b_3^{(3)},\\
&\vdots\\
a_{nn}^{(n)}x_n &= b_n^{(n)},
\end{aligned}$$

pri čemu smo stavili $a_{ij}^{(1)} = a_{ij}$, $b_i^{(1)} = b_i$.

Navedena trougaona redukcija ili kako se često kaže Gaussova eliminacija, se svodi na izračunavanje koeficijenata

$$m_{ik} = \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}, \quad a_{ij}^{(k+1)} = a_{ij}^{(k)} - m_{ik}a_{kj}^{(k)}, \quad b_i^{(k+1)} = b_i^{(k)} - m_{ik}b_k^{(k)} \quad (i,j=k+1,\ldots,n)$$

za $k = 1,2,\ldots,n-1$. Primetimo da su elementi matrice R i vektora $\vec{c}$ dati sa

$$r_{ij} = a_{ij}^{(i)}, \quad c_i = b_i^{(i)} \quad (i = 1,\ldots,n\,;\; j = i,\ldots,n).$$

Da bi navedena redukcija egzistirala, potrebno je obezbediti uslov $a_{kk}^{(k)} \neq 0$. Elementi $a_{kk}^{(k)}$ su poznati kao glavni elementi ili stožerski elementi*. Pod pretpostavkom da je matrica A sistema (2.2.2) regularna, uslove $a_{kk}^{(k)} \neq 0$ moguće je obezbediti permutacijom jednačina u sistemu(videti [2]).

*) Na engleskom jeziku pivotal element, ili prosto pivot.

Primer 2.2.1. Neka je dat sistem jednačina

$$\begin{aligned} 2.304x_1 - 1.213x_2 + 2.441x_3 &= 7.201, \\ 8.752x_1 - 5.608x_2 + 3.916x_3 &= 9.284, \\ 1.527x_1 + 4.333x_2 - 2.214x_3 &= 3.551. \end{aligned}$$

Zaokrugljujući sve medjurezultate na četiri značajne cifre, Gaussovom eliminacijom dobijamo trougaoni sistem

$$\begin{aligned} 2.304x_1 - 1.213x_2 + 2.441x_3 &= 7.201, \\ -0.9998x_2 - 5.357x_3 &= -18.07, \\ -31.36x_3 &= -94.07, \end{aligned}$$

odakle dobijamo redom $x_3 = 3.000$, $x_2 = 1.999$, $x_1 = 0.9995$. Faktori m_{ij}, u ovom slučaju, su

$$m_{21} = 3.799, \quad m_{31} = 6.628, \quad m_{32} = -5.138.$$

Napomenimo, da su tačna rešenja datog sistema $x_1 = 1$, $x_2 = 2$, $x_3 = 3$.

S obzirom na numerički postupak za trougaonu redukciju i proces zaokrugljivanja medjurezultata, javiće se greške u elementima matrice R i vektora $\vec{c}$. Naime, umesto sistema (2.2.3) dobijamo sistem $R_o\vec{x} = \vec{c}_o$, gde su $R_o = R + \Delta R$ i $\vec{c}_o = \vec{c} + \Delta\vec{c}$. Rešenje ovog sistema biće $\vec{x}_o = \vec{x} + \Delta\vec{x}$, gde je $\vec{x}$ tačno rešenje sistema (2.2.3). Nije teško ustanoviti da će greška biti veća , što je glavni element $a_{kk}^{(k)}$ manji po modulu od preostalih elemenata matrice. U vezi sa ovim navodimo jedan interesantan primer koji potiče od Wilkinsona([5,str.205]).

Primer 2.2.2. Gaussovim metodom eliminacije rešimo sistem jednačina

$$\begin{bmatrix} 0.000003 & 0.213472 & 0.332147 \\ 0.215512 & 0.375623 & 0.476625 \\ 0.173257 & 0.663257 & 0.625675 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.235262 \\ 0.127653 \\ 0.285321 \end{bmatrix}$$

uzimajući sve medjurezultate sa šest značajnih cifara. S obzirom da su faktori u ovom slučaju

$$m_{21} = 71837.3, \quad m_{31} = 57752.3, \quad m_{32} = 0.803905,$$

odgovarajući trougaoni sistem postaje

$$\begin{bmatrix} 0.000003 & 0.213472 & 0.332147 \\ & -15334.9 & -23860.0 \\ & & -0.500000 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.235262 \\ -16900.5 \\ -0.20000 \end{bmatrix},$$

odakle je $x_3 = 0.400000$, $x_2 = 0.479723$, $x_1 = -1.33333$.

Posmatrajmo sada dati sistem jednačina u kome su prva i druga jednačina permutovane. S obzirom da su sada faktori

$$m_{21} = 0.0000139203 \quad \text{i} \quad m_{31} = 0.803932,$$

posle prvog eliminacionog koraka dolazimo do sistema jednačina

$$\begin{bmatrix} 0.215512 & 0.375623 & 0.476625 \\ & 0.213467 & 0.332140 \\ & 0.361282 & 0.242501 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.127653 \\ 0.235260 \\ 0.182697 \end{bmatrix}.$$

Nadalje, kako je element na mestu (2,2) manji od elementa na mestu (3,2), u matrici poslednjeg sistema jednačina izvršimo permutaciju druge i treće vrste. Tada, s obzirom na faktor $m_{32} = 0.590860$, posle drugog eliminacionog koraka dobijamo

$$\begin{bmatrix} 0.215512 & 0.375623 & 0.476625 \\ & 0.361282 & 0.242501 \\ & & 0.188856 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.127653 \\ 0.182697 \\ 0.127312 \end{bmatrix},$$

odakle sleduje $x_3 = 0.674122$, $x_2 = 0.0532050$, $x_1 = -0.991291$.

Napomenimo da su tačna rešenja datog sistema, sa deset značajnih cifara,

$$x_3 = 0.6741214694, \quad x_2 = 0.05320393391, \quad x_1 = -0.9912894252.$$

Na osnovu poslednjeg primera vidimo da strategija izbora glavnog elementa bitno utiče na tačnost rezultata. Modifikacija Gaussovog eliminacionog metoda u ovom smislu, naziva se Gaussov metod sa izborom glavnog elementa. Dakle, prema ovom metodu za glavni element u k-tom eliminacionom koraku uzimamo element $a_{rk}^{(k)}$, za koji je $|a_{rk}^{(k)}| = \max_{k \le i \le n} |a_{ik}^{(k)}|$, uz permutaciju k-te i r-te vrste.

Ako dozvolimo i permutaciju nepoznatih najbolje je za glavni element u k-tom eliminacionom koraku uzeti element $a_{rs}^{(k)}$, za koji je $|a_{rs}^{(k)}| = \max\limits_{k\leq i,j\leq n} |a_{ij}^{(k)}|$, uz permutaciju k-te i r-te vrste i k-te i s-te kolone. Ovakav postupak se naziva metod sa totalnim izborom glavnog elementa.

Odredimo sada broj računskih operacija u Gaussovom metodu pri rešavanju sistema od n jednačina sa n nepoznatih.

Kod redukcije sistema na trougaoni oblik, u prvom eliminacionom koraku, potrebano je $n-1$ deljenja, $n(n-1)$ množenja i $n(n-1)$ oduzimanja, što iznosi $(n-1)(2n+1) = 2(n-1)^2 + 3(n-1)$. Na osnovu ovoga može se zaključiti da je potreban broj računskih operacija u k-tom eliminacionom koraku $2(n-k)^2 + 3(n-k)$, pa je ukupan broj operacija za trougaonu redukciju

$$\sum_{k=1}^{n-1} (2(n-k)^2 + 3(n-k)) = \sum_{k=1}^{n-1}(2k^2+3k) = \frac{1}{6}(4n^3+3n^2-7n).$$

Pri rešavanju sistema sa trougaonom matricom potrebno je n deljenja, $n(n-1)/2$ množenja i $n(n-1)/2$ oduzimanja, što iznosi ukupno n^2 operacija. Dakle, ukupan broj računskih operacija u Gaussovom metodu iznosi

$$N(n) = \frac{1}{6}(4n^3 + 9n^2 - 7n).$$

Za dovoljno veliko n imamo $N(n) \cong 2n^3/3$. Dugo vremena se mislilo da je Gaussov metod najoptimalniji u pogledu broja računskih operacija. U novije vreme V. Strassen ([6]) je, uvodeći iterativni algoritam za množenje i inverziju matrica, dao jedan metod za rešavanje sistema linearnih jednačina, kod koga je broj računskih operacija reda $n^{\log_2 7}$. Strassenov metod je, dakle, optimalniji od Gaussovog metoda($\log_2 7 < 3$).

<u>Primedba 2.2.1</u>. Prema Strassenovom algoritmu množenje matrica drugog reda

$$\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

izvodi se pomoću

$$s_1 = (a_{11}+a_{22})(b_{11}+b_{22}),\ s_2 = (a_{21}+a_{22})b_{11},\ s_3 = a_{22}(b_{12}-b_{11}),$$

$$s_4 = a_{22}(b_{21}-b_{11}),\ s_5 = (a_{11}+a_{12})b_{22},\ s_6 = (a_{21}-a_{11})(b_{11}+b_{12}),$$

$$s_7 = (a_{12}-a_{22})(b_{21}+b_{22})$$

i

$$c_{11} = s_1+s_4-s_5+s_7,\ c_{21} = s_2+s_4,\ c_{12} = s_3+s_5,\ c_{22} = s_1+s_3-s_2+s_6.$$

Trougaona redukcija obezbedjuje lako izračunavanje determinante sistema. Naime, važi

$$\det A = a_{11}^{(1)}a_{22}^{(2)}\dots a_{nn}^{(n)}.$$

Ukoliko je korišćen Gaussov metod sa izborom glavnog elementa treba samo voditi računa o broju permutacija vrsta(i kolona kod metoda sa totalnim izborom glavnog elementa), koje utiču na znak determinante. Ovakav način za izračunavanje determinante je veoma efikasan. Na primer, za izračunavanje determinante reda $n=30$, potrebno je 0.18 s, ako se jedna računska operacija obavlja za 10 μs.

4.2.3. Inverzija matrica pomoću Gaussovog metoda

Neka je $A = [a_{ij}]_{n\times n}$ regularna matica i neka je

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & & x_{2n} \\ \vdots & & & \\ x_{n1} & x_{n2} & & x_{nn} \end{bmatrix} = [\vec{x}_1\ \vec{x}_2\ \dots\ \vec{x}_n]$$

njena inverzna matrica. Vektori $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$ su redom prva, druga,..., n-ta kolona matrice X. Definišimo vektore $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n$ pomoću

$$\vec{e}_1 = [1\ 0\ \dots 0]^T,\ \vec{e}_2 = [0\ 1\ \dots 0]^T,\ \dots,\ \vec{e}_n = [0\ 0\ \dots 1]^T.$$

S obzirom na jednakost $AX = [A\vec{x}_1\ A\vec{x}_2 \dots A\vec{x}_n] = I = [\vec{e}_1\ \vec{e}_2 \dots \vec{e}_n]$, problem odredjivanja inverzne matrice može se svesti na rešavanje n sistema linearnih jednačina

(2.3.1) $\quad A\vec{x}_i = \vec{e}_i \quad (i=1,\ldots,n).$

Za rešavanje sistema (2.3.1) pogodno je koristiti Gauss-ov metod, s obzirom da se matrica A pojavljuje kao matrica svih sistema, pa njenu trougaonu redukciju treba izvršiti samo jednom. Pri ovome sve elementarne transformacije koje su potrebne za trougaonu redukciju matrice A treba primeniti i na jediničnu matricu $I=[\vec{e}_1\ \vec{e}_2\ \ldots\ \vec{e}_n]$. Na taj način se matrica A transformiše u trougaonu matricu R, a matrica I u matricu $C=[\vec{c}_1\ \vec{c}_2\ \ldots\ \vec{c}_n]$. Najzad, ostaje da se reše trogaoni sistemi

$$R\vec{x}_i = \vec{c}_i \quad (i=1,\ldots,n).$$

4.2.4. Faktorizacioni metodi

Faktorizacioni metodi za rešavanje sistema linearnih jednačina zasnivaju se na razlaganju matrice sistema na proizvod dve matrice čiji je oblik takav da omogućava svodjenje sistema na dva sistema jednačina koji se jednostavno sukcesivno rešavaju. U ovom odeljku ukazaćemo na metode zasnovane na LR faktorizaciji matrice(videti odeljak 4.1.2).

Neka je dat sistem jednačina

(2.4.1) $\quad A\vec{x} = \vec{b},$

sa kvadratnom matricom A, čiji su svi glavni dijagonalni minori različiti od nule. Tada, na osnovu teoreme 1.2.1, postoji faktorizacija matrice $A=LR$, gde je L donja i R gornja trougaona matrica. Faktorizacija je jednoznačno odredjena, ako se, na primer, usvoji da matrica L ima jediničnu dijagonalu. U tom slučaju, sistem (2.4.1), tj. sistem $LR\vec{x}=\vec{b}$, se može predstaviti u ekvivalentnom obliku

(2.4.2) $\quad L\vec{y} = \vec{b}, \quad R\vec{x} = \vec{y}.$

Na osnovu prethodnog, za rešavanje sistema jednačina (2.4.1), može se formulisati sledeći metod:

1° Stavimo $\ell_{ii} = 1$ $(i = 1,\ldots,n)$;

2° Odredimo ostale elemente matrice $L = [\ell_{ij}]_{n\times n}$ i matrice $R = [r_{ij}]_{n\times n}$ (videti odeljak 4.1.2);

3° Rešimo prvi sistem jednačina u (2.4.2);

4° Rešimo drugi sistem jednačina u (2.4.2).

Koraci 3° i 4° se jednostavno izvode. Naime, neka su

$$\vec{b} = [b_1\ b_2\ \ldots\ b_n]^T,\ \vec{y} = [y_1\ y_2\ \ldots\ y_n]^T,\ \vec{x} = [x_1\ x_2\ \ldots\ x_n]^T.$$

Tada je

$$y_1 = b_1,\quad y_i = b_i - \sum_{k=1}^{i-1} \ell_{ik} y_k \qquad (i = 2,\ldots,n)$$

i

$$x_n = \frac{y_n}{r_{nn}},\quad x_i = \frac{1}{r_{ii}}\left(y_i - \sum_{k=i+1}^{n} r_{ik} x_k\right) \qquad (i = n-1,\ldots,1).$$

Izloženi metod se u literaturi sreće kao metod Haleckog. U slučaju kada je matrica A normalna, tj. kada je simetrična i pozitivno definitna, metod Haleckog se može uprostiti. Naime, tada se može uzeti da je $L = R^T$. Dakle, treba odrediti faktorizaciju matrice A u obliku $A = R^T R$. Na osnovu formula iz odeljka 4.1.2 za elemente matrice R važe formule

$$r_{11} = \sqrt{a_{11}},\quad r_{1j} = \frac{a_{1j}}{r_{11}} \qquad (j = 2,\ldots,n),$$

$$\left.\begin{array}{l} r_{ii} = \sqrt{a_{ii} - \displaystyle\sum_{k=1}^{i-1} r_{ki}^2} \\ r_{ij} = \dfrac{1}{r_{ii}}\left(a_{ij} - \displaystyle\sum_{k=1}^{i-1} r_{ki} r_{kj}\right) \quad (j = i+1,\ldots,n) \end{array}\right\} (i = 2,\ldots,n).$$

U ovom slučaju sistemi (2.4.2) postaju

$$R^T\vec{y} = \vec{b},\quad R\vec{x} = \vec{y}.$$

Navedena modifikacija metoda Haleckog se naziva metod kvadratnog korena.

Primedba 2.4.1. Determinanta normalne matrice se može izračunati po metodi kvadratnog korena kao

$$\det A = (r_{11}r_{22}\cdots r_{nn})^2 .$$

Faktorizacioni metodi su naročito pogodni za rešavanje sistema linearnih jednačina, kod kojih se matrica sistema ne menja, već samo vektor slobodnih članova $\vec{b}$. Ovakvi sistemi se često javljaju u tehnici.

Sada ćemo pokazati da se Gaussov metod eliminacije može interpretirati kao LR faktorizacija matrice A. Uzmimo matricu A takvu, da prilikom eliminacije ne treba vršiti permutaciju vrsta i kolona. Polazni sistem označimo sa $A^{(1)}\vec{x}=\vec{b}^{(1)}$. Gaussov eliminacioni postupak daje n-1 ekvivalentnih sistema $A^{(2)}\vec{x}=\vec{b}^{(2)}$, $\ldots, A^{(n)}\vec{x}=\vec{b}^{(n)}$, pri čemu matrica $A^{(k)}$ ima oblik

$$A^{(k)}=\begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1k}^{(1)} & \cdots & a_{1n}^{(1)} \\ & a_{22}^{(2)} & & a_{2k}^{(2)} & & a_{2n}^{(2)} \\ & & \ddots & \vdots & & \\ & & & a_{kk}^{(k)} & & a_{kn}^{(k)} \\ & & & \vdots & & \\ & & & a_{nk}^{(k)} & & a_{nn}^{(k)} \end{bmatrix}.$$

Analizirajmo modifikaciju elementa $a_{ij}(=a_{ij}^{(1)})$ u procesu trougaone redukcije. Kako je, za $k=1,2,\ldots,n-1$,

$$a_{ij}^{(k+1)}=a_{ij}^{(k)}-m_{ik}a_{kj}^{(k)} \qquad (i,j=k+1,\ldots,n)$$

i

$$a_{i1}^{(k+1)}=a_{i2}^{(k+1)}=\ldots=a_{ik}^{(k+1)}=0 \qquad (i=k+1,\ldots,n),$$

sumiranjem dobijamo

$$\sum_{k=1}^{i-1} a_{ij}^{(k+1)}=\sum_{k=1}^{i-1} a_{ij}^{(k)}-\sum_{k=1}^{i-1} m_{ik}a_{kj}^{(k)} \qquad (i\leqq j)$$

i

$$\sum_{k=1}^{j} a_{ij}^{(k+1)}=\sum_{k=1}^{j} a_{ij}^{(k)}-\sum_{k=1}^{j} m_{ik}a_{kj}^{(k)} \qquad (i>j),$$

tj.

$$a_{ij} = a_{ij}^{(1)} = a_{ij}^{(i)} + \sum_{k=1}^{i-1} m_{ik} a_{kj}^{(k)} \qquad (i \leq j)$$

i

$$a_{ij} = a_{ij}^{(1)} = 0 + \sum_{k=1}^{j} m_{ik} a_{kj}^{(k)} \qquad (i > j).$$

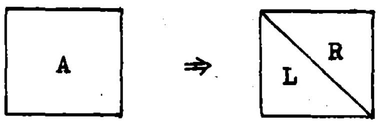
Definišući $m_{ii} = 1 (i = 1,\ldots,n)$, poslednje dve jednakosti se mogu predstaviti u obliku

$$(2.4.3) \quad a_{ij} = \sum_{k=1}^{p} m_{ik} a_{kj}^{(k)} \qquad (i,j = 1,\ldots,n),$$

gde je $p = \min(i,j)$. Jednakost (2.4.3) ukazuje da Gaussova eliminacija daje LR faktorizaciju matrice A, gde su

$$L = \begin{bmatrix} 1 & & & \\ m_{21} & 1 & & \\ & & \ddots & \\ m_{n1} & m_{n2} & \cdots & 1 \end{bmatrix}, \quad R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & r_{2n} \\ & & \ddots & \\ & & & r_{nn} \end{bmatrix}$$

i $r_{kj} = a_{kj}^{(k)}$. Pri programskoj realizaciji Gaussovog metoda u cilju dobijanja LR faktorizacije matrice A, nije potrebno koristiti nove memorijske elemente za pamćenje matrice L, već je pogodno faktore m_{ik} smeštati na mesto koeficijenata matrice A koji se anuliraju u procesu trougaone redukcije. Na taj način, posle završene trougaone redukcije, na mesto matrice A biće memorisane matrice L i R, prema sledećoj šemi



Sl. 2.4.1

Uočimo da se dijagonalni elementi matrice L, koji su svi jednaki jedinici, ne moraju memorisati.

Metod Haleckog, zasnovan na LR faktorizaciji, primenjuje se u slučajevima kada matrica A ispunjava uslove teoreme 1.2.1.

Medjutim, primenljivost ovog metoda može se proširiti i na druge sisteme sa regularnom matricom, uzimajući u obzir permutaciju jednačina u sistemu. Za faktorizaciju iskoristimo Gaussov eliminacioni metod sa izborom glavnog elementa. Pri ovome biće $LR = A'$, gde se matrica A' dobija iz matrice A konačnim brojem razmena vrsta. Ovo znači da u procesu eliminacije treba memorisati niz indeksa glavnih elemenata $I=(p_1,\ldots,p_{n-1})$, pri čemu je p_k broj vrste iz koje se uzima glavni element u k-tom eliminacionom koraku. Kod rešavanja sistema $A\vec{x}=\vec{b}$, neposredno posle faktorizacije treba, u skladu sa nizom indeksa I, permutovati koordinate vektora $\vec{b}$. Na taj način se dobija transformisani vektor $\vec{b}'$, pa se rešavanje datog sistema svodi na sukcesivno rešavanje trougaonih sistema

$$L\vec{y} = \vec{b}' \quad \text{i} \quad R\vec{x} = \vec{y}.$$

Primetimo da se pomoću indeksnog niza I može konstruisati permutaciona matrica P, takva da je $A' = PA$ i $\vec{b}' = P\vec{b}$, što znači da se, u ovom slučaju, radi o faktorizaciji $LR = PA$.

Ilustrujmo ovo jednim primerom.

<u>Primer 2.4.1.</u> Primenom Gaussovog metoda sa izborom glavnog elementa na

$$A = \begin{bmatrix} 3 & 1 & 6 \\ 2 & 1 & 3 \\ 1 & 1 & 1 \end{bmatrix}$$

dobijamo redom

$$\begin{bmatrix} 3 & 1 & 6 \\ 2 & 1 & 3 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 6 \\ 2/3 & 1/3 & -1 \\ 1/3 & 2/3 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 6 \\ 1/3 & 2/3 & -1 \\ 2/3 & 1/2 & -1/2 \end{bmatrix},$$

pri čemu je indeksni niz glavnih elemenata $I = (1,3)$. Glavni elementi su u navedenim matricama zaokruženi. Ovde je

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 2/3 & 1/2 & 1 \end{bmatrix} \quad \text{i} \quad R = \begin{bmatrix} 3 & 1 & 6 \\ 0 & 2/3 & -1 \\ 0 & 0 & -1/2 \end{bmatrix}.$$

Rešimo sada sistem $A\vec{x}=\vec{b}=[2\ 7\ 4]^T$. Kako je $\vec{b}'=[2\ 4\ 7]^T$ iz $L\vec{y}=\vec{b}'$ sleduje

$$y_1=2,\ y_2=4-\frac{1}{3}y_1=\frac{10}{3},\ y_3=7-\frac{2}{3}y_1-\frac{1}{2}y_2=4.$$

Najzad, na osnovu $R\vec{x}=\vec{y}$, nalazimo

$$x_3=-8,\ x_2=\frac{3}{2}(\frac{10}{3}-x_3)=-7,\ x_1=\frac{1}{3}(2-x_1-6x_3)=19.$$

Permutaciona matrica, u ovom slučaju, ima oblik

$$P=\begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 1\\ 0 & 1 & 0\end{bmatrix}.$$

4.2.5. Metod ortogonalizacije

Posmatrajmo sistem (2.2.1) sa regularnom matricom. Ako definišemo (n+1)-dimenzionalne vektore

$$\vec{y}=\begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_n\\ 1\end{bmatrix},\qquad \vec{a}_i=\begin{bmatrix}a_{i1}\\ a_{i2}\\ \vdots\\ a_{in}\\ -b_i\end{bmatrix}\qquad (i=1,\ldots,n),$$

tada se ovaj sistem može predstaviti u obliku

$$(2.5.1)\qquad (\vec{a}_i,\vec{y})=\vec{y}^T\vec{a}_i=0\qquad (i=1,\ldots,n).$$

Jednačine (2.5.1) ukazuju na mogućnost rešavanja sistema (2.2.1), korišćenjem uslova ortogonalnosti vektora $\vec{y}$ sa vektorima $\vec{a}_i\,(i=1,\ldots,n)$. Pomenuta ortogonalnost je ekvivalentna ortogonalnosti vektora $\vec{y}$ sa linearnim potprostorom H_n, koji je generisan vektorima $\vec{a}_i\,(i=1,\ldots,n)$, tj. $H_n=L(\vec{a}_1,\ldots,\vec{a}_n)$. Polazeći od baze $\{\vec{a}_1,\ldots,\vec{a}_n\}$, korišćenjem Gram-Schmidtovog postupka, konstruišimo ortonormiranu bazu $B_n=\{\vec{v}_1,\ldots,\vec{v}_n\}$ potprostora H_n. Vektor $\vec{y}$ je, očigledno, ortogonalan sa svim vektorima ortonormirane baze B_n.

Kako je vektor $\vec{a}_{n+1}=[0\ 0\ \ldots\ 0\ 1]^T$ linearno nezavisan u odnosu na vektore baze B_n, ortogonalizirajmo i ovaj vektor.

Dakle,

$$\vec{u} = \vec{a}_{n+1} - \sum_{i=1}^{n}(\vec{a}_{n+1},\vec{v}_i)\vec{v}_i \ .$$

Neka su koordinate vektora $\vec{u}$ redom $u_1,\ldots,u_n,u_{n+1}$, tj. $\vec{u} = [u_1 \ u_2 \ \ldots \ u_n \ u_{n+1}]^T$. Kako vektor $\vec{u}$ ispunjava uslove ortogonalnosti $(\vec{u},\vec{v}_i) = 0$ $(i = 1,\ldots,n)$, na osnovu prethodnog zaključujemo da je, takodje, $(\vec{u},\vec{a}_i) = 0$ $(i = 1,\ldots,n)$, tj.

$$\begin{aligned}
(\vec{u},\vec{a}_1) &= a_{11}u_1 + a_{12}u_2 + \ldots + a_{1n}u_n - b_1u_{n+1} = 0,\\
(\vec{u},\vec{a}_2) &= a_{21}u_1 + a_{22}u_2 + \ldots + a_{2n}u_n - b_2u_{n+1} = 0,\\
&\vdots\\
(\vec{u},\vec{a}_n) &= a_{n1}u_1 + a_{n2}u_2 + \ldots + a_{nn}u_n - b_nu_{n+1} = 0.
\end{aligned}$$

Primetimo da je $u_{n+1} \neq 0$. Naime, ako bi bilo $u_{n+1} = 0$, tada bi n-torka $(u_1,u_2,\ldots,u_n)$ bila rešenje homogenog sistema jednačina sa matricom $A = [a_{ij}]$. Medjutim, kako homogeni sistem $(\det A \neq 0)$ ima samo trivijalna rešenja, to bismo imali $u_i = 0 (i = 1,\ldots,n)$, pa bi vektor $\vec{a}_{n+1}$ bio linearna kombinacija vektora baze B_n, što je u kontradikciji sa izborom ovog vektora.

Ako sve jednačine poslednjeg sistema podelimo sa u_{n+1}, lako se zaključuje da je vektor $\vec{y}$, sa $x_i = \dfrac{u_i}{u_{n+1}}$ $(i = 1,\ldots,n)$, rešenje sistema jednačina (2.5.1). Tada je n-torka $(x_1,\ldots,x_n)$ rešenje sistema (2.2.1).

Na kraju, primetimo da navedeni metod ortogonalizacije zahteva veći broj operacija množenja i deljenja nego Gaussov metod eliminacije.

4.2.6. Analiza greške i slabouslovljeni sistemi

Posmatrajmo uticaj promene vektora $\vec{b}$ u sistemu jednačina

$$A\vec{x} = \vec{b},$$

sa regularnom matricom A, na rešenje $\vec{x}$. Neka se vektor $\vec{b}$ promeni za $\Delta\vec{b}$. Tada će se rešenje promeniti za $\Delta\vec{x}$, tj. imaćemo

$$A(\vec{x}+\Delta\vec{x}) = \vec{b}+\Delta\vec{b},$$

odakle je $A\Delta\vec{x}=\Delta\vec{b}$, tj. $\Delta\vec{x}=A^{-1}\Delta\vec{b}$. Iz poslednje jednakosti sleduje

(2.6.1) $$\|\Delta\vec{x}\| \leqq \|A^{-1}\|\cdot\|\Delta\vec{b}\| .$$

Kako je $\|\vec{b}\| = \|A\vec{x}\| \leqq \|A\|\cdot\|\vec{x}\|$, imamo

$$\frac{\|\Delta\vec{x}\|}{\|\vec{x}\|} \leqq \|A\|\cdot\|A^{-1}\|\cdot\frac{\|\Delta\vec{b}\|}{\|\vec{b}\|} = k(A)\frac{\|\Delta\vec{b}\|}{\|\vec{b}\|},$$

gde je $k(A)=\|A\|\cdot\|A^{-1}\|$. Broj $k(A)$ naziva se faktorom uslovljenosti matrice A. Faktor uslovljenosti zavisi od upotrebljene norme matrice, ali je uvek $k(A)\geqq 1$, što sleduje iz

$$\|\vec{x}\| = \|I\vec{x}\| = \|AA^{-1}\vec{x}\| = \|A\|\cdot\|A^{-1}\|\cdot\|\vec{x}\|.$$

Što je faktor $k(A)$ veći od jedinice kažemo da je matrica A slabije uslovljena. Sistem sa slabouslovljenom matricom nazivamo slabouslovljenim sistemom*.

Kada se koristi spektralna norma, faktor uslovljenosti je dat sa

$$k(A) = \sigma(A)\sigma(A^{-1}) = \sqrt{\frac{\max\lambda(A^*A)}{\min\lambda(A^*A)}} .$$

Ukoliko je matrica A hermitska, prethodni izraz se pojenostavljuje, tj. postaje

$$k(A) = \frac{\max|\lambda(A)|}{\min|\lambda(A)|} .$$

Zbog svojstva minimalnosti spektralne norme, ova vrednost broja $k(A)$ za hermitsku matricu A je najmanja u odnosu na vrednosti koje se dobijaju korišćenjem drugih normi.

Nejednakost (2.6.1) se može interpretirati i na sledeći način. Neka je $\vec{x}_p$ približno rešenje jednačine $A\vec{x}=\vec{b}$. Sa $\vec{r}(\vec{x}_p)$ označimo odgovarajući "vektor ostatak", tj.

$$\vec{r}(\vec{x}_p) = A\vec{x}_p - \vec{b} = A(\vec{x}_p-\vec{x}).$$

---

*) *U anglo-saksonskoj literaturi: ill-conditioned systems, a u ruskoj:* плохо обусловленные системы.

Dakle, $\vec{x}_p$ je tačno rešenje jednačine $A\vec{x}_p = \vec{b} - \vec{r}(\vec{x}_p)$ i za grešku važi ocena

$$\|\Delta\vec{x}\| \leq \|A^{-1}\| \, \|\vec{r}(\vec{x}_p)\| \,. \tag{2.6.2}$$

Na osnovu poslednje nejednakosti može se zaključiti da $\|\Delta\vec{x}\|$ može biti velika, i u slučajevima kada je veličina $\|\vec{r}(\vec{x}_p)\|$ dovoljno mala.

<u>Teorema 2.6.1.</u> Neka je A regularna matrica reda n, $B = A(I+F)$, $\|F\| < 1$ i vektori $\vec{x}$ i $\Delta\vec{x}$ definisani pomoću $A\vec{x} = \vec{b}$, $B(\vec{x} + \Delta\vec{x}) = \vec{b}$. Tada je

$$1^\circ \quad \frac{\|\Delta\vec{x}\|}{\|\vec{x}\|} \leq \frac{\|F\|}{1 - \|F\|} \;;$$

$$2^\circ \quad \frac{\|\Delta\vec{x}\|}{\|\vec{x}\|} \leq \frac{k(A)}{1 - k(A)\dfrac{\|B-A\|}{\|A\|}} \cdot \frac{\|B-A\|}{\|A\|} \quad \text{ako je } k(A)\frac{\|B-A\|}{\|A\|} < 1.$$

<u>Dokaz.</u> Kako, na osnovu učinjenih pretpostavki, B^{-1} egzistira imamo

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

pa je

$$\frac{\|\Delta\vec{x}\|}{\|\vec{x}\|} \leq \|B^{-1}(A-B)\| = \|-(I+F)^{-1}A^{-1}AF\| \leq \|(I+F)^{-1}\| \cdot \|F\| \leq \frac{\|F\|}{1 - \|F\|} .$$

Imajući u vidu da je $F = A^{-1}(B-A)$ i $\|F\| \leq k(A)\|B-A\|/\|A\|$, na osnovu poslednje nejednakosti dobijamo nejednakost 2°.

Pomoću teoreme 2.6.1 moguće je oceniti grešku u rešenju $\vec{x}$, pri zameni matrice A nekom drugom matricom. Naime, ako stavimo $C = (I+F)^{-1} = B^{-1}A$, $F = A^{-1}B - I$, imamo

$$\|B^{-1}A\| \leq \frac{1}{1 - \|A^{-1}B - I\|} .$$

S druge strane, iz $A^{-1} = A^{-1}BB^{-1}$ sleduje

$$\|A^{-1}\| \leq \|A^{-1}B\| \|B^{-1}\| \leq \frac{\|B^{-1}\|}{1 - \|I - B^{-1}A\|} ,$$

pri čemu smo iskoristili jednakost $A^{-1}B = (B^{-1}A)^{-1} = (I-(I-B^{-1}A))^{-1}$.

Najzad, na osnovu (2.6.2) dobijamo

$$\|\vec{x}_p - \vec{x}\| \leqq \frac{\|B^{-1}\|}{1 - \|I - B^{-1}A\|} \cdot \|\vec{r}(\vec{x}_p)\| \qquad (\|I - B^{-1}A\| < 1).$$

<u>Primer 2.6.1.</u> Jedan tipičan primer slabouslovljenog sistema je sistem(videti [7])

$$\begin{bmatrix} 121734 & 169217 & 176624 & 166662 \\ 169217 & 235222 & 245505 & 231653 \\ 176624 & 245505 & 256423 & 242029 \\ 166662 & 231653 & 242029 & 228474 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 634237 \\ 881597 \\ 920581 \\ 868818 \end{bmatrix},$$

čija su tačna rešenja $x_1 = x_2 = x_3 = x_4 = 1$. Medjutim, ako koordinate vektora slobodnih članova variraju samo za ± 1 i recimo budu

$$b_1 = 634238, \; b_2 = 881596, \; b_3 = 920580, \; b_4 = 868819,$$

rešenja sistema postaju

$$x_1 = 130214370, \; x_2 = 78645876, \; x_3 = -32701403, \; x_4 = 19395881.$$

Iz datog primera se može videti da je problem rešavanja slabouslovljenih sistema veoma složen i njemu treba pristupati obazrivo.

4.3. ITERATIVNI METODI U LINEARNOJ ALGEBRI

4.3.1. Načini formiranja iterativnih metoda za rešavanje sistema linearnih jednačina

Posmatrajmo sistem linearnih jednačina

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1,\\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2,\\
&\vdots\\
a_{n1}x_1 + a_{n1}x_2 + \ldots + a_{nn}x_n &= b_n,
\end{aligned}
\tag{3.1.1}
$$

koji se može predstaviti i u matričnom obliku

$$A\vec{x} = \vec{b}, \tag{3.1.2}$$

gde su

$$
A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & & a_{nn} \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.
$$

Uvek u ovom poglavlju, pretpostavljamo da sistem (3.1.1), tj. (3.1.2) ima jedinstveno rešenje.

Iterativni metodi za rešavanje sistema (3.1.2) imaju za cilj odredjivanje rešenja $\vec{x}$ sa unapred zadatom tačnošću. Naime, polazeći od proizvoljnog vektora $\vec{x}^{(o)}(=[x_1^{(o)}\ldots x_n^{(o)}]^T)$*, iterativnim metodom se odredjuje niz $\{\vec{x}^{(k)}\}$ $(\vec{x}^{(k)}=[x_1^{(k)}\ldots x_n^{(k)}]^T)$ takav da je

$$\lim_{k\to+\infty} \vec{x}^{(k)} = \vec{x}.$$

*) Često se za $\vec{x}^{(o)}$ uzima rešenje dobijeno nekim od direktnih metoda.

Jedan opšti iterativni metod može se predstaviti u obliku

$$\vec{x}^{(k)} = F_k(\vec{x}^{(o)}, \vec{x}^{(1)}, \ldots, \vec{x}^{(k-1)}) \quad (k=1,2,\ldots),$$

gde funkcija F_k u opštem slučaju zavisi od A, $\vec{b}$, k.

Sa stanovišta primene, najinteresantniji su iterativni metodi oblika

$$\vec{x}^{(k)} = F_k(\vec{x}^{(k-1)}) \quad (k=1,2,\ldots). \tag{3.1.3}$$

Ako F_k ne zavisi od k, metod (3.1.3) je stacionaran.

Mi ćemo razmatrati slučajeve kada je F_k linearna funkcija po $\vec{x}$, tj.

$$F_k(\vec{x}) = B_k\vec{x} + \vec{c}_k, \tag{3.1.4}$$

gde je B_k kvadratna matrica i $\vec{c}_k$ vektor.

Da bi iterativni metod, definisan funkcijom (3.1.4) bio konvergentan, potreban uslov je da F_k ima nepokretnu tačku $\vec{x} = A^{-1}\vec{b}$, tj. da je

$$A^{-1}\vec{b} = B_kA^{-1}\vec{b} + \vec{c}_k. \tag{3.1.5}$$

Iz (3.1.5) sleduje

$$\vec{c}_k = (I-B_k)A^{-1}\vec{b} = C_k\vec{b},$$

gde smo stavili $C_k = (I-B_k)A^{-1}$, tj. $C_kA + B_k = I$.

Na taj način (3.1.4) postaje

$$F_k(\vec{x}) = B_k\vec{x} + C_k\vec{b}, \tag{3.1.6}$$

ili

$$F_k(\vec{x}) = \vec{x} - C_k(A\vec{x}-\vec{b}). \tag{3.1.7}$$

Često se za C_k uzima dijagonalna matrica sa jednakim elementima na dijagonali, tj. $C_k = \mathrm{diag}(c_k,\ldots,c_k) = c_kI$ $(c_k \in R)$. Tada se (3.1.7) svodi na

$$F_k(\vec{x}) = \vec{x} - c_k(A\vec{x} - \vec{b}). \tag{3.1.8}$$

Ako se matrica A predstavi u obliku

$$A = D_k + E_k,$$

gde je D_k regularna matrica, F_k se može zadati u implicitnom obliku kao

(3.1.9) $$D_k F_k(\vec{x}) + E_k\vec{x} = \vec{b}.$$

U praktičnim primenama, za D_k se najčešće uzima dijagonalna ili trougaona matrica.

Formule (3.1.6), (3.1.7), (3.1.8), (3.1.9) se koriste za formiranje različitih iterativnih metoda za rešavanje sistema (3.1.1).

Na kraju, napomenimo da se metodom najmanjih kvadrata može dobiti niz iterativnih procesa. Ovaj metod zasniva se na minimizaciji funkcije f definisane pomoću

$$f(\vec{x}) = \|A\vec{x} - \vec{b}\|^2 .$$

U slučaju da je A realna normalna matrica, za rešavanje sistema (3.1.1) može se koristiti minimizacija funkcije odredjene sa

$$f(\vec{x}) = (A\vec{x},\vec{x}) - 2(\vec{b},\vec{x}).$$

4.3.2. Metod proste iteracije

Jedan od najprostijih stacionarnih metoda za rešavanje sistema linearnih jednačina, tzv. metod proste iteracije, zasnovan je na primeni funkcije date pomoću (3.1.6), tj.

(3.2.1) $$F(\vec{x}) = B\vec{x} + C\vec{b}.$$

Ako stavimo $C\vec{b} = \vec{\beta}$, iz (3.2.1) sleduje

(3.2.2) $$\vec{x}^{(k)} = B\vec{x}^{(k-1)} + \vec{\beta} \quad (k=1,2,\ldots).$$

Napomenimo, da je jednačina

(3.2.3) $$\vec{x} = B\vec{x} + \vec{\beta}$$

ekvivalentna sa

(3.2.4) $$A\vec{x} = \vec{b}.$$

Matrica B se naziva iterativna matrica.

Ako se podje od proizvoljnog vektora $\vec{x}^{(o)}$, pomoću (3.2.2) generiše se niz $\{\vec{x}^{(k)}\}$. Razmatraćemo uslove pod kojima ovako generisani niz konvergira ka tačnom rešenju sistema (3.2.3), tj. (3.2.4).

Ako je

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & & b_{2n} \\ \vdots & & & \\ b_{n1} & b_{n2} & & b_{nn} \end{bmatrix} \quad i \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix},$$

iterativni metod (3.2.2) može se predstaviti skalarno

$$\begin{aligned} x_1^{(k)} &= b_{11}x_1^{(k-1)} + b_{12}x_2^{(k-1)} + \ldots + b_{1n}x_n^{(k-1)} + \beta_1, \\ x_2^{(k)} &= b_{21}x_1^{(k-1)} + b_{22}x_2^{(k-1)} + \ldots + b_{2n}x_n^{(k-1)} + \beta_2, \\ &\vdots \\ x_n^{(k)} &= b_{n1}x_1^{(k-1)} + b_{n2}x_2^{(k-1)} + \ldots + b_{nn}x_n^{(k-1)} + \beta_n, \end{aligned}$$

gde je $k=1,2,\ldots$.

Dokazaćemo sledeće dve teoreme.

Teorema 3.2.1. Ako je $\vec{x}^{(o)}$ proizvoljan vektor, $\|B\| < 1$ je dovoljan uslov za konvergenciju procesa (3.2.2) ka tačnom rešenju $\vec{x}$ sistema (3.2.3).

Teorema 3.2.2. Ako su ispunjeni uslovi teoreme 3.2.1, za svako $k \in N$, važe nejednakosti

(3.2.5) $$\|(I-B)^{-1} - (I+B+\ldots+B^{k-1})\| \leqq \frac{\|B\|^k}{1-\|B\|};$$

(3.2.6) $$\|\vec{x}^{(k)} - \vec{x}\| \leqq \|B\|^k \|\vec{x}^{(o)}\| + \frac{\|\vec{\beta}\|\cdot\|B\|^k}{1-\|B\|};$$

(3.2.7) $$\|\vec{x}^{(k)} - \vec{x}\| \leqq \|B\|\cdot\|\vec{x}^{(k-1)} - \vec{x}\|;$$

(3.2.8) $$\|\vec{x}^{(k)} - \vec{x}\| \leqq \|B\|^k \|\vec{x}^{(o)} - \vec{x}\|.$$

Dokaz teoreme 3.2.1. Na osnovu (3.2.2) lako se dokazuje jednakost

(3.2.9) $$\vec{x}^{(k)} = B^k\vec{x}^{(o)} + (I+B+\ldots+B^{k-1})\vec{\beta} \qquad (k \in N).$$

Kako je $\| B\| < 1$, imamo

$$\lim_{k\to+\infty} \| B^k\| = 0, \text{ tj. } \lim_{k\to+\infty} B^k = 0$$

i

$$\lim_{k\to+\infty} (I+B+\ldots+B^{k-1}) = (I-B)^{-1}.$$

Tada, iz (3.2.9) sleduje

$$\lim_{k\to+\infty} \vec{x}^{(k)} = \lim_{k\to+\infty} B^k\vec{x}^{(o)} + \lim_{k\to+\infty} (I+B+\ldots+B^{k-1})\vec{\beta} = (I-B)^{-1}\vec{\beta},$$

odakle zaključujemo da niz $\{\vec{x}^{(k)}\}$ konvergira ka rešenju jednačine (3.2.3).

<u>Dokaz teoreme 3.2.2.</u> Iz jednakosti

$$(I-B)^{-1} = I + B + B^2 + \ldots \quad (\| B\| < 1)$$

sleduje

$$(I-B)^{-1}-(I+B+\ldots+B^{k-1}) = B^k + B^{k+1} + \ldots ,$$

odnosno

$$\begin{aligned}\| (I-B)^{-1}-(I+B+\ldots+B^{k-1})\| &\leq \| B\|^k + \| B\|^{k+1} + \ldots \\ &= \frac{\| B\|^k}{1-\| B\|}.\end{aligned}$$

Na osnovu (3.2.9), imamo

$$\begin{aligned}\vec{x}^{(k)}-\vec{x} &= B^k\vec{x}^{(o)} + (I+B+\ldots+B^{k-1})\vec{\beta} - (I-B)^{-1}\vec{\beta} \\ &= B^k\vec{x}^{(o)} - ((I-B)^{-1} - (I+B+\ldots+B^{k-1}))\vec{\beta},\end{aligned}$$

odakle je

$$\begin{aligned}\| \vec{x}^{(k)}-\vec{x}\| &\leq \| B^k\vec{x}^{(o)}\| + \|(I-B)^{-1}-(I+B+\ldots+B^{k-1})\|\cdot \|\vec{\beta}\| \\ &\leq \| B\|^k \|\vec{x}^{(o)}\| + \frac{\|\vec{\beta}\|\cdot\| B\|^k}{1-\| B\|},\end{aligned}$$

pri čemu smo iskoristili nejednakost (3.2.5).

Napomenimo da je korišćena norma matrice saglasna sa izabranom normom vektora.

Kako je $\vec{x} = B\vec{x} + \vec{\beta}$ i $\vec{x}^{(k)} = B\vec{x}^{(k-1)} + \vec{\beta}$, dobijamo

$$\vec{x}^{(k)} - \vec{x} = B(\vec{x}^{(k-1)} - \vec{x}),$$

odnosno

$$\| \vec{x}^{(k)} - \vec{x}\| \leqq \| B\| \cdot \| \vec{x}^{(k-1)} - \vec{x}\| .$$

Konačno, iteriranjem poslednje nejednakosti, dobijamo

$$\| \vec{x}^{(k)} - \vec{x}\| \leqq \| B\|^{k} \| \vec{x}^{(o)} - \vec{x}\| .$$

Ovim je teorema 3.2.2 dokazana.

Nejednakosti (3.2.7) i (3.2.8) ukazuju da iterativni proces (3.2.2) ima red konvergencije 1 (linearna konvergencija ili konvergencija tipa geometrijske progresije).

Primedba 3.2.1. Najčešće se uzima $\vec{x}^{(o)} = \vec{\beta}$. Tada se nejednakost (3.2.6), iz teoreme 3.2.1, može pooštriti. Naime, važi nejednakost

$$\| \vec{x}^{(k)} - \vec{x}\| = \| ((I-B)^{-1} - (I+B+\ldots+B^{k}))\vec{\beta}\|$$

$$\leqq \frac{\| B\|^{k+1} \| \vec{\beta}\|}{1 - \| B\|} .$$

Posledica 3.2.1. Iterativni proces (3.2.2) konvergira ako je

(a) $$\| B\|_{\infty} = \max_{i} \sum_{j=1}^{n} |b_{ij}| < 1$$

ili

(b) $$\| B\|_{1} = \max_{j} \sum_{i=1}^{n} |b_{ij}| < 1$$

ili

(c) $$\| B \|_{2} = \varepsilon(B) = \Big(\sum_{i,j} |b_{ij}|^{2}\Big)^{1/2} < 1.$$

Primetimo da su uslovi (a), (b), (c) ispunjeni ako je $|b_{ij}| < \frac{1}{n}$ $(i,j=1,\ldots,n)$.

Primedba 3.2.2. Ako je $\| B\| < 1$, važi

$$\| \vec{x}^{(k)} - \vec{x}\| \leqq \frac{\| B\|}{1 - \| B\|} \| \vec{x}^{(k)} - \vec{x}^{(k-1)}\| \qquad (k \in N).$$

Ova nejednakost je posledica jedne opštije nejednakosti, koja će biti dokazana u sledećem odeljku.

Za iterativni proces (3.2.2) u jednom specijalnom slučaju, Krasnosel'ski i i Krein([8]) odredili su ocenu za vektor greške $\vec{\varepsilon}^{(k)} = \vec{x}^{(k)} - \vec{x}$, pri prekidu iterativnog procesa kada je $\| \vec{x}^{(k+1)} - \vec{x}^{(k)} \|_E < \alpha (\alpha > 0)$ i $\| \vec{x}^{(p+1)} - \vec{x}^{(p)} \|_E \geqq \alpha$ $(p < k)$. Naime, oni su dokazali sledeću teoremu:

Teorema 3.2.3. Neka je B realna normalna matrica, čije sopstvene vrednosti ispunjavaju uslov $1 > \lambda_1 > \lambda_2 > \ldots > \lambda_n > 0$. Ako je početna greška $\vec{\varepsilon}^{(o)} = \vec{x}^{(o)} - \vec{x}$ uniformno rasporedjena u zatvorenoj kugli $\overline{S}(o,r) = \{\vec{x} \,|\, \| x \|_E \leqq r,\ \vec{x} \in R^n\}$, tada verovatnoća, da greška $\vec{\varepsilon}^{(k)}$ zadovoljava nejednakosti

$$\frac{\Theta \lambda_1 \alpha}{1-\lambda_1} \leqq \| \vec{\varepsilon}^{(k)} \|_E \leqq \frac{\alpha}{1-\lambda_1},$$

teži jedinici kada $r \to +\infty$, za svako $\Theta < 1$, tj.

$$\lim_{r \to +\infty} P\left(\frac{\Theta \lambda_1 \alpha}{1-\lambda_1} \leqq \| \vec{\varepsilon}^{(k)} \|_E \leqq \frac{\alpha}{1-\lambda_1}\right) = 1.$$

Ovaj rezultat, Peradze ([9]) je generalisao, pretpostavljajući samo uslov da je matrica B realna i regularna. Konačno, T. Yamamoto u [10] i [11] izbacio je i uslov regularnosti matrice B.

Neka je B realna matrica reda n, čije različite sopstvene vrednosti $\lambda_i (i=1,\ldots,m)$, reda višestrukosti $n_i \left(\sum_{i=1}^{m} n_i = n\right)$, ispunjavaju uslov

$$1 > \rho(B) = |\lambda_1| = |\lambda_2| = \ldots = |\lambda_s| > |\lambda_{s+1}| \geqq \ldots \geqq |\lambda_m|.$$

Navešćemo dva partikularna rezultata, do kojih je došao T. Yamamoto([10]).

Teorema 3.2.4. Neka je A realna normalna matrica i

$$\lambda = \min_{1 \leqq i \leqq s} \left|\frac{\lambda_i}{1-\lambda_i}\right|, \quad \Lambda = \max_{1 \leqq i \leqq s} \left|\frac{1}{1-\lambda_i}\right|, \quad \Lambda^* = \max_{\lambda_i \neq o} \left|\frac{1}{1-\lambda_i}\right| .$$

Ako je početna greška $\vec{\varepsilon}^{(o)}$ uniformno rasporedjena u $\bar{S}(o,r)$, tada je

$$(3.2.10)\qquad \lim_{r\to+\infty} P(\Theta_1\lambda\alpha < \| \vec{\varepsilon}^{(k)}\|_E < \Theta_2\Lambda\alpha) = 1$$

i

$$(3.2.11)\qquad \lim_{r\to+\infty} P(\Theta_1\lambda\alpha < \| \vec{\varepsilon}^{(k)}\|_E < \Lambda^*\alpha) = 1$$

za svako $\Theta_1\varepsilon(o,1)$ i svako $\Theta_2 > 1$.

<u>Teorema 3.2.5.</u> Neka je A proizvoljna realna matrica takva da je s=1 (tj. $|\lambda_1|>|\lambda_2|$). $\lambda_1 \in R$, $n_1=1$. Ako je $\vec{\varepsilon}^{(o)}$ uniformno rasporedjena u $\bar{S}(o,r)$, tada je

$$\lim_{r\to+\infty} P\left(\frac{\Theta_1|\lambda_1|}{1-\lambda_1}\alpha < \| \vec{\varepsilon}^{(k)}\|_E < \frac{\Theta_2}{1-\lambda_1}\alpha\right) = 1$$

za svako $\Theta_1\varepsilon(0,1)$ i svako $\Theta_2 > 1$.

Primetimo da je ocena (3.2.10) oštrija od (3.2.11) kada je $\Lambda < \Lambda^*$, a obrnut je slučaj kada je $\Lambda = \Lambda^*$ (inače uvek je $\Lambda \leq \Lambda^*$).

<u>Primer 3.2.1.</u> Neka je

$$B = \begin{bmatrix} -0.1 & 0.4 & 0.1 & 0.1 \\ 0.4 & -0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & -0.2 & 0.2 \\ 0.1 & 0.1 & 0.2 & -0.2 \end{bmatrix} \quad \text{i} \quad \vec{\beta} = \begin{bmatrix} 0.7 \\ 0.7 \\ 1.2 \\ -1.6 \end{bmatrix}.$$

Tada je $\vec{x} = [1\ 1\ 1\ 1]^T$, $\lambda_1 = -0.5$, $\lambda_2 = 0.4$, $\lambda_3 = -0.4$, $\lambda_4 = -0.1$ i $\lambda = \frac{1}{3}$, $\Lambda = \frac{2}{3}$, $\Lambda^* = \frac{5}{3}$.

Ako se uniformni slučajni vektor $\vec{\varepsilon}^{(o)}$ u $\bar{S}(o,r)$ generiše pomoću metoda kongruentnosti (videti [12],[10]) i iterativni proces (3.2.2) startuje sa $\vec{x}^{(o)}=\vec{x}-\vec{\varepsilon}^{(o)}$, sa sto pokušaja za fiksirano r i $\alpha = 10^{-3}$, dobijeni su sledeći rezultati koji su sredjeni u Tabeli 3.2.1. Radi jednostavnosti, uzet je granični slučaj $\Theta_1 = \Theta_2 = 1$*.

*) Sa stanovišta praktičnog računanja dopušten je izbor $\Theta_1=\Theta_2=1$, jer ako je Θ_i dovoljno blisko jedinici, tada je odgovarajući mašinski broj $fl(\Theta_i)=1$.

Tabela 3.2.1.

r	$\|\vec{\varepsilon}^{(k)}\|_E < \lambda\alpha$	$\lambda\alpha \leq \|\vec{\varepsilon}^{(k)}\|_E < \Lambda\alpha$	$\|\vec{\varepsilon}^{(k)}\|_E > \Lambda\alpha$
1	0 (puta)	93 (puta)	7 (puta)
5	0	97	3
10	0	100	0
20	0	99	1

Do sada smo razmatrali dovoljne uslove za konvergenciju iterativnog procesa (3.2.2). Sledeća teorema daje potrebne i dovoljne uslove.

Teorema 3.2.6. Neka je $\vec{x}^{(o)}$ proizvoljan početni vektor. Potreban i dovoljan uslov za konvergenciju iterativnog procesa (3.2.2) je da su sve sopstvene vrednosti matrice B po modulu manje od jedinice.

Dokaz. Kako za iterativni proces (3.2.2) važi jednakost (3.2.9), tj.

(3.2.12) $$\vec{x}^{(k)} = B^k\vec{x}^{(o)} + (I+B+\ldots+B^{k-1})\vec{\beta} \qquad (k=1,2,\ldots),$$

zaključujemo da je proces (3.2.2) ekvikonvergentan sa matričnim redom

(3.2.13) $$I + B + B^2 + \ldots = \sum_{m=o}^{+\infty} B^m.$$

S druge strane, kako su potrebni i dovoljni uslovi za konvergenciju reda (3.2.13) (videti odeljak 4.1.7)

(3.2.14) $$|\lambda_i(B)| < 1 \qquad (i=1,\ldots,n),$$

dokaz teoreme 3.2.6 je završen.

Primedba 3.2.3. Pod uslovima (3.2.14) imamo

$$\lim_{k\to+\infty} B^k = 0 \quad \text{i} \quad \lim_{k\to+\infty}(I+B+\ldots+B^{k-1}) = (I-B)^{-1}.$$

Tada iz (3.2.12) sleduje

$$\lim_{k\to+\infty} \vec{x}^{(k)} = \lim_{k\to+\infty} \left[B^k \vec{x}^{(o)} + (I+B+\ldots+B^{k-1})\vec{\beta}\right]$$

$$= \lim_{k\to+\infty} B^k \vec{x}^{(o)} + \lim_{k\to+\infty} (I+B+\ldots+B^{k-1})\vec{\beta}$$

$$= (I-B)^{-1} \vec{\beta},$$

što predstavlja tačno rešenje jednačine (3.2.3).

Dakle, iz teoreme 3.2.6 sleduje da iterativni proces (3.2.2) konvergira ako i samo ako su sve nule polinoma

$$\lambda \mapsto \det(B-\lambda I) = \begin{vmatrix} b_{11}-\lambda & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22}-\lambda & & b_{2n} \\ \vdots & & & \\ b_{n1} & b_{n2} & & b_{nn}-\lambda \end{vmatrix}$$

po modulu manje od jedinice.

Značaj uslova (3.2.14) u teorijskim razmatranjima je vrlo veliki. Medjutim, za praktičnu primenu oni nisu pogodni, jer je problem nalaženja sopstvenih vrednosti matrice dosta težak. Uslov $\|B\| < 1$, u teoremi 3.2.1, je sa stanovišta praktične primene vrlo pogodan za ispitivanje konvergencije. Nažalost, ovaj uslov je samo dovoljan, ali ne i potreban.

Primer 3.2.2. Posmatrajmo metod proste iteracije

$$\begin{aligned} x^{(k)} &= \frac{1}{3}x^{(k-1)} - \frac{1}{9}y^{(k-1)} + \frac{1}{9}, \\ y^{(k)} &= 2x^{(k-1)} + \frac{1}{3}y^{(k-1)} - \frac{4}{3}, \end{aligned} \tag{3.2.15}$$

gde je $k=1,2,\ldots$.

Na osnovu posledice 3.2.1, ništa ne možemo zaključiti o konvergenciji procesa (3.2.15), jer nijedan od uslova (a), (b), (c) za matricu

$$B = \begin{bmatrix} 1/3 & -1/9 \\ 2 & 1/3 \end{bmatrix}$$

nije ispunjen. Naime,

$$\|B\|_{\infty} = \max\left(\frac{1}{3}+\frac{1}{9},\ 2+\frac{1}{3}\right) = \frac{7}{3} > 1,$$

$$\|B\|_{1} = \max\left(\frac{1}{3}+2,\ \frac{1}{9}+\frac{1}{3}\right) = \frac{7}{3} > 1,$$

$$\|B\|_{2} = \left(\frac{1}{9}+\frac{1}{81}+4+\frac{1}{9}\right)^{1/2} = 2.057\ldots > 1.$$

Medjutim, sopstvene vrednosti matrice B su $\lambda_{1,2} = \frac{1}{3}(1\pm i\sqrt{2})$. Kako je $|\lambda_1| = |\lambda_2| = \frac{\sqrt{3}}{3} < 1$, iterativni proces (3.2.15) je konvergentan za proizvoljne vrednosti $x^{(o)}$ i $y^{(o)}$.

Uslovi (3.2.14) mogu se zameniti uslovom

$$\rho(B) < 1,$$

gde je $\rho(B)$ spektralni radijus matrice B. Za iterativnu matricu B se u ovom slučaju kaže da je konvergentna (videti [13]).

Kao kriterijum za brzinu konvergencije iterativnog procesa uzima se veličina $\rho(B)$. Naime, iterativni proces konvergira brže, ukoliko je $\rho(B)$ bliže nuli. Broj koji karakteriše brzinu konvergencije definisan je kao

$$\nu(B) = -\log \rho(B).$$

Pri ovome, za ispunjenje uslova $\|\vec{x}^{(k)}-\vec{x}\| < \varepsilon \|\vec{x}^{(o)}-\vec{x}\|$, gde je ε dovoljno mali pozitivan broj i $\rho(B) \neq 0$, potreban broj iteracija je približno dat sa

$$k \sim -\frac{\log\varepsilon}{\nu(B)}.$$

Pored brzine konvergencije jednog iterativnog procesa vrlo je bitan i broj aritmetičkih i logičkih operacija neophodnih za obavljanje jednog iterativnog koraka (iteracije). Ovaj broj često se naziva cena iteracije i označava sa C(B), gde je B odgovarajuća iterativna matrica.

Tako je ukupan broj operacija za postizanje tačnosti ε (u prethodno navedenom smislu) približno dat sa

$$N(B;\varepsilon) = k\,C(B) \sim -\frac{C(B)\log\varepsilon}{\nu(B)}.$$

Iterativni proces je efektivniji, ukoliko je ovaj broj veći.

Pokazaćemo sada kako se može dobiti jedan tzv. optimalni iterativni proces za rešavanje jednačine $A\vec{x} = \vec{b}$ u slučaju kada je A hermitska pozitivno definitna matrica. Ovaj metod je baziran na primeni formule (3.1.8).

Ovde je

$$(3.2.16)\qquad 0 < m \leqq \lambda_i(A) \leqq M < +\infty \quad (i=1,\dots,n).$$

S obzirom da su sopstvene vrednosti $\lambda_i(A)$ u malom broju slučajeva poznate, za m i M u (3.2.16) mogu se uzeti neke ocene za $\lambda_{min}(A)$ i $\lambda_{max}(A)$ respektivno.

Posmatrajmo iterativni proces

$$(3.2.17)\qquad \vec{x}^{(k)} = \vec{x}^{(k-1)} - c(A\,\vec{x}^{(k-1)} - \vec{b}),$$

tj.

$$\vec{x}^{(k)} = (I-cA)\vec{x}^{(k-1)} + c\vec{b},$$

gde je c realan broj.

Kako su sopstvene vrednosti matrice $B=I-cA$,

$$\lambda_i(B) = 1 - c\lambda_i(A),$$

spektralni radijus matrice B je

$$(3.2.18)\qquad \rho(B) = \max\,\left[1 - c\lambda_i(A)\right].$$

Za iterativni proces (3.2.17) kaže se da je optimalan, ako se parametar c odredi tako da $\rho(B)$ ima minimalnu vrednost.

S obzirom na (3.2.16), ako se uvede smena

$$\mu = \frac{2}{M-m}\lambda_i(A) - \frac{M+m}{M-m},$$

(3.2.18) se svodi na

$$\rho(B) = \frac{1}{2}|c|(M-m)\max_{-1\leqq\mu\leqq 1}\left[\mu - \frac{2-c(M+m)}{c(M-m)}\right].$$

Minimalna vrednost za $\rho(B)$ se dobija ako je $c = \frac{2}{M+m}$ i ona iznosi

$$\min_{c} \rho(B) = \frac{M-m}{M+m} .$$

Dakle, dobili smo optimalni iterativni proces

$$\vec{x}^{(k)} = \vec{x}^{(k-1)} - \frac{2}{M+m} (A\, \vec{x}^{(k-1)} - \vec{b}). \tag{3.2.19}$$

<u>Teorema 3.2.7</u>. Neka je A hermitska pozitivno definitna matrica čije sopstvene vrednosti zadovoljavaju uslov (3.2.16).

Za optimalni iterativni proces (3.2.19) važi ocena greške

$$\| \vec{x}^{(k)} - \vec{x} \|_E \leqq \left(\frac{M-m}{M+m}\right)^k \|\vec{x}^{(o)} - \vec{x}\|_E \qquad (k \in N), \tag{3.2.20}$$

gde je $\vec{x}$ tačno rešenje sistema $A\,\vec{x} = \vec{b}$.

<u>Dokaz</u>. S obzirom na učinjene pretpostavke za matricu A, matrica $B = I - \frac{2}{M+m} A$ je hermitska. Tada je, na osnovu Teoreme 1.6.2,

$$\sigma(B) = \rho(B) = \frac{M-m}{M+m} .$$

Uzimajući Euklidsku normu za normu vektora i spektralnu normu za normu matrica, nejednakost (3.2.8), iz teoreme 3.2.2, se svodi na (3.2.20).

O nekim opštijim iterativnim procesima tipa (3.2.19) može se naći u [14, str. 352-363].

Na kraju ovog odeljka, izložićemo jedan praktičan način za formiranje metoda proste iteracije.

Neka je dat sistem jednačina $A\vec{x} = \vec{b}$, gde je $A = [a_{ij}]_{n \times n}$ i $\vec{b} = [b_1 \ldots b_n]^T$, i neka je

$$D = \mathrm{diag}(A) = \begin{bmatrix} a_{11} & 0 & \ldots & 0 \\ 0 & a_{22} & & 0 \\ \vdots & & & \\ 0 & 0 & & a_{nn} \end{bmatrix}$$

regularna matrica. Tada se ovaj sistem može predstaviti u ekvivalentnom obliku

(3.2.21) $\vec{x} = D^{-1}(D-A)\vec{x} + D^{-1}\vec{b}.$

Napomenimo da je odgovarajući skalarni oblik

$$x_1 = -\frac{a_{12}}{a_{11}}x_2 - \frac{a_{13}}{a_{11}}x_3 - \dots - \frac{a_{1n}}{a_{11}}x_n + \frac{b_1}{a_{11}},$$

$$x_2 = -\frac{a_{21}}{a_{22}}x_1 - \frac{a_{23}}{a_{22}}x_3 - \dots - \frac{a_{2n}}{a_{22}}x_n + \frac{b_2}{a_{22}},$$

$$\vdots$$

$$x_n = -\frac{a_{n1}}{a_{nn}}x_1 - \frac{a_{n2}}{a_{nn}}x_2 - \dots - \frac{a_{n,n-1}}{a_{nn}}x_{n-1} + \frac{b_n}{a_{nn}}.$$

Na osnovu (3.2.21), može se formirati metod proste iteracije

(3.2.21) $\vec{x}^{(k)} = D^{-1}(D-A)\vec{x}^{(k-1)} + D^{-1}\vec{b} \qquad (k=1,2,\dots),$

koji je u literaturi poznat kao Jacobiev metod.

Kako je karakteristični polinom matrice $D^{-1}(D-A)$ dat pomoću

$$\begin{aligned} P(\lambda) &= \det\left[D^{-1}(D-A)-\lambda I\right] \\ &= -\det D^{-1}\cdot\det(\lambda D + (A-D)), \end{aligned}$$

iz teoreme 3.2.6, sleduje da Jacobiev iterativni proces konvergira ako i samo ako su svi koreni jednačine

$$\begin{vmatrix} \lambda a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & \lambda a_{22} & & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & & \lambda a_{nn} \end{vmatrix} = 0$$

po modulu manji od jedinice.

Primetimo, na kraju, da ovaj metod pripada klasi metoda koji se dobijaju na bazi primene formule (3.1.9).

4.3.3. Gauss-Seidelov metod

Gauss-Seidelov metod se dobija modifikacijom metoda proste iteracije. Kao što smo ranije videli, kod metoda proste iteracije, vrednost i-te komponente $x_i^{(k)}$ vektora $\vec{x}^{(k)}$ izračunava se na osnovu vrednosti $x_1^{(k-1)},\ldots,x_n^{(k-1)}$, tj.

$$x_i^{(k)} = \sum_{j=1}^{n} b_{ij}x_j^{(k-1)} + \beta_i \qquad (i=1,\ldots,n;\ k=1,2,\ldots).$$

Ovaj metod može se modifikovati na taj način što bi se za izračunavanje vrednosti $x_i^{(k)}$ koristile vrednosti $x_1^{(k)},\ldots,x_{i-1}^{(k)}$, $x_i^{(k-1)},\ldots,x_n^{(k-1)}$, tj.

$$(3.3.1) \qquad x_i^{(k)} = \sum_{j=1}^{i-1} b_{ij}x_j^{(k)} + \sum_{j=i}^{n} b_{ij}x_j^{(k-1)} + \vec{\beta}_i \qquad (i=1,\ldots,n;\ k=1,2,\ldots).$$

Navedena modifikacija metoda proste iteracije poznata je kao Gauss-Seidelov metod.

Iterativni proces (3.3.1) može se predstaviti i u matričnoj formi. Naime, neka je

$$B = B_1 + B_2,$$

gde su

$$B_1 = \begin{bmatrix} 0 & 0 & \ldots & 0 & 0 \\ b_{21} & 0 & & 0 & 0 \\ \vdots & & & & \\ b_{n1} & b_{n2} & & b_{n,n-1} & 0 \end{bmatrix} \quad \text{i} \quad B_2 = \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ 0 & b_{22} & & b_{2n} \\ \vdots & & & \\ 0 & 0 & & b_{nn} \end{bmatrix}.$$

Tada (3.3.1) postaje

$$(3.3.2) \qquad \vec{x}^{(k)} = B_1\vec{x}^{(k)} + B_2\vec{x}^{(k-1)} + \vec{\beta} \qquad (k=1,2,\ldots).$$

<u>Teorema 3.3.1</u>. Pri proizvoljnom vektoru $\vec{x}^{(o)}$, iterativni proces (3.3.2) konvergira ako i samo ako su svi koreni jednačine

$$\text{(3.3.3)}\quad \det\left[B_2-(I-B_1)\lambda\right] \equiv \begin{vmatrix} b_{11}-\lambda & b_{12} & \cdots & b_{1n} \\ b_{21}\lambda & b_{22}-\lambda & & b_{2n} \\ \vdots & & & \\ b_{n1}\lambda & b_{n2}\lambda & & b_{nn}-\lambda \end{vmatrix} = 0$$

po modulu manji od jedinice.

Dokaz. Kako je $\det(I-B_1)=1$, tj. matrica $I-B_1$ regularna, za (3.3.2) može se dobiti ekvivalentan metod proste iteracije. Naime, iz (3.3.2) sleduje

$$(I-B_1)\vec{x}^{(k)} = B_2\vec{x}^{(k-1)} + \vec{\beta} \qquad (k=1,2,\ldots),$$

tj.

$$\vec{x}^{(k)} = (I-B_1)^{-1}B_2\vec{x}^{(k-1)} + (I-B_1)^{-1}\vec{\beta} \qquad (k=1,2,\ldots).$$

Na osnovu teoreme 3.2.6, ovaj iterativni proces konvergira, pri proizvoljnom vektoru $\vec{x}^{(0)}$, ako i samo ako su svi koreni jednačine

$$\det\left[(I-B_1)^{-1}B_2-\lambda I\right] = 0$$

po modulu manji od jedinice.

Iz poslednje jednačine sleduje

$$\det\left[(I-B_1)^{-1}(B_2-(I-B_1)\lambda)\right] = 0,$$

tj.

$$\det(I-B_1)^{-1}\det\left[B_2-(I-B_1)\lambda\right] = 0.$$

Kako je $\det(I-B_1)^{-1}=1$, poslednja jednačina se svodi na jednačinu (3.3.3), čime je dokazana teorema 3.3.1.

Posmatrajmo sada sistem jednačina $A\vec{x}=\vec{b}$ u obliku (3.2.21). Ako stavimo

$$A - D = C_1 + C_2 ,$$

gde su

$$C_1 = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ a_{21} & 0 & & 0 & 0 \\ \vdots & & & & \\ a_{n1} & a_{n2} & & a_{n,n-1} & 0 \end{bmatrix} \quad \text{i} \quad C_2 = \begin{bmatrix} 0 & a_{12} & \dots & a_{1n} \\ 0 & 0 & & a_{2n} \\ \vdots & & & \\ 0 & 0 & & 0 \end{bmatrix},$$

može se obrazovati Gauss-Seidelov proces kao

$$\vec{x}^{(k)} = -D^{-1}C_1\vec{x}^{(k)} - D^{-1}C_2\vec{x}^{(k-1)} + D^{-1}\vec{b} \qquad (k=1,2,\dots). \tag{3.3.4}$$

Ova varijanta Gauss-Seidelovog metoda ponekad se ne naziva metod Nekrasova (videti [15]).

Iz teoreme 3.3.1 sleduje sledeća teorema:

Teorema 3.3.2. Pri proizvoljnom vektoru $\vec{x}^{(o)}$, iterativni proces (3.3.4) konvergira ako i samo ako su svi koreni jednačine

$$\det\left[C_2+(D+C_1)\lambda\right] = \begin{vmatrix} a_{11}\lambda & a_{12} & \dots & a_{1n} \\ a_{21}\lambda & a_{22}\lambda & & a_{2n} \\ \vdots & & & \\ a_{n1}\lambda & a_{n2}\lambda & & a_{nn}\lambda \end{vmatrix} = 0$$

po modulu manji od jedinice.

Kao što je i ranije napomenuto, ovi spektralni uslovi za konvergenciju iterativnih procesa, nažalost, nemaju veliki praktični značaj.

Za sisteme jednačina sa simetričnom matricom, E.Reich([16]) je dokazao sledeći rezultat:

Teorema 3.3.3. Neka je matrica A realna i simetrična i neka su joj svi dijagonalni elementi pozitivni. Iterativni proces (3.3.4) konvergira ako i samo ako su sve veličine

$$a_{11}, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix}, \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix}, \dots, \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \\ a_{1n} & & a_{nn} \end{vmatrix}$$

pozitivne.

L.Collatz ([3], [17]) dokazao je sledeći rezultat:

Teorema 3.3.4. Iterativni procesi Jacobia (3.2.21) i Gauss-Seidela (3.3.4) konvergiraju, ako matrica A reda n ispunjava sledeća dva uslova:

1^o Matrica A ne sadrži nula-submatricu tipa $m\times(n-m)$ $(1\leq m\leq n-1)$;

2^o Za svako $i\in I=\{1,\ldots,n\}$ je $|a_{ii}|\geq s_i=\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|$ i bar za jedno $i\in I$ je $|a_{ii}|>s_i$.

Primer 3.3.1. Ispitaćemo primenljivost iterativnog procesa (3.3.4) na rešavanje sistema

$$
\begin{aligned}
10x_1+3x_2-x_3&=12,\\
-x_1+5x_2-x_3&=3,\\
x_1+2x_2+10x_3&=13.
\end{aligned}
\tag{3.3.5}
$$

Kako za elemente matrice datog sistema

$$
A=\begin{bmatrix}10&3&-1\\-1&5&-1\\1&2&10\end{bmatrix}
$$

važi

$$
\begin{aligned}
|a_{11}|&=10>s_1=|a_{12}|+|a_{13}|=4,\\
|a_{22}|&=\ 5>s_2=|a_{21}|+|a_{23}|=2,\\
|a_{33}|&=10>s_3=|a_{31}|+|a_{32}|=3,
\end{aligned}
$$

i kako A ne sadrži nula-submatricu tipa 1x2 ili tipa 2x1, zaključujemo da su uslovi 1^o i 2^o u teoremi 3.3.4 ispunjeni. Dakle, iterativni proces (3.3.4) primenjen na rešavanje sistema (3.3.5) konvergira.

Polazeći od $\vec{x}^{(0)}=\vec{\beta}=[1.2\ \ 0.6\ \ 1.3]^T$, na osnovu

$$
\begin{aligned}
x_1^{(k)}&=-0.3x_2^{(k-1)}+0.1x_3^{(k-1)}+1.2,\\
x_2^{(k)}&=\ \ 0.2x_1^{(k)}+0.2x_3^{(k-1)}+0.6,\\
x_3^{(k)}&=-0.1x_1^{(k)}-0.2x_2^{(k)}+1.3,
\end{aligned}
$$

gde je k=1,2,..., dobijamo niz

$$\vec{x}^{(1)} = [1.1500000 \quad 1.0900000 \quad 0.9670000]^T,$$

$$\vec{x}^{(2)} = [0.9697000 \quad 0.9873400 \quad 1.0055620]^T,$$

$$\vec{x}^{(3)} = [1.0043542 \quad 1.0019832 \quad 0.9991680]^T,$$

$$\vec{x}^{(4)} = [0.9993219 \quad 0.9996979 \quad 1.0001283]^T,$$

itd. Primetimo da je tačno rešenje sistema (3.3.5)

$$\vec{x} = [1 \quad 1 \quad 1]^T.$$

Jedno iteresantno pitanje koje se može postaviti u vezi sa razmatranim iterativnim metodima je: Da li Gauss-Seidelov metod uvek konvergira kada konvergira odgovarajući metod proste iteracije? Odgovor na ovo pitanje nije potvrdan. Naime, u dosta velikom broju slučajeva, ako metod proste iteracije konvergira, konvergiraće i Gauss-Seidelov metod i to brže, medjutim, postoje slučajevi kada ovaj poslednji ne konvergira.

Štaviše, postoje i slučajevi kada Gauss-Seidelov metod konvergira, a metod proste iteracije divergira. Sledeći prost primer ovo lepo ilustruje.

Primer 3.3.2. Neka je

$$\vec{x} = B\vec{x} + \vec{\beta},$$

gde su

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \vec{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \quad B = \begin{bmatrix} p & q \\ -q & p \end{bmatrix} \quad (p,q \in R).$$

Odredićemo parametre p i q tako da konvergira

1^o metod proste iteracije;

2^o Gauss-Seidelov metod.

Iz

$$\begin{vmatrix} p-\lambda & q \\ -q & p-\lambda \end{vmatrix} = 0$$

sleduje $\lambda_{1,2}=p\pm iq$, odakle, za uslov konvergencije metoda proste iteracije, dobijamo

$$|\lambda_1| = |\lambda_2| = \sqrt{p^2+q^2} < 1,$$

tj.

$$p^2 + q^2 < 1.$$

Odgovarajuća karakteristična jednačina za Gauss-Seidelov metod je

$$\begin{vmatrix} p-\lambda & q \\ -q\lambda & p-\lambda \end{vmatrix} = 0,$$

tj.

$$\lambda^2 - (2p-q^2)\lambda + p^2 = 0.$$

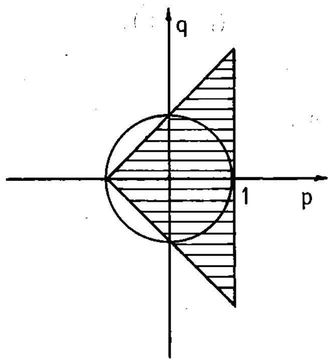
Potrebni i dovoljni uslovi da poslednja jednačina ima korene po modulu manje od jedinice su (videti [18], [19])

$$|2p-q^2| < p^2 + 1 \quad \text{i} \quad |p^2| < 1,$$

odakle, nakon elementarnih algebarskih transformacija, dobijamo

$$(3.3.6) \qquad |q| < 1 + p \quad \text{i} \quad |p| < 1.$$

U ravni pOq nejednakosti (3.3.6) definišu oblast prikazanu šrafirano na Sl.3.3.1. Ako (p,q) pripada ovoj oblasti, Gauss-Seidelov metod konvergira. S druge strane, metod proste iteracije konvergira, kao što je dokazano, ako se (p,q) nalazi u unutrašnjosti jediničnog kruga.



Sl. 3.3.1

Predjimo sada na odredjivanje greške približnog rešenja koje se dobija primenom iterativnog procesa (3.3.2).

Kako je tačno rešenje $\vec{x}$ nepoznato, greška $\vec{\varepsilon}^{(k)}=\vec{x}^{(k)}-\vec{x}$, se ne može tačno odrediti. Medjutim, kako se greška $\vec{\varepsilon}^{(k)}$ može izraziti pomoću priraštaja $\vec{\delta}^{(k)}=\vec{x}^{(k)}-\vec{x}^{(k-1)}$, moguće je dati ocenu za $\|\vec{\varepsilon}^{(k)}\|$.

Teorema 3.3.5. Ako je $\vec{x}$ rešenje jednačine

$$\vec{x} = B\vec{x} + \vec{\beta} \qquad (B=B_1+B_2)$$

i ako je $\|B\| \leqq q < 1$, važi nejednakost

$$(3.3.7) \qquad \|\vec{x}^{(k)} - \vec{x}\| \leqq \frac{\|B_2\|}{1-\|B\|}\|\vec{x}^{(k)}-\vec{x}^{(k-1)}\| \qquad (k\in N),$$

gde se niz $\{\vec{x}^{(k)}\}$ generiše pomoću (3.3.2).

Dokaz. Na osnovu (3.3.2), za svako $k\in N$, imamo

$$\vec{\varepsilon}^{(k)} = (B_1\vec{x}^{(k)}+B_2\vec{x}^{(k-1)}+\vec{\beta})-(B_1\vec{x}+B_2\vec{x}+\vec{\beta}) = B_1\vec{\varepsilon}^{(k)}+B_2\vec{\varepsilon}^{(k-1)}.$$

Kako je $\vec{\varepsilon}^{(k-1)}=\vec{x}^{(k-1)}-\vec{x} = \vec{\varepsilon}^{(k)}-\vec{\delta}^{(k)}$, iz (3.3.8) sleduje

$$\vec{\varepsilon}^{(k)} = B_1\vec{\varepsilon}^{(k)} + B_2(\vec{\varepsilon}^{(k)}-\vec{\delta}^{(k)}) \qquad (k\in N),$$

tj.

$$\vec{\varepsilon}^{(k)} = -(I-B)^{-1}B_2\vec{\delta}^{(k)} \qquad (k\in N),$$

s obzirom da egzistira $(I-B)^{-1}$ (na osnovu pretpostavke $\|B\| \leqq q < 1$).

Ako koristimo normu matrice saglasnu sa normom vektora, iz poslednje jednakosti dobijamo

$$\|\vec{\varepsilon}^{(k)}\| \leqq \|(I-B)^{-1}B_2\|\cdot\|\vec{\delta}^{(k)}\| \leqq \frac{\|B_2\|}{1-\|B\|}\|\vec{\delta}^{(k)}\| \qquad (k\in N),$$

čime smo dokazali teoremu 3.3.5. •

Primedba 3.3.1. Ako je $B_1=O(\Rightarrow B_2=B)$, (4.3.8) se svodi na nejednakost datu u primedbi 3.2.2.

4.3.4. Opšte napomene o relaksacionim metodima

Neka je $\vec{x}_p$ približno rešenje sistema jednačina

$$A\vec{x} = \vec{b}. \tag{3.4.1}$$

Kako je tačno rešenje $\vec{x}=A^{-1}\vec{b}$ nepoznato, postavlja se pitanje u kojoj meri $\vec{x}_p$ zadovoljava dati sistem jednačina. Kao mera ovoga, najčešće se koristi norma vektora ostatka

$$\vec{r} = A\vec{x}_p - \vec{b}. \tag{3.4.2}$$

Ako je $\vec{r}$ nula-vektor, lako se uočava da je $\vec{x}_p$ tačno rešenje sistema (3.4.1). Na osnovu prethodnog, može se zaključiti da je sistem (3.4.1) skoro zadovoljen ako su komponente vektora $\vec{r}$ bliske nuli. Poslednje tvrdjenje, medjutim, nije uvek u važnosti. Naime, kod slabo uslovljenih sistema, norma vektora $\vec{x}_p-A^{-1}\vec{b}$ može biti velika, i u slučajevima kada je norma vektora ostatka mala, s obzirom da je $\vec{x}_p-A^{-1}\vec{b} = A^{-1}\vec{r}$.

I pored ovog nedostatka, vektor $\vec{r}$ igra važnu ulogu u širokoj klasi iterativnih metoda, tzv. relaksacionih metoda.

Pod relaksacionim metodom podrazumeva se svaki metod kod koga se sledeća aproksimacija rešenja dobija na osnovu prethodne aproksimacije i vektora ostatka (u opštem slučaju zavisi od više prethodnih aproksimacija), koji se koristi kao indikator za veličinu korekcije.

Prve ideje o relaksacionim metodima potiču od Gaussa. Medjutim, sistematska teorijska istraživanja na ovim metodima datiraju iz poslednjih nekoliko decenija. S obzirom da su ovi metodi pogodni za rešavanje sistema sa velikim brojem jednačina, to oni sve više nalaze primenu kod rešavanja parcijalnih jednačina. O relaksacionim metodima danas postoji vrlo obimna literatura (videti posebno [20], [21], [22], [23], [24], [25].

U sledećim odeljcima obradićemo nekoliko klasa relaksacionih metoda.

4.3.5. Metod sukcesivne gornje relaksacije

U ovom odeljku razmotrićemo jedan relaksacioni metod, koji predstavlja generalizaciju jedne varijante Gauss-Seidelovog metoda za rešavanje sistema $A\vec{x}=\vec{b}$.

Neka je matrica D=diag(A) regularna. Razložimo matricu A u obliku

$$A = C_1 + D + C_2,$$

gde su

$$C_1 = \begin{bmatrix} 0 & 0 \ldots & 0 & 0 \\ a_{21} & 0 & 0 & 0 \\ \vdots & & & \\ a_{n1} & a_{n2} & a_{n,n-1} & 0 \end{bmatrix} \quad \text{i} \quad C_2 = \begin{bmatrix} 0 & a_{12} & \ldots & a_{1n} \\ 0 & 0 & & a_{2n} \\ \vdots & & & \\ 0 & 0 & & 0 \end{bmatrix}.$$

Ako za izračunavanje vektora ostatka $\vec{r}$ uzmemo za x_i poslednju izračunatu vrednost (kao kod Gauss-Seidelovog metoda), tj.

$$\vec{r} = C_1\vec{x}^{(k)} + (D+C_2)\vec{x}^{(k-1)} - \vec{b}$$

i ako stavimo $\vec{\delta}^{(k)} = \vec{x}^{(k)} - \vec{x}^{(k-1)}$, možemo formirati iterativni proces u obliku

$$D\vec{\delta}^{(k)} = -\omega\vec{r} \qquad (k=1,2,\ldots),$$

tj.

$$D\vec{x}^{(k)} = D\vec{x}^{(k-1)} + \omega\left[\vec{b} - C_1\vec{x}^{(k)} - (D+C_2)\vec{x}^{(k-1)}\right], \tag{3.5.1}$$

gde je k=1,2,... i ω realan parametar. Parametar ω nazivamo relaksacioni množilac i njega u opštem slučaju možemo menjati u toku računanja.

Iterativni proces (3.5.1) može se predstaviti i u skalarnom obliku

$$a_{ii}\overset{*}{x}_i^{(k)} = -\sum_{j<i} a_{ij}x_j^{(k)} - \sum_{j>i} a_{ij}x_j^{(k-1)} + b_i,$$

$$x_i^{(k)} = x_i^{(k-1)} + \omega(\overset{*}{x}_i^{(k)} - x_i^{(k-1)}),$$

gde su i=1,...,n i k=1,2,... .

Primetimo da se za $\omega=1$, (3.5.1) svodi na Gauss-Seidelov proces (3.3.7).

Na osnovu (3.5.1) imamo

$$(3.5.2)\qquad (D+\omega C_1)\vec{x}^{(k)} = \left[(1-\omega)D-\omega\zeta\right]\vec{x}^{(k-1)}+\omega\vec{b} \qquad (k=1,2,\ldots),$$

tj.

$$(3.5.3)\qquad \vec{x}^{(k)} = K(\omega)\vec{x}^{(k-1)} + \vec{f}(\omega) \qquad (k=1,2,\ldots),$$

gde su

$$K(\omega) = (D+\omega C_1)^{-1}\left[(1-\omega)D-\omega\zeta\right] \quad \text{i} \quad \vec{f}(\omega) = \omega(D+\omega C_1)^{-1}\vec{b}.$$

Teorema 3.5.1. Iterativni proces (3.5.1) konvergira ako i samo ako su sve sopstvene vrednosti matrice $K(\omega)$ po modulu manje od jedinice.

Dokaz. Kako je relaksacioni metod (3.5.1) ekvivalentan sa metodom proste iteracije (3.5.3), teorema 3.5.1 se dobija kao posledica teoreme 3.2.6.

Za jednu specijalnu klasu sistema, koja se vrlo često javlja u praksi, važi sledeći kriterijum ([3]).

Teorema 3.5.2. Ako je A hermitska pozitivno definitna matrica, relaksacioni iterativni proces (3.5.1) konvergira ka tačnom rešenju sistema $A\vec{x}=\vec{b}$ kada je $0<\omega<2$.

Dokaz. Na osnovu učinjenih pretpostavki za matricu A $(=C_1+D+C_2)$ imamo da je $a_{ii}>0$ $(i=1,\ldots,n)$ i $C_1=C_2^*$. Kako je, dalje,

$$(C_1-C_2)^* = C_1^*-C_2^*=C_2-C_1=-(C_1-C_2),$$

zaključujemo da je matrica C_1-C_2 kosohermitska, pa je za svako $\vec{y}$

$$(3.5.4)\qquad \vec{y}^*(C_1-C_2)\vec{y} = i\,c \qquad (c\in R).$$

Pretpostavimo sada da je λ bilo koja sopstvena vrednost matrice $K(\omega)$ i y odgovarajući sopstveni vektor. Tada je

$$K(\omega)\vec{y} = \lambda\vec{y},$$

tj.

$$\left[(1-\omega)D - \omega C_2\right]\vec{y} = \lambda(D+\omega C_1)\vec{y}. \tag{3.5.5}$$

Množenjem jednakosti (3.5.5) sleva za $2\vec{y}^*$, dobijamo

$$\vec{y}^*\left[(2-\omega)D-\omega(D+2C_2)\right]\vec{y} = \lambda\vec{y}^*\left[(2-\omega)D+\omega(D+2C_1)\right]\vec{y},$$

tj.

$$\vec{y}^*\left[(2-\omega)D-\omega A+\omega(C_1-C_2)\right]\vec{y} = \lambda\vec{y}^*\left[(2-\omega)D+\omega A+\omega(C_1-C_2)\right]\vec{y}, \tag{3.5.6}$$

s obzirom da je $A = C_1+D+C_2$.

Kako su matrice A i D pozitivno definitne imamo

$$\vec{y}^*A\vec{y} = a > 0 \quad \text{i} \quad \vec{y}^*D\vec{y} = d > 0 \quad (\forall \vec{y} \neq \vec{0}). \tag{3.5.7}$$

Na osnovu (3.5.6), (3.5.4), (3.5.7) dobijamo

$$\lambda = \frac{\xi d - a + c_i}{\xi d + a + c_i} \quad \left(\xi = \frac{2}{\omega} - 1\right),$$

odakle pod pretpostavkom da je

$$0 < \xi < +\infty, \text{ tj. } 2 > \omega > 0$$

sleduje $|\lambda| < 1$, čime je teorema 3.5.2 dokazana.

U daljem izlaganju dajemo ocenu greške za razmatrani relaksacioni metod.

Kako je $A\vec{x} = \vec{b}$, na osnovu (3.5.2) redom dobijamo

$$\omega A\vec{x}^{(k)} = \left[(1-\omega)D-\omega C_2\right](\vec{x}^{(k-1)}-\vec{x}^{(k)}) + \omega\vec{b},$$

$$\omega A(\vec{x}^{(k)}-\vec{x}) = \left[(1-\omega)D-\omega C_2\right](\vec{x}^{(k-1)}-\vec{x}^{(k)}),$$

$$\vec{x}^{(k)} - \vec{x} = A^{-1}(\gamma D+C_2)(\vec{x}^{(k)}-\vec{x}^{(k-1)}), \tag{3.5.8}$$

gde je $\gamma = 1 - \frac{1}{\omega}$.

Iz (3.5.8) sleduje:

<u>Teorema 3.5.3.</u> Kod iterativnog procesa (3.5.1) važi ocena za grešku

$$\|\vec{x}^{(k)}-\vec{x}\| \leq \|A^{-1}(\gamma D+C_2)\| \cdot \|\vec{x}^{(k)}-\vec{x}^{(k-1)}\| \quad (k \in N). \tag{3.5.9}$$

Izraz (3.5.9), koji daje ocenu greške, najopštiji je za iterativni proces (3.5.1), jer ne zahteva nikakva ograničenja za matricu sistema A (sem da je regularna). Medjutim, zbog komplikovanosti izraza $\|A^{-1}(\gamma D+C_2)\|$, ocena (3.5.9) nije pogodna. Zato ćemo, nadalje, za matricu A pretpostaviti dodatne uslove (kao u teoremi 3.5.2).

Dakle, neka je matrica A hermitska pozitivno definitna. Tada je D takodje pozitivno definitna, pa postoji matrica $D^{1/2}$ i njena inverzna matrica $D^{-1/2}$.

Uvodjenjem oznaka i transformacija

$C=C_1+C_2$, $D^{-1/2}AD^{-1/2}=B$, $D^{-1/2}CD^{-1/2}=-T$, $D^{-1/2}C_1D^{-1/2}=-T_1$, $D^{-1/2}C_2D^{-1/2}=-T_2$, $D^{1/2}\vec{x}=\vec{y}$, $D^{-1/2}\vec{b}=\vec{\beta}$,

sistem $A\vec{x}=\vec{b}$ se svodi na

$$B\vec{y}=\vec{\beta},\quad \text{tj.}\quad \vec{y}=T\vec{y}+\vec{\beta}. \tag{3.5.10}$$

Primetimo da je matrica B pozitivno definitna i T hermitska.

Greška kod relaksacionog metoda primenjenog na (3.5.10), tj. kod metoda

$$(I-\omega T_1)\vec{y}^{(k)}=\left[(1-\omega)I+\omega T_2\right]\vec{y}^{(k-1)}+\omega\vec{\beta}\qquad (k=1,2,\ldots), \tag{3.5.11}$$

je

$$\|\vec{y}^{(k)}-\vec{y}\|\leq\|B^{-1}(\gamma I-T_2)\|\cdot\|\vec{y}^{(k)}-\vec{y}^{(k-1)}\|\qquad (k\in\mathbb{N}), \tag{3.5.12}$$

gde je $B=I-T$.

Kod primene relaksacionog metoda veoma je važan izbor vrednosti relaksacionog parametra ω. U literaturi (videti [21], [23], [25]) definisan je optimalni parametar ω_{opt} pomoću

$$\rho(K(\omega_{opt}))=\min_{\omega}\rho(K(\omega)),$$

gde je sa $\rho(S)$ označen spektralni radijus matrice S. Primetimo da je kod (3.5.11)

$$K(\omega)=(I-\omega T_1)^{-1}\left[(1-\omega)I+\omega T_2\right]. \tag{3.5.13}$$

J.Albrecht ([23]) je razmatrao iterativni proces (3.5.11), pri uslovu da je hermitska matrica T(u sistemu (3.5.10)) ciklična indeksa 2 (videti [24]), tj. da je oblika

$$(3.5.14) \qquad T = \begin{array}{cc} \left[\begin{array}{c|c} O & F \\ \hline E & O \end{array}\right] & \begin{array}{l} n_1 \text{ vrsta} \\ n_2 \text{ vrsta} \end{array} \\ \begin{array}{cc} n_1 & n_2 \\ \text{kolona} & \text{kolona} \end{array} & \end{array} \qquad (n_1+n_2=n)$$

i da je $\delta(T)<1$. Matrica T ima svojstvo (A) (videti definiciju 1.1.2).

Na osnovu prethodnog imamo

$$T_1 = \left[\begin{array}{c|c} O & O \\ \hline E & O \end{array}\right] \quad \text{i} \quad T_2 = \left[\begin{array}{c|c} O & F \\ \hline O & O \end{array}\right].$$

<u>Teorema 3.5.4.</u> Neka su T i $K(\omega)$ matrice date pomoću (3.5.14) i (3.5.13) respektivno. Ako je $\{{\tau \atop \lambda}\}$ sopstvena vrednost od $\{{T \atop K(\omega)}\}$ i važi

$$(3.5.15) \qquad (\lambda+\omega-1)^2=\lambda\omega^2\tau^2,$$

tada je $\{{\lambda \atop \tau}\}$ sopstvena vrednost od $\{{K(\omega) \atop T}\}$.

Na osnovu teoreme 3.5.4, odredićemo optimalnu vrednost relaksacionog parametra ω u iterativnom procesu (3.5.11).

Neka je hermitska matrica T ciklična indeksa 2 i neka je njen spektralni radijus $\rho=\rho(T)=\max|\tau(T)|<1$.

Ako je τ sopstvena vrednost matrice T, tada je $-\tau$ takodje njena sopstvena vrednost. Na osnovu teoreme 3.5.4, ovim sopstvenim vrednostima odgovaraju dve sopstvene vrednosti matrice $K(\omega)$. Naime, iz (3.5.15) sleduje

$$\lambda_{\pm} = f_{\pm}(\omega,\tau) = \frac{1}{4}(\omega|\tau|\pm\sqrt{\omega^2\tau^2-4\omega+4})^2.$$

Kako je $f_+(\omega,\tau)f_-(\omega,\tau)=(\omega-1)^2$, zaključujemo da iterativni proces (3.5.11) divergira ako je $|\omega-1|\geq 1$, tj. $\omega\leq 0$ ili $\omega\geq 2$. Zato razmotrimo samo slučaj kada je $0<\omega<2$.

Ako je $0<\omega\leqq\omega_1$ $\left(\omega_1=\omega_1(\tau)=\dfrac{2}{1+\sqrt{1-\tau^2}}\right)$, koreni $\lambda_\pm$ su realni i pozitivni i veći od njih je

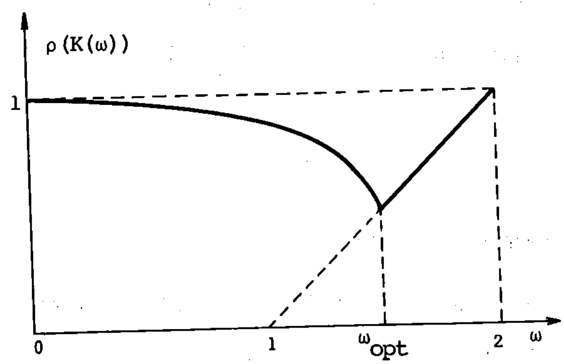
$$\lambda_+ = f_+(\omega,\tau) = \frac{1}{4}\left(\omega|\tau| + \sqrt{\omega^2\tau^2 - 4\omega + 4}\right)^2 .$$

Ako je $\omega_1<\omega<2$, koreni $\lambda_\pm$ su kompleksni sa modulom $\omega-1$.

Kako je funkcija $|\tau|\to f_+(\omega,\tau)$ rastuća na $(0,\rho)$, zaključujemo da je

$$\rho(K(\omega)) = \begin{cases} f_+(\omega,t) & \left(0<\omega\leqq\dfrac{2}{1+\sqrt{1-\rho^2}}\right), \\ \omega-1 & \left(\dfrac{2}{1+\sqrt{1-\rho^2}}\leqq\omega<2\right). \end{cases}$$

Grafik funkcije $\omega\mapsto\rho(K(\omega))$ $(0<\omega<2)$ prikazan je na slici 3.5.1.



Sl. 3.5.1

Optimalna vrednost parametra ω je

$$\omega_{opt} = \omega_1(\rho) = \frac{2}{1+\sqrt{1-\rho^2}},$$

s obzirom da funkcija $\omega\mapsto\rho(K(\omega))$ ima minimum za $\omega=\omega_{opt}$ na segmentu $(0,2)$.

Odgovarajuća vrednost za $\rho(K(\omega))$ je

$$\rho(K(\omega_{opt})) = \omega_{opt}-1 = \frac{1-\sqrt{1-\rho^2}}{1+\sqrt{1+\rho^2}} .$$

Uvedimo oznaku

$$M(\omega) = \left[(I-T)^{-1}(\gamma I-T_2)\right]^{*}\cdot\left[(I-T)^{-1}(\gamma I-T_2)\right].$$

Teorema 3.5.5. Ako je $\{{}^{\tau}_{\mu}\}$ sopstvena vrednost od $\{{}^{T}_{M(\omega)}\}$ i važi

$$\frac{1}{2}\{\frac{1-\tau^2}{\gamma^2}\mu+\frac{1}{\frac{1-\tau^2}{\gamma^2}\mu}\} = 1+\frac{1}{2}\cdot\frac{\xi^2\tau^2+\tau^4}{\gamma^2(1-\tau^2)},$$

gde su $\gamma = 1-\frac{1}{\omega}$ i $\xi = \frac{2}{\omega}-1$, tada je $\{{}^{\mu}_{\tau}\}$ sopstvena vrednost od $\{{}^{M(\omega)}_{T}\}$.

Na osnovu teoreme 3.5.5, nalazimo

$$H(\omega) = \|(I-T)^{-1}(\gamma I-T_2)\|_{sp} = \frac{1}{2}\cdot\frac{\sqrt{\rho^2(\xi^2+\rho^2)+4\gamma^2(1-\rho^2)}+\sqrt{\rho^2(\xi^2+\rho^2)}}{1-\rho^2},$$

gde je $\rho=\rho(T)=\sigma(T)<1$.

Teorema 3.5.6. Neka je hermitska matrica T ciklična indeksa 2 i neka je njen spektralni radijus $\rho=\rho(T)<1$.

Ako je $0<\omega<2$, za iterativni proces (3.5.11), važi ocena greške

$$\|\vec{y}^{(k)}-\vec{y}\|_E \leqq H(\omega)\cdot\|\vec{y}^{(k)}-\vec{y}^{(k-1)}\|_E \quad (k\in N).$$

Napomenimo da je kod Gauss-Seidelovog metoda, tj. kada je $\omega = 1$,

$$H = H(1) = \frac{\rho\sqrt{1+\rho^2}}{1-\rho^2}.$$

4.3.6. Čebiševljev semi-iterativni metod

U ovom odeljku obradićemo Čebiševljev semi-iterativni metod i ukazaćemo na njegovu vezu sa metodom sukcesivne gornje relaksacije.

Neka je dat sistem jednačina

$$\vec{x} = B\vec{x}+\vec{\beta} \tag{3.6.1}$$

sa hermitskom matricom B reda n. Ako je $\rho(B)<1$, iterativni proces

$$\vec{x}^{(k)} = B\vec{x}^{(k-1)} + \vec{\beta} \qquad (k=1,2,\ldots) \tag{3.6.2}$$

konvergira ka rešenju $\vec{x}$ sistema (3.6.1). U cilju ubrzavanja konvergencije procesa (3.6.2), posmatrajmo linearnu kombinaciju* vektora $\vec{x}^{(k)}$, tj.

$$\vec{y}^{(k)} = \sum_{i=0}^{k} c_{ki}\vec{x}^{(i)} \qquad (k=0,1,\ldots), \tag{3.6.3}$$

pod uslovom $\sum_{i=0}^{k} c_{ki} = 1$.

Ako stavimo $\vec{\varepsilon}^{(k)} = \vec{x}^{(k)} - \vec{x}$ i $\vec{\xi}^{(k)} = \vec{y}^{(k)} - \vec{x}\,(k=0,1,\ldots)$, imamo

$$\vec{\xi}^{(k)} = \sum_{i=0}^{k} c_{ki}\vec{\varepsilon}^{(i)} = \Big(\sum_{i=0}^{k} c_{ki}B^i\Big)\vec{\varepsilon}^{(o)},$$

tj.

$$\vec{\xi}^{(k)} = Q_k(B)\,\vec{\varepsilon}^{(o)}, \tag{3.6.4}$$

gde je

$$Q_k(t) = \sum_{i=0}^{k} c_{ki}t^i \quad \text{i} \quad Q_k(1) = 1. \tag{3.6.5}$$

Neka je $\{\vec{e}_i\}$ ortonormiran sistem sopstvenih vektora hermitske matrice B, gde je $B\vec{e}_i = \lambda_i\vec{e}_i\,(\lambda_i \equiv \lambda_i(B))\,(i=1,\ldots,n)$. Ako vektor $\vec{\varepsilon}^{(o)}$ razvijemo po sopstvenim vektorima $\vec{e}_i$, tj. $\vec{\varepsilon}^{(o)} = p_1\vec{e}_1 + \ldots + p_n\vec{e}_n$, jednakost (3.6.4) postaje

$$\vec{\xi}^{(k)} = \sum_{i=1}^{n} p_i Q_k(\lambda_i)\vec{e}_i\,.$$

Označimo sa $\mathcal{P}_k$ skup svih polinoma Q_k oblika (3.6.5), tj. skup svih polinoma stepena k sa normalizacijom $Q_k(1)=1$. Može se pokazati da opšta tročlana rekurentna formula kojom se generišu polinomi $Q_k \in \mathcal{P}_k$ ima oblik

$$Q_{k+1}(t) = (\alpha_k t + 1 - \alpha_k - \beta_k)Q_k(t) + \beta_k Q_{k-1}(t) \qquad (k=0,1,\ldots), \tag{3.6.6}$$

*) Često se ovakav postupak za ubrzavanje konvergencije naziva linearno ubrzavanje (videti [26]).

gde su $Q_o(t)=1$, $\beta_o=0$, α_k i β_k realni brojevi.

Kako je

$$\| \vec{\xi}^{(k)} \|_E = \left(\sum_{i=1}^{n} p_i^2 \, Q_k(\lambda_i)^2 \right)^{1/2} \leqq \max_i |Q_k(\lambda_i)| \left(\sum_{i=1}^{n} p_i^2 \right)^{1/2},$$

tj.

$$\| \vec{\xi}^{(k)} \|_E \leqq \| \vec{\varepsilon}^{(o)} \|_E \max_i |Q_k(\lambda_i)|,$$

za izbor polinoma $\{\tilde{Q}_k\}$ usvojimo kriterijum

$$\min_{Q_k \in \mathcal{P}_k} \left(\max_{-\rho \leq t \leq \rho} | Q_k(t) | \right) = \max_{-\rho \leq t \leq \rho} | \tilde{Q}_k(t) |,$$

gde je $\rho \equiv \rho(B)$. Tada dobijamo (videti

(3.6.7) $$\tilde{Q}_k(t) = \frac{T_k(t/\rho)}{T_k(1/\rho)} \qquad (k = 0,1,\ldots),$$

gde je T_k Čebiševljev polinom stepena k. Lako je pokazati da se, u ovom slučaju, rekurentna relacija (3.6.6) svodi na

(3.6.8) $$\tilde{Q}_{k+1}(t) = \alpha_k t \tilde{Q}_k(t) + (1-\alpha_k)\tilde{Q}_{k-1}(t) \qquad (k=0,1,\ldots),$$

gde su

(3.6.9) $$\alpha_o = 1, \ \beta_o = 0; \ \alpha_k = \frac{2}{\rho} \cdot \frac{T_k(1/\rho)}{T_{k+1}(1/\rho)}, \ \beta_k = 1 - \alpha_k \ (k = 1,2,\ldots).$$

Dakle, na osnovu (3.6.8) generiše se niz polinoma $\{\tilde{Q}_k\}$, a zatim, na osnovu (3.6.3) i niz $\{\vec{y}^{(k)}\}$. Ovaj postupak je poznat kao Čebiševljev semi-iterativni metod.

Pokazaćemo sada kako se izloženi metod može predstaviti i u eksplicitnom obliku (videti [25]).

Iz (3.6.8) sleduje

$$\tilde{Q}_{k+1}(t) - \tilde{Q}_{k-1}(t) = \alpha_k (t\tilde{Q}_k(t) - \tilde{Q}_{k-1}(t)).$$

S druge strane, kako je $\vec{\xi}^{(k+1)} - \vec{\xi}^{(k-1)} = (\tilde{Q}_{k+1}(B) - \tilde{Q}_{k-1}(B))\vec{\varepsilon}^{(o)}$ imamo

$$\vec{\xi}^{(k+1)} - \vec{\xi}^{(k-1)} = \alpha_k (B\tilde{Q}_k(B) - \tilde{Q}_{k-1}(B)) \vec{\varepsilon}^{(o)} = \alpha_k (B\vec{\xi}^{(k)} - \vec{\xi}^{(k-1)}),$$

tj.

(3.6.10) $\quad \vec{y}^{(k+1)} = \vec{y}^{(k-1)} + \alpha_k(B\vec{y}^{(k)} + \vec{\beta} - \vec{y}^{(k-1)}) \quad (k=1,2,\ldots),$

gde je niz $\{\alpha_k\}$ definisan pomoću (3.6.9) i

$$\vec{y}^{(o)} = \vec{x}^{(o)} \quad i \quad \vec{y}^{(1)} = B\vec{x}^{(o)} + \vec{\beta} .$$

Na osnovu prethodnog, članovi niza $\{\alpha_k\}$ su

$$\alpha_o = 1, \ \alpha_1 = \frac{2}{2-\varrho^2}, \ \alpha_k = \frac{1}{1-\frac{1}{4}\varrho^2\alpha_{k-1}} \quad (k=2,3,\ldots),$$

pri čemu je

$$2 > \alpha_1 > \alpha_2 > \ldots \geq \lim_{k\to+\infty} \alpha_k = \omega_{opt} = \frac{2}{1+\sqrt{1-\varrho^2}} > 1 .$$

Dakle, α_k teži optimalnom relaksacionom faktoru, koji je dobijen kod metoda gornje relaksacije.

Primedba 3.6.1. Ako za niz polinoma $\{\tilde{Q}_k\}$ izaberemo

$$\tilde{Q}_k(t) = \frac{1-T_{k+1}(t)}{(k+1)^2(1-t)} = \frac{\sin^2\frac{k+1}{2}\theta}{(k+1)^2\sin^2\frac{\theta}{2}} \quad (t=\cos\theta),$$

dobijamo metod Lanczosa([27]).

Primedba 3.6.2. E.Stiefel([28]) je razmatrao jedan relaksacioni metod, tzv. hipergeometrijsku relaksaciju, koristeći umesto Čebiševljevih polinoma ultrasferne(Gegenbauerove) polinome.

4.3.7. Gradijentni metodi

U klasi metoda koji se koriste kod minimizacije funkcionela posebnu ulogu igraju gradijentni metodi. U ovom odeljku obradićemo dva gradijentna metoda za rešavanje sistema linearnih jednačina

(3.7.1) $\quad A\vec{x} = \vec{b},$

gde je A normalna matrica(videti definiciju 1.4.7). Prvi od njih je metod najbržeg pada, a drugi metod konjugovanih gradijenata. Napomenimo da metod konjugovanih gradijenata u suštini nije iterativni metod. Oba navedena metoda zasnivaju se na minimizaciji funkcionele $F:R^n \to R$, definisane pomoću

$$(3.7.2) \qquad F(\vec{x}) = (A\vec{x},\vec{x}) - 2(\vec{b},\vec{x}).$$

Kako je A normalna matrica imamo

$$F(\vec{x}) - F(A^{-1}\vec{b}) = (A(\vec{x} - A^{-1}\vec{b}), \vec{x} - A^{-1}\vec{b}) \geq 0,$$

odakle zaključujemo da funkcionela F postiže minimum za $\vec{x}=A^{-1}\vec{b}$, što predstavlja rešenje sistema (3.7.1).

<u>1. Metod najbržeg pada</u>. Svaki metod oblika

$$(3.7.3) \qquad \vec{x}^{(k)} = \vec{x}^{(k-1)} - \alpha_{k-1}\operatorname{grad}F(\vec{x}^{(k-1)}) \qquad (k = 1,2,\ldots)$$

naziva se metod gradijentnog pada.

Ako se parametar α_p odredjuje iz uslova da je veličina

$$(3.7.4) \qquad F(\vec{x}^{(p)} - \alpha_p \operatorname{grad} F(\vec{x}^{(p)})) \qquad (\forall p \in N_0)$$

minimalna, metod (3.7.3) se naziva metod najbržeg pada.

Iz (3.7.2) sleduje $\operatorname{grad}F(x) = 2(Ax - b)$. Tada je

$$(3.7.5) \qquad \vec{x}^{(k)} = \vec{x}^{(k-1)} - 2\alpha_{k-1}(A\vec{x}^{(k-1)} - \vec{b}) \qquad (k = 1,2,\ldots).$$

Ako u jednakosti

$$F(\vec{x} + t\vec{r}) = F(\vec{x}) + 2t(A\vec{x} - \vec{b}, \vec{r}) + t^2(A\vec{r},\vec{r}) \qquad (t \in R)$$

izvršimo supstituciju

$$\begin{pmatrix} \vec{x} & \vec{r} & t \\ \vec{x}^{(p)} & A\vec{x}^{(p)} - \vec{b} & -2\alpha_p \end{pmatrix} \qquad (p \in N_0),$$

dobijamo da je vrednost izraza (3.7.4) data sa

$$F(\vec{x}^{(p+1)}) = F(\vec{x}^{(p)}) - 4\alpha_p(\vec{r}^{(p)},\vec{r}^{(p)}) + 4\alpha_p^2(A\vec{r}^{(p)},\vec{r}^{(p)}),$$

gde je $\vec{r}^{(p)} = A\vec{x}^{(p)} - \vec{b}$.

Kvadratni trinom $\alpha_p \mapsto \Phi(\alpha_p) = F(\vec{x}^{(p+1)})$ ima minimalnu vrednost ako je

$$(3.7.6) \qquad 2\alpha_p = \frac{(\vec{r}^{(p)},\vec{r}^{(p)})}{(A\vec{r}^{(p)},\vec{r}^{(p)})}.$$

Na osnovu (3.7.5) i (3.7.6) dobijamo metod najvećeg pada

$$(3.7.7)\qquad \vec{x}^{(k)} = \vec{x}^{(k-1)} - \frac{\|A\vec{x}^{(k-1)}-\vec{b}\|_E^2}{(A(A\vec{x}^{(k-1)}-\vec{b}),A\vec{x}^{(k-1)}-\vec{b})}(A\vec{x}^{(k-1)}-\vec{b}),$$

gde je $k=1,2,\ldots$.

<u>Teorema 3.7.1</u>. Neka je $\vec{x}$ tačno rešenje sistema (3.7.1) čija je matrica A normalna. Za metod najvećeg pada važe nejednakosti

$$(3.7.8)\qquad F(\vec{x}^{(k)}) - F(\vec{x}) \leqq \left(\frac{M-m}{M+m}\right)^2 (F(\vec{x}^{(k-1)}) - F(\vec{x})),$$

$$(3.7.9)\qquad \|\vec{x}^{(k)} - \vec{x}\|_E \leqq \left(\frac{M-m}{M+m}\right)^k \sqrt{\frac{M}{m}}\, \|\vec{x}^{(o)} - \vec{x}\|_E ,$$

gde je $\vec{x}^{(o)}$ proizvoljan početni vektor i $0 < m \leqq \lambda_i(A) \leqq M$.

<u>Dokaz</u>. Neka je $\vec{x}^{(k-1)}$ dobijeno pomoću (3.7.7), polazeći od proizvoljnog vektora $\vec{x}^{(o)}$. Stavimo $\vec{y}^{(k-1)} = \vec{x}^{(k-1)}$ i odredimo $\vec{y}^{(k)}$ koristeći optimalni metod (3.2.16), tj.

$$(3.7.10)\qquad \vec{y}^{(k)} = \vec{y}^{(k-1)} - \frac{2}{M+m}(A\vec{y}^{(k-1)} - \vec{b}).$$

Kako je (3.7.10) oblika (3.7.5) sa parametrom $\frac{2}{M+m}$ i kako je $F(\vec{x}^{(k)})$ minimalno kada je $2\alpha_{k-1}$ odredjeno pomoću (3.7.6), zaključujemo da važi nejednakost

$$(3.7.11)\qquad F(\vec{x}^{(k)}) \leqq F(\vec{y}^{(k)}).$$

Neka je $\{\vec{e}_i\}$ ortonormiran sistem sopstvenih vektora matrice A. Tada vektor $\vec{\xi}^{(k-1)} = \vec{y}^{(k-1)} - \vec{x}$ možemo predstaviti u obliku $\vec{\xi}^{(k-1)} = c_1\vec{e}_1 + \ldots + c_n\vec{e}_n$. Kako je $A\vec{e}_i = \lambda_i\vec{e}_i\ (\lambda_i \equiv \lambda_i(A))$, imamo

$$(3.7.12)\qquad (A\vec{\xi}^{(k-1)}, \vec{\xi}^{(k-1)}) = \sum_{i=1}^{n} c_i^2\lambda_i .$$

Na osnovu (3.7.10) imamo $B = I - \frac{2}{M+m}A$ i

$$\vec{\xi}^{(k)} = \vec{y}^{(k)} - \vec{x} = B(\vec{y}^{(k-1)} - \vec{x}) = B\vec{\xi}^{(k-1)} .$$

Nadalje, kako je matrica B simetrična sa sopstvenim vrednostima $1-\frac{2}{M+m}\lambda_i$ $(i=1,\dots,n)$ imamo

$$(A\vec{\xi}^{(k)},\vec{\xi}^{(k)}) = (AB\vec{\xi}^{(k-1)},B\vec{\xi}^{(k-1)}) = (BAB\vec{\xi}^{(k-1)},\vec{\xi}^{(k-1)}),$$

tj.

$$(3.7.13)\qquad (A\vec{\xi}^{(k)},\vec{\xi}^{(k)}) = \sum_{i=1}^{n} c_i^2\lambda_i\left(1-\frac{2}{M+m}\lambda_i\right)^2,$$

jer su sopstvene vrednosti matrice BAB odredjene sa

$$\lambda_i\left(1-\frac{2}{M+m}\lambda_i\right)^2 \qquad (i=1,\dots,n).$$

S druge strane, kako je

$$\left|1-\frac{2}{M+m}\lambda_i\right| \leq \frac{M-m}{M+m} \qquad (i=1,\dots,n),$$

na osnovu (3.7.12) i (3.7.13) zaključujemo da je

$$(3.7.14)\qquad (A\vec{\xi}^{(k)},\vec{\xi}^{(k)}) \leq \left(\frac{M-m}{M+m}\right)^2 (A\vec{\xi}^{(k-1)},\vec{\xi}^{(k-1)}),$$

tj.

$$(3.7.15)\qquad F(\vec{y}^{(k)}) - F(\vec{x}) \leq \left(\frac{M-m}{M+m}\right)^2 (F(\vec{y}^{(k-1)}) - F(\vec{x})).$$

Kako je $\vec{y}^{(k-1)}=\vec{x}^{(k-1)}$, iz (3.7.11) i (3.7.15) sleduje nejednakost (3.7.8).

Da bismo dokazali nejednakost (3.7.9) stavimo $\vec{\varepsilon}^{(p)}=\vec{x}^{(p)}-\vec{x}$ $(p=0,1,\dots)$. Primetimo da je $\vec{\varepsilon}^{(k-1)}=\vec{\xi}^{(k-1)}$ i

$$(A\vec{\varepsilon}^{(k)},\vec{\varepsilon}^{(k)}) = F(\vec{x}^{(k)}) - F(\vec{x}) \leq F(\vec{y}^{(k)}) - F(\vec{x}) = (A\vec{\xi}^{(k)},\vec{\xi}^{(k)}).$$

Tada iz (3.7.14) sleduje

$$(A\vec{\varepsilon}^{(k)},\vec{\varepsilon}^{(k)}) \leq \left(\frac{M-m}{M+m}\right)^2 (A\vec{\varepsilon}^{(k-1)},\vec{\varepsilon}^{(k-1)}).$$

Iteriranjem poslednje nejednakosti dobijamo

$$(A\vec{\varepsilon}^{(k)},\vec{\varepsilon}^{(k)}) \leq \left(\frac{M-m}{M+m}\right)^{2k} (A\vec{\varepsilon}^{(o)},\vec{\varepsilon}^{(o)}),$$

odakle je, s obzirom na teoremu 1.4.6,

$$m\|\vec{\varepsilon}^{(k)}\|_E^2 \leq \left(\frac{M-m}{M+m}\right)^{2k} M\,\|\vec{\varepsilon}^{(o)}\|_E^2 ,$$

tj. (3.7.9).

Ovim je dokazana teorema 3.7.1.

Na kraju napomenimo da su metod najbržeg pada i optimalni metod (3.2.16) dosta slični. Principijelna razlika ovih metoda je u tome što metod najbržeg pada ne zahteva informaciju o granicama spektra matrice A, kao optimalni metod.

2. Metod konjugovanih gradijenata. U radovima [44] i [46] predložen je metod za rešavanje jednačine (3.7.1) koji teorijski (ne uzimajući u obzir greške zaokrugljivanja) konvergira ka tačnom rešenju u najviše n iteracija(n je red matrice). Ovaj metod se sastoji u konstrukciji niza $\{\vec{x}^{(k)}\}$ pomoću

$$\vec{x}^{(k)} = \vec{x}^{(o)} + \sum_{i=o}^{k-1} \alpha_i A^i \vec{r}^{(o)} \qquad (k=1,\dots,n),$$

polazeći od proizvoljnog početnog vektora $\vec{x}^{(o)}$, pri čemu je $\vec{r}^{(o)} = A\vec{x}^{(o)} - \vec{b}$, dok se koeficijenti α_i odredjuju iz uslova da funkcija $(\alpha_o,\dots,\alpha_{k-1}) \mapsto F(\vec{x}^{(k)})$ postiže minimum. Iz praktičnih razloga,kod primene ovog metoda dobro je konstruisati niz vektora $\{\vec{p}^{(k)}\}$, koji su medjusobno konjugovani u smislu

$$(\vec{p}^{(i)}, A\vec{p}^{(j)}) = 0 \qquad (i \neq j).$$

Tada se navedeni metod može iskazati rekurzivno pomoću sledećih formula:

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} - s_k \vec{p}^{(k)} \qquad (k=0,1,\dots,n-1),$$

$$\vec{r}^{(k)} = A\vec{x}^{(k)} - \vec{b} \qquad (k=0,1,\dots,n),$$

$$\vec{p}^{(o)} = -\vec{r}^{(o)}, \ \vec{p}^{(k)} = -\vec{r}^{(k)} + q_k \vec{p}^{(k-1)} \qquad (k=1,\dots,n-1),$$

$$s_k = \frac{(\vec{p}^{(k)}, \vec{r}^{(k)})}{(\vec{p}^{(k)}, A\vec{p}^{(k)})}, \quad q_k = \frac{(\vec{r}^{(k)}, A\vec{p}^{(k-1)})}{(\vec{p}^{(k-1)}, A\vec{p}^{(k-1)})} \qquad (k=0,1,\dots,n-1).$$

Kao što je napred rečeno, teorijski za neko $m \leq n$ biće $r^{(m)} = 0$, što znači da je odredjeno tačno rešenje sistema (3.7.1).

4.3.8. Iterativni metod za inverziju matrica

S obzirom da veliki broj metoda u numeričkoj analizi zahteva inverziju matrica, ili u opštem slučaju inverziju linearnih ograničenih operatora, ovaj odeljak posvećujemo ovom problemu. Sve rezultate koje ćemo ovde izneti, odnose se na inverziju matrica, ali se mogu formalno preneti i na inverziju linearnih ograničenih operatora(videti[29], [30]).

Pretpostavimo da je A regularna matrica reda n.

Teorema 3.8.1. Neka je r prirodan broj veći od jedinice i neka je

$$F_k = I - AX_k \qquad (k = 0,1,\ldots),$$

gde je X_0 data matrica takva da je

$$(3.8.1) \qquad \|F_0\| = \|I - AX_0\| \leq q < 1 .$$

Tada niz matrica $\{X_k\}$ definisan sa

$$(3.8.2) \qquad X_k = X_{k-1}(I + F_{k-1} + \ldots + F_{k-1}^{r-1}) \qquad (k = 1,2,\ldots),$$

konvergira ka A^{-1}.

Dokaz. Na osnovu

$$\begin{aligned} F_k &= I - AX_k \\ &= I - AX_{k-1}(I + F_{k-1} + \ldots + F_{k-1}^{r-1}) \\ &= I - (I - F_{k-1})(I + F_{k-1} + \ldots + F_{k-1}^{r-1}) \\ &= F_{k-1}^{r} , \end{aligned}$$

dobijamo

$$F_k = F_0^{r^k} \qquad (k = 0,1,\ldots).$$

Iz poslednje jednakosti i uslova (3.8.1) redom sleduje

$$\|F_k\| \leq \|F_0\|^{r^k} \leq q^{r^k} \qquad (k = 0,1,\ldots)$$

i

$$\lim_{k\to+\infty} \|F_k\| = 0 ,$$

odakle zaključujemo da je

$$\lim_{k\to+\infty} F_k = \lim_{x\to+\infty} (I - AX_k) = 0, \quad \text{tj.} \quad \lim_{x\to+\infty} X_k = A^{-1}.$$

Teorema 3.8.2. Ako su ispunjeni uslovi teoreme 3.8.1, važe sledeće nejednakosti

$$(3.8.3)\qquad \|X_k - A^{-1}\| \leqq \frac{\|X_k F_k\|}{1 - \|F_k\|},$$

$$(3.8.4)\qquad \|X_k - A^{-1}\| \leqq \|F_{k-1}\|^{r-1} \frac{\|X_{k-1} F_{k-1}\|}{1 - \|F_{k-1}\|},$$

$$(3.8.5)\qquad \|X_k - A^{-1}\| \leqq \|F_o\|^{r^k} \frac{\|X_o\|}{1 - \|F_o\|}$$

za svako $k \in N$.

Dokaz. Ako stavimo, $E_k = A^{-1} - X_k$, imamo

$$E_k - E_k F_k = (A^{-1} - X_k)AX_k = X_k(I - AX_k) = X_k F_k .$$

Kako je $\|F_k\| < 1$, važi nejednakost

$$\|E_k\|(1 - \|F_k\|) \leqq \|E_k - E_k F_k\| = \|X_k F_k\|,$$

odakle neposredno sleduje (3.8.3).

Nejednakosti (3.8.4) i (3.8.5) dokazuju se na sličan način(videti[30]).

Teorema 3.8.3. Ako su ispunjeni uslovi teorema 3.8.1, iterativni proces (3.8.2) ima red konvergencije r.

Dokaz. Ako u nejednakosti (3.8.3), tj. nejednakosti

$$\|X_{k+1} - A^{-1}\| \leqq \frac{\|X_{k+1} F_{k+1}\|}{1 - \|F_{k+1}\|}$$

uvedemo smenu

$$F_{k+1} = F_k^r = (I - AX_k)^r = A^r (A^{-1} - X_k)^r$$

dobijamo

$$\|X_{k+1} - A^{-1}\| \leqq \frac{\|X_{k+1} A^r (A^{-1} - X_k)^r\|}{1 - \|F_{k+1}\|} \leqq \frac{\|X_{k+1}\| \cdot \|A\|^r}{1 - \|F_{k+1}\|} \|X_k - A^{-1}\|^r .$$

Kako je

$$\lim_{k \to +\infty} \frac{\|X_{k+1}\| \cdot \|A\|^r}{1 - \|F_{k+1}\|} = \|A^{-1}\| \cdot \|A\|^r < +\infty ,$$

imamo

$$\|X_{k+1} - A^{-1}\| = O\left(\|X_k - A^{-1}\|^r\right) \qquad (k \to +\infty),$$

što znači da iterativni proces (3.8.2) za inverziju matrice A, ima red konvergencije r, čime je dokaz teoreme 3.8.3 završen.

U specijalnom slučaju za r=2, (3.8.2) se svodi na

$$X_k = X_{k-1}(2I - AX_{k-1}) \qquad (k = 1,2,\dots).$$

Ovaj iterativni proces sa kvadratnom konvergencijom potiče od G.Schulza([31]), dok se ocene za grešku, analogne onim u teoremi 3.8.2, mogu naći u radovima [32], [33], [34], [35].

U slučaju kada je r=3, odgovarajući iterativni proces sa kubnom konvergencijom je

$$X_k = X_{k-1}(3I - 3AX_{k-1} + (AX_{k-1})^2) \qquad (k = 1,2,\dots).$$

U radu [32], J.Albrecht je ukazao na vezu ovog procesa sa Newtonovim metodom za rešavanje nelinearnnih jednačina(videti poglavlje 5.1).

4.4. LITERATURA

1. D.YOUNG: Iterative methods for solving partial difference equations of elliptic type. Trans.Amer.Math.Soc. 76(1954), 92-111.
2. D.S.MITRINOVIĆ - D.Ž.DJOKOVIĆ: Polinomi i matrice. Beograd, 1975.
3. L.COLLATZ: Functionalanalysis und numerische Mathematik. Berlin-Heidelberg-New York, 1964.
4. В.И.КРЫЛОВ, В.В.БОБКОВ, П.И.МОНАСТЫРНЫЙ: Вычислительные методы высшей математики. Том 1. Минск 1972.
5. WILKINSON: The Algebraic Eigenvalue Problem. Oxford, 1965.
6. V.STRASSEN: Gaussian elimination is not optimal. Numer.Math.13(1969), 354-356.
7. P.B.MADIĆ: Loši rešljivi sistemi linearnih algebarskih jednačina i njihovo rešavanje (Doktorska disertacija), Beograd, 1965.
8. M.A.KRASNOSEL'SKII, S.G.KREIN: Remark on the distribution of errors in the solution of a system of linear equations by means of an iterative process (Russian). Uspehi Matem.Nauk. 7(1952), 157-161.
9. D.G.PERADZE: The distribution of errors in solving a system of linear algebraic equations by the iteration method (Russian). Sakharth.SSR Mecn.Akad.Moambe. 50(1968), 289-294.
10. T.YAMAMOTO: On the distribution of errors in the iterative solution of a system of linear equations. Numer.Math. 24(1975), 71-79.
11. T.YAMAMOTO: On the distribution of errors in the iterative solution of a system of linear equations II. Ibid. 25(1976), 461-463.
12. И.М.СОБОЛЬ: Численные методы Монте-Карло. Москва 1973.
13. A.S.HAUSEHOLDER: The approximate solution of matrix problems. J.Assoc. Comp.Mach. 5(1958), 205-243.
14. Н.С.БАХВАЛОВ: Численные методы. Москва 1973.
15. М.ПЕТКОВ: Числени методи на алгебрата. София 1974.

16. E.REICH: On the Convergence of the Classical Iterative Method of Solving Linear Simultaneous Equations. Ann.Math.Statist. 20(1949), 448-451.

17. L.COLLATZ: Über die Konvergenzkriterion bei Iterationverfahren für lineare Gleichungsysteme. Math. Z. 53(1950), 149-161.

18. D.M.YOUNG, R.T.GREGORY: A Survey of Numerical Mathematics. Addison-Wesley Publ.Comp. 1973.

19. Б.П.ДЕМИДОВИЧ, И.А.МАРОН: Основы вычислительной математики.Москва, 1966.

20. R.V.SOUTHWELL: Relaxation Methods in Theoretical Physics. 2 vols. Oxford University Press. Fair Lawn. N.J. 1956.

21. N.APOSTOLATOS, U.KULISCH: Uber die Konvergenz das Relaxationsverfahrens bei nicht-negativen und diagonaldominanten Matrizen. Computing 2(1967),17-24.

22. D.M.YOUNG: Iterative Solution of Large Linear Systems. New York 1971.

23. J.ALBRECHT: Fehlerabschätzungen bei Relaxationsverfahren zur numerischen Auflösung linearer Gleichungssysteme. Numer.Math.3(1961), 188-201.

24. R.S.VARGA: Iterative Numerical Analysis. Pittsburgh 1959.

25. G.H.GOLUB and R.S.VARGA: Chebyshev semi-iterative methods, successive overrelaxation iterative methods and second order Richardson iterative methods. Numer.Math. 3(1961), 147-168.

26. G.E.FORSYTHE: Solving linear algebraic equations can be interesting. Bull. Amer.Math.Soc. 59(1953), 299-329.

27. C.LANSZOS: Solution of systems of linear equations by minimized iterations. J.Res.Nat.Bur.Standards. 49(1952), 33.

28. E.STIEFEL: Relaxationsmethoden bester Strategie zur Lösung linear Gleichungssysteme. Comm.Math.Helv.29(1955), 157-179.

29. M.ALTMAN: An optimum cubically convergent iterative method of inverting a a linear bounded operator in Hilbert space. Pacific. J.Math.10(1960), 1107-1113.

30. W.V.PETRYSHYN: On the inversion of matrices and linear operators. Proc.Amer. Math.Soc. 16(1965), 893-901.

31. G.SCHULZ: Iterative Berechnung der reziproken Matrix. ZAMM 13(1933), 57-59.

32. J.ALBRECHT: Bemerkunden zum Iterationsverfahren von Schulz zur Matrixinversion. ibid. 41(1961), 262-263.

33. W.DUCK: Fehlerabschätzungen für das Iterationsverfahren von Schulz zur Bestimmung der Inversen einer Matrix. ibid. 40(1960), 192-194.

34. H.HOTELLING: Some new methods in matrix calculation. Amer.Math.Statist. 14(1943), 1-34.

35. R.ANSORGE: Uber ein Iterationsverfahren von G.Schultz zur Ermittlung der Reziproken einer Matrix. Z.Angew.Math.Mech. 39(1959), 164-165.

36. D.A.FLANDERS and G.SHORTLEY: Numerical determination of fundamental modes. Journal of Applied Physics. 21(1950), 1326-1332).

37. J.ALBRECHT: Monotone Iteration und ihre Verwendung zur Lösung linearer Gleichungssysteme. Numer.Math. 3(1961), 345-358.

38. U.KULISCH: Uber reguläre Zerlegungen von Matrizen einige Anwendungen. ibid. 11(1968), 444-449.

39. S.SCHECHTER: Relaxation methods for linear equations. Comm.Pure and Appl.Math. 12(1959), 313-335.

40. F.B.HILDEBRAND: Introduction to Numerical Analysis. New York-London-Toronto, 1974.

41. G.FORSYTHE and C.B.MOLER: Computer Solution of Linear Algebraic Systems. New Jersey, 1967.

42. A.RALSTON: A First Course in Numerical Analysis. New York, 1965.

43. Å.BJORCK und G.DAHLQUIST: Numerische Methoden. München-Wien, 1972.

44. M.R.HESTENES, E.STIEFEL: Methods of conjugate gradients for solving linear systems. J.Res.Nat.Bur.Standards. 49(1952), 409-436.

45. A.S.HAUSEHOLDER: The Theory of Matrices in Numerical Analysis. New York-Toronto-London, 1965.

46. G.W.STEWART: Conjugate direction methods for solving systems of linear equations. Numer.Math. 21(1973), 285-297.

47. J.STOER: Einführung in die Numerische Mathematik I. Berlin-Heidelberg--New York, 1972.

48. J.STOER, R.BULIRSCH: Einführung in die Numerische Mathematik II. Berlin-Heidelberg-New York, 1973.

49. В.А.ВЕРГАСОВ, И.Г.ЖУРКИН, М.В.КРАСИКОВА и др.: Вычислительная математика. Москва 1976.

5. NELINEARNE JEDNAČINE I SISTEMI

5.1. NELINEARNE JEDNAČINE

5.1.1. Osnovne napomene

U trećoj glavi, a posebno u odeljku 3.1.3, data je opšta teorija o egzistenciji i jedinstvenosti rešenja nelinearne jednačine

$$f(x) = 0, \tag{1.1.1}$$

kao i o iterativnim procesima za nalaženje ovog rešenja. Ovo poglavlje biće posvećeno konkretnim metodima koji se najčešće primenjuju za rešavanje jednačine (1.1.1), bez obzira da li je ona algebarska ili transcendentna. Zbog specifičnosti koje poseduju algebarske jednačine i zbog važnosti koje one u primenama imaju, u literaturi je razvijen ogroman broj metoda za njihovo numeričko rešavanje. U našem izlaganju ovom problemu posvećujemo posebno poglavlje.

Svi metodi koji će biti izloženi u ovom poglavlju primenjuju se na odredjivanje izolovanih korena jednačine (1.1.1). O problemu izolovanja korena nelinearnih jednačina videti, na primer, [1], [2], [3].

5.1.2. Newtonov metod

Newtonov ili Newton-Raphsonov metod kako se često naziva, predstavlja osnovni metod za nalaženje izolovanih korena nelinearnih jednačina.

Neka je na segmentu $[\alpha,\beta]$ izolovan jedinstven koren $x=a$ jednačine (1.1.1) i neka $f\in C^1[\alpha,\beta]$ i $f'(x)\neq 0$ $(\forall x\in[\alpha,\beta])$. Izaberimo tačku $x_0\in[\alpha,\beta]$. Tada, na osnovu Taylorove formule, imamo

$$f(a) = f(x_0) + f'(x_0)(a-x_0) + \frac{1}{2}f''(\xi)(a-x_0)^2, \tag{1.2.1}$$

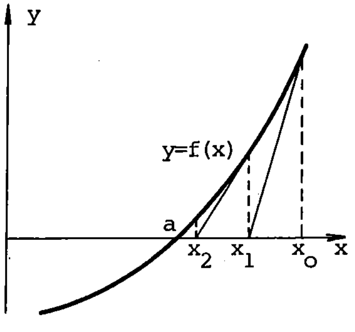
gde je $\xi = x_0 + \theta(a-x_0)$ $(0<\theta<1)$. S obzirom da je $f(a)=0$, zanemarivanjem poslednjeg člana na desnoj strani u jednakosti (1.2.1), dobijamo

$$a = x_o - \frac{f(x_o)}{f'(x_o)} .$$

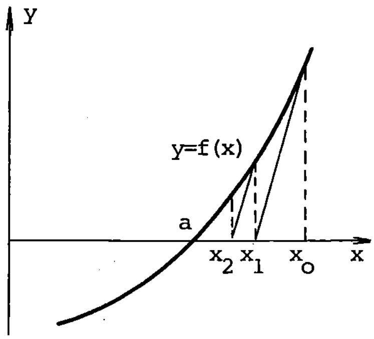
Sa x_1 označimo desnu stranu u poslednjoj približnoj jednakosti, tj.

$$x_1 = x_o - \frac{f(x_o)}{f'(x_o)} . \tag{1.2.2}$$

Geometrijski x_1 predstavlja apscisu tačke preseka tangente na krivu $y = f(x)$, u tački $(x_o, f(x_o))$, sa x-osom(videti sl. 1.2.1).



Sl. 1.2.1



Sl. 1.2.2

Jednakost (1.2.2) sugeriše konstrukciju iterativnog procesa

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \qquad (k = 0,1,\ldots), \tag{1.2.3}$$

koji je poznat kao Newtonov metod ili metod tangente.

Predjimo sada na ispitivanje konvergencije iterativnog procesa (1.2.3), uvodeći dopunsku pretpostavku za funkciju f. Naime, pretpostavimo da $f \in C^2[\alpha, \beta]$.

Kako je iterativna funkcija Φ, kod Newtonovog metoda, odredjena sa

$$\phi(x) = x - \frac{f(x)}{f'(x)},$$

diferenciranjem dobijamo

$$(1.2.4) \qquad \phi'(x) = 1 - \frac{f'(x)^2 - f(x)f''(x)}{f'(x)^2} = \frac{f(x)f''(x)}{f'(x)^2}.$$

Primetimo da je $\phi(a) = a$ i $\phi'(a) = 0$. Kako je na osnovu učinjenih pretpostavki za f, funkcija ϕ' neprekidna na $[\alpha, \beta]$ i kako je $\phi'(a) = 0$, to postoji okolina tačke $x = a$, u oznaci $U(a)$, u kojoj je

$$(1.2.5) \qquad |\phi'(x)| = \left|\frac{f(x)f''(x)}{f'(x)^2}\right| \leqq q < 1.$$

Teorema 1.2.1. Ako $x_0 \in U(a)$, niz $\{x_k\}_{k \in N_0}$ koji se generiše pomoću(1.2.3) konvergira ka tački $x = a$, pri čemu je

$$(1.2.6) \qquad \lim_{k \to +\infty} \frac{x_{k+1} - a}{(x_k - a)^2} = \frac{f''(a)}{2f'(a)}.$$

Dokaz. Iz (1.2.3) sleduje

$$x_{k+1} - a = x_k - a - \frac{f(x_k) - f(a)}{f'(x_k)},$$

tj.

$$f(x_k) - f(a) = f'(x_k)(x_k - a) - f'(x_k)(x_{k+1} - a).$$

S druge strane, na osnovu Taylorove formule, imamo

$$f(a) - f(x_k) = f'(x_k)(a - x_k) + \frac{1}{2} f''(\xi_k)(a - x_k)^2,$$

gde je ξ_k neka tačka izmedju x_k i a.

Ako poslednje dve jednakosti saberemo dobijamo

$$0 = -f'(x_k)(x_{k+1} - a) + \frac{1}{2} f''(\xi_k)(a - x_k)^2.$$

Kako je, na osnovu učinjene pretpostavke, $f'(x_k) \neq 0$, iz poslednje jednakosti sleduje

$$(1.2.7) \qquad \frac{x_{k+1} - a}{(x_k - a)^2} = \frac{f''(\xi_k)}{2f'(x_k)}.$$

Da bismo dokazali konvergenciju niza $\{x_k\}_{k\in N_0}$ dovoljno je primetiti da Φ preslikava U(a) u U(a). Tada, imajući u vidu (1.2.5), vidimo da Φ zadovoljava uslove teoreme 1.3.1 iz odeljka 3.1.3, odakle sleduje konvergencija iterativnog procesa (1.2.3), pri proizvoljnom $x_0 \in U(a)$.

Kako $x_k \to a$, kada $k \to +\infty$, i kako je f'' neprekidna funkcija, iz (1.2.7) sleduje (1.2.6), čime je dokaz teoreme završen.

Primer 1.2.1. Nadjimo rešenje jednačine

$$f(x) = x - \cos x = 0$$

na segmentu $[0,\pi/2]$ primenom Newtonovog metoda

$$x_{k+1} = x_k - \frac{x_k - \cos x_k}{1+\sin x_k} = \frac{x_k \sin x_k + \cos x_k}{1+\sin x_k} \qquad (k=0,1,\dots).$$

Primetimo da je $f'(x) = 1+\sin x > 0$ $(\forall x \in [0,\pi/2])$. Startujući sa $x_0 = 1$, kao i u primeru 2.2.1 iz odeljka 3.2.2, dobijamo

k	x_k
0	1.
1	0.750364
2	0.739133
3	0.739085
4	0.739085

Poslednje dve iteracije daju rešenje posmatrane jednačine sa šest tačnih decimala.

Primer 1.2.2. Primenom Newtonovog metoda na rešavanje jednačine $f(x) = x^n - a = 0$ $(a>0, n>1)$ dobijamo iterativnu formulu za odredjivanje n-tog korena iz pozitivnog broja a

$$x_{k+1} = x_k - \frac{x_k^n - a}{n x_k^{n-1}} = \frac{1}{n}\left[(n-1)x_k + \frac{a}{x_k^{n-1}}\right] \qquad (k=0,1,\dots).$$

Specijalan slučaj ove formule, za $n=2$, naveden je u prvoj glavi ove knjige.

Kod primene Newtonovog metoda često se nameće pitanje kako izabrati početnu vrednost x_0 pa da niz $\{x_k\}_{k\in N_0}$ bude monoton. Jedan odgovor na ovo pitanje dao je Fourier°. Naime, ako f'' ne menja znak na $[\alpha,\beta]$ i ako se x_o izabere tako da je $f(x_o)f''(x_o) > 0$, niz $\{x_k\}_{k\in N_o}$ biće monoton. Ovo tvrdjenje sleduje iz (1.2.4).

Na osnovu teoreme 1.2.1 zaključujemo da Newtonov metod primenjen na odredjivanje prostog korena $x = a$ ima kvadratnu konvergenciju ako je $f''(a) \neq 0$. U tom slučaju je faktor konvergencije(asimptotska konstanta greške)

$$C_2 = \left|\frac{f''(a)}{2f'(a)}\right| .$$

Slučaj $f''(a) = 0$ treba posebno analizirati. Naime, ako pretpostavimo da $f \in C^3[\alpha,\beta]$, može se dokazati

$$\lim_{k\to+\infty} \frac{x_{k+1} - a}{(x_k - a)^3} = \frac{f'''(a)}{3f'(a)} .$$

<u>Primer 1.2.3</u>. Posmatrajmo jednačinu

$$f(x) = x^3 - 3x^2 + 4x - 2 = 0.$$

Kako je $f(o) = -2$ i $f(1.5) = 0.625$ zaključujemo da na segmentu $[0,1.5]$ data jednačina ima koren. S druge strane, $f'(x) = 3x^2 - 6x + 4 = 3(x-1)^2 + 1 > 0$, što znači da je koren prost, pa možemo primeniti Newtonov metod.

Startujući sa $x_o = 1.5$, dobijamo

k	x_k
0	1.5
1	1.1428571
2	1.0054944
3	1.0000003.

Tačna vrednost korena je $a = 1$, s obzirom da je $f(x) = (x-1)^3 + (x-1)$. Primetimo da je u ovom slučaju $f''(1) = 0$. Ako stavimo $e_k = x_k - a$ $(k = o,1,\ldots)$ imamo

$$e_{k+1} = \frac{2e_k^3}{1 + 3e_k^2} \qquad (k = o,1,\ldots),$$

tj. $e_{k+1} \sim 2e_k^3 \quad (k \to +\infty)$.

Zaista, u ovom slučaju je $\frac{f'''(1)}{3f'(1)} = 2$.

U cilju smanjivanja broja računskih operacija, često se koristi sledeća modifikacija Newtonovog metoda

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_o)} \qquad (k = o,1,\ldots).$$

Geometrijski x_{k+1} predstavlja apscisu tačke preseka x-ose sa pravom, koja prolazi kroz tačku $(x_k, f(x_k))$ i koja je paralelna sa tangentom krive $y = f(x)$ postavljene u tački $(x_o, f(x_o))$ (videti sl.1.2.2).

Iterativna funkcija modifikovanog Newtonovog metoda je

$$\Phi_1(x) = x - \frac{f(x)}{f'(x_o)} .$$

Kako je $\Phi_1(a) = a$ i $\Phi_1'(a) = 1 - \frac{f'(a)}{f'(x_o)}$, zaključujemo da metod ima red konvergencije jedan, tj. važi

$$x_{k+1} - a \sim \left(1 - \frac{f'(a)}{f'(x_o)}\right)(x_k - a) \qquad (k \to +\infty),$$

pri čemu je uslov

$$\left|1 - \frac{f'(x)}{f'(x_o)}\right| \leqq q < 1 ,$$

analogon uslovu (1.2.5).

Newtonov metod može se razmatrati i kao specijalan slučaj tzv. uopštenog Newtonovog metoda

$$(1.2.8) \qquad x_{k+1} = x_k - \frac{\Psi(x_k) f(x_k)}{\Psi'(x_k) f(x_k) + \Psi(x_k) f'(x_k)} \qquad (k = o,1,\ldots),$$

gde je Ψ data diferencijabilna funkcija.

Za $\Psi(x) = 1$, (1.2.8) se svodi na Newtonov metod (1.2.3).

Za $\Psi(x) = x^p$, gde je p rarametar, iz (1.2.8) sleduje formula

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k) + \frac{p}{x_k} f(x_k)} \qquad (k = o,1,\ldots),$$

tj.

$$(1.2.9) \qquad x_{k+1} = x_k\left(1 - \frac{f(x_k)}{x_k f'(x_k) + pf(x_k)}\right) \qquad (k = o,1,\ldots).$$

Metod definisan formulom (1.2.9) razmatran je u radu [4]. Specijalni slučaj ovog metoda, za $p = 1 - n$, poznat je kao metod Tihonova([5]), **u slučaju kada je f algebarski polinom stepena n.**

Na kraju navedimo još jednu modifikaciju Newtonovog metoda, koja se sastoji u sukcesivnoj primeni formula

$$y_k = x_k - \frac{f(x_k)}{f'(x_k)}, \quad x_{k+1} = y_k - \frac{f(y_k)}{f'(x_k)} \qquad (k = o,1,\ldots).$$

Slično dokazu teoreme 1.2.1 mogu se dokazati tvrdjenja

$$\lim_{k\to+\infty} \frac{x_{k+1} - a}{(y_k - a)(x_k - a)} = \frac{f''(a)}{f'(a)}$$

i

$$\lim_{k\to+\infty} \frac{x_{k+1} - a}{(x_k - a)^3} = \frac{1}{2}\left(\frac{f''(a)}{f'(a)}\right)^2,$$

gde je $\lim_{k\to+\infty} x_k = a$.

5.1.3. Newtonov metod za višestruke korene

Posmatrajmo jednačinu $f(x) = 0$, koja na $[\alpha,\beta]$ ima koren $x = a$ višestrukosti m (≥ 2). Ako pretpostavimo da $f \in C^{m+1}[\alpha,\beta]$, tada je

$$f(a) = f'(a) = \ldots = f^{(m-1)}(a) = 0, \quad f^{(m)}(a) \neq 0.$$

Naime, u ovom slučaju, f se može predstaviti u obliku

$$(1.3.1) \qquad f(x) = (x-a)^m g(x),$$

gde $g \in C^{m+1}[\alpha, \beta]$ i $g(a) \neq 0$.

Iz (1.3.1) sleduje

$$f'(x) = m(x-a)^{m-1}g(x) + (x-a)^m g'(x)$$

i

$$\Delta(x) = \frac{f(x)}{f'(x)} = \frac{(x-a)g(x)}{mg(x) + (x-a)g'(x)} \qquad (x \neq a).$$

Ako stavimo $\Delta(a) \overset{\text{def}}{=} \lim\limits_{x \to a} \Delta(x)$, tada je $\Delta(a) = 0$.

Iterativna funkcija Newtonovog metoda, primenjenog na odredjivanje višestrukog korena, na osnovu prethodnog postaje

$$\phi(x) = x - \frac{(x-a)g(x)}{mg(x) + (x-a)g'(x)}.$$

Kako je $\phi(a) = a$, $\phi'(a) = 1 - \frac{1}{m}$, $\frac{1}{2} \leqq \phi'(a) < 1$ $(m \geqq 2)$ i ϕ' neprekidna funkcija, izlazi da postoji okolina korena $x = a$ u kojoj je $|\phi'(x)| \leqq q < 1$, odakle zaključujemo da je Newtonov metod i u ovom slučaju konvergentan, ali sa redom konvergencije jedan.

Ukoliko nam je unapapred poznat red višestrukosti korena, tada se Newtonov metod može modifikovati tako da ima red konvergencije dva. Naime, treba samo uzeti

$$(1.3.2) \qquad x_{k+1} = x_k - m\frac{f(x_k)}{f'(x_k)} \qquad (k = 0, 1, \ldots).$$

<u>Primedba 1.3.1.</u> Formalno, formula (1.3.2) predstavlja Newtonov metod primenjen na rešavanje jednačine

$$F(x) = \sqrt[m]{f(x)} = 0.$$

<u>Teorema 1.3.1.</u> Ako je x_0 izabrano dovoljno blisko korenu $x = a$, čiji je red višestrukosti m, tada niz $\{x_k\}_{k \in N_0}$ definisan pomoću (1.3.2) konvergira ka a, pri čemu je

$$\frac{x_{k+1} - a}{(x_k - a)^2} \sim \frac{1}{m(m+1)} \cdot \frac{f^{(m+1)}(a)}{f^{(m)}(a)} \qquad (k \to +\infty).$$

Dokaz ove teoreme može se naći, na primer, u [6].

Ukoliko red višestrukosti m nije poznat, tada se umesto jednačine $f(x)=0$ može rešavati jednačina $\frac{f(x)}{f'(x)}=0$, čiji su svi koreni prosti. Newtonov metod primenjen na ovu jednačinu daje formulu

$$x_{k+1}=x_k-\left[\frac{\frac{f(x)}{f'(x)}}{\left(\frac{f(x)}{f'(x)}\right)'}\right]_{x=x_k} \qquad (k=0,1,\ldots),$$

tj. formulu

$$x_{k+1}=x_k-\frac{f(x_k)f'(x_k)}{f'(x_k)^2-f(x_k)f''(x_k)} \qquad (k=0,1,\ldots),$$

čiji je red konvergencije dva.

5.1.4. Metod sečice

Ako se u Newtonovom metodu vrednost izvoda $f'(x_k)$ aproksimira pomoću podeljene razlike $\frac{f(x_k)-f(x_{k-1})}{x_k-x_{k-1}}$ dobija se metod sečice

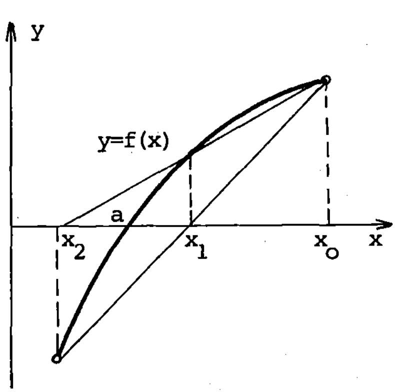
$$x_{k+1}=x_k-\frac{x_k-x_{k-1}}{f(x_k)-f(x_{k-1})}f(x_k) \qquad (k=1,2,\ldots), \tag{1.4.1}$$

koji pripada klasi metoda oblika (2.1.2)(odeljak 3.2.1). Za startovanje iterativnog procesa (1.4.1) potrebne su dve početne vrednosti x_0 i x_1. Geometrijska interpretacija metoda sečice data je na sl.1.4.1.

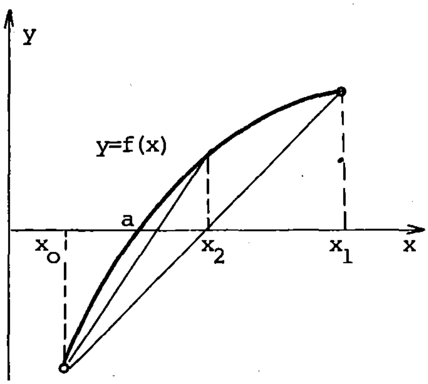
Neka na segmentu $[\alpha,\beta]$ postoji jedinstven koren $x=a$ jednačine $f(x)=0$. Za ispitivanje konvergencije iterativnog procesa (1.4.1) pretpostavimo da $f\in C^2[\alpha,\beta]$ i $f(x)\neq 0\,(\forall x\in[\alpha,\beta])$.

Ako stavimo $e_k=x_k-a$ $(k=0,1,\ldots)$, iz (1.4.1) sleduje

$$e_{k+1}=e_k-\frac{e_k-e_{k-1}}{f(x_k)-f(x_{k-1})}f(x_k). \tag{1.4.2}$$



Sl. 1.4.1



Sl. 1.4.2

Kako je

$$f(x_k) = f'(a)e_k + \frac{1}{2} f''(a)e_k^2 + O(e_k^3)$$

i

$$\frac{f(x_k) - f(x_{k-1})}{e_k - e_{k-1}} = f'(a) + \frac{1}{2}(e_k + e_{k-1})f''(a) + O(e_{k-1}^2),$$

zamenom u (1.4.2) dobijamo

$$e_{k+1} = e_k \left[1 - \frac{f'(a) + \frac{1}{2}e_k f''(a) + O(e_k^2)}{f'(a) + \frac{1}{2}(e_k + e_{k-1})f''(a) + O(e_{k-1}^2)} \right],$$

odakle je

$$e_{k+1} = e_k \left[1 - \left(1 + \frac{1}{2}e_k f''(a) + O(e_k^2)\right)\left(1 - \frac{1}{2}(e_k + e_{k-1})f''(a) + O(e_{k-1}^2)\right)\right],$$

tj.

$$e_{k+1} = e_k e_{k-1} \frac{f''(a)}{2f'(a)} (1 + O(e_{k-1})). \tag{1.4.3}$$

Da bismo odredili red konvergencije i faktor konvergencije stavimo

$$|e_{k+1}| = C_r |e_k|^r |1 + O(e_k)| . \tag{1.4.4}$$

Tada na osnovu (1.4.3) i (1.4.4) dobijamo

$$|e_{k+1}| = C_r |e_k|^r |1 + O(e_k)| = C_r (C_r |e_{k-1}|^r)^r |1 + O(e_{k-1})|$$

$$= C_r |e_{k-1}|^r |e_{k-1}| \cdot \left|\frac{f''(a)}{2f'(a)}\right| \cdot |1 + O(e_{k-1})| ,$$

odakle sleduje

$$r^2 - r - 1 = 0 \qquad \text{i} \qquad C_r = \left|\frac{f''(a)}{2f'(a)}\right|^{1/r} .$$

Red konvergencije r dobijamo kao pozitivno rešenje dobijene kvadratne jednačine, tj. $r = \frac{1}{2}(1 + \sqrt{5}) \cong 1.62$. Faktor konvergencije je

$$C_r = \left|\frac{f''(a)}{2f'(a)}\right|^{(\sqrt{5}-1)/2} .$$

Primedba 1.4.1. U literaturi se za rešavanje jednačine

$$x = g(x) \tag{1.4.5}$$

sreće metod Weigsteina ([7]), kod koga se polazeći od x_0 generiše niz $\{x_k\}_{k \in N}$ pomoću

$$x_1 = g(x_0),$$

$$x_{k+1} = g(x_k) - \frac{(g(x_k) - g(x_{k-1}))(g(x_k) - x_k)}{(g(x_k) - g(x_{k-1})) - (x_k - x_{k-1})} \qquad (k = 1, 2, \ldots). \tag{1.4.6}$$

U radu [8] je pokazano da je ovaj metod ustvari metod sečice sa početnim vrednostima x_0 i $x_1 = g(x_0)$. Naime, ako jednačinu (1.4.5) predstavimo u obliku

$$f(x) = g(x) - x = 0, \tag{1.4.7}$$

smenom (1.4.7) u (1.4.6) dobijamo (1.4.1).

Metod sečice može se modifikovati tako da je

$$x_{k+1} = x_k - \frac{x_k - x_0}{f(x_k) - f(x_0)} f(x_k) \qquad (k = 1, 2, \ldots). \tag{1.4.8}$$

Ovaj metod se često naziva metod regula falsi. Za razliku od metoda sečice, gde je dovoljno uzeti $x_1 \neq x_0$, kod ovog metoda x_1 i x_0 treba uzeti sa različitih strana u odnosu na koren $x = a$. Geometrijska interpretacija metoda regula falsi data je na sl.1.4.2.

Iterativna funkcija kod modifikovanog metoda sečice je

$$\Phi(x) = x - \frac{x - x_0}{f(x) - f(x_0)} f(x) = \frac{x_0 f(x) - x f(x_0)}{f(x) - f(x_0)} .$$

Ako pretpostavimo da $f \in C^1[\alpha,\beta]$, tada je

$$\Phi'(x) = \frac{f(x_0)}{f(x) - f(x_0)} \left[\frac{x - x_0}{f(x) - f(x_0)} f'(x) - 1 \right] .$$

Kako je $\Phi(a) = a$ i $\Phi'(a) \neq 0$, zaključujemo da iterativni proces (1.4.8), ukoliko je konvergentan, ima red konvergencije jedan. Uslov konvergencije, u ovom slučaju, dat je pomoću

$$|\Phi'(x)| \leqq q < 1 \qquad (f(x) \neq f(x_0)),$$

za svako $x \in [\alpha,\beta] \setminus \{x_0\}$.

Primer 1.4.1. Primenom teoreme 2.4.1(odeljak 3.2.4) na iterativni proces (1.4.8) dobijamo iterativni proces drugog reda

$$x_{k+1} = \frac{x_0 g(x_k) - x_k h(x_k)}{g(x_k) - h(x_k)} \qquad (k = 1,2,\ldots),$$

gde su $g(x) = \dfrac{f(x) - f(x_0)}{x - x_0}$ i $h(x) = \dfrac{f'(x) f(x_0)}{f(x)}$.

Primedba 1.4.2. Ako se izvod $f'(x_k)$ u Newtonovom metodu zameni konačnom razlikom u tački x_k, sa korakom $h = f(x_k)$, tj.

$$f'(x_k) \cong \frac{f(x_k + f(x_k)) - f(x_k)}{f(x_k)},$$

dobija se metod Steffensena

(1.4.9) $$x_{k+1} = x_k - \frac{f(x_k)^2}{f(x_k+f(x_k)) - f(x_k)} \qquad (k=0,1,\ldots).$$

Neka je $x=a$ jedinstven prost koren jednačine $f(x)=0$ na segmentu $[\alpha,\beta]$ i neka $f\in C^2[\alpha,\beta]$.

Metod Steffensena je interesantan jer ima red konvergencije dva, a da pritom iterativna funkcija

$$x \mapsto \varphi(x) = x - \frac{f(x)^2}{f(x+f(x)) - f(x)}$$

ne sadrži izvod f'. Da bismo odredili asimptotsku konstantu greške metoda (1.4.9) podjimo od Taylorove formule

$$f(x+f(x)) = f(x) + f'(x)f(x) + \frac{1}{2}f''(\xi)f(x)^2,$$

gde je $\xi=x+\theta f(x)$ $(0<\theta<1)$. Tada je

$$\varphi(x) = x - \frac{f(x)/f'(x)}{1+\frac{1}{2}f''(\xi)f(x)/f'(x)}.$$

S obzirom da postoji okolina $U(a)$, tačke $x=a$, u kojoj je $\left|\frac{1}{2}f''(\xi)f(x)/f'(x)\right| < 1$, imamo

$$\varphi(x) = x - \frac{f(x)}{f'(x)}\left(1 - \frac{1}{2}\cdot\frac{f''(\xi)}{f'(x)}f(x) + O(f(x)^2)\right) \qquad (x\in U(a)).$$

Kako je $\dfrac{f(x)}{f'(x)} = x-a - \dfrac{f''(a)}{2f'(a)}(x-a)^2 + O((x-a)^3)$ $(x\in U(a))$,

iz poslednje jednakosti sleduje

$$\varphi(x) - a \sim \frac{f''(a)}{2f'(a)}(f'(a)+1)(x-a)^2 \qquad (x\to a),$$

što znači da je asimptotska konstanta greške $C_2 = \left|\dfrac{f''(a)}{2f'(a)}(f'(a)+1)\right|$.

5.1.5. Metod polovljenja intervala

Neka je na segmentu $[\alpha,\beta]$ izolovan prost koren $x=a$ jednačine

(1.5.1) $$f(x)=0,$$

gde $f\in C[\alpha,\beta]$. Metod polovljenja intervala za rešavanje jednačine (1.5.1) sastoji se u konstrukciji niza intervala $\{(x_k,y_k)\}_{k\in N}$ takvog da je

$$y_{k+1}-x_{k+1}=\frac{1}{2}(y_k-x_k) \qquad (k=1,2,\ldots)$$

i

$$\lim_{k\to+\infty} x_k = \lim_{k\to+\infty} y_k = a.$$

Navedeni proces konstrukcije intervala se prekida, na primer, kada dužina intervala postane manja od unapred zadatog malog pozitivnog broja ε.

Metod polovljenja intervala je vrlo prost i može se iskazati kroz sledeća četiri koraka:

1° $k:=0$, $x_1:=\alpha$, $y_1:=\beta$;

2° $k:=k+1$, $z_k:=\frac{1}{2}(x_k+y_k)$;

3° Ako je

$$f(z_k)f(x_k)\ \begin{cases} <0 & \text{uzeti} \quad x_{k+1}:=x_k,\ y_{k+1}:=z_k, \\ >0 & \qquad\quad x_{k+1}:=z_k,\ y_{k+1}:=y_k, \\ =0 & \qquad\quad \text{Kraj izračunavanja } a:=z_k; \end{cases}$$

4° Ako je

$$|y_{k+1}-x_{k+1}|\ \begin{cases} \geq\varepsilon & \text{preći na } 2^\circ, \\ <\varepsilon & z_{k+1}:=\frac{1}{2}(x_{k+1}+y_{k+1}) \\ & \text{Kraj izračunavanja } a:=z_{k+1}. \end{cases}$$

Primetimo da za grešku u aproksimaciji z_{k+1} važi ocena

$$|z_{k+1}-a|\leqq\frac{1}{2^{k+1}}(\beta-\alpha).$$

5.1.6. Schröderov razvoj

Neka je funkcija $f:[\alpha,\beta]\to R$ diferencijabilna i takva da je $f'(x)\neq 0$ $(\forall x\in[\alpha,\beta])$. S obzirom da je tada f striktno monotona na $[\alpha,\beta]$, to postoji njena inverzna funkcija F koja je takodje diferencijabilna. Naime,

$$F'(y) = \frac{dx}{dy} = \frac{1}{f'(x)} \qquad (y = f(x)).$$

Štaviše, ako je f dva puta diferencijabilna funkcija na $[\alpha,\beta]$, tada je

$$F''(y) = -\frac{f''(x)}{f'(x)^3}.$$

Problem nalaženja viših izvoda funkcije F, pod pretpostavkom da je ona dovoljan broj puta diferencijabilna, može biti vrlo komplikovan. S obzirom da Schröderov razvoj, o kome će biti reči u ovom odeljku, zahteva poznavanje viših izvoda funkcije F to ćemo najpre izložiti jedan rekurzivni postupak za rešavanje pomenutog problema.

Pretpostavimo da je funkcija f (n+1) puta diferencijabilna na $[\alpha,\beta]$, kao i to da je

$$F^{(k)}(y) = \frac{X_k}{(f')^{2k-1}} \qquad (k = 1,\ldots,n+1), \tag{1.6.1}$$

gde je X_k polinom po $f', f'', \ldots, f^{(k)}$ i $f^{(i)} \equiv f^{(i)}(x)$ za $i = 1,\ldots,n+1$.

Primetimo da je formula (1.6.1) tačna za $k=1$ i $k=2$, pri čemu su $X_1 = 1$ i $X_2 = -f''$.

Pretpostavimo da je formula (1.6.1) tačna za neko $k \in \{1,\ldots,n\}$. Kako je X_k polinom po $f', \ldots, f^{(k)}$ i

$$X_k' = \frac{dX_k}{dx} = \frac{d}{dx} X_k(f', \ldots, f^{(k)}) = \sum_{i=1}^{k} \frac{\partial X_k}{\partial f^{(i)}} f^{(i+1)},$$

polinom po $f', \ldots, f^{(k+1)}$, to iz (1.6.1) sleduje

$$F^{(k+1)}(y) = \frac{d}{dx}\left(\frac{X_k}{(f')^{2k-1}}\right)\frac{dx}{dy}$$

$$= \frac{f' X_k' - (2k-1)X_k f''}{(f')^{2k+1}} = \frac{X_{k+1}}{(f')^{2k+1}},$$

gde je

$$X_{k+1} = f'X'_k - (2k-1)X_k f'' \tag{1.6.2}$$

polinom po $f', \ldots, f^{(k+1)}$.

Dakle, izvodi funkcije F su dati sa (1.6.1), pri čemu se niz $\{X_k\}_{k\in N}$ odredjuje pomoću rekurentne relacije (1.6.2) polazeći od $X_1 = 1$.

Prvih pet članova niza $\{X_k\}$ su

$$X_1 = 1,$$

$$X_2 = -f'',$$

$$X_3 = -f'f''' + 3f''^2,$$

$$X_4 = -f'^2 f^{IV} + 10f'f''f''' - 15f''^3,$$

$$X_5 = -f'^3 f^{V} + 15f'^2 f'' f^{IV} + 10f'^2 f'''^2 - 105f'f''^2 f''' + 105f''^4.$$

Pretpostavimo sada da funkcija f na segmentu $[\alpha,\beta]$ ima jednu prostu nulu $x = a$, čiju okolinu označimo sa $U(a)$. Ako stavimo $h = -\dfrac{f(x)}{f'(x)}$ $(x\in U(a))$, tada je $0 = f(x) + hf'(x)$, odakle je

$$a = F(0) = F(f + hf').$$

Neka $f\in C^{n+1}[\alpha,\beta]$. Tada na osnovu Taylorove formule imamo

$$a = \sum_{k=0}^{n} \frac{1}{k!} F^{(k)}(y)(hf')^k + \frac{F^{(n+1)}(\bar{y})}{(n+1)!}(hf')^{n+1},$$

gde je $\bar{y} = f + thf' = (1-t)f = \theta f$ $(t,\theta\in(0,1))$. Najzad, korišćenjem formule (1.6.1) dobijamo Schröderov razvoj

$$a - x = \sum_{k=1}^{n} \frac{1}{k!} X_k\left(\frac{f'}{f'}, \frac{f''}{f'}, \ldots, \frac{f^{(k)}}{f'}\right) h^k + O(f(x)^{n+1}),$$

tj.

$$a - x = h - \frac{f''}{2f'}h^2 + \frac{3f''^2 - f'f'''}{6f'^2}h^3 + \frac{10f'f''f''' - f'^2f^{\mathrm{IV}} - 15f''^3}{24f'^3}h^4 + \dots\ . \tag{1.6.3}$$

Primedba 1.6.1. Ako je funkcija analitička, u prethodnom razvoju može se uzeti da $n \to +\infty$.

5.1.7. Metodi višeg reda

U ovom odeljku ukazaćemo na neke načine za dobijanje iterativnih procesa, čiji je red konvergencije veći od dva, pri čemu pretpostavljamo da jednačina

$$f(x) = 0 \tag{1.7.1}$$

na segmentu $[\alpha,\beta]$ ima jedinstven prost koren $x = a$ kao i da je funkcija f dovoljan broj puta neprekidno-diferencijabilna na $[\alpha,\beta]$.

1. Posmatrajmo Schröderov razvoj (1.6.3). Uzimajući konačan broj prvih članova na desnoj strani ovog razvoja možemo dobiti niz iterativnih formula.

Neka je

$$\Phi_2(x) = x + h = x - \frac{f(x)}{f'(x)},$$

$$\Phi_3(x) = \Phi_2(x) - \frac{f''}{2f'}h^2 = x - \frac{f(x)}{f'(x)} - \frac{f''(x)f(x)^2}{2f'(x)^3},$$

$$\Phi_4(x) = \Phi_3(x) + \frac{3f''^2 - f'f'''}{6f'^2}h^3$$

$$= x - \frac{f(x)}{f'(x)} - \frac{f''(x)f(x)^2}{2f'(x)^3} - \frac{f(x)^3}{6f'(x)^4}\left(3\frac{f''(x)^2}{f'(x)} - f'''(x)\right),$$

itd.

Primetimo da je Φ_2 iterativna funkcija Newtonovog metoda.

Kako je, u prvoj aproksimaciji, h jednako a-x ($x \to a$), na osnovu (1.6.3) imamo

$$\Phi_m(x) - a = O(h^m) = O((x-a)^m) \qquad (m = 2,3,\ldots),$$

kada $x \to a$, što znači da iterativni proces

$$(1.7.2) \qquad x_{k+1} = \Phi_m(x_k) \qquad (k = 0,1,\ldots),$$

primenjen na odredjivanje korena jednačine (1.7.1), ima red konvergencije najmanje m. Takodje, na osnovu Schröderovog razvoja možemo dobiti i vrednosti limesa

$$L_m = \lim_{k \to +\infty} \frac{x_{k+1} - a}{(x_k - a)^m} \qquad (m = 2,3,\ldots).$$

Naime, imamo

$$L_2 = \frac{f''(a)}{2f'(a)},$$

$$L_3 = \frac{1}{2}\left(\frac{f''(a)}{f'(a)}\right)^2 - \frac{1}{6}\cdot\frac{f'(a)f'''(a)}{f'(a)^2},$$

$$L_4 = \frac{5}{8}\left(\frac{f''(a)}{f'(a)}\right)^3 + \frac{1}{24}\cdot\frac{f^{IV}(a)}{f'(a)} - \frac{5}{12}\cdot\frac{f''(a)f'''(a)}{f'(a)^2}.$$

Primedba 1.7.1. Formule (1.7.2) se često nazivaju Čebiševljeve iterativne formule.

2. Primenimo sada metod, koji je definisan teoremom 2.4.1 iz treće glave, na ubrzavanje Newtonovog procesa.

Kako je iterativna funkcija kod Newtonovog procesa data sa $\Phi(x) = x - \frac{f(x)}{f'(x)}$, imamo

$$x_{k+1} = x_k - \frac{x_k - \Phi(x_k)}{1 - \frac{1}{2}\Phi'(x_k)} \qquad (k = 0,1,\ldots),$$

tj.

$$x_{k+1} = x_k - \frac{2f(x_k)f'(x_k)}{2f'(x_k)^2 - f(x_k)f''(x_k)} \qquad (k=0,1,\ldots). \tag{1.7.3}$$

Metod (1.7.3), na osnovu pomenute teoreme, ima red konvergencije najmanje tri. Ovaj metod je u literaturi poznat kao Salehov metod tangentnih hiperbola([9]) ili kao Halleyev metod.

Za metod (1.7.3) lako se može dobiti asimptotska formula

$$\frac{x_{k+1}-a}{(x_k-a)^3} \sim \left[\left(\frac{f''(a)}{2f'(a)}\right)^2 - \frac{f'''(a)}{6f'(a)}\right] \qquad (k \to +\infty).$$

Primedba 1.7.2. U radu [10] data je sledeća modifikacija metoda (1.7.3)

$$x_{k+1} = x_k - \frac{2f(x_k)f'(x_k)}{2f'(x_k)^2 - f(x_k)\dfrac{f'(x_k)-f'(x_{k-1})}{x_k-x_{k-1}}} \qquad (k=1,2,\ldots)$$

za koju važi

$$x_{k+1} - a \sim \frac{f'''(a)}{4f'(a)}(x_k-a)^2(x_{k-1}-a) \qquad (k \to +\infty),$$

odakle izlazi da je red konvergencije ovako modifikovanog metoda $r = 1+\sqrt{2} \cong 2.414$.

3. Sukcesivna primena metoda, definisanog teoremom 2.4.4 (treća glava), na ubrzavanje Newtonovog metoda daje niz iterativnih formula, koje su po obliku slične formulama (1.7.2).

4. Izaberimo niz funkcija $\{\psi_i\}_{i=0,1,\ldots,n}$ dovoljan broj puta diferencijabilnih na $[\alpha,\beta]$.

U radu [11] je predložena konstrukcija iterativnog procesa za rešavanje jednačine (1.7.1) u obliku

$$x_{k+1} = \varphi_n(x_k) \qquad (k=0,1,\ldots), \tag{1.7.4}$$

gde su

$$\varphi_n(x) = x - \frac{D_n(x)}{\Delta_n(x)} ,$$

$$\Delta_n(x) = \begin{vmatrix} (f\psi_o)' & (f\psi_1)' & \dots & (f\psi_n)' \\ (f\psi_o)'' & (f\psi_1)'' & & (f\psi_n)'' \\ \vdots & & & \\ (f\psi_o)^{(n+1)} & (f\psi_1)^{(n+1)} & & (f\psi_n)^{(n+1)} \end{vmatrix} ,$$

$$D_n(x) = \begin{vmatrix} f\psi_o & f\psi_1 & \dots & f\psi_n \\ (f\psi_o)'' & (f\psi_1)'' & & (f\psi_n)'' \\ \vdots & & & \\ (f\psi_o)^{(n+1)} & (f\psi_1)^{(n+1)} & & (f\psi_n)^{(n+1)} \end{vmatrix} .$$

Može se pokazati da je

$$\varphi_n(a) = a, \varphi_n'(a) = \varphi_n''(a) = \dots = \varphi_n^{(n+1)}(a) = 0, \quad \varphi_n^{(n+2)}(a) \neq 0,$$

što znači da iterativni proces (1.7.4) ima red konvergencije n+2.

Navešćemo nekoliko specijalnih slučajeva formule (1.7.4).

1° Ako je $n=o$ i $\psi_o(x) \equiv \psi(x)$, (1.7.4) se svodi na uopšten Newtonov metod (1.2.8).

2° Ako je $n=1$, $\psi_o(x) \equiv 1$, $\psi_1(x) \equiv x$, (1.7.4) se svodi na (1.7.3).

3° Ako je $n=2$, $\psi_o(x) \equiv 1$, $\psi_1(x) \equiv x$, $\psi_2(x) = x^2$, iz (1.7.4) sleduje formula Domorjada-Lika([12]), čija je iterativna funkcija data sa

$$\varphi_2(x) = x - \begin{vmatrix} f'(x) & f(x) \\ \frac{1}{2}f''(x) & f'(x) \end{vmatrix} \cdot \begin{vmatrix} f'(x) & f(x) & 0 \\ \frac{1}{2}f''(x) & f'(x) & f(x) \\ \frac{1}{6}f'''(x) & \frac{1}{2}f''(x) & f'(x) \end{vmatrix}^{-1} .$$

5. Posmatrajmo formulu (1.2.9), gde je p parametar. Sa U(a) označimo okolinu korena x= a jednačine (1.7.1).

Teorema 1.7.1. Neka je funkcija f četiri puta neprekidno-diferencijabilna u U(a) i neka niz $\{x_k\}_{k\in N_0}$ $(x_0 \in U(a))$, definisan pomoću (1.2.9), konvergira ka a. Tada, pri $k\to+\infty$, važi asimptotska formula

$$x_{k+1}-a \sim \left[\frac{p}{x_k}+\frac{s_k}{2}\right]h_k^2+\left[\left(\frac{p}{x_k}\right)^2-\frac{1}{6}(3s_k^2-r_k)\right]h_k^3 \tag{1.7.5}$$
$$+\left[\left(\frac{p}{x_k}\right)^3-\frac{1}{24}(10r_ks_k-t_k-15s_k^3)\right]h_k^4,$$

gde su

$$h_k=-\frac{f(x_k)}{f'(x_k)},\quad s_k=\frac{f''(x_k)}{f'(x_k)},\quad r_k=\frac{f'''(x_k)}{f'(x_k)},\quad t_k=\frac{f^{IV}(x_k)}{f'(x_k)}.$$

Dokaz. Kako je, na osnovu (1.2.9),

$$x_{k+1}-x_k=\frac{x_kf(x_k)}{x_kf'(x_k)+pf(x_k)}=\frac{h_k}{1-\frac{p}{x_k}h_k} \tag{1.7.6}$$

i kako je $\left|\frac{p}{x_k}h_k\right|<1\ (x_k\in U(a))$, iz (1.7.6) sleduje

$$x_{k+1}-x_k=h_k+\frac{p}{x_k}h_k^2+\left(\frac{p}{x_k}\right)^2h_k^3+\left(\frac{p}{x_k}\right)^3h_k^4+O(h_k^5), \tag{1.7.7}$$

S druge strane, na osnovu (1.6.3) imamo

$$a-x_k=h_k-\frac{1}{2}s_kh_k^2+\frac{1}{6}(3s_k^2-r_k)h_k^3 \tag{1.7.8}$$
$$+\frac{1}{24}(10r_ks_k-t_k-15s_k^3)h_k^4+O(h_k^5).$$

Ako od (1.7.7) oduzmemo (1.7.8) dobijamo (1.7.5).

Posmatrajmo sada formulu (1.2.9) za dve vrednosti parametra $p(p_1$ i $p_2)$, tj.

$$x_{k+1}^{(i)}=x_k\left(1-\frac{f(x_k)}{x_kf'(x_k)+p_if(x_k)}\right)\qquad (i=1,2) \tag{1.7.9}$$

i stavimo

$$x_{k+1} = \frac{1}{2}(x_{k+1}^{(1)} + x_{k+1}^{(2)}). \tag{1.7.10}$$

Tada, na osnovu (1.7.5), imamo

$$x_{k+1} - a \sim \frac{1}{2}\left[\frac{p_1+p_2}{x_k} + s_k\right]h_k^2 + \frac{1}{2}\left[\frac{p_1^2+p_2^2}{x_k^2} - \frac{1}{3}(3s_k^2 - r_k)\right]h_k^3 \tag{1.7.11}$$

$$+ \frac{1}{2}\left[\frac{p_1^3+p_2^3}{x_k^3} - \frac{1}{12}(10r_k s_k - t_k - 15s_k^3)\right]h_k^4 .$$

Ako u formulama (1.7.9) izaberemo parametre p_1 i p_2, takve da je

$$p_1 + p_2 = -s_k x_k , \quad p_1^2 + p_2^2 = \frac{1}{3}(3s_k^2 - r_k)x_k^2 , \tag{1.7.12}$$

formula (1.7.11) postaje

$$x_{k+1} - a \sim C_k h_k^4 \qquad (k \to +\infty),$$

gde smo sa C_k označili koeficijent u formuli (1.7.11) uz h_k^4, sa vrednostima za p_1 i p_2 koje se dobijaju iz (1.7.12). Naime, p_1 i p_2 su rešenja kvadratne jednačine

$$p^2 + s_k x_k p + \frac{1}{6}x_k^2 r_k = 0,$$

tj.

$$p^2 + x_k \frac{f''(x_k)}{f'(x_k)}p + \frac{1}{6} x_k^2 \frac{f'''(x_k)}{f'(x_k)} = 0. \tag{1.7.13}$$

Na osnovu prethodnog može se formulisati i dokazati sledeća teorema (videti [13]).

<u>Teorema 1.7.2</u>. Neka je funkcija f četiri puta neprekidno-diferencijabilna u U(a) i neka su p_1 i p_2 rešenja jednačine (1.7.13). Tada iterativni proces definisan formulama (1.7.9) i (1.7.10) ima red konvergencije četiri, tj. važi

$$x_{k+1} - a \sim C(x_k - a)^4 \qquad (k \to +\infty),$$

gde je

$$C = \frac{3f''(a)^3 - 4f'(a)f''(a)f'''(a) + f'(a)^2 f^{IV}(a)}{24f'(a)^3} .$$

Primedba 1.7.3. Iterativni proces definisan formulama (1.7.9) i (1.7.10) može se eksplicitno izraziti u obliku

$$x_{k+1} = \varphi(x_k) \qquad (k = 0,1,\ldots),$$

gde je

$$\varphi(x) = x - 3f(x)\frac{2f'(x)^2 - f(x)f''(x)}{6f'(x)(f'(x)^2 - f(x)f''(x)) + f(x)^2 f'''(x)} .$$

Primetimo da je ova formula ekvivalentna sa formulom Domorjada-Lika([12]).

6. Iterativne formule trećeg reda za odredjivanje višestrukih korena jednačina razmatrane su u [14], [15], [16].

5.2. SISTEMI NELINEARNIH JEDNAČINA

5.2.1. Uvodne napomene

Ovo poglavlje posvećeno je razmatranju numeričkih metoda za rešavanje sistema nelinearnih jednačina

$$(2.1.1) \qquad f_1(x_1,\ldots,x_n) = 0 \qquad (i = 1,\ldots,n).$$

gde su $f_i:R^n \to R$ $(i = 1,\ldots,n)$ date funkcije.

Kao što je rečeno u odeljku 3.1.1, sistem jednačina (2.1.1) se može tretirati kao specijalan slučaj operatorske jednačine

$$(2.1.2) \qquad Fu = \theta,$$

gde je F operator koji preslikava Banachov prostor X u Banachov prostor Y i θ nula-vektor prostora Y. Naime, treba samo uzeti $X = Y = R^n$, $u = \vec{x} = [x_1 \ \ldots \ x_n]^T$, $\theta = [0 \ \ldots \ 0]^T$ i

$$Fu = \vec{f}(\vec{x}) = \begin{bmatrix} f_1(x_1,\ldots,x_n) \\ \vdots \\ f_n(x_1,\ldots,x_n) \end{bmatrix}$$

Osnovni metod za rešavanje sistema jednačina (2.1.1), ali i za rešavanje opšte operatorske jednačine (2.1.2) je metod Newton-Kantorovića, koji predstavlja generalizaciju Newtonovog metoda obradjenog u odeljku 5.1.2. U našem izlaganju ovaj metod biće razmatran za opšti slučaj operatorske jednačine.

S obzirom da metod Newton-Kantorovića zahteva nalaženje inverznog operatora od $F'_{(u)}$, koje ponekad može biti komplikovano, u literaturi je u poslednje vreme razradjena čitava klasa tzv. kvazi-njutnovskih metoda, kod kojih se koriste izvesne aproksimacije pomenutog operatora. U našem izlaganju zadržaćemo se samo na gradijentnom metodu, koji ćemo prezenitirati kao metod za minimizaciju izvesne funkcionele.

O pomenutim metodima postoji opširna literatura (videti, na primer, [17], [18], [19], [20], [21], [22], [23]).

5.2.2. Metod Newton-Kantoroviča

Osnovni iterativni metod za rešavanje jednačine (2.1.2) je metod Newton-Kantoroviča, koji predstavlja generalizaciju Newtonovog metoda (1.2.3). Fundamentalne rezultate vezane za egzistenciju i jedinstvenost rešenja jednačine (2.1.2) i konvergenciju metoda dao je L.V.Kantorovič ([17], [18]). Takodje, ovaj metod je razmatran i od strane drugih autora (videti, naprimer, [24], [25], [26], [27]).

Pretpostavimo da jednačina (2.1.2) ima rešenje $u = a$ i da je operator $F:X \to Y$ Frechet-diferencijabilan u konveksnoj okolini $D(\subset X)$ tačke a. Metod Newton-Kantoroviča za rešavanje jednačine (2.1.2) zasniva se na linearizaciji ove jednačine. Naime, neka je nadjeno približno rešenje u_k. Tada za nalaženje sledećeg približnog rešenja u_{k+1}, jednačinu (2.1.2) zamenimo jednačinom

$$\text{(2.2.1)} \qquad Fu_k + F'_{(u_k)}(u-u_k) = \theta .$$

Ako za operator $F'_{(u_k)}$ postoji inverzni operator $\Gamma(u_k) = [F'_{(u_k)}]^{-1}$, iz (2.2.1) dolazi se do iterativnog metoda

$$\text{(2.2.2)} \qquad u_{k+1} = u_k - \Gamma(u_k)Fu_k \qquad (k = 0,1,\ldots),$$

koji je poznat kao metod Newton-Kantoroviča. Početna vrednost u_o za generisanje niza $\{u_k\}_{k\in N_o}$ uzima se iz D, a njen izbor predstavlja dosta težak problem.

Metod (2.2.2) može se predstaviti u obliku

$$u_{k+1} = Tu_k \qquad (k = 0,1,\ldots),$$

gde je

$$Tu = u - \Gamma(u)Fu.$$

<u>Teorema 2.2.1</u>. Neka je operator F dva puta Fréchet-diferencijabilan na D, pri čemu za svako $u \in D$ postoji operator $\Gamma(u)$. Ako su operatori $\Gamma(u)$ i $F''_{(u)}$ ograničeni i $u_o \in D$ dovoljno blisko tački a, iterativni proces (2.2.2) ima red konvergencije najmanje dva.

<u>Dokaz</u>. Neka je $\|\Gamma(u)\| \leq m$, $\|F''_{(u)}\| \leq M_2$ za svako $u \in D$. Na osnovu Taylorove formule imamo

$$\text{(2.2.3)} \qquad Fa = \theta = F_u + F'_{(u)}(a-u) + W(u,a-u),$$

gde je

$$\|W(u,a-u)\| \leq \frac{1}{2} \sup_{t\in[0,1]} \|F''_{(u+t(a-u))}\| \cdot \|a-u\|^2 \leq \frac{M_2}{2} \|u-a\|^2.$$

Kako je

$$Tu-a = u-\Gamma(u)Fu-a = \Gamma(u)\left[F'_{(u)}(u-a) - Fu\right],$$

na osnovu (2.2.3), dobijamo

$$Tu-a = \Gamma(u)W(u,a-u),$$

odakle sleduje

$$\|Tu-a\| \leq \|\Gamma(u)\| \cdot \|W(u,a-u)\| \leq \frac{1}{2} mM_2\|u-a\|^2,$$

tj.

$$\|Tu-a\| = O(\|u-a\|^2).$$

Ovim je dokaz teoreme završen.

U daljem razmatranju smatraćemo da je D kugla $K[u_o,R]$, gde je u_o početna vrednost niza $\{u_k\}_{k\in N_o}$.

Ako je ispunjen Lipschitzov uslov

$$\|F'_{(u)} - F'_{(v)}\| \leq L\|u-v\| \quad (u,v \in K[u_o,R]) \tag{2.2.4}$$

iz

$$Fu-Fv-F'_{(v)}(u-v) = \int_0^1 [F'_{(v+t(u-v))}-F'_{(v)}](u-v)dt$$

sleduje nejednakost

$$\|Fu-Fv-F'_{(v)}(u-v)\| \leq \frac{L}{2}\|u-v\|^2. \tag{2.2.5}$$

<u>Teorema 2.2.2</u>. Neka je operator F Fréchet-diferencijabilan u kugli $K[u_o,R]$ i zadovoljava uslov (2.2.4) i neka su tačne nejednakosti

$$\|\Gamma_o\| \leq b_o, \ \|\Gamma_o F u_o\| \leq \eta_o, \quad h_o = b_o L \eta_o \leq \frac{1}{2}, \tag{2.2.6}$$

gde je $\Gamma_o = \Gamma(u_o)$.

Ako je

$$R \geq r_o = \frac{1-\sqrt{1-2h_o}}{h_o}\eta_o, \tag{2.2.7}$$

niz $\{u_k\}_{k\in N_o}$, koji se generiše pomoću (2.2.2), konvergira rešenju $a \in K[u_o,r_o]$ jednačine (2.1.2).

<u>Dokaz</u>. Neka su nizovi $\{b_k\}, \{\eta_k\}, \{h_k\}, \{r_k\}$ definisani sa

$$b_{k+1} = \frac{b_k}{1-h_k}, \qquad \eta_{k+1} = \frac{k}{2(1-h_k)}\eta_k,$$

$$h_{k+1} = b_{k+1}L\eta_{k+1}, \qquad r_k = \frac{1-\sqrt{1-2h_{k+1}}}{h_{k+1}}\eta_k.$$

Dokazaćemo da niz $\{u_k\}_{k\in N_o}$ postoji i da je

$$\|\Gamma(u_k)\| \leq b_k, \ \|\Gamma(u_k)Fu_k\| \leq \eta_k, \quad h_k \leq \frac{1}{2} \tag{2.2.8}$$

i

$$K[u_k,r_k] \subset K[u_{k-1}, r_{k-1}]. \tag{2.2.9}$$

Dokaz izvodimo indukcijom.

Za $k = 0$ nejednakosti (2.2.8) su tačne, jer se svode na (2.2.6).

Pretpostavimo da su tačne i za $k = m$. Kako je, na osnovu (2.2.2) i induktivne pretpostavke

$$\|u_{m+1} - u_m\| = \|\Gamma(u_m)Fu_m\| \leq \eta_m$$

i kako je $r_m > \eta_m$, sleduje da $u_{m+1} \in K[u_m, r_m]$, a tim pre $u_m \in K[u_o, R]$, odakle zaključujemo da $F'_{(u_{m+1})}$ postoji.

Operator $\Gamma(u_{m+1})$ takodje postoji, s obzirom da se može predstaviti u obliku

$$\Gamma(u_{m+1}) = [I+\Gamma(u_m)(F'_{(u_{m+1})} - F'_{(u_m)})]^{-1}\,\Gamma(u_m),$$

tj.

$$\Gamma(u_{m+1}) = \sum_{i=o}^{\infty} (-1)^i [\Gamma(u_m)(F'_{(u_{m+1})} - F'_{(u_m)})]^i\,\Gamma(u_m), \tag{2.2.10}$$

jer je, na osnovu (2.2.4)

$$\lambda = \|\Gamma(u_m)(F'_{(u_{m+1})} - F'_{(u_m)})\| \leq b_m L \|u_{m+1} - u_m\| \leq h_m \leq \frac{1}{2} < 1.$$

Iz (2.2.10) sleduje

$$\|\Gamma(u_{m+1})\| \leq \sum_{i=o}^{\infty} \lambda^i b_m = \frac{b_m}{1-h_m} = b_{m+1}, \tag{2.2.11}$$

čime je dokazana prva nejednakost u (2.2.8).

Kako je $Fu_{m+1} = Fu_{m+1} - Fu_m - F'_{(u_m)}(u_{m+1} - u_m)$, na osnovu (2.2.5) sleduje nejednakost

$$\|Fu_{m+1}\| \leq \frac{L}{2}\|u_{m+1} - u_m\|^2 \leq \frac{L}{2}\eta_m^2, \tag{2.2.12}$$

odakle, na osnovu (2.2.11) dobijamo

$$\|\Gamma(u_{m+1})Fu_{m+1}\| \leq \frac{b_m L \eta_m^2}{2(1-h_m)} = \frac{h_m}{2(1-h_m)}\eta_m = \eta_{m+1},$$

što predstavlja drugu nejednakost u (2.2.8) za $k=m+1$.

Kako je

$$h_{m+1} = B_{m+1} L \eta_{m+1} = \frac{b_m}{1-h_m} L \frac{h_m}{2(1-h_m)} \eta_m = \frac{1}{2}\left(\frac{h_m}{1-h_m}\right)^2 \leq \frac{1}{2}.$$

dokazana je i treća nejednakost u (2.2.8).

Za dokaz inkluzije (2.2.9) pretpostavimo da tačka $u \in K[u_{m+1}, r_{m+1}]$. Tada iz $\|u-u_{m+1}\| \leqq r_{m+1}$ sleduje nejednakost

$$\|u-u_m\| \leqq \|u-u_{m+1}\| + \|u_{m+1}-u_m\| \leqq r_{m+1}+\eta_m. \tag{2.2.13}$$

Kako je $r_{m+1}+\eta_m = r_m$ dokaz je završen.

Kako je $h_k \leqq \frac{1}{2}$ iz definicije niza $\{\eta_k\}$ sleduje nejednakost

$$\eta_{k+1} \leqq \frac{1}{2}\eta_k,$$

odakle zaključujemo da je $\{\eta_k\}$ nula-niz. Tada je

$$\lim_{k\to+\infty} r_k = \lim_{k\to+\infty} \frac{1-\sqrt{1-2h_k}}{h_k}\eta_k = \lim_{k\to+\infty} \frac{2\eta_k}{1+\sqrt{1-2h_k}} = 0,$$

što znači da niz $\{u_k\}$ konvergira nekoj tački $a \in K[u_o, r_o]$. Tačka a je rešenje jednačine (2.1.2), jer je na osnovu (2.2.12)

$$\lim_{m\to+\infty} \|Fu_{m+1}\| \leqq \frac{L}{2} \lim_{m\to+\infty} \eta_m^2 = 0,$$

tj.

$$\lim_{m\to+\infty} Fu_m = \theta.$$

Ovim je dokaz završen.

<u>Teorema 2.2.3</u>. Ako su ispunjeni uslovi prethodne teoreme, tada je

$$\|u_k-a\| \leqq \frac{1}{2^{k-1}}(2h_o)^{2^k-1}\eta_o \qquad (k\in N). \tag{2.2.14}$$

<u>Dokaz</u>. Primetimo najpre da je

$$\frac{h_k}{1-h_k} \leqq 2h_k \qquad (0 \leqq h_k \leqq \tfrac{1}{2}).$$

Tada je, na osnovu definicije nizova $\{\eta_k\}$ i $\{h_k\}$,

$$\eta_{k+1} \leqq h_k\eta_k \qquad \text{i} \qquad h_{k+1} \leqq 2h_k^2,$$

odakle sleduje

$$h_k \leqq \frac{1}{2}(2h_o)^{2^k}$$

i

$$\eta_k \leqq h_{k-1}h_{k-2}\eta_{k-2} \leqq \dots \leqq h_{k-1}h_{k-2}\dots h_o\eta_o$$

$$\leqq \frac{1}{2^k}(2h_o)^{1+2+\dots+2^{k-1}}\eta_o = \frac{1}{2^k}(2h_o)^{2^k-1}\eta_o.$$

Najzad, na osnovu (2.2.13), na u=a i m=k, imamo

$$||u_k-a|| \leqq r_k = \frac{2}{1+\sqrt{1-2h_k}}\eta_k \leqq 2\eta_k,$$

tj. (2.2.14).

Da bi se izbeglo odredjivanje inverznog operatora $\Gamma(u) = [F'_{(u)}]^{-1}$ pri svakom koraku, metod Newton-Kantorovića može se modifikovati na sledeći način

(2.2.15) $$u_{k+1} = u_k - \Gamma_o F u_k \qquad (k=0,1,2,\dots),$$

gde je $\Gamma_o = \Gamma(u_o)$. Uvodjenjem operatora T pomoću

(2.2.16) $$Tu = u-\Gamma_o Fu,$$

modifikovani metod (2.2.15) se može predstaviti u obliku

$$u_{k+1} = Tu_k \qquad (k=0,1,2,\dots).$$

Pretpostavimo sada da su ispunjeni sledeći uslovi:

1^o Operator F je Frechet-diferencijabilan u kugli $K[u_o,R]$,

2^o $F'_{(u)}$ zadovoljava uslov (2.2.4),

3^o Operator Γ_o postoji i

$$||\Gamma_o|| \leqq b_o, \qquad ||\Gamma_o F u_o|| \leqq \eta_o.$$

Tada važi sledeća teorema.

Teorema 2.2.4. Ako su ispunjeni uslovi

$$h_o = b_o L\eta_o < \frac{1}{2}$$

i

$$r_o = \frac{1-\sqrt{1-2h_o}}{h_o}\,\eta_o \leqq R$$

niz koji se generiše pomoću (2.2.15) konvergira ka rešavanju $a\in K[u_o,r_o]$ jednačine (2.1.2).

Dokaz. Na osnovu teoreme 2.2.2 jednačina (2.1.2) ima rešenje $a\in K[u_o,r_o]$. Dokažimo sada da je operator T, uveden pomoću (2.2.16), kontrakcija na $K[u_o,r_o]$.

Neka $u,v\in K[u_o,r_o]$. Tada iz

$$Tu-Tv = u-v-\Gamma_o(Fu-Fv)$$

(2.2.17)
$$= \Gamma_o \int_0^1 \left[F'_{(u_o)}-F'_{(v+t(u-v))}\right](u-v)\ dt$$

i uslova (2.2.4) sleduje

(2.2.18) $\|Tu-Tv\| \leqq b_o L r_o \|u-v\|$,

gde je
$$q = b_o L r_o = 1-\sqrt{1-2h_o}<1,$$

odakle zaključujemo da je T kontrakcija na kugli $K[u_o,r_o]$. Dokažimo još da je $TK[u_o,r_o] \subset K[u_o,r_o]$.

Neka $u\in K[u_o,r_o]$. Tada na osnovu (2.2.17) i (2.2.4) imamo

$$\|Tu-u_o\| \leqq \|Tu-Tu_o\| + \|Tu_o-u_o\| \leqq \frac{b_o L}{2}\|u-u_o\|^2 + \eta_o,$$

tj.

$$\|Tu-u_o\| \leqq \frac{b_o L r_o^2}{2} + \eta_o = r_o.$$

S obzirom da $T:K[u_o,r_o] \to K[u_o,r_o]$ zaključujemo da jednačina (2.1.2) ima jedinstveno rešenje $a\in K[u_o,r_o]$. Tačka a predstavlja graničnu vrednost niza, koji se generiše pomoću modifikovanog metoda Newton-Kantoroviča.

Ovim je dokaz završen.

Nejednakost (2.2.18) za v=a, postaje

$$\|Tu-a\| \leqq q\|u-a\|,$$

što znači da je iterativni proces (2.2.15) prvog reda.

Predjimo sada na slučaj sistema nelinearnih jednačina

(2.2.19) $$f_i(x_1,\ldots,x_n) = 0 \qquad (i = 1,\ldots,n).$$

Ovde je $X = Y = R^n$, $u = \vec{x} = [x_1 \ \ldots \ x_n]^T$ i F definisano pomoću (2.1.3). Ako je F Fréchet-diferencijabilan operator (videti odeljak 2.2.4) tada je

$$F'_{(u)} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \\ \frac{\partial f_n}{\partial x_1} & & \frac{\partial f_n}{\partial x_n} \end{bmatrix} = W(\vec{x}),$$

tj. $W(\vec{x})$ je Jacobieva matrica za $\vec{f}$. Ako je $\det(W(\vec{x})) \neq 0$, tj. ako je matrica $W(\vec{x})$ regularna, metod Newton-Kantoroviča za rešavanje sistema jednačina (2.2.19) je dat sa

(2.2.20) $$\vec{x}^{(k+1)} = \vec{x}^{(k)} - W^{-1}(\vec{x}^{(k)})\vec{f}(\vec{x}^{(k)}) \qquad (k=0,1,\ldots),$$

gde je $\vec{x}^{(k)} = [x_1^{(k)} \ \ldots \ x_n^{(k)}]^T$. Vrlo često se ovaj metod u literaturi sreće kao Newton-Raphsonov metod.

Kao što smo u prethodnom izlaganju videli metod (2.2.20) se može modifikovati u smislu da se inverzna matrica od $W(\vec{x})$ ne odredjuje u svakoj iteraciji, već samo u prvoj. Dakle,

(2.2.21) $$\vec{x}^{(k+1)} = \vec{x}^{(k)} - W^{-1}(\vec{x}^{(o)})\vec{f}(\vec{x}^{(k)}) \qquad (k=0,1,\ldots).$$

Primedba 2.2.1. Modifikovani metod (2.2.21) može se shvatiti i kao metod proste iteracije

$$\vec{x}^{(k+1)} = T\vec{x}^{(k)} = \vec{x}^{(k)} + \Lambda\vec{f}(\vec{x}^{(k)}) \qquad (k=0,1,\ldots),$$

sa matricom Λ odredjenom iz uslova da je izvod od T nula-operator, tj. da je $I + \Lambda W(\vec{x}^{(o)})$ nula matrica. Ako je $W(\vec{x}^{(o)})$ regularna matrica imamo $\Lambda = W^{-1}(\vec{x}^{(o)})$.

Prethodno dokazane teoreme mogu se prilagoditi za slučaj sistema nelinearnih jednačina, pri čemu se uslovi za konvergenciju procesa (2.2.20), odnosno (2.2.21), mogu iskazati na različite načine, što zavisi od uvedene norme u X.

Poseban problem kod primene metoda Newton-Kantorovića je izbor startnih vrednosti, tj. početnog vektora $\vec{x}^{(o)}$, koji bi obezbedio konvergenciju procesa. Jedno rešenje ovog problema dato je u novije vreme (1974. god.) u radu [28].

Primer 2.2.1. Neka je dat sistem jednačina

$$
\begin{aligned}
f_1(x_1,x_2) &= x_1^3 - 3x_1x_2^2 + 1 = 0,\\
f_2(x_1,x_2) &= 3x_1^2x_2 - x_2^3 \quad = 0.
\end{aligned}
$$

Tada je

$$
W(x) = \begin{bmatrix} 3(x_1^2 - x_2^2) & -6x_1x_2 \\ 6x_1x_2 & 3(x_1^2-x_2^2) \end{bmatrix}
$$

i

$$
W^{-1}(x) = \frac{1}{3(x_1^2+x_2^2)^2}\begin{bmatrix} x_1^2-x_2^2 & 2x_1x_2 \\ -2x_1x_2 & x_1^2 - x_2^2 \end{bmatrix} \qquad (x_1^2 + x_2^2 \neq 0),
$$

Skalarni oblik od (2.2.20) je

$$
x_1^{(k+1)} = x_1^{(k)} - \frac{1}{D_k}\left[(x_1^{(k)^2} - x_2^{(k)^2})f_1^{(k)} + 2x_1^{(k)}x_2^{(k)}f_2^{(k)}\right],
$$

$$
x_2^{(k+1)} = x_2^{(k)} - \frac{1}{D_k}\left[-2x_1^{(k)}x_2^{(k)}f_1^{(k)} + (x_1^{(k)^2} - x_2^{(k)^2})f_2^{(k)}\right],
$$

gde su $D_k = 3(x_1^{(k)^2} + x_2^{(k)^2})^2$, $f_i^{(k)} \equiv f_i(x_1^{(k)}, x_2^{(k)})$ $(i=1,2)$ i $k = 0,1,\ldots$. Startujući sa $x_1^{(o)}=1$ i $x_2^{(o)}=1$, primenom poslednjih formula dobijamo

k	$x_1^{(k)}$	$x_2^{(k)}$
1	0.66667	0.83333
2	0.50869	0.84110
3	0.49933	0.86627
4	0.50000	0.86602

Primetimo da je tačno rešenje datog sistema $x_1=1/2$ i $x_2=\sqrt{3}/2$.

5.2.3. Gradijentni metod

Neka je dat sistem nelinearnih jednačina (2.2.19), čiji je matični oblik (videti (2.1.3))

$$\vec{f}(\vec{x}) = \vec{0}. \tag{2.3.1}$$

Gradijentni metod za rešavanje datog sistema jednačina zasniva se na minimizaciji funkcionale

$$U(\vec{x}) = \sum_{i=1}^{n} f_i(x_1,\dots,x_n)^2 = (\vec{f}(\vec{x}),\vec{f}(\vec{x})).$$

Lako je videti da važi ekvivalencija $U(\vec{x}) = 0 \Longleftrightarrow \vec{f}(\vec{x}) = \vec{0}$.

Pretpostavimo da jednačina (2.3.1) ima jedinstveno rešenje $\vec{x} = \vec{a}$, za koje funkcionela U dostiže minimum. Neka je $\vec{x}^{(o)}$ početna aproksimacija ovog rešenja. Konstruišimo niz $\{\vec{x}^{(k)}\}$ takav da je $U(\vec{x}^{(o)}) > U(\vec{x}^{(1)}) > U(\vec{x}^{(2)}) > \dots$. Postupimo kao i kod linearnih sistema jednačina (videti odeljak 4.3.7), tj. uzmimo

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} - \lambda_k \nabla U(\vec{x}^{(k)}) \quad (k=0,1,\dots), \tag{2.3.2}$$

gde je $\nabla U(\vec{x}) = \mathrm{grad}U(\vec{x}) = \left[\frac{\partial U}{\partial x_1} \ \dots \ \frac{\partial U}{\partial x_n}\right]^T$. Parametar λ_k odredjujemo iz uslova da skalarna funkcija S, definisana pomoću $S(t) = U(\vec{x}^{(k)} - t\nabla U(\vec{x}^{(k)}))$ ima minimum u tački $t = \lambda_k$. S obzirom da je jednačina $S'(t) = 0$ nelinearna, izvršimo njenu linearizaciju u okolini $t = 0$. U tom slučaju imamo

$$L_i^{(k)} = f_i(\vec{x}^{(k)} - t\nabla U(\vec{x}^{(k)})) = f_i(\vec{x}^{(k)}) - t(\nabla f_i(\vec{x}^{(k)}), \nabla U(\vec{x}^{(k)}))$$

pa je linearizovana jednačina

$$\sum_{i=1}^{n} L_i^{(k)} \frac{d}{dt} L_i^{(k)} = -\sum_{i=1}^{n} L_i^{(k)} (\nabla f_i(\vec{x}^{(k)}), \nabla U(\vec{x}^{(k)})) = 0,$$

odakle dobijamo

$$\lambda_k = t = \sum_{i=1}^{n} H_i f_i(\vec{x}^{(k)}) \Big/ \sum_{i=1}^{n} H_i^2, \tag{2.3.3}$$

gde smo stavili $H_i = (\nabla f_i(\vec{x}^{(k)}),\ \nabla U(\vec{x}^{(k)}))$ $(i = 1,\ldots,n)$.

Kako je

$$\frac{\partial U}{\partial x_j} = \frac{\partial}{\partial x_j}\left\{\sum_{i=1}^{n} f_i(\vec{x})^2\right\} = 2\sum_{i=1}^{n} f_i(\vec{x})\,\frac{\partial f_i(\vec{x})}{\partial x_j}$$

imamo

(2.3.4) $$\nabla U(\vec{x}) = 2W^T(\vec{x})\vec{f}(\vec{x}),$$

gde je $W(\vec{x})$ Jacobieva matrica za $\vec{f}$.

Shodno prethodnom, (2.3.3) postaje

$$\lambda_k = \frac{1}{2}\cdot\frac{(\vec{f}^{(k)},W_kW_k^T\vec{f}^{(k)})}{(W_kW_k^T\vec{f}^{(k)},W_kW_k^T\vec{f}^{(k)})}\quad (\vec{f}^{(k)}=\vec{f}(\vec{x}^{(k)}),W_k=W(\vec{x}^{(k)})).$$

Tada se gradijentni metod (2.3.2) može predstaviti u obliku

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} - 2\lambda_k W_k^T\vec{f}(\vec{x}^{(k)})\qquad (k = 0,1,\ldots).$$

Metod gradijenata se uspešno koristi u raznim optimizacionim problemima.

5.3. ALGEBARSKE JEDNAČINE

5.3.1. Uvodne napomene

Jedan od najstarijih problema u matematici je rešavanje algebarskih jednačina, koji je i danas aktuelan, s obzirom da ogroman broj tehničkih problema to zahteva.

Opšti oblik nelinearne algebarske jednačine n-tog stepena je

$$P(x) = a_o x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n = 0 \qquad (a_o \neq 0), \tag{3.1.1}$$

gde je n prirodan broj i $a_i (i=0,1,\dots,n)$ realni ili kompleksni brojevi.

Koreni ove jednačine $x_1,\dots,x_n$ povezani su sa koeficijentima $a_i (i= o,1,\dots,n)$ pomoću Vièteovih formula

$$\begin{aligned}
F_1 &= \sum_i x_i = -\frac{a_1}{a_o}, \\
F_2 &= \sum_{i<j} x_i x_j = \frac{a_2}{a_o}, \\
F_3 &= \sum_{i<j<k} x_i x_j x_k = -\frac{a_3}{a_o}, \\
&\vdots \\
F_n &= x_1 x_2 \dots x_n = (-1)^n \frac{a_n}{a_o}.
\end{aligned} \tag{3.1.2}$$

Jednačina (3.1.1) je numerička ako svi njeni koeficijenti $a_i (i= o,1,\dots,n)$ imaju numeričke vrednosti. Ako bar jedan od njih ima oblik opšteg broja, onda se radi o opštoj algebarskoj jednačini. U vezi s tim se i metodi za rešavanje algebarskih jednačina mogu podeliti na metode za rešavanje opštih algebarskih jednačina i na metode za rešavanje numeričkih jednačina.

Metodi prve grupe daju rešenje u obliku formule. Dokazano je(Vanstel, Abel i drugi) da su opšte algebarske jednačine, u opštem slučaju, rešive samo za $n \leqq 4$. Medjutim, numeričke algebarske jednačine je uvek moguće rešiti sa odredjenom tačnošću. Dobijeni koreni su tada u obliku približnog broja.

Ogroman broj metoda za numeričko rešavanje algebarskih jednačina prilagodjen je digitalnim računskim mašinama i oni se uglavnom mogu podeliti u dve grupe. U prvu grupu spadaju metodi za izračunavanje korena sa zadatom tačnošću bez korišćenja približnih vrednosti korena. Druga grupa sadrži metode za izračunavanje korena na osnovu, već poznatih približnih vrednosti korena.

Metodi jedne i druge grupe koji se najviše koriste u primenama biće izloženi u ovom poglavlju. Medju ovim metodima biće onih koji se koriste za odredjivanje samo jednog korena(na primer, najvećeg po modulu), kao i onih za simultano odredjivanje više korena(na primer, dva ili svih n).

Ukoliko je nekim od metoda odredjeno $k(<n)$ korena $x_1,\ldots,x_k$, tada se za odredjivanje ostalih korena rešava jednačina nižeg stepena

$$Q(x) = \frac{P(x)}{(x-x_1)\ldots(x-x_k)} = 0 .$$

S obzirom da se koreni $x_1,\ldots,x_k$ odredjuju sa ograničenom tačnošću, jasno je da koeficijenti jednačine $Q(x)=0$ neće biti apsolutno tačni, što znači da se tačnost korena $x_{k+1},\ldots,x_n$ narušava, bez obzira na metod kojim se oni odredjuju. Naime, koreni $x_{k+1},\ldots,x_n$ neće biti, u opštem slučaju, koreni jednačine $Q(x)=0$.

Proučimo sada uticaj promene koeficijenata algebarske jednačine (3.1.1) na promenu njenog prostog korena $x=x_k$. Ne umanjujući opštost razmatranja uzmimo da je $a_0=1$. Pod pretpostavkom da koeficijenti $a_i(i=1,\ldots,n)$ postanu redom $a_i+\varepsilon_i$

$(i=1,\ldots,n)$, gde je $|\varepsilon_i| \ll |a_i|$ $(i=1,\ldots,n)$, odredićemo grešku korena x_k. Naime, u tom slučaju koren x_k postaje $x_k+\xi_k$.

Kako je

$$(x_k+\xi_k)^n + (a_1+\varepsilon_1)(x_k+\xi_k)^{n-1} + \ldots$$
$$+ (a_{n-1}+\varepsilon_{n-1})(x_k+\xi_k) + (a_n+\varepsilon_n) = 0,$$

korišćenjem aproksimacije

$$(x_k+\xi_k)^m \cong x_k^m + m x_k^{m-1}\xi_k \qquad (m=2,\ldots,n)$$

i zanemarivanjem članova oblika $\xi_k\varepsilon_i$, dobijamo

$$P(x_k) + P'(x_k)\xi_k + (\varepsilon_1 x_k^{n-1} + \varepsilon_2 x_k^{n-2} + \ldots + \varepsilon_{n-1}x_k + \varepsilon_n) \cong 0,$$

odakle, s obzirom na $P(x_k)=0$ i $P'(x_k)\neq 0$, sleduje

$$\xi_k \cong -\frac{1}{P'(x_k)}(\varepsilon_1 x_k^{n-1} + \varepsilon_2 x_k^{n-2} + \ldots + \varepsilon_{n-1}x_k + \varepsilon_n).$$

Ako pretpostavimo da su granice apsolutnih grešaka koeficijenata jednake, tj. da je $|\varepsilon_i| \leq \varepsilon$, na osnovu poslednje približne jenakosti dobijamo granicu apsolutne greške korena, u oznaci Δ,

$$\Delta \cong \frac{1}{|P'(x_k)|}(|x_k|^{n-1} + |x_k|^{n-2} + \ldots + |x_k| + 1)\varepsilon,$$

tj.

$$\Delta = \begin{cases} \dfrac{(|x_k|^n - 1)\varepsilon}{|P'(x_k)|(|x_k|-1)} & (|x_k| \neq 1), \\[2ex] \dfrac{n\,\varepsilon}{|P'(x_k)|} & (|x_k| = 1). \end{cases}$$

U radu [29], J.H.Wilkinson navodi primer jednačine

$$P(x) = (x+1)(x+2)\ldots(x+20) = 0,$$

tj.

$$P(x) = x^{20} + 210x^{19} + \ldots + 20! = 0,$$

kod koje male promene u koeficijentima izazivaju velike promene u korenima. Naime, ako koeficijent $a_1 = 210$ promenimo za $\varepsilon_1 = 2^{-23} \cong 1.2 \cdot 10^{-7}$, promene u malim korenima su neznatne, ali koren $x_{20} = -20$ postaje -20.8. Takodje, pet parova korena postaju konjigovano kompleksni. Ovo je primer tzv. slabo uslovljenog polinoma.

Pri sukcesivnom odredjivanju korena jednačine (3.1.1) i njenoj redukciji na niži stepen, preporučuje se odredjivanje korena polazeći od najmanjeg po modulu.

Poseban problem kod rešavanja algebarskih jednačina je odredjivanje višestrukih korena ili korena koji su dovoljno bliski(kvazi-višestrukost). Ovaj problem nećemo posebno razmatrati. Napomenimo samo jedan dobro poznati postupak za eliminaciju višestrukih korena koji se sastoji u sledećem: Za polinom P i njegov izvodni polinom P' odredimo najveći zajednički delilac Q i formirajmo jednačinu

$$R(x) = \frac{P(x)}{Q(x)} = 0 .$$

Iz teorije polinoma je poznato(videti, na primer [2]), da svaka višestruka nula reda k polinoma P je istovremeno i nula reda $k-1$ polinoma P', pa i polinoma Q, odakle zaključujemo da polinom R, dobijen deobom polinoma P polinomom Q, sadrži sve nule polinoma P. Naravno, sve nule polinoma R su proste, što znači da problem rešavanja jednačine $R(x)=0$ nije više tako složen.

Za odredivanje najvećeg zajedničkog delioca dva polinoma koristi se Euklidov algoritam(videti [2]). Napomenimo da je najveći zajednički delilac jedinstven sa tačnošću do jedne multiplikativne konstante.

Primer 3.1.1. Neka je

$$P(x) = x^5 + 10x^4 + 42x^3 + 92x^2 + 105x + 50 = 0.$$

S obzirom da je najveći zajednički delilac za P i P' polinom $x \mapsto Q(x) = x^2 + 4x + 5$, imamo

$$R(x) = x^3 + 6x^2 + 13x + 10 = 0.$$

5.3.2. Granice korena algebarskih jednačina

Posmatrajmo jednačinu (3.1.1) u kompleksnoj ravni, tj. neka je $x=z$. Stavimo

$$a = \max(|a_1|, |a_2|, \dots, |a_n|) \quad \text{i} \quad a' = \max(|a_0|, |a_1|, \dots, |a_{n-1}|).$$

Teorema 3.2.1. Svi koreni jednačine (3.1.1) leže u prstenu

$$\frac{|a_n|}{a' + |a_n|} < |z| < 1 + \frac{a}{|a_0|} . \tag{3.2.1}$$

Dokaz. S obzirom da je za $|z|>1$

$$|P(z)| \geq \Big| |a_0 z^n| - |a_1 z^{n-1} + \dots + a_{n-1} z + a_n| \Big| ,$$

imamo

$$|a_1 z^{n-1} + \dots + a_{n-1} z + a_n| \leq a(|z|^{n-1} + |z|^{n-2} + \dots + |z| + 1)$$

$$= a \frac{|z|^n - 1}{|z| - 1} < \frac{a|z|^n}{|z| - 1} .$$

Dakle, $|P(z)| > 0$, ako je

$$|a_0 z^n| - a \frac{|z|^n}{|z|-1} \geq 0,$$

tj.

$$|z| \geq 1 + \frac{a}{|a_0|} .$$

Kako vrednosti za z koje zadovoljavaju prethodnu nejednakost ne mogu biti koreni jednačine (3.1.1), zaključujemo da svi koreni jednačine (3.1.1) zadovoljavaju suprotnu nejednakost, tj. leže u unutrašnjosti kruga

$$|z| = 1 + \frac{a}{|a_0|} .$$

Uvodjenjem smene $z = \frac{1}{\zeta}$ jednačina (3.1.1) postaje

$$a_0 + a_1 \zeta + \dots + a_{n-1} \zeta^{n-1} + a_n \zeta^n = 0. \tag{3.2.2}$$

Na osnovu prethodnog koreni jednačine (3.2.2) zadovoljavaju nejednakost

$$|\zeta| = \frac{1}{|z|} < 1 + \frac{a'}{|a_n|},$$

tj.

$$|z| > \frac{|a_n|}{a' + |a_n|}.$$

Ovim je dokaz teoreme završen.

<u>Primer 3.2.1</u>. Ispitajmo u kojoj oblasti leže koreni jednačine

$$z^7 - (3+i)z^6 - (15-3i)z^5 + (17+15i)z^4 + (38-17i)z^3$$
$$+ (22-38i)z^2 - (60+22i)z + 60i = 0.$$

Kako je

$$a = \max(|a_1|, |a_2|, \ldots, |a_7|) = |a_6| = |-(60+22i)| \cong 63.9,$$

$$a' = \max(|a_0|, |a_1|, \ldots, |a_6|) = |a_6| \cong 63.9,$$

imamo

$$\frac{60}{63.9+60} < |z| < 1 + \frac{63.9}{1},$$

tj. $0.48 < |z| < 64.9$.

Posmatrajmo sada jednačinu

$$(3.2.3) \qquad P(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n = 0 \quad (a_0 \neq 0)$$

sa realnim koeficijentima i $a_0 > 0$.

Sledeća teorema daje granicu realnih korena jednačine (3.2.3).

<u>Teorema 3.2.2</u>. Neka je A maksimum apsolutnih vrednosti negativnih koeficijenata jednačine (3.2.3) i neka je a_m prvi negativni koeficijent u nizu $a_0, a_1, \ldots, a_n$, tada su svi pozitivni koreni ove jednačine manji od

$$R = 1 + \sqrt[m]{\frac{A}{a_0}}.$$

Posledica 3.2.3. Ako su svi koeficijenti jednačine (3.2.3) nenegativni, jednačina nema pozitivnih korena.

Primedba 3.2.1. Donja granica pozitivnih korena može se odrediti na sličan način uz prethodno uvodjenje smene $x = 1/y$ u jednačini (3.2.3).

Primer 3.2.2. Odredimo granice realnih korena jednačine

$$2x^9 + x^7 - x^4 + 19x^3 - 24x^2 + 11 = 0. \tag{3.2.4}$$

Kako je $a_0=2$, $A=24$, $m=5$, imamo

$$R = 1 + \sqrt[5]{\frac{24}{2}} < 2.65.$$

Za odredjivanje donje granice pozitivnih korena posmatrajmo jednačinu

$$11y^9 - 24y^7 + 19y^6 - y^5 + y^2 + 2 = 0.$$

Ovde je $\bar{a}_0 = 11$, $\bar{A} = 24$, $\bar{m} = 2$, pa je

$$\bar{R} = 1 + \sqrt{\frac{24}{11}} < 2.48,$$

tj. donja granica pozitivnih korena je $r = \frac{1}{\bar{R}} > 0.40$. Dakle, pozitivni koreni jednačine (3.2.4) leže u intervalu $(0.40, 2.65)$.

Smenom $x=-t$, jednačina (3.2.4) postaje

$$2t^9 + t^7 + t^4 + 19t^3 + 24t^2 - 11 = 0,$$

Primenom sličnog postupka nalazimo da pozitivni koreni ove jednačine leže u intervalu $(0.40, 2.21)$.

Dakle, negativni koreni jednačine (3.2.4) leže u intervalu $(-2.21, -0.40)$.

Primedba 3.2.2. Za nalaženje gornje granice pozitivnih korena jednačine (3.2.3) može se koristiti i Newtonov metod koji se sastoji u sledećem:

Ako za $x = a > 0$ važe nejednakosti

$$P^{(k)}(a) \geqq 0 \qquad (k=0,1,\ldots,n),$$

za gornju granicu korena može se uzeti vrednost $R=a$, što sleduje iz Taylorove formule

$$P(x) = \sum_{k=0}^{n} \frac{1}{n!} P^{(k)}(a)(x-a)^k .$$

Naime, za $x>a$ imamo $P(x)>0$.

Sledećim postupkom može se dobiti bolja granica: Predstavimo $P(x)$ u obliku

$$P(x) = Q_1(x) - Q_2(x) + Q_3(x) - \ldots + Q_{2m-1}(x) - Q_{2m}(x),$$

gde su $Q_1(x)$ suma prvih članova $P(x)$ sa pozitivnim koeficijentima, počev sa a_0x^n, $-Q_2(x)$ suma narednih članova $P(x)$ sa negativnim koeficijentima, itd., pri čemu poslednji član $-Q_{2m}(x)$ ili se sastoji iz članova sa negativnim koeficijentima ili je jednak nuli. Ako su c_j $(j=1,\ldots,m)$ pozitivni brojevi takvi da je

$$Q_{2j-1}(c_j) - Q_{2j}(c_j) > 0 \qquad (j=1,\ldots,m),$$

za gornju granicu pozitivnih korena može se uzeti broj

$$R = \max(c_1,\ldots,c_m).$$

Navedimo sada jedan rezultat, koji može biti korisniji od teoreme 3.2.1.

Neka je

$$P(z) = z^n + a_1z^{n-1} + a_2z^{n-2} + \ldots + a_{n-1}z + a_n .$$

Uvodjenjem smene $z = w - \frac{a_1}{n}$, polinom P se transformiše u polinom Q, u kome je koeficijent uz w^{n-1} jednak nuli. Dakle,

$$Q(w) = P\left(w - \frac{a_1}{n}\right) = w^n + c_2w^{n-2} + \ldots + c_{n-1}w + c_n .$$

Definišimo polinom S pomoću

$$S(w) = w^n - |c_2|w^{n-2} - \ldots - |c_{n-1}|w - |c_n| .$$

Teorema 3.2.3. Ako je bar jedan od koeficijenata $c_k \neq 0$ $(k = 2, \ldots, n)$, sve nule polinoma P leže u oblasti

$$\left| z + \frac{a_1}{n} \right| \leqq r,$$

gde je r pozitivna nula polinoma S.

Dokaz. Neka je $|w| > r$. Pod pretpostavkom da je bar jedan od koeficijenata $c_k \neq 0$ $(k=2,\ldots,n)$ i s obzirom da je $S(r) = 0$, imamo

$$\left| \frac{c_2}{w^2} + \ldots + \frac{c_n}{w^n} \right| < \frac{|c_2|}{r^2} + \ldots + \frac{|c_n|}{r^n} = 1 .$$

Kako je

$$Q(w) = w^n\left(1 + \frac{c_2}{w^2} + \ldots + \frac{c_n}{w^n}\right),$$

na osnovu prethodnog imamo

$$|Q(w)| \geqq |w|^n \left(1 - \left| \sum_{i=2}^{n} \frac{c_i}{w^i} \right| \right) > 0,$$

odakle zaključujemo da je $Q(w) \neq 0$ za svako w za koje je $|w| > r$. Drugim rečima, nule polinoma Q nalaze se u krugu $|w| \leqq r$, što je s obzirom na smenu $z = w - \frac{a_1}{n}$, ekvivalentno sa tvrdjenjem teoreme 3.2.3.

Primer 3.2.3. Neka je

$$P(z) = z^5 - 10z^4 + 43z^3 - 104z^2 + 150z - 100. \tag{3.2.5}$$

Smenom $z = w - \frac{(-10)}{5} = w + 2$ dobijamo

$$Q(w) = P(w+2) = w^5 + 3w^3 - 6w^2 + 10w$$

i

$$S(w) = w^5 - 3w^3 - 6w^2 - 10w .$$

Na osnovu teoreme 3.2.2, zaključujemo da je jedinstveni pozitivni koren $w=r$ jednačine $S(w) = 0$ manji od $R = 1 + \sqrt[3]{10} \cong 3.15$. Ova granica se može poboljšati. Naime, nije teško ustanoviti da je $r < 3$. Tada, na osnovu teoreme 3.2.3, zaključujemo da se nule polinoma (3.2.5) nalaze u krugu $|z-2| \leq r < 3$. Napomenimo da su tačne nule polinoma P redom

$$z_1 = 2, \quad z_{2,3} = 1 \pm 2i, \quad z_{4,5} = 3 \pm i.$$

Teorema 3.2.1 daje loše granice $0.4 < |z| < 151$.

5.3.3. Bernoulliev metod

U ovom odeljku izložićemo jedan vrlo jednostavan metod za odredjivanje korena algebarskih jednačina bez korišćenja približnih vrednosti ovih korena. Prve ideje o ovom metodu dao je D. Bernoulli 1728. godine. Sa pojavom digitalnih računskih mašina ovaj metod je znatno usavršen.

Neka je data jednačina

$$a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n = 0 \quad (a_0 \neq 0) \tag{3.3.1}$$

sa realnim koeficijentima. Ne umanjujući opštost možemo uzeti $a_0 = 1$. Jednačina (3.3.1) može se shvatiti kao karakteristična jednačina linearne diferencne jednačine n-tog reda

$$y_{n+k} + a_1 y_{n+k-1} + \dots + a_{n-1} y_{k+1} + a_n y_k = 0. \tag{3.3.2}$$

Rešenje ove jednačine odredjeno je korenima karakteristične jednačine (3.3.1). Naime, ako su x_i $(i=1,\dots,n)$ koreni jednačine (3.3.1), tada je

$$y_k = C_1 \varphi_1(k) + \dots + C_n \varphi_n(k)$$

rešenje diferencne jednačine (3.3.2), gde su funkcije φ_i odredjene korenima x_i $(i=1,\dots,n)$ respektivno(videti odeljak 1.3.1), a konstante C_i $(i=1,\dots,n)$ zavise od početnih uslova $y_0, y_1, \dots, y_{n-1}$.

Ne umanjujući opštost uzećemo da je

$$|x_1| \geq |x_2| \geq \dots \geq |x_n|.$$

Koren sa najvećim modulom zvaćemo dominamtnim korenom. Očigledno je da jednačina (3.3.1) može imati

1. jedan dominantan koren;
2. više dominantnih korena.

Ukoliko postoji samo jedan dominantan koren x_1, tada je on realan.

Više dominantnih korena imamo kada je

$$|x_1| = |x_2| = \dots = |x_r| > |x_{r+1}| \geq \dots \geq |x_n|,$$

pri čemu treba razlikovati slučajeve:

2.1. Dominantni koreni realni i jednaki(višestruk koren reda r), tj.

$$x_1 = \dots = x_r ;$$

2.2. Dominantni koreni realni i suprotni po znaku(jedan višestruk reda p, a drugi reda r-p), tj.

$$x_1 = \dots = x_p = -x_{p+1} = \dots = -x_r > 0 ;$$

2.3. Dominantni koreni konjugovano kompleksni(u parovima), tj.

$$x_1 = \bar{x}_2 , x_3 = \bar{x}_4 , \dots , x_{2m-1} = \bar{x}_{2m} \quad (r = 2m).$$

Primetimo da, u ovom slučaju, medju dominantnim korenima može biti i višestrukih.

2.4. Medju dominantnim korenima ima i realnih i konjugovano kompleksnih.

Bernoulliev metod omogućuje nalaženje dominantnih korena jednačine (3.3.1), pri čemu koristi niz $\{y_k\}_{k\in N_0}$, koji se generiše pomoću (3.3.2) sa početnim vrednostima $y_0, y_1, \dots, y_{n-1}$.

Razmotrimo najpre slučaj 1, tj. slučaj kada imamo jedan dominantan koren x_1. Tada je

$$(3.3.3) \qquad y_k = C_1 x_1^k + C_2 \varphi_2(k) + \dots + C_n \varphi_n(k) .$$

Nije teško pokazati da je

$$(3.3.4) \qquad \lim_{k\to +\infty} \frac{\varphi_i(k)}{x_1^k} = 0 \qquad (i = 2,\dots,n).$$

Korišćenjem niza $\{y_k\}$ konstruišimo niz $\{u_k\}$ pomoću $u_k = \frac{y_{k+1}}{y_k}$. Tada je, s obzirom na (3.3.3),

$$u_k = x_1 \frac{1 + \frac{C_2}{C_1}\cdot\frac{\varphi_2(k+1)}{x_1^{k+1}} + \dots + \frac{C_n}{C_1}\cdot\frac{\varphi_n(k+1)}{x_1^{k+1}}}{1 + \frac{C_2}{C_1}\cdot\frac{\varphi_2(k)}{x_1^{k}} + \dots + \frac{C_n}{C_1}\cdot\frac{\varphi_n(k)}{x_1^{k}}},$$

pri čemu smo pretpostavili da je $C_1 \neq 0$.

Na osnovu (3.3.4) imamo

$$\lim_{k\to+\infty} u_k = x_1, \tag{3.3.5}$$

tj. niz $\{u_k\}$ konvergira ka dominantnom korenu x_1. Za primenu ovog metoda, treba obezbediti uslov $C_1 \neq 0$, što zavisi od početnih uslova. Na primer, ovaj uslov je ispunjen ako se uzme

$$y_0 = y_1 = \dots = y_{n-2} = 0, \quad y_{n-1} = 1.$$

Ispitajmo sada konvergenciju niza $\{u_k\}$ pod uslovom da je koren x_2 realan i takav da je $|x_2| > |x_i|$ $(i = 3,\dots,n)$. Sa e_k označimo grešku $u_k - x_1$. Kako je, pod navedenim uslovom, $\varphi_2(k) = x_2^k$, za dovoljno veliko k imamo

$$e_k = \frac{y_{k+1}}{y_k} - x_1 \cong \frac{C_2}{C_1}(x_2 - x_1)\left(\frac{x_2}{x_1}\right)^k,$$

tj. $e_{k+1} \cong \frac{x_2}{x_1} e_k$, odakle zaključujemo da je konvergencija linearna.

U slučaju 2.1, umesto rešenja (3.3.3) imamo rešenje

$$y_k = (C_1 + C_2 k + \dots + C_r x^{r-1}) x_1^r + C_{r+1}\varphi_{r+1}(k) + \dots + C_n \varphi_n(k).$$

Može se pokazati da i u ovom slučaju važi (3.3.5).

<u>Primer 3.3.1.</u> Neka je data jednačina

$$x^4 - x^3 - 3x^2 - 7x - 6 = 0. \tag{3.3.6}$$

Diferencna jednačina (3.3.2), u ovom slučaju, postaje

$$y_{k+4} - y_{k+3} - 3y_{k+2} - 7y_{k+1} - 6y_k = 0.$$

Startujući sa $y_o = y_1 = y_2 = 0$, $y_3 = 1$, na osnovu prethodnog, dobijamo niz $\{y_k\}$, a zatim niz $\{u_k\}$.

k	y_k	u_k
3	1	1.
4	1	4.
5	4	3.5
6	14	2.7857143
7	39	2.9487179
8	115	3.0782609
9	354	2.9830508
10	1056	2.9895833
11	3157	3.0069686
12	9493	

Primetimo da je $x_1 = 3$ dominantan koren jednačine (3.3.6).

U slučaju 2.2, imamo

$$y_k = \left[C_1 + C_2 k + \dots + C_p k^{p-1} + (-1)^k (C_{p+1} + \dots + C_r k^{r-1})\right] x_1^k$$

$$+ C_{r+1} \varphi_{r+1}(k) + \dots + C_n \varphi_n(k) ,$$

odakle lako možemo zaključiti da niz $\{u_k\}$, gde je $u_k = \dfrac{y_{k+1}}{y_k}$, divergira. Mogućno je, medjutim, i u ovom slučaju odrediti x_1, ali pomoću niza $\{v_k\}$, gde je $v_k = \dfrac{y_{2k+2}}{y_{2k}}$. Naime, ovde je

$$\lim_{k \to +\infty} v_k = x_1^2 .$$

<u>Primer 3.3.2</u>. Posmatrajmo jednačinu

$$x^3 + 0.5x^2 - 4x - 2 = 0,$$

čiji su koreni $x_1 = -x_2 = 2$, $x_3 = -0.5$. Startujući sa $y_o = y_1 = 0$, $y_2 = 1$, nalazimo nizove $\{y_k\}$, $\{u_k\}$, $\{v_k\}$. Primetimo da niz $\{u_k\}$ divergira, a da niz $\{v_k\}$ konvergira, što je kriterijum za egzistenciju slučaja 2.2. Niz $\{v_k\}$ konvergira ka $x_1^2 = 4$.

k	y_k	u_k	$v_{k/2}$
2	1.	-0.5	4.25
3	-0.5	-0.85	
4	4.25	-0.9705882	4.25
5	-4.125	-4.3787879	
6	18.0625	-0.9429066	4.0147059
7	-17.03125	-4.2577982	
8	72.51562	-0.9412842	4.0009158
9	-68.25781	4.2504866	
10	290.12891		

Razmotrimo sada slučaj 2.3. Jednostavnosti radi uzmimo da je $r=2$, tj. da imamo samo jedan par konjugovano kompleksnih korena, koji su dominantni. Neka su to $x_1=\rho e^{i\theta}$, $x_2=\rho e^{-i\theta}$.

Tada je

$$C_1\varphi_1(k)+C_2\varphi_2(k)=\rho^k\left[A_1\cos k\theta+A_2\sin k\theta\right]$$
$$=A\rho^k\cos(k\theta+\Psi),$$

pri čemu postoji odredjena veza izmedju konstanti koje se pojavljuju u ovoj jednakosti.

Štaviše, za dovoljno veliko k, važi

$$y_k=A\rho^k\cos(k\theta+\Psi)+O(|x_3|^k). \tag{3.3.7}$$

Lako je videti da ranije definisani nizovi $\{u_k\}$ i $\{v_k\}$, u ovom slučaju divergiraju. Zato definišimo nove nizove $\{s_k\}$ i $\{t_k\}$ pomoću

$$s_k=\begin{vmatrix} y_k & y_{k+1} \\ y_{k-1} & y_k \end{vmatrix}=y_k^2-y_{k-1}y_{k+1}$$

i

$$t_k=\begin{vmatrix} y_{k+1} & y_{k+2} \\ y_{k-1} & y_k \end{vmatrix}=y_{k+1}y_k-y_{k-1}y_{k+2}\,.$$

Na osnovu (3.3.7), za dovoljno veliko k, imamo

$$s_k \cong A^2\rho^{2k}\left[\cos^2(k\theta+\Psi)-\cos((k-1)\theta+\Psi)\cos((k+1)\theta+\Psi)\right]$$

i

$$t_k \cong A^2\rho^{2k+1}\left[\cos((k+1)\theta+\Psi)\cos(k\theta+\Psi)-\cos((k-1)\theta+\Psi)\cos((k+2)\theta+\Psi)\right],$$

tj.

$$s_k \cong A^2\rho^{2k}\sin^2\theta \quad \text{i} \quad t_k \cong 2A^2\rho^{2k+1}\sin^2\theta\cos\theta,$$

odakle zaključujemo da je

$$\lim_{k\to+\infty}\frac{s_{k+1}}{s_k}=\rho^2 \quad \text{i} \quad \lim_{k\to+\infty}\frac{t_k}{2s_k}=\rho\cos\theta.$$

Dakle, u posmatranom slučaju, niz $\left\{\frac{s_{k+1}}{s_k}\right\}$ konvergira ka ρ^2, a niz $\left\{\frac{t_k}{2s_k}\right\}$ ka realnom delu dominantnog korena.

Primer 3.3.3. Neka je data jednačina

$$x^3+x^2+x-1=0,$$

kod koje je par konjugovano kompleksnih korena dominantan. Polazeći od $y_0=y_1=0$ i $y_2=1$, na osnovu

$$y_{k+3}=-y_{k+2}-y_{k+1}+y_k \qquad (k=0,1,\ldots),$$

konstruimo niz $\{y_k\}$. Dakle, imamo

$$\{y_k\}=\{0,0,1,-1,0,2,-3,1,4,-2,-1,7,-8,0,15,-23,8,30,-61,\\ 39,52,\ldots\}.$$

Kako je $s_{18}=(-61)^2-39\cdot 30=2551$, $s_{19}=39^2-(-61)\cdot 52=4693$, $t_{18}=39(-61)-52\cdot 30=-3939$ nalazimo

$$\frac{s_{19}}{s_{18}}\cong 1.8397 \quad \text{i} \quad \frac{t_{18}}{2s_{19}}\cong -0.7720,$$

odakle je

$$\begin{aligned} x_1 &\cong -0.7720+(1.8397-(-0.7720)^2)^{1/2}i \\ &\cong -0.7720+1.1152\,i \end{aligned}$$

i $x_2=\bar{x}_1$.

Slučaj tri dominantna korena, od kojih je jedan realan, a dva konjugovano kompleksna, razmatran je u radu [30].

Primedba 3.3.1. Ako se u jednačini (3.3.1) izvrši smena $x=1/z$ i na rešavanje dobijene jednačine primeni Bernoulliev metod mogu se odrediti najmanji po modulu koreni jednačine (3.3.1).

5.3.4. Dva metoda trećeg reda

Pretpostavimo da su sve nule $\xi_i (i=1,\ldots,n)$ polinoma

$$P(x)=x^n+a_1x^{n-1}+\ldots+a_{n-1}x+a_n$$

realne i različite i neka je $\xi_1<\xi_2<\ldots<\xi_n$. Tada, na osnovu Rolleove teoreme, izvodni polinom P' ima $n-1$ nula koje su takodje realne i medjusobno različite. Označimo ih redom sa $\eta_1,\ldots,\eta_{n-1}$, tako da je

$$\xi_1<\eta_1<\xi_2<\ldots<\eta_{n-1}<\xi_n .$$

Neka je $X=R\setminus\{\xi_1,\eta_1,\xi_2,\ldots,\eta_{n-1},\xi_n\}$. Definišimo funkciju $a:X\to\{\xi_1,\xi_2,\ldots,\xi_n\}$ pomoću

$$a(x)=\begin{cases}\xi_1 & (x<\xi_1)\\ \xi_i & (\xi_i<x<\eta_i)\\ \xi_{i+1} & (\eta_i<x<\xi_{i+1})\\ \xi_n & (x<\xi_n)\end{cases}\qquad (i=1,\ldots,n-1).$$

Primetimo da u svakom intervalu izmedju $a(x)$ i x, izraz $P(x)P'(x)$ ne menja znak. Naime, važi

$$\operatorname{sgn}(x-a(x))=\operatorname{sgn}(P(x)P'(x)) . \tag{3.4.1}$$

Kako je $P(x)=\prod_{i=1}^{n}(x-\xi_i)$, imamo

$$\frac{P'(x)}{P(x)}=\sum_{i=1}^{n}\frac{1}{x-\xi_i}\quad \text{i}\quad H(x)=\frac{P'(x)^2-P(x)P''(x)}{P(x)^2}=\sum_{i=1}^{n}\frac{1}{(x-\xi_i)^2}. \tag{3.4.2}$$

S obzirom da je $H(x)>0$, definišimo funkciju K pomoću

$$K(x)=\frac{\operatorname{sgn}(P(x)P'(x))}{\sqrt{H(x)}} .$$

Nije teško videti da je

$$K(x)=\frac{P(x)/P'(x)}{\sqrt{1-P(x)P''(x)/P'(x)^2}} . \tag{3.4.3}$$

Uzimajući $x_0=x\ (x\in X)$ konstruišimo iterativni proces

(3.4.4) $$x_{k+1}=x_k-K(x_k) \qquad (k=0,1,\ldots),$$

koji je u literaturi poznat kao metod kvadratnog korena(videti [6]).

Teorema 3.4.1. Niz $\{x_k\}_{k\in N_0}$ koji se generiše pomoću (3.4.4) konvergira monotono ka nuli $a=a(x)$, pri čemu je konvergencija kubna, tj. važi

(3.4.5) $$\lim_{k\to+\infty}\frac{x_{k+1}-a}{(x_k-a)^3}=\frac{3P''(a)^2-4P'(a)P'''(a)}{24P'(a)^2}.$$

Dokaz. Neka je $x_0=x<a(x)$. Tada je, s obzirom na (3.4.1), $\operatorname{sgn}(P(x)P'(x))=-1$, odakle zaključujemo da je $K(x)<0$,tj. da je $x_1=x_0-K(x_0)>x_0$. S druge strane, na osnovu druge jednakosti u (3.4.2), imamo

$$H(x)>\frac{1}{(x-a(x))^2},$$

odakle sleduje $a(x)-x>H(x)^{-1/2}=-K(x)$, tj. $x_1<a(x)$. Dakle, važi $x_0<x_1<a(x)$ i $a(x_0)=a(x_1)$. Nastavljajući sa ovakvim rezonovanjem dokazujemo da je

$$x_0<x_1<x_2<\cdots<a(x),$$

što znači da niz $\{x_k\}_{k\in N_0}$ monotono konvergira ka nekoj tački $a\le a(x)$. Kako je, u tom slučaju, $\lim\limits_{k\to+\infty}(x_{k+1}-x_k)=0$, na osnovu (3.4.4) zaključujemo da je $K(a)=0(\Rightarrow P(a)=0)$, odnosno da je $a=a(x)$.

Slučaj $x_0=x>a(x)$ dokazuje se potpuno simetrično.

Da bismo dokazali jednakost (3.4.5) stavimo $h=\dfrac{P(x)}{P'(x)}$ i pustimo da $x\to a$. Na osnovu (3.4.3) i (3.4.4) imamo

$$x_1-x=-K(x)=h\Big(1+h\frac{P''(x)}{P'(x)}\Big)^{-1/2},$$

tj.

$$x_1-x=h-\frac{1}{2}h^2\frac{P''(x)}{P'(x)}+\frac{3}{8}h^3\Big(\frac{P''(x)}{P'(x)}\Big)^2+O(h^4).$$

S druge strane, na osnovu Schröderovog razvoja (1.6.3) imamo

$$a-x = h - h^2 \frac{P''(x)}{2P'(x)} + h^3 \frac{3P''(x)^2 - P'(x)P'''(x)}{6P'(x)^2} + O(h^4)\,.$$

Ako poslednju jednakost oduzmemo od prethodne jednakosti, dobijamo

$$x_1 - x = -h^3 \frac{3P''(x)^2 - 4P'(x)P'''(x)}{24P'(x)^2} + O(h^4),$$

odakle, s obzirom da je $h \sim a-x$ $(x \to a)$ sleduje (3.4.5), uz prethodnu zamenu x sa $x_k (k \to +\infty)$.

Ovim je dokaz teoreme završen.

Dakle, metod kvadratnog korena ima kubnu konvergenciju. Medjutim, može se razmatrati slučaj kada je a višestruka nula reda p; tada je konvergencija prvog reda i važi

$$\lim_{k\to+\infty} \frac{x_{k+1}-a}{x_k-a} = 1 - \frac{1}{\sqrt{p}}\,.$$

U tesnoj vezi sa metodom kvadratnog korena je Laguerreov metod

$$x_{k+1} = x_k - \frac{nP(x_k)}{P'(x_k) + \operatorname{sgn}(P'(x_k))\sqrt{G(x_k)}} \quad (k=0,1,\ldots), \tag{3.4.6}$$

gde je $x_0 = x$ $(x \in X)$ i

$$G(t) = (n-1)\Big[(n-1)P'(t)^2 - nP(t)P''(t)\Big]\,.$$

Metod (3.4.6) se može predstaviti i u obliku

$$x_{k+1} = x_k - \frac{nP(x_k)/P'(x_k)}{1 + (n-1)\sqrt{1 - \frac{n}{n-1}P(x_k)P''(x_k)/P'(x_k)^2}} \quad (k=0,1,\ldots).$$

Red konvergencije Laguerreovog metoda je tri za proste nule, a jedan za višestruke. Slično dokazu teoreme 3.4.1 može se dokazati odgovarajuća teorema za ovaj metod. Formula (3.4.5) ostaje u važnosti, s tim što se konstanta 3(uz $P''(a)^2$) na desnoj strani ove formule, treba zameniti konstantom $3(n-2)/(n-1)$. Laguerreov metod je efikasniji od metoda kvadratnog korena, na šta ukazuje sledeći primer.

Primer 3.4.1. Neka je

$$P(x) = (x-1)(x-2)(x-3) = x^3 - 6x^2 + 11x - 6 .$$

Startujući sa $x_o = 0$, dobijamo

	pomoću (3.4.4)	pomoću (3.4.6)
x_1	0.85714286	0.98840803
x_2	0.99858769	0.99999991

U radu [14] razmatrana je jedna opštija klasa metoda, koja kao partikularne slučajeve sadrži prethodno navedena dva metoda

5.3.5. Newton-Hornerov metod

Newtonov metod izložen u odeljku 5.1.2 može se uspešno primeniti za odredjivanje nula polinoma

$$(3.5.1) \qquad P(x) = a_o x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n \qquad (a_i \in R),$$

pri čemu se vrednosti za $P(x_k)$ i $P'(x_k)$ izračunavaju po Hornerovoj šemi(videti odeljak 1.3.4).

Ako stavimo $p_j(x) = \frac{1}{j!} P^{(j)}(x)$ $(j = 0,1,\ldots,n)$, tada je, na osnovu Taylorove formule,

$$P(x) = p_o(x_k) + p_1(x_k)(x-x_k) + \ldots + p_n(x_k)(x-x_k)^n .$$

Pretpostavimo da polinom P ima realnu nulu $x = \xi$. Newtonov metod (1.2.3) za odredjivanje ove nule postaje

$$(3.5.2) \qquad x_{k+1} = x_k - \frac{p_o(x_k)}{p_1(x_k)} \qquad (k = 0,1,\ldots).$$

Kao što je poznato, vrednost za $p_o(x_k)$ može se dobiti iz n koraka Hornerovom šemom, kao $p_o(x_k) = b_n$, gde je

$$(3.5.3) \qquad b_o = a_o, \quad b_j = b_{j-1} x_k + a_j \qquad (j = 1,\ldots,n).$$

Ustvari $b_j (j = 0,1,\ldots,n-1)$ su koeficijenti polinoma dobijenog deljenjem $P(x)$ sa $x-x_k$, tj. važi

$$P(x) = p_1(x)(x-x_k) + p_o(x_k)$$

i

$$p_1(x) = b_o x^{n-1} + b_1 x^{n-2} + \ldots + b_{n-1} .$$

Slično deljenjem $p_1(x)$ sa $x-x_k$ dobijamo

$$p_2(x) = c_o x^{n-2} + \ldots + c_{n-2},$$

tj. važi

$$p_1(x) = p_2(x)(x-x_k) + p_1(x_k) .$$

Ovde je $c_o = b_o$, $c_j = c_{j-1}x_k + b_j$ $(j = 1,\ldots,n-1)$ i $p_1(x_k) = c_{n-1}$.

Navedena deljenja se šematski mogu prikazati u obliku

	a_o	a_1	$\ldots$	a_j	$\ldots$	a_{n-1}	a_n
x_k		$b_o x_k$		$b_{j-1}x_k$		$b_{n-2}x_k$	$b_{n-1}x_k$
	b_o	b_1		b_j		b_{n-1}	$\boxed{b_n = p_o(x_k)}$
x_k		$c_o x_k$		$c_{j-1}x_k$		$c_{n-2}x_k$	
	c_o	c_1		c_j		$\boxed{c_{n-1} = p_1(x_k)}$	

Dakle, korišćenjem Hornerove šeme i formule (3.5.2) može se odrediti realna nula $x = \xi$. Napomenimo da proces (3.5.2) ima kvadratnu konvergenciju ukoliko je nula prosta.

S obzirom da polinom (3.5.1) može imati i konjugovano-kompleksne nule, navedeni metod se može prilagoditi i za ovaj slučaj.

Posmatrajmo polinom (3.5.1) u kome je x zamenjeno sa $z = x + iy$. Tada razdvajanjem realnog i imaginarnog dela u $P(z)$ imamo

$$P(z) = u(x,y) + iv(x,y),$$

gde su u i v harmonijske funkcije. Izračunavanje vrednosti ovih funkcija u tački (x_k,y_k) mogućno je po istom postupku kao što je (3.5.3), s tim što b_j treba zameniti sa $\alpha_j + i\beta_j$ $(j = 0,1,\ldots,n)$. Tada je

$$\alpha_o = a_o, \quad \beta_o = 0,$$

$$\alpha_j = \alpha_{j-1}x_k - \beta_{j-1}y_k + a_j, \quad \beta_j = \alpha_{j-1}y_k + \beta_{j-1}x_k \quad (j=1,\ldots,n)$$

i $u_k = u(x_k,y_k) = \alpha_n$, $v_k = v(x_k,y_k) = \beta_n$.

Neka je $z = \xi + i\eta$ koren jednačine $P(z) = 0$. S obzirom da je ova jednačina ekvivalentna sistemu jednačina

$$u(x,y) = 0, \quad v(x,y) = 0, \tag{3.5.4}$$

to se za odredjivanje korena $z = \xi + i\eta$ može konstruisati niz $z_k = x_k + iy_k (k = 0,1,\ldots)$, na primer, pomoću metoda Newton-Kantoroviča. Imajući u vidu da funkcije u i v zadovoljavaju Cauchy-Riemannove uslove, lako se pokazuje da je metod Newton-Kantoroviča, u ovom slučaju, ekvivalentan sa kompleksnom verzijom Newtonovog metoda

$$z_{k+1} = z_k - \frac{P(z_k)}{P'(z_k)} \qquad (k = 0,1,\ldots). \tag{3.5.5}$$

Vrednost $P'(z_k) = u_k' + iv_k' = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}\Big|_{x=x_k,y=y_k}$ može se odrediti rekurzivno pomoću

$$\gamma_0 = \alpha_0 = a_0, \ \delta_0 = 0,$$

$$\gamma_j = \gamma_{j-1}x_k - \delta_{j-1}y_k + \alpha_j, \ \delta_j = \gamma_{j-1}y_k + \delta_{j-1}x_k + \beta_j \ (j=1,\ldots,n-1).$$

Naime, $u_k' = \gamma_{n-1}$ i $v_k' = \delta_{n-1}$.

<u>Primer 3.5.1</u>. Neka je $P(z) = z^3 + 3z^2 + z - 1$ i $z=1+i$. Na osnovu prethodno datih jednakosti imamo

$$\alpha_0 = 1, \ \beta_0 = 0,$$

$$\alpha_1 = 1\cdot1 - 0\cdot1 + 3 = 4, \ \beta_1 = 1\cdot1 + 0\cdot1 = 1,$$

$$\alpha_2 = 4\cdot1 - 1\cdot1 + 1 = 4, \ \beta_2 = 4\cdot1 + 1\cdot1 = 5,$$

$$\alpha_3 = 4\cdot1 - 5\cdot1 - 1 = -2, \ \beta_3 = 4\cdot1 + 5\cdot1 = 9$$

i

$$\gamma_0 = 1, \ \delta_0 = 0,$$

$$\gamma_1 = 1\cdot1 - 0\cdot1 + 4 = 5, \ \delta_1 = 1\cdot1 + 0\cdot1 + 1 = 2,$$

$$\gamma_2 = 5\cdot1 - 2\cdot1 + 4 = 7, \ \delta_2 = 5\cdot1 + 2\cdot1 + 5 = 12,$$

pa je $P(z) = -2 + 9i$, $P'(z) = 7 + 12i$.

Koririšćenjem prethodno datih rekurentnih relacija, formula (3.5.5) se svodi na

$$
(3.5.6)\qquad
\begin{aligned}
x_{k+1} &= x_k - \frac{u_k u_k' + v_k v_k'}{(u_k')^2 + (v_k')^2}\\
y_{k+1} &= y_k - \frac{v_k u_k' - u_k v_k'}{(u_k')^2 + (v_k')^2}
\end{aligned}
\qquad (k = 0, 1, \ldots).
$$

Kod primene Newton-Hornerovog metoda obično se startuje sa dovoljno malim vrednostima (po modulu) x_o i y_o i primenom formula (3.5.6) odredjuje se najmanja nula po modulu $z_1 = \xi_1 + i\eta_1$. Ako je $|\eta_1| < \varepsilon$ (ε dovolno mali pozitivni broj, na primer $\varepsilon = 10^{-7}$) nula z_1 se tretira kao realna; u protivnom nula se uzima kao kompleksna. S obzirom da su koeficijenti polinoma (3.5.1) realni, nule se javljaju u parovima, pa je na ovaj način odredjena i nula $z_2 = \xi_1 - i\eta_1$ Ponavljajući izlozeni postupak na polinomu $P(z)/(z-\xi_1)$, odnosno $P(z)/((z-\xi_1)^2 + \eta_1^2)$ odredjuju se ostale nule polinoma P.

Na kraju napomenimo da se formule (3.5.6) vrlo često sreću u programima za numeričko rešavanje algebarskih jednačina. Na primer, program POLRT(iz 1130 Scientific Subroutine Package(1130-CM-02X) firme IBM) sačinjen je po formulama (3.5.6).

5.3.6. Bairstowljev metod

Bairstowljev metod se koristi za odredjivanje kvadratnih faktora(delilaca) polinoma, na osnovu kojih se mogu odrediti kompleksne nule polinoma.

Neka je

$$P(x) = a_o x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n \qquad (a_o \neq 0)$$

polinom sa realnim koeficijentima, S obzirom da se kompleksne nule ovog polinoma javljaju kao konjugovano kompleksni parovi, to polinom P ima realne kvadratne faktore*. U cilju odredjivanja jednog takvog faktora $\hat{m}(x) = x^2 + \hat{p}x + \hat{q}$, pretpostavimo

*) Kvadratni faktori sa realnim koeficijentima.

da nam je poznata izvesna aproksimacija $m_o(x)=x^2+p_ox+q_o$ $(p_o,q_o \in R)$.

Pre nego što damo osnovne jednakosti koje definišu Bairstowljev metod izvedimo jednu rekurentnu relaciju koja se odnosi na deljenje $P(x)$ kvadratnim faktorom $m(x)=x^2+px+q$.

Neka su polinomi Q i R u jednakosti

(3.6.1) $$P(x)=m(x)Q(x)+R(x)$$

dati sa

$$Q(x)=b_2x^{n-2}+\ldots+b_{n-1}x+b_n \quad \text{i} \quad R(x)=ux+v.$$

Uporedjivanjem koeficijenata uz odgovarajuće stepene na levoj i desnoj strani jednakosti (3.6.1) dobijamo

$$a_o=b_2,\ a_1=b_3+pb_2,$$
$$a_{i-2}=b_i+pb_{i-1}+qb_{i-2} \quad (i=4,\ldots,n),$$
$$a_{n-1}=pb_n+qb_{n-1}+u,\ a_n=qb_n+v.$$

Ako stavimo $b_o=b_1=0$, tada iz prethodnih jednakosti sleduje

(3.6.2) $$b_i=a_{i-2}-pb_{i-1}-qb_{i-2} \quad (i=2,\ldots,n)$$

i

(3.6.3) $$u=a_{n-1}-pb_n-qb_{n-1}=b_{n+1},\ v=a_n-qb_n,$$

gde smo b_{n+1} definisali prirodnim produženjem rekurentne relacije (3.6.2) za $i=n+1$.

Na osnovu (3.6.2) i (3.6.3) zaključujemo da koeficijenti polinoma Q i R na jedinstven način zavise od p i q.

Dakle, za odredjivanje kvadratnog faktora polinoma P dovoljno je naći rešenje sistema jednačina

$$u(p,q)=0,\quad v(p,q)=0.$$

Bairstowljev metod se zasniva na rešavanju ovog sistema primenom metoda Newton-Kantorovića. Naime, polazeći od (p_o,q_o) generiše se niz parova $\{(p_k,q_k)\}_{k=1,2,\ldots}$ pomoću

$$(3.6.4)\qquad \begin{bmatrix} p_{k+1} \\ q_{k+1} \end{bmatrix} = \begin{bmatrix} p_k \\ q_k \end{bmatrix} - W_k^{-1} \begin{bmatrix} u(p_k, q_k) \\ v(p_k, q_k) \end{bmatrix} \qquad (k=0,1,\ldots),$$

gde je Jacobieva matrica data sa

$$W_k = \begin{bmatrix} \frac{\partial u}{\partial p} & \frac{\partial u}{\partial q} \\ \frac{\partial v}{\partial p} & \frac{\partial v}{\partial q} \end{bmatrix}_{p=p_k,\ q=q_k} .$$

Da bismo odredili elemente ove matrice uvedimo oznake

$$r_i = \frac{\partial b_i}{\partial p}, \quad t_i = \frac{\partial b_i}{\partial q} \qquad (i=0,1,\ldots,n+1)$$

i

$$s_i = r_i - t_{i+1} \qquad (i=0,1,\ldots,n).$$

Diferenciranjem po p i q prethodno dokazanih rekurentnih relacija za koeficijente b_i, dobijamo

$$(3.6.5)\qquad r_i = -b_{i-1} - p\frac{\partial b_{i-1}}{\partial p} - q\frac{\partial b_{i-2}}{\partial p} = -b_{i-1} - pr_{i-1} - qr_{i-2}$$

i

$$(3.6.6)\qquad t_i = -p\frac{\partial b_{i-1}}{\partial q} - b_{i-2} - q\frac{\partial b_{i-2}}{\partial q} = -pt_{i-1} - b_{i-2} - qt_{i-2},$$

gde je $i=2,\ldots,n+1$. Ako od jednakosti (3.6.5) oduzmemo jednakost (3.6.6) u kojoj je indeks i zamenjen sa i+1, dobijamo

$$(3.6.7)\qquad s_i = -ps_{i-1} - qs_{i-2} \qquad (i=2,\ldots,n).$$

Kako je $b_o = b_1 = 0$ i $b_2 = a_o$, imamo $r_o = r_1 = r_2 = 0$, $t_o = t_1 = t_2 = 0$, $s_o = s_1 = 0$, što zajedno sa (3.6.7) daje

$$s_i = r_i - t_{i+1} = 0, \text{ tj. } \frac{\partial b_{i+1}}{\partial q} = \frac{\partial b_i}{\partial p} \quad (i=0,1,\ldots,n).$$

Kako je na osnovu prethodnog

$$W_k = \begin{bmatrix} r_{n+1} & r_n \\ -qr_n & -(b_n+qr_{n-1}) \end{bmatrix}_{p=p_k,\ q=q_k},$$

iterativni proces (3.6.4) postaje

$$\begin{bmatrix} p_{k+1} \\ q_{k+1} \end{bmatrix} = \begin{bmatrix} p_k \\ q_k \end{bmatrix} - \frac{1}{D} \begin{bmatrix} -(b_n+qr_{n-1})b_{n+1} - r_n(a_n-qb_n) \\ qr_nb_{n+1} + r_{n+1}(a_n - qb_n) \end{bmatrix}_{p=p_k,\, q=q_k},$$

gde je $D = qr_n^2 - (b_n + qr_{n-1})r_{n+1}$.

Naravno, iterativni proces dobijen na ovaj način egzistira ako je $D \neq 0$. Red konvergencije je dva, s obzirom da se koristi metod Newton-Kantorovića.

Jedna opštija klasa metoda Bairstowljevog tipa razmatrana je u radu [31].

Kao i Newton-Hornerov metod, tako i Bairstowljev metod se vrlo često koristi u primenama.

5.3.7. Metodi za simultano odredjivanje korena algebarskih jednačina

U novije vreme razradjeni su metodi za simultano (istovremeno) odredjivanje svih korena algebarske jednačine

$$(3.7.1) \qquad P(z) = z^n + a_1 z^{n-1} + \ldots + a_{n-1}z + a_n = 0,$$

gde su $a_i (i = 1,\ldots,n)$, u opštem slučaju, kompleksni koeficijenti.

Pretpostavimo da su koreni $z_1,\ldots,z_n$ algebarske jednačine (3.7.1) medjusobno različiti i neka su njihove približne vrednosti u k-tom iterativnom koraku $z_i(k)$ ($k=1,\ldots, n$). S obzirom da je $z_i = z_i(k) + \Delta z_i(k)$ $(k = 1,\ldots,n)$, gde su $\Delta z_i(k)$ odgovarajuće greške pojedinih korena, važi identičnost

$$\prod_{i=1}^{n} (z - (z_i(k) + \Delta z_i(k))) = P(z),$$

odakle, razvijanjem u red po stepenima od $\Delta z_i(k)$, sleduje

$$\prod_{i=1}^{n}(z-z_i(k)) - \sum_i \Delta z_i(k)\Bigg(\prod_{\substack{m=1\\ m\neq i}}^{n}(z-z_m(k))\Bigg)$$

$$+\sum_{i,j}\Delta z_i(k)\Delta z_j(k)\Bigg(\prod_{\substack{m=1\\ m\neq i,j}}^{n}(z-z_m(k))\Bigg) - \ldots = P(z) .$$

Pretpostavljajući da su greške $\Delta z_i(k)$ dovoljno male po modulu, na levoj strani u poslednjej jednakosti možemo uzeti samo prva dva člana. Ako u tako dobijenoj jednakosti stavimo redom $z = z_i(k)$ $(i = 1,\ldots,n)$ dobijamo

$$\Delta z_i(k) = -\frac{P(z_i(k))}{\prod\limits_{\substack{m=1\\ m\neq i}}^{n}(z_i(k)-z_m(k))} \qquad (i=1,\ldots,n). \tag{3.7.2}$$

Na ovaj način dolazimo do simultanog iterativnog procesa

$$z_i(k+1) = z_i(k) + \Delta z_i(k) \qquad (i=1,\ldots,n;\ k=0,1,\ldots). \tag{3.7.3}$$

Ako definišemo polinom n-tog stepena Q_k pomoću

$$Q_k(z) = \prod_{m=1}^{n}(z-z_m(k)) ,$$

tada se (3.7.2) može predstaviti u obliku

$$\Delta z_i(k) = -\frac{P(z_i(k))}{Q_k'(z_i(k))} \qquad (i=1,\ldots,n).$$

Primetimo da $Q_k(z) \to P(z)$, kada $z_i(k) \to z_i$ $(i=1,\ldots,n)$. U novo uvedenoj notaciji, iterativni proces (3.7.3) postaje

$$z_i(k+1) = z_i(k) - \frac{P(z_i(k))}{Q_k'(z_i(k))} \qquad (i=1,\ldots,n;\ k=0,1,\ldots). \tag{3.7.4}$$

Iterativni proces (3.7.4) ima kvadratnu konvergenciju. Naime, on je ekvivalentan metodu Newton-Kantorovića, primenjenog na rešavanje sistema nelinearnih jednačina, koji predstavljaju Vièteove formule (3.1.2) za polinom (3.7.1).

Ako uvedemo oznake

$$\vec{f}(\vec{z}) = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{bmatrix}, \quad \vec{a} = \begin{bmatrix} a_1 \\ -a_2 \\ \vdots \\ (-1)^{n-1}a_n \end{bmatrix}, \quad \vec{z}(k) = \begin{bmatrix} z_1(k) \\ z_2(k) \\ \vdots \\ z_n(k) \end{bmatrix},$$

tada se sistem jednačina (3.1.2) može predstaviti u obliku

$$\vec{F}(\vec{z}) = \vec{f}(\vec{z}) + \vec{a} = \vec{0} ,$$

a odgovarajući metod Newton-Kantorovića u obliku

$$(3.7.5) \qquad \vec{z}(k+1) = \vec{z}(k) - \Gamma(\vec{z}(k))(\vec{f}(\vec{z}(k)) + \vec{a}) \qquad (k = 0,1,\ldots),$$

gde je $\Gamma(\vec{z}(k))$ matrica inverzna Jacobievoj matrici $W(\vec{F}(\vec{z}(k)))$.

Nije teško pokazati da se (3.7.5) svodi na (3.7.4).

Matematičkom indukcijom se može pokazati sledeći rezultat([32]).

<u>Teorema 3.7.1</u>. Ako je

$$d = \min_{i \neq j} |z_i - z_j| \quad \text{i} \quad |z_i(0) - z_i| < \rho \quad (i = 1,\ldots,n),$$

gde je

$$\rho = d\,\frac{(1+q)^{1/(n-1)} - 1}{2(1+q)^{1/(n-1)} - 1} \qquad (0 < q < 1),$$

metod (3.7.4) konvergira, pri čemu je, za svako $k \in N$,

$$|z_i(k) - z_i| \leq \rho q^{2^k - 1} \qquad (i = 1,\ldots,n).$$

Jedna od mogućnih modifikacija metoda za simultano odredjivanje nula polinoma, koja zahteva manje memorijskog prostora kod realizacije na digitalnim računskim mašinama je varijanta poput Gauss-Seidelovog metoda

$$z_i(k+1) = z_i(k) - \frac{P(z_i(k))}{\prod\limits_{m=1}^{i-1}(z_i(k) - z_m(k+1)) \prod\limits_{m=i+1}^{n}(z_i(k) - z_m(k))} .$$

Primedba 3.7.1. U radu [33] razmatran je metod za simultano odredjivanje $r(<n)$ nula polinoma (3.7.1).

Primenom metoda za ubrzavanje konvergencije iterativnih procesa (teorema 2.4.3 iz odeljka 3.2.4) na proces (3.7.4) dobija se simultani metod trećeg reda ([34])

$$z_i(k+1) = z_i(k) - e_i\left(1 - \sum_{\substack{j=1\\ j\neq i}}^{n} \frac{e_j}{z_i(k)-z_j(k)}\right) \quad (i=1,\dots,n), \tag{3.7.6}$$

gde je $e_i = \dfrac{P(z_i(k))}{Q_k'(z_k(k))}$ $(i=1,\dots,n)$.

Primer 3.7.1. Posmatrajmo sada polinom sa kompleksnim koeficijentima

$$P(z) = z^5 - (3.2+3.9i)z^4 - (13.83-1.61i)z^3 + (9.83+29.99i)z^2$$
$$- (3.63+14.79i)z + (29.43+49.09i)$$

čije su tačne nule

$$z_1 = 1.7+1.1i,\ z_2 = 4.5+2i,\ z_3 = -3,\ z_4 = -i,\ z_5 = 1.8i.$$

Polazeći sa približnim vrednostima nula

$$z_1(o) = 1+i,\ z_2(o) = 4+2.5i,\ z_3(o) = -2+0.5i,$$
$$z_4(o) = 0.5-1.1i,\ z_5(o) = -0.2+2.2i$$

i korišćenjem formula (3.7.6) dobijamo

$$\begin{aligned}
z_1(1) &= 1.86594010 + i1.16539200,\\
z_2(1) &= 4.48809503 + i1.97590059,\\
z_3(1) &= -3.13623734 - i0.25103344,\\
z_4(1) &= -0.09598915 - i1.07210234,\\
z_5(1) &= 0.07819136 + i2.08184318,
\end{aligned}$$

$$\begin{aligned}
z_1(2) &= 1.70313403 + i1.09663271,\\
z_2(2) &= 4.50008890 + i1.99997626,\\
z_3(2) &= -2.99977371 - i0.00297130,\\
z_4(2) &= -0.00219258 - i1.00216835,\\
z_5(2) &= -0.00125665 + i1.80853067,
\end{aligned}$$

$$
\begin{aligned}
z_1(3) &= 1.69999987 + i1.09999987,\\
z_2(3) &= 4.50000000 + i2.00000000,\\
z_3(3) &= -2.99999999 - i0.00000000,\\
z_4(3) &= -0.00000003 - i1.00000000,\\
z_5(3) &= 0.00000014 + i1.80000000,
\end{aligned}
$$

$$
\begin{aligned}
z_1(4) &= 1.69999999 + i1.10000000,\\
z_2(4) &= 4.50000000 + i1.99999999,\\
z_3(4) &= -3.00000000 - i0.00000000,\\
z_4(4) &= -0.00000000 - i0.99999999,\\
z_5(4) &= -0.00000000 + i1.80000000.
\end{aligned}
$$

Na osnovu dobijenih rezultata zaključujemo da su sve nule datog polinoma odredjene sa tačnošću 10^{-8}, iz četiri iterativna koraka.

U radu [35] razmatran je iterativni proces trećeg reda

$$
z_i(k+1) = z_i(k) - \frac{P(z_i(k))}{P'(z_i(k)) - P(z_i(k)) \sum_{\substack{j=1\\ j\neq i}}^{n} (z_i(k)-z_j(k))^{-1}} \qquad (i=1,\ldots,n).
$$

O metodima za simultano nalaženje nula polinoma, pojavio se u poslednje vreme veliki broj radova. Neki od tih radova citirani su u spisku literature na kraju ove glave.

5.4. LITERATURA

1. V.L.ZAGUSKIN: Spravočnik po čislennym metodam rešenija uravnenij. Moskva, 1960.
2. D.S.MITRINOVIĆ, D.Ž.DJOKOVIĆ: Polinomi i matrice. Beograd, 1975.
3. D.M.YOUNG, R.T.GREGORY: A Survey of Numerical Mathematics (Vol.I). Addison-Wesley Publishing Company, 1972.
4. L.N.DJORDJEVIĆ: An iterative solution of algebraic equations with a parameter to accelerate convergence. Univ.Beograd. Publ.Elektrotehn. Fak.Ser.Mat.Fiz. No 412 - No 460(1973), 179-182.
5. O.N.TIHONOV: O bystrom vyčislenij najbol´ših kornej mnogočlena. Zap. Leningr.gorn.in-ta 48, 3(1968), 36-41.
6. A.OSTROWSKI: Solution of Equations and Systems of Equations. New York,1966.
7. J.H.WEGSTEIN: Accelerating convergence of iterative processes. Comm.Assoc. Comput.Mach. 1(1958), 9-13.
8. V.M.VERBUK, D.I.MIL´MAN: Metod Vegstejna kak modifikacija metoda sekuščih. Ž. vyčisl. mat.i mat.fiz. 17(1977), 507-508.
9. G.S.SALEHOV: O shodimosti metoda kasatel´nyh giperbol. DAN SSSR 82(1952), 525-528.
10. L.N.HROMOVA: Ob odnoj modifikacii metoda kasatel´nyh giperbol. Taškent. ord. Trud.Kras.Znam.Gosudar.Univ. im.V.I.Lenina. 476(1975), 35-41.
11. V.A.VARJUHIN, S.A.KASJANJUK: Ob iteracionnyh metodah utočnenija kornej uravnenij. Ž.Vyčisl.mat.i mat.fiz. 9(1969), 684-687.
12. D.K.LIKA: Ob iteracionnyh metodah vyšego norjadka. Tezisy dokl. 2-j Naučno-tehn. respubl.konfer.Moldavii. Kiinev 1965, 13-16.
13. L.N.DJORDJEVIĆ and G.V.MILOVANOVIĆ: A combined iterative formula for solving equations. informatica 78, Bled, oktober 1978, (3), 207.
14. E.HANSEN and M.PATRICK: A Family of Root Finding Methods. Numer.Math. 27(1977), 257-269.

15. G.V.MILOVANOVIĆ i L.N.DJORDJEVIĆ: On some iterative formulas of the third order. Informatica 78, Bled, oktober 1978, (3) 202.

16. J.C.DAUBISSE: Sur une méthode de résolution numérique d'équations algebriques enzparticulier dans le cas de racines multiples. Univ.Beograd. Publ.Elektrotehn.Fak.Ser.Mat.Fiz. No 498 - No 541(1975), 163-166.

17. L.V.KANTOROVIČ: Funkcional'nyj analiz i prikladnaja matematika. UMN. 3(1948), 89-185.

18. L.V.KANTOROVIČ, G.P.AKILOV: Funkcional'nyj analiz v normirovannyh prostranstvah. Moskva, 1959.

19. C.G.BROYDEN: Quasi-Newton Methods and their Application to Function Minimisation. Math.of Comput. July 1967, 368-381.

20. J.M.ORTEGA and W.C.RHEINBOLOT: Iterative solution of nonlinear equations in several variables. New York-London, 1970.

21. I.PĂVĂLOIU: Introducere in teoria aproximării Solutiiilor ecuatiilor, Cluj-Napoca, 1976.

22. B.JANKÓ: Rezolvarea ecuatiilor operationale nelineare in spatii Banach. Bucuresti, 1969.

23. L.RALL: Computational Solution of nonlinear operator equations. New York, 1969.

24. I.P.MYSOVSKIH: O shodimosti metoda L.V.Kantoroviča rešenija funkcionalńyh uravnenij i ego primenenija DAN SSSR 70(1950), 565-568.

25. M.K.TABURIN: Nelinejnye funkcional'nye uravnenija i nepreryvnye analogi iterativnyh metodov. Izvestija vuzov. Matem. No 5(1958), 18-31.

26. M.A.KRASNOSEL'SKIJ, G.M.VAJNIKKO, P.P.ZABREKO, JA.B.RUTICKIJ, V.JA.STECENKO: Približennoe rešenie operatornyh uravnenij. Moskva, 1969.

27. B.A.VERTGEJM: O Nekotoryh metodah Približennogo rešenija nelinejnyh operatornyh uravnenij v prostranstvah Banaha. UMN 12(1957), 166-169.

28. O.JU.KUP'ČICKIJ, L.I.ŠIMEPEVIČ: O nahoždenii načal'nogo približenija. Ž.Vyčisl.mat. i mat.fiz. 14(1974), 1016-1018.

29. J.H.WILKINSON: The Evaluation of Zeros of Ill-conditioned Polynomials. Numer.Math. 1(1959), 150-166.

30. M.MIGNOTTE: Note sur la méthode Bernoulli. ibid. 26(1976), 325-326.

31. K.W.BRODLIE: On Bairstow's Method for the Solution of Polynomial Equations. Math.Comp. 29(1975), 816-826.

32. K.DOČEV: Vidoizmenen metod za Njuton za ednovremenno priblizitelno presmjatane na vsički koreni na dadeno algebrično uravnenie. Fiz.-Mat. spisanie. 5(1962), 136-139.

33. M.PREŠIĆ: Un procédé itératif pour déterminer k zéros d'un polynome. C.R. Acad.Sci. Paris. 273(1971), 446-449.

34. G.V.MILOVANOVIĆ: A method to accelerate iterative processes in Banach space. Univ.Beograd. Publ.Elektrotehn.Fak.Ser.Mat. Fiz. No 461 - No 497 (1974), 67-71.

35. L.W.EHRLICH: A modified Newton method for polynomials. Comm. ACM 10(1967), 107-108.

36. O.ABERTH: Iteration Methods for Finding all zeros of a Polynomial Simultaneously. Math.Comp. 27(1973), 339-344.

37. G.ALEFELD, J.HERZBERGER: On the convergence speed of some algorithms for the simultaneous approximation of polynomial rots. SIAM J. Numer.Anal. 11(1974), 237-243.

38. G.ALEFELD, J.HERZBERGER: Uber Simultanverfahren zur Bestimmung reeller Polynomwurzoln. Z.Angew.Math.Mech. 54(1974), 413-420.

39. D.BRAESS, K.P.HADELER: Simultaneous inculision of the zeros of a polynomial. Numer.Math. 21(1973), 161-165.

40. K.DOCHEV, P.BYRNEV: Certain modifications of Newton's method for the approximate solution of algebraic equations. Ž.Vyčisl.Mat.i Mat.Fiz. 4(1964), 915-920.

41. D.K.DUNAWAY: Calculation of zeros of a real polynomial through factorization using Euclid's algorithm. SIAM J. Numer.Anal.11(1974), 1087-1104.

42. J.DVORČUK: Factorization of a polynomial into quadratic factors by Newton method. Apl.Mat. 14(1969), 54-80.

43. I.O.KERNER: Ein Gesamtschrittverfahren zur Berechnung der Nullstellen von Polynomen. Numer.Math. 8(1966), 290-294.

44. M.R.FARMER and G.LOIZOU: A class of iteration functions for improving, simultaneously, approximations to the zeros of a polynomial. BIT 15 (1975), 250-258.

45. J.PETRIĆ, M.JOVANOVIĆ and S.STAMATOVIĆ: Algorithm for simultaneous determination of all roots of algebraic polynomial equations. Mat.Vesnik 9 (1972), 325-332.

46. D.Dj.TOŠIĆ and G.V.MILOVANOVIĆ: An application of Newton's method to simultaneous determination of zeros of a polynomial. Univ.Beograd. Publ. Elektrotehn.Fak.Ser.Mat.Fiz. No 412 - No 460(1973), 175-177.

47. S.B.PREŠIĆ: Jedan iterativni postupak za faktorizaciju polinoma. Mat. Vesnik 5(20)(1968), 205-216.

48. M.R.FARMER and G.LOIZOU: An algorithm for the total, of partial, factorization of a polynomial. Math.Proc.Camb.Phil.Soc.82(1977), 427-437.

6. INTERPOLACIJA I APROKSIMACIJA

6.1. OPŠTI PROBLEM APROKSIMACIJE FUNKCIJA

6.1.1. Uvod

Problemima zamene jedne funkcije f, definisane na izvesnom skupu X, drugom funkcijom Φ bavi se posebna oblast numeričke matematike-teorija aproksimacija.

Neka funkcija Φ, koju nazivamo aproksimacionom funkcijom, zavisi od n+1 parametara $a_0,a_1,\ldots,a_n$, tj.

$$\Phi(x) = \Phi(x;a_0,a_1,\ldots,a_n).$$

Problem aproksimacije funkcije f, funkcijom Φ svodi se na odredjivanje parametara $a_i = \bar{a}_i$ po nekom kriterijumu. U zavisnosti od izabranog kriterijuma razlikujemo razne vrste aproksimacija.

U našem izlaganju uvek ćemo pretpostavljati da je $f: [a,b] \to R$.

6.1.2. Tipovi aproksimacionih funkcija

Jedan od osnovnih problema u teoriji aproksimacija predstavlja izbor opšteg oblika za funkciju Φ. U ovom odeljku navešćemo nekoliko standardnih oblika, koji se najčešće koriste u praksi.

U principu sve aproksimacione funkcije mogu se podeliti na linearne i nelinearne aproksimacione funkcije.

Opšti oblik linearne aproksimacione funkcije je

$$\Phi(x) = a_0\Phi_0(x) + a_1\Phi_1(x) + \ldots + a_n\Phi_n(x), \tag{1.2.1}$$

pri čemu sistem funkcija $\{\Phi_k\}$ ispunjava odredjene osobine. Ovde se pojam linearne funkcije odnosi na linearnost po parametrima $a_i (i = 0,1,\ldots,n)$.

U specijalnom slučaju, kada je $\Phi_k(x) = x^k (k = 0,1,\ldots,n)$, tj.

$$\Phi(x) = a_0 + a_1 x + \ldots + a_n x^n,$$

govorimo o aproksimaciji algebarskim polinomima.

U slučaju, kada je $\{\Phi_k\} = \{1, \cos x, \sin x, \cos 2x, \sin 2x, \ldots\}$ govorimo o aproksimaciji trigonometrijskim polinomima(ili o trigonometrijskoj aproksimaciji).

Ako se uzme

$$\Phi_k(x) = (x-x_k)_+^m = \begin{cases} (x-x_k)^m & (x \geq x_k), \\ 0 & (x < x_k), \end{cases}$$

gde je m fiksiran prirodan broj, imamo tzv. spline aproksimaciju.

Od nelinearnih aproksimacionih funkcija, navešćemo samo dve:

1° Eksponencijalna aproksimaciona funkcija

$$\Phi(x) = \Phi(x; c_0, b_0, \ldots, c_r, b_r) = c_0 e^{b_0 x} + \ldots + c_r e^{b_r x},$$

gde je $n = 2r$.

2° Racionalna aproksimaciona funkcija

$$\Phi(x) = \Phi(x; b_0, \ldots, b_r, c_0, \ldots, c_s) = \frac{b_0 + b_1 x + \ldots + b_r x^r}{c_0 + c_1 x + \ldots + c_s x^s},$$

gde je $n = r+s+2$.

6.1.3. Kritrijumi za aproksimaciju

U ovoj glavi razmatraćemo najvažnije vrste aproksimacija koje se sreću u primenama. Tako ćemo razmotriti

1° interpolaciju,

2° Srednje-kvadratnu aproksimaciju,

3° Mini-max aproksimaciju.

Neka je funkcija f data na segmentu $[a,b]$ skupom parova (x_k,f_k) $(k=0,1,\ldots,n)$, gde je $f_k \equiv f(x_k)$. Ako je za aproksimaciju funkcije f funkcijom Φ, kriterijum za izbor parametara $a_0,a_1,\ldots,a_n$ dat sistemom jednačina

$$(1.3.1) \qquad \Phi(x_k;a_0,a_1,\ldots,a_n)=f_k \qquad (k=0,1,\ldots,n)$$

imamo tzv. problem interpolacije funkcije. Funkciju Φ, u ovom slučaju, nazivamo interpolacionom funkcijom, a tačke $x_k(k=0, 1,\ldots,n)$ interpolacionim čvorovima.

Problem interpolacije može biti znatno složeniji od pomenutog. Jedan opštiji problem se javlja kada su pored vrednosti funkcije f u interpolacionim čvorovima poznate i vrednosti njenih izvoda.

Interpolacionim problemima biće posvećeno poglavlje 6.2.

Aproksimacije 2^o i 3^o spadaju u grupu tzv. najboljih aproksimacija. Kod ovih aproksimacija kriterijum za odredjivanje parametara aproksimacione funkcije se svodi na minimizaciju norme greške. Ovim problemima posvećeno je poglavlje 6.3.

6.2. INTERPOLACIJA FUNKCIJA

6.2.1. Čebiševljevi sistemi

Neka je funkcija f data svojim vrednostima $f_k \equiv f(x_k)$ u tačkama $x_k (x_k \in [a,b]) (k=0,1,\ldots,n)$. Posmatrajmo linearnu interpolaciju funkcije f, tj. interpolaciju funkcijom (1.2.1). U tom slučaju sistem jednačina (1.3.1) se svodi na sistem linearnih jednačina po parametrima $a_i (i=0,1,\ldots,n)$

$$a_o\Phi_o(x_k) + a_1\Phi_1(x_k) + \cdots + a_n\Phi_n(x_k) = f_k \quad (k=0,1,\ldots,n)$$

tj.

$$\text{(2.1.1)} \quad \begin{bmatrix} \Phi_o(x_o) & \Phi_1(x_o) & \cdots & \Phi_n(x_o) \\ \Phi_o(x_1) & \Phi_1(x_1) & & \Phi_n(x_1) \\ \vdots & & & \\ \Phi_o(x_n) & \Phi_1(x_n) & & \Phi_n(x_n) \end{bmatrix} \cdot \begin{bmatrix} a_o \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} f_o \\ f_1 \\ \vdots \\ f_n \end{bmatrix}.$$

Da bi navedeni interpolacioni problem imao jedinstveno rešenje potrebno je da matrica sistema (2.1.1) bude regularna.

Primer 2.1.1. Neka je $\Phi_o(x)=1$ i $\Phi_1(x)=x^2$ i $X=[-1,1]$. Odgovarajuća matrica sistema (2.1.1), u ovom slučaju, je

$$\begin{bmatrix} 1 & x_o^2 \\ 1 & x_1^2 \end{bmatrix}.$$

Interpolacioni problem

(a) ima jedinstveno rešenje ako je $|x_o| \neq |x_1|$;

(b) ima beskonačno mnogo rešenja ako je $x_0 = -x_1$ i $f(x_0) = f(x_1)$;

(c) nema rešenje ako je $x_0 = -x_1$ i $f(x_0) \neq f(x_1)$.

Neka su interpolacioni čvorovi takvi da je

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

Razmotrićemo uslove pod kojima je matrica sistema (2.1.1) regularna. Uslov da je sistem funkcija $\{\Phi_k\}$ linearno nezavisan je potreban, ali ne i dovoljan, što pokazuje sledeći primer:

<u>Primer 2.1.1.</u> Sistem funkcija $\{1, \sin x\}$ je linearno nezavisan, ali je za $x_0 + x_1 = \pi$,

$$\begin{vmatrix} 1 & \sin x_0 \\ 1 & \sin x_1 \end{vmatrix} = 0.$$

Iz osobina determinanata sleduje da sistemu funkcija $\{\Phi_k\}$ treba nametnuti takve uslove, pod kojima ne postoji linearna kombinacija

$$a_0\Phi_0(x) + a_1\Phi_1(x) + \dots + a_n\Phi_n(x)$$

koja ima n+1 različitih nula na $[a,b]$. Sistemi funkcija sa ovom osobinom nazivaju se Čebiševljevi sistemi(skraćeno T-sistemi). O T-sistemima postoji izvanredna monografija [1].

<u>Teorema 2.1.1.</u> Ako su funkcije $\Phi_k: [a,b] \to R (k = 0,1,\dots,n)$ n+1 puta diferencijabilne i ako je za svako $k = 0,1,\dots,n$ Wronsky-eva determinanta W_k različita od nule, tj.

$$W_k = \begin{vmatrix} \Phi_0(x) & \Phi_1(x) & \dots & \Phi_k(x) \\ \Phi_0'(x) & \Phi_1'(x) & & \Phi_k'(x) \\ \vdots & & & \\ \Phi_0^{(k)}(x) & \Phi_1^{(k)}(x) & & \Phi_k^{(k)}(x) \end{vmatrix} \neq 0,$$

sistem funkcija $\{\Phi_k\}$ je Čebiševljev sistem.

6.2.2. Lagrangeova interpolacija

Neka je funkcija f data svojim vrednostima $f_k \equiv f(x_k)$ u tačkama $x_k (k=0,1,\dots,n)$. Ne umanjujući opštost pretpostavimo da je

$$a \leqq x_0 < x_1 < \cdots < x_n \leqq b. \tag{2.2.1}$$

Ako tačke x_k uzmemo za interpolacione čvorove i stavimo $\phi_k(x) = x^k (k=0,1,\dots,n)$ imamo problem interpolacije funkcije f algebarskim polinomom. Označimo ovaj polinom sa P_n, tj.

$$P_n(x) = \phi(x) = a_0 + a_1 x + \cdots + a_n x^n .$$

__Teorema 2.2.1.__ Polinom P_n je jedinstven i može se predstaviti u obliku

$$P_n(x) = \sum_{k=0}^{n} f(x_k) L_k(x) , \tag{2.2.2}$$

gde je

$$L_k(x) = \frac{(x-x_0)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)} .$$

__Dokaz.__ S obzirom da je determinanta matrice sistema (2.1.1), u posmatranom slučaju, Vandermondeova, tj.

$$\begin{vmatrix} 1 & x_0 & \dots & x_0^n \\ 1 & x_1 & & x_1^n \\ \vdots & & & \\ 1 & x_n & & x_n^n \end{vmatrix} = \prod_{i>j} (x_i - x_j) ,$$

i s obzirom na pretpostavku (2.2.1), zaključujemo da je polinom P_n jedinstven.

Da bismo dokazali formulu (2.2.2) dovoljno je primetiti sledeće:

1° Svi polinomi L_k su ne višeg stepena od n, tj. $\deg(L_k) \leqq n$ za svako k;

2° $L_k(x_i) = \delta_{ik}$ (δ_{ik} - Kroneckerova delta);

3° $P_n(x_k) = f(x_k)$ za svako k;

4° $\deg(P_n) \leqq n$.

Formula (2.2.2) naziva se Lagrangeova interpolaciona formula, a polinom P_n Lagrangeov interpolacioni polinom.

Definisanjem polinoma ω pomoću

(2.2.3) $$\omega(x) = (x-x_0)(x-x_1)\cdots(x-x_n),$$

polinomi L_k se mogu predstaviti u kondenzovanijem obliku

$$L_k(x) = \frac{\omega(x)}{(x-x_k)\omega'(x_k)} \qquad (k = 0,1,\ldots,n).$$

Primer 2.2.1. Za skup podataka

x_k	-1	0	2	3
$f(x_k)$	-1	2	10	35

Lagrangeov interpolacioni polinom je

$$P_3(x) = (-1)\frac{(x-0)(x-2)(x-3)}{(-1-0)(-1-2)(-1-3)} + 2\frac{(x+1)(x-2)(x-3)}{(0+1)(0-2)(0-3)}$$

$$+10\frac{(x+1)(x-0)(x-3)}{(2+1)(2-0)(2-3)} + 35\frac{(x+1)(x-0)(x-2)}{(3+1)(3-0)(3-2)},$$

tj. $P_3(x) = \frac{5}{3}x^3 - \frac{4}{3}x^2 + 2$.

Sledeća teorema daje ocenu greške kod Lagrange ove interpolacije.

<u>Teorema 2.2.2</u>. Neka $f \in C^{n+1}[a,b]$ i $x_i \in [a,b]$ $(i=0,1,\ldots,n)$. Tada postoji $\xi \in (a,b)$ takvo da se greška Lagrange ovog interpolacionog polinoma može predstaviti u obliku

$$R_n(f;x) = f(x) - P_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\,\omega(x) . \tag{2.2.4}$$

<u>Dokaz</u>. Posmatrajmo pomoćnu funkciju F, definisanu sa

$$F(x) = f(x) - P_n(x) - K_n\omega(x), \tag{2.2.5}$$

koja se anulira u tačkama $x_0, x_1, \ldots, x_n$. Neka je $\bar{x}$ proizvoljna tačka iz $[a,b]$ i takva da je $\bar{x} \neq x_i$ $(i=0,1,\ldots,n)$. Odredićemo konstantu K_n u (2.2.5) tako da se $F(x)$ anulira i u tački $\bar{x}$. S obzirom da je $\bar{x} \neq x_i$ $(i=0,1,\ldots,n)$, ovakva vrednost za K_n postoji. Naime,

$$K_n = \frac{f(\bar{x}) - P_n(\bar{x})}{\omega(\bar{x})} .$$

Kako funkcija F na $[a,b]$ ima bar $(n+2)$ različite nule, to na osnovu Rolle ove teoreme sukcesivno zaključujemo da u $(a,b$

$F'(x)$ ima bar $(n+1)$ različitih nula,

$\vdots$

$F^{(n+1)}(x)$ ima bar jednu nulu.

Neka je $\xi \in (a,b)$ nula funkcije $F^{(n+1)}$. Tada iz (2.2.5) sleduje

$$F^{(n+1)}(\xi) = f^{(n+1)}(\xi) - 0 - K_n(n+1)! = 0,$$

odakle je $K_n = \dfrac{f^{(n+1)}(\xi)}{(n+1)!}$.

Na osnovu prethodnog zaključujemo da je

$$R_n(f;\bar{x}) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\,\omega(\bar{x}) .$$

Kako je $\bar{x}$ proizvoljna tačka iz $[a,b]$, dokaz teoreme je završen.

<u>Primer 2.2.2</u>. Neka je $x \mapsto P_5(x)$ Lagrange-ov interpolacioni polinom formiran na osnovu vrednosti funkcije $x \mapsto \sin x$ u tačkama $x_k = \frac{\pi}{10} k$ $(k=0,1,2,3,4,5)$. Tada je

$$\sin x - P_5(x) = \frac{-\sin\xi}{6!}\,x\left(x-\frac{\pi}{10}\right)\left(x-\frac{2\pi}{10}\right)\left(x-\frac{3\pi}{10}\right)\left(x-\frac{4\pi}{10}\right)\left(x-\frac{5\pi}{10}\right),$$

gde je $0 < \xi < \frac{\pi}{2}$.

Neka je $M_k = \max\limits_{x\in[a,b]} |f^{(k)}(x)|$ $(k \in N)$.

<u>Teorema 2.2.3</u>. Neka $f \in C^{n+1}[a,b]$, $x_k - x_{k-1} = h = \text{const}$ $(k=1,\ldots,n)$, $x_0 = a$, $x_n = b$. Tada je

$$(2.2.6) \qquad |R_n(f;x)| = |f(x) - P_n(x)| < \frac{h^{n+1}}{4(n+1)}\,M_{n+1} .$$

<u>Dokaz</u>. Ako stavimo $x = x_0 + ph$ $(p \in [0,n])$ imamo

$$\omega(x) = p(p-1)\cdots(p-n)h^{n+1}.$$

Kako funkcija $p \mapsto F_n(p) = |p(p-1)\cdots(p-n)|$ na segmentu $[0,n]$ dostiže maksimalnu vrednost kada $p \in (0,1)$ (ili $p \in (n-1,n)$) važi

$$\max_{p\in[0,n]} F_n(p) < \max_{p\in(0,1)} |p(p-1)| \cdot \max_{p\in(0,1)} |(p-2)\cdots(p-n)| < \frac{1}{4}\,n! ,$$

odakle neposredno sleduje tvrdjenje teoreme.

U specijalnom slučaju za $n = 2$ i $n = 3$, mogu se dobiti tačnije granice za grešku $R_n(f;x)$, nego što je (2.2.6), s obzirom da je

$$\max_{p\in[0,2]} F_2(p) = \frac{2\sqrt{3}}{9} h^3 \quad \text{i} \quad \max_{p\in[0,3]} F_3(p) = h^4 .$$

Dakle, imamo

$$|R_2(f;x)| \leqq \frac{\sqrt{3}}{27} h^3 M_3 \quad \text{i} \quad |R_3(f;x)| \leqq \frac{h^4}{24} M_4 .$$

Primer 2.2.3. Odredićemo sada sa kojom tačnošću se može izračunati vrednost $\log 11.5$ korišćenjem Lagrangeovog polinoma P_3 ako su poznate vrednosti

$$\log 10 , \log 11 , \log 12 , \log 13 .$$

S obzirom da je $M_4 = \max_{x\in[10,13]} \left|- \frac{6}{x^4}\right| = 6\cdot 10^{-4}$, imamo

$$|\log 11.5 - P_3(11.5)| \leqq \frac{M_4}{4!} |(11.5-10)(11.5-11)(11.5-12)(11.5-13)|$$
$$= 1.41\cdot 10^{-5} .$$

Inače, za proizvoljno $x \in (10,13)$ važi ocena

$$|\log x - P_3(x)| \leqq \frac{1}{24}\cdot 6\cdot 10^{-4} = 2.5\cdot 10^{-5} .$$

Razmotrimo sada problem optimalnog izbora interpolacionih čvorova kod interpolacije funkcija iz klase

$$CM = \left\{ f \mid f \in C^{n+1}[-1,1] \wedge |f^{(n+1)}(x)| \leqq M_{n+1} (\forall x \in [-1,1]) \right\} .$$

Naime, odredićemo interpolacione čvorove $x_0, x_1, \ldots, x_n$ tako da veličina

(2.2.7) $$\sup_{f\in CM} \max_{x\in[-1,1]} |R_n(f;x)|$$

ima najmanju vrednost. Tada je, na osnovu (2.2.4),

(2.2.8) $$\max_{x\in[-1,1]} |R_n(f;x)| \leqq \frac{M}{(n+1)!} \max_{x\in[-1,1]} |\omega(x)| .$$

Primetimo da u poslednjoj nejednakosti nastupa jednakost ako je f polinom

$$f(x) = \frac{M}{(n+1)!}x^{n+1} + c_0 x^n + c_1 x^{n-1} + \ldots + c_n ,$$

gde su c_i $(i = 0,1,\ldots,n)$ proizvoljni realni koeficijenti.

S obzirom da desna strana u nejednakosti (2.2.8) ne zavisi od f, to je

$$\sup_{f \in CM} \max_{x\in[-1,1]} |R_n(f;x)| \leqq \frac{M}{(n+1)!} \max_{x\in[-1,1]} |\omega(x)| .$$

Problem koji smo postavili, očigledno se svodi na odredjivanje polinoma

$$x \mapsto \omega(x) = (x-x_0)(x-x_1)\ldots(x-x_n),$$

čije nule leže u $[-1,1]$ i za koji je $\max\limits_{x\in[-1,1]} |\omega(x)|$ minimalan.

Rešenje poslednjeg problema dao je Čebišev(videti [2]) i ono je

$$\omega(x) = \frac{1}{2^n} T_{n+1}(x) = \frac{1}{2^n} \cos\left[(n+1)\arccos x\right].$$

Dakle, interpolacioni čvorovi pri kojima veličina (2.2.7) ima najmanju vrednost su nule Čebiševljevog polinoma (n+1)-og stepena T_{n+1}, tj.

(2.2.9) $$x_k = \cos\frac{2k+1}{2(n+1)}\pi \qquad (k = 0,1,\ldots,n).$$

Interpolacija kod koje su interpolacioni čvorovi uzeti kao u (2.2.9) naziva se Čebiševljevom interpolacijom.

Kada nije potreban opšti izraz za interpolacioni polinom, već samo vrednost za neko konkretno x, koristi se Aitkenova šema, koja se sastoji u sukcesivnoj primeni sledećih izraza:

$$A_k = f(x_k) \qquad (k=0,1,\ldots,n);$$

$$A_{k-1,k} = \frac{1}{x_k - x_{k-1}} \begin{vmatrix} A_{k-1} & x_{k-1} - x \\ A_k & x_k - x \end{vmatrix} \qquad (k=1,\ldots,n);$$

$$A_{o,1,\dots,n} = \frac{1}{x_n - x_o}\begin{vmatrix} A_{o,1,\dots,n-1} & x_o - x \\ A_{1,2,\dots,n} & x_n - x \end{vmatrix},$$

pri čemu je

$$P_n(x) = A_{o,1,\dots,n}\,.$$

Primer 2.2.4. Odredićemo približno f(27) na osnovu sledećih podataka

x	14	17	31	35
f(x)	68.7	64.0	44.0	39.1

primenom Aitkenove sheme.

Imamo redom

$$A_{o,1} = \frac{1}{17-14}\begin{vmatrix} 68.7 & 14-27 \\ 64.0 & 17-27 \end{vmatrix} = 48.33\,,$$

$$A_{1,2} = \frac{1}{31-17}\begin{vmatrix} 64.0 & 17-27 \\ 44.0 & 31-27 \end{vmatrix} = 49.72\,,$$

$$A_{2,3} = \frac{1}{35-31}\begin{vmatrix} 44.0 & 31-27 \\ 39.1 & 35-27 \end{vmatrix} = 48.90\,,$$

$$A_{o,1,2} = \frac{1}{31-14}\begin{vmatrix} 48.33 & 14-27 \\ 49.72 & 31-27 \end{vmatrix} = 49.39\,,$$

$$A_{1,2,3} = \frac{1}{35-17}\begin{vmatrix} 49.72 & 17-27 \\ 48.90 & 35-27 \end{vmatrix} = 49.26\,.$$

Najzad,

$$f(27) \approx P_3(27) = A_{o,1,2,3} = \frac{1}{35-14}\begin{vmatrix} 49.39 & 14-27 \\ 49.26 & 35-27 \end{vmatrix} = 49.31\,.$$

Primer 2.2.5. Na osnovu vrednosti funkcije $x \mapsto f(x) = e^x$ u tačkama $x_0 = 0.40$ i $x_1 = 0.42$, Aitkenovom šemom odredićemo vrednost ove funkcije za $x = 0.411$. Odgovarajuće vrednosti funkcije u datim tačkama uzećemo sa pet značajnih cifara, tj.

$$f_0 = 1.4918 \quad \text{i} \quad f_1 = 1.5220.$$

Imamo

$$e^{0.411} \cong A_{0,1} = \frac{1}{0.42-0.40}\begin{vmatrix} 1.4918 & -0.011 \\ 1.5220 & 0.009 \end{vmatrix} \cong 1.5084.$$

U praksi se često javlja zadatak odredjivanja vrednosti argumenta na osnovu zadate vrednosti funkcije. Ovaj zadatak se rešava metodima inverzne interpolacije.

Ako je data funkcija monotona, zadatak inverzne interpolacije najjednostavnije se rešava medjusobnom zamenom vrednosti funkcije i argumenta, a zatim konstrukcijom interpolacionog polinoma.

Primer 2.2.6. Odredimo približno nulu funkcije f, čije su vrednosti date u primeru 2.2.1. Lagrangeov interpolacioni polinom za funkciju $y \mapsto f^{-1}(y)$ je

$$P_3(y) = (-1)\cdot\frac{(y-2)(y-10)(y-35)}{(-1-2)(-1-10)(-1-35)} + 0\cdot\frac{(y+1)(y-10)(y-35)}{(2+1)(2-10)(2-35)}$$

$$+2\cdot\frac{(y+1)(y-2)(y-35)}{(10+1)(10-2)(10-35)} + 3\cdot\frac{(y+1)(y-2)(y-10)}{(35+1)(35-2)(35-10)},$$

odakle, za $y=0$, dobijamo nulu funkcije f

$$x \approx P_3(0) = -0.6508 .$$

Primedba 2.2.1. U problemima inverzne interpolacije pogodno je koristiti se Aitkenovim metodom.

6.2.3. Newtonova interpolacija sa podeljenim razlikama

Za funkciju f datu svojim vrednostima $f_k \equiv f(x_k)$ u tačkama $x_k (k = 0,1,\ldots,n)$, definisaćemo najpre podeljene razlike.

Definicija 2.3.1. Količnik

$$\frac{f(x_1) - f(x_o)}{x_1 - x_o}$$

naziva se podeljena razlika prvog reda(funkcije f u tačkama x_o i x_1) i označava sa $[x_o,x_1;f]$.

Podeljena razlika reda r definiše se rekurzivno pomoću

$$(2.3.1) \qquad [x_o,x_1,\ldots,x_r;f] = \frac{[x_1,\ldots,x_r;f] - [x_o,\ldots,x_{r-1};f]}{x_r - x_o},$$

pri čemu je $[x;f] \equiv f(x)$.

Relacija (2.3.1) omogućava konstrukciju tablice podeljenih razlika

$$\begin{array}{llllll}
x_o & f_o & & & & \\
 & & [x_o,x_1;f] & & & \\
x_1 & f_1 & & [x_o,x_1,x_2;f] & & \\
 & & [x_1,x_2;f] & & [x_o,x_1,x_2,x_3;f] & \\
x_2 & f_2 & & [x_1,x_2,x_3;f] & & \\
 & & [x_2,x_3;f] & & & \\
x_3 & f_3 & & & & \\
\vdots & \vdots & & & &
\end{array}$$

Može se pokazati da podeljena razlika reda r ima osobinu linearnosti, tj. da je

$$[x_o,x_1,\ldots,x_r;c_1f+c_2g] = c_1[x_o,\ldots,x_r;f] + c_2[x_o,\ldots,x_r;g],$$

gde su c_1 i c_2 proizvoljne konstante.

Kako je, na osnovu definicije 2.3.1,

$$[x_o,x_1;f] = \frac{f(x_o)}{x_o-x_1} + \frac{f(x_1)}{x_1-x_o},$$

indukcijom se lako pokazuje da važi formula

$$[x_0, x_1, \ldots, x_r; f] = \sum_{i=0}^{r} \frac{f(x_i)}{\omega'(x_i)} ,$$

gde je $\omega(x) = (x-x_0)(x-x_1)\ldots(x-x_r)$.

Matematičkom indukcijom se takodje može dokazati sledeći rezultat:

Teorema 2.3.1. Neka $f \in C^n[a,b]$ i neka je ispunjen uslov (2.2.1). Tada, za svako $r \leq n$, važi formula

$$[x_0, x_1, \ldots, x_r; f] = \int_0^1 \int_0^{t_1} \ldots \int_0^{t_{r-1}} f^{(r)}(x_0 + \sum_{i=1}^{r}(x_i - x_{i-1})t_i)dt_1 dt_2 \ldots dt_r .$$

Primedba 2.3.1. Primenjujući teoremu o srednjoj vrednosti integrala, iz poslednje jednakosti sleduje

$$[x_0, x_1, \ldots, x_r; f] = f^{(r)}(\xi) \int_0^1 \int_0^{t_1} \ldots \int_0^{t_{r-1}} dt_1 dt_2 \ldots dt_r$$

$$= \frac{1}{r!} f^{(r)}(\xi) \qquad (a < \xi < b).$$

Uzimajući u poslednjoj jednakosti da $x_i \to x_0 (i = 1, \ldots, r)$ zaključujemo da

$$(2.3.2) \qquad [x_0, x_1, \ldots, x_r; f] \to \frac{1}{r!} f^{(r)}(x_0).$$

Izrazićemo sada vrednost funkcije $f(x_r)$ $(r \leq n)$ pomoću podeljenih razlika $[x_0, \ldots, x_i; f]$ $(i = 0, 1, \ldots, r)$.

Za $r=1$, na osnovu definicije 2.3.1. imamo

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

Slično je, za $r=2$,

$$f(x_2) = f(x_1) + (x_2 - x_1)[x_1, x_2; f]$$

$$= (f(x_0) + (x_1 - x_0)[x_0, x_1; f]) + (x_2 - x_1)([x_0, x_1; f] + (x_2 - x_0)[x_0, x_1, x_2; f]),$$

tj.

$$f(x_2) = f(x_o) + (x_2-x_o)[x_o,x_1;f] + (x_2-x_o)(x_2-x_1)[x_o,x_1,x_2;f] .$$

U opštem slučaju važi

$$(2.3.3) \quad f(x_r) = f(x_o) + (x_r-x_o)[x_o,x_1;f] + (x_r-x_o)(x_r-x_1)[x_o,x_1,x_2;f] + \dots + (x_r-x_o)(x_r-x_1)\dots(x_r-x_{r-1})[x_o,x_1,\dots,x_r;f].$$

Korišćenjem podeljenih razlika može se konstruisati interpolacioni polinom za skup podataka $(x_k,f(x_k))$ ($k=0,\dots,n$). Naime, ovaj polinom ima oblik

$$P_n(x) = f(x_o) + (x-x_o)[x_o,x_1;f] + (x-x_o)(x-x_1)[x_o,x_1,x_2;f] + \dots + (x-x_o)(x-x_1)\dots(x-x_{n-1})[x_o,x_1,\dots,x_n;f]$$

i naziva se Newtonov interpolacioni polinom.

Da bismo dokazali poslednju formulu dovoljno je primetiti da je $\deg(P_n(x)) \leq n$ i da je $P_n(x_r) = f(x_r)$ $(r=0,1,\dots,n)$. Poslednje tvrdjenje sleduje iz (2.3.3).

S obzirom na jedinstvenost algebarskog interpolacionog polinoma zaključujemo da je Newtonov interpolacioni polinom ekvivalentan sa Lagrangeovim. Primetimo da konstrukcija Newtonovog interpolacionog polinoma zahteva prethodno formiranje tablice podeljenih razlika, što nije bio slučaj kod Lagrangeove interpolacije. S druge strane, kada hoćemo da smanjimo grešku u interpolaciji uvodjenjem novog interpolacionog čvora, Newtonov polinom je znatno pogodniji od Lagrangeovog, jer ne zahteva ponavljanje celog računskog postupka. Naime, kod Newtonove interpolacije imamo

$$P_{n+1}(x) = P_n(x) + (x-x_o)(x-x_1)\dots(x-x_n)[x_o,x_1,\dots,x_{n+1};f] .$$

<u>Teorema 2.3.2</u>. Ako funkcija f ima konačne vrednosti u tačkama $x_o,x_1,\dots,x_n,x$, važi formula

$$(2.3.4) \quad R_n(x;f) = f(x) - P_n(x) = \omega(x)[x_o,x_1,\dots,x_n,x;f] ,$$

gde je $\omega(x) = (x-x_o)(x-x_1)\dots(x-x_n)$.

Dokaz. Stavimo $r=n+1$ i $x_r=x$. Tada iz (2.3.3) sleduje

$$f(x) = P_n(x) + (x-x_0)\ldots(x-x_n)[x_0,x_1,\ldots,x_n,x;f],$$

čime je dokaz završen.

Primedba 2.3.2. Kako je na osnovu primedbe 2.3.1

$$[x_0,x_1,\ldots,x_n,x;f] = \frac{1}{(n+1)!} f^{(n+1)}(\xi) \qquad (a<\xi<b),$$

ostatak (2.3.4) se može svesti na Lagrangeov oblik.

Primedba 2.3.3. Ako u Newtonovom interpolacionom polinomu P_n uzmemo da $x_i \to x_0$ $(i=1,\ldots,n)$, na osnovu (2.3.2) zaključujemo da se on svodi na Taylorov polinom.

Primer 2.3.1. Na osnovu vrednosti funkcije $x \mapsto \operatorname{ch} x$

k	0	1	2	3
x_k	0.0	0.2	0.5	1.0
$f(x_k)$	1.0000000	1.0200668	1.1276260	1.5430806

imamo sledeću tablicu podeljenih razlika

k	$[x_k,x_{k+1};f]$	$[x_k,x_{k+1},x_{k+2};f]$	$[x_k,x_{k+1},x_{k+2},x_{k+3};f]$
0			
	0.1003338		
1		0.5163938	
	0.3585307		0.0740795
2			
	0.8309093	0.5904733	
3			

odakle je Newtonov interpolacioni polinom

$$P_3(x) = 1. + 0.1003338x + 0.5163938x(x-0.2) + 0.0740795x(x-0.2)(x-0.5).$$

Na primer, za $x=0.3$, imamo

$$\operatorname{ch} 0.3 \cong P_3(0.3) = 1.0451474.$$

Dobijeni rezultat je tačan na prve tri decimale.

6.2.4. Račun konačnih razlika

U ranijem izlaganju(videti odeljak 1.3.1) koristili smo se operatorom konačne razlike, definisanog na skupu neprekidnih funkcija pomoću

$$\Delta f(x) = f(x+h) - f(x) \qquad (h = \text{const} > 0), \tag{2.4.1}$$

ili na skupu nizova pomoću

$$\Delta y_k = y_{k+1} - y_k .$$

Ako se eksplicitno želi istaći korak h, u definiciji (2.4.1), tada se u oznaci operatora u indeksu piše h, tj. Δ_h.

U odeljku 1.3.1 takodje smo videli kako se definiše iterirani operator, tj. stepen operatora Δ. Naime,

$$\Delta^k f(x) = \Delta^{k-1} f(x+h) - \Delta^{k-1} f(x) \qquad (k \in N),$$

$$\Delta^0 f(x) = f(x).$$

Na osnovu ove definicije, matematičkom indukcijom, lako se dokazuju formule

$$\Delta^k f(x) = \sum_{i=0}^{k} (-1)^i \binom{k}{i} f(x+(k-i)h), \tag{2.4.2}$$

$$f(x+kh) = \sum_{i=0}^{k} \binom{k}{i} \Delta^i f(x). \tag{2.4.3}$$

Pored operatora Δ, u numeričkoj analizi, uvodi se još nekoliko standardnih operatora, koji se nazivaju operatori konačne razlike ili diferencni operatori. Sa ovim operatorima sprovodi se formalni račun, zasnovan na pravilima algebre i analize i kao takav se često pokazuje kao elegantno sredstvo za formiranje raznih aproksimacionih formula. Ovaj račun se naziva račun konačnih razlika ili operatorski račun. Primenom ovog računa veoma jednostavno se dobijaju formule za interpolaciju u ekvidistantnim tačkama, formule za numeričko diferenciranje, formule za numeričku integraciju, itd.

Na skupu neprekidnih funkcija definisaćemo operatore:

1° $Ef(x) = f(x+h)$ (operator pomeranja);

2° $\Delta f(x) = f(x+h) - f(x)$ (operator prednje razlike);

3° $\nabla f(x) = f(x) - f(x-h)$ (operator zadnje razlike);

4° $\delta f(x) = f(x+\frac{h}{2}) - f(x-\frac{h}{2})$ (operator centralne razlike);

5° $\mu f(x) = \frac{1}{2}\left(f(x+\frac{h}{2}) + f(x-\frac{h}{2})\right)$ (operator usrednjavanja);

6° $1f(x) = f(x)$ (identički operator);

7° $Jf(x) = \int_x^{x+h} f(t)dt$ (operator integracije).

Na skupu diferencijabilnih funkcija definišimo i operator

8° $Df(x) = f'(x)$ (operator diferenciranja).

Reći ćemo da su dva operatora A i B jednaka ako je $Af(x) = Bf(x)$ za svaku funkciju f za koju su operatori A i B definisani.

Jedan od osnovnih problema koji se nameće je nalaženje veza izmedju pojedinih operatora. Pre nego što predjemo na rešavanje ovog problema, ukazaćemo na neke opšte osobine.

Sa L označimo skup prethodno definisanih operatora, tj.

$$L = \{E, \Delta, \nabla, \delta, \mu, 1, D, J\}.$$

a) Svaki operator $A \in L$ je linearan, tj. važi

$$A(\lambda_1 f(x) + \lambda_2 g(x)) = \lambda_1 Af(x) + \lambda_2 Ag(x),$$

za svako realno λ_1 i λ_2 i svako f i g iz **domena** definisanosti operatora A.

b) Za $A, B \in L$, važi komutativnost $AB = BA$, tj.

$$A(Bf(x)) = B(Af(x)).$$

c) Ako operator $A \in L$ ima inverzni operator A^{-1}, tada je

$$AA^{-1} = A^{-1}A = 1.$$

<u>Primer 2.4.1.</u> Dokažimo osobinu b), naprimer, za operatore Δ i E.

Imamo

$$\begin{aligned}\Delta(Ef(x)) &= \Delta(f(x+h)) = f(x+2h) - f(x+h) = Ef(x+h) - Ef(x)\\ &= E(f(x+h) - f(x)) = E(\Delta f(x)).\end{aligned}$$

<u>Primer 2.4.2.</u> Neka je $g(x) = Ef(x) = f(x+h)$. Tada je $f(x) = g(x-h)$ tj. $f(x) = E^{-1}g(x)$. Dakle, operator E^{-1} je pomerajući unazad.

Ako definišemo zbir operatora $C = A + B$, pomoću

$$Cf(x) = (A+B)f(x) = Af(x) + Bf(x),$$

važe pravila

$$\begin{aligned}A+B &= B+A,\\ A(B+C) &= AB + AC,\\ (A+B)+C &= A+(B+C).\end{aligned}$$

Kako je $E^kf(x) = E(E^{k-1}f(x)) = f(x+kh)$ $(k \in N)$, tj. kako iterirani operator E^k primenjen na $f(x)$ daje $f(x+kh)$, logično se nameće definisanje stepena operatora za proizvoljni izložil sledećom jednakošću

$$E^pf(x) = f(x+ph) \qquad (p \in R).$$

Predjimo sada na utvrdjivanje formalnih veza izmedju po jedinih operatora.

1. Kako je

$$\Delta f(x) = f(x+h) - f(x) = Ef(x) - 1f(x) = (E-1)f(x),$$

imamo

$$\Delta = E-1 \quad \text{ili} \quad E = 1+\Delta.$$

Primenom binomne formule na desne strane u poslednjim jednakostima dobijamo

$$\Delta^k = (E-1)^k = \sum_{i=0}^{k}(-1)^i\binom{k}{i}E^{k-i} \quad \text{i} \quad E^k = (1+\Delta)^k = \sum_{i=0}^{k}\binom{k}{i}\Delta^k,$$

odakle formalno sleduju formule (2.4.2) i (2.4.3).

2. Kako je

$$E^{1/2}f(x) = f(x+\tfrac{h}{2}) \quad \text{i} \quad E^{-1/2}f(x) = f(x-\tfrac{h}{2})$$

važe jednakosti

$$\delta f(x) = E^{1/2}f(x) - E^{-1/2}f(x) \quad \text{i} \quad \mu f(x) = \tfrac{1}{2}\left(E^{1/2}f(x) + E^{-1/2}f(x)\right),$$

tj.

$$\delta = E^{1/2} - E^{-1/2} \quad \text{i} \quad \mu = \tfrac{1}{2}(E^{1/2} + E^{-1/2}).$$

3. Važi

$$JDf(x) = J(f'(x)) = \int_x^{x+h} f'(t)dt = f(x+h) - f(x) = \Delta f(x),$$

tj.

$$JD = \Delta.$$

Na osnovu prethodne jednakosti, zaključujemo da važi

$$\Delta D^{-1} = JDD^{-1} = J.$$

4. $\nabla = 1 - E^{-1}$.

5. $\Delta^r = E^r \nabla^r = E^{r/2} \delta^r$.

6. $E^{1/2} = \left(1 + \frac{1}{4}\delta^2\right)^{1/2} + \frac{1}{2}\delta, \quad \mu = \left(1 + \frac{1}{4}\delta^2\right)^{1/2}$,

pri čemu operator $\left(1 + \frac{1}{4}\delta^2\right)^{1/2}$ treba shvatiti kao operator čiji je iterirani operator $\left[\left(1 + \frac{1}{4}\delta^2\right)^{1/2}\right]^2$ jednak $1 + \frac{1}{4}\delta^2$.

7. Za nalaženje veze izmedju E i D pretpostavimo da je funkcija f proizvoljan broj puta diferencijabilna. Tada imamo

$$Ef(x) = f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \dots ,$$

tj.

$$\begin{aligned} Ef(x) &= 1f(x) + hDf(x) + \frac{h^2}{2!}D^2 f(x) + \dots \\ &= (1 + hD + \frac{h^2}{2!}D^2 + \dots)f(x) \\ &= e^{hD} f(x) , \end{aligned}$$

odakle zaključujemo da važi

$$E = e^{hD} \quad \text{i} \quad D = \frac{1}{h}\log(1+\Delta) = \frac{1}{h}(\Delta - \frac{\Delta^2}{2} + \frac{\Delta^3}{3} - \dots).$$

8. Kako je $\delta f(x) = E^{1/2}f(x) - E^{-1/2}f(x)$ i $E = e^{hD}$, ima

$$\delta = E^{1/2} - E^{-1/2} = e^{hD/2} - e^{-hD/2} = 2\,\mathrm{sh}\frac{hD}{2} .$$

Slično,

$$\mu = 1/2(E^{1/2} + E^{-1/2}) = \mathrm{ch}\frac{hD}{2}.$$

9. $E = (1 - \nabla)^{-1} = 1 + \nabla + \nabla^2 + \dots$

10. $\left(1 + \frac{1}{4}\delta^2\right)^{1/2} = 1 + \frac{1}{8}\delta^2 - \frac{1}{128}\delta^4 + \dots$

11. $E^{1/2} = 1 + \frac{1}{2}\delta + \frac{1}{8}\delta^2 + \dots$

6.2.5. Newtonove interpolacione formule

Primenom računa konačnih razlika mogu se izvesti različite interpolacione formule u ekvidistantnim čvorovima. U ovom odeljku daćemo samo Newtonove interpolacione formule.

Neka je funkcija f data na $[a,b]$ parovima vrednosti (x_k, f_k), gde je $f_k = f(x_k)$ i $x_k = x_o + kh$ $(k=0,1,\dots,n)$.

Za dati skup podataka može se formirati tablica konačnih razlika. Na sl.2.5.1 data je tablica formirana primenom operatora Δ.

$$
\begin{array}{llllll}
x_o & f_o & & & & \\
 & & \Delta f_o & & & \\
x_1 & f_1 & & \Delta^2 f_o & & \\
 & & \Delta f_1 & & \Delta^3 f_o & \\
x_2 & f_2 & & \Delta^2 f_1 & & \Delta^4 f_o \\
 & & \Delta f_2 & & \Delta^3 f_1 & \\
x_3 & f_3 & & \Delta^2 f_2 & & \\
 & & \Delta f_3 & & & \\
x_4 & f_4 & & & & \\
\vdots & \vdots & & & &
\end{array}
$$

Sl.2.5.1

Neka je $x = x_o + ph$ $(0 \leqq p \leqq n)$, tj. $p = \dfrac{x - x_o}{h}$. Kako je

$$E^p = (1+\Delta)^p = \sum_{k=0}^{+\infty} \binom{p}{k} \Delta^k ,$$

imamo

$$E^p f_o = \sum_{k=0}^{+\infty} \binom{p}{k} \Delta^k f_o = \sum_{k=0}^{n} \binom{p}{k} \Delta^k f_o + R_n(f;x),$$

tj.

$$(2.5.1)\qquad f(x_o+ph) = \sum_{k=o}^{n} \binom{p}{k} \Delta^k f_o + R_n(f;x),$$

gde je ostatak $R_n \equiv R_n(f;x)$, s obzirom na jedinstvenost algebarskih polinoma, isti kao kod Lagrange ove interpolacione formule, tj.

$$R_n(f;x) = \frac{h^{n+1}}{(n+1)!} p(p-1)\ldots(p-n)f^{(n+1)}(\xi),$$

gde je ξ tačka iz intervala (x_o, x_n).

Polinom

$$(2.5.2)\qquad P_n(x) = \sum_{k=o}^{n} \binom{p}{k}\Delta^k f_o \qquad (ph = x - x_o),$$

dobijen na ovaj način, naziva se prvi Newton ov interpolacioni polinom. Primetimo da se ovaj polinom može definisati i rekurzivno pomoću

$$P_k(x) = P_{k-1}(x) + \binom{p}{k}\Delta^k f_o \qquad (k=1,\ldots,n),$$

polazeći od $P_o(x) = f_o$.

Polinom (2.5.2) se može predstaviti i u razvijenom obliku

$$P_n(x) = f_o + p\Delta f_o + \frac{p(p-1)}{2!}\Delta^2 f_o + \ldots + \frac{p(p-1)\ldots(p-n+1)}{n!}\Delta^n.$$

tj.

$$P_n(x) = f_o + \frac{\Delta f_o}{h}(x-x_o) + \frac{\Delta^2 f_o}{2!h^2}(x-x_o)(x-x_1) + \ldots$$

$$+ \frac{\Delta^n f_o}{n!h^n}(x-x_o)(x-x_1)\ldots(x-x_{n-1}).$$

Prvi Newton ov interpolacioni polinom se koristi u slučajevima kada se interpolacija izvodi na početku intervala, tj.

u okolini tačke x_o. Ako se koristi za približno izračunavanje vrednosti funkcije f za $x < x_o$, kažemo da se radi o ekstrapolaciji funkcije.

Formirajmo sada tablicu konačnih razlika, koristeći se operatorom ∇ (sl. 2.5.2).

$$
\begin{array}{cccccc}
\vdots & \vdots & & & & \\
x_{n-4} & f_{n-4} & & & & \\
 & & \nabla f_{n-3} & & & \\
x_{n-3} & f_{n-3} & & \nabla^2 f_{n-2} & & \\
 & & \nabla f_{n-2} & & \nabla^3 f_{n-1} & \\
x_{n-2} & f_{n-2} & & \nabla^2 f_{n-1} & & \nabla^4 f_n \\
 & & \nabla f_{n-1} & & \nabla^3 f_n & \\
x_{n-1} & f_{n-1} & & \nabla^2 f_n & & \\
 & & \nabla f_n & & & \\
x_n & f_n & & & &
\end{array}
$$

Sl. 2.5.2

Neka je $x = x_n + ph\ (0 \leqq -p \leqq n)$, tj. $p = \dfrac{x - x_n}{h}$. Kako je

$$E^p = (1-\nabla)^{-p} = \sum_{k=0}^{+\infty} (-1)^k \binom{-p}{k} \nabla^k$$

dobijamo

$$f(x_n + ph) = \sum_{k=0}^{+\infty} (-1)^k \binom{-p}{k} \nabla^k f_n$$

$$= \sum_{k=0}^{+\infty} \frac{p(p+1)\ldots(p+k-1)}{k!} \nabla^k f_n \ .$$

Za $f(x_n+ph)$ često se koristi oznaka f_{n+p}.

Koristeći se razlikama zaključno sa redom n, na osnovu poslednje jednakosti dobijamo drugu Newtonovu interpolacionu formulu

$$P_n(x) = f_n + p\nabla f_n + \frac{p(p+1)}{2!} \nabla^2 f_n + \ldots + \frac{p(p+1)\ldots(p+n-1)}{n!} \nabla^n f_n ,$$

tj.

$$P_n(x)=f_n+\frac{\nabla f_n}{h}(x-x_n)+\frac{\nabla^2 f_n}{2!h^2}(x-x_n)(x-x_{n-1})+\dots$$
$$+\frac{\nabla^n f_n}{n!h^n}(x-x_n)(x-x_{n-1})\dots(x-x_1),$$

pri čemu se ostatak može izraziti u obliku

$$R_n(f;x)=\frac{h^{n+1}}{(n+1)!}\,p(p+1)\dots(p+n)f^{(n+1)}(\xi)\qquad(\xi\in[x_0,x_n]).$$

Primer 2.5.1. Nadjimo prvi i drugi Newtonov interpolacioni polinom za skup podataka $\{(x_k,f_k)\}=\{(-1,-3),(0,-5),(1,1),(2,21)\}$.

Na osnovu tablice, formirane pomoću operatora Δ,

x_k	f_k	Δf_k	$\Delta^2 f_k$	$\Delta^3 f_k$
-1	-3			
		-2		
0	-5		8	
		6		6
1	1		14	
		20		
2	21			

nalazimo prvi Newtonov interpolacioni polinom

$$P_3(x)=-3+(-2)(x+1)+\frac{8}{2}(x+1)(x-0)+\frac{6}{3!}(x+1)(x-0)(x-1)$$
$$=x^3+4x^2+x-5.$$

Primetimo da su zaokruženi elementi u tablici korišćeni za konstrukciju ovog polinoma.

S obzirom da je

$$\nabla f_k=f_k-f_{k-1}=\Delta f_{k-1},\quad \nabla^2 f_k=\Delta^2E^{-2}f_k=\Delta^2 f_{k-2},\quad \nabla^3 f_k=\Delta^3 f_{k-3},$$

zaključujemo da nije potrebno formirati tablicu pomoću operatora ∇, jer je identična sa prethodno dobijenom tablicom, u kojoj treba samo izvršiti pomeranje indeksa. Drugi Newtonov interpolacioni polinom ima oblik

$$P_3(x)=21+20(x-2)+\frac{14}{2}(x-2)(x-1)+\frac{6}{3!}(x-2)(x-1)(x-0).$$

Na kraju proučimo kako se slučajna greška u vrednosti funkcije u nekom od interpolacionih čvorova manifestuje u tablici konačnih razlika. Neka je, na primer, umesto vrednosti f_k uzeta vrednost $f_k+\varepsilon$. Tablica dobijena primenom operatora data je na sl.2.5.3.

x_{k-4}	f_{k-4}				
		Δf_{k-4}			
x_{k-3}	f_{k-3}		$\Delta^2 f_{k-4}$		
		Δf_{k-3}		$\Delta^3 f_{k-4}$	
x_{k-2}	f_{k-2}		$\Delta^2 f_{k-3}$		$\Delta^4 f_{k-4}+\varepsilon$
		Δf_{k-2}		$\Delta^3 f_{k-3}+\varepsilon$	
x_{k-1}	f_{k-1}		$\Delta^2 f_{k-2}+\varepsilon$		$\Delta^4 f_{k-3}-4\varepsilon$
		$\Delta f_{k-1}+\varepsilon$		$\Delta^3 f_{k-2}-3\varepsilon$	
x_k	$f_k+\varepsilon$		$\Delta^2 f_{k-1}-2\varepsilon$		$\Delta^4 f_{k-2}+6\varepsilon$
		$\Delta f_k-\varepsilon$		$\Delta^3 f_{k-1}+3\varepsilon$	
x_{k+1}	f_{k+1}		$\Delta^2 f_k+\varepsilon$		$\Delta^4 f_{k-1}-4\varepsilon$
		Δf_{k+1}		$\Delta^3 f_k-\varepsilon$	
x_{k+2}	f_{k+2}		$\Delta^2 f_{k+1}$		$\Delta^4 f_k+\varepsilon$
		Δf_{k+2}		$\Delta^3 f_{k+1}$	
x_{k+3}	f_{k+3}		$\Delta^2 f_{k+2}$		
		Δf_{k+3}			
x_{k+4}	f_{k+4}				

Sl. 2.5.3

Na osnovu ove tablice, možemo zaključiti da greška koja se pojavila u vrednosti f_k čini pogrešnim razlike

$$\Delta f_{k-1},\ \Delta f_k,$$

$$\Delta^2 f_{k-2},\ \Delta^2 f_{k-1},\ \Delta^2 f_k,$$

$$\Delta^3 f_{k-3},\ \Delta^3 f_{k-2},\ \Delta^3 f_{k-1},\ \Delta^3 f_k,\ \text{itd.}$$

Štaviše, greška u razlici $\Delta^m f_{k-m+i}$ $(i=0,1,\ldots,m)$ je $\binom{m}{i}(-1)^i\varepsilon$.

Znajući ovaj zakon rasprostiranja greške u tablici konačnih razlika mogućno je, u nekim slučajevima, naći izvor greške, a samim tim i otkloniti ovu grešku(videti [3]).

Sa numeričkog stanovišta Newtonovi interpolacioni polinomi nisu naročito pogodni, pa se u praksi koriste uglavnom interpolacioni polinomi sa centralnim razlikama.

6.2.6. Interpolacione formule sa centralnim razlikama

U ovom odeljku razmotrićemo jednu opštu klasu interpolacionih formula, koja kao partikularne slučajeve sadrži Newtonove interpolacione formule.

Neka je funkcija f tabelirana na skupu tačaka

$$G_m = \{x_0-mh,\ldots,x_0-2h,x_0-h,x_0,x_0+h,x_0+2h,\ldots,x_0+mh\},$$

gde je m fiksiran prirodan broj i $h=\text{const}>0$. Sa f_k označimo vrednost funkcije u tački $x_k = x_0+kh(k=0,\pm1,\ldots,\pm m)$.

Tablicu koja sadrži sve moguće razlike(diference) funkcije f na skupu G_m nazivaćemo centralnom tablicom razlika(diferenci). Tako na skupu G_3 imamo centralnu tablicu prednjih razlika

$$\begin{array}{cccccccc}
x_{-3} & f_{-3} & & & & & & \\
 & & \Delta f_{-3} & & & & & \\
x_{-2} & f_{-2} & & \Delta^2 f_{-3} & & & & \\
 & & \Delta f_{-2} & & \Delta^3 f_{-3} & & & \\
x_{-1} & f_{-1} & & \Delta^2 f_{-2} & & \Delta^4 f_{-3} & & \\
 & & \Delta f_{-1} & & \Delta^3 f_{-2} & & \Delta^5 f_{-3} & \\
x_0 & f_0 & & \Delta^2 f_{-1} & & \Delta^4 f_{-2} & & \Delta^6 f_{-3} \\
 & & \Delta f_0 & & \Delta^3 f_{-1} & & \Delta^5 f_{-2} & \\
x_1 & f_1 & & \Delta^2 f_0 & & \Delta^4 f_{-1} & & \\
 & & \Delta f_1 & & \Delta^3 f_0 & & & \\
x_2 & f_2 & & \Delta^2 f_1 & & & & \\
 & & \Delta f_2 & & & & & \\
x_3 & f_3 & & & & & &
\end{array}$$

Sl.2.6.1

ili centralnu tablicu centralnih razlika

$$\begin{array}{cccccccc}
x_{-3} & f_{-3} & & & & & & \\
 & & \delta f_{-5/2} & & & & & \\
x_{-2} & f_{-2} & & \delta^2 f_{-2} & & & & \\
 & & \delta f_{-3/2} & & \delta^3 f_{-3/2} & & & \\
x_{-1} & f_{-1} & & \delta^2 f_{-1} & & \delta^4 f_{-1} & & \\
 & & \delta f_{-1/2} & & \delta^3 f_{-1/2} & & \delta^5 f_{-1/2} & \\
x_0 & f_0 & & \delta^2 f_0 & & \delta^4 f_0 & & \delta^6 f_0 \\
 & & \delta f_{1/2} & & \delta^3 f_{1/2} & & \delta^5 f_{1/2} & \\
x_1 & f_1 & & \delta^2 f_1 & & \delta^4 f_1 & & \\
 & & \delta f_{3/2} & & \delta^3 f_{3/2} & & & \\
x_2 & f_2 & & \delta^2 f_2 & & & & \\
 & & \delta f_{5/2} & & & & & \\
x_3 & f_3 & & & & & &
\end{array}$$

Sl.2.6.2

Imajući u vidu jednakosti

$$\Delta^r f_k = \nabla^r E^r f_k = \nabla^r f_{k+r}\,, \quad \Delta^r f_k = \delta^r E^{r/2} f_k = \delta^r f_{k+r/2} \quad (r=1,2,\ldots),$$

vidimo da se iz tablice prednjih razlika lako dolazi do tablice zadnjih razlika, ili tablice centralnih razlika, prostim pomeranjem indeksa.

Ako uvedemo smenu $x=x_o+ph$ i stavimo $f_p=f(x)$, prva Newtonova interpolaciona formula (2.5.1) se može predstaviti u obliku

$$f_p = f_o + \binom{p}{1}\Delta f_o + \binom{p}{2}\Delta^2 f_o + \binom{p}{3}\Delta^3 f_o + \ldots\,, \tag{2.6.1}$$

a pomoću operatora centralne razlike u obliku

$$f_p = f_o + \binom{p}{1}\delta f_{1/2} + \binom{p}{2}\delta^2 f_1 + \binom{p}{3}\delta^3 f_{3/2} + \ldots\,.$$

Posmatrajmo opšti član $\binom{p}{q}\Delta^q f_o$ u formuli (2.6.1). Medjusobnim množenjem jednakosti

$$\binom{p}{q} = \binom{p+1}{q+1} - \binom{p}{q+1} \quad \text{i} \quad \Delta^q f_{k+1} - \Delta^q f_k = \Delta^{q+1} f_k\,,$$

dobijamo

$$\binom{p}{q}\Delta^q f_{k+1} - \binom{p}{q}\Delta^q f_k = \binom{p+1}{q+1}\Delta^{q+1} f_k - \binom{p}{q+1}\Delta^{q+1} f_k\,,$$

tj.

$$\binom{p}{q}\Delta^q f_{k+1} + \binom{p}{q+1}\Delta^{q+1} f_k = \binom{p}{q}\Delta^q f_k + \binom{p+1}{q+1}\Delta^{q+1} f_k\,. \tag{2.6.2}$$

Poslednjoj jednakosti može se pridružiti dijagram,

$$\begin{tikzcd}
 & \Delta^q f_k \arrow[dr, dash, "\binom{p+1}{q+1}"] & \\
{} \arrow[ur, dash] \arrow[dr, dash] \arrow[rr, phantom, "\binom{p}{q}"] & & \Delta^{q+1} f_k \\
 & \Delta^q f_{k+1} \arrow[ur, dash, "\binom{p}{q+1}"'] &
\end{tikzcd}$$

koji se naziva romboidni dijagram. Smisao ovog pridruživanja je da se kretanjem od krajnjeg levog do krajnjeg desnog temena romba po bilo kojoj grani(gornjoj ili donjoj) uz usputno "težinsko" sabiranje veličina na koje nailazimo dobija isti rezultat. Član koji uzimamo u zbir formiramo na sledeći način:

Ako krećući se po grani romboidnog dijagrama naidjemo na diferencu dolazeći po uzlaznoj(silaznoj) putanji, tada tu diferencu množimo binomnim koeficijentom koji se nalazi ispod (iznad) te putanje. Tako putanji

$$\begin{array}{cccc} & \Delta^q f_k & & \binom{p+1}{q+1} \\ \diagup & & \diagdown & \\ & \binom{p}{q} & & \Delta^{q+1} f_k \end{array}$$

odgovara zbir $\binom{p}{q}\Delta^q f_k + \binom{p+1}{q+1}\Delta^{q+1} f_k$.

Od opisanih rombova mogućno je sastaviti romboidnu mrežu(sl.2.6.3).

Osim uzlaznih i silaznih putanja mogu se formirati i horizontalne putanje, na primer,

$$1\ ,\quad \Delta f_0\ ,\quad \binom{p}{2}\ ,\quad \Delta^3 f_{-1}\ ,\ \ldots\ .$$

Ako se krećemo po takvoj putanji član koji se uzima u zbir jednak je proizvodu odgovarajućeg člana na koji se naidje na putanji i aritmetičke sredine članova iznad i ispod njega. Tako za prethodno navedenu putanju imamo zbir

$$\frac{1}{2}(f_0+f_1)+\frac{1}{2}\left[\binom{p}{1}+\binom{p-1}{1}\right]\Delta f_0+\frac{1}{2}\binom{p}{2}(\Delta^2 f_{-1}+\Delta^2 f_0)+\ldots\ .$$

Svakoj putanji u romboidnoj mreži odgovara jedna interpolaciona formula. Prvoj Newtonovoj formuli odgovara putanja

$$f_0\ ,\ \Delta f_0\ ,\ \Delta^2 f_0\ ,\ \Delta^3 f_0\ ,\ \ldots\ . \tag{2.6.3}$$

Sa numeričkog stanovišta, za interpolaciju funkcije u tački x, pogodne su interpolacione formule koje koriste informaciju o vrednostima funkcije u interpolacionim čvorovima koji se nalaze, kako ispred, tako i iza vrednosti x, što nije bio slučaj sa Newtonovim formulama.

U daljem tekstu dajemo pregled važnijih interpolacionih formula:

1° Prva Newtonovova interpolaciona formula odgovara putanji (2.6.3) i data je pomoću (2.6.1).

f_{-3}	$\binom{p+3}{1}$	$\Delta^2 f_{-4}$	$\binom{p+4}{3}$	$\Delta^4 f_{-5}$	$\binom{p+5}{5}$
1	Δf_{-3}	$\binom{p+3}{2}$	$\Delta^3 f_{-4}$	$\binom{p+4}{4}$	$\Delta^5 f_{-5}$
f_{-2}	$\binom{p+2}{1}$	$\Delta^2 f_{-3}$	$\binom{p+3}{3}$	$\Delta^4 f_{-4}$	$\binom{p+4}{5}$
1	Δf_{-2}	$\binom{p+2}{2}$	$\Delta^3 f_{-3}$	$\binom{p+3}{4}$	$\Delta^5 f_{-4}$
f_{-1}	$\binom{p+1}{1}$	$\Delta^2 f_{-2}$	$\binom{p+2}{3}$	$\Delta^4 f_{-3}$	$\binom{p+3}{5}$
1	Δf_{-1}	$\binom{p+1}{2}$	$\Delta^3 f_{-2}$	$\binom{p+2}{4}$	$\Delta^5 f_{-3}$
f_0	$\binom{p}{1}$	$\Delta^2 f_{-1}$	$\binom{p+1}{3}$	$\Delta^4 f_{-2}$	$\binom{p+2}{5}$
1	Δf_0	$\binom{p}{2}$	$\Delta^3 f_{-1}$	$\binom{p+1}{4}$	$\Delta^5 f_{-2}$
f_1	$\binom{p-1}{1}$	$\Delta^2 f_0$	$\binom{p}{3}$	$\Delta^4 f_{-1}$	$\binom{p+1}{5}$
1	Δf_1	$\binom{p-1}{2}$	$\Delta^3 f_0$	$\binom{p}{4}$	$\Delta^5 f_{-1}$
f_2	$\binom{p-2}{1}$	$\Delta^2 f_1$	$\binom{p-1}{3}$	$\Delta^4 f_0$	$\binom{p}{5}$
1	Δf_2	$\binom{p-2}{2}$	$\Delta^3 f_1$	$\binom{p-1}{4}$	$\Delta^5 f_0$
f_3	$\binom{p-3}{1}$	$\Delta^2 f_2$	$\binom{p-2}{3}$	$\Delta^4 f_1$	$\binom{p-1}{5}$

Sl. 2.6.3

2^o Druga Newtonova interpolaciona formula odgovara putanji

$$f_o\,,\ \Delta f_{-1},\ \Delta^2 f_{-2},\ \Delta^3 f_{-3},\ \ldots$$

i data je pomoću

$$f_p = f_o + \binom{p}{1}\Delta f_{-1} + \binom{p+1}{2}\Delta^2 f_{-2} + \binom{p+2}{3}\Delta^3 f_{-3} + \ldots\ .$$

Primetimo da je ovde drugi Newtonov polinom konstruisan u tački x_o, a ne u x_n kao u odeljku 6.2.5.

3^o Prva Gaussova interpolaciona formula odgovara putanji

$$f_o,\ \Delta f_o,\ \Delta^2 f_{-1},\ \Delta^3 f_{-1},\ \Delta^4 f_{-2},\ \ldots$$

i data je pomoću

$$f_p = f_o + \binom{p}{1}\Delta f_o + \binom{p}{2}\Delta^2 f_{-1} + \binom{p+1}{3}\Delta^3 f_{-1} + \binom{p+1}{4}\Delta^4 f_{-2} + \ldots\ .$$

4^o Druga Gaussova interpolaciona formula odgovara putanji

$$f_o,\ \Delta f_{-1}\,,\ \Delta^2 f_{-1},\ \Delta^3 f_{-2},\ \Delta^4 f_{-2},\ \ldots$$

i može se predstaviti u obliku

$$f_p = f_o + \binom{p}{1}\Delta f_{-1} + \binom{p+1}{2}\Delta^2 f_{-1} + \binom{p+1}{3}\Delta^3 f_{-2} + \binom{p+2}{4}\Delta^4 f_{-2} + \ldots\ .$$

5^o Stirlingova interpolaciona formula odgovara putanji

$$f_o,\ \binom{p}{1},\ \Delta^2 f_{-1},\ \binom{p+1}{3},\ \Delta^4 f_{-2},\ \ldots$$

i shodno ranije datim pravilima može se predstaviti u obliku

$$f_p = f_o + \frac{1}{2}(\Delta f_{-1} + \Delta f_o)\binom{p}{1} + \frac{1}{2}\left(\binom{p+1}{2} + \binom{p}{2}\right)\Delta^2 f_{-1}$$

$$+ \frac{1}{2}(\Delta^3 f_{-2} + \Delta^3 f_{-1})\binom{p+1}{3} + \frac{1}{2}\left(\binom{p+2}{4} + \binom{p+1}{4}\right)\Delta^4 f_{-2} + \ldots\ .$$

Primetimo da se Stirlingova formula dobija kao aritmetička sredina prve i druge Gaussove interpolacione formule.

6^o Besselova interpolaciona formula odgovara horizontalnoj putanji

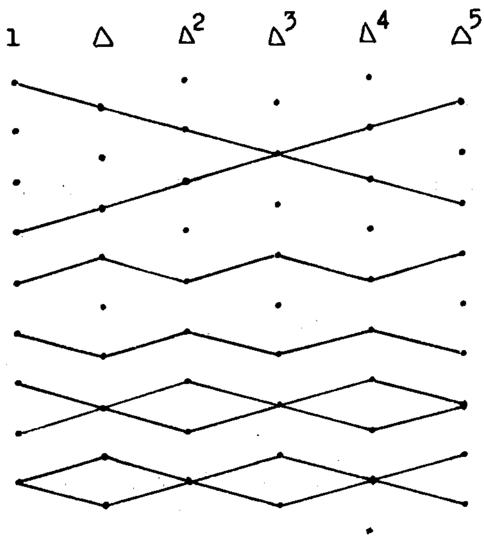
$$1,\ \Delta f_o,\ \binom{p}{2}\ ,\ \Delta^3 f_{-1},\ \binom{p+1}{4},\ \ldots$$

i ima oblik

$$f_p = \frac{1}{2}(f_o + f_1) + \frac{1}{2}\left(\binom{p}{1} + \binom{p-1}{1}\right)\Delta f_o + \frac{1}{2}(\Delta^2 f_{-1} + \Delta^2 f_o)\binom{p}{2}$$

$$+ \frac{1}{2}\left(\binom{p+1}{3} + \binom{p}{3}\right)\Delta^3 f_{-1} + \frac{1}{2}(\Delta^4 f_{-2} + \Delta^4 f_{-1})\binom{p+1}{4} + \dots .$$

Na sl. 2.6.4 navedene su putanje ovih interpolacionih formila.



Sl. 2.6.4

Za interpolaciju funkcija najčešće se koriste Stirlingova i Besselova formula. Stirlingova formula se koristi kada je $|p| \leqq 0.25$, a Besselova kada je $0.25 \leqq |p| \leqq 0.75$.

6.2.7. Hermiteova interpolacija

U ovom odeljku obradićemo jedan opštiji interpolacioni problem. Naime, neka su za funkciju $f: [a,b] \to R$ u interpolacionim čvorovima $x_i (i=0,1,\dots,m)$ poznate sledeće vrednosti:

$$f(x_i),\ f'(x_i),\ \dots,\ f^{(k_i-1)}(x_i) \qquad (i = 0,1,\dots,m).$$

Neka je broj podataka o funkciji f jednak $n+1$, tj. $n+1=k_0+k_1+\cdots+k_m$. Broj k_i se naziva višestrukost interpolacionog čvora x_i.

Posmatraćemo sada samo interpolaciju algebarskim polinomima, koja je poznata kao Hermite ova interpolacija. S obzirom na broj datih podataka o funkciji f, Hermite ov interpolacioni polinom, u opštem slučaju, biće

$$H_n(x) = a_0 + a_1x + \cdots + a_nx^n.$$

Koeficijente $a_i (i=0,1,\ldots,n)$ odredjujemo rešavanjem sistema linearnih jednačina

$$H_n^{(j)}(x_i) = f^{(j)}(x_i) \quad (i=0,1,\ldots,m;\ j=0,1,\ldots,k_i-1). \tag{2.7.1}$$

Teorema 2.7.1. Sistem jednačina (2.7.1) ima jedinstveno rešenje.

Dokaz. Za dokaz ovog tvrdjenja dovoljno je pokazati da homogeni sistem jednačina

$$H_n^{(j)}(x_i) = 0 \quad (i=0,1,\ldots,m;\ j=0,1,\ldots,k_i-1) \tag{2.7.2}$$

ima samo trivijalno rešenje $a_i=0$ $(i=0,1,\ldots,n)$, tj. $H_n(x)\equiv 0$.

U posmatranom slučaju, iz (2.7.2) sleduje da je x_i nula polinoma H_n višestrukosti najmanje k_i, što znači da zbir višestrukosti svih nula polinoma H_n nije manji od $n+1(=k_0+k_1+\cdots+k_m)$. Kako je, medjutim, H_n polinom stepena n, zaključujemo da on mora biti identički jednak nuli, čime je dokaz završen.

S obzirom da rešavanje sistema linearnih jednačina (2.7.1) može biti vrlo komplikovano, to se za konstrukciju Hermite ovog interpolacionog polinoma najčešće koristi jedan metod zasnovan na primeni Lagrange ove interpolacije, koji uključuje uslove (2.7.1). Na osnovu prethodne teoreme polinom koji se na ovaj način dobija predstavlja Hermite ov interpolacioni polinom.

Ilustrovaćemo ovaj metod na jednom primeru.

Primer 2.7.1. Nadjimo Hermite ov interpolacioni polinom za funkciju f čije su vrednosti date u tabeli

x	0	1	2
f(x)	-1	1	9
f'(x)	-1	3	
f"(x)	10		

Kako je dato šest podataka, interpolacioni polinom, u opštem slučaju, biće petog stepena. Potražimo ga u obliku

(2.7.3) $$H_5(x) = P_2(x) + x(x-1)(x-2)H_2(x),$$

gde je P_2 Lagrange ov interpolacioni polinom formiran na osnovu vrednosti funkcije f u tačkama x=0, x=1, x=2, tj.

$$P_2(x) = (-1)\cdot\frac{(x-1)(x-2)}{(0-1)(0-2)} + 1\cdot\frac{(x-0)(x-2)}{(1-0)(1-2)} + 3\cdot\frac{(x-o)(x-1)}{(2-0)(2-1)}$$

$$= 3x^2 - x - 1 ,$$

a H_2 za sada nepoznat polinom ne višeg stepena od dva. Diferenciranjem (2.7.3) dobijamo

$$H_5'(x) = 6x - 1 + (3x^2 - 6x + 2)\,H_2(x) + x(x-1)(x-2)H_2'(x),$$

odakle, s obzirom na $H_5'(0)=f'(0)=-1$ i $H_5'(1)=f'(1)=3$, sleduje $H_2(0)=0$ i $H_2(1)=2$.

Kako je, dalje,

$$H_5''(x)=6+6(x-1)H_2(x)+(6x^2-12x+4)H_2'(x)+x(x-1)(x-2)H_2''(x)$$

i $H_5''(0)=f''(0)=10$ dobijamo $H_2'(0)=1$.

Primenimo sada isti postupak na odredjivanje polinoma H_2, na osnovu podataka

$$H_2(0)=0\,,\ H_2(1)=2\ ;\ H_2'(0)=1.$$

Dakle, imamo

$$H_2(x)=P_1(x)+x(x-1)a \qquad (a=H_0(x))$$

gde je

$$P_1(x)=0\cdot\frac{x-1}{0-1}+2\cdot\frac{x-0}{1-0}=2x\,.$$

Ako stavimo $x=0$ u jednakost $H_1'(x)=2+(2x-1)a$ nalazimo $a=1$. Tada je $H_2(x)=x^2+x$ i

$$H_5(x)=x^5-2x^4-x^3+5x^2-x-1\,.$$

Sledeća teorema daje ocenu greške kod Hermite ove interpolacije.

<u>Teorema 2.7.2</u>. Neka $f\in C^{n+1}[a,b]$ i $x_i\in[a,b]$ $(i=0,1,\ldots,m)$. Tada postoji $\xi\in(a,b)$ takvo da je

$$R_n(f;x)=f(x)-H_n(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}\,\Omega_n(x), \tag{2.7.4}$$

gde je

$$\Omega_n(x)=(x-x_0)^{k_0}(x-x_1)^{k_1}\cdots(x-x_m)^{k_m}\,.$$

Dokaz. Slično kao u dokazu teoreme 2.2.2, uvedimo pomoćnu funkciju F sa

$$F(x) = f(x) - H_n(x) - K_n \Omega_n(x).$$

Neka je $\bar{x}$ proizvoljna tačka iz $[a,b]$ i takva da se ne poklapa ni sa jednim od interpolacionih čvorova $x_i (i=0,1,\ldots,m)$.

Pomoćna funkcija F ima nule u tačkama $x_0,x_1,\ldots,x_m$, čiji je red višestrukosti najmanje $k_0,k_1,\ldots,k_m$ respektivno. Konstantu K_n odredićemo iz uslova da funkcija F ima nulu i u tački $\bar{x}$. Ovakva konstanta egzistira, s obzirom da je $\Omega_n(\bar{x}) \neq 0$, i iznosi

$$K_n = \frac{f(\bar{x}) - H_n(\bar{x})}{\Omega_n(\bar{x})} . \tag{2.7.5}$$

Na osnovu Rolleove teoreme, izvod F' ima bar m nula koje se nalaze izmedju nula funkcije F. S druge strane, funkcija F' ima nule u tačkama $x_0,x_1,\ldots,x_m$, čiji je red višestrukosti najmanje $k_0-1,k_1-1,\ldots,k_m-1$ respektivno. Dakle, broj nula funkcije F' na $[a,b]$ je najmanje $n+1$ ($=(k_0-1)+(k_1-1)+\cdots+(k_m-1)+m$), pri čemu se višestruka nula uzima onoliko puta koliki je njen red višestrukosti.

Kako je $H_n^{(n+1)}(x)=0$ i $\Omega_n^{(n+1)}(x)=(n+1)!$, iz $F^{(n+1)}(\xi)=0$ sleduje

$$K_n = \frac{f^{(n+1)}(\xi)}{(n+1)!},$$

što zajedno sa (2.7.5) daje

$$f(\bar{x}) = H_n(\bar{x}) + \frac{f^{(n+1)}(\xi)}{(n+1)!}\Omega_n(\bar{x}).$$

S obzirom da je $\bar{x}$ proizvoljna tačka iz $[a,b]$, dokaz teoreme je završen.

Za $k_o = k_1 = \cdots = k_m = 2$ $(n+1 = 2m+2)$, (2.7.4) se svodi na

$$R_n(f;x) = \frac{f^{(2m+2)}(\xi)}{(2m+2)!}\,\omega(x)^2 ,$$

gde je ω definisano pomoću (2.2.3). U ovom slučaju Hermite ov interpolacioni polinom može se izraziti eksplicitno pomoću

$$H_n(x) = \sum_{i=0}^{m} U_i(x)f(x_i) + \sum_{i=0}^{m} V_i(x)f'(x_i) ,$$

gde su

$$U_i(x) = (1 - 2L_i'(x_i)(x - x_i))\,L_i(x)^2 ,$$

$$V_i(x) = (x - x_i)\,L_i(x)^2 ,$$

$$L_i(x) = \frac{\omega(x)}{(x - x_i)\,\omega'(x_i)} .$$

6.2.8. Pronyeva interpolacija

Neka je funkcija f data na skupu ekvidistantnih tačaka parovima vrednosti $\{(x_k,f_k)\}_{k=0,1,\ldots,2n-1}$. Kako je $x_k - x_{k-1} = h = \mathrm{const}\,(k = 1,\ldots,2n-1)$, smenom $x = x_o + kh$, dati skup podataka možemo zameniti skupom $\{(k,f_k)\}_{k=0,1,\ldots,2n-1}$.

Postavimo interpolacioni problem

(2.8.1) $$\varphi(k) = f_k \qquad (k = 0,1,\ldots,2n-1),$$

sa

$$\varphi(k) = c_1\varphi_1(k) + c_2\varphi_2(k) + \ldots + c_n\varphi_n(k) ,$$

gde je φ rešenje linearne diferencne jednačine reda n sa konstantnim koeficijentima

(2.8.2) $\quad \varphi(k+n)+a_n\varphi(k+n-1)+\ldots+a_2\varphi(k+1)+a_1\varphi(k)=0.$

S obzirom na (2.8.1), iz (2.8.2) za $k=0,1,\ldots,n-1$ sleduje sistem jednačina

$$\begin{aligned} f_oa_1+f_1a_2+\ldots+f_{n-1}a_n &= -f_n,\\ f_1a_1+f_2a_2+\ldots+f_na_n &= -f_{n+1},\\ &\vdots\\ f_{n-1}a_1+f_na_2+\ldots+f_{2n-2}a_n &= -f_{2n-1}. \end{aligned} \tag{2.8.3}$$

Ako je determinanta

$$\begin{vmatrix} f_o & f_1 & \cdots & f_{n-1}\\ f_1 & f_2 & & f_n\\ \vdots & & & \\ f_{n-1} & f_n & & f_{2n-2} \end{vmatrix} \neq 0,$$

rešavanjem sistema jednačina (2.8.3) nalazimo koeficijente a_i $(i=1,\ldots,n)$ i na taj način dobijamo diferencnu jednačinu (2.8.2). Odredjivanjem korena $r_1,r_2,\ldots,r_n$ karakteristične jednačine

$$r^n+a_nr^{n-1}+\ldots+a_2r+a_1=0,$$

nalazimo partikularna rešenja diferencne jednačine $\varphi_1,\varphi_2,\ldots,\varphi_n$ (videti odeljak 1.3.1). Ovim se postavljeni interpolacioni problem svodi na linearan problem. Naime, konstante $C_1,C_2,\ldots,C_n$ se mogu odrediti, na primer, iz sistema linearnih jednačina

(2.8.4) $\quad C_1\varphi_1(k)+C_2\varphi_2(k)+\ldots+C_n\varphi_n(k)=f_k \quad (k=0,1,\ldots,n-1).$

Najzad, imajući u vidu smenu $x=x_o+kh$, dobijamo interpolacionu funkciju

$$\Phi(x)=\varphi\Big(\frac{x-x_o}{h}\Big)=C_1\varphi_1\Big(\frac{x-x_o}{h}\Big)+C_2\varphi_2\Big(\frac{x-x_o}{h}\Big)+\ldots+C_n\varphi_n\Big(\frac{x-x_o}{h}\Big)$$

kojom je interpolirań skup podataka $\{(x_k,f_k)\}_{k=0,1,\ldots,2n-1}$. Ovakva interpolacija nosi naziv Pronyeva(eksponencijalna) interpolacija([4]). Problemi vezani za egzistenciju interpolacione funkcije izlaze iz okvira ove knjige.

Jedan interesantan prilaz u odredjivanju funkcije Φ dat je u [5].

<u>Primer 2.8.1.</u> Za skup podataka

k	0	1	2	3	4	5	6	7
x_k	0	0.5	1	1.5	2	2.5	3	3.5
f_k	8	2	4	0	-3	-5	-7.75	-10.375

sistem jednačina (2.8.3) postaje

$$\begin{aligned}
8a_1 + 2a_2 + 4a_3 + 0a_4 &= 3,\\
2a_1 + 4a_2 + 0a_3 - 3a_4 &= 5,\\
4a_1 + 0a_2 - 3a_3 - 5a_4 &= 7.75,\\
0a_1 - 3a_2 - 5a_3 - 7.75a_1 &= 10.375,
\end{aligned}$$

odakle je $a_1 = a_3 = 0.25$, $a_2 = 0$, $a_4 = -1.5$. Kako se karakteristična jednačina

$$r^4 - 1.5r^3 + 0.25r^2 + 0.25 = 0,$$

može predstaviti u obliku

$$(r-1)^2(r^2 + 0.5r + 0.25) = 0$$

imamo $r_1 = r_2 = 1$, $r_{3,4} = \frac{1}{4}(-1 \pm i\sqrt{3})$ pa je

$$\varphi(k) = C_1 + C_2k + 2^{-k}\left(C_3\cos\frac{2\pi k}{3} + C_4\sin\frac{2\pi k}{3}\right).$$

Najzad, iz (2.8.4) sleduje

$$C_1 = \frac{376}{49},\quad C_2 = -\frac{18}{7},\quad C_3 = \frac{16}{49},\quad C_4 = -\frac{592}{49\sqrt{3}}$$

pa je tražena interpolaciona funkcija data sa

$$\Phi(x) = \varphi(2x) = \frac{4}{49}\left(94 - 63x + 4^{1-x}\left(\cos\frac{4\pi x}{3} - \frac{37}{\sqrt{3}}\sin\frac{4\pi x}{3}\right)\right).$$

6.3. PROBLEM NAJBOLJIH APROKSIMACIJA

6.3.1. Osnovni pojmovi

Neka je funkcija $f: [a,b] \to R$ data skupom parova vrednosti (x_j, f_j) $(j = 0,1,\ldots,m)$, gde je $f_j \equiv f(x_j)$. Razmotrićemo problem aproksimacije funkcije f, linearnom aproksimacionom funkcijom (videti odeljak 6.1.2)

$$\Phi(x) = \Phi(x; a_0, \ldots, a_n) = \sum_{i=0}^{n} a_i \Phi_i(x),$$

pri čemu je $m > n$*. Postupajući kao kod interpolacije, u ovom slučaju, dolazimo do tzv. preodredjenog sistema jednačina

$$\sum_{i=0}^{n} a_i \Phi_i(x_j) = f_j \qquad (j = 0,1,\ldots,m), \tag{3.1.1}$$

koji, u opštem slučaju, nema rešenje, tj. sve jednačine sistema (3.1.1) ne mogu biti istovremeno zadovoljene.

Ako definišemo δ_n pomoću

$$\delta_n(x) = f(x) - \sum_{i=0}^{n} a_i \Phi_i(x), \tag{3.1.2}$$

moguće je tražiti "rešenje" sistema (3.1.1) takvo da je

$$\| \delta_n^* \|_r = \min_{a_i} \| \delta_n \|_r , \tag{3.1.3}$$

gde je

---

*) Za $m = n$ imamo interpolaciju.

$$\|\delta_n\|_r = \left(\sum_{j=1}^{m} |\delta_n(x_j)|^r\right)^{1/r} \qquad (r \geq 1).$$

Jednakost (3.1.3) daje kriterijum za odredjivanje parametara $a_o, a_1, \ldots, a_n$ u aproksimacionoj funkciji Φ. Veličinu $\|\delta_n^*\|_r$, koja uvek egzistira, nazivamo veličinom najbolje aproksimacije u ℓ^r. Optimalne vrednosti parametara $a_i = \bar{a}_i$ $(i = 0, 1, \ldots, n)$ u smislu (3.1.3) daju najbolju ℓ^r aproksimacionu funkciju

$$\Phi^*(x) = \sum_{i=0}^{n} \bar{a}_i \Phi_i(x).$$

Najčešće se, u praksi, uzima:

1° $r = 1$, $\|\delta_n\|_1 = \sum_{j=0}^{m} |\delta_n(x_j)|$ (najbolja ℓ^1 aproksimacija);

2° $r = 2$, $\|\delta_n\|_2 = \left(\sum_{j=0}^{m} \delta_n(x_j)^2\right)^{1/2}$ (srednje-kvadratna aproksimacija);

3° $r = +\infty$, $\|\delta_n\|_\infty = \max_{0 \leq j \leq m} |\delta_n(x_j)|$ (Čebiševljeva mini-max aproksimacija).

Analogno prethodnom, može se razmatrati problem najbolje aproksimacije funkcije f u prostoru $L^r(a,b)$. Ovde je

$$\|\delta_n\|_r = \left(\int_a^b |\delta_n(x)|^r\right)^{1/r} \qquad (1 \leq r < +\infty)$$

i

$$\|\delta_n\|_\infty = \max_{a \leq x \leq b} |\delta_n(x)| .$$

Uvedjenjem težinske funkcije $p: [a,b] \to R^+$ može se razmatrati opštiji slučaj srednje-kvadratnih aproksimacija, pri čemu su odgovarajuće norme za diskretan i kontinualan slučaj date sa

$$(3.1.4)\qquad \|\delta_n\|_2 = \|\delta_n\|_{2,p} = \left(\sum_{j=0}^{m} p(x_j)\,\delta_n(x_j)^2\right)^{1/2}$$

i

$$(3.1.5)\qquad \|\delta_n\|_2 = \|\delta_n\|_{2,p} = \left(\int_a^b p(x)\,\delta_n(x)^2\right)^{1/2}$$

respektivno.

Primer 3.1.1. Funkciju $x \mapsto f(x) = x^{1/3}$ aproksimirajmo funkcijom $x \mapsto \phi(x) = a_0 + a_1 x$ u prostoru

$$1^\circ\ L^1(0,1)\ ;\quad 2^\circ\ L^2(0,1)\ ;\quad 3^\circ\ L^\infty(0,1).$$

Ovde je $\delta_1(x) = x^{1/3} - a_0 - a_1 x$ $(0 \leqq x \leqq 1)$.

1° Najbolju $L^1(0,1)$ aproksimaciju dobijamo minimizacijom norme

$$\|\delta_1\|_1 = \int_0^1 \left|x^{1/3} - a_0 - a_1 x\right| dx .$$

Kako je $\dfrac{\partial\delta_1}{\partial a_0} = -1$ i $\dfrac{\partial\delta_1}{\partial a_1} = -x$, optimalne vrednosti parametara a_0 i a_1 odredjujemo iz sistema jednačina

$$(3.1.6)\qquad \begin{aligned} &\int_0^1 \operatorname{sgn}\delta_1(x)dx = 0,\\ &\int_0^1 x\operatorname{sgn}\delta_1(x)dx = 0. \end{aligned}$$

S obzirom da se može uzeti da δ_1 menja znak na $[0,1]$ u tačkama x_1 i x_2(videti sl.3.1.1), to se sistem jednačina (3.1.6) svodi na sistem

$$x_2 - x_1 = \frac{1}{2}, \quad x_1^2 - x_2^2 = \frac{1}{2},$$

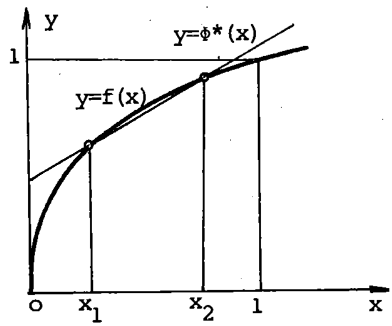
odakle sleduje $x_1 = \frac{1}{4}$ i $x_2 = \frac{3}{4}$.

Dakle, problem odredjivanja najbolje $L^1(0,1)$ aproksimacije se svodi na interpolacioni problem, tj. na odredjivanje interpolacionog polinoma Φ^* koji zadovoljava uslove

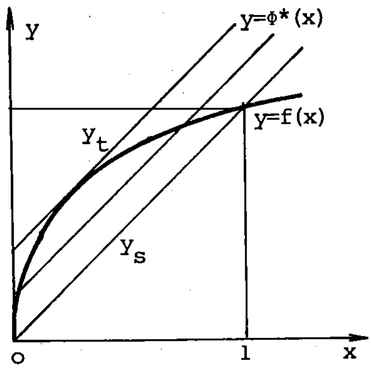
$$\Phi^*(1/4) = f(1/4) = \sqrt[3]{\frac{1}{4}} \quad \text{i} \quad \Phi^*(3/4) = f(3/4) = \sqrt[3]{\frac{3}{4}}.$$

Prema tome, imamo

$$\Phi^*(x) = \frac{2}{\sqrt[3]{4}}(\sqrt[3]{3} - 1)x + \frac{1}{2\sqrt[3]{4}}(3 - \sqrt[3]{3}) \cong 0.55720x + 0.49066.$$



Sl. 3.1.1



Sl. 3.1.2

2^o Neka je

$$I(a_o, a_1) = \|\delta_1\|_2^2 = \int_0^1 (x^{1/3} - a_o - a_1 x)^2 dx.$$

Tada iz uslova

$$\frac{\partial I}{\partial a_o} = -2 \int_0^1 (x^{1/3} - a_o - a_1 x)\, dx = 0,$$

$$\frac{\partial I}{\partial a_1} = -2 \int_0^1 x(x^{1/3} - a_o - a_1 x)\, dx = 0,$$

sleduje

$$a_o + \frac{1}{2}a_1 = \frac{3}{4}\ , \quad \frac{1}{2}a_o + \frac{1}{3}a_1 = \frac{3}{7},$$

tj. $a_o = \bar{a}_o = \frac{3}{7}$, $a_1 = \bar{a}_1 = \frac{9}{14}$. Dakle, najbolja srednje-kvadratna aproksimacija je data sa

$$\Phi^*(x) = \frac{3}{7} + \frac{9}{14}x \cong 0.42857 + 0.64286\,x\ .$$

3^o Za odredjivanje najbolje mini-max aproksimacije iskoristićemo sledeći prost geometrijski postupak. Kroz krajnje tačke krive $y = f(x) = x^{1/3}$ $(0 \leqq x \leqq 1)$ postavimo sečicu, a zatim tangentu krive koja je paralelna sa ovom sečicom (videti sl. 3.1.2). Odgovarajuće jednačine ovih pravih su redom

$$y = y_s = x \qquad \text{i} \qquad y = y_t = x + \frac{2\sqrt{3}}{9}.$$

Nije teško zaključiti da je tražena najbolja mini-max aproksimacija data sa

$$\Phi^*(x) = \frac{1}{2}(y_s + y_t) = x + \frac{\sqrt{3}}{9} \cong x + 0.19245,$$

pri čemu je veličina najbolje aproksimacije $\|\delta_1^*\|_\infty = \frac{\sqrt{3}}{9}$. Primetimo da je

$$\delta_1(0) = -\,\delta_1(\sqrt{3}/9) = \delta_1(1) = \|\delta_1^*\|_\infty\ .$$

U skladu sa prethodnim možemo tretirati opšti slučaj aproksimacije u normiranim prostorima. Na primer, neka je X Banachov prostor sa normom $\|\cdot\|$. U prostoru X uočimo sistem linearno nezavisnih elemenata $B = \{\Phi_o, \Phi_1, \ldots, \Phi_n\}$ i pomoću njih obrazujmo podprostor $X_n = L(B)$ (lineal nad B). Za fiksno $f \in X$ posmatrajmo veličinu

$$\Delta(f,\Phi) = \|f - \Phi\|,$$

gde $\Phi \in X_n$, tj. $\Phi = \sum_{i=0}^{n} c_i\Phi_i$ $(c_i \in R)$. Neka je, dalje,

$$\Delta_n(f) = \inf_{\Phi \in X_n} \Delta(f,\Phi).$$

Element $\Phi^* \in X_n$ za koji je $\Delta_n(f) = \|f - \Phi^*\|$ nazivamo elementom najbolje aproksimacije za f u podprostoru X_n, a veličinu $\Delta_n(f)$ veličinom najbolje aproksimacije.

6.3.2. Srednje-kvadratna aproksimacija

U ovom odeljku razmatraćemo problem srednje-kvadratne aproksimacije funkcije $f: [a,b] \to R$ pomoću linearne aproksimacione funkcije

$$\Phi(x) = \sum_{i=0}^{n} a_i \Phi_i(x),$$

gde je $\{\Phi_i\}$ sistem linearno nezavisnih funkcija iz prostora $L^2(a,b)$, u kome je skalarni proizvod uveden pomoću

$$(f,g) = \int_a^b p(x) f(x) g(x) dx \qquad (f,g \in L^2(a,b)),$$

gde je $p: [a,b] \to R^+$ data težinska funkcija.

Koristeći se oznakama i rezultatima iz prethodnog odeljka, za dobijanje najbolje srednje-kvadratne aproksimacije za f, zaključujemo da je potrebno minimizirati normu (3.1.5) po parametrima a_i $(i = 0,1,\ldots,n)$.

Ako stavimo $I(a_0,a_1,\ldots,a_n) = \|\delta_n\|^2 = (\delta_n,\delta_n)$, tada iz

$$\frac{\partial I}{\partial a_j} = 2 \int_a^b p(x)\Big(f(x) - \sum_{i=0}^{n} a_i \Phi_i(x)\Big)(-\Phi_j(x)) dx = 0 \quad (j=0,1,\ldots,n)$$

sleduje sistem jednačina za odredjivanje aproksimacionih parametara

$$\sum_{i=0}^{n} (\Phi_i,\Phi_j) a_i = (f,\Phi_j) \qquad (j = 0,1,\ldots,n).$$

Ovaj sistem se može predstaviti i u matričnom obliku

$$\text{(3.2.1)}\qquad \begin{bmatrix} (\Phi_o,\Phi_o) & (\Phi_1,\Phi_o) & \dots & (\Phi_n,\Phi_o) \\ (\Phi_o,\Phi_1) & (\Phi_1,\Phi_1) & & (\Phi_n,\Phi_1) \\ \vdots & & & \\ (\Phi_o,\Phi_n) & (\Phi_1,\Phi_n) & & (\Phi_n,\Phi_n) \end{bmatrix} \cdot \begin{bmatrix} a_o \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} (f,\Phi_o) \\ (f,\Phi_1) \\ \vdots \\ (f,\Phi_n) \end{bmatrix} .$$

Matrica ovog sistema je posznata kao Gramova matrica. Može se pokazati da je ova matrica regularna ako je sistem funkcija $\{\Phi_i\}$ linearno nezavisan, što znači da u tom slučaju postoji jedinstveno rešenje datog aproksimacionog problema.

Sistem jednačina (3.2.1) se može jednostavno rešiti ako je sistem funkcija $\{\Phi_i\}$ ortogonalan. Naime, tada matrica ovog sistema postaje dijagonalna pa su rešenja sistema data sa

$$\text{(3.2.2)}\qquad a_i = \bar{a}_i = \frac{(f,\Phi_i)}{(\Phi_i,\Phi_i)} \qquad (i = 0,1,\dots,n).$$

Pokažimo sada da pri ovako izabranim parametrima a_i $(i = 0,1,\dots,n)$ funkcija I, zaista dostiže minimum. Naime, kako je

$$\frac{\partial^2 I}{\partial a_j \partial a_k} = 2(\Phi_k,\Phi_j) = 2\|\Phi_k\|^2\,\delta_{kj} ,$$

gde je δ_{kj} Kroneckerova delta, imamo

$$d^2 I = 2\sum_{i=0}^{n} \|\Phi_k\|^2 da_k^2 > 0,$$

što dokazuje pomenuto tvrdjenje.

Dakle, najbolja srednje-kvadratna aproksimacija funkcije f na podprostoru $X_n = L(\Phi_o,\Phi_1,\dots,\Phi_n)$, gde je $\{\Phi_i\}$ ortogonalan sistem funkcija, data je sa

$$\Phi^*(x) = \sum_{i=0}^{n} \frac{(f,\Phi_i)}{\|\Phi_i\|^2}\,\Phi_i(x) .$$

Jedna važna klasa srednje-kvadratnih aproksimacija je aproksimacija algebarskim polinomima. U tom slučaju ortogonalnu bazu podprostora X_n konstruišemo Gram-Schmidtovim postupkom ortogonalizacije, polazeći, na primer, od prirodne baze $\{1,x,x^2,\dots,x^n\}$ (videti odeljak 2.1.4).

Primer 3.2.1. Za funkciju $x\mapsto f(x)=|x|$ na segmentu $[-1,1]$ odredimo u skupu polinoma ne višeg stepena od trećeg najbolju srednje-kvadratnu aproksimaciju sa težinskom funkcijom $x\mapsto p(x)=(1-x^2)^{3/2}$.

Sistem ortogonalnih polinoma sa ovom težinskom funkcijom konstruisan je u primeru 1.4.1(odeljak 2.1.4), pri čemu je

$$\Phi_0(x)=1\,,\ \Phi_1(x)=x\,,\ \Phi_2(x)=x^2-\frac{1}{6}\,,\ \Phi_3(x)=x^3-\frac{3}{8}x$$

i

$$\|\Phi_0\|=\sqrt{\frac{3\pi}{8}}\,,\|\Phi_1\|=\frac{\sqrt{\pi}}{4}\,,\ \|\Phi_2\|=\frac{1}{8}\sqrt{\frac{5\pi}{6}}\,,\|\Phi_3\|=\frac{\sqrt{3\pi}}{32}\,.$$

Kako je

$$(f,\Phi_0)=\frac{2}{5}\,,\ (f,\Phi_1)=0,\ (f,\Phi_2)=\frac{1}{21}\,,\ (f,\Phi_3)=0,$$

na osnovu (3.2.2) dobijamo

$$a_0=\frac{16}{15\pi}\,,\ a_1=0,\ a_2=\frac{128}{35\pi}\,,\ a_3=0,$$

što znači da je tražena aproksimacija data sa

$$\Phi^*(x)=\frac{16}{15\pi}+\frac{128}{35\pi}\left(x^2-\frac{1}{6}\right)\cong 0.14551309+1.1641047x^2\,.$$

Primetimo da je ova funkcija takodje najbolja aproksimacija i u skupu polinoma ne višeg stepena od dva.

Primedba 3.2.1. Za veličinu najbolje srednje-kvadratne aproksimacije lako se može dokazati formula

$$\|f-\Phi^*\|^2=\|f\|^2-\sum_{k=0}^{n}\bar{a}_i^2\|\Phi_i\|^2\,,$$

gde su koeficijenti $\bar{a}_i\,(i=0,1,\dots,n)$ odredjeni formulom (3.2.2).

Primer 3.2.2. U skupu polinoma ne višeg stepena od prvog, najbolja srednje-kvadratna aproksimacija za funkciju $x\mapsto\sin x$ na segmentu $[0,\pi/2]$ sa težinom $p(x)=1$ je

$$\Phi^*(x) = \frac{1}{\pi^2}(8\pi - 24) + \frac{1}{\pi^3}(96 - 24\pi)x.$$

Ukoliko se za težinsku funkciju uzme $p(x) = x$ dobija se

$$\Phi^*(x) = \frac{1}{\pi^3}(384 - 120\pi) + \frac{1}{\pi^4}(384\pi - 1152).$$

6.3.3. Ekonomizacija stepenih redova

Kao što je napomenuto u odeljku 1.3.2 za izračunavanje vrednosti funkcija vrlo često se koriste polinomski razvoji. Na primer, ako funkcija f na segmentu $[-1,1]$ ima razvoj

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots ,$$

tada se za izračunavanje vrednosti ove funkcije na segmentu $[-1,1]$ može, sa odredjenom tačnošću, koristiti polinom

$$P_n(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n . \tag{3.3.1}$$

Postupak ekonomizacije stepenih redova, koji potiče od Lanczosa, sastoji se u **snižavanju** stepena polinoma (3.3.1) uz neznatno povećanje greške, a izvodi se vrlo jednostavno uz korišćenje ortogonalnih polinoma. Najčešće se koriste Čebiševljevi i Legendreovi polinomi.

Razmotrićemo ekonomizaciju pomoću Čebiševljevih polinoma $T_k (k = 0,1,\ldots)$. Kako su

$$T_0(x) = 1,\ T_1(x) = x,\ T_2(x) = 2x^2 - 1,\ T_3(x) = 4x^3 - 3x ,$$

$$T_4(x) = 8x^4 - 8x^2 + 1,\ T_5(x) = 16x^5 - 20x^3 + 5x ,$$

$$T_6(x) = 32x^6 - 48x^4 + 18x^2 - 1,\ T_7(x) = 64x^7 - 112x^5 + 56x^3 - 7x,$$

itd., mogućno je izraziti algebarske stepene $x^k (k = 0,1,\ldots)$ pomoću Čebiševljevih polinoma na sledeći način

$$1 = T_0 ,\ x = T_1 ,\ x^2 = \frac{1}{2}(T_0 + T_2) ,\ x^3 = \frac{1}{4}(3T_1 + T_3) ,$$

$$x^4 = \frac{1}{8}(3T_0 + 4T_2 + T_4),\ x^5 = \frac{1}{16}(10T_1 + 5T_3 + T_5) ,$$

$$x^6 = \frac{1}{32}(10T_0 + 15T_2 + 6T_4 + T_6),\quad x^7 = \frac{1}{64}(35T_1 + 21T_3 + 7T_5 + T_7),$$

itd. Pomoću ovih formula polinom (3.3.1) se može predstaviti u obliku

$$P_n(x) = c_0T_0 + c_1T_1 + \dots + c_nT_n. \tag{3.3.2}$$

Označimo sa $\mathcal{P}_m$ skup svih algebarskih polinoma ne višeg stepena od m. Uzimajući samo prvih $m+1 (m < n)$ članova u razvoju (3.3.2) dobijamo polinom

$$Q_m(x) = c_0T_0 + c_1T_1 + \dots + c_mT_m, \tag{3.3.3}$$

koji predstavlja aproksimaciju za P_n u skupu $\mathcal{P}_m$.

S obzirom da Čebiševljevi polinomi zadovoljavaju nejednakost $|T_k(x)| \leqq 1$ $(|x| \leqq 1)$, za grešku aproksimacije važi

$$|P_n(x) - Q_m(x)| \leqq |c_{m+1}| + \dots + |c_n| \qquad (|x| \leqq 1).$$

Navedeni postupak aproksimacije naziva se Lanczosova ekonomizacija. Sledeća teorema daje odgovor na pitanje o kakvoj se aproksimaciji radi.

<u>Teorema 3.3.1</u>. Polinom Q_m, odredjen sa (3.3.3) predstavlja u skupu $\mathcal{P}_m$ najbolju srednje-kvadratnu aproksimaciju sa težinskom funkcijom $p(x) = (1-x^2)^{-1/2}$ za polinom P_n na segmentu $[-1,1]$.

<u>Dokaz</u>. Ako stavimo

$$(f,g) = \int_{-1}^{1} (1-x^2)^{-\frac{1}{2}} f(x)g(x)dx,$$

na osnovu (3.3.2), vidimo da za koeficijente u razvoju (3.3.3) važi

$$(P_n, T_k) = c_k(T_k, T_k) \qquad (k = 0,1,\dots,m),$$

odakle, uporedjivanjem sa (3.2.2), sleduje tvrdjenje teoreme.

Na osnovu dokazane teoreme zaključujemo da je

$$\min_{p\in\mathcal{P}_m} \int_{-1}^{1}(1-x^2)^{-\frac{1}{2}}(P_n(x)-p(x))^2dx = \int_{-1}^{1}(1-x^2)^{-\frac{1}{2}}(P_n(x)-Q_m(x))^2dx.$$

U daljem tekstu navodimo dva primera.

Primer 3.3.1. Postupkom ekonomizacije aproksimirajmo $x \mapsto x^6$ $(|x| \leq 1)$ pomoću polinoma ne višeg stepena od dva. Kako je $x^6 = \frac{1}{32}(10T_0 + 15T_2 + 6T_4 + T_6)$, odsecanjem poslednja dva člana dobijamo polinom $Q_2(x) = \frac{10}{32} + \frac{15}{32}(2x^2 - 1) = \frac{15}{16}x^2 - \frac{5}{32}$, pri čemu je apsolutna greška manja od $\frac{7}{32}$, tj. važi $|x^6 - Q_2(x)| \leq \frac{7}{32}$ $(|x| \leq 1)$.

Na osnovu razvoja po Legendreovim polinomima(srednje-kvadratna aproksimacija sa težinom $p(x) = 1$), dobija se sledeća aproksimacija

$$x^6 \sim \frac{1}{7}P_0 + \frac{10}{21}P_2 = \frac{5}{7}x^2 - \frac{2}{21} \qquad (|x| \leq 1)$$

sa apsolutnom greškom ne većom od $\frac{8}{21}$. Primetimo da je u ovom slučaju učinjena veća greška, jer je $\frac{8}{21} > \frac{7}{32}$. U primeru 3.5.2 (odeljak 6.3.5) videćemo da postoji polinom drugog stepena za koji je ova greška još manja, tačnije načeno najmanja.

Primer 3.3.2. Posmatrajmo aproksimaciju funkcije $x \mapsto e^x$ Taylor-ovim polinomom petog stepena, tj.

$$e^x \approx P_5(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4 + \frac{1}{120}x^5 .$$

Za grešku ove aproksimacije, na segmentu $[-1,1]$ važi ocena

$$|E_5(x)| = |e^x - P_5(x)| = \frac{e^{\theta x}|x|^6}{720} < \frac{e}{720} < 0.0038 ,$$

s obzirom da $\theta \in (0,1)$. Razvojem polinoma P_5 po Čebiševljevim polinomima dobijamo

$$P_5(x) = \frac{81}{64}T_0 + \frac{217}{192}T_1 + \frac{13}{48}T_2 + \frac{17}{384}T_3 + \frac{1}{192}T_4 + \frac{1}{1920}T_5 ,$$

odakle, odsecanjem poslednja dva člana, sleduje

$$Q_3(x) = \frac{1}{384}(382 + 383x + 208x^2 + 68x^3),$$

pri čemu za ukupnu grešku aproksimacije važi

$$|e^x - Q_3(x)| < \frac{e}{720} + \frac{1}{192} + \frac{1}{1920} \cong 0.0095 \qquad (|x| \leq 1).$$

Primedba 3.3.1. Ako bismo funkciju $x \mapsto e^x$ na segmentu $[-1,1]$ direktno aproksimirali srednje-kvadratnom aproksimacijom sa težinom $p(x) = (1-x^2)^{-1/2}$, polinomom trećeg stepena dobili bismo

$$e^x \approx 1.266066\,T_0 + 1.130318T_1 + 0.271495\,T_2 + 0.044337\,T_3$$

$$= 0.994571 + 0.997307x + 0.542990x^2 + 0.177348x^3 ,$$

Naime, koeficijenti u ovom razvoju su dati pomoću

$$a_k = \frac{(e^x, T_k)}{(T_k, T_k)} = \frac{1}{\|T_k\|^2}\int_{-1}^{1}(1-x^2)^{-1/2}\, e^x T_k(x)dx = \frac{1}{\|T_k\|^2}\int_0^{\pi} e^{\cos\theta}\cos k\theta d\theta ,$$

tj.

$$a_0 = I_0(1)\ ,\quad a_k = 2I_k(1) \quad (k = 1,2,3)\ ,$$

gde su I_k modifikovane Besselove funkcije prve vrste (videti, na primer, [2]).

6.3.4. Diskretna srednje-kvadratna aproksimacija

U prethodnim odeljcima proučavali smo problem najbolje aproksimacije funkcija u prostoru $L^2(a,b)$. Sada ćemo razmatrati jedan partikularni slučaj problema, sa kojim smo započeli uvodni odeljak 6.3.1. Naime, neka je funkcija $f:[a,b]\to R$ data na skupu parova vrednosti $\{(x_j, f_j)\}_{j=0,1,\ldots,m}$, gde je $f_j = f(x_j)$. Razmotrićemo problem najbolje aproksimacije ove funkcije, linearnom aproksimacionom funkcijom

$$\Phi(x) = \sum_{i=0}^{n} a_i\Phi_i(x) \qquad (n<m),$$

u smislu minimizacije norme (3.1.4), gde je $p:[a,b]\to R^+$ data težinska funkcija i δ_n definisano pomoću (3.1.2).

Ako uvedemo matričnu notaciju

$$X = \begin{bmatrix} \Phi_0(x_0) & \Phi_1(x_0) & \ldots & \Phi_n(x_0) \\ \Phi_0(x_1) & \Phi_1(x_1) & & \Phi_n(x_1) \\ \vdots & & & \\ \Phi_0(x_m) & \Phi_1(x_m) & & \Phi_n(x_m) \end{bmatrix}, \quad \vec{f} = \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_m \end{bmatrix}, \quad \vec{a} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix},$$

$$P = \text{diag}(p(x_o), p(x_1), \ldots, p(x_m)), \quad \vec{v} = \vec{f} - X\vec{a},$$

kvadrat norme, definisane pomoću (3.1.4), se može predstaviti u obliku

$$\|\delta_n\|^2 = \|\delta_n\|_2^2 = \sum_{j=o}^{m} p(x_j)\delta_n(x_j)^2 = (P\vec{v}, \vec{v}) = \vec{v}^T P \vec{v} .$$

Na osnovu

$$\frac{\partial \|\delta_n\|^2}{\partial a_i} = -2\sum_{j=o}^{m} p(x_j)\delta_n(x_j) \frac{\partial \delta_n(x_j)}{\partial a_i} = 0 \qquad (i = 0,1,\ldots,n)$$

dobijamo tzv. normalni sistem jednačina

$$\sum_{j=o}^{m} p(x_j)\delta_n(x_j)\Phi_i(x_j) = 0 \qquad (i = 0,1,\ldots,n)$$

za odredjivanje parametara $a_i (i = 0,1,\ldots,n)$. Shodno uvedenim oznakama, poslednji sistem jednačina se može predstaviti u matričnom obliku

$$X^T P \vec{v} = 0,$$

tj.

$$X^T P X \vec{a} = X^T P \vec{f}. \tag{3.4.1}$$

Primetimo da se normalni sistem jednačina (3.4.1) dobija iz preodredjenog sistema jednačina (3.1.1), predstavljenog u matričnom obliku

$$X\vec{a} = \vec{f},$$

jednostavnim množenjem matricom $X^T P$ sa leve strane.

Dijagonalna matrica P, koja se naziva težinskom matricom, ima smisao takav da se vrednostima funkcije f_j sa većom tačnošću dodeljuju veće težine $p(x_j)$. Često se, medjutim, uzima da je P jedinična matrica reda m+1. U tom slučaju (3.4.1) se svodi na

$$X^T X \vec{a} = X^T \vec{f}. \tag{3.4.2}$$

Vektor traženih koeficijenata $\vec{a}$ odredjujemo iz (3.4.1), odnosno (3.4.2). Na primer, iz (3.4.1) sleduje

$$\vec{a} = (X^T X)^{-1} X^T \vec{f}.$$

U slučaju da je sistem bazisnih funkcija izabran kao $\Phi_i(x) = x^i (i = 0,1,\dots,n)$ imamo

$$X = \begin{bmatrix} 1 & x_0 & \cdots & x_n^n \\ 1 & x_1 & & x_1^n \\ \vdots & & & \\ 1 & x_m & & x_m^n \end{bmatrix}.$$

Posmatrani slučaj diskretne srednje-kvadratne aproksimacije vrlo često se naziva metod najmanjih kvadrata.

Primer 3.4.1. Metodom najmanjih kvadrata nadjimo parametre a_0 i a_1 u aproksimacionoj funkciji $\Phi(x) = a_0 + a_1 x$ za sledeći skup parova (x_j, f_j):

$$\{(1.1,2.5),(1.9,3.2),(4.2,4.5),(6.1,6.0)\}.$$

Uzmimo da je težinska matrica P jednaka jediničnoj matrici.

Množenjem preodredjenog sistema jednačina

$$\begin{bmatrix} 1 & 1.1 \\ 1 & 1.9 \\ 1 & 4.2 \\ 1 & 6.1 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 2.5 \\ 3.2 \\ 4.5 \\ 6.0 \end{bmatrix}$$

matricom $X^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1.1 & 1.9 & 4.2 & 6.1 \end{bmatrix}$ sa leve strane dolazimo do normalnog sistema jednačina

$$\begin{bmatrix} 4 & 13.3 \\ 13.3 & 59.67 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 16.2 \\ 64.33 \end{bmatrix},$$

odakle je

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \frac{1}{61.79} \begin{bmatrix} 59.67 & -13.3 \\ -13.3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 16.2 \\ 64.33 \end{bmatrix} = \begin{bmatrix} 1.7974591 \\ 0.6774559 \end{bmatrix}.$$

Dakle, $\Phi(x) \cong 1.797 + 0.677x$.

Primer 3.4.2. Neka su rezultati merenja veličina x i f dati u tabeli

x_j	4.48	4.98	5.60	6.11	6.62	7.42
f_j	4.15	1.95	1.31	1.03	0.74	0.63

Za aproksimaciju funkcije f pretpostavimo aproksimacionu funkciju u obliku $\varphi(x) = \frac{1}{a_0 + a_1 x}$. Tada uvodjenjem smene $\varphi(x) = \frac{1}{\Phi(x)}$, metodom najmanjih kvadrata nalazimo $\Phi(x) \cong 0.468x - 1.843$, odakle je

$$\varphi(x) \cong \frac{1}{0.468x - 1.843} .$$

<u>Primedba 3.4.1</u>. Sistem funkcija $\{\phi_i\}$ moguéno je ortogonalizirati Gram-Schmidtovim postupkom, tako je matrica normalnog sistema jednačina dijagonalna.

<u>6.3.5. Čebiševljeva mini-max aproksimacija</u>

U ovom odeljku daćemo osnovne rezultate koji se odnose na mini-max aproksimaciju funkcije $f \in C[a,b]$ algebarskim polinomima. Dobar deo ovih rezultata može se preneti i na opštije tipove aproksimacionih funkcija.

Sa $\mathcal{P}_n$ označimo skup svih algebarskih polinoma ne višeg stepena od n. Pred nama se postavlja problem odredjivanja polinoma $P_n = P_n^*(\in \mathcal{P}_n)$, koji minimizira normu $\|f-P_n\|_\infty$, koju često nazivamo uniformnom normom. Naime, treba rešiti sledeći mini-max problem

$$\min_{P_n \in \mathcal{P}_n} \left(\max_{a \le x \le b} |f(x) - P_n(x)| \right) = \max_{a \le x \le b} |f(x) - P_n^*(x)| .$$

<u>Teorema 3.5.1(Weierstrass)</u>. Ako $f \in C[a,b]$, tada za svako $\varepsilon > 0$ postoji prirodan broj n i polinom $P_n \in \mathcal{P}_n$ takav da je, za svako $x \in [a,b]$,

$$|f(x) - P_n(x)| \le \varepsilon .$$

<u>Teorema 3.5.2</u>. Polinom P_n^* je najbolja mini-max aproksimacija za $f \in C[a,b]$ u skupu $\mathcal{P}_n$ ako i samo ako na $[a,b]$ postoje n+2 tačke $x_0, x_1, \ldots, x_{n+1} (x_0 < x_1 < \ldots < x_{n+1})$, takve da je

$$\delta_n^*(x_k) = -\delta_n^*(x_{k+1}) = \pm\|\delta_n^*\|_\infty, \tag{3.5.1}$$

pri čemu je $\delta_n^*(x) = f(x) - P_n^*(x)$.

Dokaz egzistencije i jedinstvenosti polinoma P_n^* može se naći, na primer, u [6]. Koristeći se teoremom 3.5.1, može se dokazati da $\|\delta_n^*\|_\infty \to 0$, kada $n \to +\infty$.

Ako za neki polinom n-tog stepena postoje n+2 tačke sa svojstvom (3.5.1) kažemo da ima Čebiševljevu alternansu za f. Ako je $\varepsilon_k = (-1)^{k+1}$, svojstvo (3.5.1) se može izraziti u obliku

$$\delta_n^*(x_o) + \varepsilon_k \delta_n^*(x_k) = 0 \qquad (k = 1,\ldots,n+1). \tag{3.5.2}$$

Primer 3.5.1. Neka je $x \mapsto f(x) = x^{n+1}$ ($|x| \leq 1$) i neka je

$$\delta_n(x) = x^{n+1} - P_n(x),$$

gde je $P_n(x) = \sum_{i=o}^{n} a_i x^i$. Kako iz uslova $T_{n+1}(x) = \cos[(n+1)\arccos x] = \pm 1$, sleduje $x_k = -\cos\frac{k\pi}{n+1}$ $(k = 0,1,\ldots,n+1)$, pri čemu je $-1 = x_o < x_1 < \ldots < x_{n+1} = 1$, zaključujemo da na $[-1,1]$ postoje n+2 tačke u kojima je $T_{n+1}(x_k) = (-1)^{n+k+1}$ $(k = 0,1,\ldots,n+1)$.

Ako stavimo $\delta_n^*(x) = \frac{1}{2^n}T_{n+1}(x)$, vidimo da je $|\delta_n^*(x)| \leq \frac{1}{2^n}$ ($|x| \leq 1$) i da $\delta_n^*(x_k) = \frac{1}{2^n}(-1)^{n+k+1}$ zadovoljava (3.5.2).

Na osnovu prethodnog i teoreme 3.5.2 zaključujemo da se najbolji mini-max polinom P_n^* za funkciju $x \mapsto f(x) = x^{n+1}$ ($|x| \leq 1$) može dobiti iz jednakosti

$$x^{n+1} - P_n^*(x) = \frac{1}{2^n}T_{n+1}(x).$$

Dakle,

$$P_n^*(x) = x^{n+1} - \frac{1}{2^n}T_{n+1}(x).$$

U specijalnom slučaju, za $n = 1,2,3$ imamo

$$x^2 \sim P_1^*(x) = \frac{1}{2}, \quad x^3 \sim P_2^*(x) = \frac{3}{4}x, \quad x^4 \sim P_3^*(x) = x^2 - \frac{1}{8}.$$

Primetimo da se kod aproksimacije funkcije $x \mapsto x^{n+1}$ $(|x| \leq 1)$ u skupu $\mathcal{P}_n$, mini-max aproksimacijom i srednje-kvadratnom aproksimacijom sa težinom $(1-x^2)^{-1/2}$, dobijaju isti aproksimacioni polinomi.

Primer 3.5.2. U skupu $\mathcal{P}_2$ odredimo najbolju mini-max aproksimaciju za $x \mapsto f(x) = x^6$ $(|x| \leq 1)$. Za odredjivanje koeficijenata polinoma najbolje mini-max aproksimacije $P_2^*(x) = a_0 + a_1 x + a_2 x^2$, potrebno je naći $n+2 = 4$ tačke x_0, x_1, x_2, x_3, takve da je

$$\delta_2^*(x_0) = -\delta_2^*(x_1) = \delta_2^*(x_2) = -\delta_2^*(x_3) = \pm\Delta, \tag{3.5.3}$$

gde su $\delta_2^*(x) = x^6 - P_2^*(x)$ i $\Delta = \|\delta_2^*\|_\infty = \max\limits_{|x| \leq 1} |\delta_2^*(x)|$.

Zbog simetrije problema može se uzeti $a_1 = 0$, a za tačke x_k $(k = 0,1,2,3)$, na primer, $x_0 = -t, x_1 = 0, x_2 = t, x_3 = 1$, gde je t $(0 < t < 1)$ tačka u kojoj δ_2^* dostiže ekstremnu vrednost. Dakle, t je pozitivan koren jednačine $6t^5 - 2a_2 t = 0$.

Kako je, na osnovu (3.5.3),

$$a_0 = t^6 - (a_0 + a_2 t^2) = a_0 + a_2 - 1,$$

lako nalazimo $a_2 = 1$ i $a_0 = -\frac{\sqrt{3}}{9}$. Prema tome, najbolja mini-max aproksimacija za $x \mapsto x^6$ $(|x| \leq 1)$ u skupu $\mathcal{P}_2$ je

$$P_2^*(x) = x^2 - \frac{\sqrt{3}}{9},$$

pri čemu je veličina najbolje aproksimacije(maksimalno odstupanje) $\Delta = \|\delta_2^*\|_\infty = \frac{\sqrt{3}}{9} \cong 0.19245$.

Primer 3.5.3. Istim postupkom kao u prethodnom primeru, nalazi se najbolji mini-max polinom u skupu $\mathcal{P}_2$ za funkciju $x \mapsto f(x) = |x|$ $(|x| \leq 1)$, $P_2^*(x) = 1/8 + x^2$, pri čemu je $\Delta = \|f - P_2^*\|_\infty = 1/8$.

Oslanjajući se na teoremu 3.5.2 konstruišu se algoritmi za odredjivanje najbolje mini-max aproksimacije date funkcije. Jedan od najprikladnijih **algoritama je Rémèsov algoritam. Jedna** varijanta Rémèsovog algoritma se može iskazati na sledeći način:

1° Izabere se skup od $n+2$ sukcesivne tačke $x_0, x_1, \ldots, x_{n+1}$ sa segmenta $[a,b]$ i odrede se koeficijenti polinoma P_n i veličina E tako da je

$$f(x_k) - P_n(x_k) = (-1)^k E \qquad (k = 0,1,\ldots,n+1). \tag{3.5.4}$$

2° Na $[a,b]$ se odredi skup od $n+2$ tačke $\hat{x}_0, \hat{x}_1, \ldots, \hat{x}_{n+1}$ u kojima $\delta_n(x) = f(x) - P_n(x)$ ima sukcesivne lokalne ekstremume sa alternativnim znacima, uključujući u ovaj skup tačku u kojoj veličina $|\delta_n(x)|$ ima najveću vrednost na $[a,b]$.

3° Za unapred zadatu tačnost ε proveravaju se uslovi $|\hat{x}_k - x_k| < \varepsilon$ $(k = 0,1,\ldots,n+1)$. Ukoliko bar jedan od ovih uslova nije zadovoljen uzima se $x_k := \hat{x}_k$ $(k = 0,1,\ldots,n+1)$ i prelazi na 1°. U slučaju da su pomenuti uslovi ispunjeni, algoritam se završava i polinom P_n se uzima kao najbolja mini-max aproksimacija P_n^*.

S obzirom da se teorema 3.5.2 o Čebiševljevoj alternansi može formulisati i za neke opštije tipove aproksimacionih funkcija, kakve su, na primer, linearna aproksimaciona funkcija i racionalna aproksimaciona funkcija, Rémesov algoritam se može primeniti i u tim slučajevima. Primetimo da u slučaju kada imamo nelinearnu aproksimacionu funkciju, sistem jednačina (3.5.4) postaje nelinearan i obično se tada rešava metodom Newton-Kantorovića. Pri ovome se po metodu Newton-Kantorovića, u cilju skraćivanja numeričkog rada, najčešće izvodi samo nekoliko prvih koraka.

Vrlo često se za nalaženje mini-max aproksimacije, problem zamenjuje odgovarajućim diskretnim mini-max problemom. Naime, segment $[a,b]$ se zamenjuje diskretnim skupom tačaka $I_m = \{x_0, x_1, \ldots, x_m\}$, gde je $m \gg n$. Neka je aproksimaciona funkcija linearna, tj.

$$\Phi(x) = a_0\Phi_0(x) + a_1\Phi_1(x) + \ldots + a_n\Phi_n(x).$$

Mini-max problem

$$\min_{a_i}\left(\max_{0\le k\le m}\left|f(x_k) - \sum_{i=0}^{n} a_i\Phi_i(x_k)\right|\right) = \max_{0\le k\le m}\left|f(x_k) - \sum_{i=0}^{n} a_i\Phi_i(x_k)\right|$$

može se zameniti ekvivalentnim problemom:

Naći $\min w$

pod ograničenjima

$$\left.\begin{aligned} w + \sum_{i=0}^{n} a_i \Phi_i(x_k) &\geq f(x_k) \\ w - \sum_{i=0}^{n} a_i \Phi_i(x_k) &\geq -f(x_k) \end{aligned}\right\} \quad (k = 0,1,\ldots,m).$$

Dobijeni problem linearnog programiranja, može se rešiti primenom simplex metoda na dualni problem([7]):

Naći $\max\left(\sum_{k=0}^{m} (s_k - t_k) f(x_k)\right)$

pod ograničenjima

$$\sum_{k=0}^{m} (s_k + t_k) \leqq 1,$$

$$\sum_{k=0}^{m} (s_k - t_k) \Phi_i(x_k) = 0 \qquad (i = 0,1,\ldots,n),$$

$$s_k \geqq 0, \quad t_k \geqq 0 \qquad (k = 0,1,\ldots,m).$$

O nekim algoritamskim modifikacijama za rešavanje poslednjeg problema može se naći u [8], dok se o aproksimaciji pomoću racionalnih funkcija može se naći u radu [9] .

6.4. LITERATURA

1. S.KARLIN and W.J.STUDDEN: Tchebysheff Systems with Applications in Analysis and Statistics. New York, 1966.
2. D.S.MITRINOVIĆ: Uvod u specijalne funkcije. Beograd, 1975.
3. B.P.DEMIDOVICH, I.A.MARON: Computational Mathematics. Noscow, 1973.
4. R.DE PRONY: Essai expérimentale et analytique... J.Ec.Polytech.Paris 1(2)(1795), 24-76.
5. D. MITROVIĆ: Algebarska metoda za analizu i sintezu sistema automatskog upravljanja. Beograd, 1966.
6. J.TODD:Introduction to the Constructive Theory of Functions.New York,1963.
7. L.M.DELVES and J.WALSH: Numerical Solution of integral equations. Oxford, 1974.
8. I.BARRODALE and C.PHILLIPS: Solution of an overdetermined system of linear equations in Chebyshev norm. Preprint.Dept.of Computational and Statistical Science. University of Liverpoll, 1973.
9. I.BARRODALE: Best rational approximation and strict quasi-convexity. SIAM J. Numer.Anal. 10(1973), 8-12.
10. D.C.HANDSCOMB: Methods of Numerical Approximation. Oxford, 1966.
11. F.B.HILDEBRAND: Introduction to Numerical Analysis. New York, 1974.
12. D.DJ.TOŠIĆ: Uvod u numeričku analizu. Beograd, 1978.
13. M.LAL, E.MOORE: An iterative technique for fitting exponentials. Int. J.Numer.Meth. Eng. 10(1976), 979-990.
14. D.W.KAMMLER: Prony's Method for Completely Monotonic Functions. J.Math. Anal.Appl. 57(1977), 560-570.
15. D.M.YOUNG, R.T.GREGORY: A Survey of Numerical Mathematics. Addison-Wesley Publ.Comp. 1973.
16. L.FOX and I.B.PARKER: Chebyshev Polynomials in Numerical Analysis. Oxford, 1972.

7. NUMERIČKO DIFERENCIRANJE I NUMERIČKA INTEGRACIJA

7.1. NUMERIČKO DIFERENCIRANJE

7.1.1. Osnovni pojmovi

U ovom poglavlju razmatramo problem numeričkog diferenciranja funkcije $f: [a,b] \to R$, za koju pretpostavljamo da je dovoljan broj puta diferencijabilna.

Potreba za numeričkim diferenciranjem javlja se u sledećim slučajevima:

1° Kada su vrednosti funkcije poznate samo na diskretnom skupu tačaka iz $[a,b]$, tj. kada je funkcija f data tabelarno;

2° Kada je analitički izraz za f dosta komplikovan.

Numeričko diferenciranje zasniva se na aproksimaciji funkcije f funkcijom Φ na $[a,b]$, a zatim se Φ diferencira odredjeni broj puta. Dakle, na osnovu

$$f(x) \sim \Phi(x) \qquad (a \leq x \leq b),$$

imamo

$$f^{(k)}(x) \sim \Phi^{(k)}(x) \qquad (a \leq x \leq b \ ; \ k=1,2,\ldots).$$

Za funkciju Φ se najčešće uzimaju algebarski interpolacioni polinomi, s obzirom da se oni jednostavno diferenciraju.

Neka je Φ interpolacioni polinom n-tog stepena, tj.

$$\Phi(x) = P_n(x).$$

Ako je poznata greška $R_n(f;x)$ u jednakosti

$$f(x) = P_n(x) + R_n(f;x) \qquad (a \leq x \leq b), \tag{1.1.1}$$

moguće je oceniti grešku i u formuli za diferenciranje. Naime, iz (1.1.1) sleduje

$$f^{(k)}(x) = P_n^{(k)}(x) + R_n^{(k)}(f;x) \qquad (a \leq x \leq b).$$

Za red izvoda ima smisla uzeti $k<n$.

Napomenimo da numeričko diferenciranje ima manju tačnost od interpolacije.

7.1.2. Formule za numeričko diferenciranje

Neka su poznate vrednosti funkcije f na skupu ekvidistantnih tačaka $\{x_o, x_1, \ldots, x_m\} \subset [a,b]$, sa korakom h. Dakle, neka je

$$f_k \equiv f(x_k) = f(x_o + kh) \qquad (k = o,1,\ldots,m).$$

Na skupu $\{x_i, x_{i+1}, \ldots, x_{i+n}\}$ $(o \leq i \leq m-n)$ konstruišimo prvi Newtonov interpolacioni polinom

$$P_n(x) = f_i + p\Delta f_i + \frac{p(p-1)}{2!}\Delta^2 f_i + \frac{p(p-1)(p-2)}{3!}\Delta^3 f_i$$
$$+ \cdots + \frac{p(p-1)\ldots(p-n+1)}{n!}\Delta^n f_i ,$$

tj.

$$P_n(x) = f_i + p\Delta f_i + \frac{p^2-p}{2}\Delta^2 f_i + \frac{p^3-3p^2+2p}{6}\Delta^3 f_i$$

(1.2.1)
$$+ \cdots + \frac{p^n - \frac{n(n-1)}{2}p^{n-1} + \cdots}{n!}\Delta^n f_i ,$$

gde je $p = \dfrac{x-x_i}{h}$.

S obzirom da je $P_n(x) = \frac{1}{h}\cdot\dfrac{dP_n(x)}{dp}$, diferenciranjem jednakosti (1.2.1) dobijamo

(1.2.2) $$P'_n(x) = \frac{1}{h}\left(\Delta f_i + \frac{2p-1}{2}\Delta^2 f_i + \frac{3p^2-6p+2}{6}\Delta^3 f_i + \cdots\right).$$

Daljim diferenciranjem (1.2.2), dobijamo redom P_n'', P_n''', itd. Na primer,

$$P_n''(x) = \frac{1}{h^2}\left(\Delta^2 f_i + (p-1)\Delta^3 f_i + \cdots\right). \tag{1.2.3}$$

Umesto prvog Newtonovog interpolacionog polinoma mogu se koristiti i ostali interpolacioni polinomi. Na primer, diferenciranjem Stirlingove formule

$$P_n(x) = f_i + \binom{p}{1}\delta\mu f_i + \frac{p}{2}\binom{p}{1}\delta^2 f_i + \binom{p+1}{3}\delta^3\mu f_i$$

$$+ \frac{p}{4}\binom{p+1}{3}\delta^4 f_i + \cdots ,$$

gde je $x = x_i + ph$, dobijamo

$$P_n'(x) = \frac{1}{h}\left(\delta\mu f_i + p\delta^2 f_i + \frac{3p^2-1}{6}\delta^3\mu f_i + \frac{2p^3-p}{12}\delta^4 f_i + \cdots\right),$$

$$P_n''(x) = \frac{1}{h^2}\left(\delta^2 f_i + p\delta^3\mu f_i + \frac{6p^2-1}{12}\delta^4 f_i + \cdots\right),$$

$$P_n'''(x) = \frac{1}{h^3}\left(\delta^3\mu f_i + p\delta^4 f_i + \cdots\right),$$

itd.

Za $x = x_i$ ($\Rightarrow p=0$) formule (1.2.2) i (1.2.3) se svode na

$$P_n'(x_i) = \frac{1}{h}\left(\Delta f_i - \frac{1}{2}\Delta^2 f_i + \frac{1}{3}\Delta^3 f_i - \cdots + \frac{(-1)^{n-1}}{n}\Delta^n f_i\right)$$

i

$$P_n''(x_i) = \frac{1}{h^2}\left(\Delta^2 f_i - \Delta^3 f_i + \frac{11}{12}\Delta^4 f_i - \cdots\right).$$

Slično, iz formula dobijenih na osnovu Stirlingove formule sleduje

$$P_n'(x_i) = \frac{1}{h}\left(\delta\mu f_i - \frac{1}{6}\delta^3\mu f_i + \cdots\right),$$

$$P_n''(x_i) = \frac{1}{h^2}\left(\delta^2 f_i - \frac{1}{12}\delta^4 f_i + \cdots\right),$$

itd.

Do formula za izvod funkcije u interpolacionim čvorovima može se doći formalno primenom operatorskog računa.

Kako je $e^{hD} = E = 1 + \Delta$, imamo

$$D = \frac{1}{h}\log(1+\Delta) = \frac{1}{h}\left(\Delta - \frac{1}{2}\Delta^2 + \frac{1}{3}\Delta^3 - \frac{1}{4}\Delta^4 + \cdots\right)$$

i

$$D^k = \frac{1}{h^k}\left\{\log(1+\Delta)\right\}^k .$$

Za dobijanje razvoja operatora D^k po stepenima od Δ, koristićemo se Stirlingovim brojevima prve vrste, koji se definišu na sledeći način:

<u>Definicija 1.2.1</u>. Koeficijenti $S_n^{(k)}$ u razvoju

$$\frac{(\log(1+x))^k}{k!} = \sum_{n=k}^{+\infty} \frac{S_n^{(k)}}{n!} x^n,$$

koji važi za dovoljno malo x, nazivaju se Stirlingovi brojevi prve vrste.

Umesto prethodne definicije može se uzeti i sledeća njoj ekvivalentna definicija.

<u>Definicija 1.2.2</u>. Koeficijenti $S_n^{(k)}$ u razvoju uopštenog stepena po algebarskim stepenima, tj.

$$x^{(n)} = x(x-1)\ldots(x-n+1) = \sum_{k=1}^{n} S_n^{(k)} x^k ,$$

nazivaju se Stirlingovi brojevi prve vrste.

<u>Primer1.2.1</u>. Dokazaćemo da za Stirlingove brojeve $S_n^{(k)}$ važi rekurentna relacija

$$(1.2.4) \qquad S_{n+1}^{(k)} = S_n^{(k-1)} - nS_n^{(k)} \qquad (k = 2,3,\ldots).$$

Kako je $x^{(n+1)} = x^{(n)}(x-n)$, zaista, iz

$$\sum_{k=1}^{n+1} S_{n+1}^{(k)} x^k = (x-n) \sum_{k=1}^{n} S_n^{(k)} x^k$$

$$= -nS_n^{(1)} x + \sum_{k=2}^{n} \left(S_n^{(k-1)} - nS_n^{(k)}\right) x^k + S_n^{(n)} x^{n+1}$$

sleduje (1.2.4), kao i jednakosti

$$S_{n+1}^{(1)} = -n\, S_n^{(1)} \quad \text{i} \quad S_{n+1}^{(n+1)} = S_n^{(n)} .$$

U sledećoj tabeli dati su Stirlingovi brojevi prve vrste $S_n^{(k)}$ za $n \leqq 7$.

n \ k	1	2	3	4	5	6	7
1	1						
2	-1	1					
3	2	-3	1				
4	-6	11	-6	1			
5	24	-50	35	-10	1		
6	-120	274	-225	85	-15	1	
7	720	-1764	1624	-735	175	-21	1

Na osnovu definicije 1.2.1. imamo

$$D^k = \frac{1}{h^k}\left\{\frac{S_k^{(k)}}{1}\Delta^k + \frac{S_{k+1}^{(k)}}{k+1}\Delta^{k+1} + \frac{S_{k+2}^{(k)}}{(k+1)(k+2)}\Delta^{k+2} + \cdots\right\},$$

tj.

$$D^k = \frac{1}{h^k}\left\{\Delta^k - \frac{k}{2}\Delta^{k+1} + \frac{k(3k+5)}{24}\Delta^{k+2} - \frac{k(k+2)(k+3)}{48}\Delta^{k+3} + \cdots\right\}.$$

Dakle, formula za k-ti izvod u tački x_i može se predstaviti u obliku

$$f^{(k)}(x_i) = D^k f_i = \frac{1}{h^k}\left(\Delta^k f_i - \frac{k}{2}\Delta^{k+1}f_i + \frac{k(3k+5)}{24}\Delta^{k+2}f_i - \cdots\right).$$

Na primer, za $k=1$, imamo

$$(1.2.5) \quad f'(x_i) = \frac{1}{h}\left(\Delta f_i - \frac{1}{2}\Delta^2 f_i + \frac{1}{3}\Delta^3 f_i - \cdots + \frac{(-1)^{n-1}}{n}\Delta^n f_i\right) + R_n'(f;x_i).$$

Ako $f \in C^{n+1}[a,b]$, greška u poslednjoj formuli može se predstaviti u obliku

$$(1.2.6) \qquad R_n'(f;x_i) = \frac{(-1)^n}{n+1} h^n f^{(n+1)}(\xi) \qquad (\xi \in (x_i, x_{n+i})).$$

Primetimo da je kod interpolacionog polinoma $R_n(f;x_i)=0$.

Primer 1.2.2. Na osnovu vrednosti funkcije $x \mapsto f(x) = x^3 - 2x - 5$ u tačkama $x_i = i\,(i=0,1,2,3,4)$ odredićemo prva tri izvoda ove funkcije u tački $x=0$.

Formirajmo najpre tablicu Δ razlika uzimajući $h=1$:

x_i	f_i	Δf_i	$\Delta^2 f_i$	$\Delta^3 f_i$	$\Delta^4 f_i$
0	-5				
		-1			
			6		
1	-6			6	
		5			
			12		0
2	-1			6	
		17			
			18		
3	16				
		35			
4	51				

Tada je

$$f'(0) = \Delta f_0 - \frac{1}{2}\Delta^2 f_0 + \frac{1}{3}\Delta^3 f_0 - \frac{1}{4}\Delta^4 f_0 = -2,$$

$$f''(0) = \Delta^2 f_0 - \Delta^3 f_0 + \frac{11}{12}\Delta^4 f_0 = 0, \quad f'''(0) = \Delta^3 f_0 - \frac{3}{2}\Delta^4 f_i = 6.$$

Na osnovu (1.2.5) i (1.2.6), za $n=1,2,3$ dobijamo respektivno

$$f'(x_i) = \frac{1}{h}(f_{i+1} - f_i) - \frac{h}{2} f''(\xi_1) \qquad (x_i < \xi_1 < x_{i+1}),$$

$$f'(x_i) = \frac{1}{2h}(-3f_i + 4f_{i+1} - f_{i+2}) + \frac{h^2}{3} f'''(\xi_2) \qquad (x_i < \xi_2 < x_{i+2}),$$

$$f'(x_i) = \frac{1}{6h}(-11f_i + 18f_{i+1} - 9f_{i+2} + 2f_{i+3})$$

$$- \frac{h^3}{4} f^{IV}(\xi_3) \qquad (x_i < \xi_3 < x_{i+3}).$$

Predjimo sada na nalaženje formula za numeričko diferenciranje u interpolacionim čvorovima primenom operatora centralne razlike.

Kako je

$$D = \frac{2}{h} \operatorname{arsh} \frac{\delta}{2} = \frac{2}{h} \log\left[\frac{1}{2}\delta + (1 + \frac{1}{4}\delta^2)^{1/2}\right],$$

tj.

$$D = \frac{1}{h}\left(\delta - \frac{1^2}{2^2 3!}\delta^3 + \frac{1^2 3^2}{2^4 5!}\delta^5 - \frac{1^2 3^2 5^2}{2^6 7!}\delta^7 + \cdots\right), \tag{1.2.7}$$

odredjivanje Df_i primenom poslednje formule, nije moguće ako ne raspolažemo sa vrednostima funkcije u središnim tačkama izmedju interpolacionih čvorova, s obzirom da je

$$\delta^{2j+1} f_i = \delta^{2j}(f_{i+(1/2)} - f_{i-(1/2)}).$$

Medjutim, formula (1.2.7) može se uspešno primeniti na odredjivanje $Df_{i+(1/2)}$, s obzirom da je

$$\delta^{2j+1} f_{i+(1/2)} = \delta^{2j}(f_{i+1} - f_i).$$

Da bismo odredili Df_i samo na osnovu vrednosti funkcije u interpolacionim čvorovima, formulu (1.2.7) treba modifikovati.

Naime, kako je $\mu = \left(1 + \frac{1}{4}\delta^2\right)^{1/2}$, imamo

$$D = \frac{2\mu}{h}\left(1 + \frac{1}{4}\delta^2\right)^{-1/2} \operatorname{arsh}\frac{\delta}{2},$$

tj.

$$D = \frac{\mu}{h}\left(\delta - \frac{1^2}{3!}\delta^3 + \frac{1^2 \cdot 2^2}{5!}\delta^5 - \frac{1^2 \cdot 2^2 \cdot 3^2}{7!}\delta^7 + \cdots\right). \tag{1.2.8}$$

Na osnovu poslednje formule dobija se niz formula za odredjivanje Df_i. Navešćemo samo prve dve:

1° $$Df_i = \frac{1}{h}\mu\delta f_i + r_1(f) = \frac{1}{2h}(f_{i+1} - f_{i-1}) + r_1(f),$$

gde je

$$r_1(f) = -\frac{h^2}{6} f'''(\xi_1) \qquad (x_{i-1} < \xi_1 < x_{i+1});$$

2° $$Df_i = \frac{1}{h}(\mu\delta f_i - \frac{1}{6}\mu\delta^3 f_i) + r_2(f) = \frac{1}{12h}(-f_{i+2} + 8f_{i+1} - 8f_{i-1} + f_{i-2}) + r_2(f),$$

gde je

$$r_2(f) = \frac{h^4}{30} f^{V}(\xi_2) \qquad (x_{i-2} < \xi_2 < x_{i+2}).$$

Za izračunavanje viših izvoda koristi se formula (1.2.7) za izvode parnog reda i formula (1.2.8) za izvode neparnog reda. Tako dobijamo

$$D^2 = \frac{1}{h^2}\left(\delta^2 - \frac{1}{12}\delta^4 + \frac{1}{90}\delta^6 - \frac{1}{560}\delta^8 + \frac{1}{3150}\delta^{10} - \cdots\right),$$

$$D^3 = \frac{\mu}{h^3}\left(\delta^3 - \frac{1}{4}\delta^5 + \frac{7}{120}\delta^7 - \cdots\right),$$

$$D^4 = \frac{1}{h^4}\left(\delta^4 - \frac{1}{6}\delta^6 + \frac{7}{240}\delta^8 - \cdots\right),$$

$$D^5 = \frac{\mu}{h^5}\left(\delta^5 - \frac{1}{3}\delta^7 + \cdots\right),$$

itd.

Na kraju, daćemo najprostiju formulu za aproksimaciju drugog izvoda, koja se u praksi najčešće koristi. Ovu formulu dobijamo uzimajući samo prvi član u razvoju za D^2, tj.

$$D^2 f_i = \frac{1}{h^2}\,\delta^2 f_i + r(f) = \frac{1}{h^2}(f_{i+1} - 2f_i + f_{i-1}) + r(f),$$

gde je $r(f) = -\frac{h^2}{12} f^{IV}(\xi) \quad (x_{i-1} < \xi < x_{i+1})$.

7.2. NUMERIČKA INTEGRACIJA

7.2.1. Uvodne napomene

Numerička integracija funkcija sastoji se u približnom izračunavanju odredjenih integrala na osnovu niza vrednosti podintegralne funkcije po odredjenoj formuli.

Formule za numeričko izračunavanje jednostrukih integrala nazivaju se kvadraturne formule. Slično, formule za dvostruke integrale nazivaju se kubaturne formule. U našem izlaganju zadržaćemo se samo na kvadraturnim formulama.

Potreba za numeričkom integracijom javlja se u velikom broju slučajeva. Naime, Newton-Leibnitzova formula

$$\int_a^b f(x)dx = F(b) - F(a), \tag{2.1.1}$$

gde je F primitivna funkcija za funkciju f, ne može se uvek uspešno primeniti*. Navešćemo neke od tih slučajeva:

1° Funkcija F se ne može predstaviti pomoću konačnog broja elementarnih funkcija(na primer, kada je $f(x) = e^{-x^2}$).

2° Primena formule (2.1.1) često dovodi do vrlo složenog izraza, čak i kod izračunavanja integrala jednostavnijih funkcija; na primer

$$\int_0^a \frac{dx}{1+x^3} = \log \sqrt[3]{|a+1|} - \frac{1}{6}\log(a^2-a+1) + \frac{1}{\sqrt{3}} \operatorname{arctg} \frac{a\sqrt{3}}{2-a} .$$

*) Ako nije drugačije naglašeno smatraćemo da je funkcija f neprekidna na [a,b].

3° Kod integracije funkcija, čije su vrednosti poznate na diskretnom skupu tačaka (dobijene, na primer, eksperimentalno), nije mogućno primeniti formulu (2.1.1).

Veliki broj kvadraturnih formula ima oblik

$$\int_a^b f(x)dx \cong \sum_{k=0}^{n} A_k f_k \,, \tag{2.1.2}$$

gde je $f_k = f(x_k)$ $(a \leqq x_0 < \cdots < x_n \leqq b)$. Ako je $x_0 = a$ i $x_n = b$, za formulu (2.1.2) kažemo da je zatvorenog tipa, dok u ostalim slučajevima kažemo da je otvorenog tipa. Napomenimo da se za integraciju diferencijabilnih funkcija koriste i formule u kojima se pored vrednosti funkcije pojavljuju i vrednosti izvoda.

Od interesa su i formule za izračunavanje integrala

$$\int_a^b p(x)f(x)dx,$$

gde je $x \mapsto p(x)$ data težinska funkcija.

Jedan prost način za konstrukciju kvadraturnih formula zasniva se na primeni interpolacije. Formule dobijene na ovaj način nazivaju se kvadraturne formule interpolacionog tipa.

Neka su vrednosti funkcije f u datim tačkama $x_0, x_1, \ldots, x_n$ $(\in [a,b])$ redom $f_0, f_1, \ldots, f_n$, tj.

$$f_k \equiv f(x_k) \qquad (k=0,1,\ldots,n).$$

Na osnovu ovih podataka, možemo konstruisati Lagrangeov interpolacioni polinom

$$P_n(x) = \sum_{k=0}^{n} f_k \frac{\omega(x)}{(x-x_k)\,\omega'(x_k)} \,,$$

gde je $\omega(x) = (x-x_0)(x-x_1)\cdots(x-x_n)$.

Tada je

$$\int_a^b p(x)f(x)dx = \int_a^b p(x)P_n(x)dx + R_{n+1}(f),$$

tj.

$$\int_a^b p(x)f(x)dx = \sum_{k=0}^{n} A_k f_k + R_{n+1}(f), \tag{2.1.3}$$

gde smo stavili

$$A_k = \int_a^b \frac{p(x)\omega(x)}{(x-x_k)\,\omega'(x_k)}\,dx \qquad (k=0,1,\ldots,n).$$

U formuli (2.1.3), veličina $R_{n+1}(f)$ naziva se ostatak kvadraturne formule i predstavlja grešku koja se čini zamenom integrala konačnom sumom. Indeks n+1 u ostatku označava da se integral približno izračunava na osnovu vrednosti podintegralne funkcije u n+1 tačaka.

Sa $\mathcal{P}_n$ označimo skup svih polinoma ne višeg stepena od n. Kako je za $f(x)= x^k(k=0,1,\ldots,n)$, $f(x)\equiv P_n(x)$, imamo

$$R_{n+1}(x^k) \equiv 0 \qquad (k=0,1,\ldots,n),$$

odakle zaključujemo da je formula (2.1.3) tačna za svako $f\in \mathcal{P}_n$, bez obzira na izbor interpolacionih čvorova $x_k(k=0,1,\ldots,n)$ i u ovom slučaju kažemo da formula (2.1.3) ima algebarski stepen tačnosti n.

Kao što ćemo videti u odeljku o Gaussovim kvadraturnim formulama, specijalnim izborom interpolacionih čvorova, algebarski stepen tačnosti kvadraturne formule (2.1.3) može se povećati najviše na 2n+1. Ovo tvrdjenje je logično, s obzirom da imamo na raspolaganju 2n+2 parametra $(A_k,x_k(k=0,1,\ldots,n))$. Ovi parametri mogu se odrediti rešavanjem sistema nelinearnih jednačina

$$\sum_{k=0}^{n} A_k x_k^m = \int_a^b p(x)x^m dx \qquad (m=0,1,\ldots,2n+1).$$

Naravno, ovakav način je vrlo komplikovan, čak i za relativno malo n.

Ukoliko su interpolacioni čvorovi unapred fiksirani, težinski koeficijenti $A_k (k=0,1,\ldots,n)$ mogu se odrediti rešavanjem sistema linearnih jednačina

$$\sum_{k=0}^{n} A_k x_k^m = \int_a^b p(x)x^m dx \qquad (m=0,1,\ldots,n).$$

Ovaj metod se naziva metod neodredjenih koeficijenata.

Primer 2.1.1. Metodom neodredjenih koeficijenata nadjimo kvadraturnu formulu interpolacionog tipa

$$\int_0^{2h} f(x)dx = h(A_0 f_0 + A_1 f_1 + A_2 f_2), \tag{2.1.4}$$

gde je $f_k = f(kh)$ $(k=0,1,2)$. Ako stavimo $f(x)=1,x,x^2$ u (2.1.4) dobijamo

$$h(A_0 + A_1 + A_2) = 2h,$$

$$h(hA_1 + 2hA_2) = 2h^2,$$

$$h(h^2A_1 + 4h^2A_2) = \frac{8}{3}h^3,$$

odakle sleduje $A_1 = \frac{4}{3}$ i $A_0 = A_2 = \frac{1}{3}$. Dakle, formula (2.1.4) postaje

$$\int_0^{2h} f(x)dx = \frac{h}{3}(f_0 + 4f_1 + f_2).$$

Ova formula je dobro poznata iz standardnih kurseva matematičke analize i naziva se Simpsonova formula.

7.2.2. Newton-Cotesove formule

U ovom odeljku izvešćemo kvadraturne formule zatvorenog tipa u kojima su interpolacioni čvorovi $x_k = x_0 + kh (k=0,1,\ldots,n)$

uzeti ekvidistantno sa korakom $h=\frac{b-a}{n}$.

Ako uvedemo smenu $x-x_0=ph$, imamo

(2.2.1) $$\omega(x) = (x-x_0)(x-x_1)\ldots(x-x_n) = h^{n+1}p(p-1)\ldots(p-n)$$

i

(2.2.2) $$\omega'(x_k) = (x_k-x_0)\ldots(x_k-x_{k-1})(x_k-x_{k+1})\ldots(x_k-x_n) = h^n(-1)^{n-k}k!(n-k)!\,.$$

Uvodjenjem oznake za uopšteni stepen $x^{(s)}=x(x-1)\ldots(x-s+1)$ na osnovu (2.2.1), (2.2.2) i rezultata iz prethodnog odeljka dobijamo

$$A_k = \int_0^n \frac{(-1)^{n-k}p^{(n+1)}h}{(p-k)!k!(n-k)!}\,dp \qquad (k=0,1,\ldots,n),$$

tj.

$$A_k = (b-a)H_k \qquad (k=0,1,\ldots,n),$$

gde smo stavili

(2.2.3) $$H_k \equiv H_k(n) = \frac{(-1)^{n-k}}{n!\,n}\binom{n}{k}\int_0^n \frac{p^{(n+1)}}{p-k}\,dp \quad (k=0,1,\ldots,n).$$

Koeficijenti H_k u literaturi (videti na primer [1] i [2]) poznati su kao **Newton-Cotesovi** koeficijenti, a odgovarajuće formule

(2.2.4) $$\int_{x_0=a}^{x_n=b} f(x)\,dx \cong (b-a)\sum_{k=0}^{n} H_k f\left(a+k\,\frac{b-a}{n}\right) \qquad (k\in N)$$

kao Newton-Cotesove formule.

<u>Primedba 2.2.1</u>. Za koeficijente H_k važe jednakosti

$$H_k = H_{n-k} \quad \text{i} \quad \sum_{k=0}^{n} H_k = 1.$$

Razmotrićemo sada dva partikularna slučaja.

a) Neka je $n=1$. Tada je, na osnovu (2.2.3), $H_o = H_1 = \frac{1}{2}$ pa je odgovarajuća Newton-Cotesova formula

$$\int_a^b f(x)dx = \frac{b-a}{2}(f(a)+f(b)) + R_2(f). \tag{2.2.5}$$

Ova formula je poznata kao trapezna formula ili trapezno pravilo.

<u>Teorema 2.2.1</u>. Ako $f \in C^2[a,b]$, za ostatak $R_2(f)$, u formuli (2.2.5), važi

$$R_2(f) = -\frac{(b-a)^3}{12} f''(\xi_1) \qquad (a < \xi_1 < b). \tag{2.2.6}$$

<u>Dokaz</u>. Definišimo funkciju $F: [o, b-a] \to R$, pomoću

$$F(h) = \int_a^{a+h} f(x)dx - \frac{h}{2}(f(a)+f(a+h)).$$

Diferenciranjem ove funkcije sukcesivno dva puta dobijamo

$$F'(h) = \frac{1}{2}(f(a+h)-f(a)) - \frac{h}{2}f'(a+h)$$

i

$$F''(h) = -\frac{h}{2} f''(a+h). \tag{2.2.7}$$

Kako je $F(0)=0$ i $F'(0)=0$, integracijom jednačine (2.2.7) dobijamo

$$F(h) = -\frac{1}{2}\int_0^h t(h-t)f''(a+t)dt.$$

S obzirom da $f \in C^2[a,b]$, primenom teoreme o srednjoj vrednosti odredjenog integrala, poslednja jednakost se svodi na

$$F(h) = -\frac{1}{2} f''(a+\theta h) \int_0^h t(h-t)dt = -\frac{h^3}{12} f''(a+\theta h), \tag{2.2.8}$$

gde je $0<\theta<1$. Najzad, ako stavimo $h=b-a$, iz (2.2.8) sleduje

$$R_2(f) = F(b-a) = -\frac{(b-a)^3}{12} f''(\xi_1) \qquad (a<\xi_1<b),$$

čime je dokaz završen.

Promenom uslova za f mogućno je dobiti i druge ocene za $R_2(f)$. Navešćemo bez dokaza neke od tih rezultata([3],[4],[5],[6]).

Teorema 2.2.2. Ako funkcija zadovoljava Lipschitzov uslov, tj.

$$|f(x) - f(y)| \leqq M|x - y| \qquad (\forall x,y\in[a,b]),$$

tada je

$$|R_2(f)| \leqq \frac{M(b-a)^2}{4} - \frac{1}{4M}(f(b) - f(a))^2.$$

Teorema 2.2.3. Ako funkcija f zadovoljava uslove $|f'(x)-f'(y)|\leqq M|x-y|$ $(\forall x,y\in[a,b])$ i $f'(a) = f'(b) = 0$, tada je

$$|R_2(f)| \leqq \frac{M(b-a)^3}{24} - \frac{1}{2M}\cdot\frac{(f(b)-f(a))^2}{b-a}.$$

b) Pretpostavimo sada da je n=2. Tada na osnovu (2.2.3) i (2.2.4), dobijamo kvadraturnu formulu

$$\int_a^b f(x)dx = \frac{b-a}{6}\left(f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right) + R_3(f), \tag{2.2.9}$$

koja je poznata kao Simpsonova formula ili pravilo(videti primer 2.2.1.).

Teorema 2.2.4. Ako $f\in C^4[a,b]$, za ostatak $R_3(f)$, u formuli (2.2.9), važi

$$R_3(f) = -\frac{(b-a)^5}{2880} f^{IV}(\xi_2) \qquad (a<\xi_2<b). \tag{2.2.10}$$

Dokaz. Slično, kao u dokazu teoreme 2.2.1, definišimo funkciju $F:[-\alpha,\alpha]\to R$, pomoću

$$F(h) = \int_{c-h}^{c+h} f(x)dx - \frac{h}{3}(f(c-h) + 4f(c) + f(c+h)),$$

gde su $c=\frac{a+b}{2}$ i $\alpha=\frac{b-a}{2}$.

Uzastopnim diferenciranjem funkcije F tri puta dobijamo

$$F'(h) = \frac{2}{3}(f(c-h) - 2f(c) + f(c+h)) - \frac{h}{3}(f'(c+h) - f'(c-h)),$$

$$F''(h) = -\frac{1}{3}(f'(c+h) - f'(c-h)) - \frac{h}{3}(f''(c+h) + f''(c-h)),$$

$$F'''(h) = -\frac{h}{3}(f'''(c+h) - f'''(c-h)).$$

Kako je $F(0)=0$, $F'(0)=0$, $F''(0)=0$, rešenje poslednje jednačine možemo predstaviti u obliku

$$F(h) = \int_0^h \frac{1}{2}(h-t)^2\left(-\frac{t}{3}\right)(f'''(c+t) - f'''(c-t))dt,$$

odakle, korišćenjem Lagrangeove teoreme, dobijamo

$$F(h) = -\frac{1}{3}\int_0^h t^2(h-t)^2 f^{IV}(c+\theta(t)t)dt \qquad (|\theta(t)|<1).$$

Najzad, primenom teoreme o srednjoj vrednosti odredjenih integrala($f\in C^4[a,b]$) i stavljanjem $h=\alpha=\frac{b-a}{2}$, iz poslednje jednakosti sleduje (2.2.10), čime je dokaz završen.

Na osnovu prethodnih razmatranja zaključujemo da je trapezna formula tačna za svako $f\in\mathcal{P}_1$, a da je Simpsonova formula tačna za svako $f\in\mathcal{P}_3$.

U daljem izlaganju daćemo pregled Newton-Cotesovih formula za $n\leqq 8$(videti [7],[8],[9]). Pri ovome koristimo oznake $h = \frac{b-a}{n}$, $f_k = f(x_k)$ $(k=0,1,\ldots,n)$.

1) $n=1$ (trapezno pravilo)

$$\int_{x_0}^{x_1} f(x)\,dx = \frac{h}{2}(f_0 + f_1) - \frac{h^3}{12} f''(\xi_1) \ ;$$

2) $n=2$ (Simpsonovo pravilo)

$$\int_{x_0}^{x_2} f(x)\,dx = \frac{h}{3}(f_0 + 4f_1 + f_2) - \frac{h^5}{90} f^{IV}(\xi_2) \ ;$$

3) $n=3$ (Simpsonovo pravilo $\frac{3}{8}$)

$$\int_{x_0}^{x_3} f(x)\,dx = \frac{3h}{8}(f_0 + 3f_1 + 3f_2 + f_3) - \frac{3h^5}{80} f^{IV}(\xi_3) \ ;$$

4) $n=4$ (Booleovo pravilo)

$$\int_{x_0}^{x_4} f(x)\,dx = \frac{2h}{45}(7f_0 + 32f_1 + 12f_2 + 32f_3 + 7f_4) - \frac{8h^7}{945} f^{(6)}(\xi_4) ;$$

5) $n=5$

$$\int_{x_0}^{x_5} f(x)\,dx = \frac{5h}{288}(19f_0 + 75f_1 + 50f_2 + 50f_3 + 75f_4 + 19f_5) - \frac{275h^7}{12096} f^{(6)}(\xi_5) \ ;$$

6) $n=6$

$$\int_{x_0}^{x_6} f(x)\,dx = \frac{h}{140}(41f_0 + 216f_1 + 27f_2 + 272f_3 + 27f_4 + 216f_5 + 41f_6) - \frac{9h^9}{1400} f^{(8)}(\xi_6) \ ;$$

7) $n=7$

$$\int_{x_0}^{x_7} f(x)\,dx = \frac{7h}{17280}(751f_0 + 3577f_1 + 1323f_2 + 2989f_3 + 2989f_4 + 1323f_5 + 3577f_6 + 751f_7) - \frac{8183h^9}{518400} f^{(8)}(\xi_7) \ ;$$

7) n=8

$$\int_{x_0}^{x_8} f(x)\,dx = \frac{4h}{14175}(989f_0 + 5888f_1 - 928f_2 + 10496f_3 - 4540f_4$$

$$+ 10496f_5 - 928f_6 + 5888f_7 + 989f_8) - \frac{2368h^{11}}{467775} f^{(10)}(\xi_8),$$

gde $\xi_k \in (x_0, x_k)$ $(k=1,\ldots,8)$.

U opštem slučaju ostatak $R_{n+1}(f)$ ima oblik(videti na primer [2], [9], [10])

$$(2.2.11) \qquad R_{n+1}(f) = C_n h^m f^{(m-1)}(\xi_n) \qquad (x_0 < \xi_n < x_n),$$

gde je $m = 2\left[\frac{n}{2}\right] + 3$. Jednakost (2.2.11) ima smisla ukoliko funkcija $f \in C^{m-1}[a,b]$.

Primedba 2.2.2. Za koeficijente $H_k(n)$ u Newton-Cotesovoj formuli za segment $[0,1]$, važe asimptotske formule(videti [9], [11])

$$H_k(n) = \frac{(-1)^{k-1}(n-1)!}{k!(n-k)!\,n\,\log^2 n}\left(\frac{1}{k} + \frac{(-1)^n}{n-k}\right)\left(1 + O\left(\frac{1}{\log n}\right)\right) \quad (1 \leq k \leq n-1)$$

$$H_0(n) = H_n(n) = \frac{1}{n^2 \log n}\left(1 + O\left(\frac{1}{\log n}\right)\right).$$

Na osnovu prethodnog uočavamo da se formule znatno komplikuju ukoliko je n veće. U praktičnim računanjima obično se koristi Simpsonova formula(n=2), s obzirom da je relativno jednostavna, a da pri tome daje dosta zadovoljavajuću tačnost ukoliko je h dovoljno malo. Uslov da h bude dovoljno malo može se obezbediti podelom segmenta $[a,b]$ na niz podsegmenata, o čemu će biti reči u narednom odeljku.

Primer 2.2.1. Primenimo Newton-Cotesove formule za n= 1,...,6 na približno izračunavanje integrala

$$(2.2.12) \qquad I = \int_0^1 \frac{dx}{1+x},$$

čija je tačna vrednost $\log 2 = 0.69314818\ldots$. Dobijene vrednosti sredjene su u narednoj tabeli.

n	Približna vrednost za I
1	0.75
2	0.6944444
3	0.69375
4	0.6931745
5	0.6931629
6	0.693148

Primedba 2.2.3. Ako Newton-Cotesovu formulu za $n=6$ predstavimo u obliku

$$\int_{x_0}^{x_6} f(x)\,dx = \frac{3h}{420}\Big\{(42-1)f_0 + (210+6)f_1 + (42-15)f_2 + (252+20)f_3 + (42-15)f_4 + (210+6)f_5 + (42-1)f_6\Big\} + R_7(f),$$

dobijamo tzv. Weddleovo pravilo

$$\int_{x_0}^{x_6} f(x)\,dx = \frac{3h}{10}(f_0 + 5f_1 + f_2 + 6f_3 + f_4 + 5f_5 + f_6) + \bar{R}_7(f), \tag{2.2.13}$$

gde je

$$\bar{R}_7(f) = -\frac{h}{140}\Delta^6 f_0 + R_7(f),$$

tj.

$$\bar{R}_7(f) = -\frac{h^7}{1400}(10f^{(6)}(\eta_6) + 9h^2 f^{(8)}(\xi_6)) \qquad (\xi_6, \eta_6 \in (x_0, x_6)).$$

Primer 2.2.2. Primenom formule (2.2.13), na integral (2.2.12), dobijamo $I = 0.6931493$, s obzirom da je

$$f_0 = 1.0000000, \quad f_1 = 0.8571428, \quad f_2 = 0.7500000,$$

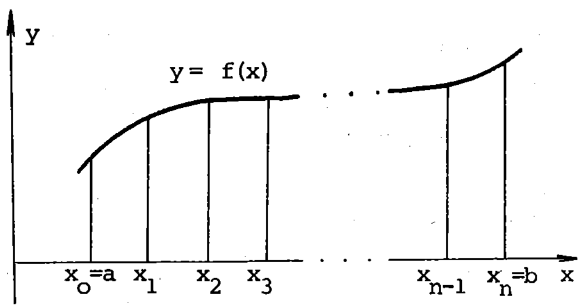
$$f_3 = 0.6666666, \quad f_4 = 0.6000000, \quad f_5 = 0.5454545,$$

$$f_6 = 0.5000000.$$

7.2.3. Uopštene kvadraturne formule

Kao što je napomenuto u prethodnom odeljku, da bismo tačnije izračunali vrednost integrala potrebno je podeliti segment $[a,b]$ na niz podsegmenata, a zatim na svakom od njih primeniti neku od kvadraturnih formula. Na taj način dobijamo uopštene ili kompozitne formule. U ovom odeljku razmotrićemo uopštene formule dobijene na bazi trapezne i Simpsonove formule, kao i neke uopštene kvadraturne formule otvorenog tipa.

Podelimo segment $[a,b]$ na niz podsegmenata $[x_{i-1},x_i]$ $(i=1,\ldots,n)$ tako da je $x_i = a+ih\,(i=0,1,\ldots,n)$ i $h=(b-a)/n$.



Sl. 2.3.1

Primenom trapezne formule na svaki od podsegmenata dobijamo

$$\int_a^b f(x)\,dx = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} f(x)\,dx = \sum_{i=1}^{n} \Big(\frac{h}{2}(f_{i-1}+f_i) - \frac{h^3}{12} f''(\xi_i)\Big),$$

tj.

$$(2.3.1) \qquad \int_a^b f(x)\,dx = T_n - \frac{h^3}{12}\sum_{i=1}^{n} f''(\xi_i),$$

gde su $T_n \equiv T_n(f;h) = h\big(\frac{1}{2}f_o + f_1 + \cdots + f_{n-1} + \frac{1}{2}f_n\big)$ i

$x_{i-1} < \xi_i < x_i \quad (i=1,\ldots,n).$

Teorema 2.3.1. Ako $f \in C^2[a,b]$ važi jednakost

$$\int_a^b f(x)dx - T_n = -\frac{(b-a)^3}{12n^2} f''(\xi) \qquad (a<\xi<b). \tag{2.3.2}$$

Dokaz. Kako uslov $f\in C^2[a,b]$ obezbedjuje egzistenciju takvog ξ $(a<\xi<b)$ da je

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

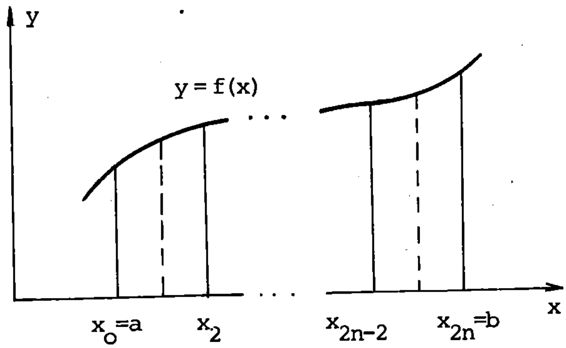
iz (2.3.1) neposredno sleduje jednakost (2.3.2).

Kvadraturna formula

$$\int_a^b f(x)dx \cong T_n(f;h) \qquad \left(h = \frac{b-a}{n} \right)$$

naziva se uopštena trapezna formula.

Pretpostavimo sada da je $h=\frac{b-a}{2n}$, tj. $x_i=a+ih(i=0,1,\dots,2n)$ (videti sl. 2.3.2), a zatim na podsegmente $[x_0,x_2]$, ..., $[x_{2n-2},x_{2n}]$



Sl. 2.3.2

primenimo Simpsonovu formulu. Na taj način dobijamo uopštenu Simpsonovu formulu

$$\int_a^b f(x)dx \cong S_n(f;h) \qquad \left(h=\frac{b-a}{2n}\right),$$

gde je

$$S_n \equiv S_n(f;h) = \frac{h}{3}\left\{f_0 + 4(f_1 + \cdots + f_{2n-1}) + 2(f_2 + \cdots + f_{2n-2}) + f_{2n}\right\}.$$

Teorema 2.3.2. Ako $f \in C^4[a,b]$ važi jednakost

$$\int_a^b f(x)\,dx - S_n = -\frac{(b-a)^5}{2880n^4} f^{IV}(\xi) \qquad (a < \xi < b).$$

Dokaz je sličan dokazu teoreme 2.3.1.

Daćemo sada jednu uopštenu kvadraturnu formulu otvorenog tipa koja se naziva pravilo srednje tačke, a koja se dobija primenom elementarne formule

$$\int_c^{c+h} f(x)\,dx \cong hf\left(c + \frac{h}{2}\right).$$

Neka je

$$M_n \equiv M_n(f;h) = h(f_{1/2} + f_{3/2} + \cdots + f_{(2n-1)/2}),$$

gde su $f_{(2k-1)/2} = f\left(a + \frac{2k-1}{2}h\right)$ $(k=1,\ldots,n)$ i $h = \frac{b-a}{n}$.

Teorema 2.3.3. Ako $f \in C^2[a,b]$ važi

$$\int_a^b f(x)\,dx - M_n = \frac{(b-a)^3}{24n^2} f''(\xi) \qquad (a < \xi < b).$$

Iz teorema 2.3.1 i 2.3.3 sleduje

Posledica 2.3.1. Ako $f \in C^2[a,b]$ i $f''(x) \geq 0$ na $[a,b]$, tada je

$$M_n \leq \int_a^b f(x)\,dx \leq T_n.$$

Sledeća teorema se može izvesti kao specijalan slučaj jednog opštijeg rezultata dobijenog u radu [12].

Teorema 2.3.4. Neka je f dva puta diferencijabilna funkcija na $[a,b]$ sa ograničenim drugim izvodom, tj. $|f''(x)| \leqq A$ $(\forall x \in [a,b])$. Tada je

$$\left| \int_a^b f(x)\,dx - \frac{1}{2}(T_n + M_n) \right| \leqq \frac{A}{48n^2} \qquad \left(h = \frac{b-a}{n}\right).$$

Primer 2.3.1. Primenom uopštene trapezne i uopštene Simpsonove formule izračunaćemo integral

$$\int_{1.0}^{1.2} \sqrt{x}\,dx,$$

čija je tačna vrednost $\frac{2}{3}\left((1.2)^{3/2} - 1\right) = 0.2096894\ldots$, ili zaokrugljena na šest značajnih cifara 0.209689. Kod izračunavanja koristićemo se tablicom vrednosti funkcije $x \mapsto f(x) = \sqrt{x}$

x	1.00	1.05	1.10	1.15	1.20
f(x)	1.00000	1.02470	1.04881	1.07238	1.09544

Imamo

$$T_4 = 0.05\left\{\frac{1}{2}\cdot 1.00000 + (1.02470 + 1.04881 + 1.07238) + \frac{1}{2}\cdot 1.09544\right\} = 0.2096805$$

i

$$S_2 = \frac{0.05}{3}\left\{1.00000 + 4(1.02470 + 1.07278) + 2\cdot 1.04881 + 1.09544\right\} = 0.2096896,$$

ili zaokrugljivanjem na šest značajnih cifara

$$T_4 = 0.209680 \quad \text{i} \quad S_2 = 0.209690,$$

pri čemu su odgovarajuće greške $9\cdot 10^{-6}$ i $1\cdot 10^{-6}$.

Ispitajmo sada kako utiču greške zaokrugljivanja u vrednostima f_k, na rezultat koji se dobija primenom uopštene trapezne i uopštene Simpsonove formule. Neka $\bar{f}_k$ označava zaokrugljenu vrednost od f_k, tj. neka je odgovarajuća greška zaokrugljivanja $e_k = \bar{f}_k - f_k$. Sem toga, neka je $|e_k| \leqq E$. Kako je

$$T_n(\bar{f};h) = T_n(f;h) + T_n(e;h) \qquad \big(e(x) = \bar{f}(x) - f(x)\big),$$

imamo

$$\left|T_n(e;h)\right| \leqq h\left(\frac{1}{2}|e_0| + \sum_{k=1}^{n-1} |e_k| + \frac{1}{2}|e_n|\right) \leqq nhE = (b-a)E.$$

Ista granica za grešku se dobija i kod Simpsonove formule. Naime, imamo

$$\begin{aligned}\left|S_n(e;f)\right| &\leqq \frac{h}{3}\Big\{|e_0| + 4(|e_1| + \cdots + |e_{2n-1}|) \\ &\quad + 2(|e_2| + \cdots + |e_{2n-2}|) + |e_{2n}|\Big\} \\ &\leqq \frac{h}{3}\Big\{1 + 4\cdot\frac{n}{2} + 2\Big(\frac{n}{2} - 1\Big) + 1\Big\}E = (b-a)E.\end{aligned}$$

Kako su vrednosti funkcije f, u primeru 2.3.1, zaokrugljene na šest značajnih cifara, tj. njihove apsolutne greške ne prelaze $5\cdot 10^{-6}$, za grešku kod izračunavanja sume T_4 važi ocena

$$\left|T_4(e;0.05)\right| \leqq (1.2 - 1)\cdot 5\cdot 10^{-6} = 10^{-6}.$$

Ista je granica i za $\left|S_2(e;0.05)\right|$.

Na kraju napomenimo da za veličine T_n, M_n, S_n važi jednakost

$$S_n(f;h) = \frac{1}{3}\big(T_n(f;2h) + 2M_n(f;2h)\big),$$

gde je $h = \frac{b-a}{2n}$.

7.2.4. Gaussove kvadraturne formule

U ovom i narednom odeljku razmotrićemo kvadraturne formule interpolacionog tipa sa najvišim algebarskim stepenom tačnosti. Posmatraćemo opšti slučaj integracije sa težinskom funkcijom. Naime, posmatraćemo kvadraturnu formulu

$$\int_a^b p(x)f(x)dx = \sum_{k=1}^{n} A_k f(x_k) + R_n(f). \tag{2.4.1}$$

Kako u ovoj formuli imamo 2n nepoznatih parametara(A_k, x_k $(k=1,\ldots,n)$), formula može imati algebarski stepen tačnosti najviše $2n-1$, što znači da je $R_n(f)=0$ za svako $f\in\mathcal{P}_{2n-1}$. Takve formule postoje i nazivaju se Gaussove kvadraturne formule.

Neka je $\{Q_k\}_{k\in N_0}$ ortogonalan niz polinoma na (a,b) sa težinskom funkcijom $x\mapsto p(x)$.

<u>Teorema 2.4.1.</u> Interpolaciona kvadraturna formula (2.4.1) ima algebarski stepen tačnosti $2n-1$, ako i samo ako su x_k $(k=1,\ldots,n)$ nule polinoma $x\mapsto Q_n(x)$.

<u>Dokaz</u>. Neka su x_k $(k=1,\ldots,n)$ nule polinoma Q_n, koje su, kao što je poznato, realne, proste i leže u (a,b). Neka je, dalje, f proizvoljan polinom iz $\mathcal{P}_{2n-1}$. Kako se ovaj polinom može predstaviti u obliku

$$f(x) = Q_n(x)u_{n-1}(x) + v_{n-1}(x), \tag{2.4.2}$$

gde su u_{n-1} i v_{n-1} polinomi iz $\mathcal{P}_{n-1}$, imamo

$$\int_a^b p(x)f(x)dx = \int_a^b p(x)Q_n(x)u_{n-1}(x)dx + \int_a^b p(x)v_{n-1}(x)dx, \tag{2.4.3}$$

tj.

$$\int_a^b p(x)f(x)dx = \int_a^b p(x)v_{n-1}(x)dx,$$

s obzirom da je prvi integral na desnoj strani u jednakosti

(2.4.3) jednak nuli zbog ortogonalnosti niza $\{Q_k\}_{k\in N_0}$.

Primena kvadraturne formule (2.4.1) na desnu stranu poslednje jednakosti daje

$$\int_a^b p(x)f(x)dx = \sum_{k=1}^{n} A_k v_{n-1}(x_k) + R_n(v_{n-1}).$$

Kako je, na osnovu (2.4.2), $f(x_k)=v_{n-1}(x_k)$ $(k=1,\ldots,n)$ i kako je $R_n(v_{n-1}) \equiv 0$ (jer je formula interpolacionog tipa), zaključujemo da kvadraturna formula (2.4.1) ima algebarski stepen tačnosti $2n-1$.

Dokažimo sada obrnuto tvrdjenje. Neka kvadraturna formula (2.4.1) ima algebarski stepen tačnosti $2n-1$. Pomoću $x_1,\ldots,x_n$ konstruišimo polinom $\tilde{Q}_n$, takav da je

$$\tilde{Q}_n(x) = (x-x_1)\ldots(x-x_n)$$

i posmatrajmo integral

$$I_m = \int_a^b p(x)\,x^m\tilde{Q}_n(x)dx \qquad (m=0,1,\ldots,n-1),$$

Kako je $x\mapsto x^m\tilde{Q}_n(x)$ polinom iz $\mathcal{P}_{n+m}$ $(\subset\mathcal{P}_{2n-1})$, imamo

$$I_m = \sum_{k=1}^{n} A_k x_k^m \tilde{Q}_n(x_k) = 0 \qquad (m=0,1,\ldots,n-1),$$

odakle zaključujemo da niz polinoma $\{\tilde{Q}_n\}_{n\in N_0}$ mora biti ortogonalan na (a,b) sa težinskom funkcijom $x\mapsto p(x)$. S obzirom na jedinstvenost ovih polinoma(do na multiplikativnu konstantu), imamo $\tilde{Q}_n(x) = c_nQ_n(x)$.

Ovim je dokaz teoreme završen.

Predjimo sada na odredjivanje težinskih koeficijenata kvadraturne formule.

Na osnovu prethodne teoreme, za interpolacione čvorove $x_k(k=1,\ldots,n)$ kod Gaussovih kvadraturnih formula treba uzeti nule polinoma Q_n. Dakle, u ovom slučaju, imamo

$$\omega(x) = (x-x_1)\ldots(x-x_n) = \frac{Q_n(x)}{a_n},$$

gde je a_n koeficijent uz najviši stepen u polinomu Q_n, tj.

$$Q_n(x) = a_n x^n + \cdots ,$$

i

$$A_k = \int_a^b \frac{p(x)\omega(x)}{(x-x_k)\omega'(x_k)}\,dx = \frac{1}{Q_n'(x_k)}\int_a^b \frac{p(x)Q_n(x)}{x-x_k}\,dx \quad (k=1,\ldots,n$$

<u>Teorema 2.4.2</u>. Težinski koeficijenti Gaussove kvadraturne formule mogu se eksplicitno izraziti pomoću

$$(2.4.4) \qquad A_k = \frac{\alpha_n}{\gamma_n}\cdot\frac{\|Q_n\|^2}{Q_{n-1}(x_k)Q_n'(x_k)} \qquad (k=1,\ldots,n),$$

gde su $x_k(k=1,\ldots,n)$ nule polinoma Q_n, a α_n i γ_n konstante koje se javljaju u rekurentnoj relaciji

$$(2.4.5) \qquad Q_{n+1}(x) = (\alpha_n x + \beta_n)Q_n(x) - \gamma_n Q_{n-1}(x) \qquad (n\in N).$$

<u>Dokaz</u>. Polazeći od Christoffel-Darbouxovog identiteta za ortogonalne polinome

$$(2.4.6) \qquad \sum_{i=0}^{n} \frac{1}{\|Q_i\|^2} Q_i(x)Q_i(t) = \frac{1}{\alpha_n\|Q_n\|^2}\,\frac{Q_{n+1}(x)Q_n(t) - Q_n(x)Q_{n+1}(t)}{x-t}$$

gde je α_n konstanta u rekurentnoj relaciji (2.4.5), možemo odrediti težinske koeficijente $A_k(k=1,\ldots,n)$. Naime, ako u (2.4.6) stavimo $t=x_k$ $(\Rightarrow Q_n(t)=0)$, a zatim dobijenu jednakost pomnožimo sa $p(x)$ i integralimo od a do b, dobijamo

$$\sum_{i=0}^{n} \frac{1}{\|Q_i\|^2} Q_i(x_k)\int_a^b p(x)Q_i(x)\,dx = \frac{-Q_{n+1}(x_k)}{\alpha_n\|Q_n\|^2}\int_a^b \frac{p(x)Q_n(x)}{x-x_k}\,dx$$

tj.

$$A_k = -\frac{\alpha_n \|Q_n\|^2}{Q_n(x_k)Q_{n+1}(x_k)} \qquad (k=1,\ldots,n), \tag{2.4.7}$$

pri čemu je iskorišćena osobina ortogonalnosti niza $\{Q_i\}_{i\in N_0}$.

Najzad, ako u (2.4.5) stavimo $x=x_k$ dobijamo

$$Q_{n+1}(x_k) = -\gamma_n Q_{n-1}(x_k),$$

što zajedno sa (2.4.7) daje (2.4.4).

Primedba 2.4.1. Kako je $x\mapsto f_k(x)=\left(\dfrac{Q_n(x)}{x-x_k}\right)^2$ polinom stepena $2n-2$, to je $R_n(f_k)=0$, tj.

$$A_k\, Q_n'(x_k)^2 = \int_a^b p(x) f_k(x)\,dx > 0,$$

odakle zaključujemo da su koeficijenti A_k $(k=1,\ldots,n)$ pozitivni.

Predjimo sada na odredjivanje ostatka $R_n(f)$ u Gaussovoj kvadraturnoj formuli.

Teorema 2.4.3. Neka su x_k i A_k $(k=1,\ldots,n)$ kao u prethodnoj teoremi i neka $f\in C^{2n}[a,b]$. Tada za ostatak $R_n(f)$ važi formula

$$R_n(f) = \frac{\|Q_n\|^2}{(2n)!\,a_n^2} f^{(2n)}(\xi) \qquad (a<\xi<b), \tag{2.4.8}$$

gde je a_n koeficijent uz najviši stepen u polinomu Q_n.

Dokaz. Posmatrajmo Hermiteov interpolacioni polinom H_{2n-1} konstruisan na skupu podataka $f(x_k), f'(x_k)$ $(k=1,\ldots,n)$. Kako je greška kod ovog polinoma (videti odeljak 6.2.3)

$$r_{2n-1}(f;x) = f(x) - H_{2n-1}(x) = \frac{f^{(2n)}(\eta)}{(2n)!}\,\omega(x)^2 \qquad (a<\eta<b),$$

gde je $\omega(x) = (x-x_1)\ldots(x-x_n)$ i kako $H_{2n-1} \in \mathcal{P}_{2n-1}$, imamo

$$\int_a^b p(x)f(x)dx = \int_a^b p(x)H_{2n-1}(x)dx + \int_a^b p(x)r_{2n-1}(f;x)dx$$

$$= \sum_{k=1}^{n} A_k H_{2n-1}(x_k) + R_n(f)$$

$$= \sum_{k=1}^{n} A_k f(x_k) + R_n(f),$$

gde je

$$R_n(f) = \int_a^b p(x)r_{2n-1}(f;x)dx = \int_a^b \frac{1}{(2n)!} p(x)\omega(x)^2 f^{(2n)}(\eta)dx.$$

S obzirom da je $\omega(x) = \dfrac{Q_n(x)}{a_n}$ i $f \in C^{2n}[a,b]$, ostatak $R_n(f)$ se može predstaviti u obliku

$$R_n(f) = \frac{f^{(2n)}(\xi)}{(2n)!\,a_n^2} \int_a^b p(x)Q_n(x)^2 dx \qquad (a < \xi < b),$$

što je ekvivalentno sa (2.4.8).

Primedba 2.4.2. Kako je

$$\alpha_n = \frac{a_{n+1}}{a_n} \quad \text{i} \quad \gamma_n = \frac{\alpha_n}{\alpha_{n-1}} \left(\frac{\|Q_n\|}{\|Q_{n-1}\|}\right)^2,$$

imamo

$$\frac{\alpha_n}{\gamma_n} \|Q_n\|^2 = \alpha_{n-1} \|Q_{n-1}\|^2 = \frac{a_{n-1}}{a_n} \|Q_{n-1}\|^2.$$

Formula (2.4.4) se tada svodi na

$$A_k = \frac{a_n}{a_{n-1}} \cdot \frac{\|Q_{n-1}\|^2}{Q_{n-1}(x_k)Q_n'(x_k)} \qquad (k=1,\ldots,n). \tag{2.4.9}$$

Primer 2.4.1. Odredimo Gaussovu kvadraturnu formulu oblika

$$\int_{-1}^{1} (1-x^2)^{3/2} f(x)\,dx = A_1 f(x_1) + A_2 f(x_2) + A_3 f(x_3) + R_3(f). \tag{2.4.10}$$

Polazeći od prirodnog bazisa $\{1, x, x^2, \dots\}$ Gram-Schmidt-ovim postupkom ortogonalizacije dobijamo niz ortogonalnih polinoma na $(-1,1)$ sa težinskom funkcijom $x \mapsto p(x) = (1-x^2)^{3/2}$:

$$Q_0(x) = 1, \quad Q_1(x) = x, \quad Q_2(x) = x^2 - \frac{1}{6}, \quad Q_3(x) = x^3 - \frac{3}{8}x,$$

$$Q_4(x) = x^4 - \frac{3}{5}x^2 - \frac{3}{80}, \text{ itd.}$$

Na osnovu teoreme 2.4.1, interpolacioni čvorovi x_1, x_2, x_3 su nule polinoma Q_3. Dakle,

$$x_1 = -\sqrt{\frac{3}{8}}, \quad x_2 = 0, \quad x_3 = \sqrt{\frac{3}{8}}.$$

Kako su koeficijenti uz najviše stepene u polinomima Q_2 i Q_3 jednaki jedinici, tj. $a_2 = 1$ i $a_3 = 1$, i kako je

$$\|Q_2\|^2 = \int_{-1}^{1} (1-x^2)^{3/2} \left(x^2 - \frac{1}{6}\right)^2 dx = \frac{5\pi}{384},$$

na osnovu (2.4.9) imamo

$$A_k = \frac{5\pi}{384} \cdot \frac{1}{3\left(x_k^2 - \frac{1}{6}\right)\left(x_k^2 - \frac{1}{8}\right)} \qquad (k=1,2,3),$$

tj.

$$A_1 = A_3 = \frac{\pi}{12} \quad \text{i} \quad A_2 = \frac{5\pi}{24}.$$

Ako $f \in C^6[-1,1]$, na osnovu teoreme 2.4.3, imamo

$$R_3(f) = \frac{f^{(6)}(\xi)}{6!} \int_{-1}^{1} (1-x^2)^{3/2} \left(x^3 - \frac{3}{8}x\right)^2 dx = \frac{3\pi f^{(6)}(\xi)}{2^{10}\, 6!},$$

gde $\xi \in (-1,1)$.

Primer 2.4.2. Primenom formule (2.4.10) odredićemo približno vrednost integrala

$$I = \int_{-1}^{1} (1-x^2)^{3/2} \cos x\, dx.$$

Imamo

$$I \simeq \frac{\pi}{12}\left(2\cos\frac{\sqrt{6}}{4} + \frac{5}{2}\right) \simeq 1.082952.$$

Medjutim, tačna vrednost integrala je

$$I = 3\pi J_2(1) \simeq 1.082935,$$

gde je J_2 Besselova funkcija.

7.2.5. Specijalni slučajevi Gaussovih formula

Razmotrićemo sada Gaussove formule u slučaju kada niz polinoma $\{Q_n\}_{n\in N_0}$ pripada klasi tzv. klasičnih ortogonalnih polinoma. Kao što je poznato, klasični ortogonalni polinomi dele se u tri podklase, čiji su kanonički oblici dati u sledećoj tabeli.

(a,b)	p(x)	Q_n	naziv
(-1,1)	$(1-x)^{\alpha}(1+x)^{\beta}$ $\alpha,\beta > -1$	$P_n^{(\alpha,\beta)}$	Jacobievi polinomi
$(0,+\infty)$	$x^s e^{-x}$ $s > -1$	L_n^s	Laguerreovi polinomi
$(-\infty,+\infty)$	e^{-x^2}	H_n	Hermiteovi polinomi

Razmotrićemo posebno ova tri slučaja.

1. Gauss-Jacobieve formule imaju oblik

$$\int_{-1}^{+1} (1-x)^{\alpha}(1+x)^{\beta} f(x)\, dx = \sum_{k=1}^{n} A_k f(x_k) + R_n(f).$$

Kako su a_n (koeficijent uz najviši stepen u polinomu Q_n) i $||Q_n||^2$ dati sa

$$a_n = \frac{1}{2^n n!}\cdot\frac{\Gamma(2n+\alpha+\beta+1)}{\Gamma(n+\alpha+\beta+1)}$$

i

$$||Q_n||^2 = 2^{\alpha+\beta+1}\frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{n!(2n+\alpha+\beta+1)\Gamma(n+\alpha+\beta+1)},$$

na osnovu (2.4.9) nalazimo težinske koeficijente

$$A_k = \frac{2^{\alpha+\beta}(2n+\alpha+\beta)\Gamma(n+\alpha)\Gamma(n+\beta)}{n!(n+\alpha+\beta)\Gamma(n+\alpha+\beta)}\cdot\frac{1}{P_{n-1}^{(\alpha,\beta)}(x_k)\frac{d}{dx}(P_n^{(\alpha,\beta)}(x_k))},$$

gde su x_k $(k=1,\ldots,n)$ nule Jacobievog polinoma $P_n^{(\alpha,\beta)}$.

Ako $f\in C^{2n}[-1,1]$ za ostatak dobijamo

$$R_n(f) = \frac{2^{2n+\alpha+\beta+1}n!\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{(2n)!(2n+\alpha+\beta+1)\Gamma(2n+\alpha+\beta+1)^2}f^{(2n)}(\xi), \tag{2.5.1}$$

gde $\xi\in(-1,1)$.

1.1. Posebno za $\alpha=\beta=0, p(x)=1$, dobijamo Gauss-Legendre ovu formulu

$$\int_{-1}^{+1} f(x)dx = \sum_{k=1}^{n} A_k f(x_k) + R_n(f),$$

pri čemu su

$$A_k = \frac{2}{n}\cdot\frac{1}{P_{n-1}(x_k)P_n'(x_k)} \quad (k=1,\ldots,n) \tag{2.5.2}$$

i x_k $(k=1,\ldots,n)$ nule Legendreovog polinoma P_n.
Formula (2.5.1) se tada svodi na

$$R_n(f) = \frac{2^{2n+1}(n!)^4}{(2n+1)((2n)!)^3}f^{(2n)}(\xi) \quad (-1<\xi<1).$$

<u>Primedba 2.5.1.</u> Za n=2,4,6 imamo

$$R_2(f) = 7.4\cdot 10^{-3}f^{(4)}(\xi_2),\quad R_4(f) = 3\cdot 10^{-7}f^{(8)}(\xi_4),$$

$$R_6(f) = 1.5\cdot 10^{-12}f^{(12)}(\xi_6),\quad (\xi_2,\xi_4,\xi_6\in(-1,1)).$$

Primedba 2.5.2. Kako je Legendreov polinom parna ili neparna funkcija u zavisnosti da li je n paran ili neparan broj, za težinske koeficijente $A_k (k=1,\ldots,n)$, koji su dati pomoću (2.5.2), važi

$$A_k = A_{n-k} \qquad (k=1,\ldots,\left[\frac{n}{2}\right]).$$

Primedba 2.5.3. Na osnovu jednakosti za Legendreove polinome

$$(1-x^2)P_n'(x) + nxP_n(x) = nP_{n-1}(x),$$

formula (2.5.2) se može predstaviti u obliku

$$A_k = \frac{2(1-x_k^2)}{n^2 P_{n-1}(x_k)^2} = \frac{2}{(1-x_k^2)P_n'(x_k)^2} \qquad (k=1,\ldots,n).$$

U sledećoj tabeli dajemo vrednosti za x_k i A_k $(k=1,\ldots,n)$ za Gauss-Legendreovu formulu kada je $n \leq 6$.

n	k	x_k	A_k
1	1	0.	2.
2	1,2	∓0.57735027	1.
3	1,3 2	∓0.77459667 0.	0.55555556 0.88888889
4	1,4 2,3	∓0.86113631 ∓0.33998104	0.34785488 0.65214516
5	1,5 2,4 3	∓0.90617985 ∓0.53846931 0.	0.23692688 0.47862868 0.56888889
6	1,6 2,5 3,4	∓0.93246951 ∓0.66120939 ∓0.23861919	0.17132450 0.36076158 0.46791394

Primer 2.5.1. Primenimo Gauss-Legendreove formule na izračunavanje integrala

$$\int_0^{\pi/2} \sin t \, dt.$$

Smenom $t = \frac{\pi}{4}(x+1)$, dati integral se svodi na $I = \int_{-1}^{1} \frac{\pi}{4} \sin \frac{\pi}{4}(x+1)dx$.

Za n=1,2,3 redom imamo

$$I \simeq 2 \cdot \frac{\pi}{4} \cdot \frac{\sqrt{2}}{2} = 1.11072,$$

$$I \simeq 1 \cdot \frac{\pi}{4}(0.325885 + 0.945409) = 0.99847,$$

$$I \simeq \frac{\pi}{4}\left(\frac{5}{9} \cdot 0.176108 + \frac{8}{9} \cdot 0.707107 + \frac{5}{9} \cdot 0.984371\right) = 1.00001,$$

pri čemu su rezultati zaokrugljeni na pet decimala. Primetimo da je tačna vrednost integrala I=1.

1.2. Za $\alpha=\beta= -\frac{1}{2}$, $p(x) = \frac{1}{\sqrt{1-x^2}}$ dobijamo Gauss-Čebiševljevu formulu

$$\int_{-1}^{+1} \frac{f(x)}{\sqrt{1-x^2}} dx = \sum_{k=1}^{n} A_k f(x_k) + R_n(f),$$

pri čemu su x_k nule Čebiševljevog polinoma T_n, koji je odredjen sa

$$T_n(x) = \frac{n!}{\left(\frac{1}{2}\right)_n} P_n^{(-1/2,\ -1/2)}(x),$$

gde je

$$(s)_n = s(s+1)\ldots(s+n-1) = \frac{\Gamma(s+n)}{\Gamma(s)}.$$

Kako se polinom T_n može predstaviti i u obliku $T_n(x) = \cos(n \arccos x)$, imamo

$$x_k = \cos \frac{\pi}{2n}(2k-1) \quad (k=1,\ldots,n).$$

Za težinske koeficijente A_k dobijamo

$$(2.5.3) \qquad A_k = \frac{2^{-1}(2n-1)\Gamma(n-\frac{1}{2})^2}{n!(n-1)\Gamma(n-1)} \cdot \frac{C_n C_{n-1}}{T_{n-1}(x_k)T_n'(x_k)},$$

gde je $C_n = \frac{n!}{\left(\frac{1}{2}\right)_n} = n! \frac{\sqrt{\pi}}{\Gamma(n+\frac{1}{2})}$.

Kako je $T_{n-1}(x_k)T_n'(x_k)=n$, (2.5.3) se svodi na $A_k = \frac{\pi}{n}$ $(k=1,\ldots,n)$.

Dakle, Gauss-Čebiševljeva kvadraturna formula je

$$\int_{-1}^{+1} \frac{f(x)}{\sqrt{1-x^2}}\,dx = \frac{\pi}{n}\sum_{k=1}^{n} f\left(\cos\frac{\pi}{2n}(2k-1)\right) + R_n(f).$$

Stavljajući $\alpha=\beta=-\frac{1}{2}$ u (2.5.1), dobijamo ostatak

$$R_n(f) = \frac{\pi}{2^{2n-1}(2n)!} f^{(2n)}(\xi) \qquad (-1<\xi<1).$$

2. <u>Gauss-Laguerreove formule</u> imaju oblik

$$\int_{0}^{+\infty} x^s e^{-x} f(x)\,dx = \sum_{k=1}^{n} A_k f(x_k) + R_n(f).$$

Kako je, u ovom slučaju,

$$a_n = (-1)^n \quad \text{i} \quad \|L_n^s\|^2 = n!\,\Gamma(n+s+1),$$

na osnovu (2.4.9) imamo

$$A_k = -\frac{(n-1)!\,\Gamma(n+s)}{L_{n-1}^s(x_k)\frac{d}{dx}L_n^s(x_k)} \qquad (k=1,\ldots,n), \tag{2.5.4}$$

gde su x_k $(k=1,\ldots,n)$ nule polinoma L_n^s.

<u>Primedba 2.5.4.</u> Kako na osnovu jednakosti

$$xL_{n-1}^{s+1}(x) = (n+s)L_{n-1}^s(x) - L_n^s(x)$$

i

$$\frac{d}{dx}L_n^s(x) = -nL_{n-1}^{s+1}(x),$$

imamo

$$\frac{d}{dx}L_n^s(x_k) = -\frac{n}{x_k}(n+s)L_{n-1}^s(x_k),$$

formula (2.5.4) se može predstaviti i u sledećim ekvivalentnim oblicima

$$A_k = \frac{(n-1)!\,\Gamma(n+s)x_k}{n(n+s)(L_{n-1}^s(x_k))^2} = \frac{n!\,\Gamma(n+s+1)}{x_k\left(\frac{d}{dx}L_n^s(x_k)\right)^2} \qquad (k=1,\ldots,n).$$

Ako $f\in C^{2n}[0,+\infty)$ za ostatak se dobija

$$R_n(f) = \frac{n!\,\Gamma(n+s+1)}{(2n)!} f^{(2n)}(\xi) \quad (0<\xi<+\infty).$$

3. Gauss-Hermiteova formula ima oblik

$$\int_{-\infty}^{+\infty} e^{-x^2} f(x)\,dx = \sum_{k=1}^{n} A_k f(x_k) + R_n(f).$$

Kako je

$$a_n = 2^n \quad \text{i} \quad \|H_n\|^2 = 2^n n!\sqrt{\pi},$$

imamo

(2.5.5) $$A_k = \frac{2^n (n-1)!\sqrt{\pi}}{H_{n-1}(x_k) H_n'(x_k)} \quad (k=1,\ldots,n),$$

gde su x_k nule Hermiteovog polinoma H_n. S obzirom na jednakost $H_n'(x) = 2nH_{n-1}(x)$, jednakost (2.5.5) se može predstaviti i u obliku

$$A_k = \frac{2^{n+1} n!\sqrt{\pi}}{H_n'(x_k)^2} = \frac{2^{n-1}(n-1)!\sqrt{\pi}}{nH_{n-1}(x_k)^2} \quad (k=1,\ldots,n).$$

Za ostatak se dobija

$$R_n(f) = \frac{n!\sqrt{\pi}}{2^n (2n)!} f^{(2n)}(\xi) \quad (-\infty<\xi<+\infty).$$

Primedba 2.5.5. Kao i kod Gauss-Legendreove formule, i u ovom slučaju je $A_k = A_{n-k}$ $(k=1,\ldots,[\frac{n}{2}])$. Neke vrednosti za x_k i A_k date su u sledećoj tabeli.

n	k	x_k	A_k
1	1	0.	1.77245385
2	1,2	∓0.70710678	0.88622693
3	1,3	∓1.22474487	0.29540898
	2	0.	1.18163590
4	1,4	∓1.65068012	0.08131284
	2,3	∓0.52464762	0.80491409

7.2.6. Modifikovane Gaussove formule

Kvadraturne formule oblika

$$\int_a^b p(x)f(x)dx = \sum_{k=1}^{n} A_k f(x_k) + \sum_{k=1}^{m} B_k f(y_k) + R_{n+m}(f), \tag{2.6.1}$$

sa algebarskim stepenom tačnosti $2n+m-1$, pri čemu su y_k unapred zadate apscise, nazivaju se modifikovane Gaussove formule. Dakle, nepoznate parametre A_k, x_k $(k=1,\ldots,n)$ i $B_k (k=1,\ldots,m)$ možemo odrediti iz uslova

$$R_{n+m}(x^i) = 0 \quad (i=0,1,\ldots,2n+m-1).$$

Primer 2.6.1. Izvedimo modifikovanu Gaussovu formulu oblika

$$\int_{-1}^{+1} f(x)dx \simeq A_1 f(x_1)+A_2 f(x_2)+B_1 f(-1)+B_2 f(1). \tag{2.6.2}$$

Kako ova formula mora biti tačna za svako $f \in P_5$, na osnovu (2.6.2), uzimajući za $f(x)$ redom $1, x, \ldots, x^5$ dobijamo sistem nelinearnih jednačina

$$\begin{aligned} A_1 + A_2 + B_1 + B_2 &= 2,\\ A_1x_1 + A_2x_2 - B_1 + B_2 &= 0,\\ A_1x_1^2 + A_2x_2^2 + B_1 + B_2 &= \frac{2}{3},\\ A_1x_1^3 + A_2x_2^3 - B_1 + B_2 &= 0,\\ A_1x_1^4 + A_2x_2^4 + B_1 + B_2 &= \frac{2}{5},\\ A_1x_1^5 + A_2x_2^5 - B_1 + B_2 &= 0. \end{aligned} \tag{2.6.3}$$

Da bismo rešili ovaj sistem jednačina uvedimo pomoćnu funkciju ω pomoću

$$\begin{aligned} \omega(x) &= (x-x_1)(x-x_2)(x+1)(x-1)\\ &= x^4 + C_3x^3 + C_2x^2 + C_1x + C_o. \end{aligned}$$

Sada, množenjem prvih pet jednačina sistema (2.6.3) sa C_o, C_1, C_2, C_3, 1 respektivno i njihovim sabiranjem dobijamo

$$A_1\omega(x_1)+A_2\omega(x_2)+B_1\omega(-1)+B_2\omega(1) = 2C_o + \frac{2}{3}C_2 + \frac{2}{5}.$$

Primenjujući isti postupak, iz poslednjih pet jednačina sistema (2.6.3) sleduje

$$A_1x_1\omega(x_1)+A_2x_2\omega(x_2)-B_1\omega(-1)+B_2\omega(1) = \frac{2}{3}C_1 + \frac{1}{5}C_3.$$

Kako je, s druge strane, $\omega(x_1)=\omega(x_2)=\omega(-1)=\omega(1)=0$, na osnovu prethodnog, dolazimo do sistema linearnih jednačina

$$C_o - C_1 + C_2 - C_3 = -1,$$

$$C_o + C_1 + C_2 + C_3 = -1,$$

$$2C_o + \frac{2}{3}C_2 = -\frac{2}{5},$$

$$\frac{2}{3}C_1 + \frac{1}{5}C_3 = 0,$$

odakle je

$$C_o = \frac{1}{5}, \quad C_1 = 0, \quad C_2 = -\frac{6}{5}, \quad C_3 = 0,$$

tj.

$$\omega(x) = x^4 - \frac{6}{5}x^2 + \frac{1}{5} = (x-1)(x+1)(x^2 - \frac{1}{5}).$$

Iz uslova $\omega(x)=0$, nalazimo $x_1 = -x_2 = -\frac{\sqrt{5}}{5}$.

Najzad, zamenom vrednosti za x_1 i x_2 u (2.6.2) i rešavanjem dobijenog sistema linearnih jednačina po A_k, B_k $(k=1,2)$ dobijamo

$$A_1 = A_2 = \frac{5}{6} \quad \text{i} \quad B_1 = B_2 = \frac{1}{6},$$

tj.

$$\int_{-1}^{+1} f(x)dx \approx \frac{5}{6}(f(-\frac{\sqrt{5}}{5})+f(\frac{\sqrt{5}}{5})) + \frac{1}{6}(f(-1)+f(1)).$$

Za modifikovane Gaussove formule može se dokazati sledeća teorema.

<u>Teorema 2.6.1.</u> Kvadraturna formula (2.6.1) ima algebarski stepen tačnosti $2n+m-1$ ako i samo ako je

1) tačna za svako $f \in P_{n+m-1}$,

2) $\int_a^b p(x)\omega(x)f(x)dx = 0$ za svako $f \in P_{n-1}$,

gde je

$$\omega(x) = \prod_{k=1}^{n} (x-x_k) \prod_{k=1}^{m} (x-y_k).$$

Navešćemo sada dva interesantna slučaja modifikovanih Gaussovih formula.

Teorema 2.6.2 (Radau). Neka su x_k $(k=1,\ldots,n-1)$ nule polinoma $x \mapsto \frac{1}{x-1}(P_n(x)+P_{n-1}(x))$ i

$$A_k = \frac{1-x_k}{n^2 P_{n-1}(x_k)^2} \quad (k=1,\ldots,n-1),$$

gde je P_n Legendreov polinom. Ako $f \in C^{2n-1}[-1,1]$ ostatak u kvadratnoj formuli

$$\int_{-1}^{1} f(x)\,dx = \frac{2}{n^2} f(-1) + \sum_{k=1}^{n-1} A_k f(x_k) + R_n(f)$$

je dat pomoću

$$R_n(f) = 2^{2n-1} n! \left[\frac{(n-1)!}{(2n-1)!}\right]^3 f^{(2n-1)}(\xi) \quad (-1 < \xi < 1).$$

Teorema 2.6.3 (Lobatto). Neka su x_k $(k=1,\ldots,n-2)$ nule polinoma $x \mapsto P'_{n-1}(x)$ (P_n Legendreov polinom) i

$$A_k = \frac{2}{n(n-1)P_{n-1}(x_k)^2} \quad (k=1,\ldots,n-2).$$

Ako $f \in C^{2n-2}[-1,1]$ važi formula

$$\int_{-1}^{+1} f(x)\,dx = \frac{2}{n(n-1)}(f(-1)+f(1)) + \sum_{k=1}^{n-2} A_k f(x_k) + R_n(f),$$

gde je

$$R_n(f) = \frac{2^{2n-1} n!(n-2)!}{(2n-1)!}\left[\frac{(n-1)!}{(2n-2)!}\right]^2 f^{(2n-2)}(\xi) \quad (-1 < \xi < 1).$$

7.2.7. Čebiševljeve kvadraturne formule

Kvadraturne formule oblika

$$\int_a^b p(x)f(x)dx = A_n \sum_{k=1}^{n} f(x_k) + R_n(f), \tag{2.7.1}$$

sa algebarskim stepenom tačnosti n, nazivaju se Čebiševljeve formule. Nepoznate parametre A_n i x_k $(k=1,\ldots,n)$ možemo odrediti rešavanjem sistema jednačina

$$A_n(x_1^m+\ldots+x_n^m) = \int_a^b p(x)x^m dx \qquad (m=0,1,\ldots,n). \tag{2.7.2}$$

Iz prve jednačine ovog sistema odredjujemo

$$A_n = \frac{1}{n}\int_a^b p(x)dx. \tag{2.7.3}$$

Primetimo da je $A_n > 0$, s obzirom da je p težinska funkcija.

Uvedimo oznake

$$s_m = x_1^m+\ldots+x_n^m \qquad (m=1,2,\ldots)$$

i

$$\omega(x) = (x-x_1)\ldots(x-x_n)$$
$$= x^n + a_1x^{n-1}+\ldots+a_{n-1}x+a_n.$$

Na osnovu (2.7.2) i (2.7.3) možemo odrediti vrednosti s_m $(m=1,\ldots,n)$ pomoću

$$s_m = \frac{1}{A_n}\int_a^b p(x)x^m dx = n\,\frac{\int_a^b p(x)x^m dx}{\int_a^b p(x)dx}.$$

Tada stavljanjem m=1,...,n u jednakosti (videti [13,str.74])

$$s_m+a_1s_{m-1}+a_2s_{m-2}+\ldots+a_{m-1}s_1+ma_m = 0$$

dolazimo do sistema linearnih jednačina

$$
(2.7.4)\qquad
\begin{aligned}
1\cdot a_1 \qquad\qquad\qquad\qquad\qquad\qquad\qquad &= -s_1,\\
s_1a_1 + 2a_2 \qquad\qquad\qquad\qquad\qquad\qquad &= -s_2,\\
s_2a_1 + s_1a_2 + 3a_3 \qquad\qquad\qquad\qquad &= -s_3,\\
&\ \vdots\\
s_{n-2}a_1+s_{n-3}a_2+s_{n-4}a_3+\dots+(n-1)a_{n-1} &= -s_{n-1},\\
s_{n-1}a_1+s_{n-2}a_2+s_{n-3}a_3+\dots+s_1a_{n-1}+na_n &= -s_n,
\end{aligned}
$$

čijim se rešavanjem nalaze nepoznati koeficijenti polinoma $x\mapsto\omega(x)$. Napomenimo da navedeni sistem linearnih jednačina ima jedinstveno rešenje s obzirom da njegova determinanta ima vrednost $n!$. Najzad, nalaženjem nula polinoma ω odredjujemo apscise x_k $(k=1,\dots,n)$.

Jasno je da se medju nulama polinoma ω mogu naći i kompleksne nule, kao i realne nule koje ne pripadaju segmentu $[a,b]$. Za integraciju realnih funkcija realnog argumenta, kakvu mi razmatramo, od interesa su samo slučajevi kada su sve nule $x_k (k=1,\dots,n)$ realne, proste i leže u $[a,b]$. Imajući u vidu ovu činjenicu može se postaviti sledeći problem (Čebiševljev kvadraturni problem):

Odrediti realno A_n i realne i različite apscise $x_1,\dots,x_n\in[a,b]$ tako da je formula (2.7.1) tačna za svako $f\in P_n$ $(n=1,2,\dots)$.

U najprostijem slučaju kada je $p(x)\equiv 1$, Čebiševljev kvadraturni problem se ne može rešiti kada je $n=8$ i $n>9$.

Za slučaj $[a,b]=[-1,1]$, J.L.Ullman ([14]) je odredio jednoparametarsku familiju težinskih funkcija oblika

$$
(2.7.5)\qquad p(x)=\frac{1}{\sqrt{1-x^2}}\cdot\frac{1+2r\,x}{1+4rx+4r^2}\qquad \left(|r|<\frac{1}{4}\right),
$$

sa osobinom da odgovarajući Čebiševljev problem ima rešenje za svako $n\in N$. Primetimo da se za $r=0$ (2.7.5) svodi na klasičnu Čebiševljevu težinsku funkciju i tada su Gauss-Čebiševljeva kvadraturna formula i odgovarajuća Čebiševljeva formula ekvivalentne.

Primer 2.7.1. Odredimo nepoznate parametre u Čebiševljevoj kvadraturnoj formuli

$$
(2.7.6)\qquad \int_{-1}^{1} f(x)dx \simeq A_3(f(x_1)+f(x_2)+f(x_3)).
$$

Uzimajući za f(x) redom $1, x, x^2, x^3$, nalazimo

$$A_3 = \frac{2}{3}, \quad s_1 = x_1+x_2+x_3 = \frac{3}{2}\int_{-1}^{1} x\,dx = 0,$$

$$s_2 = x_1^2+x_2^2+x_3^2 = \frac{3}{2}\int_{-1}^{1} x^2\,dx = 1, \quad s_3 = x_1^3+x_2^3+x_3^3 = \frac{3}{2}\int_{-1}^{1} x^3\,dx = 0.$$

Koeficijente polinoma

$$\omega(x) = (x-x_1)(x-x_2)(x-x_3) = x^3+a_1x^2+a_2x+a_3$$

odredjujemo iz sistema jednačina (2.7.4). Imamo

$$a_1 = -s_1 = 0, \quad a_2 = \frac{1}{2}(-s_2-s_1a_1) = -\frac{1}{2},$$

$$a_3 = \frac{1}{3}(-s_3-s_2a_1-s_1a_2) = 0,$$

tj.

$$\omega(x) = x^3 - \frac{1}{2}x = x\left(x+\frac{\sqrt{2}}{2}\right)\left(x-\frac{\sqrt{2}}{2}\right),$$

odakle sleduje $x_1 = -\frac{\sqrt{2}}{2}$, $x_2 = 0$, $x_3 = \frac{\sqrt{2}}{2}$. Dakle, formula (2.7.6) postaje

$$\int_{-1}^{1} f(x)\,dx \simeq \frac{2}{3}\left(f\left(-\frac{\sqrt{2}}{2}\right) + f(0) + f\left(\frac{\sqrt{2}}{2}\right)\right).$$

7.2.8. Neki metodi za ocenu ostatka u kvadraturnim formulama

Kod numeričkog izračunavanja integrala, kao što je od ranije poznato, javlja se greška ili ostatak kvadraturne formule $R_n(f)$, tj.

$$\int_a^b p(x)f(x)\,dx = \sum_{k=1}^{n} A_k f(x_k) + R_n(f).$$

Naravno, napomenimo da se kod izračunavanja sume $\sum_{k=1}^{n} A_k f(x_k)$ javlja i greška zaokrugljivanja, koja zavisi od tačnosti računara.

U ranijem izlaganju, dali smo ocenu ostatka $R_n(f)$ u nekim elementarnim kvadraturnim formulama. U opštem slučaju, veličina $R_n(f)$ zavisi od izbora kvadraturne formule i od osobina funkcije f.

U ovom odeljku, daćemo dva metoda za ocenu ostatka kvadraturnih formula. Prvi metod se odnosi na klasu neprekidno-diferencijabilnih funkcija, a drugi na jednu klasu analitičkih funkcija.

Primetimo, najpre, da je funkcionela L definisana na izvesnoj klasi funkcija X, pomoću

$$Lf = R_n(f) = \int_a^b p(x)f(x)dx - \sum_{k=1}^{n} A_k f(x_k),$$

linearna funkcionela na X.

<u>Teorema 2.8.1(Peano)</u>. Neka je $R_n(f) \equiv 0$ za svako $f \in \mathcal{P}_m$. Tada je, za svako $f \in C^{m+1}[a,b]$,

$$R_n(f) = \int_a^b K_m(t)f^{(m+1)}(t)dt, \tag{2.8.1}$$

gde je $K_m(t) = \frac{1}{m!} Lu$ i funkcija $x \mapsto u(x)$ definisana pomoću

$$u(x) = (x-t)_+^m = \begin{cases} (x-t)^m & (x \geq t) \\ 0 & (x < t). \end{cases}$$

<u>Dokaz</u>. S obzirom da se ostatak E_m u Taylorovoj formuli

$$f(x) = f(a) + f'(a)(x-a) + \cdots + \frac{1}{m!} f^{(m)}(a)(x-a)^m + E_m$$

može predstaviti u obliku

$$E_m = \frac{1}{m!} \int_a^x (x-t)^m f^{(m+1)}(t)dt = \frac{1}{m!} \int_a^b (x-t)_+^m f^{(m+1)}(t)dt,$$

primenom linearne funkcionele L na f, dobijamo

$$Lf = \sum_{i=0}^{m} \frac{1}{i!} f^{(i)}(a)L((x-a)^i) + \frac{1}{m!} L\left(\int_a^b (x-t)_+^m f^{(m+1)}(t)dt\right).$$

Kako je, medjutim, $Lf = 0$ $(\forall f \in \mathcal{P}_m)$ poslednja jednakost se svodi na

$$R_n(f) = \frac{1}{m!} \int_a^b (Lu) f^{(m+1)}(t)dt ,$$

gde smo stavili $u(x) = (x-t)_+^m$.

Ovim je dokaz završen.

<u>Primedba 2.8.1</u>. Funkcija $t \mapsto K_m(t)$ se naziva Peano ovo jezgro za funkcionelu L. Primetimo, takodje, da je

$$m!\,K_m(t) = \int_a^b p(x)(x-t)_+^m dx - \sum_{x_k>t} A_k(x_k - t)^m .$$

U slučaju kada je $p(x) \equiv 1$, poslednja jednakost se svodi na

$$m!\,K_m(t) = \frac{(b-t)^{m+1}}{m+1} - \sum_{x_k>t} A_k(x_k - t)^m .$$

<u>Posledica 2.8.1</u>. Ako jezgro K_m ne menja znak na $[a,b]$, važi

$$(2.8.2) \qquad R_n(f) = \frac{f^{(m+1)}(\xi)}{(m+1)!} R_n(x^{m+1}) \qquad (a < \xi < b).$$

<u>Dokaz</u>. Ako u (2.8.1) stavimo $f(x) = x^{m+1}$, dobijamo

$$R_n(f) = (m+1)! \int_a^b K_m(t)dt .$$

S druge strane, na osnovu teoreme o srednjoj vrednosti odredjenih integrala, jednakost (2.8.1) može se predstaviti u obliku

$$R_n(f) = f^{(m+1)}(\xi)\int_a^b K_m(t)dt \qquad (a < \xi < b),$$

s obzirom da $f \in C^{m+1}[a,b]$.

Kombinujući poslednje dve jednakosti, dobijamo (2.8.2).

Primer 2.8.1. Odredimo ostatak u trapeznoj formuli. Ovde je $n=2$, $m=1$ i

$$R_2(f) = \int_a^b f(x)dx - \frac{b-a}{2}(f(a)+f(b)).$$

Primenom Peanoove teoreme, imamo

$$K_1(t) = \int_a^b (x-t)_+dx - \frac{b-a}{2}((a-t)_+ + (b-t)_+),$$

$$= \frac{1}{2}(a-t)(b-t).$$

Kako je $K(t) \leq 0$ $(t \in [a,b])$, možemo koristiti tvrdjenje posledice 2.8.1, pa je

$$R_2(f) = \frac{1}{2} f''(\xi)R_2(x^2) = -\frac{(b-a)^3}{12} f''(\xi) \qquad (a < \xi < b).$$

Neka je, sada, X Hilbertov prostor regularnih funkcija $z \mapsto f(z)$ u disku $|z|<1$ i neprekidnih na $\Gamma = \{z \mid |z|=1\}$, sa skalarnim proizvodom

$$(f,g) = \frac{1}{2\pi}\oint_\Gamma f(z)\overline{g(z)}\,ds \qquad (f,g \in X),$$

gde je ds element luka krive Γ.

Primedba 2.8.2. Sistem funkcija $\{1,z,z^2,\dots\}$ u prostoru X je ortonormiran, tj.

$$(z^m,z^n) = \frac{1}{2\pi}\oint_\Gamma z^m\,\bar{z}^n\,ds = \begin{cases} 1 & (m=n) \\ 0 & (m\neq n), \end{cases}$$

što se lako proverava neposrednim izračunavanjem.

Ako je $-1 < a < b < 1$ i $f \in X$, daćemo ocenu ostatka

$$R_n(f) = \int_a^b f(x)dx - \sum_{k=1}^{n} A_k f(x_k).$$

Jednostavnosti radi, pretpostavićemo da $-1 < x_k < 1 (k = 1,\ldots,n)$.

Kako je, na osnovu Cauchy eve formule,

$$f(z) = \frac{1}{2\pi i} \oint_\Gamma \frac{f(\xi)}{\xi - z}\, d\xi = \frac{1}{2\pi} \oint_\Gamma \frac{f(\xi)}{1 - z\bar{\xi}}\, ds \quad (d\xi = i\xi ds),$$

imamo

$$Lf = \frac{1}{2\pi} \oint_\Gamma f(\xi)\, L\left(\frac{1}{1 - z\bar{\xi}}\right) ds, \tag{2.3.8}$$

gde je $Lf = R_n(f)$. Primenom Cauchy-Schwarz ove nejednakosti na (2.8.3) dobijamo

$$|Lf|^2 \leqq \frac{1}{4\pi^2} \oint_\Gamma |f(\xi)|^2 ds \cdot \oint_\Gamma \left|L\left(\frac{1}{1 - z\bar{\xi}}\right)\right|^2 ds . \tag{2.8.4}$$

Kako je $\|f\|^2 = (f,f) = \frac{1}{2\pi} \oint_\Gamma |f(\xi)|^2 ds$,

$$Q = L\left(\frac{1}{1 - z\bar{\xi}}\right) = \sum_{p=0}^{+\infty} L(z^p)\, \bar{\xi}^p$$

i

$$\sigma^2 = \frac{1}{4\pi^2} \oint_\Gamma |Q|^2 ds = \frac{1}{4\pi^2} \oint_\Gamma Q\bar{Q}\, ds = \frac{1}{2\pi} \sum_{p=0}^{+\infty} |L(z^p)|^2 ,$$

pri čemu je iskorišćena osobina ortonormiranosti niza $\{1, z, z^2, \ldots\}$, (2.8.4) se svodi na

$$|Lf|^2 \leqq \frac{1}{2\pi} \|f\|^2 (2\pi)^2 \sigma^2 ,$$

tj.

$$|R_n(f)| \leqq (2\pi)^{1/2} \sigma \|f\| . \tag{2.8.5}$$

Ako pretpostavimo da kvadraturna formula ima algebarski stepen tačnosti m, tada se izraz za σ^2 svodi na

$$\sigma^2 = \frac{1}{2\pi} \sum_{p=m+1}^{+\infty} |L(z^p)|^2 .$$

Za normu $\|f\|$ najčešće se koristi gruba ocena $\|f\| \leq \max_{|z|=1} |f(z)|$. i tada iz (2.8.5) sleduje

$$|R_n(f)| \leqq (2\pi)^{1/2} \sigma \max_{|z|=1} |f(z)| .$$

U radovima [15], [16], [17], Hämmerlin je odredio granice za σ u nekim standardnim kvadraturnim formulama. Tako, na primer, ako je $[a,b] = [-1/2,1/2]$, on je dokazao nejednakosti

$$(2\pi)^{1/2}\sigma < 0.189h^2 \qquad \text{(za uopštenu trapeznu formulu),}$$

$$(2\pi)^{1/2}\sigma < 0.253h^4 \qquad \text{(za uopštenu Simpsonovu formulu).}$$

Ocene za σ u formulama Gaussovog tipa odredjivao je F.Stetter([18], [19]).

7.2.9. Neki metodi za povećavanje tačnosti kvadraturnih formula

Ostatak u kvadraturnoj formuli

$$\text{(2.9.1)} \qquad \int_a^b f(x)dx = \sum_{k=0}^{n} A_k f(x_k) + R_{n+1}(f),$$

po pravilu, se može učiniti proizvoljno malim, ukoliko se uzme dovoljno veliko n. Medjutim, po svaku cenu n ne treba povećavati, s obzirom da se pri numeričkom radu javljaju greške o kojima je bilo reči u prvoj glavi. Izložićemo sada jednu ideju za povećavanje tačnosti formule (2.9.1), pri fiksiranom n, koja se sastoji u sukcesivnom izdvajanju tzv. glavnog dela ostatka.

Posmatrajmo uopštenu trapeznu formulu(videti odeljak 7.2.3)

$$\int_{x_0=a}^{x_n=b} f(x)dx = T_n(f;h) + R_{n+1}(f),$$

gde je $T_n(f;h) = h(\frac{1}{2}f_0 + f_1 + \cdots + f_{n-1} + \frac{1}{2}f_n)$ i $f_i = f(x_i)$ $= f(x_0 + ih)$ $(i = 0,1,\ldots,n)$. Ako $f \in C^2[a,b]$, ostatak se može, kao što je poznato, predstaviti u obliku

$$R_{n+1}(f) = -\frac{(b-a)h^2}{12} f''(\xi) \quad (a < \xi < b).$$

Pod uslovom da je funkcija f dovoljan broj puta diferencijabilna, ideja o sukcesivnom izdvajanju glavnog dela ostatka, može se sprovesti razvijanjem ostatka $R_{n+1}(f)$ u red po koraku h. Dodajući članove ovog reda sumi $T_n(f;h)$ dobijamo niz kvadraturnih formula, čija se tačnost povećava sa brojem dodatih članova.

Definicija 2.9.1. Koeficijenti $B_k (k = 0,1,\ldots)$ u razvoju

$$\frac{x}{e^x - 1} = \sum_{k=0}^{+\infty} \frac{B_k}{k!} x^k \quad (|x| < 2\pi) \tag{2.9.2}$$

nazivaju se Bernoullievi brojevi.

Na osnovu (2.9.2), tj. jednakosti

$$x = (x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots)(B_0 + B_1 x + \frac{1}{2!}B_2 x^2 + \cdots),$$

zaključujemo da je $B_{2i+1} = 0 \, (i=1,2,\ldots)$ i

$$B_0 = 1, \quad B_1 = -\frac{1}{2}, \quad B_2 = \frac{1}{6}, \quad B_4 = -\frac{1}{30}, \quad B_6 = \frac{1}{42},$$

$$B_8 = -\frac{1}{30}, \quad B_{10} = \frac{5}{66}, \quad B_{12} = -\frac{691}{2730}, \quad B_{14} = \frac{7}{6}, \quad \text{itd.}$$

Izrazićemo najpre operator Δ^{-1} pomoću operatora diferenciranja. Kako je $e^{hD} = 1 + \Delta$, imamo

$$\Delta^{-1} = (e^{hD} - 1)^{-1} = \frac{D^{-1}}{h} \sum_{k=0}^{+\infty} \frac{B_k}{k!} (hD)^k,$$

tj.

$$\Delta^{-1} = \frac{1}{h}D^{-1} - \frac{1}{2} + \frac{1}{12}hD - \frac{1}{720}(hD)^3 + \cdots .$$

Kako je, dalje,

$$f_n - f_o = \sum_{i=o}^{n-1}(f_{i+1} - f_i) = \sum_{i=o}^{n-1}\Delta f_i = \Delta\left(\sum_{i=o}^{n-1} f_i\right),$$

$$\Delta^{-1}(f_n - f_o) = \sum_{i=o}^{n-1} f_i ,$$

$$D^{-1}(f_n - f_o) = D^{-1}\Delta\left(\sum_{i=o}^{n-1} f_i\right) = \sum_{i=o}^{n-1} Jf_i = \int_{x_o}^{x_n} f(x)dx ,$$

primenom operatora Δ^{-1} na $f_n - f_o$, dobijamo Euler-Maclaurin-ovu sumacionu formulu

$$\sum_{i=o}^{n-1} f_i = \frac{1}{h}\int_{x_o}^{x_n} f(x)dx - \frac{1}{2}(f_n - f_o) \tag{2.9.3}$$

$$+ \sum_{k=1}^{m} \frac{B_{2k}h^{2k-1}}{(2k)!}(f_n^{(2k-1)} - f_o^{(2k-1)}) + E_m(f),$$

gde je

$$E_m(f) = n\,\frac{B_{2m+2}h^{2m+2}}{(2m+2)!} f^{(2m+2)}(\xi) \qquad (x_o < \xi < x_n).$$

Formula (2.9.3) se može predstaviti u obliku

$$\int_{x_o}^{x_n} f(x)dx = T_n(f;h) - \sum_{k=1}^{m} \frac{B_{2k}h^{2k}}{(2k)!}(f_n^{(2k-1)} - f_o^{(2k-1)}) - hE_m(f),$$

tj.

$$\int_{x_o}^{x_n} f(x)dx = T_n(f;h) - \frac{h^2}{12}(f'_n - f'_o) + \frac{h^4}{720}(f'''_n - f'''_o) - \tag{2.9.4}$$

$$-\frac{B_{2m}h^{2m}}{(2m)!}(f_n^{(2m-1)}-f_o^{(2m-1)})-hE_m(f).$$

Primer 2.8.1. Za m=1, formula (2.9.4) se svodi na

$$\int_{x_o}^{x_1} f(x)dx = T_n(f;h) - \frac{h^2}{12}(f'_n - f'_o) + \frac{nh^5}{720} f^{IV}(\xi) \quad (x_o<\xi<x_n).$$

Poslednju formulu primenimo na izračunavanje vrednosti integrala iz primera 2.3.1 sa $h=0.05$, tj. $n=4$.

Kako je $f'(x)=\frac{1}{2}x^{-1/2}$ i $T_4=0.209680$, imamo

$$\int_{1.0}^{1.2}\sqrt{x}\,dx \cong 0.209680 - \frac{0.05^2}{12}\cdot\frac{1}{2}\left(\frac{1}{\sqrt{1.2}}-1\right) = 0.209689.$$

Euler-Maclaurinova formula ima primenu i kod izračunavanja zbira $S=f_o+f_1+\cdots+f_n$.

Primer 2.9.2. Izračunaćemo zbir

$$S=\frac{1}{11^3}+\frac{1}{13^3}+\frac{1}{15^3}+\cdots+\frac{1}{51^3}$$

uzimajući u formuli (2.9.3) $f(x)=1/x^3$, $h=2$, $n=20$, $x_o=11$, $x_{2o}=51$, $m=3$.

Kako je $f'(x)=-3/x^4$, $f'''(x)=-60/x^6$, $f^{V}(x)=-2520/x^8$, imamo

$$S \cong \frac{1}{2}\int_{11}^{51}\frac{dx}{x^3}+\frac{1}{2}\left(\frac{1}{11^3}+\frac{1}{51^3}\right)+\frac{2}{12}\left(-\frac{3}{51^4}+\frac{3}{11^4}\right)$$

$$-\frac{8}{720}\left(-\frac{60}{51^6}+\frac{60}{11^6}\right)+\frac{32}{42\cdot 6!}\left(-\frac{2520}{51^8}+\frac{2520}{11^8}\right)$$

$$\cong \frac{1}{100}(0.1969999+0.0379427+0.0034077+0.0000376$$

$$+0.0000012)=0.2383891\cdot 10^{-2},$$

tj. $S \cong 0.002383891$.

Pored formule (2.9.3) postoje i druge sumacione formule. Na primer, formula

$$(2.9.5) \quad \sum_{i=0}^{n-1} f_{i+(1/2)} = \frac{1}{h}\int_{x_0}^{x_n} f(x)dx$$

$$- \sum_{k=1}^{m} \frac{(1-2^{1-2k})B_{2k}h^{2k-1}}{(2k)!}(f_n^{(2k-1)} - f_0^{(2k-1)}) + E_m',$$

gde je

$$E_m' = E_m'(f) = -n\,\frac{(1-2^{-1-2m})B_{2m+2}h^{2m+2}}{(2m+2)!} f^{(2m+2)}(\xi) \quad (x_0 < \xi < x_n),$$

koja se naziva druga Euler-Maclaurin ova formula, generališe tzv. pravilo srednje tačke(videti odeljak 7.2.3).

Navedene Euler-Maclaurin ove formule mogu se modifikovati na taj način što bi se izvodi funkcije f u tačkama x_0 i x_n zamenili odgovarajućim diferencnim formulama.

Na primer, ako izvode u tački x_0 zamenimo pomoću formula sa operatorom prednje razlike,tj.

$$h^k f_0^{(k)} = \Delta^k f_0 - \frac{k}{2}\Delta^{k+1} f_0 + \frac{k(3k+5)}{24}\Delta^{k+2} f_0 - \cdots ,$$

a izvode u tački x_n pomoću formula sa operatorom zadnje razlike, tj.

$$h^k f_n^{(k)} = \nabla^k f_n + \frac{k}{2}\nabla^{k+1} f_n + \frac{k(3k+5)}{24}\nabla^{k+2} f_n + \cdots ,$$

iz (2.9.4) sleduje Gregori eva formula

$$\int_{x_0}^{x_n} f(x)dx \sim T_n(f;h) - h\sum_{k=1}^{+\infty} G_k(\nabla^k f_n + (-1)^k \Delta^k f_0),$$

gde su

$$G_1 = \frac{1}{12},\quad G_2 = \frac{1}{24},\quad G_3 = \frac{19}{720},\quad G_4 = \frac{3}{160},\quad G_5 = \frac{864}{60480}, \text{ itd.}$$

Primedba 2.9.1. Koeficijenti G_k mogu se odrediti iz relacije

$$\sum_{k=1}^{n} \frac{1}{n+1-k} G_k = \frac{n}{2(n+1)(n+2)} .$$

7.2.10. Richardsonova ekstrapolacija i Rombergova integracija

Posmatrajmo kvadraturnu formulu

$$(2.10.1)\qquad I = \int_a^b f(x)dx = \sum_k A_k f(x_k) + R(f),$$

u kojoj su tačke $x_k(\in [a,b])$ uzete ekvidistantno sa korakom h. Sa I(h) označimo približnu vrednost integrala I koja se dobija primenom ove kvadraturne formule. Pretpostavimo, dalje, da je ostatak reda $r(r\in N)$, tj. da je oblika

$$(2.10.2)\qquad R(f) = Ch^r + O(h^{r+1}),$$

gde C ne zavisi od h.

Očigledno je da $I(h) \to I$, kada $h \to 0$. Ako za h uzmemo dve vrednosti h_1 i $h_2(h_2<h_1)$ i sa njima odredimo približne vrednosti integrala I(primenom formule (2.10.1)), sledećim postupkom, po pravilu, se može dobiti tačnija vrednost ovog integrala nego što su vrednosti $I(h_1)$ i $I(h_2)$. Naime, iz (2.10.1) i (2.10.2) sleduje

$$I - I(h_1) = Ch_1^r + O(h_1^{r+1})$$

i

$$I - I(h_2) = Ch_2^r + O(h_2^{r+1}).$$

Zanemarivanjem veličina $O(h_1^{r+1})$ i $O(h_2^{r+1})$ u poslednjim jednakostima i eliminacijom parametra C dobijamo formulu

$$(2.10.3)\qquad I \cong \frac{\left(\frac{h_1}{h_2}\right)^r I(h_2) - I(h_1)}{\left(\frac{h_1}{h_2}\right)^r - 1} = I(h_2) + \frac{I(h_2) - I(h_1)}{\left(\frac{h_1}{h_2}\right)^r - 1}\ .$$

koja je poznata kao Richardsonova ekstrapolacija*. U praktičnim računanjima najčešće se uzima $h_2 = h_1/2$ i u tom slučaju formula (2.10.3) se svodi na

$$(2.10.4)\qquad I = \frac{2^r I(h_2) - I(h_1)}{2^r - 1}\ .$$

*) Ovde se, ustvari, prognozira vrednost I(o) na osnovu vrednosti $I(h_1)$ i $I(h_2)$.

Primedba 2.10.1. Ako se kao formula (2.10.1) uzme uopštena trapezna formula ili pravilo srednje tačke imamo $r=2$, dok je kod uopštene Simpsonove formule $r=4$.

Primenom ideje koju daje Richardsonova ekstrapolacija, na Euler-Maclaurinovu sumacionu formulu može se doći do jednog rekurzivnog metoda za numeričku integraciju, koji je poznat kao Rombergova integracija([20]). Kompletna teorija u vezi sa ovim metodom data je u radu [21].

Neka je (2.10.1) uopštena trapezna formula. Tada je za $f \in C^{2m+2}[a,b]$, na osnovu Euler-Maclaurinove sumacione formule

$$(2.10.5) \quad R(f) = C_1h^2 + C_2h^4 + C_3h^6 + \dots + C_mh^{2m}$$
$$- (b-a)\frac{B_{2m+2}h^{2m+2}}{(2m+2)!} f^{(2m+2)}(\xi) \quad (\xi \in (a,b)),$$

gde su B_{2k} Bernoullievi brojevi i

$$C_k = -\frac{B_{2k}}{(2k)!}\left(f^{(2k-1)}(b) - f^{(2k-1)}(a)\right) \quad (k=1,\dots,m).$$

Ako stavimo $h = h_k = (b-a)/2^k$ i odgovarajuću trapeznu aproksimaciju $T_n(f;h)$ $(n=2^k)$ označimo sa $T_k^{(o)}$, tada primena Richardsonove ekstrapolacije na $T_k^{(o)}$ i $T_{k+1}^{(o)}$, saglasno (2.10.4), daje

$$I \cong \frac{2^2T_{k+1}^{(o)} - T_k^{(o)}}{2^2-1}.$$

Desnu stranu poslednje približne jednakosti označimo sa $T_k^{(1)}$. Odredićemo sada grešku u aproksimaciji $I \cong T_k^{(1)}$.

Kako je

$$I - T_k^{(1)} = I - \tfrac{1}{3}(4T_{k+1}^{(o)} - T_k^{(o)}) = \tfrac{4}{3}(I - T_{k+1}^{(o)}) - \tfrac{1}{3}(I - T_k^{(o)}),$$

na osnovu (2.10.5) imamo

$$I - T_k^{(1)} = \frac{4}{3}\left(C_1\left(\frac{h_k}{2}\right)^2 + C_2\left(\frac{h_k}{2}\right)^4 + \dots\right) - \frac{1}{3}(C_1h_k^2 + C_2h_k^4 + \dots)$$

tj.

$$I - T_k^{(1)} = -\frac{1}{4}C_2 h_k^4 + O(h_k^6).$$

Dakle, greška je reda h_k^4, kada $h_k \to o$.

Pod uslovom da je funkcija f dovoljan broj puta diferencijabilna, navedeni postupak sugeriše konstrukciju iterativnog procesa za odredjivanje vrednosti integrala I u obliku

$$T_k^{(m)} = \frac{h_k^2 T_{k+1}^{(m-1)} - h_{k+m}^2 T_k^{(m-1)}}{h_k^2 - h_{k+m}^2} \qquad (m=1,2,\ldots),$$

tj.

$$T_k^{(m)} = \frac{4^m T_{k+1}^{(m-1)} - T_k^{(m-1)}}{4^m - 1} \qquad (m=1,2,\ldots), \tag{2.10.6}$$

s obzirom da je $h_k = (b-a)/2^k$.

Za grešku u aproksimaciji $T_k^{(m)}$ važi ocena

$$I - T_k^{(m)} = (-1)^{m+1} \frac{h_k^{2m+2} B_{2m+2} f^{(2m+2)}(\xi)}{2^{m(m+1)}(2m+2)!} \qquad (a < \xi < b).$$

Na osnovu (2.10.6) može se konstruisati tzv. T-tabela

$$\begin{array}{ccccccc}
T_o^{(o)} & — & T_o^{(1)} & — & T_o^{(2)} & — & T_o^{(3)} \\
 & / & & / & & / & \vdots \\
T_1^{(o)} & — & T_1^{(1)} & — & T_1^{(2)} & — & \\
 & / & & / & \vdots & & \\
T_2^{(o)} & — & T_2^{(1)} & — & & & \\
 & / & \vdots & & & & \\
T_3^{(o)} & — & & & & & \\
\vdots & & & & & &
\end{array}$$

uzimajući $k = o,1,\ldots$ i $m = 1,2,\ldots$. U prvoj koloni ove tabele nalaze se redom približne vrednosti integrala I dobijene primenom trapezne formule sa $h_k = (b-a)/2^k$ $(k = o,1,\ldots)$. Druga kolona ove tabele dobija se na osnovu prve, korišćenjem formule (2.10.6), treća na osnovu druge, itd.

Iterativni proces, definisan sa (2.10.6) predstavlja standardni Rombergov metod za numeričku integraciju. Može se dokazati da nizovi $\{T_k^{(m)}\}_{k\in N_o}$ i $\{T_k^{(m)}\}_{m\in N_o}$ (po kolonama i vrstama u T-tabeli) konvergiraju ka I. Kod praktične primene Rombergove integracije, iterativni proces (2.10.6) se najčešće prekida kada je $|T_o^{(m)}-T_o^{(m-1)}|\leqq\varepsilon$, gde je ε unapred data dozvoljena greška, i tada se uzima $I\cong T_o^{(m)}$.

Primer 2.10.1. Primenom Rombergove integracije približno ćemo izračunati integral $\int_o^1 e^{-x^2}dx$. Pri računanju koristićemo se približnim vrednostima podintegralne funkcije u tačkama $x_i=i/8$ $(i=o,1,\ldots,8)$

$$f_o=1.000000,\quad f_1=0.984497,\quad f_2=0.939413,$$

$$f_3=0.868815,\quad f_4=0.778801,\quad f_5=0.676634,$$

$$f_6=0.569783,\quad f_7=0.465044,\quad f_8=0.367880.$$

Uvedimo, prethodno oznaku $A_k=T_k^{(o)}/h_k$ $(k=o,1,\ldots)$.

Primenom trapezne formule sa $h_k=1/2^k$ $(k=o,1,2,3)$ imamo redom

$$A_o=\tfrac{1}{2}f_o+\tfrac{1}{2}f_8=0.683940,\qquad T_o^{(o)}=h_oA_o=0.683940;$$

$$A_1=A_o+f_4=1.462741,\qquad T_1^{(o)}=h_1A_1=0.7313705;$$

$$A_2=A_1+f_2+f_6=2.971937,\qquad T_2^{(o)}=h_2A_2=0.7429842;$$

$$A_3=A_2+f_1+f_3+f_5+f_7=5.966927,\qquad T_3^{(o)}=h_3A_3=0.7458658,$$

a zatim primenom formule (2.10.6) na ove rezultate dobijamo T-tabelu

0.683940	0.7471806	0.7468337	0.7468241
0.7313705	0.7468554	0.7468243	
0.7429842	0.7468263		
0.7458658			

Dakle, dati integral je približno jednak 0.746824.

Rombergov metod, koji je prezentiran, sastoji se u sukcesivnom polovljenju intervala integracije, s obzirom da je $h_k = (b-a)/n_k$ i $n_k = 2^k (k=o,1,\ldots)$. Ovaj metod se može modifikovati u tom smislu da se za $\{n_k\}_{k\in N_o}$ odabere neki drugi podniz prirodnih brojeva, ali takav da je $n_{k+1}/n_k > c > 1$. Na primer, u radu [21], za $\{n_k\}$ je korišćen niz $\{1,2,3,6,9,18,27,54,\ldots\}$. Tada je

$$T_k^{(m)} = \begin{cases} \dfrac{3^m T_{k+1}^{(m-1)} - T_k^{(m-1)}}{3^m - 1} & (m \text{ parno}), \\[2ex] \dfrac{4\cdot 3^m T_{k+1}^{(m-1)} - 3T_k^{(m-1)}}{4\cdot 3^m - 3} & (m \text{ neparno, } k \text{ parno}), \\[2ex] \dfrac{3^{m+1} T_{k+1}^{(m-1)} - 4T_k^{(m-1)}}{3^{m+1} - 4} & (m \text{ neparno, } k \text{ neparno}). \end{cases}$$

Pretpostavimo sada da formula (2.10.1) predstavlja pravilo srednje tačke, tj.

$$I = \int_a^b f(x)dx = M_n(f;h) + R(f),$$

gde je $M_n(f;h) = h\sum_{i=o}^{n-1} f(a+\frac{2i+1}{2}h)$ i $h = \frac{b-a}{n}$.

Ako $f\in C^{2m+2}[a,b]$, na osnovu druge Euler-Maclaurinove sumacione formule (2.9.5), ostatak R(f) se može predstaviti u obliku

$$R(f) = -\frac{1}{2}C_1h^2 - \frac{7}{8}C_2h^4 - \frac{31}{32}C_3h^6 - \dots - \left(1-\frac{2}{4^m}\right)C_mh^{2m}$$

$$+ (b-a)\left(1-\frac{2}{4^{m+1}}\right)\frac{B_{2m+2}h^{2m+2}}{(2m+2)!}f^{(2m+2)}(\xi) \qquad (a<\xi<b),$$

gde su koeficijenti $C_k(k=1,\dots,m)$ isti kao u formuli (2.10.5). Ako stavimo $n=2^k$, tj. $h=h_k=(b-a)/2^k$ i aproksimaciju $M_n(f;h)$ označimo sa $M_k^{(o)}$, koristeći se istim postupkom kao kod uopštene trapezne formule, možemo konstruisati iterativni proces

$$M_k^{(m)} = \frac{4^mM_{k+1}^{(m-1)} - M_k^{(m-1)}}{4^m-1} \qquad (m=1,2,\dots). \tag{2.10.4}$$

Uzimajući $k=o,1,\dots$ i $m=1,2,\dots$ na osnovu (2.10.7) možemo formirati tzv. M-tabelu

$$\begin{tikzcd}
M_o^{(o)} \arrow[r, dash] & M_o^{(1)} \arrow[r, dash] & M_o^{(2)} \arrow[r, dash] & M_o^{(3)} \\
M_1^{(o)} \arrow[r, dash] \arrow[ur, dash] & M_1^{(1)} \arrow[r, dash] \arrow[ur, dash] & M_1^{(2)} \arrow[r, dash] \arrow[ur, dash] & \vdots \\
M_2^{(o)} \arrow[r, dash] \arrow[ur, dash] & M_2^{(1)} \arrow[r, dash] \arrow[ur, dash] & \vdots & \\
M_3^{(o)} \arrow[r, dash] \arrow[ur, dash] & \vdots & & \\
\vdots & & &
\end{tikzcd}$$

Primetimo da elementi T i M-tabele stoje u odredjenoj vezi. Naime, ako pretpostavimo da smo izračunali $T_{k-1}^{(o)}$ i $M_{k-1}^{(o)}$, tada se sledeći element u prvoj koloni T-tabele može odrediti pomoću

$$T_k^{(o)} = \frac{1}{2}\left(T_{k-1}^{(o)} + M_{k-1}^{(o)}\right).$$

Naravno, važi i opštija jednakost

$$T_k^{(m)} = \frac{1}{2}\left(T_{k-1}^{(m)} + M_{k-1}^{(m)}\right).$$

Korišćenjem poslednje jednakosti, (2.10.6) postaje

$$T_k^{(m)} = M_k^{(m-1)} + \frac{(2\cdot 4^{m-1} - 1)(T_k^{(m-1)} - M_k^{(m-1)})}{4^m - 1} \qquad (m=1,2,\ldots).$$

M-tabela sama za sebe nema neki praktični značaj, medjutim, ona se može uspešno koristiti za konstrukciju T-tabele, zahvaljujući napred navedenim jednakostima.

7.2.11. Konvergencija kvadraturnih procesa

U ovom odeljku ispitaćemo uslove za konvergenciju kvadraturnih procesa, tj. uslove pod kojima ostatak $R_n(f) \to 0$, kada $n \to +\infty$.

Neka je dat niz kvadraturnih formula

$$(2.11.1) \qquad I(f) = \int_a^b f(x)dx = K_n f + R_n(f) \qquad (n=1,2,\ldots),$$

gde su

$$(2.11.2) \qquad K_n = \sum_{k=1}^{n} A_k^{(n)} f(x_k^{(n)}),$$

$A_k^{(n)}$ težinski koeficijenti(konstantni za fiksirano n) i $x_k^{(n)}$ apscise koje zadovoljavaju uslov

$$a \leqq x_1^{(n)} < x_2^{(n)} < \ldots < x_n^{(n)} \leqq b.$$

Teorema 2.11.1. Potreban i dovoljan uslov da niz $\{K_n f\}_{n\in N}$ konvergira ka $I(f)$ za svako $f \in C[a,b]$ je da on konvergira za svaki algebarski polinom i da postoji pozitivna konstanta M takva da je

$$(2.11.3) \qquad \sum_{k=1}^{n} |A_k^{(n)}| \leqq M \qquad (n=1,2,\ldots).$$

Dokaz. Posmatrajmo linearnu funkcionelu K_n, definisanu pomoću (2.11.2). Kako je za svako $f \in C[a,b]$

$$|K_n f| \leqq \sum_{k=1}^{n} |A_k^{(n)}| \cdot |f(x_k^{(n)})| \leqq \left(\sum_{k=1}^{n} |A_k^{(n)}| \right) \|f\| ,$$

gde je $\|f\| = \max_{a \leqq x \leqq b} |f(x)|$, zaključujemo da je funkcionela K_n ograničena na $C[a,b]$.

Dakle, imamo

$$(2.11.4) \qquad |K_n f| \leqq \|K_n\| \cdot \|f\| \qquad (\forall f \in C[a,b])$$

i

$$(2.11.5) \qquad \|K_n\| \leqq \sum_{k=1}^{n} |A_k^{(n)}| .$$

Odredimo sada normu $\|K_n\|$. U prostoru $C[a,b]$ izaberimo funkciju $\bar{f}$, koja je odredjena sa

$$\bar{f}(x_k^{(n)}) = \operatorname{sgn} A_k^{(n)},$$

linearna na $[x_k^{(n)}, x_{k+1}^{(n)}]$ $(k = 1,\ldots,n-1)$ i konstantna na $[a, x_1^{(n)}]$ i $[x_n^{(n)}, b]$. Tada je $\|\bar{f}\| = 1$ i

$$K_n \bar{f} = \sum_{k=1}^{n} A_k^{(n)} \operatorname{sgn} A_k^{(n)} = \sum_{k=1}^{n} |A_k^{(n)}| ,$$

pa iz (2.11.4), za $f = \bar{f}$, sleduje $\|K_n\| \geqq \sum_{k=1}^{n} |A_k^{(n)}|$, što zajedno sa (2.11.5) daje $\|K_n\| = \sum_{k=1}^{n} |A_k^{(n)}|$.

Prema tome, kako je

1° Niz $\{\|K_n\|\}_{n \in N}$ ograničen(na osnovu (2.11.3));

2° $\lim_{n \to +\infty} K_n f = I(f)$ za svaki algebarski polinom;

3° Skup polinoma svuda gust u $C[a,b]$,

na osnovu Banach-Steinhausove teoreme (videti odeljak 2.2.1) zaključujemo da važi tvrdjenje teoreme 2.11.1.

Teorema 2.11.2. Ako su koeficijenti $A_k^{(n)} \geq 0$ ($k=1,\ldots,n$; $n=1,2,\ldots$) potreban i dovoljan uslov da niz $\{K_n f\}_{n\in N}$ konvergira ka $I(f)$ za svako $f \in C[a,b]$ je da on konvergira za svaki algebarski polinom.

Dokaz. Kako je $\lim_{n\to+\infty} K_n f = I(f)$, stavljajući $f(x)=1$ u (2.11.1), dobijamo

$$b-a = \sum_{k=1}^{n} A_k^{(n)},$$

što znači da je uslov (2.11.3) ispunjen sa $M=b-a$, čime je dokaz završen.

Iz teoreme 2.11.1 sleduje konvergencija Gauss ovih kvadraturnih procesa, s obzirom da je $A_k^{(n)} \geq 0$ (videti primedbu 2.4.1).

Teorema 2.11.1 nije primenljiva na niz Newton-Cotes ovih formula (videti (2.2.4)), jer su već pri $n=8$ neki od težinskih koeficijenata negativni. Napomenimo, da ovaj niz, zaista, ne konvergira za svako $f \in C[a,b]$ (videti [22] i [23]).

7.3. LITERATURA

1. Š.E.MIKELADZE: Numerical methods of mathematical analysis. Moskva, 1953 (in Russian).

2. B.P.DEMIDOVICH, I.A.MARON: Computational Mathematics. Moscow, 1973.

3. K.S.K.IYENGAR: Note on an inequality. Math.Student 6(1938), 75-76.

4. D.S.MITRINOVIĆ (with P.M.VASIĆ): Analytic Inequalities. Berlin-Heidelberg-New York, 1970.

5. P.M.VASIĆ and G.V.MILOVANOVIĆ: On an inequality of Iyengar. Univ.Beograd. Publ.Elektrotehn.Fak.Ser.Mat.Fiz. No 544 - No 576(1976), 18-24.

6. G.V.MILOVANOVIĆ and J.E.PEČARIĆ: Some considerations of Iyengar's inequality and some related applications. ibid. No 544 - No 576(1976), 166-170.

7. M.ABRAMOWITZ and I.A.STEGUN: Handbook of Mathematical Function. National Burean of Standards Applied Mathematics Series 55, 1964.

8. K.H.HUEBNER: The Finite Element Method for Engineers. New York-Sydney--Toronto, 1975.

9. P.J.DAVIS, P.RABINOWITZ: Methods of Numerical Integration. New York,1975.

10. I.S.BEREZIN, N.P.ŽIIKOV: Numerička analiza - numeričke metode. Beograd, 1963.

11. J.V.USPENSKY: Sur les valeurs asymptotiques des Coefficients de Cotes. Bull.Amer.Math.Soc. 31(1925), 145-156.

12. G.V.MILOVANOVIĆ and J.E.PEČARIĆ: On generalization of the inequality of A.Ostrowski and some related applications. Univ.Beograd. Publ.Elektrotehn. Fak.Ser.Mat.Fiz. No 544 - No 576(1976), 155-158.

13. D.S.MITRINOVIĆ, D.Ž.DJOKOVIĆ: Polinomi i matrice. Beograd, 1975.

14. J.L.ULLMAN: A class of weight functions that admit Tchebycheff quadrature. Michigan Math.J. 13(1966), 417-423.

15. G.HÄMMERLIN: Über ableitungsfreie Schranken für Quadraturfehler. Numer. Math. 5(1963), 226-233.

16. G.HÄMMERLIN: Über ableitungsfreie Schranken für Quadraturfehler II. Ergänzungen und Möglichkeiten zur Verbesserung. ibid. 7(1965), 232-237.

17. G.HÄMMERLIN: Zur Abschätzung von Quadraturfehlevn für analytische Funktionen. ibid. 8(1966), 334-344.

18. F.STETTER: Ableitungsfreie Schranken für die numerische Integration. Computing 2(1967), 257-262.

19. F.STETTER: Error bounds for Gauss-Chebyshev quadrature. Math.Comp. 22(1968), 657-659.

20. W.ROMBERG: Vereinfachte numerische Integration. Norske Vid.Selsk.Forh. (Trondheim) 28(1955), 30-36.

21. F.L.BAUER, H.RUTISHAUSER, E.L.STIEFEL: New aspects in numerical quadrature (in Experimental Arithmetic, High Speed Computing and Mathematics). Amer.Math.Soc.Providence, R.I. 1963.

22. L.COLLATZ: Funktionalanalysis und Numerische Mathematik. Berlin-Heidelberg-New York, 1968.

23. H.BRASS: Ein Gegenbeispiel zum Newton-Cotes-Verfahren. ZAMM 57(1977),609.

24. F.G.LETHER: Error bounds for fully simmetric quadrature rules. SIAM J. Numer.Anal. 11(1974), 3-9.

25. S.HABER: Midpoint Quadrature Formulas. Math.Comp. 21(1967), 719-721.