

Dr GRADIMIR V. MILOVANOVIĆ

NUMERIČKA ANALIZA

I deo

345679

Naučna Knjiga

Dr GRADIMIR V. MILOVANOVIC

NUMERICKA ANALIZA

I DEO

DRUGO IZDANJE

Naučna Knjiga

B E O G R A D, 1988.

Dr Gradimir V. Milovanović
NUMERIČKA ANALIZA, I deo

Izdavač:

IRO „Naučna knjiga“
Beograd, Uzun-Mirkova 5

Recenzenti:

Prof. dr Petar M. Vasić
Prof. dr Dobrilo Tošić

Za izdavača:

Dr Blažo Perović

Urednik:

Nikola Dončev

Tehnički urednik:

Gordana Krstić

Korice:

Miloš Majstorović

Tiraž: 1000 primeraka

ISBN 86-23-20081-0

Štampa: GRO „Prosveta“ – Niš,
Ul. Vojvode Gojka 14

Sadržaj

PREDGOVOR

VII

1.	UVOD U NUMERIČKU MATEMATIKU	1
1.1.	ZADATAK NUMERIČKE MATEMATIKE	1
1.2.	ELEMENTI TEORIJE GREŠAKA	7
1.2.1.	Klasifikacija i analiza grešaka u numeričkom rešenju	7
1.2.2.	Prostiranje grešaka	11
1.2.3.	Statistički prilaz u oceni grešaka	23
1.3.	REKURZIVNA IZRAČUNAVANJA I SUMIRANJE	25
1.3.1.	Diferencne jednačine i rekurzivna izračunavanja	26
1.3.2.	Izračunavanje vrednosti nekih elementarnih funkcija	37
1.3.3.	Izračunavanje vrednosti polinoma	42
1.3.4.	Sumiranje redova i ubrzavanje konvergencije	48
1.3.5.	Elementi teorije verižnih razlomaka	68
1.3.6.	Razvoj racionalne funkcije u verižni razlomak	65
1.3.7.	Algoritmi za izračunavanje verižnih razlomaka	68
1.3.8.	Asimptotski razvoji	72
1.4.	LITERATURA	78
2.	PROSTORI. OPERATORI. ORTOGONALNI POLINOMI	81
2.1.	LINEARNI PROSTORI	81
2.1.1.	Linearni prostor	81
2.1.2.	Banachov prostor	84
2.1.3.	Hilbertov prostor	86
2.1.4.	Ortogonalni sistemi u Hilbertovom prostoru	89
2.2.	ORTOGONALNI POLINOMI	93
2.2.1.	Moment-funkcionela i ortogonalnost	93
2.2.2.	Opšte osobine ortogonalnih polinoma	97
2.2.3.	Verižni razlomak pridružen rekurentnoj relaciji	106
2.2.4.	Osobine ortogonalnih polinoma sa parnom težinskom funkcijom	110
2.2.5.	Konstrukcija ortogonalnih polinoma	112

2.2.6.	Čebiševljev algoritam	112
2.2.7.	Stieltjesova procedura	116
2.2.8.	Klasični ortogonalni polinomi	117
2.2.9.	Osobine klasičnih ortogonalnih polinoma	121
2.2.10.	Odredjivanje norme i dva najstarija koeficijenta klasičnih ortogonalnih polinoma	126
2.2.11.	Formule za diferenciranje i integraciju	130
2.2.12.	Funkcija generatrise	132
2.2.13.	Specijalni slučajevi klasičnih ortogonalnih polinoma	135
2.2.14.	Ekstremalna svojstva Čebiševljevih polinoma	149
2.2.15.	Diskretni ortogonalni polinomi	154
2.2.16.	Polinomi ortogonalni na polukrugu	159
2.3.	OPERATORI	161
2.3.1.	Linearni operatori	161
2.3.2.	Matrica linearnog operatora na konačno-dimenzionalnim prostorima	167
2.3.3.	Bilinearni i n-linearni operatori	168
2.3.4.	Frechetova diferenciranja	169
2.3.5.	Taylorova formula	172
2.4.	LITERATURA	174

3. OPŠTA TEORIJA ITERATIVNIH PROCESA 176

3.1.	REŠAVANJE OPERATORSKIH JEDNAČINA	176
3.1.1.	Osnovne napomene o rešavanju operatorskih jednačina	176
3.1.2.	Banachov stav o nepokretnoj tački	178
3.1.3.	Iterativni procesi za rešavanje običnih jednačina	181
3.2.	KARAKTERISTIKE PROCESA	
3.2.1.	Red konvergencije iterativnih procesa	186
3.2.2.	Aitkenov Δ^2 metod	190
3.2.3.	O metodama bliskim Aitkenovom metodu	195
3.2.4.	Dva opšta metoda za ubrzavanje konvergencije iterativnih procesa	197
3.2.5.	R - red konvergencije iterativnih procesa	200
3.3.	LITERATURA	202

4. NUMERIČKI METODI U LINEARNOJ ALGEBRI 203

4.1.	ELEMENTI MATRIČNOG RAČUNA	203
4.1.1.	Operacije sa matricama razbijenim na blokove	203
4.1.2.	LR faktorizacija kvadratne matrice	207
4.1.3.	Sopstveni vektori i sopstvene vrednosti matrica	211
4.1.4.	Specijalne matrice i njihove osobine	213
4.1.5.	Jordanov kanonički oblik	215
4.1.6.	Norme vektora i matrica	217
4.1.7.	Konvergencija matričnih nizova i redova	222

4.2.	DIREKTNI METODI	226
4.2.1.	Uvodne napomene	
4.2.2.	Gaussov metod eliminacije i Gauss-Jordanov metod	228
4.2.3.	Inverzija matrica	237
4.2.4.	Faktorizacioni metodi	238
4.2.5.	Metod ortogonalizacije	244
4.2.6.	Analiza greške i slabouslovljeni sistemi	245
4.3.	ITERATIVNI METODI	248
4.3.1.	Načini formiranja iterativnih metoda	249
4.3.2.	Metod proste iteracije	251
4.3.3.	Gauss-Seidelov metod	263
4.3.4.	Opšte napomene o relaksacionim metodima	270
4.3.5.	Metod sukcesivne gornje relaksacije	271
4.3.6.	Čebiševljev semi-iterativni metod	279
4.3.7.	Gradijentni metodi	282
4.3.8.	Iterativni metodi za inverziju matrica	287
4.4.	PROBLEM SOPSTVENIH VREDNOSTI	290
4.4.1.	Lokalizacija sopstvenih vrednosti	290
4.4.2.	Metodi za određivanje karakterističnog polinoma	292
4.4.3.	Metodi za dominantne sopstvene vrednosti	300
4.4.4.	Metodi za subdominantne sopstvene vrednosti	305
4.4.5.	Jacobiev metod	312
4.4.6.	Givensov i Hauseholderov metod	319
4.4.7.	Problem sopstvenih vrednosti za simetrične trodijagonalne matrice	325
4.4.8.	LR i QR algoritmi	328
4.5.	LITERATURA	334

5.	NELINEARNE JEDNAČINE I SISTEMI	338
5.1.	NELINEARNE JEDNAČINE	
5.1.1.	Osnovne napomene	338
5.1.2.	Newtonov metod	338
5.1.3.	Newtonov metod za višestruke nule	345
5.1.4.	Metod sećice	347
5.1.5.	Metod polovljenja intervala	351
5.1.6.	Scröderov razvoj	352
5.1.7.	Metodi višeg reda	355
5.1.8.	Newtonov metod u intervalnoj aritmetici	365
5.2.	SISTEMI NELINEARNIH JEDNAČINA	373
5.2.1.	Uvodne napomene	373
5.2.2.	Metod Newton-Kantorovića	374
5.2.3.	Gradijentni metod	387
5.3.	ALGEBARSKE JEDNAČINE	390
5.3.1.	Uvodne napomene	390
5.3.2.	Granice korena algebarskih jednačina	394

-
- 5.3.3. Bernoulliev metod 399
 - 5.3.4. Dva metoda trećeg reda 406
 - 5.3.5. Newton-Hornerov metod 410
 - 5.3.6. Bairstowljev metod 413
 - 5.3.7. Metodi za simultano određivanje korena 417
- 5.4. LITERATURA 430

Predgovor

Ovaj udžbenik numeričke analize je proistekao iz predavanja koja je autor poslednjih desetak godina držao studentima redovnih i poslediplomskih studija na Elektronskom, Filozofskom (grupa za matematiku), Mašinskom i Gradjevinskom fakultetu u Nišu i Prirodno-matematičkom fakultetu u Kragujevcu, u okviru predmeta Numerička analiza i Numerička matematika. Kao osnov za izradu rukopisa ove knjige poslužila je ranija autorova knjiga: NUMERIČKA ANALIZA, I deo, Univerzitet u Nišu, Niš, 1979. U odnosu na prvo izdanje ove knjige, koje se pojavilo 1985. godine, u ovom izdanju učinjene su neznatne izmene u tekstu i ispravljene uočene štamparske greške.

Knjiga je podeljena u pet glava. 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 poglavljiju svake glave dat je spisak citirane i korišćene literature.

U prvoj glavi, koja ima uvodni karakter, daje se pored analize grešaka i poglavlje o rekurzivnim izračunavanjima i sumiranju, gde su posebno tretirani numerički aspekti rekurentnih relacija, ubrzavanje konvergencije redova, ekonomične šeme za izračunavanje vrednosti polinoma, teorija verižnih razlomaka i algoritmi za njihovo efikasno izračunavanje. Takođe, i odeljak o asimptotskim razvojima je ovde uključen.

U drugoj glavi je, pored poglavlja koje se odnosi na linearne prostore i operatore, uključeno i obimnije poglavlje, sa 16 odeljaka, u kome se tretira savremena teorija ortogonalnih polinoma, koja se široko primenjuje u numeričkoj matematici. Ovde su dati i najnoviji rezultati konstruktivne teorije ortogonalnih polinoma.

U trećoj glavi razmatra se opšta teorija iterativnih procesa za rešavanje operatorskih jednačina u Banachovom prostoru. Posebno poglavlje je posvećeno karakteristikama iterativnih procesa i metodima za ubrzavanje konvergencije. U posebnom odeljku uvodi se koncept R-reda konvergencije iterativnih procesa.

Četvrta glava je najobimnija. Pored elemenata matričnog računa, sadrži i tri poglavlja koja se odnose na direktnе metode u linearnoj algebri, iterativne

VIII PREDGOVOR

metode za rešavanje sistema linearnih jednačina i inverziju matrica, i poglavlje u kome se tretiraju problemi sopstvenih vrednosti matrica i izlažu metodi za rešavanje, uključujući i najmodernije postupke transformacije i faktorizacije, kakav je, na primer, QR algoritam.

U petoj glavi se izučavaju metodi za rešavanje nelinearnih jednačina, sistema nelinearnih jednačina, i na kraju, metodi za rešavanje numeričkih algebraških jednačina. Poseban tretman u trećem poglavlju posvećen je danas veoma popularnim metodama za simultano određivanje nula polinoma. Izlaže se pregled ovih metoda razvijenih u poslednjih dvadesetak godina. Korišćenjem koncepta R-reda konvergencije pokazano je kako se mogu tretirati konvergentna svojstva ubrzanih verzija osnovnih metoda.

Mada je knjiga namenjena prvenstveno studentima tehničkih i prirodno-matematičkih fakulteta, ona će biti od koristi i svima onima koji se bave numeričkom analizom ili koriste numeričke metode u svojim istraživanjima. Posebno, knjiga će biti korisna programerima u računskim centrima.

Svima onima koji su na neki način pomogli autoru u toku izrade rukopisa za ovu knjigu, autor se najtoplje zahvaljuje.

U Nišu, 28.09.1988.

Gradimir V. Milovanović

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 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, pri čemu su već za jednačine trećeg stepena dobro poznate Cardanoove formule prilično komplikovane. Međutim u praksi se vrlo često sreće problem rešavanja numeričkih algebarskih jednačina* visokog stepena. Takođe, često se javlja i problem rešavanja sistema linearnih algebarskih jednačina sa stotinu, pa i hiljadu nepoznatih, gde klasični aparat linearne algebre postaje neupotrebljiv. Standardna teorija diferencijalnih jednačina omogućava rešavanje (nalaženje opštih rešenja) samo nekih uskih klasa diferencijalnih jednačina. Međutim, u praksi se javljaju Cauchyevi i konturni problemi sa diferencijalnim jednačinama koje ne pripadaju ovim klasama. Posebno važnu ulogu imaju problemi koji se sude na rešavanje parcijalnih diferencijalnih jednačina.

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

* Algebarska jednačina sa numeričkim koeficijentima.

ovo prostim primerom.

Neka je potrebno odrediti pozitivan koren jednačine

$$x^2 - a = 0 \quad (a > 0).$$

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

Formirajmo sada niz $\{x_k\}$, pomoću

$$x_0 = a, \quad x_{k+1} = \frac{1}{2}(x_k + \frac{a}{x_k}) \quad (k = 0, 1, \dots).$$

Može se pokazati da niz $\{x_k\}$ konvergira vrednosti \sqrt{a} , kada $k \rightarrow +\infty$. Pri nekom konačnom n imamo $x \approx x_n$.

Ovaj postupak, kao što vidimo, omogućava nalaženje rešenja postavljenog problema sa tačnošću koja zavisi od a i n . Kako se u praktičnim izračunanjima zahteva određena tačnost ϵ , to se za x usvaja ona vrednost x_n koja zadovoljava uslov $|x - x_n| \leq \epsilon$.

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

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

odakle, uzimanjem dva, odnosno tri člana u razvoju, dobijamo

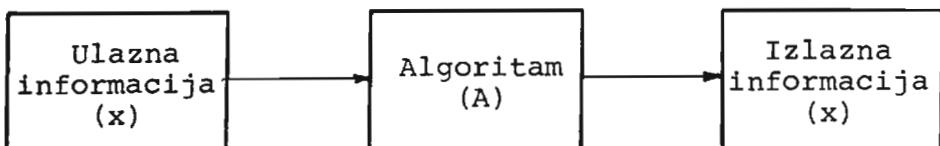
$$\sin 0.5 \approx 0.5 - \frac{0.5^3}{3!} = 0.479167$$

i

$$\sin 0.5 \approx 0.5 - \frac{0.5^3}{3!} + \frac{0.5^5}{5!} = 0.479427.$$

Tačna vrednost za $\sin 0.5$ sa šest decimala je 0.479425. Greške koje ovde nastaju 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 na- ročito važno kod primene računara s obzirom na cenu mašinskog vremena. Izbor optimalnog algoritma u prethodnom smislu pred- stavlja 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 ra- čunara. Kako računske mašine najčešće izvode samo četiri osno- vne računske operacije (sabiranje, oduzimanje, množenje i de- ljenje), to numerički metodi moraju biti takvi da se svode na konačan niz takvih operacija.

Na osnovu prethodnog, pogrešno bi bilo izvesti zaključak da se numerička matematika počela razvijati sa pojavom račun- skih mašina. Poznato je, naime, da se u delima matematičara prošlih stoljeća nalazi niz približnih metoda koji danas čine osnovu numeričkoj matematici. Međutim, nagli razvoj računske tehnike posle II svetskog rata uslovio je brzi i sistematski razvoj numeričke matematike. Tako na primer, razradjen je niz 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 kontur- nih (graničnih) problema u teoriji diferencijalnih jednačina (običnih i parcijalnih), metodi u teoriji integralnih i integro-diferencijalnih jednačina, kao i metodi za rešavanje aps-

traktnih operatorskih jednačina. Poslednjih godina dosta se radi na metodama za rešavanje graničnih problema u jednoj neiscrpoj oblasti - teoriji nelinearnih parcijalnih jednačina. Takođe, vidno mesto zauzimaju i metodi numeričke integracije.

Medju numeričkim metodama posebno su interesantni iterativni metodi, s obzirom da su veoma 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.

Problemima praktične realizacije algoritama bavi se 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.

Mada se radovi iz oblasti numeričke matematike objavljaju i u časopisima koji tretiraju opštu problematiku, ipak se u poslednjih dvadesetak godina u svetu pojavio veliki broj specijalizovanih časopisa za numeričku matematiku kao što su*:

- ACM Trans. Math. Software Association for Computing Machinery. Transactions on Mathematical Software. (New York),*
- Anal. Numér. Théor. Approx. L'Analyse Numérique et la Théorie de L'Approximation. (Cluj-Napoca),*
- Apl. Mat. Československá Akademie Věd. Aplikace Matematiky. (Prague),*
- Appl. Math. Comput. Applied Mathematics and Computation. (New York),*
- Appl. Math. Optim. Applied Mathematics and Optimization. An International Journal. (New York),*
- Applicable Anal. Applicable Analysis. An International Journal. (New York),*
- BIT BIT; Nordisk Tidskrift for Informationsbehandling (BIT) (Copenhagen),*
- Calcolo Calcolo. (Rome),*

* Naslovi časopisa su dati po abecednom redu i to onako kako se citiraju u referativnom časopisu Mathematical Reviews.

- Čisl. Metody Meh. Splošn. Sredy Akademija Nauk SSSR. Sibirskoe Otdelenie. Vyčislitel'nyi Centr. Čislennye Metody Mehaniky Splošnoi Sredy.* (Novosibirsk),
- Comput. J. The Computer Journal.* (London),
- Comput. Math. Appl. Computers & Mathematics with Applications.* An International Journal. (Oxford),
- Comput. Methods Appl. Mech. Engrg. Computer Methods in Applied Mechanics and Engineering.* (Amsterdam),
- Comput. Surveys Computing Surveys.* ACM. Association for Computing Machinery. (New York),
- Computing Computing. Archiv für Informatic und Numerik.* (Vienna),
- Differencial'nye Uravnenija i Vyčisl. Mat. Differencial'nye Uravnenija i Vyčislitel'naja Matematika.* (Saratov),
- Funct. Approx. Comment. Math. Uniwersytet im. Adama Mickiewicza w Poznańu Instytut Matematyki. Functiones et Approximatio Commentarii Mathematici.* (Poznań),
- IBM J. Res. Develop. International Business Machines Corporation. Journal of Research and Development.* (New York),
- IMA J. Numer. Anal. IMA Journal of Numerical Analysis.* (London),
- Internat. J. Comput. Math. International Journal of Computer Mathematics. Section A. Programming Theory and Methods. Section B. Computational Methods.* (New York),
- Internat. J. Numer. Methods Engrg. International Journal for Numerical Methods in Engineering.* (London),
- J. Algorithms Journal of Algorithms.* (New York),
- J. Approx. Theory Journal of Approximation Theory.* (New York),
- J. Assoc. Comput. Mach. Journal of the Association for Computing Machinery.* (New York),
- J. Comput. Appl. Math. Journal of Computational and Applied Mathematics.* (Antwerp),
- J. Comput. Chem. Journal of Computational Chemistry.* (New York),
- J. Comput. Phys. Journal of Computational Physics.* (New York),
- J. Comput. System Sci. Journal of Computer and System Sciences.* (New York),
- J. Optim. Theory Appl. Journal of Optimization Theory and Applications.* (New York),
- Math. Comp. Mathematics of Computation.* (Providence, R.I.),
- Metody Vyčisl. Leningradskii Ordena Lenina Gosudarstvennyi Universitet im. A.A. Zdanova. Kafedra Vyčislitel'noi Matematiki i Vyčislitel'nyi Centr. Metody Vyčislenii.* (Leningrad),

 1. UVOD U NUMERIČKU MATEMATIKU

- Numer. Funct. Anal. Optim.* Numerical Functional Analysis and Optimization. (New York),
- Numer. Math.* Numerische Mathematik. (Berlin),
- Numer. Math. J. Chinese Univ.* Numerical Mathematics. A Journal of Chinese Universities. (Nanking),
- Prikl. Mat. i Programmirovaniye* Akademija Nauk Moldavskoi SSR. Institut Matematiki s Vyčislitel'nym Centrom. Prikladnaja Matematika i Programmirovaniye. (Kishinev),
- RAIRO Anal. Numér.* RAIRO Analyse Numérique. (Paris),
- SIAM J. Algebraic Discrete Methods* Society for Industrial and Applied Mathematics. Journal on Algebraic and Discrete Methods. (Philadelphia, Pa.),
- SIAM J. Appl. Math.* SIAM Journal on Applied Mathematics. (Philadelphia, Pa.),
- SIAM J. Comput.* SIAM Journal on Computing. (Philadelphia, Pa.),
- SIAM J. Numer. Anal.* SIAM Journal on Numerical Analysis. (Philadelphia, Pa.),
- SIAM J. Sci. Statis. Comput.* Society for Industrial and Applied Mathematics, Journal on Scientific and Statistical Computing. (Philadelphia, Pa.),
- SIAM Rev.* SIAM Review. A Publication of the Society for Industrial and Applied Mathematics. (Philadelphia, Pa.),
- Vestnik Moskov. Univ. Ser. XV Vyčisl. Mat. Kibernet.* Vestnik Moskovskogo Universiteta. Serija XV. Vyčislitel'naja Matematika i Kibernetika. (Moscow),
- Vyčisl. Mat. i Vyčisl. Tehn. (Kharkov)* Fiziko-Tehničeskii Institut Nizkih Temperatur Akademii Nauk Ukrainskoi SSR. Vyčislitel'naja Matematika i Vyčislitel'naja Tehnika. (Kharkov),
- Vyčisl. Metody i Programmirovaniye* Moskovskii Universitet. Sbornik Rabot Vyčislitel'nogo Centra Moskovskogo Universiteta. Vyčislitel'nye Metody i Programmirovaniye. (Moscow),
- Vyčisl. Prikl. Mat. (Kiev)* Kievskii Gosudarstvennyi Universitet. Mežvedomstvennyi Naučnyi Sbornik. Vyčislitel'naja i Prikladnaja Matematika. (Kiev),
- Z. *Angew. Math. Mech.* Zeitschrift für Angewandte Mathematik und Mechanik. Ingenieurwissenschaftliche Forschungsarbeiten. (Berlin),
- Z. *Angew. Mat. Phys.* Zeitschrift für Angewandte Mathematik und Physik. ZAMP. Journal of Applied Mathematics and Physics. Journal de Mathématiques et de Physiques Appliquées. (Basel),
- Ž. *Vyčisl. Mat. i Mat. Fiz.* Akademiya Nauk SSSR. Žurnal Vyčislitel'noi Matematiki i Matematičeskoi Fiziki. (Moscow) (Translated as: U.S.S.R. Computational Math. and Math.Phys.).

Na kraju napomenimo da se u svetu svake godine organizuje sve veći broj simpozijuma posvećenih numeričkoj matematici.

1.2. ELEMENTI TEORIJE GREŠAKA

U ovom poglavlju se uvode osnovni pojmovi teorije grešaka i analizira efekat prostiranja grešaka kroz računski proces. Poseban tretman je dat mašinskim greškama zaokrugljivanja i efektu gubitaka značajnih cifara. Na kraju poglavlja su date izvesne napomene o statističkom pristupu u oceni 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

$$2.563, \quad 15.32, \quad 0.2687, \quad 0.002649$$

ima četiri značajne cifre.

Svaki broj x može se predstaviti u normalizovanom obliku
 (2.1.1)
$$x = (\pm 0.a_1 a_2 \dots a_n a_{n+1} \dots) b^k \quad (a_1 \neq 0),$$
 gde je b osnova brojnog sistema, a a_i ($i = 1, 2, \dots$) cifre brojnog sistema ($0 \leq a_i < b$).

Broj $\pm 0.a_1 a_2 \dots a_n a_{n+1} \dots$ zvaćemo mantisom i označavati

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

sa x^* . Broj k zvaćemo karakteristikom. Dakle, možemo pisati

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

Najčešće su u upotrebi binarni ($b = 2$), decimalni ($b=10$) i heksadecimalni ($b = 16$) brojni sistemi. Jednostavnosti radi, u daljem tekstu, često ćemo uzimati $b = 10$.

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 $x = 25.245\dots$ 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 absolutne greške približnog broja \bar{x} .

Definicija 2.1.3. Pod granicom absolutne greške Δ_x približnog broja \bar{x} podrazumeva se svaki broj ne manji od absolutne greške tog broja.

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

$$(2.1.2) \quad \bar{x} - \Delta_x \leq x \leq \bar{x} + \Delta_x.$$

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

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

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

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

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

$$\bar{x} = (\pm 0.a_1 a_2 \dots a_n) b^k \quad (a_1 \neq 0).$$

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

$$|e| \leq b^{k-n}.$$

Apsolutna greška koja se čini pri zameni broja x brojem \bar{x} može se smanjiti ako se koristi tzv. postupak zaokrugljivanja (zaokruživanja) brojeva. Taj postupak se sastoji u sledećem :

1) Ako je $a_{n+1} + a_{n+2}b^{-1} + \dots < \frac{1}{2}b$ koristi se prosto odsecanje;

2) Ako je $a_{n+1} + a_{n+2}b^{-1} + \dots > \frac{1}{2}b$, cifra a_n se povećava za jedinicu, a cifre a_{n+1}, a_{n+2}, \dots se odbacuju;

3) Ako je $a_{n+1} + a_{n+2}b^{-1} + \dots = \frac{1}{2}b$ ravnopravno se mogu koristiti pravila 1) i 2).

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

Napomenimo da se kod ručnog zaokrugljivanja brojeva u dekadnom sistemu ($b = 10$) u nerešenom slučaju preporučuje sledeće pravilo: Ako je cifra a_n paran broj koristiti pravilo 1), a ako je neparan broj koristiti pravilo 2).

Primer 2.1.3. Zaokrugljivanje broja $\pi = 3.141592653589793238\dots$ na 3, 5, 7 i 10 decimala daje približne brojeve:

3.142, 3.14159, 3.1415927, 3.1415926536.

Sukcesivno zaokrugljivanje poslednjeg broja daje redom brojeve:

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

Primetimo da se dobijeni približni brojevi zaokrugljeni na sedam decimala razlikuju na poslednjoj decimali, odakle zaključujemo da zaokrugljivanje ne treba izvoditi sukcesivno.

Apsolutna greška kod zaokrugljenog broja je

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

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

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

tj.

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

Po jednoj od mogućih 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.

Primer 2.1.4. Sistem jednačina (ulazna informacija)

$$\begin{aligned} 2x + 6y &= 8, \\ 2x + 6.0001y &= 8.0001 \end{aligned}$$

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

Primer 2.1.5. 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. Međutim, 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 računskih 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 dati suviše grubo, kako 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 $e_{x+y} = e_x + e_y$.

2. Kod oduzimanja, slično dobijamo

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

3. Za množenje važi

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

tj.

$$e_{xy} \approx ye_x + xe_y,$$

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

4. Kod deljenja imamo

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

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

$$\begin{aligned} \frac{\bar{x}}{\bar{y}} &= \frac{x + e_x}{y} \left(1 - \frac{e_y}{y} + \left(\frac{e_y}{y} \right)^2 - \dots \right) \\ &\approx \frac{x}{y} + \frac{e_x}{y} - \frac{x}{y^2} e_y, \end{aligned}$$

tj.

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

pri čemu su uzete pretpostavke kao i kod množenja.

Odgovarajuće relativne greške su redom

$$(2.2.1) \quad r_{x+y} = \frac{e_{x+y}}{x+y} = \frac{x}{x+y} r_x + \frac{y}{x+y} r_y,$$

$$(2.2.2) \quad r_{x-y} = \frac{e_{x-y}}{x-y} = \frac{x}{x-y} r_x - \frac{y}{x-y} r_y,$$

$$(2.2.3) \quad r_{xy} = \frac{e_{xy}}{xy} \approx r_x + r_y,$$

$$(2.2.4) \quad r_{x/y} = \frac{e_{x/y}}{x/y} \approx r_x - r_y,$$

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

Primedba 2.2.1. S obzirom da se tačne vrednosti x i y najčešće ne znaju, to se umesto njih u prethodno 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. Međutim, 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 računar radi sa pokretnom tačkom i n -razrednom mantisom. U tom slučaju, u memoriji računara, kao što je rečeno u prethodnom odeljku, svaki broj x aproksimira se postupkom zaokrugljivanja, standardnim brojem

$$(2.2.5) \quad \bar{x} = \bar{x}^* b^k,$$

gde je $\bar{x}^* = \pm 0.a_1 a_2 \dots a_n$ ($a_1 \neq 0$) mantisa, a k karakteristika (eksponent) broja \bar{x} . 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[†]. Na dalje, pretpostavimo da je $b = 10$.

Jednakost*

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

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 $* \in \{+, -, \times, \div\}$. Drugim

[†] Neka je m broj cifarskih mesta predviđenih 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 [34]).

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

rečima važi

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

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

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

Veličinu

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

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

Primer 2.2.1. 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. $f\bar{l}(\bar{x}_1 + \bar{x}_2) = 0.3961 \cdot 10^4$, 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.

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

Na primer, ako $*$ označava operaciju sabiranja, imamo

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

i

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

U konkretnom slučaju neka se operacija sabiranja izvodi u n -razrednom akumulatoru (single-precision accumulator). Za $n=3$ i brojeve $\bar{x}_1 = 0.100 \cdot 10^{-2}$, $\bar{x}_2 = -\bar{x}_3 = 0.100 \cdot 10^1$ imamo

$$f\bar{l}(f\bar{l}(\bar{x}_1 + \bar{x}_2) + \bar{x}_3) = 0 \quad \text{i} \quad f\bar{l}(\bar{x}_1 + f\bar{l}(\bar{x}_2 + \bar{x}_3)) = \bar{x}_1.$$

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

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

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

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

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

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

Odgovarajuća relativna greška je

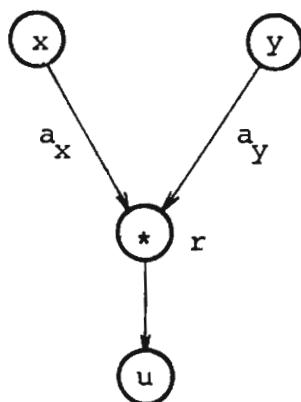
$$r_u^T = \frac{e_u^T}{u} = r_{x*y} + r ,$$

tj.

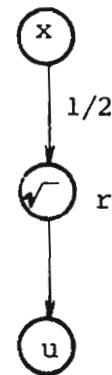
$$(2.2.7) \quad r_u^T = a_x r_x + a_y r_y + r.$$

Radi nalaženja totalne greške nekog numeričkog postupka i njene analize, predstavljajući 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 takođe 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} . Međutim, ova činjenica ne utiče na mogućnost analize greške. Naime, koristeći standardna označavanja, imamo

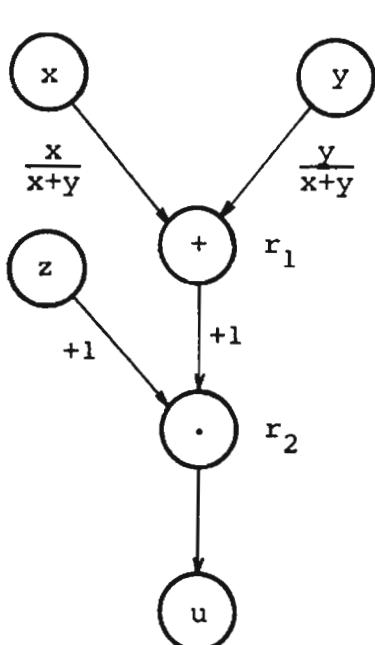
$$\bar{u} = \sqrt{\bar{x}} (1+r) = \sqrt{x(1+r_x)} (1+r) \leq \sqrt{x}(1+ \frac{1}{2}r_x)(1+r),$$

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 [21]).

Primer 2.2.2. Ako su poznate približne vrednosti brojeva x, y i z sa relativnim greškama zaokrugljivanja r_x, r_y i 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

$$\begin{aligned} r_{x+y}^T &= \frac{x}{x+y} r_x + \frac{y}{x+y} r_y + r_1, \\ r_u^T &= 1 \cdot r_{x+y}^T + 1 \cdot r_z + r_2 \\ &= \frac{x}{x+y} r_x + \frac{y}{x+y} r_y + r_z + r_1 + r_2. \end{aligned}$$

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

$$(2.2.8) \quad |r_u^T| \leq \left(\left| \frac{x}{x+y} \right| + \left| \frac{y}{x+y} \right| + 3 \right) \frac{1}{2} 10^{-n+1}.$$

Ako su brojevi x i y istog znaka,

(2.2.8) se svodi na

$$|r_u^T| \leq 2 \cdot 10^{-n+1}.$$

Primer 2.2.3. Izvršimo sada analizu greške kod izračunavanja zbiru

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

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

$$r_{x_1+x_2}^T = r_1,$$

$$r_{x_1+x_2+x_3}^T = \frac{x_1+x_2}{x_1+x_2+x_3} r_1 + r_2,$$

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

odakle je

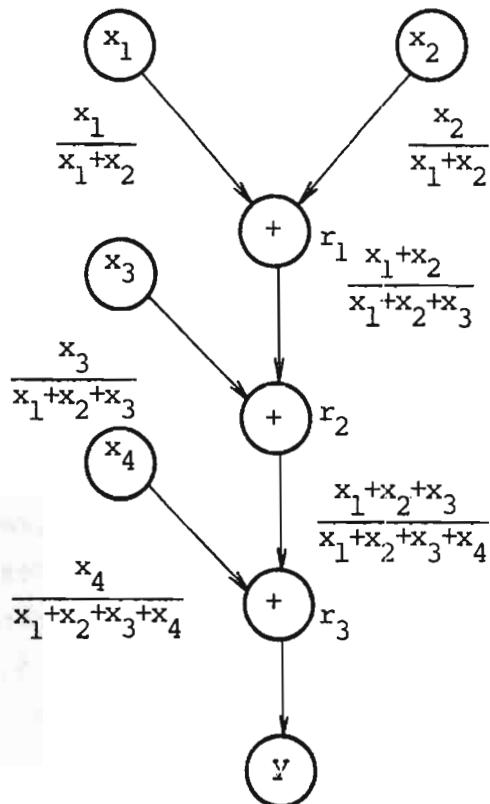
$$(2.2.10) \quad e_y^T = (x_1+x_2)r_1 + (x_1+x_2+x_3)r_2 + (x_1+x_2+x_3+x_4)r_3.$$

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

(2.2.10) sleduje

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

odakle zaključujemo da je granica absolutne 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

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

Primer 2.2.4. Nadjimo zbir brojeva

$$\begin{aligned} x_1 &= 0.1376 \cdot 10^0, & x_2 &= 0.4737 \cdot 10^0, & x_3 &= 0.7428 \cdot 10^1, \\ x_4 &= 0.6439 \cdot 10^2, & x_5 &= 0.5763 \cdot 10^3, & x_6 &= 0.2034 \cdot 10^4, \end{aligned}$$

zaokrugljujući sve medjurezultate na četiri značajne cifre.

Sabiranjem brojeva u datom redosledu, imamo redom

$$\begin{aligned} u_1 &= x_1 = 0.1376 \cdot 10^0, \\ u_2 &= f\ell(u_1+x_2) = 0.6113 \cdot 10^0, \\ u_3 &= f\ell(u_2+x_3) = 0.8039 \cdot 10^1, \\ u_4 &= f\ell(u_3+x_4) = 0.7243 \cdot 10^2, \\ u_5 &= f\ell(u_4+x_5) = 0.6487 \cdot 10^3, \\ u_6 &= f\ell(u_5+x_6) = 0.2683 \cdot 10^4. \end{aligned}$$

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

Ako sabiranje izvodimo u obrnutom redosledu, imamo

$$\begin{aligned} v_1 &= x_6 = 0.2034 \cdot 10^4, \\ v_2 &= f\ell(v_1+x_5) = 0.2610 \cdot 10^4, \\ v_3 &= f\ell(v_2+x_4) = 0.2674 \cdot 10^4, \\ v_4 &= f\ell(v_3+x_3) = 0.2681 \cdot 10^4, \\ v_5 &= f\ell(v_4+x_2) = 0.2681 \cdot 10^4, \\ v_6 &= f\ell(v_5+x_1) = 0.2681 \cdot 10^4, \end{aligned}$$

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

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

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

Primer 2.2.5. Neka su brojevi iz primera 2.2.3 pozitivni i bliški po vrednostima, tj. $x_i = x_0 + \delta_i$, $|\delta_i| \ll x_0$ ($i = 1, 2, 3, 4$). Korišćenjem rezultata iz pomenutog primera, zaključujemo da je

$$|e_y^T| \leq (9x_0 + 3|\delta_1| + 3|\delta_2| + 2|\delta_3| + |\delta_4|) \frac{1}{2} \cdot 10^{-n+1},$$

tj.

$$|e_y^T| \leq 4.5x_0 \cdot 10^{-n+1},$$

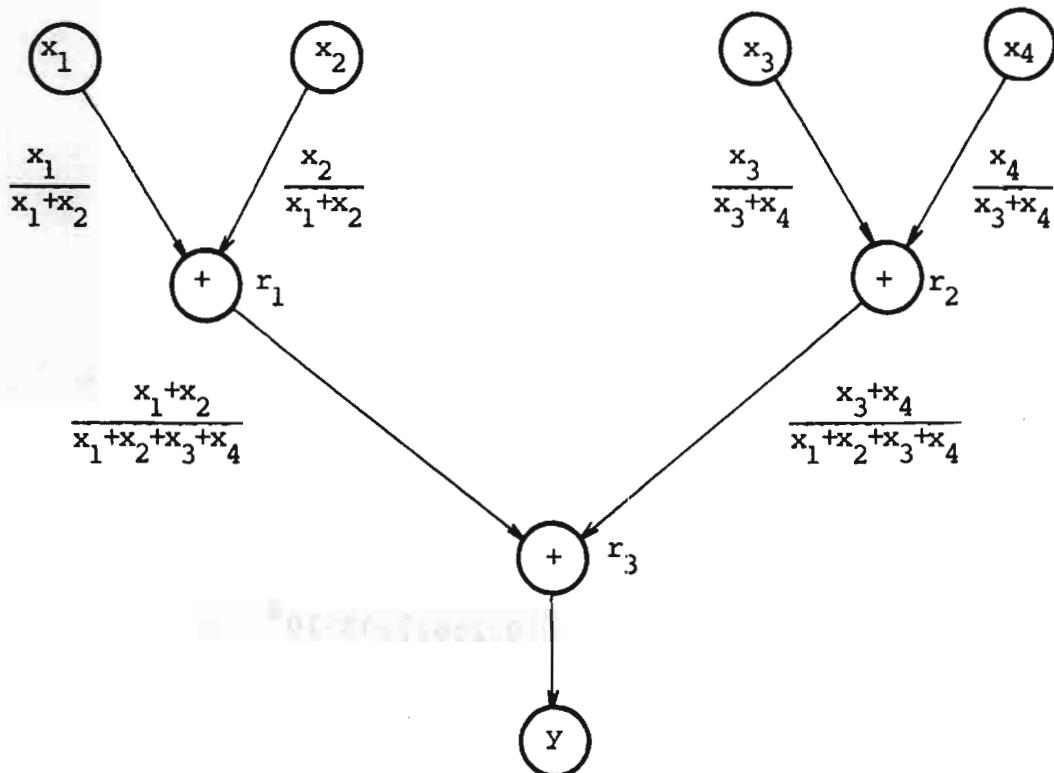
s obzirom na pretpostavku $|\delta_i| \ll x_0$ ($i = 1, 2, 3, 4$).

Izmjenimo 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 absolutne greške zbira četiri bliska pozitivna broja.

U opštem slučaju, ako imamo m^2 pozitivnih brojeva, približno jednakih po veličini, koje treba sabrati, granica absolutne greške biće utoliko 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 proja \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 za x i ako $C(p)$ označava najbliži ceo broj[†] realnom broju p , broj značajnih cifara broja \bar{x} može se oceniti pomoću

$$(2.2.11) \quad \ell = C(\log_{10} \frac{|x|}{|\bar{x} - x|}).$$

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

Posmatrajmo sada izračunavanje vrednosti

$$x = f(x_1, \dots, x_m) \quad (f: R^m \rightarrow R),$$

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

Ako je $\ell_{\min} = \min(\ell_1, \dots, \ell_m)$, za grešku rezultata može se dati približna ocena

$$(2.2.12) \quad |e| \approx \left| \sum_{i=1}^m x_i \frac{\partial f}{\partial x_i} 10^{-\ell_i} \right| \leq 10^{-\ell_{\min}} \sum_{i=1}^m \left| x_i \frac{\partial f}{\partial x_i} \right|.$$

[†] Uzećemo da je $C(k + \frac{1}{2}) = k + 1$ (k ceo broj).

Na osnovu (2.2.11) i (2.2.12) broj značajnih cifara u rezultatu može se oceniti pomoću

$$(2.2.13) \quad \ell_x = \ell_{\min} - \delta,$$

gde je

$$(2.2.14) \quad \delta = C(\log_{10} \frac{\sum_{i=1}^m |x_i \frac{\partial f}{\partial x_i}|}{|f(x_1, \dots, x_m)|}).$$

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 , najzad, može uzeti

$$\delta_x = \min \left\{ C(\log_{10} \frac{\sum_{i=1}^m |x_i \frac{\partial f}{\partial x_i}|}{|f(x_1, \dots, x_m)|}), -(n - \ell_{\min}) \right\}.$$

Ispitajmo veličinu δ u nekim standardnim izračunavanjima.

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

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

pa je $\delta = C(\log_{10} m)$. S obzirom da je, najčešće, $m < 100$, iz poslednje jednakosti, u najgorem slučaju, dobijamo $\delta = 1$. Ovo pokazuje da se gubitak značajnih cifara u ovom slučaju može zanemariti.

2º Deljenje. Uzmimo $x = f(x_1, x_2) = x_1/x_2$. Iz (2.2.14) i

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

sleduje $\delta = C(\log_{10} 2) = 0$, što znači da nema gubitka značajnih cifara.

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

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

imamo

$$(2.2.13) \quad \delta = C(\log_{10} \frac{\sum_{i=1}^m |x_i|}{|\sum_{i=1}^m x_i|}).$$

Iz poslednje jednakosti sleduje $\delta = 0$ 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 množenja, deljenja, sabiranja i oduzimanja, zaključujemo da je gubitak značajnih cifara prouzrokovani najčešće sabiranjima i oduzimanjima.

Efekat gubitka značajnih cifara jako je izražen kod oduzimanja bliskih brojeva, što ćemo ilustrovati sledećim primjerom.

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

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

Navedeni brojevi su dati sa devet značajnih cifara. Međutim, 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). Zaista,

$$\delta = C(\log_{10} \frac{|x_1| + |x_2|}{|x_1 - x_2|}) = 3.$$

Međutim, ako x izrazimo pomoću formule

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

kojom 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 istog "smera", tj. takve da obezbedjuju maksimalnu grešku izlazne informacije. Ovakav slučaj je malo verovatan.

U ovom odeljku ukazaćemo na statistički pristup 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.

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

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

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

odakle sleduje ocena

$$(2.3.1) \quad |e_u| \leq \sum_{i=1}^n |e_i| \leq \sum_{i=1}^n \Delta_i.$$

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 granice apsolutnih grešaka međusobno 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

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

Pretpostavljujući 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 okzirom na glomaznost postupka za konstrukciju ovakve funkcije, pristupićemo izvesnoj

aproksimaciji. Naime, pretpostavićemo da se suma grešaka ponaša po normalnom zakonu raspodele sa matematičkim očekivanjem 0 i standardnom devijacijom $\sigma = \Delta \sqrt{n/3}$.

Tada je

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

tj. sa verovatnoćom 0.9973 možemo očekivati da je

$$(2.3.2) \quad |e_u| < 3\sigma = \Delta \sqrt{3n} .$$

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

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

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

$$|e_u| < 5 \cdot 10^{-4} \sqrt{3 \cdot 75} = 0.75 \cdot 10^{-2},$$

tj. pet puta manja.

1.3. REKURZIVNA IZRAČUNAVANJA I SUMIRANJE

Od izuzetnog značaja u numeričkoj matematici su rekurzivna izračunavanja i postupci sumiranja. Ovo poglavlje započinjemo izlaganjem osnovne teorije linearnih diferencnih jednačina na kojima se zasniva jedna važna klasa tzv. linearnih rekursivnih postupaka. Jedna od glavnih primena ovih postupaka je u konstrukciji linearnih višekoračnih metoda za rešavanje Cauchyevog problema kod diferencijalnih jednačina. Posebna pažnja je posvećena numeričkim aspektima tročlane rekurentne relacije koja se sreće u mnogim problemima nauke i tehnike ([2]).

Poseban tretman je dat izračunavanju vrednosti nekih elementarnih funkcija, koje se sreću kao bibliotečke funkcije kod računara. Imajući u vidu da se aproksimacije funkcija najčešće daju u obliku polinoma, racionalne funkcije ili verižnog

razlomka, to se u posebnim odeljcima ovog poglavlja tretiraju algoritmi za izračunavanje vrednosti polinoma, uključujući i tzv. ekonomične šeme, algoritmi za izračunavanje verižnih razlomaka, kao i postupci za transformaciju racionalne funkcije u verižni razlomak i obrnuto. U posebnom odeljku, radi kompletnosti, izloženi su osnovni elementi verižnih razlomaka. Neki postupci za sumiranje redova i ubrzavanje konvergencije dati su u odeljku 1.3.4. Poslednji odeljak posvećen je asymptotskim razvojima koji često nalaze primenu kod izračunavanja vrednosti nekih specijalnih funkcija.

1.3.1. Diferencne jednačine i rekurzivna izračunavanja

Kako se mnogi problemi u numeričkoj matematici svode na rešavanje diferencnih jednačina, ovaj odeljak ćemo započeti izlaganjem nekih osnovnih pojmova i rezultata iz teorije ovih jednačina.

Definicija 3.1.1. Pod konačnom razlikom prvog reda funkcije f podrazumevamo izraz $\Delta f(x) = f(x+h) - f(x)$, gde je $h = \text{const} > 0$.

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

$$\Delta^n f(x) = \Delta^{n-1} f(x+h) - \Delta^{n-1} f(x) \quad (\Delta^0 f(x) = f(x)).$$

Lako se mogu dokazati sledeće formule

$$(3.1.1) \quad \Delta^n f(x) = \sum_{i=0}^n (-1)^i \binom{n}{i} f(x+(n-i)h)$$

i

$$f(x+nh) = \sum_{i=0}^n \binom{n}{i} \Delta^i f(x).$$

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

$$\Delta^0 y_k = y_k, \quad \Delta^n y_k = \Delta^{n-1} y_{k+1} - \Delta^{n-1} y_k \quad (n=1, 2, \dots).$$

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

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

i

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

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

Jednačina oblika

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

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

$$(3.1.2) \quad G(x; f(x), f(x+h), \dots, f(x+nh)) = 0.$$

Primer 3.1.2. 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 $x=x_0+kh$ i stavimo $y_k=f(x_0+kh)$, tada (3.1.2) postaje

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

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

$$(3.1.3) \quad y_{k+n} + b_1(k)y_{k+n-1} + \dots + b_n(k)y_k = Q(k).$$

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

Posmatrajmo, najpre, linearu diferencnu jednačinu prvog reda

$$(3.1.4) \quad y_{k+1} + b(k)y_k = Q(k).$$

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

$$(3.1.5) \quad y_k = (-1)^k C \prod_{i=0}^{k-1} b(i) \quad (C \text{ proizvoljna konstanta}).$$

Ako prepostavimo da C zavisi od k , tj. $C=C_k$, mogućno je polazeći od rešenja (3.1.5) naći rešenje nehomogene jednačine (3.1.4). Naime, u tom slučaju dobijamo da je rešenje nehomogene jednačine dato sa

$$y_k = (-1)^k \prod_{i=0}^{k-1} b(i) \left(C + \sum_{m=1}^k \frac{(-1)^m Q(m-1)}{\prod_{j=0}^{m-1} b(j)} \right),$$

gde je C proizvoljna konstanta.

Opšte rešenje linearne diferencne jednačine n -tog reda (3.1.3) je dato sa

$$y_k = C_1 \phi_1(k) + C_2 \phi_2(k) + \dots + C_n \phi_n(k) + f_p(k),$$

gde su $\phi_1, \phi_2, \dots, \phi_n$ linearno nezavisna partikularna rešenja odgovarajuće homogene jednačine, a f_p jedno partikularno rešenje nehomogene jednačine. Konstante C_i ($i=1, \dots, n$) su proizvoljne i u konkretnim slučajevima određuju se iz početnih uslova.

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

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

Dakle, karakteristična jednačina je

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

U zavisnosti od korena karakteristične jednačine razlikovaćemo slučajeve:

a) Ako su svi korenji $\lambda_1, \dots, \lambda_n$ prosti, opšte rešenje je

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

b) Ako su λ_p i λ_q konjugovano-kompleksni korenji, 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

$$(C_1 + C_2 k + \dots + C_{m-1} k^{m-1}) \lambda_i^k.$$

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

$$f_p(k) = A_0 k^m + A_1 k^{m+1} + \dots + A_s k^{m+s}.$$

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

$$f_p(k) = a^k (A_0 k^m + \dots + A_s k^{m+s}).$$

Primer 3.1.3. Rešimo diferencnu jednačinu

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

pod uslovima $y_0 = y_1 = y_3 = 0$ 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 ($i=1, \dots, 4$) 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

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

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

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

Nalaženje opšteg rešenja diferencne jednačine nije uvek aktuelno. Naime, daleko češće u numeričkoj matematici se sreće problem odredjivanja partikularnog rešenja koje zadovoljava odredjene uslove. Sa zadatim uslovima, diferencnu jednačinu nazivamo rekurentnom relacijom. U numeričkoj matematici od posebnog interesa su tzv. tročlane rekurentne relacije, koje se javljaju kod mnogih klasa specijalnih funkcija matematičke fizike i statistike, zatim u teoriji verižnih razlomaka i teoriji ortogonalnih polinoma, kod rešavanja linearnih diferencijalnih jednačina, kao i u mnogim problemima vezanim za konstrukciju redova. To su relacije oblika (3.1.3), sa $n=2$ i $Q(k)\equiv 0$, tj.

$$(3.1.6) \quad y_{k+1} + a_k y_k + b_k y_{k-1} = 0, \quad k=1, 2, \dots,$$

gde smo indeks k , u odnosu na (3.1.3), smanjili za jedinicu, a koeficijente b_1 ($k-1$) i b_2 ($k-1$), uprošćenja radi, zamenili sa a_k i b_k respektivno. Da bi relacija bila tročlana mora biti $b_k \neq 0$.

1^o Ako poznajemo početne uslove (vrednosti) $y_0 = f_0$ i $y_1 = f_1$, pomoću rekurentne relacije (3.1.6) možemo odrediti niz y_2, y_3, \dots, y_N , gde je N proizvoljno izabran prirodan broj ($N > 1$).

2^o Ako poznajemo vrednosti $y_N = f_N$ i $y_{N-1} = f_{N-1}$, pomoću (3.1.6) predstavljene u obliku

$$(3.1.7) \quad y_{k-1} = -\frac{1}{b_k}(y_{k+1} + a_k y_k), \quad k=N-1, N-2, \dots, 1,$$

za naznačene vrednosti k dobijamo niz $y_{N-2}, y_{N-3}, \dots, y_0$. Ovu relaciju nazivamo rekurentnom relacijom unazad.

3^o Najsloženiji uslovi su ako je dato $y_0 = f_0$ i $y_N = f_N$, tj. ako imamo tzv. konturne uslove. Tada se ovaj problem može interpretirati kao trodijagonalni sistem linearnih jednačina

$$(3.1.8) \quad \begin{bmatrix} a_1 & 1 & & & & \\ b_2 & a_2 & 1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & b_{N-2} & a_{N-2} & 1 \\ & & & & b_{N-1} & a_{N-1} \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_{N-2} \\ y_{N-1} \end{bmatrix} = \begin{bmatrix} -b_1 f_0 \\ 0 \\ \vdots \\ \vdots \\ 0 \\ -f_N \end{bmatrix}.$$

Dakle, u slučajevima 1^o i 2^o problem je eksplicitan, dok je u slučaju 3^o implicitan, jer zahteva rešavanje pomenu-tog trodijagonalnog sistema linearnih jednačina.

Iako, na prvi pogled, korišćenje rekurentne relacije je u principu jednostavno ono zaslužuje posebnu pažnju, s obzirom na numeričku nestabilnost koja se može manifestovati. Naime, svaki ciklus primene rekurentne relacije ne samo da generiše nove greške zaokrugljivanja, već i prihvata greške zaokrugljivanja iz prethodnih ciklusa. Tako, pod izvesnim - nepovoljnim uslovima, prostiranje grešaka kroz računski proces može "astronomski" ugroziti rezultat. Ovaj efekat ćemo kasnije ilustrovati na primeru izračunavanja Besselove funkcije prve

vrste i reda k , definisane pomoću

$$J_k(x) = \frac{1}{\pi} \int_0^{\pi} \cos(x \sin \theta - k\theta) d\theta.$$

Može se pokazati da $J_k(x)$ zadovoljava diferencnu jednačinu

$$(3.1.9) \quad y_{k+1} - \frac{2k}{x} y_k + y_{k-1} = 0 \quad (k=1, 2, \dots),$$

tj. da važi

$$(3.1.10) \quad J_{k+1}(x) = \frac{2k}{x} J_k(x) - J_{k-1}(x) \quad (k=1, 2, \dots).$$

Dakle, za fiksirano x , $J_k(x)$ je jedno partikularno rešenje diferencne jednačine (3.1.9), odredjeno, na primer, početnim uslovima $y_0 = J_0(x)$ i $y_1 = J_1(x)$. Funkcija $J_k(x)$ može se izraziti u obliku stepenog reda

$$J_k(x) = \left(\frac{x}{2}\right)^k \sum_{m=0}^{+\infty} \frac{(-1)^m}{(m!)^2} \left(\frac{x}{2}\right)^{2m}.$$

Primedba 3.1.1. Besselova funkcija prve vrste definiše se ne samo za realno, već i kompleksno x , a pritom red k ne mora biti nenegativan ceo broj, već može biti realan, ili čak kompleksan broj. Odgovarajuća reprezentacija u obliku reda ima oblik

$$J_v(x) = \left(\frac{x}{2}\right)^v \sum_{m=0}^{+\infty} \frac{(-1)^m}{m! \Gamma(v+m+1)} \left(\frac{x}{2}\right)^{2m}.$$

Primedba 3.1.2. Besselova funkcija druge vrste definiše se pomoću

$$Y_v(x) = \frac{J_v(x) \cos(v\pi) - J_{-v}(x)}{\sin(v\pi)}$$

ako v nije ceo broj, dok se za $v=k$ uzima

$$Y_k(x) = \lim_{v \rightarrow k} Y_v(x).$$

Inače, funkcija $Y_k(x)$ se može izraziti u obliku

$$(3.1.11) \quad Y_k(x) = -\frac{1}{\pi} \left(\frac{2}{x}\right)^k \sum_{m=0}^{k-1} \frac{(k-m-1)!}{m!} \left(\frac{x}{2}\right)^{2m} + \frac{2}{\pi} J_k(x) \log \frac{x}{2}$$

$$-\frac{1}{\pi} \left(\frac{x}{2}\right)^k \sum_{m=0}^{+\infty} (\psi(m+1) + \psi(k+m+1)) \frac{(-1)^m}{m! (k+m)!} \left(\frac{x}{2}\right)^{2m},$$

gde je psi funkcija definisana kao logaritamski izvod gama funkcije, tj. $\psi(z) = \frac{d}{dz}(\log \Gamma(z)) = \frac{\Gamma'(z)}{\Gamma(z)}$. Ako je z ceo broj, kao u formuli (3.1.11), imamo $\psi(1) = -\gamma$, $\psi(n) = -\gamma + \sum_{m=1}^{n-1} \frac{1}{m}$ ($n \geq 2$), gde je Eulerova konstanta $\gamma = 0.5772156649015328606\dots$. Formula (3.1.11) važi za $k \geq 1$. Za $k=0$ imamo

$$Y_0(x) = \frac{2}{\pi} \left(\gamma + \log \frac{x}{2}\right) J_0(x) - \frac{2}{\pi} \sum_{m=1}^{+\infty} \frac{(-1)^m}{(m!)^2} (\gamma + \psi(m+1)) \left(\frac{x}{2}\right)^{2m}.$$

Besselova funkcija druge vrste, takođe, je jedno partikularno rešenje diferencne jednačine (3.1.9), tj. važi

$$Y_{k+1}(x) = \frac{2k}{x} Y_k(x) - Y_{k-1}(x) \quad (k=1, 2, \dots).$$

S obzirom da su $J_k(x)$ i $Y_k(x)$ linearne nezavisne rešenja, to se opšte rešenje diferencne jednačine može dati u obliku

$$y_k = C_1 J_k(x) + C_2 Y_k(x),$$

gde su C_1 i C_2 proizvoljne konstante, a $x=\text{const.}$

Primedba 3.1.3. Funkcija $y(x) = C_1 J_k(x) + C_2 Y_k(x)$ je opšte rešenje Besselove diferencijalne jednačine $x^2 y'' + xy' + (x^2 - k^2)y = 0$. Ovde je $k = \text{const.}$

Vratimo se sada problemu određivanja partikularnog rešenja diferencne jednačine (3.1.9) za neke zadate početne uslove. Neka je $x=1$.

a) S obzirom da je

$$J_0(1) \approx 0.7651976866, \quad J_1(1) \approx 0.4400505857,$$

primenom rekurentne relacije (3.1.10) za $k=1, 2, \dots, 10$ dobijamo niz $\{J_k(1)\}$, koji je dat u tabeli 3.1.1. Broj u zagradi ukazuje na decimalni eksponent.

Table 3.1.1

k	$J_k(1)$	$Y_k(1)$
0	7.651976866 (-1)	8.825696420 (-2)
1	4.400505857 (-1)	-7.812128213 (-1)
2	1.149034848 (-1)	-1.650682607 (0)
3	1.95633535 (-2)	-5.821517606 (0)
4	2.4766362 (-3)	-3.327842303 (1)
5	2.497361 (-4)	-2.604058666 (2)
6	2.07248 (-5)	-2.570780243 (3)
7	-1.0385 (-6)	-3.058895705 (4)
8	-3.52638 (-5)	-4.256746185 (5)
9	-5.631823 (-4)	-6.780204939 (6)
10	-1.01020176 (-2)	-1.216180143 (8)

Primetimo da zbog oduzimanja bliskih brojeva nastupa poznati efekat gubitka značajnih cifara, tako da dobijena vrednost za $J_7(1)$ ima samo pet značajnih cifara. Štaviše, ova vrednost je pogrešna do te mere da se razlikuje i u znaku od tačne vrednosti. Tačna vrednost na deset značajnih cifara je

$$J_7(1) = 1.502325817(-6).$$

b) Uzimamo sada za početne vrednosti

$$Y_0(1) \approx 0.08825696420, \quad Y_1(1) \approx -0.7812128213.$$

Primenom iste rekurentne relacije dobijamo niz $\{Y_k(1)\}$ koji je dat, takodje, u tabeli 3.1.1. Uporedjivanjem dobijenih vrednosti sa vrednostima datim u [1 ,str.408] može se videti da su dobijeni rezultati korektni u svim ciframa.

Postavlja se sada pitanje koji su to razlozi da se u prvom slučaju generišu pogrešni rezultati, a drugom tačni. Takodje, od interesa je naći postupak za tačno generisanje niza $\{J_k(x)\}$.

U cilju odgovora na prvo pitanje primetimo da se opšte rešenje diferencne jednačine drugog reda (3.1.9), ili uopšte (3.1.6), može predstaviti kao linearne kombinacije bilo koja dva linearne nezavisna rešenja f_k i g_k , tj. $y_k = af_k + bg_k$, gde su a i b proizvoljne konstante. Nas će ovde interesovati specijalan slučaj kada postoji takav par linearne nezavisnih rešenja za koji je

$$(3.1.12) \quad \lim_{k \rightarrow +\infty} \frac{f_k}{g_k} = 0.$$

Problemi o kojima je bilo reči kod određivanja niza $\{J_k(1)\}$ nastupaju upravo kada se izračunava rešenje f_k ili cf_k , gde je c konstanta. Da bismo ovo pokazali primetimo da iz (3.1.12) sleduje

$$(3.1.13) \quad \lim_{k \rightarrow +\infty} \frac{f_k}{y_k} = 0$$

za svako y_k koje nije proporcionalno sa f_k .

Zaista, za $y_k = af_k + bg_k$ ($b \neq 0$) imamo

$$\lim_{k \rightarrow +\infty} \frac{f_k}{y_k} = \lim_{k \rightarrow +\infty} \frac{f_k/g_k}{a(f_k/g_k) + b} = 0.$$

Pretpostavimo sada da generišemo niz $\{f_k\}$, korišćenjem približnih početnih vrednosti $y_0 \approx f_0$ i $y_1 \approx f_1$ (na primer, dobijenih zaokrugljivanjem). I pod uslovom da se sve operacije u rekurentnoj relaciji obavljaju tačno (sa beskonačnom preciznošću, tj. sa beskonačnim brojem cifara u mantisi), rešenje y_k koje dobijamo biće, u opštem slučaju, linearno nezavisno od f_k . Imajući u vidu (3.1.13) zaključujemo da će relativna greška dobijenog rešenja težiti u beskonačnost, tj. imaćemo

$$\left| \frac{y_k - f_k}{f_k} \right| = \left| \frac{1 - f_k/y_k}{f_k/y_k} \right| \rightarrow +\infty \quad (k \rightarrow +\infty).$$

Prema tome, ovakav postupak za određivanje niza $\{f_k\}$ daje pogrešne rezultate. Rešenje f_k sa osobinom (3.1.13) naziva se minimalno rešenje diferencne jednačine (3.1.6). Suprotno, neminalno rešenje zovemo dominantno rešenje. Primetimo da je svako dominantno rešenje asymptotski proporcionalno sa g_k .

U prethodno posmatranom primeru imali smo $f_k = J_k(x)$ i $g_k = Y_k(x)$. Na osnovu asymptotskih formula ([1, str.365])

$$J_v(x) \sim \frac{1}{\sqrt{2\pi v}} \left(\frac{ex}{2v}\right)^v, \quad Y_v(x) \sim -\sqrt{\frac{2}{\pi v}} \left(\frac{ex}{2v}\right)^{-v},$$

pri $v \rightarrow \infty$, vidimo da $J_v(x) \rightarrow 0$ i $Y_v(x) \rightarrow -\infty$.

Dakle, rezimirajmo da se minimalno rešenje navedenim postupkom ne može izračunati, za razliku od dominantnog rešenja koje se veoma tačno određuje.

Razne metode za određivanje minimalnog rešenja tročlanе rekurentne relacije, uključujući i odgovarajuće primene, razmatrao je W.Gautschi u radu [13].

Na kraju ovog odeljka navećemo jedan postupak za dobro određivanje niza $\{J_k(x)\}$, poznat kao Millerov postupak. Na osnovu prethodno pomenute asymptotske formule za $J_k(x)$ ($x=\text{const}$) videli smo da $J_k(x)$ teži nuli kada k teži beskonačnosti. Zato izaberimo dovoljno veliko n i stavimo

$$(3.1.14) \quad \bar{J}_n(x)=0 \quad \text{i} \quad \bar{J}_{n-1}(x)=1.$$

Nešto bolji izbor je uzeti, umesto 0 i 1, vrednosti koje se dobijaju iz pomenute asymptotske relacije za $v=n$ i $v=n-1$. Mi ćemo se ipak, jednostavnosti radi, opredeliti za vrednosti date pomoću (3.1.14), a zatim ćemo primenom rekurentne relacije unazad (3.1.7), koja u našem slučaju ima oblik

$$(3.1.15) \quad \bar{J}_{k-1}(x) = \frac{2k}{x} \bar{J}_k(x) - \bar{J}_{k+1}(x),$$

odrediti redom $\bar{J}_{n-2}(x), \dots, \bar{J}_1(x), \bar{J}_0(x)$. Naravno, dobijene vrednosti ne odgovaraju tačnim vrednostima Besselovih funkcija, ali se, zbog linearnosti relacije (3.1.15) od njih razlikuju za istu multiplikativnu konstantu.

S obzirom na identitet

$$J_0(x) + 2(J_2(x) + J_4(x) + \dots) = 1$$

moguće je odrediti ovu multiplikativnu konstantu. Naime, ako odredimo sumu

$$S = \bar{J}_0(x) + 2(\bar{J}_2(x) + \bar{J}_4(x) + \dots)$$

iz proporcije $1:S = J_k(x) : \bar{J}_k(x)$ nalazimo

$$(3.1.16) \quad J_k(x) = \frac{1}{S} \bar{J}_k(x) \quad (k=0, 1, \dots).$$

Obično za n uzimamo dovoljno veliki neparan broj i tada prilikom izračunavanja sume S polazimo od najmanjeg sabirka $\bar{J}_{n-1}(x) (=1)$. Za dobijanje rezultata sa većom tačnošću treba poći od većeg broja n .

Ako uzmemo $n=11$, tj. $\bar{J}_{11}(1)=0$ i $\bar{J}_{10}(1)=1$ dobijamo rezultate koji su dati u tabeli 3.1.2. Ovde je $S=3810092281$, a $i/S=2.624608346(-10)$,

k	$\bar{J}_k(1)$	$J_k(1)$
10	1	2.624608346 (-10)
9	20	5.249216692 (-9)
8	359	9.422343962 (-8)
7	5724	1.502325817 (-6)
6	79777	2.093833800 (-5)
5	951600	2.497577302 (-4)
4	9436223	2.476638964 (-3)
3	74538184	1.956335398 (-2)
2	437792881	1.149034849 (-1)
1	1676633340	4.400505857 (-1)
0	2915473799	7.651976866 (-1)

Pomoću (3.1.16) dobijamo kolonu tabele sa vrednostima za $J_k(1)$. Uporedjivanjem sa tačnim vrednostima može se videti da su pogrešne vrednosti samo za $J_{10}(1)$, $J_9(1)$ i $J_8(1)$ (ograđene cifre su pogrešne).

Ako uzmemo $n=17$, добићемо sledeće vrednosti (sa deset značajnih cifara):

$$\begin{aligned} J_8(1) &= 9.422344172 (-8), \\ J_9(1) &= 5.249250180 (-9), \\ J_{10}(1) &= 2.630615124 (-10). \end{aligned}$$

Uporedjivanjem vidimo da su sve cifre tačne.

Millerov postupak se vrlo često koristi u numeričkoj matematici u više oblika koji su slični navedenom.

1.3.2. Izračunavanje vrednosti nekih elementarnih funkcija

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: $x \mapsto e^x$, $x \mapsto \sin x$ i $x \mapsto \log x$, korišćenjem polinomskega razvoja. U odeljku 1.3.7 razmatraćemo primenu verižnih razlomaka. Opšti metodi za aproksimaciju funkcija biće tretirani u posebnoj glavi.

I^o Za izračunavanje vrednosti funkcije $x \mapsto e^x$ može se koristiti Taylorov razvoj

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

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

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

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

sumiranjem dobijene geometrijske progresije dobijamo

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

Kako je $0 \leq x \leq 1$, iz prethodne nejednakosti sledi

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

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 $0 \leq x \leq 1$, 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} .

Primer 3.2.1. Izračunajmo broj e sa relativnom greškom koja je manja od 10^{-7} . Na osnovu prethodnog razmatranja, 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 ($k=0, 1, \dots, 10$) predstavljeni u normalizovanom obliku sa sedmorazrednom mantisom su redom

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

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 $e (=0.27182818284\dots \cdot 10^1)$ sa sedam značajnih cifara. Relativna greška je manja od 10^{-7} (tako, manja od $0.67 \cdot 10^{-7}$).

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

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

gde je $[x]$ najveće celo* od x i $z=x-[x]$. Pri ovome vrednost $e^{[x]}$ izračunava se prostim množenjem, tj.

* Na primer, $[3.14]=3$, $[-3.14]=-4$.

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

gde su

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

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

2° Za izračunavanje vrednosti funkcije $x \mapsto \sin x$ za proizvoljnu vrednost argumenta x , dovoljno je znati kako se može izračunati vrednost $\sin \frac{\pi z}{2}$ za $z \in [-1, 1]$. Naime, ako za dato x izračunamo redom veličine

$$u = \frac{2}{\pi} x, \quad v = u - 4[\frac{1}{4}(u+1)], \quad z = \begin{cases} v & (v \leq 1), \\ 2-v & (v > 1), \end{cases}$$

gde je $[p]$ oznaka za najveće celo od p , tada je $\sin x = \sin \frac{\pi z}{2}$.

Da bismo ovo dokazali primetimo da je

$$\sin x = \sin \frac{\pi u}{2} = \sin \frac{\pi}{2}(v + 4[\frac{1}{4}(u+1)]),$$

tj.

$$\sin x = \sin(\frac{\pi v}{2} + 2\pi[\frac{1}{4}(u+1)]) = \sin \frac{\pi v}{2}.$$

Kako je

$$v = 4(\frac{1}{4}(u+1) - [\frac{1}{4}(u+1)]) - 1$$

i

$$0 \leq \frac{1}{4}(u+1) - [\frac{1}{4}(u+1)] < 1$$

zaključujemo da je $-1 \leq v < 3$. Razlikujmo sada dva slučaja:

Slučaj $v \leq 1$. Tada je $z=v$ ($\in [-1, 1]$) i $\sin x = \sin \frac{\pi v}{2} = \sin \frac{\pi z}{2}$.

Slučaj v > 1. Kako je $z=2-v$ i $1 < v < 3$ zaključujemo da $z \in (-1, 1)$ i da je pritom

$$\sin x = \sin \frac{\pi v}{2} = \sin \frac{\pi}{2}(2-z) = \sin \frac{\pi z}{2}.$$

Prema tome, dovoljno je znati neku približnu formulu (aproksimaciju) za $\sin \frac{\pi z}{2}$, kada $z \in [-1, 1]$. Jedna takva formula je, na primer,

$$\sin \frac{\pi z}{2} \approx 1.57032002z - 0.64211317z^3 + 0.07186085z^5,$$

za koju apsolutna greška ne prelazi vrednost $6.8 \cdot 10^{-5}$, kada $z \in [-1, 1]$. U šestoj glavi ove knjige razmotrićemo razne postupke za aproksimaciju funkcija.

3° 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 $m \in \mathbb{Z}$. Tada je

$$\log x = m \log 2 + \log z \quad (\log 2 = 0.69314718\dots).$$

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

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

gde su

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

i

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

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

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

Uvođenjem označke $u_k = \frac{y^{2k-1}}{2k-1}$ ($k=1, \dots, n$), najzad dobijamo

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

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

1.3.3. Izračunavanje vrednosti polinoma

Jedan elementaran, ali važan problem je izračunavanje vrednosti polinoma

$$(3.3.1) \quad P(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n.$$

Ako bismo izračunavali vrednost polinoma $P(x)$, na osnovu (3.3.1), bilo bi potrebno $2n-1$ množenja i n sabiranja. Međutim, ukoliko $P(x)$ predstavimo u obliku

$$(3.3.2) \quad P(x) = (((a_0 x + a_1) x + a_2) x + a_3) x + \dots + a_{n-1} x + a_n$$

potrebno je samo n množenja i n sabiranja.

Na osnovu (3.3.2) može se formirati rekurzivni postupak za izračunavanje vrednosti polinoma za $x=x_0$

$$(3.3.3) \quad b_0 = a_0, \quad b_k = a_k + x_0 b_{k-1} \quad (k=1, \dots, n)$$

koji posle n koraka daje vrednost polinoma, tj. $P(x_0) = b_n$.

Izloženi postupak je poznat kao Hornerova šema i može se interpretirati kroz sledeću šemu:

x_0	a_0	a_1	a_2	a_3	\dots	a_{n-1}	a_n	
	$x_0 b_0$	$x_0 b_1$	$x_0 b_2$	$x_0 b_{n-2}$		$x_0 b_{n-1}$		
	b_0	b_1	b_2	b_3		b_{n-1}	$b_n = P(x_0)$	

Naime, u prvoj vrsti šeme pišemo koeficijente polinoma (3.3.1), počev od najstarijeg koeficijenta, a drugu vrstu započinjemo sa vrednošću argumenta x_0 za koji izračunavamo vrednost polinoma. U trećoj vrsti pišemo koeficijente b_k , koje

izračunavamo sabiranjem odgovarajućih elemenata prve i druge vrste, pri čemu je $b_0 = a_0$. Elemente druge vrste formiramo množenjem vrednosti x_0 sa prethodnim elementom iz treće vrste.

Primetimo da su b_k ($k=0, 1, \dots, n-1$) koeficijenti polinoma $P_1(x)$ koji se dobija deljenjem $P(x)$ sa linearним faktorom $x-x_0$. Zaista, uporedjivanjem koeficijenata uz odgovarajuće stepene u

$$\begin{aligned} a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n \\ = (x-x_0) (b_0 x^{n-1} + b_1 x^{n-2} + \dots + b_{n-1}) + b_n \end{aligned}$$

dobijamo (3.3.3).

Kao što smo prethodno rekli, Hornerova šema zahteva n množenja i n sabiranja za izračunavanje vrednosti polinoma (3.3.1). Ostrowski je dokazao da je $2n$ minimalan broj operacija za izračunavanje vrednosti polinoma ako je $n \leq 4$. Ako je $4 \leq n \leq 6$ može se dokazati ([25], [36], [18]) da se vrednost polinoma može izračunati sa $\left[\frac{1}{2}(n+3)\right]$ množenja i ne više od $n+1$ sabiranja. Oznaka $[t]$ predstavlja najveće celo od t .

Tako u slučaju $n=4$ imamo sledeći ekvivalentni oblik polinoma (3.3.1)

$$(3.3.4) \quad P(x) = ((Ax+B)^2 + Ax + C)((Ax+B)^2 + D) + E,$$

koji zahteva 3 množenja i 5 sabiranja. Vrednosti parametara u (3.3.4) mogu se odrediti pomoću

$$\begin{aligned} A &= \sqrt[4]{a_0}, \quad B = \frac{a_1 - A^3}{4A^3}, \quad D = 3B^2 + 8B^3 + \frac{a_3 A - 2a_2 B}{A^2}, \\ C &= \frac{a_2}{A^2} - 2B - 2B^2 - D, \quad E = a_4 - (C+D)B^2 - B^4 - CD, \end{aligned}$$

gde smo, ne umanjujući opštost, pretpostavili da je $a_0 > 0$.

U slučaju $n=5$ ekonomična formula je oblika

$$(3.3.5) \quad P(x) = (((Ax+B)^2 + C)(Ax+B)^2 + D)(Ax+E) + F.$$

Uporedjivanjem koeficijenata uz odgovarajuće stepene u (3.3.5) i (3.3.1) (za $n=5$) dobijamo sledeće jednakosti

$$\begin{aligned} A^5 &= a_0, \\ A^4(4B+E) &= a_1, \\ A^3(6B^2+4BE+C) &= a_2, \\ A^2(4B^3+6B^2E+2BC+CE) &= a_3, \\ A(B^4+4B^3E+B^2C+2BCE+D) &= a_4, \\ B^4E+B^2CE+DE+F &= a_5. \end{aligned}$$

Ako stavimo $c_k = \frac{a_k}{a_0} A^k$ ($k=1, 2, 3, 4, 5$), iz prethodnih jednakosti dobijamo

$$\begin{aligned} A &= \sqrt[5]{a_0}, \quad C = c_2 + 10B^2 - 4c_1B, \quad E = c_1 - 4B, \\ D &= c_4 - B^4 - 4B^3E - B^2C - 2BCE, \\ F &= c_5 - B^4E - B^2CE - DE, \end{aligned}$$

gde je B koren jednačine

$$(3.3.6) \quad 40B^3 - 24c_1B^2 + 2(2c_1^2 + c_2)B + (c_3 - c_1c_2) = 0.$$

Jednačina (3.3.6) može imati jedno ili tri realna rešenja, tako da formula (3.3.5), u opštem slučaju, nije jedinstvena.

Za izračunavanje vrednosti polinoma (3.3.5) potrebno 4 množenja i 5 sabiranja. Napomenimo da računske operacije koje obezbeđuju potrebne transformacije koeficijenata u cilju svestjenja na ekonomični oblik (3.3.5) ne ubrajamo u potrebne operacije za izračunavanje vrednosti polinoma, s obzirom da se one za konkretnom polinom izvode samo jednom.

Slično, za $n=6$ postoji ekonomična formula sa 4 množenja i 7 sabiranja (videti [12, str. 57-59]).

Primer 3.3.1. Za aproksimaciju funkcije $x \mapsto 2^x$ za $-\frac{1}{4} \leq x \leq 0$ sa relativnom greškom manjom od $2^{-28} (\approx 3.7 \times 10^{-9})$ može se koristiti sledeći polinom četvrtog stepena

$$\begin{aligned} P(x) &= 0.9999999975 + 0.6931466849x \\ &\quad + 0.2402108680x^2 + 0.0553299147x^3 + 0.0088171049x^4. \end{aligned}$$

Saglasno prethodnom imamo

$$A = \sqrt[4]{a_0} = 0.3064301558,$$

$$B = \frac{1}{4} \left(\frac{a_1}{A^3} - 1 \right) = 0.2307347358,$$

$$D = 1.3394747869, \quad C = 0.4377960868,$$

$$E = 3161295697,$$

pri čemu su rezultati zaokrugljeni na deset decimala.

Algoritam koji zahteva 3 množenja i 5 sabiranja može se iskazati u obliku

$$\alpha := A \cdot x,$$

$$\beta := \alpha + B, \quad \beta := \beta \cdot \beta,$$

$$P := (\alpha + \beta + C) \cdot (\beta + D) + E.$$

Primer 3.3.2. Posmatrajmo polinom

$$P(x) = 32x^5 + 48x^4 + 8x^3 - 12x^2 - 8x + 7,$$

za koji imamo

$$A = \sqrt[5]{a_0} = 2, \quad c_1 = \frac{48}{32} \cdot 2 = 3, \quad c_2 = \frac{8}{32} \cdot 4 = 1, \quad c_3 = \frac{-12}{32} \cdot 8 = -3,$$

$$c_4 = \frac{-8}{32} \cdot 16 = -4, \quad c_5 = \frac{7}{32} \cdot 32 = 7.$$

Jednačina (3.3.6) postaje

$$40B^3 - 72B^2 + 38B - 6 = 0,$$

tj.

$$(B-1)(2B-1)(10B-3) = 0,$$

odakle zaključujemo da postoje tri različite ekonomične formule oblika (3.3.5):

1° Za $B=1$, na osnovu prethodnog, imamo

$$C = -1, \quad E = -1, \quad D = -2, \quad F = 5,$$

tj.

$$P(x) = (((2x+1)^2 - 1)(2x+1)^2 - 2)(2x-1) + 5;$$

2^o Za $B=0.5$ imamo

$$C=-2.5, \quad E=1, \quad D=-1.4375, \quad F=9,$$

tj.

$$P(x) = (((2x+0.5)^2 - 2.5)(2x+0.5)^2 - 1.4375)(2x+1) + 9;$$

3^o Za $B=0.3$ imamo

$$C=-1.7, \quad E=1.8, \quad D=-2.2135, \quad F=11.24512,$$

tj.

$$P(x) = (((2x+0.3)^2 - 1.7)(2x+0.3)^2 - 2.2135)(2x+1.8) + 11.24512.$$

Algoritam, kao što je ranije rečeno, zahteva 4 množenja i 5 sabiranja i može se iskazati u obliku:

$$\alpha := A \cdot x, \quad \beta := \alpha + B, \quad \gamma := \beta \cdot \beta,$$

$$\gamma := (\beta + C) \cdot \beta + D,$$

$$P := \gamma \cdot (\alpha + E) + F.$$

Posmatrajmo sada polinom proizvoljnog stepena n i izražimo ga u obliku

$$(3.3.7) \quad P(x) = Q_0(x) + Q_1(x),$$

gde su

$$Q_0(x) = a_0 x^n + a_2 x^{n-2} + a_4 x^{n-4} + \dots,$$

$$Q_1(x) = a_1 x^{n-1} + a_3 x^{n-3} + a_5 x^{n-5} + \dots.$$

U zavisnosti od toga da li je n parno ili neparno definisaćemo polinome S_0 i T pomoću:

$$S_0(x^2) = Q_0(x), \quad T(x^2) = \frac{Q_1(x)}{x} \quad (n=2m+2)$$

i

$$S_0(x^2) = Q_1(x), \quad T(x^2) = \frac{Q_0(x)}{x}, \quad (n=2m+1).$$

Primetimo da je $T(t)$ polinom stepena m . Neka su $\alpha_1, \alpha_2, \dots, \alpha_m$ nule polinoma $T(t)$, koje su nam poznate.

Deljenjem polinoma $S_0(t)$ sa faktorom $t - \alpha_m$ dobijamo polinom $S_1(t)$ i odgovarajući ostatak β_m

$$S_0(t) = (t - \alpha_m) S_1(t) + \beta_m.$$

Nastavljajući proces deljenja $S_1(t)$ sa $t - \alpha_{m-1}$, itd. dobijamo rekurzivni postupak za određivanje niza $\beta_m, \beta_{m-1}, \dots, \beta_1$:

$$(3.3.8) \quad S_{i-1}(t) = (t - \alpha_{m-i+1}) S_i(t) + \beta_{m-i+1} \quad (i=1, \dots, m),$$

pri čemu je

$$\begin{aligned} S_m(t) &= a_0 t + \beta_0 & (n=2m+2) \\ &= a_1 & (n=2m+1). \end{aligned}$$

Dakle, za svako $i=1, \dots, m$ primenjujemo Hornerovu šemu.

Poznavajući nizove $(\alpha_1, \alpha_2, \dots, \alpha_m)$ i $(\beta_1, \beta_2, \dots, \beta_m)$, vrednost polinoma $P(x)$ se može odrediti pomoću $\lceil \frac{1}{2}(n+4) \rceil$ množenja i n sabiranja.

Na osnovu prethodnog, iz (3.3.7) sleduje

$$P(x) = S_0(x^2) + xT(x^2).$$

Tada korišćenjem (3.3.8) za $t=x^2$, jednostavno dobijamo

$$P(x) = (\dots ((P_0 \cdot (x^2 - \alpha_1) + \beta_1) \cdot (x^2 - \alpha_2) + \dots + \beta_{m-1} \cdot (x^2 - \alpha_m) + \beta_m,$$

gde je

$$\begin{aligned} P_0 &= a_0 x^2 + a_1 x + \beta_0 & (n=2m+2) \\ (3.3.9) \quad &= a_0 x + a_1 & (n=2m+1). \end{aligned}$$

Primetimo da je u $P(x)$ uključen i sabirak $xT(x^2) = Ax(x^2 - \alpha_1) \cdot (x^2 - \alpha_2) \dots (x^2 - \alpha_m)$, gde je $A=a_1$, ako je $n=2m+2$, i $A=a_0$, ako je $n=2m+1$. Početna vrednost P_0 je, evidentno jednaka $P_0 = S_m(x^2) + Ax$.

Prema tome, startujući sa (3.3.9) vrednost polinoma $P(x)$ možemo odrediti pomoću

$$P_i = P_{i-1} \cdot (x^2 - \alpha_i) + \beta_i \quad (i=1, \dots, m),$$

pri čemu je $P(x) = P_m$.

Navedeni metod za izračunavanje vrednosti polinoma dao je D.E. Knuth ([18]). Jednu modifikaciju ovog metoda dao je J. Eve ([11]).

1.3.4. Sumiranje redova i ubrzavanje konvergencije

Izlaganje u ovom odeljku počećemo sa jednim konkretnim primerom. Naime, razmotrićemo problem sumiranja reda

$$(3.4.1) \quad S = \sum_{k=0}^{+\infty} (-1)^k \frac{4}{2k+1}$$

čija je tačna suma $S=\pi (=3.1415926535\dots)$.

Da bismo postigli tačnost od 10^{-4} direktnim sumiranjem reda "član po član", na osnovu nejednakosti $4/(2n+1) < 10^{-4}$ zaključujemo da je potrebno uzeti oko 20000 članova reda. Praktično ovakav način za sumiranje je neizvodljiv. Za red (3.4.1) kažemo da 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

$$(3.4.2) \quad \sum_{k=0}^{+\infty} (-1)^k a_k = a_0 - a_1 + a_2 - a_3 + \dots \quad (a_k \geq 0),$$

čija je suma A. Pokazaćemo da je red

$$(3.4.3) \quad \frac{1}{2} a_0 - \frac{1}{2}(a_1 - a_0) + \frac{1}{2}(a_2 - a_1) - \dots$$

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

Pretpostavimo da su $\{A_n\}$ i $\{B_n\}$ nizovi parcijalnih sum redova (3.4.2) i (3.4.3) respektivno. Tada je

$$(3.4.4) \quad A_n - B_n = \frac{1}{2}(-1)^{n-1} a_{n-1} \quad (n=1,2,\dots).$$

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

(3.4.4) sleduje

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

Primenimo izloženu transformaciju na red (3.4.1). Tada je

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

tj.

$$(3.4.5) \quad S = 2 + \sum_{k=1}^{+\infty} \frac{4(-1)^{k-1}}{4k^2-1}.$$

Ocenimo sada koliko je članova reda (3.4.5) potrebno sabrati da bi se dobila suma sa istom tačnošću kao i ranije. Iz $4/(4n^2-1) < 10^{-4}$ sleduje $n > 100$. Dakle, sada se ista tačnost može postići ako se uzme 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

$$(3.4.6) \quad f(x) = \sum_{k=0}^{+\infty} a_k x^k,$$

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

Iz $f(x)=\sum a_k x^k$, gde je

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

sleduje

$$\begin{aligned} (1-x)g(x) &= (1-x) \sum_{k=0}^{+\infty} a_{k+1} x^k \\ &= \sum_{k=0}^{+\infty} a_{k+1} x^k - \sum_{k=0}^{+\infty} a_{k+1} x^{k+1} \\ &= a_0 + \sum_{k=0}^{+\infty} (a_{k+1} - a_k) x^k, \end{aligned}$$

odakle je

$$f(x) = a_0 + \frac{x}{1-x} (a_0 + \sum_{k=0}^{+\infty} (a_{k+1} - a_k) x^k),$$

tj.

$$f(x) = \frac{a_0}{1-x} + \frac{x}{1-x} \sum_{k=0}^{+\infty} \Delta a_k x^k.$$

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

$$\sum_{k=0}^{+\infty} (-1)^k a_k = \frac{1}{2} a_0 - \frac{1}{2} \sum_{k=0}^{+\infty} (-1)^k \Delta a_k,$$

što je ekvivalentno sa (3.4.3).

Sukcesivnom primenom izložene Euler-Abelove transformacije m puta na red (3.4.6) dobijamo

$$f(x) = \sum_{k=0}^{m-1} \frac{\Delta^k a_0 x^k}{(1-x)^{k+1}} + \left(\frac{x}{1-x}\right)^m \sum_{k=0}^{+\infty} \Delta^m a_k x^k.$$

Posebno je interesantan slučaj kada $m \rightarrow +\infty$. Tada imamo

$$(3.4.7) \quad f(x) = \frac{1}{1-x} \sum_{k=0}^{+\infty} \Delta^k a_0 \left(\frac{x}{1-x}\right)^k.$$

Primer 3.4.1. Sukcesivnom primenom Euler-Abelove transformacije dva puta na red $f(x) = \sum_{k=0}^{+\infty} \frac{x^k}{k+1}$, dobijamo

$$f(x) = \frac{1}{1-x} + \frac{\left(\frac{1}{2} - 1\right)x}{(1-x)^2} + \left(\frac{x}{1-x}\right)^2 \sum_{k=0}^{+\infty} \left(\frac{1}{k+3} - \frac{2}{k+2} + \frac{1}{k+1}\right) x^k$$

tj.

$$f(x) = \frac{1}{1-x} - \frac{x}{2(1-x)^2} + \left(\frac{x}{1-x}\right)^2 \sum_{k=0}^{+\infty} \frac{2x^k}{(k+1)(k+2)(k+3)}.$$

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

$$\log 2 \approx \frac{5}{8} + \frac{1}{2} \sum_{k=0}^n \frac{(-1)^k}{(k+1)(k+2)(k+3)},$$

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

Na ubrzavanje datog stepenog reda primenićemo sada formulu (3.4.7). Kako je $a_m = \frac{1}{m+1}$ za k -tu differencu imamo

$$\Delta^k a_m = \sum_{i=0}^k (-1)^i \binom{k}{i} a_{m+k-i} = - \sum_{i=0}^k \frac{(-1)^i \binom{k}{i}}{i-(m+k+1)} .$$

S obzirom na identitet

$$(3.4.8) \quad \sum_{i=0}^k \frac{(-1)^i \binom{k}{i}}{i+a} = \frac{k!}{a(a+1)\dots(a+k)} \quad (a \neq 0, -1, \dots, -k),$$

zaključujemo da je

$$\Delta^k a_m = \frac{(-1)^k k! m!}{(m+k+1)!} \quad i \quad \Delta^k a_0 = \frac{(-1)^k}{k+1} .$$

Primenom (3.4.7) dobijamo da je u ovom slučaju

$$f(x) = \frac{1}{1-x} \sum_{k=0}^{+\infty} \frac{(-1)^k}{k+1} \left(\frac{x}{1-x}\right)^k .$$

Najzad, za $x=-1$ imamo

$$\log 2 = \sum_{k=0}^{+\infty} \frac{(-1)^k}{k+1} = \sum_{k=1}^{+\infty} \frac{1}{2^k k} .$$

Primer 3.4.2. Posmatrajmo alternativni red

$$S_r = \sum_{k=1}^{+\infty} \frac{(-1)^{k-1}}{(2k+2r+1)\pi^{2k}} .$$

Primetimo da se za $r=0$ dobija

$$(3.4.9) \quad S_0 = \sum_{k=1}^{+\infty} \frac{(-1)^{k-1}}{(2k+1)\pi^{2k}} = 1 - \pi \operatorname{arctg} \frac{1}{\pi} .$$

Neka je potrebno odrediti narednih n članova niza (S_r) . Nije teško videti da važi rekurentna relacija

$$(3.4.10) \quad S_r = \frac{1}{2r+1} - \pi^2 S_{r-1} \quad (r=1, 2, \dots, n) .$$

Startujući sa (3.4.9) i korišćenjem rekurentne relacije (3.4.10) dobijamo rezultate koji su dati u tabeli 3.4.1, pri čemu je uzeto $n=12$. Broj u zagradi ukazuje na decimalni eksponent, na primer, $3.2(-2)=3.2 \cdot 10^{-2}$. Sva izračunavanja su sprovedena na programabilnom kalkulatoru CASIO FX-702P, sa aritmetikom od 12 dekadnih cifara u mantisi. Dobijeni članovi niza (S_r), su evidentno pogrešni, jer je $S_r > 0$ za svako r . U ovoj tabeli, rezultati su zaokrugljeni na 10 značajnih cifara.

Tabela 3.4.1

r	S_r pomoću (3.4.10)	S_r pomoću (3.4.11)
0	0.3185831012 (-1)	0.3185831012 (-1)
1	0.1890441554 (-1)	0.1890441558 (-1)
2	0.1342089717 (-1)	0.1342089682 (-1)
3	0.1039819711 (-1)	0.1039820054 (-1)
4	0.8485019141 (-2)	0.8484985269 (-2)
5	0.7165308652 (-2)	0.7165642950 (-2)
6	0.6204315120 (-2)	0.6201015723 (-2)
7	0.5432530853 (-2)	0.5465094591 (-2)
8	0.5206599001 (-2)	0.4885207787 (-2)
9	0.1244506534 (-2)	0.4416510673 (-2)
10	0.3533626045 (-1)	0.4029834444 (-2)
11	-0.3052766508 (0)	0.3705389102 (-2)
12	0.3052959776 (1)	0.3429275407 (-2)

Može se postaviti pitanje zbog čega su ovi rezultati pogrešni. Odgovor leži u tzv. slaboj uslovljenosti rekurentne relacije (3.4.10). Stavljujući $r=1,2,\dots,n$ na osnovu (3.4.10) dobijamo sistem od n linearnih jednačina

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ \pi^2 & 1 & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \pi^2 & 1 \end{bmatrix} \cdot \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_{n-1} \\ S_n \end{bmatrix} = \begin{bmatrix} 1/3 - \pi^2 S_0 \\ 1/5 \\ \vdots \\ 1/(2n-1) \\ 1/(2n+1) \end{bmatrix},$$

čija je matrica dvodijagonalna. Slično sistemu jednačina iz primera 2.1.4, i ovde imamo slabouslovljeni sistem, kod koga male promene (perturbacije) u sistemu, nastale kao posledica

zaokrugljivanja, prouzrokuju velike promene u rešenju, što vodi pogrešnom rezultatu. O slabouslovljenim sistemima biće reči u odeljku 4.2.6.

Ako se, međutim, najpre izračuna, sa potrebnom tačnošću, poslednji član niza S_n , a rekurentna relacija predstavi u obliku

$$(3.4.11) \quad S_{r-1} = \frac{1}{\pi^2} \left(\frac{1}{2r+1} - S_r \right) \quad (r=n, \dots, 1),$$

mi dobijamo dobro uslovljeni proces. Ekvivalentni sistem linearnih jednačina ima oblik

$$\begin{bmatrix} \pi^2 & 0 & \dots & 0 & 0 \\ 1 & \pi^2 & & 0 & 0 \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & & \pi^2 & 0 \\ 0 & 0 & & 1 & \pi^2 \end{bmatrix} \cdot \begin{bmatrix} S_{n-1} \\ S_{n-2} \\ \vdots \\ \vdots \\ S_1 \\ S_0 \end{bmatrix} = \begin{bmatrix} 1/(2n+1)-S_n \\ 1/(2n-1) \\ \vdots \\ \vdots \\ 1/5 \\ 1/3 \end{bmatrix}.$$

Da bismo odredili S_n , posmatraćemo stepeni red

$$(3.4.12) \quad f(x) = \sum_{k=1}^{+\infty} \frac{x^{k-1}}{2k+2n+1} = \sum_{k=0}^{+\infty} \frac{x^k}{2k+2n+3},$$

čiji je poluprečnik konvergencije jednak jedinici, i na čije sumiranje ćemo primeniti Euler-Abelovu transformaciju sa $m \rightarrow +\infty$. Kako je $a_j = \frac{1}{2j+2n+3}$, imamo

$$\Delta^k a_j = \sum_{i=0}^k (-1)^i \binom{k}{i} a_{j+k-i} = \sum_{i=0}^k \frac{(-1)^i \binom{k}{i}}{2(j+k-i)+2n+3},$$

tj.

$$\Delta^k a_j = -\frac{1}{2} \sum_{i=0}^k \frac{(-1)^i \binom{k}{i}}{i-(j+n+k+\frac{3}{2})}.$$

Korišćenjem identiteta (3.4.8), sa $a=-\frac{n+k+3}{2}=-b$ i $j=0$, dobijamo

$$\Delta^k a_0 = - \frac{1}{2} \cdot \frac{(-1)^{k+1}}{b^{(k+1)}} \frac{k!}{b},$$

gde je $p^{(s)} = p(p-1) \dots (p-s+1)$.

Dakle, na osnovu (3.4.7) imamo

$$(3.4.13) \quad f(x) = \frac{1}{2(1-x)} \sum_{k=0}^{+\infty} \frac{(-1)^k k!}{b^{(k+1)}} \left(\frac{x}{1-x}\right)^k.$$

Ako stavimo $x = -1/\pi^2$, na osnovu (3.4.12) i (3.4.13) zaključujemo da je

$$S_n = \frac{1}{\pi} f(-1/\pi^2) = \frac{1}{2} \sum_{k=0}^{+\infty} \frac{k!}{(\pi^2 + 1)^{k+1} b^{(k+1)}}.$$

Numerički red u poslednjoj jednakosti brzo konvergira. Parcijalne sume ovog reda za $n=12$ su redom

$$\begin{aligned} S_{12}^{(0)} &= 0.3407395124 \cdot 10^{-2}, \\ S_{12}^{(1)} &= 0.3429014381 \cdot 10^{-2}, \\ S_{12}^{(2)} &= 0.3429271021 \cdot 10^{-2}, \\ S_{12}^{(3)} &= 0.3429275314 \cdot 10^{-2}, \\ S_{12}^{(4)} &= 0.3429275404 \cdot 10^{-2}, \\ S_{12}^{(5)} &= 0.3429275407 \cdot 10^{-2}. \end{aligned}$$

Uzimajući $S_{12} \approx S_{12}^{(5)}$ i primenom rekurentne relacije (3.4.11) dobijamo rezultate koji su dati, takodje, u tabeli 3.4.1.

Primetimo da se dobijena vrednost za S_0 poklapa sa tačnom vrednošću (3.4.9) na svih 10 značajnih cifara. Ovo ukazuje da su sve vrednosti niza (S_r) dobijene sa visokom tačnošću.

Za ubrzavanje konvergencije redova postoje i transformacije koje se zasnivaju na ubrzavanju konvergencije nizova. Medju njima postoje kako linearne, tako i nelinearne transformacije.

Neka je $\{S_n\}$ niz parcijalnih sum alternativnog reda (3.4.2), tj.

$$(3.4.14) \quad S_n = \sum_{k=0}^{n-1} (-1)^k a_k \quad (n=1, 2, \dots).$$

Najprostija linearna transformacija je definisana aritmetičkom sredinom sukcesivnih parcijalnih sum, poznata kao Cesàro-ova transformacija,

$$S'_n = C(S_n) = \frac{1}{2}(S_n + S_{n+1}) \quad (n=1, 2, \dots).$$

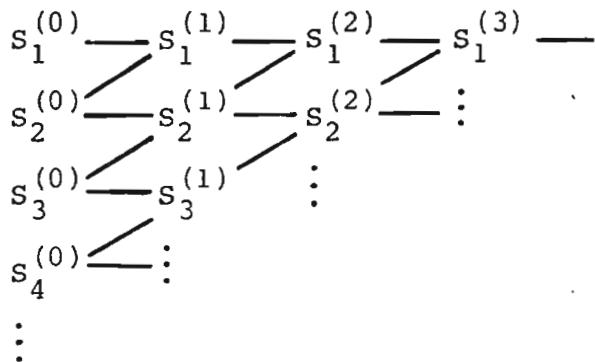
Niz $\{S'_n\}$ brže konvergira od niza $\{S_n\}$ ka istoj granici S ($\lim S'_n = \lim S_n = S$). Daljom primenom iste transformacije na niz $\{S'_n\}$ dobija se ubrzani niz $\{S''_n\}$, gde je

$$S''_n = C(S'_n) = \frac{1}{2}(S'_n + S'_{n+1}) \quad (n=1, 2, \dots).$$

Uopšte, polazeći od niza parcijalnih suma (3.4.14) mogu se sukcesivno konstruisati nizovi $\{S_n^{(m)}\}$ ($m=1, 2, \dots$) pomoću

$$(3.4.15) \quad S_n^{(m)} = C(S_n^{(m-1)}) = \frac{1}{2}(S_n^{(m-1)} + S_{n+1}^{(m-1)}) \quad (n=1, 2, \dots),$$

gde smo stavili $S_n^{(0)} = S_n$. Šematski, postupak (3.4.15) se može interpretirati kroz konstrukciju tzv. S-tabele:



Algoritamski se ovaj postupak može iskazati kroz sledeća četiri koraka:

$$1^{\circ} \quad s_1 := a_0, \quad z := 1, \quad n := 0;$$

$$2^{\circ} \quad n := n+1, \quad z := -z, \quad s_{n+1} := s_n + za_n;$$

$$3^{\circ} \quad \text{Za } k=n, \dots, 1, \quad s_k := \frac{1}{2}(s_{k+1} + s_k);$$

4^o Ako je

$$|s_1 - s_2| > \varepsilon \text{ preći na } 2^{\circ}, \\ \leq \varepsilon \text{ kraj izračunavanja } S := s_1.$$

Primetimo da ovaj postupak ne zahteva memorijski prostor za pamćenje čitave S -tabele. Naime, u n -tom koraku primene ovog postupka pamte se samo dijagonalni elementi $s_{n+1}^{(0)}, s_n^{(1)}, \dots, s_2^{(n-1)}, s_1^{(n)}$, koji su označeni redom sa $s_{n+1}, s_n, \dots, s_2, s_1$.

Sa $z := -z$ obezbedjuje se neophodna promena znaka kod izračunavanja parcijalnih suma alternativnog reda.

Može se pokazati da nizovi $\{s_n^{(m)}\}_{n \in \mathbb{N}}$ i $\{s_n^{(m)}\}_{m \in \mathbb{N}}$ (po kolonama i vrstama u S -tabeli), kao i po dijagonalama konvergiraju ka S .

Primedba 3.4.1. Kako je $s_{n+1} = s_n + (-1)^{n+1} a_{n+1}$ imamo da je

$$s_n^{(1)} = \frac{1}{2}(s_n + s_{n+1}) = s_n + \frac{1}{2}(-1)^{n+1} a_{n+1},$$

$$s_n^{(2)} = \frac{1}{2}(s_n^{(1)} + s_{n+1}^{(1)}) = s_n + \frac{1}{4}(3(-1)^{n+1} a_{n+1} + (-1)^{n+2} a_{n+2}).$$

Matematičkom indukcijom jednostavno zaključujemo da važi sledeća reprezentacija

$$(3.4.16) \quad s_n^{(m)} = s_n + \frac{1}{2^m} \sum_{k=1}^m \alpha_{mk} (-1)^{n+k} a_{n+k},$$

gde su koeficijenti α_{mk} dati rekurzivno pomoću

$$\alpha_{m1} = 2^{m-1} + \alpha_{m-1, i},$$

$$\alpha_{mk} = \alpha_{m-1, k-1} + \alpha_{m-1, k} \quad (k=2, \dots, m-1),$$

$$\alpha_{mm} = \alpha_{m-1,m-1}.$$

Primetimo da je $\alpha_{m1} = 2^m - 1$ i $\alpha_{mm} = 1$.

Na osnovu prethodnog vidimo da se elementi S-tabele mogu izračunavati i pomoću (3.4.16). Međutim, ovakav postupak bi bio numerički neefikasan zbog velikog broja računskih operacija.

Primer 3.4.3. Formirajmo S-tabelu za određivanje sume alternativnog reda

$$S = \sum_{k=0}^{\infty} \frac{(-1)^k}{k+1} = \log 2 = 0.693147\dots$$

sa tri tačne decimale. Ovde je $a_k = \frac{1}{k+1}$, $s_1^{(0)} = 1$, $s_{k+1}^{(0)} = s_k^{(0)} + (-1)^{k+1} a_{k+1}$. Primenom (3.4.15), saglasno prethodno navedenim algoritamskim koracima $1^0 - 4^0$ dobijamo S-tabelu:

1.0000	0.7500	0.7083	0.6978	0.6947	0.6936	0.6932
0.5000	0.6666	0.6874	0.6916	0.6926	0.6929	
0.8333	0.7083	0.6958	0.6937	0.6932		
0.5833	0.6833	0.6916	0.6928			
0.7833	0.7000	0.6940				
0.6167	0.6881					
0.7595						

Pri ovome smo sva izračunavanja sprovedeli tako što smo sve međurezultate zaokrugljivali na četiri decimale. Kako je

$$|s_1^{(6)} - s_2^{(5)}| = |0.6932 - 0.6929| < 0.5 \cdot 10^{-3},$$

zaključujemo da je $S \approx 0.693$. Ako se želi postići veća tačnost, izračunavanja mogu da se sprovode u aritmetici sa većim brojem cifara u mantisi. Čitaocu prepuštamo da izvrši analizu prostiranja grešaka kroz kolone S-tabele, znajući relativne mašinske greške elemenata iz prve kolone r_n ($|r_n| \leq \frac{1}{2} \cdot 10^{-\ell+1}$, gde je ℓ broj cifara mantise).

Ako sva izračunavanja sprovodimo, kao u primeru 3.4.2, sa aritmetikom od 12 cifara u mantisi, u zavisnosti od zahtevane tačnosti ϵ , dobijamo rezultate koji su navedeni u tabeli 3.4.2.

Tabela 3.4.2

ϵ	m	$s_1^{(m)}$
10^{-2}	4	0.6947016667
10^{-3}	6	0.6933779762
10^{-4}	8	0.6931842138
10^{-5}	10	0.6931536346
10^{-6}	13	0.6931476997
10^{-7}	16	0.6931472258
10^{-8}	19	0.6931471847

U drugoj koloni tabele dat je broj m koji ukazuje na broj sukcesivnih primena transformacije za postizanje zadate tačnosti ϵ . Drugim rečima, m je najmanji prirodan broj za koji je $|s_1^{(m)} - s_2^{(m-1)}| \leq \epsilon$. Prva pogrešna cifra u rezultatu $s_1^{(m)}$ je podvučena. Primetimo da direktno sumiranje reda "član po član" zahteva, prema Leibnitzovom kriterijumu, približno $\lceil 1/\epsilon \rceil$ članova. Na primer, za tačnost $\epsilon = 10^{-8}$ potrebno je sabrati fantastičnih 100 miliona članova.

Na kraju ovog odeljka napomenimo da postoji čitava klasa nelinearnih transformacija koje se primenjuju kod sumiranja redova ([3]-[10], [29]). Jedna od najprostijih je Δ^2 -transformacija definisana pomoću

$$s_n' = T(s_n) = s_n - \frac{(\Delta s_n)^2}{\Delta^2 s_n} \quad (n=1, 2, \dots)$$

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

Primedba 3.4.2. Ako se pretpostavi asimptotsko ponašanje parcijalne sume s_k u obliku $s_k \sim A + Be^{-\alpha k}$ ($\alpha > 0$), tada eliminacijom B i α iz ove asimptotske relacije za $k=n, n+1, n+2$ dobijamo

$$A \sim s_n - \frac{(s_{n+1} - s_n)^2}{s_{n+2} - 2s_{n+1} + s_n},$$

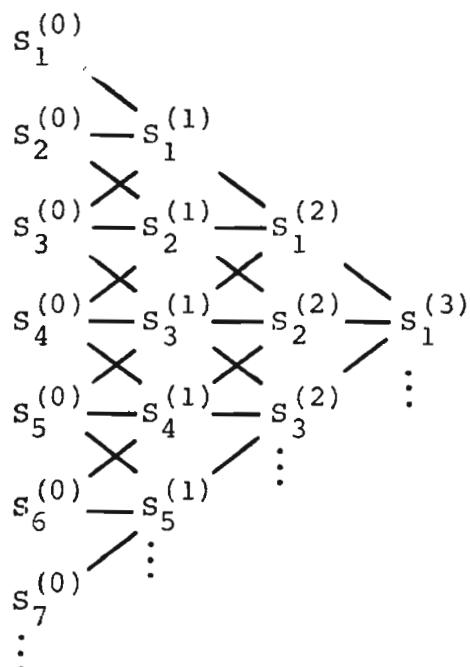
tj. $A \sim T(s_n)$. Primetimo da je

$$S = \lim_{k \rightarrow +\infty} S_k = \lim_{k \rightarrow +\infty} (A + Be^{-\alpha k}) = A.$$

Kao i Cesàro-ova transformacija i Δ^2 -transformacija se može sukcesivno primenjivati

$$S_n^{(m)} = T(S_n^{(m-1)}) = S_n^{(m-1)} - \frac{(\Delta S_n^{(m-1)})^2}{\Delta^2 S_n^{(m-1)}} \quad (n=1, 2, \dots),$$

gde smo, kao i ranije, stavili $S_n^{(0)} \equiv S_n$. Naravno, konstrukcija S-tabele, u ovom slučaju zbog nelinearnosti, je nešto složenija i može se šematski predstaviti u obliku:



1.000000				
0.500000	0.700000			
0.833333	0.690476	0.693277		
0.583333	0.694444	0.693163	0.693148	
0.783333	0.692424			
0.616667	0.693590			
0.759524				

pri čemu smo koristili aritmetiku sa šestorazrednom mantisom. Dobijeni rezultat u četvrtoj koloni ima tačnih pet decimala.

1.3.5. Elementi teorije verižnih razlomaka

Izraz oblika

$$(3.5.1) \quad a_0 + \cfrac{b_1}{a_1 + \cfrac{b_2}{a_2 + \cfrac{b_3}{a_3 + \dots}}}$$

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

$$(3.5.2) \quad \left[a_0; \frac{b_1}{a_1}, \frac{b_2}{a_2}, \frac{b_3}{a_3}, \dots \right], \quad a_0 + \frac{b_1}{|a_1|} + \frac{b_2}{|a_2|} + \frac{b_3}{|a_3|} + \dots,$$

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

pri čemu su a_k , b_k , u opštem slučaju, promenljive. U specijalnom slučaju, oni mogu biti brojevi (realni ili kompleksni), matrice, operatori, itd. Element a_k se naziva k-ti parcijalni imenilac, b_k k-ti parcijalni brojilac, a a_0 slobodan član. U našem izlaganju a_k i b_k su realne ili kompleksne veličine.

Ako je koeficijenta $a_0 = 0$, tada se u ovim notacijama, obično, on izostavlja. Na primer, pišemo samo

$$\left[\frac{b_1}{a_1}, \frac{b_2}{a_2}, \frac{b_3}{a_3}, \dots \right].$$

Verižni razlomak može biti konačan, na primer,

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

Razmotrimo sada niz konačnih verižnih razlomaka koji se dobijaju iz (3.5.1), tj. (3.5.2), uzimajući samo konačan broj članova. Tako, za $k \in \mathbb{N}$, stavimo

$$(3.5.3) \quad R_k = \frac{P_k}{Q_k} = \left[a_0; \frac{b_1}{a_1}, \frac{b_2}{a_2}, \dots, \frac{b_k}{a_k} \right].$$

Ako egzistira vrednost verižnog razlomka (3.5.1), tada se ona definiše kao

$$(3.5.4) \quad R = \lim_{k \rightarrow +\infty} R_k.$$

Razlomak (3.5.3) nazivaćemo k-tom aproksimacijom, ili k-tim konvergentom, verižnog razlomka (3.5.1).

Ako uzmemo

$$(3.5.5) \quad P_0 = a_0, \quad Q_0 = 1, \quad P_{-1} = 1, \quad Q_{-1} = 0,$$

indukcijom se jednostavno dokazuju rekurentne relacije za P_k i Q_k :

$$(3.5.6) \quad \begin{aligned} P_k &= a_k P_{k-1} + b_k P_{k-2} \\ Q_k &= a_k Q_{k-1} + b_k Q_{k-2} \end{aligned} \quad (k=1, 2, \dots).$$

Primetimo da su P_k i Q_k dva rešenja diferencne jednačine

$$y_k - a_k y_{k-1} - b_k y_{k-2} = 0.$$

Na osnovu (3.5.5) i (3.5.6) jednostavno se dokazuju jednakosti

$$(3.5.7) \quad R_k - R_{k-1} = (-1)^{k+1} \frac{b_1 b_2 \dots b_k}{Q_{k-1} Q_k} \quad (k=1, 2, \dots)$$

i

$$(3.5.8) \quad R_k - R_{k-2} = (-1)^k \frac{b_1 b_2 \dots b_{k-1} a_k}{Q_{k-2} Q_k} \quad (k=2, 3, \dots).$$

Neka su w_k proizvoljni brojevi različiti od nule. Nad verižnim razlomkom (3.5.1) možemo izvršiti ekvivalentnu transformaciju odgovarajućim množenjem sa w_k , tako da, u notaciji (3.5.2), imamo

$$(3.5.9) \quad \left[a_0; \frac{w_1 b_1}{w_1 a_1}, \frac{w_1 w_2 b_2}{w_2 a_2}, \frac{w_2 w_3 b_3}{w_3 a_3}, \dots \right].$$

Ako sa s_k i t_k označimo brojilac i imenilac u k-tom konvergentu verižnog razlomka (3.5.9) imamo

$$s_k = w_1 w_2 \dots w_k p_k \quad i \quad t_k = w_1 w_2 \dots w_k q_k.$$

Primetimo da je pri ekvivalentnoj transformaciji verižnog razlomka vrednost k-tog konvergenta invarijantna, tj.

$$\frac{s_k}{t_k} = \frac{p_k}{q_k} = r_k.$$

Ekvivalentnom transformacijom, sa specijalnim izborom brojeva w_k , verižni razlomak se može uprostiti.

Neka su $b_k \neq 0$ ($k=1, 2, \dots$). Izaberimo konstante w_k tako da su parcijalni brojioci u (3.5.9) jednaki jedinici, tj.

$$w_1 b_1 = 1, \quad w_1 w_2 b_2 = 1, \dots, \quad w_{k-1} w_k b_k = 1, \dots .$$

Tada imamo

$$w_1 = \frac{1}{b_1}, \quad w_2 = \frac{b_1}{b_2}, \quad w_3 = \frac{b_2}{b_1 b_3}, \dots,$$

ili u opštem slučaju,

$$w_{2k} = \frac{b_1 b_3 \dots b_{2k-1}}{b_2 b_4 \dots b_{2k}} \quad i \quad w_{2k+1} = \frac{b_2 b_4 \dots b_{2k}}{b_1 b_3 \dots b_{2k+1}} .$$

Zamenom ovih vrednosti u (3.5.9) dobijamo ekvivalentni verižni razlomak kod koga su parcijalni brojioci jednaki jedinicama.

Slično, ako je $a_k \neq 0$ ($k=1, 2, \dots$), stavljajući u (3.5.9) $w_k = 1/a_k$ dobijamo ekvivalentan verižni razlomak

$$(3.5.10) \quad \left[a_0; \frac{c_1}{1}, \frac{c_2}{1}, \frac{c_3}{1}, \dots \right],$$

gde su

$$c_1 = b_1/a_1, \quad c_2 = b_2/(a_1 a_2), \quad c_3 = b_3/(a_2 a_3), \text{ itd.}$$

Ako brojeve w_k izaberemo tako da je

$$w_1 a_1 = 1, \quad w_{k-1} w_k b_k + w_k a_k = 1 \quad (k=2, 3, \dots),$$

tada se (3.5.9) svodi na oblik

$$(3.5.11) \quad \left[a_0; \frac{\alpha_1}{1}, \frac{\alpha_2}{1-\alpha_2}, \dots, \frac{\alpha_k}{1-\alpha_k}, \dots \right],$$

koji je poznat kao Eulerov verižni razlomak.

Druga rekurentna relacija u (3.5.6) za razlomak (3.5.11) (umesto Q_k uzimamo ranije uvedenu oznaku T_k) postaje

$$T_1 = 1 \cdot T_0 + \alpha_1 T_{-1}, \quad T_k = (1 - \alpha_k) T_{k-1} + \alpha_k T_{k-2} \quad (k \geq 2),$$

sa $T_{-1} = 0$ i $T_0 = 1$. Primetimo da se, za svako $k \geq 1$, dobija $T_k = 1$.

Verižni razlomak (3.5.1) se može svesti na Eulerov oblik (3.5.11) ako i samo ako su $Q_i \neq 1$ ($i=1, 2, \dots$). Pri tome su

$$(3.5.12) \quad \alpha_1 = \frac{b_1}{Q_1}, \quad \alpha_i = \frac{Q_{i-2}}{Q_i} b_i \quad (i=2, 3, \dots)$$

i

$$(3.5.13) \quad R_k = a_0 + \sum_{i=1}^k (-1)^{i+1} \alpha_1 \alpha_2 \dots \alpha_i \quad (k=1, 2, \dots).$$

Pretpostavimo na dalje da je $Q_i \neq 0$ ($i=1, 2, \dots$). Na osnovu (3.5.13) imamo

$$(3.5.14) \quad \lim_{k \rightarrow \infty} R_k = a_0 + \sum_{i=1}^{+\infty} (-1)^{i+1} \alpha_1 \alpha_2 \dots \alpha_i,$$

tj. red na desnoj strani u poslednjoj jednakosti i verižni razlomak (3.5.1) su ekvikonvergentni. Ako je razlomak (3.5.1) konvergentan, tj. ako važi (3.5.4), na osnovu (3.5.13) i (3.5.14) zaključujemo da je

$$(3.5.15) \quad |R_k - R| = \left| \sum_{i=k+1}^{+\infty} (-1)^{i+1} \alpha_1 \alpha_2 \dots \alpha_i \right|.$$

Na osnovu (3.5.7) i (3.5.8) važe sledeći rezultati:

Teorema 3.5.1. Ako su $a_k, b_k > 0$ tada važe nejednakosti

$$R_1 > R_3 > \dots > R_{2k-1} > \dots, \quad R_0 < R_2 < \dots < R_{2k} < \dots,$$

$$R_{2m-1} > R_{2k} \quad (\text{za svako } m \text{ i } k).$$

Teorema 3.5.2. Ako su a_k i b_k pozitivni i takvi da je $b_k \leq a_k$ i $a_k \geq \epsilon > 0$ ($k=1,2,\dots$), gde je ϵ neka konstanta, tada je verižni razlomak (3.5.1) konvergentan.

Ako su $a_k b_k > 0$, na osnovu (3.5.12), (3.5.5), (3.5.6) zaključujemo da je

$$a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_i^{\alpha_i} = \frac{b_1 b_2 \cdots b_i}{Q_{i-1} Q_i} > 0.$$

Tada iz (3.5.15) na osnovu Leibnitzovog kriterijuma za alternativne redove dobijamo sledeću ocenu

$$|R_k - R| \leq \frac{b_1 b_2 \cdots b_{k+1}}{Q_k Q_{k+1}}.$$

Za slučaj kada su a_k i b_k kompleksni, važi sledeća teorema:

Teorema 3.5.3. Ako je $|a_k| - |b_k| \geq 1$ ($k=1,2,\dots$), verižni razlomak (3.5.1) konvergira.

Kao što smo videli, ako je $a_k \neq 0$ ($k=1,2,\dots$), ekvivalentni oblik verižnog razlomka (3.5.1) može se dati u obliku (3.5.10). Ne umanjujući opštost, na stanovišta konvergencije umesto (3.5.10), može se posmatrati verižni razlomak

$$(3.5.16) \quad \left[\frac{1}{1}, \frac{c_2}{1}, \frac{c_3}{1}, \dots \right].$$

Definicija 3.5.1. Za verižni razlomak (3.5.16) kažemo da zadovoljava fundamentalne nejednakosti, ako postoji niz nenegativnih brojeva $\{r_i\}_{i \in \mathbb{N}}$ takav da za parcijalne brojioce, za svako $i \in \mathbb{N}$, važe nejednakosti

$$r_{i-2} r_i |c_i| + |c_{i+1}| \leq r_i |1 + c_i + c_{i+1}|,$$

pri čemu treba uzeti $c_1 = 0$, $r_0 = r_{-1} = 0$.

Ako verižni razlomak (3.5.16) zadovoljava fundamentalne nejednakosti, može se dokazati (videti, na primer, [30]) da je tada $Q_k \neq 0$ ($k=1,2,\dots$) i da za brojeve

$$\alpha_i = \frac{Q_{i-2}}{Q_i} c_i$$

važe nejednakosti

$$(3.5.17) \quad |\alpha_i| \leq r_{i-1} \quad (i=2,3,\dots).$$

Dakle, u tom slučaju (3.5.16) može da se ekvivalentno transformiše na Eulerov razlomak, kod koga je $a_0=0$ i $\alpha_1=1$. Ako je red $\sum_{i=2}^{+\infty} r_1 r_2 \dots r_{i-1}$ konvergentan, na osnovu (3.5.13) i (3.5.17), zaključujemo da je verižni razlomak (3.5.16) konvergentan.

Teorema 3.5.4. Neka su c_i ($i=2,3,\dots$) funkcije jedne promenljive (ili više promenljivih), definisane u nekoj oblasti D , u kojoj je

$$|c_i| \leq \frac{1}{4} \quad (i=2,3,\dots).$$

Tada važe tvrdjenja:

- a) Verižni razlomak (3.5.16) ravnomerno konvergira u D ;
- b) Vrednosti svih konvergenata verižnog razlomka (3.5.16), kao i njegova vrednost, pripadaju oblasti kruga

$$K = \left\{ z \mid |z - \frac{4}{3}| \leq \frac{2}{3} \right\};$$

c) Konstanta $1/4$ u oblasti kruga K je najbolja mogućnost, tj. konstanta se ne može povećati, a krug K se ne može smanjiti.

1.3.6. Razvoj racionalne funkcije u verižni razlomak

Od interesa je proučiti razlaganje racionalne funkcije u verižni razlomak.

Neka je

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

Tada je

$$f(x) = \frac{1}{\frac{c_{00}}{c_{10}} + \frac{c_{00}+c_{01}x+c_{02}x^2+\dots}{\frac{c_{10}}{c_{10}+c_{11}x+c_{12}x^2+\dots} - \frac{c_{00}}{c_{10}}}} = \frac{c_{10}}{c_{00}+xf_1(x)},$$

gde je

$$f_1(x) = \frac{c_{20}+c_{21}x+c_{22}x^2+\dots}{c_{10}+c_{11}x+c_{12}x^2+\dots}$$

i

$$c_{2j} = c_{10}c_{0,j+1} - c_{00}c_{1,j+1} \quad (j=0,1,\dots).$$

Nastavljajući ovaj proces dobija se razvoj u verižni razlomak

$$f(x) = \left[\frac{c_{10}}{c_{00}}, \frac{c_{20}x}{c_{10}}, \frac{c_{30}x}{c_{20}}, \dots \right],$$

gde su

$$c_{ij} = - \begin{vmatrix} c_{i-2,0} & c_{i-2,j+1} \\ c_{i-1,0} & c_{i-1,j+1} \end{vmatrix} \quad (i=2,3,\dots).$$

Primer 3.6.1. Za funkciju $x \mapsto f(x) = \frac{1-x}{1-5x+6x^2}$ imamo

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

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

Obrnuto, verižni razlomak

$$f(x) = \frac{b_1}{x + \frac{b_2}{x + \frac{b_3}{x + \dots}}} = \left[\frac{b_1}{x}, \frac{b_2}{x}, \frac{b_3}{x}, \dots, \frac{b_n}{x} \right]$$

može se predstaviti u obliku racionalne funkcije ([31])

$$f(x) = \left(\frac{c_{2n}x + c_{4n}x^3 + c_{6n}x^5 + \dots}{c_{1n} + c_{3n}x^2 + c_{5n}x^4 + \dots} \right)^{(-1)^n},$$

pri čemu se koeficijenti c_{kn} ($k=1, \dots, n+1$) mogu izračunati po sledećem algoritmu

$$c_{11} = b_n;$$

$$c_{i+1,i} = 1 \quad (i=1, \dots, n);$$

$$c_{ki} = \begin{cases} c_{k,i-1} b_{n-i+1} & (k+i \text{ parno}), \\ c_{k,i-1} + c_{k-1,i-1} & (k+i \text{ neparno}, k > 1), \\ c_{1,i-1} & (i \text{ parno}), \end{cases}$$

pri čemu k uzima vrednost $k=i, i-1, \dots, 1$ za svako $i=2, \dots, n$.

Ilustracije radi koeficijenti c_{ki} su dati u tabeli 3.6.1.

Tabela 3.6.1

i \ k	1	2	3	4	5
1	b_n	1	0	0	0
2	b_n	b_{n-1}	1	0	0
3	$b_n b_{n-2}$	$b_n + b_{n-1}$	b_{n-2}	1	0
4	$b_n b_{n-2}$	$(b_n + b_{n-1}) b_{n-3}$	$b_n + b_{n-1} + b_{n-2}$	b_{n-3}	1

Algoritam se može uprostiti zamenom matrice $[c_{ki}]$ $(n+1) \times n$ nizom $c(k) = c_{ki}$ ($k=1, \dots, n+1$), čime se vrši uštada memorijskog prostora. Modifikovani algoritam je:

$$c(k) = 1 \quad (k=1, \dots, n+1);$$

$$c(k) = \begin{cases} c(k) b_{n-i+1} & (k+i \text{ parno}), \\ c(k) + c(k-1) & (k+i \text{ neparno}, k > 1), \end{cases}$$

pri čemu $k=i, i-1, \dots, 1$ za svako $i=1, \dots, n$.

Primer 3.6.2. Neka je

$$f(x) = \left[\frac{1}{x}, \frac{2}{x}, \frac{3}{x}, \frac{4}{x} \right],$$

tj. $b_k = k$ ($k=1, \dots, 4$). Kako je $n=4$, iz četvrte vrste tabele 3.6.1 ($i=4$) nalazimo

$$c_{14} = b_4 b_2 = 8, \quad c_{24} = (b_4 + b_3) b_1 = 7,$$

$$c_{34} = b_4 + b_3 + b_2 = 9, \quad c_{44} = b_1 = 1, \quad c_{54} = 1,$$

tj.

$$f(x) = \frac{7x^3 + x}{8 + 9x^2 + x^4}.$$

1.3.7. Algoritmi za izračunavanje verižnih razlomaka

U ovom odeljku razmotrićemo nekoliko algoritama za izračunavanje n-tog konvergenta verižnog razlomka, datog pomoću (3.5.3). Pretpostavimo da je $a_0=0$, tj. da R_n ima oblik

$$(3.7.1) \quad R_n = \frac{P_n}{Q_n} = \left[\frac{b_1}{a_1}, \frac{b_2}{a_2}, \dots, \frac{b_n}{a_n} \right].$$

Algoritam A1. Za (3.7.1) definišemo rekurentnu relaciju unazad pomoću

$$(3.7.2) \quad y_{n,k} = \frac{b_k}{a_k + y_{n,k+1}} \quad (k=n, n-1, \dots, 1),$$

pri čemu je $y_{n,n+1}=0$. Tada je $R_n=y_{n,1}$.

Algoritam A2. Algoritam se zasniva na primeni rekurentnih formula

$$c_k = \frac{b_k}{f_{k-1}}, \quad f_k = a_k + c_k, \quad q_k = -\frac{c_k q_{k-1}}{f_k}, \quad R_k = R_{k-1} + q_k \quad (k=2, \dots, n),$$

startujući sa $f_1=a_1$, $q_1=R_1=b_1/a_1$.

Algoritam A3. Ovaj algoritam se zasniva na primeni rekurentnih relacija (3.5.6), tj.

$$(3.7.3) \quad P_k = a_k P_{k-1} + b_k P_{k-2} \quad (k=1, 2, \dots, n),$$

$$Q_k = a_k Q_{k-1} + b_k Q_{k-2}$$

sa startnim vrednostima $P_0=0$, $P_{-1}=1$; $Q_0=1$, $Q_{-1}=0$. Tada je $R_n = P_n/Q_n$. Ovo je poznati Eulerov algoritam.

Algoritam A4. Startujući sa $y_1=1$ i $y_2=a_n$ odredjujemo

$$(3.7.4) \quad y_k = a_{n+2-k}y_{k-1} + b_{n+3-k}y_{k-2} \quad (k=3, \dots, n).$$

Tada je

$$(3.7.5) \quad R_n = \frac{b_1 y_n}{a_1 y_n + b_2 y_{n-1}}.$$

Poslednji algoritam je matematički ekvivalentan Eulerovom algoritmu A3, ali su oni u numeričkom smislu različiti ([24]).

Primetimo da Eulerov algoritam zahteva dvostruku primenu iste rekurentne relacije (3.7.3) (za P_k i Q_k). Ta relacija je oblika

$$S_k = a_k S_{k-1} + b_k S_{k-2} \quad (k=1, 2, \dots, n),$$

što se može predstaviti u matričnom obliku kao trougaoni sistem linearnih jednačina $\vec{As} = \vec{f}$, tj.

$$\begin{bmatrix} 1 & & & & \\ -a_2 & 1 & & & \\ -b_3 & -a_3 & 1 & & \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & & -b_n & -a_n & 1 \end{bmatrix} \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ \vdots \\ \vdots \\ S_n \end{bmatrix} = \begin{bmatrix} a_1 S_0 + b_1 S_{-1} \\ b_2 S_0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}.$$

Kako je $\vec{s} = A^{-1} \vec{f}$, a nama je potrebna samo poslednja koordinata u vektoru \vec{s} , tj. S_n , to je za njeno određivanje dovoljno poznavati samo poslednju vrstu u matrici A^{-1} . Označimo elemente ove vrste redom sa y_n, y_{n-1}, \dots, y_1 . Iz jednakosti $A^{-1}A=I$ dobijamo (3.7.4). Kako su u vektoru \vec{f} samo prve dve koordinate različite od nule imamo

$$S_n = y_n (a_1 S_0 + b_1 S_{-1}) + y_{n-1} b_2 S_0,$$

tj.

$$S_n = b_1 y_n S_{-1} + (a_1 y_n + b_2 y_{n-1}) S_0.$$

Da bismo dobili P_n , u poslednjoj jednakosti treba staviti $S_{-1}=P_{-1}=1$ i $S_0=P_0=0$, dok za Q_n treba staviti $S_{-1}=Q_{-1}=0$ i $S_0=Q_0=1$. Na taj način dobijamo (3.7.5).

U tabeli 3.7.1 dat je pregled broja operacija za napred navedena četiri algoritma.

Tabela 3.7.1

Algoritam	Broj operacija			Ukupan broj operacija
	Sabiranja	Množenja	Deljenja	
A1	n-1	0	n	2n-1
A2	2n-2	n-1	2n-1	5n-4
A3	2n-3	4n-6	1	6n-8
A4	n-1	2n-2	1	3n-2

Iz tabele vidimo da algoritam A1, definisan pomoću (3.7.2), zahteva najmanji broj operacija. Takođe, njegova programska realizacija je jednostavna. Međutim, u nekim slučajevima algoritam A4 može biti efikasniji od A1, ako je "cena deljenja znatno veća od "cene množenja". Za cenu operacije se može uzeti, na primer, procesorsko vreme potrebno za realizaciju odgovarajuće operacije.

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

Euler je dokazao razvoj

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

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

$$\tan x = \left[\frac{x}{1}, \frac{-x^2}{3}, \frac{-x^2}{5}, \dots, \frac{-x^2}{2n+1}, \dots \right]$$

dokazao Lambert. Poslednji razvoj konvergira za svako x , za koje je funkcija $x \mapsto \tan x$ neprekidna.

Primedba 3.7.1. Za izračunavanje vrednosti $\operatorname{tg} x$ pogodno je uvesti smenu $\operatorname{tg} x = \frac{x}{y}$. Tada se najpre izračunava vrednost

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

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

Primer 3.7.1. Na osnovu Eulerovog razvoja za e^x naći ćemo prvih pet aproksimacija. Ovde je $b_1 = a_1 = 1$, $b_2 = -2x$, $a_2 = 2+x$, $b_3 = b_4 = b_5 = x^2$, $a_3 = 6$, $a_4 = 10$, $a_5 = 14$. Primenom, na primer, Eulerovog algoritma dobijamo

$$R_1 = R_1(x) = \frac{1}{1},$$

$$R_2 = R_2(x) = \frac{2+x}{2-x},$$

$$R_3 = R_3(x) = \frac{12+6x+x^2}{12-6x+x^2},$$

$$R_4 = R_4(x) = \frac{120+60x+12x^2+x^3}{120-60x+12x^2-x^3},$$

$$R_5 = R_5(x) = \frac{1680+840x+180x^2+20x^3+x^4}{1680-840x+180x^2-20x^3+x^4}.$$

Primetimo da dobijene aproksimacije $R_k(x)$ zadovoljavaju uslov $R_k(x)R_k(-x)=1$. Može se pokazati da racionalna funkcija $R(x)$ ispunjava ovaj uslov ako i samo ako se ona može predstaviti u obliku

$$R(x) = 1 - \frac{2x}{T(x^2) + x},$$

gde je $T(x^2)$ racionalna funkcija po x^2 . Za dokaz ovog tvrdjenja treba, najpre, dokazati da takva racionalna funkcija mora imati reprezentaciju u obliku $R(x) = P(x)/P(-x)$, gde je P algebraiski polinom. Tada je

$$T(x^2) = x \frac{P(-x) + P(x)}{P(-x) - P(x)}.$$

Na primer, za funkciju $R_4(x)$ imamo $P(x)=120+60x+12x^2+x^3$. Odgovarajuća funkcija $T(x^2)$ je data sa

$$T(x^2) = -12 \frac{x^2 + 10}{x^2 + 60}.$$

Dakle, imamo

$$R_4(x) = 1 - \frac{2x}{x - 12 \frac{x^2 + 10}{x^2 + 60}}.$$

Ako stavimo, na primer, $x=0.5$ imamo

$$e^{0.5} \approx R_4(0.5) = 1.6487214,$$

što predstavlja tačnu vrednost na šest decimala.

1.3.8. Asimptotski razvoji

Polinomski razvoji i razvoji u verižne razlomke obično se koriste za izračunavanje vrednosti (aproksimaciju) funkcija na konačnim intervalima realne ose. Izračunavanje vrednosti funkcija za velike vrednosti argumenta moguće je kod nekih funkcija svesti na prethodni slučaj i to transformacijom argumenta, ili pak korišćenjem izvesnih svojstava funkcije, koja dozvoljavaju korišćenje aproksimacija sa konačnog segmenta. Na primer, izračunavanje $r(x)$ za $x \in (0, +\infty)$ moguće je svesti, korišćenjem funkcionalne relacije $r(x+1) = xr(x)$, na izračunavanje vrednosti ove funkcije na konačnom segmentu $[1, 2]$, na primer pomoću polinomskog razvoja

$$\begin{aligned} r(z) \approx & 1 - 0.57710166t + 0.98585399t^2 - 0.87642182t^3 \\ & + 0.83282120t^4 - 0.56847290t^5 + 0.25482049t^6 - 0.05149930t^7, \end{aligned}$$

gde je $t = z - 1$ ($0 \leq t \leq 1$).

Medjutim, kada ovi pristupi nisu mogućni, bilo bi dobro imati izvesne razvoje funkcija, tzv. asimptotske razvoje, koji bi važili za velike vrednosti argumenta.

Pretpostavimo da su funkcije f i g definisane na $(0, +\infty)$ i da je $g(x) \neq 0$ za $x > 0$.

Definicija 3.8.1. Razvoj

$$(3.8.1) \quad g(x) \sim \sum_{k=0}^{+\infty} \frac{a_k}{x^k}$$

naziva se asimptotski razvoj za $f(x)$, ako za svako $n=0, 1, 2, \dots$ važi

$$(3.8.2) \quad \left(\frac{f(x)}{g(x)} - \sum_{k=0}^n \frac{a_k}{x^k} \right) x^n \rightarrow 0, \text{ kada } x \rightarrow \infty.$$

Ovu činjenicu označavamo sa

$$f(x) \sim g(x) \sim \sum_{k=0}^{+\infty} \frac{a_k}{x^k}.$$

Asimptotski razvoj (3.8.1) može biti, u opštem slučaju, divergentan red.

Primedba 3.8.1. Ako je

$$f_1(x) - f_2(x) \sim g(x) \sim \sum_{k=0}^{+\infty} a_k x^{-k},$$

tada ćemo

$$\phi(x) = f_2(x) + g(x) \sim \sum_{k=0}^{+\infty} a_k x^{-k}$$

tretirati kao asimptotski razvoj funkcije f_1 , tj. $f_1(x) \sim \phi(x)$. Na primer, za $\log \Gamma(x)$ imamo

$$\begin{aligned} \log \Gamma(x) &\sim \left(x - \frac{1}{2} \right) \log x - x + \frac{1}{2} \log 2\pi \\ &+ \frac{1}{12x} \left(1 - \frac{1}{30x^2} \right) + \frac{1}{105x^4} - \frac{1}{140x^6} + \frac{1}{99x^8} - \dots. \end{aligned}$$

Primer 3.8.1. Nadjimo asimptotski razvoj za funkciju

$$E_1(x) = \int_x^{+\infty} (e^{-t}/t) dt.$$

Sukcesivnom primenom parcijalne integracije nalazimo

$$\begin{aligned}
 E_1(x) &= \frac{e^{-x}}{x} - \int_x^{+\infty} \frac{e^{-t}}{t^2} dt \\
 &= \frac{e^{-x}}{x} - \frac{e^{-x}}{x^2} + 2 \int_x^{+\infty} \frac{e^{-t}}{t^3} dt \\
 &\vdots \\
 &= e^{-x} \left(\frac{1}{x} - \frac{1}{x^2} + \frac{2!}{x^3} - \frac{3!}{x^4} + \dots + (-1)^{n-1} \frac{(n-1)!}{x^n} \right) + R_n,
 \end{aligned}$$

gde je

$$R_n = (-1)^n n! \int_x^{+\infty} \frac{e^{-t}}{t^{n+1}} dt.$$

Kako je za $x > 0$

$$|R_n| = n! \int_x^{+\infty} \frac{e^{-t}}{t^{n+1}} dt = n! \frac{e^{-x}}{x^{n+1}} - (n+1)! \int_x^{+\infty} \frac{e^{-t}}{t^{n+2}} dt < n! \frac{e^{-x}}{x^{n+1}},$$

imamo

$$\left| (e^x E_1(x) - \sum_{k=1}^n (-1)^{k-1} \frac{(k-1)!}{x^k} x^n) \right| < \frac{n!}{x},$$

odakle zaključujemo da je uslov (3.8.2) ispunjen. Prema tome, dobili smo asimptotski razvoj

$$(3.8.3) \quad E_1(x) \sim e^{-x} \sum_{k=1}^{+\infty} (-1)^{k-1} \frac{(k-1)!}{x^k}.$$

Primetimo da je ovaj red divergentan za svako x .

Primer 3.8.2. Slično se može naći asimptotski razvoj i za komplementarnu funkciju greške

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt.$$

Tako, primenom parcijalne integracije nalazimo

$$2 \int_x^{\infty} e^{-t^2} dt = \int_x^{\infty} \left(-\frac{1}{t} \right) \left(-2te^{-t^2} \right) dt = \frac{1}{x} e^{-x^2} - \int_x^{\infty} \frac{1}{t^2} e^{-t^2} dt.$$

Nastavljajući ovaj proces dobijamo

$$\frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt = \frac{e^{-x^2}}{\sqrt{\pi}} \left(\frac{1}{x} - \frac{1}{2x^3} + \frac{1 \cdot 3}{2^2 x^5} - \dots + (-1)^n \frac{1 \cdot 3 \cdots (2n-1)}{2^n x^{2n+1}} \right) + R_n,$$

gde je

$$R_n = (-1)^{n+1} \frac{(2n+1)!!}{2^n \sqrt{\pi}} \int_x^{\infty} \frac{1}{t^{2n+2}} e^{-t^2} dt.$$

Primetimo da je

$$R_n = (-1)^{n+1} \frac{(2n+1)!!}{2^n \sqrt{\pi}} \left(\frac{e^{-x^2}}{2x^{2n+3}} - \frac{2n+3}{2} \int_x^{\infty} \frac{e^{-t^2}}{t^{2n+4}} dt \right),$$

tj.

$$R_n = (-1)^{n+1} \frac{(2n+1)!!}{2^{n+1} \sqrt{\pi}} \cdot \frac{e^{-x^2}}{x^{2n+3}} + R_{n+1},$$

i da su R_n i R_{n+1} suprotnog znaka ($R_n R_{n+1} < 0$). Tada se može zaključiti da je

$$|R_n| < \frac{(2n+1)!!}{2^{n+1} \sqrt{\pi}} \cdot \frac{e^{-x^2}}{x^{2n+3}},$$

što pokazuje da apsolutna greška odsecanja nije veća od modula prvog odbačenog člana. Takodje, korišćenjem ove nejednakosti zaključujemo da je $\lim_{x \rightarrow \infty} (\sqrt{\pi} e^{-x^2} R_n x^{2n+1}) = 0$, za svako $n=0, 1, \dots$, što znači da je uslov (3.8.2) ispunjen. Tako smo dobili asimptotni razvoj za komplementarnu funkciju greške

$$\operatorname{erfc}(x) \sim \frac{e^{-x^2}}{x \sqrt{\pi}} \left(1 + \sum_{k=1}^{+\infty} (-1)^k \frac{(2k-1)!!}{(2x^2)^k} \right).$$

I pored toga što su asimptotski razvoji divergentni, oni mogu biti korišćeni za izračunavanje vrednosti funkcija za velike vrednosti argumenta x . Naime, odsecanjem asimptotskog razvoja, tj. uzimanjem samo konačnog broja članova, dobijamo funkciju

$$\phi_n(x) = g(x) S_n(x),$$

gde je

$$(3.8.4) \quad S_n(x) = \sum_{k=0}^n \frac{a_k}{x^k},$$

koja sa proizvoljnom tačnošću može da aproksimira $f(x)$ za dovoljno velike vrednosti argumenta x . Dakle, kada x raste, tačnost u aproksimaciji

$$f(x) \approx \phi_n(x)$$

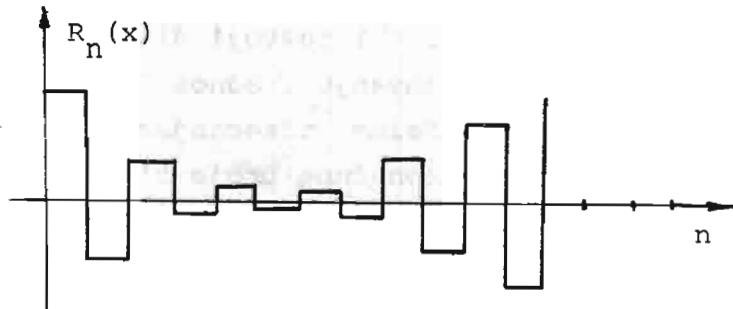
se povećava. Geometrijski ϕ_n predstavlja asimptotu od f sa dodirom najmanje reda n u beskonačnosti. Međutim, kada je x fiksno postoji jedna bitna razlika asymptotskih razvoja u odnosu na konvergentne, na primer, stepene redove. Kod poslednjih za datu vrednost argumenta x može se povećati tačnost uzimanjem većeg broja članova razvoja. Ova osobina kod divergentnih asymptotskih razvoja ne postoji, tj. uzimanje većeg broja članova (veće n u $S_n(x)$) ne dovodi uvek do povećanja tačnosti za dato $x (> 0)$.

U vezi sa (3.8.4) neka je

$$R_n(x) = S(x) - S_n(x) \quad (S(x) = f(x)/g(x)).$$

Pretpostavimo da je asymptotski razvoj alternativan tako da su $R_n(x)$ i $R_{n+1}(x)$ suprotnog znaka. S obzirom da je razvoj divergentan imaćemo situaciju kao na sl.3.8.1. Dakle, za fiksno $x > 0$, i svako $n=0,1,2,\dots$ imamo da je $S_n(x) + R_n(x) = S(x)$, pri čemu i $S_n(x)$ i $R_n(x)$ divergiraju, kada $n \rightarrow +\infty$. Međutim, kao i kod konvergentnih alternativnih redova, imaćemo da je

$$(3.8.5) \quad |S(x) - \frac{1}{2}(S_n(x) + S_{n+1}(x))| \leq \frac{1}{2}|u_{n+1}|,$$



Sl. 3.8.1

gde je $u_{n+1} = S_{n+1}(x) - S_n(x)$. U našem slučaju je $u_{n+1} = a_{n+1}/x^{n+1}$.

Na osnovu (3.8.5) vidimo da možemo uzeti

$$(3.8.6) \quad S(x) \approx \frac{1}{2}(S_n(x) + S_{n+1}(x)) = S_n(x) + \frac{1}{2} \cdot \frac{a_{n+1}}{x^{n+1}}$$

sa absolutnom greškom koja ne prelazi $\frac{1}{2}|a_{n+1}|/x^{n+1}$.

Primetimo da se u (3.8.6) na desnoj strani pojavljuje Cesàroova transformacija od $S_n(x)$. Jasno je da kod konvergentnih redova sa povećanjem n u (3.8.6) imamo bolju aproksimaciju (veću tačnost). Međutim, kod asymptotskih razvoja postoji optimalan broj n za dato x , koji se određuje iz uslova da granica absolutne greške u (3.8.6) bude najmanja. Dakle, n određujemo iz uslova

$$(3.8.7) \quad \min_k \frac{|a_{k+1}|}{x^{k+1}} = \frac{|a_{n+1}|}{x^{n+1}} .$$

Primer 3.8.3. Korišćenjem asymptotskog razvoja (3.8.3) odredićemo $E_1(5)$ i $E_1(9.5)$. U ovom slučaju imamo $|a_{k+1}| = k!$. Da bismo odredili optimalnu vrednost za broj članova koji treba uzeti u razvoju (3.8.3), posmatrajmo niz (q_k) , gde je

$$q_k = \frac{|a_k|}{x^k} = \frac{(k-1)!}{x^k} .$$

Kako je $q_{k+1}/q_k = k/x$ zaključujemo da je $q_{k+1} < q_k$ za $k < x$, dok je za $k > x$, $q_{k+1} > q_k$. Prema tome, minimum u (3.8.7) nastupa za $k=n=[x]$ (najveće celo od x). Ako je x ceo broj, tada je za $k=x$, $q_{k+1}=q_k$, što znači da minimum u (3.8.7) nastupa za dve vrednosti k , tj. za $k=n_1=x-1$ i $k=n_2=x$.

Za $x=5$ uzimamo $n=n_1=4$. Na osnovu (3.8.6) imamo

$$e^5 E_1(5) \approx S_4(5) + \frac{1}{2} \cdot \frac{4!}{5^5} = \frac{1}{5} - \frac{1}{5^2} + \frac{2}{5^3} - \frac{6}{5^4} + \frac{12}{5^5} = 0.17024$$

sa absolutnom greškom manjom od $12/5^5 = 3.84 \cdot 10^{-3}$. Primetimo da se isti rezultat dobija i za $n=n_2=5$

$$e^5 E_1(5) \approx S_5(5) - \frac{1}{2} \cdot \frac{5!}{5^6} = 0.17024 .$$

Sada imamo

$$E_1(5) \approx 0.001147,$$

pri čemu apsolutna greška nije veća od $2.6 \cdot 10^{-5}$. Tačna vrednost je $E_1(5)=0.00114829\dots$.

Za $x=9.5$ imamo $n=[9.5]=9$, pa je

$$e^{9.5} E_1(9.5) \approx S_9(9.5) - \frac{1}{2} \cdot \frac{9!}{9.5^{10}} \approx 0.0959866$$

sa apsolutnom greškom manjom od $3 \cdot 10^{-5}$. Najzad, imamo

$$E_1(9.5) \approx 7.184773 \cdot 10^{-6},$$

pri čemu je apsolutna greška manja od $2.3 \cdot 10^{-9}$. Tačna vrednost je, inače, $E_1(9.5)=7.1847746\dots \cdot 10^{-6}$.

1.4. LITERATURA

1. M.Abramowitz & I.A.Stegun: *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover New York, 1965.
2. F.L.Bauer: *Computational graphs and rounding error*. SIAM J. Numer. Anal. 11(1974), 87-96.
3. C.Brezinski: *Méthode d'accélération de la convergence en analyse numérique* (These). Grenoble, 1971.
4. C.Brezinski: *Accélération de suites à convergence logarithmique*. C.R. Acad. Sc. Paris. 273(1971), (série A), 727-730.
5. C.Brezinski: *Accélération de la convergence en analyse numérique*. Lect. Notes Math. 584, Springer, Berlin-Heidelberg-New York, 1977.
6. C.Brezinski: *A general extrapolation algorithm*. Numer. Math. 35 (1980), 175-187.
7. C.Brezinski: *Some new convergence acceleration methods*. Math. Comp. 39(1982), 133-145.
8. J.P.Delahaye: *Algorithmes pour suites non convergentes*. Numer. Math. 34(1980), 333-347.
9. J.P.Delahaye: *Optimalité du procédé Δ^2 d'Aitken pour l'accélération de la convergence linéaire*. RAIRO Anal. Numér. 15(1981), 321-330.
10. J.P.Delahaye: *Automatic selection of sequence transformations*. Math. Comp. 37(1981), 197-204.

11. J.Eve: *The evaluation of polynomials*, Numer. Math. 6(1964), 17-21.
12. C.T.Fike: Computer evaluation of mathematical functions. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1968.
13. W.Gautschi: Computational aspects of three-term recurrence relations. SIAM Review 9(1967), 24-82.
14. W.Gautschi: Minimal solutions of three-term recurrence relations and orthogonal polynomials. Math. Comp. 36(1981), 547-554.
15. W.Gautschi: On the convergence behavior of continued fractions with real elements. Math. Comp. 40(1983), 337-342.
16. A.O.Gel'fond: Isčislenie konečnyh raznostei. Nauka, Moskva, 1967.
17. W.B.Gragg: Truncation error bounds for g-fractions. Numer. Math. 11(1968), 370-379.
18. D.E.Knuth: Evaluation of polynomials by computer. Comm. ACM 5(1962), 592-599.
19. D.Levin & A.Sidi: Two new classes of nonlinear transformations for accelerating the convergence of infinite integrals and series. Appl. Math. Comput. 9(1981), 175-215.
20. H.Levy & F.Lessman: Finite difference equations. London, 1959.
21. D.D.McCracken & W.S.Dorn: Numerical methods and FORTRAN programming. Wiley, New York, 1964.
22. J.Mikloško: Investigation of algorithms for numerical computation of continued fractions. Ž. Vyčisl. Mat. i Mat. Fiz. 16(1976), 827-837.
23. J.Mikloško: A fast algorithm for repeated computation of linear recurrence relations. BIT 17(1977), 430-436.
24. J.Mikloško: An algorithm for calculating continued fractions. J. Comput. Appl. Math.
25. T.S.Motzkin: Evaluation of polynomials. Bull. Amer. Math. Soc. 61 (1955), 163.
26. K.S.Miller: An introduction to the calculus of finite differences and difference equations. New York, 1960.
27. W.Niethammer: Numerical application of Euler's series transformation and its generalizations. Numer. Math. 34(1980), 271-283.
28. N.E.Nörlund: Vorlesungen über Differenzrechnung. Springer, Berlin, 1924.
29. D.Shanks: Non-linear transformations of divergent and slowly convergent sequences. J. Math. Phys. 34(1955), 1-42.
30. V.Ja. Skorobogat'ko: Teoriya vettvjaščihsja cepnyh drobej i ee primenie v výčislitel'noi matematike. Nauka, Moskva, 1983.
31. D.V.Slavić: Transformation of the continued fraction into a rational function. Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz. N^o577 - N^o598(1977), 49-52.
32. D.A.Smith & W.F.Ford: Acceleration of linear and logarithmic convergence. SIAM J. Numer. Anal. 16(1979), 223-240.

33. D.A.Smith & W.F.Ford: *Numerical comparisons of nonlinear convergence accelerators.* Math. Comp. 38(1982), 481-499.
34. J.Stoer: *Einführung in die Numerische Mathematik I.* Springer -Verlag, Berlin-Heidelberg-New York, 1972.
35. W.J.Thron & H.Wadeland: *Truncation error bounds for limit periodic continued fractions.* Math. Comp. 40(1983), 589-597.
36. J.Todd: *Motivation for working in numerical analysis.* Comm. Pure Appl. Math. 8(1955), 97-116.
37. D.Tošić: *Uvod u numeričku analizu.* Naučna knjiga, Beograd, 1982.
38. J.H.Wilkinson: *Error analysis of floating point computation.* Numer. Math. 2(1960), 319-340.

Prostori. Operatori. Ortogonalni polinomi

2.1. LINEARNI PROSTORI

U ovom poglavlju dajemo definicije linearног, Banachovog i Hilbertovog prostora i uvodimo pojam ortogonalnosti.

2.1.1. Linearni prostor

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

1° U skupu X definisana jedna binarna operacija $+$ u odnosu na koju skup X čini Abel ovu 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:

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

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 $(u, c) \rightarrow cu$ množenje vektora skalarom (spoljašnja kompozicija).

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

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

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

$$c\theta = 0 \quad \text{za svako } c \in K.$$

Definicija 1.1.2. Vektori u_i ($i=1, \dots, n$) linearog 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

$$(1.1.2) \quad \sum_{i=1}^n c_i u_i = 0.$$

Vektori u_i ($i=1, \dots, n$) su linearno nezavisni, ako je jednakost (1.1.2) tačna samo za $c_i = 0$ ($i=1, \dots, n$).

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 vektorskem prostoru postoji n linearno nezavisnih vektora, a svaki sistem od $n+1$ vektora je linearne zavisan, kažemo da je prostor n-dimenzionalan. U protivnom slučaju kada u vektorskem prostoru postoji beskonačno mnogo linearne 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 linearne nezavisnih vektora prostora X obrazuje algebarsku ili Hamel ovu 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, međutim, 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 $B = \{u_1, \dots, u_m\}$ 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 određen 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 $R^n = \{(x_1, \dots, x_n) | x_i \in R (i=1, \dots, n)\}$. 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

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

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 ravno-pravno označke

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

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

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

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

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

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

$$\int_a^b |u(t)|^r dt < +\infty \text{ obrazuje vektorski prostor ako je unutrašnja}$$

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 sva-ko $u \in X$, koju nazivamo norma od u , takva da je

$$1^o \quad \|u\| = 0 \iff u = 0 \quad (\text{definisanost});$$

$$2^o \quad \|cu\| = |c| \cdot \|u\| \quad (\text{homogenost});$$

$$3^o \quad \|u+v\| \leq \|u\| + \|v\| \quad (\text{relacija trougla}),$$

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 $\rho: X^2 \rightarrow [0, +\infty)$ sa sledećim svojstvima

$$1^o \quad \rho(u,v) = 0 \iff u = v;$$

$$2^o \quad \rho(u,v) = \rho(v,u);$$

$$3^o \quad \rho(u,v) + \rho(v,w) \geq \rho(u,w).$$

Primer 1.2.1. Vektorski prostor \mathbb{R}^n se može normirati uvodjenjem

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

ili

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

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

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

Pošlednja norma poznata je kao Euklidova 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

$$(1.2.2) \quad \|u\| = \max_{a \leq t \leq b} |u(t)|$$

ili pomoću

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

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

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

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

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

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

Definicija 1.2.2. Neka je $\{u_n\}_{n \in \mathbb{N}}$ niz tačaka u normiranom prostoru X i neka je $u \in X$ takvo da je $\lim_{n \rightarrow +\infty} \|u_n - u\| = 0$. Tada kažemo

da ovaj niz konvergira po normi ka tački u .

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

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

$$\|u_n - u_m\| = \|(u_n - u) + (u - u_m)\| \leq \|u_n - u\| + \|u_m - u\| \rightarrow 0,$$

kada $m, n \rightarrow +\infty$. 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 unitarni prostor, ako postoji funkcija $(u,v): X^2 \rightarrow C$ koja zadovoljava sledeće uslove

$$1^\circ \quad (u,u) \geq 0;$$

$$2^\circ \quad (u,u) = 0 \Leftrightarrow u = 0;$$

$$3^\circ \quad (u+v, w) = (u, w) + (v, w);$$

$$4^\circ \quad (cu, v) = c(u, v);$$

$$5^\circ \quad (u, v) = \overline{(v, u)}$$

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

- (a) $(u, cv) = \bar{c}(u, v);$
- (b) $(u, v_1 + v_2) = (u, v_1) + (u, v_2);$
- (c) $| (u, v) |^2 \leq (u, u)(v, v).$

Dokaz. Tvrđenja (a) i (b) se jednostavno dokazuju. Da bismo dokazali tvrdjenje (c), koje je poznato kao Buniakowsky-Cauchy-Schwarzova nejednakost, uzmimo tačku $w=u+t(u, v)v$, gde je t realno i $u, v \in X$. Kako na osnovu 1° (iz definicije 1.3.1) važi

$$(w, w) = (u+t(u, v)v, u+t(u, v)v) \geq 0,$$

korišćenjem osobina 3° , 4° , 5° i (a) zaključujemo da je

$$(u, u) + 2| (u, v) |^2 t + | (u, v) |^2 (v, v) t^2 \geq 0,$$

odakle sleduje da diskriminanta D dobijenog kvadratnog trinoma mora biti manja ili jednaka nuli, tj.

$$\frac{D}{4} = | (u, v) |^4 - | (u, v) |^2 (u, u)(v, v) \leq 0.$$

Iz poslednje nejednakosti sleduje nejednakost (c).

Unitaran vektorski prostor može se normirati uvodjenjem

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

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-Hilbert-ov prostor. Ukoliko je ovaj prostor kompletan naziva se Hilbert-ov.

Primer 1.3.1. Vektorski prostor \mathbb{R}^n postaje Hilbert-ov 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 Hilbert-ov 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

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

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

Buniakowsky-Cauchy-Schwarzova nejednakost u C^n ima oblik

$$\left| \sum_{k=1}^n x_k \bar{y}_k \right|^2 \leq \left(\sum_{k=1}^n |x_k|^2 \right) \left(\sum_{k=1}^n |y_k|^2 \right).$$

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

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

dobija se Hilbert-ov prostor.

Primer 1.3.3. Sa skalarnim proizvodom

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

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

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

Uместо skalarnog proizvoda (1.3.2) može se uzeti opšitiji skalarni proizvod definisan sa

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

gde je $p: (a, b) \rightarrow \mathbb{R}^+$ data težinska funkcija (videti odeljak 2.2.1). Norma koja izvire iz skalarnog proizvoda uvedenog na ovaj način data je sa

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

U ovom slučaju nejednakost (c) postaje

$$\left| \int_a^b p(t) u(t) v(t) dt \right|^2 \leq \left(\int_a^b p(t) |u(t)|^2 dt \right) \left(\int_a^b p(t) |v(t)|^2 dt \right).$$

2.1.4. Ortogonalni sistemi u Hilbertovom prostoru

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

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

gde je δ_{nk} - Kronecker-ova 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 ortonormirani sistem i takav sistem označavaćemo sa $\{u_k^*\}_{k \in I}$.

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

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

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

i

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

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

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 $\{v_0, v_1, \dots\}$. Postupak kojim se ovom skupu vektora može pridružiti ortogonalan sistem vektora $\{u_0, u_1, \dots\}$, tako da se lineari nad ovim skupovima poklapaju, poznat je kao Gram-Schmidtov postupak ortogonalizacije i on se može iskazati na sledeći način.

Uzmimo najpre $u_0 = v_0$, a zatim u_1 predstavimo u obliku

$$u_1 = v_1 + \lambda_{10} u_0,$$

gde je λ_{10} nepoznati parametar koji određujemo iz uslova da je vektor u_1 ortogonalan na u_0 . Tada je

$$(u_1, u_0) = (v_1, u_0) + \lambda_{10} (u_0, u_0) = 0$$

sleduje

$$\lambda_{10} = - \frac{(v_1, u_0)}{(u_0, u_0)}.$$

Pretpostavimo sada da smo već konstruisali vektore u_0, u_1, \dots, u_{k-1} . Vektor u_k predstavimo u obliku

$$u_k = v_k + \lambda_{k0} u_0 + \lambda_{k1} u_1 + \dots + \lambda_{k,k-1} u_{k-1}.$$

Nepoznate parametre λ_{ki} ($i=0, 1, \dots, k-1$) odredjujemo iz uslova

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

Dakle, imamo

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

tj.

$$u_k = v_k - \sum_{i=0}^{k-1} \frac{(v_k, u_i)}{(u_i, u_i)} u_i \quad (k=1, 2, \dots).$$

Odgovarajući ortonormirani sistem vektora je $\{u_0^*, u_1^*, \dots\}$, gde je $u_k^* = \frac{u_k}{\|u_k\|}$ ($k=0, 1, \dots$).

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

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

gde je $x \mapsto p(x)$ nenegativna težinska funkcija, konstruisati ortogonalni bazis $\{Q_k\}_{k \in \mathbb{N}_0}$. Lineal nad ovim bazisom je skup svih algebarskih polinoma, koji je svuda gust* u X (videti [1, str.190]). 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 basiz 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 .

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

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

Primer 1.4.2. U prostoru $X = L^2(-1, 1)$ sa $p(x) = (1-x^2)^{3/2}$, odredićemo prvih pet članova ortogonalnog sistema $\{Q_k\}_{k \in N_0}$.

Izračunaćemo najpre integral

$$N_k = \int_{-1}^{+1} x^{2k} (1-x^2)^{3/2} dx \quad (k \in N_0).$$

Primenom parcijalne integracije na integral

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

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

$$\text{Kako je } N_0 = \frac{3\pi}{8}, \text{ imamo } N_k = 3\pi \frac{(2k-1)!!}{(2k+4)!!} \quad (k \in N).$$

Polazeći od prirodnog bazisa $\{1, x, x^2, \dots\}$, Gram-Schmidtovim postupkom ortogonalizacije nalazimo redom

$$Q_0(x) = 1,$$

$$Q_1(x) = x - \frac{(x, Q_0)}{(Q_0, Q_0)} Q_0(x) = x,$$

$$Q_2(x) = x^2 - N_1 N_0^{-1} = x^2 - \frac{1}{6},$$

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

$$\begin{aligned} Q_4(x) &= x^4 - N_2 N_0^{-1} - (N_3 - \frac{1}{6} N_2) (N_2 - \frac{1}{3} N_1 + \frac{1}{36} N_0)^{-1} (x^2 - \frac{1}{6}) \\ &= x^4 - \frac{3}{5} x^2 + \frac{3}{80}. \end{aligned}$$

Prva četiri člana odgovarajućeg ortonormiranog sistema $\{Q_k^*\}_{k \in N_0}$ su redom

$$Q_0^*(x) = \sqrt{\frac{8}{3\pi}}, \quad Q_1^*(x) = \frac{4x}{\sqrt{\pi}},$$

$$Q_2^*(x) = 8\sqrt{\frac{6}{5\pi}}(x^2 - \frac{1}{6}), \quad Q_3^*(x) = \frac{32}{\sqrt{3\pi}}(x^3 - \frac{3}{8}x),$$

Ovi polinomi su poznati kao Gegenbauerovi polinomi.

2.2. ORTOGONALNI POLINOMI

Zbog izuzetnog značaja ortogonalnih polinoma u numeričkoj matematici ovo poglavlje je posvećeno teoriji i konstrukciji ortogonalnih polinoma. Kompletnosti radi dat je pregled svih klasičnih ortogonalnih polinoma.

2.2.1. Moment-funkcionala i ortogonalnost

Definicija 2.1.1. Funkciju $x \mapsto p(x)$ definisanu na konačnom intervalu (a, b) nazivamo težinskom funkcijom ako je ona na tom intervalu nenegetivna, integrabilna i njen integral pozitivan, tj. ako su ispunjeni uslovi

$$(2.1.1) \quad p(x) \geq 0 \quad (x \in (a, b)) \quad \text{i} \quad 0 < \int_a^b p(x) dx < +\infty.$$

U slučaju kada je interval (a, b) beskonačan (na primer, $(-\infty, b)$, $(a, +\infty)$, $(-\infty, +\infty)$), pored uslova 2.1.1) potrebna je i apsolutna konvergencija integrala

$$(2.1.2) \quad C_k = \int_a^b x^k p(x) dx \quad (k=1, 2, \dots).$$

Integrale C_k nazivamo momentima težinske funkcije p .

U primeru 1.4.2 imali smo težinsku funkciju $p(x) = (1-x^2)^{3/2}$, dok su momenti težinske funkcije bili

$$C_0 = N_0 = \frac{3\pi}{8}, \quad C_{2k-1} = 0, \quad C_{2k} = N_k = 3\pi \frac{(2k-1)!!}{(2k+4)!!} \quad (k \in \mathbb{N}).$$

Kao i u odeljku 2.1.4, niz ortogonalnih polinoma na (a, b) u odnosu na težinsku funkciju $x \mapsto p(x)$, ili kako se još kaže u odnosu na skalarni proizvod

$$(2.1.3) \quad (f, g) = \int_a^b p(x) f(x) g(x) dx \quad (f, g \in L^2(a, b)),$$

označavaćemo sa $\{Q_k\}_{k \in N_0}$.

Nije teško primetiti da je i niz $\{c_k Q_k\}_{k \in N_0}$, gde su c_k ($\neq 0$) konstante, takodje ortogonalan u odnosu na isti skalarni proizvod (2.1.3).

Ortogonalni polinomi se mogu definisati znatno opštije. Naime, neka je \mathcal{P} linearни prostor svih algebarskih polinoma, $\{c_k\}_{k \in N_0}$ niz kompleksnih brojeva i $\mathcal{L}: \mathcal{P} \rightarrow \mathbb{R}$ funkcionala definisana pomoću

$$\mathcal{L}[x^k] = c_k \quad (k = 0, 1, 2, \dots),$$

$$\mathcal{L}[\alpha P(x) + \beta Q(x)] = \alpha \mathcal{L}[P(x)] + \beta \mathcal{L}[Q(x)] \quad (\forall \alpha, \beta \in \mathbb{C}; \forall P, Q \in \mathcal{P}).$$

Tada se kaže da je \mathcal{L} moment-funkcionala odredjena momentnim nizom $\{c_k\}_{k \in N_0}$, a za c_k se kaže da je moment reda k . Primetimo da je funkcionala \mathcal{L} linearna (videti odeljak 2.3.1).

Definicija 2.1.2. Niz polinoma $\{Q_k\}_{k \in N_0}$ je ortogonalan u odnosu na moment-funkcionalu \mathcal{L} ako je

- a) $dg Q_k = k$;
- b) $\mathcal{L}[Q_k(x) Q_n(x)] = 0 \quad (k \neq n)$;
- c) $\mathcal{L}[Q_k(x)^2] \neq 0$

za svako $k, n \in N_0$.

Uslovi b) i c) u prethodnoj definiciji mogu biti zamenjeni uslovom

$$\mathcal{L}[Q_k(x) Q_n(x)] = K_n \delta_{kn}, \quad K_n \neq 0,$$

gde je δ_{kn} Kroneckerova delta.

Primetimo da su ortogonalni polinomi u odnosu na moment-funkcionalu definisani sa

$$(2.1.4) \quad \mathcal{L}[f] = (f, l) = \int_a^b p(x) f(x) dx \quad (f \in \mathcal{P}),$$

upravo ortogonalni polinomi u odnosu na skalarni proizvod (2.1.3).

Tada je $K_n = \|Q_n\|^2$.

Teorema 2.1.1. Neka je \mathcal{L} moment-funkcionela i $\{Q_k\}_{k \in N_0}$ niz polinoma. Tada su sledeća tri tvrdjenja ekvivalentna:

- (a) Niz $\{Q_k\}_{k \in N_0}$ je ortogonalan u odnosu na \mathcal{L} ;
- (b) $\mathcal{L}[P(x)Q_n(x)] = 0$ za svaki polinom P stepena $m < n$ i $\mathcal{L}[P(x)Q_n(x)] \neq 0$ ako je $m = n$;
- (c) $\mathcal{L}[x^m Q_n(x)] = K_n \delta_{mn}$, gde je $K_n \neq 0$ ($m = 0, 1, \dots, n$).

Dokaz. Prepostavimo da važi (a). Kako je dg $Q_k = k$, to je $\{Q_0, Q_1, \dots, Q_m\}$ baza potprostora \mathcal{P}_m svih polinoma ne višeg stepena od m , što znači da se polinom P može da predstavi kao linearna kombinacija bazisnih polinoma

$$P(x) = \sum_{k=0}^m \alpha_k Q_k(x) \quad (\alpha_m \neq 0).$$

Tada, s obzirom na linearnost moment-funkcionele \mathcal{L} , imamo

$$\begin{aligned} \mathcal{L}[P(x)Q_n(x)] &= \sum_{k=0}^m \alpha_k \mathcal{L}[Q_k(x)Q_n(x)] = 0 \quad (m < n) \\ &= \alpha_n \mathcal{L}[Q_n(x)^2] \quad (m = n), \end{aligned}$$

što znači da važi implikacija (a) \Rightarrow (b). Za kompletiranje dokaza primetimo da (b) \Rightarrow (c) \Rightarrow (a).

U monografiji [4, str. 9] može se naći i sledeći rezultat:

Teorema 2.1.2. Neka je $\{Q_k\}_{k \in N_0}$ niz ortogonalnih polinoma u odnosu na \mathcal{L} . Tada se svaki polinom P , stepena n , može predstaviti u obliku

$$P(x) = \sum_{k=0}^n \alpha_k Q_k(x),$$

gde su koeficijenti α_k dati pomoću

$$\alpha_k = \frac{\mathcal{L}[P(x)Q_k(x)]}{\mathcal{L}[Q_k(x)^2]} \quad (k = 0, 1, \dots, n).$$

Definicija 2.1.3. Za moment-funkcionelu \mathcal{L} kažemo da je pozitivno-definitna ako je $\mathcal{L}[P(x)] > 0$ za svaki polinom P koji je nenega-

tivan za svako realno x i nije identički jednak nuli.

U vezi sa egzistencijom niza ortogonalnih polinoma uvedimo tzv. moment-determinantu

$$(2.1.5) \quad \Delta_k = \begin{vmatrix} c_0 & c_1 & \cdots & c_{k-1} \\ c_1 & c_2 & & c_k \\ \vdots & & & \\ c_{k-1} & c_k & & c_{2k-2} \end{vmatrix} \quad (k=1, 2, \dots).$$

Teorema 2.1.3. Potrebni i dovoljni uslovi za egzistenciju niza ortogonalnih polinoma u odnosu na moment-funkcionalu \mathcal{L} su

$$\Delta_k \neq 0 \quad (k = 1, 2, \dots).$$

Dokaz. Neka je $Q_n(x) = q_{n0} + q_{n1}x + \dots + q_{nn}x^n$. Na osnovu tvrdjenja (c) u teoremi 2.1.1, uslovi ortogonalnosti se mogu izraziti u obliku

$$(2.1.6) \quad \mathcal{L}[x^m Q_n(x)] = \sum_{k=0}^n q_{nk} c_{k+m} = K_n \delta_{mn}, \quad K_n \neq 0 \quad (m=0, 1, \dots, n),$$

što je ekvivalentno sistemu linearnih jednačina

$$(2.1.7) \quad \begin{bmatrix} c_0 & c_1 & \cdots & c_n \\ c_1 & c_2 & & c_{n+1} \\ \vdots & & & \\ c_n & c_{n+1} & & c_{2n} \end{bmatrix} \cdot \begin{bmatrix} q_{n0} \\ q_{n1} \\ \vdots \\ q_{nn} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ K_n \end{bmatrix}.$$

Ako pretpostavimo da niz ortogonalnih polinoma u odnosu na \mathcal{L} egzistira, tada je on jedinstveno određen konstantama K_n u (2.1.6), što znači da sistem jednačina (2.1.7) ima jedinstveno rešenje, tj. da je $\Delta_{n+1} \neq 0$ ($n \geq 0$).

Obrnuto, ako je $\Delta_k \neq 0$ ($k \geq 1$) tada za proizvoljno $K_n \neq 0$ sistem jednačina (2.1.7) ima jedinstveno rešenje, što znači da polinom Q_n , koji zadovoljava (2.1.6) egzistira. Kako je

$$q_{nn} = K_n \frac{\Delta_n}{\Delta_{n+1}} \neq 0 \quad (n \geq 1) \quad \text{i} \quad q_{00} \neq 0,$$

zaključujemo da je dg $Q_k = k$, što zajedno sa prethodnim dokazuje da je $\{Q_k\}_{k \in \mathbb{N}_0}$ niz ortogonalnih polinoma u odnosu na \mathcal{L} .

Na osnovu prethodnog rezultata, jednostavno se može izvesti sledeća eksplicitna formula

$$Q_n(x) = A_n \begin{vmatrix} c_0 & c_1 & \dots & c_n \\ c_1 & c_2 & & c_{n+1} \\ \vdots & & & \\ c_{n-1} & c_n & & c_{2n-1} \\ 1 & x & & x^n \end{vmatrix},$$

gde je $A_n = \frac{\kappa_n}{\Delta_{n+1}}$ konstanta različita od nule.

U vezi sa pozitivno-definitnim moment-funkcionelama važe sledeći rezultati (viđeti [4, str.14-16]):

Teorema 2.1.4. Ako je \mathcal{L} pozitivno-definitna funkcionala, tada ona ima sve realne momente, i njen odgovarajući niz ortogonalnih polinoma egzistira, pri čemu su svi polinomi realni.

Teorema 2.1.5. Funkcionala \mathcal{L} je pozitivno-definitna ako i samo ako su svi njeni momenti realni i $\Delta_k > 0$ ($k = 1, 2, \dots$).

Primetimo da je funkcionala \mathcal{L} definisana pomoću (2.1.4) pozitivno-definitna, što na osnovu prethodnog znači da odgovarajući niz ortogonalnih polinoma egzistira i realan je. Kao što smo videli u odeljku 2.1.4, ovi se polinomi mogu jednostavno odrediti Gram-Schmidtovim postupkom ortogonalizacije. Mi ćemo se u daljem izlaganju uglavnom baviti ovakvom klasom polinoma. Samo u poslednjem odeljku ovog poglavlja proučićemo ukratko jednu klasu kompleksnih ortogonalnih polinoma u odnosu na jednu kompleksnu moment-funkcionalu, koja nije pozitivno-definitna. Ona je zapravo kvazi-definitna, što znači da je $\Delta_k \neq 0$ za $k \geq 1$. Ovu klasu polinoma ortogonalnih na jediničnom polukrugu u kompleksnoj ravni uveli su W. Gautschi i G.V. Milovanović ([12], [13]).

2.2.2. Opšte osobine ortogonalnih polinoma

Neka je $\{Q_k\}_{k \in N_0}$ niz ortogonalnih polinoma u odnosu na moment-funkcionalu (2.1.4), tj. ortogonalan sa težinskom funkcijom p na (a, b) .

Teorema 2.2.1. Sve nule polinoma Q_k ($k=1, 2, \dots$) su realne, različite i leže u intervalu (a, b) .

Dokaz. Kako je, zbog ortogonalnosti,

$$\int_a^b p(x) Q_k(x) dx = \frac{1}{Q_0(x)} (Q_k, Q_0) = 0 \quad (k=1, 2, \dots),$$

zaključujemo da polinom Q_k menja znak bar u jednoj tački iz intervala (a, b) . Pretpostavimo da su x_1, \dots, x_m ($m \leq k$) 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 $\omega(x) = (x - x_1) \dots (x - x_m)$.

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

$$\omega(x) = \sum_{i=0}^m \alpha_i Q_i(x),$$

imamo

$$\int_a^b p(x) P(x) dx = (Q_k, \omega) = \sum_{i=0}^m \alpha_i (Q_k, Q_i),$$

tj.

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

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.

Teorema 2.2.2. Izmedju tri uzastopna polinoma u nizu $\{Q_k\}_{k \in \mathbb{N}_0}$ postoji rekurentna relacija

$$(2.2.1) \quad Q_{k+1}(x) - (\alpha_k x + \beta_k) Q_k(x) + \gamma_k Q_{k-1}(x) = 0,$$

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 2.2.1, ovaj polinom možemo predstaviti kao linearu kombinaciju polinoma Q_0, Q_1, \dots, Q_k , tj. u obliku

$$(2.2.2) \quad Q_{k+1}(x) - \alpha_k x Q_k(x) = \beta_k Q_k(x) - \sum_{j=1}^k \gamma_j Q_{j-1}(x),$$

odakle je

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

Ako je $0 \leq i \leq k-2$, korišćenjem osobine ortogonalnosti niza $\{Q_k\}_{k \in N}$, zaključujemo da je $\gamma_j = 0$ ($j=1, \dots, k-1$). Dakle, jednakost (2.2.2) se svodi na (2.2.1).

Ovim je dokaz završen.

Primedba 2.2.1. Rekurentna relacija (2.2.1) je tačna i za $k=0$, ako stavimo $Q_{-1}(x)=0$.

Primedba 2.2.2. Moguće je dokazati i obrnuto tvrdjenje nego što daje teorema 2.2.3. Naime, ako niz polinoma zadovoljava tročlanu rekurentnu relaciju oblika (2.2.1), tada postoji interval (a, b) i težinska funkcija p u odnosu na koju je ovaj niz polinoma ortogonalan na (a, b) .

Pri fiksiranoj vrednosti za x , rekurentna relacija (2.2.1) predstavlja linearu diferencnu jednačinu drugog reda

$$(2.2.3) \quad y_{k+1} - (\alpha_k x + \beta_k) y_k + \gamma_k y_{k-1} = 0,$$

čije je jedno rešenje $Q_k(x)$. Drugo linearno nezavisno rešenje ove jednačine, pri $x \notin [a, b]$, može se jednostavno naći množenjem jednakosti (2.2.1), tj. jednakosti

$$Q_{k+1}(t) - (\alpha_k t + \beta_k) Q_k(t) + \gamma_k Q_{k-1}(t) = 0,$$

sa $p(t)/(t-x)$ i integracijom po promenljivoj t u granicama od a do b , uz korišćenje identiteta $\frac{t}{x-t} = \frac{x}{x-t} - 1$. Tako, za $k \in N$, dobijamo

$$w_{k+1}(x) - (\alpha_k x + \beta_k) w_k(x) + \gamma_k w_{k-1}(x) = 0,$$

gde smo stavili

$$w_k(x) = \int_a^b \frac{p(t)}{x-t} Q_k(t) dt.$$

Funkcija $x \mapsto w_k(x)$ se naziva funkcija druge vrste u odnosu na polinom Q_k .

Pokazaćemo sada kako se koeficijenti $\alpha_k, \beta_k, \gamma_k$, koji se pojavljuju u rekurentnoj relaciji (2.2.1), mogu izraziti pomoću koeficijenata polinoma iz ortogonalnog niza $\{Q_k\}_{k \in N_0}$.

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

$$(2.2.4) \quad Q_k(x) = a_k x^k + b_k x^{k-1} + \dots .$$

Tada, na osnovu (2.2.1) i (2.2.4), imamo

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

odakle sleduje

$$(2.2.5) \quad \alpha_k = \frac{a_{k+1}}{a_k} \quad \text{i} \quad \beta_k = \alpha_k \left(\frac{b_{k+1}}{a_{k+1}} - \frac{b_k}{a_k} \right) \quad (k \in N).$$

Da bismo odredili γ_k podjimo od rekurentne relacije

$$(2.2.6) \quad Q_k(x) - (\alpha_{k-1}x + \beta_{k-1})Q_{k-1}(x) + \gamma_{k-1}Q_{k-2}(x) = 0.$$

Množenjem ove jednakosti sa $p(x)Q_k(x)$ i integracijom od a do b , uz korišćenje ortogonalnosti niza $\{Q_k\}_{k \in N_0}$, dobijamo

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

Slično, množenjem jednakosti (2.2.1) sa $p(x)Q_{k-1}(x)$ i odgovarajućom integracijom, dobijamo

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

Najzad, iz poslednje dve jednakosti sleduje

$$(2.2.7) \quad \gamma_k = \frac{\alpha_k}{\alpha_{k-1}} \left(\frac{\|Q_k\|}{\|Q_{k-1}\|} \right)^2 = \frac{a_{k+1} a_{k-1}}{a_k^2} \left(\frac{\|Q_k\|}{\|Q_{k-1}\|} \right)^2 \quad (k \in N).$$

Teorema 2.2.3. Za ortonormiran niz polinoma $\{Q_k^*\}_{k \in N_0}$ važi rekurentna relacija

$$x Q_{k-1}^*(x) = \lambda_k Q_k^*(x) + \mu_k Q_{k-1}^*(x) + \gamma_{k-1} Q_{k-2}^*(x),$$

gde su λ_k i μ_k konstante, $Q_{-1}^*(x)=0$ i $Q_0^*(x)=\left(\int_a^b p(x) dx\right)^{-1/2}$.

Dokaz. Ako u rekurentnoj relaciji (2.2.6) stavimo $Q_i(x) = \|Q_i\| Q_i^*(x)$ ($i=k-2, k-1, k$) i tako dobijenu jednakost podelimo sa $\alpha_{k-1} \|Q_{k-1}\|$, dobijamo

$$\begin{aligned} x Q_{k-1}^*(x) &= \frac{\|Q_k\|}{\alpha_{k-1} \|Q_{k-1}\|} Q_k^*(x) - \frac{\beta_{k-1}}{\alpha_{k-1}} Q_{k-1}^*(x) + \\ &+ \frac{\gamma_{k-1} \|Q_{k-2}\|}{\alpha_{k-1} \|Q_{k-1}\|} Q_{k-2}^*(x). \end{aligned}$$

Na dalje, stavimo

$$\lambda_k = \frac{\|Q_k\|}{\alpha_{k-1} \|Q_{k-1}\|} \quad \text{i} \quad \mu_k = - \frac{\beta_{k-1}}{\alpha_{k-1}}.$$

Kako je, s obzirom na (2.2.7),

$$\frac{\gamma_{k-1} \|Q_{k-2}\|}{\alpha_{k-1} \|Q_{k-1}\|} = \frac{\|Q_{k-1}\|}{\alpha_{k-2} \|Q_{k-2}\|} = \lambda_{k-1}$$

dokaz teoreme je završen.

Teorema 2.2.4. Za moničan* niz ortogonalnih polinoma $\{Q_k\}_{k \in N_0}$ važi rekurentna relacija

$$(2.2.8) \quad Q_{k+1}(x) = (x - \beta_k) Q_k(x) - \gamma_k Q_{k-1}(x),$$

gde su β_k i γ_k realne konstante i $\gamma_k > 0$.

Dokaz. Rekurentna relacija (2.2.8) direktno sleduje iz (2.2.1). Naime, kako je kod moničnih polinoma $a_k=1$, na osnovu (2.2.5) i (2.2.7) ovde imamo

* Polinom se naziva moničan ako je njegov koeficijent uz najviši stepen jednak jedinici.

$$\alpha_k = 1, \quad \beta_k = b_k - b_{k+1}, \quad \gamma_k = \left(\frac{\|Q_k\|}{\|Q_{k-1}\|} \right)^2 > 0.$$

U konstruktivnoj teoriji ortogonalnih polinoma tročlana rekurentna relacija ima fundamentalnu ulogu iz sledećih razloga ([10]):

1° Jednostavno se generiše niz ortogonalnih polinoma. Na primer, uzimajući $Q_{-1}(x)=0$ i $Q_0(x)=1$, na osnovu (2.2.8), jednostavno dobijamo $Q_1(x), Q_2(x), \dots$. Pri ovome je dovoljno poznavati dva niza $\{\beta_k\}$ i $\{\gamma_k\}$. Konstrukcija ortogonalnih polinoma pomoću koeficijenata polinoma, ili pomoću njegovih nula, zahteva poznavanje dvodimenzionalnog polja podataka (tačnije, informacija o ovim polinomima može se dati jednom trougaonom matricom).

2° Znajući β_k i γ_k jednostavno se može izračunati vrednost polinoma $Q_m(x)$, kao i njegovih izvoda, diferenciranjem jednakosti (2.2.8).

3° Konačne sume oblika $\sum_{k=0}^m c_k Q_k(x)$, koje se često javljaju u teoriji aproksimaciji, izračunavaju se, takodje, jednostavno.

4° Funkcije druge vrste

$$w_k(x) = \int_a^b \frac{p(t)}{x-t} Q_k(t) dt \quad (x \notin [a, b])$$

zadovoljavaju istu tročlanu rekurentnu relaciju.

5° Koeficijenti γ_k u (2.2.8) određuju normu polinoma Q_n

$$\|Q_n\|^2 = \int_a^b p(x) Q_n(x)^2 dx = \gamma_0 \gamma_1 \dots \gamma_n,$$

pri čemu smo uzeli $\gamma_0 = \|Q_0\|^2 = c_0 = \int_a^b p(x) dx$.

6° Od koeficijenata β_k i γ_k se može konstruisati simetrična trodijagonalna matrica, tzv. Jacobieva matrica,

$$J = \begin{bmatrix} \beta_0 & \sqrt{\gamma_1} & & \\ \sqrt{\gamma_1} & \beta_1 & \sqrt{\gamma_2} & \\ & \sqrt{\gamma_2} & \beta_2 & \sqrt{\gamma_3} \\ & & \ddots & \ddots \\ & & \ddots & \ddots \end{bmatrix}$$

koja igra značajnu ulogu u konstrukciji Gaussova kvadraturnih formula, o čemu će biti reči u poglavljima koje se odnosi na numeričku integraciju funkcija.

Primedba 2.2.3. Moguće je konstruisati ortogonalne polinome i odnosu na opštiji skalarni proizvod, nego što je (2.1.3). Razmatranja koja se odnose na tročlanu rekurentnu relaciju ostaju u važnosti ako uvedeni skalarni proizvod ispunjava uslov $(xf, g) = (f, xg)$. U protivnom, monični ortogonalni polinomi zadovoljavaju rekurentnu relaciju ([7])

$$Q_{k+1}(x) = (x - \beta_{k,k}) Q_k(x) - \beta_{k,k-1} Q_{k-1}(x) - \dots - \beta_{k,0} Q_0(x),$$

gde je $\beta_{k,i} = (x Q_k, Q_i) / (Q_i, Q_i)$.

Sledeći rezultat poznat je kao Christoffel-Darbouxov identitet.

Teorema 2.2.5. Za ortogonalan niz polinoma $\{Q_k\}_{k \in N_0}$ važi identitet

$$(2.2.9) \quad \sum_{k=0}^n \frac{Q_k(x) Q_k(t)}{\|Q_k\|^2} = \frac{1}{\alpha_n \|Q_n\|^2} \cdot \frac{Q_{n+1}(x) Q_n(t) - Q_n(x) Q_{n+1}(t)}{x - t},$$

gde je α_n konstanta koja se javlja u rekurentnoj relaciji (2.2.1)

Dokaz. Množenjem relacije (2.2.1) sa $Q_k(t)$, dobijamo

$$(2.2.10) \quad Q_{k+1}(x) Q_k(t) = (\alpha_k x + \beta_k) Q_k(x) Q_k(t) - \gamma_k Q_{k-1}(x) Q_k(t).$$

Ako u poslednjoj jednakosti zamenimo x sa t i t sa x i tako dobijenu jednakost oduzmemo od (2.2.10), 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.

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

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

i iskoristimo (2.2.6), 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 (2.2.9).

Primedba 2.2.4. Za moničan niz ortogonalnih polinoma u (2.2.9) treba uzeti $\alpha_n = 1$.

Na osnovu teorema 2.2.4 i 2.2.6 zaključujemo da važi:

Teorema 2.2.6. Za ortonormiran niz polinoma $\{Q_k^*\}_{k \in N_0}$ važi identitet

$$\sum_{k=0}^n Q_k^*(x)Q_k^*(t) = \lambda_{n+1} \frac{Q_{n+1}^*(x)Q_n^*(t) - Q_n^*(x)Q_{n+1}^*(t)}{x - t},$$

gde je λ_{n+1} konstanta koja se javlja u odgovarajućoj rekurentnoj relaciji za ortonormirane polinome.

Prelaskom na graničnu vrednost, kada $t \rightarrow x$, iz (2.2.9) sleduje sledeći rezultat:

Teorema 2.2.7. Za ortogonalan niz polinoma $\{Q_n\}_{n \in N_0}$ važi identitet

$$(2.2.11) \quad \sum_{k=0}^n \frac{Q_k(x)^2}{\|Q_k\|^2} = \frac{1}{\alpha_n \|Q_n\|^2} (Q_{n+1}'(x)Q_n(x) - Q_n'(x)Q_{n+1}(x)).$$

Označimo sada sa $x_i^{(k)}$ ($i=1, \dots, k$) nule polinoma Q_k uređjene po veličini, tj.

$$x_1^{(k)} < x_2^{(k)} < \dots < x_k^{(k)}.$$

Korišćenjem identiteta (2.2.11) može se pokazati sledeći rezultat:

Teorema 2.2.8. Za nule polinoma Q_n i Q_{n+1} važi

$$x_1^{(n+1)} < x_1^{(n)} < x_2^{(n+1)} < \dots < x_n^{(n+1)} < x_n^{(n)} < x_{n+1}^{(n+1)}.$$

Dokaz. Stavimo $x=x_i^{(n+1)}$ u (2.2.11). Tada zaključujemo da je znak proizvoda

$$Q'_{n+1}(x_i^{(n+1)}) Q_n(x_i^{(n+1)})$$

u svim nulama polinoma Q_{n+1} isti, tj. znak ne zavisi od i . S druge strane, kako na osnovu Rolleove teoreme, Q'_{n+1} menja znak u intervalu $(x_i^{(n+1)}, x_{i+1}^{(n+1)})$, zaključujemo da i Q_n , takodje, menja znak u ovom intervalu. Dakle, Q_n ima najmanje jednu nulu u intervalu $(x_i^{(n+1)}, x_{i+1}^{(n+1)})$, a s obzirom da ovakvih intervala ima n , zaključujemo da je u svakom od ovih intervala jedna i samo jedna nula polinoma Q_n , čime je završen dokaz teoreme.

Na kraju navedimo bez dokaza još jedan rezultat koji se naziva Christoffelova formula.

Teorema 2.2.9. Neka je $x \mapsto \omega(x)$ polinom m -tog stepena takav da ima proste nule x_1, x_2, \dots, x_m van segmenta $[a, b]$ i da je $\omega(x) > 0$ za svako $x \in [a, b]$. Ako je $\{Q_k\}_{k \in \mathbb{N}_0}$ niz ortogonalnih polinoma na (a, b) sa težinskom funkcijom $x \mapsto p(x)$, tada je niz polinoma $\{R_k\}_{k \in \mathbb{N}_0}$ definisan pomoću

$$(2.2.12) \quad R_k(x) = \frac{A_k}{\omega(x)} \begin{vmatrix} Q_k(x) & Q_{k+1}(x) & \dots & Q_{k+m}(x) \\ Q_k(x_1) & Q_{k+1}(x_1) & & Q_{k+m}(x_1) \\ \vdots & & & \\ Q_k(x_m) & Q_{k+1}(x_m) & & Q_{k+m}(x_m) \end{vmatrix},$$

gde je A_k proizvoljna konstanta različita od nule, ortogonalan na (a,b) sa težinskom funkcijom $x \mapsto \omega(x)p(x)$.

Primer 2.2.1. Neka je $\omega(x) = \begin{cases} x-c & (c < a) \\ c-x & (c > b) \end{cases}$ i $\{Q_k\}_{k \in N_0}$ niz ortogonalnih polinoma na (a,b) sa težinom $x \mapsto p(x)$. Tada je $\{R_k\}_{k \in N_0}$, gde je

$$R_k(x) = \frac{A_k}{x-c}(Q_{k+1}(c)Q_k(x) - Q_k(c)Q_{k+1}(x)),$$

ortogonalan niz na (a,b) sa težinskom funkcijom $x \mapsto \omega(x)p(x)$.

O primeni ovog rezultata u teoriji filtara videti rad Rakovića i Vasića [25].

2.2.3. Verižni razlomak pridružen rekurentnoj relaciji

Neka je $\{Q_k\}_{k \in N_0}$ moničan niz ortogonalnih polinoma na (a,b) u odnosu na težinsku funkciju $x \mapsto p(x)$, koji zadovoljava rekurentnu relaciju (2.2.8), tj.

$$(2.3.1) \quad Q_{k+1}(x) = (x-\beta_k)Q_k(x) - \gamma_k Q_{k-1}(x) \quad (k=0,1,\dots),$$

$$Q_{-1}(x) = 0, \quad Q_0(x) = 1,$$

gde su β_k i γ_k realne konstante i $\gamma_k > 0$. Kao i ranije, usvojimo

$$\gamma_0 = C_0 = \int_a^b p(x) dx.$$

Polinomu Q_k pridružimo polinom R_k pomoću

$$(2.3.2) \quad R_k(x) = \int_a^b p(t) \frac{Q_k(x) - Q_k(t)}{x-t} dt \quad (k=0,1,\dots),$$

čiji je stepen $k-1$. Primetimo da je

$$\begin{aligned} Q_{k+1}(x) - Q_{k+1}(t) &= \\ &= (x-\beta_k)Q_k(x) - \gamma_k Q_{k-1}(x) - (t-\beta_k)Q_k(t) + \gamma_k Q_{k-1}(t) \\ &= (x-t)Q_k(t) + (x-\beta_k)(Q_k(x) - Q_k(t)) - \gamma_k(Q_{k-1}(x) - Q_{k-1}(t)). \end{aligned}$$

Množenjem dobijene jednakosti sa $p(t)/(x-t)$ i integracijom imamo

$$R_{k+1}(x) = \int_a^b p(t)Q_k(t)dt + (x-\beta_k)R_k(x) - \gamma_k R_{k-1}(x),$$

pri čemu je integral koji se javlja na desnoj strani ove jednakosti jednak C_0 , ako je $k=0$, i jednak nuli za svako $k \in \mathbb{N}$. Prema tome, možemo zaključiti da pridruženi polinom R_k zadovoljava, takodje, rekurentnu relaciju (2.3.1) samo sa drugačijim početnim vrednostima. Naime, imamo

$$(2.3.3) \quad R_{k+1}(x) = (x-\beta_k)R_k(x) - \gamma_k R_{k-1}(x) \quad (k=1, 2, \dots),$$

$$R_0(x) = 0, \quad R_1(x) = C_0.$$

Rekurentna relacija (2.3.3) važi i za $k=0$, ako stavimo $R_{-1}(x)=-1$. Zaista, tada je $R_1(x) = (x-\beta_0) \cdot 0 - \gamma_0(-1) = \gamma_0 = C_0$.

Neka je, dalje,

$$(2.3.4) \quad F(x) = \int_a^b \frac{p(t)}{x-t} dt \quad (x \notin [a, b]).$$

Rekurentnoj relaciji (2.3.1) može se pridružiti verižni razlomak, koji takodje odgovara funkciji F (videti [9]),

$$F(x) \sim \frac{\gamma_0}{x-\beta_0 - \frac{\gamma_1}{x-\beta_1 - \frac{\gamma_2}{x-\beta_2 - \dots}}}.$$

Saglasno uvedenoj notaciji u odeljku 1.3.5 imamo

$$(2.3.5) \quad F(x) \sim \left[\frac{-\gamma_0}{x-\beta_0}, \frac{-\gamma_1}{x-\beta_1}, \dots \right].$$

Ako verižni razlomak (2.3.5) "prekinemo", tj. zaustavimo se na članu $x-\beta_{k-1}$, dobijamo da je

$$(2.3.6) \quad \left[\frac{-\gamma_0}{x-\beta_0}, \frac{-\gamma_1}{x-\beta_1}, \dots, \frac{-\gamma_{k-1}}{x-\beta_{k-1}} \right] = \frac{R_k(x)}{Q_k(x)},$$

gde su polinomi R_k i Q_k definisani pomoću (2.3.1) i (2.3.3) respektivno.

Primetimo da racionalna funkcija $x \mapsto \frac{R_k(x)}{Q_k(x)}$ ima samo proste polove u tačkama $x = x_i^{(k)}$ ($i=1, \dots, k$), koje su, u stvari, nule polinoma Q_k . Ako sa A_i ($i=1, \dots, n$) označimo ostatke u ovim polovima, imamo

$$A_i = \operatorname{Res}_{x=x_i^{(k)}} \frac{R_k(x)}{Q_k(x)} = \lim_{x \rightarrow x_i^{(k)}} (x - x_i^{(k)}) \frac{R_k(x)}{Q_k(x)} = \frac{R_k(x_i^{(k)})}{Q'_k(x_i^{(k)})},$$

odakle je, s obzirom na (2.3.2),

$$(2.3.7) \quad A_i = \frac{1}{Q'_k(x_i^{(k)})} \int_a^b p(t) \frac{Q_k(t)}{t - x_i^{(k)}} dt \quad (i=1, \dots, k).$$

Verižni razlomak (2.3.6) možemo sada jednostavno da razvijemo u "parcijalne" razlomke

$$(2.3.8) \quad \frac{R_k(x)}{Q_k(x)} = \sum_{i=1}^k \frac{A_i}{x - x_i^{(k)}}.$$

Primedba 2.3.1. Koeficijenti A_i igraju značajnu ulogu u numeričkoj integraciji, jer predstavljaju koeficijente Gaussovih kvadraturnih formula i u literaturi su poznati kao Christoffelovi brojevi. Korišćenjem Christoffel-Darbouxovog identiteta može se pokazati da je

$$A_i = - \frac{\|Q_k\|^2}{Q'_k(x_i^{(k)}) Q_{k+1}(x_i^{(k)})} \quad (i=1, \dots, k).$$

U radu [9] nalazimo sledeću definiciju:

Definicija 2.3.1. Za rešenje f_k tročlane rekurentne relacije

$$(2.3.9) \quad y_{k+1} = (x - \beta_k) y_k - \gamma_k y_{k-1}$$

kažemo da je minimalno ako postoji linearne nezavisno rešenje iste rekurentne relacije, u oznaci g_k , tako da je

$$\lim_{k \rightarrow +\infty} \frac{f_k}{g_k} = 0.$$

U istom radu [9] W. Gautschi je dokazao da je egzistencija minimalnog rešenja u uskoj vezi sa odredjenošću momentnog problema za težinsku funkciju $x \mapsto p(x)$.

Pretpostavimo da verižni razlomak (2.3.6) konvergira ka $F(x)$, tj. da je

$$(2.3.10) \quad \lim_{k \rightarrow +\infty} \frac{R_k(x)}{Q_k(x)} = F(x),$$

gde je F definisano pomoću (2.3.4). Ako stavimo

$$f_k(x) = F(x)Q_k(x) - R_k(x),$$

na osnovu (2.3.10) zaključujemo da je

$$\lim_{k \rightarrow +\infty} \frac{f_k(x)}{Q_k(x)} = F(x) - \lim_{k \rightarrow +\infty} \frac{R_k(x)}{Q_k(x)} = 0,$$

što znači da je f_k minimalno rešenje tročlane rekurentne relacije. Korišćenjem (2.3.2) i (2.3.4), f_k se može predstaviti i u obliku

$$(2.3.11) \quad f_k(x) = \int_a^b \frac{p(t)}{x-t} Q_k(t) dt \quad (x \notin [a,b], k=0,1,\dots),$$

što znači da je $f_k(x) = w_k(x)$, gde je w_k funkcija druge vrste definisana ranije. Ako stavimo $f_{-1}(x)=1$, vidimo da rekurentna relacija $f_{k+1}(x)=(x-\beta_k)f_k(x)-\gamma_k f_{k-1}(x)$ važi i za $k=0$.

Prema tome, bilo koji uslov koji garantuje konvergenciju (2.3.10) dovoljan je uslov da f_k , dato sa (2.3.11) bude minimalno rešenje rekurentne relacije (2.3.9). U radu [9] W. Gautschi je dokazao da minimalno rešenje egzistira kad god je momentni problem za težinsku funkciju $x \mapsto p(x)$ određen. U istom radu W. Gautschi je dao i algoritam za izračunavanje minimalnog rešenja $f_k(x)$ za $k=0,1,\dots,n$. Uzimajući s dovoljno veliko ($s > n$), pomoću

$$\left\{ \begin{array}{l} r_s^{(s)} = 0, \quad r_{k-1}^{(s)} = \frac{\gamma_k}{x - \beta_k - r_k^{(s)}} \quad (k=s, s-1, \dots, 1, 0), \\ f_{-1}^{(s)} = 1, \quad f_k^{(s)} = r_{k-1}^{(s)} f_{k-1}^{(s)} \quad (k=0, 1, \dots, n), \end{array} \right.$$

generišu se nizovi $\{r_k^{(s)}\}$ i $\{f_k^{(s)}\}$. Ako je $f_k(x)$ minimalno rešenje rekurentne relacije (2.3.9), tada je (videti [6] i [9])

$$\lim_{s \rightarrow +\infty} f_k^{(s)} = f_k(x) \quad (k=0, 1, \dots, n).$$

Primedba 2.3.2. Korišćenjem razvoja u "parcijalne" razlomke (2.3.11) moguće je dobiti formalni razvoj

$$\lim_{k \rightarrow +\infty} \frac{R_k(x)}{Q_k(x)} = \sum_{m=0}^{+\infty} \frac{C_m}{x^{m+1}},$$

gde su C_m momenti težinske funkcije $x \mapsto p(x)$.

2.2.4. Osobine ortogonalnih polinoma sa parnom težinskom funkcijom

Neka je $\{Q_k\}_{k \in N_0}$ niz ortogonalnih polinoma na $(-a, a)$ sa parnom težinskom funkcijom ($p(-x)=p(x)$).

Kako je, za $n \neq k$,

$$\begin{aligned} \int_{-a}^a p(x) Q_n(x) Q_k(x) dx &= \int_{-a}^a p(-x) Q_n(-x) Q_k(-x) dx \\ &= \int_{-a}^a p(x) Q_n(-x) Q_k(-x) dx = 0 \end{aligned}$$

zaključujemo da je i niz polinoma $\{Q_k(-x)\}_{k \in N_0}$, takodje, ortogonalan u odnosu na težinsku funkciju p . S druge strane, zbog jedinstvenosti niza ortogonalnih polinoma za datu težinsku funkciju do na multiplikativnu konstantu, zaključujemo da mora biti

$Q_n(-x) = c_n Q_n(x)$, odakle sleduje $c_n = (-1)^n$. Dakle, imamo

$$Q_n(-x) = (-1)^n Q_n(x),$$

tj.

$$\begin{aligned} Q_n(-x) &= Q_n(x) && (n - \text{parno}) \\ &= -Q_n(x) && (n - \text{neparno}), \end{aligned}$$

što znači da je

$$Q_{2k}(x) = U_k(x^2) \quad i \quad Q_{2k+1}(x) = x V_k(x^2),$$

gde su U_k i V_k polinomi k-tog stepena.

Neka je $n \neq k$. Tada na osnovu

$$\int_{-a}^a p(x) Q_{2n}(x) Q_{2k}(x) dx = 2 \int_0^a p(x) U_n(x^2) U_k(x^2) dx = 0,$$

smenom $x^2 = y$, dobijamo

$$(2.4.1) \quad \int_0^{a^2} \frac{p(\sqrt{y})}{\sqrt{y}} U_n(y) U_k(y) dy = 0.$$

Slično, imamo

$$\int_{-a}^a p(x) Q_{2n+1}(x) Q_{2k+1}(x) dx = 2 \int_0^{a^2} p(x) x^2 V_n(x^2) V_k(x^2) dx = 0,$$

odakle je

$$(2.4.2) \quad \int_0^{a^2} \sqrt{y} p(\sqrt{y}) V_n(y) V_k(y) dy = 0.$$

Na osnovu (2.4.1) i (2.4.2) možemo formulisati sledeći rezultat:

Teorema 2.4.1. Neka je $\{Q_k\}_{k \in \mathbb{N}_0}$ niz ortogonalnih polinoma na $(-a, a)$ sa parnom težinskom funkcijom $x \mapsto p(x)$. Tada je

1° Niz polinoma $\{Q_{2k}(\sqrt{x})\}_{k \in \mathbb{N}_0}$ ortogonalan na $(0, a^2)$ sa težinskom funkcijom $x \mapsto p(\sqrt{x})/\sqrt{x}$;

2° Niz polinoma $\{Q_{2k+1}(\sqrt{x})/\sqrt{x}\}_{k \in \mathbb{N}_0}$ ortogonalan na $(0, a^2)$ sa težinskom funkcijom $x \mapsto \sqrt{x} p(\sqrt{x})$.

2.2.5. Konstrukcija ortogonalnih polinoma

Samo za neke uske klase ortogonalnih polinoma, kakvi su, na primer, klasični ortogonalni polinomi (videti odeljak 2.2.8), moguće je analitički odrediti koeficijente tročlane rekurentne relacije (2.2.1), tj. (2.2.8). U opštem slučaju za tzv. neklasične težine to nije moguće. Pod konstrukcijom ortogonalnih polinoma podrazumevamo numeričko određivanje prvih n koeficijenata α_k i β_k ($k=0, 1, \dots, n-1$) u rekurentnoj relaciji (2.2.8).

U narednim odeljcima 2.2.6 i 2.2.7 izložićemo dva metoda za konstrukciju ortogonalnih polinoma i to:

- Čebiševljev algoritam,
- Stieltjesovu proceduru.

Detaljnije u vezi ovih metoda može se naći u radu [10].

2.2.6. Čebiševljev algoritam

Neka je $\{Q_k\}_{k \in N_0}$ moničan niz ortogonalnih polinoma koji zadovoljava rekurentnu relaciju

$$(2.6.1) \quad Q_{k+1}(x) = (x - \beta_k) Q_k(x) - \gamma_k Q_{k-1}(x) \quad (k=0, 1, \dots),$$

pri čemu je $Q_{-1}(x)=0$ i $Q_0(x)=1$. Neka je, dalje, poznato prvih $2n$ momenata težinske funkcije

$$C_k = \int_a^b x^k p(x) dx \quad (k=0, 1, \dots, 2n-1).$$

Definišimo veličine $\sigma_{k,i}$ pomoću

$$(2.6.2) \quad \sigma_{k,i} = (Q_k, x^i) = \int_a^b p(x) Q_k(x) x^i dx.$$

Tada je $\sigma_{0,i} = C_i$, $\sigma_{-1,i} = 0$, a zbog ortogonalnosti $\sigma_{k,i} = 0$, ako je $k > i$.

Kao u tački 5° odeljka 2.2.2 uzimimo $\gamma_0 = C_0$.

Na osnovu (2.6.1) za $k=0$ imamo $Q_1(x) = x - \beta_0$, odakle slijedi $(Q_1, Q_0) = 0 = (Q_0, x) - \beta_0 (Q_0, 1)$, tj. $\beta_0 = \sigma_{0,1} / \sigma_{0,0}$. Slično, za

$k \geq 1$ iz (2.6.1) sleduje

$$(Q_{k+1}, x^i) = (Q_k, x^{i+1}) - \beta_k (Q_k, x^i) - \gamma_k (Q_{k-1}, x^i),$$

tj.

$$(2.6.3) \quad \sigma_{k+1,i} = \sigma_{k,i+1} - \beta_k \sigma_{k,i} - \gamma_k \sigma_{k-1,i}.$$

Stavlјajući $i=k-1$ i $i=k$, na osnovу (2.6.3) dobijamo redom

$$0 = \sigma_{k+1,k-1} = \sigma_{k,k} - \beta_k \sigma_{k-1,k-1}$$

i

$$0 = \sigma_{k+1,k} = \sigma_{k,k+1} - \beta_k \sigma_{k,k} - \gamma_k \sigma_{k-1,k}.$$

Prema tome, imamo

$$(2.6.4) \quad \begin{aligned} \beta_0 &= \frac{\sigma_{0,1}}{\sigma_{0,0}}, & \gamma_0 &= \sigma_{0,0}, \\ \beta_k &= \frac{\sigma_{k,k+1}}{\sigma_{k,k}} - \frac{\sigma_{k-1,k}}{\sigma_{k-1,k-1}}, & \gamma_k &= \frac{\sigma_{k,k}}{\sigma_{k-1,k-1}} \quad (k=1,2,\dots). \end{aligned}$$

Veličine $\sigma_{k,i}$ sukcesivno mogu biti generisane iz momenata C_i pomoću

$$(2.6.5) \quad \begin{aligned} \sigma_{k,i} &= \sigma_{k-1,i+1} - \beta_{k-1} \sigma_{k-1,i} - \gamma_{k-1} \sigma_{k-2,i} \quad (i=k,k+1,\dots,2n-k-1), \\ \sigma_{-1,0} &= 0, & \sigma_{0,i} &= C_i. \end{aligned}$$

Formule (2.6.4) i (2.6.5) definišu tzv. Čebiševljev* algoritam ([3], [10]), koji omogućava jednostavnu konstrukciju prvih n koeficijenata $\beta_0, \dots, \beta_{n-1}$ i $\gamma_0, \dots, \gamma_{n-1}$ na osnovu prvih $2n$ datih momenata $C_0, C_1, \dots, C_{2n-1}$. Preslikavanje $K_n: \mathbb{R}^{2n} \rightarrow \mathbb{R}^{2n}$ definisano Čebiševljevim algoritmom

$(K_n: [C_0 C_1 \dots C_{2n-1}]^T \rightarrow [\beta_0 \dots \beta_{n-1} \gamma_0 \dots \gamma_{n-1}]^T)$ vrlo često pokazuje karakteristike slabe uslovjenosti, posebno kada se radi o beskonačnom intervalu ortogonalnosti. Slaba uslovjenost dolazi do izražaja sa porastom n . Uvodjenjem pogodnih modifikovanih momenata M_k (umesto momenata C_k) moguće je u izvesnim

* Čebišev je razmatrao slučaj diskretnih ortogonalnih polinoma (videti [3])

slučajevima stabilisati algoritam. Takav algoritam ćemo zvati modifikovani Čebiševljev algoritam (videti radove [10], [29], [35]), a prezentiraćemo ga u obliku kakav su dali Wheeler ([35]) i Gautschi ([10]).

Neka je $\{w_k\}_{k \in N_0}$ sistem moničnih polinoma koji zadovoljavaju tročlanu rekurentnu relaciju

$$w_{k+1}(x) = (x - b_k) w_k(x) - c_k w_{k-1}(x) \quad (k=0, 1, \dots),$$

pri čemu je $w_{-1}(x)=0$, $w_0(x)=1$, a koeficijenti b_k i c_k poznati. Umesto pomošu (2.6.2), veličine $\sigma_{k,i}$ definišimo sada pomoću

$$(2.6.6) \quad \sigma_{k,i} = (Q_k, w_i) = \int_a^b p(x) Q_k(x) w_i(x) dx$$

i pretpostavimo da nam je poznato prvih $2n$ modifikovanih momenata

$$M_k = \int_a^b p(x) w_k(x) dx \quad (k=0, 1, \dots, 2n-1).$$

Tada, slično prethodnom, imamo

$$\begin{aligned} \sigma_{-1,i} &= 0 \quad (i=1, 2, \dots, 2n-2), \quad \sigma_{0,i} = M_i \quad (i=0, 1, \dots, 2n-1), \\ \beta_0 &= b_0 + \frac{M_1}{M_0}, \quad \gamma_0 = M_0, \end{aligned}$$

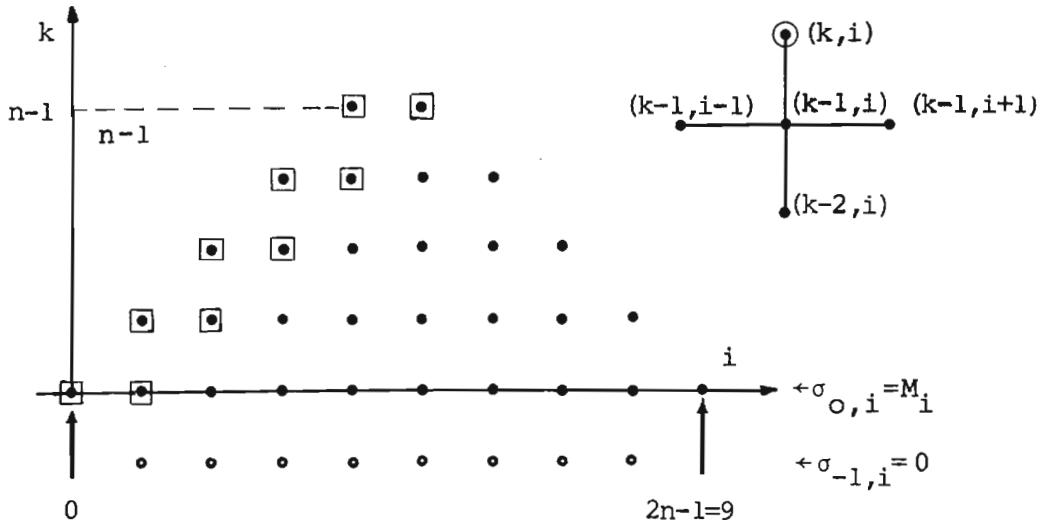
a zatim redom za $k=1, 2, \dots, n-1$

$$\begin{aligned} \sigma_{k,i} &= \sigma_{k-1,i+1} - (\beta_{k-1} - b_i) \sigma_{k-1,i} - \gamma_{k-1} \sigma_{k-2,i} + c_i \sigma_{k-1,i-1} \\ &\quad (i=k, k+1, \dots, 2n-k-1), \\ (2.6.7) \quad \beta_k &= b_k + \frac{\sigma_{k,k+1}}{\sigma_{k,k}} - \frac{\sigma_{k-1,k}}{\sigma_{k-1,k-1}}, \\ \gamma_k &= \frac{\sigma_{k,k}}{\sigma_{k-1,k-1}}. \end{aligned}$$

Primetimo da se u slučaju $w_k(x)=x^k$ ($k=0, 1, \dots, 2n-1$), tj. $b_k=c_k=0$, modifikovani Čebiševljev algoritam svodi na osnovni

algoritam definisan pomoću (2.6.4) i (2.6.5).

Na slici 2.6.1 je data šema koja ilustruje primenu modifikovanog Čebiševljevog algoritma u slučaju $n=5$. U određivanju $\sigma_{k,i}$ pomoću prve relacije u (2.6.7) učestvuju



Sl. 2.6.1

$\sigma_{k-1,i-1}$, $\sigma_{k-1,i}$, $\sigma_{k-1,i+1}$ i $\sigma_{k-2,i}$, što je na slici šematski prikazano navođenjem njihovih indeksa. U određivanju β_k i γ_k , prema drugoj i trećoj relaciji u (2.6.7), učestvuju veličine $\sigma_{k,i}$, čiji su indeksi takvi da odgovaraju poljima koja su na slici uokvirena kvadratićima. Na osnovu (2.6.6) primećimo još da su dijagonalni elementi

$$\sigma_{k,k} = (\varphi_k, \varphi_k) = \|\varphi_k\|^2 \quad (k=0,1,\dots,n-1).$$

U radu [10] W. Gautschi je detaljno analizirao uslovljenost preslikavanja $K_n: \vec{v} \rightarrow \vec{\rho}$, gde su

$$\vec{v}^T = [M_0 M_1 \dots M_{2n-1}], \vec{\rho}^T = [\beta_0 \dots \beta_{n-1} \gamma_0 \dots \gamma_{n-1}].$$

Karakteristike slabe uslovljenosti dolaze do izražaja sa porastom n , i to posebno kada se radi o beskonačnom intervalu ortogonalnosti. Tada male promene u momentima M_k (na primer, izazvane greškama zaokrugljivanja) izazivaju velike promene u vektoru koeficijenata $\vec{\rho}$.

2.2.7. Stieltjesova procedura

Neka je $\{Q_k\}_{k \in N_0}$ moničan niz ortogonalnih polinoma na (a, b) u odnosu na težinsku funkciju $x \mapsto p(x)$. Koeficijenti β_k i γ_k u rekurentnoj relaciji (2.6.1) mogu se jednostavno izraziti u obliku

$$\beta_k = \frac{(xQ_k, Q_k)}{(Q_k, Q_k)} \quad (k=0, 1, 2, \dots),$$

$$\gamma_k = \frac{(xQ_{k-1}, Q_k)}{(Q_{k-1}, Q_{k-1})} = \frac{(Q_k, Q_k)}{(Q_{k-1}, Q_{k-1})} \quad (k=1, 2, \dots),$$

tj.

$$(2.7.1) \quad \beta_k = \frac{\int_a^b p(x)xQ_k(x)^2 dx}{\int_a^b p(x)Q_k(x)^2 dx} \quad (k=0, 1, 2, \dots),$$

$$(2.7.2) \quad \gamma_k = \frac{\int_a^b p(x)Q_k(x)^2 dx}{\int_a^b p(x)Q_{k-1}(x)^2 dx} \quad (k=1, 2, \dots).$$

Kao i ranije možemo uzeti $\gamma_0 = \int_a^b p(x)dx$.

Stieltjes je ukazao da se rekurentna relacija (2.6.1) i formule (2.7.1) i (2.7.2) mogu primeniti na sukcesivno generisanje polinoma $Q_1(x), Q_2(x), \dots$. Naime, kako je $Q_0(x)=1$, pomoću (2.7.1) nalazimo

$$\beta_0 = \frac{\int_a^b p(x)x dx}{\int_a^b p(x) dx}.$$

Ako sada u (2.6.1) stavimo $k=0$, dobijamo $Q_1(x)=x-\beta_0$, jer je $Q_{-1}(x)=0$. Na dalje, na osnovu (2.7.1) i (2.7.2), za $k=1$, na-

lazimo β_1 i γ_1 , a sa ovim vrednostima, pomoću (2.6.1), polinom $Q_2(x) = (x - \beta_1)(x - \beta_0) - \gamma_1$, itd.

Navedeni postupak alternativne primene formula (2.7.1) i (2.7.2) i rekurentne relacije (2.6.1) naziva se Stieltjesova procedura ([30], [10]).

Određivanje vrednosti integrala, koji se pojavljuju u (2.7.1) i (2.7.2), moguće je izvršiti numerički, tj. primenom neke kvadraturne formule (videti poglavje numerička integracija). U tom slučaju navedeni postupak se naziva diskretizovana Stieltjesova procedura. Ovakav postupak pokazuje dobre karakteristike uslovljenosti i ima prednost nad Čebiševljevim algoritmom, pogotovo kada se radi o beskonačnom intervalu (a, b) .

2.2.8. Klasični ortogonalni polinomi

Jednu veoma važnu klasu ortogonalnih polinoma čine tzv. klasični ortogonalni polinomi. Oni se javljaju pri rešavanju niza graničnih problema matematičke fizike, kao i kod rešavanja nekih problema kvantne mehanike. Posebnu primenu nalaze u teoriji aproksimacija i numeričkoj integraciji.

Definicija 2.8.1. Neka je $\{Q_k\}_{k \in \mathbb{N}_0}$ niz ortogonalnih polinoma na (a, b) sa težinskom funkcijom $x \mapsto p(x)$. Polinomi Q_k ($k=0, 1, \dots$) nazivaju se klasičnim ako težinska funkcija zadovoljava diferencijalnu jednačinu

$$(2.8.1) \quad \frac{d}{dx}(A(x)p(x)) = B(x)p(x),$$

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

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

Na osnovu (2.8.1) možemo zaključiti da $p \in C^1(a,b)$, tj. da ima neprekidan prvi izvod za svako $x \in (a,b)$. Jedino na krajevima intervala (a,b) težinska funkcija može imati singularitete.

Teorema 2.8.1. Za svako $m=0,1,\dots$ težinska funkcija klasičnih ortogonalnih polinoma zadovoljava uslove

$$(2.8.2) \quad \lim_{x \rightarrow a^+} x^m A(x)p(x) = 0 \quad i \quad \lim_{x \rightarrow b^-} x^m A(x)p(x) = 0.$$

Dokaz ove teoreme može se naći, na primer u [22, str. 48].

Kako je rešenje diferencijalne jednačine (2.8.1) dano 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

$$(2.8.3) \quad \begin{aligned} p(x) &= (b-x)^a (x-a)^{\beta} && (a \text{ i } b \text{ konačni}), \\ &= (x-a)^s \exp(rx) && (a \text{ konačno}, b=+\infty), \\ &= (b-x)^t \exp(-rx) && (a=-\infty, b \text{ konačno}), \\ &= \exp(\int B(x) dx) && (a=-\infty, b=+\infty), \end{aligned}$$

gde su

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

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

Granični uslovi (2.8.2) zahtevaju da je

$$(2.8.4) \quad \begin{aligned} B(a) &> 0, && \text{ako je } a \text{ konačno}; \\ B(b) &< 0, && \text{ako je } b \text{ konačno}; \\ r=B'(x) &< 0. \end{aligned}$$

U slučaju kada su a i b konačni, treći uslov je posledica prva dva uslova u (2.8.4).

Na osnovu (2.8.3) 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 (2.8.2), tj. (2.8.4), 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 Jacobievi 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° Gegenbauerovi ili ultrasferni polinomi $C_k^\lambda(x)$ ($\alpha = \beta = \lambda - \frac{1}{2}$).

Primetimo da klasa Gegenbauerovih polinoma obuhvata polinome navedene pod 1° , 2° , 3° .

2. Neka je sada $p(x) = x^s e^{-x}$ i $(a,b) = (0, +\infty)$. 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 $s=0$ nazivaju se Laguerreovi polinomi i označavaju se sa $L_k(x)$ ($\equiv L_k^0(x)$).

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 tabeli 2.8.1 dat je pregled osnovnih karakteristika Jacobievih, generalisanih Laguerreovih i Hermiteovih polinoma.

Tabela 2.8.1

(a,b)	$p(x)$	$A(x)$	$B(x)$	$Q_k(x)$
$(-1, 1)$	$(1-x)^\alpha (1+x)^\beta$	$1-x^2$	$\beta-\alpha-(\alpha+\beta+2)x$	$P_k^{(\alpha, \beta)}(x)$
$(0, +\infty)$	$x^s e^{-x}$	x	$s+1-x$	$L_k^s(x)$
$(-\infty, +\infty)$	e^{-x^2}	1	$-2x$	$H_k(x)$

Primedba 2.8.1. S obzirom na jedinstvenost ortogonalnih polinoma (do na multiplikativnu konstantu), na osnovu teoreme 2.4.1 može se izvesti veza izmedju Hermiteovih i generalisanih Laguerreovih polinoma. Neka je $p(x)=e^{-x^2}$ i $a=+\infty$. Na osnovu tvrdjenja 1^o iz teoreme 2.4.1 zaključujemo da su polinomi $\{H_{2k}(\sqrt{x})\}_{k \in \mathbb{N}_0}$ ortogonalni na $(0, \infty)$ u odnosu na težinsku funkciju $p_1(x)=p(\sqrt{x})/\sqrt{x}=x^{-1/2}e^{-x}$, što znači da se radi o generalisanim Laguerreovim polinomima za $s=-1/2$. Prema tome imamo $H_{2k}(\sqrt{x})=c_k L_k^{-1/2}(x)$, tj. $H_{2k}(x)=c_k L_k^{-1/2}(x^2)$, gde je c_k konstanta. Slično, na osnovu tvrdjenja 2^o iz pomenute teoreme 2.4.1 možemo zaključiti da važi

$$H_{2k+1}(x) = d_k x L_k^{1/2}(x^2),$$

gde je d_k konstanta. U vezi sa konstantama c_k i d_k videti odeljak 2.2.13.

Klasični polinomi poseduju veći broj specifičnih svojstava, koja će biti izložena u narednim odeljcima.

2.2.9. Osobine klasičnih ortogonalnih polinoma

Teorema 2.9.1. Izvodi klasičnih ortogonalnih polinoma su, takođe, klasični ortogonalni polinomi.

Dokaz. Da bismo dokazali ovo tvrdjenje podjimo od integrala

$$I_m = \int_a^b p(x)x^{m-1}B(x)Q_k(x)dx \quad (m \in \mathbb{N}),$$

koji je za $m < k$ jednak nuli. Naime, kako je stepen polinoma $x \mapsto x^{m-1}B(x)$ jednak m , to je zbog ortogonalnosti

$$I_m = (x^{m-1}B(x), Q_k) = 0 \quad (m < k).$$

S druge strane, korišćenjem diferencijalne jednačine za težinsku funkciju (2.8.1) i primenom parcijalne integracije na poslednji integral dobijamo

$$\begin{aligned} I_m &= \int_a^b x^{m-1}Q_k(x) \frac{d}{dx}(A(x)p(x))dx \\ &= -(m-1) \int_a^b p(x)x^{m-2}A(x)Q_k(x)dx - \\ &\quad - \int_a^b A(x)p(x)x^{m-1}Q'_k(x)dx, \end{aligned}$$

pri čemu smo iskoristili uslove (1.4.2). Kako je, dalje, $x \mapsto x^{m-2}A(x)$ polinom najviše stepena m , to je i prvi integral na desnoj strani poslednje jednakosti jednak nuli.

Na osnovu prethodnog možemo zaključiti da za svako $m < k$ ($m \in \mathbb{N}$) važi

$$\int_a^b p_1(x)x^{m-1}Q'_k(x)dx = 0,$$

gde smo stavili $p_1(x) = A(x)p(x)$, što dokazuje tvrdjenje da je

niz $\{Q'_k\}_{k \in N}$ ortogonalan na (a, b) sa težinskom funkcijom $x \mapsto A(x)p(x)$. Potrebno je još dokazati da ova težinska funkcija zadovoljava diferencijalnu jednačinu oblika (2.8.1).

Kako je

$$\begin{aligned}\frac{d}{dx}(A(x)p_1(x)) &= A'(x)p_1(x) + A(x)\frac{d}{dx}(A(x)p(x)) \\ &= (A'(x) + B(x))A(x)p(x) \\ &= B_1(x)p_1(x)\end{aligned}$$

i $x \mapsto B_1(x)$ polinom prvog stepena, dokaz teoreme je završen.

Koristeći se matematičkom indukcijom lako se dokazuje sledeća opštija teorema.

Teorema 2.9.2. Niz $\{Q_k^{(m)}\}_{k=m, m+1, \dots}$ predstavlja niz klasičnih ortogonalnih polinoma na (a, b) sa težinskom funkcijom $x \mapsto p_m(x) = A(x)^m p(x)$. Odgovarajuća diferencijalna jednačina za težinsku funkciju je

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

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

Teorema 2.9.3. Polinom $x \mapsto Q_k(x)$ je jedno partikularno rešenje linearne homogene diferencijalne jednačine drugog reda

$$(2.9.1) \quad A(x)y'' + B(x)y' + \lambda_k y = 0,$$

gde je

$$(2.9.2) \quad \lambda_k = -k[\frac{1}{2}(k-1)A''(0) + B'(0)] \quad (k \in N_0).$$

Dokaz. Neka je $m < k$. Kako je niz $\{Q'_k\}_{k \in N_0}$ ortogonalan na (a, b) sa težinskom funkcijom $x \mapsto p_1(x) = A(x)p(x)$ važi jednačnost

$$I_m = \int_a^b A(x)p(x)x^{m-1}Q'_k(x)dx = 0.$$

Primenom parcijalne integracije na poslednji integral i korišćenjem uslova (2.8.2), dobijamo

$$I_m = -\frac{1}{m} \int_a^b \frac{d}{dx} (A(x)p(x)Q'_k(x)) x^m dx,$$

tj.

$$I_m = -\frac{1}{m} \int_a^b p(x) x^m \tilde{Q}_k(x) dx,$$

gde smo stavili $\tilde{Q}_k(x) = B(x)Q'_k(x) + A(x)Q''_k(x)$. S obzirom da je $I_m = 0$ za svako $m < k$ ($m \in N$) zaključujemo da je niz polinoma $\{\tilde{Q}_k\}_{k \in N}$ ortogonalan na (a, b) sa težinskom funkcijom $x \mapsto p(x)$. S druge strane, zbog jedinstvenosti ovih polinoma do na multiplikativnu konstantu, imamo

$$(2.9.3) \quad A(x)Q''_k(x) + B(x)Q'_k(x) + \lambda_k Q_k(x) = 0,$$

gde smo stavili $\tilde{Q}_k(x) = -\lambda_k Q_k(x)$.

Da bismo odredili λ_k dovoljno je uporediti koeficijente uz x^k u (2.9.3).

Ako pretpostavimo da je $Q_k(x) = a_k x^k + \dots$, imamo

$$Q'_k(x) = k a_k x^{k-1} + \dots \quad \text{i} \quad Q''_k(x) = k(k-1) a_k x^{k-2} + \dots .$$

Zamenom u (2.9.3) dobijamo

$$\frac{1}{2} A''(0)k(k-1) + kB'(0) + \lambda_k = 0,$$

tj. (2.9.2), čime je dokaz teoreme završen.

Primedba 2.9.1. Korišćenjem diferencijalne jednačine za težinsku funkciju (2.8.1) diferencijalna jednačina (2.9.1) se može predstaviti u Sturm-Liouvilleovom obliku

$$(2.9.4) \quad \frac{d}{dx} (A(x)p(x)y') + \lambda_k p(x)y = 0.$$

Supstitucijom

$$\begin{pmatrix} p(x) & B(x) & Q_k & k \\ p_m(x) & B_m(x) & Q_k^{(m)} & k-m \end{pmatrix}$$

u (2.9.4) dobijamo diferencijalnu jednačinu za $Q_k^{(m)}(x)$

$$(2.9.5) \quad \frac{d}{dx}(p_{m+1}Q_k^{(m+1)}(x)) + \lambda_{km} p_m(x) Q_k^{(m)}(x) = 0,$$

gde je

$$\begin{aligned} \lambda_{km} &= -(k-m) [\frac{1}{2}(k-m-1)A''(0) + B_m'(0)] \\ &= -(k-m) [\frac{1}{2}(k+m-1)A''(0) + B'(0)]. \end{aligned}$$

Teorema 2.9.4. Za članove niza klasičnih ortogonalnih polinoma $\{Q_k\}_{k \in N_0}$ važi formula

$$(2.9.6) \quad Q_k(x) = \frac{C_k}{p(x)} \cdot \frac{d^k}{dx^k} (A(x)^k p(x)) \quad (k=0,1,\dots),$$

gde su C_k konstante različite od nule.

Dokaz. Sukcesivnom primenom jednakosti (2.9.5) za $m=0,1,\dots,k-1$, dobijamo

$$\begin{aligned} p(x)Q_k(x) &= -\frac{1}{\lambda_{k0}} \cdot \frac{d}{dx}(p_1(x)Q_k'(x)) \\ &= \frac{(-1)^2}{\lambda_{k0}\lambda_{k1}} \cdot \frac{d^2}{dx^2}(p_2(x)Q_k''(x)) \\ &\vdots \\ &= \frac{(-1)^k}{\prod_{m=0}^{k-1} \lambda_{km}} \cdot \frac{d^k}{dx^k}(p_k(x)Q_k^{(k)}(x)). \end{aligned}$$

Pretpostavimo da je $Q_k(x) = a_k x^k + \dots$. Tada je $Q_k^{(k)}(x) = k! a_k$. Ako stavimo

$$C_k = \frac{(-1)^k k! a_k}{\prod_{m=0}^{k-1} \lambda_{km}},$$

iz prethodnog sleduje (2.9.6).

Formula (2.9.6) je poznata kao Rodriguesova formula. Rodrigues je ovu formulu dokazao samo za Legendreove polinome.

Korišćenjem Cauchyeve formule za izvod reda k regulare funkcije, formula (2.9.6) se može predstaviti u integralnom obliku

$$(2.9.7) \quad Q_k(x) = \frac{C_k}{p(x)} \frac{k!}{2\pi i} \oint_{\Gamma} \frac{A(z)^k p(z)}{(z-x)^{k+1}} dz,$$

gde je Γ zatvorena kontura i takva da $x \in \text{int}\Gamma$.

U formulama (2.9.6) i (2.9.7) konstanta C_k se može odabrat na razne načine (na primer, iz uslova da je $\|Q_k\|=1$, ili da su polinomi monični). Jedan način izbora konstante C_k , koji ima istorijski značaj, je sledeći

$$(2.9.8) \quad C_k = \begin{cases} \frac{(-1)^k}{2^k k!} & \text{za } P_k^{(\alpha, \beta)}(x), \\ 1 & \text{za } L_k^s(x), \\ (-1)^k & \text{za } H_k(x). \end{cases}$$

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

$$(2.9.9) \quad \begin{aligned} C_k^\lambda(x) &= \frac{(2\lambda)_k}{(\lambda + \frac{1}{2})_k} P_k^{(\lambda - \frac{1}{2}, \lambda - \frac{1}{2})}(x), \\ T_k(x) &= \frac{k!}{(\frac{1}{2})_k} P_k^{(-1/2, -1/2)}(x), \end{aligned}$$

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

gde je uvedena oznaka

$$(s)_k = s(s+1)\dots(s+k-1) = \frac{\Gamma(s+k)}{\Gamma(s)} \quad (\Gamma \text{ gama funkcija}).$$

2.2.10. Određivanje norme i dva najstarija koeficijenta klasičnih ortogonalnih polinoma

Norma klasičnih ortogonalnih polinoma može se jednostavno odrediti korišćenjem Rodriguesove formule (2.9.6).

Ako sa a_k označimo koeficijent uz najviši stepen u polinomu Q_k , tj. $Q_k(x) = a_k x^k + \dots$, imamo

$$\|Q_k\|^2 = \int_a^b p(x) Q_k(x)^2 dx = \int_a^b p(x) Q_k(x) (a_k x^k + \dots) dx,$$

tj.

$$\|Q_k\|^2 = a_k \int_a^b p(x) x^k Q_k(x) dx.$$

Korišćenjem Rodriguesove formule (2.9.6) dobijamo

$$\|Q_k\|^2 = a_k c_k \int_a^b x^k \frac{d^k}{dx^k} (A(x)^k p(x)) dx,$$

odakle, sukcesivnom primenom parcijalne integracije k puta, sleduje

$$(2.10.1) \quad \|Q_k\|^2 = a_k c_k (-1)^k k! \int_a^b A(x)^k p(x) dx.$$

Ostaje pitanje određivanja najstarijeg koeficijenta a_k polinoma Q_k . U tom cilju razmatraćemo, najpre, izvode

$$F_i(x; k) = \frac{d^i}{dx^i} (A(x)^k p(x)) \quad (i=0, 1, \dots, k),$$

gde je k fiksiran broj. Primetimo da je

$$F_0(x; k) = A(x)^k p(x), \quad F_1(x; k) = A(x)^{k-1} p(x) R_1(x),$$

gde je $R_1(x) = B(x) + (k-1)A'(x)$. Matematičkom indukcijom možemo dokazati da je

$$F_i(x; k) = A(x)^{k-i} p(x) R_i(x) \quad (i=0, 1, \dots, k),$$

gde je

$$(2.10.2) \quad R_i(x) = [B(x) + (k-i)A'(x)]R_{i-1}(x) + A(x)R'_{i-1}(x) \quad (i=1, \dots, k)$$

i $R_0(x) = 1$. Dakle, R_i je polinom i -tog stepena.

Na osnovu Rodriguesove formule (2.9.6) imamo

$$(2.10.3) \quad Q_k(x) = \frac{C_k}{p(x)} F_k(x; k) = C_k R_k(x).$$

Ako stavimo

$$R_i(x) = a_i^{(k)} x^i + b_i^{(k)} x^{i-1} + \dots$$

i $B(x) = B(0) + B'(0)x$, $A(x) = A(0) + A'(0)x + \frac{1}{2}A''(0)x^2$, tada upoređivanjem koeficijenata uz x^i u relaciji (2.10.2) dobijamo

$$a_i^{(k)} = (B'(0) + (k-i)A''(0))a_{i-1}^{(k)} + \frac{1}{2} A''(0)(i-1)a_{i-1}^{(k)},$$

tj.

$$a_i^{(k)} = (B'(0) + \frac{1}{2}(2k-i-1)A''(0))a_{i-1}^{(k)} \quad (a_0^{(k)} \equiv 1),$$

odakle sleduje

$$a_k^{(k)} = \prod_{i=1}^k (B'(0) + \frac{1}{2}(2k-i-1)A''(0)).$$

Najzad, na osnovu (2.10.3) imamo $a_k = C_k a_k^{(k)}$, tj.

$$(2.10.4) \quad a_k = C_k \prod_{i=1}^k (B'(0) + \frac{1}{2}(2k-i-1)A''(0)).$$

Slično se može odrediti i koeficijent $b_i^{(k)}$ u polinomu $R_i(x)$, a samim tim i koeficijent b_k u polinomu $Q_k(x) = a_k x^k + b_k x^{k-1} + \dots$. Dakle, uporedjivanjem koeficijenata uz x^{i-1} u (2.10.2) dobijamo

$$b_i^{(k)} = H(0)a_{i-1}^{(k)} + (H'(0) - \frac{1}{2}A''(0)i)b_{i-1}^{(k)} \quad (b_0^{(k)} \equiv 0),$$

gde smo stavili $H(t) = B(t) + (k-1)A'(t)$. Kombinujući ovu rekurentnu relaciju sa odgovarajućom rekurentnom relacijom za $a_i^{(k)}$, posle glomaznih transformacija dobijamo

$$b_k^{(k)} = \frac{H(0)}{H'(0)} k a_k^{(k)},$$

tj.

$$(2.10.5) \quad r_k = \frac{b_k}{a_k} = k \frac{H(0)}{H'(0)} = k \frac{B(0) + (k-1)A'(0)}{B'(0) + (k-1)A''(0)}.$$

Razmotrićemo sada konkretne slučajeve:

¹ Kod Jacobievih polinoma imamo $B'(0) = -(\alpha + \beta + 2)$ i $A''(0) = -2$. Tada, na osnovu (2.10.4), dobijamo

$$a_k = C_k (-1)^k \prod_{i=1}^k (2k+1-i+\alpha+\beta) = (-1)^k C_k \frac{\Gamma(2k+\alpha+\beta+1)}{\Gamma(k+\alpha+\beta+1)}.$$

Imajući u vidu (2.9.8), za a_k i kvadrat norme (2.10.1) dobijamo

$$(2.10.6) \quad a_k = \frac{1}{2^k k!} \cdot \frac{\Gamma(2k+\alpha+\beta+1)}{\Gamma(k+\alpha+\beta+1)} = \frac{(k+\alpha+\beta+1)_k}{2^k k!}$$

i

$$(2.10.7) \quad \|P_k^{(\alpha, \beta)}\|^2 = \frac{2^{\alpha+\beta+1} \Gamma(k+\alpha+1) \Gamma(k+\beta+1)}{k! (2k+\alpha+\beta+1) \Gamma(k+\alpha+\beta+1)}.$$

Takodje, na osnovu (2.10.5) za odnos koeficijenata b_k i a_k imamo

$$(2.10.8) \quad r_k = \frac{b_k}{a_k} = \frac{k(\alpha-\beta)}{\alpha+\beta+2k}.$$

U specijalnom slučaju kod Gegenbauerovih polinoma ($\alpha=\beta=\lambda-1/2$) imamo

$$(2.10.11) \quad a_k = \frac{(2\lambda)_k}{(\lambda + \frac{1}{2})_k} \cdot \frac{(k+2\lambda)_k}{2^k k!} = \frac{2^k}{k!} (\lambda)_k, \quad b_k = 0,$$

$$(2.10.12) \quad \|c_k^\lambda\|^2 = \sqrt{\pi} \frac{(2\lambda)_k \Gamma(\lambda + \frac{1}{2})}{(k+\lambda)k! \Gamma(\lambda)}.$$

Kod Legendreovih polinoma ($\alpha=\beta=0$), (2.10.6) i (2.10.7) se svode na

$$a_k = \frac{(2k)!}{2^k (k!)^2} \quad \text{i} \quad \|p_k\|^2 = \frac{2}{2k+1}.$$

Za Čebiševljeve polinome prve vrste ($\alpha=\beta=-1/2$) imamo

$$a_k = 2^{k-1}, \quad \|T_0\|^2 = \pi, \quad \|T_k\|^2 = \frac{\pi}{2} \quad (k \neq 0),$$

dok je kod Čebiševljevih polinoma druge vrste ($\alpha=\beta=1/2$)

$$a_k = 2^k \quad \text{i} \quad \|s_k\|^2 = \frac{\pi}{2}.$$

2° Kod generalisanih Laguerreovih polinoma imamo $B'(0)=-1$ i $A''(0)=0$. Tada je, na osnovu (2.10.4), (2.10.5) i (2.9.8)

$$(2.10.13) \quad a_k = (-1)^k, \quad r_k = -k(k+s), \quad \|L_k^s\|^2 = k! \Gamma(k+s+1).$$

Kod standardnih Laguerreovih polinoma ($s=0$) imamo

$$a_k = (-1)^k, \quad r_k = -k^2, \quad \|L_k\|^2 = k!.$$

3° Kod Hermiteovih polinoma $B'(0)=-2$ i $A''(0)=0$, pa je

$$(2.10.14) \quad a_k = 2^k, \quad r_k = 0, \quad \|H_k\|^2 = 2^k k! \sqrt{\pi}.$$

Primetimo da je kod polinoma sa parnom težinskom funkcijom (Gegenbauerovi i Hermiteovi polinomi) koeficijent r_k jednak nuli.

Ostale osobine svih pomenutih specijalnih slučajeva klasičnih ortogonalnih polinoma biće izložene u odeljku 2.2.12.

2.2.11. Formule za diferenciranje i integraciju

Kod klasičnih ortogonalnih polinoma mogućno je izraziti izvod polinoma Q_k pomoću polinoma Q_k i Q_{k-1} . Naime, važi sledeća formula

$$(2.11.1) \quad A(x)Q'_k(x) = (u_k x + v_k)Q_k(x) - w_k Q_{k-1}(x),$$

gde su

$$u_k = \frac{1}{2} k A''(0), \quad v_k = k A'(0) - \frac{1}{2} r_k A''(0), \quad r_k = \frac{b_k}{a_k},$$

(2.11.2)

$$w_k = \frac{\gamma_k}{\alpha_k} (B'(0) + (k - \frac{1}{2}) A''(0)),$$

α_k i γ_k koeficijenti u tročlanoj rekurentnoj relaciji

$$(2.11.3) \quad Q_{k+1}(x) = (\alpha_k x + \beta_k)Q_k(x) - \gamma_k Q_{k-1}(x).$$

Odnos koeficijenata r_k odredjen je u prethodnom odeljku. Korišćenjem (2.11.3) desna strana u (2.11.1) može biti izražena pomoću Q_k i Q_{k+1} .

Da bismo dokazali formulu (2.11.1) primetimo da je $F(x) = A(x)Q'_k(x) - u_k x Q_k(x)$ polinom stepena ne višeg od k , što znači da se može predstaviti kao linearna kombinacija polinoma Q_0, Q_1, \dots, Q_k . Prema tome, imamo

$$(2.11.4) \quad F(x) = v_k Q_k(x) - \sum_{i=1}^k w_i Q_{i-1}(x).$$

Slično kao u dokazu tročlane rekurentne relacije (videti teoremu 2.2.3) i ovde iz uslova ortogonalnosti zaključujemo da je $w_1 = w_2 = \dots = w_{k-1} = 0$, pri čemu smo koristili činjenicu da je

$$(2.11.5) \quad \int_a^b p(x)A(x)Q'_k(x)Q_j(x)dx = 0 \quad (0 \leq j \leq k-2).$$

Dokaz jednakosti (2.11.5) zasniva se na ideji koja je korišćena u dokazu teoreme 2.9.1, s tim što se ovde pojavljuje po-

linom $Q_j(x)$ umesto x^{m-1} .

Uporedjivanjem koeficijenata uz x^k u (2.11.4) jednostavno nalazimo koeficijent v_k . Za određivanje w_k koristimo i diferencijalnu jednačinu (2.9.4). Naime, na osnovu (2.11.1) imamo

$$\int_a^b (A(x)p(x)Q'_k(x))x^{k-1}dx = u_k(Q_k, x^k) + v_k(Q_k, x^{k-1}) - w_k(Q_{k-1}, x^{k-1}),$$

odakle, primenom parcijalne integracije na integral koji se pojavljuje na levoj strani poslednje jednakosti i korišćenjem diferencijalne jednačine za klasične ortogonalne polinome u obliku (2.9.4), dobijamo

$$\frac{\lambda_k}{ka_k} \|Q_k\|^2 = \frac{u_k}{a_k} \|Q_k\|^2 - \frac{w_k}{a_{k-1}} \|Q_{k-1}\|^2,$$

gde je a_k najstariji koeficijent u polinomu Q_k , a λ_k dato po-moću (2.9.2). Najzad, ako iskoristimo (2.2.7) iz poslednje jednakosti dobijamo koeficijent w_k .

Formula (2.11.1) je poznata kao formula za diferenciranje klasičnih ortogonalnih polinoma. Da bismo dobili odgovarajuću formulu za integraciju, diferencirajmo (2.11.1) i (2.11.3), a zatim iskoristimo diferencijalnu jednačinu (2.9.1). Pogodnom eliminacijom dobijamo formulu

$$(2.11.6) \quad \omega_k Q_k(x) = \xi_k Q'_{k-1}(x) + \eta_k Q'_k(x) + \zeta_k Q'_{k+1}(x),$$

gde su $\omega_k, \xi_k, \eta_k, \zeta_k$ konstante odredjene sa

$$(2.11.7) \quad \left\{ \begin{array}{l} \omega_k = (k+1)m_k, \quad m_k = B'(0) + \frac{1}{2}(k-2)A''(0), \\ \xi_k = -\frac{\gamma_k}{2a_k}(k+1)A''(0), \quad \zeta_k = \frac{m_k}{a_k}, \\ \eta_k = B(0) + (k-1)A'(0) - \frac{1}{2}r_k A''(0) - (r_{k+1} - r_k)m_k. \end{array} \right.$$

S obzirom da je

$$(2.11.8) \quad \omega_k \int Q_k(x) dx = \xi_k Q_{k-1}(x) + \eta_k Q_k(x) + \zeta_k Q_{k+1}(x) + \text{const},$$

formula (2.11.6), tj. (2.11.8), se naziva formula za integraciju klasičnih ortogonalnih polinoma ili Christoffelova rekurentna relacija. Slično se rekurentna relacija (2.11.3) naziva Bonnetova rekurentna relacija.

2.2.12. Funkcija generatrise

Klasični ortogonalni polinomi se mogu razmatrati kac koeficijenti Taylorovog razvoja neke analitičke funkcije.

Definicija 2.12.1. Funkciju $(x, t) \mapsto \phi(x, t)$ nazivamo funkcijom generatrise za sistem polinoma $\{Q_k\}_{k \in N_0}$ ako je, za dovoljno malo t ,

$$\phi(x, t) = \sum_{k=0}^{+\infty} \frac{\bar{Q}_k(x)}{k!} t^k,$$

gde je $\bar{Q}_k(x) = Q_k(x)/C_k$ i C_k normalizaciona konstanta koja se pojavljuje u Rodriguesovoj formuli (2.9.6).

Kako je, na osnovu (2.9.7),

$$\frac{\bar{Q}_k(x)}{k!} = \frac{1}{p(x)} \cdot \frac{1}{2\pi i} \oint_{\Gamma} \frac{A(z)^k p(z)}{(z-x)^{k+1}} dz \quad (x \in \text{intr})$$

imamo

$$\phi(x, t) = \frac{1}{2\pi i} \cdot \frac{1}{p(x)} \oint_{\Gamma} \frac{p(z)}{z-x} \left(\sum_{k=0}^{+\infty} \left(\frac{A(z)t}{z-x} \right)^k \right) dz.$$

Kako je za dovoljno malo t , $\left| \frac{A(z)t}{z-x} \right| < 1$, imamo

$$\phi(x, t) = \frac{1}{2\pi i} \cdot \frac{1}{p(x)} \oint_{\Gamma} \frac{p(z)}{z-x-A(z)t} dz.$$

Ako $t \rightarrow 0$, zaključujemo da jednačina

$$(2.12.1) \quad z - x - A(z)t = 0$$

ima koren $z=x$, a drugi koren ako postoji teži tački ∞ . Dakle, pri dovoljno malom t može se uzeti da unutar konture Γ leži samo jedan koren $z=g(x,t)$, što znači da podintegralna funkcija u unutrašnjosti konture Γ ima samo prost pol $z=g(x,t)$. Tada je

$$\begin{aligned} \Phi(x,t) &= \frac{1}{p(x)} \operatorname{Res}_{z=g(x,t)} \left\{ \frac{p(z)}{z-x-A(z)t} \right\} \\ &= \frac{1}{p(x)} \cdot \frac{p(z)}{1-A'(z)t} \Big|_{z=g(x,t)}, \end{aligned}$$

gde je $z=g(x,t)$ koren jednačine (2.12.1) koji je za dovoljno malo t blizak tački $z=x$.

Primer 2.12.1. Za Legendreove polinome imamo $p(x)=1$ i $A(x)=1-x^2$. Jednačina (2.12.1) u ovom slučaju je $z-x-(1-z^2)t=0$, odakle slijedi

$$g(x,t) = \frac{-1+\sqrt{1+4t(t+x)}}{2t}.$$

Tada, na osnovu prethodnog, imamo

$$\Phi(x,t) = \frac{1}{1+2zt} \Big|_{z=g(x,t)} = \frac{1}{\sqrt{1+4tx+4t^2}},$$

gdje je

$$\phi(x,t) = \sum_{k=0}^{+\infty} \frac{P_k(x)}{C_k k!} t^k.$$

Kako je $C_k = \frac{(-1)^k}{2^k k!}$, imamo

$$\frac{1}{\sqrt{1+4tx+4t^2}} = \sum_{k=0}^{+\infty} P_k(x) (-2t)^k,$$

tj.

$$\frac{1}{\sqrt{1-2tx+t^2}} = \sum_{k=0}^{+\infty} p_k(x) t^k.$$

Primer 2.12.2. Za generalisane Laguerreove polinome imamo $p(x)=x^s e^{-x}$ ($s > -1$) i $A(x)=x$. Iz jednačine $z-x-zt=0$ sleduje

$$g(x, t) = \frac{x}{1-t}, \text{ pa je}$$

$$\Phi(x, t) = \frac{1}{x^s e^{-x}} \left(\frac{x}{1-t}\right)^s e^{-\frac{x}{1-t}} \cdot \frac{1}{1-t} = (1-t)^{-(s+1)} e^{-\frac{xt}{1-t}}.$$

Dakle,

$$(1-t)^{-(s+1)} e^{-\frac{xt}{1-t}} = \sum_{k=0}^{+\infty} \frac{1}{k!} L_k^s(x) t^k.$$

Primer 2.12.3. Kod Hermiteovih polinoma imamo $p(x)=e^{-x^2}$ i $A(x)=1$. Jednačina (2.12.1), koja u ovom slučaju ima oblik $z-x-t=0$, ima rešenje $g(x, t)=x+t$. Tada je, na osnovu prethodnog,

$$\Phi(x, t) = e^{-(x+t)^2+x^2} = e^{-2xt-t^2}.$$

Dakle, imamo

$$e^{-2xt-t^2} = \sum_{k=0}^{+\infty} \frac{H_k(x)}{(-1)^k k!} t^k,$$

tj.

$$e^{2xt-t^2} = \sum_{k=0}^{+\infty} \frac{H_k(t)}{k!} t^k.$$

Primedba 2.12.1. U knjizi [19] ortogonalni polinomi se uvode preko funkcije generatrise.

2.2.13. Specijalni slučajevi klasičnih ortogonalnih polinoma

U ovom odeljku daćemo pregled najvažnijih rezultata koji se odnose na specijalne slučajeve klasičnih ortogonalnih polinoma. Korišćena notacija je u skladu sa prethodnim odeljcima.

1. Jacobievi polinomi $P_k^{(\alpha, \beta)}(x)$.

Ovde je $(a, b) = (-1, 1)$, $p(x) = (1-x)^\alpha (1+x)^\beta$ ($\alpha, \beta > -1$),
 $A(x) = 1-x^2$, $B(x) = \beta - \alpha - (\alpha + \beta + 2)x$.

Diferencijalna jednačina Jacobieva polinoma ima oblik

$$(2.13.1) \quad (1-x^2)y'' + (\beta - \alpha - (\alpha + \beta + 2)x)y' + k(k+\alpha+\beta+1)y = 0.$$

Tročlana rekurentna relacija (Bonnetova relacija) ima oblik

$$\begin{aligned} 2(k+1)(k+\alpha+\beta+1)(2k+\alpha+\beta)P_{k+1}^{(\alpha, \beta)}(x) &= \\ &= (2k+\alpha+\beta+1)((2k+\alpha+\beta)(2k+\alpha+\beta+2)x - \alpha^2 - \beta^2)P_k^{(\alpha, \beta)}(x) \\ &- 2(k+\alpha)(k+\beta)(2k+\alpha+\beta+2)P_{k-1}^{(\alpha, \beta)}(x), \end{aligned}$$

gde je $P_0^{(\alpha, \beta)}(x) = 1$. Rodriguesova formula (2.9.6), u ovom slučaju glasi

$$P_k^{(\alpha, \beta)}(x) = \frac{(-1)^k}{2^k k!} (1-x)^{-\alpha} (1+x)^{-\beta} \frac{d^k}{dx^k} ((1-x)^{k+\alpha} (1+x)^{k+\beta}),$$

odakle se može dobiti eksplicitan izraz

$$P_k^{(\alpha, \beta)}(x) = \frac{1}{2^k} \sum_{i=0}^k \binom{k+\alpha}{i} \binom{k+\beta}{k-i} (x-1)^{k-i} (x+1)^i.$$

Primetimo da je $P_k^{(\alpha, \beta)}(-x) = (-1)^k P_k^{(\beta, \alpha)}(x)$ i

$$P_k^{(\alpha, \beta)}(1) = \binom{k+\alpha}{k} = \frac{(\alpha+1)_k}{k!}.$$

Na osnovu prethodnog koeficijenti u Bonnetovoj relaciji su:

$$\alpha_k = \frac{(2k+\alpha+\beta+1)(2k+\alpha+\beta+2)}{2(k+1)(k+\alpha+\beta+1)},$$

$$\beta_k = \frac{(2k+\alpha+\beta+1)(\alpha^2 - \beta^2)}{2(k+1)(k+\alpha+\beta+1)(2k+\alpha+\beta)},$$

$$\gamma_k = \frac{(k+\alpha)(k+\beta)(2k+\alpha+\beta+2)}{(k+1)(k+\alpha+\beta+1)(2k+\alpha+\beta)}.$$

Konstanta C_k , kvadrat norme $\|P_k^{(\alpha, \beta)}\|^2$ i koeficijenti a_k i b_k ($r_k = b_k/a_k$) su dati pomoću (2.9.8), (2.10.7), (2.10.6) i (2.10.8) respektivno.

Na osnovu (2.11.2) i (2.11.7) nalazimo redom

$$u_k = -k, \quad v_k = \frac{k(\alpha-\beta)}{2k+\alpha+\beta}, \quad w_k = -\frac{2(k+\alpha)(k+\beta)}{2k+\alpha+\beta},$$

$$\omega_k = -(k+1)(k+\alpha+\beta), \quad \xi_k = \frac{2(k+1)(k+\alpha)(k+\beta)}{(2k+\alpha+\beta)(2k+\alpha+\beta+1)},$$

$$\eta_k = \frac{2(k+1)(k+\alpha+\beta)(\beta-\alpha)}{(2k+\alpha+\beta)(2k+\alpha+\beta+2)}, \quad \zeta_k = -\frac{2(k+1)(k+\alpha+\beta+1)(k+\alpha+\beta)}{(2k+\alpha+\beta+1)(2k+\alpha+\beta+2)}.$$

Dobijene vrednosti daju koeficijente u formulama (2.11.1) i (2.11.6).

1.a. Gegenbauerovi polinomi $C_k^\lambda(x)$.

Ovde je $(a, b) = (-1, 1)$, $\alpha = \beta = \lambda - \frac{1}{2}$ ($\lambda > -\frac{1}{2}$), $p(x) = (1-x^2)^{\lambda-1/2}$, $A(x) = 1-x^2$, $B(x) = -(2\lambda+1)x$ i, prema (2.9.9),

$$(2.13.2) \quad C_k^\lambda(x) = \frac{(2\lambda)_k}{(\lambda + \frac{1}{2})_k} P_k^{(\alpha, \alpha)}(x) \quad (\alpha = \lambda - \frac{1}{2}).$$

Slučaj $\lambda=0$ isključićemo sada iz razmatranja. Može se pokazati da je

$$(2.13.3) \quad \lim_{\lambda \rightarrow 0} \frac{C_k^\lambda(x)}{\lambda} = \frac{2}{k} T_k(x) \quad (k=1, 2, \dots),$$

gde je T_k Čebiševljev polinom prve vrste.

Koeficijenti a_k i b_k i kvadrat norme Gegenbauerovih polinoma su dati pomoću (2.10.11) i (2.10.12).

S obzirom da je težinska funkcija parna imamo da je $C_k^\lambda(-x) = (-1)^k C_k^\lambda(x)$.

Kako je $P_k^{(\alpha, \beta)}(1) = \binom{k+\alpha}{k} = \frac{(\alpha+1)_k}{k!}$, na osnovu (2.13.2) dobijamo

$$C_k^\lambda(1) = \binom{k+2\lambda-1}{k} = \frac{(2\lambda)_k}{k!}.$$

Ako stavimo $\alpha=\beta=\lambda-1/2$, diferencijalna jednačina (2.13.1) se svodi na Gegenbauerovu diferencijalnu jednačinu

$$(2.13.4) \quad (1-x^2)y'' - (2\lambda+1)xy' + k(k+2\lambda)y = 0.$$

Slično se na osnovu (2.13.2) i relacije za Jacobieve polinome dobija:

- Rekurentna relacija (Bonnetova relacija)

$$(k+1)C_{k+1}^\lambda(x) = 2(k+\lambda)x C_k^\lambda(x) - (k+2\lambda-1)C_{k-1}^\lambda(x),$$

gde su $C_0^\lambda(x)=1$ i $C_1^\lambda(x)=2\lambda x$;

- Formula diferenciranja

$$\begin{aligned} (1-x^2) \frac{d}{dx} C_k^\lambda(x) &= -kx C_k^\lambda(x) + (k+2\lambda-1) C_{k-1}^\lambda(x) \\ &= (k+2\lambda)x C_k^\lambda(x) - (k+1) C_{k+1}^\lambda(x); \end{aligned}$$

- Christoffelova rekurentna relacija (formula za integraciju)

$$2(k+\lambda) C_k^\lambda(x) = \frac{d}{dx} C_{k+1}^\lambda(x) - \frac{d}{dx} C_{k-1}^\lambda(x).$$

Ako potražimo rešenje diferencijalne jednačine (2.13.4) u obliku stepenog reda dobijamo razvoj za Gegenbauerov polinom

$$C_k^\lambda(x) = \sum_{i=0}^{\lfloor \frac{k}{2} \rfloor} \frac{(-1)^i (\lambda)_k}{i! (k-2i)!} (2x)^{k-2i}.$$

Primetimo da je

$$C_{2k}^\lambda(0) = \frac{(-1)^k (\lambda)_k}{k!} \quad i \quad C_{2k+1}^\lambda(0) = 0.$$

Takodje, važe jednakosti

$$k C_k^\lambda(x) = 2\lambda (xC_{k-1}^{\lambda+1}(x) - C_{k-2}^{\lambda+1}(x)),$$

$$(k+2\lambda) C_k^\lambda(x) = 2\lambda (C_k^{\lambda+1}(x) - xC_{k-1}^{\lambda+1}(x)),$$

$$(k+\lambda) C_k^\lambda(x) = \lambda (C_k^{\lambda+1}(x) - C_{k-2}^{\lambda+1}(x)),$$

$$\sum_{i=0}^k (2i+1+\lambda) C_{2i+1}^\lambda(x) = \lambda C_{2k+1}^{\lambda+1}(x),$$

$$\sum_{i=0}^k (2i+\lambda) C_{2i}^\lambda(x) = \lambda C_{2k}^{\lambda+1}(x).$$

Korišćenjem hipergeometrijske funkcije

$$(2.13.5) \quad F(a, b; c; z) = \sum_{i=0}^{+\infty} \frac{(a)_i (b)_i}{(c)_i} \frac{z^i}{i!},$$

moguće su sledeće reprezentacije Gegenbauerovih polinoma ([2, str. 177]):

$$C_{2k}^\lambda(x) = (-1)^k \frac{(\lambda)_k}{k!} F(-k, k+\lambda; \frac{1}{2}; x^2) = C_{2k}^\lambda(1) F(-k, k+\lambda; \lambda + \frac{1}{2}; 1-x^2),$$

$$C_{2k+1}^\lambda(x) = (-1)^k \frac{(\lambda)_{k+1}}{k!} 2x F(-k, k+\lambda+1; \frac{3}{2}; x^2) =$$

$$= C_{2k+1}^\lambda(1) x F(-k, k+\lambda+1; \lambda + \frac{1}{2}; 1-x^2).$$

Primetimo da u (2.13.5) za $a=-k$ ne postoji beskonačan razvoj, već se javljaju članovi različiti od nule samo za $i=0, 1, \dots, k$.

1.b. Legendreovi polinomi $P_k(x)$.

Ovde je $(a,b)=(-1,1)$, $p(x)=1$, $\alpha=\beta=0$, $A(x)=1-x^2$, $B(x)=-2x$. Legendreova diferencijalna jednačina ima oblik

$$(1-x^2)y'' - 2xy' + k(k+1)y = 0,$$

čije je jedno partikularno rešenje Legendreov polinom

$$P_k(x) = \frac{1}{2^k} \sum_{i=0}^{\lfloor \frac{k}{2} \rfloor} (-1)^i \binom{k}{i} \binom{2k-2i}{k} x^{k-2i},$$

za koji važi

$$P_k(-x) = P_k(x), \quad P_k(\pm 1) = (\pm 1)^k, \quad P_{2k}(0) = (-1)^k \frac{\left(\frac{1}{2}\right)_k}{k!}.$$

Primetimo da je $P_k(x) = P_k^{(0,0)}(x) = C_k^{1/2}(x)$.

Bonnetova rekurentna relacija glasi

$$P_{k+1}(x) = \frac{2k+1}{k+1} x P_k(x) - \frac{k}{k+1} P_{k-1}(x).$$

Prvih osam Legendreovih polinoma su:

$$P_0(x) = 1,$$

$$P_1(x) = x,$$

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

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

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

$$P_5(x) = \frac{1}{8}(63x^5 - 70x^3 + 15x),$$

$$P_6(x) = \frac{1}{16}(231x^6 - 315x^4 + 105x^2 - 5)$$

$$P_7(x) = \frac{1}{16}(429x^7 - 693x^5 + 315x^3 - 35x).$$

$$\text{Ovde je } a_k = \frac{(2k)!}{2^k (k!)^2}, \quad r_k = \frac{b_k}{a_k} = 0, \quad \|P_k\|^2 = \frac{2}{2k+1}.$$

Uzimajući $\alpha=\beta=0$ koeficijenti formula za diferenciranje i integraciju Jacobievih polinoma se svode na

$$u_k = -k, \quad v_k = 0, \quad w_k = -k,$$

$$\omega_k = -k(k+1), \quad \xi_k = \frac{k(k+1)}{2k+1}, \quad \eta_k = 0, \quad \zeta_k = -\frac{k(k+1)}{2k+1},$$

tako da za Legendreove polinome važe formule

$$(1-x^2)P'_k(x) = k(P_{k-1}(x) - xP_k(x)),$$

$$(2k+1)P_k(x) = P'_{k+1}(x) - P'_{k-1}(x).$$

1.c. Čebiševljevi polinomi $T_k(x)$.

Ovde je $(a,b)=(-1,1)$, $A(x)=1-x^2$, $\alpha=\beta=-\frac{1}{2}$ (tj. $\lambda=0$),
 $p(x)=\frac{1}{\sqrt{1-x^2}}$ i $B(x)=-x$. Odgovarajući polinomi su Čebiševljevi polinomi prve vrste $T_k(x)$. Saglasno uobičajenoj normalizaciji (2.9.9) imamo

$$(2.13.6) \quad T_k(x) = \frac{k!}{(\frac{1}{2})_k} P_k^{(-1/2, -1/2)}(x).$$

Odgovarajuća diferencijalna jednačina ima oblik

$$(2.13.7) \quad (1-x^2)y'' - xy' + k^2y = 0,$$

koja se sменом $x=\cos\theta$, $y(x)=z(\theta)$ ($\theta=\arccos x$) svodi na jednačinu $z'' + k^2 z = 0$. Kako je opšte rešenje ove jednačine $z(\theta) = C_1 \cos k\theta + C_2 \sin k\theta$, gde su C_1 i C_2 proizvodnje konstante za opšte rešenje Čebiševljeve jednačine (2.13.6) dobijamo

$$y = C_1 \cos(k \operatorname{arc} \cos x) + C_2 \sin(k \operatorname{arc} \cos x).$$

Ovakva reprezentacija rešenja važi za $-1 \leq x \leq 1$. Može se pokazati da je za $-1 \leq x \leq 1$, $\cos(k \operatorname{arc} \cos x) = T_k(x)$, gde je $T_k(x)$ dato pomoću (2.13.6), ili u razvijenom obliku

$$T_k(x) = \frac{k}{2} \sum_{i=0}^{\lfloor \frac{k}{2} \rfloor} \frac{(-1)^i (k-i-1)!}{i!(k-2i)!} (2x)^{k-2i} \quad (k=1, 2, \dots),$$

$$T_0(x) = 1.$$

Kako je $\operatorname{arc} \cos x = -i \operatorname{Log}(x + i\sqrt{1-x^2})$ imamo

$$\begin{aligned} T_k(x) &= \frac{1}{2} (e^{ik \operatorname{arc} \cos x} + e^{-ik \operatorname{arc} \cos x}) \\ &= \frac{1}{2} ((x + i\sqrt{1-x^2})^k + (x - i\sqrt{1-x^2})^k). \end{aligned}$$

Druge partikularno rešenje diferencijalne jednačine (2.13.7) označavamo sa $U_k(x) = \sin(k \operatorname{arc} \cos x)$ ($-1 \leq x \leq 1$).

Ako u jednakosti $\cos(k+1)\theta + \cos(k-1)\theta = 2\cos\theta\cos k\theta$ stavimo $\theta = \operatorname{arc} \cos x$ dobijamo rekurentnu relaciju

$$(2.13.8) \quad T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x).$$

Naravno, relaciju (2.13.8) možemo dobiti iz Bonnetove rekurentne relacije za Jacobieve polinome uzimajući $\alpha = \beta = -1/2$ i imajući u vidu (2.13.6). S obzirom da su koeficijenti u pomenutoj rekurentnoj relaciji za $P_k^{(-1/2, -1/2)}(x)$

$$\alpha_k = \frac{2k+1}{k+1}, \quad \beta_k = 0, \quad \gamma_k = \frac{4k^2-1}{4k(k+1)},$$

imamo

$$\frac{(\frac{1}{2})_{k+1}}{(k+1)!} T_{k+1}(x) = \frac{2k+1}{k+1} \times \frac{(\frac{1}{2})_k}{k!} T_k(x) - \frac{4k^2-1}{4k(k+1)} \cdot \frac{(\frac{1}{2})_{k-1}}{(k-1)!} T_{k-1}(x),$$

odakle sleduje (2.13.8). Slično, iz rekurentne relacije za Genbauerove polinome, uz korišćenje (2.13.3), sleduje (2.13.8). Do rekurentne relacije se može doći korišćenjem koeficijenata a_k i b_k i kvadrata norme (videti odeljak (2.2.10))

$$a_k = 2^{k-1}, \quad b_k = 0, \quad \|T_0\|^2 = \pi, \quad \|T_k\|^2 = \frac{\pi}{2} \quad (k \neq 0).$$

Naime, tada na osnovu (2.2.5) i (2.2.7) dobijamo $\alpha_k = 2$, $\beta_k = 0$, $\gamma_k = 1$, tj. rekurentnu relaciju (2.13.8).

Prvih osam Čebiševljevih polinoma su:

$$T_0(x) = 1,$$

$$T_1(x) = x,$$

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

$$T_3(x) = 4x^3 - 3x,$$

$$T_4(x) = 8x^4 - 8x^2 + 1,$$

$$T_5(x) = 16x^5 - 20x^3 + 5x,$$

$$T_6(x) = 32x^6 - 48x^4 + 18x^2 - 1,$$

$$T_7(x) = 64x^7 - 112x^5 + 56x^3 - 7x.$$

Na osnovu (2.11.2) i (2.11.7) dobijamo

$$u_k = -k, \quad v_k = 0, \quad w_k = -k,$$

$$\omega_k = -(k-1)(k+1), \quad \xi_k = \frac{1}{2}(k+1), \quad \eta_k = 0, \quad \zeta_k = -\frac{1}{2}(k-1),$$

tako da za Čebiševljeve polinome prve vrste važe formule

$$(1-x^2)T'_k(x) = k(T_{k-1}(x) - xT_k(x)),$$

$$T_k(x) = \frac{1}{2} \left(\frac{1}{k+1} T'_{k+1}(x) - \frac{1}{k-1} T'_{k-1}(x) \right).$$

Za Čebiševljeve polinome prve vrste važi

$$|T_k(x)| \leq 1 \quad (-1 \leq x \leq 1),$$

$$T_k(\pm 1) = (\pm 1)^k, \quad T_{2k}(0) = (-1)^k, \quad T_{2k+1}(0) = 0,$$

$$T'_k(\pm 1) = (\pm 1)^k k^2, \quad T_k^{(m)}(1) = \frac{k^2(k^2-1)\dots(k^2-(m-1)^2)}{(2m-1)!}.$$

Nule Čebiševljevih polinoma se mogu eksplicitno izraziti. Naime, iz $T_k(x) = \cos(k \arccos x) = 0$ sleduje

$$x_i^{(k)} = \cos \frac{2i-1}{2k} \pi \quad (i=1, \dots, k).$$

Čebiševljevi polinomi poseduju jedno važno ekstremalno svojstvo koje će biti razmatrano u posebnom odeljku.

1.d. Čebiševljevi polinomi druge vrste $S_k(x)$.

Ovde imamo $(a,b)=(-1,1)$, $A(x)=1-x^2$, $\alpha=\beta=\frac{1}{2}$ ($\lambda=1$), $p(x)=\sqrt{1-x^2}$ i $B(x)=-3x$. Čebiševljevi polinomi druge vrste su u stvari Gegenbauerovi polinomi za $\lambda=1$. Dakle, $S_k(x)=C_k^1(x)$, što se jednostavno pokazuje na osnovu (2.9.9). Imajući ovo u vidu, mi ćemo u ovom odeljku ukratko izneti rezultate za polinome $S_k(x)$, dobijajući ih direktno iz odgovarajućih za Gegenbauerove polinome. Posebno ćemo ukazati na vezu ovih polinoma sa polinomima prve vrste $T_k(x)$.

Diferencijalna jednačina čije je jedno partikularno rešenje $y=S_k(x)$ glasi

$$(1-x^2)y'' - 3xy' + k(k+2)y = 0.$$

Eksplicitan izraz za Čebiševljev polinom druge vrste je

$$S_k(x) = \sum_{i=0}^{\lfloor \frac{k}{2} \rfloor} \frac{(-1)^i (k-i)!}{i!(k-2i)!} (2x)^{k-2i},$$

a jedna standardna reprezentacija za $|x| \leq 1$

$$S_k(x) = \frac{\sin((k+1)\arccos x)}{\sqrt{1-x^2}}.$$

Polinomi S_k zadovoljavaju istu rekurentnu relaciju kao i polinomi T_k , tj.

$$S_{k+1}(x) = 2xS_k(x) - S_{k-1}(x).$$

Prvih osam polinoma su:

$$S_0(x) = 1,$$

$$S_1(x) = 2x,$$

$$S_2(x) = 4x^2 - 1,$$

$$S_3(x) = 8x^3 - 4x,$$

$$S_4(x) = 16x^4 - 12x^2 + 1,$$

$$S_5(x) = 32x^5 - 32x^3 + 6x,$$

$$S_6(x) = 64x^6 - 80x^4 + 24x^2 - 1,$$

$$S_7(x) = 128x^7 - 192x^5 + 80x^3 - 8x.$$

Jednostavno se pokazuje

$$S_k(x) = \frac{1}{k+1} T'_{k+1}(x) = \frac{1}{1-x^2} (xT_{k+1}(x) - T_{k+2}(x)),$$

$$T_k(x) = S_k(x) - xS_{k-1}(x),$$

$$(1-x^2)S'_k(x) = (k+1)S_{k-1}(x) - kxS_k(x),$$

$$2(k+1)S_k(x) = S'_{k+1}(x) - S'_{k-1}(x),$$

$$a_k = 2^k, \quad b_k = 0, \quad \|S_k\|^2 = \frac{\pi}{2}, \quad S_k(1) = k+1.$$

Primetimo da se polinom $S_k(x)$ može izraziti i u obliku

$$S_k(x) = \frac{U_{k+1}(x)}{\sqrt{1-x^2}},$$

gde je funkcija $U_{k+1}(x)$ definisana ranije.

2. Generalisani Laguerreovi polinomi $L_k^s(x)$.

Ovde je $(a,b)=(0,+\infty)$, $p(x)=x^s e^{-x}$ ($s > -1$), $A(x)=x$, $B(x)=s+1-x$.

Diferencijalna jednačina čije je jedno partikularno rešenje $y=L_k^s(x)$ ima oblik

$$xy'' + (1+s-x)y' + ky = 0.$$

Konstanta C_k , kvadrat norme $\|L_k^s\|^2$ i koeficijenti a_k i b_k ($r_k=b_k/a_k$) su dati pomoću (2.9.8) i (2.10.13). Na osnovu

(2.2.5) i (2.2.7) nalazimo

$$\alpha_k = \frac{a_{k+1}}{a_k} = -1, \quad \beta_k = \alpha_k(r_{k+1} - r_k) = 2k+s+1,$$

$$\gamma_k = \frac{\alpha_k}{\alpha_{k-1}} \left(\frac{\|L_k^s\|}{\|L_{k-1}^s\|} \right)^2 = k(k+s),$$

tako da je Bonnetova rekurentna relacija

$$L_{k+1}^s(x) = (2k+s+1-x)L_k^s(x) - k(k+s)L_{k-1}^s(x),$$

$$L_0^s(x) = 1, \quad L_1^s(x) = s+1-x.$$

Za $s=0$ dobijaju se standardni Laguerreovi polinomi $L_k(x)$. Navešćemo prvih šest polinoma:

$$L_0(x) = 1,$$

$$L_1(x) = 1-x,$$

$$L_2(x) = 2-4x+x^2,$$

$$L_3(x) = 6-18x+9x^2-x^3,$$

$$L_4(x) = 24-96x+72x^2-16x^3+x^4,$$

$$L_5(x) = 120-600x+600x^2-200x^3+25x^4-x^5.$$

Na osnovu (2.11.2) i (2.11.7) nalazimo redom

$$u_k = 0, \quad v_k = k, \quad w_k = k(k+s),$$

$$\omega_k = -(k+1), \quad \varepsilon_k = 0, \quad n_k = -(k+1), \quad \zeta_k = 1,$$

tako da važe formule

$$x \frac{d}{dx} L_k^s(x) = k(L_k^s(x) - (k+s)L_{k-1}^s(x)),$$

$$(k+1)L_k^s(x) = (k+1) \frac{d}{dx} L_k^s(x) - \frac{d}{dx} L_{k+1}^s(x).$$

Za generalisane Laguerreove polinome može se dokazati eksplicitna formula

$$L_k^s(x) = \sum_{i=0}^k (-1)^i \binom{k}{i} (s+i+1)_{k-i} x^i,$$

koja se za $s=0$ svodi na

$$L_k(x) = k! \sum_{i=0}^k (-1)^i \binom{k}{i} \frac{x^i}{i!}.$$

3. Hermiteovi polinomi $H_k(x)$.

Ovde imamo $(a,b)=(-\infty, +\infty)$, $p(x)=e^{-x^2}$, $A(x)=1$, $B(x)=-2x$.

Diferencijalna jednačina glasi

$$y'' - 2xy' + 2ky = 0.$$

Za Hermiteove polinome važi eksplicitna formula

$$H_k(x) = k! \sum_{i=0}^{\lfloor \frac{k}{2} \rfloor} \frac{(-1)^i}{i!(k-2i)!} (2x)^{k-2i}.$$

Rekurentne relacije su:

$$\begin{aligned} H_{k+1}(x) &= 2xH_k(x) - 2kH_{k-1}(x), \quad H_0(x)=1, \quad H_1(x)=2x, \\ H'_k(x) &= 2kH_{k-1}(x). \end{aligned}$$

Prvih osam Hermiteovih polinoma su:

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= 2x, \\ H_2(x) &= 4x^2 - 2, \\ H_3(x) &= 8x^3 - 12x, \\ H_4(x) &= 16x^4 - 48x^2 + 12, \\ H_5(x) &= 32x^5 - 160x^3 + 120x, \\ H_6(x) &= 64x^6 - 480x^4 + 720x^2 - 120, \\ H_7(x) &= 128x^7 - 1344x^5 + 3360x^3 - 1680x. \end{aligned}$$

Za Hermiteove polinome može se dokazati:

$$H_k(-x) = (-1)^k H_k(x), \quad H_{2k}(0) = (-1)^k \frac{(2k)!}{k!}, \quad H_{2k+1}(0) = 0;$$

$$a_k = 2^k, \quad b_k = 0, \quad \|H_k\|^2 = 2^k k! \sqrt{\pi};$$

$$\int_0^x e^{-t^2} H_k(t) dt = H_{k-1}(0) - e^{-x^2} H_{k-1}(x),$$

$$\int_{-\infty}^{+\infty} e^{-t^2} H_{2k}(xt) dt = \sqrt{\pi} \frac{(2k)!}{k!} (x^2 - 1)^k,$$

$$\int_{-\infty}^{+\infty} t e^{-t^2} t^k H_{2k+1}(xt) dt = \sqrt{\pi} \frac{(2k+1)!}{k!} x (x^2 - 1)^k,$$

$$\int_{-\infty}^{+\infty} e^{-t^2} t^k H_k(xt) dt = \sqrt{\pi} k! P_k(x),$$

gde je P_k Legendreov polinom. Takodje se može dokazati i veza Hermiteovih i generalisanih Laguerreovih polinoma (videti primedbu 2.8.1)

$$H_{2k}(x) = (-1)^k 2^{2k} L_k^{-1/2}(x^2), \quad H_{2k+1}(x) = (-1)^k 2^{2k+1} x L_k^{1/2}(x^2).$$

S obzirom na veliki značaj tročlane rekurentne relacije za monične ortogonalne polinome (videti teoremu 2.2.5)

$$Q_{k+1}(x) = (x - \beta_k) Q_k(x) - \gamma_k Q_{k-1}(x),$$

posebno kod konstrukcije Gauss-Christoffelovih kvadraturnih formula, na kraju ovog odeljka dajemo tabelu koeficijenata β_k i γ_k za klasične ortogonalne polinome.

U tabeli 2.13.1, normalizacioni koeficijent a_k za Jacobićeve polinome dat je pomoću (2.10.6).

Korišćenjem, na primer, rekurentne relacije za monične Gegenbauerove polinome \hat{C}_k^λ dobijamo niz

Tabela 2.13.1

Monični klasični ortogonalni polinomi	β_k	γ_k
Jacobievi polinomi $P_k^{(\alpha, \beta)}(x)/a_k$	$\frac{\beta^2 - \alpha^2}{(2k+\alpha+\beta)(2k+\alpha+\beta+2)}$	$\frac{4k(k+\alpha)(k+\beta)(k+\alpha+\beta)}{(2k+\alpha+\beta)^2((2k+\alpha+\beta)^2-1)}$
Gegenbauerovi polinomi $\frac{k!}{2^k (\lambda)_k} C_k^\lambda(x)$	0	$\frac{k(k+2\lambda-1)}{4(k+\lambda)(k+\lambda-1)}$
Legendreovi polinomi $\frac{2^k (k!)^2}{(2k)!} P_k(x)$	0	$\frac{k^2}{4k^2 - 1}$
Čebiševljevi polinomi prve i druge vrste $T_k(x)/2^{k-1}$ i $S_k(x)/2^k$	0	$\frac{1}{4}$ (izuzetak: $\gamma_1 = 1/2$ za polinome prve vrste.)
Generalisani Laguerre-ovi polinomi $(-1)^k L_k^s(x)$	$2k+s+1$	$k(k+s)$
Laguerreovi polinomi $(-1)^k L_k(x)$	$2k+1$	k^2
Hermiteovi polinomi $H_k(x)/2^k$	0	$\frac{k}{2}$

$$\hat{C}_0^\lambda(x) = 1,$$

$$\hat{C}_1^\lambda(x) = x,$$

$$\hat{C}_2^\lambda(x) = x^2 - \frac{1}{2(\lambda+1)},$$

$$\hat{C}_3^\lambda(x) = x^3 - \frac{3}{2(\lambda+2)}x,$$

$$\hat{C}_4^\lambda(x) = x^4 - \frac{3}{\lambda+3}x^2 + \frac{3}{4(\lambda+2)(\lambda+3)},$$

$$\hat{C}_5^\lambda(x) = x^5 - \frac{5}{\lambda+4} x^3 + \frac{15}{4(\lambda+3)(\lambda+4)} x, \text{ itd.}$$

Za $\lambda=1/2$ dobijamo monične Legendreove polinome, a za $\lambda=0$ monične Čebiševljeve polinome.

2.2.14. Ekstremalna svojstva Čebiševljevih polinoma

Čebiševljevi polinomi poseduju veoma važna ekstremalna svojstva, pogotovu kada se radi o Čebiševljevim polinomima prve vrste. Pre nego što predjemo na izučavanje ovih ekstremalnih svojstava ukazaćemo na još neke standardne osobine Čebiševljevih polinoma prve vrste.

Kao što je ranije rečeno polinom T_k je ograničen na $[-1,1]$, tj.

$$|T_k(x)| \leq 1 \quad (x \in [-1,1]).$$

Na Sl. 2.14.1 grafički su prikazani polinomi $T_0(x)$,

$T_1(x), \dots, T_5(x)$.

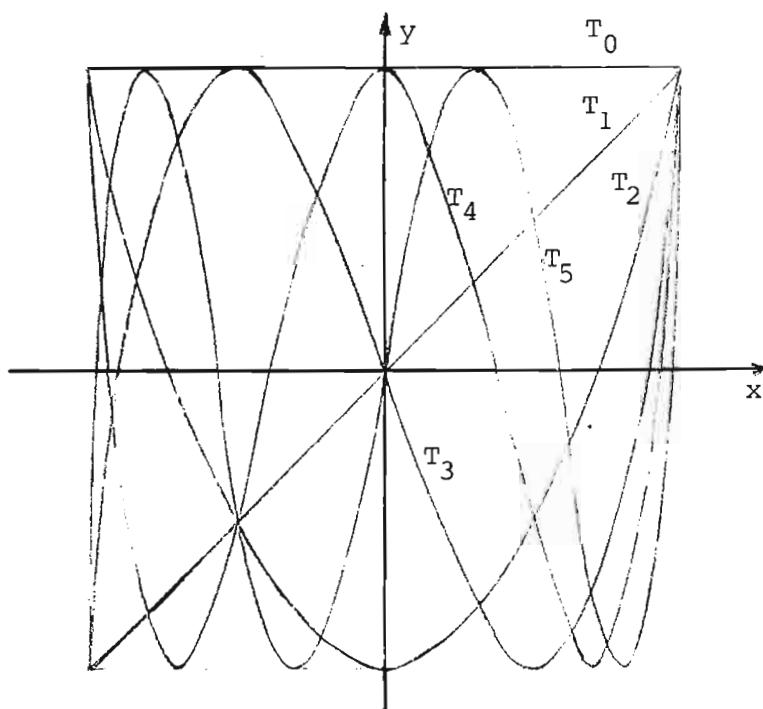
Primetimo da polinom $T_n(x)$ na $[-1,1]$ dos- tiže vrednost 1 ili -1 u $n+1$ tačaka. Označimo ove tačke sa x_k ($k=0,1,\dots,n$).

Iz $T_n(x)=\pm 1$, tj. $\cos(n \arccos x) = \pm 1$, jednostavno na- lazimo

$$x_k = \cos \frac{k\pi}{n}$$

$$(k=0,1,\dots,n).$$

Ako izvršimo prenumeraciju



Sl. 2.14.1

indeksa u smislu da indeks k zamenimo sa n-k dobijamo

$$(2.14.1) \quad x_k = -\cos \frac{k\pi}{n} \quad (k=0, 1, \dots, n),$$

tako da je sada

$$-1 = x_0 < x_1 < \dots < x_n = 1.$$

Sledeći rezultat koji se odnosi na osnovno ekstremalno svojstvo Čebiševljevih polinoma prve vrste formulisao je i dokazao poznati ruski matematičar P.L. Čebišev.

Teorema 2.14.1. Neka je Q_n proizvoljan moničan polinom stepena n ($Q_n(x) = x^n + \dots$). Tada je

$$(2.14.2) \quad \max_{x \in [-1, 1]} |Q_n(x)| \geq \max_{x \in [-1, 1]} \left| \frac{1}{2^{n-1}} T_n(x) \right| = \frac{1}{2^{n-1}},$$

pri čemu se znak jednakosti postiže ako i samo ako je

$$Q_n(x) = \frac{1}{2^{n-1}} T_n(x).$$

Dokaz. Kako je $T_n(x) = 2^{n-1}x^n + \dots$, polinom $R(x) = \frac{1}{2^{n-1}}T_n(x) - Q_n(x)$ je stepena ne višeg od n-1.

Da bismo dokazali (2.14.2) pretpostavimo suprotno, tj. da je

$$(2.14.3) \quad Q = \max_{x \in [-1, 1]} |Q_n(x)| < \frac{1}{2^{n-1}}.$$

Neka $x \in [-1, 1]$. Tada na osnovu (2.14.3) imamo

$$-\frac{1}{2^{n-1}} - Q_n(x) \leq -\frac{1}{2^{n-1}} + Q < 0$$

i

$$\frac{1}{2^{n-1}} - Q_n(x) \geq \frac{1}{2^{n-1}} - Q > 0,$$

odakle zaključujemo da polinom $R(x)$ naizmenično ima pozitivne i negativne vrednosti u tačkama x_k ($k=0, 1, \dots, n$), datih pomoću (2.14.1). Ovo znači da polinom $R(x)$ ima najmanje n nula, što

je u kontradikciji sa činjenicom da je $R(x)$ polinom ne višeg stepena od $n-1$.

Ovim je teorema dokazana.

Ekstremalno svojstvo (2.14.2) kazuje da od svih moničnih polinoma n -tog stepena polinom $\frac{1}{2^{n-1}} T_n(x)$ najmanje odstupa od nule. Odstupanje je ovde okarakterisano uobičajenom normom u prostoru neprekidnih funkcija

$$\|f\| = \max_{x \in [-1, 1]} |f(x)|.$$

Interesantno je ukazati još na jedno svojstvo Čebiševljevih polinoma prve vrste ([31, str. 80]):

Teorema 2.14.2. Neka je $P(x)$ polinom ne višeg stepena od n i takav da je $\|P\| = 1$. Tada je za svako $x \notin [-1, 1]$

$$|P(x)| \leq |T_n(x)|.$$

Drugim rečima, ako je $x = x_0$ van segmenta ortogonalnosti, tada od svih polinoma stepena ne višeg od n čije je maksimalno odstupanje na $[-1, 1]$ jednako jedinici, Čebiševljev polinom u toj tački x_0 najviše odstupa od nule.

Rezultat sličan teoremi 2.14.1 može se formulisati za polinome i u odnosu na normu

$$(2.14.4) \quad \|f\|_1 = \int_{-1}^1 |f(x)| dx,$$

ili, pak u odnosu na normu u $L^2(a, b)$

$$\|f\|_2 = \left(\int_a^b p(x) f(x)^2 dx \right)^{1/2},$$

gde je p data težinska funkcija.

Ekstremalno svojstvo Čebiševljevih polinoma druge vrste u odnosu na normu (2.14.4) otkrili su ruski matematičari A.N. Korkin i E.I. Zolotarev. U vezi sa ovim rezultatom dokazaćemo prethodno jedan pomoći rezultat:

Lema 2.14.1. Ako je s_n Čebiševljev polinom druge vrste tada važi

$$I_{n,k} = \int_{-1}^1 x^k sgn s_n(x) dx = \begin{cases} 0 & (0 \leq k \leq n-1), \\ \frac{1}{2^{n-1}} & (k=n). \end{cases}$$

Dokaz. Uvodjenjem smene $x=\cos\theta$, integral $I_{n,k}$ se transformiše na oblik

$$I_{n,k} = \int_0^\pi \cos^k \theta (\operatorname{sgn} \sin(n+1)\theta) \sin \theta d\theta.$$

Za dokaz tvrdjenja koristićemo:

1° Fourierov razvoj

$$f(t) = \operatorname{sgn} \sin t = \frac{4}{\pi} \sum_{m=0}^{+\infty} \frac{\sin(2m+1)t}{2m+1}$$

uzimajući $t=(n+1)\theta$, i rezultat (videti, na primer, [20 ,str. 169-171]):

2° Za svako $p,q \in \mathbb{N}$

$$J_{p,q} = \int_0^\pi \cos^p \theta \cos^q \theta d\theta = \begin{cases} \frac{\pi}{2^p} \left(\frac{1}{2} \binom{p}{p+q} \right) & (p+q(p \geq q) \text{ parno}), \\ 0 & (\text{u ostalim slučajevima}). \end{cases}$$

Kako je

$$f((n+1)\theta) \sin \theta = \frac{2}{\pi} \sum_{m=0}^{+\infty} \frac{1}{2m+1} (\cos(N-1)\theta - \cos(N+1)\theta),$$

gde je $N=N(m,n)=(2m+1)(n+1)$, imamo

$$(2.14.5) \quad I_{n,k} = \frac{2}{\pi} \sum_{m=0}^{+\infty} \frac{1}{2m+1} (J_{k,N-1} - J_{k,N+1}).$$

Za nas je interesantan samo slučaj kada je $0 \leq k \leq n$.

Kako je $\min_m N(m,n)=N(0,n)=n+1$, zaključujemo, na osnovu rezultata 2°, da su svi integrali koji se pojavljuju na desnoj strani u (2.14.5) jednaki nuli, sem integrala $J_{k,N-1}$ za $m=0$ i $k=n$.

Prema tome, $I_{n,k}=0$ za $0 \leq k \leq n-1$, a za $k=n$

$$I_{n,n} = \frac{2}{\pi} J_{n,N(0,n)-1} = \frac{2}{\pi} J_{n,n} = \frac{2}{\pi} \frac{\pi}{2^n} \binom{n}{n} = \frac{1}{2^{n-1}}.$$

Teorema 2.14.3. Neka je Q_n proizvoljan moničan polinom stepena n ($Q_n(x) = x^n + \dots$). Tada je

$$\|Q_n\|_1 \geq \left\| \frac{1}{2^n} S_n \right\|_1 = \frac{1}{2^{n-1}},$$

pri čemu se znak jednakosti postiže ako i samo ako je

$$Q_n(x) = \frac{1}{2^n} S_n(x).$$

Dokaz. Definišimo funkcionalu J pomoću

$$J(Q_n) = \|Q_n\|_1 = \int_{-1}^1 |Q_n(x)| dx.$$

Kako je polinom Q_n moničan, na osnovu dokazane leme zaključujemo da je

$$\int_{-1}^1 Q_n(x) \operatorname{sgn} S_n(x) dx = \frac{1}{2^{n-1}},$$

odakle sleduje nejednakost

$$J(Q_n) = \int_{-1}^1 |Q_n(x)| dx \geq \frac{1}{2^{n-1}}.$$

Kako je polinom $\frac{1}{2^n} S_n(x)$ moničan, takodje, imamo

$$\int_{-1}^1 \left| \frac{1}{2^n} S_n(x) \right| dx = \int_{-1}^1 \frac{1}{2^n} S_n(x) \operatorname{sgn} S_n(x) dx = \frac{1}{2^{n-1}}.$$

Prema tome, zaključujemo da polinom $\frac{1}{2^n} S_n(x)$ minimizira funkcionalu $J(Q_n) = \|Q_n\|_1$. Potrebno je još dokazati da je ovaj polinom jedini koji minimizira funkcionalu $J(Q_n)$. Ovo se dokazuje jednostavno pretpostavljujući da postoji još jedan moničan polinom $R_n(x)$ stepena n , za koji je, takodje, $J(R_n) = \frac{1}{2^{n-1}}$. Tada dobijamo $R_n(x) = \frac{1}{2^n} S_n(x)$.

2.2.15. Diskretni ortogonalni polinomi

Neka je $d\lambda(x)$ nenegativna mera (distribucija) na realnoj pravoj R , sa kompaktnim ili beskonačnim nosačem*, za koju svi momenti

$$(2.15.1) \quad C_k = \int_R x^k d\lambda(x) \quad (k=0,1,\dots)$$

postoje, konačni su i $C_0 > 0$. Skalarni proizvod funkcija f i g definišimo pomoću

$$(2.15.2) \quad (f,g) = \int_R f(x)g(x)d\lambda(x).$$

U specijalnom slučaju ako distribucija $d\lambda(x)$ ima oblik $d\lambda(x)=p(x)dx$, gde je p data težinska funkcija u smislu definicije 2.1.1, momenti (2.15.1) svode na momente (2.1.2), a skalarni proizvod (2.15.2) na (2.1.3).

Ortogonalni polinomi mogu se razviti i u odnosu na distribuciju $d\lambda(x)$, tj. u odnosu na skalarni proizvod (2.15.2), pri čemu ranije izloženi rezultati ostaju u važnosti i za ovaj opštiji slučaj.

U ovom odeljku razmotrićemo slučaj kada je $\lambda(x)$ funkcija ograničene varijacije sa tačno N skokova u tačkama x_0, x_1, \dots, x_{N-1} . Tačnije rečeno, neka je $\lambda(x)$ stepenasta neopadajuća funkcija koja u tačkama x_0, x_1, \dots, x_{N-1} ima redom skokove dužine p_0, p_1, \dots, p_{N-1} . Skalarni proizvod (2.15.2), kao Stieltjesov integral, tada postaje

$$(2.15.3) \quad (f,g) = [f,g]_N = \sum_{i=0}^{N-1} p_i f(x_i)g(x_i).$$

Takvu mjeru ćemo označavati sa $d\lambda_N(x)$ i nazivati diskretnom merom, a odgovarajuće polinome koji su ortogonalni u odnosu na (2.15.3) nazivaćemo diskretnim ortogonalnim polinomima. Skokove p_k nazivamo težinama. Često se uzima da je zbir težina

* Za definiciju nosača funkcije videti, na primer, [18, str. 191].

jednak jedinici, tj. $\sum_{i=0}^{N-1} p_i = 1$.

Norma koja izvire iz skalarnog proizvoda (2.15.3) je

$$(2.15.4) \quad \|f\| = \left(\sum_{i=0}^{N-1} p_i f(x_i)^2 \right)^{1/2}.$$

Primetimo da je za proizvoljnu funkciju koja je jednaka nuli u tačkama x_0, x_1, \dots, x_{N-1} , norma (2.15.4) jednaka nuli, što znači da ćemo dve funkcije f i g smatrati "identičnim" ako je $f(x_i) = g(x_i)$ ($i=0, 1, \dots, N-1$). Dakle, u ovom slučaju koristimo informacije o vrednosti funkcije samo sa diskretnog skupa tačaka $\{x_0, x_1, \dots, x_{N-1}\}$. Jasno je da ćemo ovde imati samo N polinoma koji su međusobno ortogonalni, tj. indeksni skup (videti odeljak 2.1.4) biće $I = \{0, 1, \dots, N-1\}$.

Sa $\{Q_k^{(N)}\}_{k \in I}$ označićemo ovaj skup diskretnih polinoma, ortogonalnih u odnosu na skalarni proizvod (2.15.3), pri čemu ćemo prepostaviti da su polinomi monični. Tada, saglasno prethodnom, ovi polinomi zadovoljavaju tročlanu rekurentnu relaciju

$$(2.15.5) \quad Q_{k+1}^{(N)}(x) = (x - \beta_k^{(N)}) Q_k^{(N)}(x) - \gamma_k^{(N)} Q_{k-1}^{(N)}(x),$$

$$Q_0^{(N)}(x) = 1, \quad Q_{-1}^{(N)}(x) = 0,$$

gde su $\beta_k^{(N)}$ i $\gamma_k^{(N)}$ konstante i $\gamma_k^{(N)} > 0$.

Koeficijenti $\beta_k^{(N)}$ i $\gamma_k^{(N)}$ mogu biti odredjeni, na primer, Čebiševljevim algoritmom (videti odeljak 2.2.6) ili Stieltjesovom procedurom (odeljak 2.2.7). U prvom slučaju za određivanje prvih n koeficijenata $\beta_k^{(N)}$ i $\gamma_k^{(N)}$ ($k=0, 1, \dots, n-1$) potrebno je poznavati prvih $2n$ momenata

$$c_k = \int_R x^k d\lambda_N(x) = \sum_{i=0}^{N-1} p_i x_i^k \quad (k=0, 1, \dots, 2n-1).$$

U drugom slučaju, tj. primenom Stieltjesove procedure, imamo

$$(2.15.6) \quad \beta_k^{(N)} = \frac{[\mathbf{x}Q_k^{(N)}, Q_k^{(N)}]_N}{[Q_k^{(N)}, Q_k^{(N)}]_N} = \frac{\sum_{i=0}^{N-1} p_i x_i Q_k^{(N)}(x_i)^2}{\sum_{i=0}^{N-1} p_i Q_k^{(N)}(x_i)^2},$$

$$(2.15.7) \quad \gamma_k^{(N)} = \frac{[Q_k^{(N)}, Q_k^{(N)}]_N}{[Q_{k-1}^{(N)}, Q_{k-1}^{(N)}]_N} = \frac{\sum_{i=0}^{N-1} p_i Q_k^{(N)}(x_i)^2}{\sum_{i=0}^{N-1} p_i Q_{k-1}^{(N)}(x_i)^2},$$

za $k=0, 1, \dots, N-2$, pri čemu uzimamo

$$(2.15.8) \quad \gamma_0^{(N)} = \sum_{i=0}^{N-1} p_i.$$

Posebno je interesantan slučaj kada su tačke x_i ekvidistantne, tj. kada je $x_i - x_{i-1} = h = \text{const}$. Tada možemo staviti $x_i = a + ih$ ($i=0, 1, \dots, N-1$), gde je $a = x_0$. Uvodjenjem smene $x = a + th$, tačke x_i se preslikavaju na tačke $t_i = i$ ($i=0, 1, \dots, N-1$). Prema tome, ne umanjujući opštost, možemo odmah razmatrati slučaj kada su date tačke $x_i = i$ ($i=0, 1, \dots, N-1$). U tom slučaju za skalarni proizvod (2.15.3) imamo

$$[f, g]_N = \sum_{i=0}^{N-1} p_i f(i) g(i).$$

Ako su težine p_i takve da ispunjavaju uslov $p_i = p_{N-1-i}$ za svako i , tada odgovarajući diskretni polinomi $Q_k^{(N)}(x)$ ispunjavaju uslov

$$(2.15.9) \quad Q_k^{(N)}(N-1-x) = (-1)^k Q_k^{(N)}(x),$$

odakle možemo zaključiti da je $Q_{2m+1}^{(N)}(\frac{N-1}{2}) = 0$.

U slučaju kada su sve težine jednake, na primer $p_i = 1$, ili kako se češće uzima $p_i = \frac{1}{N}$ (zbog uslova $\sum_{i=0}^{N-1} p_i = 1$), odgovarajući polinomi se nazivaju Čebiševljevi diskretni polinomi.

Kako je, s obzirom na (2.15.9),

$$\sum_{i=0}^{N-1} i Q_k^{(N)}(i)^2 = \sum_{i=0}^{N-1} (N-1-i) Q_k^{(N)}(N-1-i)^2 \\ = \sum_{i=0}^{N-1} (N-1-i) Q_k^{(N)}(i)^2,$$

na osnovu (2.15.6) zaključujemo da je koeficijent $\beta_k^{(N)}$ u reku-rentnoj relaciji (2.15.5) jednak

$$\beta_k^{(N)} = \frac{\sum_{i=0}^{N-1} i Q_k^{(N)}(i)^2}{\sum_{i=0}^{N-1} Q_k^{(N)}(i)^2} = \frac{N-1}{2}.$$

Na osnovu (2.15.8) i (2.15.7) imamo

$$\gamma_0^{(N)} = 1, \quad \gamma_1^{(N)} = \frac{N^2 - 1}{12}, \quad \gamma_2^{(N)} = \frac{N^2 - 4}{15}.$$

Dalja primena formula (2.15.7) i (2.15.5) u cilju odredjivanja $\gamma_k^{(N)}$ postaje komplikovana za proizvoljno N. Programske se, međutim, za fiksirano N jednostavno određuju koefficijenti $\gamma_k^{(N)}$.

Primer 2.15.1. Neka je $N=5$ i $p_i = \frac{1}{5}$ ($i=0, 1, \dots, 4$). Odredićemo skup diskretnih ortogonalnih polinoma $\{Q_k^{(5)}\}_{k=0, 1, \dots, 4}$. Imamo $\beta_0^{(5)} = 2$ i

$$Q_0^{(5)}(x) = 1, \quad Q_1^{(5)}(x) = x - \beta_1^{(5)} = x - 2.$$

Kako je $\gamma_1^{(5)} = 2$ nalazimo

$$Q_2^{(5)}(x) = (x - \beta_2^{(5)}) Q_1^{(5)}(x) - \gamma_1^{(5)} Q_0^{(5)}(x) = x^2 - 4x + 2.$$

Nadalje, $\gamma_2^{(5)} = \frac{21}{15} = \frac{7}{5}$ i

$$Q_3^{(5)}(x) = (x - \beta_3^{(5)}) Q_2^{(5)}(x) - \gamma_2^{(5)} Q_1^{(5)}(x) = (x-2)(x^2 - 4x + 2) - \frac{7}{5}(x-2),$$

tj.

$$Q_3^{(5)}(x) = x^3 - 6x^2 + \frac{43}{5}x - \frac{6}{5}.$$

Najzad, možemo naći

$$\gamma_3^{(5)} = \frac{\frac{1}{5} \sum_{i=0}^4 Q_3^{(5)}(i)^2}{\frac{1}{5} \sum_{i=0}^4 Q_2^{(5)}(i)^2} = \frac{2.88}{2.8} = \frac{36}{35}$$

i

$$Q_4^{(5)}(x) = (x-2)(x^3 - 6x^2 + \frac{43}{5}x - \frac{6}{5}) - \frac{36}{35}(x^2 - 4x + 2),$$

tj.

$$Q_4^{(5)}(x) = x^4 - 8x^3 + \frac{137}{7}x^2 - \frac{100}{7}x + \frac{12}{35}.$$

Ovim smo konstruisali sve članove ortogonalnog niza.

Ukoliko bismo nastavili sa ovim postupkom dobili bismo

$\gamma_4^{(5)} = \frac{4}{7}$ i $Q_5^{(5)}(x) = x(x-1)(x-2)(x-3)(x-4)$, za koji je evidentno $\|Q_5^{(5)}\| = 0$, što znači da $Q_5^{(5)}(x)$ kao "nula-element" (u smislu uvedene norme) ne pripada skupu diskretnih ortogonalnih polinoma.

U opštem slučaju za proizvoljno N može se pokazati da je

$$\gamma_0^{(N)} = 1, \quad \gamma_k^{(N)} = \frac{k^2(N^2-k^2)}{4(4k^2-1)} \quad (k=1, 2, \dots, N-2),$$

tako da rekurentna relacija za diskrete monične Čebiševljeve polinome glasi

$$Q_{k+1}^{(N)}(x) = (x - \frac{N-1}{2})Q_k^{(N)}(x) - \gamma_k^{(N)}Q_{k-1}^{(N)}(x) \quad (k=0, 1, \dots, N-2),$$

pri čemu je $Q_{-1}^{(N)}(x)=0$ i $Q_0^{(N)}(x)=1$.

Za kvadrat norme polinoma $Q_k^{(N)}(x)$ jednostavno dobijamo

$$\|Q_k^{(N)}\|^2 = \gamma_0^{(N)} \gamma_1^{(N)} \cdots \gamma_k^{(N)} = \prod_{i=1}^k \frac{i^2(N^2-i^2)}{4(4i^2-1)},$$

tj.

$$\|Q_k^{(N)}\|^2 = \frac{(k!)^4 (N^2-1^2)(N^2-2^2)\cdots(N^2-k^2)}{(2k+1)((2k)!)^2}.$$

Diskretni ortogonalni polinomi nalaze primenu kod tzv. diskretnih srednje-kvadratnih aproksimacija.

2.2.16. Polinomi ortogonalni na polukrugu

Neka je \mathcal{P} skup svih algebarskih polinoma. G. Szegö ([32]) je definisao i proučavao polinome ortogonalne na jediničnom krugu u odnosu na skalarni proizvod

$$(f, g) = \int_{-\pi}^{\pi} f(e^{i\theta}) \overline{g(e^{i\theta})} d\sigma(\theta), \quad d\sigma(\theta) \geq 0 \quad (f, g \in \mathcal{P}).$$

Ovako definisani skalarni proizvod je pozitivno-definitan jer je

$$(f, f) = \int_{-\pi}^{\pi} |f(e^{i\theta})|^2 d\sigma(\theta) > 0.$$

U ovom odeljku izložićemo jedan drugi tip ortogonalnosti, tzv. ortogonalnost na polukrugu, gde je skalarni proizvod definisan pomoću

$$(2.16.1) \quad (f, g) = \int_0^{\pi} f(e^{i\theta}) g(e^{i\theta}) d\sigma(\theta), \quad d\sigma(\theta) \geq 0,$$

pri čemu ćemo se zadržati samo na najprostijem slučaju kada je $d\sigma(\theta) = d\theta$ (konstantna težinska funkcija).

Skalarni proizvod (2.16.1) se može predstaviti i u obliku

$$(2.16.2) \quad (f, g) = \int_{\Gamma} (iz)^{-1} f(z) g(z) dz,$$

gde je Γ polukrug $\{z \mid z = e^{i\theta}, 0 \leq \theta \leq \pi\}$. Primetimo da drugi faktor u (2.16.1), tj. (2.16.2), nije konjugovan, tako da skalarni proizvod nije pozitivno-definitan. Međutim, ortogonalni polinomi u ovom slučaju ne samo da egzistiraju jedinstveno, već poseduju i niz interesantnih osobina. Ove polinome su definisali i izučavali W. Gautschi i G.V. Milovanović ([12], [13]).

Pomoću skalarnog proizvoda (2.16.1) definišimo moment-funcionelu

$$(2.16.3) \quad \langle z^k \rangle = C_k, \quad C_k = (1, z^k) = \int_0^{\pi} e^{ik\theta} d\theta = \begin{cases} \pi, & k=0, \\ 2i/k, & k \text{ neparno,} \\ 0, & k \neq 0 \text{ parno.} \end{cases}$$

Na osnovu teoreme 2.1.3, ortogonalni polinomi u odnosu na ovu funkcionalnu egzistiraju jedinstveno ako je ona kvazi-definitna, tj. ako je $\Delta_k \neq 0$ za svako $k \geq 1$, gde je moment-determinanta Δ_k definisana pomoću (2.1.5). Direktnim izračunavanjem determinante Δ_k nalazimo da je $\Delta_k > 0$ za svako $k \geq 1$ (videti [12]), što znači da ortogonalni niz jedinstveno egzistira. U daljem tekstu navešćemo najvažnije osobine ovih polinoma.

Teorema 2.16.1. Monični kompleksni polinomi $\{\pi_k\}_{k \in N_0}$ ortogonalni u odnosu na skalarni proizvod (2.16.2) zadovoljavaju tročlanu rekurentnu relaciju

$$\begin{aligned}\pi_{k+1}(z) &= (z - i\alpha_k)\pi_k(z) - \beta_k\pi_{k-1}(z) \quad (k = 0, 1, \dots), \\ \pi_{-1}(z) &= 0, \quad \pi_0(z) = 1,\end{aligned}$$

gde su

$$\alpha_0 = \theta_0, \quad \alpha_k = \theta_k - \theta_{k-1}, \quad \beta_k = \theta_{k-1}^2 \quad (k=1, 2, \dots),$$

a θ_k dato pomoću

$$\theta_k = \frac{2}{2k+1} \left(\frac{\Gamma((k+2)/2)}{\Gamma((k+1)/2)} \right)^2 \quad (k = 0, 1, \dots).$$

Teorema 2.16.2. Neka je $\{\hat{P}_k\}$ niz moničnih Legendreovih polinoma. Tada važi

$$\pi_n(z) = \hat{P}_n(z) - i\theta_{n-1}\hat{P}_{n-1}(z) \quad (n \geq 1),$$

gde je θ_k dato u prethodnoj teoremi.

Teorema 2.16.3. Ako je $\zeta (\in C)$ nula polinoma π_n , tada je $i - \bar{\zeta}$ takođe njegova nula.

Teorema 2.16.4. Sve nule polinoma π_n su proste i nalaze se u polu disku $D_+ = \{z \mid |z| < 1 \wedge \operatorname{Im} z > 0\}$.

Teorema 2.16.5. Polinom $\pi_n(z)$ je jedno partikularno rešenje linearne diferencijalne jednačine drugog reda

$$A(z)y'' + B(z)y' + C(z)y = 0,$$

gde su

$$A(z) = (1-z^2) \left[n^2 - (2n-1)^2 \theta_{n-1}^2 - 2n(2n-1)zi\theta_{n-1} \right],$$

$$B(z) = 2 \left[n(2n-1)(z+1)i\theta_{n-1} - (n^2 - (2n-1)^2 \theta_{n-1}^2)z \right],$$

$$C(z) = n \left[(n+1)n^2 - (n-1)(2n-1)^2 \theta_{n-1}^2 - 2(2n-1)n^2 z i\theta_{n-1} \right]$$

i θ_k kao u teoremi 2.16.1.

2.3. OPERATORI

U ovom poglavlju se daju osnovni elementi teorije operatora. Kompletnija teorija se može naći u [1], [5], [15], [17].

2.3.1. Linearni operatori

Neka su X i Y Banachovi prostori. Pod operatorom* $T:X \rightarrow Y$ podrazumeva se preslikavanje

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

Definicija 3.1.1. Operator $T:X \rightarrow Y$ je homogen ako je

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

Definicija 3.1.2. Operator $T:X \rightarrow Y$ je aditivan ako je

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

Definicija 3.1.3. Operator $T:X \rightarrow Y$ je linearan ako je istovremeno homogen i aditivan, tj.

$$T(c_1 u_1 + c_2 u_2) = c_1 Tu_1 + c_2 Tu_2 \quad (\forall u_1, u_2 \in X; \forall c_1, c_2 \in K).$$

Definicija 3.1.4. Operator $T:X \rightarrow Y$ je neprekidan ako za svaki niz $\{u_n\}_{n \in \mathbb{N}}$ iz X važi

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

Definicija 3.1.5. Linearni operator $T:X \rightarrow Y$ je ograničen ako postoji nenegativan broj M takav da je

*) Operator $T:X \rightarrow R$ nazivamo funkcionalom.

$$(3.1.1) \quad \|Tu\| \leq M \|u\| \quad (\forall u \in X).$$

Infimum brojeva M za koje važi (3.1.1) označava se sa $\|T\|$ i naziva norma operatora T .

Teorema 3.1.1. Linearni operator $T:X \rightarrow Y$ je ograničen ako i samo ako je neprekidan.

Dokaz. Ako je operator T ograničen važi

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

odakle pri $u_n \rightarrow u$ sledi 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

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

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

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

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

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

Definicija 3.1.6. 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 linearни 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 3.1.7. 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

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

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

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

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

Za operatore T^n i T^m ($n, m \in \mathbb{N}_0$) važi jednakost

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

Ako je T ograničen operator tada je

$$\|T^n\| \leq \|T\|^n \quad (n \in \mathbb{N}_0).$$

Definicija 3.1.8. Neka je $T: X \rightarrow Y$ linearan operator. Ako postoje preslikavanje T^{-1} takvo da je

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

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

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

što se (kod linearnih operatora) svodi na

$$Tu = \theta \Rightarrow u = \theta.$$

Teorema 3.1.2. Ako linearan operator $T:X \rightarrow Y$ 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 3.1.1. 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.

$$(3.1.2) \quad g'(x) = \alpha'(x) - \int_a^x e^{-u} f(t) dt + \int_x^b e^u f(t) dt.$$

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 (3.1.2) sleduje

$$(3.1.3) \quad g'(a) = \alpha'(a)g(a) \quad i \quad g'(b) = -\alpha'(b)g(b),$$

zaključujemo da je inverzan operator T^{-1} u prostoru $C^2[a, b]$ definisan za funkcije koje zadovoljavaju granične uslove (3.1.3).

Teorema 3.1.3. Neka je X Banachov prostor, I identički operator u X i $T: X \rightarrow X$ ograničen linearни operator kod koga je $\|T\| \leq q < 1$. Tada postoji operator $(I-T)^{-1}$ za koji važi:

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

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

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

$$(3.1.4) \quad \sum_{k=0}^{+\infty} \|T^k\| \leq \sum_{k=0}^{+\infty} \|T\|^k \leq \frac{1}{1-q} < +\infty.$$

Kako je, dalje, prostor X kompletan, to iz konvergenčije 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 $n \in \mathbb{N}$, a s obzirom na

$$\|T^{n+1}\| \leq \|T\|^n n! \rightarrow 0 \Rightarrow T^{n+1} \rightarrow 0 \quad (n \rightarrow +\infty),$$

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 (3.1.4) neposredno dobijamo

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

Teorema 3.1.4. Neka su $A:X \rightarrow X$ i $B:X \rightarrow X$ linearni operatori za koje postoje neprekidni inverzni operatori A^{-1} i B^{-1} . Tada je

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

gde je I identički operator u X .

Definicija 3.1.9. Neka je $T:X \rightarrow X$ 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 3.1.5 (Banach-Steinhaus). Neka su X i Y Banachovi prostori i neka je $\{A_n\}_{n \in \mathbb{N}}$ niz ograničenih linearnih operatora koji preslikavaju X i Y . Potrebni i dovoljni uslovi za konvergenciju niza $\{A_n u\}$ ka Au za svako $u \in X$, gde je $A:X \rightarrow Y$ ograničen linearan operator, su

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

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

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

2.3.2. Matrica linearog operatora na konačno-dimenzionalnim prostorima

Neka su X i Y konačno-dimenzionalni prostori sa bazama $B_u = \{u_1, \dots, u_n\}$ i $B_v = \{v_1, \dots, v_m\}$ respektivno i neka je $A: X \rightarrow Y$ 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 ($i=1, \dots, n$). Razložimo ove vektore po vektorima baze B_v . Tada imamo

$$(3.2.1) \quad \begin{aligned} Au_1 &= a_{11}v_1 + a_{21}v_2 + \dots + a_{m1}v_m, \\ Au_2 &= a_{12}v_1 + a_{22}v_2 + \dots + a_{m2}v_m, \\ &\vdots \\ Au_n &= a_{1n}v_1 + a_{2n}v_2 + \dots + a_{mn}v_m. \end{aligned}$$

Na osnovu (3.2.1) formirajmo matricu

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

Definicija 3.2.1. Za A_{vu} kažemo da je matrica operatora $A: X \rightarrow Y$ 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 $A:X \rightarrow X$ 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.

2.3.3. Bilinearni i n -linearjni 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 linearног operatora T u smislu definicije 3.1.5.

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

Definicija 3.3.1. Operator B je bilinearан 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:

$$B(c_1 u_1 + c_2 u_2, u') = c_1 B(u_1, u') + c_2 B(u_2, u') ,$$

$$B(u, c_1 u'_1 + c_2 u'_2) = c_1 B(u, u'_1) + c_2 B(u, u'_2) .$$

Ako postoji pozitivan broj M takav da je

$$(3.3.1) \quad \|B(u, u')\| \leq M \|u\| \cdot \|u'\|$$

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

Infimum brojeva M za koje važi (3.3.1) 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 3.3.2. Operator $N: X^n \rightarrow Y$ je n -linearan ako svakom uredjenom sistemu elemenata (u_1, u_2, \dots, u_n) iz X odgovara element $g = N(u_1, u_2, \dots, u_n) \in Y$, pri čemu je on linearan po svakom u_i , pri fiksiranim ostalim elementima $u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_n$.

Ako postoji pozitivan broj M takav da je

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

operator N je ograničen.

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

2.3.4. Fréchetova diferenciranja

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

Definicija 3.4.1. Operator T je Fréchet-diferencijabilan u tački $u \in D$, ako postoji takav ograničeni linearan operator $A \in L(X, Y)$ da je

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

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

Teorema 3.4.1. 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

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

i neka je $A = A_1 - A_2$.

Korišćenjem

$$\|Ah\| = \|A_1 h - A_2 h\| \leq \|T(u+h) - Tu - A_1 h\| + \|T(u+h) - Tu - A_2 h\|$$

nalazimo

$$(3.4.1) \quad \lim_{\|h\| \rightarrow 0} \frac{\|Ah\|}{\|h\|} = L_1 + L_2 = 0.$$

Medjutim, ako je za neko h

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

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

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

odakle zaključujemo da je jednakost (3.4.1) nemoguća kada je A nenula-operator.

Dakle, (3.4.1) važi ako je A nula-operator, tj. $A_1 = A_2$, čime je dokaz završen.

Teorema 3.4.2. Svaki linearan ograničeni operator je Fréchet-diferencijabilan, pri čemu je $T_{(u)}^* = T^*$ za svako $u \in D$.

Dokaz. Kako je $T \in L(X, Y)$, na osnovu definicije 3.4.1 i prethodne teoreme neposredno sleduje

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

Sada ćemo dati neka pravila diferencijalnog računa:

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

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

2° Neka su X, Y, Z Banachovi prostori, a $T: X \rightarrow Y$ i $S: Y \rightarrow Z$ 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° Ako je A proizvoljan linearan ograničeni operator, a T Fréchet-diferencijabilan operator važi

$$(AT)'(u) = AT'(u) \quad i \quad (tA)'(u) = T'(Au)^A.$$

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

Primer 3.4.1. Neka je $X=R^n$, $Y=R$, $\vec{x} = f(\vec{x}) = f(x_1, \dots, x_n)$, $\vec{x} = [x_1 \dots x_n]^T$, $\vec{h} = [h_1 \dots h_n]^T$, $\|\vec{x}\|_\infty = \max_k |x_k|$. 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

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

tj.

$$(3.4.2) \quad |f(\vec{x} + \vec{h}) - f(\vec{x}) - (\text{grad } f)^T \vec{h}| \leq M(f, \vec{c}) \left(\max_k |h_k| \right)^2,$$

gde smo stavili

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

S obzirom na neprekidnost parcijalnih izvoda drugog reda funkcije f , veličina $M(f, \vec{c})$ je konačna, pa na osnovu (3.4.2) imamo

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

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

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

Primer 3.4.2. Neka je $X=R^n$, $Y=R^m$, $\vec{x} \in X$, $\vec{y} \in Y$, $\|\vec{x}\|_\infty = \max_{1 \leq k \leq n} |x_k|$, $\|\vec{y}\|_\infty = \max_{1 \leq k \leq m} |y_k|$, $\vec{h} = [h_1 \dots h_n]^T \in X$ i neka je operator $T: X \rightarrow Y$ definisan pomoću

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

Ako pretpostavimo da funkcije f_k ($k=1, \dots, m$) 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.

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

2.3.5. Taylorova formula

Najpre uvodimo pojam L-metričkog prostora.

Definicija 3.5.1. Metrički prostor X naziva se L-metrički, ako za svako u X postoji linearna ograničena funkcionala $L:X \rightarrow \mathbb{R}$, takva da je

$$(3.5.1) \quad \|L\| = 1 \quad i \quad Lu = \rho(u, \theta).$$

Može se pokazati da Banachov prostor poseduje svojstvo (3.5.1) (videti [1]).

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 3.5.1. Ako su u i $u+h$ zadate tačke iz D , tada je

$$(3.5.2) \quad T(u+h) = \sum_{k=0}^n \frac{1}{k!} T^{(k)}(u) (\underbrace{h, h, \dots, h}_k \text{ puta}) + W(u, h),$$

gde je

* Skup D je konveksan ako važi implikacija $u, u+h \in D \Rightarrow u+th \in D$ ($0 \leq t \leq 1$).

$$(3.5.3) \quad \|W(u,h)\| \leq \frac{1}{(n+1)!} \sup_{t \in [0,1]} \|T_{(u+th)}^{(n+1)}\| \cdot \|h\|^{n+1}.$$

Dokaz. S obzirom da je X L-metrički prostor to se za element $W(u,h)$ nalazi linearna ograničena funkcionala L , takva da je

$$\|L\| = 1 \quad i \quad LW = \rho(W, \theta) = \|W(u,h)\|.$$

Uvedimo sada pomoćnu funkciju

$$(3.5.4) \quad t \mapsto F(t) = LT(u+th),$$

čiji su izvodi redom jednaki

$$\begin{aligned} F'(t) &= LT'_{(u+th)} h, \\ &\vdots \\ F^{(n+1)}(t) &= LT^{(n+1)}_{(u+th)} \underbrace{(h, h, \dots, h)}_{n+1 \text{ puta}}, \end{aligned}$$

s obzirom da je $(LT)'(a) = LT'(a)$ i $\frac{d}{dt}(u+th) = h$.

Sada imamo

$$\|W(u,h)\| = LW = LT(u+h) - LTu - \frac{1}{1!} LT'_{(u)} h - \dots - \frac{1}{n!} LT^{(n)}_{(u)} \underbrace{(h, \dots, h)}_{n \text{ puta}},$$

tj.

$$\|W(u,h)\| = F(1) - F(0) - \frac{1}{1!} F'(0) - \dots - \frac{1}{n!} F^{(n)}(0).$$

Kako za funkciju (3.5.4) važi klasična ocena ostatka u Taylorovoj formuli

$$|F(1) - F(0) - \frac{1}{1!} F'(0) - \dots - \frac{1}{n!} F^{(n)}(0)| \leq \frac{1}{(n+1)!} \max_{t \in [0,1]} |F^{(n+1)}(t)|,$$

dobijamo

$$\|W(u,h)\| \leq \frac{1}{(n+1)!} \sup_{t \in [0,1]} (\|L\| \cdot \|T_{(u+th)}^{(n+1)}\|) \|h\|^{n+1},$$

odakle, na osnovu (3.5.1), zaključujemo da važi (3.5.3), čime je dokaz završen.

Formula (3.5.2) naziva se Taylorova formula za operatore.

Stavljanjem $n=0$ u (3.5.3) 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.

2.4. LITERATURA

1. S.Aljančić: Uvod u realnu i funkcionalnu analizu. Gradjevinska knjiga, Beograd, 1979.
2. H.Bateman & A.Erdélyi: Higher transcendental functions. Vol. II., Mc. Graw-Hill, New York, 1953 (rusko izdanje: Nauka, Moskva, 1974).
3. P.L.Chebyshev: *Sur l'interpolation par la méthode des moindres carres.* Mém. Acad. Impér. Sci. St. Pétersbourg.(7)1(1859), no 15, 1 - 24.
4. T.S.Chihara: An introduction to orthogonal polynomials. Gordon & Breach, New York, 1978.
5. L.Collatz: Functionalanalysis und numerische Mathematik. Springer-Verlag, Berlin-Heidelberg-New York, 1964.
6. W.Gautschi: Computational aspects of three-term recurrence relations. SIAM Review. 9(1967), 24-82
7. W.Gautschi: Construction of Gauss-Christoffel quadrature formulas. Math. Comp. 22(1968), 251-270.
8. W.Gautschi: On the construction of Gaussian quadrature rules from modified moments. Math. Comp. 24(1970), 245-260.
9. W.Gautschi: Minimal solutions of three-term recurrence relations and orthogonal polynomials. Math. Comp. 36(1981), 547-554.
10. W.Gautschi: On generating orthogonal polynomials. SIAM J. Sci. Stat. Comput. 3(1982), 289-317.
11. W.Gautschi: Polynomials orthogonal with respect to the reciprocal gamma function. BIT 22(1982), 387-389.
12. W.Gautschi & G.V.Milovanović: Polynomials orthogonal on the semi-circle. J. Approx. Theory 46(1986), 230-250.
13. W.Gautschi & G.V.Milovanović: Polynomials orthogonal on the semi-circle. Rend. Sem. Mat. Univ. Politec. Torino (1985), 179-185.
14. V.C.L.Hutson & J.S.Pym: Applications of functional analysis and operator theory. Academic Press, London, 1980.
15. L.V.Kantorović i G.P.Akilov: Funkcional'nyi analiz. Nauka, Moskva, 1977.

16. S.Kurepa: Konačno-dimenzionalni vektorski prostori i primjene. Tehnička knjiga, Zagreb, 1967.
17. S.Kurepa: Funkcionalna analiza - Elementi teorije operatora. Školska knjiga, Zagreb, 1981.
18. S.Mardešić: Matematička analiza u n-dimenzionalnom realnom prostoru. II deo Integral i mjera. Školska knjiga, Zagreb, 1977.
19. D.S.Mitrinović: Uvod u specijalne funkcije. Gradjevinska knjiga, Beograd, 1975.
20. D.S.Mitrinović (saradnik J.D.Kečkić): Kompleksna analiza - Zbornik zadataka i problema. Naučna knjiga, Beograd, 1972.
21. P.G.Nevai: Orthogonal polynomials. Mem. Amer. Math. Soc. No 213, 1979.
22. A.F.Nikiforov i V.B.Uvarov: Osnovy teorii special'nyh funkciǐ. Nauka, Moskva, 1974.
23. S.Paškovskii: Vyčislitel'nye primenenija mnogočlenov i rjadov Čebyševa. Nauka, Moskva, 1983.
24. B.D.Rakovich & P.M.Vasić: Some new sums for orthogonal polynomials involved in filter synthesis. Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz. N°544 - N°576(1976), 8-12.
25. B.D.Raković i P.M.Vasić: Primena ekstremalnih svojstava ortogonalnih polinoma u sintezi električnih kola. XXII Jugoslovenska konferencija za ETAN, 12-16. juna 1978, Zadar, I.265-I.272.
26. B.D.Rakovich & P.M.Vasić: Some extremal properties of orthogonal polynomials. Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz. N°634 - N°677(1979), 25-32.
27. B.D.Rakovich & P.M.Vasić: Some extremal properties of orthogonal polynomials. Aeq. Math. 26(1983), 125-137.
28. T.J.Rivlin: The Chebyshev polynomials. Wiley, New York, 1974.
29. R.A.Sack & A.F.Donovan: An algorithm for Gaussian quadrature given modified moments. Numer. Math. 18(1971/72), 465-478.
30. T.J.Stieltjes: Quelques recherches sur la théorie des quadratures di-tes mécaniques. Ann. Sci. Ecole Norm. Paris Sér. 3, 1(1884), 409-426.
31. P.K.Suetin: Klassičeskie ortogonal'nye mnogočleny. Nauka, Moskva, 1976.
32. G.Szegő: Über trigonometrische und harmonische Polynome. Math. Ann. 79(1918), 323-339.
33. G.Szegő: Orthogonal polynomials. AMS Colloquium Publications 23, Providence, R.I., 1975.
34. N.I.Vasil'ev, Ju.A.Klokov, A.Ja.Škerstena: Primenenie polinomov Čebyševa v čislennom analize. Zinatne, Riga, 1984.
35. J.C.Wheeler: Modified moments and Gaussian quadrature. Rocky Mountain J. Math. 4(1974), 287-296.

Opšta teorija iterativnih procesa

3.1. REŠAVANJE OPERATORSKIH JEDNAČINA

Ovo poglavlje je posvećeno problemu egzistencije rešenja operatorskih jednačina u Banachovom prostoru i definiciji iterativnog procesa.

3.1.1. Osnovne napomene o rešavanju operatorskih jednačina

Neka su X i Y Banachovi prostori, D konveksan podskup prostora X i $F:D \rightarrow Y$. Posmatrajmo operatorsku jednačinu

$$(1.1.1) \quad Fu = \theta,$$

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 primera.

Primer 1.1.1. Ako je $X = Y = \mathbb{R}$, $u = x$, $F = f$, jednačine

$$f(x) = x - \cos x = 0$$

i

$$f(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n = 0$$

su oblika (1.1.1).

Primer 1.1.2. Ako je $X = Y = \mathbb{R}^n$, $u = \vec{x} = [x_1 \dots x_n]^T$ i

$$Fu = F(\vec{x}) = \begin{bmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_n(x_1, \dots, x_n) \end{bmatrix},$$

gde su $f_i: \mathbb{R}^n \rightarrow \mathbb{R}$ date funkcije, jednačina (1.1.1) predstavlja sistem jednačina

$$f_i(x_1, \dots, x_n) = 0 \quad (i = 1, \dots, n).$$

Ako je F linearan operator, na primer, $F(\vec{x}) = A\vec{x} - \vec{b}$, gde su matrica A i vektor \vec{b} dati sa

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & & a_{nn} \end{bmatrix} \quad i \quad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix},$$

jednačina (1.1.1) predstavlja sistem linearnih algebarskih jednačina

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = b_i \quad (i = 1, \dots, n).$$

Primer 1.1.3. Neka su $X = C^2[a, b]$, $Y = C[a, b] \times \mathbb{R}$, $u \equiv u(t)$,

$$Fu = \begin{bmatrix} f_1(u) \\ f_2(u) \end{bmatrix},$$

$$f_1(u)(t) = u''(t) - f(t, u(t), u'(t)) \quad (t \in [a, b]),$$

$$f_2(u) = g(u(a), u(b)),$$

gde su $F: \mathbb{R}^3 \rightarrow \mathbb{R}$ i $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ date funkcije.

Tada rešenje operatorske jednačine (1.1.1) predstavlja rešenje konturnog problema

$$u''(t) = f(t, u(t), u'(t)) \quad (t \in [a, b]),$$

$$g(u(a), u(b)) = 0.$$

Predmet našeg razmatranja je rešavanje jednačine (1.1.1),

tj. nalaženje takve tačke u ϵD , koja zadovoljava (1.1.1). U tom cilju ovu jednačinu predstavimo u ekvivalentnom obliku

$$(1.1.2) \quad u = Tu$$

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

$$(1.1.3) \quad u_{k+1} = Tu_k \quad (k = 0, 1, \dots)$$

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 $\{u_k\}_{k \in N}$ zvemo iterativnim procesom.

Pored iterativnih procesa oblika (1.1.3) postoje i opštiji iterativni procesi oblika

$$u_{k+1} = S(u_k, u_{k-1}, \dots, u_{k-m+1}) \quad (k = m-1, m, \dots),$$

pri čemu $S: X^m \rightarrow X$ i kod kojih se startuje sa m početnih vrednosti $u_0, u_1, \dots, u_{m-1} \in D$.

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

$$(1.2.1) \quad u = Tu$$

zovu se nepokretne tačke operatora T . Dovoljne uslove za egzistenciju jedinstvene nepokretne tačke operatora T dao je Banach.

Definicija 1.2.1. Operator $T:X \rightarrow X$ naziva se kontrakcijom ako postoji pozitivan broj $q < 1$ takav da je za bilo koje dve tačke $u, v \in X$

$$\|Tu - Tv\| \leq q \|u - v\|.$$

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 $u_0 \in X$ konstruišimo niz $\{u_k\}_{k \in \mathbb{N}}$ pomoću

$$u_{k+1} = Tu_k \quad (k = 0, 1, \dots).$$

Kako je za $k \geq 1$

$$\|u_{k+1} - u_k\| = \|Tu_k - Tu_{k-1}\| \leq q \|u_k - u_{k-1}\|,$$

imamo

$$\|u_{k+1} - u_k\| \leq q^k \|u_1 - u_0\|.$$

Ako je $m > k$, tada je

$$\begin{aligned}
 (1.2.2) \quad \|u_m - u_k\| &= \|(u_{k+1} - u_k) + (u_{k+2} - u_{k+1}) + \dots + (u_m - u_{m-1})\| \\
 &\leq \|u_{k+1} - u_k\| + \|u_{k+2} - u_{k+1}\| + \dots + \|u_m - u_{m-1}\| \\
 &\leq (q^k + q^{k+1} + \dots + q^{m-1}) \|u_1 - u_0\| \\
 &< \frac{q^k}{1-q} \|u_1 - u_0\| \quad (q < 1),
 \end{aligned}$$

odakle je $\lim_{k \rightarrow +\infty} \|u_m - u_k\| = 0$, što znači da je $\{u_k\}_{k \in \mathbb{N}_0}$ Cauchyev niz.

S obzirom da je X Banachov prostor, to je on kompletan pa postoji

$$(1.2.3) \quad \lim_{k \rightarrow +\infty} u_k = a \quad (a \in X).$$

S druge strane, kako je

$$\|u_{k+1} - Ta\| = \|Tu_k - Ta\| \leq q \|u_k - a\|,$$

imamo

$$\lim_{k \rightarrow +\infty} \|u_{k+1} - Ta\| = 0,$$

tj. $\lim_{k \rightarrow +\infty} u_{k+1} = Ta$.

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

$$\|Ta - Tb\| = \|a - b\| \leq q \|a - b\|,$$

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 građnična vrednost niza, koji se generiše pomoću

$$(1.2.4) \quad u_{k+1} = Tu_k \quad (k = 0, 1, \dots),$$

gde je u_0 proizvoljna tačka iz X .

Ako se zadržimo na k -toj aproksimaciji u_k , koja je određena pomoću (1.2.4), iz (1.2.2), pri $m \rightarrow +\infty$, sleduje

$$\|a - u_k\| < \frac{q^k}{1-q} \|u_1 - u_0\|,$$

što znači da se ova aproksimacija nalazi u kugli sa centrom u tački a i poluprečnikom $r_k = \frac{q^k}{1-q} \|u_1 - u_0\|$.

3.1.3. Iterativni procesi za rešavanje običnih jednačina

Neka je data jednačina

$$(1.3.1) \quad f(x) = 0,$$

gde $f: [\alpha, \beta] \rightarrow \mathbb{R}$, i neka je

$$(1.3.2) \quad x = \phi(x)$$

njen ekvivalentni oblik.

Teorema 1.3.1. Pretpostavimo da neprekidna funkcija ϕ zadovoljava uslove

$$1^{\circ} \quad \phi : [\alpha, \beta] \rightarrow [\alpha, \beta],$$

2° ϕ ima izvod u svakoj tački $x \in [\alpha, \beta]$, takav da je $|\phi'(x)| \leq q < 1$.

Tada jednačina (1.3.2), tj. jednačina (1.3.1), ima jedinstveno rešenje $a \in [\alpha, \beta]$ i ono se može odrediti iterativnim procesom

$$(1.3.3) \quad x_{k+1} = \phi(x_k) \quad (k = 0, 1, \dots)$$

sa proizvoljnim $x_0 \in [\alpha, \beta]$.

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

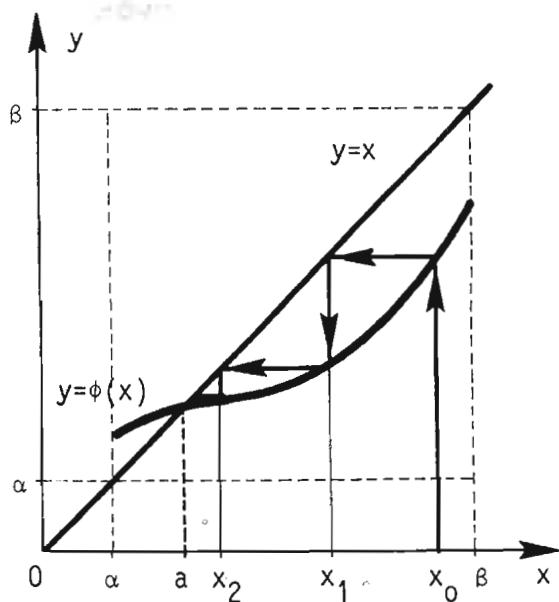
$$\Phi(x) - \Phi(y) = \Phi'(\xi)(x - y),$$

gde su x i y proizvoljne tačke iz $[\alpha, \beta]$ i $\xi = y + \theta(x - y)$ ($0 < \theta < 1$), sleduje

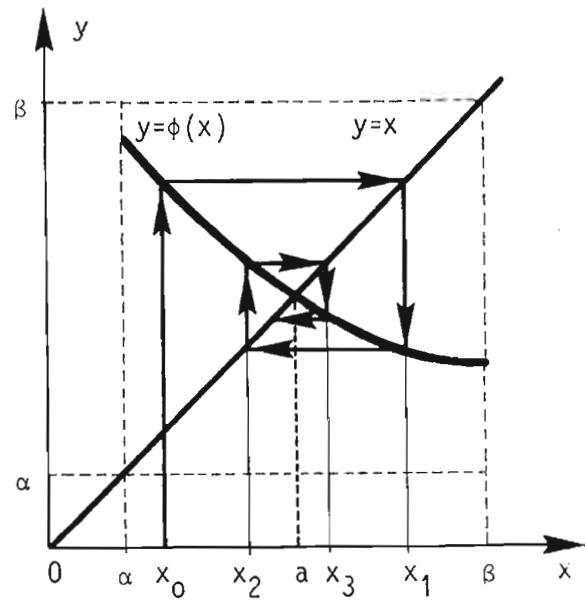
$$|\Phi(x) - \Phi(y)| = |\Phi'(\xi)| \cdot |x - y| \leq q|x - y|,$$

zaključujemo da je Φ kontrakcija na $[\alpha, \beta]$, čime je dokaz teoreme završen.

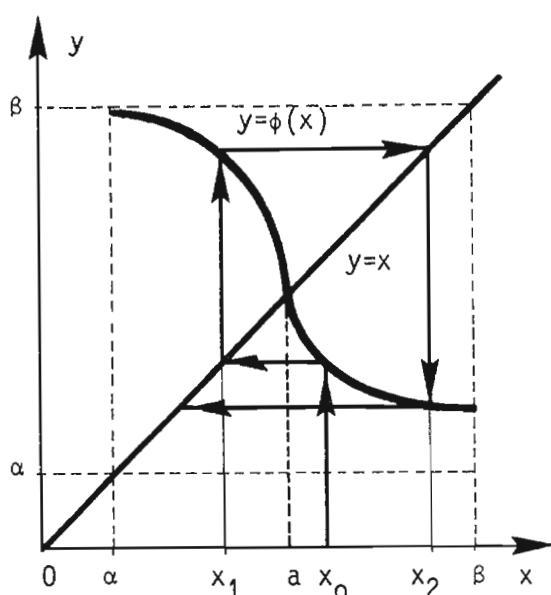
Dakle, da bi iterativni proces (1.3.3) konvergirao ka tački a , funkcija Φ (često se naziva iterativna funkcija) mora ispunjavati odredjene uslove. Na slikama 1.3.1 do 1.3.4 data je geometrijska interpretacija iterativnih procesa oblika (1.3.3). Prva dva procesa (videti sl. 1.3.1 i sl. 1.3.2) su konvergentna. 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° , a kod drugog uslov 1° iz teoreme 1.3.1.



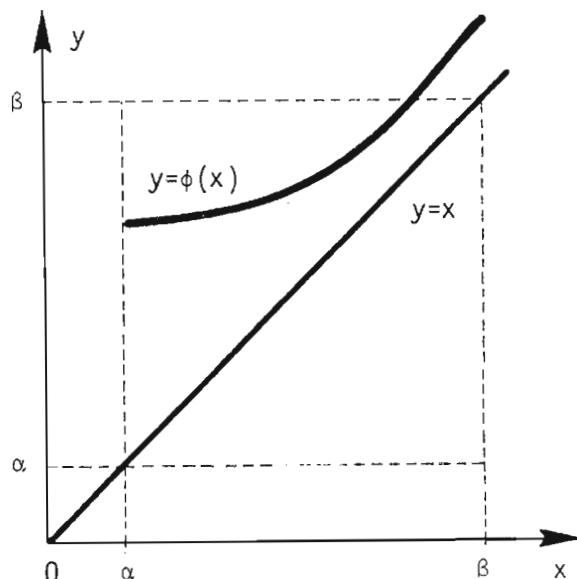
Sl. 1.3.1



Sl. 1.3.2



Sl. 1.3.3



Sl. 1.3.4

Primer 1.3.1. Posmatrajmo jednačinu

$$(1.3.4) \quad x^3 + x - 60 = 0.$$

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

$$|\phi'_2(x)| = \left| \frac{-1}{3\sqrt[3]{(60-x)^2}} \right| \leq \frac{1}{3(56)^{2/3}} \approx 0.022$$

kada $x \in [3,4]$.

Dakle, koren jednačine (1.3.4) može se odrediti iterativnim procesom

$$(1.3.5) \quad x_{k+1} = \sqrt[3]{60 - x_k} \quad (k = 0, 1, \dots).$$

Polazeći od $x_0 = 4$, pomoću (1.3.5) dobijamo

k	x_k
0	4.
1	3.8258623
2	3.8298239
3	3.8297338
4	3.8297359
5	3.8297358

tj. $a \approx 3.8297358$.

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

$$(1.3.6) \quad 0 < m \leq f'(x) \leq M \quad (\alpha \leq x \leq \beta).$$

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

$$(1.3.7) \quad 0 \leq \Phi'(x) = 1 - \lambda f'(x) \leq q < 1$$

za $x \in [\alpha, \beta]$. Tada na osnovu (1.3.6), imamo

$$0 \leq 1 - \lambda M \leq 1 - \lambda m \leq q,$$

odakle zaključujemo da će (1.3.7) biti ispunjeno ako je $\lambda = \frac{1}{M}$. U tom slučaju je

$$q = \max_{\alpha \leq x \leq \beta} (1 - \lambda f'(x)) = 1 - \frac{m}{M} < 1,$$

što znači da se može uzeti $\Phi(x) = x - \frac{1}{M} f(x)$.

Primedba 1.3.1. Ako je $|\Phi'(x)| \geq p > 1$, kada $x \in [\alpha, \beta]$, iterativni proces (1.3.3) divergira. Međutim, ako se jednačina (1.3.2) napiše u obliku

$$x = \Psi(x),$$

gde je Ψ inverzna funkcija od Φ , iterativni proces

$$x_{k+1} = \Psi(x_k) \quad (k = 0, 1, \dots)$$

biće konvergentan, jer je

$$|\Psi'(x)| = \left| \frac{1}{\Phi'(\Psi(x))} \right| \leq \frac{1}{p} = q < 1.$$

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

$$(1.3.8) \quad \vec{Ax} = \vec{b},$$

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

$$(1.3.9) \quad \vec{x} = \vec{Bx} + \vec{\beta},$$

gde su $B = [b_{ij}]_{n \times n}$ i $\vec{\beta} = [\beta_1 \dots \beta_n]^T$, može se konstruisati iterativni proces sa operatom $\vec{x} \mapsto T(\vec{x}) = \vec{Bx} + \vec{\beta}$.

S obzirom da je T linearan operator, na osnovu

$$\|T\vec{x} - T\vec{y}\| = \|B(\vec{x} - \vec{y})\| \leq \|B\| \cdot \|\vec{x} - \vec{y}\|$$

zaključujemo da je $\|B\| \leq q < 1$ dovoljan uslov da je T kontrakcija.

Dakle, ako je $\|B\| \leq q < 1$, sistem jednačina (1.3.8), tj. (1.3.9), ima jedinstveno rešenje i odgovarajući iterativni proces

$$\vec{x}_{k+1} = T(\vec{x}_k) \quad (k = 0, 1, \dots)$$

konvergira ka $\vec{a} = A^{-1}\vec{b}$. Napomenimo da vrednost $\|B\|$ zavisi od izbora metrike u R^n .

3.2. KARAKTERISTIKE PROCESA

U ovom poglavlju se definišu osnovne karakteristike iterativnih procesa - red konvergencije i asymptotska konstanta greške, a zatim se daju neki postupci za ubrzavanje konvergencije. U poslednjem odeljku uveden je pojam R-reda konvergencije.

3.2.1. Red konvergencije iterativnih procesa

Neka su X Banachov prostor i operator $T:X \rightarrow X$ i neka niz

$$(2.1.1) \quad u_{k+1} = Tu_k \quad (k=0,1,\dots)$$

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

$$(2.1.2) \quad u_{k+1} = S(u_k, u_{k-1}, \dots, u_{k-m+1}) \quad (k=m-1, m, \dots),$$

pri čemu $S:X^m \rightarrow X$.

Definicija 2.1.1. Za iterativni proces koji konvergira ka a , kaže se da ima red konvergencije r ako je

$$(2.1.3) \quad \|u_{k+1}-a\| = O(\|u_k-a\|^r),$$

tj. ako postoji konstanta A takva da je za dovoljno veliko k

$$\|u_{k+1}-a\| \leq A \|u_k-a\|^r.$$

Označimo sa $U(a)$ konveksnu okolinu tačke a . Tada, za iterativni proces (2.1.1), umesto (2.1.3) u prethodnoj definiciji, možemo uzeti

$$\|Tu - a\| = O(\|u-a\|^r) \quad (u \in U(a)).$$

Teorema 2.1.1. Ako je operator $T:X \rightarrow X$ r-puta Fréchet-diferencijabilan u konveksnoj okolini $U(a)$, iterativni proces (2.1.1) je reda r ako su ispunjeni sledeći uslovi:

$$1^{\circ} \quad Ta = a,$$

$$2^{\circ} \quad T'_{(a)}, T''_{(a)}, \dots, T^{(r-1)}_{(a)} \text{ su nula-operatori,}$$

$$3^{\circ} \quad T^{(r)}_{(u)} \text{ je nenula-operator sa normom koja je ograničena na } U(a).$$

Dokaz. Neka $u \in U(a)$, $\|T^{(r)}_{(u)}\| \leq M_r$ i

$$q = Ta + \frac{1}{1!} T'_{(a)}(u-a) + \dots + \underbrace{\frac{1}{(r-1)!} T^{(r-1)}_{(a)}(u-a, \dots, u-a)}_{k-1 \text{ puta}}.$$

Tada na osnovu Taylorove formule, imamo

$$\|Tu - q\| \leq \frac{1}{r!} \sup_{t \in [0,1]} \|T^{(r)}_{(a+t(u-a))}\| \cdot \|u-a\|^r,$$

tj.

$$\|Tu - q\| \leq \frac{M_r}{r!} \|u-a\|^r.$$

Kako je, na osnovu pretpostavki teoreme, $q = Ta = a$, imamo

$$\|Tu - a\| \leq A \|u - a\|^r \quad (A = \frac{M_r}{r!}),$$

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 prirodan broj, a da je kod procesa tipa (2.1.2) realan broj (≥ 1).

U slučaju da je $r = 1$, dovoljan uslov za konvergenciju iterativnog procesa je $0 \leq A < 1$.

Posmatrajmo sada slučaj kada je $X = R$.

Teorema 2.1.2. Neka je

$$x_{k+1} = \phi(x_k) \quad (k = 0, 1, \dots)$$

iterativni proces reda r , gde je $x \mapsto \phi(x)$ r puta neprekidno-diferencijabilna funkcija u okolini tačke $a = \lim_{k \rightarrow +\infty} x_k$. Tada je

$$(2.1.4) \quad \lim_{k \rightarrow +\infty} \frac{x_{k+1} - a}{(x_k - a)^r} = \frac{\phi'(a)}{r!}.$$

Dokaz. Kako je $\phi(a) = a$ i $\phi^{(i)}(a) = 0$ ($i = 1, \dots, r-1$), na osnovu Taylor ove 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 $a = \lim_{k \rightarrow +\infty} x_k$ i $\phi^{(r)}$ neprekidna funkcija,

deobom jednakosti (2.1.5) sa $(x_k - a)^r$ i prelazom na graničnu vrednost dobijamo (2.1.4).

Definicija 2.1.2. Veličina

$$c_r = \lim_{k \rightarrow +\infty} \frac{|x_{k+1} - a|}{|x_k - a|^r}$$

naziva se faktor konvergencije ili asymptotska konstanta greške.

Primedba 2.1.1. Iterativni proces kod koga je $r = 1$ naziva se proces sa linearnom konvergencijom. Njegov 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

$$x_{k+1} = \phi(x_k) \quad (k = 0, 1, \dots),$$

gde je

$$\phi(x) = x \frac{(m-1)x^m + (m+1)N}{(m+1)x^m + (m-1)N}.$$

Sukcesivnim diferenciranjem jednakosti

$$(2.1.6) \quad g(x)\phi(x) = (m-1)x^{m+1} + (m+1)Nx,$$

gde je

$$g(x) = (m+1)x^m + (m-1)N,$$

nalazimo redom

$$(2.1.7) \quad g(x)\phi'(x) = -m(m+1)x^{m-1}\phi(x) + (m-1)(m+1)x^m + (m+1)N,$$

$$(2.1.8) \quad g(x)\phi''(x) = -2m(m+1)x^{m-1}\phi'(x) - m(m-1)(m+1)x^{m-2}\phi(x) \\ + m(m-1)(m+1)x^{m-1},$$

$$(2.1.9) \quad g(x)\phi'''(x) = -3m(m+1)x^{m-1}\phi''(x) - 3m(m-1)(m+1)x^{m-2}\phi'(x) \\ - m(m-2)(m-1)(m+1)x^{m-3}\phi(x) + m(m-1)^2(m+1)x^{m-2}.$$

Kako je $a^m = N$, iz (2.1.6) do (2.1.9) sleduje

$$(2.1.10) \quad \phi(a) = a, \quad \phi'(a) = 0, \quad \phi''(a) = 0, \quad \phi'''(a) = \frac{m^2-1}{2a^2}.$$

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

$$\lim_{k \rightarrow +\infty} \frac{x_{k+1}-a}{(x_k-a)^3} = \frac{\phi'''(a)}{3!} = \frac{m^2-1}{12a^2}.$$

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 $\{x_k\}_{k \in \mathbb{N}_0}$, koji se generiše pomoću

$$(2.2.1) \quad x_{k+1} = \phi(x_k) \quad (k = 0, 1, \dots),$$

pri čemu je $\lim_{k \rightarrow +\infty} x_k = a$.

Pretpostavimo da je iterativni proces (2.2.1) sa linearnom konvergencijom, tj. da je

$$(2.2.2) \quad x_{k+1} - a = C_k(x_k - a) \quad (k = 0, 1, \dots),$$

gde su $C_k = C + \delta_k$ ($k = 0, 1, \dots$), C konstanta takva da je $|C| < 1$ i $\delta_k \rightarrow 0$, kada $k \rightarrow +\infty$.

Iz asimptotskih relacija

$$x_{k+2} - a \sim C(x_{k+1} - a) \quad i \quad x_{k+1} - a \sim C(x_k - a) \quad (k \rightarrow +\infty),$$

koje se dobijaju na osnovu (2.2.2), eliminacijom konstante C , sleduje

$$a \sim \frac{x_{k+2}x_k - x_{k+1}^2}{x_{k+2} - 2x_{k+1} + x_k} \quad (k \rightarrow +\infty).$$

Dobijena asimptotska relacija sugerije konstrukciju niza $\{x_k^*\}_{k \in \mathbb{N}_0}$ pomoću

$$(2.2.3) \quad x_k^* = \frac{x_{k+2}x_k - x_{k+1}^2}{x_{k+2} - 2x_{k+1} + x_k} \quad (k = 0, 1, \dots).$$

Formula (2.2.3), međutim, nije pogodna za primenu sa numeričkog stanovišta, pa se zato koriste njeni ekvivalentni matematički oblici

$$(2.2.4) \quad x_k^* = x_k - \frac{(x_{k+1} - x_k)^2}{x_{k+2} - 2x_{k+1} + x_k},$$

$$(2.2.5) \quad x_k^* = x_{k+1} - \frac{(x_{k+1} - x_k)(x_{k+2} - x_{k+1})}{x_{k+2} - 2x_{k+1} + x_k},$$

$$(2.2.6) \quad x_k^* = x_{k+2} - \frac{(x_{k+2} - x_{k+1})^2}{x_{k+2} - 2x_{k+1} + x_k},$$

gde je $k = 0, 1, \dots$

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 formulama (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 poređenju 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

$$(2.2.7) \quad x_k^* = x_k - \frac{(\Delta x_k)^2}{2\Delta x_k} \quad (k = 0, 1, \dots).$$

Sledeća teorema ukazuje na to da niz $\{x_k^*\}_{k \in \mathbb{N}_0}$ brže konvergira ka a, nego niz $\{x_k\}_{k \in \mathbb{N}_0}$.

Teorema 2.2.1. Neka je za niz $\{x_k\}_{k \in \mathbb{N}_0}$, $x_k \neq a$ i

$$x_{k+1} - a = (C + \delta_k)(x_k - a) \quad (|C| < 1, \lim_{k \rightarrow +\infty} \delta_k = 0),$$

tada je

$$(2.2.8) \quad \lim_{k \rightarrow +\infty} \frac{x_k^* - a}{x_k - a} = 0,$$

gde je $a = \lim_{k \rightarrow +\infty} x_k$.

Dokaz. Neka je $e_k = x_k - a$. Tada je $e_{k+1} = (C + \delta_k)e_k$ i

$$\Delta x_k = \Delta e_k = e_k((C-1) + \delta_k),$$

$$\begin{aligned}\Delta^2 x_k &= \Delta^2 e_k = e_k ((C + \delta_{k+1}) (C + \delta_k) - 2(C + \delta_k) + 1) \\ &= e_k ((C-1)^2 + \gamma_k),\end{aligned}$$

gde je $\gamma_k = C(\delta_{k+1} + \delta_k) - 2\delta_k + \delta_k \delta_{k+1}$. Primetimo da uslov $\lim_{k \rightarrow +\infty} \delta_k = 0$ implicira uslov $\lim_{k \rightarrow +\infty} \gamma_k = 0$.

Da bismo dokazali (2.2.8) podjimo od (2.2.7).

Kako je $e_k \neq 0$, $C \neq 1$ i $\gamma_k \rightarrow 0$ ($k \rightarrow +\infty$), imamo $\Delta^2 x_k = 0$, što znači da je niz $\{x_k^*\}_{k \in \mathbb{N}_0}$ 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.

$$\lim_{k \rightarrow +\infty} \frac{x_k^* - a}{x_k - a} = \lim_{k \rightarrow +\infty} \frac{\gamma_k - 2(C-1)\delta_k - \delta_k^2}{(C-1)^2 + \gamma_k} = 0,$$

čime je dokaz završen.

Navedeni postupak ubrzavanja konvergencije niza $\{x_k\}_{k \in \mathbb{N}_0}$ 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 $\omega_k \rightarrow \omega$, kada $k \rightarrow +\infty$. 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, međutim, da svi iterativni procesi sa linearnom konvergencijom ne poseduju ovu osobinu.

Teorema 2.2.2. Neka su ispunjeni uslovi teoreme 2.2.1 sa $\delta_k = \omega_k e_k$ ($\omega_k \rightarrow \omega$, kada $k \rightarrow +\infty$). Tada je

$$(2.2.8) \quad \lim_{k \rightarrow +\infty} \frac{x_k^* - a}{(x_k^* - a)^2} = \frac{C\omega}{C-1}.$$

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, $\lim_{k \rightarrow +\infty} \alpha_k = \frac{C\omega}{C-1}$, iz prethodnog neposredno sleduje (2.2.8).

Teorema 2.2.3. Ako su ispunjeni uslovi prethodne teoreme važi

$$(2.2.9) \quad \lim_{k \rightarrow +\infty} \frac{x_{k+1}^* - a}{x_k^* - a} = C^2.$$

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 $\{x_k^*\}_{k \in \mathbb{N}_0}$ sa linearnom konvergencijom, ali je njegov faktor konvergencije manji nego kod niza $\{x_k\}_{k \in \mathbb{N}_0}$ ($C^2 < |C| < 1$).

Primedba 2.2.2. Posmatrajmo iterativni proces $x_{k+1} = \phi(x_k)$ ($k = 0, 1, \dots$), gde $x_0 \in [\alpha, \beta]$. Ako za iterativnu funkciju $\phi: [\alpha, \beta] \rightarrow [\alpha, \beta]$ prepostavimo uslove:

$$1^{\circ} \quad \Phi \in C^2[\alpha, \beta],$$

$$2^{\circ} \quad |\Phi'(x)| < 1 \text{ za svako } x \in [\alpha, \beta],$$

na osnovu Taylor ove formule, imamo

$$x_{k+1} = \Phi(x_k) = a + \Phi'(a)e_k + \frac{1}{2}\Phi''(\xi_k)e_k^2,$$

$$\text{qde su } a = \lim_{k \rightarrow +\infty} x_k, \quad e_k = x_k - a \quad \text{i} \quad \xi_k = a + \theta_k(x_k - a) \quad (0 \leq \theta \leq 1).$$

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

$$\lim_{k \rightarrow +\infty} \frac{x_k^* - a}{(x_k - a)^2} = \frac{1}{2} \cdot \frac{\Phi''(a)\Phi'(a)}{\Phi'(a) - 1}.$$

Primer 2.2.1. Primenom formule (2.2.6) na niz koji se generiše pomoću procesa $x_{k+1} = \cos x_k$ ($k=0, 1, \dots$) 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 decimalne.

3.2.3. O metodima bliskim Aitkenovom metodu

Posmatrajmo iterativni proces

$$(2.3.1) \quad x_{k+1} = \phi(x_k) \quad (k=0,1,\dots).$$

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

$$(2.3.2) \quad x_k^* = \psi(x_k) \quad (k=0,1,\dots),$$

gde je

$$(2.3.3) \quad \psi(x) = \frac{x\phi^2(x) - \phi(x)^2}{\phi^2(x) - 2\phi(x) + x}.$$

Funkciju ψ , definisanu sa (2.3.3), Steffensen (videti na primer [8, 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

$$x_{k+1} = \psi(x_k) \quad (k=0,1,\dots).$$

Opštiji metod za ubrzavanje konvergencije dao je Hauseholder (videti [3]), posmatrajući dva iterativna procesa reda r

$$x_{k+1}^{(j)} = \phi_j(x_k^{(j)}) \quad (j=1,2; k=0,1,\dots),$$

koji konvergiraju ka $x=a$ i definisanjem funkcije ψ_1 pomoću

$$(2.3.4) \quad \psi_1(x) = \frac{x\phi_1(\phi_2(x)) - \phi_1(x)\phi_2(x)}{x - \phi_1(x) - \phi_2(x) + \phi_1(\phi_2(x))}.$$

Sledeći rezultat odnosi se na iterativni proces

$$(2.3.5) \quad x_{k+1} = \psi_1(x_k) \quad (k=0,1,\dots).$$

Teorema 2.3.1 (Householder). 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

$$x_{k+1}^{(j)} = \phi_j(x_k^{(j)}) \quad (j=1,\dots,m; \quad k=0,1,\dots),$$

čiji je red konvergencije r, u radu [9] dato je sledeće uopšteњe prethodnih rezultata.

Teorema 2.3.2. Ako je

$$\phi_{p,q,\dots,s,t}(x) = \phi_p(\phi_q(\dots\phi_s(\phi_t(x)))\dots),$$

$$D_n(x) = \begin{vmatrix} x & \phi_{i_1,i_2,\dots,i_n}(x) \\ \phi_j(x) & \phi_{j_1,j_2,\dots,j_{n+1}}(x) \end{vmatrix},$$

$$\Delta_n(x) = \begin{vmatrix} x - \phi_{i_1,i_2,\dots,i_n}(x) & 1 \\ \phi_j(x) - \phi_{j_1,j_2,\dots,j_{n+1}}(x) & 1 \end{vmatrix}$$

$$(j, i_\alpha, j_\beta \in \{1, 2, \dots, m\}, \alpha = 1, 2, \dots, n, \beta = 1, 2, \dots, n+1),$$

i

$$1 - \phi'_j(a) - \phi'_{i_1}(a)\phi'_{i_2}(a)\dots\phi'_{i_n}(a) + \phi'_{j_1}(a)\phi'_{j_2}(a)\dots\phi'_{j_{n+1}}(a) \neq 0,$$

iterativni proces definisan pomoću $x_{k+1} = D_n(x_k) \Delta_n^{-1}(x_k)$ ($k=0,1,\dots$),

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 $\{x_k\}_{k \in \mathbb{N}_0}$ koji konvergira ka tački a generiše pomoću

$$(2.4.1) \quad x_{k+1} = \phi(x_k) \quad (k = 0, 1, \dots).$$

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 $T:X \rightarrow X$ i neka niz koji se generiše pomoću

$$(2.4.2) \quad u_{k+1} = Tu_k \quad (k = 0, 1, \dots)$$

konvergira ka $a \in X$. Jednakost (2.4.2) definiše iterativni proces za rešavanje operatorske jednačine $u = Tu$. Sa $U(a)$ označimo konveksnu okolinu granične tačke a .

U pomenutom redu [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 $T:X \rightarrow X$ ($r+1$) - puta Fréchet-diferencijabilan u okolini $U(a)$ i neka postoji inverzan operator $[I - \frac{1}{r}T'(u)]^{-1}$ kada $u \in U(a)$. Tada je

$$(2.4.3) \quad u_{k+1} = Gu_k = u_k - \left[I - \frac{1}{r}T'(u_k) \right]^{-1} (u_k - Tu_k)$$

iterativni proces najmanje reda $r+1$.

Osnovni nedostatak metoda, definisanog poslednjom teoremom, je što zahteva nalaženje inverznog operatora $[I - \frac{1}{r}T'(u)]^{-1}$, što je u većini slučajeva vrlo složeno.

Sada ćemo izložiti jedan metod za ubrzavanje konvergencije iterativnih procesa, koji ne zahteva nalaženje inverznog operatora (videti [6]). 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 $T:X \rightarrow X$ ($r+1$) - puta Fréchet-diferencijabilan u okolini $U(a)$. Tada je

$$(2.4.4) \quad u_{k+1} = Fu_k = Tu_k - \frac{1}{r} T'(u_k) (u_k - Tu_k) \quad (k = 0, 1, \dots)$$

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

$$(2.4.5) \quad T'(u) = \frac{1}{(r-1)!} T^{(r)}(a) \underbrace{(u-a, \dots, u-a)}_{r-1 \text{ puta}} + W(a, u-a),$$

$$(2.4.6) \quad Tu = a + \frac{1}{r!} T^{(r)}(a) \underbrace{(u-a, \dots, u-a)}_{r \text{ puta}} + w(a, u-a),$$

$$\|w(a, u-a)\| = O(\|u-a\|^r) \quad \text{i} \quad \|w(a, u-a)\| = O(\|u-a\|^{r+1}).$$

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

$$\|Fu-a\| \leq \|w(a, u-a)\| + \frac{1}{r} \|w(a, u-a)\| \cdot \|u-a\| + \frac{1}{r} \|T'_{(u)}\| \cdot \|Tu-a\|.$$

Kako je $\|T'_{(u)}\| = O(\|u-a\|^{r-1})$ i $\|Tu-a\| = O(\|u-a\|^r)$

pri $r \geq 2$, 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

$$G_u - a = w(a, u-a) - \frac{1}{r} W(a, u-a)(u-a) + \frac{1}{r} T'(u)(G_u - a).$$

Iz poslednje jednakosti sledi

$$(1 - \frac{1}{r} \|T'(u)\|) \|G_u - a\| \leq \|w(a, u-a)\| + \frac{1}{r} \|W(a, u-a)\| \cdot \|u-a\|,$$

tj.

$$\|G_u - a\| = O(\|u-a\|^{r+1}),$$

čime je dokaz završen.

Teorema 2.4.4. Neka je (2.4.1) iterativni proces sa konvergencijom reda $r (\geq 2)$ i funkcija $\phi(r+1)$ -puta diferencijabilna u okolini granične tačke a . Tada je

$$x_{k+1} = \phi(x_k) - \frac{1}{r} \phi'(x_k)(x_k - \phi(x_k)) \quad (k = 0, 1, \dots)$$

iterativni proces najmanje reda $r+1$.

Ova teorema je posledica teoreme 2.4.3.

3.2.5. R-red konvergencije iterativnih procesa

U ovom odeljku ukazaćemo na jedan poseban tretman konvergencije iterativnih procesa u prostoru $X = \mathbb{R}^n$, koji je uveden od Ortege i Rheinboldta ([7]).

Definicija 2.5.1. Neka je $\{\vec{x}^{(k)}\}$ proizvoljan niz u \mathbb{R}^n koji konvergira ka \vec{a} . Tada se brojevi

$$R_p \{\vec{x}^{(k)}\} = \begin{cases} \limsup_{k \rightarrow +\infty} \|\vec{x}^{(k)} - \vec{a}\|^{1/k}, & \text{ako je } p=1, \\ \limsup_{k \rightarrow +\infty} \|\vec{x}^{(k)} - \vec{a}\|^{1/p^k}, & \text{ako je } p > 1, \end{cases}$$

nazivaju faktori konvergencije po korenu, ili kraće R-faktori niza $\{\vec{x}^{(k)}\}$. Ako je IP iterativni proces sa graničnom tačkom \vec{a} i $C(IP, \vec{a})$ skup svih nizova generisanih pomoću IP, koji konvergiraju ka \vec{a} , tada se veličina

$$R_p(IP, \vec{a}) = \sup \{ R_p \{\vec{x}^{(k)}\} \mid \{\vec{x}^{(k)}\} \in C(IP, \vec{a}) \} \quad (1 \leq p < +\infty)$$

naziva R - faktorom iterativnog procesa u tački \vec{a} .

Primedba 2.5.1. Ako niz $\{\vec{x}^{(k)}\}$ konvergira ka \vec{a} , tada postoji k_0 takvo da je

$$0 \leq \|\vec{x}^{(k)} - \vec{a}\| \leq 1 \quad \text{za svako } k \geq k_0,$$

odakle zaključujemo da je $0 \leq R_p \{\vec{x}^{(k)}\} \leq 1$ za svako $p \geq 1$.

Sada navodimo bez dokaza dve teoreme:

Teorema 2.5.1. Neka je $\{\vec{x}^{(k)}\}$ proizvoljan niz u R^n koji konvergira ka \vec{a} . Faktor $R_p \{\vec{x}^{(k)}\}$ ne zavisi od izbora norme u R^n ni za jedno $p \in [1, +\infty)$. Takodje, R - faktor $R_p (IP, \vec{a})$ iterativnog procesa je nezavisan od izbora norme.

Teorema 2.5.2. Neka je IP iterativni proces sa graničnom tačkom \vec{a} . Tada važi jedan od sledećih uslova:

- a) $R_p (IP, \vec{a}) = 0$ za svako $p \in [1, +\infty)$;
- b) $R_p (IP, \vec{a}) = 1$ za svako $p \in [1, +\infty)$;
- c) Postoji $p_0 \in [1, +\infty)$ takvo da je $R_p (IP, \vec{a}) = 0$ za svako $p \in [1, p_0)$ i $R_p (IP, \vec{a}) = 1$ za svako $p \in [p_0, +\infty)$.

Na osnovu prethodnog može se uvesti tzv. R - red konvergencije iterativnih procesa:

Definicija 2.5.2. R - red konvergencije iterativnog procesa u tački \vec{a} je veličina

$$\Omega_R (IP, \vec{a}) = \begin{cases} +\infty & (\text{ako je } R_p (IP, \vec{a}) = 0 \text{ za svako } p \in [1, +\infty)), \\ \inf \{p \in [1, +\infty) \mid R_p (IP, \vec{a}) = 1\} & (\text{u ostalim slučajevima}). \end{cases}$$

Primetimo da ovako definisani R - red konvergencije iterativnog procesa ne zavisi od izbora norme u R^n . Takodje, primetimo sledeće činjenice:

- Ako je $R_p (IP, \vec{a}) < 1$ za neko $p \in [1, +\infty)$, tada je R - red ne manji od p, tj. važi $\Omega_R (IP, \vec{a}) \geq p$;
- Ako je $R_q (IP, \vec{a}) > 0$ za neko $q \in [1, +\infty)$, tada za R - red važi $\Omega_R (IP, \vec{a}) \leq q$;
- Ako je $0 < R_p (IP, \vec{a}) < 1$ za neko $p \in [1, +\infty)$, tada je $\Omega_R (IP, \vec{a}) = p$.

U slučaju kada je $0 < R_1 (IP, \vec{a}) < 1$ kažemo da je konver-

gencija procesa u tački \vec{a} R-linearna. Ako je, međutim, $R_1(IP, \vec{a}) = 0$ ili $R_1(IP, \vec{a}) = 1$, za konvergenciju kažemo da je R-podlinearna, odnosno R-nadlinearna.

Kada uporedjujemo konvergenciju dva iterativna procesa IP_1 i IP_2 postupamo na sledeći način: Najpre, uporedimo R-red, tj. veličine $O_R(IP_1, \vec{a})$ i $O_R(IP_2, \vec{a})$, pri čemu je brži onaj proces koji ima veći R-red. Međutim, ako je $O_R(IP_1, \vec{a}) = O_R(IP_2, \vec{a}) = \bar{p}$, tada uporedjujemo R-faktore za $p = \bar{p}$. Ako je, na primer, $R_{\bar{p}}(IP_1, \vec{a}) < R_{\bar{p}}(IP_2, \vec{a})$ tada je IP_2 brži od IP_1 .

3.3. LITERATURA

1. L.Collatz: Functional analysis und numerische Mathematik. Springer Verlag, Berlin-Heidelberg-New York, 1968.
2. O.Hadžić: Fixed point theory in topological vector spaces. Univ. Novi Sad, Faculty of Science Institute of Mathematics, Novi Sad, 1984.
3. A.S.Hausholder: Principles of numerical analysis. Dover, New York, 1953.
4. B.Jovanović: A method for obtaining iterative formulas of higher order. Mat. Vesnik 9(24) (1972), 365-369.
5. L.V.Kantorović i G.P.Akilov: Funkcional'nyi analiz. Nauka, Moskva, 1977.
6. G.V.Milovanović: A method to accelerate iterative processes in Banach space. Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz. N-461 - N-497 (1974), 67-71.
7. J.M.Ortega & W.C.Rheinboldt: Iterative solution of nonlinear equations in several variables. Academic Press, New York, 1970.
8. A.Ostrowski: Solution of equations and systems of equations. Academic Press, New York, 1966.
9. A.V.Prokopčenko: Ob iteracionnyh processah vysših porjadkov. Ž. Vychisl. Mat. i Mat. Fiz. 14 (1974), 230-233.
10. J.Stoer: Einführung in die Numerische Mathematik I. Springer Verlag, Berlin-Heidelberg-New York, 1972.
11. D.Tošić: Uvod u numeričku analizu. Naučna knjiga, Beograd, 1982.
12. J.F.Traub: Iterative methods for the solution of equations. Prentice-Hall, New Jersey, 1964.

Numerički metodi u linearnoj algebri

4.1. ELEMENTI Matričnog računa

U ovom poglavlju se daju osnovni elementi matričnog računa koji su neophodni za praćenje izlaganja u sledećim poglavljima.

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 razbijena na blokove.

Blokovi su matrice A_{ij} tipa $m_i \times n_j$, $\sum_{i=1}^p m_i = m$, $\sum_{j=1}^q n_j = n$.

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.

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1q} \\ \vdots & & \\ A_{pl} & & A_{pq} \end{bmatrix} \quad \text{i} \quad B = \begin{bmatrix} B_{11} & \cdots & B_{1q} \\ \vdots & & \\ B_{pl} & & B_{pq} \end{bmatrix},$$

gde su A_{ij} i B_{ij} matrice istog tipa. Tada je

$$cA = \begin{bmatrix} cA_{11} & \cdots & cA_{1q} \\ \vdots & & \vdots \\ cA_{p1} & & cA_{pq} \end{bmatrix} \quad i \quad A + B = \begin{bmatrix} A_{11} + B_{11} & \cdots & A_{1q} + B_{1q} \\ \vdots & & \vdots \\ A_{p1} + B_{p1} & & A_{pq} + B_{pq} \end{bmatrix},$$

gde je c skalar.

Teorema 1.1.2. Neka je

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1q} \\ \vdots & & \vdots \\ A_{p1} & & A_{pq} \end{bmatrix} \quad i \quad B = \begin{bmatrix} B_{11} & \cdots & B_{1s} \\ \vdots & & \vdots \\ B_{q1} & & B_{qs} \end{bmatrix},$$

gde je broj kolona bloka A_{ij} jednak broju vrsta bloka B_{jk} ($i=1, \dots, p$; $j=1, \dots, q$; $k=1, \dots, s$). Tada je

$$AB = \begin{bmatrix} C_{11} & \cdots & C_{1s} \\ \vdots & & \vdots \\ C_{p1} & & C_{ps} \end{bmatrix},$$

gde je $C_{ik} = \sum_{j=1}^q A_{ij}B_{jk}$ ($i=1, \dots, p$; $k=1, \dots, s$).

Primer 1.1.1. Neka su

$$A = \begin{bmatrix} 5 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 8 & 3 \\ 0 & 0 & 5 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -2 & -1 & 0 & 0 \\ -2 & 5 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & -3 \\ 0 & 0 & 0 & -5 & 8 \end{bmatrix}$$

Ako stavimo

$$A_{11} = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix}, \quad A_{22} = \begin{bmatrix} 8 & 3 \\ 5 & 2 \end{bmatrix}, \quad B_{11} = \begin{bmatrix} 1 & -2 & -1 \\ -2 & 5 & 2 \end{bmatrix}, \quad B_{22} = \begin{bmatrix} 2 & -3 \\ -5 & 8 \end{bmatrix},$$

imamo

$$AB = \begin{bmatrix} A_{11}B_{11} & 0 \\ 0 & A_{22}B_{22} \end{bmatrix}.$$

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

$$AB = \begin{bmatrix} 1 & 0 & -1 & | & 0 & 0 \\ 0 & -1 & 0 & | & 0 & 0 \\ 0 & 0 & 0 & | & 1 & 0 \\ 0 & 0 & 0 & | & 0 & 1 \end{bmatrix}.$$

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,

$$A^{-1} = \left[\begin{array}{cc|cc} -1 & 3 & -7 & 20 \\ -7 & -3 & 5 & -10 \\ \hline 9 & 3 & -3 & 6 \\ 3 & 3 & -3 & 6 \end{array} \right]$$

Definicija 1.1.2. Kvadratna matrica A ima svojstvo (A) ako se permutacijom vrsta i kolona može svesti na oblik

$$\left[\begin{array}{ccccccc} D_1 & F_1 & 0 & 0 & \dots & 0 & 0 \\ E_1 & D_2 & F_2 & 0 & & 0 & 0 \\ 0 & E_2 & D_3 & F_3 & & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & 0 & & E_{m-1} & D_m \end{array} \right],$$

gde su D_j ($j=1, \dots, m$) dijagonalne matrice, a E_j i F_j ($j = 1, \dots, m-1$) pravougaone.

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

$$\Delta_k = \begin{vmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \\ a_{kk} & & a_{kk} \end{vmatrix} \quad (k = 1, \dots, n-1)$$

različite od nule, matrica $A = [a_{ij}]_{n \times n}$ može se predstaviti u obliku

$$(1.2.1) \quad A = LR,$$

gde je L donja i R gornja trougaona matrica.

Trougaone matrice L i R reda n, imaju oblike:

$$(1.2.2) \quad L = [\ell_{ij}]_{n \times n} \quad (\ell_{ij} = 0 \leq i < j),$$

$$(1.2.3) \quad R = [r_{ij}]_{n \times n} \quad (r_{ij} = 0 \leq i > j).$$

Razlaganje (1.2.1), poznato kao LR faktorizacija (dekompozicija), nije jedinstveno, s obzirom na jednakost

$$LR = (cL) \left(\frac{1}{c}R\right) \quad (\forall c \neq 0).$$

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

$$a_{ij} = \sum_{k=1}^{\min(i,j)} \ell_{ik} r_{kj} \quad (i, j = 1, \dots, n)$$

elementi matrica L i R mogu se lako odrediti rekurzivnim postupkom, ukoliko se unapred zadaju elementi $r_{ii} (\neq 0)$ ili $\ell_{ii} (\neq 0)$ ($i = 1, \dots, n$).

Tako, na primer, neka su dati brojevi r_{ii} ($\neq 0$) ($i=1, \dots, n$).

Tada važi

$$(1) \quad \ell_{11} = \frac{a_{11}}{r_{11}},$$

$$\left. \begin{array}{l} r_{li} = \frac{a_{li}}{\ell_{11}} \\ \ell_{il} = \frac{a_{il}}{r_{11}} \end{array} \right\} (i = 2, \dots, n);$$

$$(i) \quad \left. \begin{array}{l} \ell_{ii} = \frac{1}{r_{ii}} (a_{ii} - \sum_{k=1}^{i-1} \ell_{ik} r_{ki}), \\ r_{ij} = \frac{1}{\ell_{ii}} (a_{ij} - \sum_{k=1}^{i-1} \ell_{ik} r_{kj}) \\ \ell_{ji} = \frac{1}{r_{ii}} (a_{ji} - \sum_{k=1}^{i-1} \ell_{jk} r_{ki}) \end{array} \right\} (j=i+1, \dots, n) \quad \left. \right\} (i=2, \dots, n).$$

Slično bismo mogli iskazati i rekursivni postupak za odredjivanje elemenata matrica L i R ako su unapred dati brojevi ℓ_{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 rekursivnog postupka, imamo redom:

$$(1) \quad \ell_{11} = 1,$$

$$r_{12} = 4, \quad \ell_{21} = 0, \quad r_{13} = 1, \quad \ell_{31} = 3, \quad r_{14} = 3, \quad \ell_{41} = 1;$$

$$(2) \quad \ell_{22} = -1$$

$$r_{23} = -2, \quad \ell_{32} = 2, \quad r_{24} = 1, \quad \ell_{42} = -2;$$

$$(3) \quad \ell_{33} = 5,$$

$$r_{34} = -2, \quad \ell_{43} = -3;$$

$$(4) \quad \ell_{44} = 2.$$

Dakle, dobili smo

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 3 & 2 & 5 & 0 \\ 1 & -2 & -3 & 2 \end{bmatrix} \quad i \quad R = \begin{bmatrix} 1 & 4 & 1 & 3 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

b) Polazeći od $\ell_{ii} = 1$ ($i = 1, \dots, 4$), imamo redom

$$(1) \quad r_{11} = 1,$$

$$r_{12} = 4, \quad \ell_{21} = 0, \quad r_{13} = 1, \quad \ell_{31} = 3, \quad r_{14} = 3, \quad \ell_{41} = 1;$$

$$(2) \quad r_{22} = -1,$$

$$r_{23} = 2, \quad \ell_{32} = -2, \quad r_{24} = -1, \quad \ell_{42} = 2;$$

$$(3) \quad r_{33} = 5,$$

$$r_{34} = -10, \quad \ell_{43} = -3/5;$$

$$(4) \quad r_{44} = 2,$$

tj.

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 3 & -2 & 1 & 0 \\ 1 & 2 & -3/5 & 1 \end{bmatrix} \quad i \quad R = \begin{bmatrix} 1 & 4 & 1 & 3 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & 5 & -10 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

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 $|i-j| \leq 1$ i $a_{ij} = 0$ za $|i-j| > 1$, matrica je trodijagonalna. Obično elemente ovakve matrice predstavljamo vektorima (a_2, \dots, a_n) , (b_1, \dots, b_n) , (c_1, \dots, c_{n-1}) , tj.

$$(1.2.4) \quad A = \begin{bmatrix} b_1 & c_1 & 0 & \dots & 0 & 0 \\ a_2 & b_2 & c_2 & 0 & 0 & 0 \\ 0 & a_3 & b_3 & 0 & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & a_n & b_n & \end{bmatrix}.$$

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

$$L = \begin{bmatrix} \beta_1 & 0 & 0 & \dots & 0 & 0 \\ \alpha_2 & \beta_2 & 0 & & 0 & 0 \\ 0 & \alpha_3 & \beta_3 & & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & & \alpha_n & \beta_n \end{bmatrix} \quad (\beta_1 \beta_2 \dots \beta_n \neq 0)$$

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 određivanje elemenata α_i ,

β_i , γ_i :

$$\beta_1 = b_1, \quad \gamma_1 = \frac{c_1}{\beta_1},$$

$$(1.2.5) \quad \alpha_i = a_i, \quad \beta_i = b_i - \alpha_i \gamma_{i-1}, \quad \gamma_i = \frac{c_i}{\beta_i} \quad (i=2, \dots, n-1),$$

$$\alpha_n = a_n, \quad \beta_n = b_n - \alpha_n \gamma_{n-1}.$$

4.1.3. Sopstveni vektori i sopstvene vrednosti matrica

Definicija 1.3.1. Neka je A kompleksna kvadratna matrica reda n . Svaki vektor $\vec{x} \in \mathbb{C}^n$, koji je različit od nula-vektora, naziva se sopstveni vektor matrice A ako postoji skalar $\lambda \in \mathbb{C}$ takav da je

$$(1.3.1) \quad A\vec{x} = \lambda \vec{x} .$$

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 netrivialna 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 $A = [a_{ij}]_{n \times n}$. Karakteristični polinom može se predstaviti u obliku

$$P(\lambda) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & & a_{nn} - \lambda \end{vmatrix}$$

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.

$$p_k = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \det(A_{i_1 i_2 \dots i_k}^{i_1 i_2 \dots i_k}).$$

Primetimo da je

$$p_1 = \sum_{i=1}^n a_{ii} = \text{tr } A \quad i \quad p_n = \det(A).$$

Često se umesto karakterističnog polinoma P koristi tzv. normirani karakteristični polinom H, definisan pomoću

$$H(\lambda) = (-1)^n P(\lambda) = \lambda^n - p_1 \lambda^{n-1} + p_2 \lambda^{n-2} - \dots + (-1)^n p_n.$$

Sopstvene vrednosti matrice A (tj. nule polinoma P) λ_i ($i=1, \dots, n$) označavaćemo sa $\lambda_i(A)$.

Definicija 1.3.3. Skup svih sopstvenih vrednosti kvadratne matrice A naziva se spektrom te matrice i označava se sa $\text{Sp}(A)$.

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 $\lambda_1, \dots, \lambda_n$ sopstvene vrednosti matrice A reda n i neka je $x \mapsto Q(x)$ skalarni polinom stepena m. Tada su

$$Q(\lambda_1), \dots, Q(\lambda_n)$$

sopstvene vrednosti matrice $Q(A)$.

Teorema 1.3.3. Neka su $\lambda_1, \dots, \lambda_n$ sopstvene vrednosti regularne matrice A reda n. Tada su

$$\lambda_1^{-1}, \dots, \lambda_n^{-1}$$

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

$$A_k = \begin{bmatrix} b_1 & c_1 & 0 & \dots & 0 \\ a_2 & b_2 & c_2 & & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & & b_k \end{bmatrix} \text{ i } H_k(\lambda) = (-1)^k \det(A_k - \lambda I).$$

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 $H_0(\lambda) = 1$ 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

$$B = C^{-1} A C.$$

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 $A = [a_{ij}]_{m \times n}$ ($a_{ij} \in \mathbb{C}$). Sa \bar{A} označićemo konjugovanu matricu matrici A , tj. $\bar{A} = [\bar{a}_{ij}]_{m \times n}$, 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

$$(A\vec{x}, \vec{y}) = (\vec{x}, A\vec{y}) \quad (\forall \vec{x}, \vec{y} \in \mathbb{C}^n).$$

Slično se uslov za kosohermitsku matricu ($A = -A^*$) može predstaviti u obliku

$$(A\vec{x}, \vec{y}) + (\vec{x}, A\vec{y}) = 0 \quad (\forall \vec{x}, \vec{y} \in \mathbb{C}^n).$$

Definicija 1.4.6. Hermitska matrica A naziva se pozitivno definitna ako je za svako $\vec{x} \neq \vec{0}$ ispunjen uslov

$$(1.4.1) \quad (A\vec{x}, \vec{x}) = \vec{x}^* A \vec{x} > 0.$$

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 Sylvesterovi uslovi

$$\Delta_k = \begin{vmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \\ a_{k1} & & a_{kk} \end{vmatrix} > 0 \quad (k = 1, \dots, n).$$

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

$$1^\circ \quad \lambda(A^*A) = \lambda(A)^2,$$

$$2^\circ \quad \lambda_{\min}(A)(\vec{x}, \vec{x}) \leq (A\vec{x}, \vec{x}) \leq \lambda_{\max}(A)(\vec{x}, \vec{x}).$$

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

$$(1.5.1) \quad J_r(\lambda) = \begin{cases} [\lambda] & (r=1), \\ \begin{bmatrix} \lambda & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & 0 & & 0 & 0 \\ 0 & 0 & \lambda & 1 & & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & 0 & & \lambda & 1 \\ 0 & 0 & 0 & 0 & & 0 & \lambda \end{bmatrix} & (r \geq 2), \end{cases}$$

naziva se Jordanov blok.

Definicija 1.5.2. Kvazidijagonalna matrica oblika

$$(1.5.2) \quad J = \begin{bmatrix} J_{r_1}(\lambda) & 0 & \dots & 0 \\ 0 & J_{r_2}(\lambda) & & 0 \\ \vdots & & & \\ 0 & 0 & & J_{r_m}(\lambda) \end{bmatrix}$$

naziva se Jordanova matrica.

Jordanova matrica često se predstavlja i u obliku

$$(1.5.3) \quad J = J_{r_1}(\lambda_1) + J_{r_2}(\lambda_2) + \dots + J_{r_m}(\lambda_m).$$

Pod Jordanovim kanoničkim oblikom matrice A podrazumeva se ona Jordanova matrica koja je slična matrici A i njena egzistencija sledi iz sledeće teoreme (videti, na primer, [42]).

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 $x \mapsto Q(x) = a_0 x^k + a_1 x^{k-1} + \dots + a_k$ 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

$$Q(J) = Q(J_{r_1}(\lambda_1)) + Q(J_{r_2}(\lambda_2)) + \dots + Q(J_{r_m}(\lambda_m)).$$

Teorema 1.5.2. Za proizvoljnu hermitsku matricu A, postoji unutrašnja 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 \mathbb{R}^n (kompleksan prostor \mathbb{C}^n) sa nula-vektorom $\vec{0}([0 \dots 0]^T)$. Za vektor $\vec{x}([x_1 \dots x_n]^T)$ definiše se norma saglasno definiciji 1.2.1 (odeljak 2.1.2). Od svih mogućih normi vektora od interesa su norme oblika

$$(1.6.1) \quad \|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p} \quad (1 \leq p < +\infty).$$

U graničnom slučaju, kada $p \rightarrow +\infty$, iz (1.6.1) sleduje

$$(1.6.2) \quad \|\vec{x}\|_\infty = \max_i |x_i|.$$

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.

$$(1.6.3) \quad \|\vec{x}\|_1 = \sum_{i=1}^n |x_i|$$

$$(1.6.4) \quad \|\vec{x}\|_2 = \|\vec{x}\|_E = \left(\sum_{i=1}^n |x_i|^2 \right)^{1/2}$$

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 nejednosti

$$\|\vec{x}\|_\infty \leq \|\vec{x}\|_1 \leq n \|\vec{x}\|_\infty,$$

$$\|\vec{x}\|_\infty \leq \|\vec{x}\|_2 \leq \sqrt{n} \|\vec{x}\|_\infty,$$

$$\frac{1}{\sqrt{n}} \|\vec{x}\|_1 \leq \|\vec{x}\|_2 \leq \|\vec{x}\|_1.$$

Neka je X_M linearan realan ili kompleksan prostor kvadratnih matrica reda n , sa nula-matricom $0 \in X_M$.

Definicija 1.6.1. Pod normom matrice $A \in X_M$ podrazumeva se nenegativan broj $\|A\|$, takav da je

$$1^\circ \quad \|A\| = 0 \Leftrightarrow A = 0 \quad (\text{definisanost}),$$

$$2^{\circ} \|cA\| = |c| \cdot \|A\| \quad (\text{homogenost}),$$

$$3^{\circ} \|A+B\| \leq \|A\| + \|B\| \quad (\text{relacija trougla}),$$

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.

Definicija 1.6.2. Norma $\|A\|$ saglasna je sa normom $\|\vec{x}\|$, ako je

$$\|A\vec{x}\| \leq \|A\| \cdot \|\vec{x}\| \quad (\forall A \in X_M \wedge \forall \vec{x} \in X)$$

i

$$\|AB\| \leq \|A\| \cdot \|B\| \quad (\forall A, B \in X_M).$$

Definicija 1.6.3. Neka je norma $\|A\|$ saglasna sa normom $\|\vec{x}\|$. Ako se za svako $A \in X_M$ može naći $\vec{x} (\neq \vec{0})$, takvo da je $\|A\vec{x}\| = \|A\| \cdot \|\vec{x}\|$, 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, [14]) da za svaku normu vektora $\|\vec{x}\|$ postoji bar jedna potčinjena norma $\|A\|$. Na primer, norma matrice A , definisana pomoću

$$(1.6.5) \quad \|A\| = \max_{\|\vec{x}\|=1} \|A\vec{x}\| = \max_{\vec{x} \neq \vec{0}} \frac{\|A\vec{x}\|}{\|\vec{x}\|},$$

potčinjena je upotrebljenoj normi vektora.

Na osnovu (1.6.5) možemo, na primer, naći potčinjenu normu matrice, za Euklidsku normu vektora. Tako imamo

$$\|A\| = \max_{\vec{x} \neq \vec{0}} \frac{\|A\vec{x}\|_E}{\|\vec{x}\|_E}.$$

Kako je $\|Ax\|_E^2 = (\vec{Ax}) * (\vec{Ax}) * (\vec{Ax}) = \vec{x} * A^* A \vec{x}$ i matrica $A^* A$ hermitska, imamo

$$(1.6.6) \quad \|Ax\|_E^2 \leq \lambda_{\max} \|\vec{x}\|_E^2,$$

gde je λ_{\max} najveća sopstvena vrednost matrice $A^* A$ (videti teoremu 1.4.6). Nejednakost (1.6.6) sugerije 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

$$\sigma(A^{-1}) = +\sqrt{\frac{1}{\min \lambda(A^*A)}}.$$

Teorema 1.6.2. Spektralni radijus $\rho(A)$ matrice A nije veći od njene spektralne norme, tj. važi $\rho(A) \leq \sigma(A)$. 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 mogućih normi $\|A\|$, saglasnih sa nekom normom vektora.

Pored spektralne norme u upotrebi su i sledeće norme matrice $A = [a_{ij}]_{n \times n}$:

$$1^o \quad \|A\|_1 = \max_j \left(\sum_{i=1}^n |a_{ij}| \right),$$

$$2^o \quad \|A\|_2 = \epsilon(A) = \left(\sum_{i,j=1}^n |a_{ij}|^2 \right)^{1/2} \text{ (norma Schmidta)},$$

$$3^o \quad \|A\|_\infty = \max_i \left(\sum_{j=1}^n |a_{ij}| \right).$$

Za norme 1^o, 2^o i 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 $\epsilon(A)$ za matricu reda n saglasna je sa Euklidskom normom (1.6.4), ali joj nije potčinjena za $n > 1$.

Teorema 1.6.7. Norma $\|A\|$ potčinjena je normi vektora $\|\vec{x}\|_\infty$.

Lako se može dokazati da za proizvoljnu kvadratnu maticu reda n važe sledeće nejednakosti

$$\frac{1}{n}m(A) \leq \|A\|_p \leq m(A) \quad (p = 1, 2, \infty),$$

$$\frac{1}{\sqrt{n}} \epsilon(A) \leq \|A\|_p \leq \sqrt{n} \epsilon(A) \quad (p = 1, \infty),$$

$$\frac{1}{\sqrt{n}} \sigma(A) \leq \|A\|_p \leq \sqrt{n} \sigma(A) \quad (p = 1, \infty),$$

$$m(A) \leq n \sigma(A), \quad \epsilon(A) \leq \sqrt{n} \sigma(A),$$

$$\frac{1}{n} \|A\|_\infty \leq \|A\|_1 \leq n \|A\|_\infty,$$

gde je $m(A) = n \max_{i,j} |a_{ij}|$.

Neka je $A = [a_{ij}]_{n \times n}$ data kvadratna matrica i $\vec{b} = [b_1 \dots b_n]^T$ dati n -dimenzionalni vektor. Sledeća teorema se odnosi na sistem linearnih jednačina $A\vec{x} = \vec{b}$.

Teorema 1.6.8. Ako je

$$(1.6.7) \quad \min_{1 \leq i \leq n} \{ |a_{ii}| - \sum_{j \neq i} |a_{ij}| \} = q > 0$$

tada sistem jednačina $A\vec{x} = \vec{b}$ ima jedinstveno rešenje i važi nejednakost

$$(1.6.8) \quad \|\vec{x}\|_\infty \leq \frac{1}{q} \|\vec{b}\|_\infty$$

Dokaz. Neka je ispunjen uslov (1.6.7) i neka je $\|\vec{x}\|_\infty = \max_{1 \leq i \leq n} |x_i| = |x_m|$. Tada imamo

$$\|\vec{b}\|_\infty \geq |a_{mm}| \cdot |x_m| - \sum_{j \neq m} |a_{mj}| \cdot |x_m| \geq q \|\vec{x}\|_\infty,$$

što daje (1.6.8). Da je rešenje \vec{x} jedinstveno sleduje iz (1.6.8).

Zaista, ako je \vec{x} nula vektor, na osnovu (1.6.8) zaključujemo da je $\|\vec{x}\|_\infty = 0$, tj. $\vec{x} = \vec{0}$, što znači da sistem ima samo trivijalno rešenje. Odavde, dalje, zaključujemo da je determinanta sistema različita od nule, pa sistem ima jedinstveno rešenje za bilo koji vektor \vec{b} .

Ako je za matricu A ispunjen uslov (1.6.7) kažemo da matrica ima dominantnu dijagonalu.

U nejednakosti (1.6.8) norma $\|\cdot\|_\infty$ može biti zamjenjena normom $\|\cdot\|_1$ ili $\|\cdot\|_2$, ali u tim slučajevima i uslov (1.6.7) treba biti zamjenjen za

$$\min_{1 \leq i \leq n} |a_{ii}| - \max_{1 \leq j \leq n} \left\{ \sum_{i \neq j} |a_{ij}| \right\} = q > 0$$

ili

$$\min_{1 \leq i \leq n} |a_{ii}| - \max_{1 \leq j \leq n} \left(\sum_{i \neq j} |a_{ij}| \right)^{1/2} \min_{1 \leq i \leq n} \left(\sum_{j \neq i} |a_{ij}| \right)^{1/2} = q > 0,$$

respektivno.

Pomoću regularne matrice H, moguće 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.

$$H = \begin{bmatrix} h_1 & & & \\ & \ddots & & 0 \\ & 0 & \ddots & \\ & & & h_n \end{bmatrix} = \text{diag}(h_1, \dots, h_n) \quad (h_i > 0 \quad (i = 1, \dots, n)).$$

Na ovaj način, norme $\|\vec{x}\|_1$, $\|A\|_1$ i norme $\|\vec{x}\|_\infty$, $\|A\|_\infty$ transformišu se u

$$\|\vec{x}\|_1^H = \sum_{i=1}^n h_i^{-1} |x_i|, \quad \|A\|_1^H = \max h_j \left(\sum_{i=1}^n h_i^{-1} |a_{ij}| \right)$$

i

$$\|\vec{x}\|_{\infty}^H = \max_i h_i^{-1} |x_i|, \|A\|_{\infty}^H = \max_i h_i^{-1} \left(\sum_{j=1}^n h_j |a_{ij}| \right)$$

respektivno.

4.1.7. Konvergencija matričnih nizova i redova

Posmatrajmo niz vektora $\{\vec{x}^{(k)}\}_{k \in \mathbb{N}}$, gde je

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

Definicija 1.7.1. Ako postoje konačne granične vrednosti

$$a_i = \lim_{k \rightarrow +\infty} x_i^{(k)} \quad (i = 1, \dots, n),$$

kažemo da niz $\{\vec{x}^{(k)}\}_{k \in \mathbb{N}}$ konvergira ka vektoru $\vec{a} = [a_1 \dots a_n]^T$, a vektor \vec{a} nazivamo graničnom vrednošću niza $\{\vec{x}^{(k)}\}_{k \in \mathbb{N}}$.

Slično definišemo i konvergenciju niza matrica $\{A^{(k)}\}_{k \in \mathbb{N}}$ gde je $A^{(k)} = [a_{ij}^{(k)}]_{n \times n}$.

Definicija 1.7.2. Ako postoje konačne granične vrednosti

$$a_{ij} = \lim_{k \rightarrow +\infty} a_{ij}^{(k)} \quad (i, j = 1, \dots, n),$$

kažemo da niz $\{A^{(k)}\}_{k \in \mathbb{N}}$ konvergira ka matrici $A = [a_{ij}]_{n \times n}$. Matricu A nazivamo graničnom vrednošću niza $\{A^{(k)}\}_{k \in \mathbb{N}}$.

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 $\vec{x}^{(k)} \rightarrow \vec{a}$ ($k \rightarrow +\infty$), ako $\|\vec{x}^{(k)} - \vec{a}\| \rightarrow 0$, kada $k \rightarrow +\infty$. Slično, kažemo da $A^{(k)} \rightarrow A$ ($k \rightarrow +\infty$), ako $\|A^{(k)} - A\| \rightarrow 0$, kada $k \rightarrow +\infty$. Može se, međutim, pokazati da je definicija konvergencije po normi ekvivalentna prethodnoj datoј definiciji po koordinatama (videti, na primer, [36]).

Primedba 1.7.1. Na osnovu nejednakosti

$$\left| \|\vec{x}^{(k)}\| - \|\vec{a}\| \right| \leq \|\vec{x}^{(k)} - \vec{a}\|$$

zaključujemo da iz $\vec{x}^{(k)} \rightarrow \vec{a} (k \rightarrow +\infty)$ sleduje $\|\vec{x}^{(k)}\| \rightarrow \|\vec{a}\| (k \rightarrow +\infty)$.
Takođe važi

$$A^{(k)} \rightarrow A (k \rightarrow +\infty) \Rightarrow \|A^{(k)}\| \rightarrow \|A\| (k \rightarrow +\infty).$$

Teorema 1.7.1. Niz matrica $\{A^k\}_{k \in \mathbb{N}}$ konvergira ka nula matrici 0 , 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 $k \in \mathbb{N}$,

$$A^k = CJ^kC^{-1},$$

odakle zaključujemo da niz $\{A^k\}_{k \in \mathbb{N}}$ konvergira (ne konvergira) ka nula matrici ako i samo ako niz $\{J^k\}_{k \in \mathbb{N}}$ konvergira (ne konvergira) ka nula matrici.

Neka su $\lambda_1, \dots, \lambda_m (m \leq n)$ medjusobno različite sopstvene vrednosti matrice A (matrice J). Kako na osnovu primedbe 1.5.1 (sa $Q(x) = x^k$) važe ekvivalencije

$$\lim_{k \rightarrow +\infty} J_r(\lambda)^k = 0 \Leftrightarrow |\lambda| < 1$$

i

$$\lim_{k \rightarrow +\infty} J^k = 0 \Leftrightarrow \lim_{k \rightarrow +\infty} J_{r_i}(\lambda_i)^k = 0 \quad (\forall i),$$

dokaz je završen.

Sledeća teorema daje dovoljan uslov za konvergenciju niza $\{A^k\}_{k \in \mathbb{N}}$ ka nula matrici.

Teorema 1.7.2. Ako je bilo koja norma matrice A manja od jedinice važi $A^k \rightarrow 0 (k \rightarrow +\infty)$.

Dokaz. Kako je

$$\|A^k - 0\| = \|A^k\| = \|AA^{k-1}\| \leq \|A\| \cdot \|A^{k-1}\| \leq \dots \leq \|A\|^k$$

i kako je po pretpostavci $\|A\| < 1$, imamo $\|A^k\| \rightarrow 0$ ($k \rightarrow +\infty$), tj.
 $A^k \rightarrow 0$ ($k \rightarrow +\infty$).

Ovim je dokaz završen.

Sada ćemo dokazati jedan važan rezultat koji je u vezi sa teoremmama 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 definišemo matricu B pomoću

$$B = \frac{1}{\|A\| + \epsilon} A.$$

Kako je $\|B\| = \frac{1}{\|A\| + \epsilon} \|A\| < 1$, iz teoreme 1.7.2 sleduje $B^k \rightarrow 0$ ($k \rightarrow +\infty$). 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

$$|\lambda_i(B)| = \frac{1}{\|A\| + \epsilon} |\lambda_i(A)| < 1,$$

tj. $|\lambda_i(A)| < \|A\| + \epsilon$. 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

$$(1.7.1) \quad \sum_{m=0}^{+\infty} B^{(m)} = \lim_{k \rightarrow +\infty} \sum_{m=0}^k B^{(m)}$$

gde su $B^{(m)}$ ($m = 0, 1, \dots$) matrice istog reda.

Teorema 1.7.4. Ako matrični red (1.7.1) konvergira tada je

$$\lim_{k \rightarrow +\infty} B^{(k)} = 0.$$

Dokaz. Neka je $S^{(k)} = \sum_{m=0}^k B^{(m)}$ ($k = 0, 1, \dots$) i neka je

suma reda (1.7.1) S , tj. $\lim_{k \rightarrow +\infty} S^{(k)} = S$. Tada je

$$S^{(k)} - S^{(k-1)} = B^{(k)},$$

tj.

$$\lim_{k \rightarrow +\infty} S^{(k)} - \lim_{k \rightarrow +\infty} S^{(k-1)} = \lim_{k \rightarrow +\infty} B^{(k)},$$

odakle neposredno sleduje tvrdjenje teoreme 1.7.4.

U daljem tekstu daćemo potrebne i dovoljne uslove za konvergenciju matrične geometrijske progresije

$$(1.7.2) \quad \sum_{m=0}^{+\infty} A^m = I + A + A^2 + \dots,$$

gde je A kvadratna matrica reda n . Ovde je $B^{(m)} = A^m$ ($m = 0, 1, \dots$).

Teorema 1.7.5. Matrični red (1.7.2) konvergira ako i samo ako je

$$(1.7.3) \quad \lim_{k \rightarrow +\infty} A^k = 0.$$

Štaviše, tada je

$$(1.7.4) \quad I + A + A^2 + \dots = (I - A)^{-1}.$$

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. $|\lambda_i(A)| < 1$ ($i = 1, \dots, n$). Kako je

$$\det(I - A) = \prod_{i=1}^n (1 - \lambda_i(A)) \neq 0,$$

zaključujemo da postoji matrica $(I - A)^{-1}$. Množenjem identiteta

$$(I + A + A^2 + \dots + A^k)(I - A) = I - A^{k+1},$$

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

$$\begin{aligned}\lim_{k \rightarrow +\infty} (I + A + A^2 + \dots + A^k) &= (I - A)^{-1} - \lim_{k \rightarrow +\infty} A^{k+1}(I - A)^{-1} \\ &= (I - A)^{-1},\end{aligned}$$

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

Numerički problemi u linearnoj algebri mogu se klasificovati u nekoliko grupa:

1° Rešavanje sistema linearnih algebarskih jednačina

$$\vec{A}\vec{x} = \vec{b}$$

sa regularnom matricom A, izračunavanje determinante od A i inverzija matrice A;

2° Rešavanje proizvoljnog sistema linearnih jednačina metodom najmanjih kvadrata;

3° Određivanje sopstvenih vrednosti i sopstvenih vektora date kvadratne matrice;

4° Rešavanje zadatka linearog programiranja.

Za rešavanje ovih problema razvijen je čitav niz metoda, koji se mogu podeliti na direktne i iterativne. Ovo poglavlje posvećujemo direktnim metodama za rešavanje problema iz tačke 1°.

4.2.1. Uvodne napomene

Osnovna karakteristika direktnih metoda, ili tačnih metoda kako se ponekad nazivaju, 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 zao-krugljivanjem medjurezultata, krajnji rezultat je ograničene tačnosti. Zbog toga je od velikog značaja analiza grešaka kod ovih metoda. Jasno je da će uticaj grešaka biti veći ukoliko je broj operacija kod primenjenog metoda veći.

Posmatrajmo sistem linearnih algebarskih jednačina

$$\vec{Ax} = \vec{b},$$

gde su

$$(2.1.1) \quad A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & & a_{nn} \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Pretpostavimo da sistem ima jedinstveno rešenje, tj. da je $\det A \neq 0$. Tada se rešenja mogu izraziti pomoću Cramerovih formula

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

gde je A_i matrica dobijena iz matrice A zamenom i-te kolone vektorom \vec{b} . Međutim, 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 po definiciji

$$\det A = \sum (-1)^i a_{1i_1} a_{2i_2} \dots a_{ni_n},$$

gde se sumiranje obavlja preko svih permutacija (i_1, i_2, \dots, i_n) osnovnog skupa $(1, 2, \dots, n)$ (i je broj inverzija u permutaciji (i_1, i_2, \dots, i_n)), potrebno je izvršiti $S_n = n! - 1$ sabiranja i $M_n = (n-1)n!$ množenja, što ukupno iznosi $P_n = S_n + M_n \approx n \cdot n!$. Pod pretpostavkom da je za obavljanje jedne računske operacije potrebno

10 μ s, što je slučaj kod većine računara, to bi za izračunavanje vrednosti jedne determinante tridesetog reda bilo potrebno oko

$$\frac{30 \cdot 30! \cdot 10 \cdot 10^{-6}}{3600 \cdot 24 \cdot 365} = 2.5 \cdot 10^{21} \text{ godina.}$$

Uopšteno govoreći ovakav postupak je praktično ne primenljiv već za determinante reda $n > 5$.

4.2.2. Gaussov metod eliminacije i Gauss-Jordanov metod

Neka je dat sistem linearnih jednačina

$$(2.2.1) \quad \begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2, \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_n, \end{aligned}$$

koji ima jedinstveno rešenje. Za rešavanje ovog sistema, tj.

$$(2.2.2) \quad A\vec{x} = \vec{b},$$

gde su A, \vec{b}, \vec{x} dati pomoću (2.1.1), postoji veliki broj direktnih metoda. Jedan od najpogodnijih je svakako Gaussov metod eliminacije, koji ima više varijanata. U suštini Gaussov metod se zasniva na redukciji sistema (2.2.2), primenom elementarnih transformacija, na trougaoni sistem jednačina

$$(2.2.3) \quad R\vec{x} = \vec{c},$$

gde su

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ & r_{22} & & r_{2n} \\ & & \ddots & \\ & & & r_{nn} \end{bmatrix} \quad \text{i} \quad \vec{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

Sistem (2.2.3) se rešava sukcesivno polazeći od poslednje jednačine. Naime,

$$x_n = \frac{c_n}{r_{nn}},$$

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

Napomenimo da su koeficijenti $r_{ii} \neq 0$, jer po pretpostavci sistema (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

$$m_{il} = \frac{a_{il}}{a_{11}} \quad (i = 2, \dots, n),$$

a zatim množenjem prve jednačine u sistemu (2.2.1) sa m_{11} i oduzimanjem od i-te jednačine, dobijamo sistem od n-1 jednačina

$$(2.2.4) \quad \begin{aligned} a_{22}^{(2)} x_2 + \dots + a_{2n}^{(2)} x_n &= b_2^{(2)}, \\ &\vdots \\ a_{n2}^{(2)} x_2 + \dots + a_{nn}^{(2)} x_n &= b_n^{(2)}, \end{aligned}$$

gde su

$$a_{ij}^{(2)} = a_{ij} - m_{11} a_{1j}, \quad b_i^{(2)} = b_i - m_{11} b_1 \quad (i, j = 2, \dots, n).$$

Pod pretpostavkom da je $a_{22}^{(2)} \neq 0$, primenjujući isti postupak na (2.2.4) sa $m_{i2} = a_{i2}^{(2)} / a_{22}^{(2)}$ ($i = 3, \dots, n$) dobijamo sistem od n-2 jednačine

$$\begin{aligned} a_{33}^{(3)} x_3 + \dots + a_{3n}^{(3)} x_n &= b_3^{(3)}, \\ &\vdots \\ a_{n3}^{(3)} x_3 + \dots + a_{nn}^{(3)} x_n &= b_n^{(3)}, \end{aligned}$$

gde su

$$a_{ij}^{(3)} = a_{ij}^{(2)} - m_{i2} a_{2j}^{(2)}, \quad b_i^{(3)} = b_i^{(2)} - m_{i2} b_2^{(2)} \quad (i, j = 3, \dots, n).$$

Nastavljajući ovaj postupak, posle n-1 koraka dolazimo do jednačine

$$a_{nn}^{(n)} x_n = b_n^{(n)}.$$

Iz dobijenih sistema, uzimanjem prvih jednačina, dolazi do sistema jednačina

$$a_{11}^{(1)} x_1 + a_{12}^{(1)} x_2 + a_{13}^{(1)} x_3 + \dots + a_{1n}^{(1)} x_n = b_1^{(1)},$$

$$a_{22}^{(2)} x_2 + a_{23}^{(2)} x_3 + \dots + a_{2n}^{(2)} x_n = b_2^{(2)},$$

$$a_{33}^{(3)} x_3 + \dots + a_{3n}^{(3)} x_n = b_3^{(3)},$$

⋮

$$a_{nn}^{(n)} x_n = b_n^{(n)},$$

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

$$m_{ik} = \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}, \quad a_{ij}^{(k+1)} = a_{ij}^{(k)} - m_{ik} a_{kj}^{(k)}, \quad b_i^{(k+1)} = b_i^{(k)} - m_{ik} b_k^{(k)} \quad (i, j = k+1, \dots, n)$$

za $k = 1, 2, \dots, n-1$. Primetimo da su elementi matrice R i vektora \vec{c} dati sa

$$r_{ij} = a_{ij}^{(i)}, \quad c_i = b_i^{(i)} \quad (i = 1, \dots, n; \quad j = i, \dots, n).$$

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 obezrediti permutacijom jednačina u sistemu (videti [42]).

* 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

$$m_{21} = 3.799, m_{31} = 6.628, m_{32} = -5.138.$$

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 greška u elementima matrice R i vektora ĉ. 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 ([63, str. 205]).

Primer 2.2.2. Gaussovim metodom eliminacije rešimo sistem jednačina

$$\begin{bmatrix} 0.000003 & 0.213472 & 0.332147 \\ 0.215512 & 0.375623 & 0.476625 \\ 0.173257 & 0.663257 & 0.625675 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.235262 \\ 0.127653 \\ 0.285321 \end{bmatrix}$$

uzimajući sve medjurezultate sa šest značajnih cifara. S obzirom

da su faktori u ovom slučaju

$$m_{21} = 71837.3, m_{31} = 57752.3, m_{32} = 0.803905,$$

odgovarajući trougaoni sistem postaje

$$\begin{bmatrix} 0.000003 & 0.213472 & 0.332147 \\ -15334.9 & -23860.0 & \\ -0.500000 & & \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.235262 \\ -16900.5 \\ -0.20000 \end{bmatrix},$$

$$\text{odakle je } x_3 = 0.400000, x_2 = 0.479723, x_1 = -1.33333.$$

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

$$\begin{bmatrix} 0.215512 & 0.375623 & 0.476625 \\ 0.361282 & 0.242501 & \\ 0.188856 & & \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.127653 \\ 0.182697 \\ 0.127312 \end{bmatrix},$$

$$\text{odakle sleduje } x_3 = 0.674122, x_2 = 0.0532050, x_1 = -0.991291.$$

Napomenimo da su tačna rešenja datog sistema, sa deset značajnih cifara,

$$x_3 = 0.6741214694, x_2 = 0.05320393391, x_1 = -0.9912894252.$$

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 $|a_{rk}^{(k)}| = \max_{k \leq i \leq n} |a_{ik}^{(k)}|$, 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 $|a_{rs}^{(k)}| = \max_{k \leq i, j \leq n} |a_{ij}^{(k)}|$, 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, potrebno 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 $2(n-k) - 3(n-k)$, 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 zadovoljno veliko n imamo $N(n) \asymp 2n^3/3$. Dugo vremena se mislio da je Gaussov metod optimalan u pogledu broja računskih operacija. U novije vreme V. Strassen ([60]) 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, efi-

kasniji od Gaussovog metoda ($\log_2 7 < 3$).

Primedba 2.2.1. 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

$$s_1 = (a_{11} + a_{22})(b_{11} + b_{22}), \quad s_2 = (a_{21} + a_{22})b_{11}, \quad s_3 = a_{22}(b_{12} - b_{21}),$$

$$s_4 = a_{22}(b_{21} - b_{11}), \quad s_5 = (a_{11} + a_{12})b_{22}, \quad s_6 = (a_{21} - a_{11})(b_{11} + b_{12}),$$

$$s_7 = (a_{12} - a_{22})(b_{21} + b_{22})$$

i

$$c_{11} = s_1 + s_4 - s_5 + s_7, \quad c_{21} = s_2 + s_4, \quad c_{12} = s_3 + s_5, \quad c_{22} = s_1 + s_3 - s_2 + s_6.$$

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 0.18s, ako se jedna računska operacija obavlja za 10 μs.

Pored izloženih varijanti Gaussovog eliminacionog metoda za rešavanje sistema linearnih jednačina postoje i drugi direktni metodi koji transformišu sistem (2.2.2) na neki sistem, poput (2.2.3), koji je pogodan za rešavanje. Na primer, ako je transformisani sistem dijagonalan, tj.

$$(2.2.5) \quad \vec{D}\vec{x} = \vec{c} \quad (D = \text{diag}(d_{11}, \dots, d_{nn})),$$

njegovo rešavanje je veoma jednostavno

$$x_k = \frac{c_k}{d_{kk}} \quad (k = 1, 2, \dots, n).$$

Štaviše, ako je D jedinična matrica onda je već postupkom eliminacije završeno i rešavanje sistema jednačina (2.2.5), $\vec{x} = \vec{c}$.

Izložićemo sada postupak za transformaciju sistema (2.2.2) na oblik (2.2.5), koji je poznat kao Gauss-Jordanov metod. Prvi korak u ovom metodu poklapa se sa Gaussovim metodom. Stavimo, kao i ranije, $a_{ij}^{(1)} = a_{ij}$, $b_i^{(1)} = b_i$. Pod pretpostavkom da je $a_{11}^{(1)} \neq 0$, posle prvog eliminacionog koraka dobijamo sistem od $n-1$ jednačina koji pridružimo prvoj jednačini.

Tako imamo

$$a_{11}^{(1)}x_1 + a_{12}^{(2)}x_2 + a_{13}^{(2)}x_3 + \dots + a_{1n}^{(2)}x_n = b_1^{(2)},$$

$$a_{22}^{(2)}x_2 + a_{23}^{(2)}x_3 + \dots + a_{2n}^{(2)}x_n = b_2^{(2)},$$

$$a_{32}^{(2)}x_2 + a_{33}^{(2)}x_3 + \dots + a_{3n}^{(2)}x_n = b_3^{(2)},$$

⋮

$$a_{n2}^{(2)}x_2 + a_{n3}^{(2)}x_3 + \dots + a_{nn}^{(2)}x_n = b_n^{(2)},$$

gde smo dodatno izvršili prenumeraciju $a_{1j}^{(2)} = a_{1j}^{(1)}$ ($j=2, \dots, n$) i $b_1^{(2)} = b_1^{(1)}$.

Pod pretpostavkom da je $a_{22}^{(2)} \neq 0$, u drugom eliminacionom koraku eliminišemo x_2 (tj. anuliramo koeficijente uz x_2) iz svih jednačina sem iz druge. Naime, kod Gaussovog metoda smo eliminaciju sprovedeli samo u jednačinama koje se nalaze posle druge jednačine. Ovde, dakle, eliminaciju sprovodimo i u prvoj jednačini. Tako sa faktorima $m_{i2} = a_{i2}^{(2)} / a_{22}^{(2)}$ ($i \neq 2$) dobijamo

$$a_{11}^{(1)}x_1 + a_{13}^{(3)}x_3 + \dots + a_{1n}^{(3)}x_n = b_1^{(3)},$$

$$a_{22}^{(2)}x_2 + a_{23}^{(3)}x_3 + \dots + a_{2n}^{(3)}x_n = b_2^{(3)},$$

$$a_{33}^{(3)}x_3 + \dots + a_{3n}^{(3)}x_n = b_3^{(3)},$$

⋮

$$a_{n3}^{(3)}x_3 + \dots + a_{nn}^{(3)}x_n = b_n^{(3)},$$

gde su $a_{2j}^{(3)} = a_{2j}^{(2)}$ ($j = 3, \dots, n$), $b_2^{(3)} = b_2^{(2)}$ i
 $a_{ij}^{(3)} = a_{ij}^{(2)} - m_{i2} a_{2j}^{(2)}$, $b_i^{(3)} = b_i^{(2)} - m_{i2} b_2^{(2)}$ ($i \neq 2; j=3, \dots, n$).

Nastavljajući ovaj postupak posle n koraka dobijamo ekvivalentni dijagonalni sistem

$$(2.2.6) \quad \begin{aligned} a_{11}^{(1)} x_1 &= b_1^{(n+1)}, \\ a_{22}^{(2)} x_2 &= b_2^{(n+1)}, \\ &\vdots \\ a_{nn}^{(n)} x_n &= b_n^{(n+1)}. \end{aligned}$$

Primetimo da je ovde neophodno n koraka, za razliku od Gaussovog metoda gde je bilo dovoljno $n-1$ koraka. Naime, ovde se posle $n-1$ koraka dobija sistem jednačina oblika

$$\begin{aligned} a_{11}^{(1)} x_1 + a_{1n}^{(n)} x_n &= b_1^{(n)}, \\ a_{22}^{(2)} x_2 + a_{2n}^{(n)} x_n &= b_2^{(n)}, \\ &\vdots \\ a_{nn}^{(n)} x_n &= b_n^{(n)}. \end{aligned}$$

U poslednjem n -tom koraku treba eliminisati x_n iz prvih $n-1$ jednačina.

Gauss-Jordanov metod zahteva više operacija od Gaussovog metoda. Naime, ovde je u k -tom eliminacionom koraku potrebno $n-1$ deljenja, $(n-1)(n-k+1)$ množenja i isto toliko oduzimanja, što znači da je za redukciju sistema na dijagonalni oblik potrebno $n(n-1)$ deljenja i po $\frac{1}{2}n(n^2 - 1)$ množenja i oduzimanja. Rešavanje dijagonalnog sistema zahteva još n deljenja. Prema tome, ukupno je potrebno n^2 deljenja i po $\frac{1}{2}n(n^2 - 1)$ množenja i oduzimanja, što sve iznosi

$$\bar{N}(n) = n^3 + n^2 - n.$$

Ovo je jasno veće u poređenju sa Gaussovim metodom. Naime, za dovoljno veliko n imamo da je $\bar{N}(n) \approx \frac{3}{2} N(n)$, gde je $N(n)$ broj operacija kod Gaussovog metoda.

Na osnovu (2.2.6) vidimo da su dijagonalni elementi u matričnom sistemu (2.2.5) $d_{kk} = a_{kk}^{(k)}$ i da su $c_k = b_k^{(n+1)}$.

I ovde, kao i kod Gaussovog metoda na potpuno istovetan način se može izvesti izbor glavnih elemenata i pomoću njih sprovesti eliminacioni postupak.

Ako se pre svakog eliminacionog koraka izvrši deljenje odgovarajuće jednačine sa glavnim elementom, tj. pre prvog koraka podeli prva jednačina, pre drugog koraka podeli druga jednačina, itd., transformisani sistem imaće jediničnu matricu $D = I$. Naravno, ovakav postupak zahteva veći broj operacija.

4.2.3. Inverzija matrica

Neka je $A = [a_{ij}]_{n \times n}$ regularna matrica i neka je

$$\mathbf{x} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & & x_{2n} \\ \vdots & & & \\ x_{n1} & x_{n2} & & x_{nn} \end{bmatrix} = [\vec{x}_1 \ \vec{x}_2 \ \dots \ \vec{x}_n]$$

njena inverzna matrica. Vektori $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$... 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 određivanja inverzne matrice može se svesti na rešavanje n sistema linearnih jednačina

$$(2.3.1) \quad A\vec{x}_i = \vec{e}_i \quad (i = 1, \dots, n).$$

Za rešavanje sistema (2.3.1) pogodno je koristiti Gaussov 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 $I = [\vec{e}_1 \ \vec{e}_2 \ \dots \ \vec{e}_n]$. Na taj način se matrica A transformiše u trougaonu matricu R , a matrica I u matricu $C = [\vec{c}_1 \ \vec{c}_2 \ \dots \ \vec{c}_n]$. Najzad,

ostaje da se reše trougaoni sistemi $\vec{R}\vec{x}_i = \vec{c}_i$ ($i = 1, \dots, n$). Dakle, primena Gaussovog metoda može se iskazati kao transformacija:

$$[A | I] \rightarrow [R | C].$$

Za inverziju matrica vrlo često se koristi Gauss-Jordanov metod i to u varijanti sa jediničnom dijagonalom. U ovom slučaju, odgovarajuća transformacija je

$$[A | I] \rightarrow [I | A^{-1}],$$

što znači da se na mesto jedinične matrice pojavljuje inverzna matrica.

Primer 2.3.1. Neka je data matrica

$$A = \begin{bmatrix} 3 & 1 & 6 \\ 2 & 1 & 3 \\ 1 & 1 & 1 \end{bmatrix}.$$

Primenom Gauss-Jordanovog metoda na $[A | I]$ imamo redom

$$\begin{array}{l} \left[\begin{array}{ccc|ccc} 1 & 1/3 & 2 & 1/3 & 0 & 0 \\ 2 & 1 & 3 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array} \right] \rightarrow \left[\begin{array}{ccc|ccc} 1 & 1/3 & 2 & 1/3 & 0 & 0 \\ 0 & 1/3 & -1 & -2/3 & 1 & 0 \\ 0 & 2/3 & -1 & -1/3 & 0 & 1 \end{array} \right] \rightarrow \\ \rightarrow \left[\begin{array}{ccc|ccc} 1 & 1/3 & 2 & 1/3 & 0 & 0 \\ 0 & 1 & -3 & -2 & 3 & 0 \\ 0 & 2/3 & -1 & -1/3 & 0 & 1 \end{array} \right] \rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 1 & -1 & 0 \\ 0 & 1 & -3 & -2 & 3 & 0 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array} \right] \rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -2 & 5 & -3 \\ 0 & 1 & 0 & 1 & -3 & 3 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array} \right]. \end{array}$$

Dakle,

$$A^{-1} = \begin{bmatrix} -2 & 5 & -3 \\ 1 & -3 & 3 \\ 1 & -2 & 1 \end{bmatrix}.$$

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 $\ell_{ii} = 1$ ($i = 1, \dots, n$),

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

$$\vec{b} = [b_1 \ b_2 \ \dots \ b_n]^T, \quad \vec{y} = [y_1 \ y_2 \ \dots \ y_n]^T, \quad \vec{x} = [x_1 \ x_2 \ \dots \ x_n]^T.$$

Tada je

$$y_1 = b_1, \quad y_i = b_i - \sum_{k=1}^{i-1} \ell_{ik} y_k \quad (i = 2, \dots, n)$$

i

$$x_n = \frac{y_n}{r_{nn}}, \quad x_i = \frac{1}{r_{ii}} \left(y_i - \sum_{k=i+1}^n r_{ik} x_k \right) \quad (i = n-1, \dots, 1).$$

Izloženi metod se u literaturi sreće kao metod Haleckoq. U slučaju kada je matrica A normalna, tj. kada je simetrična i pozitivno definitna, metod Haleckoq 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

$$\left. \begin{aligned} r_{11} &= \sqrt{a_{11}}, \quad r_{1j} = \frac{a_{1j}}{r_{11}} \quad (j = 2, \dots, n) \\ r_{ii} &= \sqrt{a_{ii} - \sum_{k=1}^{i-1} r_{ki}^2} \\ r_{ij} &= \frac{1}{r_{ii}} \left(a_{ij} - \sum_{k=1}^{i-1} r_{ki} r_{kj} \right) \quad (j = i+1, \dots, n) \end{aligned} \right\} (i=2, \dots, n).$$

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 metodom kvadratnog korena

$$\det A = (r_{11} r_{22} \dots r_{nn})^2.$$

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 maticu 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 $A^{(2)} \vec{x} = \vec{b}^{(2)}, \dots, A^{(n)} \vec{x} = \vec{b}^{(n)}$, pri čemu matrica $A^{(k)}$ ima oblik takav da su svi njeni elementi ispod glavne dijagonale i ispred k-te kolone jednaki nuli, tj.

* Na primer, kod izračunavanja prenosnih karakteristika elektronskih kola.

$$A^{(k)} = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \dots & a_{1k}^{(1)} & \dots & a_{1n}^{(1)} \\ a_{22}^{(2)} & & a_{2k}^{(2)} & & a_{2n}^{(2)} & \\ \vdots & \ddots & \vdots & & & \\ a_{kk}^{(k)} & & a_{kn}^{(k)} & & & \\ \vdots & & & & & \\ a_{nk}^{(k)} & & a_{nn}^{(k)} & & & \end{bmatrix}.$$

Analizirajmo modifikaciju elementa a_{ij} ($= a_{ij}^{(1)}$) u procesu trougaone redukcije. Kako je, za $k = 1, 2, \dots, n-1$,

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - m_{ik} a_{kj}^{(k)} \quad (i, j = k+1, \dots, n)$$

$$a_{i1}^{(k+1)} = a_{i2}^{(k+1)} = \dots = a_{ik}^{(k+1)} = 0 \quad (i = k+1, \dots, n),$$

sumiranjem dobijamo

$$\sum_{k=1}^{i-1} a_{ij}^{(k+1)} = \sum_{k=1}^{i-1} a_{ij}^{(k)} - \sum_{k=1}^{i-1} m_{ik} a_{kj}^{(k)} \quad (i \leq j)$$

$$\sum_{k=1}^j a_{ij}^{(k+1)} = \sum_{k=1}^j a_{ij}^{(k)} - \sum_{k=1}^j m_{ik} a_{kj}^{(k)} \quad (i > j),$$

tj.

$$a_{ij} = a_{ij}^{(1)} = a_{ij}^{(i)} + \sum_{k=1}^{i-1} m_{ik} a_{kj}^{(k)} \quad (i \leq j),$$

$$a_{ij} = a_{ij}^{(1)} = 0 + \sum_{k=1}^j m_{ik} a_{kj}^{(k)} \quad (i > j).$$

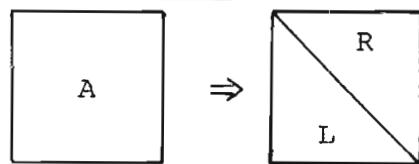
Definišući $m_{ii} = 1$ ($i = 1, \dots, n$), poslednje dve jednakosti se mogu predstaviti u obliku

$$(2.4.3) \quad a_{ij} = \sum_{k=1}^p m_{ik} a_{kj}^{(k)} \quad (i, j = 1, \dots, n),$$

gde je $p = \min(i, j)$. Jednakost (2.4.3) ukazuje da Gaussova eliminacija daje LR faktorizaciju matrice A, gde su

$$L = \begin{bmatrix} 1 & & & \\ m_{21} & 1 & & \\ & \ddots & \ddots & \\ m_{n1} & m_{n2} & \dots & 1 \end{bmatrix}, \quad R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{22} & & & r_{2n} \\ & \ddots & & \\ & & & r_{nn} \end{bmatrix}$$

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. Međutim, 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 $I = (p_1, \dots, p_{n-1})$, pri čemu je p_k broj vrste iz koje se uzima glavni element u k-tom eliminacionom koraku. Kod rešavanja sistema $\vec{Ax} = \vec{b}$, 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.

Primer 2.4.1. Primenom Gaussovog metoda sa izborom glavnog elementa na

$$A = \begin{bmatrix} 3 & 1 & 6 \\ 2 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix}$$

dobijamo redom

$$\left[\begin{array}{ccc} (3) & 1 & 6 \\ 2 & 1 & 3 \\ 1 & 1 & 1 \end{array} \right] \Rightarrow \left[\begin{array}{ccc|cc} 3 & 1 & 6 & & \\ \hline 2/3 & 1/3 & -1 & 1/3 & \\ 1/3 & 2/3 & -1 & 2/3 & \\ \hline & 1/2 & -1/2 & & \end{array} \right],$$

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

$$y_1 = 2, \quad y_2 = 4 - \frac{1}{3}y_1 = \frac{10}{3}, \quad y_3 = 7 - \frac{2}{3}y_1 - \frac{1}{2}y_2 = 4.$$

Najzad, na osnovu $R\vec{x} = \vec{y}$, nalazimo

$$x_3 = -8, \quad x_2 = \frac{3}{2} \left(\frac{10}{3} + x_3 \right) = -7, \quad x_1 = \frac{1}{3} (2 - x_2 - 6x_3) = 19.$$

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

$$y = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ 1 \end{bmatrix}, \quad \vec{a}_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \\ -b_i \end{bmatrix} \quad (i = 1, \dots, n),$$

tada se ovaj sistem može predstaviti u obliku

$$(2.5.1) \quad (\vec{a}_i, \vec{y}) = \vec{y}^T \vec{a}_i = 0 \quad (i = 1, \dots, n).$$

Jednačine (2.5.1) ukazuju na mogućnost rešavanja sistema (2.2.1), korišćenjem uslova ortogonalnosti vektora \vec{y} sa vektorima \vec{a}_i ($i = 1, \dots, n$). Pomenuta ortogonalnost je ekvivalentna ortogonalnosti vektora \vec{y} sa linearnim potprostorom H_n , koji je generisan vektorima \vec{a}_i ($i = 1, \dots, n$), tj. $H_n = L(\vec{a}_1, \dots, \vec{a}_n)$. Polazeći od baze $\{\vec{a}_1, \dots, \vec{a}_n\}$, korišćenjem Gram-Schmidtovog postupka, konstruišimo ortonormirani bazu $B_n = \{\vec{v}_1, \dots, \vec{v}_n\}$ potprostora H_n . Vektor \vec{y} je, očigledno, ortogonalan sa svim vektorima ortonormirane baze B_n .

Kako je vektor $\vec{a}_{n+1} = [0 \ 0 \ \dots \ 0 \ 1]^T$ linearno nezavisran 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, \dots, u_n, u_{n+1} , tj.
 $\vec{u} = [u_1 \ u_2 \ \dots \ u_n \ u_{n+1}]^T$. Kako vektor \vec{u} ispunjava uslove ortogonalnosti $(\vec{u}, \vec{v}_i) = 0$ ($i=1, \dots, n$), na osnovu prethodnog zaključujemo da je, takodje, $(\vec{u}, \vec{a}_i) = 0$ ($i=1, \dots, n$), tj.

$$(\vec{u}, \vec{a}_1) = a_{11}u_1 + a_{12}u_2 + \dots + a_{1n}u_n - b_1u_{n+1} = 0,$$

$$(\vec{u}, \vec{a}_2) = a_{21}u_1 + a_{22}u_2 + \dots + a_{2n}u_n - b_2u_{n+1} = 0,$$

⋮

$$(\vec{u}, \vec{a}_n) = a_{n1}u_1 + a_{n2}u_2 + \dots + a_{nn}u_n - b_nu_{n+1} = 0.$$

Primetimo da je $u_{n+1} \neq 0$. Naime, ako bi bilo $u_{n+1} = 0$, tada bi n-torka (u_1, u_2, \dots, u_n) bila rešenje homogenog sistema jednačina sa matricom $A = [a_{ij}]$. Međutim, kako homogeni sistem ($\det A \neq 0$) ima samo trivijalna rešenja, to bismo imali $u_i = 0$ ($i=1, \dots, n$), 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 $x_i = \frac{u_i}{u_{n+1}}$ ($i=1, \dots, n$), rešenje sistema jednačina (2.5.1). Tada je n-torka (x_1, \dots, 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 slabouslovjeni 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 ove jednakosti sleduje

$$(2.6.1) \quad \|\Delta\vec{x}\| \leq \|A^{-1}\| \cdot \|\Delta\vec{b}\|.$$

Kako je $\|\vec{b}\| = \|A\vec{x}\| \leq \|A\| \cdot \|\vec{x}\|$, imamo

$$\frac{\|\Delta\vec{x}\|}{\|\vec{x}\|} \leq \|A\| \cdot \|A^{-1}\| \cdot \frac{\|\Delta\vec{b}\|}{\|\vec{b}\|} = k(A) \frac{\|\Delta\vec{b}\|}{\|\vec{b}\|},$$

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 $k(A) \geq 1$, što sleduje iz

$$\|\vec{x}\| = \|I\vec{x}\| = \|AA^{-1}\vec{x}\| \leq \|A\| \cdot \|A^{-1}\| \cdot \|\vec{x}\|.$$

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 pojednostavljuje, 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

$$(2.6.2) \quad \|\Delta\vec{x}\| \leq \|A^{-1}\| \cdot \|\vec{r}(\vec{x}_p)\|.$$

Na osnovu ove nejednakosti može se zaključiti da $\|\Delta\vec{x}\|$ može biti velika, i u slučajevima kada je veličina $\|\vec{r}(\vec{x}_p)\|$ dovoljno mala.

Teorema 2.6.1. 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

$$1^o \quad \frac{\|\Delta\vec{x}\|}{\|\vec{x}\|} \leq \frac{\|F\|}{1-\|F\|},$$

$$2^o \quad \frac{\|\Delta\vec{x}\|}{\|\vec{x}\|} \leq \frac{k(A)}{1-k(A)} \cdot \frac{\|B-A\|}{\|A\|} \text{ ako je } k(A) \cdot \frac{\|B-A\|}{\|A\|} < 1.$$

Dokaz. Kako, na osnovu učinjenih pretpostavki, B^{-1} egzistira imamo

$$\Delta\vec{x} = B^{-1}\vec{b} - A^{-1}\vec{b} = B^{-1}(A-B)A^{-1}\vec{b}, \vec{x} = A^{-1}\vec{b},$$

$$\frac{\|\Delta\vec{x}\|}{\|\vec{x}\|} \leq \|B^{-1}(A-B)\| = \|(I+F)^{-1}A^{-1}AF\| \leq \|(I+F)^{-1}\| \cdot \|F\| \leq \frac{\|F\|}{1-\|F\|}.$$

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^o .

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

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

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

$$\|\vec{x}_p - \vec{x}\| \leq \frac{\|B^{-1}\|}{1 - \|I - B^{-1}A\|} \|\vec{r}(\vec{x}_p)\| \quad (\|I - B^{-1}A\| < 1).$$

Primer 2.6.1. Jedan tipičan primer slabouslovljenog sistema je sistem (videti [40])

$$\begin{bmatrix} 121734 & 169217 & 176624 & 166662 \\ 169217 & 235222 & 245505 & 231653 \\ 176624 & 245505 & 256423 & 242029 \\ 166662 & 231653 & 242029 & 228474 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 634237 \\ 881597 \\ 920581 \\ 868818 \end{bmatrix},$$

čija su tačna rešenja $x_1 = x_2 = x_3 = x_4 = 1$. Međutim, 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

Drugu važnu klasu metoda u linearnoj algebri čine iterativni metodi, kod kojih se teorijski rezultat dobija posle beskonačnog broja koraka. Praktično, međutim, za nalaženje rešenja sa dovoljnom tačnošću potrebno je izvršiti konačan broj koraka. Broj koraka, jasno, zavisi od zahtevane tačnosti.

Kao početne vrednosti rešenja kod primene iterativnih metoda, često se koriste rezultati dobijeni nekim od direktnih

metoda izloženih u prethodnom poglavlju.

U ovom poglavlju razmatraćemo samo iterativne metode za rešavanje sistema linearnih jednačina i metode za inverziju matriča. Iterativni metodi za rešavanje problema sopstvenih vrednosti (kao uostalom i direktni metodi) biće razmatrani u posebnom poglavlju. Napomenimo ovde da se za rešavanje sistema sa velikim brojem jednačina, kakvi se javljaju pri rešavanju parcijalnih diferencijalnih jednačina, uglavnom koriste iterativni metodi.

4.3.1. Načini formiranja iterativnih metoda

Posmatrajmo sistem linearnih jednačina

$$(3.1.1) \quad \begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2, \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_n, \end{aligned}$$

koji se može predstaviti i u matričnom obliku

$$(3.1.2) \quad \vec{A}\vec{x} = \vec{b},$$

gde su

$$\vec{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & & a_{nn} \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

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 određivanje rešenja \vec{x} sa unapred zadatom tačnošću. Naime, polazeći od proizvoljnog vektora $\vec{x}^{(0)} (= [x_1^{(0)} \dots x_n^{(0)}]^T)$, iterativnim metodom

se određuje niz $\{\vec{x}^{(k)}\}$ ($\vec{x}^{(k)} = [x_1^{(k)} \dots x_n^{(k)}]^T$) takav da je

$$\lim_{k \rightarrow +\infty} \vec{x}^{(k)} = \vec{x}.$$

Jedan opšti iterativni metod može se predstaviti u obliku

$$\vec{x}^{(k)} = F_k(\vec{x}^{(0)}, \vec{x}^{(1)}, \dots, \vec{x}^{(k-1)}) \quad (k=1, 2, \dots),$$

gde funkcija F_k u opštem slučaju zavisi od A, \vec{b}, k .

Sa stanovišta primene, najinteresantniji su iterativni metodi oblika

$$(3.1.3) \quad \vec{x}^{(k)} = F_k(\vec{x}^{(k-1)}) \quad (k=1, 2, \dots).$$

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.

$$(3.1.4) \quad F_k(\vec{x}) = B_k \vec{x} + \vec{c}_k,$$

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

$$(3.1.5) \quad A^{-1}\vec{b} = B_k A^{-1}\vec{b} + \vec{c}_k.$$

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. $C_k A + B_k = I$.

Na taj način (3.1.4) postaje

$$(3.1.6) \quad F_k(\vec{x}) = B_k \vec{x} + C_k \vec{b},$$

ili

$$(3.1.7) \quad F_k(\vec{x}) = \vec{x} - C_k(A\vec{x} - \vec{b}).$$

Često se za C_k uzima dijagonalna matrica sa jednakim elementima na dijagonali, tj. $C_k = \text{diag}(c_k, \dots, c_k) = c_k I$ ($c_k \in \mathbb{R}$). Tada se (3.1.7) svodi na

$$(3.1.8) \quad F_k(\vec{x}) = \vec{x} - c_k(A\vec{x} - \vec{b}).$$

Ako se matrica A predstavi u obliku

$$A = D_k + E_k,$$

gde je D_k regularna matrica, tada se F_k može zadati u implicitnom obliku kao

$$(3.1.9) \quad D_k F_k(\vec{x}) + E_k \vec{x} = \vec{b}.$$

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 se zasniva na minimizaciji funkcionele 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 funkcionele 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.

$$(3.2.1) \quad F(\vec{x}) = B\vec{x} + C\vec{b}.$$

Ako stavimo $\vec{C}\vec{b} = \vec{\beta}$, iz (3.2.1) sleduje

$$(3.2.2) \quad \vec{x}^{(k)} = \vec{B}\vec{x}^{(k-1)} + \vec{\beta} \quad (k=1, 2, \dots).$$

Napomenimo, da je jednačina

$$(3.2.3) \quad \vec{x} = \vec{B}\vec{x} + \vec{\beta}$$

ekvivalentna sa

$$(3.2.4) \quad \vec{A}\vec{x} = \vec{b}.$$

Matrica B se naziva iterativna matrica.

Ako se podje od proizvoljnog vektora $\vec{x}^{(0)}$, pomoću (3.2.2) generiše se niz $\{\vec{x}^{(k)}\}$. Razmatraćemo uslove pod kojima generisani niz konvergira ka tačnom rešenju sistema (3.2.3).

Ako je

$$B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & & b_{2n} \\ \vdots & & & \\ b_{n1} & b_{n2} & & b_{nn} \end{bmatrix} \text{ i } \vec{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix},$$

iterativni metod (3.2.2) može se predstaviti skalarno

$$x_1^{(k)} = b_{11}x_1^{(k-1)} + b_{12}x_2^{(k-1)} + \dots + b_{1n}x_n^{(k-1)} + \beta_1,$$

$$x_2^{(k)} = b_{21}x_1^{(k-1)} + b_{22}x_2^{(k-1)} + \dots + b_{2n}x_n^{(k-1)} + \beta_2,$$

\vdots

$$x_n^{(k)} = b_{n1}x_1^{(k-1)} + b_{n2}x_2^{(k-1)} + \dots + b_{nn}x_n^{(k-1)} + \beta_n,$$

gde je $k=1, 2, \dots$.

Dokazaćemo sledeće dve teoreme.

Teorema 3.2.1. Ako je $\vec{x}^{(0)}$ proizvoljan vektor, $\|B\|<1$ je dovoljan uslov zakonvergenciju procesa (3.2.2) ka tačnom rešenju \vec{x} sistema (3.2.3).

Teorema 3.2.2. Ako je $\vec{x}^{(0)}$ proizvoljan vektor i $\|B\| < 1$, tada za svako $k \in \mathbb{N}$, važe nejednakosti

$$(3.2.5) \quad \|(\mathbf{I}-B)^{-1} - (\mathbf{I}+B+\dots+B^{k-1})\| \leq \frac{\|B\|^k}{1-\|B\|};$$

$$(3.2.6) \quad \|\vec{x}^{(k)} - \vec{x}\| \leq \|B\|^k \|\vec{x}^{(0)}\| + \frac{\|\vec{\beta}\| \cdot \|B\|^k}{1-\|B\|};$$

$$(3.2.7) \quad \|\vec{x}^{(k)} - \vec{x}\| \leq \|B\| \cdot \|\vec{x}^{(k-1)} - \vec{x}\|;$$

$$(3.2.8) \quad \|\vec{x}^{(k)} - \vec{x}\| \leq \|B\|^k \|\vec{x}^{(0)} - \vec{x}\|.$$

Dokaz teoreme 3.2.1. Na osnovu (3.2.2) matematičkom indukcijom lako dokazujemo jednakost

$$(3.2.9) \quad \vec{x}^{(k)} = B^k \vec{x}^{(0)} + (\mathbf{I}+B+\dots+B^{k-1}) \vec{\beta} \quad (k \in \mathbb{N}).$$

Kako je $\|B\| < 1$, imamo

$$\lim_{k \rightarrow +\infty} \|B^k\| = 0, \text{ tj. } \lim_{k \rightarrow +\infty} B^k = \mathbf{0}$$

i

$$\lim_{k \rightarrow +\infty} (\mathbf{I}+B+\dots+B^{k-1}) = (\mathbf{I}-B)^{-1}.$$

Tada, iz (3.2.9) sleduje

$$\lim_{k \rightarrow +\infty} \vec{x}^{(k)} = \lim_{k \rightarrow +\infty} B^k \vec{x}^{(0)} + \lim_{k \rightarrow +\infty} (\mathbf{I}+B+\dots+B^{k-1}) \vec{\beta} = (\mathbf{I}-B)^{-1} \vec{\beta},$$

odakle zaključujemo da niz $\{\vec{x}^{(k)}\}$ konvergira ka rešenju jednačine (3.2.3).

Dokaz teoreme 3.2.2. Iz jednakosti

$$(\mathbf{I}-B)^{-1} = \mathbf{I} + B + B^2 + \dots \quad (\|B\| < 1)$$

sleduje

$$(\mathbf{I}-B)^{-1} - (\mathbf{I}+B+\dots+B^{k-1}) = B^k + B^{k+1} + \dots,$$

odnosno

$$\begin{aligned} \|(I-B)^{-1} - (I+B+\dots+B^{k-1})\| &\leq \|B\|^k + \|B\|^{k+1} + \dots \\ &= \frac{\|B\|^k}{1 - \|B\|}. \end{aligned}$$

Na osnovu (3.2.9), imamo

$$\begin{aligned} \vec{x}^{(k)} - \vec{x} &= B^k \vec{x}^{(0)} + (I+B+\dots+B^{k-1}) \vec{\beta} - (I-B)^{-1} \vec{\beta} \\ &= B^k \vec{x}^{(0)} - ((I-B)^{-1} - (I+B+\dots+B^{k-1})) \vec{\beta}, \end{aligned}$$

odakle je

$$\begin{aligned} \|\vec{x}^{(k)} - \vec{x}\| &\leq \|B^k \vec{x}^{(0)}\| + \|(I-B)^{-1} - (I+B+\dots+B^{k-1})\| \cdot \|\vec{\beta}\| \\ &\leq \|B\|^k \|\vec{x}^{(0)}\| + \frac{\|\vec{\beta}\| \cdot \|B\|^k}{1 - \|B\|}, \end{aligned}$$

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

$$\|\vec{x}^{(k)} - \vec{x}\| \leq \|B\| \cdot \|\vec{x}^{(k-1)} - \vec{x}\|.$$

Konačno, iteriranjem poslednje nejednakosti, dobijamo

$$\|\vec{x}^{(k)} - \vec{x}\| \leq \|B\|^k \|\vec{x}^{(0)} - \vec{x}\|.$$

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 $\vec{x}^{(0)} = \vec{0}$. Tada se nejednakost (3.2.6), iz teoreme 3.2.1, može pooštiti. Naime, važi nejednakost

$$\begin{aligned}\|\vec{x}^{(k)} - \vec{x}\| &= \|((I-B)^{-1} - (I+B+\dots+B^k))\vec{\beta}\| \\ &\leq \frac{\|B\|^{k+1} \|\beta\|}{1 - \|B\|}.\end{aligned}$$

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

$$(c) \|B\|_2 = \varepsilon(B) = (\sum_{i,j} |b_{ij}|^2)^{1/2} < 1.$$

Primetimo da su uslovi (a), (b), (c) ispunjeni ako je $|b_{ij}| < \frac{1}{n}$ ($i, j = 1, \dots, n$).

Primedba 3.2.2. Ako je $\|B\| < 1$, važi

$$\|\vec{x}^{(k)} - \vec{x}\| \leq \frac{\|B\|}{1 - \|B\|} \|\vec{x}^{(k)} - \vec{x}^{(k-1)}\| \quad (k \in \mathbb{N}).$$

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 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 $\|\vec{x}^{(p+1)} - \vec{x}^{(p)}\|_E \geq \alpha$ ($p < k$). Naime, oni su dokazali sledeću teoremu:

Teorema 3.2.3. Neka je B realna normalna matrica, čije sopstvene vrednosti ispunjavaju uslov $1 > \lambda_1 > \lambda_2 \dots > \lambda_n > 0$. Ako je početna greška $\vec{\varepsilon}^{(0)} = \vec{x}^{(0)} - \vec{x}$ uniformno rasporedjena u zatvorenoj kugli

$\bar{S}(0, r) = \left\{ \vec{x} \mid \|x\|_E \leq r, \vec{x} \in \mathbb{R}^n \right\}$, tada verovatnoća, da greška $\vec{\varepsilon}^{(k)}$ zadovoljava nejednakosti

$$\frac{\theta \lambda_1^\alpha}{1 - \lambda_1} \leq \|\vec{\varepsilon}^{(k)}\|_E \leq \frac{\alpha}{1 - \lambda_1},$$

teži jedinici kada $r \rightarrow +\infty$, za svako $\theta < 1$, tj.

$$\lim_{r \rightarrow +\infty} P\left(\frac{\theta \lambda_1^\alpha}{1 - \lambda_1} \leq \|\vec{\varepsilon}^{(k)}\|_E \leq \frac{\alpha}{1 - \lambda_1}\right) = 1.$$

Ovaj rezultat, Peradze [45]) je generalisao, pretpostavljajući samo uslov da je matrica B realna i regularna. Konačno, T.Yamamoto u [65] i [66] izbacio je i uslov regularnosti matrice B.

Neka je B realna matrica reda n, čije različite sopstvene vrednosti λ_i ($i=1, \dots, m$), reda višestrukosti n_i ($\sum_{i=1}^m n_i = n$), ispunjavaju uslov

$$1 > \rho(B) = |\lambda_1| = |\lambda_2| = \dots = |\lambda_s| > |\lambda_{s+1}| \geq \dots \geq |\lambda_m|.$$

Navešćemo dva partikularna rezultata, do kojih je došao T.Yamamoto ([65]).

Teorema 3.2.4. Neka je A realna normalna matrica i

$$\lambda = \min_{1 \leq i \leq s} \left| \frac{\lambda_i}{1 - \lambda_i} \right|, \Lambda = \max_{1 \leq i \leq s} \left| \frac{1}{1 - \lambda_i} \right|, \Lambda^* = \max_{\lambda_i \neq 0} \left| \frac{1}{1 - \lambda_i} \right|.$$

Ako je početna greška $\vec{\varepsilon}^{(0)}$ uniformno rasporedjena u $\bar{S}(0, r)$, tada je

$$(3.2.10) \quad \lim_{r \rightarrow +\infty} P(\theta_1 \lambda \alpha < \|\vec{\varepsilon}^{(k)}\|_E < \theta_2 \Lambda \alpha) = 1$$

i

$$(3.2.11) \quad \lim_{r \rightarrow +\infty} P(\theta_1 \lambda \alpha < \|\vec{\varepsilon}^{(k)}\|_E < \Lambda^* \alpha) = 1$$

za svako $\theta_1 \in (0, 1)$ i svako $\theta_2 > 1$.

Teorema 3.2.5. Neka je A proizvoljna realna matrica takva da je $s = 1$ (tj. $|\lambda_1| > |\lambda_2|$), $\lambda_1 \in \mathbb{R}$, $n_1 = 1$. Ako je $\vec{\varepsilon}^{(0)}$ uniformno rasporedjena u $\bar{S}(0, r)$, tada je

$$\lim_{r \rightarrow +\infty} P\left(\frac{\theta_1 |\lambda_1|}{1 - \lambda_1} \alpha < \|\vec{\varepsilon}^{(k)}\|_E < \frac{\theta_2}{1 - \lambda_1} \alpha\right) = 1$$

za svako $\theta_1 \in (0, 1)$ i svako $\theta_2 > 1$.

Primetimo da je ocena (3.2.10) oštija od (3.2.11) kada je $\Lambda < \Lambda^*$, a obrnut je slučaj kada je $\Lambda = \Lambda^*$ (inače uvek je $\Lambda \leq \Lambda^*$).

Primer 3.2.1. Neka je

$$B = \begin{bmatrix} -0.1 & 0.4 & 0.1 & 0.1 \\ 0.4 & -0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & -0.2 & 0.2 \\ 0.1 & 0.1 & 0.2 & -0.2 \end{bmatrix} \quad \text{i } \vec{\beta} = \begin{bmatrix} 0.7 \\ 0.7 \\ 1.2 \\ -1.6 \end{bmatrix}$$

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 $\vec{\varepsilon}^{(0)}$ u $\bar{S}(0, r)$ generiše pomoću metoda kongruentnosti (videti [53], [65]) i iterativni proces (3.2.2) startuje sa $\vec{x}^{(0)} = \vec{x} - \vec{\varepsilon}^{(0)}$, 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 $\theta_1 = \theta_2 = 1^*$.

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

* Sa stanovišta praktičnog računanja dopušten je izbor $\theta_1 = \theta_2 = 1$, jer je θ_i dovoljno blisko jedinici, tada je odgovarajući mašinski broj $fl(\theta_i) = 1$.

Do sada smo razmatrali dovoljne uslove za konvergenciju iterativnog procesa (3.2.2). Sledеća teorema daje potrebne i dovoljne uslove.

Teorema 3.2.6. Neka je $\vec{x}^{(0)}$ 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.

$$(3.2.12) \quad \vec{x}^{(k)} = B^k \vec{x}^{(0)} + (I+B+\dots+B^{k-1}) \vec{\beta} \quad (k=1, 2, \dots),$$

zaključujemo da je proces (3.2.2) ekvikonvergentan sa matričnim redom

$$(3.2.13) \quad I + B + B^2 + \dots = \sum_{m=0}^{+\infty} B^m.$$

S druge strane, kako su potrebni i dovoljni uslovi za konvergenciju reda (3.2.13) (videti odeljak 4.1.7)

$$(3.2.14) \quad |\lambda_i(B)| < 1 \quad (i=1, \dots, n),$$

dokaz teoreme 3.2.6 je završen.

Primedba 3.2.3. Pod uslovima (3.2.14) imamo

$$\lim_{k \rightarrow +\infty} B^k = 0 \quad \text{i} \quad \lim_{k \rightarrow +\infty} (I+B+\dots+B^{k-1}) = (I-B)^{-1}.$$

Tada iz (3.2.12) sleduje

$$\begin{aligned} \lim_{k \rightarrow +\infty} \vec{x}^{(k)} &= \lim_{k \rightarrow +\infty} [B^k \vec{x}^{(0)} + (I+B+\dots+B^{k-1}) \vec{\beta}] \\ &= \lim_{k \rightarrow +\infty} B^k \vec{x}^{(0)} + \lim_{k \rightarrow +\infty} (I+B+\dots+B^{k-1}) \vec{\beta} \\ &= (I-B)^{-1} \vec{\beta}, \end{aligned}$$

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

$$\lambda \mapsto \det(B - \lambda I) = \begin{vmatrix} b_{11} - \lambda & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} - \lambda & & b_{2n} \\ \vdots & & & \\ b_{n1} & b_{n2} & & b_{nn} - \lambda \end{vmatrix}$$

po modulu manje od jedinice.

Značaj uslova (3.2.14) u teorijskim razmatranjima je vrlo veliki. Međutim, 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

$$\begin{aligned} x^{(k)} &= \frac{1}{3} x^{(k-1)} - \frac{1}{9} y^{(k-1)} + \frac{1}{9}, \\ (3.2.15) \quad y^{(k)} &= 2x^{(k-1)} + \frac{1}{3} y^{(k-1)} - \frac{4}{3}, \end{aligned}$$

gde je $k = 1, 2, \dots$.

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,

$$\begin{aligned} \|B\|_\infty &= \max \left(\frac{1}{3} + \frac{1}{9}, 2 + \frac{1}{3} \right) = \frac{7}{3} > 1, \\ \|B\|_1 &= \max \left(\frac{1}{3} + 2, \frac{1}{9} + \frac{1}{3} \right) = \frac{7}{3} > 1, \\ \|B\| &= \left(\frac{1}{9} + \frac{1}{81} + 4 + \frac{1}{9} \right)^{1/2} = 2.057\dots > 1. \end{aligned}$$

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 $x^{(0)}$ i $y^{(0)}$.

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 [28]).

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

$$v(B) = -\log \rho(B).$$

Pri ovome, za ispunjenje uslova $\|\vec{x}^{(k)} - \vec{x}\| < \epsilon \|\vec{x}^{(0)} - \vec{x}\|$, gde je ϵ dovoljno mali pozitivan broj i $\rho(B) \neq 0$, potreban broj iteracija je približno dat sa $k \approx -\log \epsilon / v(B)$.

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 (u prethodno navedenom smislu) približno dat sa

$$N(B, \epsilon) = k C(B) \approx -\frac{C(B) \log \epsilon}{v(B)}.$$

Iterativni proces je efektivniji, ukoliko je ovaj broj manji.

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

$$(3.2.16) \quad 0 < m \leq \lambda_i(A) \leq M < +\infty \quad (i = 1, \dots, n).$$

S obzirom da se sopstvene vrednosti $\lambda_i(A)$ obično neznaju to se za m i M u (3.2.16) mogu uzeti neke ocene za $\lambda_{\min}(A)$ i $\lambda_{\max}(A)$ respektivno.

Posmatrajmo iterativni proces

$$(3.2.17) \quad \vec{x}^{(k)} = \vec{x}^{(k-1)} - c(A\vec{x}^{(k-1)} - \vec{b}),$$

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

$$(3.2.18) \quad \rho(B) = \max |1 - c\lambda_i(A)|.$$

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

$$\rho(B) = \frac{1}{2} |c| (M-m) \max_{-1 \leq \mu \leq 1} \left| \mu - \frac{2-c(M+m)}{c(M-m)} \right|.$$

Minimalna vrednost za $\rho(B)$ se dobija ako je $c = \frac{2}{M+m}$ i ona iznosi $\min \rho(B) = \frac{M-m}{M+m}$.

Dakle, dobili smo optimalni iterativni proces

$$(3.2.19) \quad \vec{x}^{(k)} = \vec{x}^{(k-1)} - \frac{2}{M+m}(A\vec{x}^{(k-1)} - \vec{b}).$$

Teorema 3.2.7. 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

$$(3.2.20) \quad \|\vec{x}^{(k)} - \vec{x}\|_E \leq \left(\frac{M-m}{M+m}\right)^k \|\vec{x}^{(0)} - \vec{x}\|_E \quad (k \in \mathbb{N}),$$

gde je \vec{x} tačno rešenje sistema $A\vec{x} = \vec{b}$.

Dokaz. 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 [7, str. 352-363].

Na kraju ovog odeljka, izložićemo jedan praktičan način za formiranje metoda proste iteracije.

Neka je dat sistem jednačina $\vec{Ax} = \vec{b}$, gde je $A = [a_{ij}]_{n \times n}$ i $\vec{b} = [b_1 \dots b_n]^T$, i neka je

$$D = \text{diag}(A) = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & & 0 \\ \vdots & & & \\ 0 & 0 & & a_{nn} \end{bmatrix}$$

regуларна матрица. Тада се овај систем може представити у еквивалентном облику

$$\vec{x} = D^{-1}(D-A)\vec{x} + D^{-1}\vec{b}.$$

Напоменимо да је одговарајући скаларни облик

$$x_1 = -\frac{a_{12}}{a_{11}} x_2 - \frac{a_{13}}{a_{11}} x_3 - \dots - \frac{a_{1n}}{a_{11}} x_n + \frac{b_1}{a_{11}},$$

$$x_2 = -\frac{a_{21}}{a_{22}} x_1 - \frac{a_{23}}{a_{22}} x_3 - \dots - \frac{a_{2n}}{a_{22}} x_n + \frac{b_2}{a_{22}},$$

⋮

$$x_n = -\frac{a_{n1}}{a_{nn}} x_1 - \frac{a_{n2}}{a_{nn}} x_2 - \dots - \frac{a_{n,n-1}}{a_{nn}} x_{n-1} + \frac{b_n}{a_{nn}}.$$

Na osnovu prethodnog, može se formirati metod proste iteracije

$$(3.2.21) \quad \vec{x}^{(k)} = D^{-1}(D - A)\vec{x}^{(k-1)} + D^{-1}\vec{b} \quad (k=1, 2, \dots),$$

koji je u literaturi poznat kao Jacobiev metod.

Kako je karakteristični polinom matrice $D^{-1}(D - A)$ dat sa

$$P(\lambda) = \det(D^{-1}(D - A) - \lambda I) = -\det(D^{-1})\det(\lambda D + (A - D)),$$

iz teoreme 3.2.6 sleduje da Jacobiev iterativni proces konvergira ako i samo ako su svi korenji jednačine

$$\begin{vmatrix} \lambda a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & \lambda a_{22} & & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & & \lambda a_{nn} \end{vmatrix} = 0$$

po modulu manji od jedinice.

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 $\vec{x}_i^{(k)}$ vektora $\vec{x}^{(k)}$ izračunava se na osnovu vrednosti $x_1^{(k-1)}, \dots, x_n^{(k-1)}$, tj.

$$x_i^{(k)} = \sum_{j=1}^n b_{ij} x_j^{(k-1)} + \beta_i \quad (i = 1, \dots, n; \quad k = 1, 2, \dots).$$

Ovaj metod može se modifikovati na taj način što bi se za izračunavanje vrednosti $x_i^{(k)}$ koristile vrednosti $x_i^{(k)}, \dots, x_{i-1}^{(k)}$, $x_i^{(k-1)}, \dots, x_n^{(k-1)}$, tj.

$$x_i^{(k)} = \sum_{j=1}^{i-1} b_{ij} x_j^{(k)} + \sum_{j=i}^n b_{ij} x_j^{(k-1)} + \vec{\beta}, \quad (i = 1, \dots, n; \quad k = 1, 2, \dots).$$

Navedena modifikacija metoda proste iteracije poznata je kao Gauss-Seidelov metod.

Iterativni proces (3.3.1) može se predstaviti i u matičnoj formi. Naime, neka je

$$B = B_1 + B_2,$$

gde su

$$B_1 = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ b_{21} & 0 & & 0 & 0 \\ \vdots & & & & \\ b_{n1} & b_{n2} & & b_{n,n-1} & 0 \end{bmatrix} \text{ i } B_2 = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ 0 & b_{22} & & b_{2n} \\ \vdots & & & \\ 0 & 0 & & b_{nn} \end{bmatrix}$$

Tada (3.3.1) postaje

$$(3.3.2) \quad \vec{x}^{(k)} = B_1 \vec{x}^{(k)} + B_2 \vec{x}^{(k-1)} + \vec{\beta} \quad (k=1,2,\dots).$$

Teorema 3.3.1. Pri proizvoljnom vektoru $\vec{x}^{(0)}$, iterativni proces (3.3.2) konvergira ako i samo ako su svi koreni jednačine

$$(3.3.3) \quad \det[B_2 - (I - B_1)\lambda] \equiv \begin{vmatrix} b_{11}^{-\lambda} & b_{12} & \dots & b_{1n} \\ b_{21}^{\lambda} & b_{22}^{-\lambda} & & b_{2n} \\ \vdots & & & \\ b_{n1}^{\lambda} & b_{n2}^{\lambda} & & b_{nn}^{-\lambda} \end{vmatrix} = 0$$

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

$$(I - B_1) \vec{x}^{(k)} = B_2 \vec{x}^{(k-1)} + \vec{\beta} \quad (k=1,2,\dots),$$

tj.

$$\vec{x}^{(k)} = (I - B)^{-1} B_2 \vec{x}^{(k-1)} + (I - B_1)^{-1} \vec{\beta} \quad (k=1,2,\dots)$$

Na osnovu teoreme 3.2.6, ovaj iterativni proces konvergira, pri proizvoljnom vektoru $\vec{x}^{(0)}$, ako i samo ako su svi koreni jednačine $\det[(I - B_1)^{-1} B_2 - \lambda I] = 0$ po modulu manji od jedinice.

Iz poslednje jednačine sleduje

$$\det[(I-B_1)^{-1}(B_2 - (I-B_1)\lambda)] = 0,$$

tj.

$$\det(I-B_1)^{-1} \det[B_2 - (I-B_1)\lambda] = 0.$$

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

$$C_1 = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ a_{21} & 0 & & 0 & 0 \\ \vdots & & & & \\ a_{n1} & a_{n2} & a_{n,n-1} & 0 \end{bmatrix} \quad i \quad C_2 = \begin{bmatrix} 0 & a_{12} & \dots & a_{1n} \\ 0 & 0 & & a_{2n} \\ 0 & 0 & & 0 \end{bmatrix},$$

može se obrazovati Gauss-Seidelov proces kao

$$(3.3.4) \quad \vec{x}^{(k)} = -D^{-1}C_1\vec{x}^{(k)} - D^{-1}C_2\vec{x}^{(k-1)} + D^{-1}\vec{b} \quad (k = 1, 2, \dots).$$

Ova varijanta Gauss-Seidelovog metoda ponekad se ne naziva metod Nekrasova (videti [46]).

Iz teoreme 3.3.1 sleduje sledeća teorema:

Teorema 3.3.2. Pri proizvoljnom vektoru $\vec{x}^{(0)}$, iterativni proces (3.3.4) konvergira ako i samo ako su svi korenji jednačine

$$\det[C_2 + (D+C_1)\lambda] = \begin{vmatrix} a_{11}\lambda & a_{12} & \dots & a_{1n} \\ a_{21}\lambda & a_{22}\lambda & & a_{2n} \\ \vdots & & & \\ a_{n1}\lambda & a_{n2}\lambda & & a_{nn}\lambda \end{vmatrix} = 0$$

po modulu manji od jedinice.

Kao što je i ranije napomenuto, ovi spektralni uslovi za konvergenciju iterativnih procesa, nažalost, nemaju veliki praktični značaj.

Za sistem jednačina sa simetričnom matricom, E.Reich ([49]) 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

$$a_{11}, \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix}, \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix}, \dots, \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{1n} & \dots & a_{nn} \end{vmatrix}$$

pozitivne.

L.Collatz ([13],[14]) dokazao je sledeći rezultat:

Teorema 3.3.4. Iterativni procesi Jacobia (3.2.21) i Gauss-Seidel (3.3.4) konvergiraju, ako matrica A reda n ispunjava sledeća dva uslova:

1° Matrica A ne sadrži nula-submatricu tipa $m \times (n-m)$ ($1 \leq m \leq n-1$),

2° za svako $i \in I = \{1, \dots, n\}$ je $|a_{ii}| \geq s_i = \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|$,

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

$$(3.3.5) \quad \begin{aligned} 10x_1 + 3x_2 - x_3 &= 12, \\ -x_1 + 5x_2 - x_3 &= 3, \\ x_1 + 2x_2 + 10x_3 &= 13. \end{aligned}$$

Kako elementi matrice datog sistema

$$A = \begin{bmatrix} 10 & 3 & -1 \\ -1 & 5 & -1 \\ 1 & 2 & 10 \end{bmatrix}$$

važi

$$|a_{11}| = 10 > s_1 = |a_{12}| + |a_{13}| = 4,$$

$$|a_{22}| = 5 > s_2 = |a_{21}| + |a_{23}| = 2,$$

$$|a_{33}| = 10 > s_3 = |a_{31}| + |a_{32}| = 3,$$

i kako A ne sadrži nula-submatricu tipa 1×2 ili tipa 2×1 , zaključujemo da su uslovi 1° i 2° 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

$$x_1^{(k)} = -0.3x_2^{(k-1)} + 0.1x_3^{(k-1)} + 1.2,$$

$$x_2^{(k)} = 0.2x_1^{(k)} + 0.2x_3^{(k-1)} + 0.6,$$

$$x_3^{(k)} = -0.1x_1^{(k)} - 0.2x_2^{(k)} + 1.3,$$

gde je $k = 1, 2, \dots$, dobijamo niz

$$\vec{x}^{(1)} = [1.1500000 \ 1.0900000 \ 0.9670000]^T,$$

$$\vec{x}^{(2)} = [0.9697000 \ 0.9873400 \ 1.0055620]^T,$$

$$\vec{x}^{(3)} = [1.0043542 \ 1.0019832 \ 0.9991680]^T,$$

$$\vec{x}^{(4)} = [0.9993219 \ 0.9996979 \ 1.0001283]^T,$$

itd. Primetimo da je tačno rešenje sistema (3.3.5)

$$\vec{x} = [1 \ 1 \ 1]^T.$$

Jedno interesantno pitanje koje se može postaviti u vezi sa razmatranim iterativnim metodama je: Da li Gauss-Seidelov metod uvek konvergira kada konvergira odgovarajući metod proste

iteracije? Odgovor na ovo pitanje nije potvrđan. Naime, u doista velikom broju slučajeva, ako metod proste iteracije konvergira, konvergiraće i Gauss-Seidelov metod i to brže, međutim, 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

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \vec{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \quad B = \begin{bmatrix} p & q \\ -q & p \end{bmatrix} \quad (p, q \in \mathbb{R}).$$

- Odredićemo parametre p i q tako da konvergira
- 1° metod proste iteracije,
 - 2° 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 [16], [69])

$$|2p-q^2| < p^2 + 1 \quad \text{i} \quad |p^2| < 1,$$

odakle, nakon elementarnih algebarskih transformacija, dobijamo

$$(3.3.6) \quad |q| < 1 + p \quad \text{i} \quad |p| < 1.$$

U ravni p_0q , nejednakosti (3.3.6) definišu oblast prikaza- nu 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.

Pridjimo sada na određivanje greške približnog rešenja koje se dobija primenom iterativnog procesa (3.3.2).

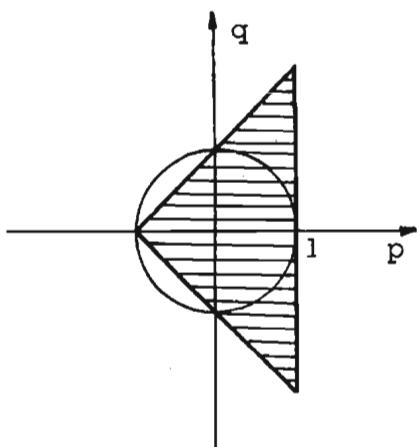
Sl. 3.3.1

Kako je tačno rešenje \vec{x} nepoznato, greška $\vec{\epsilon}^{(k)} = \vec{x}^{(k)} - \vec{x}$, se ne može tačno odrediti. Međutim, kako se greška $\vec{\epsilon}^{(k)}$ može izražiti pomoću priraštaja $\vec{\delta}^{(k)} = \vec{x}^{(k)} - \vec{x}^{(k-1)}$, moguće je dati ocenu za $\|\vec{\epsilon}^{(k)}\|$.

Teorema 3.3.5. Ako je \vec{x} rešenje jednačine

$$\vec{x} = B\vec{x} + \vec{\beta} \quad (B = B_1 + B_2)$$

i ako je $\|B\| \leq q < 1$, važi nejednakost



$$(3.3.7) \quad \|\vec{x}^{(k)} - \vec{x}\| \leq \frac{\|B_2\|}{1-\|B\|} \|\vec{x}^{(k)} - \vec{x}^{(k-1)}\| \quad (k \in \mathbb{N}),$$

gde se niz $\{x^{(k)}\}$ generiše pomoću (3.3.2).

Dokaz. Na osnovu (3.3.2), za svako $k \in \mathbb{N}$, 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

$$\vec{\varepsilon}^{(k)} = B_1 \vec{\varepsilon}^{(k)} + B_2 (\vec{\varepsilon}^{(k)} - \vec{\delta}^{(k)}) \quad (k \in \mathbb{N}),$$

tj.

$$\vec{\varepsilon}^{(k)} = -(I-B)^{-1} B_2 \vec{\delta}^{(k)} \quad (k \in \mathbb{N}),$$

s obzirom da egzistira $(I-B)^{-1}$ (na osnovu pretpostavke $\|B\| \leq q < 1$).

Ako koristimo normu matrice saglasnu sa normom vektora, iz poslednje jednakosti dobijamo

$$\|\vec{\varepsilon}^{(k)}\| \leq \|(I-B)^{-1} B_2\| \cdot \|\vec{\delta}^{(k)}\| \leq \frac{\|B_2\|}{1-\|B\|} \|\vec{\delta}^{(k)}\| \quad (k \in \mathbb{N}),$$

čime smo dokazali teoremu 3.3.5.

Primedba 3.3.1. Ako je $B_1 = 0$ ($\Rightarrow B_2 = B$), (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

$$(3.4.1) \quad A\vec{x} = \vec{b}.$$

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

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, međutim, 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 metodama potiču od Gaussa. Međutim, 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 [1], [6], [25], [54], [61], [68]).

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 $D = \text{diag}(A)$ regularna, Razložimo matricu A u obliku

$$A = C_1 + D + C_2,$$

gde su

$$C_1 = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ a_{21} & 0 & & 0 & 0 \\ \vdots & & & & \\ a_{n1} & a_{n2} & & a_{n,n-1} & 0 \end{bmatrix} \quad i \quad C_2 = \begin{bmatrix} 0 & a_{12} & \dots & a_{1n} \\ 0 & 0 & & a_{2n} \\ \vdots & & & \\ 0 & 0 & & 0 \end{bmatrix}.$$

Ako za izračunavanje vektora oстатка \vec{r} uzmemos za x_i последњу израчунату вредност (како код Gauss-Seidelovog метода), тј.

$$\vec{r} = C_1 \vec{x}^{(k)} + (D+C_2) \vec{x}^{(k-1)} - \vec{b}$$

i ако ставимо $\vec{\delta}^{(k)} = \vec{x}^{(k)} - \vec{x}^{(k-1)}$, можемо формирати итеративни процес у облику

$$B\vec{\delta}^{(k)} = -\omega \vec{r} \quad (k = 1, 2, \dots),$$

тј.

$$(3.5.1) \quad D\vec{x}^{(k)} = D\vec{x}^{(k-1)} + \omega \left[\vec{b} - C_1 \vec{x}^{(k)} - (D+C_2) \vec{x}^{(k-1)} \right],$$

где је $k = 1, 2, \dots$ и ω реалан параметар. Параметар ω називамо рељасациони мноžилак и нега у општем случају можемо менјати у токурачунанja.

Итеративни процес (3.5.1) може се представити и у скаларном облику

$$a_{ii} x_i^{*(k)} = - \sum_{j < i} a_{ij} x_j^{(k)} - \sum_{j > i} a_{ij} x_j^{(k-1)} + b_i,$$

$$x_i^{(k)} = x_i^{(k-1)} + \omega (x_i^{*(k)} - x_i^{(k-1)})$$

где су $i = 1, \dots, n$ и $k = 1, 2, \dots$.

Приметимо да се за $\omega = 1$, (3.5.1) своди на Gauss-Seidelов процес (3.3.7).

На основу (3.5.1) имамо

$$(3.5.2) \quad (D + \omega C_1) \vec{x}^{(k)} = [(1-\omega)D - \omega C_2] \vec{x}^{(k-1)} + \omega \vec{b} \quad (k = 1, 2, \dots),$$

tj.

$$(3.5.3) \quad \vec{x}^{(k)} = K(\omega) \vec{x}^{(k-1)} + \vec{f}(\omega) \quad (k = 1, 2, \dots),$$

gde su

$$K(\omega) = (D + \omega C_1)^{-1} [(1-\omega)D - \omega \zeta] \quad i \quad \vec{f}(\omega) = \omega (D + \omega C_1)^{-1} \vec{b}.$$

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 ([14]).

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 $a_{ii} > 0$ ($i = 1, \dots, n$) 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 svaku \vec{y}

$$(3.5.4) \quad \vec{y}^* (C_1 - C_2) \vec{y} = ic \quad (c \in \mathbb{R}).$$

Prepostavimo sada da je λ bilo koja sopstvena vrednost matrice $K(\omega)$ i \vec{y} odgovarajući sopstveni vektor. Tada je

$$K(\omega) \vec{y} = \lambda \vec{y},$$

tj.

$$(3.5.5) \quad [(1-\omega)D - \omega C_2] \vec{y} = \lambda (D + \omega C_1) \vec{y}.$$

Množenjem jednakosti (3.5.5) sleva sa \vec{y}^* , dobijamo

$$\vec{y}^* [(2-\omega)D - \omega(D+2C_2)] \vec{y} = \lambda \vec{y}^* [(2-\omega)D + \omega(D+2C_1)] \vec{y},$$

tj.

$$(3.5.6) \quad \vec{y}^* [(2-\omega)D - \omega A + \omega(C_1 - C_2)] \vec{y} = \lambda \vec{y}^* [(2-\omega)D + \omega A + \omega(C_1 - C_2)] \vec{y},$$

s obzirom da je $A = C_1 + D + C_2$.

Kako su matrice A i D pozitivno definitne imamo

$$(3.5.7) \quad \vec{y}^* A \vec{y} = a > 0 \quad \text{i} \quad \vec{y}^* D \vec{y} = d > 0 \quad (\forall \vec{y} \neq \vec{0}).$$

Na osnovu (3.5.6), (3.5.4), (3.5.7) dobijamo

$$\lambda = \frac{\xi d - a + c_i}{\xi d + a + c_i} \quad (\xi = \frac{2}{\omega} - 1),$$

odakle pod pretpostavkom da je

$$0 < \xi < +\infty, \quad \text{tj.} \quad 2 > \omega > 0$$

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

$$(3.5.8) \quad \begin{aligned} \omega A \vec{x}^{(k)} &= [(1-\omega)D - \omega C_2] (\vec{x}^{(k-1)} - \vec{x}^{(k)}) + \omega \vec{b}, \\ \omega A (\vec{x}^{(k)} - \vec{x}) &= [(1-\omega)D - \omega C_2] (\vec{x}^{(k-1)} - \vec{x}^{(k)}), \\ \vec{x}^{(k)} - \vec{x} &= A^{-1} (\gamma D + C_2) (\vec{x}^{(k)} - \vec{x}^{(k-1)}), \end{aligned}$$

gde je $\gamma = 1 - \frac{1}{\omega}$.

Iz (3.5.8) sleduje:

Teorema 3.5.3. Kod iterativnog procesa (3.5.1) važi ocena za grešku

$$(3.5.9) \quad \|\vec{x}^{(k)} - \vec{x}\| \leq \|A^{-1}(\gamma D + C_2)\| \cdot \|\vec{x}^{(k)} - \vec{x}^{(k-1)}\| \quad (k \in \mathbb{N}).$$

Izraz (3.5.9), koji daje ocenu greške, najopštiji je za iterativni proces (3.5.1), jer ne zahteva nikakva ograničenja za matričnu sistem A (sem da je regularna). Međutim, zbog komplikovanosti izraza $\|A^{-1}(\gamma D + C_2)\|$, ocena (3.5.9) nije pogodna. Zato ćemo, nadalje, za matricu A prepostaviti dodatne uslove (kao u teoremi 3.5.2).

Dakle, neka je matrica A hermitska pozitivno definitna. Tada je D takođe pozitivno definitna, pa postoji matrica $D^{1/2}$ i njena inverzna matrica $D^{-1/2}$.

Uvodjenjem oznaka i transformacija

$$C = C_1 + C_2, \quad D^{-1/2} A D^{-1/2} = B, \quad D^{-1/2} C D^{-1/2} = -T,$$

$$D^{-1/2} C_1 D^{-1/2} = -T_1, \quad D^{-1/2} C_2 D^{-1/2} = -T_2, \quad D^{1/2} \vec{x} = \vec{y}, \quad D^{-1/2} \vec{B} = \vec{\beta},$$

sistem $A\vec{x} = \vec{B}$ se svodi na

$$(3.5.10) \quad \vec{B}\vec{y} = \vec{\beta}, \quad \text{tj.} \quad \vec{y} = T\vec{y} + \vec{\beta}.$$

Primetimo da je matrica B pozitivno definitna i T hermitska.

Greška kod relaksacionog metoda primenjenog na (3.5.10), tj. kod metoda

$$(3.5.11) \quad (I - \omega T_1) \vec{y}'^{(k)} = [(1-\omega)I + \omega T_2] \vec{y}^{(k-1)} + \omega \vec{\beta} \quad (k = 1, 2, \dots),$$

je

$$(3.5.12) \quad \|\vec{y}^{(k)} - \vec{y}\| \leq \|B^{-1}(\gamma I - T_2)\| \cdot \|\vec{y}^{(k)} - \vec{y}^{(k-1)}\| \quad (k \in \mathbb{N}),$$

gde je $B = I - T$.

Kod primene relaksacionog metoda veoma je važan izbor vrednosti relaksacionog parametra ω . U literaturi (videti [1]),

[6], [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)

$$(3.5.13) \quad K(\omega) = (I - \omega T_1)^{-1} [(1-\omega)I + \omega T_2].$$

J. Albrecht ([1]) je razmatrao iterativni proces (3.5.11), pri uslovu da je hermitska matrica T (u sistemu (3.5.10)) ciklična indeksa 2 (videti [61]), tj. da je oblika

$$(3.5.14) \quad T = \begin{bmatrix} 0 & | & F \\ \cdots & | & \cdots \\ E & | & 0 \\ \hline n_2 & & n_2 \end{bmatrix} \quad \begin{array}{l} n_1 \text{ vrsta} \\ (n_1 + n_2 = n) \\ n_2 \text{ vrsta} \\ \hline \text{kolona} & \text{kolona} \end{array}$$

i da je $\rho(T) < 1$. Matrica T ima svojstvo (A) (videti definiciju 1.1.2).

Na osnovu prethodnog imamo

$$T_1 = \begin{bmatrix} 0 & | & 0 \\ \cdots & | & \cdots \\ E & | & 0 \end{bmatrix} \quad i \quad T = \begin{bmatrix} 0 & | & F \\ \cdots & | & \cdots \\ 0 & | & 0 \end{bmatrix}.$$

Teorema 3.5.4. Neka su T i $K(\omega)$ matrice date pomoću (3.5.14) i (3.5.13) respektivno. Ako je $\{\lambda\}$ sopstvena vrednost od $\{K(\omega)\}$ i važi

$$(3.5.15) \quad (\lambda + \omega - 1)^2 = \lambda \omega^2 \tau^2,$$

tada je $\{\frac{\lambda}{\tau}\}$ sopstvena vrednost od $\{T\}$.

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 $-\tau$ takođe 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 $0 < \omega \leq \omega_1$ ($\omega_1 = \omega_1(\tau) = \frac{1}{1+\sqrt{1-\tau^2}}$), korenii λ_{\pm} su realni i pozitivni i veći od njih je

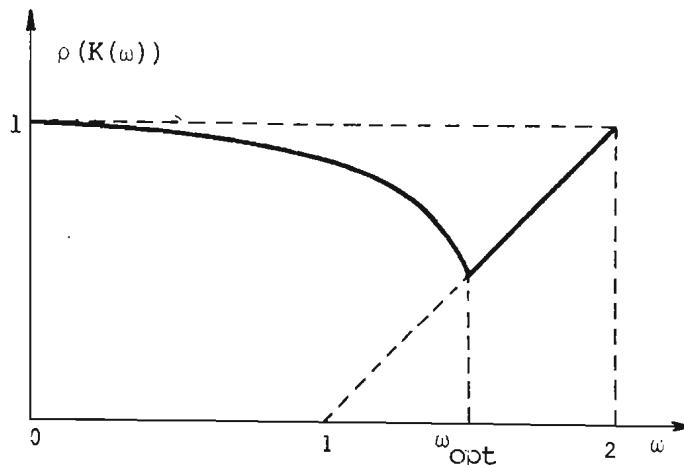
$$\lambda_+ = f_+(\omega, \tau) = \frac{1}{4}(\omega + \tau + \sqrt{\omega^2 \tau^2 - 4\omega + 4})^2.$$

Ako je $\omega_1 < \omega < 2$, korenii λ_{\pm} su kompleksni sa modulom $\omega - 1$.

Kako je funkcija $|\tau| \mapsto f_+(\omega, \tau)$ rastuća na $(0, \rho)$, zaključujemo da je

$$\rho(K(\omega)) = \begin{cases} f_+(\omega, \tau) & (0 < \omega < \frac{2}{1+\sqrt{1-\rho^2}}), \\ \omega - 1 & (\frac{2}{1+\sqrt{1-\rho^2}} < \omega < 2). \end{cases}$$

Grafik funkcije $\omega \mapsto \rho(K(\omega))$ ($0 < \omega < 2$) prikazan je na slici 3.5.1.



S1. 3.5.1

Optimalna vrednost parametra ω je

$$\omega_{\text{opt}} = \omega_1(\rho) = \frac{2}{1+\sqrt{1-\rho^2}},$$

s obzirom da funkcija $\omega \mapsto \rho(K(\omega))$ ima minimum za $\omega = \omega_{\text{opt}}$ na segmentu $(0, 2)$. Odgovarajuća vrednost za $\rho(K(\omega))$ je

$$\rho(K(\omega_{\text{opt}})) = \omega_{\text{opt}}^{-1} = \frac{1-\sqrt{1-\rho^2}}{1+\sqrt{1+\rho^2}}.$$

Uvedimo oznaku

$$M(\omega) = [(\mathbf{I}-T)^{-1}(\gamma\mathbf{I}-T_2)]^* \cdot [(\mathbf{I}-T)^{-1}(\gamma\mathbf{I}-T_2)].$$

Teorema 3.5.5. Ako je $\{\frac{\tau}{\mu}\}$ sopstvena vrednost od $\{M(\omega)\}$ i važi

$$\frac{1}{2} \left\{ \frac{1-\tau^2}{\gamma^2} \mu + \frac{1}{1-\tau^2} \right\} = 1 + \frac{1}{2} \frac{\xi^2 \tau^2 + \tau^4}{\gamma^2 (1-\tau^2)},$$

gde su $\gamma = 1 - \frac{1}{\omega}$ i $\xi = \frac{2}{\omega} - 1$, tada je $\{\frac{\mu}{\tau}\}$ sopstvena vrednost od $\{\frac{M(\omega)}{T}\}$.

Na osnovu teoreme 3.5.5., nalazimo

$$H(\omega) = \|(\mathbf{I}-T)^{-1}(\gamma\mathbf{I}-T_2)\|_{\text{sp}} = \frac{1}{2} \frac{\sqrt{\frac{2}{\rho} (\xi^2 + \rho^2) + 4\gamma^2 (1-\rho^2)} + \sqrt{\frac{2}{\rho} (\xi^2 + \rho^2)}}{1-\rho^2},$$

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

$$\|\vec{y}^{(k)} - \vec{y}\|_E \leq H(\omega) \cdot \|\vec{y}^{(k)} - \vec{y}^{(k-1)}\|_E \quad (k \in \mathbb{N}).$$

Napomenimo da je kod Gauss-Seidelovog metoda, tj. kada je relaksacioni parametar $\omega = 1$,

$$H = H(1) = \frac{\rho \sqrt{1 - \rho^2}}{1 - \rho^2}.$$

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

$$(3.6.1) \quad \vec{x} = B\vec{x} + \vec{\beta}$$

sa hermitskom matricom B reda n . Ako je $\rho(B) < 1$, iterativni proces

$$(3.6.2) \quad \vec{x}^{(k)} = B\vec{x}^{(k-1)} + \vec{\beta} \quad (k = 1, 2, \dots)$$

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.

$$(3.6.3) \quad \vec{y}^{(k)} = \sum_{i=0}^k c_{ki} \vec{x}^{(i)} \quad (k = 0, 1, \dots),$$

pod uslovom $\sum_{i=0}^k c_{ki} = 1$.

Ako stavimo $\vec{\epsilon}^{(k)} = \vec{x}^{(k)} - \vec{x}$ i $\vec{\zeta}^{(k)} = \vec{y}^{(k)} - \vec{x}$ ($k = 0, 1, \dots$), imamo

$$\vec{\zeta}^{(k)} = \sum_{i=0}^k c_{ki} \vec{\epsilon}^{(i)} = \left(\sum_{i=0}^k c_{ki} B^i \right) \vec{\epsilon}^{(0)},$$

tj.

$$(3.6.4) \quad \vec{\zeta}^{(k)} = Q_k(B) \vec{\epsilon}^{(0)},$$

*) Često se ovakav postupak za ubrzavanje konvergencije naziva linearno ubrzavanje (videti [20]).

gde je

$$(3.6.5) \quad Q_k(t) = \sum_{i=0}^k c_{ki} t^i \quad i \quad Q_k(1) = 1.$$

Neka je $\{\vec{e}_i\}$ ortonormiran sistem sopstvenih vektora hermitske matrice B , gde je $B\vec{e}_i = \lambda_i \vec{e}_i$ ($\lambda_i \in \lambda(B)$) ($i = 1, \dots, n$). Ako vektor $\vec{\varepsilon}^{(0)}$ razvijamo po sopstvenim vektorima \vec{e}_i , tj.
 $\vec{\varepsilon}^{(0)} = p_1 \vec{e}_1 + \dots + p_n \vec{e}_n$, jednakost (3.6.4) postaje

$$\vec{\zeta}(k) = \sum_{i=1}^n p_i Q_k(\lambda_i) \vec{e}_i.$$

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

$$(3.6.6) \quad Q_{k+1}(t) = (\alpha_k t + 1 - \alpha_k - \beta_k) Q_k(t) + \beta_k Q_{k-1}(t) \quad (k=0, 1, \dots),$$

gde su $Q_0(t) = 1$, $\beta_0 = 0$, α_k i β_k realni brojevi.

Kako je

$$\|\vec{\zeta}^{(k)}\|_E = \left(\sum_{i=1}^n p_i^2 Q_k(\lambda_i)^2 \right)^{1/2} \leq \max_i |Q_k(\lambda_i)| \left(\sum_{i=1}^n p_i^2 \right)^{1/2},$$

tj.

$$\|\vec{\zeta}^{(k)}\|_E \leq \|\vec{\varepsilon}^{(0)}\|_E \max_i |Q_k(\lambda_i)|,$$

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 $\rho \in \rho(B)$. Tada dobijamo (videti [1] i [25])

$$(3.6.7) \quad \tilde{Q}_k(t) = \frac{T_k(t/\rho)}{T_k(1/\rho)} \quad (k = 0, 1, \dots),$$

gde je T_k Čebiševljev polinom stepena k. Lako je pokazati da se, u ovom slučaju, rekurentna relacija (3.6.6) svodi na

$$(3.6.8) \quad \hat{Q}_{k+1}(t) = \alpha_k t \hat{Q}_k(t) + (1-\alpha_k) \hat{Q}_{k-1}(t) \quad (k=0,1,\dots),$$

gde su

$$(3.6.9) \quad \alpha_0 = 1, \quad \beta_0 = 0, \quad \alpha_k = \frac{2}{\rho} \cdot \frac{T_k(1/\rho)}{T_{k+1}(1/\rho)}, \quad \beta_k = 1 - \alpha_k \quad (k=1,2,\dots).$$

Dakle, na osnovu (3.6.8) generiše se niz polinoma $\{\hat{Q}_k\}$, 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

$$\hat{Q}_{k+1}(t) - \hat{Q}_{k-1}(t) = \alpha_k (t \hat{Q}_k(t) - \hat{Q}_{k-1}(t)).$$

S druge strane, kako je $\vec{\zeta}^{(k+1)} - \vec{\zeta}^{(k-1)} = (\hat{Q}_{k+1}(B) - \hat{Q}_{k-1}(B)) \vec{\epsilon}^{(0)}$

imamo

$$\vec{\zeta}^{(k+1)} - \vec{\zeta}^{(k-1)} = \alpha_k (B \hat{Q}_k(B) - \hat{Q}_{k-1}(B)) \vec{\epsilon}^{(0)} = \alpha_k (B \vec{\zeta}^{(k)} - \vec{\zeta}^{(k-1)}),$$

tj.

$$(3.6.10) \quad \vec{y}^{(k+1)} = \vec{y}^{(k-1)} + \alpha_k (B \vec{y}^{(k)} + \vec{\beta} - \vec{y}^{(k-1)}) \quad (k=1,2,\dots),$$

gde je niz $\{\alpha_k\}$ definisan pomoću (3.6.9) i

$$\vec{y}^{(0)} = \vec{x}^{(0)} \quad \text{i} \quad \vec{y}^{(1)} = B \vec{x}^{(0)} + \vec{\beta}.$$

Na osnovu prethodnog, članovi niza $\{\alpha_k\}$ su

$$\alpha_0 = 1, \quad \alpha_1 = \frac{2}{2-\rho}, \quad \alpha_k = \frac{1}{1-\frac{1}{4}\rho^2 \alpha_{k-1}} \quad (k=2,3,\dots),$$

pri čemu je

$$2 > \alpha_1 > \alpha_2 > \dots \geq \lim_{k \rightarrow +\infty} \alpha_k = \omega_{\text{opt}} = \frac{2}{1+\sqrt{1-\rho^2}} > 1.$$

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 ([39]).

Primedba 3.6.2. E.Stiefel ([56]) 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 funkcionala 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 $F: \mathbb{R}^n \rightarrow \mathbb{R}$, definisane pomoću

$$(3.7.2) \quad F(\vec{x}) = (A\vec{x}, \vec{x}) - 2(\vec{b}, \vec{x}).$$

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 funkcionala F postiže minimum za $\vec{x} = A^{-1}\vec{b}$, što predstavlja rešenje sistema (3.7.1).

1. Metod najbržeg pada. Svaki metod oblika

$$(3.7.3) \quad \vec{x}^{(k)} = \vec{x}^{(k-1)} - \alpha_{k-1} \text{grad}F(\vec{x}^{(k-1)}) \quad (k = 1, 2, \dots)$$

naziva se metod gradijentnog pada.

Ako se parametar α_p određuje iz uslova da je veličina

$$(3.7.4) \quad F(\vec{x}^{(p)}) - \alpha_p \text{grad}F(\vec{x}^{(p)}) \quad (\forall p \in N_0)$$

minimalna, metod (3.7.3) se naziva metod najbržeg pada.

Iz (3.7.2) sledi $\text{grad}F(x) = 2(A\vec{x} - \vec{b})$. Tada je

$$(3.7.5) \quad \vec{x}^{(k)} = \vec{x}^{(k-1)} - 2\alpha_{k-1}(A\vec{x}^{(k-1)} - \vec{b}) \quad (k = 1, 2, \dots).$$

Ako u jednakosti

$$F(\vec{x} + t\vec{r}) = F(\vec{x}) + 2t(A\vec{x} - \vec{b}, \vec{r}) + t^2(A\vec{r}, \vec{r}) \quad (t \in \mathbb{R})$$

izvršimo supstituciju

$$\begin{pmatrix} \vec{x} & \vec{r} & t \\ \vec{x}^{(p)} & A\vec{x}^{(p)} - \vec{b} & -2\alpha_p \end{pmatrix} \quad (p \in N_0),$$

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

$$(3.7.6) \quad 2\alpha_p = \frac{(\vec{r}^{(p)}, \vec{r}^{(p)})}{(A\vec{r}^{(p)}, \vec{r}^{(p)})}.$$

Na osnovu (3.7.5) i (3.7.6) dobijamo metod najvećeg pada

$$(3.7.7) \quad \vec{x}^{(k)} = \vec{x}^{(k-1)} - \frac{\|A\vec{x}^{(k-1)} - \vec{b}\|_E^2}{(A(A\vec{x}^{(k-1)} - \vec{b}), A\vec{x}^{(k-1)} - \vec{b})} (A\vec{x}^{(k-1)} - \vec{b}),$$

gde je $k = 1, 2, \dots$.

Teorema 3.7.1. Neka je \vec{x} tačno rešenje sistema (3.7.1) čija je matrica A normalna. Za metod najvećeg pada važe nejednakosti

$$(3.7.8) \quad F(\vec{x}^{(k)}) - F(\vec{x}) \leq \left(\frac{M-m}{M+m} \right)^2 (F(\vec{x}^{(k-1)}) - F(\vec{x})),$$

$$(3.7.9) \quad \|\vec{x}^{(k)} - \vec{x}\|_E \leq \left(\frac{M-m}{M+m} \right)^k \sqrt{\frac{M}{m}} \|\vec{x}^{(0)} - \vec{x}\|_E,$$

gde je $\vec{x}^{(0)}$ proizvoljan početni vektor i $0 < m \leq \lambda_1(A) \leq M$.

Dokaz. Neka je $\vec{x}^{(k-1)}$ dobijeno pomoću (3.7.7), polazeći od proizvoljnog vektora $\vec{x}^{(0)}$. Stavimo $\vec{y}^{(k-1)} = \vec{x}^{(k-1)}$ i odredimo $\vec{y}^{(k)}$ koristeći optimalni metod (3.2.16), tj.

$$(3.7.10) \quad \vec{y}^{(k)} = \vec{y}^{(k-1)} - \frac{2}{M+m} (A\vec{y}^{(k-1)} - \vec{b}).$$

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 $2\alpha_{k-1}$ određeno pomoću (3.7.6), zaključujemo da važi nejednakost

$$(3.7.11) \quad F(\vec{x}^{(k)}) \leq F(\vec{y}^{(k)}).$$

Neka je $\{\vec{e}_i\}$ ortonormirani sistem sopstvenih vektora matrice A. Tada vektor $\vec{y}^{(k-1)} = \vec{y}^{(k-1)} - \vec{x}$ možemo predstaviti u obliku $\vec{y}^{(k-1)} = c_1 \vec{e}_1 + \dots + c_n \vec{e}_n$. Kako je $A\vec{e}_i = \lambda_i \vec{e}_i$ ($\lambda_i \leq \lambda_1(A)$), imamo

$$(3.7.12) \quad (A\vec{y}^{(k-1)}, \vec{y}^{(k-1)}) = \sum_{i=1}^n c_i^2 \lambda_i.$$

Na osnovu (3.7.10) imamo $B = I - \frac{2}{M+m} A$ i

$$\vec{y}^{(k)} = \vec{y}^{(k)} - \vec{x} = B(\vec{y}^{(k-1)} - \vec{x}) = B\vec{y}^{(k-1)}.$$

Nadalje, kako je matrica B simetrična sa sopstvenim vre-

dnostima $1 - \frac{2}{M+m} \lambda_i$ ($i = 1, \dots, n$) imamo

$$(A\zeta^{(k)}, \zeta^{(k)}) = (AB\zeta^{(k-1)}, B\zeta^{(k-1)}) = (BAB\zeta^{(k-1)}, \zeta^{(k-1)}),$$

tj.

$$(3.7.13) \quad (A\zeta^{(k)}, \zeta^{(k)}) = \sum_{i=1}^n c_i^2 \lambda_i (1 - \frac{2}{M+m} \lambda_i)^2,$$

jer su sopstvene vrednosti matrice BAB odredjene sa

$$\lambda_i (1 - \frac{2}{M+m} \lambda_i)^2 \quad (i = 1, \dots, n).$$

S druge strane, kako je

$$\left| 1 - \frac{2}{M+m} \lambda_i \right| \leq \frac{M-m}{M+m} \quad (i = 1, \dots, n),$$

na osnovu (3.7.12) i (3.7.13) zaključujemo da je

$$(3.7.14) \quad (A\zeta^{(k)}, \zeta^{(k)}) \leq \left(\frac{M-m}{M+m} \right)^2 (A\zeta^{(k-1)}, \zeta^{(k-1)}),$$

tj.

$$(3.7.15) \quad F(\vec{y}^{(k)}) - F(\vec{x}) \leq \left(\frac{M-m}{M+m} \right)^2 (F(\vec{y}^{(k-1)}) - F(\vec{x})).$$

Kako je $\vec{y}^{(k-1)} = \vec{x}^{(k-1)}$, iz (3.7.11) i (3.7.15) sledi 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 $\vec{\varepsilon}^{(k-1)} = \vec{\zeta}^{(k-1)}$ i

$$(A\vec{\varepsilon}^{(k)}, \vec{\varepsilon}^{(k)}) = F(\vec{x}^{(k)}) - F(\vec{x}) \leq F(\vec{y}^{(k)}) - F(\vec{x}) = (A\zeta^{(k)}, \zeta^{(k)}).$$

Tada iz (3.7.14) sledi

$$(A\vec{\varepsilon}^{(k)}, \vec{\varepsilon}^{(k)}) \leq \left(\frac{M-m}{M+m} \right)^2 (A\vec{\varepsilon}^{(k-1)}, \vec{\varepsilon}^{(k-1)}).$$

Iteriranjem poslednje nejednakosti dobijamo

$$(A\vec{\varepsilon}^{(k)}, \vec{\varepsilon}^{(k)}) \leq \left(\frac{M-m}{M+m}\right)^{2k} (A\vec{\varepsilon}^{(0)}, \vec{\varepsilon}^{(0)}),$$

odakle je, s obzirom na teoremu 1.4.6,

$$m \|\vec{\varepsilon}^{(k)}\|_E^2 \leq \left(\frac{M-m}{M+m}\right)^{2k} m \|\vec{\varepsilon}^{(0)}\|_E^2,$$

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 [31] i [55] 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

$$\vec{x}^{(k)} = \vec{x}^{(0)} + \sum_{i=0}^{k-1} \alpha_i A^i \vec{r}^{(0)} \quad (k = 1, \dots, n),$$

polazeći od proizvoljnog početnog vektora $\vec{x}^{(0)}$, pri čemu je $\vec{r}^{(0)} = A\vec{x}^{(0)} - \vec{b}$, dok se koeficijenti α_i određuju iz uslova da funkcija $(\alpha_0, \dots, \alpha_{k-1}) \mapsto F(\vec{x}^{(k)})$ postiže minimum. Iz praktičnih razloga, kod primene ovog metoda dobro je konstruisati niz vektora $\{\vec{p}^{(k)}\}$, koji su međusobno konjugovani u smislu

$$(\vec{p}^{(i)}, A\vec{p}^{(j)}) = 0 \quad (i \neq j).$$

Tada se navedeni metod može iskazati rekurzivno pomoću sledećih formula:

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} - c_k \vec{p}^{(k)} \quad (k = 0, 1, \dots, n-1),$$

$$\vec{r}^{(k)} = A\vec{x}^{(k)} - \vec{b} \quad (k = 0, 1, \dots, n),$$

$$\vec{p}^{(0)} = -\vec{r}^{(0)}, \quad \vec{p}^{(k)} = -\vec{r}^{(k)} + q_k \vec{p}^{(k-1)} \quad (k = 1, \dots, n-1),$$

$$c_k = \frac{(\vec{p}^{(k)}, \vec{r}^{(k)})}{(\vec{p}^{(k)}, A\vec{p}^{(k)})}, \quad q_k = \frac{(\vec{r}^{(k)}, A\vec{p}^{(k-1)})}{(\vec{p}^{(k-1)}, A\vec{p}^{(k-1)})} \quad (k = 0, 1, \dots, n-1).$$

Kao što je napred rečeno, za neko $m \leq n$ teorijski biće $\vec{r}^{(m)} = 0$, što znači da je odredjeno tačno rešenje sistema (3.7.1).

4.3.8. Iterativni metodi 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 [4], [47]).

Prepostavimo da je A regularna matrica reda n .

Teorema 3.8.1. Neka je r prirodan broj veći od jedinice i neka je

$$F_k = I - AX_k \quad (k = 0, 1, \dots),$$

gde je X_0 data matrica takva da je

$$(3.8.1) \quad \|F_0\| = \|I - AX_0\| \leq q < 1.$$

Tada niz matrica $\{X_k\}$ definisan sa

$$(3.8.2) \quad X_k = X_{k-1}(I + F_{k-1} + \dots + F_{k-1}^{r-1}) \quad (k = 1, 2, \dots),$$

konvergira ka A^{-1} .

Dokaz. Na osnovu

$$\begin{aligned} F_k &= I - AX_k \\ &= I - AX_{k-1}(I + F_{k-1} + \dots + F_{k-1}^{r-1}) \\ &= I - (I - F_{k-1})(I + F_{k-1} + \dots + F_{k-1}^{r-1}) \\ &= F_{k-1}^r, \end{aligned}$$

dobijamo

$$F_k = F_0^{r^k} \quad (k = 0, 1, \dots).$$

Iz poslednje jednakosti i uslova (3.8.1) redom sleduje

$$\|F_k\| \leq \|F_0\|^{r^k} \leq q^{r^k} \quad (k = 0, 1, \dots),$$

i

$$\lim_{k \rightarrow +\infty} \|F_k\| = 0,$$

odakle zaključujemo da je

$$\lim_{k \rightarrow +\infty} F_k = \lim_{k \rightarrow +\infty} (I - AX_k) = 0, \quad \text{tj.} \quad \lim_{k \rightarrow +\infty} X_k = A^{-1}.$$

Teorema 3.8.2. Ako su ispunjeni uslovi teoreme 3.8.1, za svako $k \in \mathbb{N}$ važe nejednakosti

$$(3.8.3) \quad \|X_k - A^{-1}\| \leq \frac{\|X_k F_k\|}{1 - \|F_k\|},$$

$$(3.8.4) \quad \|X_k - A^{-1}\| \leq \|F_{k-1}\|^{r-1} \frac{\|X_{k-1} F_{k-1}\|}{1 - \|F_{k-1}\|},$$

$$(3.8.5) \quad \|X_k - A^{-1}\| \leq \|F_0\|^{r^k} \frac{\|X_0\|}{1 - \|F_0\|}$$

za svako $k \in \mathbb{N}$.

Dokaz. Ako stavimo, $E_k = A^{-1} - X_k$, imamo

$$E_k - E_k F_k = (A^{-1} - X_k) A X_k = X_k (I - A X_k) = X_k F_k.$$

Kako je $\|F_k\| < 1$, važi nejednakost

$$\|E_k\|(1 - \|F_k\|) \leq \|E_k - E_k F_k\| = \|X_k F_k\|,$$

odakle neposredno sleduje (3.8.3).

Nejednakosti (3.8.4) i (3.8.5) dokazuju se na sličan način (videti [47]).

Teorema 3.8.3. Ako su ispunjeni uslovi teoreme 3.8.1, iterativni proces (3.8.2) ima red konvergencije r .

Dokaz. Ako u nejednakosti (3.8.3), tj. nejednakosti

$$\|x_{k+1} - A^{-1}\| \leq \frac{\|x_{k+1} F_{k+1}\|}{1 - \|F_{k+1}\|},$$

uveđemo smenu

$$F_{k+1} = F_k^r = (I - AX_k)^r = A^r (A^{-1} - X_k)^r$$

dobijamo

$$\|x_{k+1} - A^{-1}\| \leq \frac{\|x_{k+1} A^r (A^{-1} - X_k)^r\|}{1 - \|F_{k+1}\|} \leq \frac{\|x_{k+1}\| \cdot \|A\|^r}{1 - \|F_{k+1}\|} \|x_k - A^{-1}\|^r.$$

Kako je

$$\lim_{k \rightarrow +\infty} \frac{\|x_{k+1}\| \cdot \|A\|^r}{1 - \|F_{k+1}\|} = \|A^{-1}\| \cdot \|A\|^r < +\infty,$$

imamo

$$\|x_{k+1} - A^{-1}\| = O(\|x_k - A^{-1}\|^r) \quad (k \rightarrow +\infty),$$

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

$$x_k = x_{k-1} (2I - AX_{k-1}) \quad (k = 1, 2, \dots).$$

Ovaj iterativni proces sa kvadratnom konvergencijom potiče od G.Schulza ([52]), dok se ocene za grešku, analogne onim u teoremi 3.8.2, mogu naći u radovima [3], [5], [18], [33].

U slučaju kada je $r = 3$, odgovarajući iterativni proces sa kubnom konvergencijom je

$$x_k = x_{k-1} (3I - 3AX_{k-1} + (AX_{k-1})^2) \quad (k = 1, 2, \dots).$$

U radu [3], J.Albrecht je ukazao na vezu ovog procesa sa Newtonovim metodom za rešavanje nelinearnih jednačina (videti poglavljje 5.1).

4.4. PROBLEM SOPSTVENIH VREDNOSTI

U ovom poglavlju se izlažu metodi za rešavanje problema sopstvenih vrednosti. Podela je učinjena tako što su, najpre, dati metodi za određivanje karakterističnog polinoma matrice, zatim metodi za određivanje dominantnih i subdominantnih sopstvenih vrednosti i odgovarajućih sopstvenih vektora, i na kraju metodi za rešavanje kompletног problema sopstvenih vrednosti. Posebna pažnja je posvećena simetričnim trodijagonalnim matricama zbog važnosti koje one imaju u numeričkoj matematici.

4.4.1. Lokalizacija sopstvenih vrednosti

U odeljku 4.1.3 date su osnovne definicije i stavovi koji se odnose na sopstvene vrednosti i sopstvene vektore kvadratne matrice $A = [a_{ij}]_{n \times n}$. Ovo poglavlje biće posvećeno numeričkim metodima za njihovo određivanje, s obzirom na veliki značaj koji oni imaju u numeričkoj matematici. Naime, veliki broj problema se svodi na rešavanje problema sopstvenih vrednosti. U ovom odeljku daćemo neke rezultate koji se odnose na lokalizaciju sopstvenih vrednosti u kompleksnoj ravni.

Teorema 4.1.1(Gershgorin). Neka je $A = [a_{ij}]_{n \times n}$ kvadratna matrica reda n i neka su C_i ($i=1, \dots, n$) diskovi u kompleksnoj ravni sa centrima u a_{ii} i poluprečnicima $r_i = \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|$, tj.

$$C_i = \{z \mid |z - a_{ii}| \leq r_i\} \quad (i=1, \dots, n).$$

Ako sa C označimo uniju ovih diskova, tada se sve sopstvene vrednosti matrice A nalaze u C .

Dokaz. Neka je λ sopstvena vrednost matrice A , a \vec{x} odgovarajući sopstveni vektor normalizovan tako da je $\|\vec{x}\|_\infty = \max_i |x_i| = |x_m| = 1$. Tada je $\lambda \vec{x} = A \vec{x}$, tj.

$$(\lambda - a_{ii})x_i = \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij}x_j \quad (i=1, \dots, n),$$

odakle za $i=m$ imamo

$$|\lambda - a_{mm}| \leq \sum_{\substack{j=1 \\ j \neq m}}^n |a_{ij}| \cdot |x_j| \leq \sum_{\substack{j=1 \\ j \neq m}}^n |a_{ij}| = r_m.$$

Dakle, sopstvena vrednost λ leži u disku C_m . Kako je λ proizvoljna sopstvena vrednost zaključujemo da se sve sopstvene vrednosti nalaze u uniji diskova, tj. u C .

Primedba 4.1.1. S obzirom da matrica A^T ima iste sopstvene vrednosti kao i matrica A , na osnovu prethodne teoreme može se zaključiti da se sve sopstvene vrednosti matrice A nalaze i u uniji D diskova

$$D_j = \{ z \mid |z - a_{jj}| \leq s_j \} \quad (j=1, \dots, n),$$

$$\text{gde je } s_j = \sum_{\substack{i=1 \\ i \neq j}}^n |a_{ij}|.$$

Na osnovu prethodnog zaključujemo da sve sopstvene vrednosti matrice A leže u preseku skupova C i D .

Teorema 4.1.2. Ako m diskova iz teoreme 4.1.1 čini povezanu oblast, koja je izolovana od ostalih diskova, tada se u toj povezanoj oblasti nalazi tačno m sopstvenih vrednosti matrice A .

Dokaz ove teoreme može se naći, na primer, u izvanrednoj monografiji Wilkinsona ([63]).

Primer 4.1.1. Neka je

$$A = \begin{bmatrix} 1 & 0.1 & -0.1 \\ 0 & 2 & 0.4 \\ -0.2 & 0 & 3 \end{bmatrix}.$$

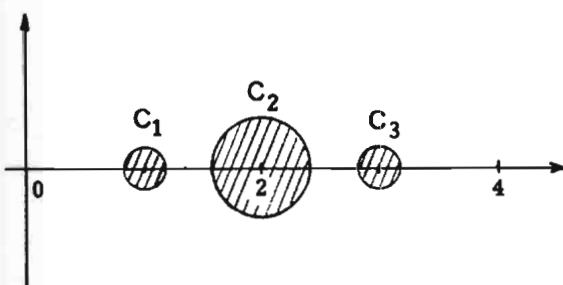
Na osnovu teoreme 4.1.1 sopstvene vrednosti se nalaze u diskovima (videti sl. 4.1.1)

$$C_1 = \{ z \mid |z-1| \leq 0.2 \}, \quad C_2 = \{ z \mid |z-2| \leq 0.4 \}, \quad C_3 = \{ z \mid |z-3| \leq 0.2 \}.$$

Primetimo da na osnovu primedbe 4.1.1 sleduje da diskovi D_1 , D_2 , D_3 imaju redom poluprečnike 0.2, 0.1, 0.5. Inače, tačne vrednosti sopstvenih vrednosti matrice A , na sedam decimala, su $\lambda_1 = 0.9861505$, $\lambda_2 = 2.0078436$, $\lambda_3 = 3.0060058$, dok je

normirani karakteristični polinom

$$H(\lambda) = \lambda^3 - 6\lambda^2 + 10.98\lambda - 5.952.$$



Sl. 4.1.1

Teorema o lokalizaciji sopstvenih vrednosti ima i teorijski i praktični značaj (na primer, za određivanje početnih vrednosti kod iterativnih metoda, za analizu kod perturbacionih problema, itd.).

Za određivanje sopstvenih vrednosti postoji veliki broj metoda, pri čemu neki od metoda omogućavaju nalaženje svih sopstvenih vrednosti, a drugi samo neke od sopstvenih vrednosti, na primer, dominantne, tj. one sa maksimalnim modulom. Neki od metoda omogućavaju samo nalaženje koeficijenata karakterističnog polinoma, tako da se za određivanje sopstvenih vrednosti mora primenjivati neki od metoda za rešavanje algebarskih jednačina (videti poglavlje 5.3). Ovakav pristup u određivanju sopstvenih vrednosti se ne preporučuje, jer je u većini slučajeva numerički nestabilan, tj. slabo uslovljen. Naime, kako su koeficijenti karakterističnog polinoma opterećeni, u opštem slučaju, greškama zaokrugljivanja, usled slabe uslovljenosti karakterističnog polinoma dolazi do velikih grešaka u sopstvenim vrednostima. O uticaju promene koeficijenata na promenu nula kod slabo uslovljenih polinoma videti odeljak 5.3.1.

4.4.2. Metodi za određivanje karakterističnog polinoma

U ovom odeljku ukratko ćemo navesti nekoliko metoda za određivanje karakterističnog polinoma matrice $A = [a_{ij}]_{n \times n}$,

$$(4.2.1) \quad P(\lambda) = \det(A - \lambda I).$$

Kao što je rečeno na kraju prethodnog odeljka nije preporučljivo koristiti ovako dobijeni polinom za određivanje sopstvenih vrednosti i sopstvenih vektora matrice A , sem u slučajevima kada je karakteristični polinom dobro uslovljen.

1. Metod Krilova. Umesto (4.2.1) posmatraćemo normalizovani karakteristični polinom

$$(4.2.2) \quad H(\lambda) = (-1)^n P(\lambda) = \lambda^n - p_1 \lambda^{n-1} + p_2 \lambda^{n-2} - \dots + (-1)^n p_n.$$

Na osnovu Cayley-Hamiltonove teoreme (teorema 1.3.1) imamo

$$H(A) = A^n - p_1 A^{n-1} + p_2 A^{n-2} - \dots + (-1)^n p_n I = 0.$$

Neka je sada $\vec{y}^{(0)}$ proizvoljan n -dimenzionalni vektor, sa kojim pomnožimo prethodnu jednakost s desne strane. Tada dobijamo

$$p_1 A^{n-1} \vec{y}^{(0)} - p_2 A^{n-2} \vec{y}^{(0)} + \dots + (-1)^{n-1} p_n \vec{y}^{(0)} = A^n \vec{y}^{(0)},$$

odakle, korišćenjem koordinatne reprezentacije, dolazimo da sistema linearnih jednačina

$$(4.2.3) \quad \begin{bmatrix} y_1^{(n-1)} & -y_1^{(n-2)} & \dots & (-1)^{n-1} y_1^{(0)} \\ y_2^{(n-1)} & -y_2^{(n-2)} & & (-1)^{n-1} y_2^{(0)} \\ \vdots & & & \\ y_n^{(n-1)} & -y_n^{(n-2)} & & (-1)^{n-1} y_n^{(0)} \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} = \begin{bmatrix} y_1^{(n)} \\ y_2^{(n)} \\ \vdots \\ y_n^{(n)} \end{bmatrix},$$

gde smo stavili $\vec{y}^{(k)} = A^k \vec{y}^{(0)} = [y_1^{(k)} \quad y_2^{(k)} \quad \dots \quad y_n^{(k)}]^T$ ($k = 1, 2, \dots, n$). Primetimo da stepene matrice A^k ne treba izračunavati, već treba koristiti rekurzivni postupak

$$(4.2.4) \quad \vec{y}^{(k)} = A \vec{y}^{(k-1)} \quad (k = 1, \dots, n).$$

Pod uslovom da je matrica dobijenog sistema linearnih jednačina (4.2.3) regularna, rešavanjem ovog sistema dobijamo koeficijente p_1, p_2, \dots, p_n . Ako je, međutim, matrica ovog sistema singularna treba promeniti početni vektor $\vec{y}^{(0)}$.

Ilustrovaćemo primenu ovog metoda na nalaženje karakterističnog polinoma jedne matrice četvrtog reda.

Primer 4.2.1. Neka je

$$A = \begin{bmatrix} 3 & 2 & -2 & -1 \\ -1 & 3 & -1 & 0 \\ 1 & -2 & 4 & 1 \\ 3 & 0 & 1 & 3 \end{bmatrix}.$$

Ako uzmemo $\vec{y}^{(0)} = [1 \ 0 \ 0 \ 0]^T$, pomoću (4.2.4) nalazimo redom

$$\vec{y}^{(1)} = A\vec{y}^{(0)} = \begin{bmatrix} 3 \\ -1 \\ 1 \\ 3 \end{bmatrix}, \quad \vec{y}^{(2)} = A\vec{y}^{(1)} = \begin{bmatrix} 2 \\ -7 \\ 12 \\ 19 \end{bmatrix},$$

$$\vec{y}^{(3)} = A\vec{y}^{(2)} = \begin{bmatrix} -51 \\ -35 \\ 83 \\ 75 \end{bmatrix}, \quad \vec{y}^{(4)} = A\vec{y}^{(3)} = \begin{bmatrix} -464 \\ -137 \\ 426 \\ 155 \end{bmatrix}.$$

Sistem (4.2.3) postaje

$$\begin{bmatrix} -51 & -2 & 3 & -1 \\ -35 & 7 & -1 & 0 \\ 83 & -12 & 1 & 0 \\ 75 & -19 & 3 & 0 \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix} = \begin{bmatrix} -464 \\ -137 \\ 426 \\ 155 \end{bmatrix},$$

odakle dobijamo

$$p_1 = 13, \ p_2 = 67, \ p_3 = 151, \ p_4 = 120.$$

Prema tome imamo

$$H(\lambda) = P(\lambda) = \lambda^4 - 13\lambda^3 + 67\lambda^2 - 151\lambda + 120.$$

2. Leverrierov metod i modifikacija Faddeeva. Leverrierov metod se bazira na poznatim Newtonovim formulama za sume stepena svih nula polinoma. Neka su $\lambda_1, \dots, \lambda_n$ sopstvene vrednosti matrice A , tj. nule polinoma (4.2.2), pri čemu se svaka nula uzima onoliko puta kolika je njena višestrukost. Tada za sume

$$s_m = \lambda_1^m + \lambda_2^m + \dots + \lambda_n^m \quad (m=0, 1, \dots, n)$$

važe Newtonove formule (videti, na primer, [42,str. 241-242])

$$s_m - p_1 s_{m-1} + p_2 s_{m-2} - \dots + (-1)^{m-1} p_{m-1} s_1 + (-1)^m p_m = 0 \\ (m=1, \dots, n),$$

odakle imamo

$$p_1 = s_1,$$

$$p_2 = -\frac{1}{2}(s_2 - p_1 s_1),$$

$$p_3 = \frac{1}{3}(s_3 - p_1 s_2 + p_2 s_1),$$

⋮

$$p_n = \frac{(-1)^{n-1}}{n} (s_n - p_1 s_{n-1} + \dots + (-1)^{n-1} p_{n-1} s_1).$$

Prema tome, ako su sume s_m poznate možemo naći koeficijente karakterističnog polinoma. Primetimo da je

$$s_1 = \lambda_1 + \lambda_2 + \dots + \lambda_n = \operatorname{tr} A = \sum_{i=1}^n a_{ii}.$$

Kako su $\lambda_1^m, \lambda_2^m, \dots, \lambda_n^m$ sopstvene vrednosti matrice A^m (videti teoremu 1.3.2) zaključujemo da je

$$s_m = \operatorname{tr} A^m.$$

Dakle, ako je $A^m = [a_{ij}^{(m)}]_{n \times n}$ imamo $s_m = \sum_{i=1}^n a_{ii}^{(m)}$, pri čemu stepene matrice određujemo redom pomoću

$$A^m = A \cdot A^{m-1}.$$

Primer 4.2.2. Za matricu iz prethodnog primera imamo redom

$$A^2 = A \cdot A = \begin{bmatrix} 2 & 16 & -17 & -8 \\ -7 & 9 & -5 & 0 \\ 12 & -12 & 17 & 6 \\ 19 & 4 & 1 & 7 \end{bmatrix}, \quad A^3 = A \cdot A^2 = \begin{bmatrix} -51 & 86 & -98 & -43 \\ -35 & 23 & -15 & 2 \\ 83 & -46 & 62 & 23 \\ 75 & 48 & -31 & 3 \end{bmatrix},$$

$$A^4 = A \cdot A^3 = \begin{bmatrix} -464 & 348 & -411 & -174 \\ -137 & 29 & -11 & 26 \\ 426 & -96 & 151 & 48 \\ 155 & 356 & -319 & -97 \end{bmatrix}.$$

Kako je

$$\begin{aligned}s_1 &= \text{tr } A = 3 + 3 + 4 + 3 = 13, \\ s_2 &= \text{tr } A^2 = 2 + 9 + 17 + 7 = 35, \\ s_3 &= \text{tr } A^3 = -51 + 23 + 62 + 3 = 37, \\ s_4 &= \text{tr } A^4 = -464 + 29 + 151 - 97 = -381,\end{aligned}$$

imamo redom

$$\begin{aligned}p_1 &= 13, \\ p_2 &= -(35 - 13 \cdot 13)/2 = 67, \\ p_3 &= (37 - 13 \cdot 35 + 67 \cdot 13)/3 = 151, \\ p_4 &= -(-381 - 13 \cdot 37 + 67 \cdot 35 - 151 \cdot 13)/4 = 120.\end{aligned}$$

Jednu modifikaciju Leverrierovog metoda dao je Faddeev. Ta modifikacija zahteva manji broj numeričkih operacija i sastoji se u sledećem:

Uместо stepena A^m izračunava se niz matrica A_m ($m=1, \dots, n$) pomoću formula

$$A_m = AB_{m-1}, \quad q_m = \frac{1}{m} \text{tr } A_m, \quad B_m = A_m - q_m I,$$

pri čemu se za B_0 uzima jedinična matrica.

Matematičkom indukcijom se može dokazati da je $q_m = (-1)^{m-1} p_m$ ($m = 1, \dots, n$).

Primer 4.2.3. Primenom metoda Faddeeva za matricu iz prethodnih primera dobijamo redom

$$A_1 = A, \quad q_1 = \text{tr } A_1 = 13, \quad B_1 = A_1 - 13I = \begin{bmatrix} -10 & 2 & -2 & -1 \\ -1 & -10 & -1 & 0 \\ 1 & -2 & -9 & 1 \\ 3 & 0 & 1 & -10 \end{bmatrix},$$

$$A_2 = AB_1 = \begin{bmatrix} -37 & -10 & 9 & 5 \\ 6 & -30 & 8 & 0 \\ -1 & 14 & -35 & -7 \\ -20 & 4 & -12 & 32 \end{bmatrix}, \quad q_2 = \frac{1}{2} \text{tr } A_2 = -67,$$

$$B_2 = A_2 + 67I = \begin{bmatrix} 30 & -10 & 9 & 5 \\ 6 & 37 & 8 & 0 \\ -1 & 14 & 32 & -7 \\ -20 & 4 & -12 & 35 \end{bmatrix}, \quad A_3 = AB_2 = \begin{bmatrix} 124 & 12 & -9 & -6 \\ -11 & 107 & -17 & 2 \\ -6 & -24 & 109 & 12 \\ 29 & -4 & 23 & 113 \end{bmatrix},$$

$$q_3 = \frac{1}{3} \operatorname{tr} A_3 = 151, \quad B_3 = A_3 - 151I = \begin{bmatrix} -27 & 12 & -9 & -6 \\ -11 & -44 & -17 & 2 \\ -6 & -24 & -42 & 12 \\ 29 & -4 & 23 & -38 \end{bmatrix},$$

$$A_4 = AB_3 = \begin{bmatrix} -120 & 0 & 0 & 0 \\ 0 & -120 & 0 & 0 \\ 0 & 0 & -120 & 0 \\ 0 & 0 & 0 & -120 \end{bmatrix}, \quad q_4 = \frac{1}{4} \operatorname{tr} A_4 = -120.$$

Dakle, $p_1 = q_1 = 13$, $p_2 = -q_2 = 67$, $p_3 = q_3 = 151$, $p_4 = -q_4 = -120$.

3. Metod Danilevskog. Ovaj metod se zasniva na transformaciji matrice A na tzv. Frobeniusov oblik

$$(4.2.5) \quad F = \begin{bmatrix} f_1 & f_2 & \dots & f_{n-1} & f_n \\ 1 & 0 & & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & & 1 & 0 \end{bmatrix},$$

pri čemu su matrice A i F slične, tj. postoji regularna matrica C takva da je $F = C^{-1}AC$ (videti definiciju 1.3.5). S obzirom da slične matrice imaju identične karakteristične polinome, jednostavno se, na osnovu (4.2.5), dobija karakteristični polinom matrice A. Naime, ako $\det(F - \lambda I)$ razvijemo po elementima prve kolone dobijamo

$$P(\lambda) = (f_1 - \lambda)(-\lambda)^{n-1} - f_2(-\lambda)^{n-2} + \dots + (-1)^{n-1}f_n,$$

tj.

$$P(\lambda) = (-1)^n(\lambda^n - f_1\lambda^{n-1} - f_2\lambda^{n-2} - \dots - f_n).$$

Dakle, $p_m = (-1)^{m-1}f_m$ ($m = 1, \dots, n$).

Za vektor-vrstu $\vec{z} = [z_1 \ z_2 \ \dots \ z_n]$ definišimo kvadratnu matricu n-tog reda

$$G_r(\vec{z}) = \begin{bmatrix} I_{n-r} & | & O_{n-r,r} \\ \hline P_{r,n-r}(\vec{z}) & | & R_r(\vec{z}) \end{bmatrix},$$

gde je I_{n-r} jedinična matrica reda $n-r$, $O_{n-r,r}$ nula matrica tipa $r \times (n-r)$, $P_{r,n-r}(\vec{z})$ pravougaona matrica tipa $r \times (n-r)$, koja u prvoj vrsti ima redom elemente z_1, z_2, \dots, z_{n-r} , dok su svi ostali njeni elementi jednaki nuli, i nazad, $R_r(\vec{z})$ je kvadratna matrica reda r koja se razlikuje od jedinične matrice samo u prvoj vrsti, gde sadrži preostale elemente vektora \vec{z} , tj. ima redom elemente z_{n-r+1}, \dots, z_n .

Pretpostavimo da je $z_{n-r+1} \neq 0$. Na osnovu teoreme 1.1.3 lako se nalazi inverzna matrica za $G_r(\vec{z})$

$$G_r^{-1}(\vec{z}) = \begin{bmatrix} I_{n-r} & | & O_{n-r,r} \\ \hline P_{r,n-r}(\vec{z}) & | & -\frac{R_r^{-1}(\vec{z})}{z_{n-r+1}} \end{bmatrix},$$

gde je

$$R_r^{-1}(\vec{z}) = \begin{bmatrix} \frac{1}{z_{n-r+1}} & -\frac{z_{n-r+2}}{z_{n-r+1}} & \dots & -\frac{z_n}{z_{n-r+1}} \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}.$$

Neka je $A^{(k)}$ matrica n-tog reda čije su vrste vektori $\vec{a}_1^{(k)}, \dots, \vec{a}_n^{(k)}$, tj.

$$A^{(k)} = \begin{bmatrix} \vec{a}_1^{(k)} \\ \vdots \\ \vec{a}_n^{(k)} \end{bmatrix}.$$

Stavimo sada da je $A^{(1)} = A$, gde je A data matrica čiju Frobeniusovu formu tražimo.

Direktnim množenjem može se pokazati da se pomoću niza transformacija

$$A^{(2)} = G_2(\vec{a}_n^{(1)}) A^{(1)} G_2^{-1}(\vec{a}_n^{(1)}) \quad (a_{n,n-1}^{(1)} \neq 0),$$

$$A^{(3)} = G_3(\vec{a}_{n-1}^{(2)}) A^{(2)} G_3^{-1}(\vec{a}_{n-1}^{(2)}) \quad (a_{n-1,n-2}^{(2)} \neq 0),$$

$$\vdots$$

$$A^{(n)} = G_n(\vec{a}_2^{(n-1)}) A^{(n-1)} G_n^{-1}(\vec{a}_2^{(n-1)}) \quad (\vec{a}_{2,1}^{(n-1)} \neq 0),$$

dolazi do Frobeniusovog oblika $F = A^{(n)}$. Pri ovome je

$$C = G_2^{-1}(\vec{a}_n^{(1)}) G_3^{-1}(\vec{a}_{n-1}^{(2)}) \dots G_n^{-1}(\vec{a}_2^{(n-1)})$$

i

$$C^{-1} = G_n(\vec{a}_2^{(n-1)}) G_{n-1}(\vec{a}_3^{(n-2)}) \dots G_2(\vec{a}_n^{(1)}).$$

Prethodne transformacije egzistiraju pod navedenim uslovima. Ovi uslovi mogu biti obezbedjeni dodatnom permutacijom vrsta i kolona. U slučaju kada je vektor-vrsta $\vec{a}_{n-k+1}^{(k)}$ nula vektor problem se jednostavno redukuje na problem niže dimenzije.

Primer 4.2.4. Za matricu iz prethodnog primera imamo redom

$$G_2 = G_2(3, 0, 1, 3) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 3 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad G_2^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & -3 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$A^{(2)} = G_2 A G_2^{-1} = \begin{bmatrix} 9 & 2 & -2 & 5 \\ 2 & 3 & -1 & 3 \\ 16 & 4 & 1 & 4 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$G_3 = G_3(16, 4, 1, 4) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 16 & 4 & 1 & 4 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad G_3^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -4 & 1/4 & -1/4 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$A^{(3)} = G_3 A^{(2)} G_3^{-1} = \begin{bmatrix} 1 & 1/2 & -5/2 & 3 \\ -24 & 12 & -43 & 48 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$G_4 = G_4(-24, 12, -43, 48) = \begin{bmatrix} -24 & 12 & -43 & 48 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, G_4^{-1} = \begin{bmatrix} -\frac{1}{24} & \frac{1}{2} & -\frac{43}{24} & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$F = A^{(4)} = G_4 A^{(3)} G_4^{-1} = \begin{bmatrix} 13 & -67 & 151 & -120 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

4.4.3. Metodi za dominantne sopstvene vrednosti

Vrlo često se u nekim primenama zahteva nalaženje samo maksimalne po modulu sopstvene vrednost i njoj odgovarajućeg sopstvenog vektora.

Neka su $\lambda_1, \dots, \lambda_n$ sopstvene vrednosti i $\vec{x}_1, \dots, \vec{x}_n$ odgovarajući sopstveni vektori matrice $A = [a_{ij}]_{n \times n}$. Ako je

$$|\lambda_1| = \dots = |\lambda_r| > |\lambda_{r+1}| \geq \dots \geq |\lambda_n|$$

kažemo da su $\lambda_1, \dots, \lambda_r$ dominantne sopstvene vrednosti matrice A . U ovom odeljku razmatraćemo jedan metod za određivanje dominantne sopstvene vrednosti i odgovarajućeg sopstvenog vektora, kao i neke modifikacije ovog metoda. Prepostavimo, pritom, da su sopstveni vektori $\vec{x}_1, \dots, \vec{x}_n$ linearno nezavisni, pa kao takvi oni čine jednu bazu u R^n . Dakle, proizvoljan ne-nula vektor \vec{v}_0 može se izraziti pomoću

$$(4.3.1) \quad \vec{v}_0 = \sum_{i=1}^n \alpha_i \vec{x}_i,$$

gde su α_i neki skaliari. Definišimo sada iterativni proces

$$\vec{v}_k = A\vec{v}_{k-1} \quad (k = 1, 2, \dots).$$

Tada je

$$\vec{v}_k = A\vec{v}_{k-1} = A^2\vec{v}_{k-2} = \dots = A^k\vec{v}_0 = \sum_{i=1}^n \alpha_i A^k \vec{x}_i,$$

ili, s obzirom na (4.3.1) i tvrdjenje teoreme 1.3.2,

$$(4.3.2) \quad \vec{v}_k = \sum_{i=1}^n \alpha_i \lambda_i^k \vec{x}_i.$$

Posebno je ovde interesantan slučaj kada imamo jednu dominantnu sopstvenu vrednost λ_1 ($r = 1$). Pod pretpostavkom da je $\alpha_1 \neq 0$, na osnovu (4.3.2) imamo

$$\vec{v}_k = \alpha_1 \lambda_1^k (\vec{x}_1 + \sum_{i=2}^n \frac{\alpha_i}{\alpha_1} \left(\frac{\lambda_i}{\lambda_1} \right)^k \vec{x}_i) = \alpha_1 \lambda_1^k (\vec{x}_1 + \vec{\epsilon}_k),$$

gde vektor $\vec{\epsilon}_k \rightarrow 0$, kada $k \rightarrow +\infty$.

Uvedimo sada oznaku $(\vec{y})_i$ za i-tu koordinatu nekog vektora \vec{y} . Tada je i-ta koordinata vektora \vec{v}_k

$$(\vec{v}_k)_i = \alpha_1 \lambda_1^k ((\vec{x}_1)_i + (\vec{\epsilon}_k)_i).$$

Kako je

$$\vec{v}_{k+1} = \alpha_1 \lambda^{k+1} (\vec{x}_1 + \vec{\epsilon}_{k+1}),$$

na osnovu prethodnog, za svako i ($1 \leq i \leq n$) imamo

$$\frac{(\vec{v}_{k+1})_i}{(\vec{v}_k)_i} = \lambda_1 \frac{(\vec{x}_1)_i + (\vec{\epsilon}_{k+1})_i}{(\vec{x}_1)_i + (\vec{\epsilon}_k)_i} \rightarrow \lambda_1 \quad (k \rightarrow +\infty).$$

Na osnovu ove činjenice može se formulisati metod za određivanje dominantne sopstvene vrednosti λ_1 , koji je poznat kao metod stepenovanja*. Vektor \vec{v}_k je pri ovome jedna aproksimacija za nenormirani sopstveni vektor† koji odgovara dominantnoj sopstvenoj vrednosti. Pri praktičnoj realizaciji ovog metoda ide se na normiranje sopstvenog vektora, tj. na normiranje vektora \vec{v}_k posle svakog iterativnog koraka. Normiranje se sprovodi deljenjem vektora \vec{v}_k sa svojom koordinatom maksimalnog modula. Tako se metod stepenovanja može izraziti pomoću formula

$$\vec{z}_k = A\vec{v}_{k-1}, \quad \vec{v}_k = \vec{z}_k / \gamma_k,$$

gde je γ_k koordinata vektora \vec{z}_k sa najvećim modulom, tj.

* Na engleskom jeziku power method.

† Ako je \vec{x} sopstveni vektor, tada je i $c\vec{x}$ ($c \neq 0$), takodje, sopstveni vektor koji odgovara istoj sopstvenoj vrednosti.

$\gamma_k = (\vec{z}_k)_i$ i $|(\vec{z}_k)_i| = \|\vec{z}_k\|$. Primetimo da $\gamma_k \rightarrow \lambda_1$ i $\vec{v}_k \rightarrow \frac{\vec{x}_1}{\|\vec{x}_1\|_\infty}$ kada $k \rightarrow +\infty$.

Brzina konvergencije ovog metoda zavisi od količnika $|\lambda_2/\lambda_1|$. Naime, važi

$$(4.3.3) \quad |\lambda_1 - \gamma_k| = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right).$$

Primetimo da smo u izvodjenju ovog metoda prepostavili da je $\alpha_1 \neq 0$, što znači da metod konvergira ako je λ_1 dominantna sopstvena vrednost i ako početni vektor \vec{v}_0 ima komponentu u pravcu koji odgovara sopstvenom vektoru \vec{x}_1 . O ponašanju metoda bez ovih prepostavki među se naći u monografiji Wilkinsa [63, str. 570] i radu Parleta i Poolea [44]. Praktično, zbog pojave grešaka zaokrugljivanja u iterativnom procesu, uslov $\alpha_1 \neq 0$ biće zadovoljen posle nekoliko koraka, iako polazna prepostavka za vektor \vec{v}_0 nije ispunjena. Pri primeni ovog metoda često se unapred poznaje neka ocena za sopstveni vektor \vec{x}_1 , pa se onda ona uzima kao početni vektor \vec{v}_0 . Primenom perturbacione teorije mogu se dati izvesne ocene za (4.3.3) (videti monografiju Goluba i Van Loana [24, str. 210-211]).

Prethodno razmatranje se odnosi na slučaj jedne dominantne sopstvene vrednosti. Moguće je razmatrati i slučajeve kada ima više dominantnih sopstvenih vrednosti. Ova razmatranja su slična sa odgovarajućim pristupom kod Bernoullievog metoda za određivanje dominantnih korena algebarskih jednačina (videti odeljak 5.3.3).

Primer 4.3.1. Neka je

$$A = \begin{bmatrix} -261 & 209 & -49 \\ -530 & 422 & -98 \\ -800 & 631 & -144 \end{bmatrix}$$

čije su sopstvene vrednosti $\lambda_1 = 10$, $\lambda_2 = 4$, $\lambda_3 = 3$ ([24, str. 210]).

Uzimajući za početni vektor $\vec{v}_0 = [0 \ 0 \ -1]^T$, primenom metoda stepenovanja dobijamo rezultate koji su prikazani u tabeli 4.3.1.

Tabela 4.3.1

k	γ_k	$(\vec{v}_k)_1$	$(\vec{v}_k)_2$	$(\vec{v}_k)_3$
1	144.	0.340278	0.680556	1.
2	13.2083	0.334911	0.669821	1.
3	10.7287	0.333774	0.667549	1.
4	10.2038	0.333463	0.666926	1.
5	10.0599	0.333372	0.666744	1.
6	10.0179	0.333345	0.666690	1.
7	10.0054	0.333337	0.666674	1.
8	10.0016	0.333334	0.666669	1.
9	10.0005	0.333334	0.666667	1.
10	10.0001	0.333333	0.666667	1.
11	10.0000	0.333333	0.666667	1.

S obzirom da je konvergencija metoda stepenovanja linearna za ubrzavanje konvergencije može da se koristi Aitkenov Δ^2 -metod.

Jedan prost način za ubrzavanje konvergencije metoda stepenovanja, posebno kod matrica sa realnim sopstvenim vrednostima

$$(4.3.4) \quad \lambda_1 > \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_n,$$

je primena istog metoda na određivanje dominantne sopstvene vrednosti matrice $A-pI$, gde je parametar p na pogodan način izabran*. Naime, sopstvene vrednosti ovakve matrice su $\lambda_i - p$, dok sopstveni vektori ostaju nepromenjeni. Pod uslovom (4.3.4) dominantna sopstvena vrednost će biti $\lambda_1 - p$ ili $\lambda_n - p$. Neka je to prva od njih. Kako brzina konvergencije metoda stepenovanja zavisi od količnika $|\lambda_2/\lambda_1|$ to će brzina procesa primjenjenog na $A-pI$ zavisiti od $|(\lambda_2-p)/(\lambda_1-p)|$ ili $|(\lambda_n-p)/(\lambda_1-p)|$ u zavisnosti od toga koji je od ovih količnika veći. Vrednosti parametra p za koje su ovi količnici manji od $|\lambda_2/\lambda_1|$ daje bržu konvergenciju metoda, nego u standardnom slučaju kada je $p=0$. Optimalna vrednost za p se jednostavno dobija, ako se poznaju λ_2 i λ_n , kao

$$p = \frac{1}{2}(\lambda_n + \lambda_2).$$

Naravno, mi najčešće ove vrednosti ne poznajemo, ali je moguće, na primer, da znamo izvesne granice za ove sopstvene vrednosti,

* Ovakva modifikacija se u anglo-saksonskoj literatiri sreće kao shift of origin.

na osnovu kojih možemo približno odrediti parametar p .

Primer 4.3.2. U primeru 4.3.1 brzina konvergencije je, s obzirom na (4.3.3), dirigovana sa $(2/5)^k$. Ako za p uzmemo vrednost $(\lambda_1 + \lambda_2)/2 = 3.5$, brzina konvergencije će biti dirigovana sa $(1/13)^k$, što predstavlja značajno ubrzanje. Startujući od istog početnog vektora \vec{v}_0 , sada se dobijaju rezultati koji su dati u tabeli 4.3.2.

Tabela 4.3.2

k	γ_k	$(\vec{v}_k)_1$	$(\vec{v}_k)_2$	$(\vec{v}_k)_3$
1	147.5000	0.332203	0.664407	1.
2	5.9780	0.333428	0.666856	1.
3	6.5437	0.333326	0.666652	1.
4	6.4967	0.333334	0.666668	1.
5	6.5003	0.333333	0.666667	1.
6	6.5000	0.333333	0.666667	1.

Dakle, ista tačnost kao u primeru 4.3.1 je postignuta sa 6 iteracija. Primetimo da $\gamma_k \rightarrow \lambda_1 - 3.5 = 6.5$.

Na kraju razmotrimo slučaj kada je realna matrica $A = [a_{ij}]_{n \times n}$ simetrična. Tada su njene sopstvene vrednosti realni brojevi za koje se može pretpostaviti distribucija data sa (4.3.4).

Definicija 4.3.1. Ako je \vec{x} ne nula vektor, veličina

$$r(\vec{x}) = \frac{(A\vec{x}, \vec{x})}{(\vec{x}, \vec{x})} = \frac{\vec{x}^T A \vec{x}}{\vec{x}^T \vec{x}}$$

se naziva Rayleighov količnik.

Na osnovu teoreme 1.4.6 za Rayleighov količnik važi

$$\lambda_n \leq r(\vec{x}) \leq \lambda_1.$$

Nalaženje najveće sopstvene vrednosti λ_1 može se interpretirati kao optimizacioni problem

$$(4.3.5) \quad \lambda_1 = \max_{\vec{x} \neq \vec{0}} r(\vec{x}),$$

s obzirom da se maksimum u (4.3.5) postiže kada je x sopstveni vektor koji odgovara sopstvenoj vrednosti λ_1 . Rešavanje optimizacionog problema (4.3.5) se može sprovesti tako da se za \vec{x}

uzima vektor \vec{v}_k koji se generiše pomoću metoda stepenovanja.

Neka je $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$ sistem ortonormiranih sopstvenih vektora, tj. $(\vec{x}_i, \vec{x}_j) = \vec{x}_j^T \vec{x}_i = \delta_{ij}$, gde je δ_{ij} Kroneckerova delta.

Kako je na osnovu prethodnoq

$$\vec{v}_k = \sum_{i=1}^n \alpha_i \lambda_i^k \vec{x}_i \quad i \quad A\vec{v}_k = \sum_{i=1}^n \alpha_i \lambda_i^{k+1} \vec{x}_i,$$

imamo

$$(\vec{v}_k, \vec{v}_k) = \sum_{i=1}^n \alpha_i^2 \lambda_i^{2k} \quad i \quad (A\vec{v}_k, \vec{v}_k) = \sum_{i=1}^n \alpha_i^2 \lambda_i^{2k+1},$$

pa je

$$r(\vec{v}_k) = \frac{\sum_{i=1}^n \alpha_i^2 \lambda_i^{2k+1}}{\sum_{i=1}^n \alpha_i^2 \lambda_i^{2k}} = \lambda_1 + O\left(\left(\frac{\lambda_2}{\lambda_1}\right)^{2k}\right).$$

Prema tome, $r(\vec{v}_k) \rightarrow \lambda_1$ brže, nego γ_k kod osnovnog metoda stepenovanja.

4.4.4. Metodi za subdominantne sopstvene vrednosti

Pretpostavimo da su sopstvene vrednosti matrice A uređene, tj. da važi

$$|\lambda_1| > |\lambda_2| > \dots > |\lambda_n|.$$

U ovom odeljku biće razmatrani metodi za određivanje subdominantnih sopstvenih vrednosti, tj. $\lambda_2, \dots, \lambda_m$ ($m < n$). Izložićemo tri metoda.

1. Metod ortogonalizacije. Pretpostavimo, najpre, da je matrica A simetrična, i neka je, na primer, metodom stepenovanja određen sopstveni vektor \vec{x}_1 koji odgovara dominantnoj sopstvenoj vrednosti λ_1 ($|\lambda_1| > |\lambda_i|$, $i=2, \dots, n$). Polažeći od proizvoljnog vektora \vec{z} formirajmo vektor \vec{v}_0 koji je ortogonalan na \vec{x}_1 . Tako imamo (videti Gram-Schmidtov

postupak ortogonalizacije, odeljak 2.1.4)

$$(4.4.1) \quad \vec{v}_0 = \vec{z} - \frac{(\vec{z}, \vec{x}_1)}{(\vec{x}_1, \vec{x}_1)} \vec{x}_1.$$

Kako je $(\vec{v}_0, \vec{x}_1) = 0$, teorijski gledano niz $\vec{v}_k = A\vec{v}_{k-1}$ ($k=1, 2, \dots$) u metodi stepenovanja, mogao bi da se iskoristi za određivanje λ_2 i odgovarajućeg sopstvenog vektora \vec{x}_2 . Međutim, bez obzira što \vec{v}_0 nema komponentu u pravcu sopstvenog vektora \vec{x}_1 , metod stepenovanja bi zbođ prisustva grešaka zaokrugljivanja počeo, posle izvesnog broja koraka, da konvergira ka sopstvenom vektoru \vec{x}_1 . O ovoj činjenici je bilo reči i u prethodnom odeljku.

Ovaj uticaj grešaka zaokrugljivanja moguće je odstraniti tzv. periodičnim "čišćenjem" vektora \vec{v}_0 od komponente u pravcu \vec{x}_1 . Ovo znači da posle, recimo r koraka, ponovo izračunamo \vec{v}_0 korišćenjem \vec{v}_r umesto \vec{z} u (4.4.1), tj. pomoću

$$\vec{v}_0 = \vec{v}_r - \frac{(\vec{v}_r, \vec{x}_1)}{(\vec{x}_1, \vec{x}_1)} \vec{x}_1.$$

Na ovaj način, ako je perioda "čišćenja" dovoljno mala da ne dodje do značajne akumulacije grešaka zaokrugljivanja, metodom stepenovanja možemo odrediti sopstvenu vrednost λ_2 i sopstveni vektor \vec{x}_2 .

Nastavljujući ovakav pristup možemo dalje odrediti λ_3 i \vec{x}_3 .

U opštem slučaju, ako smo odredili $\lambda_1, \dots, \lambda_v$ i odgovarajuće vektore $\vec{x}_1, \dots, \vec{x}_v$ ($v < m$) moguće je odrediti λ_{v+1} i \vec{x}_{v+1} metodom stepenovanja, formirajući vektor \vec{v}_0 ortogonalan na $\vec{x}_1, \dots, \vec{x}_v$. Dakle, polazeći od proizvoljnog vektora \vec{z} imamo

$$(4.4.2) \quad \vec{v}_0 = \vec{z} - \sum_{i=1}^v \frac{(\vec{z}, \vec{x}_i)}{(\vec{x}_i, \vec{x}_i)} \vec{x}_i,$$

što znači da vektor \vec{v}_0 ima komponente samo u pravcu preostalih sopstvenih vektora, tj. da je

$$\vec{v}_0 = \alpha_{v+1} \vec{x}_{v+1} + \dots + \alpha_n \vec{x}_n.$$

Metod stepenovanja primenjen na \vec{v}_0 daje \vec{x}_{v+1} i λ_{v+1} , ukoliko greške zaokrugljivanja nisu prisutne. S obzirom da ovo nije

tačno, potrebno je često "čišćenje" vektora \vec{v}_k od komponenata u pravcu $\vec{x}_1, \dots, \vec{x}$. Drugim rečima, posle r koraka treba ponovo određivati \vec{v}_0 pomoću (4.4.2), uz korišćenje \vec{v}_r umesto \vec{z} .

I u slučaju kada matrica A nije simetrična, ali ima potpun sistem sopstvenih vektora, prethodni orogonalizacioni postupak se može primeniti.

2. Metod inverzne iteracije. Ovaj metod se primenjuje na opštu matricu A i bazira se na rešavanju sistema jednačina

$$(4.4.3) \quad (A - pI)\vec{v}_k = \vec{v}_{k-1},$$

gde je p konstanta, a \vec{v}_0 proizvoljan vektor. Sistem jednačina (4.4.3) se obično rešava Gaussovim metodom eliminacije ili metodom Haleckog uz LR faktorizaciju matrice $B = A - pI$. Primetimo da je metod inverzne iteracije ekvivalentan metodu stepenovanja primenjenoq na B. Prema tome, primenom metoda inverzne iteracije dobija se dominantna sopstvena vrednost matrice B, tj. $\mu_v = 1/(\lambda_v - p)$, za koju važi

$$\min_j |\lambda_j - p| = |\lambda_v - p|.$$

Sopstvena vrednost λ_v je najbliža sopstvena vrednost matrice A broju p. Sopstveni vektor koji se pritom dobija isti je za matricu B i matricu A.

Pogodnim izborom parametra p mogu se, u principu, odrediti sve sopstvene vrednosti matrice A.

Kao i kod metoda stepenovanja, i ovde je pogodno izvršiti normiranje vektora \vec{v}_k tako da imamo

$$(4.4.4) \quad B\vec{z}_k = \vec{v}_{k-1}, \quad \vec{v}_k = \vec{z}_k / \gamma_k,$$

gde je γ_k koordinata vektora \vec{z}_k sa najvećim modulom.

Primer 4.4.1. Metodom inverzne iteracije za matricu

$$A = \begin{bmatrix} 4 & 1 & 4 \\ 1 & 10 & 1 \\ 4 & 1 & 10 \end{bmatrix}$$

odredićemo sopstvenu vrednost koja je najbliža broju p=9, kao i odgovarajući sopstveni vektor.

Korišćenjem faktorizacije pomoću Gaussovog metoda sa izborom glavnog elementa za matricu $B = A - 9I$ dobijamo

$$LR = PB,$$

gde su

$$L = \begin{bmatrix} 1 & 0 & 0 \\ -4/5 & 1 & 0 \\ -1/5 & 2/3 & 1 \end{bmatrix}, \quad R = \begin{bmatrix} -5 & 1 & 4 \\ 0 & 9/5 & 21/5 \\ 0 & 0 & -1 \end{bmatrix}$$

i permutaciona matrica P odredjena indeksnim nizom $I = (1, 3)$.

Sada se metod inverzne iteracije (4.4.4) može izraziti u obliku

$$L\vec{y}_k = P\vec{v}_{k-1}, \quad R\vec{z}_k = \vec{y}_k, \quad \vec{v}_k = \vec{z}_k / \gamma_k,$$

čijom primenom se dobijaju rezultati koji su dati u tabeli 4.4.1.

Tabela 4.4.1

k	$(\vec{v}_k)_1$	$(\vec{v}_k)_2$	$(\vec{v}_k)_3$	β_k
1	0.	1.	-1.	6.
2	-0.2	1.	-0.5	9.3
3	-0.17241	1.	-0.48276	9.34483
4	-0.17200	1.	-0.48000	9.34800
5	-0.17185	1.	-0.47980	9.34835
6	-0.17184	1.	-0.47977	9.34838

Za startni vektor smo uzeli $\vec{v}_0 = [1 \ 0 \ 0]^T$. U poslednjoj koloni tabele data je veličina $\beta_k = p + 1/\gamma_k$, koja daje aproksimaciju za odgovarajuću sopstvenu vrednost λ . Vidimo da je ta sopstvena vrednost približno jednaka 9.34838.

3. Metodi deflacijske. Metodi ove klase se sastoje u konstrukciji niza matrica $A_n (=A), A_{n-1}, \dots, A_1$, čiji je red jednak indeksu i pritom

$$\text{Sp}(A_n) \supset \text{Sp}(A_{n-1}) \supset \dots \supset \text{Sp}(A_1),$$

gde je $\text{Sp}(A_k)$ spektar matrice A_k .

Opisacemo sada jedan specijalan, ali važan slučaj metoda deflacijske kada je matrica A hermitska.

Neka je $\vec{x} = [x_1 \ x_2 \ \dots \ x_n]^T$ sopstveni vektor matrice A koji odgovara sopstvenoj vrednosti λ i takav da je normiran

$$(\vec{x}, \vec{x}) = \vec{x}^* \vec{x} = \|\vec{x}\|_E^2 = 1$$

i da mu je prva koordinata x_1 nenegativna.

Posmatrajmo matricu

$$(4.4.5) \quad P = I - 2\vec{w}\vec{w}^*,$$

gde je vektor $\vec{w} = [w_1 \ w_2 \ \dots \ w_n]^T$ definisan pomoću prvog vektora $\vec{e}_1 = [1 \ 0 \ \dots \ 0]^T$ iz prirodnog bazisa prostora R^n na sledeći način

$$(4.4.6) \quad \vec{w}^* \vec{w} = \|\vec{x}\|_E^2 = 1, \quad w_1 \geq 0,$$

$$(4.4.7) \quad P\vec{e}_1 = \vec{x}.$$

Matrica P ima oblik

$$(4.4.8) \quad P = \begin{bmatrix} 1 - 2w_1 \bar{w}_1 & -2w_1 \bar{w}_2 & \dots & -2w_1 \bar{w}_n \\ -2w_2 \bar{w}_1 & 1 - 2w_2 \bar{w}_2 & & -2w_2 \bar{w}_n \\ \vdots & & & \\ -2w_n \bar{w}_1 & -2w_n \bar{w}_2 & & 1 - 2w_n \bar{w}_n \end{bmatrix}.$$

Primetimo da je $P^* = P$ što znači da je i matrica P hermit-ska. Štaviše, s obzirom na (4.4.6), direktnim množenjem vidimo da je $P^*P = P^2 = I$ pa zaključujemo da je matrica P unitarna (videti definiciju 1.4.4).

Na osnovu (4.4.7) nalazimo koordinate vektora w. Tako iz $1 - 2w_1 \bar{w}_1 = x_1$ i $-2w_k \bar{w}_1 = x_k$ ($k=2, \dots, n$) sleduje

$$w_1 = \sqrt{\frac{1-x_1}{2}} \quad \text{i} \quad w_k = -\frac{x_k}{2w_1} \quad (k=2, \dots, n).$$

Primetimo da je $\bar{w}_1 = w_1 > 0$.

Sada, na osnovu (4.4.7) i $A\vec{x} = \lambda\vec{x}$ nalazimo da je

$$AP\vec{e}_1 = P\vec{e}_1,$$

odakle zaključujemo da je

$$P^*AP\vec{e}_1 = \lambda\vec{e}_1,$$

tj. da je \vec{e}_1 sopstveni vektor matrice $B = P^*AP = PAP$. Primetimo, takodje, da je prva kolona u matrici B upravo vektor $\lambda \vec{e}_1$, tj.

$$B = \begin{bmatrix} \lambda & b_{12} & b_{13} & \cdots & b_{1n} \\ 0 & b_{22} & b_{23} & & b_{2n} \\ 0 & b_{32} & b_{33} & & b_{3n} \\ \vdots & & & & \\ 0 & b_{n2} & b_{n3} & & b_{nn} \end{bmatrix} = \begin{bmatrix} \lambda & | & \vec{b}_{n-1}^T \\ \hline \vec{0}_{n-1} & | & A_{n-1} \end{bmatrix},$$

gde smo sa A_{n-1} označili matricu reda $n-1$ koja se poklapa sa ogradjenim blokom, $\vec{0}_{n-1}$ je nula-vektor reda $n-1$, i najzad, $\vec{b}_{n-1}^T = [b_{12} \ b_{13} \ \cdots \ b_{1n}]^T$.

S obzirom da je matrica B slična (kažemo još i unitarno slična) sa matricom A , zaključujemo da je

$$\text{Sp}(A_{n-1}) = \text{Sp}(A_n) \setminus \{\lambda\} \quad (A_n = A).$$

Da bismo dobili matricu A_{n-2} postupamo na sličan način. Umesto dosadašnje matrice P koristimo matricu

$$P_1 = \begin{bmatrix} 1 & | & \vec{0}_{n-1}^T \\ \hline \vec{0}_{n-1} & | & Q \end{bmatrix},$$

gde je Q matrica reda $n-1$ oblika (4.4.5) i zadovoljava uslove (4.4.6) i (4.4.7) u odnosu na sopstveni vektor \vec{y} i sopstvenu vrednost μ matrice A_{n-1} . Kako je $P_1^{-1} = P_1^* = P_1$ zaključujemo da je i matrica P_1 unitarna.

Sada matrica $C = P_1 B P_1 = P_1 P A P P_1$ ima oblik

$$C = \begin{bmatrix} \lambda & c_{12} & c_{13} & \cdots & c_{1n} \\ 0 & \mu & c_{23} & & c_{2n} \\ 0 & 0 & c_{33} & & c_{3n} \\ \vdots & & & & \\ 0 & 0 & c_{n3} & & c_{nn} \end{bmatrix} = \begin{bmatrix} \lambda & c_{12} & c_{13} & \cdots & c_{1n} \\ 0 & \mu & c_{23} & & c_{2n} \\ 0 & 0 & \hline \vdots & & A_{n-2} \\ 0 & 0 & & & \end{bmatrix},$$

gde je A_{n-2} matrica reda $n-2$. Nastavljajući ovakav postupak do-

lazimo do gornje trougaone matrice koja je unitarno slična sa polaznom matricom A. Imajući u vidu da je matrica A hermitska zaključujemo da je ona unitarno slična sa dijagonalnom matricom.

Izloženi postupak zahteva pre svakog koraka određivanje jedne sopstvene vrednosti i njoj odgovarajućeg sopstvenog vektora, što se može učiniti nekim od prethodno izloženih metoda. Tako pre prvog koraka treba odrediti sopstvenu vrednost λ i sopstveni vektor \vec{x} matrice A, pre drugog koraka sopstvenu vrednost μ i sopstveni vektor \vec{y} matrice $(n-1)$ -og reda A_{n-1} , itd.

Jasno je da su sopstvene vrednosti matrice A dijagonalni elementi dobijene trougaone matrice, tj. $\lambda_1 = \lambda$, $\lambda_2 = \mu$, itd. Ostaje pitanje šta je sa sopstvenim vektorima matrice A? Jasno je za sada da je $\vec{x}_1 = \vec{x}$. Pokazaćemo kako se na osnovu dobijenih rezultata može naći drugi sopstveni vektor matrice A.

Neka su koordinate sopstvenog vektora \vec{y} redom y_2, \dots, y_n . U cilju nalaženja, najpre, sopstvenog vektora \vec{y}' matrice B stavimo $\vec{y}' = [y_1 \ y_2 \ \dots \ y_n]^T$ i pokušajmo da odredimo y_1 .

Kako je

$$B\vec{y}' = \begin{bmatrix} \lambda & & \vec{b}_{n-1}^T \\ 0 & \ddots & \vdots \\ 0 & & A_{n-1} \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ \vec{y} \end{bmatrix} = \begin{bmatrix} \lambda y_1 + \vec{b}_{n-1}^T \vec{y} \\ A_{n-1} \vec{y} \end{bmatrix},$$

tj.

$$B\vec{y}' = \begin{bmatrix} \lambda y_1 + \vec{b}_{n-1}^T \vec{y} \\ \mu \vec{y} \end{bmatrix}$$

izlazi da mora biti

$$y_1 + \vec{b}_{n-1}^T \vec{y} = y_1.$$

Ako je $\lambda \neq \mu$, na osnovu prethodne jednakosti dobijamo

$$y_1 = \frac{1}{\lambda - \mu} \vec{b}_{n-1}^T \vec{y} = \frac{1}{\lambda - \mu} (b_{12} y_2 + \dots + b_{1n} y_n).$$

Sada jednostavno nalazimo sopstveni vektor \vec{x}_2 matrice A koji odgovara sopstvenoj vrednosti $\lambda_2 \neq \mu$. Žaista, kako je $PAP\vec{y}' = \mu \vec{y}'$, tj. $A(P\vec{y}') = \mu (P\vec{y}')$ zaključujemo da je $\vec{x}_2 = P\vec{y}'$.

Slično se može postupiti i za određivanje ostalih sopstvenih vektora.

4.4.5. Jacobiev metod

Ovim odeljkom počinjemo izlaganje metoda za rešavanje kompletног problema sopstvenih vrednosti, tj. za određivanje svih sopstvenih vrednosti i odgovarajućih sopstvenih vektora. Metod koji izlažemo u ovom odeljku potiče od C.G.J. Jacobia (1846), a primenljiv je za hermitske matrice. Jacobiev metod se zasniva na transformaciji hermitske matrice A na dijagonalnu matricu D , čije su sopstvene vrednosti dijagonalni elementi. Na osnovu teoreme 1.5.2 postoji unitarna matrica H takva da je

$$(4.5.1) \quad H^*AH = D.$$

Kako je $H^* = H^{-1}$, na osnovu (4.5.1) matrice A i D su slične pa imaju iste sopstvene vrednosti.

Razmotrimo najpre slučaj transformacije

$$(4.5.2) \quad B = R^*AR,$$

gde je R proizvoljna unitarna matrica.

Teorema 4.5.1. Matrice B i A iz (4.5.2) imaju jednake Schmidtovе norme, tj. $\varepsilon(B) = \varepsilon(A)$.

Dokaz. Matrice A i B su slične. Neka su λ_i ($i = 1, 2, \dots, n$) njihove sopstvene vrednosti. Kako je

$$\varepsilon(A)^2 = \sum_{i,j=1}^n |a_{ij}|^2 = \text{tr}(A^*A) = \text{tr}(A^2) = \sum_{i=1}^n \lambda_i^2,$$

zaključujemo da je $\varepsilon(B) = \varepsilon(A)$.

Osnovna ideja u Jacobievom metodu je u konstrukciji niza sličnih matrica $\{A_k\}_{k \in \mathbb{N}}$, startujući od $A_1 = A$, sa strategijom minimizacije veličine $\text{vd}(A_k)$ na svakom koraku, gde je sa $\text{vd}(A)$ označena "norma" vandijagonalnih elemenata matrice A , tj.

$$\text{vd}(A) = \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|^2.$$

Primetimo da je $\text{vd}(A) + d(A) = \varepsilon(A)$, gde je $d(A) = \sum_{i=1}^n |a_{ii}|^2$.

Strategija minimizacije treba da obezbedi

$$\text{vd}(A_{k+1}) \leq \text{vd}(A_k) \quad \text{i} \quad \lim_{k \rightarrow +\infty} \text{vd}(A_k) = 0.$$

Ostaje pitanje kako generisati niz A_k , tj. kakve unitarne matrice koristiti za transformaciju A_k u A_{k+1} .

Posmatrajmo tzv. matricu rotacije $R = R(p, q)$ čiji su elementi

$$r_{pp} = e^{i\alpha} \cos \theta, \quad r_{pq} = e^{i\beta} \sin \theta,$$

$$r_{qp} = -e^{-i\beta} \sin \theta, \quad r_{qq} = e^{-i\alpha} \cos \theta,$$

$$r_{ij} = \delta_{ij} \quad (\text{u ostalim slučajevima}),$$

gde su θ, α, β realni brojevi. Direktnim množenjem pokazujemo da je $R^*R = I$, tj. da je matrica rotacije unitarna. Obično ovu matricu nazivamo elementarnom matricom rotacije po uglu θ u ravni (p, q) .

Mada je Jacobiev metod primenljiv za hermitske matrice, jednostavnosti radi, izložićemo ga za slučaj kada je A realna simetrična matrica. U prilog ovome ide i činjenica da se problem sopstvenih vrednosti za hermitsku matricu A reda n može svesti na problem sopstvenih vrednosti za jednu simetričnu matricu čiji je red $2n$. Naime, tada se A može razložiti na dve realne matrice S i K u obliku $A = S + iK$, gde je S simetrična, a K koso-simetrična matrica. Sopstvena vrednost λ i sopstveni vektor $\vec{v} = \vec{x} + i\vec{y}$ matrice A zadovoljavaju jednakost

$$(4.5.3) \quad \begin{bmatrix} S & -K \\ K & S \end{bmatrix} \cdot \begin{bmatrix} \vec{x} \\ \vec{y} \end{bmatrix} = \lambda \begin{bmatrix} \vec{x} \\ \vec{y} \end{bmatrix},$$

pri čemu ako je λ prosta sopstvena vrednost matrice A , onda je ona dvostruka za matricu reda $2n$ koja se pojavljuje u (4.5.3).

Iz prethodno navedenih razloga na dalje pretpostavljamo da je A realna simetrična matrica. U matrici rotacije možemo tada uzeti $\alpha = \beta = 0$, tako da je ova matrica sada ortogonalna. Prema tome, imamo

$$r_{pp} = \cos \theta = c, \quad r_{pq} = \sin \theta = s,$$

$$r_{qp} = -\sin \theta = -s, \quad r_{qq} = \cos \theta = c,$$

$$r_{ij} = \delta_{ij} \quad (\text{u ostalim slučajevima}).$$

Primetimo da će se elementi matrice B u transformaciji

$$(4.5.4) \quad B = R^T A R$$

poklapati sa odgovarajućim elementima matrice A , sem onih koji se nalaze u p -toj i q -toj vrsti i koloni. Štaviše, imamo

$$(4.5.5) \quad \begin{bmatrix} b_{pp} & b_{pq} \\ b_{qp} & b_{qq} \end{bmatrix} = \begin{bmatrix} c & -s \\ s & c \end{bmatrix} \begin{bmatrix} a_{pp} & a_{pq} \\ a_{qp} & a_{qq} \end{bmatrix} \begin{bmatrix} c & s \\ -s & c \end{bmatrix},$$

odakle, s obzirom na tvrdjenje teoreme 4.5.1, nalazimo

$$b_{pp}^2 + 2b_{pq}^2 + b_{qq}^2 = a_{pp}^2 + 2a_{pq}^2 + a_{qq}^2.$$

Kako je opet na osnovu teoreme 4.5.1, $\epsilon(A) = \epsilon(B)$ imamo

$$vd(B) = \epsilon(B) - d(B) = \epsilon(A) - d(B),$$

tj.

$$\begin{aligned} vd(B) &= vd(A) + d(A) - d(B) \\ &= vd(A) + a_{pp}^2 + a_{qq}^2 - (b_{pp}^2 + b_{qq}^2) \\ &= vd(A) + 2b_{pq}^2 - 2a_{pq}^2, \end{aligned}$$

odakle vidimo da će $vd(B)$ biti minimizirano ako je $b_{pq}=0$. Iz ovog uslova odredićemo $c=\cos\theta$ i $s=\sin\theta$ u matrici rotacije.

Kako je na osnovu (4.5.5)

$$b_{pq} = (c^2 - s^2)a_{pq} + cs(a_{pp} - a_{qq}),$$

nalazimo

$$(\cos^2\theta - \sin^2\theta)a_{pq} + \cos\theta\sin\theta(a_{pp} - a_{qq}) = 0,$$

tj.

$$\tau = \cotg(2\theta) = \frac{a_{qq} - a_{pp}}{2a_{pq}} \quad (a_{pq} \neq 0).$$

Ako je $a_{pq} = 0$, imamo $c=1$ i $s=0$, što znači da je R jedinična matrica. Stavimo $\tg\theta = t$. Kako je $\cotg(2\theta) = (1-t^2)/2t = \tau$, za određivanje t treba rešiti kvadratnu jednačinu

$$t^2 + 2\tau t - 1 = 0.$$

Obično se uzima rešenje ove jednačine sa manjim modulom, tj.

$$t = \frac{\operatorname{sgn} \tau}{|\tau| + \sqrt{\tau^2 + 1}},$$

što obezbeđuje uslov da je ugao rotacije $|\theta| \leq \pi/4$. Tada su elementi matrice rotacije

$$c = \frac{1}{\sqrt{1+t^2}} \quad i \quad s = tc.$$

Sa ovako određenom matricom rotacije R "norma" vandijagonalnih elemenata matrice B u (4.5.4) se minimizira, pri čemu se element na mestu (p,q) anulira. Ovakav postupak se ponavlja, tj. generiše se niz sličnih matrica

$$(4.5.6) \quad A_{k+1} = R_k^T A_k R_k \quad (k = 1, 2, \dots),$$

startujući sa $A_1 = A$, pri čemu još ostaje otvoreno pitanje kako birati matricu rotacije $R_k = R_k(p,q)$, tj. parove (p,q) , na svakom koraku u (4.5.6). Kod klasičnog Jacobievog metoda za (p,q) se uzima pozicija dominantnog elementa, tj. vandijagonalnog elementa matrice $A_k = [a_{ij}^{(k)}]$ sa najvećim modulom. Dakle, (p,q) odredjujemo iz uslova

$$|a_{pq}^{(k)}| = \max_{i < j} |a_{ij}^{(k)}|.$$

Elementi matrice A_k koji se menjaju pri transformaciji (4.5.6) su samo oni koji se nalaze u p -toj i q -toj vrsti i koloni:

$$\left. \begin{aligned} a_{ip}^{(k+1)} &= a_{pi}^{(k+1)} = c a_{ip}^{(k)} - s a_{iq}^{(k)} \\ a_{iq}^{(k+1)} &= a_{qi}^{(k+1)} = s a_{ip}^{(k)} + c a_{iq}^{(k)} \\ a_{pp}^{(k+1)} &= c^2 a_{pp}^{(k)} - 2cs a_{pq}^{(k)} + s^2 a_{qq}^{(k)}, \\ a_{qq}^{(k+1)} &= s^2 a_{pp}^{(k)} + 2cs a_{pq}^{(k)} + c^2 a_{qq}^{(k)}, \\ a_{pq}^{(k+1)} &= a_{qp}^{(k+1)} = 0. \end{aligned} \right\} \quad (i \neq p, q),$$

Za ovako konstruisani metod imamo $A_k \rightarrow D = [\lambda_i \delta_{ij}]$, kada $k \rightarrow +\infty$. Zaista, s obzirom na nejednakost

$$2(a_{pq}^{(k)})^2 \geq \frac{1}{N} \text{vd}(A_k) \quad (N = \frac{1}{2}n(n-1)),$$

imamo

$$\text{vd}(A_{k+1}) = \text{vd}(A_k) - 2(a_{pq}^{(k)})^2 \leq \left(1 - \frac{1}{N}\right) \text{vd}(A_k),$$

tj.

$$\text{vd}(A_k) \leq \left(1 - \frac{1}{N}\right)^{k-1} \text{vd}(A),$$

odakle sleduje $\lim_{k \rightarrow +\infty} \text{vd}(A_k) = 0$, odnosno $\lim_{k \rightarrow +\infty} A_k = D$. Konvergencija Jacobievog metoda je kvadratna u smislu da postoji konstanta $M (> 0)$ takva da je za dovoljno veliko k

$$\text{vd}(A_{k+N}) \leq M(\text{vd}(A_k))^2.$$

Pomoću (4.5.6) možemo pisati

$$(4.5.7) \quad A_{k+1} = H_k^T A H_k,$$

gde je $H_k = R_1 R_2 \dots R_k$. Za dovoljno veliko k , matrica H_k se može tretirati kao dovoljno dobra aproksimacija za unitarnu matricu H koja se javlja u (4.5.1). Jacobiev iterativni proces (4.5.6) se obično prekida kada je $\text{vd}(A_{k+1}) \leq \delta^2$, gde je δ unapred data tačnost.

U klasičnom Jacobievom metodu dosta se vremena može utrošiti na određivanje pozicije dominantnog elementa. Jedna modifikacija klasičnog Jacobievog metoda, poznata kao ciklični Jacobiev metod, uzima za (p, q) redom parove:

$$(1, 2), (1, 3), \dots, (1, n); (2, 3), \dots, (2, n); \dots; (n-1, n); (1, 2), \dots$$

Može se pokazati da ovaj metod ima, takodje, kvadratnu konvergenciju.

Ako je k dovoljno veliko u (4.5.7), tako da je zadovoljen kriterijum za zaustavljanje procesa, tada možemo smatrati da je

$$H_k^T A H_k = A_{k+1} = D = [\lambda_i \delta_{ij}]_{n \times n},$$

odakle sleduje

$$A H_k = H_k D,$$

što znači da su kolone matrice H_k sopstveni vektori matrice A .

Štaviše, ovi sopstveni vektori čine ortogonalni skup. Matrica $H_k = [h_{ij}^{(k)}]_{n \times n}$ se može generisati rekursivno pomoću matrice rotacije

$$H_k = H_{k-1} R_k (p, q) \quad (H_0 = I).$$

U skalarnom obliku imamo

$$\left. \begin{array}{l} h_{ip}^{(k+1)} = c h_{ip}^{(k)} - s h_{iq}^{(k)} \\ h_{iq}^{(k+1)} = s h_{ip}^{(k)} + c h_{iq}^{(k)} \end{array} \right\} \quad (i = 1, 2, \dots, n),$$

$$h_{ij}^{(k+1)} = h_{ij}^{(k)} \quad (\text{u ostalim slučajevima}).$$

Primer 4.5.1. Na matricu A iz primera 4.4.1 primenićemo klasičan Jacobiev metod. Stavimo

$$A_1 = A = \begin{bmatrix} 4 & 1 & 4 \\ 1 & 10 & 1 \\ 4 & 1 & 10 \end{bmatrix}.$$

Dominantni element je na poziciji $(p, q) = (1, 3)$. Tada je

$$\tau = \frac{a_{33} - a_{11}}{2a_{13}} = \frac{3}{4}, \quad t = \frac{1}{2}, \quad c = c_1 = \frac{2}{\sqrt{5}}, \quad s = s_1 = \frac{1}{\sqrt{5}},$$

pa je

$$R_1 = R_1(1, 3) = \begin{bmatrix} 2/\sqrt{5} & 0 & 1/\sqrt{5} \\ 0 & 1 & 0 \\ -1/\sqrt{5} & 0 & 2/\sqrt{5} \end{bmatrix},$$

$$H_1 = R_1, \quad A_2 = \begin{bmatrix} 2 & 1/\sqrt{5} & 0 \\ 1/\sqrt{5} & 10 & 3/\sqrt{5} \\ 0 & 3/\sqrt{5} & 12 \end{bmatrix}.$$

Nastavljujući ovaj postupak dobijamo sledeće rezultate:

a) Za elementarne matrice rotacije:

k	(p, q)	c _k	s _k
2	(2, 3)	0.89376	0.44855
3	(1, 2)	0.99852	0.05431
4	(1, 3)	0.99982	0.01872

b) Za nizove $\{H_k\}$ i $\{A_k\}$ (matrice A_k su simetrične; elementi donjeg trougla nisu navedeni):

$$H_2 = \begin{bmatrix} 0.89443 & -0.20060 & 0.39970 \\ 0. & 0.89376 & 0.44855 \\ -0.44721 & -0.40119 & 0.79940 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 2. & 0.39970 & 0.20060 \\ & 9.32668 & 0. \\ & & 12.67332 \end{bmatrix},$$

$$H_3 = \begin{bmatrix} 0.90400 & -0.15172 & 0.39970 \\ -0.04854 & 0.89244 & 0.44855 \\ -0.42476 & -0.42489 & 0.79940 \end{bmatrix},$$

$$A_4 = \begin{bmatrix} 1.97826 & 0. & 0.20030 \\ & 9.34842 & 0.01089 \\ & & 12.67332 \end{bmatrix},$$

$$H_4 = \begin{bmatrix} 0.89636 & -0.15172 & 0.41655 \\ -0.05693 & 0.89244 & 0.44756 \\ -0.43965 & -0.42489 & 0.79131 \end{bmatrix},$$

$$A_5 = \begin{bmatrix} 1.97451 & -0.00020 & 0. \\ & 9.34842 & 0.01089 \\ & & 12.67707 \end{bmatrix}.$$

Svi rezultati su zaokrugljeni na 5 decimala.

Na osnovu A_5 i H_4 imamo da su sopstvene vrednosti matrice A

$$\lambda_1 \approx 1.97451, \quad \lambda_2 \approx 9.34842, \quad \lambda_3 \approx 12.67707,$$

i njima odgovarajući sopstveni vektori

$$\vec{x}_1 \approx \begin{bmatrix} 0.89636 \\ -0.05693 \\ -0.43965 \end{bmatrix}, \quad \vec{x}_2 \approx \begin{bmatrix} -0.15172 \\ 0.89244 \\ -0.42489 \end{bmatrix}, \quad \vec{x}_3 \approx \begin{bmatrix} 0.41655 \\ 0.44756 \\ 0.79131 \end{bmatrix}.$$

Primetimo da smo u primeru 4.4.1 odredili, metodom inverzne matrice, sopstvenu vrednost λ_2 i odgovarajući sopstveni

vektor \vec{x}_2 (normiran u odnosu na koordinatu sa najvećim modulom).

Primedba 4.5.1. Ukoliko se ne zahteva rešavanje kompletног problema sopstvenih vrednosti, matricu H_k nije potrebno generisati. Na primer, ako je potreban samo jedan sopstveni vektor, recimo \vec{x}_m , tada se on može jednostavno dobiti primenom rotacija na vektor \vec{e}_m , čija je m -ta koordinata jednaka jedinici, a sve ostale su jednake nuli.

4.4.6. Givensov i Hauseholderov metod

Kao što smo videli u prethodnom odeljku, Jacobiev metod transformiše simetričnu (ili, uopšte hermitsku) matricu na dijagonalnu posle beskonačno koraka. Vandijagonalni elementi (a_{pq} i a_{qp}) koji se anuliraju na odredjenom koraku primene Jacobievog metoda, mogu u kasnijim koracima da postanu takvi da znatno odstupaju od nule, što je posebno izraženo kada je red matrice visok. Ova činjenica usporava algoritam. U ovom odeljku izložićemo dva metoda kod kojih je navedeni nedostatak uklonjen. Prvi od ovih metoda razvio je W.J. Givens (1954), a drugi A.S. Householder (1958). Metodi su takvi da kroz konačan broj koraka transformišu polaznu realnu simetričnu matricu na simetričnu trodijagonalnu matricu. U narednom odeljku razmatraćemo problem sopstvenih vrednosti za simetrične trodijagonalne matrice. I Givensov i Hauseholderov metod se mogu jednostavno preneti na hermitiske matrice. Štaviše, njihova primena na opšte matrice dovodi do redukcije matrice na tzv. Hessenbergovu formu ($a_{ij} = 0$ za $i \geq j+2$).

Kao i u prethodnom odeljku, i ovde ćemo izložiti pomenute metode za slučaj kada je data matrica A realna i simetrična.

Givensov metod. Ovaj metod redukcije se zasniva na sukcesivnoj primeni tzv. Givensove transformacije, pri čemu se posle konačnog broja rotacija matrica A transformiše na trodijagonalnu matricu.

Elemente matrice $A = [a_{ij}]_{n \times n}$ za čije indekse važi nejednakost $|i-j| > 1$ zvaćemo vantridijagonalnim elementima. Za njih ćemo uvesti takvo uredjenje da ćemo reći da je a_{ij} s-ti vantridijagonalni element ako je par (i,j) s-ti član cikličnog indeksnog niza $(1,3), (1,4), \dots, (1,n), (2,4), (2,5), \dots, (2,n), \dots, (n-2,n)$.

Primetimo da ovaj niz sadrži ukupno

$$M = (n-2) + (n-1) + \dots + 1 = \frac{1}{2}(n-1)(n-2)$$

parova.

Neka je $A_1 = A$ i

$$(4.6.1) \quad A_{k+1} = G_k^T A_k G_k \quad (k = 1, 2, \dots, M),$$

gde su matrice G_k , tzv. dvodimenzionalne rotacije, izabrane tako da prvih k vantridijagonalnih elemenata matrice A_{k+1} bude jednako nuli.

Teorema 4.6.1 (Givens). Neka je A realna simetrična matrica, niz $A_k = [a_{ij}^{(k)}]_{n \times n}$ definisan pomoću (4.6.1) i neka je $(p-1, q)$ k -ti par cikličnog indeksnog niza vantridijagonalnih elemenata. Ako se matrica rotacije $G_k = [g_{ij}^{(k)}]_{n \times n}$ definiše pomoću

$$g_{pp}^{(k)} = g_{qq}^{(k)} = c, \quad g_{pq}^{(k)} = -g_{qp}^{(k)} = -s,$$

$$g_{ij}^{(k)} = \delta_{ij} \quad (\text{u ostalim slučajevima}),$$

gde su

$$c = \cos \theta = \frac{1}{S} a_{p-1,p}^{(k)}, \quad s = \sin \theta = \frac{1}{S} a_{p-1,q}^{(k)},$$

$$S = \sqrt{(a_{p-1,p}^{(k)})^2 + (a_{p-1,q}^{(k)})^2},$$

ili sa $G_k = I$, ako je $a_{p-1,q}^{(k)} = 0$, imamo:

- (a) Matrice A_{k+1} su realne i simetrične;
- (b) Prvih k vantridijagonalnih elemenata matrice A_{k+1} su nula ($k = 1, 2, \dots, M$);
- (c) Matrica A_{M+1} je trodijagonalna.

Dokaz se može dati matematičkom indukcijom. Primetimo suštinsku razliku izmedju Givensovog i Jacobievog metoda. Naime, kod Jacobievog metoda sa matricom rotacije $R_k = R_k(p, q)$ se anulira samo element na poziciji (p, q) . U Givensovom metodu imamo da je u matrici A_k prvih $k-1$ vantridijagonalnih elemenata jednako nuli. Sa Givensovom rotacijom $G_k = G_k(p, q)$, odgovarajući elementi u matri-

ci A_{k+1} ostaju nepromenjeni, tj. jednaki nuli. Kako je, međutim,

$$a_{p-1,q}^{(k+1)} = -a_{p-1,p}^{(k)} \sin \theta + a_{p-1,q}^{(k)} \cos \theta,$$

izborom

$$\tan \theta = \frac{a_{p-1,q}^{(k)}}{a_{p-1,p}^{(k)}}$$

imamo da je i k-ti element jednak nuli, tj. $a_{p-1,q}^{(k+1)} = 0$. Upravo ovakav izbor $\tan \theta$ je uzet za određivanje elemenata c i s u matrici rotacije G_k .

Na osnovu (4.6.1) imamo

$$a_{pp}^{(k+1)} = c^2 a_{pp}^{(k)} + 2cs a_{pq}^{(k)} + s^2 a_{qq}^{(k)},$$

$$a_{pq}^{(k+1)} = a_{qp}^{(k+1)} = (c^2 - s^2) a_{pq}^{(k)} + cs(a_{qq}^{(k)} - a_{pp}^{(k)}),$$

$$a_{qq}^{(k+1)} = s^2 a_{pp}^{(k)} - 2cs a_{pq}^{(k)} + c^2 a_{qq}^{(k)},$$

$$a_{ip}^{(k+1)} = a_{pi}^{(k+1)} = c a_{ip}^{(k)} + s a_{iq}^{(k)} \quad \left. \right\} \quad (i \neq p, q),$$

$$a_{iq}^{(k+1)} = a_{qi}^{(k+1)} = -s a_{ip}^{(k)} + c a_{iq}^{(k)} \quad \left. \right\}$$

$$a_{ij}^{(k+1)} = a_{ji}^{(k+1)} = a_{ij}^{(k)} \quad (\text{u ostalim slučajevima}).$$

Primetimo da Givensov algoritam zahteva ukupno M korenovanja i aproksimativno $\frac{4}{3}n^3$ množenja.

Primer 4.6.1. Primenićećem Givensov metod na redukciju matrice

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{bmatrix}$$

na trodijagonalni oblik. Na osnovu prethodnog, za to je potrebno $M = 3$ koraka. Givensove rotacije su, pritom, odredjene pomoću:

k	(p, q)	c _k	q _k
1	(2, 3)	0.707107	0.707107
2	(2, 4)	0.816496	0.577350
3	(3, 4)	0.397360	0.917663

dok je odgovarajući niz simetričnih matrica A_k ($k=2,3,4$) (elementi donjeg trougla nisu navedeni):

$$A_2 = \begin{bmatrix} 1. & 1.414214 & 0. & 1. \\ & 7. & 2. & 9.899495 \\ & & 1. & 4.242641 \\ & & & 20. \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 1. & 1.732051 & 0. & 0. \\ & 20.666667 & 4.082493 & 9.428090 \\ & & 1. & 2.309401 \\ & & & 6.333333 \end{bmatrix},$$

$$A_4 = \begin{bmatrix} 1. & 1.732051 & 0. & 0. \\ & 20.666667 & 10.274023 & 0. \\ & & 7.175439 & 0.364642 \\ & & & 0.157895 \end{bmatrix}.$$

Svi rezultati su zaokrugljeni na šest decimala.

Hauseholderov metod. Ovaj metod se zasniva na korišćenju niza ortogonalnih transformacija oblika

$$H = I - 2\vec{w}\vec{w}^T, \quad \vec{w}^T\vec{w} = 1$$

sa pogodno izabranim vektorima \vec{w} . Nije teško pokazati da su ove matrice ortogonalne (videti odeljak 4.4.4, gde je razmatran opštiji slučaj).

Neka je A realna simetrična matrica reda n . Stavimo $A_1 = A$ i definišimo niz

$$(4.6.2) \quad A_{k+1} = H_k^T A_k H_k \quad (k = 1, 2, \dots, n-2),$$

gde je

$$(4.6.3) \quad H_k = I - 2\vec{w}\vec{w}^T.$$

Specijalni izbor vektora \vec{w} na svakom koraku u Hauseholderovom metodu obezbedjuje da je matrica A_{n-1} trodijagonalna. Inače, sve matrice u nizu $\{A_k\}$ su realne i simetrične.

Strategija Hauseholderovog metoda je da se posle prvog koraka anuliraju vantridijagonalni elementi u prvoj vrsti (i koloni), posle drugog koraka vantridijagonalni elementi u drugoj vr-

sti (i koloni), itd. Pri ovome, prethodno anulirani elementi se ne menjaju. Dakle, za matricu $A_k = [a_{ij}^{(k)}]$ ($k = 2, \dots, n-1$) imamo

$$a_{ij}^{(k)} = 0 \quad (1 \leq i \leq k-1 \wedge |i-j| > 1).$$

Da bi se ovakva transformacija obezbedila, za vektor \vec{w} u (4.6.3) treba uzeti:

$$\vec{w} = \vec{0}, \text{ ako je } m_k = \sum_{j=k+2}^n (a_{kj}^{(k)})^2 = 0,$$

i

$$\vec{w} = \beta \vec{v}, \text{ ako je } m_k \neq 0.$$

Pri ovome, za koordinate vektora $\vec{v} = [v_1 \dots v_n]^T$ treba uzeti

$$v_i = 0 \quad (i \leq k), \quad v_{k+1} = 2Sy^2, \quad v_i = a_{ki}^{(k)} \quad (i \geq k+2),$$

gde su

$$S^2 = \sum_{j=k+1}^n (a_{kj}^{(k)})^2 \quad (S = \operatorname{sgn}(a_{k,k+1}^{(k)})\sqrt{S^2}, \text{ ako je } a_{k,k+1}^{(k)} \neq 0),$$

$$y = \frac{1}{2K}(S + a_{k,k+1}^{(k)}), \quad 2K^2 = S^2 + a_{k,k+1}^{(k)}S, \quad \beta = \frac{1}{2Sy}.$$

Odredjivanje matrice A_{k+1} u (4.6.2) može se uprostiti, s obzirom na činjenicu da je

$$\begin{aligned} A_{k+1} &= (I - 2\vec{w}\vec{w}^T)A_k(I - 2\vec{w}\vec{w}^T) \\ &= A_k - 2\vec{w}\vec{w}^T A_k - 2A_k \vec{w}\vec{w}^T + 4\vec{w}\vec{w}^T A_k \vec{w}\vec{w}^T, \end{aligned}$$

tj.

$$A_{k+1} = A_k - 2\beta^2 (\vec{v}\vec{u}^T + \vec{u}\vec{v}^T),$$

gde su

$$\vec{u} = \vec{\xi} - a\vec{v}, \quad \vec{\xi} = A_k \vec{v}, \quad a = \beta^2 \vec{v}^T \vec{\xi}.$$

Primetimo da Hauseholderov metod zahteva $n-2$ korenovanja i aproksimativno $\frac{2}{3}n^3$ množenja, što je dva puta manje od broja množenja u Givensovom algoritmu.

Primer 4.6.2. Primenimo Hauseholderov metod na transformaciju matrice A iz prethodnog primera. Sada je potrebno samo dva koraka ($n-2 = 2$).

Prvi korak (k=1): Imamo $m_1 = 2 \neq 0$, $s^2 = 3$, $s = 1.732051$, $K = 1.538189$, $y = 0.888074$, $\beta = 0.325058$, $a = 10.479274$,

$$\vec{v} = \begin{bmatrix} 0. \\ 2.732051 \\ 1. \\ 1. \end{bmatrix}, \quad \vec{\xi} = \begin{bmatrix} 4.732051 \\ 12.464102 \\ 24.196152 \\ 40.928203 \end{bmatrix}, \quad \vec{u} = \begin{bmatrix} 4.732051 \\ -16.165808 \\ 13.716878 \\ 30.448929 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} 1. & -1.732051 & 0. & 0. \\ & 20.666667 & -1.503206 & -10.163460 \\ & & 0.202565 & 0.666667 \\ & & & 7.130768 \end{bmatrix};$$

Drugi korak (k=2): Sada imamo $m_2 = 103.295919 \neq 0$,

$s^2 = 105.555556$, $s = -10.274023$, $K = 7.778160$,

$y = -0.757070$, $\beta = 0.064283$, $a = 3.818322$,

$$\vec{v} = \begin{bmatrix} 0. \\ 0. \\ -11.777230 \\ -10.163460 \end{bmatrix}, \quad \vec{\xi} = \begin{bmatrix} 0. \\ 120.999532 \\ -9.161295 \\ -80.324767 \end{bmatrix}, \quad \vec{u} = \begin{bmatrix} 0. \\ 120.999532 \\ 35.819733 \\ -41.507243 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 1. & -1.732051 & 0. & 0. \\ & 20.666667 & 10.274023 & 0. \\ & & 7.175439 & -0.364642 \\ & & & 0.157895 \end{bmatrix}.$$

Elementi donjeg trougla u simetričnim matricama A_2 i A_3 nisu navedeni.

Primetimo da se dobijena trodijagonalna matrica razlikuje od one koja je dobijena Givensovim algoritmom. Naravno, ove dve matrice su slične, jer egzistira transformacija sličnosti sa dijagonalnom matricom $D = \text{diag}(1, -1, -1, 1)$.

Napomenimo da se Givensovim i Hauseholderovim metodom hermitske matrice transformišu na hermitske trodijagonalne matrice. Štaviše, opšte kompleksne matrice se transformišu na Hessenbergov oblik.

4.4.7. Problem sopstvenih vrednosti za simetrične trodijagonalne matrice

Neka je A realna simetrična trodijagonalna matrica reda n čije će ne-nula elemente označiti sa

$$\begin{aligned} a_{ii} &= b_i \quad (i = 1, \dots, n), \\ a_{i,i-1} &= a_{i-1,i} = c_i \quad (i = 2, \dots, n). \end{aligned}$$

Sa $p_k(\lambda)$ označimo glavni minor reda k matrice $A - \lambda I$, tj.

$$p_k(\lambda) = \begin{vmatrix} b_1 - \lambda & c_2 & & & & & & O \\ c_2 & b_2 - \lambda & c_3 & & & & & \\ & \ddots & \ddots & \ddots & & & & \\ & & \ddots & \ddots & \ddots & & & \\ O & & & c_{k-1} & b_{k-1} - \lambda & c_k & & \\ & & & & c_k & b_k - \lambda & & \end{vmatrix}$$

i definišimo $p_0(\lambda) = 1$. Primetimo da je $p_1(\lambda) = b_1 - \lambda$.

Razvijanjem determinante $p_k(\lambda)$ po elementima poslednje vrste dobijamo

$$p_k(\lambda) = (b_k - \lambda)p_{k-1}(\lambda) - c_k^2 p_{k-2}(\lambda).$$

Vrednost karakterističnog polinoma matrice A za vrednost λ , tj. $p_n(\lambda) = \det(A - \lambda I)$, možemo jednostavno odrediti, na osnovu prethodnog, korišćenjem tročlane rekurentne relacije

$$(4.7.1) \quad \begin{aligned} p_k(\lambda) &= (b_k - \lambda)p_{k-1}(\lambda) - c_k^2 p_{k-2}(\lambda) \quad (k=2, \dots, n), \\ p_0(\lambda) &= 1, \quad p_1(\lambda) = b_1 - \lambda. \end{aligned}$$

Jedan jednostavan metod za određivanje sopstvenih vrednosti simetričnih trodijagonalnih matrica zasniva se na korišćenju rekurentne relacije (4.7.1), metoda polovljenja intervala (videti odeljak 5.1.5) i tvrdjenju sledeće teoreme, koja se jednostavno dokazuje:

Teorema 4.7.1 (Givens). Neka su kod simetrične trodijagonalne matrice A reda n svi elementi $c_k \neq 0$. Tada važi:

- (1) Nule svakog polinoma p_k ($k=2, \dots, n$) su realne, razli-

čite i razdvojene nulama polinoma p_{k-1} :

(2) Ako je $p_n(\lambda) \neq 0$, broj sopstvenih vrednosti matrice A manjih od λ je jednak broju promene znaka $s(\lambda)$ u nizu

$$(4.7.2) \quad p_0(\lambda), p_1(\lambda), \dots, p_n(\lambda).$$

Ako je neko $p_k(\lambda) = 0$, onda se na tom mestu u nizu (4.7.2) može uzeti bilo koji znak, s obzirom da je $p_{k-1}(\lambda)p_{k+1}(\lambda) < 0$.

Primetimo da u teoremi egzistira uslov $c_k \neq 0$ za svako $k = 2, \dots, n$. Ako je, na primer, za neko $k=m$, $c_m=0$, tada se problem pojednostavljuje, jer se raspada na dva problema nižeg reda (m i $n-m$). Naime, matrica A postaje

$$A = \begin{bmatrix} A' & O \\ O & A'' \end{bmatrix},$$

gde su A' i A'' trodijagonalne simetrične matrice reda m i $n-m$ respectivno, i u tom slučaju je

$$\det(A - \lambda I) = \det(A' - \lambda I) \det(A'' - \lambda I).$$

Korišćenjem više vrednosti za λ moguće je sistematskom primenom teoreme 4.7.1, odrediti disjunktne intervale u kojima leže sopstvene vrednosti matrice A. Dakle, ako nadjemo da je

$$s(\lambda_1) = s_1 \quad i \quad s(\lambda_2) = s_2 = s_1 + 1 \quad (\lambda_1 < \lambda_2),$$

na osnovu teoreme 4.7.1 imamo da u intervalu (λ_1, λ_2) leži samo jedna sopstvena vrednost matrice A. Tada se za njeno određivanje može iskoristiti jednostavni metod polovljenja intervala (metod bisekcije), sužavajući ovaj polazni interval do zahtevane tačnosti (videti odeljak 5.1.5).

Za određivanje intervala u kojima leže sopstvene vrednosti može se koristiti i Gershgorinova teorema (videti odeljak 4.4.1), na osnovu koje su ti intervali

$$[b_1 - |c_2|, b_1 + |c_2|],$$

$$[b_i - |c_i| - |c_{i+1}|, b_i + |c_i| + |c_{i+1}|] \quad (i=2, \dots, n-1),$$

$$[b_n - |c_n|, b_n + |c_n|].$$

Nažalost, ovi intervali nisu disjunktni, i u opštem slučaju ne sadrže samo po jednu sopstvenu vrednost matrice A.

Primer 4.7.1. Za datu matricu

$$A = \begin{bmatrix} 1 & 1 & & \\ 1 & 3 & 2 & \\ 2 & 5 & 3 & \\ 3 & 7 & & \end{bmatrix}$$

imamo

$$p_0(\lambda) = 1, \quad p_1(\lambda) = 1 - \lambda, \quad p_2(\lambda) = (3 - \lambda)p_1(\lambda) - p_0(\lambda),$$

$$p_3(\lambda) = (5 - \lambda)p_2(\lambda) - 4p_1(\lambda), \quad p_4(\lambda) = (7 - \lambda)p_3(\lambda) - 9p_2(\lambda).$$

Neka je $\lambda = 0$. Tada imamo $p_0(0) = 1, p_1(0) = 1, p_2(0) = 2,$
 $p_3(0) = 6, p_4(0) = 24$. Znaci u nizu (4.7.2) su

+ + + + + ,

što znači da nema promene znaka, tj. da je $s(0) = 0$. Prema teoremi 4.7.1, matrica A nema negativnih sopstvenih vrednosti, tj. ona je pozitivno definitna.

Uzimajući za λ redom vrednosti $1, 2, 4, 5, 7, 9, 10$ dobijamo sledeće rezultate:

Tabela 4.7.1

λ	$p_0(\lambda)$	$p_1(\lambda)$	$p_2(\lambda)$	$p_3(\lambda)$	$p_4(\lambda)$	$s(\lambda)$
1	1	0	-1	-4	-15	1
2	1	-1	-2	-2	8	2
4	1	-3	2	14	24	2
5	1	-4	7	16	-31	3
7	1	-6	23	-22	-207	3
9	1	-8	47	-156	-111	3
10	1	-9	62	-274	264	4

Na osnovu vrednosti za $s(\lambda)$ zaključujemo da se u intervalima $(0, 1), (1, 2), (4, 5), (9, 10)$ nalazi po jedna sopstvena vrednost matrice A. Te sopstvene vrednosti na šest decimala su

$$\lambda_1 \approx 0.322548, \quad \lambda_2 \approx 1.745761, \quad \lambda_3 \approx 4.536620, \quad \lambda_4 \approx 9.395071.$$

Napomenimo da su ovo nule Laguerreovog polinoma L_4 .

Odredjivanje $p_k(\lambda)$ pomoću (4.7.1), u aritmetici sa pokretnom tačkom, je stabilno. Nažalost, i kod matrica ne tako visokog reda, vrednosti $p_k(\lambda)$, za k koje je blisko n, mogu da izadju izvan opsega brojeva, tj. da ove vrednosti postanu veće od maksimal-

nog broja koji može biti zapisan u memoriji računara, ili pak , manje od minimalnog broja, što se onda tretira kao nula. Ova činjenica ne dozvoljava normiranje vrednosti za $p_k(\lambda)$, jer se praktično javljaju obe vrste prekoračenja. Navedena teškoća dolazi posebno do izražaja kada matrica A ima bliske sopstvene vrednosti.

Jedan bolji pristup ovoj problematici ([8], [64]) je konstrukcija niza $q_k(\lambda)$, definisanog pomoću

$$q_k(\lambda) = p_k(\lambda)/p_{k-1}(\lambda) \quad (k = 1, \dots, n).$$

Broj $s(\lambda)$ je sada broj negativnih članova niza $\{q_k(\lambda)\}$. Na osnovu (4.7.1) dobijamo rekurentnu relaciju za $q_k(\lambda)$

$$q_k(\lambda) = b_k - \lambda - c_k^2/q_{k-1}(\lambda) \quad (k = 2, \dots, n),$$

gde je $q_1(\lambda) = b_1 - \lambda$.

Jedan efikasan metod, tzv. QR algoritam, može se uspešno primeniti, u specijalnom slučaju, i na trodijagonalne matrice, o čemu će biti reči u narednom odeljku.

4.4.8. LR i QR algoritmi

Ovaj odeljak posvećujemo tzv. faktorizacionim metodama. Prvi takav algoritam za rešavanje problema sopstvenih vrednosti za proizvoljnu matricu A opisao je H. Rutishauser ([50]) 1958.godine, nazivajući ga LR transformacijom. Metod se sastoji u konstrukciji niza matrica $\{A_k\}_{k \in N}$, startujući od $A_1 = A$, na sledeći način: Matrica A_k se faktorizuje na donju trougaonu matricu L_k sa jediničnom dijagonalom i gornju trougaonu matricu R_k , tj.

$$(4.8.1) \quad A_k = L_k R_k,$$

a zatim se naredni član niza određuje množenjem dobijenih faktora u obrnutom redosledu, tj.

$$A_{k+1} = R_k L_k.$$

Primetimo da su matrice A_{k+1} i A_k slične, jer su povezane transformacijom sličnosti

$$(4.8.2) \quad A_{k+1} = L_k^{-1} A_k L_k.$$

Faktorizacija (4.8.1) se može izvesti Gaussovim metodom eliminacije.

Ako stavimo

$$L^{(k)} = L_1 \dots L_k \quad i \quad R^{(k)} = R_k \dots R_1 ,$$

na osnovu (4.8.2) imamo

$$L^{(k)} A_{k+1} = A L^{(k)} ,$$

odakle je

$$L^{(k)} R^{(k)} = L^{(k-1)} A_k R^{(k-1)} = A L^{(k-1)} R^{(k-1)} .$$

Iterirajući poslednju jednakost dobijamo

$$L^{(k)} R^{(k)} = A L^{(k-2)} R^{(k-2)} = \dots = A^k ,$$

što znači da je $L^{(k)} R^{(k)}$ faktorizacija matrice A^k . Koristeći ove činjenice, Rutishauser (videti, takodje, [63]) je pokazao da pod određenim uslovima niz matrica $\{A_k\}$ konvergira ka nekoj gornje trougaonoj matrici, čiji elementi na glavnoj dijagonali daju sopstvene vrednosti matrice A. Obično se LR metod primenjuje na matrice prethodno svedene na gornju Hessenbergovu formu ($a_{ij} = 0$ za $i \geq j+2$). Ako je nekim od metoda opšta matrica svedena na donju Hessenbergovu formu, LR metod primenjujemo na transponovanu matricu koja ima iste sopstvene vrednosti. Sve matrice u nizu $\{A_k\}$ imaju Hessenbergov oblik. Ubrzavanje konvergencije niza $\{A_k\}$ može biti učinjeno uvodjenjem pogodnog pomeranja (šifta) p_k , tako da umesto A_k , faktorizujemo $B_k = A_k - p_k I = L_k R_k$, pri čemu je, dalje, $A_{k+1} = p_k I + R_k L_k$.

Nažalost, LR algoritam ima više nedostataka (videti monografiju Wilkinsona [63]). Na primer, faktorizacija ne egzistira za svaku matricu. Jedan bolji faktorizacioni metod razvili su J. G.F. Francis ([22]) i V.N. Kublanovskaya ([37]), u kome je matrica L zamenjena sa unitarnom matricom Q. Tako se dobija QR algoritam definisan pomoću

$$(4.8.3) \quad A_k = Q_k R_k, \quad A_{k+1} = R_k Q_k \quad (k = 1, 2, \dots),$$

startujući od $A_1 = A$. Primetimo da je $A_{k+1} = Q_k^* A_k Q_k$.

Ako stavimo

$$(4.8.4) \quad Q^{(k)} = Q_1 \dots Q_k \quad i \quad R^{(k)} = R_k \dots R_1 ,$$

slično kao kod LR metoda, nalazimo

$$(4.8.5) \quad Q^{(k)} A_{k+1} = A Q^{(k)} \quad i \quad Q^{(k)} R^{(k)} = A^k .$$

Teorema 4.8.1. Ako je matrica A regularna, tada egzistira dekompozicija $A = QR$, gde je Q unitarna, a R gornje trougaona matrica. Štavice, ako su dijagonalni elementi matrice R pozitivni, dekompozicija je jedinstvena.

QR faktorizacija (4.8.3) se može izvesti korišćenjem unitarnih matrica oblika $I - 2\vec{w}\vec{w}^*$. Tako, u cilju transformacije A_k na R_k , tj. redukcije kolona u A_k , imamo

$$(4.8.6) \quad (I - 2\vec{w}_{n-1}\vec{w}_{n-1}^*) \dots (I - 2\vec{w}_1\vec{w}_1^*) A_k = R_k.$$

Matrica Q_k je tada

$$(4.8.7) \quad Q_k = (I - 2\vec{w}_1\vec{w}_1^*) \dots (I - 2\vec{w}_{n-1}\vec{w}_{n-1}^*).$$

QR algoritam je efikasan ako polazna matrica ima (gornju) Hessenbergovu formu. Tada se prethodno pomenute unitarne matrice svode na dvodimenzionalne rotacije. Sve matrice A_k imaju Hessenbergovu formu. Dakle, problem sopstvenih vrednosti za opštu matricu je najpogodnije rešavati kroz dva koraka. Najpre, svesti matricu na Hessenbergov oblik, a zatim primeniti QR algoritam.

U specijalnom slučaju kada je polazna matrica trodijagonalna, matrice A_k u QR algoritmu su takođe trodijagonalne. U tom slučaju, uz korišćenje pogodno odabranog pomeraja p_k , QR algoritam postaje veoma efikasan za rešavanje problema sopstvenih vrednosti za trodijagonalne matrice.

Uvodjenjem pomeraja p_k , formule (4.8.3) postaju

$$(4.8.8) \quad A_k - p_k I = Q_k R_k, \quad A_{k+1} = p_k I + R_k Q_k \quad (k = 1, 2, \dots).$$

Prepostavimo da je $A (=A_1)$ simetrična trodijagonalna realna matrica. Sve ostale matrice $A_k = [a_{ij}^{(k)}]_{n \times n}$ su isto takve. Uprošćenja radi, uvedimo notaciju iz odjeljka 4.4.7, tj.

$$a_{ii}^{(k)} = b_i^{(k)} \quad (i = 1, \dots, n),$$

$$a_{i,i-1}^{(k)} = a_{i-1,i}^{(k)} = c_i^{(k)} \quad (i = 2, \dots, n),$$

i prepostavimo da su $c_i^{(1)} \neq 0$ ($i = 2, \dots, n$). Tada matrica A_1 ima sve različite sopstvene vrednosti.

Postoje dva načina za izbor pomeraja p_k . Prvi način je da se za p_k uzme vrednost elementa iz donjeg desnog ugla matrice A_k ,

tj. $p_k = b_n^{(k)}$. Drugi način izbora p_k je takav da se za p_k uzima ona sopstvena vrednost matrice tipa 2×2

$$(4.8.9) \quad \begin{bmatrix} b_{n-1}^{(k)} & c_n^{(k)} \\ c_n^{(k)} & b_n^{(k)} \end{bmatrix}$$

koja je bliža vrednosti $b_n^{(k)}$. Ovakav izbor

$$(4.8.10) \quad p_k = b_n^{(k)} + d - \operatorname{sgn}(d) \sqrt{d^2 + (c_n^{(k)})^2}, \quad d = \frac{1}{2}(b_{n-1}^{(k)} - b_n^{(k)})$$

potiče od Wilkinsona i daje bržu konvergenciju QR algoritma u odnosu na izbor $p_k = b_n^{(k)}$. U oba slučaja, inače imamo konvergenciju takvu da je

$$c_n^{(k)} c_{n-1}^{(k)} \rightarrow 0 \quad (k \rightarrow +\infty).$$

Ako $c_n^{(k)}$ postane zanemarljivo malo (na primer, mašinska nula), možemo uzeti da je $b_n^{(k)}$ jedna sopstvena vrednost matrice A i da pritom poslednju vrstu i kolonu u A_k izostavimo, tako da na dalje, rešavamo problem dimenzije $n-1$. Međutim, ako je $c_{n-1}^{(k)}$ zanemarljivo malo, dok je $c_n^{(k)}$ značajno, mi možemo odmah odrediti dve sopstvene vrednosti matrice A. To su, zapravo, sopstvene vrednosti matrice (4.8.9). Proces dalje nastavljamo tako što u matrici A_k izostavljamo poslednje dve vrste i kolone, i primenjujemo algoritam (4.8.8) na problem dimenzije $n-2$. Na ovaj način, QR algoritam postaje veoma efikasan jer proizvodi deflaciju reda matrice.

U praktičnoj primeni QR algoritma, dekompoziciju (4.8.3), odnosno (4.8.8), nije potrebno eksplicitno izračunavati. Naime, moguće je direktno odrediti (4.8.5), tj.

$$A_{k+1} = Q^{(k)T} A Q^{(k)},$$

gde je sada $Q^{(k)}$ ortogonalna matrica odredjena sa (4.8.4). Primenimo da matrica $Q^{(k)}$ konvergira ka ortogonalnoj matrici sopstvenih vektora matrice A (uporediti sa slučajem kod Jacobievog metoda, odeljak 4.4.5).

Kao što je već rečeno, unitarne (ortogonalne) matrice koje se pojavljuju u (4.8.6) i (4.8.7), svode se na dvodimenzionalne rotacije.

Neka je $z_p = z_p^{(k)} = I - \vec{w}_p \vec{w}_p^T$, $\vec{w}_p = [w_{p1} \dots w_{pn}]^T$, gde smo za koordinate vektora w_p uzeli

$$w_{pp} = \sin \frac{\theta}{2}, \quad w_{p,p+1} = \cos \frac{\theta}{2}, \quad w_{pi} = 0 \quad (i \neq p, p+1).$$

Na osnovu (4.4.8) imamo

$$Z_p = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & \ddots & 1 & & \\ & & & c & s & \\ & & & s & -c & \\ & & & & & 1 \\ & & & & & \ddots \\ & & & & & 1 \end{bmatrix} \begin{array}{l} \\ \\ \\ \leftarrow p \\ \leftarrow p+1 \end{array}$$

gde su $c = \cos \theta$ i $s = \sin \theta$.

Jednostavnosti radi, u matrici A_k izostavimo indeks iteracije, tj. uzmimo da je $b_i^{(k)} = b_i$ i $c_i^{(k)} = c_i$. Izaberimo sada ugao $\theta = \theta_1^{(k)} = \theta_1$, tako da su elementi prve kolone ispod glavne dijagonale u matrici $Z_1 A_k$ jednaki nuli. Tada imamo $\operatorname{ctg} \theta_1 = b_1/c_2$, tj.

$$c = \frac{b_1}{\sqrt{b_1^2 + c_2^2}} \quad i \quad s = \frac{c_2}{\sqrt{b_1^2 + c_2^2}} .$$

U matrici

$$z_1 A_k z_1 = \begin{bmatrix} b'_1 & c'_2 & d'_1 & 0 & \dots & 0 & 0 \\ c'_2 & b'_2 & c'_3 & 0 & & 0 & 0 \\ d'_1 & c'_3 & b'_3 & c'_4 & & 0 & 0 \\ 0 & 0 & c'_4 & b'_4 & & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & 0 & b'_{n-1} & c'_n & \\ 0 & 0 & 0 & 0 & c'_n & b'_n & \end{bmatrix}$$

elementi označeni sa primom su oni elementi matrice A_k koji se menjaju pri navedenoj transformaciji. Matrica A_{k+1} se dobija pomoću

$$A_{k+1} = z_{n-1} \dots z_2 z_1 A_k z_1 z_2 \dots z_{n-1},$$

gde se z_2, \dots, z_{n-1} konstruišu na sličan način kao i z_1 , tako da matrica A_{k+1} postane trodijagonalna. Proizvod svih ortogonalnih (dvodimenzionalnih) rotacija

$$Z = \prod_{k=1}^{\infty} (z_1^{(k)} \ z_2^{(k)} \ \dots \ z_{n-1}^{(k)})$$

daje matricu sopstvenih vektora. Naime, ovde je

$$\prod_{j=1}^k (z_1^{(j)} \ z_2^{(j)} \ \dots \ z_{n-1}^{(j)}) = Q^{(k)}.$$

Sada možemo rekurzivno formulisati QR algoritam za određivanje jedne sopstvene vrednosti λ i odgovarajućeg ortonormiranog vektora $\vec{x} = Z \vec{e}_1$ ($\vec{x}^T \vec{x} = 1$, $\vec{e}_1 = [1 \ 0 \ \dots \ 0]^T$) realne simetrične trodijagonalne matrice.

Neka su $\lambda^{(k)}$ i $\vec{y}^{(k)} = [y_1^{(k)} \ y_2^{(k)} \ \dots \ y_n^{(k)}]^T$ aproksimacije za sopstvenu vrednost λ i sopstveni vektor \vec{x} u k-tom iterativnom koraku ($k=1, 2, \dots$).

Startujući od $\vec{y}^{(1)} = \vec{e}_1$, tj. $y_1^{(1)} = 1$, $y_i^{(1)} = 0$ ($i=2, \dots, n$), k-ti iterativni korak se može iskazati na sledeći način ([26]):

Za $p = 1, 2, \dots, n-1$ određujemo

$$\alpha := [(\bar{c}_p^{(k)})^2 + (d_p^{(k)})^2]^{1/2}, \quad c := \bar{c}_p^{(k)} / \alpha, \quad s := d_p^{(k)} / \alpha,$$

$$b_p^{(k+1)} := c^2 \bar{b}_p^{(k)} + 2cs \tilde{c}_{p+1}^{(k)} + s^2 b_{p+1}^{(k)},$$

$$\bar{b}_{p+1}^{(k)} := s^2 \bar{b}_p^{(k)} - 2cs \tilde{c}_{p+1}^{(k)} + c^2 b_{p+1}^{(k)},$$

$$c_p^{(k+1)} := c \bar{c}_p^{(k)} + s d_p^{(k)} = \alpha,$$

$$\bar{c}_p^{(k)} := (\bar{b}_p^{(k)} - b_{p+1}^{(k)})cs + \tilde{c}_{p+1}^{(k)}(s^2 - c^2),$$

$$\tilde{c}_{p+2}^{(k)} := -c c_{p+2}^{(k)}, \quad d_{p+1}^{(k)} := s c_{p+2}^{(k)},$$

$$y_p^{(k+1)} := c \bar{y}_p^{(k)} + s y_{p+1}^{(k)}, \quad \bar{y}_{p+1}^{(k)} := s \bar{y}_p^{(k)} - c y_{p+1}^{(k)},$$

pomoću

$$d_1^{(k)} := c_2^{(k)}, \quad \bar{c}_1^{(k)} := b_1^{(k)} - \lambda^{(k)},$$

$$\bar{b}_1^{(k)} := b_1^{(k)}, \quad \tilde{c}_2^{(k)} := c_2^{(k)}, \quad \bar{y}_1^{(k)} := y_1^{(k)}$$

i $\lambda_k := p_k$, gde je p_k sopstvena vrednost matrice (4.8.9) odredjena

pomoću (4.8.10).

Iterativni proces se prekida kada, na primer, $c_n^{(k)}$ postane dovoljno malo. Kao što smo ranije rekli, u tom slučaju uzimamo da je $\lambda := b_n^{(k)}$ i $\vec{x} := \vec{y}^{(k)}$, odbacujemo poslednju vrstu i poslednju kolonu u matrici A_{k+1} , a zatim kompletan iterativni postupak ponavljamo nad ovom matricom reda $n-1$. Tada odredjujemo drugu sopstvenu vrednost i odgovarajući sopstveni vektor (videti metod deflacijske, odeljak 4.4.4), itd. Slično postupamo ukoliko je $c_{n-1}^{(k)}$ zanemarljivo, a $c_n^{(k)}$ značajno. U tom slučaju odredjujemo dve sopstvene vrednosti istovremeno.

Primetimo da nizovi $\bar{b}_p^{(k)}, \bar{c}_p^{(k)}, \bar{\tilde{c}}_p^{(k)}, \bar{\tilde{y}}_p^{(k)}$ ne zahtevaju dodatni memorijski prostor. Naime, oni se mogu memorisati na istim mestima gde se memorišu nizovi $b_p^{(k)}, c_p^{(k)}, y_p^{(k)}$.

Slično QR algoritmu razvijen je i QL algoritam ([12]), gde je L donja trougaona matrica, a Q unitarna matrica. Takođe, razvijen je i tzv. implicitni QL algoritam ([17]). U oba od pomenu-tih radova dati su i odgovarajući programi na ALGOL jeziku.

4.5. LITERATURA

1. J. Albrecht: *Fehlerabschätzungen bei Relaxationsverfahren zur numerischen Auflösung linearer Gleichungssysteme*. Numer. Math. 3 (1961), 188 - 201.
2. J. Albrecht: *Monotone Iteration und ihre Verwendung zur Lösung linearer Gleichungssysteme*. Numer. Math. 3 (1961), 345-358.
3. J. Albrecht: *Bemerkungen zum Iterationsverfahren von Schulz zur Matrix-inversion*. Z. Angew. Math. Mech. 41 (1961), 262-263.
4. M. Altman: *An optimum cubically convergent iterative method of inverting a linear bounded operator in Hilbert space*. J. Math. 10 (1960), 1107 - 1113.
5. R. Ansorge: *Über ein Iterationsverfahren von G. Schulz zur Ermittlung der Reziproken einer Matrix*. Z. Angew. Math. Mech. 39 (1959), 164-165.
6. N. Apostolatos und U. Kulisch: *Über die Konvergenz des Relaxationsverfahrens bei nicht-negativen und diagonaldominanten Matrizen*. Computing 2 (1967), 17-24.
7. N. S. Bahvalov: *Čislennye metody*. Nauka, Moskva, 1973.
8. W. Barth, R. S. Martin, J. H. Wilkinson: *Calculation of the eigenvalues of a symmetric tridiagonal matrix by the method bisection*. Numer. Math. 9 (1967), 386-393.
9. Å. Björck und G. Dahlquist: *Numerische Methoden*. R. Oldenbourg Verlag, München-Wien, 1972.

10. Z.Bohte: Numeričke metode. Državna založba Slovenije, Ljubljana, 1978.
11. Z.Bohte & M.Petkovšek: Gaussian elimination for diagonally dominant matrices. In: Numerical Methods and Approximation Theory (ed. by G.V.Milovanović), Faculty of Electronic Engineering, Niš, 1984, pp. 1-6.
12. H.Bowdler, R.S.Martin, C.Reinsch, J.H.Wilkinson: *The QR and QL algorithms for symmetric matrices*. Numer. Math. 11(1968), 293-306.
13. L.Collatz: Über die Konvergenzkriterion bei Iterationsverfahren für lineare Gleichungssysteme. Math. Z. 53(1950), 149-161.
14. L.Collatz: Funktionanalysis und numerische Mathematik. Springer-Verlag, Berlin-Heidelberg-New York, 1964.
15. Lj.Cvetković & D.Herceg: Some sufficient conditions for convergence of AOR-method. In: Numerical Methods and Approximation Theory (ed. by G.V.Milovanović), Faculty of Electronic Engineering, Niš, 1984, pp. 13-18.
16. B.P.Demidovič i I.A.Maron: Osnovy vyčislitel'noj matematiki. Nauka, Moskva, 1966.
17. A.Dubrulle, R.S.Martin, J.H.Wilkinson: *The implicit QL algorithm*. Numer. Math. 12(1968), 377-383.
18. W.Dück: Fehlerabschätzungen für das Iterationsverfahren von Schulz zur Bestimmung der Inversen einer Matrix. Z. Angew. Math. Mech. 40 (1960), 192-194.
19. D.A.Flanders & G.Shortley: Numerical determination of fundamental modes. J. Appl. Phys. 21(1950), 1326-1332.
20. G.E.Forsythe: Solving linear algebraic equations can be interesting. Bull. Amer. Math. Soc. 59(1953), 299-329.
21. G.Forsythe & C.B.Moler: Computer solution of linear algebraic systems. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1967.
22. J.G.F.Francis: The QR transformation - a unitary analogue to the LR transformation. Comput. J. 4(1961/62), 265-271, 332-345.
23. W.Givens: A method of computing eigenvalues and eigenvectors suggested by classical results on symmetric matrices. Appl. Math. Ser. Nat. Bur. Stand. 29(1953), 117-122.
24. G.H.Golub & C.F.Van Loan: Matrix computations. The Johns Hopkins University Press, Baltimore, Maryland, 1984.
25. G.H.Golub & 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.H.Golub & J.H.Welsch: Calculation of Gauss quadrature rules. Math. Comp. 23(1969), 221-230.
27. A.R.Gourlay & G.A.Watson: Computational methods for matrix eigenproblems. John Wiley & Sons, Chichester-New York, 1973.
28. A.S.Hauselholder: The approximate solution of matrix problems. J.Assoc. Comp. Mach. 5(1958), 205-243.
29. A.Hauselholder: The theory of matrices in numerical analysis. Blaisdell Publ. Co., New York, 1964.

30. D.Herceg i Z.Stojaković: Numeričke metode linearne algebre - zbirka zadataka. Gradjevinska knjiga, Beograd, 1981.
31. M.R.Hestenes & E.Stiefel: *Methods of conjugate gradients for solving linear systems.* J. Res. Nat. Bur. Stand. 49(1952), 409-436.
32. F.B.Hildebrand: *Introduction to numerical analysis.* McGraw-Hill, Inc., New York, 1974.
33. H.Hotelling: *Some new methods in matrix calculation.* Amer. Math. Statist. 14(1943), 1-34.
34. E.Isaacson & H.B.Keller: *Analysis of numerical methods.* John Wiley & Sons, Inc., New York, 1966.
35. M.A.Krasnosel'skiĭ i 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). Uspehy Matem. Nauk. 7(1952), 157-161.
36. V.I.Krylov, V.V.Bobkov, P.I.Monastyrnyi: *Vyčislitel'nye metody vysšej matematiki.* Tom 1. Vyšešaja škola, Minsk, 1972.
37. V.N.Kublanovskaja: *O nekotoryh algorifmah dlja rešenija polnoj problemy sobstvennyh značenii.* Ž. Vyčisl. Mat. i Mat. Fiz. 1(1961), 555-570.
38. U.Kulisch: *Über reguläre Zerlegungen von Matrizen einige Anwendungen.* Numer. Math. 11(1968), 444-449.
39. C.Lanszos: *Solution of systems of linear equations by minimized iterations.* J. Res. Nat. Bur. Stand. 49(1952), 33.
40. P.B.Madić: Loše rešljivi sistemi linearnih algebarskih jednacina i njihovo rešavanje (Doktorska disertacija).Beograd, 1965.
41. G.V.Milovanović i Dj.R.Djordjević: Programiranje numeričkih metoda na FORTRAN jeziku. Institut za dokumentaciju zaštite na radu "Edvard Kardelj", Niš, 1981.
42. D.S.Mitrinović i D.Ž.Djoković: Polinomi i matrice. ICS, Beograd, 1975.
43. B.N.Parlett: *The symmetric eigenvalue problem.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1980.
44. B.N.Parlett & W.G.Poole: *A geometric theory for the QR, LU, and power iterations.* SIAM J. Numer. Anal. 10(1973), 389-412.
45. D.G.Peradze: *The distribution of errors in solving a system of linear algebraic equations by the iteration method.* (Russian). Sakhartsh. SSR Mecn. Akad. Moambe. 50(1968), 289-294.
46. M.Petkov: Čisleni metodi na algebrata. Nauka i izkustvo, Sofia, 1974.
47. W.V.Petryshyn: *On the inversion of matrices and linear operators.* Proc. Amer. Math. Soc. 16(1965), 893-901.
48. A.Ralston: *A first course in numerical analysis.* McGraw-Hill, New York, 1965.
49. E.Reich: *On the convergence of the classical iterative method of solving linear simultaneous equations.* Ann. Math. Statist. 20(1949), 448-451.
50. H.Rutishauser: *Solution of eigenvalue problems with the LR-transformation.* Appl. Math. Ser. Nat. Bur. Stand. 49(1958), 47-81.

51. S.Schechter: *Relaxation methods for linear equations.* Comm. Pure and Appl. Math. 12(1959), 313-335.
52. G.Schulz: *Iterative Berechnung der reziproken Matrix.* Z. Angew. Math. Mech. 13(1933), 57-59.
53. I.M.Sobol': Čislennye metody Monte-Karlo. Nauka, Moskva, 1973.
54. R.V.Soutwell: *Relaxation methods in theoretical physics.* 2 vols. Oxford University Press. Fair Lawn., New Jersey, 1956.
55. G.W.Stewart: *Conjugate direction methods for solving systems of linear equations.* Numer. Math. 21(1973), 285-297.
56. E.Stiefel: *Relaxationsmethoden bester Strategie zur Lösung linearer Gleichungssysteme.* Comm. Math. Helv. 29(1955), 157-179.
57. J.Stoer: *Einführung in die Numerische Mathematik I.* Springer-Verlag, Berlin-Heidelberg-New York, 1973.
58. J.Stoer und R.Bulirsch: *Einführung in die Numerische Mathematik II.* Springer-Verlag, Berlin-Heidelberg-New York, 1973.
59. Z.Stojaković i D.Herceg: Numeričke metode linearne algebре. Gradjevinska knjiga, Beograd, 1982.
60. V.Strassen: *Gaussian elimination is not optimal.* Numer. Math. 13 (1969), 354-356.
61. R.S.Varga: *Matrix iterative analysis.* Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1962.
62. V.A.Vergasov, I.G.Žurkin, M.V.Krasikova i dr.: Vyčislitel'na-ja matematika. Nedra, Moskva, 1976.
63. J.H.Wilkinson: *The algebraic eigenvalue problem.* Oxford Uni-versity Press, London, 1965.
64. J.H.Wilkinson & C.Reinsch: *Handbook for automatic computati-on.* Vol. II Linear algebra. Springer-Verlag, Berlin-Heidel-berg-New York, 1971.
65. T.Yamamoto: *On the distribution of errors in the iterative solution of a system of linear equations.* Numer. Math. 24(1975), 71-79.
66. T.Yamamoto: *On the distribution of errors in the iterative solution of a system of linear equations II.* Numer. Math. 25(1976), 461-463.
67. D.Young: *Iterative methods for solving partial difference equations of elliptic type.* Trans. Amer. Math. Soc. 76(1954), 92-111.
68. D.M.Young: *Iterative solution of large systems.* Academic Pre-ss, New York, 1971.
69. D.M.Young & R.T.Gregory: *A survey of numerical mathematics.* Vol. II. Addison-Wesley Publ. Co., Reading, Massachusetts, 1973.

Nelinearne jednačine i sistemi

5.1. NELINEARNE JEDNAČINE

Ovo poglavlje je posvećeno opštim metodima za rešavanje nelinearnih jednačina oblika $f(x) = 0$, bez obzira da li su one algebarske ili transcendentne. Zbog specifičnosti koje poseđuju algebarske jednačine i zbog važnosti koje one u primenama imaju, u literaturi je razvijen ogroman broj metoda za njihovo rešavanje. U našem izlaganju ovom problemu posvećujemo posebno poglavlje.

5.1.1. Osnovne napomene

U trećoj glavi, a posebno u odeljku 3.1.3, data je opšta teorija o egzistenciji i jedinstvenosti rešenja nelinearnih jednačina oblika

$$(1.1.1) \quad f(x) = 0,$$

kao i o iterativnim procesima za nalaženje ovih rešenja. U ovom poglavlju razmatraćemo konstrukciju konkretnih iterativnih metoda za rešavanje jednačine (1.1.1). Svi metodi koji će biti izloženi primenjuju se na određivanje izolovanih korena jednačine (1.1.1). O problemu izolovanja korena nelinearnih jednačina videti, na primer, [56], [103], [104].

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 prost 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

$$(1.2.1) \quad f(a) = f(x_0) + f'(x_0)(a - x_0) + O((a - x_0)^2),$$

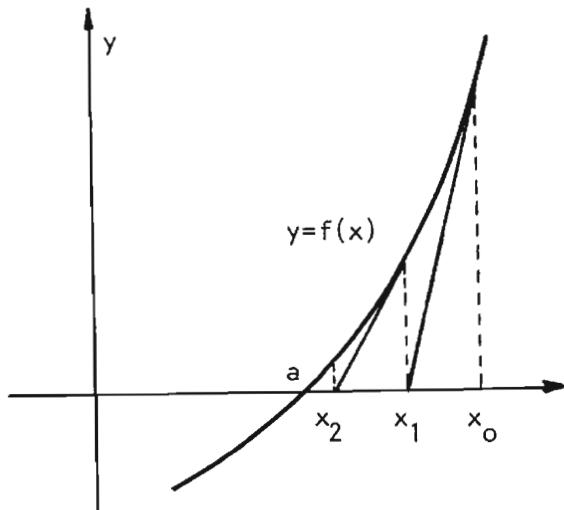
gde je $\xi = x_0 + \theta(a - x_0)$ ($0 < \theta < 1$). Imajući u vidu da je $f(a) = 0$, zanemarivanjem poslednjeg člana na desnoj strani u jednakosti (1.2.1), dobijamo

$$a \cong x_0 - \frac{f(x_0)}{f'(x_0)}.$$

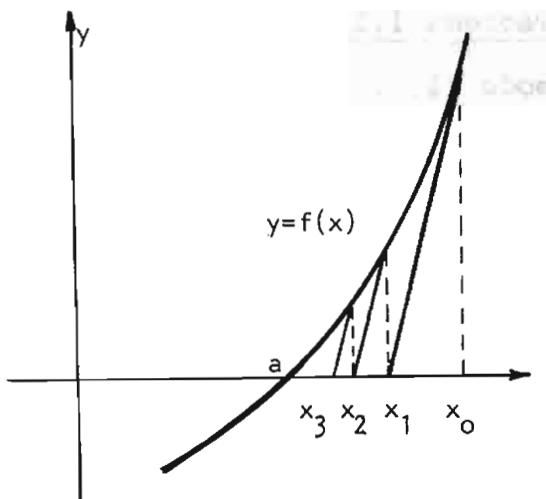
Sa x_1 označimo desnu stranu u poslednjoj približnoj jednakosti, tj.

$$(1.2.2) \quad x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Geometrijski x_1 predstavlja apscisu tačke preseka tangente na krivu $y = f(x)$, u tački $(x_0, f(x_0))$, sa x-osom (videti sl. 1.2.1).



Sl. 1.2.1



Sl. 1.2.2

Jednakost (1.2.2) sugerije konstrukciju iterativnog procesa

$$(1.2.3) \quad x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \quad (k = 0, 1, \dots),$$

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

$$(1.2.4) \quad \phi'(x) = 1 - \frac{f'(x)^2 - f(x)f''(x)}{f'(x)^2} = \frac{f(x)f''(x)}{f'(x)^2}.$$

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

$$(1.2.5) \quad |\phi'(x)| = \left| \frac{f(x)f''(x)}{f'(x)^2} \right| \leq q < 1.$$

Teorema 1.2.1. Ako $x_0 \in U(a)$, niz $\{x_k\}_{k \in \mathbb{N}_0}$ koji se generiše pomoću (1.2.3) konvergira ka tački $x = a$, pri čemu je

$$(1.2.6) \quad \lim_{k \rightarrow +\infty} \frac{x_{k+1} - a}{(x_k - a)^2} = \frac{f''(a)}{2f'(a)}.$$

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

$$(1.2.7) \quad \frac{x_{k+1} - a}{(x_k - a)^2} = \frac{f''(\xi_k)}{2f'(x_k)}.$$

Da bismo dokazali konvergenciju niza $\{x_k\}_{k \in \mathbb{N}_0}$ 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 $x_k \rightarrow a$, kada $k \rightarrow +\infty$, 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

$$x_{k+1} = x_k - \frac{x_k - \cos x_k}{1 + \sin x_k} = \frac{x_k \sin x_k + \cos x_k}{1 + \sin x_k} \quad (k = 0, 1, \dots).$$

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 određivanje n -tog korena iz pozitivnog broja a

$$x_{k+1} = x_k - \frac{x_k^n - a}{nx_k^{n-1}} = \frac{1}{n} \left\{ (n-1)x_k + \frac{a}{x_k^{n-1}} \right\} \quad (k = 0, 1, \dots).$$

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 $\{x_k\}_{k \in \mathbb{N}}$ bude monotон. Jedan odgovor na ovo pitanje dao je Fourier. Naime, ako f'' ne menja znak na $[\alpha, \beta]$ i ako se x_0 izabere tako da je $f(x_0)f''(x_0) > 0$, niz $\{x_k\}_{k \in \mathbb{N}}$ biće monotон. Ovo tvrdjenje sleduje iz (1.2.4).

Na osnovu teoreme 1.2.1 zaključujemo da Newtonov metod primenjen na određivanje 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

$$\lim_{k \rightarrow +\infty} \frac{x_{k+1} - a}{(x_k - a)^3} = \frac{f'''(a)}{3f'(a)}.$$

Primer 1.2.3. Posmatrajmo jednačinu

$$f(x) = x^3 - 3x^2 + 4x - 2 = 0.$$

Kako je $f(0) = -2$ 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 mo-

žemo primeniti Newtonov metod.

Startujući sa $x_0 = 1.5$, 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 $e_k = x_k - a$ ($k = 0, 1, \dots$) imamo

$$e_{k+1} = \frac{2e_k^3}{1 + 3e_k^2} \quad (k = 0, 1, \dots),$$

tj. $e_{k+1} \sim 2e_k^3$ ($k \rightarrow +\infty$).

Zaista, u ovom slučaju je $\frac{f'''(1)}{3f'(1)} = 2$.

U cilju smanjivanja broja računskih operacija, često se koristi sledeća modifikacija Newtonovog metoda

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \quad (k = 0, 1, \dots).$$

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 $(x_0, f(x_0))$ (videti sl. 1.2.2).

Iterativna funkcija modifikovanog Newtonovog metoda je

$$\phi_1(x) = x - \frac{f(x)}{f'(x_0)}.$$

Kako je $\phi_1(a) = a$ i $\phi'_1(a) = 1 - \frac{f'(a)}{f'(x_0)}$, zaključujemo da metod ima red konvergencije jedan, tj. važi

$$x_{k+1} - a \sim \left(1 - \frac{f'(a)}{f'(x_0)}\right) (x_k - a) \quad (k \rightarrow +\infty),$$

pri čemu je uslov

$$\left|1 - \frac{f'(x)}{f'(x_0)}\right| \leq q < 1,$$

analogon uslovu (1.2.5).

Newtonov metod može se razmatrati i kao specijalan slučaj tzv. uopštenog Newtonovog metoda

$$(1.2.8) \quad x_{k+1} = x_k - \frac{\psi(x_k) f(x_k)}{\psi'(x_k) f(x_k) + \psi(x_k) f'(x_k)} \quad (k = 0, 1, \dots),$$

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 parametar, iz (1.2.8) sleduje formula

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k) + \frac{p}{x_k} f(x_k)} \quad (k = 0, 1, \dots),$$

tj.

$$(1.2.9) \quad x_{k+1} = x_k \left\{ 1 - \frac{f(x_k)}{x_k f'(x_k) + p f(x_k)} \right\} \quad (k = 0, 1, \dots).$$

Metod definisan formulom (1.2.9) razmatran je u radu [4]. Specijalni slučaj ovog metoda, za $p = 1 - n$, poznat je kao metod Tihonova, u slučaju kada je f algebarski polinom stepena n .

Na kraju navedimo još jednu modifikaciju Newtonovog metoda, koja se sastoji u sukcesivnoj primeni formula

$$(1.2.10) \quad y_k = x_k - \frac{f(x_k)}{f'(x_k)}, \quad x_{k+1} = y_k - \frac{f(y_k)}{f'(y_k)} \quad (k = 0, 1, \dots).$$

Slično dokazu teoreme 1.2.1 mogu se dokazati tvrdjenja

$$\lim_{k \rightarrow +\infty} \frac{x_{k+1} - a}{(y_k - a)(x_k - a)} = \frac{f''(a)}{f'(a)}$$

i

$$\lim_{k \rightarrow +\infty} \frac{x_{k+1} - a}{(x_k - a)^3} = \frac{1}{2} \left(\frac{f''(a)}{f'(a)} \right)^2,$$

gde je $\lim_{k \rightarrow +\infty} x_k = a$. Dakle, iterativni proces definisan formулама (1.2.10) ima kubnu konvergenciju.

5.1.3. Newtonov metod za višestruke nule

Posmatrajmo jednačinu $f(x) = 0$, koja na $[a, b]$ ima koren $x = a$ višestrukosti $m (\geq 2)$. Ako pretpostavimo da $f \in C^{m+1} [a, b]$, tada je

$$f(a) = f'(a) = \dots = f^{(m-1)}(a) = 0, \quad f^{(m)}(a) \neq 0.$$

Naime, u ovom slučaju, f se može predstaviti u obliku

$$(1.3.1) \quad f(x) = (x-a)^m g(x),$$

gde $g \in C^{m+1} [a, b]$ 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

$$\Delta(x) = \frac{f(x)}{f'(x)} = \frac{(x-a)g(x)}{mg(x) + (x-a)g'(x)} \quad (x \neq a).$$

Ako stavimo $\Delta(a) \stackrel{\text{def}}{=} \lim_{x \rightarrow a} \Delta(x)$, tada je $\Delta(a) = 0$.

Iterativna funkcija Newtonovog metoda, primjenjenog na određivanje višestrukog korena, na osnovu prethodnog postaje

$$\phi(x) = x - \frac{(x-a)g(x)}{mg(x) + (x-a)g'(x)}.$$

Kako je $\phi(a) = a$, $\phi'(a) = 1 - \frac{1}{m}$, $\frac{1}{2} \leq \phi'(a) < 1$ ($m \geq 2$)

i ϕ' neprekidna funkcija, izlazi da postoji okolina korena $x = a$ u kojoj je $|\phi'(x)| \leq q < 1$, odakle zaključujemo da je Newtonov metod i u ovom slučaju konvergentan, ali se redom konvergencije jedan.

Ukoliko nam je unapred poznat red višestrukosti korena, tada se Newtonov metod može modifikovati tako da ima red konvergencije dva. Naime, treba samo uzeti

$$(1.3.2) \quad x_{k+1} = x_k - m \frac{f(x_k)}{f'(x_k)} \quad (k = 0, 1, \dots).$$

Primedba 1.3.1. Formalno, formula (1.3.2) predstavlja Newtonov metod primenjen na rešavanje jednačine

$$F(x) = \sqrt[m]{f(x)} = 0.$$

Teorema 1.3.1. Ako je x_0 izabрано dovoljno blisko korenu $x = a$, čiji je red višestrukosti m , tada niz $\{x_k\}_{k \in \mathbb{N}_0}$ definisan pomoću (1.3.2) konvergira ka a , pri čemu je

$$\frac{x_{k+1} - a}{(x_k - a)^2} \sim \frac{1}{m(m+1)} \cdot \frac{f^{(m+1)}(a)}{f^{(m)}(a)} \quad (k \rightarrow +\infty).$$

Dokaz ove teoreme može se naći, na primer, u [63].

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

$$x_{k+1} = x_k - \left[\frac{\frac{f(x)}{f'(x)}}{\left(\frac{f(x)}{f'(x)} \right)'} \right]_{x=x_k} \quad (k = 0, 1, \dots),$$

tj. formulu

$$x_{k+1} = x_k - \frac{f(x_k) f'(x_k)}{f'(x_k)^2 - f(x_k) f''(x_k)} \quad (k = 0, 1, \dots),$$

čiji je red konvergencije dva. Primetimo da se ova funkcija dobija iz (1.2.8) uzimajući $\Psi(x) = 1/f'(x)$.

5.1.4. Metod sečice

Ako se u Newtonovom metodu vrednost izvoda $f'(x_k)$ aproksimira pomoću podeljene razlike $\frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}$ dobijamo metod sečice

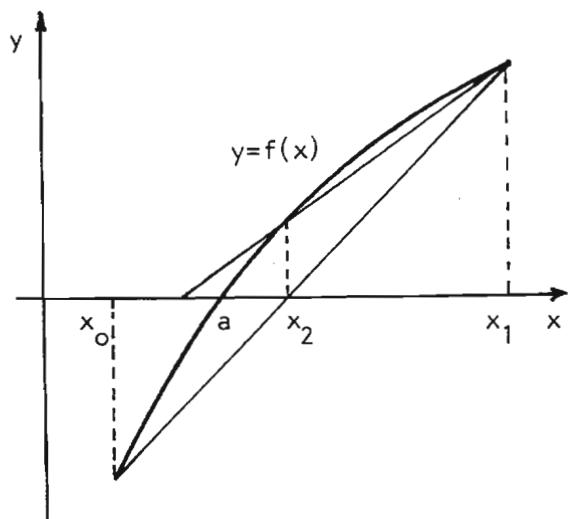
$$(1.4.1) \quad x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f(x_k) - f(x_{k-1})} f(x_k) \quad (k = 1, 2, \dots),$$

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 dala je na sl. 1.4.1.

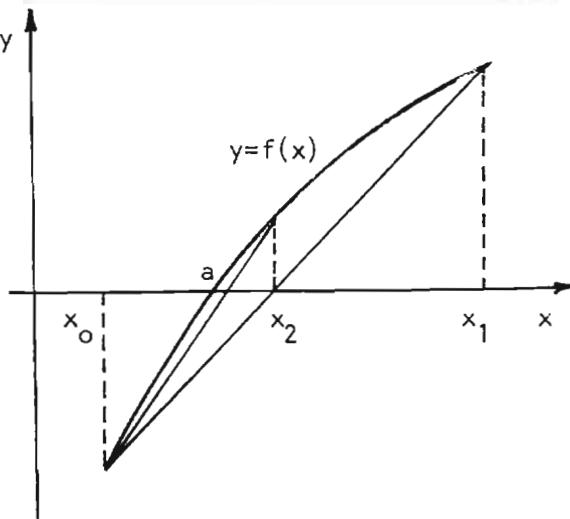
Neka na segmentu $[\alpha, \beta]$ postoji jedinstven koren $x = a$ jednačine $f(x) = 0$. Za ispitivanje konvergencije iterativnog procesa (1.4.1) pretpostavimo da $f \in C^2 [\alpha, \beta]$ i $f'(x) \neq 0 \ (\forall x \in [\alpha, \beta])$.

Ako stavimo $e_k = x_k - a$ ($k = 0, 1, \dots$), iz (1.4.1) sleduje

$$(1.4.2) \quad e_{k+1} = e_k - \frac{e_k - e_{k-1}}{f(x_k) - f(x_{k-1})} f(x_k).$$



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

$$e_{k+1} = e_k \left[1 - \left(1 + \frac{1}{2} e_k \frac{f''(a)}{f'(a)} + O(e_k^2) \right) \left(1 - \frac{1}{2} (e_k + e_{k-1}) \frac{f''(a)}{f'(a)} + O(e_{k-1}^2) \right) \right],$$

tj.

$$(1.4.3) \quad e_{k+1} = e_k e_{k-1} \frac{f''(a)}{2f'(a)} (1 + O(e_{k-1})).$$

Da bismo odredili red konvergencije i faktor konvergencije stavimo

$$(1.4.4) \quad |e_{k+1}| = C_r |e_k|^r |1 + O(e_k)|.$$

Tada na osnovu (1.4.3) i (1.4.4) dobijamo

$$\begin{aligned} |e_{k+1}| &= C_r |e_k|^r |1 + O(e_k)| = C_r (C_r |e_{k-1}|^r)^r |1 + O(e_{k-1})| \\ &= C_r |e_{k-1}|^r |e_{k-1}| \cdot \left| \frac{f''(a)}{2f'(a)} \right| \cdot |1 + O(e_{k-1})|, \end{aligned}$$

odakle sleduje

$$r^2 - r - 1 = 0 \quad \text{i} \quad C_r = \left| \frac{f''(a)}{2f'(a)} \right|^{1/r}.$$

Red konvergencije r dobijamo kao pozitivno rešenje dobijene kvadratne jednačine, tj. $r = \frac{1}{2}(1 + \sqrt{5}) \approx 1.62$. 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

$$(1.4.5) \quad x = g(x)$$

sreće metod Weigsteina ([98]), kod koga se polazeći od x_0 gene-riše niz $\{x_k\}_{k \in \mathbb{N}}$ pomoću

$$(1.4.6) \quad x_1 = g(x_0),$$

$$x_{k+1} = g(x_k) - \frac{(g(x_k) - g(x_{k-1})) (g(x_k) - x_k)}{(g(x_k) - g(x_{k-1})) - (x_k - x_{k-1})} \quad (k = 1, 2, \dots).$$

U radu [96] 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

$$(1.4.7) \quad f(x) = g(x) - x = 0$$

smenom (1.4.7) u (1.4.6) dobijamo (1.4.1).

Metod sečice može se modifikovati tako da je

$$(1.4.8) \quad x_{k+1} = x_k - \frac{x_k - x_0}{f(x_k) - f(x_0)} f(x_k) \quad (k = 1, 2, \dots).$$

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

$$|\phi'(x)| \leq q \leq 1 \quad (f(x) \neq f(x_0)),$$

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

$$x_{k+1} = \frac{x_0 g(x_k) - x_k h(x_k)}{g(x_k) - h(x_k)} \quad (k = 1, 2, \dots),$$

$$\text{gde su } g(x) = \frac{f(x) - f(x_0)}{x - x_0} \quad \text{i} \quad h(x) = \frac{f'(x) f(x_0)}{f(x)}.$$

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

$$(1.4.9) \quad x_{k+1} = x_k - \frac{f(x_k)^2}{f(x_k + f(x_k)) - f(x_k)} \quad (k = 0, 1, \dots).$$

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

$$x \mapsto \phi(x) = x - \frac{f(x)^2}{f(x + f(x)) - f(x)}$$

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

$$\phi(x) = x - \frac{f(x)/f'(x)}{1 + \frac{1}{2}f''(\xi)f(x)/f'(x)}.$$

S obzirom da postoji okolina $U(a)$, tačke $x = a$, u kojoj je

$$\left| \frac{1}{2}f''(\xi)f(x)/f'(x) \right| < 1, \text{ imamo}$$

$$\phi(x) = x - \frac{f(x)}{f'(x)} \left(1 - \frac{1}{2} \cdot \frac{f''(\xi)}{f'(x)} f(x) + O(f(x)^2) \right) \quad (x \in U(a)).$$

$$\text{Kako je } \frac{f(x)}{f'(x)} = x - a - \frac{f''(a)}{2f'(a)}(x-a)^2 + O((x-a)^3) \quad (x \in U(a)),$$

iz poslednje jednakosti sleduje

$$\phi(x) - a \sim \frac{f''(a)}{2f'(a)}(f'(a)+1)(x-a)^2 \quad (x \rightarrow a),$$

što znači da je asimptotska konstanta greške $C_2 = \left| \frac{f''(a)}{2f'(a)}(f'(a)+1) \right|$.

5.1.5. Metod polovljenja intervala

Neka je na segmentu $[\alpha, \beta]$ izolovan prost koren $x = a$ jednačine

$$(1.5.1) \quad f(x) = 0,$$

gde $f \in C[\alpha, \beta]$. Metod polovljenja intervala za rešavanje jednačine (1.5.1) sastoji se u konstrukciji niza intervala $\{(x_k, y_k)\}_{k \in \mathbb{N}}$ takvog da je

$$y_{k+1} - x_{k+1} = \frac{1}{2} (y_k - x_k) \quad (k = 1, 2, \dots)$$

i pritom da je $\lim_{k \rightarrow +\infty} x_k = \lim_{k \rightarrow +\infty} y_k = a$.

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:

$$1^{\circ} \quad k := 0, \quad x_1 := \alpha, \quad y_1 := \beta;$$

$$2^{\circ} \quad k := k+1, \quad z_k := \frac{1}{2}(x_k + y_k);$$

3^o Ako je

$$\begin{array}{ll} f(z_k) f(x_k) < 0 & \text{uzeti } x_{k+1} := x_k, \quad y_{k+1} := z_k, \\ > 0 & x_{k+1} := z_k, \quad y_{k+1} := y_k, \\ = 0 & \text{Kraj izračunavanja } a := z_k; \end{array}$$

4^o Ako je

$$\begin{array}{ll} |y_{k+1} - x_{k+1}| \geq \epsilon & \text{preći na } 2^{\circ}, \\ < \epsilon & z_{k+1} := \frac{1}{2}(x_{k+1} + y_{k+1}) \\ & \text{Kraj izračunavanja } a := z_{k+1}. \end{array}$$

Primetimo da za grešku u aproksimaciji z_{k+1} važi ocena

$$|z_{k+1} - a| \leq \frac{1}{2^{k+1}} (\beta - \alpha).$$

5.1.6. Schröderov razvoj

Neka je funkcija $f: [\alpha, \beta] \rightarrow \mathbb{R}$ 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 takođe diferencijabilna. Naime,

$$F'(y) = \frac{dx}{dy} = \frac{1}{f'(x)} \quad (y = f(x)).$$

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.

Prepostavimo da je funkcija f diferencijabilna $(n+1)$ puta na $[\alpha, \beta]$, kao i to da je

$$(1.6.1) \quad F^{(k)}(y) = \frac{x_k}{(f')^{2k-1}} \quad (k = 1, \dots, n+1),$$

gde je x_k polinom po $f', f'', \dots, f^{(k)}$ i $f^{(i)} \equiv f^{(i)}(x)$ za $i = 1, \dots, n+1$.

Primetimo da je formula (1.6.1) tačna za $k = 1$ i $k = 2$, pri čemu su $x_1 = 1$ i $x_2 = -f''$.

Prepostavimo da je formula (1.6.1) tačna za neko $k \in \{1, \dots, n\}$. Kako je x_k polinom po $f', \dots, f^{(k)}$ i

$$x'_k = \frac{dx_k}{dx} = \frac{d}{dx} x_k(f', \dots, f^{(k)}) = \sum_{i=1}^k \frac{\partial x_k}{\partial f^{(i)}} f^{(i+1)}$$

polinom po $f', \dots, f^{(k+1)}$, to iz (1.6.1) sleduje

$$\begin{aligned} F^{(k+1)}(y) &= \frac{d}{dx} \left(\frac{x_k}{(f')^{2k-1}} \right) \frac{dy}{dx} \\ &= \frac{f' x'_k - (2k-1)x_k f''}{(f')^{2k+1}} = \frac{x_{k+1}}{(f')^{2k+1}}, \end{aligned}$$

gde je

$$(1.6.2) \quad x_{k+1} = f' x'_k - (2k-1)x_k f''$$

polinom po $f', \dots, f^{(k+1)}$.

Dakle, izvodi funkcije F su dati sa (1.6.1), pri čemu se

niz $\{x_k\}_{k \in \mathbb{N}}$ određuje 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 $h = -\frac{f(x)}{f'(x)}$ ($x \in U(a)$), 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

$$a - x = \sum_{k=1}^n \frac{1}{k!} x_k \left(\frac{f'}{f''}, \frac{f''}{f'}, \dots, \frac{f^{(k)}}{f'} \right) h^k + O(f(x)^{n+1}),$$

tj.

$$(1.6.3) \quad a - x = h - \frac{f''}{2f} h^2 + \frac{3f''^2 - f' f'''}{6f'^2} h^3$$

$$+ \frac{10f' f'' f''' - f'^2 f^{IV} - 15f''^3}{24f'^3} h^4 + \dots .$$

Primedba 1.6.1. Ako je funkcija analitička, u prethodnom razvoju može se uzeti da $n \rightarrow +\infty$.

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

$$(1.7.1) \quad f(x) = 0$$

na segmentu $[a, b]$ ima jedinstven prost koren $x = a$ kao i da je funkcija f dovoljan broj puta neprekidno-diferencijabilna na $[a, b]$.

1. Posmatrajmo Schröderov razvoj (1.6.3). Uzimajući koničan broj prvih članova na desnoj strani ovog razvoja možemo dobiti niz iterativnih formula.

Neka je

$$\phi_2(x) = x + h = x - \frac{f(x)}{f'(x)},$$

$$\phi_3(x) = \phi_2(x) - \frac{f''}{2f}, \quad h^2 = x - \frac{f(x)}{f'(x)} - \frac{f''(x)f(x)^2}{2f'(x)^3},$$

$$\begin{aligned} \phi_4(x) &= \phi_3(x) + \frac{3f''^2 - f'f'''}{6f'^2} h^3 \\ &= x - \frac{f(x)}{f'(x)} - \frac{f''(x)f(x)^2}{2f'(x)^3} - \frac{f(x)^3}{6f'(x)^4} \left(3\frac{f''(x)^2}{f'(x)} - f'''(x) \right), \end{aligned}$$

itd.

Primetimo da je ϕ_2 iterativna funkcija Newtonovog metoda.

Kako je, u prvoj aproksimaciji, h jednako $a-x$ ($x \rightarrow a$), na osnovu (1.6.3) imamo

$$\phi_m(x) - a = O(h^m) = O((x-a)^m) \quad (m = 2, 3, \dots),$$

kada $x \rightarrow a$, što znači da iterativni proces

$$(1.7.2) \quad x_{k+1} = \phi_m(x_k) \quad (k = 0, 1, \dots),$$

primenjen na određivanje korena jednačine (1.7.1), ima red konvergencije najmanje m . Takođe, na osnovu Schröderovog razvoja možemo dobiti i vrednosti limesa

$$L_m = \lim_{k \rightarrow \infty} \frac{x_{k+1} - a}{(x_k - a)^m} \quad (m = 2, 3, \dots).$$

Naime, imamo

$$L_2 = \frac{f''(a)}{2f'(a)},$$

$$L_3 = \frac{1}{2} \left(\frac{f''(a)}{f'(a)} \right)^2 - \frac{1}{6} \cdot \frac{f'''(a)}{f'(a)},$$

$$L_4 = \frac{5}{8} \left(\frac{f''(a)}{f'(a)} \right)^3 + \frac{1}{24} \cdot \frac{f^{IV}(a)}{f'(a)} - \frac{5}{12} \cdot \frac{f''(a)f'''(a)}{f'(a)^2}.$$

Formule (1.7.2) se često nazivaju Čebiševljeve iterativne formule.

Korišćenjem Hermiteove interpolacione formule (videti odeljak 6.2.8) za funkciju f u tačkama $x=x_{k-1}$ i $x=x_k$ može se dobiti iterativna formula (videti [86])

$$(1.7.3) \quad x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} - \frac{f(x_k)^2}{2f'(x_k)} \bar{f}''(x_k) \quad (k = 1, 2, \dots),$$

gde je $\bar{f}''(x_k) = -\frac{6}{\varepsilon_k^2}(f(x_k) - f(x_{k-1})) + \frac{2}{\varepsilon_k^2}(2f'(x_k) + f'(x_{k-1}))$ i

$\varepsilon_k = x_k - x_{k-1}$. Red konvergencije ovog procesa je $r = 1 + \sqrt{3} \approx 2.73$. Primetimo da iterativna funkcija ovog procesa predstavlja jednu modifikaciju Čebiševljeve funkcije ϕ_3 .

U radu [52] Milovanović i Petković su razmatrali modifikaciju funkcije ϕ_3 korišćenjem aproksimacije

$$f''(x_k) \cong \frac{f'(x_k + \varepsilon_k) - f'(x_k)}{\varepsilon_k},$$

pri čemu $\varepsilon_k \rightarrow 0$, kada $k \rightarrow +\infty$. Odgovarajući iterativni proces je tada

$$(1.7.4) \quad x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} - \frac{f(x_k)^2}{2f'(x_k)^3} \cdot \frac{f'(x_k + \varepsilon_k) - f'(x_k)}{\varepsilon_k}.$$

Teorema 1.7.1. Neka $f \in C^3[\alpha, \beta]$ i neka je $\varepsilon_k = x_{k-1} - x_k$. Tada iterativni proces (1.7.4) ima red konvergencije $r = 1 + \sqrt{2}$, tj. važi

$$|x_{k+1} - a| \sim C_r |x_k - a|^r \quad (k \rightarrow +\infty),$$

gde je $C_r = |f'''(a)/(4f'(a))|^{1/\sqrt{2}}$.

Dokaz. Za $\varepsilon_k = x_{k-1} - x_k$, formula (1.7.4) postaje

$$(1.7.5) \quad x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} - \frac{f(x_k)^2}{2f'(x_k)^3} \cdot \frac{f'(x_k) - f'(x_{k-1})}{x_k - x_{k-1}}.$$

Na osnovu Schröderovog razvoja (1.6.3), za x_k koje je dovoljno blisko korenu $x=a$, imamo

$$a - x_k = h_k - \frac{h_k^3 s_k}{2} + \frac{h_k^3}{6} (3s_k^2 - r_k) + O(h_k^4),$$

gde smo uveli notaciju $h_k = -f(x_k)/f'(x_k)$, $s_k = f''(x_k)/f'(x_k)$, $r_k = f'''(x_k)/f'(x_k)$. S druge strane, primenom Taylorove formule imamo

$$f'(x_{k-1}) = f'(x_k) + \varepsilon_k f''(x_k) + \frac{1}{2} \varepsilon_k^2 f'''(\zeta_k),$$

gde je $\zeta_k = x_k + \theta \varepsilon_k$ ($0 < \theta < 1$).

$$\text{Kako je } \varepsilon_k = x_{k-1} - x_k = \frac{f(x_{k-1})}{f'(x_{k-1})} + O(h_{k-1}^2) = -h_{k-1} + O(h_{k-1}^2),$$

na osnovu prethodnog imamo

$$x_{k+1} - a = \frac{h_k^2}{2f'(x_k)} \left[f''(x_k) + \frac{f'(x_k) - f'(x_{k-1})}{x_{k-1} - x_k} + O(h_k) \right],$$

tj.

$$\begin{aligned}x_{k+1} - a &= \frac{h_k^2}{4} \left[\frac{f'''(\zeta_k)}{f'(x_k)} e_k + O(h_k) \right] \\&= \frac{h_k^2}{4} \left[\frac{f'''(\zeta_k)}{f'(x_k)} (-h_{k-1} + O(h_{k-1}^2)) + O(h_k) \right].\end{aligned}$$

Neka je dalje $e_k = x_k - a$. Kako je $e_k = -h_k + O(h_k^2)$, zaključujemo da je

$$h_k = -e_k + O(e_k^2) \quad \text{i} \quad h_{k-1} = -e_{k-1} + O(e_{k-1}^2).$$

Sada imamo

$$\begin{aligned}e_{k+1} &= \frac{f'''(\zeta_k)}{4f'(x_k)} (e_k + O(e_k^2))^2 (e_{k-1} + O(e_{k-1}^2) + O(e_k)) \\&= \frac{f'''(\zeta_k)}{4f'(x_k)} (e_k^2 e_{k-1} + O(e_k^2 e_{k-1}^2)) \sim \frac{f'''(\zeta_k)}{4f'(x_k)} e_k^2 e_{k-1}.\end{aligned}$$

Kada $k \rightarrow +\infty$, $x_k \rightarrow a$ pri čemu je

$$|e_{k+1}| \sim \left| \frac{f'''(a)}{4f'(a)} \right| |e_k|^2 |e_{k-1}|.$$

Iz poslednje asimptotske relacije, slično kao kod metoda sečice (odeljak 5.1.4), nalazimo red konvergencije iz jednačine $r^2 - 2r - 1 = 0$. Dakle, $r = 1 + \sqrt{2}$. Asimptotska kon-

stanta greške je $C_r = \left| \frac{f'''(a)}{4f'(a)} \right|^{1/\sqrt{2}}$.

Primedba 1.7.1. Iterativni proces (1.7.3) ima veći red konvergencije nego proces (1.7.5), ali je s druge strane iterativna formula procesa (1.7.5) jednostavnija i zahteva manji broj računskih operacija.

Slično prethodnoj teoremi, u pomenutom radu [52] dokazan je i sledeći rezultat:

Teorema 1.7.2. Neka $f \in C^3[\alpha, \beta]$ i neka je $e_k = f(x_k)$. Tada iterativni proces (1.7.4) ima red konvergencije tri, tj. važi

$$|x_{k+1} - a| \sim C_3 |x_k - a|^3 \quad (k \rightarrow +\infty),$$

gde je

$$C_3 = \frac{1}{12f'(a)^2} |3f'(a)^2 f'''(a) + 2f'(a)f''(a) - 6f''(a)^2|.$$

Uvodjenjem aproksimacija za $f'(x)$ i $f''(x)$ pomoću

$$(1.7.6) \quad \begin{cases} f'(x) \cong \bar{f}'(x) = \frac{f(x+f(x))-f(x-f(x))}{2f(x)} \\ f''(x) \cong \bar{f}''(x) = \frac{f(x+f(x))-2f(x)+f(x-f(x))}{f(x)^2} \end{cases}$$

u radu [51] dobijena je iterativna funkcija

$$\phi_3^*(x) = x - \frac{f(x)}{\bar{f}'(x)} - \frac{f(x)^2 \bar{f}''(x)}{2\bar{f}'(x)^3}$$

i dokazan sledeći rezultat:

Teorema 1.7.3. Neka $f \in C^3[a, b]$. Iterativni proces $x_{k+1} = \phi_3^*(x_k)$ ($k = 0, 1, \dots$) ima red konvergencije $r=3$, tj. važi

$$L_3^* = \lim_{k \rightarrow +\infty} \frac{x_{k+1} - a}{(x_k - a)^3} = L_3 + \frac{1}{6} f'(a) f'''(a),$$

gde je L_3 dato ranije.

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

$$x_{k+1} = x_k - \frac{x_k - \phi(x_k)}{1 - \frac{1}{2}\phi'(x_k)} \quad (k = 0, 1, \dots),$$

tj.

$$(1.7.7) \quad x_{k+1} = x_k - \frac{2f(x_k)f'(x_k)}{2f'(x_k)^2 - f(x_k)f''(x_k)} \quad (k = 0, 1, \dots).$$

Metod (1.7.7), na osnovu pomenute teoreme, ima red konvergencije

najmanje tri. Ovaj metod je u literaturi poznat kao Salehov metod tangentnih hiperbola ([87]) ili kao Halleyev metod.

Za metod (1.7.7) lako se može dobiti asymptotska formula

$$\frac{x_{k+1} - a}{(x_k - a)^3} \sim \left(\frac{f''(a)}{2f'(a)} \right)^2 - \frac{f'''(a)}{6f'(a)} \quad (k \rightarrow +\infty)$$

Primedba 1.7.2. U radu [38] data je modifikacija metoda (1.7.7) korišćenjem aproksimacije $f''(x_k) \approx (f'(x_k) - f'(x_{k-1})) / (x_k - x_{k-1})$. Može se dokazati da je

$$x_{k+1} - a \sim \frac{f'''(a)}{4f'(a)} (x_k - a)^2 (x_{k-1} - a) \quad (k \rightarrow +\infty),$$

odakle izlazi da je red konvergencije ovako modifikovanog metoda $r = 1 + \sqrt{2} \approx 2.414$.

Korišćenjem aproksimacija (1.7.6), u radu [51] Milovanović i Kovačević su razmatrali iterativnu funkciju

$$\psi_3^*(x) = x - \frac{2f(x)\bar{f}'(x)}{2\bar{f}'(x) - f(x)\bar{f}''(x)},$$

koja predstavlja jednu modifikaciju Halleyevog metoda (1.7.7). Konvergencija procesa je kubna.

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 $\{\psi_i\}_{i=0,1,\dots,n}$ dovoljan broj puta diferencijabilnih na $[\alpha, \beta]$.

U radu [95] je predložena konstrukcija iterativnog procesa za rešavanje jednačine (1.7.1) u obliku

$$(1.7.8) \quad x_{k+1} = \phi_n(x_k) \quad (k = 0, 1, \dots),$$

gde su

$$\phi_n(x) = x - \frac{D_n(x)}{\Delta_n(x)},$$

$$\Delta_n(x) = \begin{vmatrix} (f\psi_0)' & (f\psi_1)' & \dots & (f\psi_n)' \\ (f\psi_0)'' & (f\psi_1)'' & & (f\psi_n)'' \\ \vdots & & & \\ (f\psi_0)^{(n+1)} & (f\psi_1)^{(n+1)} & & (f\psi_n)^{(n+1)} \end{vmatrix}$$

$$D_n(x) = \begin{vmatrix} f\psi_0 & f\psi_1 & \dots & f\psi_n \\ (f\psi_0)'' & (f\psi_1)'' & & (f\psi_n)'' \\ \vdots & & & \\ (f\psi_0)^{(n+1)} & (f\psi_1)^{(n+1)} & & (f\psi_n)^{(n+1)} \end{vmatrix}.$$

Može se pokazati da je

$$\phi_n(a) = a, \phi'_n(a) = \phi''_n(a) = \dots = \phi_n^{(n+1)}(a) = 0, \phi_n^{(n+2)}(a) \neq 0,$$

što znači da iterativni proces (1.7.8) ima red konvergencije $n+2$.

Navećemo nekoliko specijalnih slučajeva formule (1.7.8).

1° Ako je $n=0$ i $\psi_0(x) \equiv \psi(x)$, (1.7.8) se svodi na uopšten Newtonov metod (1.2.8).

2° Ako je $n=1$, $\psi_0(x) \equiv 1$, $\psi_1(x) \equiv x$, (1.7.8) se svodi na (1.7.7).

3° Ako je $n=2$, $\psi_0(x) \equiv 1$, $\psi_1(x) \equiv x$, $\psi_2(x) \equiv x^2$, iz (1.7.8) sleduje formula Domorjada-Lika ([44]), čija je iterativna funkcija data sa

$$\phi_2(x) = x - \begin{vmatrix} f'(x) & f(x) \\ \frac{1}{2}f''(x) & f'(x) \end{vmatrix} \cdot \begin{vmatrix} f'(x) & f(x) & 0 \\ \frac{1}{2}f''(x) & f'(x) & f(x) \\ \frac{1}{6}f'''(x) & \frac{1}{2}f''(x) & f'(x) \end{vmatrix}^{-1}.$$

5. Posmatrajmo formulu (1.2.9), gde je p parametar. Sa $U(a)$ označimo okolinu korena $x = a$ jednačine (1.7.1).

Teorema 1.7.4. Neka je funkcija f četiri puta neprekidno-diferencijabilna u $U(a)$ i neka niz $\{x_k\}_{k \in \mathbb{N}_0}$ ($x_0 \in U(a)$), definisan pomoću (1.2.9), konvergira ka a . Tada, pri $k \rightarrow +\infty$, važi asymptotska formula

$$x_{k+1} - a \sim \left[\frac{p}{x_k} + \frac{s_k}{2} \right] h_k^2 + \left[\left(\frac{p}{x_k} \right) - \frac{1}{6} (3s_k^2 - r_k) \right] h_k^3$$

$$+ \left[\left(\frac{p}{x_k} \right)^3 - \frac{1}{24} (10r_k s_k - t_k - 15s_k^3) \right] h_k^4,$$

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),

$$x_{k+1} - x_k = \frac{x_k f(x_k)}{x_k f'(x_k) + p f(x_k)} = \frac{h_k}{1 - \frac{p}{x_k} h_k}$$

i kako je $|\frac{p}{x_k} h_k| < 1$ ($x_k \in U(a)$), nalazimo

$$x_{k+1} - x_k = h_k + \frac{p}{x_k} h_k^2 + \left(\frac{p}{x_k} \right)^2 h_k^3 + \left(\frac{p}{x_k} \right)^3 h_k^4 + o(h_k^5).$$

S druge strane, na osnovu (1.6.3) imamo

$$a - x_k = h_k - \frac{1}{2} s_k h_k^2 + \frac{1}{6} (3s_k^2 - r_k) h_k^3$$

$$+ \frac{1}{24} (10r_k s_k - t_k - 15s_k^3) h_k^4 + o(h_k^5).$$

Na osnovu poslednjih jednakosti dobijamo tvrdjenje teoreme.

Posmatrajmo sada formulu za $p = p_1$ i $p = p_2$, tj.

$$(1.7.9) \quad x_{k+1}^{(i)} = x_k \left(1 - \frac{f(x_k)}{x_k f'(x_k) + p_i f(x_k)} \right) \quad (i = 1, 2)$$

i stavimo

$$(1.7.10) \quad x_{k+1} = \frac{1}{2}(x_{k+1}^{(1)} + x_{k+1}^{(2)}).$$

Tada, na osnovu teoreme 1.7.4, imamo

$$(1.7.11) \quad x_{k+1} - a \sim \frac{1}{2} \left[\frac{p_1 + p_2}{x_k} + s_k \right] h_k^2 + \frac{1}{2} \left[\frac{p_1^2 + p_2^2}{x_k^2} - \frac{1}{3}(3s_k^2 - r_k) \right] h_k^3 \\ + \frac{1}{2} \left[\frac{p_1^3 + p_2^3}{x_k^3} - \frac{1}{12}(10r_k s_k - t_k - 15s_k^3) \right] h_k^4.$$

Ako u formulama (1.7.9) izaberemo parametre p_1 i p_2 , takve da je

$$(1.7.12) \quad p_1 + p_2 = -s_k x_k, \quad p_1^2 + p_2^2 = \frac{1}{3}(3s_k^2 - r_k) x_k^2,$$

formula (1.7.11) postaje

$$x_{k+1} - a \sim C_k h_k^4 \quad (k \rightarrow +\infty),$$

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.

$$(1.7.13) \quad p^2 + x_k \frac{f''(x_k)}{f'(x_k)} p + \frac{1}{6} x_k^2 \frac{f'''(x_k)}{f'(x_k)} = 0.$$

Na osnovu prethodnog može se formulisati i dokazati slede-

ča teorema (videti [15]).

Teorema 1.7.5. 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

$$x_{k+1} - a \sim C(x_k - a)^4 \quad (k \rightarrow +\infty),$$

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

$$x_{k+1} = \phi(x_k) \quad (k = 0, 1, \dots),$$

gde je

$$\phi(x) = x - 3f(x) \frac{2f'(x)^2 - f(x)f''(x)}{6f'(x)(f'(x)^2 - f(x)f''(x)) + f(x)^2 f'''(x)}.$$

Primetimo da je ova formula ekvivalentna sa formulom Domorjada-Lika ([44]).

Iterativni proces definisan formulama (1.7.9) i (1.7.10) pripada tzv. klasi metoda dvostranog približavanja. Naime, kod ovakvih metoda jedna od aproksimacija $x_{k+1}^{(i)}$ ($i = 1, 2$) je manja, a druga veća od korena $x = a$. U opštem slučaju, ako je

$$(1.7.14) \quad x_{k+1}^{(i)} = \phi_i(x_k) \quad (i = 1, 2)$$

i

$$(1.7.15) \quad x_{k+1} = \frac{1}{2} (x_{k+1}^{(1)} + x_{k+1}^{(2)}) \quad (k = 0, 1, \dots)$$

i pritom

$$(\phi_1(x_k) - a)(\phi_2(x_k) - a) < 0 \quad (k = 0, 1, \dots),$$

imamo metod dvostranog približavanja.

Jedan bolji izbor funkcija ϕ_i ($i = 1, 2$), nego što je (1.7.9), daje takođe metod četvrtog reda, ali sa manjim brojem numeričkih operacija ([66], [74]). Naime, ako uzmemo

$$(1.7.16) \quad \phi_1(x_k) = x_k + \frac{h_k}{1+h_k s_k}$$

i

$$(1.7.17) \quad \phi_2(x_k) = x_k + h_k \left(1 - \frac{1}{3} h_k^2 r_k\right),$$

gde su veličine h_k , s_k , r_k definisane u teoremi 1.7.4, dobijamo metod dvostranog približavanja za koji važi sledeći rezultat ([66, str. 82-83]):

Teorema 1.7.6. Neka je funkcija f četiri puta neprekidno-diferencijabilna u $U(a)$. Tada iterativni proces, definisan pomoću (1.7.14)-(1.7.17), ima red konvergencije četiri, tj. važi

$$x_{k+1} - a \sim C(x_k - a)^4 \quad (k \rightarrow +\infty),$$

gde je

$$C = \frac{10f'(a)f''(a)f'''(a) - f'(a)^2 f^{IV}(a) - 27f''(a)^3}{24f'(a)^3}$$

6. Iterativne formule trećeg reda za određivanje višestrukih korena jednačina razmatrane su u [13], [30], [50].

5.1.8. Newtonov metod u intervalnoj aritmetici

Numerički algoritmi za nalaženje korena jednačina oblika $f(x) = 0$, koji koriste iterativne metode, u stvarnosti nikad ne daju tačne vrednosti korena posmatrane jednačine. Ovo je rezultat nužnih ograničenja koja se javljaju pri korišćenju računskih mašina. Naime, pri realizaciji iterativnog procesa koji

* Docent dr M. S. Petković sastavio je tekst ovog odeljka.

teorijski konvergira u beskonačno mnogo koraka, potrebno je uvesti ograničenje za broj iteracija. Samim tim dolazi do greške te je jedino moguće odrediti aproksimativnu vrednost tražene veličine. Osim toga, greška se povećava zbog zaokrugljivanja prilikom izvodjenja računskih operacija. Dalje, poznato je da se u računskim mašinama jedna neprekidna funkcija $x \mapsto f(x)$ može predstaviti samo diskretno jer se argument x izražava pomocu racionalnog broja u obliku niza bitova konačne dužine. Grafik funkcije f je samo više ili manje dobro aproksimiran pomocu skupa tačaka unutar izvesnog pojasa. Zbog toga je praktično nemoguće odrediti tačno rešenje jednačine $f(x) = 0$ bez detaljne informacije o nastaloj greški zaokrugljivanja.

Razvoj intervalne aritmetike doveo je do novog pristupa analizi grešaka i do primene metoda za lokalizaciju nula funkcije. Ovo je omogućilo rešavanje napred iznetih problema razvijanjem metoda za određivanje intervala koji sadrže nule funkcije. Upotrebom mašinske intervalne aritmetike (videti, na primer [5], [11], [102]) može se postići da se greške zaoštruju automatski uzimajući u obzir pri numeričkim izračunavanjima.

U ovom odeljku razmatraćemo intervalnu verziju Newtonovog iterativnog procesa (1.2.3), tj.

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \quad (k = 0, 1, \dots).$$

Intervalni Newtonov metod predložio je R. E. Moore 1966. godine ([57]). Kasnije se pojavilo više modifikacija Mooreovog rezultata (videti, na primer, [2], [12], [28], [29], [37], [42], [67], [68]) sa ciljem da se otklone nedostaci ovog metoda i poboljšaju njegove konvergentne osobine.

Pre nego što izložimo osnovnu varijantu Newtonovog intervalnog metoda daćemo kratak pregled osnovnih osobina intervalne aritmetike.

Pod realnim intervalom $[a, b]$ podrazumevamo zatvoren ograničen skup realnih brojeva x sa osobinom $a \leq x \leq b$, tj.

$$[a, b] := \{x \mid a \leq x \leq b, a, b \in \mathbb{R}\}.$$

Skup svih realnih intervala označavamo sa $I(\mathbb{R})$.

Dužina, sredina i apsolutna vrednost intervala definišu se redom pomoću formula

$$d([a,b]) := b - a, \quad m([a,b]) := \frac{a+b}{2},$$

$$| [a,b] | := \max_{x \in [a,b]} |x| = \max(|a|, |b|).$$

Presek dva intervala $A = [a_1, a_2]$ i $B = [b_1, b_2]$ je prazan, $A \cap B = \emptyset$, ako je $a_1 > b_2$ ili $b_1 > a_2$. U suprotnom, presek intervala A i B je opet interval dat sa

$$A \cap B := [\max(a_1, b_1), \min(a_2, b_2)].$$

Očigledno je

$$(1.8.1) \quad A \cap B \subset A \quad \text{i} \quad A \cap B \subset B.$$

Od posebnog interesa su operacije sa intervalima koje odgovaraju operacijama sabiranja, oduzimanja, množenja i deljenja u skupu realnih brojeva. Neka je $*$ jedna od ovih operacija, tj. $* \in \{+, -, \cdot, :\}$. U skupu $I(\mathbb{R})$ binarna operacija $*$ definiše se pomoću

$$A * B := \{a * b \mid a \in A, b \in B\} \quad (A, B \in I(\mathbb{R})).$$

Na osnovu ovog za intervale $[a_1, a_2]$ i $[b_1, b_2]$ imamo (podrazumevajući da $0 \notin B$ ako je $*$ deljenje)

$$[a_1, a_2] + [b_1, b_2] = [a_1 + b_1, a_2 + b_2],$$

$$[a_1, a_2] - [b_1, b_2] = [a_1 - b_2, a_2 - b_1],$$

$$\begin{aligned} [a_1, a_2] \cdot [b_1, b_2] &= [\min(a_1 b_1, a_1 b_2, a_2 b_1, a_2 b_2), \\ &\quad \max(a_1 b_1, a_1 b_2, a_2 b_1, a_2 b_2)], \end{aligned}$$

$$[a_1, a_2] : [b_1, b_2] = [a_1, a_2] \cdot \left[\frac{1}{b_2}, \frac{1}{b_1} \right] \quad (0 \notin [b_1, b_2]).$$

Ovakvu intervalnu aritmetiku uveli su T. Sunaga ([90]) i R.E. Moore ([57]), i ona se najčešće naziva realnom ili običnom intervalnom aritmetikom.

Neka je funkcija f definisana na uniji svih realnih intervala iz $D \subset I(\mathbb{R})$. Tačna intervalna funkcija f definiše se kao

$$\bar{F}(X) = \{f(x) \mid x \in X, X \in D \subset I(\mathbb{R})\}.$$

U praksi je teško, ili čak nemoguće, odrediti tačnu oblast $\bar{F}(X)$. Zbog toga se uvodi tzv. intervalno proširenje F takvo da je

$$F(X) \supseteq \bar{F}(X) \quad \text{za svako } X \in D,$$

$$F(x) = f(x) \quad (\text{realna restrikcija}).$$

Neka je F intervalna funkcija definisana na skupu intervala $D \subset I(\mathbb{R})$. Ako važi implikacija

$$A \subseteq B \Rightarrow F(A) \subseteq F(B) \quad (A, B \in D),$$

kažemo da je funkcija F inkluzivno izotona. U specijalnom slučaju, ako je $y=f(x)$ i $x \in [a,b]$, tada $y \in f([a,b])$.

Neka je $x \mapsto f(x)$ realna funkcija realne promenljive x , diferencijabilna na $[a,b]$ i neka je ξ nula funkcije f koja pripada intervalu $[a,b]$. Na osnovu teoreme o srednjoj vrednosti za $x, \xi \in [a,b]$ imamo

$$f(\xi) = f(x) + f'(x+t(\xi-x))(\xi - x) \quad (t \in [0,1]),$$

odakle je

$$\xi = x - \frac{f(x)}{f'(x+t(\xi-x))}.$$

Na osnovu osobine inkluzivne izotonosti iz poslednje jednakosti sleduje

$$\xi \in x - \frac{f(x)}{f'([a,b])}.$$

Ovo sugerira sledeći intervalni metod:

Neka je F' intervalno proširenje funkcije f' ; tada je $f'(x) \in F'(X)$ i $F'([x,x]) = f'(x)$ ($x \in X$). Neka je $m(X)$ sredina intervala X i neka $X \subset [a,b]$. R.E. Moore je u [57, pogl. 7] uveo intervalnu funkciju N (Newtonov operator) pomoću

$$(1.8.2) \quad N(X) = m(X) - \frac{f(m(X))}{F'(X)}.$$

Polazeći od intervala X_0 koji sadrži realnu nulu funkcije f , Moore je definisao intervalni niz $\{X_k\}$ pomoću

$$(1.8.3) \quad x_{k+1} = N(x_k) \cap x_k \quad (k = 0, 1, \dots).$$

Formule (1.8.2) i (1.8.3) definišu intervalnu verziju Newtonovog metoda. Pod pogodnim početnim uslovima niz dužina intervala $\{d(x_k)\}$ konvergira ka nuli kvadratno, pri čemu je

$$x_{k+1} \subset x_k \subset \dots \subset x_1 \subset x_0 \quad (\text{na osnovu (1.8.1)}),$$

$$\xi \in x_0 \Rightarrow \xi \in x_k \quad (k = 1, 2, \dots).$$

Dakle, Newtonov intervalni metod omogućuje određivanje realnog korena jednačine sa automatskom granicom greške (ako je $m(x_k)$ jedna aproksimacija za nulu ξ , tada je $|\xi - m(x_k)| \leq d(x_k)/2$).

Red konvergencije običnog i intervalnog Newtonovog metoda jednak je dva za proste korene. Osim mogućnosti određivanja granica greške pri traženju korena jednačine, Newtonov intervalni metod ima prednost nad običnim i u pogledu konvergentnih osobina. Ovo ćemo ilustrovati sledećim primerima.

Primer 1.8.1. Približna vrednost realne nule polinoma $f(x) = x^3 - 3x - 2.001$ je $\xi \approx 2.0001111$. Primenjujući Newtonov intervalni metod sa početnim intervalom $x_0 = [-3, 3]$, posle šeste iteracije nalazimo interval $x_6 = [2.0001108, 2.0001114]$ koji sadrži tačnu nulu. Sastavno drugačije ponašanje pokazuje običan Newtonov metod. Ukoliko početna vrednost x_0 nije veća od 1, ovaj metod generiše vrlo nestabilan niz. Na primer, startujući sa $x_0 = 0$, dobijamo

$$\begin{aligned} x_1 &= -0.667, \quad x_2 = -0.84519, \quad x_3 = -0.92593, \quad x_4 = -0.96577, \\ x_5 &= -0.987, \quad x_6 = -1.00789, \quad x_7 = -0.98292, \quad x_8 = -1.00133, \dots \end{aligned}$$

koji oscilira oko tačke lokalnog ekstremuma funkcije f , u kojoj f ima vrlo malu vrednost. Naravno, ovde x_0 nije izabran u dovoljno maloj okolini nule ξ .

Primer 1.8.2. Funkcija $f(x) = x^{1/3} = (\operatorname{sgn} x) |x|^{1/3}$ ima jedinstvenu realnu nulu $\xi = 0$. Može se pokazati (videti [58,str.68]) da ako intervalni metod (1.8.3) startuje sa intervalom x_0 koji sadrži tačnu nulu $\xi = 0$, za svaku iteraciju će važiti $\xi \in x_k$ i $d(x_k) \leq 2^{-k} d(x_0)$, tj. metod (1.8.3) konvergira ka ξ . Ovaj rezultat je u potpunoj suprotnosti sa ponašanjem običnog Newtonovog metoda. Za funkciju $f(x) = x^{1/3}$ imamo

$$x_{k+1} = x_k - f(x_k)/f'(x_k) = x_k - 3x_k = -2x_k.$$

Dakle, metod je divergentan u svim slučajevima (nezavisno od x_0) izuzev ako početna vrednost nije nula funkcije. Neprimenljivost Newtonovog metoda u ovom slučaju je očigledna zbog nediferencijabilnosti funkcije f u $x = \xi = 0$.

U radu [29] E. Hansen je dao jedno poboljšanje konvergentnih osobina intervalnog metoda (1.8.3) zamjenjujući realnim brojevima izvesne intervale koji se javljaju u iterativnoj formuli.

Neka je f funkcija u kojoj se x javlja bar p puta ($p \geq 2$). Ako se x zameni sa x_1, x_2, \dots, x_p na jednom ili više mesta, dobija se funkcija oblika $g(x_1, x_2, \dots, x_p)$, pri čemu je očigledno $g(x, x, \dots, x) = f(x)$. Neka je X interval koji sadrži realnu nulu $x = \xi$. Koristeći razvoj funkcije g u Taylorov red (videti [27]), Hansen je u radu [29] definisao Newtonov operator N_p pomoću

$$(1.8.4) \quad N_p(X) = x - \frac{f(x)}{g'(X)} \quad (x \in X),$$

gde je

$$(1.8.5) \quad g'(X) = \frac{\partial}{\partial x_1} g(X, x, \dots, x) + \frac{\partial}{\partial x_2} g(X, x, \dots, x) + \dots + \frac{\partial}{\partial x_p} g(X, x, \dots, x).$$

Za Newtonov operator N_p važi $\xi \in N_p(X)$ tako da se iterativni postupak koji koristi (1.8.4) može konstruisati na sledeći način:

Polazeći od početnog intervala X_0 , koji sadrži nulu funkcije f , formira se intervalni niz $\{X_k\}$ po formuli

$$(1.8.6) \quad X_{k+1} = N_p(X_k) \cap X_k \quad (k = 0, 1, \dots).$$

Izloženi metod odgovara Mooreovom obliku (1.8.3) pri čemu $g'(X)$ predstavlja jedan razvijeni oblik intervalne funkcije $f'(X)$. Prednost intervalnog operatora N_p nad N je u tome što se mnogi argumenti javljaju kao realne promenljive umesto intervala. U opštem slučaju ovo smanjuje broj operacija sa intervalima i kao rezultat toga dobijaju se uži intervali. Osim toga, Hansenova modifikacija je primenljiva pod slabijim uslovima.

Primer 1.8.3. Mooreovu i Hansenovu verziju Newtonowog intervalnog metoda ilustrovaćemo na primeru funkcije $f(x) = x^4 - 1$. Neka je $X = [0.5, 3.5]$ početni interval koji sadrži tačnu nulu $\xi = 1$. S obzirom da je $m(X) = 2$, $f'(x) = 4x^3$ i $F'(X) = 4X^3$, za Mooreov metod dobijamo iz (1.8.2)

$$N(X) = 2 - \frac{f(2)}{4X^3} = [-28, 1.913].$$

Napišimo sada f u obliku $g(x_1, x_2, x_3, x_4) = x_1 x_2 x_3 x_4 - 1$.

Koristeći (1.8.4) i (1.8.5) nalazimo

$$N_4(X) = 2 - \frac{f(2)}{x^3 + x^2 X + x X + x^3} = [0.588, 1.832].$$

Dužina intervala $N(X)$ je 29.91, dok je dužina $d(N_4(X))$ samo 1.244, što ukazuje na prednost Hansenove modifikacije.

Nedostatak Mooreovog metoda leži u zahtevu da $F'(X)$ ni u jednoj iteraciji ne sme biti nula-interval. Uslov $0 \notin F'(X)$ je dosta strog jer i u slučaju kada je $|f'(x)| > 0$ za $x \in X$, interval $F'(X)$ može sadržati 0 (zbog $F'(X) \supseteq \{f'(x) | x \in X\}$).

Ovaj nedostatak se može otkloniti korišćenjem proširene intervalne aritmetike (videti [32]). Neka je $F'(X_k) = [a_k, b_k]$. Ako $0 \in F'(X_k)$ (tj. $a_k b_k \leq 0$), definišimo

$$(1.8.7) \quad \frac{1}{F'(X_k)} = \begin{cases} [1/b_k, +\infty] & (a_k = 0), \\ [-\infty, 1/a_k] & (b_k = 0), \\ [-\infty, 1/a_k] \cup [1/b_k, +\infty] & (a_k b_k < 0). \end{cases}$$

Ako je $f(x_k) = 0$, tada je x_k nula funkcije f . Isključujući ovaj slučaj, imamo dve mogućnosti:

a) Ako je $f(x_k) > 0$, tada iz (1.8.2) i (1.8.7) dobijamo

$$(1.8.8) \quad N_0(X_k) = \begin{cases} [-\infty, q_k] & (a_k = 0), \\ [p_k, +\infty] & (b_k = 0), \\ [-\infty, q_k] \cup [p_k, +\infty] & (a_k b_k < 0), \end{cases}$$

gde je $p_k = x_k - f(x_k)/a_k$, $q_k = x_k - f(x_k)/b_k$.

b) Ako je $f(x_k) < 0$, tada je

$$(1.8.9) \quad N_0(x_k) = \begin{cases} [q_k, +\infty] & (a_k = 0), \\ [-\infty, p_k] & (b_k = 0), \\ [-\infty, p_k] \cup [q_k, +\infty] & (a_k b_k < 0). \end{cases}$$

U (1.8.8) i (1.8.9) operator N_0 primenjuje se na interval koji sadrži nulu. Novi interval x_{k+1} se dobija kao

$$(1.8.10) \quad x_{k+1} = N_0(x_k) \cap x_k \quad (k = 0, 1, \dots).$$

Dakle, ako $0 \in F'(x_k)$, interval x_{k+1} biće ograničen skup. U radu [28] pokazano je da metod (1.8.10) globalno konvergira u smislu definicija (1.8.7), (1.8.8) i (1.8.9). Konvergencija je kvadratna za proste nule, a linearna za višestruke nule.

Razdvajanje i ograničavanje proste i višestruke nule biće prikazano u sledećem primeru.

Primer 1.8.4. Posmatrajmo polinom $f(x) = (x-1)^2(x+1) = x^3 - x^2 - x + 1$ koji ima prostu nulu $\xi_1 = -1$ i dvostruku nulu $\xi_2 = 1$. Funkciju f možemo predstaviti u obliku

$$g(x_1, x_2, x_3) = x_1 x_2 x_3 - x_1 x_2 - x_1 + 1,$$

pri čemu je $g(x, x, x) = f(x)$. Koristeći (1.8.4), dobijamo

$$g'(X) = x^2 - x - 1 + X(X + x - 1) \quad (x = m(X)).$$

Uzmimo za početni interval $x_0 = [-2, 2.5]$ koji sadrži i prostu i dvostruku nulu. Tada je $x_0 = m(x_0) = 0.25$. Korišćenjem izraza za $g'(X)$ nalazimo $g'(x_0) = [-8.07, 4.32]$. Kako $0 \notin g'(x_0)$, na osnovu (1.8.9) dobijamo dva neograničena intervala

$$N_0^{(1)}(x_0) = [-\infty, 0.087], \quad N_0^{(2)}(x_0) = [0.337, +\infty].$$

Na osnovu (1.8.10) imamo preseke

$$x_1^{(1)} = [-2, 0.087], \quad x_1^{(2)} = [0.337, 2.5].$$

Posmatrajmo najpre proces sa intervalom $x_1^{(1)}$. Kako $0 \notin F'(x_1^{(1)})$, može se, zbog jednostavnosti, primeniti Mooreov metod (1.8.3). Tada nalazimo da je $x_2^{(1)} = [-1.28, -0.975]$. Posle pete iteracije dobijamo interval

$$x_5^{(1)} = [-1.000000038, -0.999999996],$$

koji sadrži prostu nulu $\xi_1 = -1$.

U cilju nalaženja inkluzivnog intervala za dvostruku nulu $\xi_2 = 1$, nastavljamo iterativni postupak sa intervalom $x_1^{(2)}$. Konvergencija je u ovom slučaju linearna. Posle 33 iteracija dobijamo interval

$$x_{33}^{(2)} = [0.999995768, 1.000005406]$$

koji sadrži nulu $\xi_2 = 1$.

5.2. SISTEMI NELINEARNIH JEDNAČINA

Ovo poglavlje je posvećeno problemu rešavanja sistema nelinearnih jednačina. Opštosti radi, tretiran je slučaj rešavanja operatorskih jednačina u Banachovom prostoru.

5.2.1. Uvodne napomene

Kao što je rečeno u odeljku 3.1.1, sistem nelinearnih jednačina

$$(2.1.1) \quad f_i(x_1, \dots, x_n) = 0 \quad (i = 1, \dots, n),$$

gde su $f_i : R^n \rightarrow R$ ($i = 1, \dots, n$) date funkcije, može se tretirati kao specijalan slučaj operatorske jednačine

$$(2.1.2) \quad Fu = \theta,$$

gde je F operator koji preslikava Banachov prostor X u Banachov prostor Y , i θ nula-vektor prostora Y . Naime, treba samo uzeti da je $X = Y = R^n$, $u = \vec{x} = [x_1 \dots x_n]^T$, $\theta = [0 \dots 0]^T$,

$$Fu = \vec{f}(\vec{x}) = \begin{bmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_n(x_1, \dots, x_n) \end{bmatrix}.$$

Osnovni metod za rešavanje operatorske jednačine (2.1.2), ali i za rešavanje sistema jednačina (2.1.1), je metod Newton-Kantoroviča, koji predstavlja uopštenje Newtonovog metoda (1.2.3).

U našem izlaganju ovaj metod Newton-Kantoroviča biće razmatran za opšti slučaj operatorske jednačine (2.1.2).

S obzirom da metod Newton-Kantoroviča zahteva nalaženje inverznog operatora $F'_{(u)}^{-1}$, koje ponekad može biti komplikovano, ili čak nemoguće, u literaturi je u poslednje vreme razradjena čitava klasa tzv. kvazi-njutnovskih metoda, kod kojih se koriste izvesne aproksimacije pomenutog operatora. U našem izlaganju, od ovih metoda obradićemo samo gradijentni metod, i njega ćemo prezentirati kao metod za minimizaciju izvesne funkcionele.

O pomenutim metodima postoji opširna literatura (videti, na primer, [10], [39], [62], [85]).

5.2.2. Metod Newton-Kantoroviča

Osnovni iterativni metod za rešavanje jednačine (2.1.2) je metod Newton-Kantoroviča, koji predstavlja generalizaciju klasičnog 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 (videti [39]). Takođe, ovaj metod je razmatran i od strane drugih autora (videti, na primer, [41], [55], [91], [97]).

Pretpostavimo da jednačina (2.1.2) ima rešenje $u=a$ i da je operator $F:X \rightarrow Y$ Fréchet-diferencijabilan u konveksnoj okolini D tačke a . Metod Newton-Kantoroviča se zasniva na linearizaciji jednačine (2.1.2). Naime, neka je nadjeno približno rešenje u_k . Tada za nalaženje sledeće aproksimacije u_{k+1} , jednačinu (2.1.2) zamениmo jednačinom

$$(2.2.1) \quad Fu_k + F'_{(u_k)}(u - u_k) = \theta .$$

Ako za operator $F'_{(u_k)}$ postoji inverzni operator $\Gamma(u_k) = (F'_{(u_k)})^{-1}$, iz (2.2.1) dolazimo do iterativnog metoda

$$(2.2.2) \quad u_{k+1} = u_k - \Gamma(u_k)Fu_k \quad (k = 0, 1, \dots),$$

koji je poznat kao metod Newton-Kantoroviča. Početna vrednost u_0 za generisanje niza $\{u_k\}_{k \in \mathbb{N}}$ uzima se iz D , a njen izbor predstavlja dosta težak problem.

Metod (2.2.2) može se predstaviti u obliku

$$u_{k+1} = Tu_k \quad (k = 0, 1, \dots),$$

gde je

$$Tu = u - \Gamma(u)Fu.$$

Teorema 2.2.1. 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 $u_0 \in D$ dovoljno blisko tački a , iterativni proces (2.2.2) ima red konvergencije najmanje dva.

Dokaz. Neka je $\|\Gamma(u)\| \leq m$, $\|F''(u)\| \leq M_2$ za svako $u \in D$.

Na osnovu Taylorove formule imamo

$$(2.2.3) \quad Fa = \theta = F_u + F'_u(a-u) + W(u, a-u),$$

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

$$Tu-a = u-\Gamma(u)Fu-a = \Gamma(u)[F'_u(u-a) - Fu],$$

na osnovu (2.2.3), dobijamo

$$Tu-a = \Gamma(u)W(u, a-u),$$

odakle sleduje

$$\|Tu-a\| \leq \|\Gamma(u)\| \cdot \|W(u, a-u)\| \leq \frac{1}{2} m M_2 \|u-a\|^2,$$

tj.

$$\|Tu-a\| = 0 (\|u-a\|^2).$$

Ovim je dokaz teoreme završen.

U daljem razmatranju smatraćemo da je D kugla $K[u_0, R]$, gde je u_0 početna vrednost niza $\{u_k\}_{k \in N_0}$.

Ako je ispunjen Lipschitzov uslov

$$(2.2.4) \quad \|F'(u) - F'(v)\| \leq L \|u-v\| \quad (u, v \in K[u_0, R])$$

iz

$$Fu - Fv - F'(v)(u-v) = \int_0^1 [F'(v+t(u-v)) - F'(v)](u-v) dt$$

sleduje nejednakost

$$(2.2.5) \quad \|Fu - Fv - F'(v)(u-v)\| \leq \frac{L}{2} \|u-v\|^2.$$

Teorema 2.2.2. Neka je operator F Fréchet-diferencijabilan u kugli $K[u_0, R]$ i zadovoljava uslov (2.2.4) i neka su tačne nejednakosti

$$(2.2.6) \quad \|\Gamma_0\| \leq b_0, \quad \|\Gamma_0 Fu_0\| \leq n_0, \quad h_0 = b_0^{-1} n_0 \leq \frac{1}{2},$$

gde je $\Gamma_0 = \Gamma(u_0)$.

Ako je

$$(2.2.7) \quad R \geq r_0 = \frac{1-\sqrt{1-2h_0}}{h_0} n_0,$$

niz $\{u_k\}_{k \in N_0}$, koji se generiše pomoću (2.2.2), konvergira rešenju $a \in K[u_0, r_0]$ jednačine (2.1.2).

Dokaz. Neka su nizovi $\{b_k\}, \{n_k\}, \{h_k\}, \{r_k\}$ definisani sa

$$b_{k+1} = \frac{b_k}{1-h_k}, \quad n_{k+1} = \frac{h_k}{2(1-h_k)} n_k,$$

$$h_{k+1} = b_{k+1}^{-1} n_{k+1}, \quad r_k = \frac{1-\sqrt{1-2h_{k+1}}}{h_{k+1}} n_k.$$

Dokazaćemo da niz $\{u_k\}_{k \in N_0}$ postoji i da je

$$(2.2.8) \quad \|\Gamma(u_k)\| \leq b_k, \quad \|\Gamma(u_k) Fu_k\| \leq n_k, \quad h_k \leq \frac{1}{2}$$

i

$$(2.2.9) \quad K[u_k, r_k] \subset K[u_{k-1}, r_{k-1}].$$

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 $u_m \in K[u_0, R]$, 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.

$$(2.2.10) \quad \Gamma(u_{m+1}) = \sum_{i=0}^{\infty} (-1)^i [\Gamma(u_m)(F'(u_{m+1})^{-F'(u_m)})]^i \Gamma(u_m),$$

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

$$(2.2.11) \quad \|\Gamma(u_{m+1})\| \leq \sum_{i=0}^{\infty} \lambda^i b_m = \frac{b_m}{1-h_m} = b_{m+1},$$

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

$$(2.2.12) \quad \|Fu_{m+1}\| \leq \frac{L}{2} \|u_{m+1} - u_m\|^2 \leq \frac{L}{2} \eta_m^2,$$

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 $\|u - u_{m+1}\| \leq r_{m+1}$ sleduje nejednakost

$$(2.2.13) \quad \|u - u_m\| \leq \|u - u_{m+1}\| + \|u_{m+1} - u_m\| \leq r_{m+1} + \eta_m.$$

Kako je $r_{m+1} + \eta_m = r_m$ dokaz je završen.

Kako je $h_k \leq \frac{1}{2}$ iz definicije niza $\{\eta_k\}$ sleduje nejednakost

$$\eta_{k+1} \leq \frac{1}{2} \eta_k,$$

odakle zaključujemo da je $\{\eta_k\}$ nula-niz. Tada je

$$\lim_{k \rightarrow +\infty} r_k = \lim_{k \rightarrow +\infty} \frac{1 - \sqrt{1 - 2h_k}}{h_k} \eta_k = \lim_{k \rightarrow +\infty} \frac{2\eta_k}{1 + \sqrt{1 - 2h_k}} = 0,$$

što znači da niz $\{u_k\}$ konvergira nekoj tački $a \in K[u_0, r_0]$.

Tačka a je rešenje jednačine (2.1.2), jer je na osnovu (2.2.12)

$$\lim_{m \rightarrow +\infty} \|Fu_{m+1}\| \leq \frac{L}{2} \quad \lim_{m \rightarrow +\infty} \eta_m^2 = 0,$$

tj.

$$\lim_{m \rightarrow +\infty} Fu_m = 0.$$

Ovim je dokaz završen.

Teorema 2.2.3. Ako su ispunjeni uslovi prethodne teoreme, tada je

$$(2.2.14) \quad \|u_k - a\| \leq \frac{1}{2^{k-1}} (2h_0)^{2^k - 1} \eta_0 \quad (k \in \mathbb{N}).$$

Dokaz. Primetimo najpre da je

$$\frac{h_k}{1-h_k} \leq 2h_k \quad (0 \leq h_k \leq \frac{1}{2}).$$

Tada je, na osnovu definicije nizova $\{\eta_k\}$ i $\{h_k\}$,

$$\eta_{k+1} \leq h_k \eta_k \quad i \quad h_{k+1} \leq 2h_k^2,$$

odakle sleduje

$$h_k \leq \frac{1}{2} (2h_0)^{2^k}$$

i

$$\eta_k \leq h_{k-1} h_{k-2} \eta_{k-2} \leq \dots \leq h_{k-1} h_{k-2} \dots h_0 \eta_0$$

$$\leq \frac{1}{2^k} (2h_0)^{1+2+\dots+2^{k-1}} \eta_0 = \frac{1}{2^k} (2h_0)^{2^k - 1} \eta_0.$$

Najzad, na osnovu (2.2.13), na $u=a$ i $m=k$, imamo

$$\|u_k - a\| \leq r_k = \frac{2}{1 + \sqrt{1 - 2h_k}} \eta_k \leq 2\eta_k,$$

tj. (2.2.14).

Da bi se izbeglo odredjivanje inverznog operatora $\Gamma(u) = [F'_{(u)}]^{-1}$ pri svakom koraku, metod Newton-Kantoroviča može se modifikovati na sledeći način

$$(2.2.15) \quad u_{k+1} = u_k - \Gamma_0 F u_k \quad (k=0,1,2,\dots),$$

gde je $\Gamma_0 = \Gamma(u_0)$. Uvodjenjem operatora T pomoću

$$(2.2.16) \quad Tu = u - \Gamma_0 F u,$$

modifikovani metod (2.2.15) se može predstaviti u obliku

$$u_{k+1} = Tu_k \quad (k=0,1,2,\dots).$$

Prepostavimo sada da su ispunjeni sledeći uslovi:

1° Operator F je Frechet-diferencijabilan u kugli $K[u_0, R]$,

2° $F'_{(u)}$ zadovoljava uslov (2.2.4),

3° Operator Γ_0 postoji i

$$\|\Gamma_0\| \leq b_0, \quad \|\Gamma_0 F u_0\| \leq \eta_0.$$

Tada važi sledeća teorema.

Teorema 2.2.4. Ako su ispunjeni uslovi

$$h_o = b_o L n_o < \frac{1}{2}$$

i

$$r_o = \frac{1 - \sqrt{1 - 2h_o}}{h_o} n_o \leq R$$

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

$$Tu - Tv = u - v - \Gamma_o (Fu - Fv)$$

$$(2.2.17) \quad = \Gamma_o \int_0^1 [F'(u_o) - F'(v + t(u-v))] (u-v) dt$$

i uslova (2.2.4) sleduje

$$(2.2.18) \quad \|Tu - Tv\| \leq b_o L r_o \|u - v\|,$$

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

$$\|Tu - u_o\| \leq \|Tu - Tu_o\| + \|Tu_o - u_o\| \leq \frac{b_o L}{2} \|u - u_o\|^2 + n_o,$$

tj.

$$\|Tu - u_o\| \leq \frac{b_o L r_o^2}{2} + n_o = r_o.$$

S obzirom da $T: K[u_o, r_o] \rightarrow K[u_o, r_o]$ 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

$$\|Tu - a\| \leq q \|u - a\|,$$

što znači da je iterativni proces (2.2.15) prvog reda.

Predjimo sada na slučaj sistema nelinearnih jednačina

$$(2.2.19) \quad f_i(x_1, \dots, x_n) = 0 \quad (i = 1, \dots, n).$$

Ovde je $X = Y = \mathbb{R}^n$, $u = \vec{x} = [x_1 \dots x_n]^T$ i F definisano pomoću (2.1.3). Ako je F Fréchet-diferencijabilan operator (videti odeljak 2.2.4) tada je

$$F'(u) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} = W(\vec{x}),$$

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

$$(2.2.20) \quad \vec{x}^{(k+1)} = \vec{x}^{(k)} - W^{-1}(\vec{x}^{(k)}) \vec{f}(\vec{x}^{(k)}) \quad (k=0,1,\dots),$$

gde je $\vec{x}^{(k)} = [x_1^{(k)} \dots x_n^{(k)}]^T$. Vrlo često se ovaj metod u literaturi sreće kao Newton-Raphsonov metod.

Do metoda (2.2.20) možemo doći i elementarnije linearizacijom sistema jednačina (2.1.1) u okolini aproksimacije $\vec{x}^{(k)}$. Neka je $\vec{a} = [a_1 \dots a_n]^T$ tačno rešenje ovog sistema. Korišćenjem Taylorovog razvoja za funkcije koje se pojavljuju u (2.1.1) dobijamo

$$\begin{aligned} f_i(a_1, \dots, a_n) &= f_i(x_1^{(k)}, \dots, x_n^{(k)}) + \frac{\partial f_i}{\partial x_1}(a_1 - x_1^{(k)}) + \dots \\ &\quad + \frac{\partial f_i}{\partial x_n}(a_n - x_n^{(k)}) + r_i^{(k)} \quad (i = 1, \dots, n), \end{aligned}$$

gde se parcijalni izvodi na desnoj strani ovih jednakosti izračunavaju u tački $\vec{x}^{(k)}$. Veličina $r_i^{(k)}$ je odgovarajući ostatak u Taylorovoj formuli.

Kako je $f_i(a_1, \dots, a_n) = 0$ ($i=1, \dots, n$), prethodni sistem jednakosti se može predstaviti u matričnom obliku

$$\vec{0} = \vec{f}(\vec{x}^{(k)}) + W(\vec{x}^{(k)})(\vec{a} - \vec{x}^{(k)}) + \vec{r}^{(k)},$$

gde je $\vec{r}^{(k)} = [r_1^{(k)} \dots r_n^{(k)}]^T$. Ako je Jacobieva matrica za \vec{x} regularna, tada imamo

$$\vec{a} = \vec{x}^{(k)} - W^{-1}(\vec{x}^{(k)}) \vec{f}(\vec{x}^{(k)}) - W^{-1}(\vec{x}^{(k)}) \vec{r}^{(k)}.$$

Zanemarivanjem poslednjeg člana na desnoj strani ove jednakošti, umesto vektora \vec{a} dobićemo njegovu novu aproksimaciju koju ćemo označiti sa $\vec{x}^{(k+1)}$. Tako dobijamo (2.2.20).

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 određuje u svakoj iteraciji, već samo u prvoj. Dakle,

$$(2.2.21) \quad \vec{x}^{(k+1)} = \vec{x}^{(k)} - W^{-1}(\vec{x}^{(0)}) \vec{f}(\vec{x}^{(k)}) \quad (k=0,1,\dots).$$

Primedba 2.2.1. Modifikovani metod (2.2.21) može se shvatiti i kao metod proste iteracije

$$\vec{x}^{(k+1)} = T\vec{x}^{(k)} = \vec{x}^{(k)} + \Lambda \vec{f}(\vec{x}^{(k)}) \quad (k=0,1,\dots),$$

sa matricom Λ određenom iz uslova da je izvod od T nula-operator, tj. da je $I + \Lambda W(\vec{x}^{(0)})$ nula matrica. Ako je $W(\vec{x}^{(0)})$ regularna matrica imamo $\Lambda = W^{-1}(\vec{x}^{(0)})$.

Prethodno dokazane teoreme mogu se prilagoditi za slučaj sistema nelinearnih jednačina, pri čemu se uslovi za konvergenciju procesa (2.2.20) i (2.2.21) mogu iskazati na različite načine, što zavisi od uvedene norme u X . Tako na primer, uzimajući za normu u R^n

$$\|\vec{x}\| = \|\vec{x}\|_\infty = \max_i |x_i|$$

i prepostavljajući da $f \in C^2(D)$, gde je D kugla $K[x^{(0)}, R]$, iz teoreme 2.2.2 sleduje:

Posledica 2.2.1. Neka su u D ispunjeni sledeći uslovi:

$$(2.2.22) \quad s_{ij} = \sum_{k=1}^n \left| \frac{\partial^2 f_i}{\partial x_j \partial x_k} \right| \leq N \quad (i,j = 1, \dots, n);$$

$$(2.2.23) \quad \|\vec{f}(\vec{x}^{(0)})\| \leq Q, \quad \|W^{-1}(\vec{x}^{(0)})\| \leq b;$$

$$(2.2.24) \quad \Delta_0 = \det W(\vec{x}^{(0)}) \neq 0, \quad h = nNQb^2 \leq \frac{1}{2}.$$

Tada, ako je $R \geq r = \frac{1 - \sqrt{1-2h}}{h} Qb$, metod Newton-Kantoroviča (2.2.20) konvergira ka rešenju $a \in K[\vec{x}^{(0)}, r]$.

Dokaz. Neka je $a_{ij} = \left. \frac{\partial f_i}{\partial x_j} \right|_{\vec{x}=\vec{x}^{(0)}}$ i neka su sa A_{ij} označeni kofaktori elemenata a_{ij} Jacobieve matrice $W(\vec{x}^{(0)}) = [a_{ij}]$. Tada je, na osnovu (2.2.23)

$$\|r_0\| = \|W^{-1}(\vec{x}^{(0)}) \vec{f}(\vec{x}^{(0)})\| = \max_i \frac{1}{|\Delta_0|} \sum_{j=1}^n |A_{ji}| \leq b.$$

Dalje, na osnovu saglasnosti norme matrice sa normom vektora, imamo

$$\|W^{-1}(\vec{x}^{(0)}) \vec{f}(\vec{x}^{(0)})\| \leq \|W^{-1}(\vec{x}^{(0)})\| \cdot \|\vec{f}(\vec{x}^{(0)})\| \leq Qb.$$

Primetimo da pod uslovom (2.2.22), $W(\vec{x})$ zadovoljava Lipschitzov uslov (2.2.4) sa konstantom $L = nN$. Zaista, za dva proizvoljna vektora $\vec{x} = [x_1 \dots x_n]^T$ i $\vec{y} = [y_1 \dots y_n]^T$ imamo

$$\begin{aligned} \|W(\vec{x}) - W(\vec{y})\| &= \max_i \sum_{j=1}^n \left| \frac{\partial f_i}{\partial x_j}(\vec{x}) - \frac{\partial f_i}{\partial x_j}(\vec{y}) \right| \\ &= \max_i \sum_{j=1}^n \left| \sum_{k=1}^n \frac{\partial^2 f_i}{\partial x_j \partial x_k}(\xi) (x_k - y_k) \right| \\ &= \max_i \sum_{j=1}^n (N \|\vec{x} - \vec{y}\|) = nN \|\vec{x} - \vec{y}\|, \end{aligned}$$

gde je $\xi = \vec{y} + \theta(\vec{x} - \vec{y})$, $0 < \theta < 1$.

Na osnovu prethodnog vidimo da su pod uslovima (2.2.24), ispunjeni uslovi (2.2.6) u teoremi 2.2.2, pa je ovaj rezultat posledica teoreme 2.2.2.

Primedba 2.2.2. Kako za $0 < h \leq 1/2$ važi $(1 - \sqrt{1-2h})/h \leq 2$, za r u posledici 2.2.1 možemo uzeti $r = 2Qb$.

Primedba 2.2.3. Modifikovani metod Newton-Kantoroviča (2.2.21) konvergira, takodje, pod uslovima datim u posledici 2.2.1.

Primer 2.2.1. Posmatrajmo sistem nelinearnih jednačina

$$f_1(x_1, x_2) = 9x_1^2 x_2 + 4x_2^2 - 36 = 0,$$

$$f_2(x_1, x_2) = 16x_2^2 - x_1^4 + x_2 + 1 = 0,$$

koji ima rešenje u prvom kvadrantu ($x_1, x_2 > 0$). Da bismo se u to uverili dati sistem jednačina predstavimo u obliku

$$x_1^2 = \frac{4}{9x_2}(9 - x_2^2), \quad x_1^4 = 1 + x_2 + 16x_2^2,$$

odakle nalazimo grane implicitnih funkcija f_1 i f_2 u prvom kvadrantu:

$$x_1 = x_1(f_1) = \frac{2}{3} \sqrt{\frac{9 - x_2^2}{x_2}} \quad (0 < x_2 \leq 3),$$

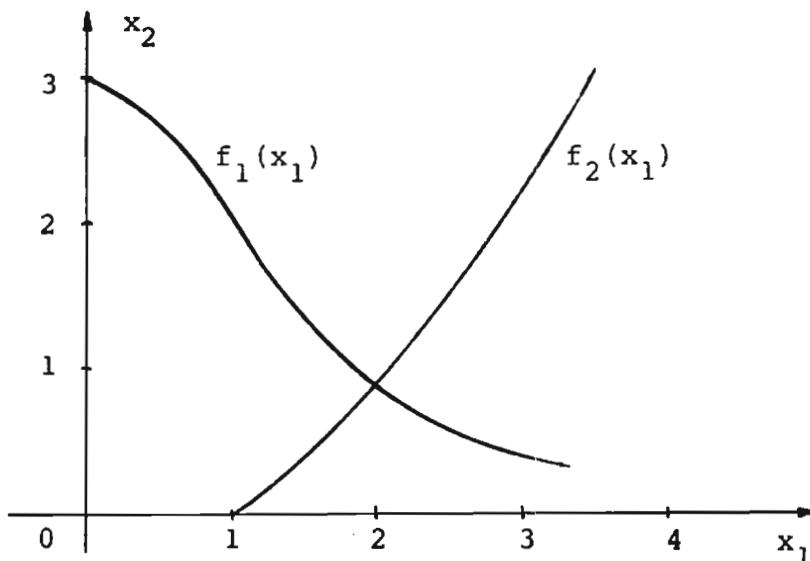
$$x_1 = x_1(f_2) = \sqrt[4]{1 + x_2 + 16x_2^2} \quad (x_2 \geq 0).$$

Na osnovu tabele vrednosti

x_2	0.0	0.5	1.0	1.5	2.0	2.5	3.0
$x_1(f_1)$		2.79	1.89	1.41	1.05	0.70	0.00
$x_1(f_2)$	1.00	1.53	2.06	2.49	2.86	3.19	3.49

možemo približno skicirati grafike ovih grana (videti sl.2.2.1).

Kako se rešenje nalazi u okolini tačke $(2,1)$, za početni



sl. 2.2.1

vektor čemo uzeti $\vec{x}^{(0)} = [2 \ 1]^T$, tj. $x_1^{(0)} = 2$ i $x_2^{(0)} = 1$.

Kako je

$$\frac{\partial f_1}{\partial x_1} = 18x_1x_2, \quad \frac{\partial f_1}{\partial x_2} = 9x_1^2 + 8x_2, \quad \frac{\partial f_2}{\partial x_1} = -4x_1^3, \quad \frac{\partial f_2}{\partial x_2} = 32x_2 + 1;$$

$$\frac{\partial^2 f_1}{\partial x_1^2} = 18x_2, \quad \frac{\partial^2 f_1}{\partial x_1 \partial x_2} = 18x_1, \quad \frac{\partial^2 f_1}{\partial x_2^2} = 8,$$

$$\frac{\partial^2 f_2}{\partial x_1^2} = -12x_1^2, \quad \frac{\partial^2 f_2}{\partial x_1 \partial x_2} = 0, \quad \frac{\partial^2 f_2}{\partial x_2^2} = 32,$$

imamo

$$W(x) = \begin{bmatrix} 18x_1x_2 & 9x_1^2 + 8x_2 \\ -4x_1^3 & 32x_2 + 1 \end{bmatrix},$$

$$W^{-1}(\vec{x}) = \frac{1}{\Delta(\vec{x})} \begin{bmatrix} 32x_2 + 1 & -(9x_1^2 + 8x_2) \\ 4x_1^3 & 18x_1x_2 \end{bmatrix},$$

$$\Delta(\vec{x}) = 18x_1x_2(32x_2 + 1) + 4x_1^3(9x_1^2 + 8x_2) > 0,$$

$$s_{11} = 18(x_1 + x_2), \quad s_{12} = 2(9x_1 + 4), \quad s_{21} = 12x_1^2, \quad s_{22} = 32.$$

Na dalje, stavimo $f_i^{(k)} \equiv f_i(x_1^{(k)}, x_2^{(k)})$ i $\Delta_k = \Delta(x^{(k)})$ ($i = 1, 2$; $k = 0, 1, \dots$). Skalarni oblik metoda Newton-Kantoroviča (2.2.20) je

$$x_1^{(k+1)} = x_1^{(k)} - \frac{1}{\Delta_k} \{ (32x_2^{(k)} + 1) f_1^{(k)} - (9x_1^{(k)})^2 + 8x_2^{(k)} f_2^{(k)} \},$$

$$x_2^{(k+1)} = x_2^{(k)} - \frac{1}{\Delta_k} \{ 4x_1^{(k)} f_1^{(k)} + 18x_1^{(k)} x_2^{(k)} f_2^{(k)} \}.$$

Kako je

$$\|\vec{f}(\vec{x}^{(0)})\| = \max(|f_1^{(0)}|, |f_2^{(0)}|) = \max(4, 2) = 4,$$

$$W^{-1}(\vec{x}^{(0)}) = \frac{1}{2596} \begin{bmatrix} 33 & -44 \\ 32 & 36 \end{bmatrix},$$

$$\| W^{-1}(\vec{x}^{(0)}) \| = \frac{1}{2596} \max(33+44, 32+36) = \frac{77}{2596} < 0.03,$$

tj. $Q = 4$, $b = 0.03$, saglasno primedbi 2.2.2 za r se može uzeti $r = 2Qb = 0.24$. Tada uslov da se \vec{x} nadje u kugli $K[\vec{x}^{(0)}, r]$, tj. uslov $\| \vec{x} - \vec{x}^{(0)} \| = \max_i |x_i - x_i^{(0)}| \leq r = 0.24$, daje

$$1.76 \leq x_1 \leq 2.24, \quad 0.76 \leq x_2 \leq 1.24,$$

pa imamo

$$s_{11} \leq 18(1.24 + 2.24) = 62.64, \quad s_{12} = 2(9 \cdot 2.24 + 4) = 48.32,$$

$$s_{21} \leq 12 \cdot 2.24^2 = 60.2112, \quad s_{22} = 32.$$

Dakle, možemo uzeti $N = 63$.

Kako je $h = nNQb^2 = 2 \cdot 63 \cdot 4 \cdot 0.03^2 = 0.4536 < 1/2$, zaključujemo da će prethodno dati iterativni proces, za izabrani početni vektor, konvergirati ka rešenju. Rezultati dobijeni primenom ovog procesa su dati u tabeli 2.2.1. Brojevi u zagradama označavaju decimalni eksponent. Primetimo da se poslednje dve iteracije poklapaju na svim ciframa.

Tabela 2.2.1

k	$x_1^{(k)}$	$x_2^{(k)}$	$f_1^{(k)}$	$f_2^{(k)}$
0	2.	1.	4.	2.
1	1.983050847	0.922958397	7.3(-2)	8.8(-2)
2	1.983707109	0.920743215	-2.9(-5)	6.8(-5)
3	1.983708734	0.920742637	9. (-10)	-1.4(-9)
4	1.983708734	0.920742637		

Poseban problem kod primene metoda Newton-Kantorovića je izbor startnih vrednosti, tj. početnog vektora $\vec{x}^{(0)}$, koji bi obezbedio konvergenciju procesa. Jedno rešenje ovog problema dato je 1974. godine u radu [43]. Naime, za njegovo načaštenje predlaže se iterativni proces

$$(2.2.25) \quad \vec{x}_{m+1} = \vec{x}_m - W^{-1}(\vec{x}_m)(\vec{f}(\vec{x}_m) - \alpha_m \vec{f}(\vec{x}_0)) \quad (m=0, 1, \dots),$$

gde je

$$\alpha_m = \max \left[0, 1 - \frac{1}{2nN \|\vec{f}(\vec{x}_0)\|} \left(\frac{1}{\|W^{-1}(\vec{x}_m)\|^2} + \frac{3}{4} \sum_{i < m} \frac{1}{\|W^{-1}(\vec{x}_i)\|^2} \right) \right],$$

startujući sa nekom grubom aproksimacijom \vec{x}_0 . Proces (2.2.25) se može interpretirati kao iteracija metoda Newton-Kantoroviča iz tačke \vec{x}_m na sistem

$$\vec{f}(\vec{x}) = \alpha_m \vec{f}(\vec{x}_0).$$

Primetimo da $\alpha_m \in [0, 1]$. Ako je $\det W(\vec{x}_0) \neq 0$, tada se može pokazati da je, takodje, $\det W(\vec{x}_m) \neq 0$ i da je

$$2nN \|W^{-1}(\vec{x}_m)\|^2 \|\vec{f}(\vec{x}_m) - \alpha_m \vec{f}(\vec{x}_0)\| \leq 1,$$

gde je $\{\alpha_m\}$ nerastući niz, tj. $\alpha_{m+1} \leq \alpha_m$.

Neka je $G (\subset \mathbb{R}^n)$ konveksna oblast koja sadrži rešenje \vec{a} . Pod pretpostavkama

$$\vec{f}(\vec{x}) \in C^2(G), \det W(\vec{x}) \neq 0 \quad (\forall \vec{x} \in G), \quad \vec{f}(\vec{x}) \neq \vec{f}(\vec{y}) \quad (\forall \vec{x}, \vec{y} \in G)$$

u pomenutom radu [43] dokazana je sledeća teorema:

Teorema 2.2.5. Za svako $\vec{x}_0 \in G$, posle konačnog broja iteracija s u (2.2.25), dolazi se do \vec{x}_s za koji su ispunjeni uslovi za konvergenciju metoda Newton-Kantoroviča, pri čemu je $\alpha_m = 0$ za svako $m \geq s$.

Dakle, kao početni vektor $\vec{x}^{(0)}$ za metod Newton-Kantoroviča može se uzeti vektor \vec{x}_s , tj. $\vec{x}^{(0)} = \vec{x}_s$.

5.2.3. Gradijentni metod

Neka je dat sistem nelinearnih jednačina (2.2.19), čiji je matrični oblik (videti (2.1.3))

$$(2.3.1) \quad \vec{f}(\vec{x}) = \vec{0}.$$

Gradijentni metod za rešavanje datog sistema jednačina zasniva se na minimizaciji funkcionele

$$U(\vec{x}) = \sum_{i=1}^n f_i(x_1, \dots, x_n)^2 = (\vec{f}(\vec{x}), \vec{f}(\vec{x})).$$

Lako je videti da važi ekvivalencija $U(\vec{x}) = 0 \Leftrightarrow \vec{f}(\vec{x}) = \vec{0}$.

Pretpostavimo da jednačina (2.3.1) ima jedinstveno rešenje $\vec{x} = \vec{a}$, za koje funkcionala U dostiže minimum. Neka je $\vec{x}^{(0)}$ početna aproksimacija ovog rešenja. Konstruišimo niz $\{\vec{x}^{(k)}\}$ takav da je $U(\vec{x}^{(0)}) > U(\vec{x}^{(1)}) > U(\vec{x}^{(2)}) > \dots$. Postupimo kao i kod linearnih sistema jednačina (videti odeljak 4.3.7), tj. uzmimo

$$(2.3.2) \quad \vec{x}^{(k+1)} = \vec{x}^{(k)} - \lambda_k \nabla U(\vec{x}^{(k)}) \quad (k=0, 1, \dots),$$

gde je $\nabla U(\vec{x}) = \text{grad } U(\vec{x}) = \left[\frac{\partial U}{\partial x_1} \dots \frac{\partial U}{\partial x_n} \right]^T$. Parametar λ_k odre-

djujemo 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

$$(2.3.3) \quad \lambda_k = t = \frac{\sum_{i=1}^n H_i f_i(\vec{x}^{(k)})}{\sum_{i=1}^n H_i^2},$$

gde smo stavili $H_i = (\nabla f_i(\vec{x}^{(k)}), \nabla U(\vec{x}^{(k)}))$ ($i = 1, \dots, n$).

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

$$(2.3.4) \quad \nabla U(\vec{x}) = 2W^T(\vec{x}) \vec{f}(\vec{x}),$$

gde je $\vec{W}(\vec{x})$ Jacobieva matrica.

Shodno prethodnom, (2.3.3) se svodi na

$$\lambda_k = \frac{1}{2} \cdot \frac{(\vec{f}(k), \vec{w}_k \vec{w}_k^T \vec{f}(k))}{(\vec{w}_k \vec{w}_k^T \vec{f}(k), \vec{w}_k \vec{w}_k^T \vec{f}(k))},$$

gde su $\vec{f}(k) = \vec{f}(\vec{x}^{(k)})$ i $\vec{w}_k = \vec{w}(\vec{x}^{(k)})$. Tada se gradijentni metod (2.3.2) može predstaviti u obliku

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} - 2\lambda_k \vec{w}_k^T \vec{f}(\vec{x}^{(k)}) \quad (k=0, 1, \dots).$$

Kao što vidimo, ulogu inverzne matrice $\vec{w}^{-1}(\vec{x}^{(k)})$, koja se javlja kod metoda Newton-Kantoroviča, preuzeala je matrica $2\lambda_k \vec{w}_k^T$.

Primer 2.3.1. Na rešavanje sistema jednačina iz primera 2.2.1 primenićemo gradijentni metod startujući opet sa istim početnim vektorom $\vec{x}^{(0)} = [2 \ 1]^T$. Odgovarajući rezultati su dati u tabeli 2.3.1.

Tabela 2.3.1

k	$x_1^{(k)}$	$x_2^{(k)}$	$2\lambda_k$
0	2.	1.	3.05787395 (-4)
1	1.975537008	0.9259994504	5.387476894 (-4)
2	1.983210179	0.9201871306	3.395536228 (-4)
3	1.983643559	0.9207840032	5.355965389 (-4)
4	1.983705230	0.9207387845	3.393286035 (-4)
5	1.983708270	0.9207429317	5.355733542 (-4)
6	1.983708709	0.9207426096	3.393325162 (-4)
7	1.983708731	0.9207426391	5.359906237 (-4)
8	1.983708734	0.9207426368	3.377933010 (-4)
9	1.983708734	0.9207426370	

Primetimo da je konvergencija ovde znatno sporija nego kod metoda Newton-Kantoroviča, s obzirom da je gradijentni metod prvega reda.

Metod gradijenata se uspešno koristi u mnogim optimizacionim problemima nelinearnog programiranja. U literaturi je poslednjih godina razradjen veliki broj metoda, posebno gradijentnog tipa, na osnovu kojih je formiran veći broj paketa programa za rešavanje zadataka nelinearnog programiranja.

5.3. ALGEBARSKE JEDNAČINE

Zbog važnosti koje zauzimaju algebarske jednačine u nauci i tehnici, ovo poglavlje posvećujemo numeričkim metodama za rešavanje numeričkih algebarskih jednačina. Poseban odeljak posvećujemo lokalizaciji korena jednačina, imajući pritom u vidu izbor početnih rešenja u iterativnim procesima. Pored standardnih metoda kakvi su Bernoulliev i Newton-Hornerov metod, izložićemo i dva metoda trećeg reda, kao i metode za simultano nalaženje korena.

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

$$(3.1.1) \quad P(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n = 0 \quad (a_0 \neq 0),$$

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 ($i = 0, 1, \dots, n$) pomoću Vièteovih formula

$$(3.1.2) \quad \begin{aligned} F_1 &= \sum_i x_i = -\frac{a_1}{a_0}, \\ F_2 &= \sum_{i < j} x_i x_j = \frac{a_2}{a_0}, \\ F_3 &= \sum_{i < j < k} x_i x_j x_k = -\frac{a_3}{a_0}, \\ &\vdots \\ &\vdots \\ F_n &= x_1 x_2 \dots x_n = (-1)^n \frac{a_n}{a_0}. \end{aligned}$$

Jednačina (3.1.1) je numerička ako svi njeni koeficijenti a_i ($i = 0, 1, \dots, n$) 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 $n \leq 4$. Međutim, numeričke algebarske jednačine je uvek moguće rešiti sa određenom 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. Među ovim metodama biće onih koji se koriste za određivanje samo jednog korena (na primer, najvećeg po modulu), kao i onih za simultano određivanje više korena (na primer, dva ili svih n).

Ukoliko je nekim od metoda određeno $k (< n)$ korena x_1, x_2, \dots, x_k , tada se za određivanje ostalih korena rešava jednačina nižeg stepena

$$Q(x) = \frac{P(x)}{(x-x_1)(x-x_2)\dots(x-x_k)} = 0.$$

S obzirom da se koreni x_1, \dots, x_k određuju sa ograničenom tačnošću, jasno je da koeficijenti jednačine $Q(x) = 0$ neće biti apsolutno tačni, što znači da se tačnost korena x_{k+1}, \dots, x_n narušava, bez obzira na metod kojim se oni određuju. Naime, koreni x_{k+1}, \dots, 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 pretpostav-

kom da koeficijenti a_i ($i=1, \dots, n$) postanu redom $a_i + \varepsilon_i$, gde je $|\varepsilon_i| \ll |a_i|$ ($i=1, \dots, n$), odredićemo grešku korena x_k . Naime, u tom slučaju koren x_k postaje $x_k + \xi_k$.

Kako je

$$(x_k + \xi_k)^n + (a_1 + \varepsilon_1)(x_k + \xi_k)^{n-1} + \dots + (a_{n-1} + \varepsilon_{n-1})(x_k + \xi_k) + (a_n + \varepsilon_n) = 0,$$

korišćenjem aproksimacije

$$(x_k + \xi_k)^m \approx x_k^m + mx_k^{m-1}\xi_k \quad (m = 2, \dots, n)$$

i zanemarivanjem članova oblika $\xi_k \varepsilon_i$, dobijamo

$$P(x_k) + P'(x_k)\xi_k + (\varepsilon_1 x_k^{n-1} + \varepsilon_2 x_k^{n-2} + \dots + \varepsilon_{n-1} x_k + \varepsilon_n) \approx 0,$$

odakle, s obzirom na $P(x_k) = 0$ i $P'(x_k) \neq 0$, sleduje

$$\xi_k \approx -\frac{1}{P'(x_k)} (\varepsilon_1 x_k^{n-1} + \varepsilon_2 x_k^{n-2} + \dots + \varepsilon_{n-1} x_k + \varepsilon_n).$$

Ako pretpostavimo da su granice apsolutnih grešaka koeficijenata jednake, tj. da je $|\varepsilon_i| \leq \varepsilon$, na osnovu poslednje približne jednakosti dobijamo granicu apsolutne greške korena, u oznaci Δ ,

$$\Delta \approx \frac{1}{|P'(x_k)|} (|x_k|^{n-1} + |x_k|^{n-2} + \dots + |x_k| + 1) \varepsilon,$$

tj.

$$\Delta = \begin{cases} \frac{(|x_k|^n - 1) \varepsilon}{|P'(x_k)| (|x_k| - 1)} & (|x_k| \neq 1), \\ \frac{n \varepsilon}{|P'(x_k)|} & (|x_k| = 1). \end{cases}$$

U radu [101], J.H.Wilkinson navodi primer jednačine

$$P(x) = (x+1)(x+2)\dots(x+20) = 0,$$

tj.

$$P(x) = x^{20} + 210x^{19} + \dots + 20! = 0,$$

kod koje male promene u koeficijentima izazivaju velike promene u korenima. Naime, ako koeficijent $a_1 = 210$ promenimo za $\epsilon_1 = 2^{-23} \approx 1.2 \cdot 10^{-7}$, promene u malim korenima su neznatne, ali koren $x_{20} = -20$ postaje -20.8 . Takođe, pet pari korena postaju konjugovano kompleksni. Ovo je primer tzv. slabo uslovljenog polinoma.

Pri sukcesivnom određivanju korena jednačine (3.1.1) i njenoj redukciji na niži stepen, preporučuje se određivanje korena polazeći od najmanjeg po modulu.

Poseban problem kod rešavanja algebarskih jednačina je određivanje 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 formiramo jednačinu

$$R(x) = \frac{P(x)}{Q(x)} = 0.$$

Iz teorije polinoma je poznato (videti, na primer [56]) 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 određivanje najvećeg zajedničkog delioca dva polinoma koristi se Euklidov algoritam (videti [56]). Napomenimo da je najveći zajednički delilac jedinstven sa tačnošću do jedne mnoštvene konstante.

Primer 3.1.1. Neka je dat polinom

$$P(x) = x^5 + 10x^4 + 42x^3 + 92x^2 + 105x + 50 .$$

S obzirom da je najveći zajednički delilac za P i P' polinom Q , definisan sa $Q(x) = x^2 + 4x + 5$, imamo

$$R(x) = x^3 + 6x^2 + 13x + 10 .$$

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

$$(3.2.1) \quad \frac{|a_n|}{a' + |a_n|} < |z| < 1 + \frac{a}{|a_0|} .$$

Dokaz. Pretpostavimo da je $|z| > 1$. Kako je

$$|P(z)| \geq \left| |a_0 z^n| - |a_1 z^{n-1}| - \dots - |a_{n-1} z| - |a_n| \right| ,$$

i

$$\begin{aligned} |a_1 z^{n-1} + \dots + a_{n-1} z + a_n| &\leq a (|z|^{n-1} + |z|^{n-2} + \dots + |z| + 1) \\ &= a \frac{|z|^{n-1}}{|z| - 1} < \frac{a |z|^n}{|z| - 1}, \end{aligned}$$

zaključujemo da je $|P(z)| > 0$, 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 ove jednačine zadovoljavaju suprotnu nejednakost, tj. leže u unutrašnjosti kruga

$$|z| = 1 + \frac{a}{|a_0|} .$$

Uvodjenjem smene $z = 1/\zeta$, jednačina (3.1.1) postaje

$$(3.2.2) \quad a_0 + a_1 \zeta + \dots + a_{n-1} \zeta^{n-1} + a_n \zeta^n = 0.$$

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.

Primer 3.2.1. Ispitajmo u kojoj oblasti leže koreni jednačine

$$\begin{aligned} z^7 - (3+i)z^6 - (15-3i)z^5 + (17+15i)z^4 + (38-17i)z^3 \\ + (22-38i)z^2 - (60+22i)z + 60i = 0. \end{aligned}$$

Kako je

$$a = \max(|a_1|, |a_2|, \dots, |a_7|) = |a_6| = |-(60+22i)| \approx 63.9,$$

$$a' = \max(|a_0|, |a_1|, \dots, |a_6|) = |a_6| \approx 63.9,$$

imamo

$$\frac{60}{63.9 + 60} < |z| < 1 + \frac{63.9}{1},$$

$$\text{tj. } 0.48 < |z| < 64.9.$$

Posmatrajmo sada jednačinu

$$(3.2.3) \quad P(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n = 0 \quad (a_0 \neq 0)$$

sa realnim koeficijentima i $a_0 > 0$.

Sledeća teorema daje granicu realnih korena ove jednačine:

Teorema 3.2.2. 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, \dots, 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) negativni, 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

$$(3.2.4) \quad 2x^9 + x^7 - x^4 + 19x^3 - 24x^2 + 11 = 0.$$

Kako je $a_0 = 2$, $A = 24$, $m = 5$, imamo

$$R = 1 + \sqrt[5]{\frac{24}{2}} < 2.65.$$

Za određivanje 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 $r = 1/\bar{R} > 0.40$. 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

$$P^{(k)}(a) \geq 0 \quad (k = 0, 1, \dots, n),$$

za gornju granicu korena može se uzeti vrednost $R = a$, što sleduje iz Taylorove formule

$$P(x) = \sum_{k=0}^n \frac{1}{k!} P^{(k)}(a) (x-a)^k.$$

Naime, za $x > a$ imamo $P(x) > 0$.

Sledećim postupkom može se dobiti bolja granica. Predstavimo $P(x)$ u obliku

$$P(x) = Q_1(x) - Q_2(x) + Q_3(x) - \dots + Q_{2m-1}(x) - Q_{2m}(x),$$

gde su $Q_1(x)$ suma prvih članova $P(x)$ sa pozitivnim koeficijentima, počev sa $a_0 x^n$, $-Q_2(x)$ 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 ($j = 1, \dots, m$) pozitivni brojevi takvi da je

$$Q_{2j-1}(c_j) - Q_{2j}(c_j) > 0 \quad (j=1, \dots, m),$$

za gornju granicu pozitivnih korena može se uzeti broj

$$R = \max(c_1, \dots, c_m).$$

Navedimo sada jedan rezultat, koji može biti korisniji od teoreme 3.2.1.

Neka je

$$P(z) = z^n + a_1 z^{n-1} + a_2 z^{n-2} + \dots + a_{n-1} z + a_n.$$

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,

$$Q(w) = P\left(w - \frac{a_1}{n}\right) = w^n + c_2 w^{n-2} + \dots + c_{n-1} w + c_n.$$

Definišimo polinom S pomoću

$$S(w) = w^n - |c_2|w^{n-2} - \dots - |c_{n-1}|w - |c_n|.$$

Teorema 3.2.3. Ako je bar jedan od koeficijenata $c_k \neq 0$ ($k = 2, \dots, n$), sve nule polinoma P leže u oblasti

$$\left|z + \frac{a_1}{n}\right| \leq r,$$

gde je r pozitivna nula polinoma S .

Dokaz. Neka je $|w| > r$. Pod pretpostavkom da je bar jedan od koeficijenata $c_k \neq 0$ ($k = 2, \dots, n$) i s obzirom da je $S(r) = 0$, imamo

$$\left| \frac{c_2}{w^2} + \dots + \frac{c_n}{w^n} \right| < \frac{|c_2|}{r^2} + \dots + \frac{|c_n|}{r^n} = 1.$$

Kako je

$$Q(w) = w^n \left(1 + \frac{c_2}{w^2} + \dots + \frac{c_n}{w^n}\right),$$

na osnovu prethodnog imamo

$$|Q(w)| \geq |w|^n \left(1 - \left| \sum_{i=2}^n \frac{c_i}{w^i} \right| \right) > 0,$$

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 $|w| \leq r$, 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

$$(3.2.5) \quad P(z) = z^5 - 10z^4 + 43z^3 - 104z^2 + 150z - 100.$$

$$\text{Smenom } z = w - \frac{(-10)}{5} = w + 2 \text{ dobijamo}$$

$$Q(w) = P(w+2) = w^5 + 3w^3 - 6w^2 + 10w$$

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 $R = 1 + \sqrt{10} \approx 4.16$. 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 određivanje 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

$$(3.3.1) \quad a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n = 0 \quad (a_0 \neq 0)$$

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

$$(3.3.2) \quad y_{n+k} + a_1 y_{n+k-1} + \dots + a_{n-1} y_{k+1} + a_n y_k = 0.$$

Rešenje ove jednačine odredjeno je korenima karakteristične jednačine (3.3.1). Naime, ako su x_i ($i = 1, \dots, n$) korenji jednačine (3.3.1), tada je

$$y_k = c_1 \phi_1(k) + \dots + c_n \phi_n(k)$$

rešenje diferencne jednačine (3.3.2), gde su funkcije ϕ_i određene 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 dominantnim 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

$$(3.3.3) \quad y_k = C_1 x_1^k + C_2 \phi_2(k) + \dots + C_n \phi_n(k).$$

Nije teško pokazati da je

$$(3.3.4) \quad \lim_{k \rightarrow +\infty} \frac{\phi_i(k)}{x_1^k} = 0 \quad (i = 2, \dots, n).$$

Korišćenjem niza $\{y_k\}$ konstruišemo niz $\{u_k\}$ pomoću

$$u_k = \frac{y_{k+1}}{y_k}. \quad \text{Tada je, s obzirom na (3.3.3),}$$

$$u_k = x_1 \frac{1 + \frac{c_2}{c_1} \cdot \frac{\phi_2(k+1)}{x_1^{k+1}} + \dots + \frac{c_n}{c_1} \cdot \frac{\phi_n(k+1)}{x_1^{k+1}}}{1 + \frac{c_2}{c_1} \cdot \frac{\phi_2(k)}{x_1^k} + \dots + \frac{c_n}{c_1} \cdot \frac{\phi_n(k)}{x_1^k}},$$

pri čemu smo pretpostavili da je $c_1 \neq 0$.

Na osnovu (3.3.4) imamo

$$(3.3.5) \quad \lim_{k \rightarrow +\infty} u_k = x_1,$$

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, $\phi_2(k) = x_2^k$, za dovoljno veliko k imamo

$$e_k = \frac{y_{k+1}}{y_k} - x_1 \underset{k \rightarrow +\infty}{\approx} \frac{c_2}{c_1} (x_2 - x_1) \left(\frac{x_2}{x_1} \right)^k,$$

tj. $e_{k+1} \underset{k \rightarrow +\infty}{\approx} \frac{x_2}{x_1} e_k$, odakle zaključujemo da je konvergencija linearna.

U slučaju 2.1, umesto rešenja (3.3.3) imamo rešenje

$$y_k = (c_1 + c_2 k + \dots + c_r k^{r-1}) x_1^r + c_{r+1} \phi_{r+1}(k) + \dots + c_n \phi_n(k).$$

Može se pokazati da i u ovom slučaju važi (3.3.5).

Primer 3.3.1. Neka je data jednačina

$$(3.3.6) \quad x^4 - x^3 - 3x^2 - 7x - 6 = 0.$$

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 $y_0 = y_1 = y_2 = 0$, $y_3 = 1$, 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

$$y_k = [c_1 + c_2 k + \dots + c_p k^{p-1} + (-1)^k (c_{p+1} + \dots + c_r k^{r-p-1})] x_1^k + c_{r+1} \phi_{r+1}(k) + \dots + c_n \phi_n(k),$$

odakle lako možemo zaključiti da niz $\{u_k\}$, gde je $u_k = \frac{y_{k+1}}{y_k}$, divergira. Moguće je, međutim, i u ovom slučaju odrediti x_1 , ali pomoću niza $\{v_k\}$, gde je $v_k = \frac{y_{2k+2}}{y_{2k}}$. Naime, ovde je

$$\lim_{k \rightarrow +\infty} v_k = x_1^2.$$

Primer 3.3.2. Posmatrajmo jednačinu $x^3 + 0.5x^2 - 4x - 2 = 0$, čiji su koreni $x_1 = -x_2 = 2$, $x_3 = -0.5$. Startujući sa $y_0 = y_1 = 0$,

$y_2 = 1$, 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

$$\begin{aligned} C_1 \phi_1(k) + C_2 \phi_2(k) &= \rho^k [A_1 \cos k\theta + A_2 \sin k\theta] \\ &= A\rho^k \cos(k\theta + \psi), \end{aligned}$$

pri čemu postoji odredjena veza izmedju konstanti koje se pojavljuju u ovoj jednakosti.

Štaviše, za dovoljno veliko k , važi

$$(3.3.7) \quad y_k = A\rho^k \cos(k\theta + \psi) + O(|x_3|^k).$$

Lako je videti da ranije definisani nizovi $\{u_k\}$ i $\{v_k\}$, u ovom slučaju divergiraju. Zato definji o nove nizove $\{s_k\}$ i $\{t_k\}$ pomocu

$$s_k = \begin{vmatrix} y_k & & & \\ & \ddots & & \\ & & y_{k+1} & \\ y_{k-1} & & y_k & \end{vmatrix} = y_k^2 - y_{k-1} y_{k+1},$$

$$t_k = \begin{vmatrix} y_{k+1} & y_{k+2} \\ y_{k-1} & y_k \end{vmatrix} = y_{k+1}y_k - y_{k-1}y_{k+2}.$$

Na osnovu (3.3.7), za dovoljno veliko k , imamo

$$s_k \cong A^2 \rho^{2k} [\cos^2(k\theta + \psi) - \cos((k-1)\theta + \psi) \cos((k+1)\theta + \psi)]$$

i

$$t_k \cong A^2 \rho^{2k+1} [\cos((k+1)\theta + \psi) \cos(k\theta + \psi) - \cos((k-1)\theta + \psi) \cos((k+2)\theta + \psi)],$$

tj.

$$s_k \cong A^2 \rho^{2k} \sin^2 \theta \quad i \quad t_k \cong 2A^2 \rho^{2k+1} \sin^2 \theta \cos \theta,$$

odakle zaključujemo da je

$$\lim_{k \rightarrow +\infty} \frac{s_{k+1}}{s_k} = \rho^2 \quad i \quad \lim_{k \rightarrow +\infty} \frac{t_k}{2s_k} = \rho \cos \theta.$$

Dakle, u posmatranom slučaju, niz $\{\frac{s_{k+1}}{s_k}\}$ konvergira ka ρ^2 , a niz $\{\frac{t_k}{2s_k}\}$ 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

$$y_{k+3} = -y_{k+2} - y_{k+1} + y_k \quad (k = 0, 1, \dots),$$

konstruišimo niz $\{y_k\}$. Dakle, imamo

$$\{y_k\} = \{0, 0, 1, -1, 0, 2, -3, 1, 4, -2, -1, 7, -8, 0, 15, -23, 8, 30, -61, 39, 52, \dots\}.$$

Kako je $s_{18} = (-61)^2 - 39 \cdot 30 = 2551$, $s_{19} = 39^2 - (-61) \cdot 52 = 4693$, $t_{18} = 39 \cdot (-61) - 52 \cdot 30 = -3939$ nalazimo

$$\frac{s_{19}}{s_{18}} \cong 1.8397 \quad i \quad \frac{t_{18}}{2s_{18}} \cong -0.7720,$$

odakle je

$$\begin{aligned}x_1 &\approx -0.7720 + (1.8397 - (-0.7720)^2)^{1/2} i \\&\approx -0.7720 + 1.1152 i\end{aligned}$$

$$i \ x_2 = \bar{x}.$$

Slučaj tri dominantna korena, od kojih je jedan realan, a dva konjugovano kompleksna, razmatran je u radu [48].

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 ξ_i ($i = 1, \dots, n$) polinoma

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

realne i različite i neka je $\xi_1 < \xi_2 < \dots < \xi_n$. Tada, na osnovu Rolleove teoreme, izvodni polinom P' ima $n-1$ nula koje su, takođe, realne i međusobno različite. Označimo ih redom sa $\eta_1, \dots, \eta_{n-1}$, tako da je

$$\xi_1 < \eta_1 < \xi_2 < \dots < \eta_{n-1} < \xi_n.$$

Neka je $X = \mathbb{R} \setminus \{\xi_1, \eta_1, \xi_2, \dots, \eta_{n-1}, \xi_n\}$. Definišimo funkciju $a: X \rightarrow \{\xi_1, \xi_2, \dots, \xi_n\}$ pomoću

$$a(x) = \begin{cases} \xi_1 & (x < \xi_1) \\ \xi_1 & (\xi_i < x < \eta_i) \\ \xi_{i+1} & (\eta_i < x < \xi_{i+1}) \\ \xi_n & (x > \xi_n) \end{cases} \quad (i = 1, \dots, n-1).$$

Primetimo da u svakom intervalu izmedju $a(x)$ i x , izraz $P(x)P'(x)$ ne menja znak. Naime, važi

$$(3.4.1) \qquad \text{sgn}(x-a(x)) = \text{sgn}(P(x)P'(x)).$$

Kako je $P(x) = \prod_{i=1}^n (x - \xi_i)$, imamo

$$(3.4.2) \quad \frac{P'(x)}{P(x)} = \sum_{i=1}^n \frac{1}{x - \xi_i}, \quad H(x) = \frac{P'(x)^2 - P(x)P''(x)}{P(x)^2} = \sum_{i=1}^n \frac{1}{(x - \xi_i)^2}.$$

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

$$(3.4.3) \quad K(x) = \frac{P(x)/P'(x)}{\sqrt{1 - P(x)P''(x)/P'(x)^2}}.$$

Uzimajući $x_0 = x$ ($x \in X$) konstruišimo iterativni proces

$$(3.4.4) \quad x_{k+1} = x_k - K(x_k) \quad (k = 0, 1, \dots),$$

koji je u literaturi poznat kao metod kvadratnog korena ([63]).

Teorema 3.4.1. Niz $\{x_k\}_{k \in \mathbb{N}_0}$ koji se generiše pomoću (3.4.4) konvergira monotono ka nuli $a = a(x)$, pri čemu je konvergencija kubna, tj. važi

$$(3.4.5) \quad \lim_{k \rightarrow +\infty} \frac{x_{k+1} - a}{(x_k - a)^3} = \frac{3P''(a)^2 - 4P'(a)P'''(a)}{24P'(a)^2}.$$

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 $a(x_0) = a(x_1)$. Nastavljujući sa ovakvim rezul-

novanjem dokazujemo da je

$$x_0 < x_1 < x_2 < \dots < a(x),$$

što znači da niz $\{x_k\}_{k \in \mathbb{N}}$ monotono konvergira ka nekoj tački $a \leq a(x)$. Kako je, u tom slučaju, $\lim_{k \rightarrow +\infty} (x_{k+1} - x_k) = 0$, 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 $h = \frac{P(x)}{P'(x)}$ i pustimo da $x \rightarrow a$. Na osnovu (3.4.3) i (3.4.4) imamo

$$x_1 - x = -K(x) = h \left(1 + h \frac{P''(x)}{P'(x)}\right)^{-1/2},$$

tj.

$$x_1 - x = h - \frac{1}{2} h^2 \frac{P''(x)}{P'(x)} + \frac{3}{8} h^3 \left(\frac{P''(x)}{P'(x)}\right)^2 + O(h^4).$$

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 $h \sim a - x (x \rightarrow a)$ sleduje (3.4.5), uz prethodnu zamenu x sa $x_k (k \rightarrow +\infty)$.

Ovim je dokaz teoreme završen.

Dakle, metod kvadratnog korena ima kubnu konvergenciju. Međutim, može se razmatrati slučaj kada je a višestruka nula reda p; tada je konvergencija prvog reda i važi

$$\lim_{k \rightarrow +\infty} \frac{x_{k+1} - a}{x_k - a} = 1 - \frac{1}{\sqrt{p}}.$$

U tesnoj vezi sa metodom kvadratnog korena je Laguerreov metod

$$(3.4.6) \quad x_{k+1} = x_k - \frac{n P(x_k)}{P'(x_k) + \operatorname{sgn}(P'(x_k)) \sqrt{G(x_k)}} \quad (k = 0, 1, \dots),$$

gde je $x_0 = x$ ($x \in X$) i

$$G(t) = (n-1) [(n-1)P'(t)^2 - nP(t)P''(t)].$$

Metod (3.4.6) se može predstaviti i u obliku

$$x_{k+1} = x_k - \frac{n P(x_k) / P'(x_k)}{1 + (n-1) \sqrt{1 - \frac{n}{n-1} P(x_k) P''(x_k) / P'(x_k)^2}} \quad (k = 0, 1, \dots).$$

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 $3(\text{uz } P''(a))^2$ na desnoj strani ove formule, treba zameniti konstantom $3(n-2)/(n-1)$. Laguerreov metod je efikasniji od metoda kvadratnog korena, na što 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 $x_0 = 0$, dobijamo

	pomoću (3.4.4)	pomoću (3.4.6)
x_1	0.85714286	0.98840803
x_2	0.99858769	0.99999991

U radu [30] 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 određivanje nula polinoma

$$(3.5.1) \quad P(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n \quad (a_i \in \mathbb{R}),$$

pri čemu se vrednosti za $P(x_k)$ i $P'(x_k)$ izračunavaju po Hornerovoј šemi (videti odeljak 1.3.3).

Ako stavimo $p_j(x) = \frac{1}{j!} P^{(j)}(x)$ ($j = 0, 1, \dots, n$), tada je, na osnovu Taylorove formule,

$$P(x) = p_0(x_k) + p_1(x_k)(x-x_k) + \dots + p_n(x_k)(x-x_k)^n.$$

Pretpostavimo da polinom P ima realnu nulu $x = \xi$. Newtonov metod (1.2.3) za određivanje ove nule postaje

$$(3.5.2) \quad x_{k+1} = x_k - \frac{p_0(x_k)}{p_1(x_k)} \quad (k = 0, 1, \dots).$$

Kao što je poznato, vrednost za $p_0(x_k)$ može se dobiti iz n koraka Hornerovom šemom, kao $p_0(x_k) = b_n$, gde je

$$(3.5.3) \quad b_0 = a_0, \quad b_j = b_{j-1} x_k + a_j \quad (j = 1, \dots, n).$$

Ustvari b_j ($j = 0, 1, \dots, n-1$) su koeficijenti polinoma dobijenog deljenjem $P(x)$ sa $x - x_k$, tj. važi

$$P(x) = p_1(x)(x-x_k) + p_0(x_k)$$

i

$$p_1(x) = b_0 x^{n-1} + b_1 x^{n-2} + \dots + b_{n-1}.$$

Slično deljenjem $p_1(x)$ sa $x - x_k$ dobijamo

$$p_2(x) = c_0 x^{n-2} + \dots + c_{n-2},$$

tj. važi

$$p_1(x) = p_2(x)(x-x_k) + p_1(x_k).$$

Ovde je $c_0 = b_0$, $c_j = c_{j-1}x_k + b_j$ ($j = 1, \dots, n-1$) i $p_1(x_k) = c_{n-1}$.

Navedena deljenja se šematski mogu prikazati u obliku

x_k	a_0	a_1	\dots	a_j	\dots	a_{n-1}	a_n	
	$b_0 x_k$			$b_{j-1} x_k$		$b_{n-2} x_k$	$b_{n-1} x_k$	
x_k	b_0	b_1		b_j		b_{n-1}		$b_n = p_0(x_k)$
		$c_0 x_k$		$c_{j-1} x_k$		$c_{n-2} x_k$		
	c_0	c_1		c_j				$c_{n-1} = p_1(x_0)$

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) moguće je po istom postupku kao što je (3.5.3), s tim što b_j treba zameniti sa $\alpha_j + i\beta_j$ ($j = 0, 1, \dots, n$). Tada je

$$\alpha_0 = a_0, \quad \beta_0 = 0,$$

$$\alpha_j = \alpha_{j-1}x_k - \beta_{j-1}y_k + a_j, \quad \beta_j = \alpha_{j-1}y_k + \beta_{j-1}x_k \quad (j = 1, \dots, n),$$

$$u_k = u(x_k, y_k) = \alpha_n, \quad v_k = v(x_k, y_k) = \beta_n.$$

Neka je $z = \xi + i\eta$ koren jednačine $P(z) = 0$. S obzirom da je ova jednačina ekvivalentna sistemu jednačina

$$(3.5.4) \quad u(x, y) = 0, \quad v(x, y) = 0,$$

to se za određivanje korena $z = \xi + i\eta$ može konstruisati niz $z_k = x_k + iy_k$ ($k = 0, 1, \dots$), na primer, pomoću metoda Newton-Kantorovića. Imajući u vidu da funkcije u i v zadovoljavaju Cauchy-Riemannove uslove, lako se pokazuje da je metod Newton-Kantorovića, u ovom slučaju, ekvivalentan sa kompleksnom verzijom Newtonovog metoda

$$(3.5.5) \quad z_{k+1} = z_k - \frac{P(z_k)}{P'(z_k)} \quad (k = 0, 1, \dots).$$

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 rekursivno pomoću

$$\gamma_0 = \alpha_0 = a_0, \quad \delta_0 = 0,$$

$$\gamma_j = \gamma_{j-1}x_k - \delta_{j-1}y_k + \alpha_j, \quad \delta_j = \gamma_{j-1}y_k + \delta_{j-1}x_k + \beta_j \quad (j = 1, \dots, n-1).$$

Naime, $u'_k = \gamma_{n-1}$ i $v'_k = \delta_{n-1}$.

Primer 3.5.1. Neka je $P(z) = z^3 + 2z^2 + z - 1$ i $z = 1+i$. Na osnovu prethodno datih jednakosti imamo

$$\alpha_0 = 1, \quad \beta_0 = 0,$$

$$\alpha_1 = 1 \cdot 1 - 0 \cdot 1 + 3 = 4, \quad \beta_1 = 1 \cdot 1 + 0 \cdot 1 = 1,$$

$$\alpha_2 = 4 \cdot 1 - 1 \cdot 1 + 1 = 4, \quad \beta_2 = 4 \cdot 1 + 1 \cdot 1 = 5,$$

$$\alpha_3 = 4 \cdot 1 - 5 \cdot 1 - 1 = -2, \quad \beta_3 = 4 \cdot 1 + 5 \cdot 1 = 9$$

i

$$\gamma_0 = 1, \quad \delta_0 = 0,$$

$$\gamma_1 = 1 \cdot 1 - 0 \cdot 1 + 4 = 5, \quad \delta_1 = 1 \cdot 1 + 0 \cdot 1 + 1 = 2$$

$$\gamma_2 = 5 \cdot 1 - 2 \cdot 1 + 4 = 7, \quad \delta_2 = 5 \cdot 1 + 2 \cdot 1 + 5 = 12,$$

pa je $P(z) = -2 + 9i$, $P'(z) = 7 + 12i$.

Korišćenjem prethodno datih rekurentnih relacija, formula (3.5.5) se svodi na

$$(3.5.6) \quad \left\{ \begin{array}{l} x_{k+1} = x_k - \frac{u_k u'_k + v_k v'_k}{(u'_k)^2 + (v'_k)^2} \\ y_{k+1} = y_k - \frac{v_k u'_k - u_k v'_k}{(u'_k)^2 + (v'_k)^2} \end{array} \right. \quad (k = 0, 1, \dots).$$

Kod primene Newton-Hornerovog metoda obično se startuje sa dovoljno malim vrednostima (po modulu) x_0 i y_0 i primenom formule (3.5.6) određuje se najmanja nula po modulu $z_1 = \xi_1 + i n_1$. Ako je $|n_1| < \epsilon$ (ϵ dovoljno mali pozitivni broj, na primer $\epsilon = 10^{-7}$) 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 određena i nula $z_2 = \xi_1 - i n_1$. Ponavljamajući izloženi postupak na polinomu $P(z)/(z - \xi_1)$, odnosno $P(z)/((z - \xi_1)^2 + n_1^2)$ određuju 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 određivanje kvadratnih faktora (delilaca) polinoma, na osnovu kojih se mogu odrediti kompleksne nule polinoma.

Neka je

$$P(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n \quad (a_0 \neq 0)$$

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 određivanja jednog takvog faktora $\hat{m}(x) = x^2 + \hat{p}x + \hat{q}$, pretpostavimo da nam je poznata izvesna aproksimacija $m_0(x) = x^2 + p_0x + q_0$ ($p_0, q_0 \in \mathbb{R}$).

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

$$(3.6.1) \quad P(x) = m(x)Q(x) + R(x)$$

dati sa

$$Q(x) = b_2x^{n-2} + \dots + b_{n-1}x + b_n \quad \text{i} \quad R(x) = ux + v.$$

Uporedjivanjem koeficijenata uz odgovarajuće stepene na levoj i desnoj strani jednakosti (3.6.1) dobijamo

$$a_0 = b_2, \quad a_1 = b_3 + pb_2,$$

$$a_{i-2} = b_i + pb_{i-1} + qb_{i-2} \quad (i = 4, \dots, n),$$

$$a_{n-1} = pb_n + qb_{n-1} + u, \quad a_n = qb_n + v.$$

Ako stavimo $b_0 = b_1 = 0$, tada iz prethodnih jednakosti sleduje

$$(3.6.2) \quad b_i = a_{i-2} - pb_{i-1} - qb_{i-2} \quad (i = 2, \dots, n)$$

i

$$(3.6.3) \quad u = a_{n-1} - pb_n - qb_{n-1} = b_{n+1}, \quad v = a_n - qb_n,$$

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 .

* Kvadratni faktori sa realnim koeficijentima.

Dakle, za određivanje 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_0, q_0) generiše se niz parova $\{(p_k, q_k)\}_{k=1,2,\dots}$ pomoću

$$(3.6.4) \quad \begin{bmatrix} p_{k+1} \\ q_{k+1} \end{bmatrix} = \begin{bmatrix} p_k \\ q_k \end{bmatrix} - W_k^{-1} \begin{bmatrix} u(p_k, q_k) \\ v(p_k, q_k) \end{bmatrix} \quad (k = 0, 1, \dots),$$

gde je Jacobieva matrica data sa

$$W_k = \begin{bmatrix} \frac{\partial u}{\partial p} & \frac{\partial u}{\partial q} \\ \frac{\partial v}{\partial p} & \frac{\partial v}{\partial q} \end{bmatrix} \quad p=p_k, \quad q=q_k$$

Da bismo odredili elemente ove matrice uvedimo oznake

$$r_i = \frac{\partial b_i}{\partial p}, \quad t_i = \frac{\partial b_i}{\partial q} \quad (i = 0, 1, \dots, n+1)$$

i

$$s_i = r_i - t_{i+1} \quad (i = 0, 1, \dots, n).$$

Diferenciranjem po p i q prethodno dokazanih rekurentnih relacija za koeficijente b_i , dobijamo

$$(3.6.5) \quad r_i = -b_{i-1} - p \frac{\partial b_{i-1}}{\partial p} - q \frac{\partial b_{i-2}}{\partial p} = -b_{i-1} - p r_{i-1} - q r_{i-2}$$

i

$$(3.6.6) \quad t_i = -p \frac{\partial b_{i-1}}{\partial q} - b_{i-2} - q \frac{\partial b_{i-2}}{\partial q} = -p t_{i-1} - b_{i-2} - q t_{i-2},$$

gde je $i = 2, \dots, n+1$. Ako od jednakosti (3.6.5) oduzmemos jednakošć (3.6.6) u kojoj je indeks i zamenjen sa $i+1$, dobijamo

$$(3.6.7) \quad s_i = -ps_{i-1} - qs_{i-2} \quad (i = 2, \dots, n).$$

Kako je $b_0 = b_1 = 0$ i $b_2 = a_0$, imamo $r_0 = r_1 = r_2 = 0$, $t_0 = t_1 = t_2 = 0$, $s_0 = s_1 = 0$, što zajedno sa (3.6.7) daje

$$s_i = r_i - t_{i+1} = 0, \text{ tj. } \frac{\partial b_{i+1}}{\partial q} = \frac{\partial b_i}{\partial p} \quad (i = 0, 1, \dots, n).$$

Kako je na osnovu prethodnog

$$w_k = \begin{bmatrix} r_{n+1} & r_n \\ \vdots & \\ -qr_n & -(b_n + qr_{n-1}) \end{bmatrix} \quad p=p_k, q=q_k$$

iterativni proces (3.6.4) postaje

$$\begin{bmatrix} p_{k+1} \\ q_{k+1} \end{bmatrix} = \begin{bmatrix} p_k \\ q_k \end{bmatrix} - \frac{1}{D} \begin{bmatrix} -(b_n + qr_{n-1})b_{n+1} - r_n(a_n - qb_n) \\ qr_n b_{n+1} + r_{n+1}(a_n - qb_n) \end{bmatrix} \quad p=p_k, q=q_k$$

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-Kantoroviča.

Jedna opštija klasa metoda Bairstowljevog tipa razmatrana je u radu [9]. Naime, pokazano je da je Bairstowljev metod samo jedan od metoda iz čitave familije algoritama za određivanje kvadratnih faktora polinoma. U radu se sugerise izbor pogodnog metoda iz ove familije za rešavanje konkretnog problema.

Kao i Newton-Hornerov metod, tako i Bairstowljev metod se često koristi u primenama.

5.3.7. Metodi za simultano odredjivanje korena

U novije vreme sve više se proučavaju metodi za simultano (istovremeno) odredjivanje svih korena algebarske jednačine

$$(3.7.1) \quad P(z) = z^n + a_1 z^{n-1} + \dots + a_{n-1} z + a_n = 0,$$

gde su a_i ($i = 1, \dots, n$), u opštem slučaju, kompleksni koeficijenti.

Prepostavimo da su koreni r_1, \dots, r_n algebarske jednačine (3.7.1) međusobno različiti i neka su njihove približne vrednosti u k -tom iterativnom koraku $z_i^{(k)}$ ($k = 1, \dots, n$). S obzirom da je $r_i = z_i^{(k)} + \Delta z_i^{(k)}$ ($k = 1, \dots, n$), 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

$$\begin{aligned} & \prod_{i=1}^n (z - z_i^{(k)}) - \sum_{i=1}^n \Delta z_i^{(k)} \left(\prod_{\substack{m=1 \\ m \neq i}}^n (z - z_m^{(k)}) \right) \\ & + \sum_{i,j} \Delta z_i^{(k)} \Delta z_j^{(k)} \left(\prod_{\substack{m=1 \\ m \neq i, j}}^n (z - z_m^{(k)}) \right) - \dots = P(z). \end{aligned}$$

Prepostavljajući da su greške $\Delta z_i^{(k)}$ dovoljno male po modulu, na levoj strani u poslednjoj jednakosti možemo uzeti samo prva dva člana. Ako u tako dobijenoj jednakosti stavimo redom $z := z_i^{(k)}$ ($i = 1, \dots, n$), dobijamo

$$(3.7.2) \quad \Delta z_i^{(k)} = - \frac{P(z_i^{(k)})}{\prod_{\substack{m=1 \\ m \neq i}}^n (z_i^{(k)} - z_m^{(k)})} \quad (i = 1, \dots, n).$$

Na ovaj način dolazimo do simultanog iterativnog procesa

$$(3.7.3) \quad z_i^{(k+1)} = z_i^{(k)} + \Delta z_i^{(k)} \quad (i = 1, \dots, n; k = 0, 1, \dots).$$

Ako definišemo polinom n -toga stepena $Q_k(z) = \prod_{m=1}^n (z - z_m^{(k)})$, tada se (3.7.2) može predstaviti u obliku koji podseća na Newton-

ovu korekciju, tj.

$$z_i^{(k)} = - \frac{P(z_i^{(k)})}{Q'_k(z_i^{(k)})} \quad (i = 1, \dots, n).$$

Primetimo da $Q_k(z) \rightarrow P(z)$, kada $z_i^{(k)} \rightarrow r_i$ ($i=1, \dots, n$). U novo uvedenoj notaciji, iterativni proces (3.7.3) postaje

$$(3.7.4) \quad z_i^{(k+1)} = z_i^{(k)} - \frac{P(z_i^{(k)})}{Q'_k(z_i^{(k)})} \quad (i = 1, \dots, n; k=0, 1, \dots).$$

Jedno kraće izvodjenje formula (3.7.4) može se dati polazeći od faktorizacije

$$P(z) = \prod_{m=1}^n (z - r_m) = (z - r_i) \prod_{\substack{m=1 \\ m \neq i}}^n (z - r_m),$$

tj. od identiteta koji iz njega sleduje

$$(3.7.5) \quad r_i = z - \frac{P(z)}{\prod_{\substack{m=1 \\ m \neq i}}^n (z - r_m)}.$$

Uzimajući za r_m ($m \neq i$) približne vrednosti $z_m^{(k)}$, tj. $r_m := z_m^{(k)}$ i $z := z_i^{(k)}$, desna strana u (3.7.5), naravno, neće biti jednaka tačnoj nuli r_i , već će to biti neka nova aproksimacija za tu nulu, koju ćemo označiti sa $z_i^{(k+1)}$. Tako dobijamo iterativni proces (3.7.4).

Formule (3.7.4) su otkrivane više puta od strane većeg broja autora. Zato se mogu i sresti više načina izvodjenja ovih formula. Primetimo da formule (3.7.4) podsećaju na klasičan Newtonov metod, ali se od njega suštinski razlikuju. Naime, ovde imamo n iterativnih formula koje generišu n nizova koji su medjusobno zavisni; ($k+1$)-vi član jednog niza zavisi k -tih članova svih nizova. Napomenimo da se ove formule mogu naći u sabranim delima Weierstrassa. Naime, Weierstrass ([99]) ih je koristio u vezi sa dokazom osnovne teoreme algebre. Medjutim, za određivanje nula ove formule se koriste u poslednjih dvadesetak godina.

Iterativni proces (3.7.4) ima kvadratnu konvergenciju. Naime, on je ekvivalentan metodu Newton-Kantoroviča, primjenjenog na rešavanje sistema nelinearnih jednačina, koji predstavljaju Viète-

ove formule (3.1.2) za polinom (3.7.1).

Ako uvedemo oznake

$$\vec{f}(\vec{z}) = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{bmatrix}, \quad \vec{a} = \begin{bmatrix} a_1 \\ -a_2 \\ \vdots \\ (-1)^{n-1} a_n \end{bmatrix}, \quad \vec{z}^{(k)} = \begin{bmatrix} z_1^{(k)} \\ z_2^{(k)} \\ \vdots \\ z_n^{(k)} \end{bmatrix},$$

sistem jednačina (3.1.2) se može predstaviti u obliku

$$\vec{F}(\vec{z}) = \vec{f}(\vec{z}) + \vec{a} = \vec{0}.$$

Odgovarajući metod Newton-Kantorovića je tada

$$(3.7.6) \quad \vec{z}^{(k+1)} = \vec{z}^{(k)} - \Gamma(\vec{z}^{(k)}) (\vec{f}(\vec{z}^{(k)}) + \vec{a}) \quad (k = 0, 1, \dots),$$

gde je $\Gamma(\vec{z}^{(k)})$ inverzna matrica Jacobievoj matrici $W(\vec{F}(\vec{z}^{(k)}))$.

Direktnim nalaženjem $\Gamma(\vec{z}^{(k)})$ može se pokazati da je (3.7.6) ekvivalentno sa (3.7.4) (videti [93]).

Da bi simultani metod (3.7.4) bio definisan neophodno je da sve startne vrednosti budu medjusobno različite, tj. $z_i^{(0)} \neq z_j^{(0)}$ ($i \neq j$). Može se pokazati (videti, na primer, [16]) da metod ima jedno interesantno svojstvo. Naime, bez obzira na izbor startnih vrednosti (sem da budu medjusobno različite), posle prve iteracije imamo da je

$$(3.7.7) \quad \sum_{i=1}^n z_i^{(1)} = -a_1,$$

gde je a_1 koeficijent u jednačini (3.7.1). Naravno, nije teško primetiti da je posle svake iteracije zbir aproksimacija $z_i^{(k)}$ jednak $-a_1$. Daćemo sada jedan interesantan dokaz jednakosti (3.7.7):

Kako su startne vrednosti medjusobno različite, zaključujemo da racionalna funkcija $z \mapsto P(z)/Q_O(z)$ ima samo proste polove. Neka je $r > \max_i |z_i^{(0)}|$. Tada je

$$I = \frac{1}{2\pi i} \oint_{|z|=r} \frac{P(z)}{Q_O(z)} dz = \sum_{i=1}^n \text{Res} \frac{P(z)}{Q_O(z)} = \sum_{i=1}^n \frac{P(z_i^{(0)})}{Q_O'(z_i^{(0)})}.$$

S druge strane, zamenom $w=1/z$, poslednji integral postaje

$$I = \frac{1}{2\pi i} \oint_{|w|=1/r} \frac{1}{w^2} \cdot \frac{P(1/w)}{Q_O(1/w)} dw = \lim_{w \rightarrow 0} \frac{d}{dw} \left\{ \frac{w^n P(1/w)}{w^n Q_O(1/w)} \right\},$$

tj.

$$I = a_1 + \sum_{i=1}^n z_i^{(o)}.$$

Sumiranjem (3.7.4) za $i = 1, \dots, n$, pri $k=0$, na osnovu prethodnog dobijamo

$$\sum_{i=1}^n z_i^{(1)} = \sum_{i=1}^n z_i^{(o)} - (a_1 + \sum_{i=1}^n z_i^{(o)}) = -a_1.$$

Navešćemo sada dve teoreme koje se odnose na konvergenciju metoda (3.7.4). Dokazi ovih rezultata mogu se dati matematičkom indukcijom (videti [16], [82]).

Teorema 3.7.1(K. Dočev). Ako je

$$d = \min_{i \neq j} |r_i - r_j| \quad i \quad |z_i^{(o)} - r_i| < \rho \quad (i = 1, \dots, n),$$

gde je

$$\rho = d \frac{(1+q)^{1/(n-1)} - 1}{2(1+q)^{1/(n-1)} - 1} \quad (0 < q < 1),$$

metod (3.7.4) konvergira, pri čemu je, za svako $k \in \mathbb{N}$,

$$|z_i^{(k)} - r_i| \leq \rho q^{2^{k-1}} \quad (i = 1, \dots, n).$$

Teorema 3.7.2(M.D. Prešić). Neka su početne vrednosti $z_i^{(o)}$ medju-sobno različiti kompleksni brojevi za koje važe nejednakosti

$$\frac{|P(z_i^{(o)})|}{|Q'_O(z_i^{(o)})|} \leq \frac{|Q'_O(z_i^{(o)})|}{6nM(0)^{n-2}} \quad (i = 1, \dots, n),$$

gde je $M(0) = \max_{i \neq j} |z_i^{(o)} - z_j^{(o)}|$, tada simultani iterativni proces

(3.7.4) konvergira, tj. $\lim_{k \rightarrow +\infty} z_i^{(k)} = r_i$ ($i = 1, \dots, n$). Štaviše, za

svako $i = 1, \dots, n$, svi članovi niza $\{z_i^{(k)}\}_{k \in \mathbb{N}}$ se nalaze u krugu

$$K_i = K(z_i^{(o)}) = \{ z \mid |z - z_i^{(1)}| \leq |z_i^{(1)} - z_i^{(o)}| \}.$$

U radovima [83] i [84] S.B. Prešić je razmatrao postupak za faktorizaciju polinoma, koji se u slučaju linearnih faktora svodi

na metod (3.7.4). U radu [80] (videti, takodje, [81]) M.D. Prešić je razmatrala metod za određivanje $m(< n)$ nula polinoma (3.7.1). Jedan drugačiji način izvodjenja ovog metoda i generalizaciju za slučaj višestrukih nula polinoma dao je G. Loizou ([45]).

Polazeći od iterativne formule (3.7.4) i disjunktnih realnih intervala $x_i^{(o)} = [x_i^{(o)} - \varepsilon_i^{(o)}, x_i^{(o)} + \varepsilon_i^{(o)}]$ koji sadrže nule r_i ($i=1, \dots, n$), J. Herzberger ([36], [37]) je formulisao intervalni metod za simultano određivanje nula polinoma P :

$$x_i^{(k+1)} = \left\{ x_i^{(k)} - \frac{P(x_i^{(k)})}{\prod_{\substack{m=1 \\ m \neq i}}^n (x_i^{(k)} - x_m^{(k)})} \right\} \cap x_i^{(k)} \quad (i=1, \dots, n; k=0, 1, \dots).$$

dokazujući da za svako $i=1, \dots, n$ i svako $m \in N_o$ važi:

$$(a) \quad x_i^{(k+1)} \subset x_i^{(k)}; \quad (b) \quad r_i \in x_i^{(k)};$$

$$(c) \quad \text{postoji konstanta } q (> 0) \text{ takva da je } \varepsilon^{(k+1)} \leq q(\varepsilon^{(k)})^2,$$

gde je $\varepsilon^{(k)} = \max_i \varepsilon_i^{(k)}$.

Opšti slučaj kompleksnih nula razmatrao je M.S. Petković ([65]) korišćenjem kompleksne kružne aritmetike*. Naime, neka su nule polinoma P izolovane u disjunktnim kružnim intervalima $z_i^{(o)} = \{z_i^{(o)}; \varepsilon_i^{(o)}\}$ ($i=1, \dots, n$) i neka je u (3.7.5) uzeto $z=z_i^{(o)}$. Imajući u vidu da $r_m \in z_m^{(o)}$ ($m=1, \dots, n$), na osnovu osobine inkluzivne izotonosti sleduje

$$r_i \in z_i^{(1)} = \left\{ z_i^{(o)} - \frac{P(z_i^{(o)})}{\prod_{\substack{m=1 \\ m \neq i}}^n (z_i^{(o)} - z_m^{(o)})} \right\} \cap z_i^{(o)}.$$

* Kompleksni kružni interval $Z = \{c; r\}$ je disk u kompleksnoj ravni sa centrom u tački c i poluprečnikom r , tj.

$$Z = \{ z \mid |z - c| \leq r \}.$$

Tačka a se predstavlja pomoću $\{a; 0\}$. Kompleksna kružna aritmetika je definisana nad kružnim intervalima $z_k = \{c_k; r_k\}$ ($k=1, 2$) pomoću

$$z_1 \pm z_2 = \{c_1 \pm c_2; r_1 \pm r_2\}, \quad z_1 z_2 = \{c_1 c_2; |c_2|r_1 + |c_1|r_2 + r_1 r_2\},$$

$$z_1 \div z_2 = z_1 z_2^{-1} \quad (0 \notin z_2), \quad z^{-1} = \{\bar{a}\bar{c}; \bar{a}r\} \quad (a = |c|^2 - r^2; 0 \notin z_2).$$

Za detalje videti [34], [64], [66].

Poslednja formula sugerije intervalni iterativni metod

$$(3.7.9) \quad z_i^{(k+1)} = \left\{ z_i^{(k)} - \frac{P(z_i^{(k)})}{\prod_{\substack{m=1 \\ m \neq i}}^n (z_i^{(k)} - z_m^{(k)})} \right\} \cap z_i^{(k)} \quad (i=1, \dots, n; k=0, 1, \dots).$$

Korišćenjem kompleksne kružne aritmetike, u pomenutom radu [65], iterativni metod (3.7.9) je sведен na oblik

$$z_i^{(k+1)} = \left\{ z_i^{(k)} - \frac{P(z_i^{(k)})}{\prod_{\substack{m=1 \\ m \neq i}}^n (z_i^{(k)} - z_m^{(k)})} ; \frac{a^{(k)} (n-1) |P(z_i^{(k)})| \varepsilon^{(k)}}{(\rho^{(k)})^n} \right\} \cap z_i^{(k)},$$

gde su

$$\varepsilon^{(k)} = \max_i \varepsilon_i^{(k)}, \quad \rho^{(k)} = \min_{i \neq j} \{ |z_i^{(k)} - z_j^{(k)}| - \varepsilon_j^{(k)} \}$$

i $a^{(k)}$ pozitivan broj za koji važi $a^{(k)} \geq (1 + \varepsilon^{(k)}) / \rho^{(k)})^{n-1}$. Kao posledica navedenog iterativnog procesa dobija se uopštenje metoda (3.7.8).

Pri praktičnoj realizaciji simultanih metoda najčešće se koristi Gauss-Seidelov pristup, tj. korišćenje aproksimacija dobijenih u istoj iteraciji. Ovakav postupak ne samo da ubrzava konvergenciju osnovnog metoda bez dodatnih izračunavanja, već je i pogodan sa stanovišta zauzeća memorijskog prostora. Analiza konvergencije ovako modifikovanih metoda je omogućena uvođenjem koncepta R-reda konvergencije (videti odeljak 3.2.5).

Osnovna modifikacija procesa (3.7.4) dobijena pomoću Gauss-Seidelovog pristupa je

$$(3.7.10) \quad z_i^{(k+1)} = z_i^{(k)} - \frac{P(z_i^{(k)})}{\prod_{\substack{m=1 \\ m \neq i}}^{i-1} (z_i^{(k)} - z_m^{(k+1)}) \prod_{m=i+1}^n (z_i^{(k)} - z_m^{(k)})} \quad (i=1, \dots, n).$$

Programska realizacija ove varijante je jednostavnija u odnosu na osnovni oblik (3.7.4). Novo izračunata aproksimacija $z_i^{(k)}$ se odmah smešta u istu poziciju gde je bila vrednost $z_i^{(k)}$, tako da po obavljenoj iteraciji (po svim nulama) nema zamene $z_i^{(k)} := z_i^{(k+1)}$ ($i = 1, \dots, n$).

Izložićemo sada još neke simultane metode koji se javljaju u primenama. Najpre, uvedimo izvesna označavanja koja omogućavaju bolju preglednost kod definisanja dve susedne iteracije:

1° Aproksimacije nula u k-toj iteraciji $z_1^{(k)}, \dots, z_n^{(k)}$ označavaćemo prosto sa z_1, \dots, z_n , a nove aproksimacije koje dobijamo primenom simultanog metoda redom sa $\hat{z}_1, \dots, \hat{z}_n$;

$$2^o \quad Q(z) = \prod_{m=1}^n (z - z_m);$$

3° $w_i = P(z_i)/Q'(z_i)$ (Weierstrassova korekcija, tj. korekcija kod metoda (3.7.4));

4° $N_i = P(z_i)/P'(z_i)$ (Newtonova korekcija).

U novo uvedenoj notaciji, osnovni metod (3.7.4) postaje

$$(3.7.11) \quad \hat{z}_i = z_i - w_i \quad (i = 1, \dots, n),$$

dok je metod (3.7.10)

$$(3.7.12) \quad \hat{z}_i = z_i - \frac{P(z_i)}{\prod_{m=1}^{i-1} (z_i - \hat{z}_m) \prod_{m=i+1}^n (z_i - z_m)} \quad (i = 1, \dots, n).$$

Ranije smo rekli da osnovni metod (3.7.11) ima kvadratnu konvergenciju. U radu [3] dokazano je da je R-red konvergencije iterativnog procesa (3.7.12), u zavisnosti od stepena polinoma, jednak $r(n) = 1 + \sigma_n$, gde je σ_n jedinstven pozitivan koren jednačine $\sigma^n - \sigma - 1 = 0$.

Jedna druga modifikacija procesa (3.7.11), sa kubnom konvergencijom, koja koristi popravku w_i (videti, Nourein [60]) data pomoću

$$(3.7.13) \quad \hat{z}_i = z_i - \frac{P(z_i)}{\prod_{\substack{m=1 \\ m \neq i}}^n (z_i - z_m + w_m)} \quad (i = 1, \dots, n).$$

Dalje ubrzavanje konvergencije može se postići kombinacijom formula (3.7.12) i (3.7.13) (videti rad Petkovića i Milovanovića [73]):

$$(3.7.14) \quad \hat{z}_i = z_i - \frac{\frac{P(z_i)}{n}}{\prod_{m=1}^n (z_i - \hat{z}_m) \prod_{m=i+1}^n (z_i - z_m + w_m)} \quad (i=1, \dots, n).$$

R-red konvergencije ovog procesa je $r(n) = 1 + \sigma_n$, gde je σ_n jedinstven pozitivan koren jednačine $\sigma^n - \sigma - \sum_{k=0}^{n-1} \sigma^k = 0$.

Jedan drugačiji pristup modifikaciji osnovne formule (3.7.11) dao je W. Börsch-Supan ([7]):

$$(3.7.15) \quad z_i = z_i - \frac{w_i}{1 + \sum_{\substack{m=1 \\ m \neq i}}^n \frac{w_m}{z_i - z_m}} \quad (i = 1, \dots, n).$$

Konvergencija ovog procesa je kubna. Do formule (3.7.15) može se jednostavno doći konstrukcijom Lagrangeovog interpolacionog polinoma (videti odeljak 6.2.2) za funkciju(polinom) $z \mapsto P(z) - Q(z)$ u čvorovima z_1, \dots, z_n , tj.

$$P(z) - Q(z) = \sum_{m=1}^n P(z_m) \frac{Q(z)}{(z - z_m) Q'(z_m)}.$$

Ako za z uzmememo bilo koju nulu polinoma r_i , iz poslednje jednakosti dobijamo

$$(3.7.16) \quad r_i = z_i - \frac{w_i}{1 - \sum_{\substack{m=1 \\ m \neq i}}^n \frac{w_m}{z_m - r_i}} \quad (i=1, \dots, n).$$

Najzad, stavljajući $r_i := z_i$, desna strana u (3.7.16) daje aproksimaciju za nulu r_i . Označavajući ovu aproksimaciju sa \hat{z}_i dobijamo formulu (3.7.15).

Slično konstrukciji formule (3.7.13), na osnovu (3.7.15) se može dobiti formula četvrtog reda (videti, Nourein [61]):

$$(3.7.17) \quad \hat{z}_i = z_i - \frac{w_i}{1 + \sum_{\substack{m=1 \\ m \neq i}}^n \frac{w_m}{z_i - w_i - z_m}} \quad (i = 1, \dots, n).$$

Koristeći logaritamski izvod polinoma P , Maehly [47] (kasnije Börsch-Supan [6], Ehrlich [20], Aberth [1], i drugi) je izveo formulu

$$(3.7.18) \quad \hat{z}_i = z_i - \frac{1}{\frac{1}{N_i} - \sum_{\substack{m=1 \\ m \neq i}}^n \frac{1}{z_i - z_m}} \quad (i = 1, \dots, n),$$

koja ima kubnu konvergenciju. Naime, na osnovu

$$\frac{P'(z)}{P(z)} = \sum_{m=1}^n \frac{1}{z - r_m},$$

važi

$$r_i = z - \frac{1}{\frac{P'(z)}{P(z)} - \sum_{\substack{m=1 \\ m \neq i}}^n \frac{1}{z - r_m}},$$

odakle zamenom $z := z_i$, $r_m := z_m$ ($m \neq i$), na desnoj strani u prethodnoj jednakosti dobijamo novu aproksimaciju za r_i . Ako ovu aproksimaciju označimo sa \hat{z}_i dobijamo formulu (3.7.18).

Korišćenjem Gauss-Seidelovog pristupa dobijamo poboljšani metod

$$(3.7.19) \quad \hat{z}_i = z_i - \frac{1}{\frac{1}{N_i} - \sum_{m=1}^{i-1} \frac{1}{z_i - \hat{z}_m} - \sum_{m=i+1}^n \frac{1}{z_i - z_m}} \quad (i=1, \dots, n),$$

koji ima R-red konvergencije $r(n)=2 + \sigma_n$ (> 3), gde je σ_n jedinstveni pozitivni koren jednačine $\sigma^n - \sigma - 2 = 0$ (videti, Alefeld i Herzberger [3]).

Uzimajući Newtonovu korekciju, Nourein ([60]) je dobio sledeću modifikaciju formule (3.7.18):

$$(3.7.20) \quad \hat{z}_i = z_i - \frac{1}{\frac{1}{N_i} - \sum_{\substack{m=1 \\ m \neq i}}^n \frac{1}{z_i - z_m + N_m}} \quad (i = 1, \dots, n),$$

čiji je red konvergencije četiri.

Analogno formuli (3.7.14) može se dobiti ubrzani iterativni proces bez dodatnih izračunavanja (videti rad Milovanovića i

Petković [53]):

$$(3.7.21) \quad \hat{z}_i = z_i - \frac{1}{\frac{1}{N_i} - \sum_{m=1}^{i-1} \frac{1}{z_j - \hat{z}_m} - \sum_{m=i+1}^n \frac{1}{z_i - z_m + N_m}} \quad (i=1, \dots, n)$$

čiji je R-red konvergencije $r(n) = 2(1 + \sigma_n)$, gde je $\sigma_n \in (1, 2)$ jedinstveni pozitivni koren jednačine $\sigma^n - \sigma - 1 = 0$.

Ilustrovaćemo određivanje R-rede konvergencije na primjeru metoda (3.7.21). Neka je prethodno

$$d = \min_{i \neq j} |r_i - r_j|, \quad v_j^{(k)} = z_j^{(k)} - r_j, \quad \hat{g}_i^{(k)} = \sum_{j=1}^n \frac{z_i^{(k)} - r_i}{z_i^{(k)} - r_j}, \quad j \neq i$$

$$g_i^{(k)} = \sum_{j=1}^{i-1} \frac{(z_i^{(k)} - r_i)(r_j - z_j^{(k+1)})}{(z_i^{(k)} - r_j)(z_i^{(k)} - z_j^{(k+1)})} + \sum_{j=i+1}^n \frac{(z_i^{(k)} - r_i)(r_j - w_j^{(k)})}{(z_i^{(k)} - r_j)(z_i^{(k)} - w_j^{(k)})},$$

gde je $w_j^{(k)}$ aproksimacija dobijena po Newtonovom metodu, tj.

$$w_j^{(k)} = z_j^{(k)} - p(z_j^{(k)}) / p'(z_j^{(k)}).$$

Nije teško pokazati da se Newtonov metod može predstaviti u obliku

$$(3.7.22) \quad w_j^{(k)} - r_j = \frac{\hat{g}_j^{(k)}}{1 + \hat{g}_j^{(k)}} (z_j^{(k)} - r_j) = \frac{\hat{g}_j^{(k)}}{1 + \hat{g}_j^{(k)}} v_j^{(k)}.$$

Slično, za iterativni proces (3.7.21) imamo

$$z_1^{(k+1)} = r_1 + \frac{g_1^{(k)}}{1 + g_1^{(k)}} (z_1^{(k)} - r_1) \quad (i=1, \dots, n; k=0, 1, \dots),$$

odakle je

$$(3.7.23) \quad v_i^{(k+1)} = \frac{g_i^{(k)}}{1 + g_i^{(k)}} v_i^{(k)} \quad (i=1, \dots, n; k=0, 1, \dots).$$

Pretpostavimo da su početne aproksimacije odabrane tako da su zadovoljeni uslovi

$$(3.7.24) \quad |v_i^{(0)}| < \frac{1}{q} = \frac{d}{2n-1} \quad (i=1, \dots, n)$$

Tada za $i \neq j$ imamo

$$|z_1^{(0)} - r_j| \geq |r_1 - r_j| - |z_1^{(0)} - r_1| > d - \frac{d}{2n-1},$$

$$|z_i^{(0)} - z_j^{(0)}| \geq |z_i^{(0)} - r_j| - |z_j^{(0)} - r_j| > (d - \frac{d}{2n-1}) - \frac{d}{2n-1},$$

odakle je

$$|z_i^{(0)} - r_j| > \frac{2(n-1)}{q} \quad i \quad |z_i^{(0)} - z_j^{(0)}| > \frac{2n-3}{q} \geq \frac{1}{q}.$$

Odredimo sada ocenu za $|g_i^{(0)}|$. Na osnovu (3.7.24) i prethodnih nejednakosti imamo

$$(3.7.25) \quad |\hat{g}_i^{(0)}| \leq |v_i^{(0)}| \sum_{\substack{j=1 \\ j \neq i}}^n \frac{1}{|z_i^{(0)} - r_j|} < \frac{q}{2} |v_i^{(0)}| < \frac{1}{2}.$$

Tada iz (3.7.22), za $k=0$, sleduje

$$(3.7.26) \quad |r_j^{(0)} - w_j^{(0)}| \leq \frac{|\hat{g}_j^{(0)}|}{1 - |\hat{g}_j^{(0)}|} |v_j^{(0)}| < q |v_i^{(0)}|^2 < \frac{1}{q}.$$

Kako je, za $i \neq j$

$$|z_i^{(0)} - w_j^{(0)}| \geq |z_i^{(0)} - r_j| - |r_j - w_j^{(0)}| > \frac{2n-3}{q} \geq \frac{1}{q},$$

korisćenjem prethodnje ocena nalazimo

$$(3.7.27) \quad \frac{|r_j - w_j^{(0)}|}{|z_i^{(0)} - r_j| |z_i^{(0)} - w_j^{(0)}|} < \frac{q^3}{2(n-1)} |v_j^{(0)}|^2.$$

Nadjimo sada ocene za $|v_i^{(1)}|$ i $|g_i^{(0)}|$. Na osnovu (3.7.23) dobijamo

$$(3.7.28) \quad |v_i^{(1)}| \leq \frac{|g_i^{(0)}|}{1 - |g_i^{(0)}|} |v_i^{(0)}| \quad (i = 1, \dots, n).$$

Kako je

$$g_1^{(0)} = \left| \sum_{j=2}^n \frac{(z_1^{(0)} - r_1)(r_j - w_j^{(0)})}{(z_1^{(0)} - r_j)(z_1^{(0)} - w_j^{(0)})} \right|,$$

korisćenjem (3.7.24) i (3.7.27), nalazimo da je

$$|g_1^{(0)}| < \frac{q^3}{2(n-1)} \cdot \frac{n-1}{q^2} |v_1^{(0)}| < \frac{1}{2},$$

odakle, s obzirom na (3.7.28), zaključujemo da je $|v_1^{(1)}| < |v_1^{(0)}| < \frac{1}{q}$.

Na osnovu prethodnog razmatranja i nejednakosti

$$|z_2^{(0)} - z_1^{(1)}| \geq |z_2^{(0)} - r_1| - |v_1^{(1)}| > \frac{1}{q},$$

sukcesivno za $i = 2, \dots, n$, dobijamo sledeće ocene

$$\begin{aligned} |g_i^{(0)}| &< |v_i^{(0)}| \left(\frac{q^2}{2(n-1)} \sum_{j=1}^{i-1} |v_j^{(1)}| + \frac{q^3}{2(n-1)} \sum_{j=i+1}^n |v_j^{(0)}|^2 \right) < \frac{1}{2}, \\ |v_i^{(1)}| &< \frac{q^2}{n-1} |v_i^{(0)}|^2 \left(\sum_{j=1}^{i-1} |v_j^{(1)}| + q \sum_{j=i+1}^n |v_j^{(0)}|^2 \right) < \frac{1}{q}. \end{aligned}$$

Matematičkom indukcijom se sada može dokazati da za svako $k=0, 1, \dots$ važi

$$|v_i^{(k+1)}| < \frac{q^2}{n-1} |v_i^{(k)}|^2 \left(\sum_{j=1}^{i-1} |v_j^{(k+1)}| + q \sum_{j=i+1}^n |v_j^{(k)}|^2 \right) < \frac{1}{q}.$$

Uvodjenjem smene $q|v_i^{(k)}| = h_i^{(k)}$, poslednja nejednakost postaje

$$(3.7.29) \quad h_i^{(k+1)} < \frac{1}{n-1} (h_i^{(k)})^2 \left(\sum_{j=1}^{i-1} h_i^{(k+1)} + \sum_{j=i+1}^n (h_i^{(k)})^2 \right),$$

gde je $i=1, \dots, n$ i $k=0, 1, \dots$. Na osnovu uslova (3.7.24) imamo da je $h_i^{(0)} < 1$ ($i=1, \dots, n$). Ako stavimo $h = \max_i h_i^{(0)}$, tada je

$$h_i^{(0)} \leq h < 1 \quad (i=1, \dots, n).$$

Na osnovu (3.7.29) zaključujemo da je iterativni proces (3.7.21) konvergentan. Stavimo dalje da je

$$h_i^{(k+1)} \leq h_i^{(k+1)} \quad (i=1, \dots, n; k=0, 1, \dots).$$

Ako definišemo matricu B pomoću $B = 2A$, gde je

$$A = \begin{bmatrix} 1 & 1 & & & \\ & 1 & 1 & \dots & 0 \\ & & \ddots & \ddots & \ddots \\ & 0 & & \ddots & \ddots \\ & & & 1 & 1 \\ 1 & 1 & 0 & \dots & 0 & 1 \end{bmatrix},$$

vektori $\vec{u}^{(k)} = [u_1^{(k)} \dots u_n^{(k)}]^T$ mogu biti sukcesivno izračunati po-

moću

$$\vec{u}^{(k+1)} = B \vec{u}^{(k)} \quad (k = 0, 1, \dots),$$

startujući sa $\vec{u}^{(0)} = [1 \dots 1]^T$. Ovo se može dokazati matematičkom indukcijom. Slično razmatranju u radu [3] pokazuje se da za R-red konvergencije iterativnog procesa (3.7.21) važi

$$O_R((3.7.21), \vec{r}) \geq \rho(B) = 2\rho(A) \quad (\vec{r} = [r_1 \dots r_n]^T),$$

gde su $\rho(A)$ i $\rho(B)$ spektralni radijusi matrica A i B.

Karakteristični polinom matrice A je

$$f_n(\lambda) = (\lambda - 1)^n - (\lambda - 1) - 1.$$

Ako stavimo $\lambda = \sigma - 1$ dobijamo da je $\bar{f}_n(\sigma) = f_n(1+\lambda) = \sigma^n - \sigma - 1$. Kako je $\bar{f}_n(1) = -1 < 0$ i $\bar{f}_n(2) = 2^n - 3 > 0$, zaključujemo da postoji nula σ_n u intervalu $(1, 2)$. Lako je pokazati da je ovo jedina pozitivna nula polinoma \bar{f}_n , i da je $\rho(A) = 1 + \sigma_n$. Najzad, na osnovu (3.7.30) dobijamo

$$O_R((3.7.21), \vec{r}) \geq 2(1 + \sigma_n),$$

što je trebalo i dokazati.

Za iterativne procese (3.7.12), (3.7.14), (3.7.19), (3.7.21) kođa kojih je R-red konvergencije zavisan od n, u tabeli 3.7.1 dame vrednosti za $r(n)$ ($n=3, \dots, 10$).

Tabela 3.7.1

n	(3.7.12)	(3.7.14)	(3.7.19)	(3.7.21)
3	2.325	3.148	3.521	4.649
4	2.221	3.066	3.353	4.441
5	2.167	3.032	3.267	4.335
6	2.135	3.016	3.215	4.269
7	2.113	3.008	3.180	4.226
8	2.097	3.004	3.154	4.194
9	2.085	3.002	3.135	4.170
10	2.076	3.001	3.125	4.152

Primetimo da $r(n)$ teži redu konvergencije osnovnog(ne ubrzanoг) metoda, kada $n \rightarrow +\infty$.

U radu [72], M.S. Petković je izveo jednu familiju metoda za simultano određivanje nula polinoma, korišćenjem razvoja ra-

cionalne funkcije u parcijalne razlomke. Svi metodi ove familije imaju kubnu konvergenciju za proste nule. U više radova ([26], [23], [66], [69], [75]) razmatrani su i simultani intervalni metodi, uključujući i generalizaciju metoda iz odeljka (5.3.4).

Takodje, u nekoliko radova ([25], [54], [71], [89]) razmatrani su i simultani metodi za odredjivanje višestrukih nula polinoma.

5.4. LITERATURA

1. O.Aberth: *Iteration methods for finding all zeros of a polynomial simultaneously.* Math. Comp. 27(1973), 339-344.
2. G.Alefeld: *Eine Modifikation des Newtonverfahrens zur Bestimmung der reellen Nullstellen einer reellen Funktion.* Z. Angew. Math. Mech. 50 (1970), 32-33.
3. G.Alefeld & J.Herzberger: *On the convergence speed of some algorithms for the simultaneous approximation of polynomial roots.* SIAM J. Numer. Anal. 11(1974), 237-243.
4. G.Alefeld und J.Herzberger: *Über Simultanverfahren zur Bestimmung reeller Polynomwurzeln.* Z. Angew. Math. Mech. 54(1974), 413-420.
5. N.Apostolatos und U.Kulisch: *Grundlagen einer Maschinenintervallarithmetik.* Computing 2(1967), 458-467.
6. W.Börsch-Supan: *A posteriori error bounds for the zeros of polynomials.* Numer. Math. 5(1963), 380-398.
7. W.Börsch-Supan: *Residuenabschätzung für Polynom-Nullstellen mittels Lagrange-Interpolation.* Numer. Math. 14(1970), 287-296.
8. D.Braess & K.P.Hadeler: *Simultaneous inclusion of the zeros of a polynomial.* Numer. Math. 21(1973), 161-165.
9. K.W.Brodlie: *On Bairstow's method for the solution of polynomial equations.* Math. Comp. 29(1975), 816-826.
10. C.G.Broyden: *Quasi-Newton methods and their application to function minimisation.* Math. Comp. 21(1967), 368-381.
11. H.Christ: *Realisierung einer Maschinenintervall arithmetic mit beliebigen Algol 60 - Compilern.* Elektron. Rechenanl. 10(1968), 217-222.
12. R.Dargel, F.Loscalzo, T.Witt: *Automatic error bounds on real zeros of rational functions.* Comm. ACM 9(1966), 806-809.
13. J.C.Daubisse: *Sur une méthode de résolution numérique d'équations algébriques en particulier dans le cas de racines multiples.* Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz. N°498 - N°541 (1975), 163-166.
14. L.N.Djordjević: *An iterative solution of algebraic equations with a parameter to accelerate convergence.* Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz. N° 412 - N°460 (1973), 179-182.
15. L.N.Djordjević & G.V.Milovanović: *A combined iterative formula for solving equations.* Informatica 78, Bled 1978, 3(207).

16. K.Dočev: *A modified Newton method for the simultaneous approximation of all roots of the given algebraic equation.* (Bulgarian). *Fiz.-Mat. Spis. Bulgar. Akad. Nauk* 5(1962), 136-139.
17. 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), 915-912.
18. D.K.Dunaway: *Calculation of zeros of a real polynomial through factorization using Euclid's algorithm.* SIAM J. Numer. Anal. 11(1974), 1087-1104.
19. J.Dvorčuk: *Factorization of a polynomial into quadratic factors by Newton method.* Apl. Mat. 14(1969), 54-80.
20. L.W.Ehrlich: *A modified Newton method for polynomials.* Comm. ACM 10 (1967), 107-108.
21. M.R.Farmer & G.Loizou: *A class of iteration functions for improving, simultaneously approximations to the zeros of a polynomial.* BIT 15 (1975), 250-258.
22. M.R.Farmer & G.Loizou: *An algorithm for the total, of partial, factorization of a polynomial.* Math. Proc. Camb. Phil. Soc. 82(1977), 427-437.
23. I.Gargantini: *Parallel Laguerre iterations: Complex case.* Numer. Math. 26(1976), 317-323.
24. I.Gargantini: *Further applications of circular arithmetic: Schroeder-like algorithm with error bounds for finding zeros of polynomial.* SIAM J. Numer. Anal. 15(1978), 497-510.
25. I.Gargantini: *Parallel square-root iterations for multiple roots.* Comput. Math. Appl. 6(1980), 279-288.
26. I.Gargantini & P.Henrici: *Circular arithmetic and the determination of polynomial zeros.* Numer. Math. 18(1972), 305-320.
27. E.Hansen: *On solving systems of equations using interval arithmetic.* Math. Comp. 22(1968), 374-384.
28. E.Hansen: *A globally convergent interval method for computing and bounding real roots.* BIT 18(1978), 415-424.
29. E.Hansen: *Interval forms of Newton method.* Computing 20(1978), 153-163.
30. E.Hansen & M.Patrick: *A family of root finding methods.* Numer. Math. 27(1977), 257-269.
31. E.Hansen, M.Patrick, J.Rusnak: *Some modifications of Laguerre's method.* BIT 17(1977), 409-417.
32. R.Hanson: *Automatic error bounds for real roots of polynomials having interval coefficients.* Comput. J. 13(1970), 284-288.
33. P.Henrici: *Uniformly convergent algorithms for the simultaneous determination of all zeros of a polynomial.* Studies in Numer. Anal., 1969, 1-8.
34. P.Henrici: *Circular arithmetic and the determination of polynomial zeros.* Conference on Application of Numerical Analysis. In: Lect. Notes Math. 228, Springer-Verlag, Berlin-Heidelberg-New York, 1971, pp.86-92.

35. P.Henrici: Applied and computational complex analysis. Vol.I, Wiley - Interscience, New York, 1974.
36. J.Herzberger: Über ein Verfahren zur Bestimmung reeller Nullstellen mit Anwendung auf Parallelrechnung. Elektron. Rechenanal. 14(1972), 250-254.
37. J.Herzberger: Bemerkungen zu einem Verfahren von R.E. Moore. Z. Angew. Math. Mech. 53(1973), 356-358.
38. L.N.Hromova: Ob odnoš modifikacij metoda kasatel'nyh giperbol. Taškent. Ord.Trud. Kras.Znam.Gos. Univ. im. V.I.Lenina 476(1975), 35-41.
39. L.V.Kantorovič: Funkcional'nyj analiz i prikladnaja matematika. Uspehi Mat. Nauk 3(1948), 89-185.
40. I.O.Kerner: Ein Gesamtstschrittverfahren zur Berechnung der Nullstellen von Polynomen. Numer. Math. 8(1966), 290-294.
41. M.A.Krasnosel'skiĭ, G.M.Vaňniko, P.P.Zabreiko, Ja.B.Rutickiĭ, V.Ja.Stecenko: Približennoe rešenie operatornyh uravnenii. Nauka, Moskva, 1969.
42. R.Krawczyk: Newton-Algorithmen zur Bestimmung von Nullstellen mit Fehlerabschranken. Computing 4(1969), 187-201.
43. O.Ju.Kul'čickiĭ i L.I.Šimelevič: O nahoždenii načal'nogo približenija. Ž. Vyčisl. Mat. i Mat. Fiz. 14(1974), 1016-1018.
44. D.K.Lika: Ob iteracionnyh metodah vysšego porjadka. Tezisi Dokl. 2-ї Nauč.-tehn. respubl. konf. Moldavii. Kišinev, 1965, 13-16.
45. G.Loizou: Une Note sur le procédé itératif de M^{me} Marica D. Prešić. C. R. Acad. Sci. Paris, 295(1982), 707-710.
46. G.Loizou: Higher-order iteration functions for simultaneously approximating polynomial zeros. Internat. J. Comput. Math. 14(1983), 45-58.
47. H.J.Maehly: Zur iterativen Auflösung algebraischer Gleichungen. Z. Angew. Math. Phys. 5(1954), 260-263.
48. M.Mignotte: Note sur la méthode Bernoulli. Numer. Math. 325-326.
49. G.V.Milovanović: A method to accelerate iterative processes in Banach space. Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz. N^o461 - N^o497(1974), 67-71.
50. G.V.Milovanović & L.N.Djordjević: On some iterative formulas of the third order. Informatica 78, Bled 1978, 3(202).
51. G.V.Milovanović i M.A.Kovačević: Dva iterativna procesa trećeg reda u kojima se ne pojavljuju izvodi. Zbornik radova sa IV znanstvenog skupa PPPR, Stubičke Toplice 1982, 473-478.
52. G.V.Milovanović & M.S.Petković: On some modifications of a third order method for solving equations. Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz. N^o678 - N^o715(1980), 63-67.
53. G.V.Milovanović & M.S.Petković: On the convergence order of a modified method for simultaneous finding polynomial zeros. Computing 30 (1983), 171-178.
54. G.V.Milovanović i M.S.Petković: Metodi visokog reda za simultano određivanje višestrukih nula polinoma. Zbornik radova sa V znanst-

- venog skupa PPPR, Stubičke Toplice 1983, 95-100.
55. I.P.Mysovskih: *O shodimosti metoda L.V. Kantoroviča rešenija funkcional-n'ih uravnenii i ego primenenija*. Dokl. Akad. Nauk SSSR 70(1950), 565-568.
56. D.S.Mitrinović i D.Ž.Djoković: Polinomi i matrice. ICS, Beograd, 1975.
57. R.E.Moore: Interval analysis. Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1966.
58. R.E.Moore: Methods and applications of interval analysis. SIAM Studies in Applied Mathematics. Philadelphia, 1979.
59. A.W.Nourein: *An iteration formula for simultaneous determination of the zeroes of a polynomial*. J. Comput. Appl. Math. 1(1975), 251-254.
60. A.W.Nourein: *An improvement on two iteration methods for simultaneous determination of the zeroes of a polynomial*. Internat. J. Comput. Math. 3(1977), 241-252.
61. A.W.Nourein: *An improvement of Nourein's method for the simultaneous determination of the zeroes of a polynomial*. J.Comput. Appl. Math. 2(1977), 109-112.
62. J.M.Ortega & W.C.Rheinboldt: Iterative solution of nonlinear equations in several variables. Academic Press, New York, 1970.
63. A.Ostrowski: Solution of equations and system equations. Academic Press, New York, 1966.
64. Lj.D.Petković: O nekim aproksimacijama pomoću diskova (Doktorska disertacija). Kragujevac, 1984.
65. M.S.Petković: *Some interval methods of the second order for the simultaneous approximation of polynomial roots*. Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz. N°634 - N°677(1979), 74-81.
66. M.S.Petković: Neki iterativni intervalni metodi za rešavanje jednačina (Doktorska disertacija). Niš, 1980.
67. M.S.Petković: *On an interval Newton's method derived from exponential curve fitting*. Z. Angew. Math. Mech. 61(1981), 117-119.
68. M.S.Petković: *A note on Moore's method*. Freiburger Intervall-Berichte. 8(1981), 31-35.
69. M.S.Petković: *On a generalisation of the root iterations for polynomial complex zeros in circular interval arithmetic*. Computing 27(1981), 37-55.
70. M.S.Petković: *On an iterative method for simultaneous inclusion of polynomial complex zeros*. J. Comput. Appl. Math. 8(1982), 51-56.
71. M.S.Petković: *Generalized root iterations for the simultaneous determination of multiple complex zeros*. Z. Angew. Math. Mech. 62(1982), 627-630.
72. M.S.Petković: *A family of simultaneous method for the determination of polynomial complex zeros*. Internat. J. Comput. Math. 11(1982), 285-296.

73. M.S.Petković & G.V.Milovanović: *A note on some improvements of the simultaneous method for determination of polynomial zeros.* J. Comput. Appl. Math. 9(1983), 65-69.
74. M.S.Petković & Lj.D.Petković: *On a method of two-sided approaching for solving equations.* Freiburger Intervall-Berichte 10(1980), 1-10.
75. M.Petković & L.Stefanović: *On the convergence order of accelerated root iterations.* Numer. Math. 44(1984), 463-476.
76. M.S.Petković & L.V.Stefanović: *The numerical stability of the generalised root iterations for polynomial zeros.* Comput. Math. Appl. 10 (1984), 97-106.
77. J.Petrić, M.Jovanović, S.Stamatović: *Algorithm for simultaneous determination of all roots of algebraic polynomial equations.* Mat. Vesnik 9(24)(1972), 325-332.
78. D.B.Popovski: *A family of one-point iteration formulae for finding roots.* Internat. J. Comput. Math. 8(1980), 85-88.
79. D.B.Popovski: *Sixth order methods for solving equations.* Z. Angew. Math. Phys. 33(1982), 434-438.
80. M.D.Prešić: *Un procédé itératif pour déterminer les zéros d'un polynome.* C. R. Acad. Sci. Paris. 273(1971), 446-449.
81. M.D.Prešić: *Jedan iterativni postupak za jednovremeno određivanje k realnih rešenja jednačine na polju realnih brojeva.* Mat. Vesnik 10(25) (1973), 299-308.
82. M.D.Prešić: *A convergence theorem for a method for simultaneous determination of all zeros of a polynomial.* Publ. Inst. Math. 28(42)(1980), 159-168.
83. S.B.Prešić: *Un procédé itératif pour la factorisation des polynomes.* C. R. Acad. Sci. Paris. 262(1966), 862-863.
84. S.B.Prešić: *Jedan iterativni postupak za faktorizaciju polinoma.* Mat. Vesnik 5(20)(1968), 205-216.
85. L.Rall: Computational solution of nonlinear operator equations. New York, 1969.
86. A.Ralston: A first course in numerical analysis. McGraw-Hill, New York, 1965.
87. G.S.Salehov: *O shodimosti metoda kasatel'nyh giperbol.* Dokl. Akad. Nauk SSSR 82(1952), 525-528.
88. R.M.Sekulovski: *Uopštenje iterativne metode S. Prešića za faktorizaciju polinoma.* Mat. Vesnik 9(24)(1972), 257-264.
89. H.Semeržiev: *A method for simultaneous finding all roots of algebraic equations with given multiplicity (Bulgarian).* C.D. Acad. Bulgare Sci. 35(1982), 1057-1060.
90. T.Sunaga: *Theory of an interval algebra and its applications to numerical analysis.* RAAG Memoirs 2(1958), 29-46.
91. M.K.Taburin: *Nelineinie funkcional'nye uravnenija i nepreryvnye analogi iterativnyh metodov.* Izv. Vysš.Uč. Zaved. Matematika 5(1958), 18-31.

92. O.N.Tihonov: *O bystrom vyčisleniš naibol'sih komei mnogočlena.* Zap. Leningrad. gorn. in-ta 48, 3(1968), 36-41.
93. D.Dj.Tošić & 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. N-412 - N-460(1973), 175-177.
94. J.F.Traub: *Iterative methods for the solution of equations.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1964.
95. V.A.Varjuhin i S.A.Kasjanjuk: *Ob iteracionnyh metodah utočnenija komei uravneniš.* Ž. Vyčisl. Mat. i Mat. Fiz. 9(1969), 684 - 687.
96. V.M.Verbuš i D.I.Mil'man: *Metod Vegsteina kak modifikacija metoda sekuščih.* Ž. Vyčisl. Mat. i Mat. Fiz. 17(1977), 507-508.
97. B.A.Vertgeim: *O nekotoryh metodah približennogo rešenija nelineiných operatornyh uravnenii v prostranstvah Banaha.* Uspehi Mat. Nauk 12 (1957), 166-169.
98. J.H.Wegstein: *Accelerating convergence of iterative processes.* Comm. ACM 1(1958), 9-13.
99. K.Weierstrass: *Neuer Beweis des Satzes, dass jede Ganze Rationale Function einer Veränderlichen dargestellt werden kann als ein Product aus Linearen Functionen darstellen Veränderlichen.* Ges. Werke 3(1903), 251-296.
100. W.Werner: *Iterative solution of nonlinear systems of equations.* In: Lect. Notes Math. 953, Springer-Verlag, Berlin-Heidelberg-New York, pp. 188-202.
101. J.H.Wilkinson: *The evaluation of zeros of ill-conditioned polynomials.* Numer. Math. 1(1959), 150-166.
102. H.W.Wippermann: *Realisierung einer Intervallarithmetik in einem Algol-60 System.* Elektron. Rechenanl. 9(1967), 224-233.
103. D.M.Young & R.T.Gregory: *A survey of numerical mathematics.* Addison-Wesley Publ. Co., Reading, Massachusetts, 1972.
104. V.L.Zaguskin: *Spravočnik po čislennym metodam rešenija uravneniš.* Fizmatgiz, Moskva, 1960.