Tomislav Došlić

Numerička matematika

Građevinski fakultet
Sveučilište u Zagrebu

Sadržaj

1 Uvod **1**
1.1 Apsolutne i relativne pogrješke . . . 1
1.2 Osnovni izvori pogrješaka . . . 1

2 Približno rješavanje jednadžbi **2**
2.1 Izoliranje rješenja . . . 2
2.2 Metoda raspolavljanja (bisekcije) . . . 4
2.3 Metoda sekante (regula falsi) . . . 4
2.4 Newtonova metoda (metoda tangente) . . . 6
2.5 Metoda iteracije (metoda fiksne točke) . . . 7

3 Aproksimacija i interpolacija **9**
3.1 Problem aproksimacije funkcija . . . 9
3.2 Osnovni tipovi aproksimacijskih funkcija . . . 10
3.3 Kriteriji optimalnosti . . . 11
3.4 Polinomijalna interpolacija . . . 13
3.4.1 Rješivost problema interpolacije . . . 13
3.4.2 Lagrangeov oblik interpolacijskog polinoma . . . 13
3.4.3 Newtonov oblik interpolacijskog polinoma . . . 15
3.4.4 Problemi s polinomijalnom interpolacijom . . . 16
3.4.5 Splineovi . . . 17
3.4.6 Zaključak . . . 19

4 Numeričko integriranje **20**
4.1 Uvod . . . 20
4.2 Newton-Cotesove kvadraturne formule . . . 22
4.3 Trapezna formula . . . 23
4.4 Simpsonova formula . . . 24
4.5 Newton-Cotesove formule viših redova . . . 25
4.6 Poopćena trapezna formula . . . 26
4.7 Poopćena Simpsonova formula . . . 26
4.8 Gaussove kvadraturne formule . . . 27
4.9 Mogući problemi . . . 29
4.10 Kubaturne formule i višestruki integrali . . . 30

5 Obične diferencijalne jednadžbe **31**
5.1 Eulerova metoda . . . 32
5.2 Poboljšana Eulerova metoda . . . 33
5.3 Metode Runge-Kutta . . . 33

6 Matrice i linearni sustavi **36**
6.1 Izvori problema . . . 36
6.2 Tipovi matrica . . . 36
6.3 Tipovi problema . . . 37

6.4 Tipovi metoda 38
6.5 Gausove eliminacije 38
6.5.1 Modifikacije Gausovih eliminacija - pivotiranje 39
6.5.2 Gauss-Jordanove eliminacije 40
6.6 Simetrične matrice 40
6.7 Analiza pogrješke izravnih metoda 41
6.8 Jacobijeva metoda 42
6.9 Gauss-Seidelova metoda 43
6.10 Ocjena pogrješke iteracijskih metoda 44
6.11 OR metode 45
6.12 Problem svojstvenih vrijednosti 45
6.12.1 Karakteristični polinom 47
6.12.2 Krylovljeva metoda 47
6.12.3 Metoda neodređenih koeficijenata 48
6.12.4 Lokalizacija nul-točaka 48

Predgovor

Ovaj nastavni materijal namijenjen je studentima druge godine svih smjerova diplomskog studija koji su se tijekom svog obrazovanja već upoznali s mnogim matematičkim modelima kojima opisujemo mehaničke probleme. Ti modeli opisuju odnose veličine koja nas zanima (recimo progiba grede) i vanjskih i unutarnjih parametara o kojima ta veličina ovisi (kao što su svojstva materijala, raspodjela vanjskih opterećenja i sl.). U idealnom slučaju, rješenje takvog modela trebala bi biti formula koja bi na kompaktan način prikazala ovisnost promatrane veličine o parametrima modela. Iz iskustva znamo da je to više iznimka nego pravilo - rješenja koja se daju iskazati elegantnim matematičkim formulama dobivaju se samo za vrlo specijalne slučajeve. U pravilu je riječ o područjima visoke simetrije (segment, kvadrat, pravokutnik, krug) i o modelima koji zanemarivanjem nekih parametara značajno pojednostavnjuju fizikalnu situaciju (mali progibi, konstantni parametri i sl.). Kad god imamo područje nepravilne geometrije i/ili model koji ne zanemaruje nezgodna svojstva parametara, moramo rješenje modela tražiti numeričkim metodama.

Numerička matematika, tj. grana matematike koja se bavi razvojem i proučavanjem metoda za približno rješavanje problema, doživjela je proteklih desetljeća iznimno brz razvoj. To je, u sinergiji s istim takvim razvojem računala, operacijskih sustava i programskih jezika, dovelo do toga da se danas mogu uspješno modelirati složeni nelinearni problemi iz hidrodinamike i mehanike deformabilnih krutih tijela. Razvijeni su i standardni programski alati za pojedina područja koji omogućuju rutinsko modeliranje netrivijalnih problema i osobama koje nisu specijalisti. Takvi alati, osim nesumnjivih i očitih prednosti, nose i potencijalne opasnosti, kao što je njihova primjena na probleme za koje nisu predviđeni. Kako bi se izbjegle negativne posljedice i poboljšao učinak standardnih programskih paketa, bilo bi dobro i poželjno da njihovi korisnici imaju određeno razumijevanje matematičkih ideja i pojmova na kojima se oni temelje. S tim je ciljem i koncipiran kolegij Numerička matematika i napisan ovaj nastavni materijal.

Ni skripta ni kolegij nisu zamišljeni kao iscrpan prikaz numeričke matematike i njenih metoda. Namjera im je informirati i zainteresirati studente i čitatelje. Oni koji budu o tome željeli znati više mogu potražiti dodatne informacije u literaturi navedenoj na kraju skripte.

Za nastanak ove skripte u velikoj su mjeri zaslužni slušači prve generacije kojoj je kolegij predavan, Marina Alagušić, Luka Božić, Andrea Klarendić, Janko Košćak, Dejan Stjepanović i Gregor Turkalj. Zahvalan sam im što su s velikim strpljenjem i upornošću utipkali ne uvijek čitljiv i uredan tekst mojih bilješki. Posebno zahvaljujem Janku Košćaku koji je izradio slike.

Zahvaljujem i asistentici Ani Martinčić i višem asistentu Nikoli Sandriću koji su pozorno pročitali prvu verziju teksta i ukazali mi na mnoge pogrješke.

Odgovornost za sve preostale činjenične, pravopisne i stilske nedostatke je isključivo moja. Bit ću zahvalan svim čitateljima koji mi na njih ukažu.

U Zagrebu, 5. II 2014.

Tomislav Došlić

1 Uvod

1.1 Apsolutne i relativne pogrješke

Neka je $A \in \mathbb{R}$ točna vrijednost nekog broja. Njegova **približna vrijednost** je broj $a \in \mathbb{R}$, koji se malo razlikuje od A. Ako znamo da je $a < A$, kažemo da je a **donja aproksimacija** od A; ako znamo da je $a > A$, onda je a **gornja aproksimacija** od A.

Pogrješka Δa približnog broja a je razlika točne i približne vrijednosti: $\Delta a = A - a$. U većini slučajeva predznak pogrješke nije poznat. Zato se često radi s njenom apsolutnom vrijednošću. **Apsolutna pogrješka** Δ približnog broja a je $\Delta = |A - a|$.

Točna vrijednost veličine Δ u pravilu nije poznata pa se umjesto točne vrijednosti apsolutne pogrješke moramo zadovoljiti njenim ocjenama ili ogradama: $\Delta = |A - a| \leq \Delta_a$. Imamo li takvu ogradu, znamo u kojem se intervalu oko a nalazi točna vrijednost A: $a - \Delta_a \leq A \leq a + \Delta_a$. Često skraćeno pišemo $A = a \pm \Delta_a$.

Apsolutna pogrješka (ili njena ograda) ne daje potpun opis točnosti rezultata; ista apsolutna pogrješka od 1 cm znači puno više, ako je točna vrijednost $A = 1$ m, nego ako je $A = 1$ km. Stoga uvodimo pojam **relativne pogrješke**, $\delta = \frac{\Delta}{|A|}$. U pravilu radimo s ocjenama (ogradama) oblika $\delta \leq \delta_a$.
$\frac{\Delta}{|A|} \leq \delta_a \Rightarrow \Delta \leq |A|\,\delta_a$, pa možemo uzeti $\Delta_a = |A|\,\delta_a$, odnosno, zbog $A \simeq a$, $\Delta_a = |a|\,\delta_a$. Odatle možemo dobiti interval oko a u kojem leži A, $a(1 - \delta_a) \leq A \leq a(1 + \delta_a)$, ili skraćeno, $A = a(1 \pm \delta_a)$.

1.2 Osnovni izvori pogrješaka

1. Pogrješke modela - dolaze od idealizacija i aproksimacija u formulaciji matematičkog modela. Primjer - linearizacija.
2. Rezidualne pogrješke - dolaze od zamjena beskonačnih procesa konačnima.
   Primjer - računanje e^x.
3. Ulazne (početne) pogrješke - dolaze od netočnih vrijednosti parametara modela.
   Primjer - fizikalne konstante, $g \simeq 9.81$.
4. Pogrješke zaokruživanja - dolazi od problema s predstavljanjem brojeva u danom brojevnom sustavu.
   Primjer - $\frac{1}{3} = 0.333$ unosi pogrješku $\Delta \simeq 3 \times 10^{-4}$.
5. Pogrješke operacija - netočni ulazi daju netočne rezultate.

Na neke izvore i tipove pogrješaka ne možemo utjecati, a neke možemo smanjiti ili izbjeći pažljivim pristupom. Često se događa da smanjenje pogrješke jednog tipa vodi do povećanja pogrješke drugog tipa. Treba balansirati cijenu (u vremenu i opremi) i kvalitetu rezultata.

Primjer 1.1:
Koliko je $(\sqrt{2})^2$? Koliko je star fosil dinosaura?

2 Približno rješavanje jednadžbi

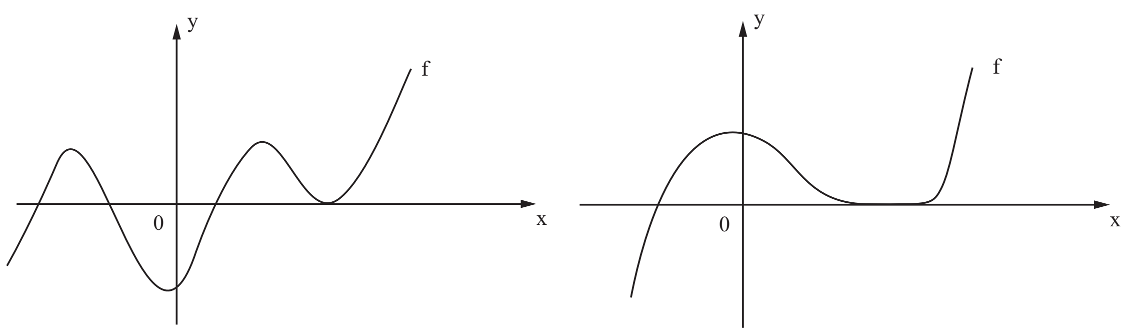
Mnogi inženjerski problemi svode se na rješavanje jednadžbi oblika $f(x) = 0$. Samo u najjednostavnijim slučajevima moguće je naći rješenja takve jednadžbe u obliku formule. (To je moguće ako je $f(x)$ polinom stupnja najviše 4, no i u tom slučaju su formule nespretne i nepraktične. Za ostale funkcije $f(x)$ je rješenje pomoću formule moguće samo za specijalne kombinacije koeficijenata. Npr., $\sin x - \frac{1}{2} = 0$ ima rješenje $x = \frac{\pi}{6} + 2k\pi,\ x = \frac{5\pi}{6} + 2k\pi$, no $\sin x - \frac{1}{3} = 0$?) Za neke važnije jednadžbe su rješenja tabelirana i proglašena standardnim funkcijama (npr. $e^x - c = 0$, $\tan x - c = 0$ itd.).

U mnogim slučajevima ni koeficijenti nisu točno, već samo približno poznati, pa nema ni smisla govoriti o točnom ili egzaktnom rješenju. Stoga je važno biti u stanju naći približno rješenje jednadžbi.

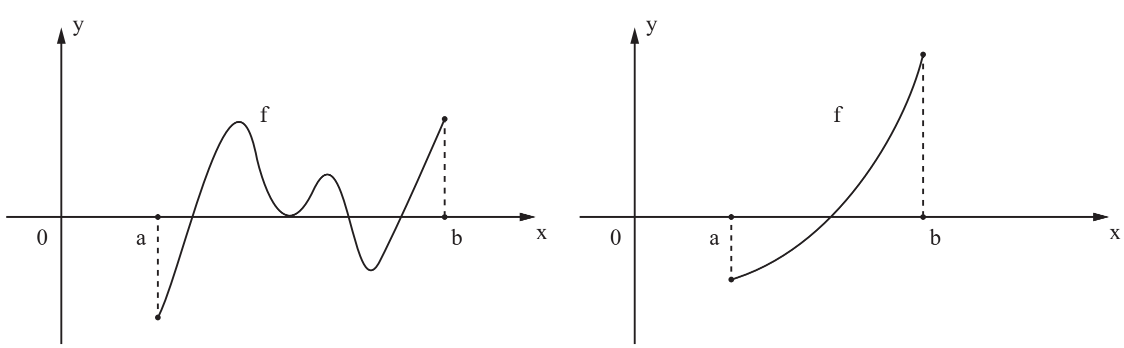
2.1 Izoliranje rješenja

Promatramo jednadžbu $f(x) = 0$, gdje je funkcija f definirana i neprekidna na nekom intervalu $\langle a, b\rangle$ (koji može biti i beskonačan). Ponekad ćemo zahtijevati od f i određenu glatkost, tj. neprekidnost od f, f' i f''.

Svaka vrijednost ξ za koju je $f(\xi) = 0$ je **korijen** ili **rješenje** jednadžbe $f(x) = 0$. Kažemo još da je ξ **nul-točka** funkcije f. Pretpostavljamo da f ima samo **izolirane** nul-točke, tj. da oko svake nul-točke postoji okolina na kojoj je $f(x) \neq 0$.



Slika 1: Primjeri funkcija s izoliranim (lijevo) i neizoliranim (desno) nul-točkama.



Slika 2: Primjeri intervala s više nultočaka i s jednom nul-točkom funkcije.

Približno rješavanje jednadžbe $f(x) = 0$ uključuje dva koraka:

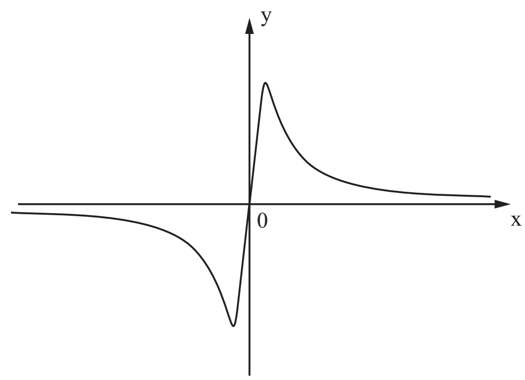
1. Izoliranje korijena, tj. određivanje intervala $[a, b]$ koji sadrže jedan i samo jedan korijen.
2. Nalaženje korijena u tim intervalima uz zadanu točnost.

Teorem 1. *Ako neprekidna funkcija $f(x)$ na krajevima intervala $[a, b]$ ima vrijednosti suprotnih predznaka, $f(a)f(b) < 0$, onda u tom intervalu postoji barem jedna njena nul-točka.* □

Nul-točka će biti **jedinstvena**, ako $f'(x)$ postoji i ne mijenja predznak na $\langle a, b\rangle$.

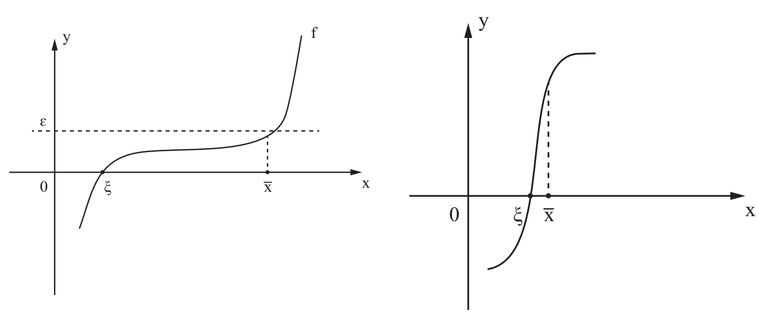
Danas je postupak izolacije rješenja bitno olakšan postojanjem softwarea za crtanje grafova funkcija. Potrebno je pritom paziti na moguće lažne korijene.

Primjer 2.1:
$f(x) = \frac{1}{x}$ na $[-1000, 1000]$. Loš software (ili loš matematičar) nacrtat će ovakav graf.



Slika 3: Loš software (ili loš matematičar) ...

Kvaliteta približnog rješenja - što je dobar kriterij? Jednom kad smo našli približno rješenje $\overline{x}$, zanima nas koliko je ono dobro, tj. koliko je ono blizu pravom rješenju ξ. Prva ideja, provjera koliko je $f(\overline{x})$ daleko od nule, nije dobra.



Slika 4: Što je dobra mjera kvalitete aproksimacije?

Teorem 2. *Neka je ξ točno, a $\overline{x}$ približno rješenje jednadžbe $f(x) = 0$ u intervalu $[a, b]$. Ako je u tom intervalu $|f'(x)| \geq m_1 > 0$, onda je*

$$|\overline{x} - \xi| \leq \frac{|f(\overline{x})|}{m_1}.$$

□

2.2 Metoda raspolavljanja (bisekcije)

Neka je funkcija $f(x)$ neprekidna na $[a, b]$ i neka je $f(a)f(b) < 0$. Dakle u $[a, b]$ postoji rješenje jednadžbe $f(x) = 0$.

Promatrajmo $f\left(\frac{a+b}{2}\right)$. Ako je $f\left(\frac{a+b}{2}\right) = 0$, našli smo rješenje. Ako nije, tada za točno jedan od intervala $\left[a, \frac{a+b}{2}\right]$ i $\left[\frac{a+b}{2}, b\right]$, funkcija ima različite predznake u rubovima. Nazovimo taj interval $[a_1, b_1]$ i ponovimo postupak. Nakon određenog broja koraka doći ćemo ili do točne vrijednosti rješenja ξ ili do tako malog intervala $[a_n, b_n]$ koji ju sadrži da nam je ta točnost dovoljna. Ako ne naiđemo na točno rješenje, dobivamo (beskonačan) niz ugniježđenih intervala $[a_1, b_1],\ [a_2, b_2],\ \ldots,\ [a_n, b_n],\ \ldots$ takvih da je

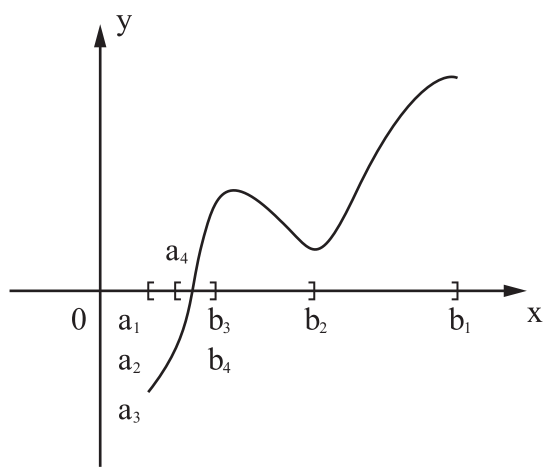
$$f(a_n)f(b_n) < 0 \quad \text{i} \quad b_n - a_n = \frac{1}{2^n}(b - a).$$

Lijevi rubovi intervala, $a_1,\ a_2,\ \ldots,\ a_n,\ \ldots$ čine monotono neopadajući niz i svi su manji od b; dakle niz (a_n) je monoton i ograničen, pa i konvergentan. Slično niz (b_n) je monoton (nerastući), ograničen odozdo s a, pa je i on konvergentan.

Kako je $a_n \le \xi \le b_n$, iz $b_n - a_n = \frac{1}{2^n}(b - a) \to 0$ slijedi da je

$$\xi = \lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n.$$

Očito je $0 \le \xi - a_n \le \frac{1}{2^n}(b - a)$ pa imamo i ocjenu pogrješke. Metoda raspolavljanja je gruba,



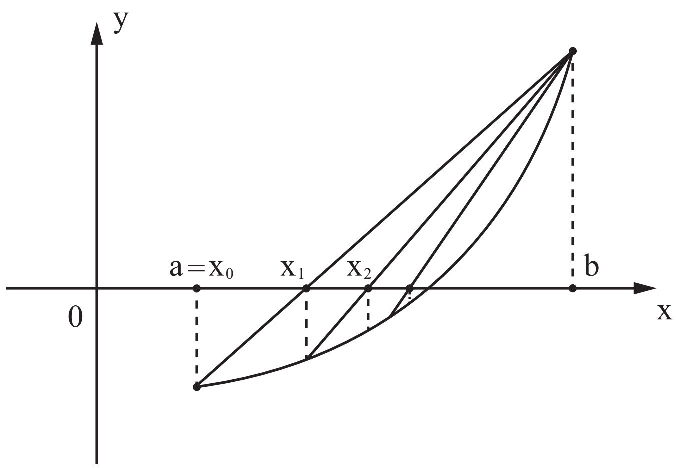
Slika 5: Metoda raspolavljanja.

robustna i laka za implementaciju na računalu. Ne zahtijeva glatkost funkcije i neosjetljiva je na strminu. Loša strana je polagana konvergencija. Koliko koraka treba za točnost od 10^{-3} za $b - a = 1$?

2.3 Metoda sekante (regula falsi)

Zamijenimo komad grafa funkcije između a i b za koje je $f(a)f(b) < 0$ sekantom i pogledajmo gdje ta sekanta siječe os x. U situaciji kao na slici to vodi na niz aproksimacija
$a = x_0 < x_1 < x_2 < \ldots < x_n < x_{n+1} < \ldots < \xi < b$, čiji je opći član definiran formulom

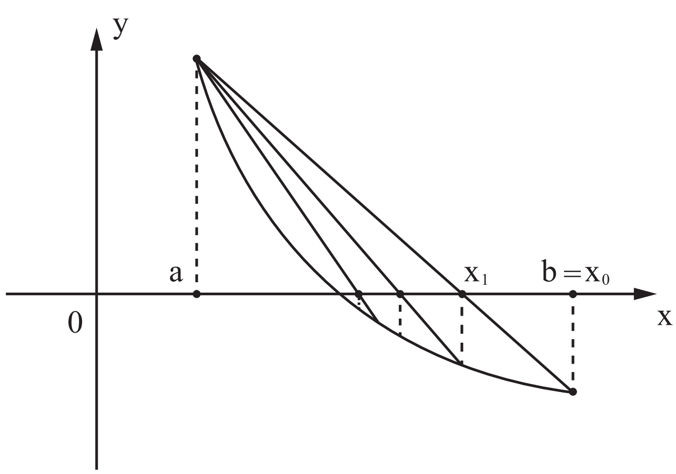
$$x_{n+1} = x_n - \frac{f(x_n)}{f(b) - f(x_n)}(b - x_n).$$



Slika 6: Metoda sekante.

Uz pretpostavku da na $[a, b]$ postoji samo jedno rješenje, gornji niz aproksimacija (x_n) konvergira (monoton je i ograničen) prema rješenju ξ.

Vidimo da rub b ostaje fiksan u ovim iteracijama. Moguće su i situacije u kojima rub a ostaje fiksan. Ako postoji $f''(x)$, onda je pravilo da ostaje fiksan onaj rub u kojem je $f(x)$ istog predznaka kao i $f''(x)$. Sve aproksimacije su s iste strane korijena ξ, i to s one strane na kojoj je predznak od $f(x)$ suprotan predznaku od $f''(x)$.



Slika 7: Metoda sekante.

Za ocjenu pogrješke imamo $|x_n - \xi| \leq \frac{f(x_n)}{m_1}$, gdje je $m_1 \leq |f'(x)|$ za $a \leq x \leq b$. Ako znamo ograde $m_1 \leq |f'(x)| \leq M_1$ na $[a, b]$, možemo dati i ocjenu pogrješke oblika

$$|\xi - x_n| \leq \frac{M_1 - m_1}{m_1} |x_n - x_{n-1}| .$$

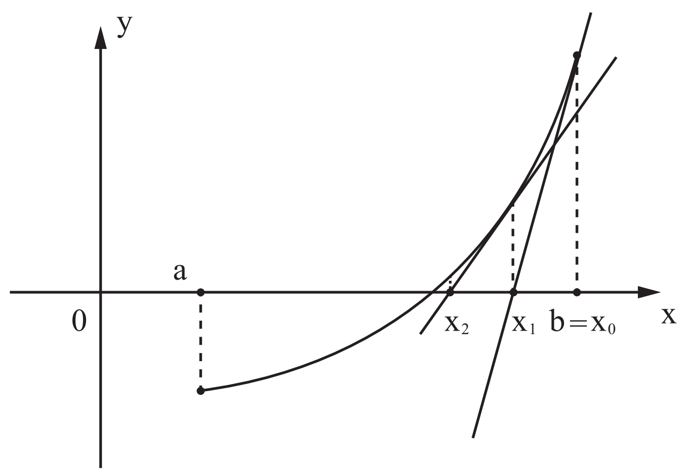
To nam daje kriterij zaustavljanja.

2.4 Newtonova metoda (metoda tangente)

Neka je korijen ξ izoliran u $[a,b]$ i neka su $f'(x)$ i $f''(x)$ neprekidne i ne mijenjaju predznak na $[a,b]$. Znamo da od svih pravaca kroz neku točku na krivulji tu krivulju najbolje aproksimiramo tangentom. Povucimo tangentu kroz točku na desnom kraju intervala $[a,b]$ (u slučaju na slici kroz $(b, f(b))$) i pogledajmo gdje ona siječe os x. Nazovimo tu točku x_1 i povucimo tangentu kroz $(x_1, f(x_1))$. Ona siječe os x u x_2. Nastavljajući taj postupak dobivamo niz

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)},$$

koji izgleda kao da konvergira prema ξ. Sa slike vidimo da bi $x_0 = a$ odnijelo x_1 izvan intervala



Slika 8: Metoda tangente.

$[a,b]$. Pravilo je uzeti x_0 za koji je $f(x_0)f''(x_0) > 0$.

Teorem 3. *Neka je $f(a)f(b) < 0$ i neka su $f'(x)$, $f''(x)$ neprekidne, različite od 0 i ne mijenjaju predznak na $[a,b]$. Tada, polazeći od x_0 za koji je $f(x_0)f''(x_0) > 0$, niz $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ konvergira prema rješenju ξ jednadžbe $f(x) = 0$.* $\square$

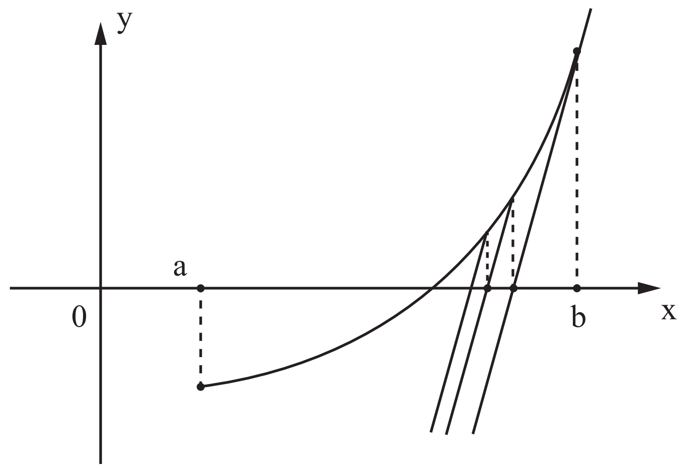
Newtonova metoda je neprikladna za računanje nul-točaka u čijoj je okolini funkcija f malog nagiba.

Ako je $m_1 \leq |f'(x)|$ i $|f''(x)| \leq M_2$ na $[a,b]$, imamo ocjenu

$$|\xi - x_n| \leq \frac{M_2}{2m_1} |x_n - x_{n-1}|^2 .$$

Kažemo da je konvergencija Newtonove metode **kvadratična**.

Nekad, dok je računanje bilo skupo, običavalo se Newtonovu metodu modificirati tako da se ne računa derivacija u svakom koraku, već se jednom izračunata vrijednost koristila u nekoliko koraka kao na slici. To rezultira sporijom konvergencijom.



Slika 9: Modificirana metoda tangente.

2.5 Metoda iteracije (metoda fiksne točke)

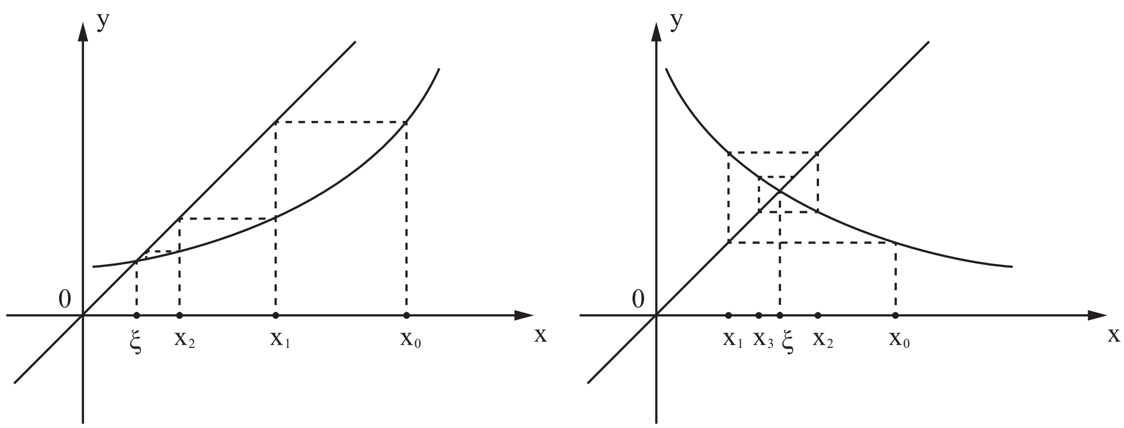
Metoda se još zove i metoda sukcesivnih aproksimacija. Promatramo jednadžbu $f(x) = 0$ i zapišemo ju u obliku $x = \varphi(x)$, gdje je φ neprekidna funkcija. Uzmimo neku početnu aproksimaciju x_0. Kad bi to bilo rješenje, vrijedilo bi $x_0 = \varphi(x_0)$. Kako nije, označimo $\varphi(x_0)$ s x_1 i ponovimo postupak:

$$x_1 = \varphi(x_0),\ x_2 = \varphi(x_1), \ldots,\ x_{n+1} = \varphi(x_n).$$

Ako taj niz konvergira, onda je

$$\underbrace{\lim_{n\to\infty} x_{n+1}}_{\xi=\varphi(\xi)} = \lim_{n\to\infty} \varphi(x_n) = \varphi(\underbrace{\lim_{n\to\infty} x_n}_{\xi}),$$

tj. onda on konvergira prema rješenju jednadžbe $f(x) = 0$.



Slika 10: Različiti načini konvergencije metode iteracija: “stubište” (lijevo) i “spirala” (desno).

Teorem 4. *Neka je funkcija $\varphi(x)$ definirana i derivabilna na $[a, b]$ i neka su joj sve vrijednosti u $[a, b]$. Ako postoji $0 < q < 1$ za koji je $|\varphi'(x)| \leq q < 1$ za sve $x \in [a, b]$, onda za svaki $x_0 \in [a, b]$ niz*

$$x_n = \varphi(x_{n-1})$$

konvergira prema (jedinom) rješenju ξ jednadžbe $x = \varphi(x)$ u $[a, b]$. □

Funkcija φ koja zadovoljava uvjete Teorema 4 je **kontrakcija**. Za ocjenu pogrješke imamo izraze

$$|\xi - x_n| \leq \frac{q^n}{1-q} \left| x_1 - x_0 \right|;$$

$$|\xi - x_n| \leq \frac{q}{1-q} \left| x_n - x_{n-1} \right|.$$

Dokaz bi slijedio iz teorema srednje vrijednosti.

3 Aproksimacija i interpolacija

3.1 Problem aproksimacije funkcija

Promatramo funkciju $f : I \to \mathbb{R}$. Ponekad je potrebno umjesto funkcije f promatrati neku drugu funkciju φ koja joj je u izvjesnom smislu bliska. Ako aproksimacijska funkcija φ osim o x ovisi još i o parametrima $a_0, a_1, \ldots, a_n$, tj.

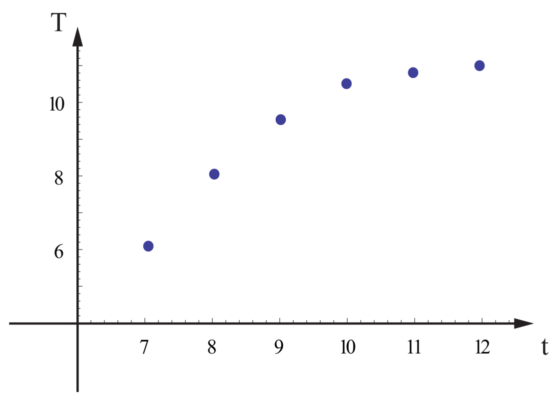
$$\varphi(x) = \varphi(x, a_0, a_1, \ldots, a_n),$$

onda se problem aproksimacije funkcije f funkcijom φ svodi na određivanje parametara $a_0, \ldots, a_n$ prema nekom zadanom kriteriju. U ovisnosti o odabranom kriteriju imamo razne vrste aproksimacija.

Tipična situacija u kojoj imamo gornji problem je kad su vrijednosti nepoznate funkcije f poznate (ili dostupne) samo na diskretnom skupu $x_0, x_1, \ldots, x_n$, a trebaju nam (približne) vrijednosti te funkcije u točkama koje nisu iz tog skupa.

Primjer 3.1:
Temperatura zraka Kolika je bila temperatura u 9 sati i 45 minuta? A u 16 sati i 15 minuta?

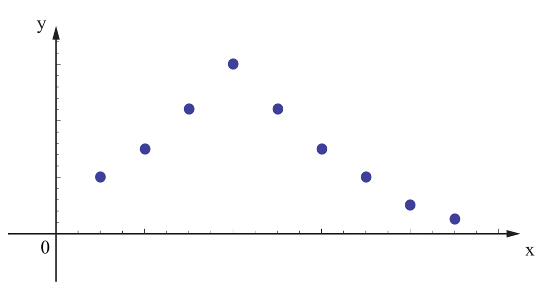
t	7	8	9	10	11	12
T	6	8	9.5	10.5	10.8	11



Slika 11: Tablični i grafički prikaz podataka o temperaturi.

Druga tipična situacija je kad je funkcija f komplicirana i/ili skupa za računanje. Recimo vrijednost funkcije u točki x_0 zahtijeva 2 tjedna računa.

Kvaliteta (i smislenost) aproksimacije ovise o izboru oblika funkcije φ. Izbor neodgovarajuće funkcije daje, u pravilu, besmislen ili, u najboljem slučaju, jako loš rezultat.



Slika 12: Kakva je funkcija pogodna za aproksimaciju ovih podataka?

3.2 Osnovni tipovi aproksimacijskih funkcija

Osnovna podjela je na **linearne** i **nelinearne** aproksimacijske funkcije. Opći oblik linearne aproksimacijske funkcije je

$$\varphi(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \ldots + a_n\varphi_n(x)$$

pri čemu funkcije $\varphi_0,\ \varphi_1,\ \ldots,\ \varphi_n$ zadovoljavaju određene uvjete. Linearnost se odnosi na parametre $a_0,\ \ldots,\ a_n$, oni u formulu ulaze linearno.

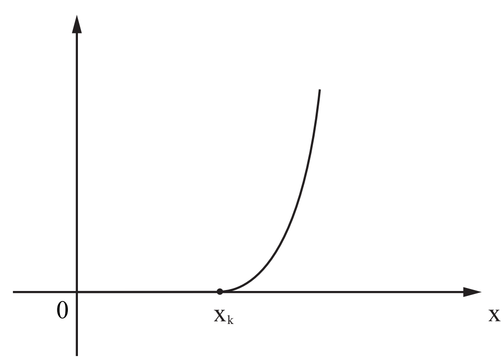
Za $\varphi_k(x) = x^k$ imamo $\varphi(x) = a_0 + a_1 x + \ldots + a_n x^n$, dakle aproksimaciju (algebarskim) polinomima.

Za $\{\varphi_k(x)\} = \{1,\ \cos x,\ \sin x,\ \cos 2x,\ \sin 2x, \ldots\}$ imamo aproksimaciju trigonometrijskim polinomima.

Uzmemo li

$$\varphi_k(x) = (x - x_k)_+^m = \begin{cases} (x - x_k)^m & , \quad x \geq x_k \\ 0 & , \quad x < x_k \end{cases},$$

imamo aproksimaciju splineovima. Funkcije $(x-x_k)_+^m$ zovu se **prikraćene potencije** (truncated powers).



Slika 13: Prikraćena potencija.

Od nelinearnih aproksimacijskih funkcija često se koriste **eksponencijalna**

$$\begin{aligned}\varphi(x) =& \varphi(x; c_0,\ b_0, \ldots,\ c_r, b_r) = c_0 e^{b_0 x} + \ldots + c_r e^{b_r x} \\ & (n + 1 = 2r + 2, \text{tj. } n = 2r + 1)\end{aligned}$$

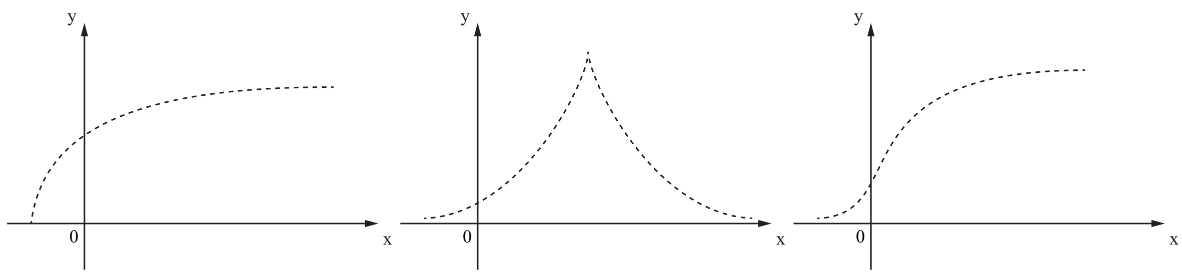
i **racionalna aproksimacija**

$$\varphi(x) = \varphi(x;\ b_0, \ldots,\ b_r,\ c_0, \ldots,\ c_s) = \frac{b_0 + b_1 x + \ldots + b_r x^r}{c_0 + c_1 x + \ldots + c_s x^s}$$

za koju je $n = r + s + 1$.

Izbor tipa aproksimacijske funkcije ovisi o poznavanju prirode problema koji generira podatke i iskustvu.

Primjer 3.2:

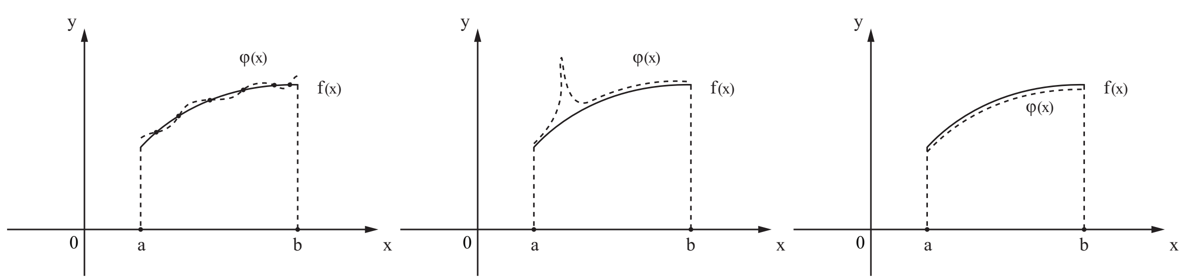


Slika 14: Kakve su funkcije pogodne za aproksimaciju ovih podataka?

3.3 Kriteriji optimalnosti

Postoje različiti načini mjerenja "bliskosti" dviju funkcija. U nekim slučajevima nam je bitno podudaranje vrijednosti na nekom skupu točaka. U nekim drugim slučajevima su bolje druge mjere.

Primjer 3.3:



Slika 15: Različiti kriteriji kvalitete aproksimacije.

Kvaliteta aproksimacije se mjeri malošću razlike funkcije i aproksimacije u odabranoj normi u funkcijskom prostoru kojem pripadaju.

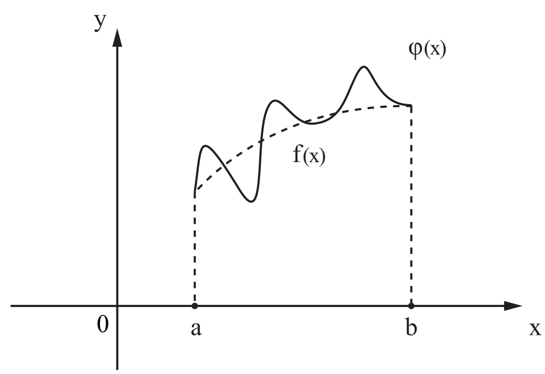
Ako od aproksimacijske funkcije zahtijevamo podudaranje sa zadanim vrijednostima na konačnom (dakle diskretnom) skupu točaka, onda je kriterij za izbor parametara zadovoljavanje sustava jednadžbi

$$\varphi(x_k;\ a_0, \ldots,\ a_n) = f(x_k), \quad k = 0, \ldots,\ n.$$

Takva aproksimacija se zove **interpolacija** funkcije f. Funkciju φ zovemo **interpolacijskom funkcijom**, a točke $x_k,\ k = 0, \ldots,\ n$, u kojima se njene vrijednosti podudaraju sa zadanima, zovemo **čvorovima interpolacije**.

Kriterij kod interpolacije je poništavanje razlike funkcije i njene aproksimacije na zadanom diskretnom (štoviše, konačnom) skupu. Aproksimacija koja je dobra po tom kriteriju ne mora biti dobra ako gledamo odstupanja u drugim točkama ili razliku površina ispod grafova funkcije i njene aproksimacije.

Primjer 3.4:



Slika 16: Interpolacija.

Odstupanje u čvorovima je 0, no što ako nas zanima $|f(x) - \varphi(x)|$ na $[a, b]$? Što ako nas zanima $\int_a^b f(x)dx$? U tom slučaju je bolje uzeti druge mjere bliskosti:

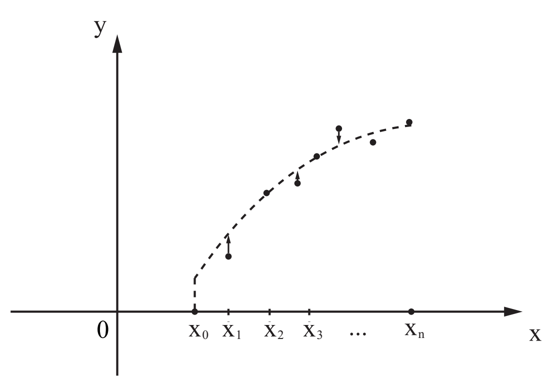
$$\max_{x\in[a,b]} |f(x) - \varphi(x)|, \text{ ili } \int_a^b (f(x) - \varphi(x))^2 dx$$

Te mjere daju tzv. min-max aproksimaciju i srednje-kvadratnu aproksimaciju, redom. Te se mjere mogu definirati i na diskretnim skupovima.

Popularna je metoda aproksimacije zvana diskretna metoda najmanjih kvadrata, u kojoj se minimizira suma kvadrata odstupanja u zadanim točkama.

Primjer 3.5:

Odabrati funkciju zadanog oblika tako da



Slika 17: Diskretna metoda najmanjih kvadrata.

$$\sum_{k=0}^{n} (f(x_k) - \varphi(x_k;\ a_0, \dots,\ a_n))^2 \to \min.$$

Koliko su ovakvi problemi rješivi? Pokazuje se da se svaka funkcija koja je neprekidna na $[a, b]$ može po volji dobro aproksimirati polinomom.

Teorem 5. *(Weierstrass) Ako je funkcija $f \in C[a, b]$, onda $(\forall \varepsilon > 0)(\exists n \in \mathbb{N})$ i polinom $P_n(x)$ stupnja n takav da je $\forall x \in [a, b]$*

$$|f(x) - P_n(x)| \leq \varepsilon.$$

□

Teorem ne kaže ništa o tome kako naći takav polinom, čak ni kojeg je stupnja.

3.4 Polinomijalna interpolacija

3.4.1 Rješivost problema interpolacije

Promatramo skup čvorova $x_k,\ k = 0, \ldots, n$ u intervalu $[a, b]$ u kojima su zadane vrijednosti (nepoznate) funkcije f, $f(x_k) = f_k$. Promatramo problem interpolacije linearnom aproksimacijskom funkcijom oblika $\varphi(x) = \varphi(x;\ a_0, \ldots, a_n)$. Dobivamo sustav jednadžbi

$$\begin{array}{c}
a_0\varphi_0(x_0) + a_1\varphi_1(x_0) + \ldots + a_n\varphi_n(x_0) = f_0 \\
a_0\varphi_0(x_1) + a_1\varphi_1(x_1) + \ldots + a_n\varphi_n(x_1) = f_1 \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
a_0\varphi_0(x_n) + a_1\varphi_1(x_n) + \ldots + a_n\varphi_n(x_n) = f_n
\end{array}$$

Taj sustav ima jedinstveno rješenje ako i samo ako mu je matrica regularna. Matricu čine vrijednosti baznih funkcija $\varphi_0, \ldots, \varphi_n$ u čvorovima $x_0, \ldots,\ x_n$. Nepoznanice su parametri $a_0, \ldots,\ a_n$.

Pokazuje se da za $a \leq x_0 < x_1 < \ldots < x_n \leq b$ bazne funkcije $\varphi_k(x) = x^k$ uvijek daju regularnu matricu. Dakle problem interpolacije (algebarskim) polinomima uvijek ima jedinstveno rješenje.

Teorem 6. *Polinom $P_n(x)$ koji interpolira funkciju f u točkama $x_0 < x_1 < \ldots < x_n$ je jedinstven i može se prikazati u obliku*

$$P_n(x) = \sum_{k=0}^{n} f(x_k)L_k(x),$$

gdje je

$$L_k(x) = \frac{(x - x_0)(x - x_1)\cdots(x - x_{k-1})(x - x_{k+1})\cdots(x - x_n)}{(x_k - x_0)(x_k - x_1)\cdots(x_k - x_{k-1})(x_k - x_{k+1})\cdots(x_k - x_n)}.$$

□

3.4.2 Lagrangeov oblik interpolacijskog polinoma

Uvedemo li oznaku

$$\omega(x) = \prod_{k=0}^{n}(x - x_k) = (x - x_0)\cdot\cdots\cdot(x - x_n),$$

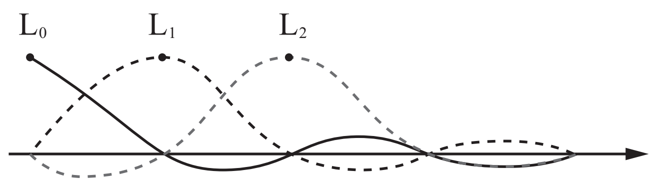
imamo

$$L_k(x) = \frac{\omega(x)}{(x - x_k)\omega'(x_k)}.$$

Zašto? Polinomi $L_k(x)$ imaju svojstvo

$$L_k(x_j) = \delta_{kj} = \begin{cases} 1 & , \quad j = k \\ 0 & , \quad j \neq k \end{cases} .$$

Zovemo ih **bazni polinomi**.



Slika 18: Bazni polinomi za Lagrangeovu interpolaciju.

Znamo li maksimum apsolutne vrijednosti $(n+1)$-ve derivacije od f, $M_{n+1} = \max_{x\in[a,b]} \left| f^{(n+1)}(x) \right|$, možemo dati i ocjenu pogrješke:

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

Problem je da za tabelirane vrijednosti (x_k i f_k) ne znamo što bi mogla biti $f^{(n+1)}(x)$.

Zadatak 3.1:

Odredimo Lagrangeov interpolacijski polinom za podatke iz tablice

x_k	-1	0	2	3
$f(x_k)$	-1	2	10	35

Rješenje 3.1:

Ovdje je $x_0 = -1$, $x_1 = 0$, $x_2 = 2$ i $x_3 = 3$.

$$\begin{aligned} L_0(x) &= \frac{(x-0)(x-2)(x-3)}{(-1-0)(-1-2)(-1-3)} = -\frac{1}{12}x(x-2)(x-3) \\ L_1(x) &= \frac{(x+1)(x-2)(x-3)}{(0-(-1))(0-2)(0-3)} = \frac{1}{6}(x+1)(x-2)(x-3), \quad \text{itd.} \end{aligned} \tag{3.1}$$

Konačno, $P_3(x) = \frac{5}{3}x^3 - \frac{4}{3}x^2 + 2$. □

Ako su čvorovi ekvidistantni, $x_k - x_{k-1} = h$, ocjenu greške možemo dati u obliku $|f(x) - P_n(x)| < \frac{M_{n+1}}{(n+1)} h^{n+1}$.

Lagrangeov oblik je konceptualno jednostavan, ali nije pogodan za numeriku. Posebno, dodavanje novog čvora znači računanje svega ispočetka.

3.4.3 Newtonov oblik interpolacijskog polinoma

Promatramo funkciju f zadanu svojim vrijednostima u čvorovima $x_k, k = 0, \ldots, n$. **Podijeljena razlika** (reda 1, prva) funkcije f u točkama x_0 i x_1 je kvocijent

$$f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0}.$$

Podijeljene razlike višeg reda definiraju se rekurzivno:

$$f[x_0, \ldots, x_r] = \frac{f[x_1, \ldots, x_r] - f[x_0, \ldots, x_{r-1}]}{x_r - x_0}.$$

Počinjemo s $f[x_k] = f_k$, $k = 0, \ldots, n$. Podijeljene razlike imaju svojstvo linearnosti, tj.

$$(\alpha f + \beta g)[x_0, \ldots, x_r] = \alpha f[x_0, \ldots, x_r] + \beta g[x_0, \ldots, x_r].$$

Dalje, iz definicije podijeljene razlike vidimo da je

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

Odatle se indukcijom može pokazati da vrijedi

$$f[x_0, \ldots, x_r] = \sum_{i=0}^{r} \frac{f(x_i)}{\omega'(x_i)},$$

gdje je $\omega(x) = (x - x_0) \cdots (x - x_r)$.

Pomoću podijeljenih razlika možemo dati alternativni zapis interpolacijskog polinoma za vrijednosti $f(x_k)$ u čvorovima x_k:

$$\begin{aligned} P_n(x) = {} & f[x_0] + (x - x_0) f[x_0, x_1] + (x - x_0)(x - x_1) f[x_0, x_1, x_2] + \ldots \\ & + (x - x_0)(x - x_1) \cdot \ldots \cdot (x - x_{n-1}) f[x_0, \ldots, x_n]. \end{aligned}$$

To je Newtonov oblik interpolacijskog polinoma. Za njegovo računanje nam treba tablica podijeljenih razlika:

$$\begin{array}{cccccc}
x_0 & \underline{f[x_0]} & & & & \\
 & & \underline{f[x_0, x_1]} & & & \\
x_1 & f[x_1] \; \cdot^{\cdot^{\cdot}} & & \underline{f[x_0, x_1, x_2]} & & \\
 & & f[x_1, x_2] \; \cdot^{\cdot^{\cdot}} & & \underline{f[x_0, x_1, x_2, x_3]} & \\
x_2 & f[x_2] \; \cdot^{\cdot^{\cdot}} & & f[x_1, x_2, x_3] \; \cdot^{\cdot^{\cdot}} & & \ddots \\
 & & f[x_2, x_3] \; \cdot^{\cdot^{\cdot}} & & \vdots & \underline{f[x_0, x_1, \ldots, x_n]} \\
x_3 & f[x_3] \; \cdot^{\cdot^{\cdot}} & & \vdots & & \cdot^{\cdot^{\cdot}} \\
\vdots & \vdots & \vdots & \vdots & \cdot^{\cdot^{\cdot}} & \\
\vdots & \vdots & \vdots & f[x_{n-2}, x_{n-1}, x_n] & & \\
\vdots & \vdots & f[x_{n-1}, x_n] \; \cdot^{\cdot^{\cdot}} & & & \\
x_n & f[x_n] \; \cdot^{\cdot^{\cdot}} & & & &
\end{array}$$

Za računanje interpolacijskog polinoma nam trebaju samo podvučene podijeljene razlike iz gornje stranice trokuta.

Računamo podijeljene razlike po uzlaznim dijagonalama. Na taj način kod dodavanja nove točke možemo koristiti već izračunate koeficijente:

$$P_{n+1}(x) = P_n(x) + (x - x_0)(x - x_1) \cdot \ldots \cdot (x - x_n) f[x_0, x_1, \ldots, x_{n+1}].$$

Ako su f_k vrijednosti funkcije kojoj znamo $(n+1)$-vu derivaciju, možemo dati ocjenu pogrješke:

$$R_n(x, f) = f(x) - P_n(x) = \omega(x) f[x_0, x_1, \ldots, x_n, x].$$

Što dobijemo kad pustimo $x_k \to x_0$ za sve k? Taylorov polinom! Iz podijeljenih razlika se za ekvidistantne čvorove dobiju konačne razlike. Pomoću konačnih razlika organiziranih u trokutastu tablicu može se izvesti više interpolacijskih formula, npr. Gaussova, Besselova, Stirlingova, druga Newtonova itd.

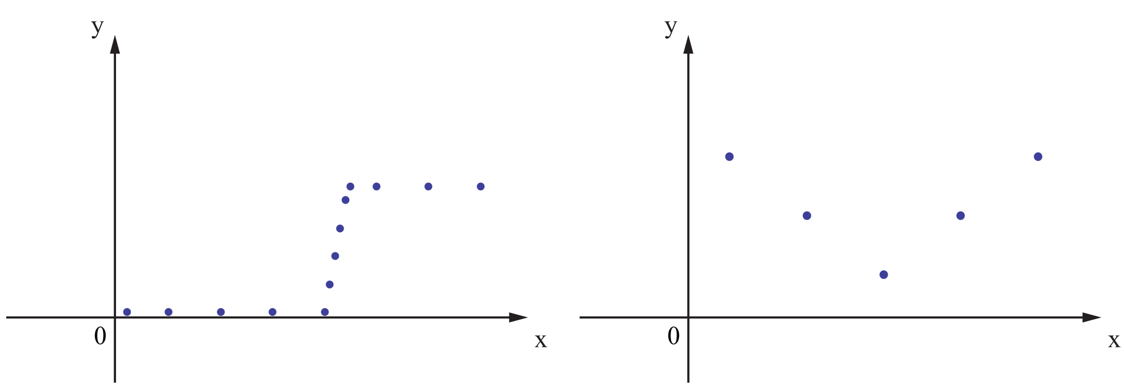
Mogu se gledati i interpolacijski problemi u kojima su u pojedinim čvorovima zadane ne samo vrijednosti funkcije već i njenih derivacija. Polinomijalna interpolacija u takvom slučaju zove se Hermiteova interpolacija.

3.4.4 Problemi s polinomijalnom interpolacijom

Interpolacijski polinomi visokog stupnja ($n > 3$) daju jako velika odstupanja između točaka; veliko je $|f(x) - P_n(x)|$. Velike su oscilacije pogotovo blizu rubova intervala. Vrlo su strmi, što ih čini beskorisnima za procjenu derivacije od f. Osjetljivi su na "outliere", tj. slučajnu kontaminaciju podataka.

Ekstrapolacija je aproksimacija vrijednosti nepoznate funkcije izvan intervala u kojem su nam neke njene vrijednosti poznate. Polinomijalna ekstrapolacija u pravilu ne daje dobre rezultate.

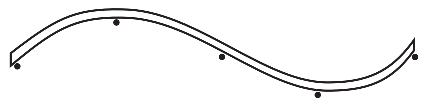
Primjer 3.6:



Slika 19: Primjeri skupova podataka koji se loše interpoliraju polinomima.

Rješenje - nepolinomijalna ili po dijelovima polinomijalna interpolacija. To nas vodi do **splineova**.

3.4.5 Splineovi



Slika 20: Spline.

Ime dolazi od elastične drvene ili metalne letvice (engl. *spline*) koja je nekad korištena za crtanje glatkih krivulja koje prolaze kroz zadane točke. Prolazimo od pretpostavke da letvica zauzima oblik u kojem se njena potencijalna energija minimizira. Potencijalna energija dolazi od elastičnosti i proporcionalna je integralu po splineu kvadrata njegove zakrivljenosti. Ako je oblik splinea opisan krivuljom $\Gamma = f(x)$, onda je potencijalna energija proporcionalna integralu:

$$\int_\Gamma \kappa(s)^2 ds = \int_a^b \frac{f''(x)^2}{(\sqrt{1+f'(x)^2})^5} dx = E[f].$$

Oblik koji spline zauzima je funkcija f koja minimizira funkciju E uz nametnuta ograničenja.

Za male vrijednosti od $f'(x)$ možemo zanemariti član $f'(x)^2$, pa u nazivniku dobijemo 1. Dakle minimiziramo integral (funkcional)

$$E[f] = \int_a^b f''(x)^2 dx.$$

Matematička teorija splineova i njihova primjena počinju se razvijati krajem prve polovice 20-og stoljeća (Collatz, Courant, Schoenberg). Mi ćemo promatrati **polinomske splineove**, tj. splineove koji su na svakom segmentu $[x_{i-1}, x_i]$ polinomi zadanog sutpnja i koji se u čvorovima sastaju tako da se čuva zadana glatkost.

Funkcija $S_m^k(x)$ je **polinomski spline** stupnja m i defekta k ($1 \le k \le m$) s čvorovima $a = x_0 < x_1 < \ldots < x_n = b$, ako vrijedi

(i) $S_m^k(x)$ je polinom stupnja (najviše) m na svim $[x_{i-1}, x_i]$

(ii) $S_m^k(x) \in C^{m-k}[a, b]$.

Obično se gledaju polinomski splineovi defekta 1. To znači da im je m-ta derivacija možda prekidna u čvorovima. Najjednostavniji slučaj su splineovi stupnja 0 i 1, prikazani na slici 21, dok se u praksi najčešće pojavljuju kubični i B-splineovi.

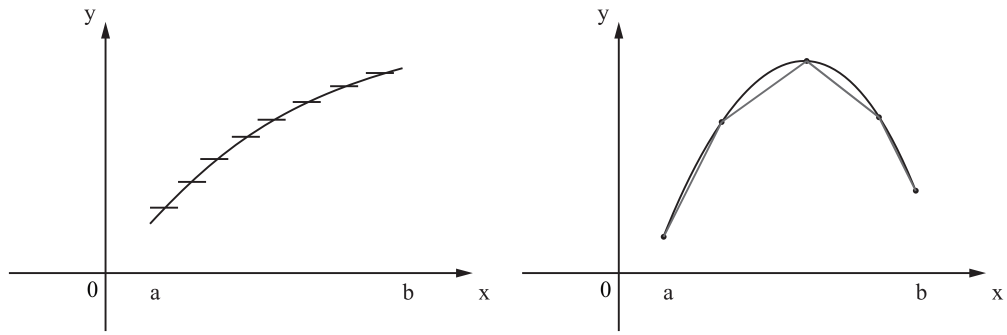
Kubični interpolacijski spline za funkciju f na čvorovima $a = x_0 < x_1 < \ldots < x_n = b$ je funkcija $S_3(x)$ koja zadovoljava sljedeće uvjete

(i) $S_3(x)$ je polinom stupnja najviše 3 na svim $[x_{i-1}, x_i]$, $\quad i = 1, \ldots, n$;

(ii) $S_3(x) \in C^2[a, b]$;

(iii) $S_3(x_i) = f(x_i), \quad i = 0, \ldots, n.$

Uvjet (iii) ga čini interpolacijskim splineom.



Slika 21: Splineovi stupnja 0 (lijevo) i 1 (desno).

Kubični spline interpolira funkciju f u čvorovima, neprekidan je zajedno sa svojom 1. i 2. derivacijom i na svakom je podintervalu polinom stupnja ne većeg od 3. Uvjet neprekidnosti 2. derivacije je vrlo bitan u mehaničkim primjenama (ubrzanje, tj. sila, ne smije imati skokove) i u računalnoj grafici.

Kubični spline se konstruira iz zadanih vrijednosti u čvorovima i uvjeta neprekidnosti derivacija. Kad se oni svi uzmu u obzir, ostaju dva slobodna parametra koje se fiksira dodatnim uvjetima. To su obično uvjeti tipa

(i) $S_3'(a) = S_3'(b), \quad S_3''(a) = S_3''(b)$ - periodički spline

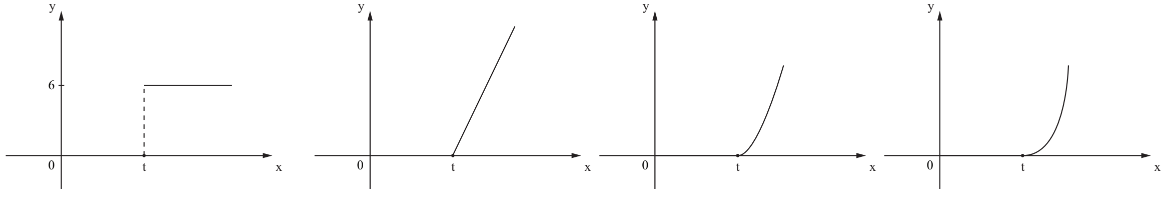
(ii) $S_3'(a) = a, \quad S_3'(b) = b$

(iii) $S_3''(a) = A, \quad S_3''(b) = B.$

U uvjetima tipa (iii) često se stavlja $S_3''(a) = f''(a)$, $S_3''(b) = f''(b)$, ako su vrijednosti druge derivacije na krajevima poznate. U mehaničkim modelima je često $A = B = 0$ pa se kubični spline s takvim uvjetima zove **prirodni kubični spline**.

Kako se splineovi matematički prikazuju? Koristi se baza **prikraćenih** potencija

$$(x-t)_+^k = \begin{cases} (x-t)^k & , \quad x \geq t \\ 0 & , \quad x < t \end{cases}$$

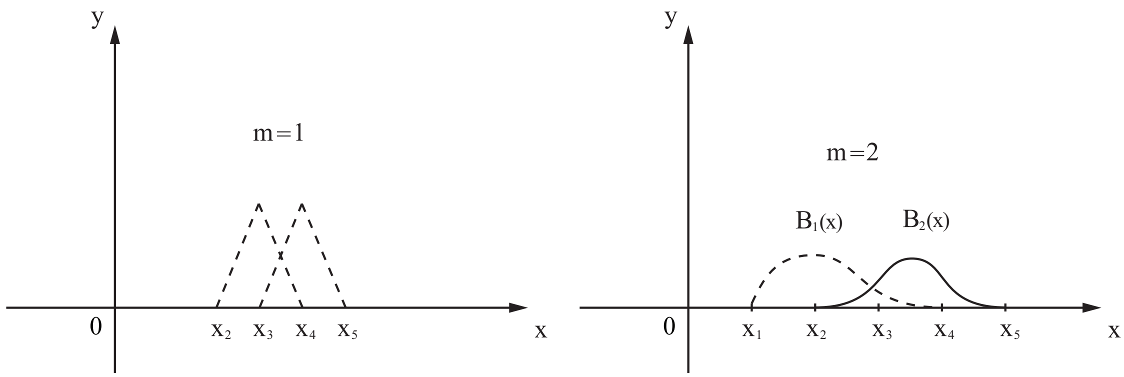


(a) $6(x-t)_+^0$ (b) $6(x-t)_+^1$ (c) $3(x-t)_+^2$ (d) $(x-t)_+^3$

Slika 22: Baza prikraćenih potencija.

Interpolacija kubičnim splineovima je globalna - spline svuda ovisi o vrijednosti u svakom čvoru. Alternativni način je prikaz pomoću B-splineova. **B-spline** je po dijelovima polinom stupnja m koji je različit od 0 samo na $m+1$ podintervalu (lokalnost).

Primjer 3.7:



Slika 23: B-splineovi stupnja 1 (lijevo) i 2 (desno).

$S(x) = \sum_i a_i \cdot B_i(x)$ je prikaz splinea $S(x)$ u bazi B-splineova. Prikaz preko B-splineova je kompaktan, numerički stabilan, lako se računa (lako za računalo). Mana im je da su komplicirani za manipulaciju - korisnici se moraju osloniti na standardne programske pakete.

Osim u interpolaciji, splineovi se koriste i u drugim tipovima aproksimacija - srednje kvadratnoj, metodi konačnih elemenata itd. Poopćavaju se na dvije i više dimenzija.

3.4.6 Zaključak

Polinomijalnom interpolacijom se služimo u tri standardne situacije:

1. Izvodi formula - numeričko integriranje, deriviranje, približno rješavanje jednadžbi.

2. Lokalna zamjena/manipulacija podatcima - treba imati neku ideju o tome kako se funkcija (ili tabelirani podatci) lokalno ponaša - treba formulirati model. Polinomi nekog stupnja su obično dobar model ako u blizini nema singulariteta i/ili kontaminacija.

3. Globalna zamjena/manipulacija podatcima - polinomi su rijetko kad dobar globalni model. Splineovi su bolji, no čak i njima je teško interpolirati mnoge situacije koje se često javljaju. U globalnom je slučaju bolje ići na druge tipove aproksimacija (npr. metoda najmanjih kvadrata i sl.).

4 Numeričko integriranje

4.1 Uvod

Za funkciju f neprekidnu na $[a,b]$ za koju znamo primitivnu funkciju F, određeni integral $\int_a^b f(x)dx$ računamo po Newton-Leibnizovoj formuli:

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

(Za primitivnu funkciju $F(x)$ vrijedi $F'(x) = f(x)$).

U mnogim slučajevima primitivnu funkciju nije moguće izraziti preko elementarnih funkcija; u nekim drugim slučajevima, čak i kad primitivna funkcija može biti izražena preko elementarnih funkcija, formule mogu biti prekomplicirane i/ili nepraktične. Osim toga, u praksi je $f(x)$ često zadana tablično pa se i ne može govoriti o primitivnoj funkciji. Stoga je bitno razviti metode za približno računanje određenih integrala.

Kako je računanje određenog integrala, ustvari, računanje površine nekog lika, numerička integracija se često zove i **numerička (mehanička) kvadratura**. Formule za približno računanje određenih integrala zovu se i **kvadraturne formule**. Kad je riječ o računanju dvostrukih integrala, imamo mehaničku kubaturu i **kubaturne formule**.

Problem numeričke kvadrature je problem računanja određenog integrala na temelju niza (konačnog) vrijednosti integranda. Standardni pristup je zamjena integranda $f(x)$ na segmentu $[a,b]$ interpolacijskom ili aproksimacijskom funkcijom $\varphi(x)$ tako da bude približno zadovoljena jednakost

$$\int_a^b f(x)dx \approx \int_a^b \varphi(x)dx.$$

Pri tome se funkcija $\varphi(x)$ odabire tako da se $\int_a^b \varphi(x)dx$ može izravno i točno izračunati. Ako je $f(x)$ zadana analitički, dobro bi bilo znati procijeniti pogrješku.

Pretpostavimo da za funkciju $f(x)$ znamo vrijednost u $n+1$ točki $x_0, x_2, \ldots, x_n \in [a,b]$. Koristeći vrijednost $y_i = f(x_i)$ možemo konstruirati Lagrangeov interpolacijski polinom

$$P_n(x) = \sum_{i=0}^{n} L_i(x)y_i,$$

gdje je

$$L_i(x) = \frac{\omega(x)}{(x-x_i)\omega'(x_i)}.$$

Odatle,

$$\int_a^b f(x)dx = \int_a^b P_n(x)dx + R_n[f],$$

gdje je $R_n[f]$ pogrješka. Približna formula je

$$\int_a^b f(x)dx = \sum_{i=0}^{n} A_i y_i,$$

gdje je

$$A_i = \int_a^b L_i(x)dx, \quad i = 0, 1 \ldots, n.$$

Ako su granice a, b među interpolacijskim čvorovima, ove su kvadraturne formule **zatvorenog tipa**; inače su **otvorenog tipa.**

Kako računamo A_i? Uočimo sljedeće:

(i) A_i su neovisni o funkciji f, ovise samo o čvorovima interpolacije.

(ii) Gornja formula mora biti **točna** za polinome stupnja najviše n (jer je interpolacijski polinom jedinstven). Odatle slijedi da formula mora biti točna za sve x^k, $k = 0, \ldots, n$, tj. $R_n[x^k] = 0$ za $k = 0, \ldots, n$.

Uvrštavajući $f(x) = x^k$, $k = 0, \ldots, n$, u kvadraturnu formulu, dobivamo sustav od $n + 1$ jednadžbi za nepoznate koeficijente od A_i:

$$\begin{aligned} \sum_{i=0}^{n} A_i &= I_0 \\ \sum_{i=0}^{n} A_i x_i &= I_1 \\ &\vdots \\ \sum_{i=0}^{n} A_i x_i^n &= I_n, \end{aligned}$$

gdje su desne strane točni integrali,

$$I_k = \int_a^b x^k dx = \frac{b^{k+1} - a^{k+1}}{k+1}.$$

Gornji sustav uvijek ima rješenje (za čvorove x_i koji su svi različiti), jer je determinanta matrice tog sustava Vandermondeova determinanta:

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

Uočimo da se ne mora eksplicitno računati sam interpolacijski polinom $P_n(x)$.

Primjer 4.1:

Nađimo kvadraturnu formulu oblika

$$\int_0^1 f(x)dx = A_0 f\left(\frac{1}{4}\right) + A_1 f\left(\frac{1}{2}\right) + A_2 f\left(\frac{3}{4}\right).$$

Rješenje 4.1:
Uvrstimo $x^0,\ x^1$ i x^2 u formulu gore i uvrštavajući $\int_0^1 x^k dx = \frac{1}{k+1}$, dobivamo sustav jednadžbi

$$\begin{aligned} A_0 + A_1 + A_2 &= 1 \\ \frac{1}{4}A_0 + \frac{1}{2}A_1 + \frac{3}{4}A_2 &= \frac{1}{2} \\ \frac{1}{16}A_0 + \frac{1}{4}A_1 + \frac{9}{16}A_2 &= \frac{1}{3}. \end{aligned}$$

Rješenje je $A_0 = \frac{2}{3}$, $A_1 = -\frac{1}{3}$ i $A_2 = \frac{2}{3}$. Dakle je

$$\int_0^1 f(x)dx = \frac{2}{3}f\left(\frac{1}{4}\right) - \frac{1}{3}f\left(\frac{1}{2}\right) + \frac{2}{3}f\left(\frac{3}{4}\right) = \frac{1}{3}\left(2f\left(\frac{1}{4}\right) - f\left(\frac{1}{2}\right) + 2f\left(\frac{3}{4}\right)\right).$$

Dobivena formula je otvorenog tipa i točna je za sve polinome stupnja najviše 2. Što se događa uvrstimo li u nju $f(x) = x^3$?

$$\int_0^1 x^3 dx = \frac{1}{3}\left(\frac{2}{64} - \frac{1}{8} + \frac{54}{64}\right) = \frac{1}{3}\left(\frac{48}{64}\right) = \frac{1}{4} \quad !$$

Vidimo da je formula točna i za polinome 3. stupnja! □

4.2 Newton-Cotesove kvadraturne formule

Formule kakve smo izveli u prijašnjem primjeru pripadaju formulama Newton-Cotesovog tipa. Primjenjuju se kad je funkcija zadana u fiksnim čvorovima koje ne možemo birati po volji.

Promatrajmo $\int_a^b f(x)dx$. Podijelimo $[a, b]$ na n podintervala ekvidistantnim točkama $x_0 = a, x_1, \ldots, x_n = b$. Vrijedi $x_i - x_{i-1} = h$, tj. $x_i = x_0 + ih$, $i = 0, \ldots, n$. Tada je $h = \frac{b-a}{n}$. Označimo $y_i = f(x_i)$, $i = 0, \ldots, n$. Zamjenjujući $f(x)$ Lagrangeovim interpolacijskim polinomom dobivamo

$$\int_{x_0}^{x_n} f(x)dx \approx \sum_{i=0}^{n} A_i y_i.$$

Uvodeći oznaku $q = \frac{x-x_0}{h}$ možemo prijeći na novu varijablu integracije. Primijetimo da za $x = x_i$ imamo $q = i$. Dalje, $q^{[n+1]} = q(q-1)\cdots(q-n)$, je **padajuća potencija** od q. Lagrangeovi interpolacijski polinom u varijabli q sada ima oblik

$$P_n(x) = \sum_{i=0}^{n} \frac{(-1)^{n-i}}{i!(n-i)!} \frac{q^{[n+1]}}{q-i} y_i$$

Uvrštavajući to u kvadraturnu formulu dobivamo

$$A_i = \int_{x_0}^{x_n} \frac{(-1)^{n-i}}{i!(n-i)!} \frac{q^{[n+1]}}{q-i} dx$$

Prijeđemo na integraciju po q: $q = \frac{x-x_0}{h}$, $dq = \frac{dx}{h}$, $dx = hdq$

$$A_i = \int_0^n h \frac{(-1)^{n-i}}{i!(n-i)!} \frac{q^{[n+1]}}{q-i} dq = h \frac{(-1)^{n-i}}{i!(n-i)!} \int_0^n \frac{q^{[n+1]}}{q-i} dq, \quad i = 0, 1, \ldots, n.$$

Iz $h = \frac{b-a}{n}$ stavimo $A_i = (b-a)H_i$, gdje su

$$H_i = \frac{1}{n}\frac{(-1)^{n-i}}{i!(n-i)!}\int_0^n \frac{q^{[n+1]}}{q-i}dq, \quad i = 0, 1, \ldots, n,$$

Cotesovi koeficijenti.Dakle je

$$\int_a^b f(x)dx = (b-a)\sum_{i=0}^{n} H_i y_i,$$

gdje je $h = \frac{b-a}{n}$, $y_i = f(x_0 + ih)$,

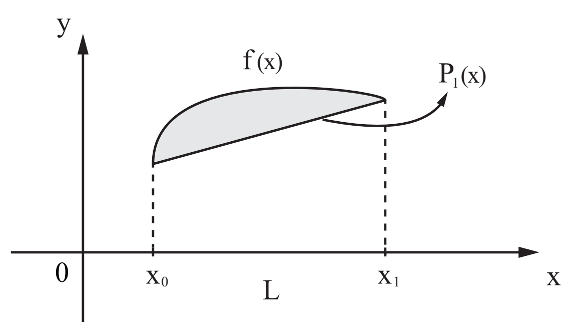
$$\boxed{H_i = \frac{1}{n}\frac{(-1)^{n-i}}{i!(n-i)!}\int_0^n \frac{q^{[n+1]}}{q-i}dq, \quad i = 0, 1, \ldots, n.}$$

Za Cotesove koeficijente vrijedi

$$\sum_{i=0}^{n} H_i = 1, \quad H_i = H_{n-i}.$$

Iz općenitog računa koji smo proveli sada možemo za različite n dobiti odgovarajuće kvadraturne formule jednostavnim uvrštavanjem n i računanjem Cotesovih koeficijenata. Najjednostavniji je slučaj $n = 1$.

4.3 Trapezna formula



Slika 24: Trapezna formula.

Uvrštavajući $n = 1$ u formulu za H_i, $i = 0, 1$, dobivamo

$$H_0 = -\int_0^1 \frac{q(q-1)}{q}dq = \frac{1}{2}$$
$$H_1 = \int_0^1 qdq = \frac{1}{2}.$$

Dakle je

$$\int_{x_0}^{x_1} f(x)dx = \frac{h}{2}(y_0 + y_1) \ - \ \text{trapezna formula.}$$

Što možemo reći o ostatku (pogrješci) ove formule?

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

Ako je $f(x)$ dovoljno glatka, možemo izvesti ocjenu za R. Promatrajmo R kao funkciju od h, $R(h)$.

$$R = \int_{x_0}^{x_1} f(x)dx - \frac{h}{2}\Big(f(x_0) + f(x_0 + h)\Big).$$

Deriviramo ovo po h, dvaput:

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

Znamo da je $R(0) = 0$, $R'(0) = 0$. Integrirajmo gornju relaciju

$$\begin{aligned}
R'(h) &= R'(0) + \int_0^h R''(t)dt = -\frac{1}{2}\int_0^h tf''(x_0 + t)dt = \text{(TSV)} \\
&= -\frac{1}{2}f''(\xi_1)\int_0^h tdt = -\frac{h^2}{4}f''(\xi_1), \quad \text{za neki } \xi_1 \in \langle x_0, x_0 + h\rangle \\
R(h) &= R(0) + \int_0^h R'(t)dt = -\frac{1}{4}\int_0^h t^2 f''(\xi)dt = -\frac{1}{4}f''(\xi)\frac{h^3}{3}, \ \xi \in \langle x_0, x_0 + h\rangle.
\end{aligned}$$

$$\boxed{R(h) = -\frac{h^3}{12}f''(\xi)}$$

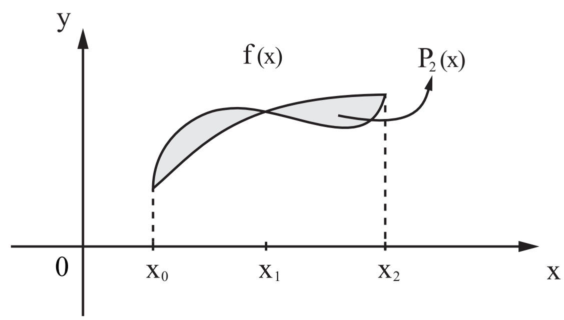
Ako je f konveksna na $[x_0, x_1]$ imamo preveliku, ako je konkavna, imamo premalu vrijednost integrala. Znamo li $M_2 = \max_{x\in[x_0,x_1]} |f''(x)|$, možemo ocijeniti R:

$$\boxed{|R| \leq \frac{M_2}{12}h^3}$$

4.4 Simpsonova formula

Simpsonovu kvadraturnu formulu dobivamo računajući Cotesove koeficijente za $n = 2$ Uvrštavajući $n = 2$ u formulu za H_i, $i = 0, 1, 2$, dobivamo

$$\begin{aligned}
H_0 &= \frac{1}{2}\frac{1}{2}\int_0^2 (q-1)(q-2)dq = \frac{1}{6} \\
H_1 &= -\frac{1}{2}\frac{1}{1}\int_0^2 q(q-2)dq = \frac{2}{3} \\
H_2 &= \frac{1}{2}\frac{1}{2}\int_0^2 q(q-1)dq = \frac{1}{6} \quad - \ \text{mogli smo i iz simetrije.}
\end{aligned}$$



Slika 25: Simpsonova formula.

Odatle, zbog $x_2 - x_0 = 2h$, imamo

$$\boxed{\int_{x_0}^{x_2} f(x)dx = \frac{h}{3}\Big(f(x_0) + 4f(x_1) + f(x_2)\Big)} \quad - \text{ Simpsonova formula}$$

Računom sličnim onom koji smo proveli kod trapezne formule može se pokazati da je

$$\boxed{R = -\frac{h^5}{90} f^{(4)}(\xi)} \quad \text{za neki } \xi \in \langle x_0, x_2 \rangle.$$

Vidimo da je Simpsonova formula točna i za polinome stupnja 3 - slično primjeru iz uvoda. Ponovo imamo bonus, što Simpsonovu formulu čini vrlo pogodnom za praktičnu primjenu - uz malo točaka daje (dosta) visoku točnost.

4.5 Newton-Cotesove formule viših redova

Za $n = 3$ računanjem Cotesovih koeficijenata dobivamo tzv. **Newtonovu kvadraturnu formulu** (poznatu još i kao formula $3/8$):

$$\boxed{\int_{x_0}^{x_3} f(x)dx = \frac{3h}{8}\Big(f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3)\Big)}$$

Za njen ostatak dobivamo $R = -\frac{3h^5}{80} f^{(4)}(\xi)$ za $\xi \in \langle x_0, x_3 \rangle$. Vidimo da je on istog reda kao i za Simpsonovu formulu.

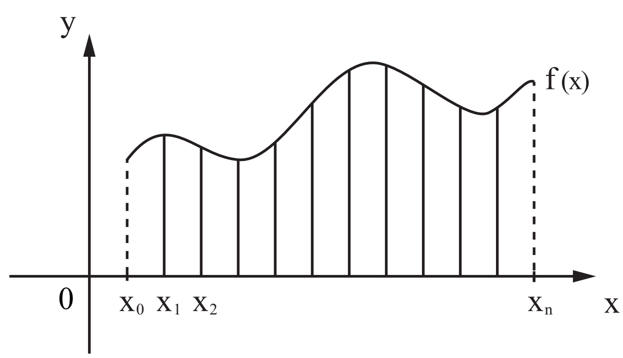
Newton-Cotesove formule postaju nepraktične za $n \sim 8$ i više. Ostatci su reda veličine $R = \mathcal{O}\left(h^{2\lfloor n/2 \rfloor + 3}\right)$, gdje je $\lfloor n/2 \rfloor$ najveće cijelo u $\frac{n}{2}$. Vidimo da kvaliteta raste skokovito za 2 pri prijelazu s parnog na neparni broj točaka i ostaje ista pri prijelazu s neparnog na sljedeći parni broj točaka.

n	$\hat{H}_0$	$\hat{H}_1$	$\hat{H}_2$	$\hat{H}_3$	$\hat{H}_4$	$\hat{H}_5$	$\hat{H}_6$	$\hat{H}_7$	D_n
1	1	1							2
2	1	4	1						6
3	1	3	3	1					8
4	7	32	12	32	7				90
5	19	75	50	50	75	19			288
6	41	216	27	272	27	216	41		840
7	751	3577	1323	2989	2989	1323	3577	751	17280

Newton-Cotesove koeficijente H_i za zadani n dobivamo iz gornje tablice kao kvocijente $H_i = \frac{\hat{H}_i}{D_n}$ za $i = 0, 1, \ldots, n$.

Problemi s Newton-Cotesovim formulama visokog reda rješavaju se tako da se područje integracije podijeli na podintervale na kojima se onda primjenjuju N-C formule nižeg reda. To vodi do tzv. poopćenih kvadraturnih formula.

4.6 Poopćena trapezna formula



Slika 26: Poopćena trapezna formula.

Podijelimo područje integracije $[a, b]$ na n podintervala točkama $a = x_0, x_1, \ldots, x_n = b$, gdje je $x_i = x_0 + ih$, $h = \frac{b-a}{n}$ i na svakom intervalu $[x_{i-1}, x_i]$ primijenimo trapeznu formulu.

$$\int_a^b f(x)dx = \frac{h}{2}(y_0 + y_1) + \frac{h}{2}(y_1 + y_2) + \frac{h}{2}(y_2 + y_3) + \ldots + \frac{h}{2}(y_{n-1} + y_n)$$

$$\boxed{\int_a^b f(x)dx = \frac{h}{2}\Big[f(x_0) + 2f(x_1) + \ldots + 2f(x_{n-1}) + f(x_n)\Big]}$$

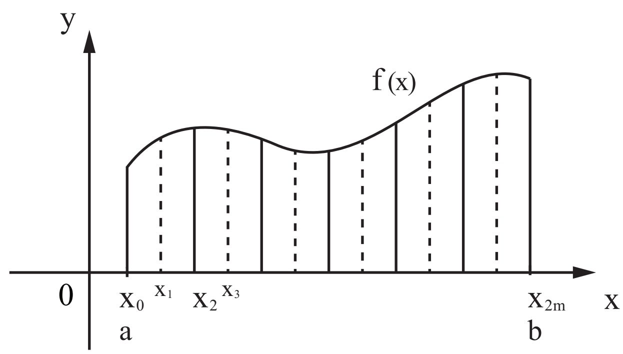
Za ostatak se može izvesti formula

$$R = -\frac{nh^3}{12}f''(\xi) = -\frac{b-a}{12}h^2 f''(\xi).$$

4.7 Poopćena Simpsonova formula

Ako je broj podintervala paran, možemo na dva po dva zamijeniti graf funkcije interpolacijskom parabolom, tj. integrirati pomoću Simpsonove formule

$$h = \frac{b-a}{n} = \frac{b-a}{2m}$$



Slika 27: Poopćena Simpsonova formula.

$$\int_a^b f(x)dx = \frac{h}{3}(y_0 + 4y_1 + y_2) + \frac{h}{3}(y_2 + 4y_3 + y_4) + \ldots + \frac{h}{3}(y_{2m-2} + y_{2m-1} + y_{2m}) =$$

$$= \frac{h}{3}\Big[y_0 + y_{2m} + 4\underbrace{(y_1 + y_3 + \ldots + y_{2m-1})}_{\sigma_1} + 2\underbrace{(y_2 + y_4 + \ldots + y_{2m-2})}_{\sigma_2}\Big]$$

$$\boxed{\int_a^b f(x)dx = \frac{h}{3}\Big(y_0 + y_{2m} + 4\sigma_1 + 2\sigma_2\Big)}$$

Za ostatak se dobiva

$$R = -\frac{mh^5}{90} f^{(4)}(\xi) = -\frac{b-a}{180} h^4 f^{(4)}(\xi) \quad \text{za } \xi \in \langle a, b \rangle.$$

4.8 Gaussove kvadraturne formule

Promatramo funkciju $f : [-1, 1] \to \mathbb{R}$. Želimo odabrati točke $t_1, t_2, \ldots, t_n$ i koeficijente $A_1, A_2, \ldots, A_n$ tako da formula

$$\int_{-1}^{1} f(t)dt = \sum_{i=1}^{n} A_i f(t_i)$$

bude točna za polinome što je moguće višeg stupnja N. Koji je najviši mogući stupanj N za koji se možemo nadati točnosti? Imamo $2n$ slobodnih parametara, u najboljem slučaju možemo dobiti formulu točnu za polinome do stupnja $2n-1$. Uvrštavajući funkcije $f(t) = 1, t, t^2, \ldots, t^{2n-1}$ u gornju formulu i računajući egzaktno integrale $\int_{-1}^{1} t^k dt$ dobivamo sustav od $2n$ jednadžbi s

$2n$ nepoznanica

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

Ovaj sustav je **nelinearan**. Može se pokazati da se uzimanjem t_i kao nul-točaka Legendreovih polinoma taj sustav može svesti na linearan sustav za A_i, čija je matrica regularna (ima Vandermondeovu determinantu).

Legendreovi polinomi su ortogonalni na $[-1, 1]$ i zadovoljavaju rekurzivne relacije

$$P_0(x) = 1,\ P_1(x) = x, \quad nP_n(x) = (2n-1)xP_{n-1}(x) - (n-1)P_{n-2}(x).$$

Ortogonalni su, ali nisu ortonormirani:

$$\int_{-1}^{1} P_m(x)P_n(x)dx = \frac{2}{2n+1}\delta_{mn}.$$

Prvih nekoliko je

$$\begin{aligned}
P_0(x) &= 1 \\
P_1(x) &= x \\
P_2(x) &= \frac{1}{2}(3x^2 - 1) \\
P_3(x) &= \frac{1}{2}(5x^3 - 3x) \\
P_4(x) &= \frac{1}{8}(35x^4 - 30x^2 + 3).
\end{aligned}$$

Osim toga, $P_n(1) = 1$, $P_n(-1) = (-1)^n$ i $\int_{-1}^{1} P_n(x)Q_k(x) = 0$, za sve polinome $Q_k(x)$ stupnja manjeg od n.

Primjer 4.2:

Gaussova kvadraturna formula za $n = 3$. Znamo da su t_i nultočke Legendreovog polinoma $P_3(t) = \frac{1}{2}t(5t^2 - 3)$. Dakle je $t_1 = -\sqrt{\frac{3}{5}},\ t_2 = 0,\ t_3 = \sqrt{\frac{3}{5}}$. Koeficijenti A_i se određuju iz (linearnog) sustava

$$\left.\begin{aligned}
A_1 + A_2 + A_3 &= 2 \\
-\sqrt{\frac{3}{5}}A_1 + \sqrt{\frac{3}{5}}A_2 &= 0 \\
\frac{3}{5}A_1 + \frac{3}{5}A_2 &= \frac{2}{3}
\end{aligned}\right\} \qquad A_1 = A_3 = \frac{5}{9},\ A_2 = \frac{8}{9}$$

Dakle imamo

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

Može se pokazati da je ostatak dan formulom

$$R_3 = \frac{1}{15+50} f^{(6)}(\xi), \text{ za neki } \xi \in [-1,1].$$

□

Vidimo da je točnost formule bolja od točnosti Simpsonove koja koristi isti broj točaka.

Općenito, za $\int_a^b f(x)dx$ prvo transformiramo interval integracije na $[-1,1]$ supstitucijom $x = \frac{b+a}{2} + \frac{b-a}{2}t$ i onda primjenjujemo Gaussove formule

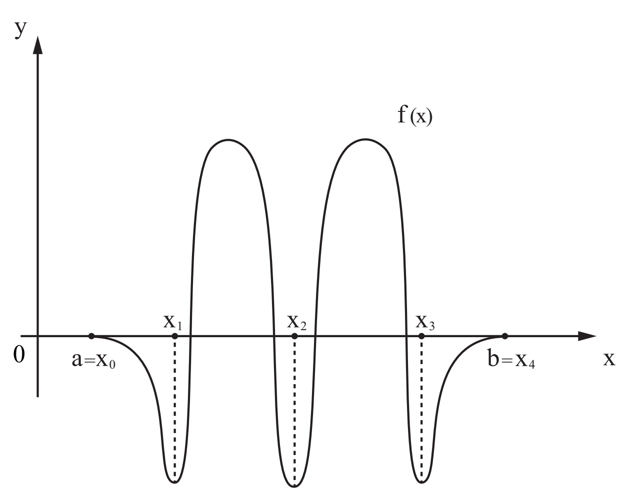
$$\int_a^b f(x)dx = \frac{b-a}{2}\sum_{i=1}^{n} A_i f(x_i),$$

gdje je $x_i = \frac{b+a}{2} + \frac{b-a}{2}t_i$, a t_i su nul-točke Legendreovog polinoma $P_n(t)$. Ostatak je oblika

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

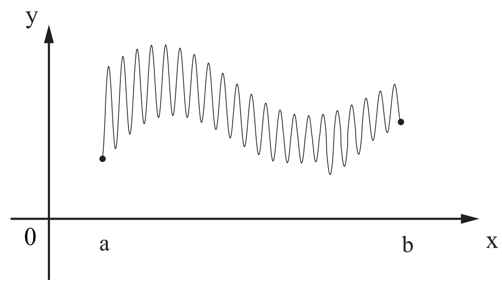
Postoje i drugi tipovi Gaussovih kvadraturnih formula. U svima se pojavljuju nul-točke ortogonalnih polinoma (Čebiševljevih, Laguerreovih, Hermiteovih). Obično su tabelirane, zajedno s pripadajućim koeficijentima A_i.

4.9 Mogući problemi



Slika 28: Patološki primjer za ekvidistantne čvorove.

Prije računanja integrala treba nastojati upoznati ponašanje funkcije na području integracije. U primjeru na slici bi kvadraturna formula s ekvidistantnim čvorovima $x_0, \ldots, x_n$ dala $\int_a^b f(x)dx < 0$, dok je, očito, prava vrijednost pozitivna. Notorno teške za integrirati su brzo-oscilirajuće funkcije, čak i kad ne mijenjaju predznak, jer se gubi sva informacija o ponašanju



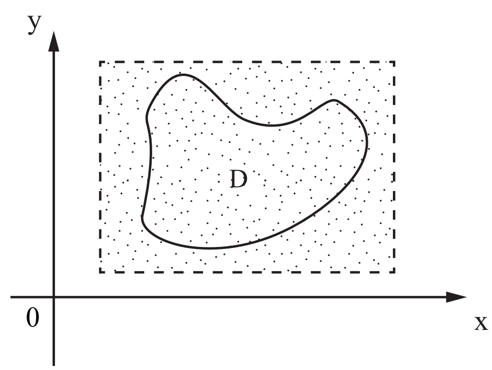
Slika 29: Brzooscilirajuća funkcija.

između čvorova. U takvim je situacijama potrebno integrirati posebno pozitivni i posebno negativni dio (ako je moguće). Ako je moguće treba uzeti više točaka. Najbolje je koristiti posebne metode prilagođene za brzo varirajuće funkcije.

Dakle: Saznati što je više moguće o ponašanju funkcije.
Nacrtati graf (ako je moguće).
Odabrati čvorove primjereno ponašanju funkcije.
Usporediti rezultate različitih metoda - nominalno točnija ne mora nužno biti bolja.
Koristiti specijalno prilagođene metode.

4.10 Kubaturne formule i višestruki integrali

Kod računanja višedimenzionalnih integrala broj točaka (pa onda i broj računskih operacija) vrlo brzo rastu - tzv. kombinatorna eksplozija. Za niske dimenzije (2,3) kubaturne formule Newton-Cotesovog i Gaussovog tipa još daju prihvatljive rezultate. Za više dimenzije je uglavnom zgodnije koristiti tzv. Monte Carlo metode. Ideja Monte Carlo metode je ilustrirana na slici 30. Že-



Slika 30: Metoda Monte Carlo za dvostruke integrale.

limo li izračunati površinu lika D, uzmemo neki pravokutnik koji sadrži D i generiramo slučajne točke koje su uniformno raspodijeljene u tom pravokutniku. Imamo li dovoljno mnogo točaka, razumno je očekivati da će omjer broja onih koje su u D i njihovog ukupnog broja biti dobra aproksimacija omjera površine lika D i površine pravokutnika.

Kubaturne formule se javljaju, npr. kod računanja elemenata matrice krutosti za 2-D i 3-D konačne elemente.

5 Obične diferencijalne jednadžbe

Promatramo Cauchyjev problem za obične diferencijalne jednadžbe (ODJ) 1. reda

$$\left.\begin{aligned} y' &= f(x,y) \\ y(x_0) &= y_0 \end{aligned}\right\}$$

Većina metoda koje rade za ovaj problem izravno se poopćava i na sustave ODJ.

Primjer 5.1:

$$\begin{aligned} y_1' &= y_1 + y_2 \quad , \quad y_1(0) = 0 \\ y_2' &= -y_1 + y_2 \quad , \quad y_2(0) = 1 \end{aligned}$$

□

ODJ višeg reda mogu se svesti na sustav ODJ 1.reda.

Primjer 5.2:

Promatramo ODJ 4. reda

$$y^{(4)} + (x+1)y' + y + x + 1 = 0$$

Uvodimo nove varijable $z,\ u$ i v kao $y' = z$, $z' = u$, $u' = v$.

$$\begin{aligned} y' &= z \\ z''' &+ (x+1)z + y + x + 1 = 0 \end{aligned} \qquad \begin{aligned} y' &= z \\ z' &= u \\ u'' &+ (x+1)z + y + x + 1 = 0 \end{aligned} \qquad \begin{aligned} y' &= z \\ z' &= u \\ u' &= v \\ v' &+ (x+1)z + y + x + 1 = 0 \end{aligned}$$

To je ekvivalentno sustavu

$$\underbrace{\begin{bmatrix} y \\ z \\ u \\ v \end{bmatrix}'}_{\vec{\omega}' \,=\, \underbrace{A(x)\vec{\omega} + \vec{f}_0(x)}_{\vec{f}(x,\vec{\omega})}} = \begin{bmatrix} z \\ u \\ v \\ -(x+1)z - y - x - 1 \end{bmatrix}$$

U našem slučaju,

$$A(x) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & -(x+1) & 0 & 0 \end{bmatrix}, \quad \vec{f}_0(x) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -(x+1) \end{bmatrix}$$

□

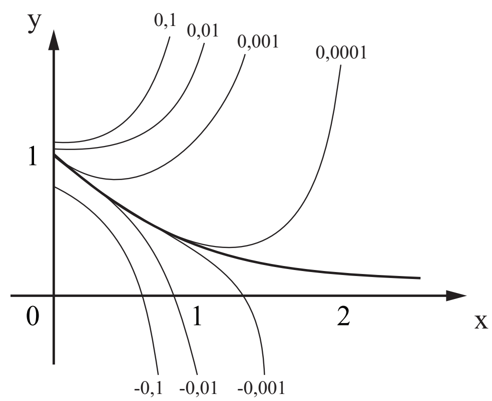
Ako je polazna jednadžba n-tog reda, dobije se sustav n-tog reda. Ako je jednažba linearna, sustav je linearan.

Pri numeričkom rješavanju ODJ javljaju se problemi stabilnosti i krutosti.

Primjer 5.3:

$$y' = 4y - 5e^{-x}, \quad y(0) = 1$$

Rješenje je $y = Ce^{4x} + e^{-x}$, početni uvjet daje $C = 0$.



Slika 31: Nestabilnost rješenja u ovisnosti o koraku h.

Rješavamo li našu jednadžbu numerički, pogrješke u ulaznim podatcima uzrokuju odstupanje od točnog rješenja i komad e^{4x} postaje dominantan. Rješenje je **nestabilno**, mala perturbacija na ulazu dovodi do velikih odstupanja u rezultatu. Nestabilnost je uzrokovana velikim (> 1) korijenom karakteristične jednadžbe ($\lambda - 4 = 0$). Čak i za jednadžbe koje imaju stabilna rješenja može doći do nestabilnosti zbog svojstava numeričke metode. Drugi problem je **krutost**. Javlja se pogotovo kod sustava ODJ kad se razne funkcije (nepoznanice) različito ponašaju - jedna se brzo mijenja, druga sporo. Korak diskretizacije koji je dobar za onu koja se sporo mijenja je besmislen za onu drugu, korak dobar za onu koja se brzo mijenja vodi do ogromnog utroška računalnih resursa.

5.1 Eulerova metoda

Promatramo Cauchyjev problem

$$\left.\begin{aligned} y' &= f(x,y) \\ y(x_0) &= y_0 \end{aligned}\right\}$$

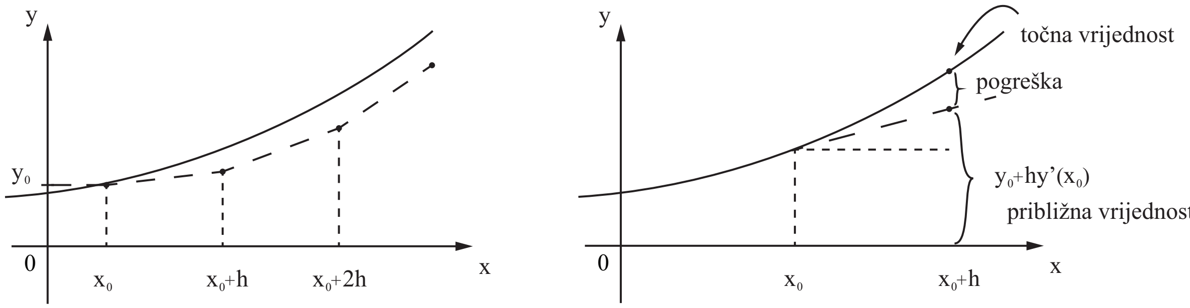
Najjednostavnija ideja je iz poznate vrijednosti y_0 za $x = x_0$ dobiti vrijednost u $x_0 + h$ kao

$$y(x_0 + h) \approx y_0 + h\,\underbrace{f(x_0, y_0)}_{y'(x_0)}$$

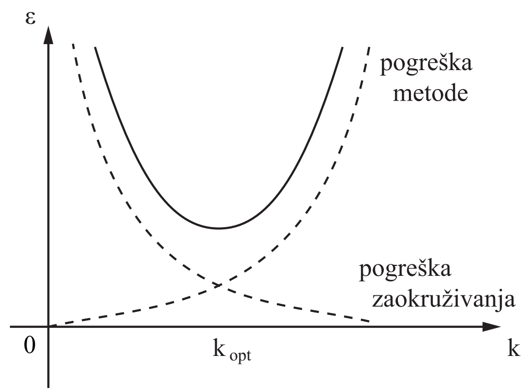
Općenito, za $y(x_i) = y_i$ vrijednosti u x_{i+1} dobivamo kao

$$y_{i+1} \approx y_i + hf(x_i, y_i) \; - \text{ Eulerova metoda}$$

Eulerova metoda je vrlo jednostavna no nije jako dobra u praksi jer brzo akumulira pogrješke. Pogrješka se može smanjiti smanjenjem koraka h, no to vodi do rasta pogrješke zaokruživanja. Eulerova metoda odgovara integraciji metodom pravokutnika.



Slika 32: Eulerova metoda.



Slika 33: Odnos pogrješke metode i pogrješke zaokruživanja.

5.2 Poboljšana Eulerova metoda

Modifikacija Eulerove metode u kojoj se nagib pod kojim se kreće iz x_i korigira koristeći informaciju o (procijenjenom) nagibu u x_{i+1}, zove se poboljšana Eulerova metoda.

$$
\begin{aligned}
y_{i+1}^* &= y_i + hf(x_i, y_i) \\
y_{i+1} &= y_i + \frac{1}{2}h\Big[f(x_i, y_i) + f(x_{i+1}, y_{i+1}^*)\Big]
\end{aligned}
$$

Korigirani nagib je aritmetička sredina nagiba u (x_i, y_i) (tj. vrijednosti $y'(x_i) = f(x_i, y_i)$) i procijenjenog nagiba u (x_{i+1}, y_{i+1}). Zašto moramo procjenjivati nagib u (x_{i+1}, y_{i+1})? Zašto ne uzeti točno $f(x_{i+1}, y_{i+1})$? Zato što ne znamo y_{i+1}. Procijenimo ga linearnom aproksimacijom iz (x_i, y_i).

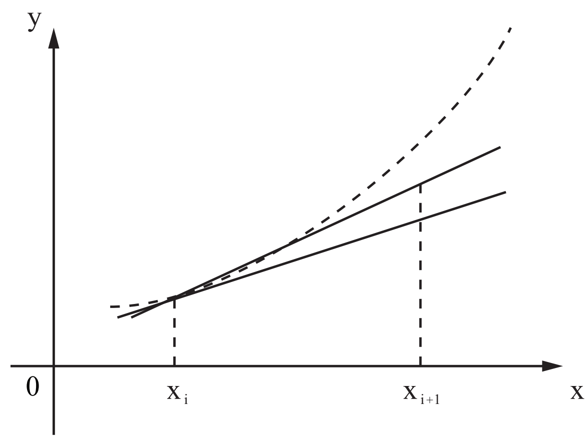
Što ako zanemarimo da ne znamo y_{i+1}? Dobivamo formulu

$$
y_{i+1} = y_i + \frac{1}{2}h\Big[f(x_i, y_i) + f(x_{i+1}, y_{i+1})\Big].
$$

To je **implicitna Adamsova formula 2. reda** Ponekad se Adamsova formula može eksplicitno riješiti po y_{i+1}, recimo kad je $f(x, y)$ linearno u y. Može se i iterativno rješavati.

5.3 Metode Runge-Kutta

Eulerova metoda temelji se na linearnoj ekstrapolaciji funkcije $y(x)$. Pokušajmo uvidjeti što možemo dobiti kvadratnom ekstrapolacijom. Uzmimo $x_0 = 0$, pretpostavimo da je $y(x) =$



Slika 34: Poboljšana Eulerova metoda.

$\alpha + \beta x + \gamma x^2$ u okolini 0 i odredimo nepoznate koeficijente $\alpha,\ \beta$ i γ tako da jednadžba bude zadovoljena u okolini 0.

$$y' = \beta + 2\gamma x = f(x, \alpha + \beta x + \gamma x^2)$$

Razvijemo desnu stranu u Taylorov red (ili ju aproksimiramo Taylorovim polinomom)

$$\beta + 2\gamma x = f(0,\alpha) + x\frac{\partial f}{\partial x}(0,\alpha) + (\beta x + \gamma x^2)\frac{\partial f}{\partial y}(0,\alpha) + \dots$$

Zbog $y(0) = y_0$ mora biti $\alpha = y_0$. Izjednačavajući koeficijente uz potencije od x koliko se god može (ovdje do x^1) dobivamo

$$\beta = f(0,\alpha), \quad 2\gamma = \frac{\partial f}{\partial x}(0,\alpha) + \beta\frac{\partial f}{\partial y}(0,\alpha).$$

Stavimo li $\gamma = 0$ dobijemo Eulerovu metodu, linearnu ekstrapolaciju

$$y(h) = y(0) + hf(0,y_0) + \frac{h^2}{2}\left[\frac{\partial f}{\partial x}(0,y_0) + f(0,y_0)\frac{\partial f}{\partial y}(0,y_0)\right].$$

Sada treba aproksimirati $\frac{\partial f}{\partial x}(0,y_0)$ i $\frac{\partial f}{\partial y}(0,y_0)$. Dovoljno je uzeti najjednostavniju aproksimaciju konačnim razlikama jer su već množene sa h^2, što je malo.

$$\begin{aligned}\frac{\partial f}{\partial x}(0,y_0) &= \frac{1}{h}\left[f(h,y_0) - f(0,y_0)\right]\\ \frac{\partial f}{\partial x}(0,y_0) &= \frac{1}{h}\left[f(j,y_0+k) - f(j,y_0)\right].\end{aligned}$$

Ovdje su j i k veličine sličnog reda kao i h. Možemo ih odabrati po volji. Neki od mogućih izbora su:

$$\begin{array}{llllll}(1) & j=0, & k = hf(0,y_0) & (3) & j=h, & k=hf(0,y_0)\\ (2) & j=0, & k=hf(h,y_0) & (4) & j=h, & k=hf(h,y_0)\end{array}$$

Na primjer, treći izbor daje

$$y = y_0 + hf(0, y_0) + \frac{h^2}{2}\left[\frac{1}{h}\left(f(h, y_0) - f(0, y_0)\right) + f(0, y_0)\frac{f(h, y_0 + h) - f(h, y_0)}{hf(0, y_0)}\right]$$

$$y = y_0 + \frac{h}{2}\underbrace{f(0, y_0)}_{k_1} + \frac{h}{2}\underbrace{f(h, y_0 + hf(0, y_0))}_{k_2}$$

Uz gornje oznake imamo

$$\begin{aligned} k_1 &= hf(0, y_0) \\ k_2 &= hf\left(h, y_0 + \frac{k_1}{2}\right) \\ y(h) &= y_0 + \frac{1}{2}k_1 + \frac{1}{2}k_2 \end{aligned}$$

Općenito iz (x_i, y_i) dobivamo (x_{i+1}, y_{i+1}) formulama

$$\left.\begin{aligned} k_1 &= hf(x_i, y_i) \\ k_2 &= hf\left(x_{i+1}, y_i + \frac{1}{2}k_1\right) \\ y_{i+1} &= y_i + \frac{1}{2}k_1 + \frac{1}{2}k_2 \end{aligned}\right\}$$

Runge-Kutta metoda drugog reda (= poboljšana Eulerova metoda). Odgovara integraciji trapeznom formulom u kojoj je $y(h)$ ocijenjen jednostavnom Eulerovom metodom.

$$y(h) \approx y_0 + \frac{h}{2}\left[f(0, h) + f(h, \underbrace{y(h)}_{y_0 + hf(0, y_0)})\right].$$

Korištenjem Taylorovih polinoma višeg stupnja dobiju se formule za Runge-Kutta metode viših redova. Najčešće se koriste Runge-Kutta metode četvrtog reda:

$$\left.\begin{aligned} k_1 &= hf(x_i, y_i) \\ k_2 &= hf\left(x_i + \frac{h}{2}, y_i + \frac{k_1}{2}\right) \\ k_3 &= hf\left(x_i + \frac{h}{2}, y_i + \frac{k_2}{2}\right) \\ k_4 &= hf(x_{i+1}, y_i + k_3) \\ y_{i+1} &= y_i + \frac{1}{6}k_1 + \frac{1}{3}k_2 + \frac{1}{3}k_3 + \frac{1}{6}k_4 \end{aligned}\right\}$$

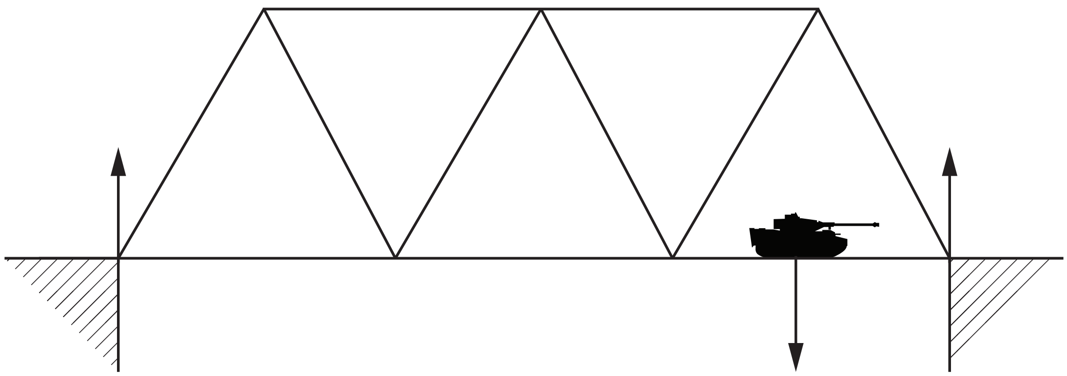
Runge-Kutta metoda četvrtog reda. Koeficijenti se dobivaju kao (nejednoznačna) rješenja nelinearnih sustava.

Sve opisane metode su **jednokoračne**. Postoje i **višekoračne** metode, prediktor-korektor formule itd.

6 Matrice i linearni sustavi

6.1 Izvori problema

Brojni fizikalni (posebno građevinski) problemi vode na sustave linearnih jednadžbi. Npr., sile koje djeluju na most na slici (težina mosta i vozila) uravnotežene su silama na krajevima mosta. Te se sile propagiraju duž nosača i u svakom čvoru njihova rezultanta mora biti jednaka nuli (inače bi se most počeo gibati). Razložimo li ih u horizontalne i vertikalne komponente, zbroj



Slika 35: Ovo bi moglo voditi na linearni sustav ...

komponenata svakog tipa mora biti 0 u svakom čvoru; dakle u svakom čvoru imamo dvije jednadžbe. Za m čvorova dobijemo $\sim 2m$ jednadžbi. Poznate veličine (težine mosta i vozila) idu na desnu stranu, ostale se slažu u matricu sustava koja obično ima specijalnu strukturu.

Drugi tip izvora problema s linearnim sustavima su metode diskretizacije i aproksimacije. Tako se numeričko rješavanje običnih i parcijalnih diferencijalnih jednadžbi svodi na rješavanja linearnih sustava, u kojima su nepoznanice vrijednosti rješenja u čvorovima (metoda konačnih razlika) ili koeficijenti u prikazu rješenja pomoću baznih funkcija (metode konačnih elemenata). I matrice takvih sustava obično imaju specijalnu strukturu.

6.2 Tipovi matrica

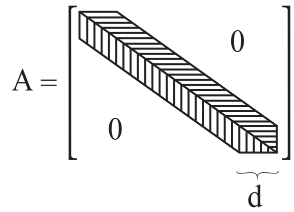
Najčešće ćemo promatrati **kvadratne** matrice $A \in M_n$. U praksi se često javljaju matrice visokog reda često i do $n \sim 100000$.

Kod takvih matrica bi i sam smještaj na računalu bio velik problem. Sretna okolnost je da je većina matrica koje se javljaju u praksi **rijetka**. Matrica je **rijetka** (engl. *sparse*) ako joj je najveći dio elemenata jednak nuli. Obično se smatra da je matrica rijetka ako je broj ne-nul elemenata reda veličine $\mathbb{O}(n)$, ili barem raste sporije od $\mathbb{O}(n^2)$.

Rijetkost matrice je najčešće posljedica **lokalnosti** modela, tj. činjenice da udaljeni dijelovi modela ne interagiraju ili dovoljno slabo utječu jedan na drugoga da to možemo zanemariti. Npr., izvori na modelu mosta interagiraju samo ako su povezani nosačem. Konačni elementi daju ne-nul element u matrici krutosti samo ako su blizu.

Osim rijetkosti, lokalnost najčešće diktira i **vrpčastu** ili **trakastu** strukturu matrice (engl. *band matrix*). Matrica A je vrpčasta ako postoji neki $d > 0$ takav da je $a_{ij} = 0$ za $|i - j| > d$.

Ne moraju svi elementi u vrpci širine d biti različiti od nule.



Slika 36: Vrpčasta matrica.

Primjere vrpčastih matrica smo vidjeli kod metode konačnih razlika i metoda konačnih elemenata. Najjednostavnije (netrivijalne) vrpčaste matrice su **trodijagonalne** matrice.

Matrica A je **simetrična** ako je $A = A^T$, tj. $a_{ij} = a_{ji}, \forall i, j$.

Mnogi fizikalni problemi, posebno ravnotežni, vode na simetrične matrice. Takve matrice trebaju manje memorije i (obično) manje računanja.

Ako za svaki $\vec{x} \neq 0$ vektor vrijedi $\vec{x}^T A\vec{x} > 0$ (≥ 0), matrica A je **pozitivno definitna** (**pozitivno semidefinitna**). Ovdje oznaka $\vec{x}^T A\vec{x}$ znači $(\vec{x}^T A)\vec{x}$. Simetrične matrice su često i pozitivno (semi)definitne. Pozitivno definitne matrice su regularne. Dobra vijest je da mnoge metode diskretizacije vode na pozitivno definitne matrice, što znači da su sustavi jednoznačno rješivi. Za simetrične i pozitivno definitne matrice postoje metode koje su efikasnije od metoda za općenite matrice.

Matrica A je **donja** (**gornja**) **trokutasta** ako je $a_{ij} = 0$ za $i < j$ $(i > j)$. Sustavi s takvim matricama se vrlo lako rješavaju pa se one često javljaju kao matrice na koje se svode općenite matrice.

Matrica A je **ortogonalna** ako je $A^{-1} = A^T$, odnosno $AA^T = I$. Sustav $A\vec{x} = \vec{b}$ se trivijalno rješava kao $\vec{x} = A^T\vec{b}$. Ortogonalne matrice ne mijenjaju duljinu vektora koje množe.

Matrica A je **dijagonalna** ako je $a_{ij} = 0$ za $i \neq j$. Računanje s dijagonalnim matricama je trivijalno. Nalaženje baze u kojoj će prikaz matrice A biti dijagonalan vodi na problem svojstvenih vrijednosti.

Matrica A je **dijagonalno dominantna** ako vrijedi

$$|(a_{ii})| \geq \sum_{i=1,\ j\neq i}^{n} |a_{ij}|\,,\ \forall i = 1, \dots, n,$$

i za bar jedan i vrijedi stroga nejednakost. A je **strogo dijagonalno dominantna** ako stroga nejednakost vrijedi za sve $i = 1, \dots, n$.

6.3 Tipovi problema

U praksi se najčešće javljaju **linearni sustavi**, tj. **sustav linearnih jednadžbi**:

$$A\vec{x} = \vec{b}.$$

Matrična jednadžba $AX = B$ u kojoj treba naći nepoznatu matricu X tipa (u, p) nije ništa drugo nego skup od p linearnih sustava, $A\vec{x}_1 = \vec{b}_1,\ A\vec{x}_2 = \vec{b}_2, \dots,\ A\vec{x}_p = \vec{b}_p$. Znamo li riješiti linearni sustav, znamo i matričnu jednadžbu.

U teoriji, rješenje sustava $A\vec{x} = \vec{b}$ (ako postoji i jedinstveno je) dano je formulom $\vec{x} = A^{-1}\vec{b}$. U praksi, gotovo se nikad ne radi tako, jer je računanje inverzne matrice neefikasno.

Također, u praksi malu važnost ima i determinanta. **Cramerovo pravilo** je iznimno neefikasan način rješavanja sustava. Ni kriterij $\det A \neq 0$ nije ključan, jer se u praksi uvijek javljaju greške zaokruživanja.

Glavna ideja kod rješavanja sustava je manipulacijom svesti njegovu matricu na oblik iz kojeg se rješenja lako dobivaju. Dakle sustav $A\vec{x} = \vec{b}$ svodi se na $T\vec{x} = \vec{c}$, gdje je T trokutasta matrica ili na neki drugi pogodan oblik.

Pokušamo li matricu sustava svesti na najjednostavniji oblik, tj. na dijagonalan, imamo problem svojstvenih vrijednosti. Ako za matricu A postoji matrica E takva da je $E^{-1}AE = D$, gdje je D dijagonalna matrica, onda stupce od E zovemo **svojstvenim vektorima**, a (dijagonalne) elemente od D **svojstvenim vrijednostima**. Iz $AE = ED$ vidimo da to znači $A\vec{e}_i = d_i\vec{e}_i$, što je upravo definicija svojstvene vrijednosti d_i i pripadnog svojstvenog vektora $\vec{e}_i$. Ako je A simetrična, onda su sve njene svojstvene vrijednosti realne.

6.4 Tipovi metoda

Metode za rješavanje linearnih sustava dijele se na **izravne** (direktne) i **iteracijske**.

Izravnim metodama se poslije konačnog broja računskih operacija, ako nema pogrješaka zaokruživanja, dolazi do točnog rješenja sustava. Kod iteracijskih metoda rješenje sustava dobivamo kao limes beskonačnog niza aproksimacija.

Zašto bi itko htio koristiti iteracijske metode kad daju samo približno rješenje? Dva razloga: U praksi se i izravnim metodama (u pravilu) dobivaju približna rješenja, zbog pogrješaka zaokruživanja. Osim toga, za mnoge matrice sa specijalnom strukturom iteracijske metode konvergiraju brzo i ukupan broj računskih operacija može biti osjetno manji nego za izravnu metodu, a točnost je još uvijek prihvatljiva. U pravilu je izravnim metodama potrebno $\sim n^3$ operacija. Taj se broj smanjuje za rijetke matrice, no red veličine ostaje. Iteracijske metode za rijetke matrice obično trebaju puno manje od n^2 operacija za jedan korak iterativnog postupka, što znači da i uz puno koraka još mogu biti brže od izravnih. Manje su osjetljive i na grešku zaokruživanja. Činjenica da je konvergencija zajamčena samo za dijagonalno dominantne simetrične matrice nije veliki problem jer su mnoge matrice koje se javljaju u nama zanimljivim primjenama upravo toga tipa. Iteracijski postupci obično zahtijevaju i manje memorije.

Najpoznatije izravne metode su Gaussove eliminacije (s modifikacijama kap što su Gauss-Jordanove eliminacije i uz djelomično ili potpuno pivotiranje) te rastav Choleskog. Najčešće korištene iteracijske metode su Jacobijeva, Gauss-Seidelova i SOR metoda.

6.5 Gausove eliminacije

Gaussove eliminacije koriste se za rješavanje linearnih sustava još od antike. Glavna ideja je svesti sustav $A\vec{x} = \vec{b}$ na ekvivalentan sustav s jednostavnijom matricom (dva sustava su ekvivalentna ako je svako rješenje jednog sustava ujedno i rješenje drugog sustava i obratno). To se postiže konačnim brojem elementarnih transformacija:

(i) zamjena dviju jednadžbi;

(ii) množenje jedne jednadžbe brojem različitim od nule;

(iii) dodavanje jedne jednadžbe pomnožene nekim brojem nekim drugim brojem nekoj drugoj jednadžbi

Za sustav $A\vec{x} = \vec{b}$ s kvadratnom matricom koja je regularna, elementarne transformacije se rade tako da se matrica A svede na gornju trokutastu. Ako matrica A nije regularna, elementarnim transformacijama se dobiva ekvivalentan sustav koji nema rješenja. Radi se s proširenom matricom sustava

$$\left[\begin{array}{cccc:c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{array}\right] \sim \ldots \sim \left[\begin{array}{cccccc:c} u_{11} & \cdots & u_{1p} & u_{1,p+1} & \cdots & u_{1n} & c_1 \\ \vdots & \ddots & \vdots & & & & \vdots \\ 0 & & u_{pp} & u_{p,p+1} & \cdots & u_{pn} & c_p \\ 0 & \cdots & 0 & 0 & \cdots & 0 & c_{p+1} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & c_n \end{array}\right],$$

gdje su $u_{ii} \neq 0$ za $i = 1, \ldots, p$.

Ako je $p < n$ i barem jedan od $c_{p+1}, \ldots, c_n \neq 0$, sustav nema rješenja. Ako je $p = n$, dobili smo ekvivalentan sustav $U\vec{y} = \vec{c}$, u kojem je matrica sustava gornja trokutasta, a vektor nepoznanica $\vec{y}$ je neka permutacija vektora nepoznanica $\vec{x}$ (ako se rade transformacije samo nad retcima, $\vec{y} = \vec{x}$).

Elementarne transformacije nad retcima mogu se zapisati pomoću matričnog množenja. Primjena transformacija zamjene i-tog i j-tog retka može se zapisati kao množenje matrice A s lijeva matricom P_{ij} koja se dobiva od jedinične matrice I zamjenom i-tog i j-tog retka (matrica P_{ij} je **permutacijska** matrica). Matrice kojima opisujemo elementarne transformacije zovemo **elementarnim matricama**.

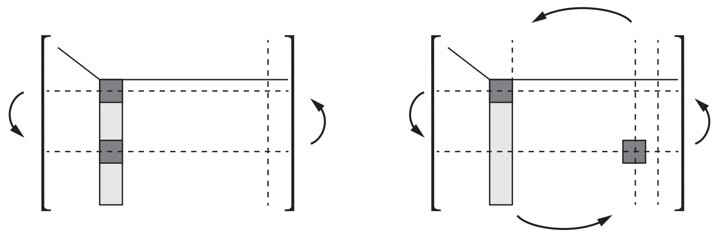
Može se pokazati da je produkt svih elementarnih matrica kojima smo množili A, kako bismo dobili U, regularna donja trokutasta matrica. Označimo ju s E. Imamo, dakle, $EA = U$, odnosno $A = E^{-1}U$, pri čemu je E^{-1} i sama donja trokutasta. Označimo ju s L. Imamo $A = LU$, tj. matrica sustava A je prikazana kao produkt donje (L) i gornje (U) trokutaste matrice. Kažemo da Gaussove eliminacije daju **LU - faktorizaciju** matrice A.

Broj operacija potrebnih za rješavanje sustava $A\vec{x} = \vec{b}$ je reda veličine $\mathbb{O}(n^3)$. Točnije:

Teorem 7. *Neka je A regularna matrica reda n. Tada se sustav linearnih jednadžbi $A\vec{x} = \vec{b}$ može riješiti Gaussovim postupkom eliminacije s $\frac{n}{3}(n^2 + 3n - 1)$ množenja i $\frac{n}{6}(2n^2 + 3n - 5)$ zbrajanja.* □

6.5.1 Modifikacije Gausovih eliminacija - pivotiranje

Zbog numeričke stabilnosti algoritma dobro je birati **ključni element** (onaj kojim radimo eliminacije) tako da bude što je moguće veći (po apsolutnoj vrijednosti). To možemo postići tako da na glavnu dijagonalu u k-tom koraku dovedemo najveći (po modulu) element k-tog stupca (zamjenom redaka) ili najveći (po modulu) element dolje desno od k (zamjenom odgovarajućih redaka i stupaca). Prvi postupak se zove **parcijalno** (djelomično), a drugi **potpuno pivotiranje**. Teorijski se realiziraju množenjem permutacijskim matricama. To poskupljuje račun, posebno potpuno pivotiranje kod kojeg je potrebno i prenumerirati nepoznanice. U praksi se pokazuje da je djelomično pivotiranje dobar kompromis, tj. da uz razumno produljenje računa dobivamo zadovoljavajuću numeričku stabilnost.



Slika 37: Djelomično (lijevo) i potpuno (desno) pivotiranje.

6.5.2 Gauss-Jordanove eliminacije

Ako Gaussovim eliminacijama matricu sustava ne svodimo prvo na gornju trokutastu već odmah na dijagonalnu, dobivamo Gauss-Jordanove eliminacije.

$$\left[\begin{array}{ccc:c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} & b_n \end{array}\right] \sim \left[\begin{array}{ccc|ccc:c} a'_{11} & & 0 & & & & b'_1 \\ & \ddots & & & \mathbf{0} & & \vdots \\ 0 & & a'_{kk} & & & & \vdots \\ \hline & & & a'_{kn} & & & \vdots \\ & \mathbf{0} & & & \ddots & & \vdots \\ & & & & & & b'_k \end{array}\right] \sim \left[\begin{array}{ccc:c} a'_{11} & & 0 & c_1 \\ & \ddots & & \vdots \\ 0 & & a'_{nn} & c_n \end{array}\right]$$

Gauss-Jordanov postupak treba oko 50% više operacija za rješavanje linearnog sustava ($\sim \frac{n^3}{2}$), no ravnopravan je Gaussovim eliminacijama za problem nalaženja inverzne matrice.

Teorem 8. *Inverzna matrica A^{-1} regularne matrice A reda n, može se odrediti Gaussovim (ili Gauss-Jordanovim) eliminacijama uz n^3 množenja i $n(n-1)^2$ zbrajanja.* □

6.6 Simetrične matrice

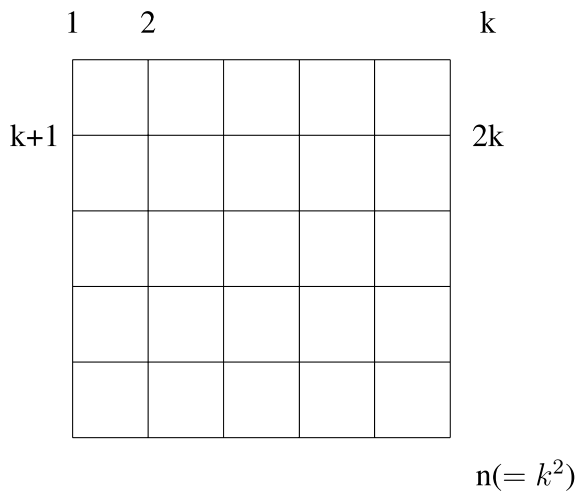
Za simetrične matrice je moguće sačuvati simetriju i u njihovom LU-rastavu tako da se on dobije u obliku LL^T.

Teorem 9. *Simetrična realna matrica$A$ reda n može se faktorizirati kao $A = LL^T$, gdje je L regularna realna donja trokutasta matrica, ako i samo ako je matrica A pozitivno definitna.* □

Ako matrica A nije pozitivno definitna u jednom se trenutku u računu pojavljuju kompleksni brojevi. Gornji rezultat daje i najjeftiniji kriterij provjere pozitivne definitnosti.

Metoda rješavanja linearnog sustava temeljem LL^T rastava zove se još i **metoda drugog korijena** ili **metoda Choleskog**. Broj operacija se reducira (otprilike) na pola u odnosu na Gaussove eliminacije.

Daljnje smanjenje broja potrebnih operacija moguće je za rijetke i/ili vrpčaste matrice. Primjerice, za matrice s vrpcom širine d, broj operacija je $\sim d^2n$. Mnoge matrice koje se javljaju u diskretizacijama PDJ imaju $d \sim \sqrt{n}$ pa za njih Gaussove eliminacije trebaju $\sim n^2$ operacija.



Zašto je $d \sim \sqrt{n}$?

6.7 Analiza pogrješke izravnih metoda

Glavni izvor pogrješaka kod izravnih metoda su početne greške u koeficijentima i greške zaokruživanja tijekom postupka.

Sustav $A\vec{x} = \vec{b}$ je loše uvjetovan ako male relativne promjene u elementima od A i $\vec{b}$ mogu dovesti do velikih relativnih promjena u rješenju $\vec{x}$. U suprotnom je sustav **dobro uvjetovan**

Primjer 6.1:
Promatramo sustave

$$\begin{aligned} 2x_1 - x_2 &= 3 \\ x_1 + x_2 &= 3 \end{aligned} \qquad\qquad \begin{aligned} 2x_1 - x_2 &= 3 \\ 2x_1 - 1.0001x_2 &= 2.9999 \end{aligned}$$

Rješenje oba sustava je $x_1 = 2$, $x_1 = 1$. Promijenimo sada slobodni član druge jednadžbe u oba sustava za $\Delta = 0.0003$:

$$\begin{aligned} 2x_1 - x_2 &= 3 \\ x_1 + x_2 &= 3.0003 \end{aligned} \qquad\qquad \begin{aligned} 2x_1 - x_2 &= 3 \\ 2x_1 - 1.0001x_2 &= 3.0002 \end{aligned}$$

Rješenje prvog sustava je $x_1 = 2.0001$, $x_2 = 1.0002$ dok je rješenje drugog sustava $x_1 = 0.5$, $x_2 = -2$! Drugi sustav je primjer loše uvjetovanog sustava. □

Treba nam nešto za matrice što bi odgovaralo pojmu duljine (norme) vektora. Time ćemo moći ocijeniti učinjenu pogrješku iz "ostatka", tj. iz toga koliko dobro približno rješenje zadovoljava jednadžbu.

Primjer 6.2:
Neka je x' približno rješenje linearne jednadžbe $ax = b$. Neka je $r = b - ax'$. Tada je

$$|x - x'| \le |a^{-1}||r|.$$

Tvrdnja slijedi iz $r = b - ax' = ax - ax' = a(x - x')$, tj.

$$|r| = |a||x - x'|.$$

Vidimo da "kvaliteta" rješenja ovisi o "kvaliteti" koeficijenta a. Nešto slično će vrijediti i za linearne sustave.

Teorem 10. *Neka je $\vec{x}'$ približno rješenje sustava $A\vec{x} = \vec{b}$ i neka je $\vec{r} = \vec{b} - A\vec{x}'$. Tada je*

$$||\vec{x} - \vec{x}'|| \leq ||A^{-1}|| \cdot ||\vec{r}||.$$

□

Ovdje je s $||\vec{r}||$ označena neka norma vektora $\vec{r}$, npr. $||\vec{r}|| = \sqrt{\vec{r} \cdot \vec{r}}$. Odgovarajuća norma $||A||$ matrice A definira se kao

$$||A|| = \sup_{\vec{x} \neq \vec{0}} \frac{||A\vec{x}||}{||\vec{x}||}.$$

Zove se **prirodna matrična norma** inducirana vektorskom normom $||\ ||$. Ima puno vektorskih (pa onda i matričnih) normi, no može se pokazati da su one sve ekvivalentne.

Uvjetovanost $k(A)$ kvadratne regularne matrice A za danu prirodnu normu $||\ ||$ je broj $k(A) = ||A|| \cdot ||A^{-1}||$. Uvjetovanost je uvijek veća ili jednaka od 1, jer je za svaku prirodnu normu $||I|| = 1$, a za sve norme vrijedi $||AB|| = ||A|| \cdot ||B||$. Poželjna uvjetovanost je mala, iako velika uvjetovanost ne znači nužno da je i sustav loše uvjetovan.

6.8 Jacobijeva metoda

Jacobijeva metoda spada u iteracijske metode. Kod iteracijskih metoda se rješenje dobiva kao limes niza približnih rješenja $\vec{x}^0, \vec{x}^1, \ldots, \vec{x}^k, \ldots$. Postavlja se pitanje konvergencije.

Iteracijski postupci imaju zajamčenu konvergenciju za dosta usku klasu matrica (može se dogoditi da postupak konvergira i za matricu izvan te klase, no ne možemo se osloniti na to). Srećom, matrice koje se javljaju u primjenama kao što su diskretizacija rubnih problema, često spadaju baš u tu klasu.

Niz vektora $\vec{x}^0, \vec{x}^1, \ldots, \vec{x}^k, \ldots$ konvergira prema vektoru $\vec{x}$ ako i samo ako je $\lim\limits_{k\to\infty} x_i^k = x_i$, $\forall i = 1, \ldots, n$. Dakle konvergenciju niza vektora definiramo po komponentama. Indeks niza pišemo kao $\vec{x}^k$, $k = 1, 2, \ldots$. Oznaku n čuvamo za red sustava, tj. duljinu (dimenziju) vektora. Ne postoji opasnost od konfuzije s potencijom.

$$\vec{x} = \lim_{k\to\infty} \vec{x}^k \Leftrightarrow x_i = \lim_{k\to\infty} x_i^k.$$

Slično se definira i konvergencija niza matrica.

Promatrajmo sustav linearnih jednadžbi $A\vec{x} = \vec{b}$, gdje je $A \in M_n$ regularna matrica i $a_{ii} \neq 0$, $\forall i = 1, \ldots, n$. Rastavimo matricu A kao $D - B$, gdje je D dijagonalna matrica s elementima a_{ii}:

$$A\vec{x} = \vec{b} \Leftrightarrow (D - B)\vec{x} = \vec{b} \Rightarrow D\vec{x} = B\vec{x} + \vec{b}.$$

Uzmimo neki $\vec{x}^0 \in \mathbb{R}^n$ i definirajmo niz $\vec{x}^k \in \mathbb{R}^n$.

$$D\vec{x}^{k+1} = B\vec{x}^k + \vec{b}, \quad \text{za } k = 0, 1, 2, \ldots.$$

Zbog $a_{ii} \neq 0$ možemo lako naći $D^{-1} = \operatorname{diag}\left[\frac{1}{a_{ii}}\right]$ pa je $\vec{x}^{k+1} = D^{-1}B\vec{x}^k + \vec{d}$, odnosno

$$\boxed{\vec{x}^{k+1} = B_J\vec{x}^k + \vec{d}}$$

gdje je $B_J = D^{-1}B$, $\vec{d} = D^{-1}\vec{b}$. Matricu B_J možemo lako eksplicitno izračunati:

$$B_J = \begin{bmatrix} 0 & -\frac{a_{12}}{a_{11}} & -\frac{a_{13}}{a_{11}} & \cdots & -\frac{a_{1n}}{a_{11}} \\ -\frac{a_{21}}{a_{22}} & 0 & -\frac{a_{23}}{a_{22}} & \cdots & -\frac{a_{2n}}{a_{22}} \\ \vdots & \vdots & \vdots & & \vdots \\ -\frac{a_{n1}}{a_{nn}} & -\frac{a_{n2}}{a_{nn}} & -\frac{a_{n3}}{a_{nn}} & \cdots & 0 \end{bmatrix}$$

Raspisano po komponentama:

$$x_i^{k+1} = -\frac{1}{a_{ii}} \sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ij} x_j^k + \frac{b_i}{a_{ii}}, \quad i = 1, 2, \ldots, n.$$

Uz koje uvjete na A (dakle na B_J) niz $\vec{x}^k$ konvergira? Može se pokazati da je dovoljno da neka matrična norma matrice B_J bude manja od 1, $||B_J|| < 1$. To je uvijek zadovoljeno za dvije klase dijagonalno dominantnih matrica.

Matrica A je **razloživa**, ako se može presložiti tako da je

$$PAP^T = \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix},$$

gdje je P permutacijska matrica reda n, a B_{11} i B_{22} su kvadratne matrice reda manjeg od n.

Matrica A je **nerazloživa** ako nije razloživa. Svaka matrica čiji su svi elementi različiti od nule je nerazloživa.

Teorem 11. *neka je A strogo dijagonalno dominantna matrica ili nerazloživa dijagonalno dominantna matrica. Tada Jacobijev iteracijski postupak za rješavanje linearnog sustava $A\vec{x} = \vec{b}$ konvergira za svaki početni vektor $\vec{x}^0$.* □

6.9 Gauss-Seidelova metoda

Promatramo sustav $A\vec{x} = \vec{b}$ s regularnom matricom $A \in M_n$ za koju je $a_{ii} \neq 0$ za sve $1 \leq i \leq n$. Prikažimo matricu A u obliku $A = N - P$, gdje su

$$N = \begin{bmatrix} a_{11} & & & \\ a_{21} & a_{22} & & \mathbf{0} \\ \vdots & \vdots & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \quad P = \begin{bmatrix} 0 & -a_{12} & -a_{13} & \cdots & -a_{1n} \\ & 0 & -a_{23} & \cdots & -a_{2n} \\ & & 0 & \ddots & \vdots \\ & \mathbf{0} & & & -a_{n-1,n} \\ & & & & 0 \end{bmatrix}$$

Sada iz $A\vec{x} = \vec{b}$ slijedi $N\vec{x} = P\vec{x} + \vec{b}$. Za neki $\vec{x}^0 \in \mathbb{R}^n$ definiramo niz $\vec{x}^k$ iterativno formulom.
$N\vec{x}^{k+1} = P\vec{x}^k + \vec{b}$, $k = 0, 1, \ldots$.

Zbog regularnosti matrice N ($\det N = \prod_{i=1}^{n} a_{ii} \neq 0$) možemo množiti gornju relaciju s lijeva s N^{-1} pa dobivamo $\vec{x}^{k+1} = N^{-1}P\vec{x}^k + N^{-1}\vec{b}$, odnosno

$$\boxed{\vec{x}^{k+1} = B_G\vec{x}^k + \vec{d},}$$

gdje je $B_G = N^{-1}P$, $\vec{d} = N^{-1}\vec{b}$.

Formule za raspis po komponentama su složenije nego za Jacobijevu metodu:

$$x_1^{k+1} = -\frac{1}{a_{11}}\sum_{j=2}^{n} a_{ij}x_j^k + \frac{b_1}{a_{11}}$$

$$x_i^{k+1} = -\frac{1}{a_{ii}}\left[\sum_{j=1}^{i-1} a_{ij}x_j^{k+1} + \sum_{j=i+1}^{n} a_{ij}x_j^k\right] + \frac{b_i}{a_{ii}}, \quad 2 \le i \le n-1$$

$$x_n^{k+1} = -\frac{1}{a_{nn}}\sum_{j=1}^{n-1} a_{ij}x_j^{k+1} + \frac{b_n}{a_{nn}}.$$

Vidimo da se pri računanju i-te komponente vektora $\vec{x}^{k+1}$ koriste već izračunate komponente (od 1. do $(i-1)$-ve) tog vektora.

Uvjeti konvergencije su slični kao za Jacobijevu metodu

Teorem 12. *Neka je A strogo dijagonalno dominantna ili nerazloživa dijagonalno dominanta matrica. Tada Gasuss-Seidelov iteracijski postupak za rješavanje sustava $A\vec{x} = \vec{b}$ konvergira za svaki izbor početne aproksimacije $\vec{x}^0$.* □

Teorem 13. *Ako je matrica $A \in M_n$ simetrična i pozitivno definitna, onda Gauss-Seidelov iteracijski postupak za rješavanje sustava $A\vec{x} = \vec{b}$ konvergira za sve $\vec{x}^0 \in \mathbb{R}^n$.* □

Kako možemo prepoznati pozitivno definitnu matricu?

Teorem 14. *Neka je A simetrična strogo dijagonalno dominantna ili simetrična nerazloživa dijagonalana dominantna matrica s pozitivnim dijagonalnim elementima. Tada je matrica A pozitivno definitna.* □

U većini slučajeva Gauss-Seidelov postupak konvergira brže od Jacobijevog, no ne nužno. Postoje sustavi za koje jedan postupak konvergira, a drugi ne i obratno. Takvi primjeri (?) u pravilu narušavaju neki od dozvoljenih uvjeta iz gornjih teorema. Za trodijagonalnu matricu A ili obje metode konvergiraju ili obje divergiraju. Ako obje konvergiraju, Gauss-Seidelova je brža.

6.10 Ocjena pogrješke iteracijskih metoda

Promatrajmo iteracijsku shemu $\vec{x}^{k+1} = M\vec{x}^k + \vec{d}$, $k = 0, 1, \ldots$ za koju znamo da konvergira za proizvoljan $\vec{x}^0$. Dakle vrijedi $\vec{x} = \lim\limits_{k\to\infty} \vec{x}^k$. Koliko je ta konvergencija brza? Kad smo dovoljno blizu točnom rješenju $\vec{x}$? Kad možemo stati s iteracijskim postupkom?

Teorem 15. *Neka za matricu M vrijedi*

$$||M\vec{y} - M\vec{z}|| \le L||\vec{y} - \vec{z}||,$$

za sve $\vec{y}, \vec{z} \in \mathbb{R}^n$ i za neki $L \in [0, 1\rangle$. Ako je $\vec{x}^{k+1} = M\vec{x}^k + \vec{d}$, $k = 0, 1, \ldots$ uz proizvoljni $\vec{x}^0 \in \mathbb{R}^n$, onda je

$$||\vec{x} - \vec{x}^k|| \le \frac{L}{1-L}||\vec{x}^k - \vec{x}^{k-1}||, \; k = 1, 2, \ldots,$$

gdje je $\vec{x}$ traženo rješenje sustava $\vec{x} = M\vec{x} + \vec{d}$. □

Ovo nam daje kriterij zaustavljanja na temelju razlike dviju uzastopnih aproksimacija.

Ocjena greške iz Teorema 15 je **aposteriorna**, jer je dobivena poslije računanja $\vec{x}^k$. Može se dokazati i **apriorna** ocjena

$$||\vec{x} - \vec{x}^k|| \leq \frac{L^k}{1-L}||\vec{x}^1 - \vec{x}^0||, \; k = 1, 2, \ldots$$

Apriorna je jer se njome može ocjenjivati $||\vec{x} - \vec{x}^k||$ prije računanja samog $\vec{x}^k$. Na temelju apriorne ocjene može se odrediti potreban broj iteracija za postizanje zadane točnosti ε.

Gornja analiza ne uzima u obzir greške zaokruživanja. Iteracijskim postupcima mogu se korigirati približna rješenja linearnih sustava i poboljšavati približno izračunate inverzne matrice.

6.11 OR metode

OR je zajednički naziv za klasu iteracijskih metoda koje poopćuju Jacobijevu i Gauss-Seidelovu metodu uvođenjem parametra čijim se variranjem i optimalnim izborom ubrzava konvergencija. OR dolazi od **over-relaxation**. Postoje razne verijante - mi ćemo ovdje izložiti poopćenje Gauss-Seidelove metode poznate kao SOR (**successive over-relaxation**).

Promatramo sustav $A\vec{x} = \vec{b}$ za čiju matricu A vrijedi: $a_{ii} \neq 0, \; 1 \leq i \leq n$. Rastavimo matricu A kao $A = N - P$, pri čemu je

$$N = \frac{1}{\omega}D - T, \quad P = \frac{1-\omega}{\omega}D + S,$$

za neki $\omega \in \mathbb{R}$, $\omega \neq 0$. Ovdje je $D = \mathrm{diag}[a_{11}, \ldots, a_{nn}]$, T je (strogo) donji trokut od $-A$, a S je (strogo) gornji trokut od $-A$ (dakle vrijedi $A = D - T - S$). Realni broj ω zove se **relaksacijski parametar**. Uz ovu supstituciju imamo

$$\begin{aligned} A\vec{x} = \vec{b} \;\to N\vec{x} &= P\vec{x} + \vec{b} \\ \vec{x} &= M\vec{x} + \vec{d}, \end{aligned}$$

gdje je $M = N^{-1}P = (D - \omega T)^{-1}\left[(1-\omega)D + \omega S\right]$, $\vec{d} = N^{-1}\vec{b} = \omega(D - \omega T)^{-1}\vec{b}$. Stavimo $L = D^{-1}T$, $U = D^{-1}S$ pa imamo $M = (I - \omega L)^{-1}\left[(1-\omega)I + \omega U\right]$. Za $\omega = 1$ dobijemo $B_G = (I - L)^{-1}U$, matricu Gauss-Seidelevog postupka.

Teorem 16. *Neka je A simetrična pozitivno definitna matrica. Tada SOR postupak konvergira za sve $\omega \in \langle 0, 2\rangle$* □

6.12 Problem svojstvenih vrijednosti

Realni broj λ je **svojstvena vrijednost** (**vlastita vrijednost**) kvadratne matrice A, ako postoji ne-nul vektor $\vec{x} \in \mathbb{R}^n$ takav da je $A\vec{x} = \lambda\vec{x}$. Vektor $\vec{x}$ je **svojstveni** (**vlastiti**) **vektor** koji odgovara svojstvenoj vrijednosti λ.

Iz definicije slijedi da vektor $\vec{x}$ mora biti netrivijalno rješenje sustava $(A - \lambda I)\vec{x} = \vec{0}$. Znamo da homogeni sustavi ima netrivijalno rješenje samo ako mu je matrica singularna. Dakle takva rješenja postoje za one vrijednosti λ za koje je

$$\det(A - \lambda I) = 0$$

Ovo je svojstvena ili **karakteristična jednadžba** matrice A. Funkcija $\det(A - \lambda I)$ je polinom u varijabli λ čiji je stupanj jednak redu matrice A. To je **karakteristični polinom** matrice (determinantu $\det(A - \lambda I)$ ponekad još zovu i **sekularnom determinantom**).

Polinom n-tog stupnja ima točno n korijena u $\mathbb{C}$, pri čemu se svaki korijen broji sa svojom kratnošću. Dakle matrica $A \in M_n$ ima općenito, n kompleksnih svojstvenih vrijednosti.

Primjer 6.3:

$$A = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}, \quad |A - \lambda I| = \begin{vmatrix} \cos\varphi - \lambda & -\sin\varphi \\ \sin\varphi & \cos\varphi - \lambda \end{vmatrix}$$

$$|A - \lambda I| = \cos^2\varphi - 2\lambda\cos\varphi + \lambda^2 + \sin^2\varphi = \lambda^2 - 2\lambda\cos\varphi + 1$$

$$\begin{aligned} |A - \lambda I| = 0 &\Leftrightarrow \lambda^2 - 2\lambda\cos\varphi + 1 = 0 \\ \lambda_{1,2} &= \cos\varphi \pm \sqrt{\cos^2\varphi - 1} \\ &= \cos\varphi \pm i\sin\varphi \end{aligned}$$

Vidimo da A ima realne svojstvene vrijednosti samo za $\varphi = k\pi$. Zašto? □

Skup svih svojstvenih vrijednosti matrice A naziva se **spektar** od A i označava se sa $\sigma(A)$. Najveća po modulu svojstvena vrijednost od A zove se **spektralni radijus** od A i označava se s $\rho(A)$

$$\rho(A) = \max_{\lambda \in \sigma(A)} |\lambda|.$$

Vrijedi

$$\sigma(A) \subset [-\rho(A), \rho(A)].$$

Spektralni radijus nosi puno informacija o ponašanju velikih potencija matrice. Primjerice, nužan i dovoljan uvjet konvergencije Jacobijeve i Gauss-Seidelove metode je $\rho(M) < 1$.

Teorem 17. *(Perron) Ako su svi elementi realne kvadratne matrice pozitivni, tada je i njena najveća svojstvena vrijednost pozitivna i jednostruka. Svojstveni vektor koji pripada toj svojstvenoj vrijednosti ima sve komponente pozitivne.* □

Matrice A i B su **slične** ako postoji regularna matrica S takva da je $B = S^{-1}AS$. Slične matrice imaju iste karakteristične polinome (a onda i isti spektar).

Za razliku od rješavanja linearnih sustava, ne postoje izravne metode kojima se točno (egzaktno) dobivaju svojstvene vrijednosti općenite matrice. To slijedi iz činjenice da za polinome visokog stupnja ne postoje formule za njihove nul-točke (visokog stupnja znači stupnja većeg od četiri). Problemi nalaženja svojstvenih vrijednosti i vektora se u pravilu rješavaju numerički, programskim paketima tipa EISPACK i sličnima.

Dva su osnovna pristupa: U jednom se, egzaktno ili približno, prvo odrede koeficijenti svojstvenog polinoma pa se onda (približnim metodama) određuju njihove nul-točke. Druga klasa metoda, koja se najčešće koristi samo za dobivanje najveće (ili k najvećih) svojstvene vrijednosti ne računa eksplicitno karakteristični polinom, već se tražene svojstvene vrijednosti dobivaju iterativnim postupkom iz matrice i neke početne aproksimacije. Ovdje ćemo se ograničiti na metode prvog tipa.

6.12.1 Karakteristični polinom

Koeficijenti karakterističnog polinoma vezani su sa svojstvenim vrijednostima Vieteovim formulama:

$$\det(A - \lambda I) = (-1)^n[\lambda^n - \sigma_1\lambda^{n-1} + \sigma_2\lambda^{n-2} - \ldots + (-1)^n\sigma_n]$$

Vrijedi:

$$\lambda_1 + \lambda_2 + \ldots + \lambda_n = -\sigma_1$$
$$\lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_n = \sigma_n = \det A$$

Ponekad se u literaturi umjesto $\det(A - \lambda I) = 0$ uzima $\det(\lambda I - A) = 0$, što rezultira karakterističnim polinomom s jedinicom kao vodećim koeficijentom, $\lambda^n - \sigma_1\lambda^{n-1} + \ldots + (-1)^n\sigma_n$.

Teorem 18. *(Hamilton-Cayley) Svaka kvadratna matrica poništava svoj karakteristični polinom.* □

To znači da je

$$A^n - \sigma_1 A^{n-1} + \ldots + (-1)^n\sigma_n I = 0.$$

Gornji rezultat je polazište za Krylovljevu metodu računanja koeficijenata karakterističnog polinoma.

Teorem 19. *Slične matrice imaju isti karakteristični polinom.* □

Na teoremu 19 se temelji metoda Danilevskog, kod koje se zadana matrica A transformacijama sličnosti svodi na matricu oblika

$$B = \begin{bmatrix} b_{n-1} & b_{n-2} & \cdots & b_1 & b_0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & & & 1 & 0 \end{bmatrix},$$

gdje su $b_{n-1}, b_{n-2}, \ldots, b_1, b_0$ koeficijenti karakterističnog polinoma.

6.12.2 Krylovljeva metoda

Neka je $p(\lambda) = \lambda^n + b_{n-1}\lambda^{n-1} + \ldots + b_n\lambda + b_0$ karakteristični polinom matrice A. Tada je, po Hamilton-Cayleyjevom teoremu

$$A^n + b_{n-1}A^{n-1} + \ldots + b_0 I = 0.$$

Lijeva strana je kvadratna matrica. Djelujemo li njome na proizvoljan vektor $\vec{y}$ dobivamo

$$A^n\vec{y} + b_{n-1}A^{n-1}\vec{y} + \ldots + b_0 I\vec{y} = \vec{0}.$$

To je sustav od n linearnih jednadžbi s nepoznanicama $b_0, b_1, \ldots, b_{n-1}$ (vektor $A^n\vec{y}$ prebacimo na desnu stranu i imamo sustav $b_{n-1}A^{n-1}\vec{y} + \ldots + b_0 I\vec{y} = -A^n\vec{y}$). Ako taj sustav ima jedinstveno rješenje, onda su njegove komponente koeficijenti karakterističnog polinoma. Ako je matrica tog sustava singularna, pokušamo s drugim vektorom.

U ovom postupku nije važno eksplicitno računati potencije $A^2, A^3, \ldots, A^n$. Uz oznake $\vec{y}^0 = \vec{y}$, $\vec{y}^k = A\vec{y}^{k-1}$, vektore $\vec{y}^k$ dobivamo iterativno iz $\vec{y}$. Sustav možemo pisati kao

$$\begin{bmatrix} y_1^{k-1} & y_1^{k-2} & \cdots & y_1^0 \\ y_2^{k-1} & y_2^{k-2} & \cdots & y_2^0 \\ \vdots & \vdots & & \vdots \\ y_n^{k-1} & y_n^{k-2} & \cdots & y_n^0 \end{bmatrix} \begin{bmatrix} b_{n-1} \\ b_{n-2} \\ \vdots \\ b_0 \end{bmatrix} = \begin{bmatrix} -y_1^k \\ -y_2^k \\ \vdots \\ -y_n^k \end{bmatrix}$$

Uočimo da eksponenti gore ne označavaju potenciranje, već se odnose na djelovanje odgovarajuće potencije od A na početni vektor $\vec{y}$, odnosno na korak u iteracijskom procesu.

Krylovljeva metoda zahtijeva oko n^3 množenja i $n(n-1)^2$ zbrajanja za računanje svih vektora $\vec{y}^0, \ldots, \vec{y}^n$. Nakon toga nam treba još oko $\frac{n^3}{3}$ množenja i isto toliko zbrajanja za rješavanje sustava i onda tek imamo karakteristični polinom.

Metodom Krylova se mogu računati i svojstveni vektori, no to ne ćemo raditi.

6.12.3 Metoda neodređenih koeficijenata

Polinom n-tog stupnja je jednoznačno određen svojim vrijednostima za $n+1$ realnih argumenata. Znamo li da mu je vodeći koeficijent jednak 1, ostaju nam njegova vrijednosti za n različitih argumenata.

Uvrstimo $\lambda = 0, 1, 2, \ldots, n-1$ u $p(\lambda) = \lambda^n + b_{n-1}\lambda^{n-1} + \ldots + b_0$.

$$\left.\begin{aligned} b_0 &= p(0) \\ 1^n + b_{n-1}1^{n-1} + \ldots + b_1 1 + b_0 &= p(1) \\ 2^n + b_{n-1}2^{n-1} + \ldots + b_1 2 + b_0 &= p(2) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ (n-1)^n + b_{n-1}(n-1)^{n-1} + \ldots + b_1(n-1) + b_0 &= p(n-1) \end{aligned}\right\}\text{presložimo}$$

$$\left.\begin{aligned} b_{n-1} + b_{n-2} + \ldots + b_1 &= p(1) - p(0) - 1^n \\ 2^{n-1}b_{n-1} + 2^{n-1}b_{n-2} + \ldots + 2b_1 &= p(2) - p(1) - 2^n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ (n-1)^{n-1}b_{n-1} + (n-1)^{n-2}b_{n-2} + \ldots + (n-1)b_1 &= p(n-1) - p(0) - (n-1)^n \end{aligned}\right\}\begin{array}{l}\text{linearni sustav za} \\ b_{n-1}, \ldots, b_1\end{array}$$

Dobili smo linearni sustav

$$C\vec{b} = \vec{d},$$

gdje je

$$C = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 2^{n-1} & 2^{n-1} & \cdots & 2 \\ \vdots & \vdots & & \vdots \\ (n-1)^{n-1} & (n-1)^{n-2} & \cdots & (n-1) \end{bmatrix}, \quad \vec{d} = \begin{bmatrix} p(1) - p(0) - 1^n \\ p(2) - p(0) - 2^n \\ \vdots \\ p(n-1) - p(0) - (n-1)^n \end{bmatrix}.$$

Matrica C je uvijek regularna (zašto?).

6.12.4 Lokalizacija nul-točaka

Svojstvene vrijednosti realne matrice su, općenito, kompleksni brojevi. Njihov položaj u kompleksnoj ravnini moguće je procijeniti iz odnosa dijagonalnih i nedijagonalnih elemenata matrice.

Teorem 20. *(Geršgorin) Neka je A matrica reda n i neka je C_i krug u kompleksnoj ravnini sa središtem u a_{ii} i polumjerom $r_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|$, $i = 1, 2, \ldots, n$. Tada sve svojstvene vrijednosti matrice A leže u skupu*

$$S = \bigcup_{i=1}^{n} C_i.$$

$\square$

Kako matrice A i A^T imaju isti spektar, gornja tvrdnja mora vrijediti i ako polumjere krugova računamo zbrajajući apsolutnu vrijednost nedijagonalnih elemenata po stupcima matrice A, dakle s polumjerima $s_j = \sum_{\substack{i=1 \\ i \neq j}}^{n} |a_{ij}|$, $j = 1, 2, \ldots, n$.

Ako je $T = \cup D_j$, gdje je D_j krug u $\mathbb{C}$ sa središtem a_{ii} i polumjerom s_j, onda se sve svojstvene vrijednosti matrice A moraju nalaziti u $S \cap T$.

Ako je realna matrica A simetrična, onda su sve njene svojstvene vrijednosti realne. U tom nam slučaju Geršgorinov teorem daje lokaciju svojstvenih vrijednosti u određenim intervalima u $\mathbb{R}$.

Za simetrične matrice se svojstvene vrijednosti i vektori mogu odrediti specijalnim metodama (Jacobijeve i Givensove rotacije itd.).

Literatura

[1] B. P. Demidovič, I. A. Maron, *Computational Mathematics*, Mir, Moscow, 1987

[2] Z. Drmač, V. Hari, M. Marušić, M. Rogina, S. Singer, S. Singer, *Numerička analiza*, dostupno na *www.math.hr/znanost/iprojekti/numat*

[3] J. R. Rice, *Numerical Methods, Software, and Analysis*, McGraw-Hill, Tokyo, 1983

[4] G. U. Milovanović, *Numerička analiza II deo*, Naučna knjiga, Beograd, 1985

[5] Z. Stojaković, D. Herceg, *Numeričke metode linearne algebre*, Građevinska knjiga, Beograd, 1982

[6] W. H. Preuss, S. A. Teukolsky, *Numerical Recipes*, Cambridge University Press, Cambridge, 1992