SVEUČILIŠTE U ZAGREBU
PMF – MATEMATIČKI ODJEL

Mladen Rogina Sanja Singer Saša Singer

Numerička matematika

Predavanja i vježbe (2. dio)

Zagreb, 2008.

Sadržaj

1. Aproksimacija funkcija (nastavak)

1.1. Diskretna metoda najmanjih kvadrata

Ponovno, neka je funkcija f zadana na diskretnom skupu točaka $x_0, \ldots, x_n$. Također, pretpostavljamo da je tih točaka mnogo više nego nepoznatih parametara aproksimacione funkcije.

Aproksimaciona funkcija

$$\varphi(x, a_0, \ldots, a_m)$$

određuje se iz uvjeta da je 2-norma vektora pogrešaka u čvorovima aproksimacije najmanja moguća, tj. tako da minimiziramo

$$S = \sum_{k=0}^{n} (f(x_k) - \varphi(x_k))^2 \to \min .$$

Ovu funkciju S (kvadrat 2-norme vektora greške) interpretiramo kao funkciju nepoznatih parametara

$$S = S(a_0, \ldots, a_m).$$

Očito je uvijek $S \geq 0$, bez obzira kakvi su parametri. Dakle, zadatak je minimizirati funkciju S kao funkciju više varijabli $a_0, \ldots, a_m$. Ako je S dovoljno glatka funkcija, a ova je (jer je funkcija u parametrima a_k), nužni uvjet ekstrema je

$$\frac{\partial S}{\partial a_k} = 0, \quad k = 0, \ldots, m.$$

Takav pristup vodi na tzv. **sustav normalnih jednadžbi**.

1.1.1. Linearni problemi i linearizacija

Ilustrirajmo to na najjednostavnijem primjeru, kad je aproksimaciona funkcija pravac.

Primjer 1.1.1. *Zadane su točke* $(x_0, f_0), \ldots, (x_n, f_n)$*, koje po diskretnoj metodi najmanjih kvadrata aproksimiramo pravcem*

$$\varphi(x) = a_0 + a_1 x.$$

Greška aproksimacije u čvorovima koju minimiziramo je

$$S = S(a_0, a_1) = \sum_{k=0}^{n} (f_k - \varphi(x_k))^2 = \sum_{k=0}^{n} (f_k - a_0 - a_1 x_k)^2 \to \min .$$

Nađimo parcijalne derivacije po parametrima a_0 *i* a_1*:*

$$0 = \frac{\partial S}{\partial a_0} = -2 \sum_{k=0}^{n} (f_k - a_0 - a_1 x_k),$$
$$0 = \frac{\partial S}{\partial a_1} = -2 \sum_{k=0}^{n} (f_k - a_0 - a_1 x_k) x_k.$$

Dijeljenjem s -2 *i sređivanjem po nepoznanicama* a_0*,* a_1*, dobivamo linearni sustav*

$$a_0(n+1) + a_1 \sum_{k=0}^{n} x_k = \sum_{k=0}^{n} f_k$$
$$a_0 \sum_{k=0}^{n} x_k + a_1 \sum_{k=0}^{n} x_k^2 = \sum_{k=0}^{n} f_k x_k.$$

Uvedemo li standardne skraćene oznake

$$s_\ell = \sum_{k=0}^{n} x_k^\ell, \quad t_\ell = \sum_{k=0}^{n} f_k x_k^\ell, \quad \ell \geq 0,$$

onda linearni sustav možemo pisati kao

$$\begin{aligned} s_0 a_0 + s_1 a_1 &= t_0 \\ s_1 a_0 + s_2 a_1 &= t_1. \end{aligned} \tag{1.1.1}$$

Nije teško pokazati da je matrica sustava regularna, što slijedi iz linearne nezavisnosti vektora

$$(1, 1, \ldots, 1)^T \quad i \quad (x_0, x_1, \ldots, x_n)^T,$$

uz uvjet da imamo barem dvije različite točke x_k *(prirodan uvjet za pravac), pa postoji jedinstveno rješenje sistema. Samo rješenje dobiva se rješavanjem linearnog sustava (1.1.1).*

Ostaje još pitanje da li smo dobili minimum, ali i to nije teško pokazati, korištenjem drugih parcijalnih derivacija (dovoljan uvjet minimuma je pozitivna definitnost Hesseove matrice). Ipak, provjera da je to minimum, može i puno lakše. Budući da se radi o zbroju kvadrata, S *predstavlja paraboloid s otvorom prema gore u varijablama* a_0*,* a_1*, pa je jasno da takvi paraboloidi imaju minimum. Zbog toga se nikad ni ne provjerava da li je dobiveno rješenje minimum za* S*.*

Za funkciju φ mogli bismo uzeti i polinom višeg stupnja,

$$\varphi(x) = a_0 + a_1 x + \cdots + a_m x^m,$$

ali postoji opasnost da je za malo veće m ($m \approx 10$) dobiveni sustav vrlo loše uvjetovan (blizak singularnom), pa dobiveni rezultati mogu biti jako pogrešni. Zbog toga se to nikada, ovako direktno, ne radi. Ako se uopće koriste aproksimacije polinomima viših stupnjeva, onda se to radi korištenjem ortogonalnih polinoma.

Linearni model diskretnih najmanjih kvadrata je potpuno primjenjiv na opću linearnu funkciju

$$\varphi(x) = a_0\varphi_0(x) + \cdots a_m\varphi_m(x),$$

gdje su $\varphi_0, \ldots, \varphi_m$ poznate (zadane) funkcije. Ilustrirajmo to ponovno na općoj linearnoj funkciji s 2 parametra.

Primjer 1.1.2. *Zadane su točke $(x_0, f_0), \ldots, (x_n, f_n)$, koje po diskretnoj metodi najmanjih kvadrata aproksimiramo funkcijom oblika*

$$\varphi(x) = a_0\varphi_0(x) + a_1\varphi_1(x).$$

Postupak je potpuno isti kao u prošlom primjeru. Opet minimiziramo kvadrat 2-norme vektora pogrešaka aproksimacije u čvorovima

$$S = S(a_0, a_1) = \sum_{k=0}^{n} (f_k - \varphi(x_k))^2 = \sum_{k=0}^{n} (f_k - a_0\varphi_0(x_k) - a_1\varphi_0(x_k))^2 \to \min.$$

Sređivanjem parcijalnih derivacija

$$0 = \frac{\partial S}{\partial a_0} = -2\sum_{k=0}^{n} (f_k - a_0\varphi_0(x_k) - a_1\varphi_1(x_k))\,\varphi_0(x_k),$$
$$0 = \frac{\partial S}{\partial a_1} = -2\sum_{k=0}^{n} (f_k - a_0\varphi_0(x_k) - a_1\varphi_1(x_k))\,\varphi_1(x_k),$$

po varijablama a_0, a_1, uz dogovor da je

$$s_0 = \sum_{k=0}^{n} \varphi_0^2(x_k), \qquad s_1 = \sum_{k=0}^{n} \varphi_0(x_k)\varphi_1(x_k), \qquad s_2 = \sum_{k=0}^{n} \varphi_1^2(x_k),$$
$$t_0 = \sum_{k=0}^{n} f_k\varphi_0(x_k), \qquad t_1 = \sum_{k=0}^{n} f_k\varphi_1(x_k),$$

dobivamo potpuno isti oblik linearnog sustava

$$s_0 a_0 + s_1 a_1 = t_0$$
$$s_1 a_0 + s_2 a_1 = t_1.$$

Ovaj sustav ima ista svojstva kao i u prethodnom primjeru. Pokažite to!

Što ako φ nelinearno ovisi o parametrima? Dobili bismo nelinearni sustav jednadžbi, koji se relativno teško rješava. Uglavnom, problem postaje ozbiljan optimizacijski problem, koji se, recimo, može rješavati metodama pretraživanja ili nekim drugim optimizacijskim metodama, posebno prilagođenim upravo za rješavanje nelinearnog problema najmanjih kvadrata (na primjer, Levenberg–Marquardt metoda).

Postoji i drugi pristup. Katkad se jednostavnim transformacijama problem može transformirati u linearni problem najmanjih kvadrata.

Nažalost, rješenja lineariziranog problema najmanjih kvadrata i rješenja originalnog nelinearnog problema, u principu, **nisu** jednaka. Problem je u različitim mjerama za udaljenost (grešku).

Ilustrirajmo, ponovno, nelinearni problem najmanjih kvadrata na jednom jednostavnom primjeru.

Primjer 1.1.3. *Zadane su točke $(x_0, f_0), \ldots, (x_n, f_n)$, koje po diskretnoj metodi najmanjih kvadrata aproksimiramo funkcijom oblika*

$$\varphi(x) = a_0 e^{a_1 x}.$$

Greška aproksimacije u čvorovima (koju minimiziramo) je

$$S = S(a_0, a_1) = \sum_{k=0}^{n} (f_k - \varphi(x_k))^2 = \sum_{k=0}^{n} (f_k - a_0 e^{a_1 x_k})^2 \to \min .$$

Parcijalnim deriviranjem po varijablama a_0 i a_1 dobivamo

$$0 = \frac{\partial S}{\partial a_0} = -2 \sum_{k=0}^{n} (f_k - a_0 e^{a_1 x_k})\, e^{a_1 x_k},$$
$$0 = \frac{\partial S}{\partial a_1} = -2 \sum_{k=0}^{n} (f_k - a_0 e^{a_1 x_k}) a_0\, x_k e^{a_1 x_k},$$

što je nelinearan sustav jednadžbi.

S druge strane, ako logaritmiramo relaciju

$$\varphi(x) = a_0 e^{a_1 x},$$

dobivamo

$$\ln \varphi(x) = \ln(a_0) + a_1 x.$$

Moramo logaritmirati još i vrijednosti funkcije f u točkama x_k, pa uz supstitucije

$$h(x) = \ln f(x), \quad h_k = h(x_k) = \ln f_k, \quad k = 0, \ldots, n,$$

i

$$\psi(x) = \ln \varphi(x) = b_0 + b_1 x,$$

gdje je

$$b_0 = \ln a_0, \quad b_1 = a_1,$$

dobivamo linearni problem najmanjih kvadrata

$$\widetilde{S} = \widetilde{S}(b_0, b_1) = \sum_{k=0}^{n} (h_k - \psi(x_k))^2 = \sum_{k=0}^{n} (h_k - b_0 - b_1 x_k)^2 \to \min .$$

Na kraju, iz rješenja b_0 i b_1 ovog problema, lako očitamo a_0 i a_1

$$a_0 = e^{b_0}, \quad a_1 = b_1.$$

Uočite da ovako dobiveno rješenje uvijek daje pozitivan a_0, tj. linearizacijom dobivena funkcija $\varphi(x)$ će uvijek biti veća od 0. Nekako je odmah jasno da to nije "pravo" rješenje za sve početne podatke (x_k, f_k)! No, možemo li na ovako opisani način linearizirati sve početne podatke? Očito je **ne**, *jer mora biti $f_k > 0$ da bismo mogli logaritmirati.*

Ipak, i kad su neki $f_k \leq 0$, nije teško, korištenjem translacije svih podataka dobiti $f_k +$ translacija > 0, pa onda nastaviti postupak linearizacije. Pokušajte korektno formulirati linearizaciju!

Konačno, evo i popisa nekoliko funkcija koje su često u upotrebi i njihovih standardnih linearizacija u problemu najmanjih kvadrata.

(a) Funkcija

$$\varphi(x) = a_0 x^{a_1}$$

linearizira se logaritmiranjem

$$\psi(x) = \log \varphi(x) = \log(a_0) + a_1 \log x, \quad h_k = \log f_k, \quad k = 0, \ldots, n.$$

Drugim riječima, dobili smo linearni problem najmanjih kvadrata

$$\widetilde{S} = \widetilde{S}(b_0, b_1) = \sum_{k=0}^{n} (h_k - b_0 - b_1 \log(x_k))^2 \to \min,$$

gdje je

$$b_0 = \log(a_0), \quad b_1 = a_1.$$

U ovom slučaju, da bismo mogli provesti linearizaciju, moraju biti i $x_k > 0$ i $f_k > 0$.

(b) Funkcija

$$\varphi(x) = \frac{1}{a_0 + a_1 x}$$

linearizira se na sljedeći način

$$\psi(x) = \frac{1}{\varphi(x)} = a_0 + a_1 x, \quad h_k = \frac{1}{f_k}, \quad k = 0, \ldots, n.$$

Pripadni linearni problem najmanjih kvadrata je

$$\widetilde{S} = \widetilde{S}(a_0, a_1) = \sum_{k=0}^{n} (h_k - a_0 - a_1 x_k)^2 \to \min .$$

(c) Funkciju

$$\varphi(x) = \frac{x}{a_0 + a_1 x}$$

možemo linearizirati na više načina. Prvo, možemo staviti

$$\psi(x) = \frac{1}{\varphi(x)} = a_0 \frac{1}{x} + a_1, \quad h_k = \frac{1}{f_k}, \quad k = 0, \ldots, n.$$

Pripadni linearni problem najmanjih kvadrata je

$$\widetilde{S} = \widetilde{S}(a_0, a_1) = \sum_{k=0}^{n} (h_k - a_0 \frac{1}{x_k} - a_1)^2 \to \min .$$

Može i ovako

$$\psi(x) = \frac{x}{\varphi(x)} = a_0 + a_1 x, \quad h_k = \frac{x_k}{f_k}, \quad k = 0, \ldots, n.$$

Pripadni linearni problem najmanjih kvadrata je

$$\widetilde{S} = \widetilde{S}(a_0, a_1) = \sum_{k=0}^{n} (h_k - a_0 - a_1 x_k)^2 \to \min .$$

(d) Funkcija

$$\varphi(x) = \frac{1}{a_0 + a_1 e^{-x}}$$

linearizira se stavljanjem

$$\psi(x) = \frac{1}{\varphi(x)} = a_0 + a_1 e^{-x}, \quad h_k = \frac{1}{f_k}, \quad k = 0, \ldots, n.$$

Pripadni linearni problem najmanjih kvadrata je

$$\widetilde{S} = \widetilde{S}(a_0, a_1) = \sum_{k=0}^{n} (h_k - a_0 - a_1 e^{-x_k})^2 \to \min .$$

1.1.2. Matrična formulacija linearnog problema najmanjih kvadrata

Da bismo formirali matrični zapis linearnog problema najmanjih kvadrata, moramo preimenovati nepoznanice, naprosto zato da bismo matricu, vektor desne strane i nepoznanice u linearnom sustavu pisali u uobičajenoj formi (standardno su nepoznanice $x_1, \ldots, x_m$, a ne $a_0, \ldots, a_m$).

Pretpostavimo da imamo skup mjerenih podataka (t_k, y_k), $k = 1, \ldots, n$, i da želimo taj model aproksimirati funkcijom oblika $\varphi(t)$. Ako je $\varphi(t)$ linearna, tj. ako je

$$\varphi(t) = x_1\varphi_1(t) + \cdots + x_m\varphi_m(t),$$

onda bismo željeli pronaći parametre x_j tako da mjereni podaci (t_k, y_k) zadovoljavaju

$$y_k = \sum_{j=1}^{m} x_j\varphi_j(t_k), \quad k = 1, \ldots, n.$$

Ako označimo

$$a_{kj} = \varphi_j(t_k), \quad b_k = y_k,$$

onda prethodne jednadžbe možemo u matričnom obliku pisati kao

$$Ax = b.$$

Ako je mjerenih podataka više nego parametara, tj. ako je $n > m$, onda ovaj sustav jednadžbi ima više jednadžbi nego nepoznanica, pa je preodređen.

Kao što smo već u uvodu rekli, postoji mnogo načina da se odredi "najbolje" rješenje, ali zbog statističkih razloga to je često metoda najmanjih kvadrata, tj. određujemo x tako da minimizira grešku $r = Ax - b$ (r se često zove rezidual)

$$\min_x \|r\|_2 = \min_x \|Ax - b\|_2, \quad A \in \mathbb{R}^{n\times m}, \quad b \in \mathbb{R}^n. \tag{1.1.2}$$

Ako je $\operatorname{rang}(A) < m$, onda rješenje x ovog problema očito **nije** jedinstveno, jer mu možemo dodati bilo koji vektor iz nul-potprostora od A, a da se rezidual ne promijeni. S druge strane, među svim rješenjima x problema najmanjih kvadrata uvijek postoji jedinstveno rješenje x najmanje norme, tj. koje još minimizira i $\|x\|_2$.

1.1.3. Karakterizacija rješenja

Prvo, karakterizirajmo skup svih rješenja problema najmanjih kvadrata.

Teorem 1.1.1. *Skup svih rješenja problema najmanjih kvadrata (1.1.2) označimo s*

$$\mathcal{S} = \{x \in \mathbb{R}^m \mid \|Ax - b\|_2 = \min\}.$$

Tada je $x \in \mathcal{S}$ ako i samo ako vrijedi sljedeća relacija ortogonalnosti

$$A^T(b - Ax) = 0. \tag{1.1.3}$$

Dokaz:

Pretpostavimo da $\hat{x}$ zadovoljava

$$A^T\hat{r} = 0, \quad \hat{r} = b - A\hat{x}.$$

Tada za bilo koji $x \in \mathbb{R}^m$ imamo

$$r = b - Ax = \hat{r} + A\hat{x} - Ax = \hat{r} - A(x - \hat{x}).$$

Ako označimo

$$e = x - \hat{x},$$

onda je

$$\|r\|_2^2 = r^T r = (\hat{r} - Ae)^T(\hat{r} - Ae) = \hat{r}^T\hat{r} + \|Ae\|_2^2,$$

što je minimizirano kad je $x = \hat{x}$.

S druge strane, pretpostavimo da je

$$A^T\hat{r} = z \neq 0$$

i uzmimo

$$x = \hat{x} + \varepsilon z.$$

Tada je

$$r = \hat{r} - \varepsilon Az$$

i

$$\|r\|_2^2 = r^T r = \hat{r}^T\hat{r} - 2\varepsilon z^T z + \varepsilon^2(Az)^T(Az) < \hat{r}^T\hat{r}$$

za dovoljno mali ε, pa $\hat{x}$ nije rješenje u smislu najmanjih kvadrata. ■

Relacija (1.1.3) često se zove sustav normalnih jednadžbi i uobičajeno se piše u obliku

$$A^T Ax = A^T b.$$

Matrica $A^T A$ je simetrična i pozitivno semidefinitna, a sustav normalnih jednadžbi je uvijek konzistentan, jer je

$$A^T b \in \mathcal{R}(A^T) = \mathcal{R}(A^T A).$$

Čak štoviše, vrijedi i sljedeći teorem.

Teorem 1.1.2. *Matrica $A^T A$ je pozitivno definitna ako i samo ako su stupci od A linearno nezavisni, tj. ako je* $\operatorname{rang}(A) = m$.

Dokaz:

Ako su stupci od A linearno nezavisni, tada za svaki $x \neq 0$ vrijedi $Ax \neq 0$ (definicija linearne nezavisnosti), pa je za takav x

$$x^T A^T A x = \|Ax\|_2^2 > 0,$$

tj. $A^T A$ je pozitivno definitna.

S druge strane, ako su stupci linearno zavisni, tada postoji $x_0 \neq 0$ takav da je $Ax_0 = 0$, pa je za takav x_0

$$x_0^T A^T A x_0 = 0.$$

Ako je x takav da je $Ax \neq 0$, onda je $x^T A^T A x > 0$, pa je $A^T A$ pozitivno semidefinitna. ■

Iz prethodnog teorema slijedi, da ako je $\text{rang}(A) = m$, onda postoji jedinstveno rješenje problema najmanjih kvadrata, koje je dano s

$$x = (A^T A)^{-1} A^T b, \quad r = b - A(A^T A)^{-1} A^T b.$$

Ako je $S \subset \mathbb{R}^n$ potprostor, onda je $P_S \in \mathbb{R}^{n\times n}$ **ortogonalni projektor** na S, ako je $\mathcal{R}(P_S) = S$ i

$$P_S^2 = P_S, \quad P_S^T = P_S.$$

Nadalje, vrijedi i

$$(I - P_S)^2 = I - P_S, \quad (I - P_S)P_S = 0,$$

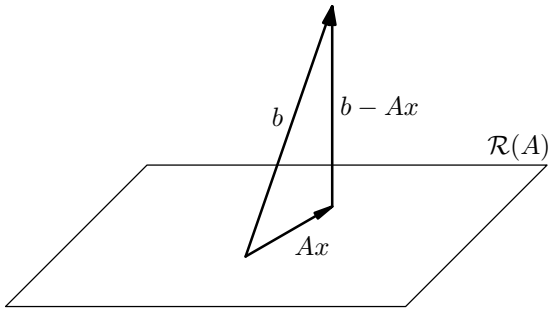
pa je $I - P_S$ projektor na ortogonalni komplement od S.

Tvrdimo da postoji jedinstveni ortogonalni projektor na S. Pretpostavimo da postoje dva ortogonalna projektora P_1 i P_2. Za sve $z \in \mathbb{R}^n$, onda vrijedi

$$\begin{aligned}\|(P_1 - P_2)z\|_2^2 &= z^T(P_1 - P_2)^T(P_1 - P_2)z = z^T(P_1^T P_1 - P_2^T P_1 - P_1^T P_2 + P_2^T P_2)z \\ &= z^T(P_1 - P_2P_1 - P_1P_2 + P_2)z \\ &= z^T P_1(I - P_2)z + z^T P_2(I - P_1)z = 0.\end{aligned}$$

Odatle odmah slijedi da je $P_1 = P_2$, tj. ortogonalni je projektor jedinstven.

Iz geometrijske interpretacije problema najmanjih kvadrata odmah vidimo da je Ax ortogonalna projekcija vektora b na $\mathcal{R}(A)$.



Također

$$r = (I - P_{\mathcal{R}(A)})b$$

i u slučaju punog ranga matrice A vrijedi

$$P_{\mathcal{R}(A)} = A(A^TA)^{-1}A^T.$$

Ako je $\text{rang}(A) < m$, onda A ima netrivijalni nul-potprostor i rješenje problema najmanjih kvadrata nije jedinstveno. Istaknimo jedno od rješenja $\hat{x}$. Skup svih rješenja $\mathcal{S}$ onda možemo opisati kao

$$\mathcal{S} = \{x = \hat{x} + z \mid z \in \mathcal{N}(A)\}.$$

Ako je $\hat{x} \perp \mathcal{N}(A)$, onda je

$$\|x\|_2^2 = \|\hat{x}\|_2^2 + \|z\|_2^2,$$

pa je $\hat{x}$ jedinstveno rješenje problema najmanjih kvadrata koje ima minimalnu 2-normu.

1.2. QR faktorizacija

U mnogim je primjenama simetrična matrica A reda n zadana svojim, generalno, pravokutnim faktorom G, tako da je

$$A = G^TG.$$

Na prvi je pogled jasno da je tako definirana A simetrična. Lako dokazujemo da je tako definirana matrica pozitivno semidefinitna, jer je

$$x^TAx = (x^TG^T)(Gx) = (Gx)^T(Gx) \geq 0. \tag{1.2.1}$$

Da bi A bila pozitivno definitna, potrebni su još neki uvjeti na oblik faktora G.

- Ako je G tipa $m \times n$, onda mora biti $m \geq n$. U protivnom, kad bi bilo $m < n$, onda bi bio $\text{rang}(G) \leq m$, što povlači da je i $\text{rang}(A) \leq m$, tj. A je singularna.
- G mora imati puni stupčani rang, tj. mora biti $\text{rang}(G) = n$. U tom je slučaju nul-potprostor od G (tj. svi oni vektori za koje je $Gx = 0$) trivijalan (samo $x = 0$), pa za sve $x \neq 0$ vrijedi $Gx \neq 0$ i u (1.2.1) izlazi $x^TAx > 0$.

Dakle, pretpostavimo da je pozitivno definitna matrica A zadana svojim faktorom G tipa $m \times n$, $m \geq n$, $\text{rang}(G) = n$. Znamo da se svaka takva matrica A može rastaviti faktorizacijom Choleskog u $A = R^TR$. Pitamo se može li se to napraviti i ako je A implicitno zadana faktorom G. Očito je da može i to eksplicitnim formiranjem matrice G. S numeričke strane takvo eksplicitno formiranje matrice A

nije jako poželjno, prvo zato jer formiranjem elemenata od A radimo neke greške zaokruživanja, a drugo, takav proces predugo traje.

Kad bismo mogli matricu G odmah faktorizirati tako da je

$$G = QR = Q \begin{bmatrix} R_0 \\ 0 \end{bmatrix}, \tag{1.2.2}$$

gdje je Q ortogonalna matrica reda m a R_0 gornjetrokutasta matrica reda n s pozitivnim dijagonalnim elementima, onda bismo lako pročitali faktorizaciju Choleskog matrice A, jer je

$$A = G^T G = R^T Q^T Q R = R^T R = R_0^T R_0,$$

pa bi R_0 bio baš traženi faktor.

Faktorizacija (1.2.2) za G punog stupčanog ranga uvijek postoji i zove se **QR faktorizacija**. Primijetite da smo (1.2.2) mogli pisati i u jednostavnijoj formi, ako prvih n stupaca matrice Q označimo s Q_0 (pazite Q_0 je tipa $m \times n$), onda je

$$G = QR = Q_0 R_0, \quad Q_0^T Q_0 = I_n.$$

Ostaje samo pokazati da QR faktorizacija matrice G postoji.

Teorem 1.2.1. *Neka je $G \in \mathbb{R}^{m\times n}$, $m \geq n$ i neka je* $\operatorname{rang}(G) = n$. *Tad postoji jedinstvena faktorizacija oblika*

$$G = Q_0 R_0,$$

pri čemu je Q_0 tipa $m \times n$, $Q_0^T Q_0 = I_n$, a R_0 gornjetrokutasta s pozitivnim dijagonalnim elementima.

Dokaz:

Najjednostavniji je dokaz ovog teorema je korištenjem Gram-Schmidtove ortogonalizacije. Ako stupce matrice $G = [g_1, g_2, \ldots, g_n]$ ortogonaliziramo slijeva udesno, dobit ćemo ortonormalni niz vektora q_1 do q_n koji razapinje isti potprostor kao i stupci od G. Stavimo li $Q_0 = [q_1, q_2, \ldots, q_n]$, dobili smo $m \times n$ ortogonalnu matricu. Također Gram–Schmidtov postupak ortogonalizacije računa i koeficijente $r_{ji} = q_j^T g_i$ koji polazni stupac g_i izražavaju kao linearnu kombinaciju prvih i vektora q_j ortonormirane baze, tako da je

$$g_i = \sum_{j=1}^{i} r_{ji} q_j.$$

Elementi r_{ji} su elementi matrice R_0. Iz Gram–Schmidtovog algoritma bit će jasno da se može uzati $r_{ii} > 0$. ■

Iako je ovaj dokaz ortogonalizacijom elegantan, u praksi se **nikad** ne koristi Gram–Schmidtov (CGS) postupak ortogonalizacije, jer je nestabilan kad su stupci

od G skoro linearno zavisni. Umjesto toga, može se koristiti tzv. modificirani Gram–Schmidtov postupak (MGS) koji je mnogo stabilniji, ali i kod njega se može dogoditi da je izračunati Q_0 vrlo daleko od ortogonalnog, tj. $\|Q_0^T Q_0 - I\| \gg u$ kad je G loše uvjetovana.

Potpunosti radi, dajemo i CGS i MGS algoritam.

Algoritam 1.2.1. (Klasični i modificirani Gram–Schmidt)

for $i := 1$ **to** n **do**
  {`Nađi` i`-ti stupac od` Q `i` R}
begin
$q_i = g_i$;
for $j := 1$ **to** $i - 1$ **do**
  {`Oduzmi komponentu u` q_j `u smjeru` g_i}
  begin
  $r_{ji} := q_j^T g_i$; {`kod CGS-a`} `ili` $r_{ji} := q_j^T q_i$; {`kod MGS-a`}
  $q_i := q_i - r_{ji} q_j$;
  end;
$r_{ii} := \|q_i\|_2$;
if $r_{ii} = 0$ **do**
  begin
  error := *true*;
  exit;
  end
else
  begin
  error := *false*;
  $q_i := q_i / r_{ii}$;
  end;
end;

Pokažite da su dvije formule za r_{ji} koje koriste CGS i MGS matematički ekvivalentne.

Kad želimo ortogonalan Q, koristimo ili **Givensove rotacije** ili **Householderove reflektore** kojima poništavamo odgovarajuće elemente u matrici G, što će nam, ponovno, dati konstruktivni dokaz teorema 1.2.1.

1.2.1. Givensove rotacije

Matrica

$$R(\varphi) = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$$

je Givensova rotacija koja svaki vektor $x \in \mathbb{R}^2$ rotira obrnuto od kazaljke na satu za kut φ.

Možemo definirati i Givensovu rotaciju u (i, j) ravnini s

$$R(i,j,\varphi) = \begin{bmatrix} & & & \overset{i}{\downarrow} & & & \overset{j}{\downarrow} & & & \\ 1 & & & & & & & & & \\ & \ddots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & \cos\varphi & & & -\sin\varphi & & & \leftarrow i \\ & & & & 1 & & & & & \\ & & & & & \ddots & & & & \\ & & & & & & 1 & & & \\ & & & \sin\varphi & & & \cos\varphi & & & \leftarrow j \\ & & & & & & & 1 & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & 1 \end{bmatrix}$$

Za zadani vektor $x \in \mathbb{R}^m$, lako poništavamo njegovu j-tu komponentu x_j korištenjem rotacije $R(i, j, \varphi)$. Primijetite da množenjem matrice $R(i, j, \varphi)$ na x mijenjamo samo i-tu i j-tu komponentu u x, pa dobivamo

$$\begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} x_i \\ x_j \end{bmatrix} = \begin{bmatrix} x_i' \\ 0 \end{bmatrix}.$$

Drugi redak u matričnoj jednadžbi je

$$\sin\varphi\, x_i + \cos\varphi\, x_j = 0.$$

Ako je $x_j = 0$ nemamo što poništavati, a ako nije, dobivamo

$$\operatorname{ctg}\varphi = -\frac{x_i}{x_j},$$

odakle, korištenjem trigonometrijskog identiteta

$$1 + \operatorname{ctg}^2\varphi = \frac{1}{\sin^2\varphi}$$

slijedi

$$\sin^2\varphi = \frac{x_j^2}{x_i^2 + x_j^2}, \quad \cos^2\varphi = 1 - \sin^2\varphi = \frac{x_i^2}{x_i^2 + x_j^2}.$$

Sada možemo izabrati predznake za $\sin\varphi$ i $\cos\varphi$, tako da x_i' bude pozitivan. Ako stavimo

$$\sin\varphi = -\frac{x_j}{\sqrt{x_i^2 + x_j^2}}, \quad \cos\varphi = \frac{x_i}{\sqrt{x_i^2 + x_j^2}},$$

dobivamo

$$x_i' = \frac{x_i}{\sqrt{x_i^2 + x_j^2}}\, x_i + \frac{x_j}{\sqrt{x_i^2 + x_j^2}}\, x_j = \frac{x_i^2 + x_j^2}{\sqrt{x_i^2 + x_j^2}} = \sqrt{x_i^2 + x_j^2} > 0.$$

Primijetite da je element x_i' dobiven nakon transformacije upravo norma i-te i j-te komponente polaznog vektora.

Sistematskim poništavanjem elemenata, konstruirat ćemo QR faktorizaciju matrice G. Počnimo s prvim stupcem. Redom, možemo poništavati elemente g_{j1}, $j = 2, \ldots m$ korištenjem rotacija $R(1, j, \varphi)$, tj. rotacija koje "nabacuju" normu prvog stupca na prvi element u stupcu. Zatim to možemo ponoviti za drugi, treći i svaki daljnji stupac. Primijetite, da time nećemo "pokvariti" već sređene nule u prethodnim stupcima. Grafički, za jednu matricu tipa 5×3 to izgleda ovako

$$\begin{bmatrix} x & x & x \\ x & x & x \\ x & x & x \\ x & x & x \\ x & x & x \end{bmatrix} \rightarrow \begin{bmatrix} x & x & x \\ 0 & x & x \\ 0 & x & x \\ 0 & x & x \\ 0 & x & x \end{bmatrix} \rightarrow \begin{bmatrix} x & x & x \\ 0 & x & x \\ 0 & 0 & x \\ 0 & 0 & x \\ 0 & 0 & x \end{bmatrix} \rightarrow \begin{bmatrix} x & x & x \\ 0 & x & x \\ 0 & 0 & x \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Za ocjenu greške zaokruživanja postoji i bolji raspored poništavanja elemenata, jer se ovakvom transformacijom, pri "sređivanju" prvog stupca prvi redak mijenja $m - 1$ puta. Kad bismo ujednačili da se svaki redak podjednak broj puta transformira, onda bi naša ocjena greške bila bolja. To možemo postići korištenjem niza nezavisnih rotacija koje ne zahvaćaju iste retke. Osim toga, takav raspored odvijanja rotacija dozvoljava paralelizaciju algoritma.

Naravno, na kraju algoritma, na mjestu matrice G piše matrica R. Kako ćemo pronaći Q? Ako promatramo transformacije koje obavljamo, dobivamo

$$R(n, m, \varphi_{nm}) \cdot R(n, m-1, \varphi_{n,m-1}) \cdots R(1, 2, \varphi_{12}) G := Q^{-1} G = R.$$

Primijetimo da smo matricu G slijeva množili produktom ortogonalnih matrica, koji je i sam ortogonalna matrica, a možemo je označiti s Q^{-1}, pa je $G = QR$.

1.2.2. Householderovi reflektori

Umjesto da elemente u stupcu poništavamo jedan po jedan, korištenjem Householderovih reflektora, možemo odjednom poništiti sve elemente, osim jednog, u dijelu odgovarajućeg stupca.

Matrica H definirana s

$$H = I - 2uu^T, \quad \|u\|_2 = 1$$

zove se Householderov reflektor. Matrica H je simetrična, što je očito, i ortogonalna. Vrijedi

$$\begin{aligned} HH^T = H^2 &= (I - 2uu^T)(I - 2uu^T) = I - 4uu^T + 4uu^T uu^T \\ &= I - 4uu^T + 4u(u^T u)u^T = I - 4uu^T + 4\|u\|_2^2 uu^T = I. \end{aligned}$$

Zašto baš ime reflektor? Promatrajmo hiperravninu koja je okomita na u i prolazi ishodištem. Reflektor H sve vektore x preslikava u simetrični obzirom na tu ravninu.

Ako imamo zadan vektor x, jednostavno je pronaći u za Householderov reflektor tako da poništimo sve osim prve komponente vektora x, tj. tražimo da je

$$Hx = \begin{bmatrix} c \\ 0 \\ \vdots \\ 0 \end{bmatrix} = c \cdot e_1.$$

Raspišimo tu jednadžbu

$$Hx = (I - 2uu^T)x = x - 2u(u^T x) = c \cdot e_1.$$

($u^T x$ je broj!) Odatle, premještanjem pribrojnika, slijedi

$$u = \frac{1}{2(u^T x)}(x - ce_1),$$

tj. u mora biti linearna kombinacija od x i ce_1, tj. mora biti

$$u = \alpha(x - ce_1).$$

Također, zbog unitarne invarijantnosti mora biti

$$\|x\|_2 = \|Hx\|_2 = |c|,$$

pa u mora biti paralelan s vektorom

$$\tilde{u} = x \pm \|x\|_2 e_1, \tag{1.2.3}$$

a jedinične norme, pa je

$$u = \frac{\tilde{u}}{\|\tilde{u}\|_2}.$$

Oba izbora znakova u (1.2.3) zadovoljavaju $Hx = ce_1$, sve dok je $\tilde{u} \neq 0$, ali se u praksi, zbog stabilnosti koristi

$$\tilde{u} = x + \operatorname{sign}(x_1)\|x\|_2 e_1,$$

jer to znači da nema kraćenja pri računanju prve komponente od $\tilde{u}$, koja je jednaka

$$\tilde{u}_1 = x_1 + \text{sign}(x_1)\|x\|_2,$$

tj. oba su pribrojnika istog znaka.

Primijetite da se računanje u može izbjeći, ako definiramo

$$H = I - 2\frac{\tilde{u}\tilde{u}^T}{\tilde{u}^T\tilde{u}}.$$

Ponovno, sustavna primjena Householderovih reflektora na poništavanje elemenata prvog stupca, zatim elemenata drugog stupca od dijagonalnog mjesta nadolje daje konstrukciju QR faktorizacije.

1.2.3. Numeričko rješavanje problema najmanjih kvadrata

Postoji nekoliko načina rješavanja problema najmanjih kvadrata u praksi. Obično se koristi jedna od sljedećih metoda:

1. sustav normalnih jednadžbi,
2. QR faktorizacija,
3. dekompozicija singularnih vrijednosti,
4. transformacija u linearni sustav.

Sustav normalnih jednadžbi

Prva od navedenih metoda je najbrža, ali je najmanje točna. Koristi se kad je A^TA pozitivno definitna i kad je njena uvjetovanost mala. Matrica A^TA rastavi se faktorizacijom Choleskog, a zatim se riješi linearni sustav

$$A^TAx = A^Tb.$$

Ukupan broj aritmetičkih operacija za računanje A^TA, A^Tb, te zatim faktorizaciju Choleskog je $nm^2+\frac{1}{3}m^3+O(m^2)$. Budući da je $n \geq m$, onda je prvi član dominantan u ovom izrazu, a potječe od formiranja A^TA.

Korištenje QR faktorizacije u problemu najmanjih kvadrata

Ponovno, pretpostavimo da je A^TA pozitivno definitna. Polazimo od rješenja problema najmanjih kvadrata dobivenog iz sustava normalnih jednadžbi

$$x = (A^TA)^{-1}A^Tb.$$

Zatim napišemo QR faktorizaciju matrice A

$$A = QR = Q_0 R_0,$$

gdje je Q_0 ortogonalna matrica tipa (n, m), a R_0 trokutasta tipa (m, m) i uvrstimo u rješenje. Dobivamo

$$\begin{aligned} x &= (A^T A)^{-1} A^T b = (R_0^T Q_0^T Q_0 R_0)^{-1} R_0^T Q_0^T b \\ &= (R_0^T R_0)^{-1} R_0^T Q_0^T b = R_0^{-1} R_0^{-T} R_0^T Q_0^T b = R_0^{-1} Q_0^T b, \end{aligned}$$

tj. x se dobiva primjenom "invertirane" skraćene QR faktorizacije od A na b (po analogiji s rješavanjem linearnih sustava, samo što A ne mora imati inverz).

Preciznije, da bismo našli x, rješavamo trokutasti linearni sustav

$$R_0 x = Q_0^T b.$$

Na ovakav se način najčešće rješavaju problemi najmanjih kvadrata. Nije teško pokazati da je cijena računanja $2nm^2 - \frac{2}{3}m^3$, što je dvostruko više nego za sustav normalnih jednadžbi kad je $n \gg m$, a približno jednako za $m = n$.

QR faktorizacija može se koristiti i za problem najmanjih kvadrata kad matrica A nema puni stupčani rang, ali tada se koristi QR faktorizacija sa stupčanim pivotiranjem (na prvo mjesto dovodi se stupac čiji "radni dio" ima najveću normu). Zašto baš tako? Ako matrica A ima rang $r < m$, onda njena QR faktorizacija ima oblik

$$A = QR = Q \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

gdje je R_{11} nesingularna reda r, a R_{12} neka $r \times (m - r)$ matrica. Zbog grešaka zaokruživanja, umjesto pravog R, izračunamo

$$R' = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \\ 0 & 0 \end{bmatrix}.$$

Naravno, željeli bismo da je $\|R_{22}\|_2$ vrlo mala, reda veličine $\varepsilon\|A\|_2$, pa da je možemo "zaboraviti", tj. staviti $R_{22} = 0$ i tako odrediti rang od A. Nažalost, to nije uvijek tako. Na primjer, bidijagonalna matrica

$$A = \begin{bmatrix} \frac{1}{2} & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \frac{1}{2} \end{bmatrix}$$

je skoro singularna ($\det(A) = 2^{-n}$), njena QR faktorizacija je $Q = I$, $R = A$, i nema niti jednog R_{22} koji bi bio po normi malen.

Zbog toga koristimo pivotiranje, koje R_{11} pokušava držati što bolje uvjetovanim, a R_{22} po normi što manjim.

Dekompozicija singularnih vrijednosti i problem najmanjih kvadrata

Vjerojatno jedna od najkorisnijih dekompozicija i s teoretske strane (za dokazivanje činjenica) i s praktične strane je dekompozicija singularnih vrijednosti (engl. "singular value decomposition") ili, skraćeno, SVD.

Teorem 1.2.2. *Neka je A proizvoljna matrica tipa $n \times m$, $n \geq m$. Tada se A može dekomponirati kao*

$$A = \widehat{U} \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^* = U \Sigma V^*,$$

gdje je

$$\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_m), \quad \sigma_1 \geq \cdots \geq \sigma_m \geq 0,$$

$\widehat{U} = [U, U_0]$ je $n \times n$, a V je $m \times m$ unitarna matrica. Stupce matrice U (u oznaci u_i) zovemo lijevi singularni vektori, stupce matrice V (u oznaci v_i) desni singularni vektori, a dijagonalne elemente matrice Σ singularne vrijednosti. Ako je $n < m$, dekompozicija singularnih vrijednosti definira se za A^*. Ako je A realna, U i V su, također, realne.

Umjesto dokaza (v. proširene verzije skripte), objasnimo značenje dekompozicije. Ako o matrici A razmišljamo kao zapisu operatora koji preslikava vektor $x \in \mathbb{R}^m$ u vektor $y = Ax \in \mathbb{R}^n$, onda možemo izabrati ortogonalni koordinatni sustav u $\mathbb{R}^m$ (osi su mu jedinični vektori stupci u V) i drugi ortogonalni koordinatni sustav u $\mathbb{R}^n$ (osi su mu jedinični vektori stupci u U), takve da je zapis tog operatora u tom paru baza dijagonalna matrica.

Drugim riječima, A preslikava vektor

$$x = \sum_{i=1}^{m} \beta_i v_i$$

u

$$y = Ax = \sum_{i=1}^{m} \sigma_i \beta_i u_i,$$

tj. svaka se matrica može "dijagonalizirati" u **paru** baza, ako smo joj za domenu i sliku izabrali odgovarajuće ortogonalne koordinatne sustave (baze).

Veza između SVD-a i problema najmanjih kvadrata dana je sljedećom tvrdnjom.

Teorem 1.2.3. *Neka je $A = U \Sigma V^T$ dekompozicija singularnih vrijednosti (SVD) realne matrice A tipa $n \times m$, $n \geq m$.*

Tvrdnja 1. *Ako A ima puni rang, onda je rješenje problema najmanjih kvadrata*

$$\min_x \|Ax - b\|_2$$

jednako

$$x = V\Sigma^{-1}U^T b,$$

tj. dobiva se primjenom "invertiranog" skraćenog SVD-a od A na b.

Dokaz:

Vrijedi

$$\|Ax - b\|_2^2 = \|U\Sigma V^T x - b\|_2^2.$$

Budući da je A punog ranga, to je i Σ. Zbog unitarne ekvivalencije 2-norme, vrijedi

$$\begin{aligned}\|U\Sigma V^T x - b\|_2^2 = \|\widehat{U}^T(U\Sigma V^T x - b)\|_2^2 &= \left\| \begin{bmatrix} U^T \\ U_0^T \end{bmatrix} (U\Sigma V^T x - b) \right\|_2^2 \\ &= \left\| \begin{bmatrix} \Sigma V^T x - U^T b \\ -U_0^T b \end{bmatrix} \right\|_2^2 = \|\Sigma V^T x - U^T b\|_2^2 + \|U_0^T b\|_2^2.\end{aligned}$$

Prethodni izraz se minimizira ako je prvi član jednak 0, tj. ako je

$$x = V\Sigma^{-1}U^T b.$$

Usput dobivamo i vrijednost minimuma $\min\limits_x \|Ax - b\|_2 = \|U_0^T b\|_2$. ■

Uočite da u prethodnom teoremu piše rješenje problema najmanjih kvadrata kad je matrica A punog ranga. Uobičajeno se SVD primjenjuje u metodi najmanjih kvadrata i kad matrica A nema puni stupčani rang. Rješenja su istog oblika (sjetite se, više ih je), samo što moramo znati izračunati "inverz" matrice Σ kad ona nije regularna, tj. kad ima neke nule na dijagonali. Takav inverz zove se generalizirani inverz i označava sa Σ^+ ili $\Sigma^\dagger$. U slučaju da je

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix},$$

pri čemu je Σ_1 regularna, onda je

$$\Sigma^+ = \begin{bmatrix} \Sigma_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Još preciznije, za problem najmanjih kvadrata tada vrijedi sljedeća propozicija.

Propozicija 1.2.1. *Neka matrica A ima rang $r < m$. Rješenje x koje minimizira $\|Ax - b\|_2$ može se karakterizirati na sljedeći način. Neka je $A = U\Sigma V^T$ SVD od A i neka je*

$$A = U\Sigma V^T = [U_1, U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} [V_1, V_2]^T = U_1\Sigma_1 V_1^T,$$

gdje je Σ_1 nesingularna, reda r, a matrice U_1 i V_1 imaju r stupaca. Neka je

$$\sigma := \sigma_{\min}(\Sigma_1),$$

najmanja ne-nula singularna vrijednost od A. Tada se sva rješenja problema najmanjih kvadrata mogu napisati u formi

$$x = V_1\Sigma_1^{-1}U_1^T b + V_2 z,$$

gdje je z proizvoljni vektor. Rješenje x koje ima minimalnu 2-normu je ono za koje je $z = 0$, tj.

$$x = V_1\Sigma_1^{-1}U_1^T b,$$

i vrijedi ocjena

$$\|x\|_2 \leq \frac{\|b\|_2}{\sigma}.$$

Dokaz:

Nadopunimo matricu $[U_1, U_2]$ stupcima matrice U_3 do ortogonalne matrice reda n, i označimo je s $\widehat{U}$. Korištenjem unitarne invarijantnosti 2-norme, dobivamo

$$\|Ax-b\|_2^2 = \|\widehat{U}^T(Ax-b)\|_2^2 = \left\| \begin{bmatrix} U_1^T \\ U_2^T \\ U_3^T \end{bmatrix} (U_1\Sigma_1 V_1^T x - b) \right\|_2^2$$
$$= \left\| \begin{bmatrix} \Sigma_1 V_1^T x - U_1^T b \\ -U_2^T b \\ -U_3^T b \end{bmatrix} \right\|_2^2 = \|\Sigma_1 V_1^T x - U_1^T b\|_2^2 + \|U_2^T b\|_2^2 + \|U_3^T b\|_2^2.$$

Očito, izraz je minimiziran kad je prva od tri norme u posljednjem redu jednaka 0, tj. ako je

$$\Sigma_1 V_1^T x = U_1^T b,$$

ili

$$x = V_1\Sigma_1^{-1}U_1^T b.$$

Stupci matrica V_1 i V_2 su međusobno ortogonalni, pa je $V_1^T V_2 z = 0$ za sve vektore z. Odavde vidimo da x ostaje rješenje problema najmanjih kvadrata i kad mu dodamo $V_2 z$, za bilo koji z, tj. ako je

$$x = V_1\Sigma_1^{-1}U_1^T b + V_2 z.$$

To su ujedno i sva rješenja, jer stupci matrice V_2 razapinju nul-potprostor $\mathcal{N}(A)$ (tvrdnja 7 teorema 1.2.3. u proširenoj skripti). Osim toga, zbog spomenute ortogonalnosti vrijedi i

$$\|x\|_2^2 = \|V_1\Sigma_1^{-1}U_1^T b\|_2^2 + \|V_2 z\|_2^2,$$

a to je minimalno za $z = 0$. Na kraju, za to minimalno rješenje vrijedi ocjena

$$\|x\|_2 = \|V_1\Sigma_1^{-1}U_1^T b\|_2 = \|\Sigma_1^{-1}U_1^T b\|_2 \leq \frac{\|U_1^T b\|_2}{\sigma} = \frac{\|b\|_2}{\sigma}.$$

Primjerom se lako pokazuje da je ova ocjena dostižna. ■

Rješenje problema najmanjih kvadrata korištenjem SVD-a je najstabilnije, a može se pokazati da je, za $n \gg m$, njegovo trajanje približno jednako kao i trajanje rješenja korištenjem QR-a. Za manje n, trajanje je približno $4nm^2 - \frac{4}{3}m^3 + O(m^2)$.

Transformiranje problema najmanjih kvadrata na linearni sustav

Ako matrica A ima puni rang po stupcima, onda problem najmanjih kvadrata možemo transformirati i na linearni sustav različit od sustava normalnih jednadžbi. Simetrični linearni sustav

$$\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} r \\ x \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix},$$

ekvivalentan je sustavu normalnih jednadžbi. Ako napišemo prvu i drugu blok-komponentu

$$r + Ax = b, \quad A^T r = 0,$$

onda uvrštavanjem r-a iz prve blok-jednadžbe u drugu dobivamo sustav

$$A^T(b - Ax) = 0.$$

Prvi sustav ima bitno manji raspon elemenata od sustava normalnih jednadžbi. Osim toga, ako je matrica A loše uvjetovana, kod tog sustava možemo lakše koristiti iterativno profinjavanje rješenja.

1.3. Opći oblik metode najmanjih kvadrata

Nakon što smo napravili osnovni oblik diskretne metode najmanjih kvadrata, na sličan način možemo riješiti i opći problem aproksimacije po metodi najmanjih kvadrata, tj. u 2-normi. Dovoljno je uočiti da je diskretna 2-norma generirana običnim euklidskim skalarnim produktom na konačno dimenzionalnim prostorima. Po istom principu, u općem slučaju, radimo na nekom unitarnom prostoru s nekim skalarnim produktom, a pripadna norma je generirana tim skalarnim produktom.

Na početku zgodno je uvesti oznake koje nam omogućavaju da diskretni i neprekidni slučaj analiziramo odjednom, u istom općem okruženju unitarnih prostora.

1.3.1. Težinski skalarni produkti

Unitarni prostor $\mathcal{U}$ je vektorski prostor na kojem je definiran skalarni produkt.

1.4. Familije ortogonalnih funkcija

Za dvije funkcije reći ćemo da su ortogonalne, ako je njihov skalarni produkt jednak 0. Na primjer, za neprekidnu ili diskretnu mjeru $d\lambda$, te funkcije u i v koje imaju konačnu normu možemo definirati skalarni produkt kao

$$\int_{\mathbb{R}} u(x)\, v(x)\, d\lambda.$$

Postoji mnogo familija ortogonalnih funkcija. Evo nekoliko primjera takvih familija (sistema).

- Ortogonalni polinomi;
- Trigonometrijski polinomi.

1.5. Neka svojstva ortogonalnih polinoma

Ortogonalni polinomi imaju još i niz dodatnih "dobrih" svojstava, zbog kojih se mogu konstruktivno primijeniti u raznim granama numeričke matematike. Sljedeći niz teorema sadrži samo neka osnovna svojstva koja ćemo kasnije iskoristiti za konstrukciju algoritama. Sva ta svojstva su direktna posljedica ortogonalnosti polinoma i ne ovise bitno o tome da li je skalarni produkt diskretan ili kontinuiran.

Međutim, na ovom mjestu je zgodno napraviti razliku između diskretnih i kontinuiranih skalarnih produkata, prvenstveno radi jednostavnosti iskaza, dokaza i kasnijeg pozivanja na ove teoreme. Pažljivije čitanje će samo potvrditi da bitne razlike nema.

Standardno ćemo promatrati neprekidni skalarni produkt

$$\langle u, v\rangle = \int_a^b w(x)u(x)v(x)\, dx$$

generiran težinskom funkcijom $w \geq 0$ na $[a, b]$. Ako svi polinomi pripadaju odgovarajućem prostoru kvadratno integrabilnih funkcija, onda postoji pripadna familija ortogonalnih polinoma koju označavamo s $\{p_n(x) \mid n \geq 0\}$. Dogovorno smatramo da je stupanj polinoma p_n baš jednak n, za svaki $n \geq 0$.

Paralelno ćemo promatrati i diskretni skalarni produkt

$$\langle u, v\rangle = \sum_{i=0}^{n} w_i u(x_i) v(x_i)$$

generiran međusobno različitim čvorovima $x_0, \ldots, x_n$ i pripadnim pozitivnim težinama $w_0, \ldots, w_n$. Pripadni unitarni prostor "funkcija" na zadanoj mreži čvorova (izomorfno) sadrži sve polinome stupnja manjeg ili jednakog n, pa sigurno postoji pripadna baza ortogonalnih polinoma koju označavamo s $\{p_k(x) \mid 0 \leq k \leq n\}$. Opet uzimamo je stupanj polinoma p_k baš jednak k, za svaki $k \in \{0, \ldots, n\}$.

Teorem 1.5.1. *Neka je $\{p_n(x) \mid n \geq 0\}$ familija ortogonalnih polinoma na intervalu $[a, b]$ s težinskom funkcijom $w(x) \geq 0$. Ako je f polinom stupnja m, tada vrijedi*

$$f = \sum_{n=0}^{m} \frac{\langle f, p_n\rangle}{\langle p_n, p_n\rangle} p_n.$$

Dokaz:

Prvo, pokažimo da se svaki polinom može napisati kao kombinacija ortogonalnih polinoma stupnja manjeg ili jednakog njegovom.

Dokaz ide korištenjem Gram–Schmidtove ortogonalizacije. Pokažimo, redom, da se monomi $\{1, x, x^2, \ldots\}$ mogu prikazati pomoću ortogonalnih polinoma.

Ako je stupanj ortogonalnog polinoma 0, on je nužno konstanta različita od nule, tj. vrijedi

$$p_0(x) = c_{0,0}, \quad c_{0,0} \neq 0,$$

pa se prvi monom 1 može napisati kao

$$1 = \frac{1}{c_{0,0}} p_0(x).$$

Za polinome stupnja jedan, konstrukcija slijedi iz Gram–Schmidtovog procesa ortogonalizacije sustava funkcija $\{1, x\}$

$$p_1(x) = c_{1,1}x + c_{1,0}p_0(x), \quad c_{1,1} \neq 0,$$

tj. vrijedi

$$x = \frac{1}{c_{1,1}} [p_1(x) - c_{1,0}p_0(x)].$$

Korištenjem indukcije u Gram–Schmidtovom procesu na $\{1, x, x^2, \ldots, x^n\}$, dobivamo

$$p_n(x) = c_{n,n}x^n + c_{n,n-1}p_{n-1}(x) + \cdots + c_{n,0}p_0(x), \quad c_{n,n} \neq 0,$$

gdje su $p_0, p_1, \ldots, p_{n-1}$ dobiveni ortogonalizacijom iz $\{1, x, \ldots, x^{n-1}\}$, pa je

$$x^n = \frac{1}{c_{n,n}}[p_n(x) - c_{n,n-1}p_{n-1}(x) - \cdots - c_{n,0}p_0(x)].$$

Neka je f bilo koji polinom stupnja (manjeg ili jednakog) m, za neki $m \in \mathbb{N}_0$. Tada se f može napisati kao linearna kombinacija monoma $\{1, x, \ldots, x^m\}$, prikazom u standardnoj bazi. Budući da se svaki monom može napisati kao linearna kombinacija ortogonalnih polinoma stupnja manjeg ili jednakog od stupnja tog monoma, odmah slijedi da se i f može napisati kao neka linearna kombinacija ortogonalnih polinoma stupnjeva manjih ili jednakih m, tj. da vrijedi

$$f = \sum_{j=0}^{m} b_j p_j.$$

Ostaje samo odrediti koeficijente b_j. Množenjem prethodne relacije težinskom funkcijom w, pa polinomom p_n, a zatim integriranjem na $[a, b]$, tj. skalarnim množenjem s p_n, dobivamo

$$\langle f, p_n\rangle = \sum_{j=0}^{m} b_j \langle p_j, p_n\rangle = b_n \langle p_n, p_n\rangle,$$

koristeći ortogonalnost p_j i p_n, za $j \neq n$. Odatle odmah slijedi da je

$$b_n = \frac{\langle f, p_n\rangle}{\langle p_n, p_n\rangle},$$

jer je $\|p_n\|^2 = \langle p_n, p_n\rangle > 0$. ■

Razvoj polinoma f stupnja m iz prethodnog teorema možemo napisati i tako da suma ide do ∞, a ne do m, samo su svi dodatni koeficijenti $b_n = 0$, za $n > m$. To je posljedica sljedeće tvrdnje.

Korolar 1.5.1. *Ako je f polinom stupnja manjeg ili jednakog $m-1$, onda je*

$$\langle f, p_m\rangle = 0,$$

tj. p_m je okomit na f. Dakle, p_m je okomit na sve polinome stupnja strogo manjeg od m.

Dokaz:

Po prethodnom teoremu, f se može razviti po ortogonalnim polinomima stupnja manjeg ili jednakog $m-1$

$$f(x) = \sum_{n=0}^{m-1} b_n p_n(x).$$

Množenjem s $w(x)p_m(x)$, te integriranjem, dobivamo da je

$$\langle f, p_m \rangle = \sum_{n=0}^{m-1} b_n \langle p_n, p_m \rangle = 0,$$

zbog svojstva ortogonalnosti ortogonalnih polinoma $\langle p_n, p_m \rangle = 0$, za $n \neq m$. ■

Teorem 1.5.2. *Neka je $\{p_n(x) \mid n \geq 0\}$ familija ortogonalnih polinoma na intervalu $[a, b]$ s težinskom funkcijom $w(x) \geq 0$. Tada svaki polinom p_n ima točno n različitih (jednostrukih) realnih nultočaka na otvorenom intervalu (a, b).*

Dokaz:

Neka su $x_1, x_2, \ldots, x_m$ sve nultočke polinoma p_n za koje vrijedi:

- $a < x_i < b$,
- $p_n(x)$ mijenja predznak u x_i.

Budući da je p_n stupnja n, po osnovnom teoremu algebre, polinom p_n ima ukupno n nultočaka, pa onih koje zadovoljavaju prethodna dva svojstva ima manje ili jednako n. Pretpostavimo da je nultočaka koje zadovoljavaju tražena dva svojstva striktno manje od n, tj. $m < n$. Pokažimo da je to nemoguće.

Definiramo polinom

$$B(x) = (x - x_1) \cdots (x - x_m).$$

Po definiciji točaka $x_1, \ldots, x_m$, polinom

$$p_n(x)B(x) = (x - x_1) \cdots (x - x_m)p_n(x)$$

ne mijenja znak prolaskom kroz točke $x_1, \ldots, x_m$, tj. čitav polinom ne mijenja znak na (a, b). Preciznije, to implicira oblik funkcije p_n

$$p_n(x) = h(x)(x - x_1)^{r_1} \cdots (x - x_m)^{r_m},$$

pri čemu moraju biti svi r_i neparni, a $h(x)$ ne smije promijeniti predznak na (a, b). Množenjem s $B(x)$, dobivamo

$$p_n(x)B(x) = h(x)(x - x_1)^{r_1+1} \cdots (x - x_m)^{r_m+1}.$$

Nadalje, vrijedi

$$\int_a^b w(x)B(x)p_n(x)\,dx \neq 0,$$

budući da je to integral nenegativne funkcije. S druge je strane taj integral skalarni produkt od B (polinom stupnja $m < n$) i sa p_n (polinom stupnja n), pa je po prethodnom korolaru

$$\int_a^b w(x)B(x)p_n(x)\,dx = \langle B, p_n\rangle = 0.$$

To je, očito kontradikcija, pa je pretpostavka o stupnju polinoma B bila pogrešna, tj. mora biti $m = n$. Budući da p_n ima točno n nultočaka $x_1, \ldots, x_n$ u kojima mijenja predznak, one moraju biti jednostruke, jer je $p_n'(x_i) \neq 0$. ■

Neka je ponovno zadana familija ortogonalnih polinoma na intervalu $[a, b]$ i neka su prva dva koeficijenta funkcije p_n jednaki

$$p_n(x) = A_n x^n + B_n x^{n-1} + \cdots.$$

Također, tada p_n možmo napisati i kao

$$p_n(x) = A_n(x - x_{n,1})(x - x_{n,2}) \cdots (x - x_{n,n}).$$

Definiramo također i

$$a_n = \frac{A_{n+1}}{A_n}, \quad \gamma_n = \langle p_n, p_n\rangle > 0.$$

Teorem 1.5.3. (tročlana rekurzija) *Neka je $\{p_n(x) \mid n \geq 0\}$ familija ortogonalnih polinoma na intervalu $[a, b]$ s težinskom funkcijom $w(x) \geq 0$. Tada za $n \geq 1$ vrijedi rekurzija*

$$p_{n+1}(x) = (a_n x + b_n)p_n(x) - c_n p_{n-1}(x),$$

pri čemu su

$$b_n = a_n\left(\frac{B_{n+1}}{A_{n+1}} - \frac{B_n}{A_n}\right), \quad c_n = \frac{A_{n+1}A_{n-1}}{A_n^2} \cdot \frac{\gamma_n}{\gamma_{n-1}}.$$

Dokaz:

Promatrajmo polinom

$$\begin{aligned} G(x) &= p_{n+1}(x) - a_n x p_n(x) \\ &= (A_{n+1}x^{n+1} + B_{n+1}x^n + \cdots) - \frac{A_{n+1}}{A_n}x(A_n x^n + B_n x^{n-1} + \cdots) \\ &= \left(B_{n+1} - \frac{A_{n+1}B_n}{A_n}\right)x^n + \cdots. \end{aligned}$$

Očito, polinom G je stupnja manjeg ili jednakog n, pa ga možemo napisati kao linearnu kombinaciju ortogonalnih polinoma stupnja manjeg ili jednakog n, tj.

$$G(x) = d_n p_n(x) + \cdots + d_0 p_0(x)$$

za neki skup konstanti d_i. Računanjem d_i izlazi

$$d_i = \frac{\langle G, p_i\rangle}{\langle p_i, p_i\rangle} = \frac{1}{\gamma_i}(\langle p_{n+1}, p_i\rangle - a_n\langle xp_n, p_i\rangle).$$

Budući da je $\langle p_{n+1}, p_i\rangle = 0$ za $i \leq n$ i da za $i \leq n-2$ vrijedi

$$\langle xp_n, p_i\rangle = \int_a^b w(x)p_n(x)xp_i(x)\,dx = 0,$$

zaključujemo da je stupanj polinoma $xp_i(x)$ manji ili jednak $n-1$. Kombiniranjem ta dva rezultata, dobivamo

$$d_i = 0 \quad \text{za } 0 \leq i \leq n-2,$$

pa je zbog toga

$$G(x) = d_n p_n(x) + d_{n-1}p_{n-1}(x)$$
$$p_{n+1}(x) = (a_n x + d_n)p_n(x) + d_{n-1}p_{n-1}(x).$$

Ostaje još samo pokazati koliki su koeficijenti d_{n-1} i d_n. Iz prve od dvije prethodne relacija, uspoređivanjem vodećih koeficijenata funkcije G i vodećih koeficijenata funcije s desne strane, dobivamo relaciju za d_n. ■

Teorem 1.5.4. (Christoffel–Darbouxov identitet) *Neka je* $\{p_n(x) \mid n \geq 0\}$ *familija ortogonalnih polinoma na intervalu* $[a, b]$ *s težinskom funkcijom* $w(x) \geq 0$. *Za njih vrijedi sljedeći identitet*

$$\sum_{k=0}^{n} \frac{p_k(x)p_k(y)}{\gamma_k} = \frac{p_{n+1}(x)p_n(y) - p_n(x)p_{n+1}(y)}{a_n\gamma_n(x-y)}.$$

Dokaz:

Manipulacijom tročlane rekurzije. ■

1.5.1. Klasični ortogonalni polinomi

U aproksimacijama i rješavanju diferencijalnih jednadžbi najčešće se susrećemo s pet tipova klasičnih ortogonalnih polinoma. Za polinome

$$\{p_0, p_1, p_2, \ldots, p_n, \ldots\},$$

pri čemu indeks polinoma označava njegov stupanj, reći ćemo da su ortogonalni obzirom na težinsku funkciju w, $w(x) \geq 0$ na intervalu $[a, b]$, ako vrijedi

$$\int_a^b w(x)\,p_m(x)\,p_n(x)\,dx = 0, \quad \text{za } m \neq n.$$

Težinska funkcija određuje sistem polinoma do na konstantni faktor u svakom od polinoma. Izbor takvog faktora zove se još i standardizacija ili normalizacija.

Čebiševljevi polinomi prve vrste

Čebiševljevi polinomi prve vrste obično se označavaju s T_n. Oni su ortogonalni na intervalu $[-1, 1]$ obzirom na težinsku funkciju

$$w(x) = \frac{1}{\sqrt{1-x^2}}.$$

Vrijedi

$$\int_{-1}^{1} \frac{T_m(x)\,T_n(x)}{\sqrt{1-x^2}}\,dx = \begin{cases} 0, & \text{za } m \neq n, \\ \pi, & \text{za } m = n = 0, \\ \pi/2, & \text{za } m = n \neq 0. \end{cases}$$

Oni zadovoljavaju rekurzivnu relaciju

$$T_{n+1}(x) - 2xT_n(x) + T_{n-1}(x) = 0,$$

uz start

$$T_0(x) = 1, \quad T_1(x) = x.$$

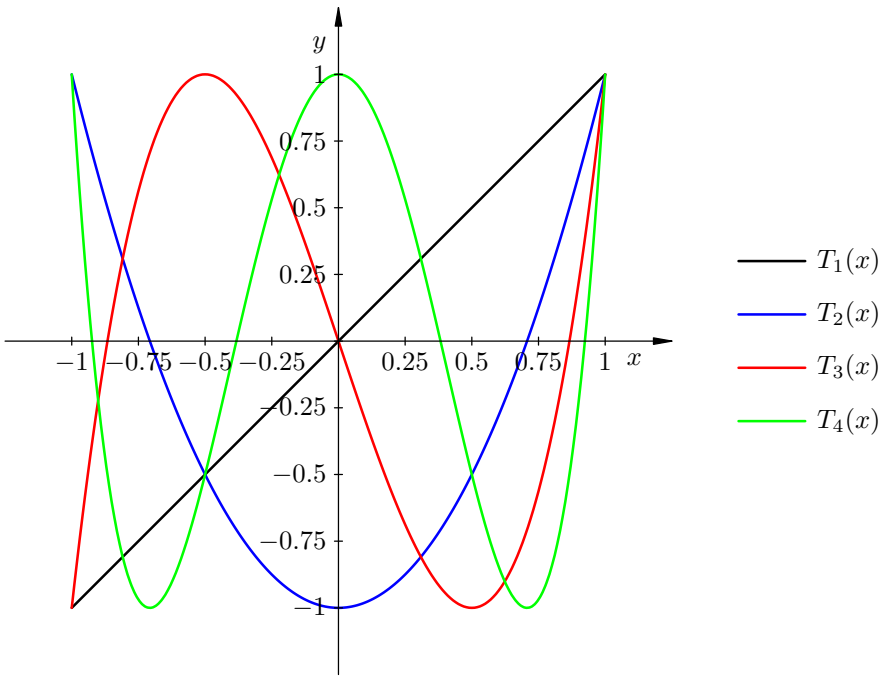
Za njih postoji i eksplicitna formula

$$T_n(x) = \cos(n \arccos x).$$

Osim toga, n-ti Čebiševljev polinom prve vrste T_n zadovoljava diferencijalnu jednadžbu

$$(1-x^2)y'' - xy' + n^2 y = 0.$$

Graf prvih par polinoma izgleda ovako.



Katkad se koriste i Čebiševljevi polinomi prve vrste transformirani na interval $[0,1]$, u oznaci T_n^*. Korištenjem linearne (preciznije, afine) transformacije

$$[0,1] \ni x \mapsto \xi := 2x - 1 \in [-1,1]$$

dolazimo do svih svojstava tih polinoma. Na primjer, relacija ortogonalnosti tada postaje

$$\int_0^1 \frac{T_m^*(x)\, T_n^*(x)}{\sqrt{x - x^2}}\, dx = \begin{cases} 0, & \text{za } m \neq n, \\ \pi, & \text{za } m = n = 0, \\ \pi/2, & \text{za } m = n \neq 0, \end{cases}$$

a rekurzivna relacija

$$T_{n+1}^*(x) - 2(2x-1)T_n^*(x) + T_{n-1}^*(x) = 0,$$

uz start

$$T_0^*(x) = 1, \quad T_1^*(x) = 2x - 1.$$

Čebiševljevi polinomi druge vrste

Čebiševljevi polinomi druge vrste obično se označavaju s U_n. Oni su ortogonalni na intervalu $[-1,1]$ obzirom na težinsku funkciju

$$w(x) = \sqrt{1 - x^2}.$$

Vrijedi

$$\int_{-1}^1 \sqrt{1 - x^2}\, U_m(x)\, U_n(x)\, dx = \begin{cases} 0, & \text{za } m \neq n, \\ \pi/2, & \text{za } m = n. \end{cases}$$

Oni zadovoljavaju istu rekurzivnu relaciju kao Čebiševljevi polinomi prve vrste

$$U_{n+1}(x) - 2xU_n(x) + U_{n-1}(x) = 0,$$

samo uz malo drugačiji start

$$U_0(x) = 1, \quad U_1(x) = 2x.$$

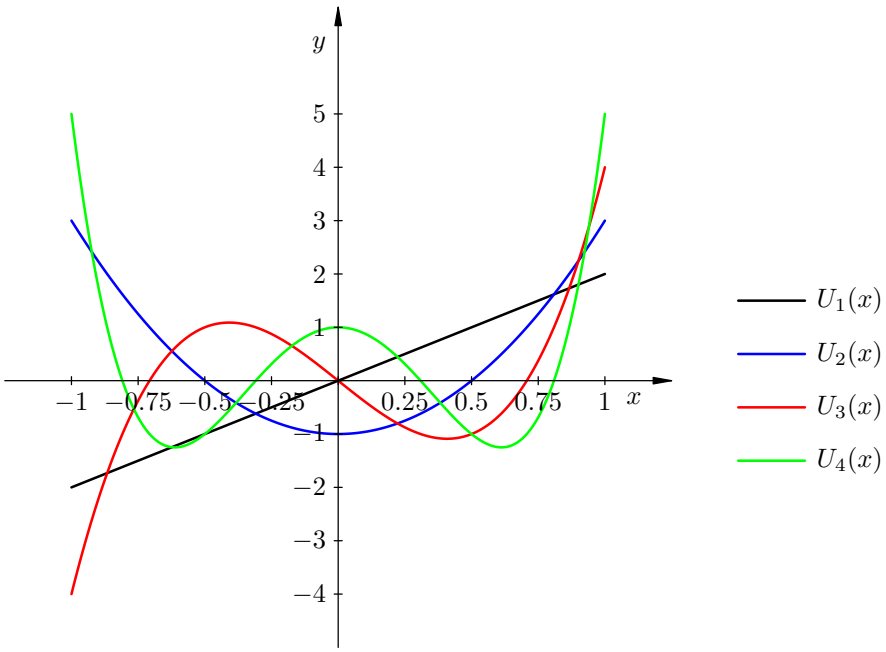
Za njih postoji i eksplicitna formula

$$U_n(x) = \frac{\sin((n+1)\arccos x)}{\sin(\arccos x)}.$$

Osim toga, n-ti Čebiševljev polinom druge vrste U_n zadovoljava diferencijalnu jednadžbu

$$(1 - x^2)y'' - 3xy' + n(n+2)y = 0.$$

Graf prvih par polinoma izgleda ovako.



Legendreovi polinomi

Legendreovi polinomi obično se označavaju s P_n. Oni su ortogonalni na intervalu $[-1, 1]$ obzirom na težinsku funkciju

$$w(x) = 1.$$

Vrijedi

$$\int_{-1}^{1} P_m(x)\, P_n(x)\, dx = \begin{cases} 0, & \text{za } m \neq n, \\ 2/(2n+1), & \text{za } m = n. \end{cases}$$

Oni zadovoljavaju rekurzivnu relaciju

$$(n+1)P_{n+1}(x) - (2n+1)xP_n(x) + nP_{n-1}(x) = 0,$$

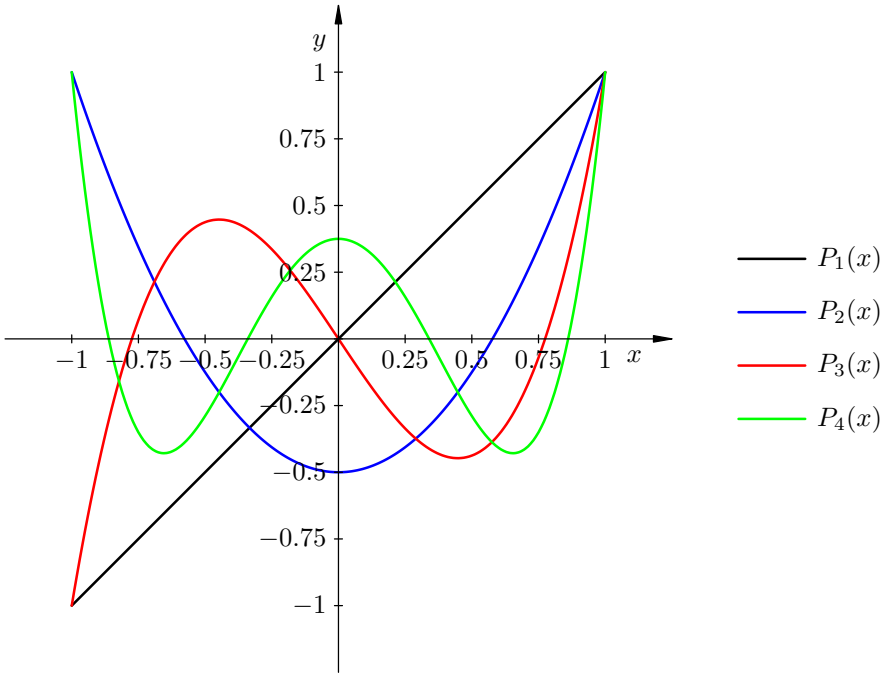
uz start

$$P_0(x) = 1, \quad P_1(x) = x.$$

Osim toga, n-ti Legendreov polinom P_n zadovoljava diferencijalnu jednadžbu

$$(1-x^2)y'' - 2xy' + n(n+1)y = 0.$$

Graf prvih par polinoma izgleda ovako.



Laguerreovi polinomi

Laguerreovi polinomi obično se označavaju s L_n. Oni su ortogonalni na intervalu $[0, \infty)$ obzirom na težinsku funkciju

$$w(x) = e^{-x}.$$

Vrijedi

$$\int_0^\infty e^{-x} L_m(x) L_n(x)\, dx = \begin{cases} 0, & \text{za } m \neq n, \\ 1, & \text{za } m = n. \end{cases}$$

Oni zadovoljavaju rekurzivnu relaciju

$$(n+1)L_{n+1}(x) + (x - 2n - 1)L_n(x) + nL_{n-1}(x) = 0,$$

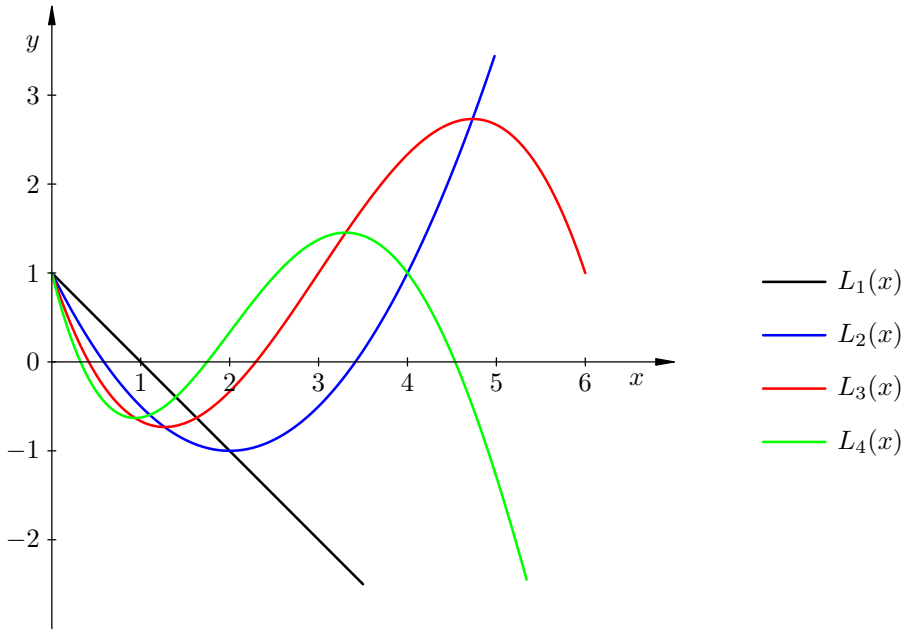
uz start

$$L_0(x) = 1, \quad L_1(x) = 1 - x.$$

Osim toga, n-ti Laguerreov polinom L_n zadovoljava diferencijalnu jednadžbu

$$xy'' + (1 - x)y' + ny = 0.$$

Graf prvih par polinoma izgleda ovako.



U literaturi se često nailazi na još jednu rekurziju za Laguerreove polinome

$$\widetilde{L}_{n+1}(x) + (x - 2n - 1)\widetilde{L}_n(x) + n^2\widetilde{L}_{n-1}(x) = 0,$$

uz jednaki start

$$\widetilde{L}_0(x) = 1, \quad \widetilde{L}_1(x) = 1 - x.$$

Uspoređivanjem ove i prethodne rekurzije dobivamo da je

$$\widetilde{L}_n(x) = n!\, L_n(x),$$

tj. radi se samo o drugačijoj normalizaciji ortogonalnih polinoma. Lako je pokazati da vrijedi

$$\int_0^\infty e^{-x}\, \widetilde{L}_m(x)\, \widetilde{L}_n(x)\, dx = \begin{cases} 0, & \text{za } m \neq n, \\ (n!)^2, & \text{za } m = n. \end{cases}$$

Hermiteovi polinomi

Hermiteovi polinomi obično se označavaju s H_n. Oni su ortogonalni na intervalu $(-\infty, \infty)$ obzirom na težinsku funkciju

$$w(x) = e^{-x^2}.$$

Vrijedi

$$\int_{-\infty}^\infty e^{-x^2}\, H_m(x)\, H_n(x)\, dx = \begin{cases} 0, & \text{za } m \neq n, \\ 2^n\, n!\sqrt{\pi}, & \text{za } m = n. \end{cases}$$

Oni zadovoljavaju rekurzivnu relaciju

$$H_{n+1}(x) - 2xH_n(x) + 2nH_{n-1}(x) = 0,$$

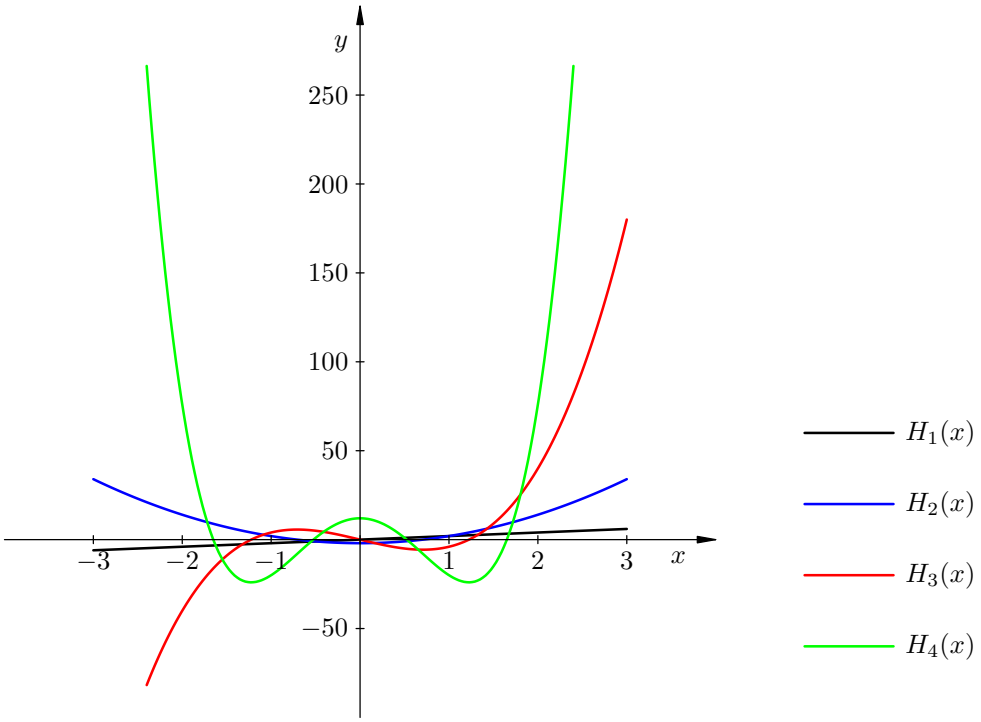
uz start

$$H_0(x) = 1, \quad H_1(x) = 2x.$$

Osim toga, n-ti Hermiteov polinom H_n zadovoljava diferencijalnu jednadžbu

$$y'' - 2xy' + 2ny = 0.$$

Graf prvih par polinoma izgleda ovako.



1.6. Trigonometrijske funkcije

Trigonometrijske funkcije

$$\{1, \cos x, \cos 2x, \cos 3x, \ldots, \sin x, \sin 2x, \sin 3x, \ldots\}$$

čine ortogonalnu familiju funkcija na intervalu $[0, 2\pi]$ uz mjeru

$$d\lambda = \begin{cases} dx & \text{na } [0, 2\pi], \\ 0 & \text{inače.} \end{cases}$$

Pokažimo da je to zaista istina. Neka su $k, \ell \in \mathbb{N}_0$. Tada vrijedi

$$\int_0^{2\pi} \sin kx \cdot \sin \ell x \, dx = -\frac{1}{2} \int_0^{2\pi} (\cos(k+\ell)x - \cos(k-\ell)x) \, dx.$$

U slučaju da je $k = \ell$, onda je prethodni integral jednak

$$-\frac{1}{2}\left[\frac{\sin(k+\ell)x}{k+\ell} - x\right]\Bigg|_0^{2\pi} = \pi.$$

Ako je $k \neq \ell$, onda je jednak

$$-\frac{1}{2}\left[\frac{\sin(k+\ell)x}{k+\ell} - \frac{\sin(k-\ell)x}{k-\ell}\right]\Bigg|_0^{2\pi} = 0.$$

Drugim riječima, vrijedi

$$\int_0^{2\pi} \sin kx \cdot \sin \ell x \, dx = \begin{cases} 0, & k \neq \ell, \\ \pi, & k = \ell, \end{cases} \qquad k, \ell = 1, 2, \ldots,$$

Na sličan način, pretvaranjem produkta trigonometrijskih funkcija u zbroj, možemo pokazati da je

$$\int_0^{2\pi} \cos kx \cdot \cos \ell x \, dx = \begin{cases} 0, & k \neq \ell, \\ 2\pi, & k = \ell = 0, \\ \pi, & k = \ell > 0, \end{cases} \qquad k, \ell = 0, 1, \ldots,$$

te, također, da je

$$\int_0^{2\pi} \sin kx \cdot \cos \ell x \, dx = 0, \quad k = 1, 2, \ldots, \quad \ell = 0, 1, \ldots,$$

Ako periodičku funkciju f osnovnog perioda duljine 2π želimo aproksimirati redom oblika

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx),$$

onda, množenjem odgovarajućim trigonometrijskim funkcijama i integriranjem, za koeficijente u redu formalno dobivamo

$$a_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \, \cos kx \, dx, \quad b_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \, \sin kx \, dx.$$

Prethodni red poznat je pod imenom Fourierov red, a koeficijenti kao Fourierovi koeficijenti.

Posebno, ako Fourierov red odsiječemo za $k = m$ i dobijemo trigonometrijski polinom, koji je najbolja L_2 aproksimacija za f u klasi trigonometrijskih polinoma stupnja manjeg ili jednakog m, obzirom na normu

$$\|u\|_2 = \left(\int_0^{2\pi} |u(t)|^2 \, dt \right)^{1/2}.$$

1.6.1. Diskretna ortogonalnost trigonometrijskih funkcija

Umjesto neprekidne, za pripadnu mjeru možemo uzeti i diskretnu mjeru, pa umjesto integrala, dobivamo sume. Ako pogodno izaberemo točke, opet dobivamo da su trigonometrijske funkcije međusobno ortogonalne, ali u diskretnom smislu.

Za dokaz pripadnih relacija diskretne ortogonalnosti trebamo sljedeću pomoćnu tvrdnju.

Lema 1.6.1. *Neka je $n \in \mathbb{N}$ zadani prirodni broj. Za bilo koji cijeli broj $m \in \mathbb{Z}$ promatramo trigonometrijske sume*

$$C_m := \sum_{j=0}^{n-1} \cos\left(j \, \frac{2m\pi}{n}\right), \quad S_m := \sum_{j=0}^{n-1} \sin\left(j \, \frac{2m\pi}{n}\right).$$

Ove sume, u principu, ovise o n, ali to nećemo posebno označiti, radi preglednosti. Za ove sume vrijede relacije

$$C_m = \begin{cases} n, & \text{za } m \bmod n = 0, \\ 0, & \text{za } m \bmod n \neq 0, \end{cases} \qquad S_m = 0, \quad \text{za svaki } m \in \mathbb{Z}.$$

Dakle, samo C_m stvarno ovisi o n.

Dokaz:

Bitno najlakši dokaz dobivamo prijelazom na kompleksne brojeve u trigonometrijskom (polarnom) obliku i korištenjem De Moivreove formule za potenciranje kompleksnih brojeva.

Za zadani $n \in \mathbb{N}$, neka je ω osnovni n-ti korijen iz jedinice

$$\omega := \cos\frac{2\pi}{n} + i \sin\frac{2\pi}{n} = e^{i2\pi/n}.$$

Nadalje, za bilo koji fiksni $m \in \mathbb{Z}$, definiramo kompleksni broj q relacijom

$$q := \omega^m = \cos\frac{2m\pi}{n} + i \sin\frac{2m\pi}{n} = e^{i2m\pi/n}.$$

Sad promatramo kompleksni broj $Z_m := C_m + iS_m$. Kad uvrstimo sume za C_m i S_m, dobivamo redom

$$\begin{aligned} Z_m &= \sum_{j=0}^{n-1} \cos\left(j\,\frac{2m\pi}{n}\right) + i\sum_{j=0}^{n-1} \sin\left(j\,\frac{2m\pi}{n}\right) \\ &= \sum_{j=0}^{n-1} \left(\cos\left(j\,\frac{2m\pi}{n}\right) + i\sin\left(j\,\frac{2m\pi}{n}\right)\right) \\ &= \sum_{j=0}^{n-1} e^{ij2m\pi/n} = \sum_{j=0}^{n-1} \left(e^{i2m\pi/n}\right)^j = \sum_{j=0}^{n-1} q^j. \end{aligned}$$

Ova geometrijska suma se lako računa, samo treba paziti je li $q = 1$ ili ne.

Ako je $m \bmod n = 0$, onda je $q = 1$, pa je

$$Z_m = \sum_{j=0}^{n-1} 1 = n.$$

Prijelazom na realni i imaginarni dio od Z_m slijedi

$$C_m = n, \quad S_m = 0, \quad \text{za } m \bmod n = 0.$$

Ako je $m \bmod n \neq 0$, onda je $q \neq 1$, pa koristimo poznatu formulu za geometrijsku sumu

$$Z_m = \sum_{j=0}^{n-1} q^j = \frac{q^n - 1}{q - 1} = \frac{e^{i2nm\pi/n} - 1}{e^{i2m\pi/n} - 1} = \frac{e^{i2m\pi} - 1}{e^{i2m\pi/n} - 1} = \frac{1 - 1}{e^{i2m\pi/n} - 1} = 0.$$

Prijelazom na realni i imaginarni dio od Z_m slijedi

$$C_m = 0, \quad S_m = 0, \quad \text{za } m \bmod n \neq 0.$$

Usput smo dokazali da za sumu Z_m vrijedi ista formula kao i za C_m. ■

Trigonometrijske funkcije zadovoljavaju sljedeće relacije diskretne ortogonalnosti, iz kojih onda dobivamo aproksimacije u smislu diskretne metode najmanjih kvadrata.

Teorem 1.6.1. *Za trigonometrijske funkcije, na mreži od $N+1$ točaka $0, 1, \ldots, N$, uz oznake*

$$x_j = j, \quad x_{k,j} = \frac{2\pi}{N+1} k x_j, \quad x_{\ell,j} = \frac{2\pi}{N+1} \ell x_j, \quad j = 0, \ldots, N,$$

vrijede sljedeće relacije ortogonalnosti

$$\sum_{j=0}^{N} \cos x_{k,j} \cos x_{\ell,j} = \begin{cases} 0, & k \neq \ell, \\ (N+1)/2, & k = \ell \neq 0, \\ N+1, & k = \ell = 0, \end{cases}$$

$$\sum_{j=0}^{N} \sin x_{k,j} \sin x_{\ell,j} = \begin{cases} 0, & k \neq \ell \ i \ k = \ell = 0, \\ (N+1)/2, & k = \ell \neq 0, \end{cases}$$

$$\sum_{j=0}^{N} \sin x_{k,j} \cos x_{\ell,j} = 0$$

uz uvjet da je $k + \ell \leq N$.

Dokaz:

Za dokaz koristimo standardne formule za pretvaranje produkta trigonometrijskih funkcija u zbroj ili razliku

$$\cos\alpha \cdot \cos\beta = \frac{1}{2}\Big(\cos(\alpha-\beta) + \cos(\alpha+\beta)\Big),$$
$$\sin\alpha \cdot \sin\beta = \frac{1}{2}\Big(\cos(\alpha-\beta) - \cos(\alpha+\beta)\Big),$$
$$\sin\alpha \cdot \cos\beta = \frac{1}{2}\Big(\sin(\alpha-\beta) + \sin(\alpha+\beta)\Big).$$

Stvaljanjem $\alpha = x_{k,j}$ i $\beta = x_{\ell,j}$, tražene sume iz tvrdnje dobivaju oblik

$$\begin{aligned}
\sum_{j=0}^{N} \cos x_{k,j} \cos x_{\ell,j} &= \sum_{j=0}^{N} \cos\left(\frac{2\pi}{N+1} k x_j\right) \cdot \cos\left(\frac{2\pi}{N+1} \ell x_j\right) \\
&= \frac{1}{2} \sum_{j=0}^{N} \left[\cos\left(j\,\frac{2(k-\ell)\pi}{N+1}\right) + \cos\left(j\,\frac{2(k+\ell)\pi}{N+1}\right)\right], \\
\sum_{j=0}^{N} \sin x_{k,j} \sin x_{\ell,j} &= \sum_{j=0}^{N} \sin\left(\frac{2\pi}{N+1} k x_j\right) \cdot \sin\left(\frac{2\pi}{N+1} \ell x_j\right) \\
&= \frac{1}{2} \sum_{j=0}^{N} \left[\cos\left(j\,\frac{2(k-\ell)\pi}{N+1}\right) - \cos\left(j\,\frac{2(k+\ell)\pi}{N+1}\right)\right], \\
\sum_{j=0}^{N} \sin x_{k,j} \cos x_{\ell,j} &= \sum_{j=0}^{N} \sin\left(\frac{2\pi}{N+1} k x_j\right) \cdot \cos\left(\frac{2\pi}{N+1} \ell x_j\right) \\
&= \frac{1}{2} \sum_{j=0}^{N} \left[\sin\left(j\,\frac{2(k-\ell)\pi}{N+1}\right) + \sin\left(j\,\frac{2(k+\ell)\pi}{N+1}\right)\right].
\end{aligned}$$

Ako označimo $n = N+1$, onda možemo iskoristiti oznake za sume kosinusa i sinusa

iz leme 1.6.1., pa dobivamo

$$\sum_{j=0}^{N} \cos x_{k,j} \cos x_{\ell,j} = \frac{1}{2}\Big(C_{k-\ell} + C_{k+\ell}\Big),$$
$$\sum_{j=0}^{N} \sin x_{k,j} \sin x_{\ell,j} = \frac{1}{2}\Big(C_{k-\ell} - C_{k+\ell}\Big),$$
$$\sum_{j=0}^{N} \sin x_{k,j} \cos x_{\ell,j} = \frac{1}{2}\Big(S_{k-\ell} + S_{k+\ell}\Big).$$

Prema tvrdnji leme 1.6.1., za sume vrijedi

$$C_m = \begin{cases} N+1, & \text{za } m \bmod (N+1) = 0, \\ 0, & \text{za } m \bmod (N+1) \neq 0, \end{cases} \qquad S_m = 0, \quad \text{za svaki } m \in \mathbb{Z}.$$

Iz druge relacije odmah slijedi zadnja relacija ortogonalnosti

$$\sum_{j=0}^{N} \sin x_{k,j} \cos x_{\ell,j} = 0,$$

za bilo koje k i ℓ.

Kod nas je $m = k - \ell$ ili $m = k + \ell$, s tim da je $k, \ell \geq 0$ i $k + \ell \leq N$. Onda mora vrijediti i $-N \leq k - \ell \leq N$, što pokazuje da je uvijek $-N \leq m \leq N$. Dakle, vidimo da je $m \bmod (N+1) = 0$ ako i samo ako je $m = 0$. To se događa ako i samo ako je $m = k - \ell$ i $k = \ell$, s tim da smije biti i $m = k + \ell = 0$, kad je $k = \ell = 0$.

Za $k \neq \ell$ je $C_{k-\ell} = C_{k+\ell} = 0$, pa je

$$\sum_{j=0}^{N} \cos x_{k,j} \cos x_{\ell,j} = \sum_{j=0}^{N} \sin x_{k,j} \sin x_{\ell,j} = 0.$$

Za $k = \ell > 0$ je $C_{k-\ell} = N + 1$ i $C_{k+\ell} = 0$, pa je

$$\sum_{j=0}^{N} \cos x_{k,j} \cos x_{k,j} = \sum_{j=0}^{N} \sin x_{k,j} \sin x_{k,j} = \frac{N+1}{2}.$$

Konačno, za $k = \ell = 0$ je $C_{k-\ell} = C_{k+\ell} = N + 1$, pa dobivamo

$$\sum_{j=0}^{N} \cos x_{k,j} \cos x_{k,j} = N + 1, \qquad \sum_{j=0}^{N} \sin x_{k,j} \sin x_{k,j} = 0.$$

Time su dokazane sve relacije ortogonalnosti iz tvrdnje. ■

Ovo znači da restrikcije funkcija

$$\cos \frac{2\pi}{N+1} kx, \quad \sin \frac{2\pi}{N+1} kx, \tag{1.6.1}$$

na mreži $\{0, \ldots, N\}$, možemo koristiti kao ortogonalnu familiju, pri čemu su dozvoljeni $k \in \mathbb{N}_0$ za kosinuse i $k \in N$ za sinuse. Linearne kombinacije funkcija (1.6.1) zvat ćemo **trigonometrijskim polinomima**.

Nažalost, baze takvih trigonometrijskih polinoma ovise o parnosti broja N.

Neparan broj točaka

Neka je zadan neparan broj točaka $\mathcal{M} = \{0, 1, \ldots, N = 2L\}$. Za bazu se tada uzima prvih $L + 1$ kosinusa (prvi je konstanta) i prvih L sinusa, a pripadna trigonometrijska aproksimacija ima oblik

$$T_N(x) = \frac{a_0}{2} + \sum_{k=1}^{L} (a_k \cos x_k + b_k \sin x_k), \tag{1.6.2}$$

pri čemu je

$$x_k = \frac{2\pi}{N+1} kx := \frac{2\pi}{2L+1} kx.$$

Koeficijenti trigonometrijskog polinoma određuju se iz relacija ortogonalnosti na uobičajeni način, množenjem lijeve i desne strane u (1.6.2) izabranom funkcijom baze uz koju je odgovarajući koeficijent. Ako trigonometrijski polinom T_N interpolira funkciju f u $x \in \mathcal{M}$, tj. ako je $T_n(x) = f(x)$ onda množenjem (1.6.2) s $\cos x_\ell$, $\ell \geq 0$, i upotrebom relacija ortogonalnosti dolazimo do koeficijenata a_ℓ

$$f(x) \cos x_\ell = \frac{a_0}{2} \cos x_\ell + \sum_{k=1}^{L} a_k \cos x_k \cos x_\ell + \sum_{k=1}^{L} b_k \sin x_k \cos x_\ell.$$

Zbrajanjem po svim x dobivamo

$$\begin{aligned} \sum_{x=0}^{2L} f(x) \cos x_\ell &= \frac{a_0}{2} \sum_{x=0}^{2L} \cos 0 \cos x_\ell + \sum_{k=1}^{L} a_k \sum_{x=0}^{2L} \cos x_k \cos x_\ell + \sum_{k=1}^{L} b_k \sum_{x=0}^{2L} \sin x_k \cos x_\ell \\ &= \frac{2L+1}{2} a_\ell. \end{aligned}$$

Odatle odmah zaključujemo da je (pišući k umjesto ℓ)

$$a_k = \frac{2}{2L+1} \sum_{x=0}^{2L} f(x) \cos x_k, \quad k = 0, \ldots, L.$$

Na sličan način, množenjem sa $\sin x_\ell$, $\ell > 0$, i zbrajanjem po svim x dobivamo

$$\begin{aligned} \sum_{x=0}^{2L} f(x) \sin x_\ell &= \frac{a_0}{2} \sum_{x=0}^{2L} \cos 0 \sin x_\ell + \sum_{k=1}^{L} a_k \sum_{x=0}^{2L} \cos x_k \sin x_\ell + \sum_{k=1}^{L} b_k \sum_{x=0}^{2L} \sin x_k \sin x_\ell \\ &= \frac{2L+1}{2} b_\ell. \end{aligned}$$

Slično kao kod a_k, imamo

$$b_k = \frac{2}{2L+1} \sum_{x=0}^{2L} f(x) \sin x_k, \quad k = 1, \ldots, L.$$

Dakle, u slučaju neparnog broja točaka koeficijenti u (1.6.2) su

$$a_k = \frac{2}{2L+1} \sum_{x=0}^{2L} f(x) \cos x_k, \quad k = 0, \ldots, L,$$
$$b_k = \frac{2}{2L+1} \sum_{x=0}^{2L} f(x) \sin x_k, \quad k = 1, \ldots, L.$$

Zadatak 1.6.1. *Pokažite da za bilo koju točku x^*, ne nužno iz $\mathcal{M}$ vrijedi*

$$T_N(x^*) = \frac{1}{2L+1} \sum_{x=0}^{2L} f(x) \Big(\sum_{k=0}^{2L} \cos \Big(\frac{2\pi}{2L+1} k(x - x^*) \Big) \Big).$$

Paran broj točaka

Neka je zadan paran broj točaka $\mathcal{M} = \{0, 1, \ldots, N = 2L-1\}$. Za bazu se tada uzima prvih $L+1$ kosinusa (prvi je konstanta) i prvih $L-1$ sinusa, a pripadna trigonometrijska aproksimacija ima oblik

$$T_N(x) = \frac{a_0}{2} + \sum_{k=1}^{L-1} (a_k \cos x_k + b_k \sin x_k) + \frac{1}{2} a_L \cos x_L, \tag{1.6.3}$$

pri čemu je

$$x_k = \frac{2\pi}{N+1} kx := \frac{\pi}{L} kx.$$

Na sličan način kao kod neparnog broja točaka, koeficijenti u (1.6.3) su

$$a_k = \frac{1}{L} \sum_{x=0}^{2L-1} f(x) \cos x_k, \quad k = 0, \ldots, L,$$
$$b_k = \frac{1}{2L} \sum_{x=0}^{2L-1} f(x) \sin x_k, \quad k = 1, \ldots, L-1.$$

Zadatak 1.6.2. *Pokažite da i u slučaju neparnog i u slučaju parnog broja točaka, T_N ima period $N+1$. Zbog toga se jednostavno koristi za interpolaciju trigonometrijskih funkcija, a dovoljno je zadati samo točke x iz jednog perioda.*

Primjer 1.6.1. *Funkcija f ima period 3 i zadana je tablično s*

x_k	0	1	2
f_k	0	1	1

Nađimo trigonometrijski polinom koji interpolira f u svim točkama iz $\mathbb{Z}$, a zatim izračunajmo $T_N(1/2)$ i $T_N(3/2)$.

Budući da je $N = 2$, broj točaka je neparan, pa je

$$T_2(x) = \frac{1}{2}a_0 + a_1 \cos\frac{2\pi}{3}x + b_1 \sin\frac{2\pi}{3}x.$$

Prema formulama za koeficijente, dobivamo

$$\begin{aligned} a_0 &= \frac{2}{3}\left(0\cos 0 + 1\cdot\cos 0 + 1\cdot\cos 0\right) = \frac{4}{3} \\ a_1 &= \frac{2}{3}\left(0\cos 0 + 1\cdot\cos\frac{2\pi}{3} + 1\cdot\cos\frac{4\pi}{3}\right) = -\frac{2}{3} \\ b_1 &= \frac{2}{3}\left(0\sin 0 + 1\cdot\sin\frac{2\pi}{3} + 1\cdot\sin\frac{4\pi}{3}\right) = 0. \end{aligned}$$

Prma tome, trigonometrijski polinom koji interpolira zadane točke je

$$T_2(x) = \frac{2}{3} - \frac{2}{3}\cos\frac{2\pi}{3}x.$$

Odatle se odmah može izračunati da je

$$\begin{aligned} T_2(1/2) &= \frac{2}{3} - \frac{2}{3}\cos\frac{\pi}{3} = \frac{1}{3} \\ T_2(3/2) &= \frac{2}{3} - \frac{2}{3}\cos\pi = \frac{4}{3}. \end{aligned}$$

Metoda majnanjih kvadrata za trigonometrijske funkcije

I za metodu najmanjih kvadrata možemo koristiti trigonometrijske polinome, jer je dovoljno uzeti podskup baze prostora. Slično kao kod interpolacije biramo početni dio baze (1.6.1). Također moramo paziti na parnost/neparnost broja točaka N i na parnost/neparnost stupnja trigonometrijskog polinoma M, $M \leq N$.

Ilustrirajmo to na slučaju $N = 2L$ paran (broj točaka neparan) i $M = 2m$ paran (neparna dimenzija potprostora). Trigonometrijski polinom odgovarajućeg stupnja je

$$T_M(x) = \frac{1}{2}A_0 = \sum_{k=1}^{m}(A_k \cos x_k + B_k \sin x_k), \tag{1.6.4}$$

gdje je

$$x_k = \frac{2\pi}{N+1}kx := \frac{2\pi}{2L+1}kx.$$

Metoda najmanjih kvadrata minimizira kvadrat greške

$$S = \sum_{x=0}^{2L} (f(x) - T_M(x))^2 \to \min .$$

Tvrdimo da je rješenje problema mininizacije trigonometrijski interpolacijski polinom kojemu je

$$A_k = a_k, \quad k = 0, \ldots, m$$
$$B_k = b_k, \quad k = 1, \ldots, m,$$

a koeficijenti a_k i b_k se računaju po formulama za interpolaciju. Primijetite da u točkama interpolacije x, $x = 0, \ldots, 2L$ interpolacijski polinom ima istu vrijednost kao funkcija f, pa je dovoljno (u točkama interpolacije) uspoređivati interpolacijski trigonometrijski polinom T_N, $N = 2L$ i trigonometrijski polinom T_M, $M = 2m$ dobiven metodom najmanjih kvadrata. Vrijedi

$$T_N(x) - T_M(x) = \frac{1}{2}(a_0 - A_0) + \sum_{k=1}^{m} ((a_k - A_k)\cos x_k + (b_k - B_k)\sin x_k) + \sum_{k=m+1}^{L} (a_k \cos x_k + b_k \sin x_k).$$

Dakle, u točkama interpolacije x vrijedi

$$f(x) - T_M(x) = T_N(x) - T_M(x).$$

Greška S koju minimiziramo dobiva se upotrebom relacija ortogonalnosti. Izlazi

$$\begin{aligned}
S &:= \sum_{x=0}^{2L} (T_N(x) - T_M(x))^2 \\
&= \sum_{x=0}^{2L} \frac{1}{4}(a_0 - A_0)^2 + \sum_{x=0}^{2L} \sum_{k=1}^{m} ((a_k - A_k)\cos x_k + (b_k - B_k)\sin x_k)^2 \\
&\quad + \sum_{x=0}^{2L} \sum_{k=m+1}^{L} (a_k \cos x_k + b_k \sin x_k)^2 \\
&= \frac{1}{4}(a_0 - A_0)^2 \cdot (2L+1) + \sum_{x=0}^{2L} \sum_{k=1}^{m} \Big[(a_k - A_k)^2 \cos^2 x_k \\
&\qquad + 2(a_k - A_k)(b_k - B_k)\cos x_k \sin x_k + (b_k - B_k)^2 \sin^2 x_k\Big] \\
&\quad + \sum_{x=0}^{2L} \sum_{k=m+1}^{L} (a_k^2 \cos^2 x_k + 2a_k b_k \cos x_k \sin x_k + b_k^2 \sin^2 x_k) \\
&= \frac{1}{4}(a_0 - A_0)^2 \cdot (2L+1) + \frac{2L+1}{2} \sum_{k=1}^{m} (a_k - A_k)^2 + (b_k - B_k)^2) \\
&\quad + \frac{2L+1}{2} \sum_{k=m+1}^{L} (a_k^2 + b_k^2).
\end{aligned}$$

Prema tome, odmah je vidljivo da je greška S minimalna ako je

$$A_k = a_k, \quad k = 0, \ldots, m$$
$$B_k = b_k, \quad k = 1, \ldots, m,$$

i njena minimalna vrijednost jednaka je

$$S_{\min} = \frac{2L+1}{2} \sum_{k=m+1}^{L} (a_k^2 + b_k^2).$$

Ovaj oblik minimalne greške nije praktičan, jer uobičajeno ne znamo a_k, b_k za $k > m$.

Zadatak 1.6.3. *Dokažite da vrijedi*

$$S_{\min} = \sum_{x=0}^{2L} (f(x))^2 - \frac{2L+1}{4} a_0^2 - \frac{2L+1}{2} \sum_{k=1}^{m} (a_k^2 + b_k^2)$$

korištenjem relacija ortogonalnosti. Prethodni oblik greške često se koristi za detekciju stupnja trigonometrijskog polinoma, jer nagli pad greške pri dizanju stupnja trigonometrijskog polinoma znači da smo otkrili stupanj polinoma. Greška pritom ne mora biti 0*, jer je pojava mogla imati slučajne greške koje smo ionako željeli maknuti.*

Zadatak 1.6.4. *Izvedite metodu najmanjih kvadrata za tri preostala slučaja:*

1. *broj točaka paran $N = 2L - 1$, dimenzija prostora neparna $M = 2m$,*
2. *broj točaka neparan $N = 2L$, dimenzija prostora parna $M = 2m - 1$,*
3. *broj točaka paran $N = 2L - 1$, dimenzija prostora parna $M = 2m - 1$.*

Zadatak 1.6.5. *Neka je funkcija f zadana na mreži točaka $\mathcal{M} = \{0, 1, \ldots, P-1\}$, P neparan (tj. točaka je paran broj) i neka je P period funkcije f, tj.*

$$f(x + P) = f(x).$$

Pokažite da su tada

$$a_k = \frac{2}{P} \sum_{x=-L+1}^{L} f(x) \cos \frac{2\pi}{P} kx, \quad k = 0, \ldots, L$$
$$b_k = \frac{2}{P} \sum_{x=-L+1}^{L} f(x) \sin \frac{2\pi}{P} kx, \quad k = 1, \ldots, L-1.$$

Ako je f neparna funkcija $f(-x) = -f(x)$, pokažite da je

$$a_k = 0, \quad k = 0, \ldots, L$$
$$b_k = \frac{4}{P} \sum_{x=1}^{L-1} f(x) \sin \frac{2\pi}{P} kx, \quad k = 1, \ldots L-1.$$

Ako je f parna funkcija $f(-x) = f(x)$, pokažite da je

$$a_k = \frac{2}{P}(f(0) + f(L)\cos k\pi) + \frac{4}{P}\sum_{x=1}^{L-1} f(x)\cos\frac{2\pi}{P}kx, \quad k = 0, \ldots, L$$
$$b_k = 0, \quad k = 1, \ldots L-1.$$

Zadatak 1.6.6. *Riješite prethodni zadatak uz uvjet da je $P = 2L+1$, tj. da funkcija ima neparan period.*

1.7. Diskretne ortogonalnosti polinoma T_n

Budući da su Čebiševljevi polinomi T_n zapravo kosinusi, onda oni zadovoljavaju relacije diskretne ortogonalnosti vrlo slične onima koje zadovoljavaju trigonometrijske funkcije.

Neka su x_j sve različite nultočke Čebiševljevog polinoma T_{N+1}, tj. neka je

$$T_{N+1}(x_j) = \cos(N+1)\vartheta_j = 0.$$

Nije teško izračunati da je tada

$$x_j = \cos\vartheta_j, \quad \vartheta_j = \frac{(2j+1)\pi}{2(N+1)}, \quad j = 0, \ldots, N.$$

Teorem 1.7.1. *Za Čebiševljeve polinome, u nultočkama polinoma T_{N+1} vrijede sljedeće relacije ortogonalnosti*

$$U_{k,\ell} = \sum_{j=0}^{N} T_k(x_j)T_\ell(x_j) = \sum_{j=0}^{N} \cos(k\vartheta_j)\cos(\ell\vartheta_j),$$

gdje je

$$U_{k,\ell} = \begin{cases} 0 & k \neq \ell, \text{ uz } k, \ell \leq N, \\ (N+1)/2 & k = \ell, \text{ uz } 0 < k \leq N, \\ N+1 & k = \ell = 0. \end{cases}$$

Dokaz:

Za dokaz ovih relacija koristimo formulu za pretvaranje produkta dva kosinusa u zbroj trigonometrijskih funkcija. Vrijedi

$$\cos\alpha \cdot \cos\beta = \frac{1}{2}\Big(\cos(\alpha-\beta) + \cos(\alpha+\beta)\Big).$$

Onda je

$$\begin{aligned} U_{k,\ell} &= \sum_{j=0}^{N} \cos(k\vartheta_j)\cos(\ell\vartheta_j) = \frac{1}{2}\sum_{j=0}^{N}\Big(\cos(k-\ell)\vartheta_j + \cos(k+\ell)\vartheta_j\Big) \\ &= \frac{1}{2}\sum_{j=0}^{N}\Big[\cos\Big((k-\ell)\,\frac{(2j+1)\pi}{2(N+1)}\Big) + \cos\Big((k+\ell)\,\frac{(2j+1)\pi}{2(N+1)}\Big)\Big] \\ &= \frac{1}{2}\Big[\sum_{j=0}^{N}\cos\Big((k-\ell)\,\frac{(2j+1)\pi}{2(N+1)}\Big) + \sum_{j=0}^{N}\cos\Big((k+\ell)\,\frac{(2j+1)\pi}{2(N+1)}\Big)\Big]. \end{aligned}$$

Iz pretpostavke $0 \le k, \ell \le N$ slijedi

$$-N \le k-\ell \le N, \quad 0 \le k+\ell \le 2N.$$

Dakle, za nastavak dokaza trebamo izračunati vrijednost sume

$$C'_m = \sum_{j=0}^{N}\cos\Big(m\,\frac{(2j+1)\pi}{2(N+1)}\Big) = \sum_{j=0}^{N}\cos\Big((2j+1)\frac{m\pi}{2(N+1)}\Big),$$

za razne vrijednosti m, s tim da možemo uzeti $-N \le m \le 2N$.

Ova suma C'_m je slična sumi C_m iz leme 1.6.1., ali se ne može svesti na nju. Za $m = 1$, ranija suma C_1 obilazi sve višekratnike od $2\pi/n$ po cijeloj jediničnoj kružnici. Za razliku od toga, uz $n = N+1$, ova suma C'_1 obilazi samo neparne višekratnike od $\pi/(2n)$ i to po "gornjoj" polovini jedinične kružnice (kutevi manji od π). Svejedno, dokaz se provodi na sličan način kao i ranije.

Prvo uočimo da je

$$C'_m = \sum_{j=0}^{N}\cos\Big((2j+1)\frac{m\pi}{2(N+1)}\Big) = \operatorname{Re}\sum_{j=0}^{N}\exp\Big(i(2j+1)\frac{m\pi}{2(N+1)}\Big).$$

Za bilo koji fiksni m, definiramo kompleksni broj q relacijom

$$q := \cos\frac{m\pi}{2(N+1)} + i\sin\frac{m\pi}{2(N+1)} = e^{im\pi/(2(N+1))}.$$

Sad promatramo kompleksni broj $Z'_m := C'_m + iS'_m$, gdje je S'_m zbroj odgovarajućih sinusa

$$S'_m = \sum_{j=0}^{N}\sin\Big((2j+1)\frac{m\pi}{2(N+1)}\Big) = \operatorname{Im}\sum_{j=0}^{N}\exp\Big(i(2j+1)\frac{m\pi}{2(N+1)}\Big).$$

Onda imamo redom

$$Z'_m = \sum_{j=0}^{N}\exp\Big(i(2j+1)\frac{m\pi}{2(N+1)}\Big) = \sum_{j=0}^{N}\Big(e^{im\pi/(2(N+1))}\Big)^{2j+1} = \sum_{j=0}^{N}q^{2j+1} = q\sum_{j=0}^{N}q^{2j}.$$

Ovo je geometrijska suma s faktorom q^2, pa treba paziti je li $q^2 = 1$ ili ne.

Zbog $q^2 = e^{i2m\pi/(2(N+1))}$, vidimo da je $q^2 = 1$ ako i samo ako je m cjelobrojni višekratnik od $2(N+1)$. No, iz ograničenja $-N \le m \le 2N$ odmah slijedi da je to moguće ako i samo ako je $m = 0$.

Dakle, ako je $m = 0$, onda je i $q = 1$, pa je

$$Z'_0 = 1 \cdot \sum_{j=0}^{N} 1 = N + 1,$$

odakle slijedi $C'_0 = N + 1$ i $S'_0 = 0$.

Ako je $-N \le m \le 2N$ i $m \neq 0$, onda je $q^2 \neq 1$, pa dobivamo

$$Z'_m = q \sum_{j=0}^{N} q^{2j} = q\,\frac{q^{2(N+1)} - 1}{q^2 - 1}.$$

Odmah vidimo da je

$$q^{2(N+1)} = e^{i2(N+1)m\pi/(2(N+1))} = e^{im\pi} = \begin{cases} 1, & \text{za } m \text{ paran,} \\ -1, & \text{za } m \text{ neparan.} \end{cases}$$

Ako je m paran (i $m \neq 0$), onda izlazi $Z'_m = 0$, pa vrijedi i $C'_m = 0$, $S'_m = 0$.

Za neparne m dobivamo $Z'_m = -2q/(q^2 - 1)$. Za nastavak dokaza pretvaramo nazivnik u realan broj, tako da brojnik i nazivnik pomnožimo konjugiranim nazivnikom $\bar{q}^2 - 1$. Osim toga, koristimo i činjenicu da je $|q| = 1$.

$$\begin{aligned} Z'_m &= -2\,\frac{q}{q^2 - 1} = -2\,\frac{q}{q^2 - 1} \cdot \frac{\bar{q}^2 - 1}{\bar{q}^2 - 1} = -2\,\frac{q(\bar{q}^2 - 1)}{|\bar{q}^2 - 1|^2} \\ &= -2\,\frac{q \cdot \bar{q} \cdot \bar{q} - q}{|\bar{q}^2 - 1|^2} = -2\,\frac{|q|^2\bar{q} - q}{|\bar{q}^2 - 1|^2} = -2\,\frac{\bar{q} - q}{|\bar{q}^2 - 1|^2}. \end{aligned}$$

Nazivnik je realan, a u brojniku ostaje samo imaginarni dio, jer se realni dijelovi u $\bar{q} - q$ skrate. Dakle, mora biti $\operatorname{Re} Z'_m = C_m = 0$. Usput, ovdje je $S'_m \neq 0$, ali nas ta suma ne zanima.

Time smo dokazali da za $-N \le m \le 2N$ vrijedi

$$C'_m = \begin{cases} N + 1, & \text{za } m = 0, \\ 0, & \text{za } m \neq 0. \end{cases}$$

Sad se vratimo u polazne relacije za $U_{k,\ell}$.

$$U_{k,\ell} = \frac{1}{2}\,(C'_{k-\ell} + C'_{k+\ell}).$$

Za $k = \ell = 0$ dobivamo

$$U_{0,0} = \frac{1}{2}\Big((N+1) + (N+1)\Big) = N+1.$$

Za $k = \ell \neq 0$, zbog $k + \ell \neq 0$, dobivamo

$$U_{k,k} = \frac{1}{2}\Big((N+1) + 0\Big) = \frac{N+1}{2}.$$

Konačno, za $k \neq \ell$ imamo $k - \ell \neq 0$ i $k + \ell \neq 0$, pa odmah dobivamo $U_{k,\ell} = 0$. ■

Sada možemo funkciju razviti po Čebiševljevim polinomima koristeći prethodnu relaciju diskretne ortogonalnosti. Te relacije su namjerno zapisane u nultočkama polinoma T_{N+1}, što je zgodno za paralelu s trigonometrijskim funkcijama.

Međutim, za formulaciju rezultata o koeficijentima pripadnog razvoja, zgodnije je promatrati nultočke polinoma T_n, tj. uzeti da je $N + 1 = n$. Tada je

$$x_j = \cos(\vartheta_j), \quad \vartheta_j = \frac{(2j+1)\pi}{2n}, \quad j = 0, \ldots, n-1.$$

Relacije diskretne ortogonalnosti tada glase

$$U_{k,\ell} = \sum_{j=0}^{n-1} T_k(x_j) T_\ell(x_j) = \sum_{j=0}^{n-1} \cos(k\vartheta_j)\cos(\ell\vartheta_j),$$

gdje je

$$U_{k,\ell} = \begin{cases} 0 & k \neq \ell, \text{ uz } k, \ell < n, \\ n/2 & k = \ell, \text{ uz } 0 < k < n, \\ n & k = \ell = 0. \end{cases}$$

Uz te oznake, može se pokazati da vrijedi sljedeći teorem.

Teorem 1.7.2. *Neka je $f_n(x)$ aproksimacija za $f(x)$,*

$$f_n(\cos\vartheta) = \frac{d_0}{2} + \sum_{k=1}^{n-1} d_k \cos k\vartheta,$$

ili

$$f_n(x) = \frac{d_0}{2} + \sum_{k=1}^{n-1} d_k T_k(x). \tag{1.7.1}$$

Tada je

$$d_k = \frac{2}{n}\sum_{j=0}^{n-1} f(\cos\vartheta_j)\cos k\vartheta_j = \frac{2}{n}\sum_{j=0}^{n-1} f(x_j)T_k(x_j).$$

Pretpostavimo da je f' neprekidna na $[-1,1]$, osim najviše u konačno mnogo točaka, gdje ima ograničene skokove. Tada se f može razviti u konvergentan red oblika

$$f(x) = \frac{c_0}{2} + \sum_{k=1}^{\infty} c_k T_k(x), \tag{1.7.2}$$

gdje je

$$c_k = \frac{2}{\pi}\int_{-1}^{1} \frac{f(x)T_k(x)}{\sqrt{1-x^2}}\,dx = \frac{2}{\pi}\int_{0}^{\pi} f(\cos\vartheta)\cos k\vartheta\,d\vartheta.$$

Osim toga, postoji veza između koeficijenata u diskretnom i kontinuiranom razvoju:

$$\begin{aligned} d_0 &= c_0 + 2\sum_{r=1}^{\infty}(-1)^r c_{2rn} \\ d_k &= c_k + \sum_{r=1}^{\infty}(-1)^r c_{2rn-k} + \sum_{r=1}^{\infty}(-1)^r c_{2rn+k}, \quad k = 1,\ldots,n-1. \end{aligned}$$

Sljedeći teorem govori o greškama koje smo napravili aproksimacijom f_n obzirom na f.

Teorem 1.7.3. *Neka je*

$$\epsilon_n(x) = f(x) - f_n(x),$$

pri čemu su f_n i f zadani s (1.7.1) i (1.7.2). Za grešku ϵ_n tada vrijedi

$$\begin{aligned} \epsilon_n(\cos\vartheta) = \cos n\vartheta&\Big(c_n + 2\sum_{r=1}^{2n-1} c_{n+r}\cos r\vartheta + c_{3n}\cos 2n\vartheta\Big) \\ &- \sin 2n\vartheta\Big(c_{3n}\sin n\vartheta + 2\sum_{r=1}^{2n-1} c_{3n+r}\sin(n+r)\vartheta + c_{5n}\sin 3n\vartheta\Big) \\ &+ \cos 3n\vartheta\Big(c_{5n}\cos 2n\vartheta + 2\sum_{r=1}^{2n-1} c_{5n+r}\cos(2n+r)\vartheta + c_{7n}\cos 4n\vartheta\Big) - \cdots, \end{aligned}$$

odnosno, približno

$$\epsilon_n(\cos\vartheta) \approx c_n\cos n\vartheta\Big(1 + \frac{2c_{n+1}}{c_n}\cos\vartheta\Big).$$

Posebno, vrijedi

$$\epsilon_n(\cos\vartheta_j) = 0.$$

Iz prethodnih teorema uočavamo da x_j leže u unutrašnjosti intervala. U mnogim je primjenama je korisno dozvoliti aproksimaciju i u rubnim točkama ± 1 i točkama koje leže u sredini među ϑ_j. Primijetite da su to ekstremi odgovarajućeg Čebiševljevog polinoma. Sada možemo napraviti sličan niz tvrdnji kao za diskretnu ortogonalnost u nultočkama.

Neka su x_j sve različite točke ekstrema Čebiševljevog polinoma T_n na $[-1,1]$, tj. neka je

$$x_j = \cos\psi_j, \quad \psi_j = \frac{j\pi}{n}, \quad j = 0,\ldots,n.$$

Teorem 1.7.4. *Za Čebiševljeve polinome, u ekstremima polinoma T_n vrijede sljedeće relacije ortogonalnosti*

$$\begin{aligned} V_{k,\ell} &= \frac{1}{2}(T_k(x_0)T_\ell(x_0) + T_k(x_n)T_\ell(x_n)) + \sum_{j=1}^{n-1} T_k(x_j)T_\ell(x_j) \\ &= \frac{1}{2}(\cos(k\psi_0)\cos(\ell\psi_0) + \cos(k\psi_n)\cos(\ell\psi_n)) + \sum_{j=1}^{n-1} \cos(k\psi_j)\cos(\ell\psi_j), \end{aligned}$$

gdje je

$$V_{k,\ell} = \begin{cases} 0 & k \neq \ell, \text{ uz } k, \ell < n, \\ n/2 & k = \ell, \text{ uz } 0 < k < n, \\ n & k = \ell = 0 \text{ ili } k = \ell = n. \end{cases}$$

Sada možemo funkciju razviti po Čebiševljevim polinomima koristeći prethodnu relaciju diskretne ortogonalnosti. Može se pokazati da vrijedi sljedeći teorem.

Teorem 1.7.5. *Neka je $f_n(x)$ aproksimacija za $f(x)$,*

$$f_n(\cos\psi) = \frac{e_0}{2} + \sum_{k=1}^{n-1} e_k \cos k\psi + \frac{e_n}{2}\cos n\psi,$$

ili

$$f_n(x) = \frac{e_0}{2} + \sum_{k=1}^{n-1} e_k T_k(x) + \frac{e_n}{2} T_n(x). \tag{1.7.3}$$

Tada je

$$\begin{aligned} e_k &= \frac{2}{n}\Big(\frac{f(1) + (-1)^k f(-1)}{2} + \sum_{j=1}^{n-1} f(\cos\psi_j)\cos k\psi_j\Big) \\ &= \frac{2}{n}\Big(\frac{f(1) + (-1)^k f(-1)}{2} + \sum_{j=1}^{n-1} f(x_j)T_k(x_j)\Big). \end{aligned}$$

Osim toga, postoji veza između koeficijenata u diskretnom i kontinuiranom razvoju:

$$\begin{aligned} e_0 &= c_0 + 2\sum_{r=1}^{\infty} c_{2rn} \\ e_n &= 2c_n + 2\sum_{r=1}^{\infty} c_{(2r+1)n} \\ e_k &= c_k + \sum_{r=1}^{\infty} c_{2rn-k} + \sum_{r=1}^{\infty} c_{2rn+k}, \quad k = 1, \ldots, n-1. \end{aligned}$$

Sljedeći teorem govori o greškama koje smo napravili aproksimacijom f_n obzirom na f.

Teorem 1.7.6. *Neka je*

$$\delta_n(x) = f(x) - f_n(x),$$

pri čemu su f_n i f zadani s (1.7.3) i (1.7.2). Za grešku δ_n tada vrijedi

$$\delta_n(\cos\psi) = -2\sin n\psi \sum_{r=1}^{\infty} c_{n+r}\sin r\psi$$

odnosno, približno

$$\delta_n(\cos\psi) \approx -2\sin n\psi \sin\psi c_{n+1}\left(1 + \frac{2c_{n+2}}{c_{n+1}}\cos\psi\right).$$

Posebno, vrijedi

$$\delta_n(\cos\psi_j) = 0.$$

Ako se c_k iz razvoja f integrira po trapeznoj formuli (vidjeti kasnije), onda se takvom aproksimacijom dobivaju koeficijenti e_k. I d_k su koeficijenti koji se dobivaju približnom integracijom c_k (modificiranom trapeznom formulom, odnosno, tzv. "midpoint" pravilom).

2. Izvrednjavanje funkcija

Jedan od osnovnih zadataka koji se javlja u numeričkoj matematici je izračunavanje vrijednosti funkcije u nekoj točki ili na nekom skupu točaka (tzv. izvrednjavanje funkcije). Zašto baš to?

Efikasno računanje možemo raditi samo s onim vrstama funkcija za koje imamo dobar algoritam za izvrednjavanje. Pri tome moramo voditi računa o tome da aritmetika računala stvarno podržava samo četiri osnovne aritmetičke operacije, pa samo njih možemo koristiti u algoritmima. Osim toga, i tada računanje nije egzaktno, već u svakoj operaciji imamo greške zaokruživanja. Zbog toga, pri konstrukciji algoritama imamo dva cilja. Dobar algoritam za izvrednjavanje mora (kao, uostalom, i svaki numerički algoritam) zadovoljavati dva uvjeta:

- efikasnost ili brzina, tj. imati što manji broj aritmetičkih operacija;
- točnost, u smislu stabilnosti ili osjetljivosti na greške zaokruživanja.

Oba zahtjeva su posebno bitna baš kod izvrednjavanja, jer se ono obično puno puta koristi, pa i mala lokalna ubrzanja daju velike ukupne uštede u vremenu, a isto vrijedi i za ukupni efekt grešaka zaokruživanja. Općenito, očekujemo da brži algoritam ima i manju grešku, jer imamo manje operacija koje unose grešku u račun. Međutim, ovo **ne mora** biti istina! U mnogim slučajevima možemo drastično popraviti stabilnost algoritma tako da žrtvujemo dio efikasnosti, a katkad čak i bez toga, uz pametnu reformulaciju algoritma.

U ovom poglavlju ćemo više pažnje posvetiti efikasnosti, a manje stabilnosti, osim tamo gdje je nestabilnost opasna. Cilj nam je konstruirati efikasne algoritme i opravdati njihovu efikasnost, a ne analizirati ili dokazivati njihovu stabilnost. U skladu s tim, potencijalne nestabilnosti izlažemo opisno i ilustriramo na primjerima, bez strogih dokaza.

Pretpostavimo da je zadana funkcija $f : D \to \mathbb{R}$, gdje je $D \subseteq \mathbb{R}$ neka domena. Naš zadatak je izračunati vrijednosti te funkcije f u zadanoj točki $x_0 \in D$. Preciznije, moramo sastaviti algoritam koji računa $f(x_0)$. Naravno, točka x_0 može biti bilo koja i naš algoritam mora raditi za sve ulaze $x_0 \in D$.

Trenutno zanemarimo pitanje kako se **zadaje** funkcija f. Naime, ako je f ulaz u algoritam, onda f mora biti zadana s najviše konačno mnogo podataka o f, i ti

podaci moraju jednoznačno odrediti f. Ovo je, očito, fundamentalno ograničenje i bitno smanjuje klasu funkcija koje uopće možemo algoritamski izračunati. Odgovore na takva pitanja daje tzv. teorija izračunljivosti u okviru matematičke logike i osnova matematike.

U praksi odmah dobivamo i bitno jača ograničenja. Naime, ako imamo na raspolaganju samo 4 osnovne aritmetičke operacije, onda su **racionalne** funkcije jedine funkcije f kojima možemo izračunati vrijednost u bilo kojoj točki $x_0 \in D$. A takve funkcije možemo jednoznačno zadati konačnim brojem parametara — na primjer, ponašanjem (vrijednostima) u konačnom broju točaka (vidjeti poglavlje o interpolaciji) ili koeficijentima u nekom prikazu.

Dakle, sigurno trebamo efikasne algoritme za računanje vrijednosti racionalnih funkcija. Ako se sjetimo da racionalnu funkciju možemo napisati kao kvocijent dva polinoma, onda je zgodno imati i algoritme za izvrednjavanje polinoma. Osim toga, polinomi su još jednostavnije funkcije, jer nema dijeljenja, definirane su na cijeloj domeni (nema problema s nultočkama nazivnika), a koriste se "svagdje".

I da završimo ovo filozofiranje o izvrednjavanju funkcija. Strogo govoreći, sve ostale funkcije moramo **aproksimirati** na neki način — ako ni zbog čega drugog, onda zato jer naša aritmetika nije dovoljno jaka da bismo izračunali njihovu vrijednost u točki. To nipošto nije jedini razlog za aproksimacije funkcija, ali i on pokazuje zašto je aproksimacija centralni problem numeričke analize. Uostalom, u nastavku se gotovo isključivo bavimo raznim metodama za nalaženje različitih vrsta aproksimacija (i to, uglavnom, funkcija)!

Međutim, u praksi možemo pretpostaviti da za neke osnovne matematičke funkcije f **već imamo** dobre aproksimacije za približno računanje vrijednosti $f(x_0)$ u zadanoj točki x_0:

- procesor računala ("hardware") ima ugrađene aproksimacije i odgovarajuće instrukcije za njihov poziv (izvršavanje), ili
- koristimo neki gotovi (pot)program ("software") koji to radi.

Standardno, to su: $\sqrt{x}$ (a katkad i opće potenciranje x^α), trigonometrijske, eksponencijalne, hiperboličke i njima inverzne funkcije. Kako se nalaze takve aproksimacije za razne funkcije f za "hardware" ili "software" implementaciju — o tome kasnije, kod aproksimacija (jasno je da su polinomne ili racionalne)!

Bez obzira na realizaciju, u oba slučaja, bitno je samo to da ih možemo direktno koristiti u našim algoritmima i da znamo da izračunata vrijednost tražene funkcije f u zadanoj točki x_0 ima malu relativnu grešku u odnosu na točnost računanja. Dakle, možemo pretpostaviti da za $f\ell(f(x_0))$ vrijede iste ili slične ocjene kao i za osnovne aritmetičke operacije (vidjeti (1–2.6.2), ali i primjer 1–4.1.1.).

Reklo bi se: "gotova priča", dalje se ne trebamo brinuti oko toga. Međutim, nije baš tako. Računanje $f(x_0)$, u principu, traje **dulje**, a katkad i puno dulje, nego

što je trajanje jedne osnovne aritmetičke operacije (čak i ako sve četiri operacije nemaju isto ili podjednako trajanje). Za osnovne funkcije, taj omjer može biti 10 pa i više puta. Zato ponekad izbjegavamo puno poziva takvih funkcija, pogotovo ako se to može razumno izbjeći, tj. efikasno i bez većeg gubitka točnosti.

U mnogim slučajevima se to može napraviti i u ovom poglavlju ćemo pokazati neke opće algoritme tog tipa. Ideja je da se iskoriste neke rekurzivne relacije koje zadovoljavaju takve i slične, a za aproksimaciju važne funkcije. Na primjer, u teoriji se obično koriste ortogonalni sustavi funkcija za aproksimaciju, a funkcije takvog sustava zadovoljavaju tročlanu rekurziju, što je ključno za efikasno računanje.

2.1. Hornerova shema

Polinomi su najjednostavnije algebarske funkcije. Možemo ih definirati nad bilo kojim prstenom R u obliku

$$p(x) = \sum_{i=0}^{n} a_i x^i, \quad n \in \mathbb{N}_0,$$

gdje su $a_i \in R$ koeficijenti iz tog prstena, a x je simbolička "varijabla". Polinomi, kao simbolički objekti, također, imaju algebarsku strukturu prstena.

Međutim, polinome možemo interpretirati i kao funkcije, koje možemo izvrednjavati u svim točkama x_0 iz tog prstena R, uvrštavanjem x_0 umjesto simboličke varijable x. Dobiveni rezultat $p(x_0)$ je opet u R. Zanimaju nas efikasni algoritmi za računanje te vrijednosti.

Složenost očito ovisi o broju članova u sumi. Da broj članova ne bi bio umjetno prevelik, standardno uzimamo da je $p \neq 0$ i da je vodeći koeficijent $a_n \neq 0$, tako da je n stupanj tog polinoma p. Kada želimo naglasiti stupanj, polinom označavamo s p_n.

Algoritmi koje ćemo napraviti u principu rade nad bilo kojim prstenom R, ali neki rezultati o njihovoj složenosti vrijede samo za beskonačna neprebrojiva polja, poput $\mathbb{R}$ i $\mathbb{C}$, što su ionako najvažniji primjeri u praksi. Zbog toga, u nastavku možemo uzeti da radimo isključivo s polinomima nad $\mathbb{R}$ ili $\mathbb{C}$.

Kako zadajemo polinom? Pretpostavljamo da je polinom zadan stupnjem n i koeficijentima $a_0, \ldots, a_n$ u nekoj bazi vektorskog prostora polinoma stpunja ne većeg od n. Na početku koristimo standardnu bazu $1, x, x^2, \ldots, x^n$, a kasnije ćemo modificirati algoritme i za neke druge baze.

Složenost, naravno, mjerimo brojem osnovnih aritmetičkih operacija. Kad radimo nad $\mathbb{R}$, stvar je čista, jer aritmetika računala modelira upravo te operacije, pa je brojanje korektno. No, kad radimo nad $\mathbb{C}$, treba voditi računa o tome da

se kompleksne aritmetičke operacije realiziraju putem realnih, što znači da tek broj realnih operacija daje pravu mjeru složenosti. Baš na tu temu, u nekim kompleksnim algoritmima možemo ostvariti značajne uštede, pažljivim promatranjem realnih operacija.

2.1.1. Računanje vrijednosti polinoma u točki

Zadan je polinom stupnja n

$$p_n(x) = \sum_{i=0}^{n} a_i x^i, \quad a_n \neq 0$$

kojemu treba izračunati vrijednost u točki x_0. To se može napraviti na više načina. Prvo, napravimo to direktno po zapisu, potencirajući. Krenemo li od nulte potencije $x^0 = 1$, svaka sljedeća potencija dobiva se rekurzivno kao

$$x^k = x \cdot x^{k-1}.$$

Imamo li zapamćen x^{k-1}, lako je izračunati x^k korištenjem samo jednog množenja.

Algoritam 2.1.1. (Vrijednost polinoma s pamćenjem potencija)

$sum := a_0$;
$pot := 1$;
for $i := 1$ **to** n **do**
  begin
  $pot := pot * x_0$;
  $sum := sum + a_i * pot$;
  end;
{ Na kraju je $p_n(x_0) = sum$. }

Prebrojimo zbrajanja i množenja koja se javljaju u tom algoritmu. U unutarnjoj petlji javljaju se 2 množenja i 1 zbrajanje. Budući da se petlja izvršava n puta, ukupno imamo

$$2n \text{ množenja } + n \text{ zbrajanja.}$$

Naravno, kad smo nad $\mathbb{C}$, ove operacije su kompleksne.

Izvrednjavanje polinoma u točki može se izvesti i s manje množenja. Ako polinom p_n zapišemo u obliku

$$p_n(x) = (\cdots((a_n x + a_{n-1})x + a_{n-2})x + \cdots + a_1)x + a_0.$$

Algoritam koji po prethodnoj relaciji izvrednjava polinom zove se Hornerova shema. Predložio ga je W. G. Horner, 1819. godine, ali sličan zapis je koristio i Isaac Newton, još 1669. godine.

Algoritam 2.1.2. (Hornerova shema)

$sum := a_n$;
for $i := n - 1$ **downto** 0 **do**
  $sum := sum * x_0 + a_i$;
{ Na kraju je $p_n(x_0) = sum.$ }

Odmah je očito da smo korištenjem ovog algoritma broj množenja prepolovili, tj. da je njegova složenost

$$n \text{ množenja } + n \text{ zbrajanja.}$$

2.1.2. Hornerova shema je optimalan algoritam

Bitno je napomenuti da se Hornerovom shemom izvrednjavaju opći polinomi za koje znamo da imaju većinu nenula koeficijenata. Na primjer, polinom

$$p_{100}(x) = x^{100} + 1$$

besmisleno je izvrednjavati Hornerovom shemom, jer bi to predugo trajalo (binarno potenciranje je brže). Ili, kad izvrednjavamo polinom koji ima samo parne koeficijente

$$p_{2n}(x) = \sum_{i=0}^{n} a_{2i}\, x^{2i},$$

treba modificirati Hornerovu shemu tako da koristi samo parne potencije. Isto vrijedi i za polinom koji ima samo neparne potencije. Sastavite pripadne algoritme.

Za Hornerovu shemu može se pokazati da je optimalan algoritam.

Teorem 2.1.1. (Borodin, Munro) *Za opći polinom n-tog stupnja potrebno je barem n aktivnih množenja. Pod aktivnim množenjem podrazumjevamo množenje između a_i i x.*

Dakle, Hornerova shema ima optimalan broj množenja.

Rezultat prethodnog teorema može se poboljšati samo ako jedan te isti polinom izvrednjavamo u mnogo točaka. Tada se koeficijenti polinoma prije samog izvrednjavanja **adaptiraju** ili **prekondicioniraju**, tako da bismo kasnije imali što manje operacija po svakoj pojedinoj točki.

Zanimljivo je da u slučaju polinoma stupnjeva $n = 1$, 2 i 3, Hornerova shema je optimalna čak i kad računamo vrijednost polinoma u više točaka. Pokažimo jedan primjer adaptiranja koeficijenata za polinom stupnja 4.

Primjer 2.1.1. *Uzmimo opći polinom stupnja* 4

$$p_4(x) = a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0$$

i promatrajmo shemu računananja u formi

$$y = (x + c_0)x + c_1,$$
$$p_4(x) = ((y + x + c_2)y + c_3)c_4.$$

Primijetite da ona ima 3 *množenja i* 5 *zbrajanja, ako uspijemo odrediti* c_i *u ovisnosti o* a_i*.*

Izrazimo li ovaj oblik za p_4 *u potencijama od* x*, dobivamo*

$$p_4(x) = c_4x^4 + (2c_0c_4 + c_4)x^3 + (c_0^2 + 2c_1 + c_0c_4 + c_2c_4)x^2 + (2c_0c_1c_4 + c_1c_4 + c_0c_2c_4)x + (c_1^2c_4 + c_1c_2c_4 + c_3c_4).$$

Uočite da veza između a_i *i* c_i *nije linearna. Rješavanjem po* a_i*, dobivamo*

$$\begin{aligned} c_4 &= a_4 & c_1 &= a_1/a_4 - c_0b \\ c_0 &= (a_3/a_4 - 1)/2 & c_2 &= b - 2c_1 \\ b &= a_2/a_4 - c_0(c_0 + 1) & c_3 &= a_0/a_4 - c_1(c_1 + c_2). \end{aligned}$$

Ove relacije zahtjevaju dosta računanja, ali se to obavlja samo jednom, pa će izvrednjavanje u dovoljno točaka zahtjevati manje množenja.

Dapače, V. Pan je pokazao da vrijedi sljedeći teorem.

Teorem 2.1.2. (Pan) *Za bilo koji polinom* p_n *stupnja* $n \geq 3$ *postoje realni brojevi* c*,* d_i*,* e_i*, za* $0 \leq i \leq \lceil n/2 \rceil - 1$*, takvi da se* p_n *može izračunati korištenjem*

$$(\lfloor n/2 \rfloor + 2) \text{ množenja } + n \text{ zbrajanja}$$

po sljedećoj shemi

$$\begin{aligned} y &= x + c \\ w &= y^2 \\ z &:= \begin{cases} (a_n y + d_0)y + e_0, & n \text{ paran}, \\ a_n y + e_0, & n \text{ neparan}, \end{cases} \\ z &:= z(w - d_i) + e_i, \quad \text{za } i = 1, 2, \ldots \lceil n/2 \rceil - 1. \end{aligned}$$

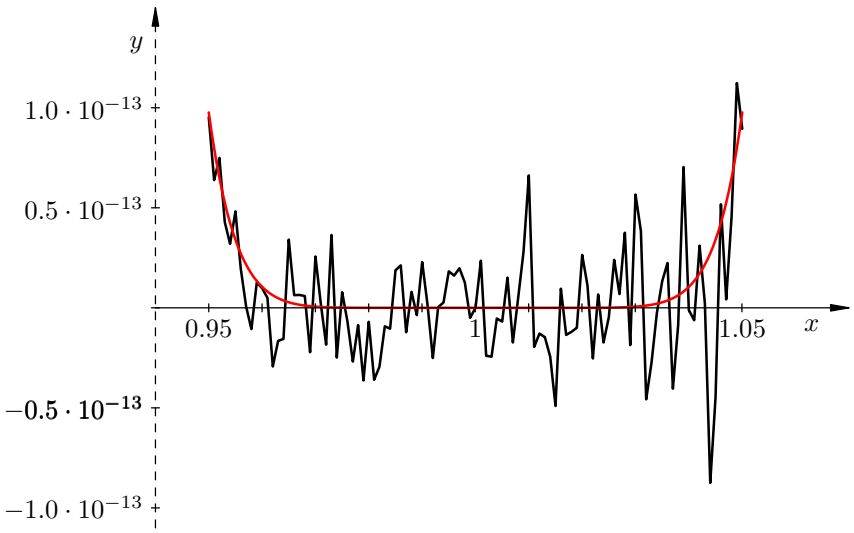
U prethodnom teoremu nismo ništa rekli o tome koliko nam je operacija potrebno za računanje c, d_i i e_i.

Teorem 2.1.3. (Motzkin, Belaga) *Slučajno odabrani polinom stupnja* n *ima vjerojatnost* 0 *da ga se može izračunati za strogo manje od* $\lceil (n + 1)/2 \rceil$ *množenja/dijeljenja ili za strogo manje od* n *zbrajanja/oduzimanja.*

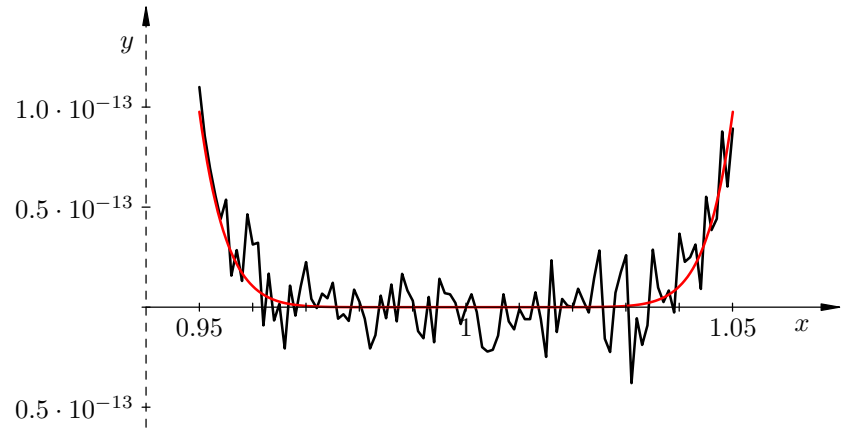
Što to znači? To znači da je red veličine broja operacija u Hornerovoj shemi optimalan za gotovo sve polinome.

2.1.3. Stabilnost Hornerove sheme

Sada znamo da je Hornerova shema optimalan algoritam u smislu efikasnosti. Pažljivom analizom grešaka zaokruživanja nije teško pokazati da je Hornerova shema i stabilan algoritam.



Izvrednjavanje polinoma $(x-1)^{10}$ razvijenog po potencijama od x: korištenjem direktne sumacije u double precision aritmetici.



Izvrednjavanje polinoma $(x-1)^{10}$ razvijenog po potencijama od x: korištenjem Hornerove sheme u double precision aritmetici.

2.1.4. Dijeljenje polinoma linearnim faktorom oblika $x-x_0$

Kako se praktično zapisuje Hornerova shema kad se radi "na ruke"? Napravi se tablica na sljedeći način. U gornjem se redu popišu svi koeficijenti polinoma p_n.

Donji se red formira na sljedeći način: vodeći se koeficijent prepiše, a svi ostali se računaju tako da se posljednji izračunati koeficijent pomnoži s x_0, a zatim mu se doda koeficijent iznad. Na kraju, ispod koeficijenta a_0 piše vrijednost polinoma u točki x_0. Pokažimo kako to funkcionira na konkretnom primjeru.

Primjer 2.1.2. *Izračunajmo vrijednost polinoma*

$$p_5(x) = 2x^5 - x^3 + 4x^2 + 1$$

u točki $x_0 = -1$.

Formirajmo tablicu:

	2	0	−1	4	0	1
−1	2	−2	1	3	−3	4

.

Dakle, $p_5(-1) = 4$.

Pogledajmo općenito što je značenje koeficijenata c_i koji se javljaju u donjem redu tablice

	a_n	a_{n-1}	$\cdots$	a_1	a_0
x_0	c_{n-1}	c_{n-2}	$\cdots$	c_0	r_0

.

Primijetite da prema algoritmu 2.1.2. (pravilu za popunjavanje tablice) vrijedi da je

$$\begin{aligned} &c_{n-1} = a_n, \\ &c_{i-1} := c_i * x_0 + a_{i-1}, \quad i = n, \ldots, 1. \end{aligned} \tag{2.1.1}$$

Očito je $r_0 = p_n(x_0)$. Promatrajmo polinom koji dobijemo dijeljenjem polinoma p_n linearnim faktorom $x - x_0$. Nazovimo taj polinom q_{n-1}. Tada vrijedi

$$p_n(x) = (x - x_0)q_{n-1}(x) + r_0. \tag{2.1.2}$$

Znamo da je q_{n-1} polinom stupnja $n - 1$ s koeficijentima

$$q_{n-1}(x) = \sum_{i=0}^{n-1} b_{i+1}x^i. \tag{2.1.3}$$

Dodatno, označimo s $b_0 = r_0$.

Uvrstimo li (2.1.3) u (2.1.2) i sredimo koeficijente uz odgovarajuće potencije, dobivamo

$$p_n(x) = b_n x^n + (b_{n-1} - x_0\, b_n)x^{n-1} + \cdots + (b_1 - x_0\, b_2)x + b_0 - x_0\, b_1.$$

Za vodeći koeficijent vrijedi $b_n = a_n$, a za a_i, uz $i < n$, je

$$a_i = b_i - x_0 \cdot b_{i+1},$$

odnosno, b_i možemo izračunati iz b_{i+1}

$$b_i = a_i + x_0 \cdot b_{i+1}.$$

Primijetite da je to relacija istog oblika kao (2.1.1), samo s pomaknutim indeksima, pa je

$$b_i = c_{i-1}, \quad i = 1, \ldots, n,$$

tj. koeficijenti koje dobijemo u Hornerovoj shemi su baš koeficijenti kvocijenta i ostatka pri dijeljenju polinoma p_n linearnim faktorom $x - x_0$.

Primjer 2.1.3. *Podijelimo*

$$p_5(x) = 2x^5 - x^3 + 4x^2 + 1$$

linearnim polinomom $x + 1$.

Primijetite da je to ista tablica kao u primjeru 2.1.2., *pa imamo*

	2	0	−1	4	0	1
−1	2	−2	1	3	−3	4

.

Odatle lako čitamo

$$2x^5 - x^3 + 4x^2 + 1 = (x + 1)\left(2x^4 - 2x^3 + x^2 + 3x - 3\right) + 4.$$

Konačno, napišimo algoritam koji nalazi koeficijente pri dijeljenju polinoma linearnim polinomom.

Algoritam 2.1.3. (Dijeljenje polinoma linearnim faktorom $(x - x_0)$**)**

$b_n := a_n$;
for $i := n - 1$ **downto** 0 **do**
  $b_i := b_{i+1} * x_0 + a_i$;

2.1.5. Potpuna Hornerova shema

Što se događa ako postupak dijeljenja polinoma linearnim faktorom nastavimo, tj. ponovimo više puta?

Vrijedi

$$\begin{aligned} p_n(x) &= (x - x_0)q_{n-1}(x) + r_0 \\ &= (x - x_0)[(x - x_0)q_{n-2}(x) + r_1] + r_0 \\ &= (x - x_0)^2 q_{n-2}(x) + r_1(x - x_0) + r_0 \\ &= \cdots \\ &= r_n(x - x_0)^n + \cdots + r_1(x - x_0) + r_0. \end{aligned}$$

Dakle, polinom p_n napisan je razvijeno po potencijama od $(x-x_0)$. Koja su značenja r_i? Usporedimo dobiveni oblik s Taylorovim polinomom oko x_0

$$p_n(x) = \sum_{i=0}^{n} \frac{p_n^{(i)}(x_0)}{i!} (x - x_0)^i.$$

Odatle odmah izlazi relacija za koeficijente

$$r_i = \frac{p_n^{(i)}(x_0)}{i!},$$

tj. potpuna Hornerova shema računa sve derivacije polinoma u zadanoj točki.

Primjer 2.1.4. *Nađite sve derivacije polinoma*

$$p_5(x) = 2x^5 - x^3 + 4x^2 + 1$$

u točki -1.

Formirajmo potpunu Hornerovu tablicu.

	2	0	-1	4	0	1
-1	2	-2	1	3	-3	4
-1	2	-4	5	-2	-1	
-1	2	-6	11	-13		
-1	2	-8	19			
-1	2	-10				
-1	2					

Odatle lako čitamo

$$\begin{aligned} p_5(-1) &= 4, & p_5^{(1)}(-1) &= -1 \cdot 1! = -1, \\ p_5^{(2)}(-1) &= -13 \cdot 2! = -26, & p_5^{(3)}(-1) &= 19 \cdot 3! = 114, \\ p_5^{(4)}(-1) &= -10 \cdot 4! = -240, & p_5^{(5)}(-1) &= 2 \cdot 5! = 240. \end{aligned}$$

Algoritam koji nalazi koeficijente r_i, odnosno derivacije zadanog polinoma u točki, može se napisati u jednom jedinom polju.

Algoritam 2.1.4. (Taylorov razvoj)

for $i := 0$ **to** n **do**
  $r_i := a_i$;
for $i := 1$ **to** n **do**
  for $j := n - 1$ **downto** $i - 1$ **do**
    $r_j := r_j + x_0 * r_{j+1}$;

2.1.6. "Hornerova shema" za interpolacijske polinome

Kao što ćemo vidjeti, kod izvrednjavanja interpolacijskog polinoma u Newtonovoj formi, treba izračunati izraz oblika

$$p_n(x) = a_n (x - x_0)(x - x_1) \cdots (x - x_{n-1}) + a_{n-1} (x - x_0)(x - x_1) \cdots (x - x_{n-2}) + \cdots + a_1 (x - x_0) + a_0,$$

pri čemu su točke x_i točke interpolacije, a x točka u kojoj želimo izračunati vrijednost polinoma.

Algoritam kojim se vrši izvrednjavanje je vrlo sličan Hornerovoj shemi. Označimo s $y_i = x - x_i$. Ponovno, postavimo zagrade kao u Hornerovoj shemi. Vrijedi

$$p_n(x) = (\cdots((a_n y_{n-1} + a_{n-1})y_{n-2} + a_{n-2})y_{n-3} + \cdots + a_1)y_0 + a_0.$$

Algoritam 2.1.5. ("Hornerova shema" za interpolacijske polinome)

$sum := a_n$;
for $i := n - 1$ **downto** 0 **do**
  $sum := sum * (x - x_i) + a_i$;

Dakle, Hornerovu shemu možemo iskoristiti i za ovakav prikaz polinoma, koji više nije u standardnoj bazi.

2.2. Generalizirana Hornerova shema

U prošlom odjeljku napravili smo nekoliko algoritama za izvrednjavanje polinoma i njegovih derivacija u zadanoj točki. Te algoritme, koji su egzaktni u egzaktnoj aritmetici, možemo koristiti i kao **približne** algoritme za izvrednjavanje redova potencija, tj. analitičkih funkcija.

Pretpostavimo da se funkcija f u okolini neke točke x_0 (u $\mathbb{R}$ ili $\mathbb{C}$) može razviti u red potencija oblika

$$f(x) = \sum_{n=0}^{\infty} a_n (x - x_0)^n, \tag{2.2.1}$$

s tim da znamo taj red konvergira prema f na toj okolini od x_0. Dodatno, pretpostavimo da znamo sve koeficijente a_n u ovom razvoju, u smislu da ih možemo brzo i točno izračunati. Naravno, ovu beskonačnu sumu ne možemo efektivno algoritamski izračunati, jer zahtijeva beskonačan broj aritmetičkih operacija.

Međutim, konačne komade ovog razvoja možemo iskoristiti za aproksimaciju funkcije f na toj okolini. Iz konvergencije razvoja po točkama odmah slijedi da, za

bilo koju unaprijed zadanu točnost $\varepsilon > 0$, postoji $N \in \mathbb{N}$ takav da je

$$f_N(x) = \sum_{n=0}^{N} a_n (x - x_0)^n, \tag{2.2.2}$$

aproksimacija za $f(x)$ s greškom manjom od ε. Nije bitno da li grešku mjerimo u apsolutnom ili relativnom smislu, osim ako je $f(x) = 0$. Sasvim općenito, potrebna duljina razvoja N ovisi i o ε i o x. No, ako se sjetimo da redovi potencija konvergiraju uniformno na kompaktima, možemo postići i uniformnu aproksimaciju s točnošću ε na takvim kompaktima, pa N onda ovisi samo o ε.

Kad uzmemo u obzir da ionako **približno** računamo u aritmetici računala, ovim pristupom možemo bitno povećati klasu funkcija s kojima možemo računati. Ako je greška ε dovoljno mala, recimo reda veličine osnovne greške zaokruživanja u, onda je pripadna aproksimacija $f_N(x)$ gotovo jednako dobra kao i $f(x)$.

Algoritam za računanje $f_N(x)$ u zadanoj točki x bitno ovisi u tome da li N znamo unaprijed ili ne. Ako ga **ne znamo**, onda se obično koristi sumacija unaprijed, sve dok se izračunata suma ne stabilizira na zadanu točnost. Koliko to može biti opasno, već smo vidjeli u primjeru za $\sin x$. Zbog toga se sumacija unaprijed koristi samo kao "zadnje utočište".

Vrlo često se N može unaprijed odrediti iz analitičkih svojstava funkcije f, tako da dobijemo uniformnu aproksimaciju s točnošću ε na nekom kompaktu. Obično se za taj kompakt uzima neki segment u $\mathbb{R}$, odnosno neki krug u $\mathbb{C}$. Čak nije jako bitno da N bude "savršen", tj. najmanji mogući, ako je takav N teško izračunati. Katkad je sasvim dobra i približna vrijednost za N. Tada je f_N polinom poznatog stupnja N i možemo koristiti Hornerovu shemu i njene varijacije za računanje $f_N(x)$.

Trenutno ne ulazimo u to kako se nalaze takve aproksimacije. Tome ćemo posvetiti punu pažnju u poglavlju o aproksimacijama. Zasad recimo samo to da se izbjegava direktno korištenje redova potencija (2.2.1) i pripadnih polinomnih aproksimacija u obliku (2.2.2), zbog loše uvjetovanosti sustava funkcija

$$\{1, (x - x_0), (x - x_0)^2, \ldots, (x - x_0)^n, \ldots\}$$

i nejednolikog rasporeda pogreške $e(x) = f(x) - f_N(x)$ na domeni aproksimacije.

Umjesto reda potencija (2.2.1), standardno se koriste razvoji oblika

$$f(x) = \sum_{n=0}^{\infty} a_n p_n(x), \tag{2.2.3}$$

gdje je $\{p_n \mid n \in \mathbb{N}_0\}$ neki **ortogonalni** sustav funkcija na domeni aproksimacije. U aproksimaciji elementarnih i "manje elementarnih" tzv. specijalnih funkcija vrlo često se koriste tzv. Čebiševljevi polinomi, zbog skoro jednolikog rasporeda greške na domeni. Kasnije ćemo pokazati i algoritam za nalaženje takve "kvazi–uniformne" aproksimacije iz poznatog reda potencija (tzv. Čebiševljeva ekonomizacija).

Razvoj funkcije f u red oblika (2.2.3) je očita generalizacija reda potencija. Njega, također, po istom principu, možemo iskoristiti za aproksimaciju funkcije f, ako znamo da on konvergira prema f na nekoj domeni. "Rezanjem" reda dobivamo aproksimaciju funkcije f

$$f_N(x) = \sum_{n=0}^{N} a_n p_n(x), \tag{2.2.4}$$

što je očita generalizacija polinoma iz (2.2.2). Naravno, da bismo izračunali $f_N(x)$ moramo znati sve koeficijente a_n i sve funkcije p_n. Međutim, u većini primjena **nemamo** direktnu "formulu" za računanje vrijednosti $p_n(x)$ u zadanoj točki x, za sve $n \in \mathbb{N}_0$. Umjesto toga, **znamo** da funkcije p_n zadovoljavaju neku, relativno jednostavnu rekurziju po n. Funkcije p_n ne moraju biti polinomi. Dovoljno je da ih možemo rekurzivno računati!

Pristup računanju vrijednosti $f_N(x)$ je isti kao i ranije. Ako unaprijed ne znamo N, onda se sumacija vrši unaprijed, a $p_n(x)$ računa redom iz rekurzije. S druge strane, iz teorije aproksimacija, vrlo često je moguće unaprijed naći koliko članova N treba uzeti za (uniformnu) zadanu točnost. Tada bi bilo zgodno koristiti neku generalizaciju Hornerove sheme za brzo izvrednjavanje f_N oblika (2.2.4) i to je cilj ovog odjeljka.

2.2.1. Izvrednjavanje rekurzivno zadanih funkcija

Budući da ortogonalni polinomi zadovoljavaju tročlane, homogene rekurzije, a vrlo se često koriste, posebnu pažnju posvetit ćemo baš takvim rekurzijama. Osim toga, tročlane rekurzije istog općeg oblika vrijede i za mnoge specijalne funkcije koje ne moraju biti ortogonalne. Zato pretpostavljamo da funkcije p_n, za $n \in \mathbb{N}_0$, zadovoljavaju rekurziju oblika

$$p_{n+1}(x) + \alpha_n(x)p_n(x) + \beta_n(x)p_{n-1}(x) = 0, \quad n = 1, 2, \ldots, \tag{2.2.5}$$

s tim da su poznate "početne" funkcije p_0 i p_1, i sve funkcije α_n, β_n, za $n \in \mathbb{N}$, koje su obično jednostavnog oblika.

Primijetite da potencije $p_n(x) = x^n$ zadovoljavaju dvočlanu homogenu rekurziju

$$p_n(x) - xp_{n-1}(x) = 0, \quad n \in \mathbb{N},$$

uz $p_0(x) = 1$, pa je (2.2.5) zaista generalizacija polinomnog slučaja. Sličan algoritam za brzo izvrednjavanje f_N može se napraviti i kad p_n zadovoljavaju četveročlane ili višečlane rekurzije, ali se takve rekurzije rijetko pojavljuju u praksi.

Algoritam je vrlo sličan izvrednjavanju realnog polinoma u kompleksnoj točki.

Definiramo rekurziju za koeficijente

$$\begin{aligned} B_{N+2} &= B_{N+1} = 0, \\ B_n &= a_n - \alpha_n(x)B_{n+1} - \beta_{n+1}(x)B_{n+2}, \quad n = N, \ldots, 0. \end{aligned} \tag{2.2.6}$$

Uvrštavanjem u formulu (2.2.4) za $f_N(x)$, dobivamo

$$\begin{aligned} f_N(x) &= \sum_{n=0}^{N} a_n p_n(x) = \sum_{n=0}^{N} \left(B_n + \alpha_n(x)B_{n+1} + \beta_{n+1}(x)B_{n+2}\right) p_n(x) \\ &= \sum_{n=-1}^{N-1} B_{n+1}p_{n+1}(x) + \sum_{n=0}^{N} \alpha_n(x)B_{n+1}p_n(x) + \sum_{n=1}^{N+1} \beta_n(x)B_{n+1}p_{n-1}(x) \\ &= \sum_{n=1}^{N-1} B_{n+1}(p_{n+1}(x) + \alpha_n(x)p_n(x) + \beta_n(x)p_{n-1}(x)) \\ &\qquad + B_0p_0(x) + B_1p_1(x) + \alpha_0(x)B_1p_0(x) \\ &= B_0p_0(x) + B_1p_1(x) + \alpha_0(x)B_1p_0(x). \end{aligned}$$

Pripadni silazni algoritam izvrednjavanja ima sljedeći oblik.

Algoritam 2.2.1. (Generalizirana Hornerova shema za $f_N(x)$)

$B_1 := 0;$
$B_0 := a_N;$
for $k := N - 1$ **downto** 0 **do**
  begin;
  $B_2 := B_1;$
  $B_1 := B_0;$
  $B_0 := a_k - \alpha_k(x) * B_1 - \beta_{k+1}(x) * B_2;$
  end;
$f_N(x) := B_0 * p_0(x) + B_1 * (p_1(x) + \alpha_0(x) * p_0(x));$

Ako trebamo izračunati i derivaciju $f'_N(x)$, do pripadnog algoritma možemo doći deriviranjem relacije (2.2.4)

$$f'_N(x) = \sum_{n=0}^{N} a_n p'_n(x),$$

i deriviranjem rekurzije (2.2.5), tako da dobijemo i rekurziju za funkcije p'_n. Pokušajte to napraviti sami.

Međutim, postoji i jednostavniji put, deriviranjem rekurzije (2.2.6), slično kao u algoritmu Bairstowa. Ovdje je to još bitno jednostavnije, jer imamo samo jednu varijablu. Koeficijente B_n shvatimo kao funkcije od x, što oni zaista i jesu. Zatim deriviramo (2.2.6), s tim da B'_n označava derivaciju od B_n po x, u točki x. Takvim

"formalnim" deriviranjem dobivamo rekurziju za koeficijente B'_n.

$$\begin{aligned}
&B_{N+2} = B_{N+1} = 0,\\
&B'_{N+2} = B'_{N+1} = 0,\\
&\quad B_n = a_n - \alpha_n(x)B_{n+1} - \beta_{n+1}(x)B_{n+2}, \quad n = N, \ldots, 0.\\
&\quad B'_n = -\alpha'_n(x)B_{n+1} - \alpha_n(x)B'_{n+1}\\
&\qquad\qquad - \beta'_{n+1}(x)B_{n+2} - \beta_{n+1}(x)B'_{n+2}, \quad n = N, \ldots, 0.
\end{aligned}$$

Odavde odmah vidimo da je i $B'_N = 0$. Uz standardnu oznaku

$$b_n = -\alpha'_n(x)B_{n+1} - \beta'_{n+1}(x)B_{n+2}, \quad n = N, \ldots, 0,$$

s tim da je očito $b_N = 0$, rekurziju za B'_n možemo napisati u obliku

$$B'_n = b_n - \alpha_n(x)B'_{n+1} - \beta_{n+1}(x)B'_{n+2}, \quad n = N, \ldots, 0,$$

što ima skoro isti oblik kao i rekurzija za B_n, osim zamjene a_n u b_n. Konačni rezultat, također, dobivamo deriviranjem ranijeg konačnog rezultata

$$f_N(x) = B_0 p_0(x) + B_1(p_1(x) + \alpha_0(x)p_0(x)),$$

odakle slijedi

$$\begin{aligned}
f'_N(x) = {} & B_0 p'_0(x) + B'_0 p_0(x) + B_1(p'_1(x) + \alpha'_0(x)p_0(x) + \alpha_0(x)p'_0(x)),\\
& + B'_1(p_1(x) + \alpha_0(x)p_0(x)).
\end{aligned}$$

Dakle, da bismo izračunali $f'_N(x)$, dovoljno je znati samo derivacije "početnih" funkcija p'_0 i p'_1, koje su obično jednostavne. Naravno, treba znati i derivacije α'_n, β'_n funkcija iz polazne tročlane rekurzije, ali i one su obično jednostavne. Rekurzija za derivacije p'_n nas uopće ne zanima, iako ju nije teško napisati.

Vidimo da nam za računanje $f'_N(x)$ treba i rekurzija za računanje $f_N(x)$, pa se te dvije vrijednosti obično zajedno računaju, a ne svaka posebno. Tada rekurzije za B_n i B'_n provodimo u istoj petlji. Konačni rezultati izgledaju komplicirano, ali kad u njih uvrstimo konkretne objekte, vrlo rijetko ostanu svi članovi. Obično se te formule svedu na

$$f_N(x) = B_0, \quad f'_N(x) = B'_0.$$

Algoritam 2.2.2. (Generalizirana Hornerova shema za $f_N(x)$ i $f'_N(x)$)

$B_1 := 0;$
$B_0 := a_N;$
$B'_1 := 0;$
$B'_0 := 0;$
for $k := N - 1$ **downto** 0 **do**

begin;
$B_2 := B_1$;
$B_1 := B_0$;
$B_0 := a_k - \alpha_k(x) * B_1 - \beta_{k+1}(x) * B_2$;
$B_2' := B_1'$;
$B_1' := B_0'$;
$b := -\alpha_k'(x) * B_1 - \beta_{k+1}'(x) * B_2$;
$B_0' := b - \alpha_k(x) * B_1' - \beta_{k+1}(x) * B_2'$;
end;
$f_N(x) := B_0 * p_0(x) + B_1 * (\alpha_0(x) * p_0(x) + p_1(x))$;
$f_N'(x) := B_0 * p_0'(x) + B_0' * p_0(x) + B_1 * (p_1'(x) + \alpha_0'(x) * p_0(x) + \alpha_0(x) * p_0'(x))$
$\quad + B_1' * (p_1(x) + \alpha_0(x) * p_0(x))$;

Istim putem možemo izvesti i rekurzije za računanje viših derivacija $f_N^{(k)}(x)$, za $k \geq 2$. Zanimljivo je da u praksi to gotovo nikada nije potrebno. Razlog leži u činjenici da gotovo sve "korisne" familije funkcija p_n, $n \in \mathbb{N}$, zadovoljavaju neke diferencijalne jednadžbe **drugog** reda, s parametrom n. Jasno je da tada treba koristiti odgovarajuću diferencijalnu jednadžbu za računanje $f_N''(x)$, ali i to je vrlo rijetko potrebno.

Čak i algoritam za derivacije se rijetko koristi. Naime, ako znamo naći, tj. izračunati koeficijente a_n u prikazu

$$f(x) = \sum_{n=0}^{\infty} a_n p_n(x),$$

s dovoljnom točnošću, za $n \leq N$, tako da je pripadni $f_N(x)$ dovoljno dobra aproksimacija, onda se **ne isplati** koristiti

$$f_N'(x) = \sum_{n=0}^{N} a_n p_n'(x)$$

kao aproksimaciju za $f'(x)$, jer ona obično ima manju točnost od aproksimacije za f. Puno je bolje izračunati koeficijente a_n' (to nisu derivacije) u pravom razvoju derivacije f' po **istim** funkcijama p_n, a ne po njihovim derivacijama. Dakle, za f' koristimo aproksimaciju oblika

$$f_{N'}'(x) = \sum_{n=0}^{N'} a_n' p_n(x),$$

koja ne mora imati istu duljinu, ali zato ima željenu točnost.

Složenost ovih algoritama ključno ovisi o složenosti računanja svih potrebnih funkcija — p_0, p_1, α_n, β_n, i njihovih derivacija, pa je besmisleno brojati pojedinačne aritmetičke operacije na nivou općeg algoritma.

U praktičnim aproksimacijama se najčešće koriste tzv. ortogonalne familije funkcija p_n, koje čine ortogonalnu bazu u nekom prostoru funkcija, obzirom na neki skalarni produkt na tom prostoru. Vrlo često je p_n polinom stupnja n, za svaki $n \in \mathbb{N}_0$. Neke primjere klasičnih ortogonalnih polinoma i pripadnih rekurzija dajemo nešto kasnije.

Međutim, već smo rekli da funkcije p_n ne moraju biti polinomi i prvi primjer je baš tog tipa.

2.2.2. Izvrednjavanje Fourierovih redova

Za aproksimaciju periodičkih funkcija standardno koristimo Fourierove redove. Pretpostavimo, radi jednostavnosti, da je f periodička funkcija na segmentu $[-\pi, \pi]$. Tada, uz relativno blage pretpostavke, funkciju f možemo razviti u Fourierov red oblika

$$\sum_{n=0}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx).$$

Umjesto a_0, standardno se piše $a_0/2$, ali to nije bitna razlika. Zanemarimo trenutno pitanje konvergencije ovog reda i značenja njegove sume. Uočimo samo da ove trigonometrijske funkcije tvore ortogonalan sustav funkcija, obzirom na skalarni produkt definiran integralom.

Pretpostavimo da su nam koeficijenti a_n i b_n poznati. Naš zadatak je izračunati aproksimaciju oblika

$$f_N(x) = \sum_{n=0}^{N} a_n \cos(nx) + \sum_{n=1}^{N} b_n \sin(nx),$$

gdje je N unaprijed zadan. Ovakav izraz se često zove i trigonometrijski polinom. Vidimo da se on sastoji iz dva dijela, kosinusnog i sinusnog, pa ćemo tako i sastaviti algoritam. Usput, sjetimo se da Fourierov red parne funkcije $f(x) = f(-x)$ ima samo kosinusni dio, a Fourierov red neparne funkcije $f(x) = -f(-x)$ ima samo sinusni dio razvoja.

Pretpostavimo stoga da je f parna funkcija i trebamo izračunati aproksimaciju oblika

$$f_N(x) = \sum_{n=0}^{N} a_n \cos(nx).$$

U direktnoj sumaciji trebamo N računanja funkcije cos, za $\cos(nx)$, uz $n \geq 1$. Iako to danas više ne traje pretjerano dugo, možemo naći i bolji algoritam, koji treba samo jedno jedino računanje funkcije cos.

Da bismo dobili polazni oblik aproksimacije (2.2.4) iz generalizirane Hornerove sheme, očito treba definirati

$$p_n(x) = \cos(nx).$$

Nedostaje nam još samo tročlana homogena rekurzija za ove funkcije. Međutim, i to ide lako, ako se sjetimo formule koja sumu kosinusa pretvara u produkt

$$\cos a + \cos b = 2\cos\left(\frac{a+b}{2}\right)\cos\left(\frac{a-b}{2}\right).$$

Dovoljno je uzeti $a = (n+1)x$ i $b = (n-1)x$. Dobivamo

$$\cos((n+1)x) + \cos((n-1)x) = 2\cos(nx)\cos x,$$

pa tražena rekurzija ima oblik

$$p_{n+1}(x) - 2\cos x\, p_n(x) + p_{n-1}(x) = 0, \quad n \in \mathbb{N},$$

odakle slijedi da u općoj rekurziji (2.2.5) treba uzeti

$$\alpha_n(x) = -2\cos x, \quad \beta_n(x) = 1, \quad n \in \mathbb{N}.$$

Vidimo da $\alpha_n(x)$ i $\beta_n(x)$ ne ovise o n, a $\beta_n(x)$ ne ovisi ni o x, već je konstanta.

Rekurzija (2.2.6) za B_n ima oblik

$$\begin{aligned} B_{N+2} &= B_{N+1} = 0, \\ B_n &= a_n + 2\cos x\, B_{n+1} - B_{n+2}, \quad n = N, \ldots, 0. \end{aligned}$$

Početne funkcije su $p_0(x) = 1$ i $p_1(x) = \cos x$, pa je konačni rezultat

$$\begin{aligned} f_N(x) &= B_0 p_0(x) + B_1(p_1(x) + \alpha_0(x)p_0(x)) \\ &= B_0 \cdot 1 + B_1(\cos x - 2\cos x \cdot 1) \\ &= B_0 - B_1 \cos x. \end{aligned}$$

Sad imamo sve elemente za generaliziranu Hornerovu shemu.

Algoritam 2.2.3. (Fourierov "red" parne funkcije)

$B_1 := 0;$
$B_0 := a_N;$
$alpha := 2 * \cos x;$
for $k := N - 1$ **downto** 0 **do**
  begin;
  $B_2 := B_1;$
  $B_1 := B_0;$
  $B_0 := a_k + alpha * B_1 - B_2;$
  end;
$f_N(x) := B_0 - 0.5 * alpha * B_1;$

Ovaj algoritam zaista "troši" jedan jedini kosinus, pod cijenu jednog množenja s 0.5. Što se stabilnosti tiče, on je podjednako stabilan kao i direktna sumacija. Male vrijednosti $\cos(nx)$ ionako ne dobivamo s malom relativnom, već malom apsolutnom greškom.

Ako trebamo izračunati i derivaciju $f'_N(x)$, za pripadni algoritam trebamo

$$\alpha'_n(x) = 2\sin x, \quad \beta'_n(x) = 0, \quad n \in \mathbb{N}.$$

Onda je

$$b_n = -2\sin x\, B_{n+1}, \quad n = N, \ldots, 0,$$

i

$$B'_n = b_n + 2\cos x\, B'_{n+1} - B'_{n+2}, \quad n = N, \ldots, 0,$$

a formalnim deriviranjem $f_N(x) = B_0 - B_1 \cos x$ dobivamo

$$f'_N(x) = B'_0 - B'_1 \cos x + B_1 \sin x.$$

Dakle, cijeli taj algoritam treba još samo jedan sinus. I tog bismo mogli izbaciti, tako da sinus izrazimo preko kosinusa,

$$\sin x = \pm\sqrt{1 - \cos^2 x},$$

ali to se već ne isplati, jer moramo paziti na znak, a oduzimanje može dovesti do nepotrebnog gubitka točnosti u sinusu.

Pretpostavimo sad da je f neparna funkcija. Trebamo izračunati aproksimaciju oblika

$$f_N(x) = \sum_{n=1}^{N} b_n \sin(nx).$$

Suma ovdje ide od 1, pa treba biti malo oprezan. Zgodniji je zapis

$$f_N(x) = \sum_{n=0}^{N-1} b_{n+1} \sin((n+1)x).$$

Nije baš lijepo ostaviti indeks N u f_N, ali sad je očito da treba definirati

$$p_n(x) = \sin((n+1)x).$$

Zatim koristimo formulu

$$\sin a + \sin b = 2\sin\left(\frac{a+b}{2}\right)\cos\left(\frac{a-b}{2}\right),$$

i uzmemo $a = (n+2)x$ i $b = nx$. Dobivamo

$$\sin((n+2)x) + \sin(nx) = 2\sin((n+1)x)\cos x,$$

pa tražena rekurzija ima oblik

$$p_{n+1}(x) - 2\cos x\, p_n(x) + p_{n-1}(x) = 0, \quad n \in \mathbb{N},$$

što je potpuno isti oblik kao i za parne funkcije, odnosno za $p_n(x) = \cos(nx)$. Dakle, rekurzija za pripadne B_n ima isti oblik, samo starta od $N-1$

$$\begin{aligned} B_{N+1} &= B_N = 0, \\ B_n &= b_{n+1} + 2\cos x\, B_{n+1} - B_{n+2}, \quad n = N-1, \ldots, 0. \end{aligned}$$

Početne funkcije su $p_0(x) = \sin x$ i $p_1(x) = \sin(2x) = 2\sin x \cos x$, pa je konačni rezultat

$$\begin{aligned} f_N(x) &= B_0 p_0(x) + B_1(p_1(x) + \alpha_0(x)p_0(x)) \\ &= B_0 \cdot \sin x + B_1(2\sin x \cos x - 2\cos x \cdot \sin x) \\ &= B_0 \sin x. \end{aligned}$$

Algoritam možete i sami napisati.

Za opći Fourierov red koji ima i parni i neparni dio, treba spojiti prethodne algoritme. Jedina je neugoda što je neparni za 1 kraći, jer starta s $N-1$. Ako nas to baš jako smeta, onda možemo i malo drugačije postupiti u neparnom dijelu. Umjetno definiramo da je $b_0 = 0$ i pišemo

$$f_N(x) = \sum_{n=0}^{N} b_n \sin(nx).$$

Zatim uzmemo

$$p_n(x) = \sin(nx).$$

Rekurzija za p_n, naravno, ostaje ista, a za B_n sad vrijedi "produljena" rekurzija

$$\begin{aligned} B_{N+1} &= B_N = 0, \\ B_n &= b_n + 2\cos x\, B_{n+1} - B_{n+2}, \quad n = N, \ldots, 0. \end{aligned}$$

Početne funkcije su $p_0(x) = 0$ i $p_1(x) = \sin x$, pa je konačni rezultat

$$\begin{aligned} f_N(x) &= B_0 p_0(x) + B_1(p_1(x) + \alpha_0(x)p_0(x)) \\ &= B_0 \cdot 0 + B_1(\sin x - 2\cos x \cdot 0) \\ &= B_1 \sin x. \end{aligned}$$

To pokazuje da B_0 uopće ne treba računati, ali baš to i očekujemo, kad smo rekurziju pomakli za jedan indeks naviše!

Spomenimo na kraju da obje funkcije $\cos(nx)$ i $\sin(nx)$ zadovoljavaju istu diferencijalnu jednadžbu drugog reda

$$y'' + n^2 y = 0.$$

3. Numerička integracija

3.1. Općenito o integracionim formulama

Zadana je funkcija $f : I \to \mathbb{R}$, gdje je I obično interval (može i beskonačan). Želimo izračunati integral funkcije f na intervalu $[a, b]$,

$$I(f) = \int_a^b f(x)\,dx. \tag{3.1.1}$$

Svi znamo da je deriviranje (barem analitički) jednostavan postupak, dok integriranje to nije, pa se integrali analitički u "lijepoj formi" mogu izračunati samo za malen skup funkcija f. Zbog toga, u većini slučajeva ne možemo iskoristiti osnovni teorem integralnog računa, tj. Newton–Leibnitzovu formulu za računanje $I(f)$ preko vrijednosti primitivne funkcije F od f u rubovima intervala

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

Drugim riječima, jedino što nam preostaje je približno, numeričko računanje $I(f)$.

Osnovna ideja numeričke integracije je izračunavanje $I(f)$ korištenjem vrijednosti funkcije f na nekom konačnom skupu točaka. Recimo odmah da postoje i integracione formule koje koriste i derivacije funkcije f, ali o tome kako se one dobivaju i čemu služe, bit će više riječi nešto kasnije.

Opća integraciona formula ima oblik

$$I(f) = I_m(f) + E_m(f),$$

pri čemu je $m+1$ broj korištenih točaka, $I_m(f)$ pripadna aproksimacija integrala, a $E_m(f)$ pritom napravljena greška. Ovakve formule za približnu integraciju funkcija jedne varijable (tj. na jednodimenzionalnoj domeni) često se zovu i **kvadraturne** formule, zbog interpretacije integrala kao površine ispod krivulje.

Ako koristimo samo funkcijske vrijednosti za aproksimaciju integrala, onda aproksimacija $I_m(f)$ ima oblik

$$I_m(f) = \sum_{k=0}^{m} w_k^{(m)} f(x_k^{(m)}), \tag{3.1.2}$$

pri čemu je m neki unaprijed zadani prirodni broj. Koeficijenti $x_k^{(m)}$ zovu se čvorovi integracije, a $w_k^{(m)}$ težinski koeficijenti.

U općem slučaju, za fiksni m, moramo nekako odrediti $2m + 2$ nepoznatih koeficijenata. Uobičajen način njihovog određivanja je zahtjev da su integracione formule egzaktne na vektorskom prostoru **polinoma** što višeg stupnja. Zašto baš tako? Ako postoji Taylorov red za funkciju f i ako on konvergira, onda bi to značilo da integraciona formula egzaktno integrira početni komad Taylorovog reda, tj. Taylorov polinom. Drugim riječima, greška bi bila mala, tj. jednaka integralu greške koji nastaje kad iz Taylorovog reda napravimo Taylorov polinom.

Zbog linearnosti integrala kao funkcionala

$$\int (\alpha f(x) + \beta g(x))\, dx = \alpha \int f(x)\, dx + \beta \int g(x)\, dx, \tag{3.1.3}$$

dovoljno je gledati egzaktnost tih formula na nekoj bazi vektorskog prostora, recimo na

$$\{1, x, x^2, x^3, \ldots, x^m, \ldots\},$$

jer svojstvo (3.1.3) onda osigurava egzaktnost za sve polinome do najvišeg stupnja baze.

Ako su čvorovi fiksirani, recimo ekvidistantni, onda dobivano tzv. Newton–Cotesove formule, za koje moramo odrediti $m + 1$ nepoznati koeficijent (težine). Uvjeti egzaktnosti na vektorskom prostoru polinoma tada vode na sustav linearnih jednadžbi. Kasnije ćemo pokazati da se te formule mogu dobiti i kao integrali interpolacionih polinoma stupnja m za funkciju f na zadanoj (ekvidistantnoj) mreži čvorova.

S druge strane, možemo fiksirati samo neke čvorove, ili dozvoliti da su svi čvorovi "slobodni". Ove posljednje formule zovu se formule Gaussovog tipa. U slučaju Gaussovih formula (ali može se i kod težinskih Newton–Cotesovih formula) uobičajeno je (3.1.1) zapisati u obliku

$$I(f) = \int_a^b w(x) f(x)\, dx, \tag{3.1.4}$$

pri čemu je funkcija $w \geq 0$ tzv. težinska funkcija. Ona ima istu ulogu "gustoće" mjere kao i kod metode najmanjih kvadrata. Ideja je "razdvojiti" podintegralnu

funkciju na dva dijela, tako da singulariteti budu uključeni u w. Gaussove se formule nikad ne računaju "direktno" iz uvjeta egzaktnosti, jer to vodi na nelinearni sustav jednadžbi. Pokazat ćemo da postoji veza Gaussovih formula, funkcije w i ortogonalnih polinoma obzirom na funkciju w na intervalu $[a, b]$, koja omogućava efikasno računanje svih parametara za Gaussove formule.

Na kraju ovog uvoda spomenimo još da postoje primjene u kojima je korisno tražiti egzaktnost integracionih formula na drugačijim sustavima funkcija, koji nisu prostori polinoma do određenog stupnja.

3.2. Newton–Cotesove formule

Newton–Cotesove formule zatvorenog tipa imaju ekvidistantne čvorove, s tim da je prvi čvor u točki $x_0 := a$, a posljednji u $x_m := b$. Preciznije, za zatvorenu (to se često ispušta) Newton–Cotesovu formulu s $(m + 1)$-nom točkom čvorovi su

$$x_k^{(m)} = x_0 + kh_m, \quad k = 0, \ldots, m, \quad h_m = \frac{b-a}{m}.$$

Drugim riječima, osnovni je oblik Newton–Cotesovih formula

$$\int_a^b f(x)\,dx \approx I_m(f) = \sum_{k=0}^{m} w_k^{(m)} f(x_0 + kh_m). \tag{3.2.1}$$

3.2.1. Trapezna formula

Izvedimo najjednostavniju (zatvorenu) Newton–Cotesovu formulu za $m = 1$.

Za $m = 1$, aproksimacija integrala (3.2.1) ima oblik

$$I_1(f) = w_0^{(1)} f(x_0) + w_1^{(1)} f(x_0 + h_1),$$

pri čemu je

$$h := h_1 = \frac{b-a}{1} = b - a,$$

pa je $x_0 = a$ i $x_1 = b$. Da bismo olakšali pisanje, kad znamo da je $m = 1$, možemo izostaviti gornje indekse u $w_k^{(1)}$, tj., radi jednostavnosti, pišemo $w_k := w_k^{(1)}$. Dakle, moramo pronaći težine w_0 i w_1, tako da integraciona formula egzaktno integrira polinome što višeg stupnja na intervalu $[a, b]$, tj. da za polinome f što višeg stupnja bude

$$\int_a^b f(x)\,dx = I_1(f) = w_0 f(a) + w_1 f(b).$$

Stavimo, redom, uvjete na bazu vektorskog prostora polinoma. Ako je f neki od polinoma baze vektorskog prostora, morat ćemo izračunati njegov integral. Zbog toga je zgodno odmah izračunati integrale oblika

$$\int_a^b x^k \, dx, \quad k \geq 0,$$

a zatim rezultat koristiti za razne k. Vrijedi

$$\int_a^b x^k \, dx = \frac{x^{k+1}}{k+1} \Bigg|_a^b = \frac{b^{k+1} - a^{k+1}}{k+1}. \tag{3.2.2}$$

Za $f(x) = 1 = x^0$ dobivamo

$$b - a = \int_a^b x^0 \, dx = w_0 \cdot 1 + w_1 \cdot 1.$$

Odmah je očito da iz jedne jednadžbe ne možemo odrediti dva nepoznata parametra, pa moramo zahtjevati da integraciona formula bude egzaktna i na polinomima stupnja 1.

Za $f(x) = x$ izlazi

$$\frac{b^2 - a^2}{2} = \int_a^b x \, dx = w_0 \cdot a + w_1 \cdot b.$$

Sada imamo dvije jednadžbe s dvije nepoznanice

$$\begin{aligned} w_0 + w_1 &= b - a \\ aw_0 + bw_1 &= \frac{b^2 - a^2}{2}. \end{aligned}$$

Pomnožimo li prvu jednadžbu s $-a$ i dodamo drugoj, dobivamo

$$(b-a)w_1 = \frac{b^2 - a^2}{2} - a(b-a) = \frac{b^2 - 2ab + a^2}{2} = \frac{(b-a)^2}{2}.$$

Budući da je $a \neq b$, dijeljenjem s $b - a$, dobivamo

$$w_1 = \frac{1}{2}(b-a) = \frac{h}{2}.$$

Drugu težinu w_0 lako izračunamo iz prve jednadžbe linearnog sustava

$$w_0 = b - a - w_1 = \frac{1}{2}(b-a) = \frac{h}{2},$$

pa je $w_0 = w_1$.

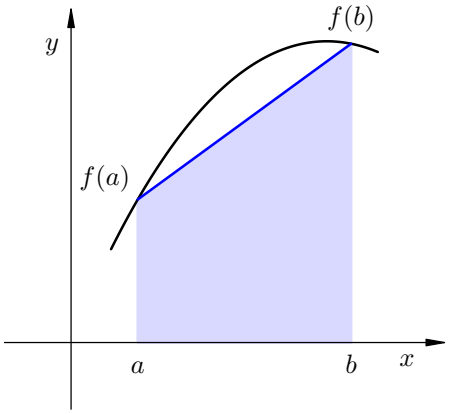
Vidimo da je integraciona formula $I_1(f)$ dobivena iz egzaktnosti na svim polinomima stupnja manjeg ili jednakog 1, i glasi

$$\int_a^b f(x)\,dx \approx \frac{h}{2}\,(f(a) + f(b)).$$

Ta formula zove se trapezna formula. Odakle joj ime? Napišemo li je na malo drugačiji način, kao

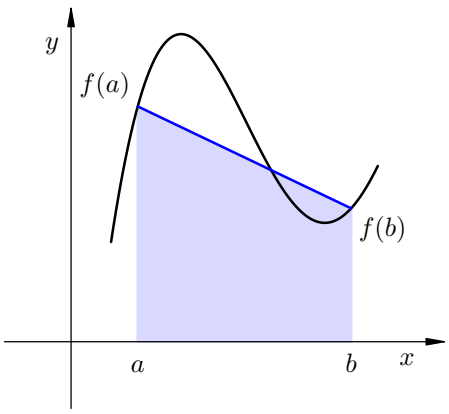
$$\int_a^b f(x)\,dx \approx \frac{f(a) + f(b)}{2}\,(b - a),$$

odmah ćemo vidjeti da je $(f(a) + f(b))/2$ srednjica, a $b - a$ visina trapeza sa slike.



Drugim riječima, površinu ispod krivulje zamijenili smo (tj. aproksimirali) površinom trapeza.

Koliko je ta zamjena dobra? Ovisi o funkciji f. Sve dok pravac razumno aproksimira oblik funkciju f, greška je mala. Na primjer, za funkciju



pravac nije dobra aproksimacija za oblik funkcije f. Da smo nacrtali funkciju f "simetričnije" oko sjecišta, moglo bi se dogoditi da je greška vrlo mala, jer bi se ono što je previše uračunato u površinu s jedne strane "skratilo" s onim što je premalo uračunato s druge strane. S numeričkog stanovišta, takav pristup je opasan.

Trapezna integraciona formula neće egzaktno integrirati sve polinome stupnja 2. To nije teško pokazati, jer već za

$$f(x) = x^2$$

vrijedi

$$\frac{b^3 - a^3}{3} = \int_a^b x^2\,dx \neq I_1(x^2) = \frac{a^2 + b^2}{2}\,(b - a).$$

Slika nas upućuje na još jednu činjenicu. Povučemo li kroz $(a, f(a))$, $(b, f(b))$ linearni interpolacioni polinom, a zatim ga egzaktno integriramo od a do b, dobivamo trapeznu formulu. Pokažimo da je to tako.

Interpolacioni polinom stupnja 1 koji prolazi kroz zadane točke je

$$p_1(x) = f(a) + f[a, b]\,(x - a).$$

Njegov integral na $[a, b]$ je

$$\begin{aligned}\int_a^b p_1(x)\,dx &= \left(f(a)x - a\,f[a, b]x + f[a, b]\,\frac{x^2}{2}\right)\Bigg|_a^b \\ &= (b - a)f(a) + \frac{(b - a)^2}{2}\,f[a, b] = (b - a)\,\frac{f(a) + f(b)}{2}.\end{aligned}$$

Ovaj nam pristup omogućava i ocjenu greške integracione formule, preko ocjene greške interpolacionog polinoma, uz uvjet da možemo ocijeniti grešku interpolacionog polinoma (tj. ako f ima dovoljan broj neprekidnih derivacija).

Neka je funkcija $f \in C^2[a, b]$. Greška interpolacionog polinoma stupnja 1 koji funkciju f interpolira u točkama $(a, f(a))$, $(b, f(b))$ na intervalu $[a, b]$ jednaka je

$$e_1(x) = f(x) - p_1(x) = \frac{f''(\xi)}{2}\,(x - a)\,(x - b).$$

Drugim riječima, vrijedi

$$E_1(f) = \int_a^b \frac{f''(\xi)}{2}\,(x - a)\,(x - b)\,dx.$$

Ostaje samo izračunati $E_1(f)$. Iskoristit ćemo generalizaciju teorema srednje vrijednosti za integrale. Ako su funkcije g i w integrabilne na $[a,b]$ i ako je $w(x) \geq 0$ na $[a,b]$, a

$$m = \inf_{x\in[a,b]} g(x), \quad M = \sup_{x\in[a,b]} g(x),$$

onda vrijedi

$$m \int_a^b w(x)\,dx \leq \int_a^b w(x)g(x)\,dx \leq M \int_a^b w(x)\,dx.$$

Prethodna formula lako se dokazuje, jer je

$$m \leq g(x) \leq M \implies mw(x) \leq g(x)w(x) \leq Mw(x),$$

pa je

$$m \int_a^b w(x)\,dx \leq \int_a^b w(x)g(x)\,dx \leq M \int_a^b w(x)\,dx. \tag{3.2.3}$$

Digresija za nematematičare. inf (čitati infimum) je minimum funkcije koji se ne mora dostići. Na primjer, funkcija

$$g(x) = x \quad \text{na} \quad (0,1) \tag{3.2.4}$$

nema minimum, ali je

$$\inf_{x\in(0,1)} x = 0.$$

Slično vrijedi i za sup (čitati supremum). Supremum je maksimum funkcije koji se ne mora dostići. Na primjer, funkcija iz relacije (3.2.4) nema ni maksimum, ali je

$$\sup_{x\in(0,1)} x = 1.$$

■

Korištenjem relacije (3.2.3), lako dokazujemo integralni teorem srednje vrijednosti s težinama.

Teorem 3.2.1. *Neka su funkcije g i w integrabilne na $[a,b]$ i neka je*

$$m = \inf_{x\in[a,b]} g(x), \quad M = \sup_{x\in[a,b]} g(x).$$

Nadalje, neka je $w(x) \geq 0$ na $[a,b]$. Tada postoji broj μ, $m \leq \mu \leq M$ takav da vrijedi

$$\int_a^b w(x)g(x)\,dx = \mu \int_a^b w(x)\,dx.$$

Posebno, ako je g neprekidna na $[a,b]$, onda postoji broj ζ takav da je

$$\int_a^b w(x)g(x)\,dx = g(\zeta)\int_a^b w(x)\,dx.$$

Dokaz:

Ako je

$$\int_a^b w(x)\,dx = 0,$$

onda je po (3.2.3) i

$$\int_a^b w(x)g(x)\,dx = 0,$$

pa za μ možemo uzeti proizvoljan realan broj. Ako je

$$\int_a^b w(x)\,dx > 0,$$

onda dijeljenjem formule (3.2.3) s prethodnim integralom dobivamo

$$m \le \frac{\int_a^b w(x)g(x)\,dx}{\int_a^b w(x)\,dx} \le M,$$

pa za μ možemo uzeti

$$\mu = \frac{\int_a^b w(x)g(x)\,dx}{\int_a^b w(x)\,dx}.$$

Posljednji zaključak teorema slijedi iz činjenice da neprekidna funkcija na segmentu postiže sve vrijednosti između minimuma i maksimuma, pa mora postići i μ. Drugim riječima, postoji ζ takav da je $\mu = g(\zeta)$. ■

Prisjetite se, već smo pokazali da je

$$E_1(f) = \int_a^b \frac{f''(\xi)}{2}\,(x-a)\,(x-b)\,dx.$$

Primijetite da je funkcija

$$\frac{(x-a)\,(x-b)}{2} \le 0 \quad \text{na} \quad [a,b],$$

pa možemo uzeti

$$w(x) = -\frac{(x-a)\,(x-b)}{2}, \quad g(x) = -f''(\xi).$$

Po generaliziranom teoremu srednje vrijednosti, ako je $f \in C^2[a,b]$, (što znači da je $f'' \in C^0[a,b]$), vrijedi da je

$$E_1(f) = -f''(\zeta) \int_a^b -\frac{(x-a)\,(x-b)}{2}\,dx.$$

Ovaj se integral jednostavno računa. Integriranjem dobivamo

$$\int_a^b \frac{(x-a)\,(x-b)}{2}\,dx = -\frac{(b-a)^3}{12} = -\frac{h^3}{12},$$

pa je

$$E_1(f) = -f''(\zeta)\frac{h^3}{12}.$$

3.2.2. Simpsonova formula

Izvedimo sljedeću (zatvorenu) Newton–Cotesovu formulu za $m = 2$, poznatu pod imenom Simpsonova formula.

Za $m = 2$, aproksimacija integrala (3.2.1) ima oblik

$$I_2(f) = w_0^{(2)} f(x_0) + w_1^{(2)} f(x_0 + h_2) + w_2^{(2)} f(x_0 + 2h_2),$$

pri čemu je

$$h := h_2 = \frac{b-a}{2}.$$

Ponovno, da bismo olakšali pisanje, kad znamo da je $m = 2$, možemo, radi jednostavnosti, izostaviti gornje indekse u $w_k := w_k^{(2)}$. Oprez, to nisu isti w_k i h kao u trapeznoj formuli! Kad uvrstimo značenje h u aproksimacionu formulu, dobivamo

$$I_2(f) = w_0 f(a) + w_1\, f\Big(\frac{a+b}{2}\Big) + w_2 f(b).$$

Stavimo uvjete na egzaktnost formule na vektorskom prostoru polinoma što višeg stupnja. Moramo postaviti najmanje tri jednadžbe, jer imamo tri nepoznata koeficijenta. Za $f(x) = 1$ dobivamo

$$b - a = \int_a^b x^0\,dx = w_0 \cdot 1 + w_1 \cdot 1 + w_2 \cdot 1.$$

Za $f(x) = x$ izlazi

$$\frac{b^2 - a^2}{2} = \int_a^b x\,dx = w_0 \cdot a + w_1 \frac{a+b}{2} + w_2 \cdot b.$$

Konačno, za $f(x) = x^2$ dobivamo

$$\frac{b^3 - a^3}{3} = \int_a^b x^2\,dx = w_0 \cdot a^2 + w_1 \frac{(a+b)^2}{4} + w_2 \cdot b^2.$$

Sada imamo linearni sustav s tri jednadžbe i tri nepoznanice

$$\begin{aligned} w_0 + w_1 + w_2 &= b - a \\ aw_0 + \frac{a+b}{2} w_1 + bw_2 &= \frac{b^2 - a^2}{2} \\ a^2 w_0 + \frac{(a+b)^2}{4} w_1 + b^2 w_2 &= \frac{b^3 - a^3}{3}. \end{aligned}$$

Rješavanjem ovog sustava, dobivamo

$$w_0 = w_2 = \frac{h}{3} = \frac{b-a}{6}, \quad w_1 = \frac{4h}{3} = \frac{4(b-a)}{6}.$$

Drugim riječima, integraciona formula $I_2(f)$ dobivena je iz egzaktnosti na svim polinomima stupnja manjeg ili jednakog 2, i glasi

$$\int_a^b f(x)\,dx \approx \frac{h}{3}\Big(f(a) + 4f\Big(\frac{a+b}{2}\Big) + f(b)\Big).$$

Simpsonova formula ima još jednu prednost. Iako je dobivena iz uvjeta egzaktnosti na vektorskom prostoru polinoma stupnja manjeg ili jednakog 2, ona egzaktno integrira i sve polinome stupnja 3. Dovoljno je pokazati da egzaktno integrira

$$f(x) = x^3.$$

Egzaktni integral jednak je

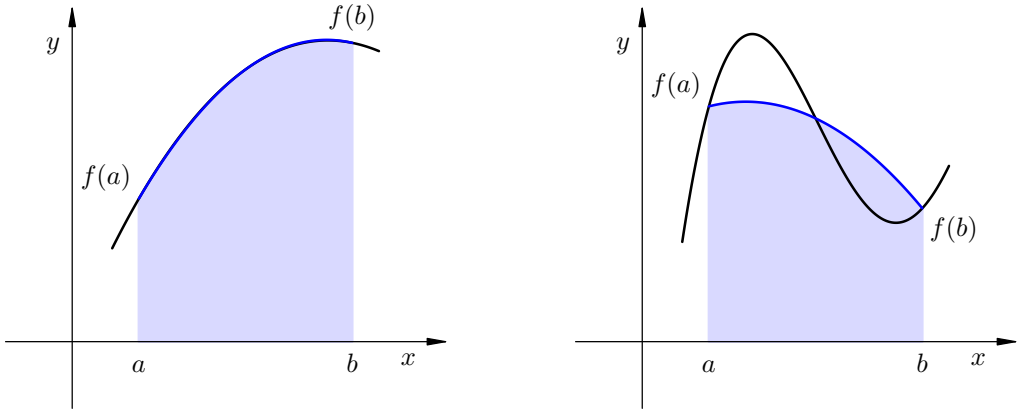
$$\int_a^b x^3\,dx = \frac{b^4 - a^4}{4},$$

a po Simpsonovoj formuli, za $f(x) = x^3$ dobivamo

$$\begin{aligned} I_2(x^3) &= \frac{b-a}{6}\Big(a^3 + 4\Big(\frac{a+b}{2}\Big)^3 + b^3\Big) \\ &= \frac{b-a}{4}\,(a^3 + a^2 b + ab^2 + b^3) = \frac{b^4 - a^4}{4}. \end{aligned}$$

Ponovno, nije teško pokazati da je i ova formula interpolaciona. Ako povučemo kvadratni interpolacioni polinom kroz $(a, f(a))$, $(\frac{a+b}{2}, f(\frac{a+b}{2}))$ i $(b, f(b))$, a zatim ga egzaktno integriramo od a do b, dobivamo Simpsonovu formulu.

Ako pogledamo kako ona funkcionira na funkcijama koje smo već integrirali trapeznom formulom, vidjet ćemo da joj je greška bitno manja. Posebno, na prvom primjeru, kvadratni interpolacioni polinom tako dobro aproksimira funkciju f, da se one na grafu ne razlikuju.



Grešku Simpsonove formule računamo slično kao kod trapezne, integracijom greške kvadratnog interpolacionog polinoma

$$e_2(x) = f(x) - p_2(x) = \frac{f'''(\xi)}{6}\,(x-a)\Big(x - \frac{a+b}{2}\Big)(x-b).$$

Dakle, za grešku Simpsonove formule vrijedi

$$E_2(f) = \int_a^b e_2(x)\,dx.$$

Nažalost, funkcija

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

nije više fiksnog znaka na $[a,b]$, pa ne možemo direktno primijeniti generalizirani teorem srednje vrijednosti. Pretpostavimo da je $f \in C^4[a,b]$. Označimo

$$c := \frac{a+b}{2}$$

i definiramo

$$w(x) = \int_a^x (t-a)\,(t-c)\,(t-b)\,dt.$$

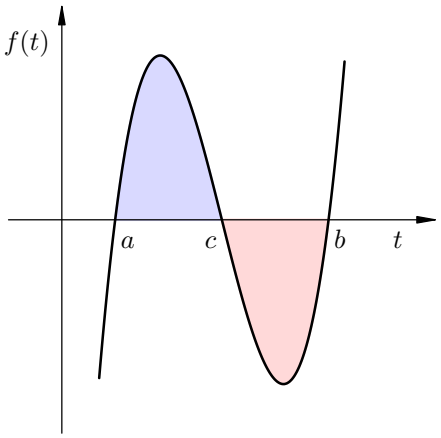
Tvrdimo da vrijedi

$$w(a) = w(b) = 0, \quad w(x) > 0, \quad x \in (a, b). \tag{3.2.5}$$

Skiciramo li funkciju

$$f(t) = (t - a)(t - c)(t - b)$$

odmah vidimo da je ona centralno simetrična oko srednje točke



pa će integral rasti od 0 do svog maksimuma (plava površina), a zatim padati (kad dođe u crveno područje) do 0.

Ostaje samo još napisati grešku interpolacionog polinoma kao podijeljenu razliku. To smo pokazali općenito u poglavlju o Newtonovom interpolacionom polinomu, a posebno za $n = 3$ vrijedi

$$f[a, b, c, x] = \frac{f'''(\xi)}{6}.$$

Uz oznaku (3.2.5), grešku Simpsonove formule, onda možemo napisati kao

$$E_2(f) = \int_a^b w'(x) f[a, b, c, x]\, dx.$$

Parcijalnom integracijom ovog integrala dobivamo

$$E_2(f) = w(x) f[a, b, c, x] \Big|_a^b - \int_a^b w(x) \frac{d}{dx} f[a, b, c, x]\, dx.$$

Prvi član je očito jednak 0, jer je $w(a) = w(b) = 0$. Ostaje još "srediti" drugi član. Kod splajnova smo objašnjavali da je podijeljena razlika s dvostrukim čvorom jednaka derivaciji funkcije. Na sličan je način derivacija treće podijeljene razlike

$f[a,b,c,x]$ po x, četvrta podijeljena razlika s dvostrukim čvorom x. Prema tome, dobivamo formulu greške u obliku

$$E_2(f) = -\int_a^b w(x) f[a,b,c,x,x]\,dx.$$

Sad je funkcija w nenegativna i možemo primijeniti generalizirani teorem srednje vrijednosti. Izlazi

$$E_2(f) = -f[a,b,c,\eta,\eta]\int_a^b w(x)\,dx,$$

gdje je $a \le \eta \le b$. Napišemo li $f[a,b,c,\eta,\eta]$ kao derivaciju, dobivamo

$$E_2(f) = -\frac{f^{(4)}(\zeta)}{4!}\int_a^b w(x)\,dx.$$

Ostaje još samo integrirati funkciju w. Vrijedi

$$\begin{aligned} w(x) &= \int_a^x (t-a)\,(t-c)\,(t-b)\,dt = \text{zamjena varijable } y = t-c \\ &= \int_{-h}^{x-c} (y-h)y(y+h)\,dy = \int_{-h}^{x-c} (y^3 - h^2 y)\,dy \\ &= \left(\frac{y^4}{4} - h^2\frac{y^2}{2}\right)\Bigg|_{-h}^{x-c} = \frac{(x-c)^4}{4} - h^2\frac{(x-c)^2}{2} + \frac{h^4}{4}. \end{aligned}$$

Nadalje je

$$\begin{aligned} \int_a^b w(x)\,dx &= \int_a^b \left(\frac{(x-c)^4}{4} - h^2\frac{(x-c)^2}{2} + \frac{h^4}{4}\right)dx = \text{zamjena varijable } y = x-c \\ &= \int_{-h}^{h} \left(\frac{y^4}{4} - h^2\frac{y^2}{2} + \frac{h^4}{4}\right)dy = \left(\frac{y^5}{20} - h^2\frac{y^3}{6} + \frac{h^4 y}{4}\right)\Bigg|_{-h}^{h} \\ &= 2\left(\frac{h^5}{20} - \frac{h^5}{6} + \frac{h^5}{4}\right) = \frac{4}{15}\,h^5. \end{aligned}$$

Kad to uključimo u formulu za grešku, dobivamo

$$E_2(f) = -\frac{f^{(4)}(\zeta)}{24}\cdot\frac{4}{15}\,h^5 = -\frac{h^5}{90}f^{(4)}(\zeta).$$

Primijetite, greška je za red veličine bolja no što bi po upotrijebljenom interpolacionom polinomu trebala biti.

3.2.3. Produljene formule

Nije teško pokazati da su sve Newton–Cotesove formule integrali interpolacionih polinoma na ekvidistantnoj mreži. Ako ne valja dizanje stupnjeva interpolacionih polinoma na ekvidistantnoj mreži, onda neće biti dobri niti njihovi integrali.

Pokažimo to na primjeru Runge. Prava vrijednost integrala je

$$\int_{-5}^{5} \frac{dx}{1+x^2} = 2\,\mathrm{arctg}\,5 \approx 2.74680153389003172.$$

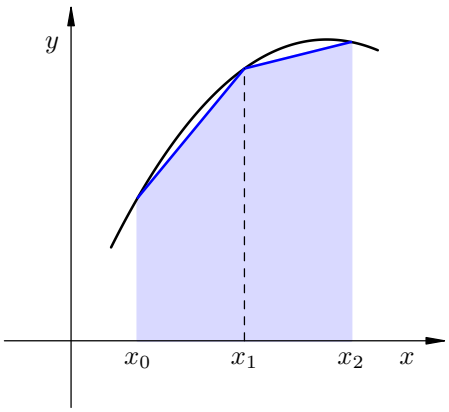
Sljedeća tablica pokazuje aproksimacije integrala izračunate Newton–Cotesovim formulama raznih redova i pripadne greške.

Red formule m	Aproksimacija integrala	Greška
1	0.38461538461538462	2.36218614927464711
2	6.79487179487179487	−4.04807026098176315
3	2.08144796380090498	0.66535357008912674
4	2.37400530503978780	0.37279622885024392
5	2.30769230769230769	0.43910922619772403
6	3.87044867347079978	−1.12364713958076805
7	2.89899440974837875	−0.15219287585834703
8	1.50048890712791179	1.24631262676211993
9	2.39861789784183472	0.34818363604819700
10	4.67330055565349876	−1.92649902176346704
11	3.24477294027858525	−0.49797140638855353
12	−0.31293651575343889	3.05973804964347061
13	1.91979721683238891	0.82700431705764282
14	7.89954464085193082	−5.15274310696189909
15	4.15555899270655713	−1.40875745881652541
16	−6.24143731477308329	8.98823884866311501
17	0.26050944143760372	2.48629209245242800
18	18.87662129010920670	−16.12981975621917490
19	7.24602608588196936	−4.49922455199193763
20	−26.84955208882447960	29.59635362271451140

Očito je da aproksimacije **ne** konvergiraju prema pravoj vrijednosti integrala. Potpunije opravdanje ovog ponašanja dajemo nešto kasnije.

I što sad? Ne smijemo dizati red formula, jer to postaje opasno. Rješenje je vrlo slično onome što smo primijenili kod interpolacije. Umjesto da dižemo red

formule, podijelimo interval $[a, b]$ na više dijelova, recimo, jednake duljine, i na svakom od njih primijenimo odgovarajuću integracionu formulu niskog reda. Tako dobivene formule zovu se **produljene** formule. Na primjer, za funkciju koju smo već razmatrali, produljena trapezna formula s 2 podintervala izgledala bi ovako.



Općenito, produljenu trapeznu formulu dobivamo tako da cijeli interval $[a, b]$ podijelimo na n podintervala oblika $[x_{k-1}, x_k]$, za $k = 1, \ldots, n$, s tim da je

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b,$$

i na svakom od njih upotrijebimo "običnu" trapeznu formulu. Znamo da je tada

$$\int_a^b f(x)\,dx = \sum_{k=1}^{n} \int_{x_{k-1}}^{x_k} f(x)\,dx,$$

pa na isti način zbrojimo i "obične" trapezne aproksimacije u produljenu trapeznu aproksimaciju.

Najjednostavniji je slučaj kad su točke x_k ekvidistantne, tj. kad je svaki podinterval $[x_{k-1}, x_k]$ iste duljine h. To znači da je

$$x_k = a + kh, \quad k = 0, \ldots, n, \quad h = \frac{b-a}{n}.$$

Aproksimacija produljenom trapeznom formulom je

$$\int_a^b f(x)\,dx = h\Big(\frac{1}{2}f_0 + f_1 + \cdots + f_{n-1} + \frac{1}{2}f_n\Big) + E_n^T(f),$$

pri čemu je $E_n^T(f)$ greška produljene formule. Nju možemo zapisati kao zbroj grešaka osnovnih trapeznih formula na podintervalima

$$E_n^T(f) = \sum_{k=1}^{n} -f''(\zeta_k)\frac{h^3}{12}.$$

Greška ovako napisana nije naročito lijepa i korisna, pa ju je potrebno napisati malo drugačije

$$E_n^T(f) = -\frac{h^3 n}{12}\Big(\frac{1}{n}\sum_{k=1}^{n} f''(\zeta_k)\Big).$$

Izraz u zagradi je aritmetička sredina vrijednosti drugih derivacija u točkama ζ_k. Taj se broj sigurno nalazi između najmanje i najveće vrijednosti druge derivacije funkcije f na intervalu $[a, b]$. Budući da je f'' neprekidna na $[a, b]$, onda je broj u zagradi vrijednost druge derivacije u nekoj točki $\xi \in [a, b]$, pa formulu za grešku možemo pisati kao

$$E_n^T(f) = -\frac{h^3 n}{12} f''(\xi) = -\frac{(b-a)h^2}{12} f''(\xi).$$

Iz ove formule izvodimo važnu ocjenu za broj podintervala potrebnih da se postigne zadana točnost za produljenu trapeznu metodu

$$|E_n^T(f)| \leq \frac{(b-a)h^2}{12} M_2 = \frac{(b-a)^3}{12n^2} M_2, \quad M_2 = \max_{x\in[a,b]} |f''(x)|.$$

Želimo li da je $|E_n^T(f)| \leq \varepsilon$, onda je dovoljno tražiti da bude

$$\frac{(b-a)^3}{12n^2} M_2 \leq \varepsilon,$$

odnosno da je

$$n \geq \sqrt{\frac{(b-a)^3 M_2}{12\varepsilon}}, \quad n \text{ cijeli broj.}$$

Na sličan se način izvodi i produljena Simpsonova formula. Primijetite, osnovna Simpsonova formula ima 3 točke, tj. 2 podintervala, pa produljena formula mora imati, također, paran broj podintervala. Pretpostavimo stoga da je n paran broj. Ograničimo se samo na ekvidistantni slučaj. Onda je ponovno

$$h = \frac{b-a}{n}, \quad x_k = a + kh, \quad k = 0, \ldots, n.$$

Aproksimaciju integrala produljenom Simpsonovom formulom dobivamo iz

$$\int_a^b f(x)\,dx = \sum_{k=1}^{n/2} \int_{x_{2k-2}}^{x_{2k}} f(x)\,dx,$$

tako da na svakom podintervalu $[x_{2k-2}, x_{2k}]$, duljine $2h$, primijenimo običnu Simpsonovu formulu, za $k = 1, \ldots, n/2$. Zbrajanjem izlazi

$$\int_a^b f(x)\,dx = \frac{h}{3}\Big(f_0 + 4f_1 + 2f_2 + 4f_3 + 2f_4 + \cdots + 4f_{n-1} + f_n\Big) + E_n^S(f),$$

pri čemu je $E_n^S(f)$ greška produljene formule. Nju možemo zapisati kao zbroj grešaka osnovnih Simpsonovih formula na podintervalima

$$E_n^S(f) = \sum_{k=1}^{n/2} -f^{(4)}(\zeta_k)\frac{h^5}{90}.$$

Opet je grešku korisno napisati malo drugačije

$$E_n^S(f) = -\frac{h^5(n/2)}{90}\Big(\frac{2}{n}\sum_{k=1}^{n/2} f^{(4)}(\zeta_k)\Big).$$

Sličnim zaključivanjem kao kod trapezne formule, izraz u zagradi možemo zamijeniti s $f^{(4)}(\xi)$, $\xi \in [a,b]$, pa dobivamo

$$E_n^S(f) = -\frac{h^5 n}{180}f^{(4)}(\xi) = -\frac{(b-a)h^4}{180}f^{(4)}(\xi).$$

Ponovno, iz ove formule izvodimo ocjenu za broj podintervala potrebnih da se postigne zadana točnost za Simpsonovu metodu

$$|E_n^S(f)| \le \frac{(b-a)h^4}{180}M_4 = \frac{(b-a)^5}{180n^4}M_4, \quad M_4 = \max_{x\in[a,b]}|f^{(4)}(x)|.$$

Želimo li da je $|E_n^S(f)| \le \varepsilon$, onda je dovoljno tražiti da bude

$$\frac{(b-a)^5}{180n^4}M_4 \le \varepsilon,$$

odnosno da je

$$n \ge \sqrt[4]{\frac{(b-a)^5 M_4}{180\varepsilon}}, \quad n \text{ paran cijeli broj.}$$

3.2.4. Primjeri

Primjer 3.2.1. *Izračunajte vrijednost integrala*

$$\int_1^2 xe^{-x}\,dx$$

korištenjem (produljene) Simpsonove formule tako da greška bude manja ili jednaka 10^{-6}. *Nađite pravu vrijednost integrala i pogreške. Koliko je podintervala potrebno za istu točnost korištenjem (produljene) trapezne formule?*

Prmo, moramo ocijeniti pogrešku za produljenu trapeznu i produljenu Simpsonovu formulu. Za to su nam potrebni maksimumi apsolutnih vrijednosti druge i četvrte derivacije na zadanom intervalu. Derivacije su redom

$$f^{(1)}(x) = (1-x)e^{-x}, \quad f^{(2)}(x) = (x-2)e^{-x}, \quad f^{(3)}(x) = (3-x)e^{-x},$$
$$f^{(4)}(x) = (x-4)e^{-x}, \quad f^{(5)}(x) = (5-x)e^{-x}.$$

Nađimo maksimume apsolutnih vrijednosti derivacija na zadanom intervalu.

Prvo ocijenimo grešku za produljenu trapeznu formulu. Na intervalu $[1,2]$ *je* $f^{(3)}(x) > 0$*, što znači da* $f^{(2)}$ *raste. Uočimo još da je na zadanom intervalu* $f^{(2)}(x) \leq 0$*, pa je maksimum apsolutne vrijednosti druge derivacije u lijevom rubu, tj.*

$$M_2 = \max_{x\in[1,2]} |f^{(2)}(x)| = |f^{(2)}(1)| = e^{-1} \approx 0.367879441171.$$

Broj podintervala n_T *za produljenu trapeznu formulu je*

$$n_T \geq \sqrt{\frac{(b-a)^3 M_2}{12\varepsilon}} = \sqrt{\frac{e^{-1}}{12\cdot 10^{-6}}} \approx 175.09,$$

pa je najmanji broj podintervala $n_T = 176$*.*

Sada ocijenimo grešku za produljenu Simpsonovu formulu. Na intervalu $[1,2]$ *je* $f^{(5)}(x) > 0$*, što znači da* $f^{(4)}$ *raste. Također je i* $f^{(4)}(x) < 0$*, što znači da je njen maksimum po apsolutnoj vrijednosti ponovno u lijevom rubu, tj.*

$$M_4 = \max_{x\in[1,2]} |f^{(4)}(x)| = |f^{(4)}(1)| = 3\cdot e^{-1} \approx 1.103638323514.$$

Za grešku produljene Simpsonove formule imamo

$$n_S \geq \sqrt[4]{\frac{(b-a)^5 M_4}{180\varepsilon}} = \sqrt[4]{\frac{3\cdot e^{-1}}{180\cdot 10^{-6}}} \approx 8.85,$$

tj. treba najmanje $n_S = 10$ *podintervala.*

Sad možemo upotrijebiti produljenu Simpsonovu formulu s 10 *podintervala (*11

čvorova). Imamo

k	x_k	$f(x_k)$
0	1.0	0.3678794412
1	1.1	0.3661581921
2	1.2	0.3614330543
3	1.3	0.3542913309
4	1.4	0.3452357495
5	1.5	0.3346952402
6	1.6	0.3230344288
7	1.7	0.3105619909
8	1.8	0.2975379988
9	1.9	0.2841803765
10	2.0	0.2706705665

Sada je

$$\begin{aligned}
&S_0 = f(x_0) + f(x_{10}) = 0.63855000765,\\
&S_1 = 4(f(x_1) + f(x_3) = f(x_5) + f(x_7) + f(x_9)) = 6.5995485226,\\
&S_2 = 2(f(x_2) + f(x_4) + f(x_6) + f(x_8)) = 2.6544824628.
\end{aligned}$$

Vrijednost integrala po Simpsonovoj formuli je

$$I_s = \frac{0.1}{3}(S_0 + S_1 + S_2) = 0.3297526998.$$

U ovom konkretnom slučaju možemo bez puno napora izračunati i egzaktnu vrijednost integrala. Jedina korist od toga je da vidimo koliko je zaista ocjena za Simpsonovu metodu bliska sa stvarnom greškom. Parcijalna integracija daje

$$\begin{aligned}
\int_1^2 xe^{-x}\,dx &= \left\{ \begin{array}{ll} u = x, & du = dx \\ dv = e^{-x}\,dx, & v = -e^{-x} \end{array} \right\} = -xe^{-x}\Big|_1^2 + \int_1^2 e^{-x}\,dx \\
&= e^{-1} - 2e^{-2} - e^{-x}\Big|_1^2 = e^{-1} - 2e^{-2} - e^{-2} + e^{-1} \\
&= 2e^{-1} - 3e^{-2} \approx 0.3297530326.
\end{aligned}$$

Drugim riječima, prava pogreška je

$$I - I_S = 0.3297530326 - 0.3297526998 = 3.328 \cdot 10^{-7},$$

tj. ocjena greške nije daleko od prave pogreške.

3.2.5. Midpoint formula

Ako u Newton–Cotesovim formulama ne interpoliramo (pa onda niti ne integriramo) jednu ili obje rubne točke, dobili smo otvorene Newton–Cotesove formule. Ako definiramo $x_{-1} := a$, $x_{m+1} := b$ i

$$h_m = \frac{b-a}{m+2},$$

onda otvorene Newton–Cotesove formule imaju oblik

$$\int_a^b f(x)\,dx \approx I_m(f) = \sum_{k=0}^{m} w_k^{(m)} f(x_0 + kh_m). \tag{3.2.6}$$

Vjerojatno najkorištenija i najpoznatija otvorena Newton–Cotesova formula je ona najjednostavnija za $m = 0$, poznata pod imenom "midpoint formula" (formula srednje točke).

Dakle za bismo odredili midpoint formulu, moramo naći koeficijent $w_0 := w_0^{(0)}$ takav da je

$$\int_a^b f(x)\,dx = w_0 f\Big(\frac{a+b}{2}\Big)$$

egzaktna na vektorskom prostoru polinoma što višeg stupnja.

Za $f(x) = 1$, imamo

$$b - a = \int_a^b 1\,dx = w_0,$$

odakle odmah slijedi da je

$$\int_a^b f(x)\,dx = (b-a) f\Big(\frac{a+b}{2}\Big).$$

Greška te integracione formule je integral greške interpolacionog polinoma stupnja 0 (konstante), koji interpolira funkciju f u srednjoj točki. Ako definiramo

$$w(x) = \int_a^x (t-c)\,dt, \quad c := \frac{a+b}{2},$$

onda koristeći istu tehniku kao kod izvoda greške za Simpsonovu formulu, izlazi da je greška midpoint formule

$$E_0(f) = \int_a^b e_0(x)\,dx = f''(\xi)\frac{(b-a)^3}{24}.$$

Da bismo izveli produljenu formulu, podijelimo interval $[a, b]$ na n podintervala i na svakom upotrijebimo midpoint formulu. Tada vrijedi

$$I_n(f) = h(f_1 + \cdots + f_n) + E_n^M(f), \quad h = \frac{b-a}{n}, \quad x_k = a + \Big(k - \frac{1}{2}\Big)h,$$

pri čemu je $E_n^M(f)$ ukupna greška koja je jednaka

$$E_n^M(f) = \sum_{k=1}^{n} f''(\xi_k)\frac{h^3}{24} = \frac{h^3 n}{24}\Big(\frac{1}{n}\sum_{k=1}^{n} f''(\xi_k)\Big) = \frac{h^3 n}{24} f''(\xi) = \frac{h^2(b-a)}{24} f''(\xi).$$

3.3. Rombergov algoritam

Pri izvodu Rombergovog algoritma koristimo se sljedećim principima:

- udvostručavanjem broja podintervala u produljenoj trapeznoj metodi,
- eliminacijom člana greške iz dvije susjedne produljene formule. Ponovljena primjena ovog principa zove se Richardsonova ekstrapolacija.

Asimptotski razvoj ocjene pogreške za trapeznu integraciju daje Euler–Mac-Laurinova formula.

Teorem 3.3.1. *Neka je $m \geq 0$, $n \geq 1$, m, n cijeli brojevi. Definiramo ekvidistantnu mrežu s n podintervala na $[a, b]$, tj.*

$$h = \frac{b-a}{n}, \quad x_k = a + kh, \quad k = 0, \ldots, n.$$

Pretpostavimo da je $f \in C^{(2m+2)}[a, b]$. Za pogrešku produljene trapezne metode vrijedi

$$E_n(f) = \int_a^b f(x)\,dx - I_n^T(f) = \sum_{i=1}^{m} \frac{d_{2i}}{n^{2i}} + F_{n,m},$$

gdje su koeficijenti

$$d_{2i} = -\frac{B_{2i}}{(2i)!}\,(b-a)^{2i}\,(f^{(2i-1)}(b) - f^{(2i-1)}(a)),$$

a ostatak je

$$F_{n,m} = \frac{(b-a)^{2m+2}}{(2m+2)!n^{2m+2}} \cdot \int_a^b \overline{B}_{2m+2}\Big(\frac{x-a}{h}\Big)\, f^{(2m+2)}(x)\,dx.$$

Ovdje su B_{2i} Bernoullijevi brojevi,

$$B_i = -\int_0^1 B_i(x)\,dx, \quad i \geq 1,$$

a $\overline{B}_i$ je periodičko proširenje običnih Bernoullijevih polinoma

$$\overline{B}_i(x) = \begin{cases} B_i(x), & za\ 0 \leq x \leq 1, \\ \overline{B}_i(x-1), & za\ x \geq 1. \end{cases}$$

Ovo je jedan od klasičnih teorema numeričke analize i njegov se dokaz može naći u mnogim knjigama.

Umjesto dokaza, nekoliko objašnjenja. Bernoullijevi polinomi zadani su implicitno funkcijom izvodnicom

$$\frac{t(e^{xt}-1)}{e^t-1} = \sum_{i=0}^{\infty} B_i(x)\frac{t^i}{i!}.$$

Prvih nekoliko Bernullijevih polinoma su:

$$B_0(x) = 0 \qquad B_1(x) = x \qquad B_2(x) = x^2 - x$$
$$B_3(x) = x^3 - \frac{3x^2}{2} + \frac{x}{2} \qquad B_4(x) = x^2(1-x)^2.$$

Uvijek vrijedi $B_i(0) = 0$ za $i \geq 0$. Rekurzivne relacije su

$$B_i'(x) = \begin{cases} iB_{i-1}(x), & \text{za } i \text{ paran i } i \geq 4, \\ i(B_{i-1}(x) + B_{i-1}), & \text{za } i \text{ neparan i } i \geq 3. \end{cases}$$

Iz prethodne se formule integracijom mogu dobiti $B_i(x)$, jer je slobodni član jednak 0.

Bernoullijevi brojevi također su definirani implicitno

$$\frac{t}{e^t-1} = \sum_{i=0}^{\infty} B_i\,\frac{t^i}{i!},$$

odakle se integracijom na $[0,1]$ po x u rekurziji za $B_i(x)$ dobiva

$$B_i = -\int_0^1 B_i(x)\,dx, \quad i \geq 1.$$

Prvih nekoliko Bernoullijevih brojeva:

$$B_0 = 1, \qquad B_1 = -\frac{1}{2}, \quad B_2 = \frac{1}{6}, \qquad B_4 = -\frac{1}{30}, \quad B_6 = \frac{1}{42},$$
$$B_8 = -\frac{1}{30}, \quad B_{10} = \frac{5}{66}, \quad B_{12} = -\frac{691}{2730}, \quad B_{14} = \frac{7}{6}, \qquad B_{16} = -\frac{3617}{510}$$

i dalje vrlo brzo rastu po apsolutnoj vrijednosti

$$B_{60} \approx -2.139994926 \cdot 10^{34}.$$

Napomena 3.3.1. *U literaturi se može naći i malo drugačija definicija Bernoullijevih polinoma, označimo ih s $B_i^*(x)$. Oni su zadani implicitno funkcijom izvodnicom*

$$\frac{te^{xt}}{e^t-1}=\sum_{i=0}^{\infty}B_i^*(x)\frac{t^i}{i!}.$$

Veza između jednih i drugih Bernoullijevih polinoma je $B_i^*(x)=B_i(x)+B_i$, za $i\geq 0$.

Rombergov algoritam dobivamo tako da eliminiramo član po član iz reda za ocjenu greške na osnovu vrijednosti integrala s duljinom koraka h i $h/2$.

Za podintegralne funkcije koje nisu dovoljno glatke, također, se može (uz blage pretpostavke) asimptotski dobiti razvoj pogreške. Posebno to vrijedi za funkcije s algebarskim (x^α) i/ili logaritamskim ($\ln x$) singularitetima.

Izvedimo sad Rombergov algoritam. Označimo s $I_n^{(0)}$ trapeznu formulu s duljinom intervala $h=(b-a)/n$. Iz Euler–MacLaurinove formule, ako je n paran, za asimptotski razvoj greške imamo

$$\begin{aligned}I-I_n^{(0)}&=\frac{d_2^{(0)}}{n^2}+\frac{d_4^{(0)}}{n^4}+\cdots+\frac{d_{2m}^{(0)}}{n^{2m}}+F_{n,m}\\ I-I_{n/2}^{(0)}&=\frac{4d_2^{(0)}}{n^2}+\frac{16d_4^{(0)}}{n^4}+\cdots+\frac{2^{2m}d_{2m}^{(0)}}{n^{2m}}+F_{n/2,m}.\end{aligned}$$

Ako prvi razvoj pomnožimo s 4 i oduzmemo mu drugi razvoj, skratit će se prva greška s desne strane $d_2^{(0)}$, tj. dobit ćemo

$$4(I-I_n^{(0)})-(I-I_{n/2}^{(0)})=-\frac{12d_4^{(0)}}{n^4}-\frac{60d_6^{(0)}}{n^6}+\cdots.$$

Izlučivanjem članova koji imaju I na lijevu stranu, a zatim dijeljenjem, dobivamo

$$I=\frac{4I_n^{(0)}-I_{n/2}^{(0)}}{3}-\frac{4d_4^{(0)}}{n^4}-\frac{20d_6^{(0)}}{n^6}+\cdots.$$

Prvi član zdesna možemo uzeti kao bolju, popravljenu aproksimaciju integrala, u oznaci

$$I_n^{(1)}=\frac{4I_n^{(0)}-I_{n/2}^{(0)}}{3},\quad n \text{ paran, } n\geq 2.$$

Niz $I_n^{(2)}$, $I_n^{(4)}$, $I_n^{(6)}$ je novi integracijski niz. Njegova je greška

$$I-I_n^{(1)}=\frac{d_4^{(1)}}{n^4}+\frac{d_6^{(1)}}{n^6}+\cdots,$$

gdje je

$$d_4^{(1)}=-4d_4^{(0)},\quad d_6^{(1)}=-20d_6^{(0)}.$$

Nađimo eksplicitnu formulu za $I_n^{(1)}$. Zbog podjele na odgovarajući broj podintervala, ako je h duljina podintervala za $I_n^{(0)}$, onda je $h_1 := 2h$ duljina podintervala za $I_{n/2}^{(0)}$, pa vrijede sljedeće formule

$$I_n^{(0)} = \frac{h}{2}(f_0 + 2f_1 + \cdots 2f_{n-1} + f_n)$$
$$I_{n/2}^{(0)} = \frac{h_1}{2}(f_0 + 2f_2 + \cdots 2f_{n-2} + f_n).$$

Uvrštavanjem u $I_n^{(1)}$, dobivamo

$$\begin{aligned} I_n^{(1)} &= \frac{4h}{3}\Big(\frac{1}{2}f_0 + 2f_1 + \cdots 2f_{n-1} + \frac{1}{2}f_n\Big) - \frac{2h}{3}\Big(\frac{1}{2}f_0 + 2f_1 + \cdots 2f_{n-1} + \frac{1}{2}f_n\Big) \\ &= \frac{h}{3}(f_0 + 4f_2 + 2f_4 + \cdots 4f_{n-2} + f_n), \end{aligned}$$

što je Simpsonova formula s n podintervala.

Sličan argument kao i prije možemo upotrijebiti i dalje. Vrijedi

$$I - I_{n/2}^{(1)} = \frac{16d_4^{(1)}}{n^4} + \frac{64d_6^{(1)}}{n^6} + \cdots.$$

Tada je

$$16(I - I_n^{(1)}) - (I - I_{n/2}^{(1)}) = \frac{-48d_6^{(1)}}{n^6} + \cdots,$$

odnosno

$$I = \frac{16I_n^{(1)} - I_{n/2}^{(1)}}{15} - \frac{-48d_6^{(1)}}{15n^6} + \cdots.$$

Ponovno, prvi član s desne strane proglasimo za novu aproksimaciju integrala

$$I_n^{(2)} = \frac{16I_n^{(1)} - I_{n/2}^{(1)}}{15}, \quad n \text{ djeljiv s } 4, \quad n \geq 4.$$

Induktivno, ako nastavimo postupak, dobivamo Richardsonovu ekstrapolaciju

$$I_n^{(k)} = \frac{4^k I_n^{(k-1)} - I_{n/2}^{(k-1)}}{4^k - 1}, \quad n \geq 2^k,$$

pri čemu je greška jednaka

$$E_n^{(k)} = I - I_n^{(k)} = \frac{d_{2k+2}^{(k)}}{n^{2k+2}} + \cdots = \beta_k(b-a)h^{2k+2}f^{(2k+2)}(\xi), \quad a \leq \xi \leq b.$$

Sada možemo definirati Rombergovu tablicu

$$\begin{array}{llll} I_1^{(0)} & & & \\ I_2^{(0)} & I_2^{(1)} & & \\ I_4^{(0)} & I_4^{(1)} & I_4^{(2)} & \\ \vdots & \vdots & \vdots & \ddots \end{array}.$$

Ako pogledamo omjere grešaka članova u stupcu, uz pretpostavku dovoljne glatkoće, onda dobivamo

$$\frac{E_n^{(k)}}{E_{2n}^{(k)}} = 2^{2k+2},$$

tj. omjeri pogrešaka u stupcu se moraju ponašati kao

$$\begin{array}{llll} 1 & & & \\ 4 & 1 & & \\ 4 & 16 & 1 & \\ 4 & 16 & 64 & 1 \\ \vdots & \vdots & \vdots & \vdots \end{array} \quad \ddots.$$

Pokažimo na primjeru da prethodni omjeri pogrešaka u stupcu vrijede samo ako je funkcija dovoljno glatka.

Primjer 3.3.1. *Rombergovim algoritmom s točnošću* 10^{-12} *nađite vrijednosti integrala*

$$\int_0^1 e^x\,dx, \quad \int_0^1 x^{3/2}\,dx, \quad \int_0^1 \sqrt{x}\,dx$$

i pokažite kako se ponašaju omjeri pogrešaka u stupcima.

Pogledajmo redom funkcije. Eksponencijalna funkcija ima beskonačno mnogo neprekidnih derivacija, pa bi se računanje integrala morala ponašati po predviđanju. Kao vrijednost, nakon 2^5 *podintervala u trapeznoj formuli, dobivamo umjesto prave vrijednosti integrala* I*, približnu vrijednost*

$$\begin{aligned} I_5 &= 1.71828182845904524 \\ I &= e - 1 = 1.71828182845904524 \\ I - I_5 &= 0. \end{aligned}$$

Pokažimo omjere pogrešaka u stupcima,

```
0 1.0000
1 3.9512  1.0000
2 3.9875 15.6517  1.0000
3 3.9969 15.9913 62.4639   1.0000
4 3.9992 15.9777 63.6087 249.7197    1.0000
5 3.9998 15.9944 63.9017 254.4010 1000.5738 1.0000
```

a zatim samo eksponente omjera pogrešaka (eksponenti od 2*, koji bi ako je funkcija glatka morali biti* $2k+2$*).*

```
0 1.0000
1 1.9823 1.0000
2 1.9955 3.9682 1.0000
3 1.9989 3.9920 5.9650 1.0000
4 1.9997 3.9980 5.9912 7.9642 1.0000
5 1.9999 3.9995 5.9978 7.9910 9.9666 1.0000
```

Što je s drugom funkcijom? Funkciji $f(x) = x^{3/2}$ *puca druga derivacija u* 0*, pa bi zanimljivo ponašanje moralo početi veću drugom stupcu (za trapez je funkcija dovoljno glatka za ocjenu pogreške). Kao vrijednost, nakon* 2^{15} *podintervala u trapeznoj formuli, dobivamo umjesto prave vrijednosti integrala* I*, približnu vrijednost*

$$\begin{aligned} I_{15} &= 0.40000000000004512 \\ I &= 2/5 = 0.40000000000000000 \\ I - I_{15} &= -0.00000000000004512. \end{aligned}$$

Primijetite da je broj intervala poprilično velik! Što je s omjerima pogrešaka?

```
 0  1.0000
 1  3.7346 1.0000
 2  3.8154 5.4847 1.0000
 3  3.8721 5.5912 5.6484 1.0000
 4  3.9112 5.6331 5.6559 5.6566 1.0000
 5  3.9381 5.6484 5.6568 5.6568 5.6569 1.0000
 6  3.9567 5.6539 5.6568 5.6569  ···  5.6569 1.0000
 ⋮     ⋮      ⋮                          ⋱     ⋱
15  3.9981 5.6569  ···                   ···  5.6569 1.0000
```

Primjećujemo da su se nakon prvog stupca omjeri pogrešaka stabilizirali. Bit će nam mnogo lakše provjeriti što se događa ako napišemo samo eksponente omjera

pogrešaka.

$$
\begin{array}{cccccccc}
0 & 1.0000 & & & & & & \\
1 & 1.9010 & 1.0000 & & & & & \\
2 & 1.9318 & 2.4554 & 1.0000 & & & & \\
3 & 1.9531 & 2.4832 & 2.4978 & 1.0000 & & & \\
4 & 1.9676 & 2.4939 & 2.4998 & 2.4999 & 1.0000 & & \\
5 & 1.9775 & 2.4978 & 2.5000 & 2.5000 & 2.5000 & 1.0000 & \\
6 & 1.9843 & 2.4992 & 2.5000 & 2.5000 & 2.5000 & 2.5000 & 1.0000 \\
\vdots & \vdots & \vdots & & & & \ddots & \ddots \\
15 & 1.9993 & 2.5000 & \cdots & & & \cdots & 2.5000 \quad 1.0000
\end{array}
$$

Primijetite da su eksponenti omjera pogrešaka od drugog stupca nadalje točno za 1 *veći od eksponenta same funkcije (integriramo!).*

Situacija s funkcijom $f(x) = \sqrt{x}$ *mora biti još gora, jer njoj puca prva derivacija u* 0*. Nakon* 2^{15} *podintervala u trapeznoj formuli (što je ograničenje zbog veličine polja u programu), ne dobivamo željenu točnost*

$$
\begin{aligned}
I_{15} &= 0.66666665510837633 \\
I &= 2/3 = 0.66666666666666667 \\
I - I_{15} &= 0.00000001155829033.
\end{aligned}
$$

Omjeri pogrešaka u tablici su:

$$
\begin{array}{cccccccc}
0 & 1.0000 & & & & & & \\
1 & 2.6408 & 1.0000 & & & & & \\
2 & 2.6990 & 2.8200 & 1.0000 & & & & \\
3 & 2.7393 & 2.8267 & 2.8281 & 1.0000 & & & \\
4 & 2.7667 & 2.8281 & 2.8284 & 2.8284 & 1.0000 & & \\
5 & 2.7854 & 2.8284 & \cdots & \cdots & 2.8284 & 1.0000 & \\
\vdots & \vdots & \vdots & & & & \ddots & \ddots \\
15 & 2.8271 & 2.8284 & \cdots & & & \cdots & 2.8284 \quad 1.0000
\end{array}
$$

Pripadni eksponenti su

$$
\begin{array}{cccccccc}
0 & 1.0000 & & & & & & \\
1 & 1.4010 & 1.0000 & & & & & \\
2 & 1.4324 & 1.4957 & 1.0000 & & & & \\
3 & 1.4538 & 1.4991 & 1.4998 & 1.0000 & & & \\
4 & 1.4681 & 1.4998 & 1.5000 & 1.5000 & 1.0000 & & \\
5 & 1.4779 & 1.5000 & \cdots & \cdots & 1.5000 & 1.0000 & \\
\vdots & \vdots & \vdots & & & & \ddots & \ddots \\
15 & 1.4993 & 1.5000 & \cdots & & & \cdots & 1.5000 \quad 1.0000
\end{array}
$$

Ipak, u ova dva jednostavna primjera, može se Rombergovom algoritmu "pomoći" tako da supstitucijom u integralu dobijemo glatku funkciju. U oba slučaja, ako stavimo supstituciju $x = t^2$, podintegralna će funkcija imati beskonačno mnogo neprekidnih derivacija, pa će se algoritam ponašati po ocjeni pogreške.

U literaturi postoji i malo drugačija oznaka za aproksimacije integrala u Rombergovoj tablici

$$T_m^{(k)} = \frac{4^m T_{m-1}^{(k+1)} - T_{m-1}^{(k)}}{4^m - 1}.$$

Sama tablica ima oblik

$$\begin{array}{llll} T_0^{(0)} & & & \\ T_0^{(1)} & T_1^{(0)} & & \\ T_0^{(2)} & T_1^{(1)} & T_2^{(0)} & \\ \vdots & \vdots & \vdots & \ddots \end{array}.$$

Pokažimo sad nekoliko primjera kako treba, odnosno ne treba koristiti Rombergov algoritam.

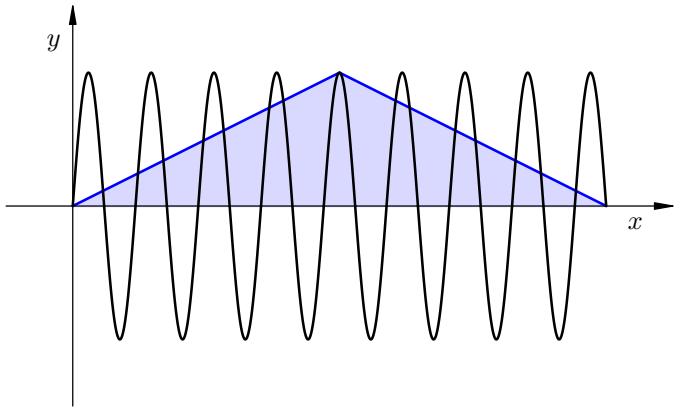
Primjer 3.3.2. *Izračunajte korištenjem Rombergovog algoritma približnu vrijednost integrala*

$$\int_0^1 \sin(17\pi x)\, dx$$

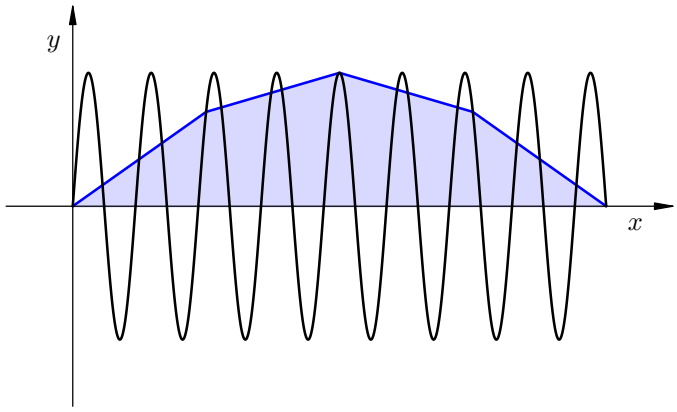
Tako da greška bude manja ili jednaka 10^{-4}. *Napišimo tablicu (samo prvih par decimala, ostale pamtimo u računalu, ali nemamo prostora za ispis)*

0	0.00000							
1	0.50000	0.66667						
2	0.60355	0.63807	0.63616					
3	0.62841	0.63671	0.63661	0.63662				
4	−0.00616	−0.21768	−0.27464	−0.28910	−0.29273			
5	0.02832	0.03982	0.05698	0.06225	0.06362	0.06397		
6	0.03525	0.03756	0.03741	0.03710	0.03700	0.03697	0.03697	
7	0.03690	0.03745	0.03745	0.03745	0.03745	0.03745	0.03745	0.03745

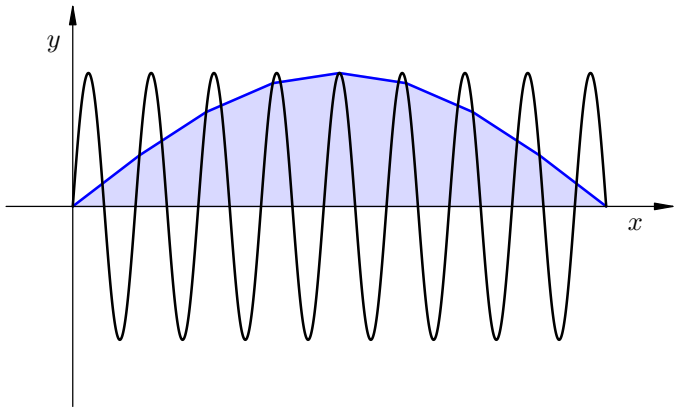
Što je razlog stabilizacije oko jedne, pa oko druge vrijednosti? Nedovoljan broj podintervala u trapezu, koji ne opisuju dobro ponašanje funkcije.



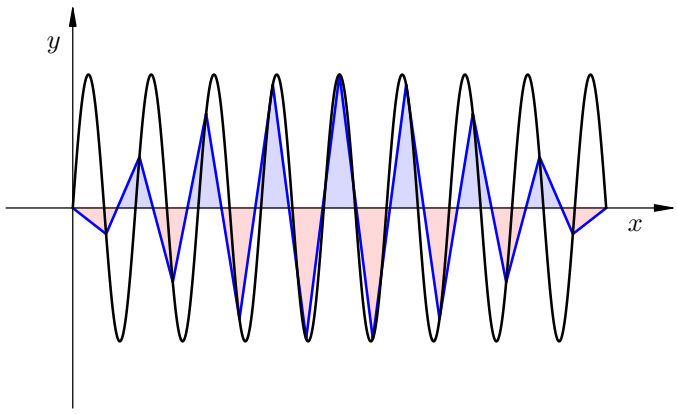
Produljena trapezna formula s 2 *podintervala.*



Produljena trapezna formula s 4 *podintervala.*



Produljena trapezna formula s 8 *podintervala.*



Produljena trapezna formula sa 16 *podintervala.*

3.4. Težinske integracione formule

Dosad smo detaljno analizirali samo nekoliko osnovnih Newton–Cotesovih integracionih formula s malim brojem točaka i pripadne produljene formule. U ovom odjeljku napravit ćemo opću konstrukciju i analizu točnosti za neke klase integracionih formula, uključujući opće Newton–Cotesove i Gaussove formule.

Želimo (približno) izračunati vrijednost integrala

$$I_w(f) = \int_a^b f(x)w(x)\,dx, \tag{3.4.1}$$

gdje je w pozitivna (ili barem nenegativna) "težinska" funkcija za koju pretpostavljamo da je integrabilna na (a, b), s tim da dozvoljavamo da w nije definirana u rubovima a i b. Interval integracije može biti konačan, ali i beskonačan. Drugim riječima, promatramo opći problem jednodimenzionalne integracije zadane funkcije f po zadanoj neprekidnoj mjeri $d\lambda$ generiranoj težinskom funkcijom w na zadanoj domeni. Katkad koristimo i skraćenu oznaku $I(f)$, umjesto $I_w(f)$, za integral u (3.4.1), ako je $w(x) = 1$ na cijelom $[a, b]$, ili kad je težinska funkcija jasna iz konteksta, da skratimo pisanje.

Kao i ranije, ovaj integral aproksimiramo "težinskom" sumom funkcijskih vrijednosti funkcije f na konačnom skupu točaka, Za razliku od ranijih oznaka, ovdje je zgodnije točke numerirati od 1, a ne od 0. Dakle, opća težinska integraciona ili kvadraturna formula za aproksimaciju integrala $I_w(f)$ ima oblik

$$I_n(f) = \sum_{k=1}^{n} w_k^{(n)} f(x_k^{(n)}), \tag{3.4.2}$$

gdje je n prirodni broj. Kao i prije, gornje indekse (n) za čvorove i težine često ne pišemo, ako su očiti iz konteksta, ali ne treba zaboraviti na ovisnost o n.

Dakle, sasvim općenito možemo pisati

$$I_w(f) = \int_a^b f(x)w(x)\,dx = I_n(f) + E_n(f), \tag{3.4.3}$$

gdje je $E_n(f)$ greška aproksimacije.

Osnovnu podlogu za konstrukciju integracionih formula i ocjenu greške $E_n(f)$ daje sljedeći rezultat.

Teorem 3.4.1. *Ako je $I_w(f)$ iz (3.4.1) Riemannov integral, i ako je $\widehat{f}$ bilo koja druga funkcija za koju postoji $I_w(\widehat{f})$, onda vrijedi ocjena*

$$|I_w(f) - I_w(\widehat{f})| \le \|w\|_1 \|f - \widehat{f}\|_\infty, \tag{3.4.4}$$

i postoji funkcija $\widehat{f}$ za koju se ova ocjena dostiže.

Dokaz:

Prvo uočimo da w ne mora biti nenegativna, jer je riječ o Riemannovom integralu, ali zato treba pretpostaviti da je $|w|$ integrabilna.

Ocjena izlazi direktno iz osnovnih svojstava Riemannovog integrala jer podintegralne funkcije moraju biti ograničene. Dobivamo

$$\begin{aligned} |I_w(f) - I_w(\widehat{f})| &= \Big| \int_a^b f(x)w(x)\,dx - \int_a^b \widehat{f}(x)w(x)\,dx \Big| \\ &\le \int_a^b |w(x)| \cdot |f(x) - \widehat{f}(x)|\,dx. \end{aligned}$$

Iskoristimo ocjenu

$$|f(x) - \widehat{f}(x)| \le \sup_{x\in[a,b]} |f(x) - \widehat{f}(x)| = \|f - \widehat{f}\|_\infty, \quad \forall x \in [a,b],$$

i definiciju L_1 norme funkcije w (koja je apsolutno integrabilna po pretpostavci)

$$\|w\|_1 = \int_a^b |w(x)|\,dx,$$

pa dobivamo traženu ocjenu. Ako za perturbiranu funkciju $\widehat{f}$ uzmemo

$$\widehat{f}(x) := f(x) + c\,\mathrm{sign}(w(x)),$$

gdje je $c > 0$ bilo koja konstanta, onda u ocjeni (3.4.4) dobivamo jednakost, uz $\|f - \widehat{f}\|_\infty = c$. ■

U ovoj formulaciji, za klasični Riemannov integral, domena $[a, b]$ integracije mora biti konačna. Teorem onda kaže da je apsolutni broj uvjetovanosti za $I_w(f)$ upravo jednak $\|w\|_1$ i ne ovisi o f, već samo o I_w.

Ovaj rezultat može se proširiti i na neprave Riemannove integrale (beskonačna domena, singulariteti funkcija), i tada više ne vrijedi zaključak o broju uvjetovanosti. Međutim, trenutno nam to nije bitno, već je ključna malo drugačija interpretacija ocjene (3.4.4).

Zamislimo da je $\widehat{f}$ neka aproksimacija (a ne perturbacija) funkcije f, koju želimo iskoristiti za približno računanje integrala. Onda (3.4.4) daje ocjenu (apsolutne) pogreške u integralu, preko greške aproksimacije funkcije u uniformnoj (L_∞) normi na $[a, b]$.

Ono što stvarno želimo dobiti je **niz** aproksimacija integrala koji konvergira prema $I_w(f)$. Jedan od puteva da to postignemo je izbor odgovarajućeg niza aproksimacija $\widehat{f}_n$, $n \in \mathbb{N}$, za funkciju f. Prethodna ocjena upućuje na to da, u ovisnosti o n, za aproksimacione funkcije $\widehat{f}_n$ treba uzimati takve funkcije za koje znamo da možemo postići po volji dobru **uniformnu** aproksimaciju funkcije f, jer tada

$$\|f - \widehat{f}_n\|_\infty \to 0 \Longrightarrow |I_w(f) - I_w(\widehat{f}_n)| \to 0, \quad n \to \infty.$$

Uočimo da ove aproksimacije, naravno, ovise o konkretnoj funkciji f. Da ne bismo za svaki novi f posebno konstruirali odgovarajući niz aproksimacija, poželjno je da bilo koju funkciju f, za koju postoji integral $I_w(f)$, možemo dovoljno dobro aproksimirati nekim prostorom funkcija. Tj. umjesto niza pojedinačnih aproksimacija, koristimo niz vektorskih prostora aproksimacionih funkcija V_n, a za svaki pojedini f nađemo pripadnu aproksimaciju $\widehat{f}_n \in V_n$.

Weierstraßov teorem o uniformnoj aproksimaciji neprekidnih funkcija polinomima na konačnom intervalu $[a, b]$ sugerira da treba uzeti V_n kao prostor polinoma $\mathcal{P}_d$ stupnja manjeg ili jednakog d, gdje d ovisi o n (i raste s n). Kao što ćemo vidjeti, korisno je dozvoliti da bude $d \neq n$.

Isti princip koristimo i za beskonačne domene, samo treba osigurati da su polinomi integrabilni s težinom w. To postižemo dodatnim zahtjevom na težinsku funkciju w, tako da pretpostavimo da svi momenti težinske funkcije

$$\mu_k := \int_a^b x^k w(x)\, dx, \quad k \in \mathbb{N}_0, \tag{3.4.5}$$

postoje i da su konačni. U nastavku pretpostavljamo da težinska funkcija w zadovoljava ovu pretpostavku. Takve težinske funkcije obično zovemo (polinomno) dopustivima.

Napomenimo odmah da se ovaj pristup može generalizirati i na bilo koji drugi sustav funkcija aproksimacionih funkcija $\{\widehat{f}_n \mid n \in \mathbb{N}\}$ koji je gust u prostoru $C[a,b]$ neprekidnih funkcija na $[a,b]$. Pripadni prostori V_n generirani su početnim komadima ovog sustava funkcija (kao linearne ljuske).

Za praktičnu primjenu ovog pristupa moramo moći efektivno izračunati integral $I_w(\widehat{f}_n)$ aproksimacione funkcije, i to za bilo koju funkciju f. To se najlakše postiže tako da konstruiramo pripadnu integracionu formulu I_n koja je egzaktna na cijelom prostoru $V_n = \mathcal{P}_d$ aproksimacionih funkcija. Dakle, uvjet egzaktnosti za I_n je

$$I_w(f) = I_n(f) \quad \text{ili} \quad E_n(f) = 0, \quad \text{za sve } f \in V_n.$$

Iz relacija (3.4.3) i (3.4.4) odmah dobivamo i ocjenu greške pripadne integracione formule $I_n(f)$, za bilo koji f

$$|E_n(f)| = |I_w(f) - I_n(f)| = |I_w(f) - I_w(\widehat{f}_n)| \leq \|w\|_1 \|f - \widehat{f}_n\|_\infty.$$

3.5. Gaussove integracione formule

Kao što smo već rekli, Gaussove formule imaju dvostruko više slobodnih parametara nego Newton–Cotesove, pa bi zbog toga trebale egzaktno integrirati polinome približno dvostruko većeg stupnja od Newton–Cotesovih.

Za razliku od Newton–Côtesovih formula, **Gaussove integracijske formule** su oblika

$$\int_a^b f(x)\,dx \approx \sum_{i=1}^{n} w_i f(x_i),$$

u kojima točke integracije x_i nisu unaprijed poznate, nego se izračunaju tako da greška takve formule bude najmanja. Motivirani praktičnim razlozima, promatrat ćemo malo općenitije integracijske formule oblika

$$\int_a^b w(x)\, f(x)\,dx \approx \sum_{i=1}^{n} w_i f(x_i),$$

gdje je w **težinska funkcija**, pozitivna na otvorenom intervalu (a,b). Koeficijente w_i zovemo **težinski koeficijenti** ili, skraćeno, **težine** integracione formule. Gornji specijalni slučaj u kojem je $w \equiv 1$ čine formule koje se zovu **Gauss–Legendreove**. Težinska funkcija u općem slučaju utječe na težine i točke integracije, ali se ne pojavljuje eksplicitno u Gaussovoj formuli.

Bitno je znati da se za neke težinske funkcije na određenim intervalima, čvorovi

i težine standardno tabeliraju u priručnicima. To su

težinska funkcija w	interval	formula Gauss–
1	$[-1,1]$	Legendre
$\dfrac{1}{\sqrt{1-x^2}}$	$[-1,1]$	Čebišev
$\sqrt{1-x^2}$	$[-1,1]$	Čebišev 2. vrste
e^{-x}	$[0,\infty)$	Laguerre
e^{-x^2}	$(-\infty,\infty)$	Hermite

Glavni rezultat je sljedeći: ako zahtijevamo da formula integrira egzaktno polinome što je moguće većeg stupnja, onda su točke integracije x_i nultočke polinoma koji su ortogonalni na intervalu (a,b) obzirom na težinsku funkciju w, a težine w_i mogu se eksplicitno izračunati po formuli

$$w_i = \int_a^b w(x)\,\ell_i(x)\,dx, \quad i = 1,\ldots,n.$$

Pritom je ℓ_i poseban polinom Lagrangeove baze kojeg smo razmatrali u poglavlju o polinomnoj interpolaciji, definiran uvjetom $\ell_i(x_j) = \delta_{ij}$ (v. (1–7.2.16)). Primijetimo samo da je kod numeričke integracije zgodnije čvorove numerirati od x_1 do x_n, (za razliku od numeracije x_0 do x_n u poglavlju o interpolaciji), pa je i ℓ_i polinom stupnja $n-1$.

Kao što se Newton–Côtesove formule mogu dobiti integracijom Lagrangeovog interpolacijskog polinoma, tako se i Gaussove formule mogu dobiti integracijom Hermiteovog interpolacijskog polinoma. Takav pristup ekvivalentan je s pristupom u kojem zahtijevamo da Gaussove formule integriraju egzaktno polinome što je moguće višeg stupnja, tj. da vrijedi

$$\int_a^b w(x)\,x^j\,dx = \sum_{i=1}^{n} w_i x_i^j, \quad j = 0,1,\ldots,2n-1.$$

Mogli bismo iskoristiti ovu relaciju da napišemo $2n$ jednadžbi za $2n$ nepoznanica x_i i w_i, međutim nepoznanice x_i ulaze u sistem nelinearno, pa je ovakav pristup teži. Čak i dokaz da taj nelinearni sistem ima jedinstveno rješenje nije jednostavan.

Napišimo još jednom formulu za Hermiteov interpolacijski polinom h_{2n-1}, stupnja $2n-1$, koji u čvorovima integracije x_i interpolira vrijednosti $f_i = f(x_i)$

i $f_i' = f'(x_i)$, za $i = 1, \ldots, n$. Iz relacija (1–7.2.19) i (1–7.2.20) dobivamo

$$\begin{aligned} h_{2n-1}(x) &= \sum_{i=1}^{n} \Big(h_{i,0}(x)\, f_i + h_{i,1}(x)\, f_i' \Big) \\ &= \sum_{i=1}^{n} \Big([1 - 2(x - x_i)\ell_i'(x_i)]\, \ell_i^2(x)\, f_i + (x - x_i)\, \ell_i^2(x)\, f_i' \Big). \end{aligned}$$

Integracijom dobijemo

$$\int_a^b w(x)\, h_{2n-1}(x)\, dx = \sum_{i=1}^{n} \Big(A_i f_i + B_i f_i' \Big), \tag{3.5.1}$$

gdje su

$$\begin{aligned} A_i &= \int_a^b w(x)\, [1 - 2(x - x_i)\ell_i'(x_i)]\, \ell_i^2(x)\, dx, \\ B_i &= \int_a^b w(x)\, (x - x_i)\, \ell_i^2(x)\, dx. \end{aligned} \tag{3.5.2}$$

Integraciona formula (3.5.1) sliči na Gaussovu integracionu formulu, osim što ima dodatne članove $B_i f_i'$, koji koriste i derivacije funkcije f u čvorovima integracije.

Kad bi, kao u Newton–Cotesovim formulama, čvorovi x_i bili unaprijed zadani, iz uvjeta egzaktne integracije polinoma trebalo bi odrediti $2n$ parametara — težinskih koeficijenata A_i, B_i. Zato očekujemo da ovakva formula egzaktno integrira polinome do stupnja $2n-1$ (dimenzija prostora je $2n$). No, za upotrebu ove formule trebamo znati ne samo funkcijske vrijednosti $f(x_i)$ u čvorovima, već i vrijednosti derivacije $f'(x_i)$ funkcije u tim čvorovima.

Zato je ideja da probamo izbjeći korištenje derivacija, tako da izborom čvorova x_i **poništimo** koeficijente B_i uz derivacije f_i'. Točnost integracione formule mora ostati ista (egzaktna integracija polinoma stupnja do $2n - 1$), ali tako dobivena formula koristila bi samo funkcijske vrijednosti u čvorovima, tj. postala bi Gaussova integraciona formula.

Zaista, odgovarajućim izborom čvorova x_i može se postići da težinski koeficijenti B_i uz derivacije budu jednaki nula. Da bismo to dokazali, uvodimo posebni "polinom čvorova" (engl. "node polynomial") ω_n, koji ima nultočke u svim čvorovima integracije

$$\omega_n := (x - x_1)(x - x_2) \cdots (x - x_n).$$

Taj polinom smo već susreli u poglavlju o Lagrangeovoj interpolaciji. Sljedeći rezultat govori o tome kako treba izabrati čvorove.

Lema 3.5.1. *Ako je* $\omega_n(x) = (x - x_1) \cdots (x - x_n)$ **ortogonalna s težinom** w *na sve polinome nižeg stupnja, tj. ako vrijedi*

$$\int_a^b w(x)\, \omega_n(x)\, x^k\, dx = 0, \quad k = 0, 1, \ldots, n-1, \tag{3.5.3}$$

onda su svi koeficijenti B_i *u* (3.5.2) *jednaki nula.*

Dokaz:

Lagano provjerimo identitet

$$(x - x_i)\ell_i(x) = \frac{\omega_n(x)}{\omega_n'(x_i)}. \tag{3.5.4}$$

Supstitucijom u izraz (3.5.2) za B_i slijedi

$$B_i = \frac{1}{\omega_n'(x_i)} \int_a^b w(x)\, \omega_n(x)\, \ell_i(x)\, dx.$$

Kako je ℓ_i polinom stupnja $n-1$, i po pretpostavci je ω_n ortogonalna s težinom w na sve takve polinome, tvrdnja slijedi. ■

Lako se vidi da vrijedi i obrat ove tvrdnje, tj. da su svi koeficijenti $B_i = 0$ u (3.5.1), ako i samo ako je polinom čvorova ω_n ortogonalan na sve polinome nižeg stupnja (do $n-1$), s težinskom funkcijom w. Razlog tome je što su funkcije ℓ_i, $i = 1, \ldots, n$, Lagrangeove baze zaista baza prostora $\mathcal{P}_{n-1}$ (zadatak 1–7.2.2.).

Iz ranijih rezultata o ortogonalnim polinomima znamo da ortogonalni polinom stupnja n obzirom na w postoji i jednoznačno je određen do na (recimo) vodeći koeficijent. Da bismo dobili Gaussovu integracionu formulu u (3.5.1), polinom čvorova ω_n mora biti ortogonalni polinom s vodećim koeficijentom 1, tj. ω_n postoji i jedinstven je.

Nadalje, uvjet ortogonalnosti (3.5.3) **jednoznačno** određuje raspored čvorova za Gaussovu integraciju. Iz teorema 1.5.2. slijedi da ω_n ima n jednostrukih nultočaka u otvorenom intervalu (a, b) (što nam baš odgovara za integraciju). Njegove nultočke $x_1, \ldots, x_n$ možemo samo permutirati (drugačije indeksirati), a uz standardni dogovor $x_1 < \cdots < x_n$, one su jednoznačno određene.

Time smo dokazali da postoji jedinstvena Gaussova integraciona formula oblika

$$\int_a^b w(x)\, f(x)\, dx \approx \sum_{i=1}^{n} w_i f(x_i),$$

Čvorovi integracije x_i su nultočke ortogonalnog polinoma stupnja n na $[a, b]$ s težinskom funkcijom w, a težinske koeficijente možemo izračunati iz (3.5.2), budući da je tada $w_i = A_i$, za $i = 1, \ldots, n$.

Iskoristimo li pretpostavku ortogonalnosti iz leme 3.5.1., možemo pojednostavniti i izraze za koeficijente $w_i = A_i$ u (3.5.2). Sasvim općenito, koristeći relaciju za B_i, koeficijent A_i možemo napisati u obliku

$$A_i = \int_a^b w(x)\,[1 - 2(x - x_i)\ell_i'(x_i)]\,\ell_i^2(x)\,dx = \int_a^b w(x)\,\ell_i^2(x)\,dx - 2\ell_i'(x_i)B_i.$$

Uz uvjet ortogonalnosti (Gaussova integracija) je $B_i = 0$ i $A_i = w_i$, pa je

$$w_i = \int_a^b w(x)\,\ell_i^2(x)\,dx.$$

Podintegralna funkcija je nenegativna i ℓ_i^2 je polinom stupnja $2(n-1)$ koji nije nul-polinom, pa desna strana mora biti pozitivna. Dakle, slijedi da su svi težinski koeficijenti u Gaussovoj integraciji pozitivni, $w_i > 0$, za $i = 1, \ldots, n$, što je vrlo bitno za numeričku stabilnost i konvergenciju.

Pokažimo još da vrijedi i

$$w_i = \int_a^b w(x)\,\ell_i^2(x)\,dx = \int_a^b w(x)\,\ell_i(x)\,dx.$$

Očito, to je isto kao i dokazati

$$\int_a^b w(x)\,\ell_i^2(x)\,dx - \int_a^b w(x)\,\ell_i(x)\,dx = \int_a^b w(x)\,\ell_i(x)\,(\ell_i(x) - 1)\,dx = 0.$$

Ali polinom $\ell_i(x) - 1$ se poništava u točki $x = x_i$, po definiciji polinoma ℓ_i, jer je $\ell_i(x_j) = \delta_{ij}$. Znači da $\ell_i(x) - 1$ mora sadržavati $x - x_i$ kao faktor, tj. možemo napisati

$$\ell_i(x) - 1 = (x - x_i)q(x),$$

gdje je q neki polinom stupnja $n-2$, za jedan manje od stupnja polinoma ℓ_i. Dakle,

$$\ell_i(x)\,(\ell_i(x) - 1) = \frac{\omega_n(x)}{\omega_n'(x_i)(x - x_i)}\,(\ell_i(x) - 1) = \frac{1}{\omega_n'(x_i)}\,\omega_n(x)\,q(x),$$

pa je zbog ortogonalnosti ω_n na sve polinome nižeg stupnja

$$\int_a^b w(x)\,\ell_i(x)\,(\ell_i(x) - 1)\,dx = \frac{1}{\omega_n'(x_i)} \int_a^b w(x)\,\omega_n(x)\,q(x)\,dx = 0.$$

■

Pokazali smo da Gaussovu integracionu formulu možemo dobiti kao integral Hermiteovog interpolacijskog polinoma, uz odgovarajući izbor čvorova, a za težinske koeficijente vrijedi

$$w_i = \int_a^b w(x)\,\ell_i(x)\,dx. \tag{3.5.5}$$

Primijetimo da je ova formula za koeficijente ista kao i ona u Newton–Côtesovim formulama, što je ovdje posljedica pretpostavke o ortogonalnosti. U oba slučaja do integracionih formula dolazimo interpolacijom funkcije u čvorovima.

Pokažimo i primjerom da ortogonalnost produkta korijenskih faktora, tj. funkcije $\omega_n(x)$ na sve polinome nižeg stupnja zapravo određuje točke integracije x_i.

Primjer 3.5.1. *Neka je $w(x) = 1$ i $n = 3$. Odredimo točke integracije iz uvjeta ortogonalnosti. Uobičajeno je da za interval integracije uzmemo $(-1, 1)$, budući da integrale na drugim intervalima možemo lagano računati, ako podintegralnu funkciju transformiramo linearnom supstitucijom. Problem se dakle svodi na to da odredimo nultočke kubične funkcije $\omega_3(x) = a + bx + cx^2 + x^3$ za koju vrijedi*

$$\int_{-1}^{1} \omega_3(x)\,x^k\,dx = 0, \quad k = 0, 1, 2.$$

Nakon integracije dobivamo sustav jednadžbi za koeficijente a, b, c

$$2a + \frac{2}{3}c = 0, \qquad \frac{2}{3}b + \frac{2}{5} = 0, \qquad \frac{2}{3}a + \frac{2}{5}c = 0,$$

odakle nađemo $a = c = 0$ i $b = -3/5$. Dobivamo

$$\omega_3(x) = x^3 - \frac{3}{5}x = \left(x + \sqrt{\frac{3}{5}}\right)x\left(x - \sqrt{\frac{3}{5}}\right),$$

odakle slijedi da su točke integracije $x_i = -\sqrt{3/5},\ 0,\ \sqrt{3/5}$.

Teorijski, ovaj pristup možemo iskoristiti za sve moguće intervale integracije i razne težinske funkcije. Za veće n potrebno je odrediti nule polinoma visokog stupnja, što je egzaktno nemoguće, a numerički u najmanju ruku neugodno. Stoga je potrebno za specijalne težine i intervale integracije doći do dodatnih informacija o ortogonalnim polinomima. Na kraju, bilo bi dobro izračunati formulom i težinske faktore w_i u Gaussovim formulama. Analitički je moguće doći do ovakvih rezultata za mnoge specijalne težine $w(x)$ koje se pojavljuju u primjenama. Riješimo na početku važnu situaciju $w \equiv 1$, $a = -1$, $b = 1$. Pripadne formule nazvali smo Gauss–Legendreovima; u gornjem primjeru izračunali smo točke integracije za Gauss–Legendreovu formulu reda 3.

Zadatak 3.5.1. *Iz uvjeta egzaktnosti i poznatih točaka integracije za $n = 3$ izračunajte težinske koeficijente w_i. Primijetite da je sustav jednadžbi linearan, pa stoga računanje ovih faktora ne predstavlja veće probleme.*

3.5.1. Gauss–Legendreove integracione formule

Prepostavimo u daljnjem da je $w \equiv 1$ na intervalu $(-1, 1)$ i izvedimo specijalnu Gaussovu formulu, tj. Gauss–Legendre-ovu formulu

$$\int_{-1}^{1} f(x)\,dx \approx \sum_{i=1}^{n} w_i f(x_i).$$

Kao što znamo, Legendreov polinom stupnja n definiran je **Rodriguesovom formulom**

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

Tako definirani polinomi čine **ortogonalnu bazu** u prostoru polinoma stupnja n, tj. oni su linearno nezavisni i ortogonalni obzirom na skalarni produkt

$$\langle P, Q\rangle := \int_{-1}^{1} P(x)\,Q(x)\,dx. \tag{3.5.6}$$

Pojavljuju se prirodno u parcijalnim diferencijalnim jednadžbama, kod metode separacije varijabli za Laplaceovu jednadžbu u kugli. Za nas je bitno samo jedno specijalno svojstvo, iz kojeg slijede sva ostala:

Lema 3.5.2. *Legendreov polinom stupnja n ortogonalan je na sve potencije x^k nižeg stupnja, tj. vrijedi*

$$\int_{-1}^{1} x^k P_n(x)\,dx = 0, \quad \text{za} \quad k = 0, 1, \ldots, n-1,$$

i vrijedi

$$\int_{-1}^{1} x^n P_n(x)\,dx = \frac{2^{n+1}(n!)^2}{(2n+1)!}.$$

Dokaz:

Uvrštavanjem Rodriguesove formule, nakon k $(k < n)$ parcijalnih integracija dobivamo

$$\int_{-1}^{1} x^k \frac{d^n}{dx^n}(x^2-1)^n\,dx = \underbrace{x^k \frac{d^{n-1}}{dx^{n-1}}(x^2-1)^n\Big|_{-1}^{1}}_{=0} - \int_{-1}^{1} kx^{k-1}\frac{d^{n-1}}{dx^{n-1}}(x^2-1)^n\,dx$$

$$= \cdots = (-1)^k k! \int_{-1}^{1} \frac{d^{n-k}}{dx^{n-k}}(x^2-1)^n\,dx = 0,$$

pa smo dokazali prvu formulu. Za $k = n$, na isti način imamo

$$\int_{-1}^{1} x^n \frac{d^n}{dx^n}(x^2-1)^n\,dx = (-1)^n n! \int_{-1}^{1} (x^2-1)^n\,dx = 2n! \int_{0}^{1} (1-x^2)^n\,dx$$
$$= \{x = \sin t\} = 2n! \int_{0}^{\pi/2} \cos^{2n+1} t\,dt.$$

Za zadnji integral parcijalnom integracijom izlazi

$$\int_{0}^{\pi/2} \cos^{2n+1} t\,dt = \underbrace{\frac{\cos^{2n} t\,\sin t}{2n+1}\Big|_{0}^{\pi/2}}_{=0} + \frac{2n}{2n+1} \int_{0}^{\pi/2} \cos^{2n-1} t\,dt$$
$$= \cdots = \frac{2n(2n-2)\cdots 2}{(2n+1)(2n-1)\cdots 3} \int_{0}^{\pi/2} \cos t\,dt,$$

pa je stoga

$$\int_{-1}^{1} x^n \frac{d^n}{dx^n}(x^2-1)^n\,dx = 2n!\,\frac{2n(2n-2)\cdots 2}{(2n+1)(2n-1)\cdots 3}.$$

Pomnožimo li brojnik i nazivnik s $2n(2n-2)\cdots 2 = 2^n n!$, a zatim, zbog definicije Legendreovog polinoma P_n, sve podijelimo s $2^n n!$, slijedi

$$\int_{-1}^{1} x^n P_n(x)\,dx = \frac{1}{2^n n!}\,2n!\,\frac{2^n n! \cdot 2^n n!}{(2n+1)!} = \frac{2^{n+1}(n!)^2}{(2n+1)!}.$$

■

Lema 3.5.3. *Legendreovi polinomi su ortogonalni na intervalu* $(-1, 1)$ *obzirom na skalarni produkt* (3.5.6)

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

Norma Legendreovog polinoma je

$$\|P_n\|^2 := \int_{-1}^{1} [P_n(x)]^2\,dx = \frac{2}{2n+1}.$$

Dokaz:

Prva tvrdnja je direktna posljedica dokazane ortogonalnosti na potencije nižeg

stupnja. Druga tvrdnja slijedi iz

$$\int_{-1}^{1} [P_n(x)]^2\,dx = \int_{-1}^{1} \left[\frac{1}{2^n n!}\,\frac{(2n)!}{n!}\,x^n + \cdots\right] P_n(x)\,dx.$$

Potencije manje od x^n ne doprinose integralu, pa druga tvrdnja leme 3.5.2. povlači

$$\int_{-1}^{1} [P_n(x)]^2\,dx = \frac{(2n)!}{2^n (n!)^2}\,\frac{2^{n+1}(n!)^2}{(2n+1)!} = \frac{2}{2n+1}.$$

■

Lema 3.5.4. *Legendreovi polinomi P_n imaju n nultočaka, koje su sve realne i različite, i nalaze se u otvorenom intervalu $(-1, 1)$.*

Dokaz:

Dokaz ide iz definicije Legendreovih polinoma

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

induktivnom primjenom Rolleovog teorema. Polinom $(x^2 - 1)^n$ je stupnja $2n$ i ima višestruke (n-terostruke) nultočke u rubovima intervala ± 1. Prema Rolleovom teoremu, prva derivacija ima jednu nultočku u intervalu $(-1, 1)$. Međutim, prva derivacija je, također, nula u ± 1, pa ukupno mora imati tri nultočke u zatvorenom intervalu $[-1, 1]$. Druga derivacija stoga ima dvije unutarnje nule po Rolleovom teoremu, i dvije u ± 1, pa ima ukupno četiri nule u $[-1, 1]$. I tako redom, vidimo da $n - 1$-a derivacija ima $n - 1$ unutarnju nultočku i još dvije u ± 1. Na kraju zaključimo da n-ta derivacija, koja je do na multiplikativni faktor jednaka P_n, ima n unutarnjih nultočaka. ■

Na taj način smo zapravo našli točke integracije u Gauss–Legendreovoj formuli i bez eksplicitnog rješavanja nelinearnog sistema jednadžbi za w_i i x_i, iz uvjeta egzaktne integracije potencija najvećeg mogućeg stupnja. Taj rezultat rezimiran je u sljedećem teoremu.

Teorem 3.5.1. *Čvorovi integracije u Gauss–Legendreovoj formuli reda n su nultočke Legendreovog polinoma P_n, za svaki n.*

Dokaz:

Znamo da su točke integracije x_i nultočke polinoma ω_n po konstrukciji. Zbog uvjeta ortogonalnosti (3.5.3) polinom ω_n, s vodećim koeficijentom 1, proporcionalan je Legendreovom polinomu P_n. Vodeći koeficijent u P_n lako izračunamo iz Rodriguesove formule, odakle je

$$\omega_n(x) = \frac{2^n (n!)^2}{(2n)!}\,P_n(x),$$

pa vidimo da su sve nultočke polinoma ω_n zapravo nultočke od P_n (lema 3.5.4.). ■

Primjer 3.5.2. *Iz Rodriguesove formule možemo izračunati nekoliko prvih Legendreovih polinoma.*

$$
\begin{aligned}
P_0(x) &= 1,\\
P_1(x) &= \frac{1}{2}\,\frac{d}{dx}(x^2-1) = x,\\
P_2(x) &= \frac{1}{8}\,\frac{d^2}{dx^2}(x^2-1)^2 = \frac{1}{2}\,(3x^2-1),\\
P_3(x) &= \frac{1}{48}\,\frac{d^3}{dx^3}(x^2-1)^3 = \frac{1}{2}\,(5x^3-3x),\\
P_4(x) &= \frac{1}{16\cdot 24}\,\frac{d^4}{dx^4}(x^2-1)^4 = \frac{1}{8}\,(35x^4-30x^2+3),\\
P_5(x) &= \frac{1}{8}\,(63x^5-70x^3+15x),\\
P_6(x) &= \frac{1}{16}\,(231x^6-315x^4+105x^2-5),\\
P_7(x) &= \frac{1}{16}\,(429x^7-693x^5+315x^3-35x),\\
P_8(x) &= \frac{1}{128}\,(6435x^8-12012x^6+6930x^4-1260x^2+35).
\end{aligned}
$$

Vidimo, na primjer, da su nultočke od P_3 identične s točkama integracije koje smo dobili u primjeru 3.5.1.*, direktno iz uvjeta ortogonalnosti.*

Računanje nultočaka Legendreovih polinoma (na mašinsku točnost!) nije jednostavan problem, budući da egzaktne formule postoje samo za male stupnjeve. Napomenimo za sad samo toliko, da postoje specijalni algoritmi, te da je dovoljno tabelirati te nultočke jednom, pa brzina algoritma nije važna, nego samo preciznost. Tabelirane nultočke (kao i težine w_i) moguće je naći u gotovo svim standardnim knjigama i tablicama iz područja numeričke analize.

Postoji lakši način za računanje $P_n(x)$, zasnovan na činjenici da Legendreovi polinomi zadovoljavaju tročlanu rekurziju, čiji se koeficijenti mogu eksplicitno izračunati. Ova rekurzivna formula igra važnu ulogu i u konstrukciji spomenutog specijalnog algoritma za traženje nultočaka.

Lema 3.5.5. *Legendreovi polinomi zadovoljavaju rekurzivnu formulu*

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x), \quad n \geq 1,$$

s početnim vrijednostima $P_0(x) = 1$, $P_1(x) = x$.

Dokaz:

Kako je $xP_n(x)$ polinom stupnja $n+1$ i $\{P_i\}_{i=0}^{n+1}$ baza za prostor polinoma stupnja do $n+1$, postoje koeficijenti c_i tako da vrijedi

$$xP_n(x) = \sum_{i=0}^{n+1} c_i P_i(x).$$

Pomnožimo li obje strane s $P_k(x)$ i integriramo od -1 do 1, zbog ortogonalnosti (lema 3.5.3.) slijedi

$$\int_{-1}^{1} xP_k(x)\, P_n(x)\, dx = c_k \int_{-1}^{1} P_k^2(x)\, dx. \tag{3.5.7}$$

Ali za $k < n-1$ je $xP_k(x)$ polinom stupnja manjeg ili jednakog $n-1$, pa je $P_n(x)$ ortogonalan na njega (lema 3.5.2.). Stoga je $c_k = 0$ za $k < n-1$, a u sumi za $xP_n(x)$ ostaju samo zadnja tri člana

$$xP_n(x) = c_{n+1}P_{n+1}(x) + c_nP_n(x) + c_{n-1}P_{n-1}(x). \tag{3.5.8}$$

Treba još izračunati koeficijente c_{n+1}, c_n i c_{n-1}. Kako je

$$P_n(x) = \frac{(2n)!}{2^n(n!)^2}\,\omega_n(x) = \frac{(2n)!}{2^n(n!)^2}\,x^n + \text{niže potencije od } x,$$

usporedimo li koeficijente uz x^{n+1} u (3.5.8), dobivamo da je

$$\frac{(2n)!}{2^n(n!)^2} = c_{n+1}\,\frac{(2n+2)!}{2^{n+1}[(n+1)!]^2},$$

odakle slijedi da je

$$c_{n+1} = \frac{n+1}{2n+1}.$$

Lagano se vidi (iz Rodriguesove formule) da se u Legendreovim polinomima pojavljuju samo alternirajuće potencije, tj. P_{2n} je linearna kombinacija parnih potencija x^{2k}, $k = 0, \ldots, n$, a P_{2n+1} je linearna kombinacija neparnih potencija x^{2k+1}, $k = 0, \ldots, n$. Iz rekurzije (3.5.8) na osnovu toga zaključimo da je $c_n = 0$, pa preostaje samo izračunati c_{n-1}. Za $k = n-1$, iz (3.5.7) imamo da je

$$\int_{-1}^{1} xP_{n-1}(x)\, P_n(x)\, dx = c_{n-1} \int_{-1}^{1} P_{n-1}^2(x)\, dx.$$

Zbog

$$xP_{n-1}(x) = \frac{(2(n-1))!}{2^{n-1}[(n-1)!]^2}\,x^n + \text{niže potencije od } x$$

i ortogonalnosti P_n na sve niže potencije od x, dobivamo

$$\frac{(2n-2)!}{2^{n-1}[(n-1)!]^2}\int_{-1}^{1} x^n P_n(x)\,dx = c_{n-1}\int_{-1}^{1} P_{n-1}^2(x)\,dx.$$

Ovi integrali su poznati (lema 3.5.2. i lema 3.5.3.), pa slijedi

$$c_{n-1} = \frac{n}{2n+1}.$$

Tako smo našli sve nepoznate koeficijente u linearnoj kombinaciji (3.5.8), odakle odmah slijedi tročlana rekurzija. Primijetimo da smo usput dokazali i formulu

$$\int_{-1}^{1} xP_{n-1}(x)\,P_n(x)\,dx = \frac{n}{2n+1}\,\frac{2}{2n-1} = \frac{2n}{4n^2-1}. \tag{3.5.9}$$

■

Zadatak 3.5.2. *Budući da Legendreovi polinomi zadovoljavaju tročlanu rekurziju, moguće je napisati algoritam za brzu sumaciju parcijalnih suma redova oblika*

$$\sum_{n=0}^{\infty} a_n P_n(x),$$

poznat pod nazivom **generalizirana Hornerova shema**. *Koristeći rekurziju iz leme* 3.5.5., *napišite eksplicitno taj algoritam. Razvoji po Legendreovim polinomima pojavljuju se često kod rješavanja Laplaceove jednadžbe u sfernim koordinatama.*

Sljedeće dvije leme korisne su za dobivanje ekplicitnih formula za težine u Gauss–Legendreovim formulama.

Lema 3.5.6. (Christoffel–Darbouxov identitet) *Za Legendreove polinome P_n vrijedi*

$$(t-x)\sum_{k=0}^{n}(2k+1)P_k(x)P_k(t) = (n+1)\,[P_{n+1}(t)P_n(x) - P_n(t)P_{n+1}(x)].$$

Dokaz:

Pomnožimo li rekurziju iz leme 3.5.5. (uz zamjenu $n \mapsto k$) s $P_k(t)$, dobijemo

$$(2k+1)xP_k(x)P_k(t) = (k+1)P_{k+1}(x)P_k(t) + kP_{k-1}(x)P_k(t).$$

Zamijenimo li x i t imamo

$$(2k+1)tP_k(t)P_k(x) = (k+1)P_{k+1}(t)P_k(x) + kP_{k-1}(t)P_k(x).$$

Odbijanjem prve relacije od druge, slijedi

$$(2k+1)(t-x)P_k(x)P_k(t) = (k+1)\left[P_{k+1}(t)P_k(x) - P_k(t)P_{k+1}(x)\right] \\ - k\left[P_k(t)P_{k-1}(x) - P_{k-1}(t)P_k(x)\right].$$

Sumiramo li po k od 1 do n, sukcesivni članovi u sumi na desnoj strani se krate, pa ostaju samo prvi i zadnji

$$(t-x)\sum_{k=1}^{n}(2k+1)P_k(x)P_k(t) = (n+1)\left[P_{n+1}(t)P_n(x) - P_n(t)P_{n+1}(x)\right] - (t-x).$$

Zadnji član možemo prebaciti na lijevu stranu kao nulti član u sumi, a to je baš Christoffel–Darbouxov identitet. ■

Lema 3.5.7. *Derivacija Legendreovih polinoma može se rekurzivno izraziti pomoću samih Legendreovih polinoma, formulom*

$$(1-x^2)P_n'(x) + nxP_n(x) = nP_{n-1}(x), \quad n \geq 1.$$

Dokaz:

Polinom $(1-x^2)P_n' + nxP_n$ je očito stupnja manjeg ili jednakog od $n+1$. Napišimo P_n kao linearnu kombinaciju potencija od x (pojavljuje se samo svaka druga potencija)

$$P_n(x) = a_n x^n + a_{n-2}x^{n-2} + \cdots,$$

pa je

$$P_n'(x) = na_n x^{n-1} + (n-2)a_{n-2}x^{n-3} + \cdots.$$

No, onda je

$$(1-x^2)P_n'(x) + nxP_n(x) = (-na_n + na_n)x^{n+1} + O(x^{n-1}),$$

tj. polinom $(1-x^2)P_n' + nxP_n$ je zapravo stupnja $n-1$. Kao i u dokazu rekurzivne formule, moraju postojati koeficijenti c_i takovi da vrijedi

$$(1-x^2)P_n'(x) + nxP_n(x) = \sum_{i=0}^{n-1} c_i P_i(x).$$

Pomnožimo ovu relaciju s $P_k(x)$ i integriramo od -1 do 1. Zbog ortogonalnosti, na desnoj strani ostaje samo jedan član

$$\frac{2}{2k+1}\,c_k = \int_{-1}^{1}(1-x^2)\,P_n'(x)\,P_k(x)\,dx + n\int_{-1}^{1} xP_n(x)\,P_k(x)\,dx.$$

Prvi integral integriramo parcijalno, pa kako se faktor $(1-x^2)$ poništava na granicama integracije, slijedi

$$\frac{2}{2k+1}\,c_k = -\int_{-1}^{1} P_n(x)\,\frac{d}{dx}[(1-x^2)P_k(x)]\,dx + n\int_{-1}^{1} xP_n(x)\,P_k(x)\,dx.$$

Za $k < n-1$, oba integranda su oblika $P_n(x) \times$ (polinom stupnja najviše $n-1$), pa su svi ovi integrali jednaki nula (lema 3.5.2.), tj. $c_k = 0$ za $k < n-1$. Za $k = n-1$ treba izračunati dva integrala u prethodnoj relaciji. Drugi je jednostavan

$$n\int_{-1}^{1} xP_n(x)\,P_{n-1}(x)\,dx = (3.5.9) = \frac{2n^2}{4n^2-1}.$$

U prvom integralu

$$-\int_{-1}^{1} P_n(x)\,\frac{d}{dx}[(1-x^2)P_{n-1}(x)]\,dx,$$

zbog prve tvrdnje u lemi 3.5.2. (ortgonalnost), doprinos daje samo vodeći član u $(1-x^2)P_{n-1}(x)$, pa je taj integral jednak

$$\int_{-1}^{1} P_n(x)\,\frac{d}{dx}\left\{x^2\,\frac{(2n-2)!}{2^{n-1}[(n-1)!]^2}\,x^{n-1}\right\}dx,$$

a zbog druge tvrdnje u lemi, integral se svodi na

$$\frac{(2n-2)!}{2^{n-1}[(n-1)!]^2}\,(n+1)\,\frac{2^{n+1}(n!)^2}{(2n+1)!} = \frac{2n(n+1)}{(2n+1)(2n-1)}.$$

Na kraju je

$$c_{n-1} = \frac{2n-1}{2}\left[\frac{2n(n+1)}{(2n+1)(2n-1)} + \frac{2n^2}{(2n+1)(2n-1)}\right] = n,$$

što smo i htjeli dokazati. ■

Lema 3.5.8. *Težinski faktori u Gauss–Legendreovim formulama mogu se eksplicitno izračunati formulama*

$$w_i = \frac{2(1-x_i^2)}{n^2[P_{n-1}(x_i)]^2},$$

gdje su x_i, $i = 0, \ldots, n$, *nultočke Legendreovog polinoma* P_n.

Dokaz:

Neka je x_i nultočka polinoma P_n. Stavimo li $t = x_i$ u Christoffel–Darbouxov identitet (lema 3.5.6.), dobivamo

$$\frac{(n+1)P_{n+1}(x_i)P_n(x)}{x - x_i} = -\sum_{k=0}^{n}(2k+1)P_k(x)P_k(x_i).$$

Kad integriramo ovu jednakost od -1 do 1 i uzmemo u obzir da je k-ti Legendreov polinom ortogonalan na konstantu $P_k(x_i)$, na desnoj strani preostane samo član za $k = 0$

$$\int_{-1}^{1} \frac{P_n(x)}{(x - x_i)}\,dx = \frac{-2}{(n+1)P_{n+1}(x_i)}.$$

Tročlana rekurzija iz leme 3.5.5. u nultočki x_i Legendreovog polinoma P_n ima oblik $(n+1)P_{n+1}(x_i) = -nP_{n-1}(x_i)$, pa je stoga

$$\int_{-1}^{1} \frac{P_n(x)}{(x - x_i)}\,dx = \frac{2}{nP_{n-1}(x_i)}.$$

Za težinske koeficijente w_i vrijede relacije (3.5.5) i (3.5.4)

$$w_i = \int_{-1}^{1} \ell_i(x)\,dx = \int_{-1}^{1} \frac{\omega_n(x)}{\omega_n'(x_i)(x - x_i)}\,dx = \int_{-1}^{1} \frac{P_n(x)}{P_n'(x_i)(x - x_i)}\,dx,$$

pa je dakle

$$w_i = \frac{2}{nP_n'(x_i)\,P_{n-1}(x_i)}. \tag{3.5.10}$$

Primijetimo da je Christoffel–Darbouxov identitet potreban jedino zato da se izračuna neugodan integral

$$\int_{-1}^{1} \frac{P_n(x)}{(x - x_i)}\,dx,$$

u kojem podintegralna funkcija ima uklonjivi singularitet.

Na kraju, iskoristimo rekurzivnu formulu za derivacije Legendreovog polinoma iz leme 3.5.7. u specijalnom slučaju kada je $x = x_i$. Dobivamo da vrijedi

$$(1 - x_i^2)P_n'(x_i) = nP_{n-1}(x_i).$$

Uvrstimo li taj rezultat u (3.5.10), tvrdnja slijedi. ■

U dokazu prethodne leme 3.5.8. pokazali smo (usput) da u nultočki x_i Legendreovog polinoma P_n vrijedi

$$(1 - x_i^2)P_n'(x_i) = nP_{n-1}(x_i) = -(n+1)P_{n+1}(x_i).$$

Ovu relaciju možemo iskoristiti na različite načine u (3.5.10), što daje pet raznih formula za težinske koeficijente u Gauss–Legendreovim formulama

$$\begin{aligned} w_i &= \frac{2(1-x_i^2)}{[nP_{n-1}(x_i)]^2} = \frac{2(1-x_i^2)}{[(n+1)P_{n+1}(x_i)]^2} \\ &= \frac{2}{nP_n'(x_i)\,P_{n-1}(x_i)} = -\frac{2}{(n+1)P_n'(x_i)\,P_{n+1}(x_i)} \\ &= \frac{2}{(1-x_i^2)\,[P_n'(x_i)]^2}. \end{aligned} \tag{3.5.11}$$

Sljedeći teorem rezimira prethodne rezultate, i ujedno daje ocjenu greške za Gauss–Legendreovu integraciju.

Teorem 3.5.2. *Za funkciju $f \in C^{2n}[-1,1]$ Gauss–Legendreova formula integracije glasi*

$$\int_{-1}^{1} f(x)\,dx = \sum_{i=1}^{n} w_i f(x_i) + E_n(f),$$

gdje su x_i nultočke Legendreovog polinoma P_n i koeficijenti w_i dani u (3.5.11). *Za grešku $E_n(f)$ vrijedi*

$$E_n(f) = \frac{2^{2n+1}(n!)^4}{(2n+1)\,[(2n)!]^3}\, f^{(2n)}(\xi), \quad \xi \in (-1,1).$$

Dokaz:

Treba samo dokazati formulu za ocjenu greške. Kako je Gauss–Legendreova formula zapravo integral Hermiteovog interpolacijskog polinoma, treba integrirati grešku kod Hermiteove interpolacije, koju smo procijenili u teoremu 1–7.2.5., i uvrstiti odgovarajući ω_n. Integracijom i primjenom teorema srednje vrijednosti za integrale, dobivamo

$$E_n(f) = \frac{f^{(2n)}(\xi)}{(2n)!} \int_{-1}^{1} \omega_n^2(x)\,dx,$$

za neki $\xi \in (-1,1)$. Kako je

$$\omega_n(x) = \frac{2^n(n!)^2}{(2n)!}\, P_n(x),$$

zbog poznatog kvadrata norme Legendreovog polinoma (lema 3.5.3.), imamo

$$E_n(f) = \frac{f^{(2n)}(\xi)}{(2n)!} \left[\frac{2^n(n!)^2}{(2n)!}\right]^2 \frac{2}{2n+1} = \frac{2^{2n+1}(n!)^4}{(2n+1)\,[(2n)!]^3}\, f^{(2n)}(\xi).$$

■

Navedeni izraz za grešku nije lagano primijeniti, budući da je potrebno naći neku ogradu za vrlo visoku derivaciju funkcije f (red derivacije je dva puta veći nego kod Newton–Côtesovih formula). Član uz $f^{(2n)}(\xi)$ vrlo brzo pada s porastom n. Na primjer, za $n = 6$, greška je oblika

$$1.6 \cdot 10^{-12}\, f^{(12)}(\xi).$$

Da ocjena greške za Gaussove formule može biti previše pesimistična, pokazuje sljedeći primjer.

Primjer 3.5.3. *Primijenimo Gauss–Legendreovu formulu na integral*

$$\int_0^{\pi/2} \log(1+t)\,dt = \left(1+\frac{\pi}{2}\right)\left[\log\left(1+\frac{\pi}{2}\right)-1\right]+1.$$

Zamjena varijable $t = \pi(x+1)/4$ *prebacuje integral na standardnu formu*

$$\int_{-1}^{1} \frac{\pi}{4}\,\log\left(1+\frac{\pi(x+1)}{4}\right)dx.$$

U ovom slučaju možemo lagano izračunati bilo koju derivaciju podintegralne funkcije, koja raste s faktorijelima. Zapravo, sve ocjene greške formula za numeričku integraciju pokazuju slično ponašanje (usporedite, na primjer, trapeznu i Simpsonovu formulu), ali Gaussove formule naročito, budući da uključuju visoke derivacije. Tako je, na primjer, osma derivacija, koja je potrebna za Gaussovu formulu s četiri točke jednaka

$$\left(\frac{\pi}{4}\right)^9 \cdot \frac{-7!}{(1+t)^8},$$

pa je greška 7! *puta veća nego da smo, recimo, integrirali trigonometrijsku funkciju* sin *ili* cos*, koje imaju ograničene derivacije. Ipak, lagano vidimo da već sa šest točaka dobivamo* 6 *znamenaka točno, iako ocjena greške uključuje faktor od* 11!*. Simpsonovoj formuli treba* 64 *točke za istu točnost. Možemo slutiti, da je za analitičke funkcije moguća bolja ocjena greške.*

Korolar 3.5.1. (Uvjeti egzaktnosti) *Gauss–Legendreova formula egzaktno integrira polinome stupnja* $2n-1$.

Dokaz:

Očito, budući da se greška, koja uključuje $2n$-tu derivaciju, poništava na takvim polinomima. ■

Svojstvo iz gornjeg korolara može se upotrijebiti za alternativni dokaz teorema 3.5.2., kao što smo napomenuli na početku. Hermiteova interpolacija poslužila

je kao "trik", da izbjegnemo rješavanje nelinearnog sistema koji proizilazi iz uvjeta egzaktnosti.

Rekurziju za derivacije Legendreovih polinoma iz leme 3.5.7. možemo koristiti i za računanje vrijednosti $P_n'(x)$

$$(1-x^2)P_n'(x) = n(P_{n-1}(x) - xP_n(x)), \quad n \geq 1.$$

Nažalost, ova formulu ne možemo upotrijebiti u rubnim točakama $x = \pm 1$, zbog dijeljenja s nulom. Međutim, Legendreovi polinomi zadovoljavaju i mnoge druge rekurzivne relacije. Neke od njih dane su u sljedećem zadatku.

Zadatak 3.5.3. *Dokažite da za Legendreove polinoma vrijedi $P_n(1) = 1$, za $n \geq 0$, što opravdava izbor normalizacije. Također, dokažite da za $n \geq 1$ vrijede rekurzivne relacije*

$$\begin{aligned} P_n'(x) - xP_{n-1}'(x) &= nP_{n-1}(x), \\ xP_n'(x) - P_{n-1}'(x) &= nP_n(x), \\ P_{n+1}'(x) - P_{n-1}'(x) &= (2n+1)P_n(x) \\ \int_{-1}^{x} P_n(t)\,dt &= \frac{1}{2n+1}\left(P_{n+1}(x) - P_{n-1}(x)\right). \end{aligned}$$

Na kraju, primijetimo da Gaussove formule možemo shvatiti i kao rješenje optimizacijskog problema: naći točke integracije tako da egzaktno integriramo polinom što većeg stupnja sa što manje čvorova. Rezultat su formule visoke točnosti, koje se lagano implementiraju, i imaju vrlo mali broj izvrednjavanja podintegralne funkcije. Cijenu smo platili time što ocjena greške zahtijeva vrlo glatku funkciju, ali također i time što upotreba takvih formula na "finijoj" mreži zahtijeva ponovno računanje funkcije u drugim čvorovima, koji s čvorovima formule nižeg reda nemaju ništa zajedničko. Kod profinjavanja mreže čvorova za formule Newton–Côtesovog tipa (na primjer, raspolavljanjem h), naprotiv, jedan dio čvorova ostaje zajednički, pa već izračunate funkcijske vrijednosti možemo iskoristiti (kao u Rombergovom algoritmu).

3.5.2. Druge Gaussove integracione formule

U praksi se često javljaju specijalni integrali koji uključuju težinske funkcije poput e^{-x}, e^{-x^2} i mnoge druge, na specijalnim intervalima, često neograničenim. Jednostavnom linearnom supstitucijom nije moguće takve intervale i/ili težinske funkcije prebaciti na interval $(-1, 1)$ i jediničnu težinsku funkciju — situaciju u kojoj možemo primijeniti Gauss–Legendreove formule.

Alternativa je iskoristiti odgovarajuće Gaussove formule s "prirodnom" težinskom funkcijom. Iz prethodnog odjeljka znamo da za čvorove integracije treba uzeti

nultočke funkcije $\omega_n(x) = (x - x_1) \cdots (x - x_n)$, s tim da vrijede relacije ortogonalnosti (3.5.3). Težine w_i onda možemo odrediti rješavanjem linearnog sistema, a možda u specijalnim slučajevima možemo doći i do eksplicitnih formula, kao što smo to učinili u slučaju Gauss–Legendreovih formula. Postavlja se pitanje da li možemo doći do formula za polinome koji su ortogonalni (obzirom na težinsku funkciju w) na polinome nižeg stupnja, uključivo i ostale formule na koje smo se oslanjali, poput tročlane rekurzije i slično (v. lema 3.5.2.).

U mnogim važnim slučajevima, ali ne i uvijek, moguće je analitički doći do formula sličnim onima u slučaju Gauss–Legendreove integracije. U drugim slučajevima, koji nisu pokriveni egzaktnim formulama, u principu je moguće generirati ortogonalne polinome i numerički. Poznati postupci (Stieltjesov i Čebiševljev algoritam) ne pokrivaju, međutim, sve moguće situacije, tj. nisu uvijek numerički stabilni, što ostavlja postora za daljnja istraživanja. Slučajevi tzv. **klasičnih ortogonalnih polinoma** uglavnom se mogu karakterizirati na osnovu sljedeća dva teorema, od kojih je prvi egzistencijalni, i vezan uz teoriju rubnih problema za obične diferencijalne jednadžbe.

Teorem 3.5.3. (Generalizirana Rodriguesova formula)

Na otvorenom intervalu (a, b) postoji, do na multiplikativnu konstantu, jedinstvena funkcija $U_n(x)$ koja zadovoljava diferencijalnu jednadžbu

$$D^{n+1}\left(\frac{1}{w(x)}\, D^n U_n(x)\right) = 0$$

i rubne uvjete

$$U_n(a) = DU_n(a) = \cdots = D^{n-1}U_n(a) = 0,$$
$$U_n(b) = DU_n(b) = \cdots = D^{n-1}U_n(b) = 0.$$

Ovdje opet koristimo oznaku D za operator deriviranja funkcije f jedne varijable, kad je iz konteksta očito po kojoj varijabli se derivira, jer ta oznaka znatno skraćuje zapis nekih dugih formula. Onda n-tu derivaciju funkcije f u točki x možemo pisati u bilo kojem od sljedeća tri oblika

$$D^n f(x) = \frac{d^n}{dx^n}\, f(x) = f^{(n)}(x).$$

Budući da nas interesiraju rješenja koja se mogu eksplicitno konstruirati, nećemo dokazivati ovaj teorem. U svakom konkretnom slučaju, za zadane a, b i $w(x)$, konstruirat ćemo funkciju U_n formulom. Napomenimo još da teorem 3.5.3. vrijedi i na neograničenim i poluograničenim intervalima, tj. u slučajevima $a = -\infty$ i/ili $b = \infty$.

Funkcije U_n iz prethodnog teorema generiraju familiju ortogonalnih polinoma na (a, b) s težinskom funkcijom w.

Teorem 3.5.4. *Uz pretpostavke teorema* 3.5.3., *funkcije*

$$p_n(x) = \frac{1}{w(x)}\, D^n U_n(x)$$

su polinomi stupnja n koji su ortogonalni na sve polinome nižeg stupnja na intervalu (a, b) *obzirom na težinsku funkciju* $w(x)$, *tj. vrijedi*

$$\int_a^b w(x)\, p_n(x)\, x^k\, dx = 0, \quad \text{za} \quad k = 0, 1, \ldots, n-1.$$

Dokaz:

Funkcija p_n je očito polinom stupnja n, jer je $D^{n+1} p_n(x) = 0$. Da dokažemo ortogonalnost, pretpostavimo da je $n \geq 1$. Za $k = 0$ imamo odmah po Newton–Lebnitzovoj formuli

$$\int_a^b w(x)\, p_n(x)\, dx = \int_a^b D^n U_n(x)\, dx = (n \geq 1) = D^{n-1} U_n(x)\Big|_a^b = 0,$$

zbog rubnih uvjeta $D^{n-1} U_n(a) = D^{n-1} U_n(b) = 0$.

Za $1 \leq k \leq n-1$, integriramo parcijalno k puta i iskoristimo opet rubne uvjete koje zadovoljava funkcija U_n. Dobivamo redom

$$\begin{aligned}
\int_a^b w(x) p_n(x)\, x^k\, dx &= \int_a^b x^k\, D^n U_n(x)\, dx \\
&= \underbrace{x^k\, D^{n-1} U_n(x)\Big|_a^b}_{=0} - k \int_a^b x^{k-1}\, D^{n-1} U_n(x)\, dx \\
&= -k \left(\underbrace{x^{k-1}\, D^{n-2} U_n(x)\Big|_a^b}_{=0} - (k-1) \int_a^b x^{k-2}\, D^{n-2} U_n(x)\, dx \right) \\
&= \cdots = (-1)^{k-1} k(k-1) \cdots 2 \left(\underbrace{x\, D^{n-k} U_n(x)\Big|_a^b}_{=0} - \int_a^b D^{n-k} U_n(x)\, dx \right) \\
&= (-1)^k k(k-1) \cdots 2 \cdot 1 \left(\underbrace{D^{n-k-1} U_n(x)\Big|_a^b}_{=0} \right) = 0,
\end{aligned}$$

jer je $n-k-1 \geq 0$. Primijetimo da smo za dokaz ortogonalnosti iskoristili sve rubne uvjete na funkciju U_n. ■

Ovaj teorem u mnogim slučajevima omogućava efektivnu konstrukciju ortogonalnih polinoma.

Primjer 3.5.4. *Neka je $w(x) = 1$ na intervalu $(-1, 1)$. Nađimo pripadne ortogonalne polinome. Prema teoremu* 3.5.3.*, prvi korak je rješavanje diferencijalne jednadžbe*

$$D^{n+1}(D^n U_n(x)) = D^{2n+1} U_n(x) = 0,$$

uz rubne uvjete

$$U_n(\pm 1) = DU_n(\pm 1) = \cdots = D^{n-1} U_n(\pm 1) = 0.$$

Polinom $2n$-tog stupnja koji se poništava u krajevima mora, zbog simetrije, biti oblika $U_n(x) = C_n(x^2 - 1)^n$, gdje je C_n proizvoljna multiplikativna konstanta (različita od nule). Tradicionalno, konstanta C_n uzima se u obliku

$$C_n = \frac{1}{2^n n!}.$$

Pripadni ortogonalni polinomi su tada, prema teoremu 3.5.4.*, dani formulom*

$$P_n(x) = \frac{1}{2^n n!} D^n (x^2 - 1)^n,$$

tj. dobivamo, očekivano, Legendreove polinome.

Zadatak 3.5.4. *Pokažite da je multiplikativna konstanta C_n odabrana tako da vrijedi $P_n(1) = 1$, za svako n. Također, pokažite da vrijedi $|P_n(x)| \leq 1$, za svaki $x \in [-1, 1]$ i svaki $n \geq 0$. To znači da P_n dostiže ekstreme u rubovima intervala, što je dodatno opravdanje za izbor normalizacije, jer je $\|P_n\|_\infty = 1$ na $[-1, 1]$.*

Primjer 3.5.5. *Neka je $w(x) = e^{-\alpha x}$ na intervalu $(0, \infty)$, za neki $\alpha > 0$. Nađimo pripadne ortogonalne polinome. Prema teoremu* 3.5.3.*, trebamo prvo riješiti diferencijalnu jednadžbu*

$$D^{n+1}(e^{\alpha x} D^n U_n(x)) = 0,$$

uz rubne uvjete

$$U_n(0) = DU_n(0) = \cdots = D^{n-1} U_n(0) = 0,$$
$$U_n(\infty) = DU_n(\infty) = \cdots = D^{n-1} U_n(\infty) = 0.$$

Očito je rješenje oblika

$$U_n(x) = e^{-\alpha x}(c_0 + c_1 x + \cdots + c_n x^n) + d_0 + d_1 x + \cdots + d_{n-1} x^{n-1}.$$

Rubni uvjet u točki ∞ povlači $d_0 = \cdots = d_{n-1} = 0$, a rubni uvjet u točki 0 povlači $c_0 = \cdots = c_{n-1} = 0$, pa je

$$U_n(x) = C_n x^n e^{-\alpha x}.$$

Polinomi za koje je $\alpha = 1$ i $C_n = 1$ zovu se tradicionalno **Laguerreovi polinomi***, u oznaci $\widetilde{L}_n$. Njihova Rodriguesova formula je dakle*

$$\widetilde{L}_n(x) = e^x D^n (x^n e^{-x}).$$

U općem slučaju, za $\alpha \neq 1$, uz $C_n = 1$, lagano vidimo da je $p_n(x) = \widetilde{L}_n(\alpha x)$. Tada vrijede relacije ortogonalnosti

$$\int_0^\infty e^{-\alpha x}\, \widetilde{L}_m(\alpha x)\, \widetilde{L}_n(\alpha x)\, dx = 0, \quad m \neq n.$$

Napomenimo još da oznaku L_n koristimo za ortonormirane Laguerreove polinome. Njih dobivamo izborom normalizacione konstante $C_n = 1/n!$, pa je $\widetilde{L}_n(x) = n!\, L_n(x)$.

Primjer 3.5.6. *Neka je $w(x) = e^{-\alpha^2 x^2}$ na intervalu $(-\infty, \infty)$, za neki $\alpha \neq 0$. Nađimo pripadne ortogonalne polinome. Prema teoremu* 3.5.3.*, trebamo prvo riješiti diferencijalnu jednadžbu*

$$D^{n+1}(e^{\alpha^2 x^2}\, D^n U_n(x)) = 0,$$

uz rubne uvjete

$$U_n(\pm\infty) = DU_n(\pm\infty) = \cdots = D^{n-1}U_n(\pm\infty) = 0.$$

Lagano pogodimo da je

$$U_n(x) = C_n\, e^{-\alpha^2 x^2}.$$

Odaberemo li $\alpha^2 = 1$ i multiplikativnu konstantu $C_n = (-1)^n$, dolazimo do klasičnih polinoma, koji nose ime **Hermiteovi polinomi***, u oznaci H_n, s Rodriguesovom formulom*

$$H_n(x) = (-1)^n e^{x^2}\, D^n(e^{-x^2}).$$

U općem slučaju, za $\alpha^2 \neq 1$, uz $C_n = (-\alpha)^n$, lagano vidimo da su polinomi koje tražimo oblika

$$p_n(x) = H_n(\alpha x) = (-\alpha)^n\, e^{\alpha^2 x^2}\, D^n(e^{-\alpha^2 x^2}).$$

Pripadne relacije ortogonalnosti su

$$\int_{-\infty}^\infty e^{-\alpha^2 x^2}\, H_m(\alpha x)\, H_n(\alpha x)\, dx = 0, \quad m \neq n.$$

U literaturi se ponekad može naći još jedna definicija za klasične Hermiteove polinome, koja odgovara izboru $\alpha^2 = 1/2$, uz $C_n = (-1)^n$.

Svi ortogonalni polinomi zadovoljavaju tročlane rekuzije (v. izvod Stieltjesovog algoritma uz metodu najmanjih kvadrata). Za Laguerreove i Hermiteove polinome mogu se analitički izračunati koeficijenti u rekurziji, postupkom koji je vrlo sličan onom kojeg smo u detalje proveli u slučaju Legendreovih polinoma. Primijetimo, također, da i Čebiševljevi polinomi prve vrste zadovoljavaju relacije ortogonalnosti i tročlanu rekurziju, i da smo taj slučaj do kraja proučili. Kako su čvorovi Gaussove

formule integracije reda n nultočke odgovarajućeg ortogonalnog polinoma p_n, preostaje još samo izračunati težine w_i po formuli (3.5.5). Sasvim općenito, može se pokazati da vrijedi

$$w_i = \int_a^b w(x)\,\ell_i(x)\,dx = \frac{1}{p_n'(x_i)} \int_a^b w(x)\,\frac{p_n(x)}{x - x_i}\,dx,$$

gdje su ℓ_i polinomi Lagrangeove baze, i te integrale treba naći egzaktno. Formule za težine mogu se dobiti za cijeli niz klasičnih ortogonalnih polinoma, ali njihovo računanje ovisi o specijalnim svojstvima, posebnim rekurzijama i identitetima oblika Christoffel–Darbouxovog. Obzirom na duljinu ovih izvoda, zadovoljimo se s kratkim opisom nekoliko najpoznatijih Gaussovih formula.

Gauss–Laguerreove formule

Formule oblika

$$\int_0^\infty e^{-x}\,f(x)\,dx \approx \sum_{i=1}^n w_i f(x_i)$$

zovu se **Gauss–Laguerreove formule**. Čvorovi integracije su nultočke polinoma $\widetilde{L}_n$ definiranih Rodriguesovom formulom

$$\widetilde{L}_n(x) = e^x\,\frac{d^n}{dx^n}(x^n e^{-x}),$$

a težine u Gaussovoj formuli su

$$\begin{aligned} w_i &= \frac{[(n-1)!]^2 x_i}{[n\widetilde{L}_{n-1}(x_i)]^2} = \frac{(n!)^2 x_i}{[\widetilde{L}_{n+1}(x_i)]^2} \\ &= -\frac{[(n-1)!]^2}{\widetilde{L}_n'(x_i)\,\widetilde{L}_{n-1}(x_i)} = \frac{(n!)^2}{\widetilde{L}_n'(x_i)\,\widetilde{L}_{n+1}(x_i)} \\ &= \frac{(n!)^2}{x_i[\widetilde{L}_n'(x_i)]^2}. \end{aligned}$$

Greška kod numeričke integracije dana je formulom

$$E_n(f) = \frac{(n!)^2}{(2n)!}\,f^{(2n)}(\xi), \quad \xi \in (0,\infty).$$

Gauss–Hermiteove formule

Formule oblika

$$\int_{-\infty}^\infty e^{-x^2}\,f(x)\,dx \approx \sum_{i=1}^n w_i f(x_i)$$

zovu se **Gauss–Hermiteove formule**. Čvorovi integracije su nultočke polinoma H_n definiranih Rodriguesovom formulom

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n}(e^{-x^2}),$$

a težine u Gaussovoj formuli su

$$\begin{aligned} w_i &= \frac{2^{n-1}(n-1)!\sqrt{\pi}}{n[H_{n-1}(x_i)]^2} = \frac{2^{n+1}n!\sqrt{\pi}}{[H_{n+1}(x_i)]^2} \\ &= \frac{2^n(n-1)!\sqrt{\pi}}{H_n'(x_i)\,H_{n-1}(x_i)} = -\frac{2^{n+1}n!\sqrt{\pi}}{H_n'(x_i)\,H_{n+1}(x_i)} \\ &= \frac{2^{n+1}n!\sqrt{\pi}}{[H_n'(x_i)]^2}. \end{aligned}$$

Greška kod numeričke integracije dana je formulom

$$E_n(f) = \frac{n!\sqrt{\pi}}{2^n(2n)!} f^{(2n)}(\xi), \quad \xi \in (-\infty, \infty).$$

Gauss–Čebiševljeve formule

Formule oblika

$$\int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} f(x)\,dx \approx \sum_{i=1}^{n} w_i f(x_i)$$

zovu se **Gauss–Čebiševljeve formule**. Čvorovi integracije su nultočke Čebiševljevih polinoma $T_n(x) = \cos(n \arccos(x))$. Izuzetno, te se nultočke mogu eksplicitno izračunati

$$x_i = \cos\left(\frac{(2i-1)\pi}{2n}\right).$$

Sve težine w_i su jednake

$$w_i = \frac{\pi}{n}.$$

Greška kod numeričke integracije dana je formulom

$$E_n(f) = \frac{\pi}{2^{2n-1}(2n)!} f^{(2n)}(\xi), \quad \xi \in (-1, 1).$$

Zadatak 3.5.5. *Neka je težinska funkcija $w(x) = (x-a)^\alpha (b-x)^\beta$ na intervalu (a, b), gdje su $\alpha > -1$ i $\beta > -1$. Nađite funkciju U_r i napišite Rodriguesovu formulu! Pridruženi ortogonalni polinomi zovu se* **Jacobijevi polinomi**. *Legendreovi i Čebiševljevi polinomi specijalni su slučaj.*

Pomoću Gaussovih formula možemo jednostavno računati neke određene integrale analitički, kao što se vidi iz sljedećih primjera.

Primjer 3.5.7. *Ako Gauss–Laguerreovom formulom reda $n = 1$ računamo integral*

$$\int_0^\infty e^{-x}\,dx,$$

imamo približnu formulu

$$\int_0^\infty e^{-x}\,f(x)\,dx \approx f(1),$$

budući da je $\widetilde{L}_1(x) = 1 - x$, pa je $x_1 = 1$ i $w_1 = 1/[\widetilde{L}'(1)]^2 = 1$. Kako formula egzaktno integrira konstante, za $f(x) = 1$ imamo

$$\int_0^\infty e^{-x}\,dx = f(1) = 1.$$

Slično, za $f(x) = ax + b$, budući da formula egzaktno integrira i linearne funkcije,

$$\int_0^\infty e^{-x}\,(ax + b)\,dx = f(1) = a + b.$$

Primjer 3.5.8. *Ako Gauss–Čebiševljevom formulom računamo*

$$\int_{-1}^{1} \frac{x^4}{\sqrt{1 - x^2}}\,dx$$

zgodno je upotrijebiti formulu Gauss–Čebiševa reda 3, koja zahtijeva nultočke polinoma $T_3(x) = 4x^3 - 3x$, a to su $x_1 = 0$, $x_{2,3} = \pm\sqrt{3}/2$. Formula vodi na egzaktan rezultat

$$\int_{-1}^{1} \frac{x^4}{\sqrt{1 - x^2}}\,dx = \frac{\pi}{3}\left(0 + \frac{9}{16} + \frac{9}{16}\right) = \frac{3\pi}{8}.$$

4. Rješavanje nelinearnih jednadžbi

4.1. Općenito o iterativnim metodama

Računanje nultočaka nelinearnih funkcija jedan je od najčešćih zadataka primijenjene matematike. Općenito, neka je zadana funkcija

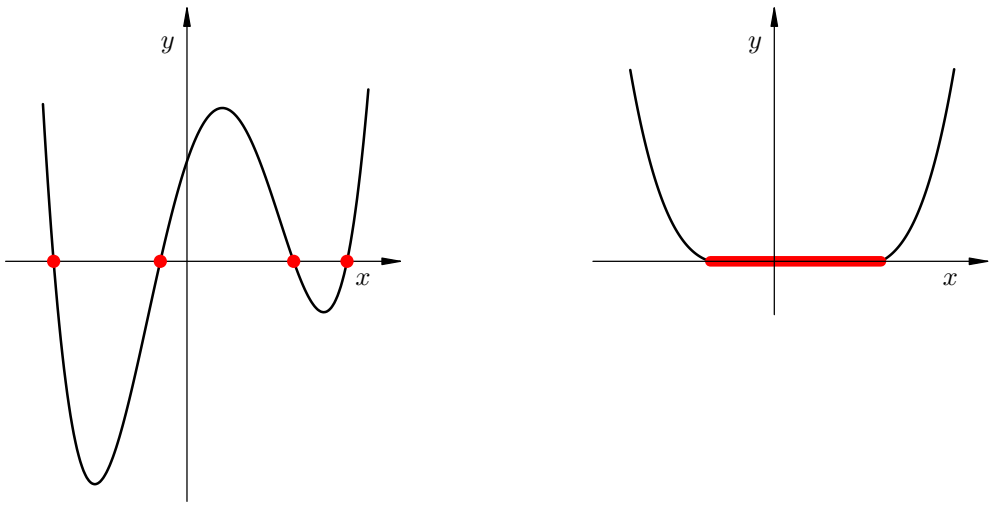
$$f : I \to \mathbb{R},$$

gdje je I neki interval. Tražimo sve one $x \in I$ za koje je

$$f(x) = 0.$$

Takvi x-evi zovu se rješenja, korijeni pripadne jednadžbe ili nultočke funkcije f.

U pravilu, pretpostavljamo da je f **neprekidna** na I i da su joj nultočke izolirane. U protivnom postojao bi problem konvergencije.



Traženje nultočki na zadanu točnost sastoji se od dvije faze.

1. Izolacije jedne ili više nultočki, tj. nalaženje intervala I unutar kojeg se nalazi bar jedna nultočka. Ovo je teži dio posla i obavlja se na temelju analize toka funkcije.
2. Iterativno nalaženje nultočke na traženu točnost.

Postoji mnogo metoda za nalaženje nultočaka nelinearnih funkcija na zadanu točnost. One se bitno razlikuju po tome hoće li uvijek konvergirati, tj. imamo li sigurnu konvergenciju ili ne i po brzini konvergencije.

Uobičajen je slučaj da brze metode nemaju sigurnu konvergenciju, dok je sporije metode imaju.

Definirajmo brzinu konvergencije metode

Definicija 4.1.1. *Niz iteracija* $(x_n, n \in \mathbb{N}_0)$ *konvergira prema točki* α *s redom konvergencije* p, $p \geq 1$ *ako vrijedi*

$$|\alpha - x_n| \leq c\,|\alpha - x_{n-1}|^p, \quad n \in \mathbb{N} \tag{4.1.1}$$

za neki $c > 0$. *Ako je* $p = 1$, *kažemo da niz konvergira linearno prema* α. *U tom je slučaju nužno da je* $c < 1$ *i obično se* c *naziva faktor linearne konvergencije.*

Relacija (4.1.1) katkad nije zgodna za linearne iterativne algoritme. Ako u (4.1.1) upotrijebimo indukciju za $p = 1$, $c < 1$, onda dobivamo da je

$$|\alpha - x_n| \leq c^n |\alpha - x_0|, \quad n \in \mathbb{N}. \tag{4.1.2}$$

Katkad će biti mnogo lakše pokazati (4.1.2) nego (4.1.1). I u slučaju (4.1.2), reći ćemo da niz iteracija konvergira linearno s faktorom c.

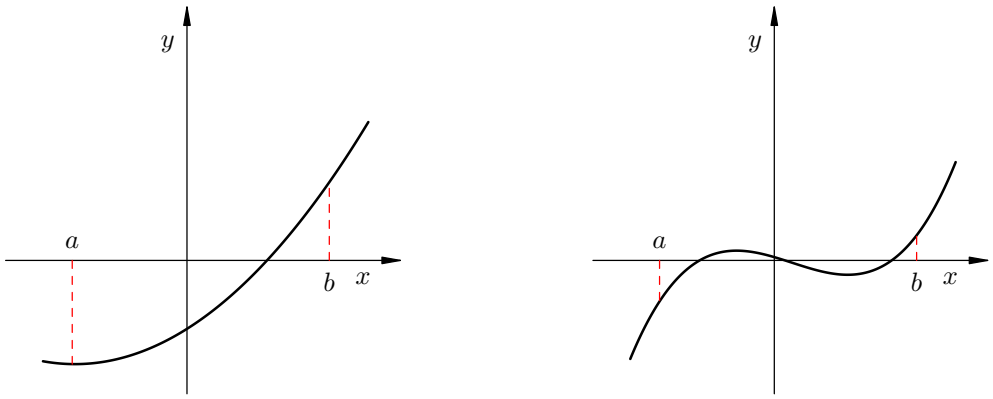
4.2. Metoda raspolavljanja (bisekcije)

Najjednostavnija metoda nalaženja nultočaka funkcije je metoda raspolavljanja. Ona funkcionira za neprekidne funkcije, ali zbog toga ima i najlošiju ocjenu pogreške.

Osnovna pretpostavka za početak algoritma raspolavljanja je **neprekidnost** funkcije f na intervalu $[a, b]$ uz pretpostavku da vrijedi

$$f(a) \cdot f(b) < 0.$$

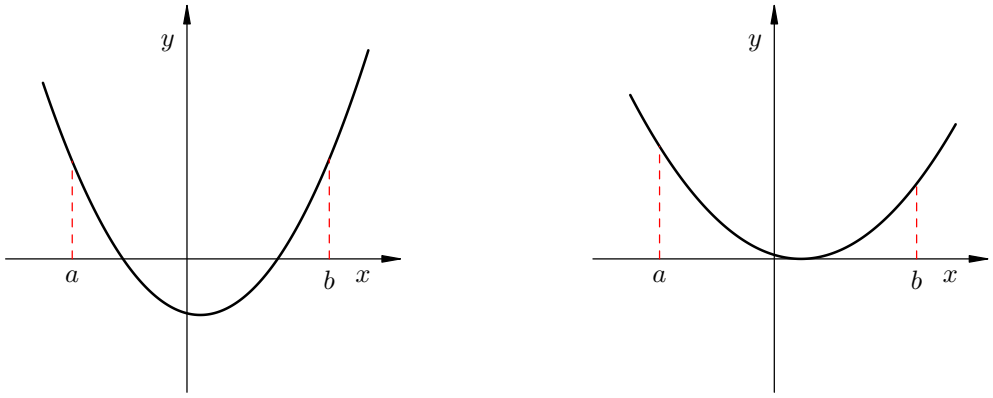
Prethodna relacija znači da funkcija f ima na intervalu $[a, b]$ **bar jednu** nultočku.



Obratno, ako je

$$f(a) \cdot f(b) > 0,$$

to **ne mora** značiti da f nema unutar $[a, b]$ nultočku. Na primjer, moglo se dogoditi da smo loše separirali nultočke i da f ima unutar $[a, b]$ paran broj nultočaka, ili nultočku parnog reda.



Dok je za prvi primjer s prethodne slike lako, boljom separacijom nultočki postići $f(a) \cdot f(b) < 0$, za drugi je primjer to nemoguće! Dakle, nultočke parnog reda nemoguće je naći metodom bisekcije.

Ako vrijede startne pretpostavke metode, metoda raspolavljanja konvergirat će prema nekoj nultočki iz intervala $[a, b]$.

Algoritam raspolavljanja je vrlo jednostavan. Označimo s α pravu nultočku funkcije, a zatim s $a_0 := a$, $b_0 := b$ i x_0 polovište $[a_0, b_0]$, tj.

$$x_0 = \frac{a_0 + b_0}{2}.$$

U n-tom koraku algoritma konstruiramo interval $[a_n, b_n]$ kojemu je duljina polovina duljine prethodnog intervala, ali tako da je nultočka ostala unutar intervala $[a_n, b_n]$.

Konstrukcija intervala $[a_n, b_n]$ sastoji se u raspolavljanju intervala $[a_{n-1}, b_{n-1}]$ točkom x_{n-1} i to tako da je

$$\begin{aligned} a_n = x_{n-1}, b_n = b_{n-1} \quad &\text{ako je} \quad f(a_{n-1}) \cdot f(x_{n-1}) > 0, \\ a_n = a_{n-1}, b_n = x_{n-1} \quad &\text{ako je} \quad f(a_{n-1}) \cdot f(x_{n-1}) < 0. \end{aligned}$$

Postupak zaustavljamo kad je

$$|\alpha - x_n| \leq \varepsilon.$$

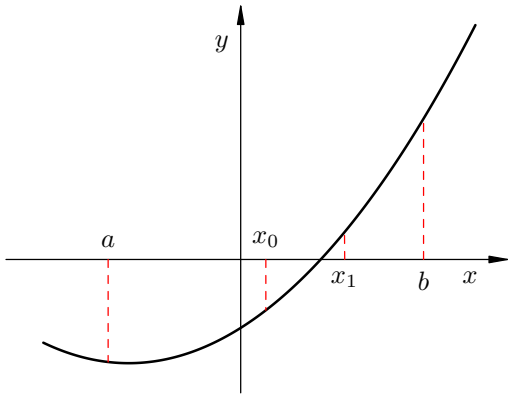
Pitanje je kako ćemo znati da je prethodna relacija ispunjena ako ne znamo α? Jednostavno! Budući da je x_n polovište intervala $[a_n, b_n]$, a $\alpha \in [a_n, b_n]$, onda je

$$|\alpha - x_n| \leq b_n - x_n,$$

pa je dovoljno staviti zahtjev

$$b_n - x_n \leq \varepsilon.$$

Grafički, metoda raspolavljanja izgleda ovako



Algoritam za metodu raspolavljanja je sljedeći.

Algoritam 4.2.1. (Metoda raspolavljanja)

$x := (a + b)/2;$
while $b - x > \varepsilon$ **do**
  begin;
  if $f(x) * f(b) < 0.0$ **then**
    $a := x$
  else
    $b := x;$
  $x := (a + b)/2;$
  end;
{ Na kraju je $x \approx \alpha$. }

Iz konstrukcije metode lako se izvodi pogreška n-te aproksimacije nultočke α. Vrijedi

$$|\alpha - x_n| \leq b_n - x_n = \frac{1}{2}(b_n - a_n) = \frac{1}{2^2}(b_{n-1} - a_{n-1}) = \cdots = \frac{1}{2^{n+1}}(b-a). \quad (4.2.1)$$

Primijetite da je

$$\frac{b-a}{2} = b - x_0,$$

pa bismo korištenjem te relacije (4.2.1) mogli pisati kao

$$|\alpha - x_n| \leq \frac{1}{2^n}(b - x_0).$$

Ova relacija podsjeća na (4.1.2), ali zdesna se nigdje ne pojavljuje $|\alpha - x_0|$. Ipak desna strana daje nam naslutiti da će konvergencija biti dosta spora.

Relacija (4.2.1) omogućava sa unaprijed odredimo koliko je koraka raspolavljanja potrebno da bismo postigli točnost ε. Da bismo postigli da je $|\alpha - x_n| \leq \varepsilon$, dovoljno je zahtijevati da je

$$\frac{1}{2^{n+1}}(b-a) \leq \varepsilon.$$

Množenjem prethodne jednadžbe s 2^{n+1} i dijeljenjem s ε dobivamo

$$\frac{b-a}{\varepsilon} \leq 2^{n+1},$$

a zatim logaritmiranjem dobivamo

$$\log(b-a) - \log\varepsilon \leq (n+1)\log 2,$$

odnosno

$$n \geq \frac{\log(b-a) - \log\varepsilon}{\log 2} - 1, \quad n \in \mathbb{N}_0.$$

Ako je funkcija f još i klase $C^1[a,b]$, tj. ako f ima i neprekidnu prvu derivaciju, može se dobiti dinamička ocjena za udaljenost aproksimacije nultočke od prave nultočke.

Po Teoremu srednje vrijednosti za funkciju f imamo

$$f(x_n) = f(\alpha) + f'(\xi)(x_n - \alpha),$$

pri čemu je ξ između x_n i α. Prvo iskoristimo da je α nultočka, tj. $f(\alpha) = 0$, a zatim uzmemo apsolutne vrijednosti obje strane. Dobivamo

$$|f(x_n)| = |f'(\xi)|\,|\alpha - x_n|. \quad (4.2.2)$$

Primijetite da je

$$|f'(\xi)| \geq m_1, \quad m_1 = \min_{x\in[a,b]} |f'(x)|.$$

Ako je $m_1 > 0$, uvrštavanjem prethodne ocjene u (4.2.2) izlazi

$$|\alpha - x_n| \leq \frac{|f(x_n)|}{m_1}.$$

Drugim riječima, ako želimo da je $|\alpha - x_n| \leq \varepsilon$, dovoljno je zahtijevati da je

$$\frac{|f(x_n)|}{m_1} \leq \varepsilon,$$

odnosno da vrijedi

$$|f(x_n)| \leq m_1\varepsilon.$$

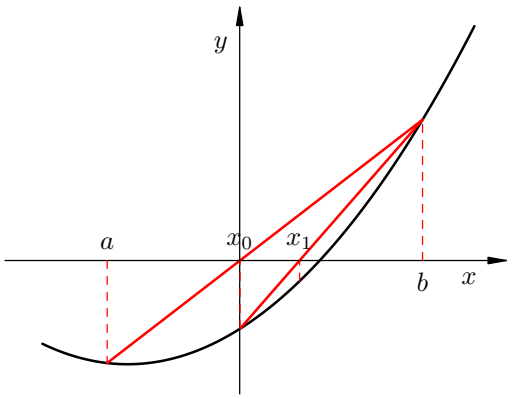
4.3. Regula falsi (metoda pogrešnog položaja)

U prethodnom poglavlju opisali smo metodu raspolavljanja, koja ima sigurnu konvergenciju, ali je vrlo spora. Prirodan je pokušaj ubrzavanja te metode je *regula falsi*. Konstruirat ćemo metodu koja će, ponovno biti konvergentna, čim se nultočka nalazi unutar $[a, b]$.

Pretpostavimo da je funkcija $f : [a, b] \to \mathbb{R}$ neprekidna na $[a, b]$ i da vrijedi

$$f(a) \cdot f(b) < 0.$$

Aproksimirajmo funkciju f pravcem koji prolazi točkama $(a, f(a))$, $(b, f(b))$. Nultočku α tada možemo aproksimirati nultočkom tog pravca, točkom x_0. Nakon toga, pomaknemo ili točku a ili točku b u x_0, ali tako da je unutar novodobivenog intervala ostala nultočka. Postupak ponavljamo sve dok nismo postigli željenu točnost.



Točka x_0 dobiva se jednostavno iz jednadžbe pravca, pa je

$$x_0 = b - f(b)\,\frac{b-a}{f(b)-f(a)}. \tag{4.3.1}$$

Postoji nekoliko ozbiljnih problema s ovom metodom, iako je aproksimacija pravcem i zatvaranje nultočke u određeni interval sasvim dobra ideja.

Izvedimo red konvergencije metode. Iskoristimo relaciju (4.3.1) za x_0, pomnožimo je s -1 i dodajmo α s obje strane. Odatle, uz oznaku ($f[a,b]$ je prva podijeljena razlika)

$$f[a,b] = \frac{f(b)-f(a)}{b-a},$$

izlazi

$$\begin{aligned}
\alpha - x_0 &= \alpha - b + \frac{f(b)}{f[a,b]} = (\alpha-b)\Big(1 + \frac{f(b)}{(\alpha-b)f[a,b]}\Big) \\
&= (\alpha-b)\left(1 + \frac{f(b)-f(\alpha)}{(\alpha-b)f[a,b]}\right) = (\alpha-b)\left(1 + (b-\alpha)\,\frac{f[b,\alpha]}{(\alpha-b)f[a,b]}\right) \\
&= (\alpha-b)\left(1 - \frac{f[b,\alpha]}{f[a,b]}\right) = (\alpha-b)\,\frac{f[a,b]-f[b,\alpha]}{f[a,b]} \\
&= -(\alpha-b)\,(\alpha-a)\frac{f[a,b,\alpha]}{f[a,b]},
\end{aligned}$$

pri čemu je po definiciji $f[a,b,\alpha]$ druga podijeljena razlika

$$f[a,b,\alpha] = \frac{f[b,\alpha]-f[a,b]}{\alpha-a}.$$

Ako je f klase $C^1[a,b]$, onda po Teoremu srednje vrijednosti imamo

$$f[a,b] = f'(\xi), \quad \xi \in [a,b].$$

Na sličan način, ako je f klase $C^2[a,b]$, lako je dokazati da je

$$f[a,b,\alpha] = \frac{1}{2}\,f''(\zeta),$$

gdje se ζ nalazi između minimima i maksimuma vrijednosti a, b, α. Iskoristimo li te dvije relacije, za funkcije klase $C^2[a,b]$ dobivamo sljedeću ocjenu

$$\alpha - x_0 = -(\alpha-b)\,(\alpha-a)\frac{f''(\zeta)}{2f'(\xi)}. \tag{4.3.2}$$

Da bismo pojednostavnili analizu, pretpostavimo da je $f''(\alpha) \neq 0$ i α je jedini korijen unutar $[a,b]$. Također, pretpostavimo da je $f''(a) \geq 0$ za sve $x \in [a,b]$. Razlikujemo dva slučaja:

Slučaj $f'(x) > 0$.

U tom je slučaju f konveksna rastuća funkcija, a spojnica točaka $(a, f(a))$ i $(b, f(b))$ se uvijek nalazi **iznad** funkcije f. Uvrštavanjem podataka o prvoj i drugoj derivaciji u (4.3.2), dobivamo da je desna strana (4.3.2) veća od 0, tj. $\alpha > x_0$, pa će se u sljedećem koraku pomaknuti a. Isto će se dogoditi u svim narednim koracima. Drugim riječima, α neprestano ostaje desno od aproksimacija x_n. Promatramo li (4.3.2), to znači da je b fiksan, pa za proizvoljnu iteraciju x_n dobivamo

$$\alpha - x_n = -(\alpha - b)\,(\alpha - a_n)\frac{f''(\zeta_n)}{2f'(\xi_n)}.$$

Uzimanjem apsolutnih vrijednosti zdesna i slijeva, slijedi da je u tom slučaju konvergencija *regule falsi* linearna.

Pogled na sličnu ocjenu za metodu bisekcije, odmah kaže da ne bi trebalo biti preteško konstruirati primjere kad je metoda bisekcije brža no *regula falsi*.

Slučaj $f'(x) < 0$.

U ovom slučaju je aproksimacija nultočke uvijek desno od α, a uvijek se pomiče b. Analiza ovog slučaja vrlo je slična prethodnom.

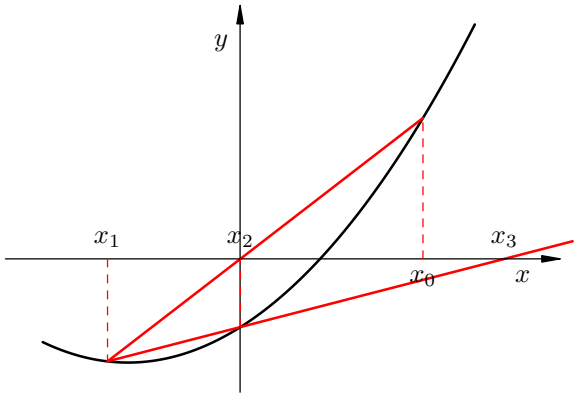
4.4. Metoda sekante

Ako graf funkcije f aproksimiramo sekantom, slično kao kod *regule falsi*, samo ne zahtijevamo da nultočka funkcije f ostane "zatvorena" unutar posljednje dvije iteracije, dobili smo metodu sekante. Time smo izgubili svojstvo sigurne konvergencije, ali se nadamo da će metoda, kad konvergira konvergirati brže nego *regula falsi*.

Počinjemo s dvije početne točke x_0 i x_1 i povlačimo sekantu kroz $(x_0, f(x_0))$, $(x_1, f(x_1))$. Ta sekanta siječe os x u točki x_2. Postupak nastavljamo povlačenjem sekante kroz posljednje dvije točke $(x_1, f(x_1))$ i $(x_2, f(x_2))$. Formule za metodu sekante dobivaju se iteriranjem početne formule za *regulu falsi*, tako da dobivamo

$$x_{n+1} = x_n - f(x_n)\,\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}. \tag{4.4.1}$$

Grafički to izgleda ovako.



Primijetite da je treće iteracija izašla izvan početnog intervala, pa metoda sekante ne mora konvergirati. Jednako tako, da smo "prirodno" numerirali prve dvije točke, tako da je $x_0 < x_1$, imali bismo konvergenciju prema rješenju.

Iskoristimo li ocjenu (4.3.2) za svaki n, dobit ćemo rad konvergencije metode sekante, uz odgovarajuće pretpostavke. Imamo

$$\alpha - x_{n+1} = -(\alpha - x_n)\,(\alpha - x_{n-1})\frac{f''(\zeta_n)}{2f'(\xi_n)}. \tag{4.4.2}$$

Teorem 4.4.1. *Neka su f, f' i f'' neprekidne za sve x u nekom intervalu koji sadrži jednostruku nultočku α ($f'(\alpha) \neq 0$). Ako su početne aproksimacije x_0 i x_1 izabrane dovoljno blizu α, niz iteracija x_n konvergirat će prema α s redom konvergencije p, gdje je*

$$p = \frac{1+\sqrt{5}}{2} \approx 1.618.$$

Dokaz:

Budući da je $f'(\alpha) \neq 0$, u nekoj okolini nultočke α, $I = [\alpha - \varepsilon, \alpha + \varepsilon]$, $\varepsilon > 0$, možemo definirati

$$M = \frac{\max\limits_{x\in I} |f''(x)|}{2\min\limits_{x\in I} |f'(x)|}.$$

Za sve $x_0, x_1 \in I$, korištenjem (4.4.2), dobivamo

$$|\alpha - x_2| \leq |\alpha - x_1|\,|\alpha - x_0|\,M.$$

Da bismo skratili zapis, označimo s $e_n = \alpha - x_n$ grešku n-te iteracije (aproksimacije nultočke). Množenjem prethodne nejednakosti s M dobivamo

$$M|e_2| \leq M|e_1|\,M|e_0|.$$

Nadalje, pretpostavimo da su x_0 i x_1 izabrani tako da je

$$\delta = \max\{M|e_0|, M|e_1|\} < 1.$$

Odatle odmah slijedi da je

$$M|e_2| \leq \delta^2 < \delta.$$

Odatle zaključujemo da je

$$|e_2| < \frac{\delta}{M} = \max\{|e_0|, |e_1|\} \leq \varepsilon,$$

odnosno

$$x_2 \in [\alpha - \varepsilon, \alpha + \varepsilon] = I.$$

Primijenimo li induktivno taj argument, dobivamo

$$M|e_3| \leq M|e_2|M|e_1| \leq \delta^2 \cdot \delta = \delta^3$$
$$M|e_4| \leq M|e_3|M|e_2| \leq \delta^5.$$

Općenito, ako je

$$M|e_{n-1}| \leq \delta^{q_{n-1}}, \quad M|e_n| \leq \delta^{q_n},$$

onda je

$$M|e_{n+1}| \leq M|e_n|M|e_{n-1}| \leq \delta^{q_n+q_{n-1}} = \delta^{q_{n+1}},$$

pa je

$$q_{n+1} = q_n + q_{n-1}, \quad n \geq 1,$$

s $q_0 = q_1 = 1$. Prethodna rekurzija je rekurzija za Fibonaccijeve brojeve i lako se računa njeno eksplicitno rješenje – tj. dovoljno je riješiti diferencijsku jednadžbu

$$q_{n+1} - q_n - q_{n-1} = 0,$$

uz zadane početne $q_0 = q_1 = 1$.

Karakteristična jednadžba je

$$k^2 - k - 1 = 0,$$

pa su njena rješenja

$$k_{1,2} = \frac{1 \pm \sqrt{5}}{2}.$$

Označimo li

$$r_0 = \frac{1 + \sqrt{5}}{2}, \quad r_1 = \frac{1 - \sqrt{5}}{2},$$

onda je opće rješenje te diferencijske jednadžbe

$$q_n = c_0 r_0^n + c_1 r_1^n.$$

Konstante c_0 i c_1 određujemo iz početnih uvjeta. Dobivamo

$$1 = q_0 = c_0 + c_1$$
$$1 = q_1 = c_0 r_0 + c_1 r_1.$$

Rješavanjem ovog para jednadžbi, dobivamo

$$c_0 = \frac{1}{\sqrt{5}} r_0, \quad c_1 = -\frac{1}{\sqrt{5}} r_1,$$

pa je

$$q_n = \frac{1}{\sqrt{5}} (r_0^{n+1} - r_1^{n+1}), \quad n \geq 0.$$

Budući da je

$$r_0 \approx 1.618, \quad r_1 \approx -0.618,$$

onda za velike n $r_1^{n+1} \to 0$, pa je

$$q_n \approx \frac{1}{\sqrt{5}} (1.618)^{n+1}.$$

Vratimo se na e_n. Ovim smo pokazali da je

$$|e_n| \leq \frac{1}{M} \delta^{q_n}, \quad n \geq 0.$$

Budući da $q_n \to \infty$ za $n \to \infty$, dobivamo da $x_n \to \alpha$.

Ovaj "kvazidokaz" (jer su svugdje gornje ograde) daje nam samo ideju o redu konvergencije, koji je zaista $p = r_0$, ali je pravi dokaz mnogo teži. ■

Kod metode sekante postoji nekoliko problema. Prvi je da može divergirati ako početne aproksimacije nisu dobro odabrane.

Drugi problem koji se može javiti je kraćenje u kvocijentu

$$\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}$$

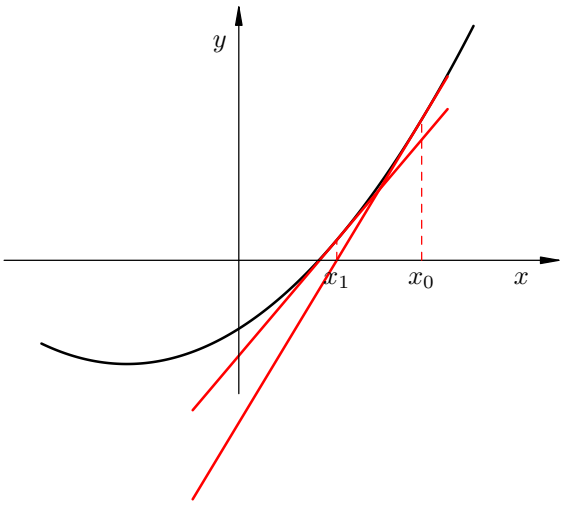
kad $x_n \to \alpha$. Osim toga, budući da iteracije ne "zatvaraju" nultočku nije lako reći kad treba zaustaviti iterativni proces.

Konačno, primijetite da je za svaku iteraciju metode sekante potrebno samo jednom izvrednjavati funkciju f i to u točki x_n, jer $f(x_{n-1})$ čuvamo od prethodne iteracije.

4.5. Metoda tangente (Newtonova metoda)

Ako graf funkcije f umjesto sekantom, aproksimiramo tangentom, dobili smo metodu tangente ili Newtonovu metodu. Slično kao i kod sekante, time smo izgubili svojstvo sigurne konvergencije, ali se nadamo da će metoda brzo konvergirati.

Pretpostavimo da je zadana početna točka x_0. Ideja metode je povući tangentu u točki $(x_0, f(x_0))$ i definirati novu aproksimaciju x_0 u točki gdje ona siječe os x.



Geometrijski izvod je jednostavan. U točki x_n napiše se jednadžba tangente i pogleda se gdje siječe os x. Jednadžba tangente je

$$y - f(x_n) = f'(x_n)(x - x_n),$$

odakle izlazi da je nova aproksimacija $x_{n+1} := x$

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

Primijetite da je prethodna formula usko vezana uz metodu sekante, jer je

$$f'(x_n) \approx \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}.$$

Do Newtonove metode može se doći i na drugačiji način. Pretpostavimo li da je funkcija f dva puta neprekidno derivabilna (na nekom području oko α), onda je možemo razviti u Taylorov red oko x_n do uključivo prvog člana. Dobivamo

$$f(x) = f(x_n) + f'(x_n)(x - x_n) + \frac{f''(\xi_n)}{2}(x - x_n)^2,$$

pri čemu je ξ_n između x i x_n. Uvrštavanjem $x = \alpha$, dobivamo

$$0 = f(\alpha) = f(x_n) + f'(x_n)(\alpha - x_n) + \frac{f''(\xi_n)}{2}(\alpha - x_n)^2.$$

Premještanjem, uz pretpostavku $f'(x_n) \neq 0$, izlazi

$$\alpha = x_n - \frac{f(x_n)}{f'(x_n)} - (\alpha - x_n)^2 \frac{f''(\xi_n)}{2f'(x_n)}.$$

Primijetite da prva dva člana zdesna daju x_{n+1}, pa dobivamo

$$\alpha - x_{n+1} = -(\alpha - x_n)^2 \frac{f''(\xi_n)}{2f'(x_n)}. \tag{4.5.1}$$

Iz (4.5.1), odmah čitamo da je Newtonova metoda, kad konvergira kvadratično konvergentna. Ipak, treba biti oprezan, jer takav zaključak vrijedi samo ako $f'(x_n)$ ne teži u 0 tijeku cijelog procesa, tj. ako je $f'(\alpha) \neq 0$ (drugim riječima, ako je nultočka jednostruka).

Slično, kao kod metode sekante, možemo dokazati sljedeći teorem o konvergenciji Newtonove metode.

Teorem 4.5.1. *Neka su f, f' i f'' neprekidne za sve x u nekom intervalu koji sadrži jednostruku nultočku α ($f'(\alpha) \neq 0$). Ako je početna aproksimacija x_0 izabrana dovoljno blizu α, niz iteracija x_n konvergirat će prema α s redom konvergencije $p = 2$. Čak štoviše, vrijedi*

$$\lim_{n\to\infty} \frac{\alpha - x_{n+1}}{(\alpha - x_n)^2} = -\frac{f''(\alpha)}{2f'(\alpha)}.$$

Dokaz:

Izaberimo interval $I = [\alpha - \varepsilon, \alpha + \varepsilon]$ i neka je

$$M = \frac{\max\limits_{x\in I} |f''(x)|}{2 \min\limits_{x\in I} |f'(x)|}.$$

Za sve $x_0 \in I$, korištenjem (4.5.1), dobivamo

$$|\alpha - x_1| \leq M|\alpha - x_0|^2,$$

odnosno

$$M|\alpha - x_1| \leq (M|\alpha - x_0|)^2.$$

Izaberimo $|\alpha - x_0| \leq \varepsilon$ i $M|\alpha - x_0| < 1$. Tada je

$$M|\alpha - x_1| \leq M|\alpha - x_0|,$$

što pokazuje da je

$$|\alpha - x_1| \leq |\alpha - x_0| \leq \varepsilon.$$

Primjenom istog argumenta, induktivno dobivamo

$$|\alpha - x_n| \leq \varepsilon, \quad M|\alpha - x_n| < 1$$

za sve $n \geq 1$. Da bismo pokazali konvergenciju iskoristimo (4.5.1). Imamo

$$|\alpha - x_{n+1}| \leq M|\alpha - x_n|^2, \quad M|\alpha - x_{n+1}| \leq (M|\alpha - x_n|)^2,$$

i induktivno

$$M|\alpha - x_n| \leq (M|\alpha - x_0|)^{2^n}, \quad |\alpha - x_n| \leq \frac{1}{M}(M|\alpha - x_0|)^{2^n}.$$

Budući da je $M|\alpha - x_0| < 1$, to pokazuje da $x_n \to \alpha$ za $n \to \infty$.

Budući da u (4.5.1) ξ_n leži između x_n i α, onda mora biti $\xi_n \to \alpha$ za $n \to \infty$. Zbog toga je

$$\lim_{n\to\infty} \frac{\alpha - x_{n+1}}{(\alpha - x_n)^2} = -\lim_{n\to\infty} \frac{f''(\xi_n)}{2f'(x_n)} = -\frac{f''(\alpha)}{2f'(\alpha)}.$$

■

Jednostavnim riječima, ovaj teorem daje dovoljne uvjete za tzv. **lokalnu** konvergenciju Newtonove metode prema jednostrukoj nultočki. Lokalnost se odnosi na to da početna aproksimacija mora biti dovoljno blizu nultočke

$$|\alpha - x_0| \leq \varepsilon.$$

Veličina ε određena je drugim uvjetom $M|\alpha - x_0| < 1$ koji daje sigurnu konvergenciju. Naravno, tada je

$$|\alpha - x_0| < \frac{1}{M},$$

pa bi ispalo da treba uzeti $\varepsilon = 1/M$. To, nažalost, ne mora vrijediti, jer M općenito ovisi o ε. Ipak, u nekim situacijama možemo iskoristiti sličan uvjet za osiguranje konvergencije Newtonove metode.

Pretpostavimo da smo locirali nultočku funkcije f u segmentu $[a, b]$ i znamo da je f klase C^2 na tom segmentu. Neka je

$$M_2 = \max_{x\in[a,b]} |f''(x)|, \quad m_1 = \min_{x\in[a,b]} |f'(x)|.$$

Ako je f još i strogo monotona na $[a, b]$, onda je $m_1 > 0$ (a vrijedi i obrat). Tada f ima jedinstvenu jednostruku nultočku α u $[a, b]$. Umjesto "lokalnog" M, izračunamo "globalnu" veličinu

$$M' := \frac{M_2}{2m_1}.$$

Ako vrijedi

$$\frac{b-a}{2} < \frac{1}{M'},$$

onda možemo uzeti $\varepsilon = (b-a)/2$, a startna točka je polovište intervala $x_0 := (a+b)/2$. Zbog

$$|x_0 - \alpha| \leq \varepsilon < 1/M',$$

imamo sigurnu konvergenciju iteracija prema nultočki. Ako vrijedi i jači uvjet

$$b - a < \frac{1}{M'},$$

onda bilo koja startna točka $x_0 \in [a, b]$ daje sigurnu konvergenciju.

Naravno, to možemo iskoristiti samo ako imamo dovoljno informacija o funkciji f da možemo izračunati M', odnosno M_2 i m_1. Umjesto M_2, možemo uzeti i neku gornju ogradu za M_2, a umjesto m_1, neku pozitivnu donju ogradu za m_1.

Ove dvije veličine M_2 i m_1 daju i lokalne ocjene greške iteracija u Newtonovoj metodi, uz uvjet da su sve iteracije u $[a, b]$. Iz ranije relacije (4.5.1)

$$\alpha - x_n = -\frac{f''(\xi_{n-1})}{2f'(x_{n-1})}\,(\alpha - x_{n-1})^2,$$

gdje je ξ_{n-1} između α i x_{n-1}, odmah slijedi

$$|\alpha - x_n| \leq \frac{M_2}{2m_1}\,(\alpha - x_{n-1})^2.$$

Ova ocjena nije naročito korisna za praksu, jer α ne znamo. Uočite da smo sličnu ocjenu već imali u prethodnom teoremu, samo s M umjesto M'.

Za dvije susjedne iteracije u Newtonovoj metodi također vrijedi veza preko Taylorove formule

$$f(x_n) = f(x_{n-1}) + f'(x_{n-1})(x_n - x_{n-1}) + \frac{f''(\xi_{n-1})}{2}(x_n - x_{n-1})^2,$$

pri čemu je ξ_{n-1} između x_{n-1} i x_n. Po definiciji iteracija u Newtonovoj metodi vrijedi i

$$f(x_{n-1}) + f'(x_{n-1})(x_n - x_{n-1}) = 0,$$

pa je

$$f(x_n) = \frac{f''(\xi_{n-1})}{2}(x_n - x_{n-1})^2.$$

Koristeći pretpostavku x_{n-1}, $x_n \in [a, b]$, dobivamo

$$|f(x_n)| \leq \frac{M_2}{2}(x_n - x_{n-1})^2.$$

Kao i kod metode bisekcije, ako je $m_1 > 0$, iz (4.2.2) slijedi ocjena

$$|\alpha - x_n| \leq \frac{|f(x_n)|}{m_1}.$$

Kombinacijom ovih ocjena dobivamo

$$|\alpha - x_n| \leq \frac{M_2}{2m_1}(x_n - x_{n-1})^2,$$

što se može iskoristiti u praksi. Ako je ε tražena točnost, onda test

$$\frac{M_2}{2m_1}(x_n - x_{n-1})^2 \leq \varepsilon$$

garantira da je $|\alpha - x_n| \leq \varepsilon$, do na greške zaokruživanja. Naravno, možemo koristiti i raniji test

$$\frac{|f(x_n)|}{m_1} \leq \varepsilon.$$

U ovim ocjenama greške koristili smo pretpostavku da je f strogo monotona na $[a, b]$. Ako i druga derivacija ima fiksni predznak na tom intervalu, onda možemo dobiti i **globalnu** konvergenciju Newtonove metode.

Teorem 4.5.2. *Neka je $f \in C^2[a, b]$ i $f(a) \cdot f(b) < 0$. Ako f' i f'' nemaju nultočku u $[a, b]$, tj. ako f' i f'' imaju fiksni predznak na $[a, b]$, onda Newtonova metoda konvergira prema (jedinstvenoj jednostrukoj) nultočki α funkcije f, za svaku startnu aproksimaciju $x_0 \in [a, b]$ za koju vrijedi*

$$f(x_0) \cdot f''(x_0) > 0.$$

Dokaz:

Pretpostavimo, na primjer, da je $f' > 0$ i $f'' > 0$ na cijelom $[a, b]$. Tada, jer f raste, mora biti $f(a) < 0$ i $f(b) > 0$. Zbog $f'' > 0$, startna aproksimacija mora zadovoljavati $f(x_0) > 0$. U praksi možemo uzeti $x_0 = b$, jer je to jedina točka za koju sigurno znamo da vrijedi $f(x_0) > 0$.

Neka je $(x_n, n \in \mathbb{N}_0)$, niz iteracija generiran Newtonovom metodom iz startne točke x_0 za koju je $f(x_0) > 0$

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

Za početak, znamo da je $x_0 > \alpha$. Tvrdimo da je $\alpha < x_n \leq x_0$ za svaki $n \in N_0$. Dokaz ide indukcijom, a bazu već imamo. Pretpostavimo da je $\alpha < x_n \leq x_0$. Onda je $f(x_n) > 0$ i $f'(x_n) > 0$, pa je

$$x_{n+1} < x_n \leq x_0,$$

što pokazuje da (x_n) monotono pada. Iz Taylorove formule je

$$0 = f(\alpha) = f(x_n) + f'(x_n)(\alpha - x_n) + \frac{f''(\xi_n)}{2}(\alpha - x_n)^2,$$

pri čemu je $\xi_n \in (\alpha, x_n) \subset [a, b]$. Zbog toga je $f''(\xi_n) > 0$, pa je

$$f(x_n) + f'(x_n)(\alpha - x_n) < 0,$$

odakle slijedi

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

Dakle, niz (x_n) je odozdo ograničen s α i monotono pada pa postoji limes

$$\alpha' := \lim_{n\to\infty} x_n.$$

Odmah znamo i da je $\alpha \leq \alpha' \leq x_0$, tj. $\alpha' \in [a, b]$. Prijelazom na limes u formuli za Newtonove iteracije dobivamo

$$\alpha' = \alpha' - \frac{f(\alpha')}{f'(\alpha')},$$

odakle, koristeći $f'(\alpha') \neq 0$, slijedi $f(\alpha') = 0$. No, znamo da f ima jedinstvenu nultočku α u $[a, b]$ pa mora biti $\alpha = \alpha'$.

Preostala tri slučaja za predznake prve i druge derivacije se dokazuju potpuno analogno. ■

Uvjet $f(x_0) \cdot f''(x_0) > 0$ na izbor startne točke u prethodnom teoremu ima vrlo jednostavnu geometrijsku interpretaciju. Ako pogledamo graf funkcije f na $[a, b]$, startnu točku x_0 treba odabrati na "strmijoj" strani funkcije.

Primijetite da računanje u Newtonovoj metodi, iako ima veći red konvergencije nego sekanta, može trajati dulje (naravno uz istu točnost rezultata). Objašnjenje leži u činjenici da se za svaki korak Newtonove metode mora izračunati i vrijednost funkcije i vrijednost derivacije u točki. Ako se derivacija komplicirano računa, sekanta će biti brža.

Prethodni teoremi daju samo dovoljne uvjete konvergencije pojedinih iterativnih metoda. U praktičnom računanju često imamo samo interval $[a, b]$ u kojem smo locirali nultočku funkcije f, a **nemamo** dodatne informacije o funkciji f iz kojih bismo mogli izvući zaključak o konvergenciji bržih iterativnih metoda. Zbog toga se ove metode katkad kombiniraju s metodom bisekcije na sljedeći način. Prvo izračunamo novu iteraciju po bržoj metodi i ako ona ostaje u trenutnom intervalu, onda ju prihvaćamo i s njom nastavljamo iteracije i skraćujemo interval. U protivnom, radimo korak bisekcije za smanjivanje intervala.

4.6. Metoda jednostavne iteracije

Pretpostavimo da tražimo α, rješenje jednadžbe

$$x = g(x). \tag{4.6.1}$$

Definiramo jednostavnu iteracionu funkciju (iteracionu funkciju koja "pamti" samo jednu prethodnu točku) s

$$x_{n+1} = g(x_n), \quad n \geq 0,$$

uz x_0 kao početnu aproksimaciju za α. Primijetite da Newtonova metoda pripada klasi jednostavnih iteracija, jer je

$$g(x) = x - \frac{f(x)}{f'(x)}.$$

Rješenja, tj. točke za koje je $x = g(x)$, zovu se **fiksne točke** od g. Uobičajeno, mi smo zainteresirani $f(x) = 0$, pa taj problem treba reformulirati na problem (4.6.1). Postoji mnogo načina za tu reformulaciju.

Primjer 4.6.1. *Reformulirajmo problem*

$$x^2 - a = 0, \quad a > 0$$

u oblik (4.6.1). Na primjer, to možemo napraviti na jedan od sljedećih načina:

1. $x = x^2 + x - a$, *ili općenitije* $x = x + c(x^2 - a)$ *za neki* $c \neq 0$,
2. $x = a/x$,
3. $x = 0.5\,(x + a/x)$.

Prirodno je pitanje kako se različite jednostavne iteracije ponašaju. Odgovor ćemo dobiti nizom sljedećih tvrdnji.

Lema 4.6.1. *Neka je funkcija g neprekidna na intervalu $[a, b]$ i neka je*

$$a \leq g(x) \leq b, \quad \forall x \in [a, b],$$

u oznaci $g([a, b]) \subseteq [a, b]$. Tada jednostavna iteracija $x = g(x)$ ima bar jedno rješenje na $[a, b]$.

Dokaz:

Za neprekidnu funkciju $g(x) - x$ na intervalu $[a, b]$ vrijedi

$$g(a) - a \geq 0, \quad g(b) - b \leq 0.$$

Drugim riječima, funkcija $g(x) - x$ je promijenila predznak na intervalu $[a, b]$, a to može samo prolaskom kroz nultočku (neprekidna je!). ■

Lema 4.6.2. *Neka je funkcija g neprekidna na $[a,b]$ i neka je*

$$g([a,b]) \subseteq [a,b].$$

Nadalje, pretpostavimo da postoji konstanta λ, $0 < \lambda < 1$, takva da vrijedi

$$|g(x) - g(y)| \leq \lambda\,|x - y|, \quad \forall x, y \in [a,b].$$

Tada $x = g(x)$ ima jedinstveno rješenje α unutar $[a,b]$. Također, niz iteracija

$$x_n = g(x_{n-1}), \quad n \geq 1$$

konvergira prema α za proizvoljni $x_0 \in [a,b]$.

Dokaz:

Prema prethodnoj lemi, postoji bar jedno rješenje $\alpha \in [a,b]$. Pokažimo da ne postoji više od jednog rješenja. Da bismo to pokazali, pretpostavimo suprotno, tj. postoje barem dva rješenja. Uzmimo bilo koja dva od tih rješenja i nazovimo ih α i β iz $[a,b]$. Budući da su to rješenja, vrijedi

$$g(\alpha) = \alpha \quad \text{i} \quad g(\beta) = \beta.$$

Po pretpostavci, uvažavajući prethodne jednakosti, dobivamo

$$|\alpha - \beta| = |g(\alpha) - g(\beta)| \leq \lambda|\alpha - \beta|,$$

ili drugim riječima

$$(1 - \lambda)\,|\alpha - \beta| \leq 0.$$

Budući da je $1 - \lambda > 0$, mora biti $\alpha = \beta$.

Dokažimo još konvergenciju jednostavnih iteracija za proizvoljnu startnu točku $x_0 \in [a,b]$. Prvo, uočimo da $x_{n-1} \in [a,b]$ povlači da je $x_n = g(x_{n-1}) \in [a,b]$. Nadalje, vrijedi

$$|\alpha - x_n| = |g(\alpha) - g(x_{n-1})| \leq \lambda|\alpha - x_{n-1}|,$$

odnosno indukcijom po n dobivamo

$$|\alpha - x_n| \leq \lambda^n|\alpha - x_0|, \quad n \geq 1.$$

Ako pustimo $n \to \infty$, onda $\lambda^n \to 0$, pa vrijedi $x_n \to \alpha$. ■

Ako je g derivabilna na $[a,b]$, onda je po Teoremu srednje vrijednosti

$$g(x) - g(y) = g'(\xi)(x - y), \quad \xi \text{ između } x \text{ i } y$$

za sve x, $y \in [a,b]$. Definiramo

$$\lambda = \max_{x \in [a,b]} |g'(x)|, \tag{4.6.2}$$

onda možemo pisati

$$|g(x) - g(y)| = \lambda|x - y|, \quad \forall x \in [a,b].$$

Primijetite λ može biti veći od 1!

Teorem 4.6.1. *Neka je funkcija g neprekidno diferencijabilna na $[a,b]$, neka je*

$$g([a,b]) \subseteq [a,b],$$

i neka za λ iz (4.6.2) vrijedi

$$\lambda < 1. \tag{4.6.3}$$

Tada vrijedi:

1. *$x = g(x)$ ima točno jedno rješenje na $\alpha \in [a,b]$,*
2. *za proizvoljni $x_0 \in [a,b]$, za jednostavnu iteraciju $x_{n+1} = g(x_n)$, $n \geq 0$ vrijedi*

$$\lim_{n\to\infty} x_n = \alpha,$$

$$|\alpha - x_n| \leq \lambda^n |\alpha - x_0|$$

i

$$\lim_{n\to\infty} \frac{\alpha - x_{n+1}}{\alpha - x_n} = g'(\alpha).$$

Dokaz:

Sve tvrdnje ovog teorema dokazane su u prethodne dvije leme, osim posljednje relacije o brzini konvergencije.

Vrijedi

$$\alpha - x_{n+1} = g(\alpha) - g(x_n) = g'(\xi_n)(\alpha - x_n), \quad n \geq 0,$$

gdje je ξ_n neki broj između α i x_n. Budući da $x_n \to \alpha$, onda i $\xi_n \to \alpha$, pa vrijedi

$$\lim_{n\to\infty} \frac{\alpha - x_{n+1}}{\alpha - x_n} = \lim_{n\to\infty} g'(\xi_n) = g'(\alpha).$$

■

Pokažimo koliko je pretpostavka (4.6.3) značajna, tj. pretpostavimo da je $|g'(\alpha)| > 1$. Tada, ako imamo niz $x_{n+1} = g(x_n)$ i rješenje $\alpha = g(\alpha)$, vrijedi

$$\alpha - x_{n+1} = g(\alpha) - g(x_n) = g'(\xi_n)(\alpha - x_n).$$

Za x_n dovoljno blizu α, onda je i i $|g'(\xi_n)| > 1$, pa je $|\alpha - x_{n+1}| \geq |\alpha - x_n|$, pa konvergencija metode nije moguća.

Prethodni teorem se može malo i pojednostavniti, tako da se ne navodi eksplicitno interval $[a,b]$.

Teorem 4.6.2. *Neka je α rješenje jednostavne iteracije $x = g(x)$ i neka je g neprekidno diferencijabilna na nekoj okolini od α i neka je $|g'(\alpha)| < 1$. Tada vrijede svi rezultati Teorema 4.6.1., uz pretpostavku da je x_0 dovoljno blizu α.*

Dokaz:

Uzmimo $I = [\alpha - \varepsilon, \alpha + \varepsilon]$ takav da je

$$\max_{x \in I} |g'(x)| \leq \lambda < 1.$$

Tada je $g(I) \subseteq I$, jer $|\alpha - x| \leq \varepsilon$ povlači

$$|\alpha - g(x)| = |g(\alpha) - g(x)| = |g'(\xi)||\alpha - x| \leq \lambda |\alpha - x| \leq \varepsilon.$$

Sada možemo primijeniti prethodni teorem za $[a, b] = I$. ■

Primjer 4.6.2. *U primjeru 4.6.1., definirali smo tri iteracione funkcije.*

1. *Ako je* $g(x) = x^2 + x - a$, *onda je* $g'(x) = 2x + 1$ *i u nultočki* $\alpha = \sqrt{a}$ *je*

$$g'(\sqrt{a}) = 2\sqrt{a} + 1 > 1,$$

pa ta iteraciona funkcija neće konvergirati. U općenitijem je slučaju $g(x) = x + c(x^2 - a)$, *pa je* $g'(x) = 1 + 2cx$ *i*

$$g'(\sqrt{a}) = 1 + 2c\sqrt{a}.$$

Da bismo osigurali konvergenciju, mora biti

$$-1 < 1 + 2c\sqrt{a} < 1,$$

odnosno

$$-\frac{1}{\sqrt{a}} < c < 0.$$

2. *Ako je* $g(x) = a/x$, *onda je* $g'(x) = -a/x^2$, *pa je*

$$g'(\sqrt{a}) = -1.$$

3. *Ako je* $g(x) = 0.5\,(x + a/x)$, *onda je* $g'(x) = 0.5\,(1 - a/x^2)$, *pa je*

$$g'(\sqrt{a}) = 0.$$

Ovaj odjeljak završit ćemo promatranjem jednostavnih iteracionih funkcija, ali višeg reda konvergencije, kao što je, na primjer Newtonova metoda.

Teorem 4.6.3. *Neka je* α *rješenje od* $x = g(x)$ *i neka je* g p *puta neprekidno diferencijabilna za sve* x *u okolini* α, *za neki* $p \geq 2$. *Nadalje, pretpostavimo da je*

$$g'(\alpha) = \cdots = g^{(p-1)}(\alpha) = 0. \tag{4.6.4}$$

Ako je startna vrijednost x_0 *dovoljno blizu* α, *iteraciona funkcija*

$$x_{n+1} = g(x_n), \quad n \geq 0$$

imat će red konvergencije p *i*

$$\lim_{n \to \infty} \frac{\alpha - x_{n+1}}{(\alpha - x_n)^p} = (-1)^{p-1} \frac{g^{(p)}(\alpha)}{p!}.$$

Dokaz:

Razvijmo $g(x)$ u okolini α do uključivo $(p-1)$-ve potencije i napišimo ostatak. Zatim, uvrstimo $x = x_n$, pa dobivamo

$$x_{n+1} = g(x_n) = g(\alpha) + g'(\alpha)(x_n - \alpha) + \cdots + \frac{g^{(p-1)}(\alpha)}{(p-1)!}(x_n - \alpha)^{p-1} + \frac{g^{(p)}(\xi_n)}{p!}(x_n - \alpha)^p,$$

za neki ξ_n između x_n i α. Iskoristimo li da je $g(\alpha) = \alpha$ i pretpostavku (4.6.4), slijedi

$$x_{n+1} = \alpha + \frac{g^{(p)}(\xi_n)}{p!}(x_n - \alpha)^p,$$

odnosno

$$\alpha - x_{n+1} = -\frac{g^{(p)}(\xi_n)}{p!}(x_n - \alpha)^p.$$

Sada možemo primijeniti prethodni Teorem, koji pokazuje da će iteraciona funkcija konvergirati. Nadalje, to znači da $x_n \to \alpha$, pa i $\xi_n \to \alpha$, što daje traženu relaciju. ■

Korištenjem prethodnog teorema možemo analizirati i Newtonovu metodu za koju je

$$g(x) = x - \frac{f(x)}{f'(x)}.$$

Deriviranjem dobivamo da je

$$g'(x) = 1 - \frac{(f'(x))^2 - f(x)f''(x)}{(f'(x))^2} = \frac{f(x)f''(x)}{(f'(x))^2},$$

pa je

$$g(\alpha) = 0,$$

uz pretpostavku da je $f'(\alpha) \neq 0$. Na sličan način, dobivamo

$$g''(\alpha) = \frac{f''(\alpha)}{f'(\alpha)},$$

pa ako je $f''(\alpha) \neq 0$, možemo pokazati da je red konvergencije Newtonove metode jednak 2. Ako je $f'(\alpha) \neq 0$, $f''(\alpha) = 0$, onda će red konvergencije biti barem 3.

4.7. Newtonova metoda za višestruke nultočke

Promotrimo što će se dogoditi s konvergencijom Newtonove metode, ako funkcija f ima neprekidnih prvih $p+1$ derivacija i p-struku, $p \geq 2$ nultočku u α. Tada vrijedi

$$f(\alpha) = f'(\alpha) = \cdots = f^{(p-1)}(\alpha) = 0, \quad f^{(p)}(\alpha) \neq 0.$$

Samu funkciju f možemo napisati i u obliku

$$f(x) = (x-\alpha)^p\, h(x), \quad h(\alpha) \neq 0. \tag{4.7.1}$$

Ograničimo se samo na cjelobrojne p i promatrajmo Newtonovu metodu kao jednostavnu iteraciju,

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

Deriviranjem (4.7.1) dobivamo jednostavniji oblik za derivaciju

$$f'(x) = p(x-\alpha)^{p-1}h(x) + (x-a)^p h'(x),$$

pa je

$$g(x) = x - \frac{(x-\alpha)h(x)}{ph(x) + (x-\alpha)h'(x)}.$$

Deriviranjem funkcije g dobivamo

$$g'(x) = 1 - \frac{h(x)}{ph(x)+(x-\alpha)h'(x)} - (x-\alpha)\frac{d}{dx}\Big(\frac{h(x)}{ph(x)+(x-\alpha)h'(x)}\Big),$$

tako da je

$$g'(\alpha) = 1 - \frac{1}{p} \neq 0 \quad \text{za} \quad p > 1,$$

što pokazuje linearnu konvergenciju. Prema teoremu 4.6.1., faktor konvergencije bit će $g'(\alpha) = 1 - 1/p$, što je vrlo sporo. U prosjeku to je podjednako brzo kao bisekcija za $p = 2$ ili čak lošije od bisekcije za $p \geq 3$.

Kako možemo popraviti (ubrzati) Newtonovu metodu za p-struke nultočke, $p \geq 2$. Prvo pretpostavimo da znamo p. Definiramo iteracionu funkciju

$$g(x) = x - p\,\frac{f(x)}{f'(x)}.$$

Tada je

$$g'(x) = 1 - p\,\frac{(f'(x))^2 - f(x)f''(x)}{(f'(x))^2} = 1 - p + p\,\frac{f(x)f''(x)}{(f'(x))^2}.$$

Iskoristimo li oblik funkcije f, dobivamo

$$\begin{aligned} f(x) &= (x-\alpha)^p h(x) \\ f'(x) &= (x-\alpha)^{p-1}[ph(x) + (x-\alpha)h'(x)] \\ f''(x) &= (x-\alpha)^{p-2}[p(p-1)h(x) + 2p(x-\alpha)h'(x) + (x-\alpha)^2 h''(x)], \end{aligned}$$

pa je

$$\lim_{x\to\alpha} \frac{f(x)f''(x)}{(f'(x))^2} = 1 - \frac{1}{p}.$$

Odatle odmah slijedi

$$\lim_{x\to\alpha} g'(x) = 0,$$

što pokazuje da ova modifikacija osigurava barem kvadratično konvergentnu metodu.

Što ćemo napraviti ako unaprijed ne znamo p? Primijetimo da funkcija

$$u(x) = \frac{f(x)}{f'(x)} = \frac{(x-\alpha)^p h(x)}{(x-\alpha)^{p-1}[ph(x) + (x-\alpha)h'(x)]} = \frac{(x-\alpha)h(x)}{ph(x) + (x-\alpha)h'(x)}$$

ima jednostruku nultočku u α. Drugim riječima, obična Newtonova metoda, ali primijenjena na $u(x)$ konvergirat će kvadratično,

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

gdje je

$$u'(x) = \frac{(f'(x))^2 - f(x)f''(x)}{(f'(x))^2} = 1 - \frac{f''(x)}{f'(x)}\,u(x),$$

što pokazuje da ćemo dobiti kvadratičnu konvergenciju, iako ne znamo red nultočke, ali uz računanje još jedne derivacije funkcije (f'').

Slično vrijedi i za metodu sekante, koju ćemo ubrzati, kao da radimo s jednostrukim nultočkama, ako primijenimo metodu sekante za funkciju u

$$x_{n+1} = x_n - u(x_n)\,\frac{x_n - x_{n-1}}{u(x_n) - u(x_{n-1})}.$$

I u ovom slučaju postoji "cijena", a to je računanje f'.

4.8. Hibridna Brent–Dekkerova metoda

Brent–Dekkerova metoda smišljena je kao metoda koja će imati sigurnu konvergenciju, a nadamo se da će konvergirati brže nego metoda sekante, u najboljem slučaju kvadratično. Ona **ne zahtijeva** računanje derivacija, pa ako joj je red konvergencije u prosjeku bolji od sekante, možemo očekivati da će metoda po brzini biti slična Newtonovoj, ali će imati sigurnu konvergenciju.

Metoda se sastoji od tri dijela, koje grubo možemo opisati kao inverznu kvadratnu interpolaciju, metodu sekante i metodu bisekcije. Algoritam počinje metodom sekante koja generira treću točku. Ako se prema nekim kriterijima ta točka prihvaća kao dobra, možemo nastaviti raditi s kvadratnom interpolacijom kroz posljednje tri točke, ali inverznom (uloga x i y zamijenjena) i time dobivamo četvrtu točku.

Ako je treća točka odbačena kao loša, radi se jedan korak metode bisekcije. Drugim riječima, metoda se "vrti" između svoja tri sastavna dijela, a mi se nadamo da će rijetko koristiti bisekciju.

Točni parametri kad se neka aproksimacija nultočke prihvaća kao dobra, odnosno odbacuje kao loša su dosta složeni. Metoda je sastavni dio velikih numeričkih biblioteka programa, kao što je IMSL.

4.9. Primjeri

Prije konkretnih primjera, zanimljivo je napomenuti da se u praksi može sasvim dobro numerički procijeniti red konvergencije iterativne metode i taj podatak iskoristiti kao dodatna informacija o konvergenciji metode.

Kao najjednostavniji primjer za usporedbu metoda za nalaženje nultočaka uzmimo da treba izračunati $\sqrt[3]{1.5}$. Taj problem možemo interpretirati i kao traženje realne pozitivne nultočke funkcije $f(x) = x^3 - 1.5$.

Primjer 4.9.1. *Nultočka funkcije*

$$f(x) = \operatorname{arctg}(x)$$

je $x = 0$*, ali Newtonova metoda neće konvergirati iz svake startne točke* x_0*. Naći ćemo točku* β *za koju vrijedi*

$$\begin{cases} |x_0| < |\beta| & \textit{Newtonova metoda sa startom } x_0 \textit{ konvergira,} \\ |x_0| > |\beta| & \textit{Newtonova metoda sa startom } x_0 \textit{ divergira,} \\ |x_0| = |\beta| & \textit{Newtonova metoda sa startom } x_0 \textit{ ciklira.} \end{cases}$$

Kako ćemo naći točku "cikliranja"? Funkcija $f(x) = \operatorname{arctg} x$ *je neparna, pa da bismo dobili cikliranje, dovoljno je da tangenta na funkciju u točki* β *presiječe os* x *u točki* $-\beta$*. Jednadžba tangente na* arctg *u točki* β *je*

$$y - \operatorname{arctg}\beta = \frac{1}{1+\beta^2}(x - \beta),$$

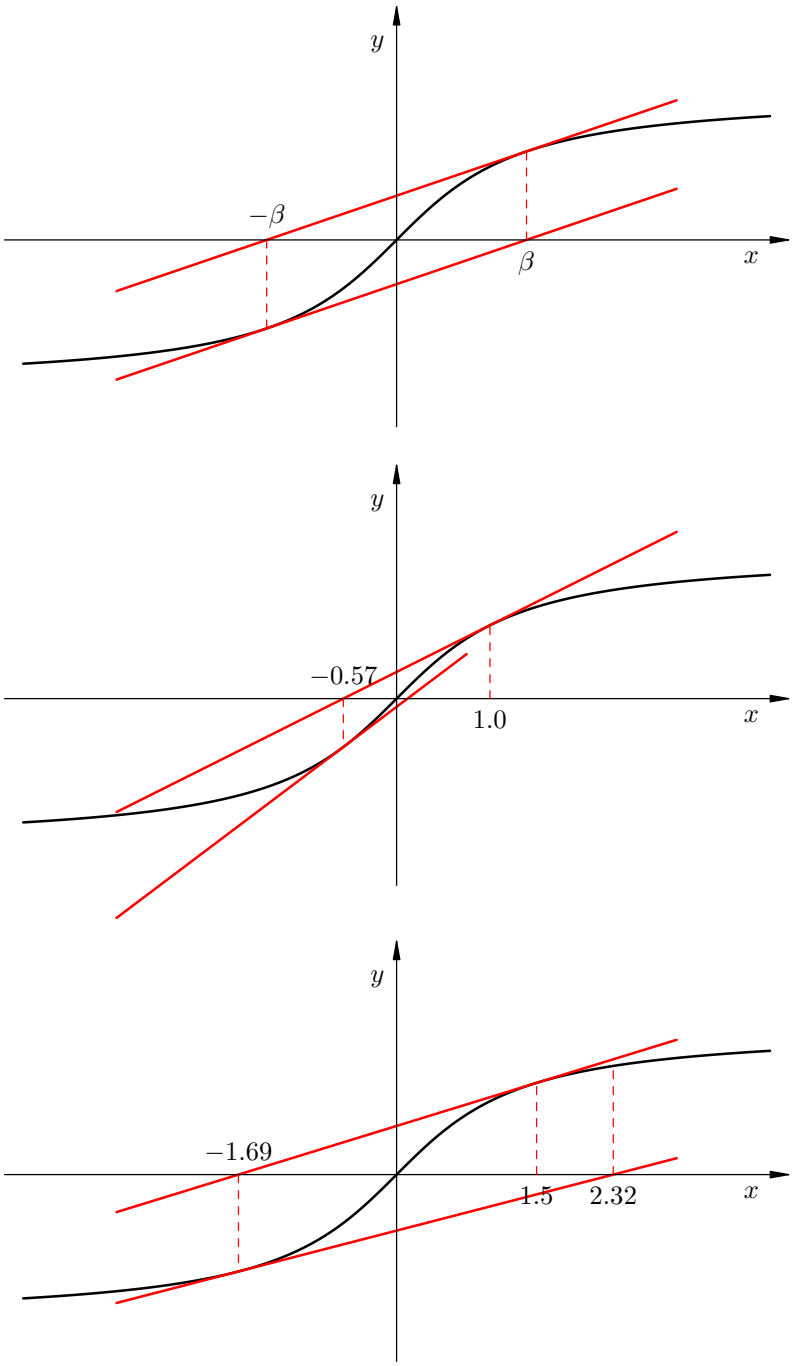
pa će tangenta sijeći os x *u* $-\beta$*, ako je*

$$\operatorname{arctg}\beta = \frac{2\beta}{1+\beta^2},$$

čime smo dobili nelinearnu jednadžbu po β*. Očito, postoje dva rješenja, suprotnih predznaka, i nije ih teško izračunati metodom bisekcije*

$$\beta = \pm 1.39174520027073489.$$

Nacrtajmo grafove Newtonove metode za sve tri mogućnosti za x_0, recimo za $x_0 = 1$, $x_0 = \beta$ i $x_0 = 1.5$.



Literatura

[1] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, D. Sorensen, *LAPACK Users' Guide*, Third edition, SIAM, Philadelpia, 1999.

[2] K. E. Atkinson, *An Introduction to Numerical Analysis (2^{nd} edition)*, John Wiley & Sons, New York, 1989.

[3] W. Gautschi, *Numerical Analysis (An Introduction)*, Birkhäuser, Boston, 1997.

[4] D. Goldberg, *What every computer scientist should know about floating-point arithmetic*, ACM Computing Surveys, vol. 23, no. 1, March 1991.

[5] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, 1996.

[6] M. L. Overton, *Numerical Computing with IEEE Floating Point Arithmetic*, SIAM, Philadelpia, 2001.

[7] A. Ralston, P. Rabinowitz, *A First Course in Numerical Analysis*, McGraw–Hill, Singapore, 1978.

[8] G. W. Stewart, J.-G. Sun, *Matrix Perturbation Theory*, Academic Press, 1990.

[9] J. H. Wilkinson, *Rounding Errors in Algebraic Processes*, Notes on Applied Science No. 32, Her Majesty's Stationery Office, London, 1963. (Also published by Prentice–Hall, Englewood Cliffs, NJ, USA. Reprinted by Dover, New-York, 1994, ISBN 0–486–67999–5.)