1

Numeričko deriviranje

1.1 Osnovne formule

Formule za numeričko deriviranje daju aproksimaciju derivacije funkcije izračunatu pomoću vrijednosti funkcije u konačno mnogo točaka. Najjednostavnije takve formule dobivamo polazeći od definicije derivacije

$$u'(x) = \lim_{h \to 0} \frac{u(x+h) - u(x)}{h} \tag{1.1}$$

iz koje se vidi da $u'(x)$ možemo aproksimirati izrazom

$$\frac{u(x+h) - u(x)}{h}.$$

Ovakva formula može dati dobru aproksimaciju derivacije ukoliko vrijedi:

- Korak h je dovoljno malen;
- Derivacija $u'(x)$ je neprekidna.

Lako je vidjeti da formula može dati neprihvatljive rezultate i za vrlo male vrijednosti koraka h ako derivacija ima skokove (pokažite primjerom).

Formule za numeričko deriviranje koristimo kada nam analitički izraz za derivaciju funkcije nije poznat. Isto tako, ako umjesto funkcije poznamo samo tabelu funkcijskih vrijednosti u konačno mnogo točaka tada derivaciju možemo "računati" samo numerički. U takvoj situaciji pretpostavljamo da kroz tabulirane vrijednosti prolazi neprekidno derivabilna funkcija.

Na temelju formule (1.1) dobivamo sljedeće formule za numeričko deriviranje:

$$u'(x) \approx \frac{u(x+h) - u(x)}{h} = \Delta^+ u(x) \quad \text{diferencija unaprijed},$$
$$u'(x) \approx \frac{u(x) - u(x-h)}{h} = \Delta^- u(x) \quad \text{diferencija unazad},$$
$$u'(x) \approx \frac{u(x+h) - u(x-h)}{2h} = \Delta^0 u(x) \quad \text{centralna diferencija}.$$

Točnost ovih formula lako se nalazi pomoću Taylorovog razvoja uz pretpostavku glatkoće funkcije. Na primjer, po teoremu srednje vrijednosti imamo

$$u(x+h) = u(x) + u'(x)h + \frac{1}{2}u''(\xi)h^2$$

za neki $\xi \in J(x, x+h)$ gdje smo uveli oznaku $J(a,b) = (\min\{a,b\}, \max\{a,b\})$. Odavde izlazi

$$u'(x) = \frac{u(x+h) - u(x)}{h} - \frac{1}{2}u''(\xi)h = \Delta^+ u(x) - \frac{1}{2}u''(\xi)h.$$

Izraz

$$e_d(x;h) = u'(x) - \frac{u(x+h) - u(x)}{h}$$

nazivamo greška diskretizacije ili greška odsjecanja (*engl. truncation error*). Za diferenciju unaprijed $e_d(x;h) = -\frac{1}{2}u''(x)h$. Posve analogna formula vrijedi i za diferenciju unazad.

Ukoliko grešku diskretizacije možemo ocijeniti na sljedeći način:

$$|e_d(x;h)| \leq Mh^n,$$

gdje M ne ovisi o h i x, onda kažemo da formula numeričkog deriviranja ima red točnost jednak n. Evidentno diferencija unaprijed (i unazad) aproksimira derivaciju s točnošću prvog reda ako je druga derivacija funkcije ograničena.

Za centralnu diferenciju imamo:

$$u(x+h) = u(x) + u'(x)h + \frac{1}{2}u''(x)h^2 + \frac{1}{6}u'''(\xi^+)h^3, \quad \xi^+ \in J(x, x+h),$$

$$u(x-h) = u(x) - u'(x)h + \frac{1}{2}u''(x)h^2 - \frac{1}{6}u'''(\xi^-)h^3, \quad \xi^- \in J(x, x-h).$$

Oduzimanjem dobivamo

$$\frac{u(x+h) - u(x-h)}{2h} = u'(x) + \frac{1}{12}[u'''(\xi^+) + u'''(\xi^-)]h^2.$$

Dakle, u ovom slučaju je

$$e_d(x;h) = \frac{1}{12}[u'''(\xi^+) + u'''(\xi^-)]h^2$$

odnosno formula ima drugi red točnosti ako je treća drerivacija funkcije ograničena. Ukoliko je funkcija u''' neprekidna, onda izraz za grešku diskretizacije možemo pojednostaviti. Tada, naime postoji točka $\xi \in J(\xi^-, \xi^+)$ takva da je $u'''(\xi) = \frac{1}{2}(u'''(\xi^+) + u'''(\xi^-))$ pa imamo

$$e_d(x;h) = \frac{1}{6}u'''(\xi)h^2.$$

Aproksimaciju druge derivacije možemo dobiti pomoću već izvedenih aproksimacija prve derivacije. Na primjer,

$$\Delta^2 u(x) = \frac{\Delta^+ u(x) - \Delta^- u(x)}{h} = \frac{u(x+h) - 2u(x) + u(x-h)}{h^2}.$$

To je centralna diferencija drugog reda i ona je drugog reda točnosti (Zadatak Z1.1).

Kod numeričkog deriviranja važnu ulogu imaju greške zaokruživanja. Analizirajmo njihov utjecaj na primjeru diferencije unaprijed. Pretpostavimo da su vrijednosti $u(x)$ i $u(x+h)$ poznate s nekim greškama δ_1, odnosno δ_2, $|\delta_1|$, $|\delta_2| \leq \delta$ Tada imamo

$$(\Delta^+ u(x))_{izr} = \frac{[u(x+h) + \delta_1] - [u(x) + \delta_2]}{h} = \Delta^+ u(x) + \frac{\delta_1 - \delta_2}{h}.$$

Izraz $e_z(h) = (\delta_1 - \delta_2)/h$ je greška zaokruživanja, a ukupna greška je

$$|e_t(x;h)| = |e_d(x;h) + e_z(h)| \leq \frac{1}{2}Mh + \frac{2\delta}{h}, \quad M = \max|u''(x)|.$$

Funkcija $G(h) = Mh/2 + 2\delta/h$ postiže svoj minimum u

$$h_{min} = 2\sqrt{\frac{\delta}{M}}, \quad G(h_{min}) = 2\sqrt{\delta M}. \tag{1.2}$$

Smanjivanjem koraka h ispod h_{min} greške zaokruživanja dominiraju nad greškama diskretizacije i ukupna greška se povećava. (Tu nismo uzeli u obzir dokidanje značajnih znamenki.)

1.2 Opći pristup izvođenju diferencijskih formula

Da bismo našli formule za numeričko (približno) deriviranje funkcije koja nam je poznata samo u konačno mnogo točaka možemo se poslužiti sljedećim postupkom: formiramo interpolacijski polinom za zadane funkcijske vrijednosti i derivaciju funkcije aproksimiramo derivacijom interpolacijskog polinoma.

Tim postupkom odmah dobivamo i ocjenu greške. Preciznije, neka je f zadana funkcija definirana na nekom segmentu $[a,b]$ i neka su $x_0, x_1, \ldots, x_n$ točke u kojima poznamo vrijednosti funkcije. Ako je p_n interpolacijski polinom provučen kroz te točke onda je

$$f(x) = p_n(x) + f[x_0, x_1, \ldots, x_n, x]w_n(x), \quad w_n(x) = \prod_{j=0}^{n}(x - x_j).$$

Za k-tu derivaciju imamo

$$f^{(k)}(x) = p_n^{(k)}(x) + \frac{d^k}{dx^k}(f[x_0, x_1, \ldots, x_n, x]w_n(x)).$$

M. Jurak 16. studenog 2005.

Greška koju smo učinili aproksimirajući $f^{(k)}(x)$ sa $p_n^{(k)}(x)$ jednaka je $\frac{d^k}{dx^k}R_n(x)$ gdje je $R_n(x) = f[x_0, x_1, \ldots, x_n, x]w_n(x)$ greška interpolacije. Iako ovako izvedene formule daju aproksimativnu derivaciju funkcije u cijeloj domeni $[a, b]$ u praksi se najčešće takve formule koriste samo u interpolacijskim točkama x_j. U tim točkama se izraz za grešku formule numeričke derivacije pojednostavljuje.

Ilustrirajmo taj postupak na jednom primjeru. Funkcija f zadana je u tri točke x_0, x_1, x_2. Želimo naći formulu za približno računanje prve derivacije. Interpolacijski polinom kroz te točke je dan formulom

$$p_2(x) = f(x_0) + (x - x_0)f[x_0, x_1] + (x - x_0)(x - x_1)f[x_0, x_1, x_2],$$

a formula za grešku interpolacije nam daje

$$f(x) = p_2(x) + (x - x_0)(x - x_1)(x - x_2)f[x_0, x_1, x_2, x].$$

Deriviranjem ove formule dobivamo

$$f'(x) = p_2'(x) + [(x - x_1)(x - x_2) + (x - x_0)(x - x_2) + (x - x_0)(x - x_1)] \times f[x_0, x_1, x_2, x] + (x - x_0)(x - x_1)(x - x_2)\frac{d}{dx}f[x_0, x_1, x_2, x].$$

Član $p_2'(x)$ je aproksimacija derivacije $f'(x)$ a posljednja dva člana predstavljaju grešku aproksimacije. Očito je da se izraz za grešku pojednostavljuje ako za x uzmemo jednu od interpolacijskih točaka x_0, x_1, x_2. Uzmimo na primjer $x = x_0$; tada je

$$f'(x_0) \approx p_2'(x_0) = f[x_0, x_1] + (x_0 - x_1)f[x_0, x_1, x_3],$$

dok je greška te formule

$$(x_0 - x_1)(x_0 - x_2)f[x_0, x_1, x_2, x_0].$$

Tu smo pretpostavili da je izraz $\frac{d}{dx}f[x_0, x_1, x_2, x]$ ograničen u okolini točke x_0. Ukoliko f ima neprekidnu treću derivaciju, onda znamo da je

$$f[x_0, x_1, x_2, x_0] = \frac{f^{(3)}(\xi)}{3!}$$

za neki ξ iz intervala koji sadrži točke x_0, x_1 i x_2. U slučaju ekvidistantnih točaka $x_1 = x_0 + h$, $x_2 = x_0 + 2h$ dobivamo sljedeću formulu

$$f'(x_0) = \frac{-3f(x_0) + 4f(x_0 + h) - f(x_0 + 2h)}{2h} + \frac{h^2}{3}f^{(3)}(\xi), \tag{1.3}$$

za neki $\xi \in (x_0, x_0 + 2h)$. Računajući pak derivaciju u točki x_2 možemo dobiti formulu numeričke derivacije koja koristi točke lijevo od zadane točke (zadatak Z1.4)

Drugi postupak za nalaženje formula višeg reda za numeričko deriviranje bazira se na Taylorovom razvoju i metodi neodređenih koeficijenata. Ilustrirajmo to na primjeru formule četvrtog reda točnosti za prvu derivaciju.

Uzmimo da imamo ekcidistantno raspoređene točke $x_i = x_o + ih$, $i = 0, 1, 2, \ldots$, gdje je $h > 0$ korak mreže. Želimo naći formulu za aproksimaciju derivacije funkcije $f(x)$ u točki $x = x_i$, koristeći vrijednosti funkcije u točkama x_{i-2}, x_{i-1}, x_i, x_{i+1} i x_{i+2}. Radi jednostavnosti zapisa pisat ćemo $f_i = f(x_i)$, $f_i' = f'(x_i)$, $f_i'' = f''(x_i)$ itd. Prvo raspišemo Taylorove razvoje:

$$
\begin{aligned}
f_{i+2} &= f_i + 2hf_i' + \frac{4h^2}{2}f_i'' + \frac{8h^3}{3!}f_i''' + \frac{16h^4}{4!}f_i^{(iv)} + \frac{32h^5}{5!}f_i^{(v)} + \cdots \\
f_{i+1} &= f_i + hf_i' + \frac{h^2}{2}f_i'' + \frac{h^3}{3!}f_i''' + \frac{h^4}{4!}f_i^{(iv)} + \frac{h^5}{5!}f_i^{(v)} + \cdots \\
f_{i-1} &= f_i - hf_i' + \frac{h^2}{2}f_i'' - \frac{h^3}{3!}f_i''' + \frac{h^4}{4!}f_i^{(iv)} - \frac{h^5}{5!}f_i^{(v)} + \cdots \\
f_{i-2} &= f_i - 2hf_i' + \frac{4h^2}{2}f_i'' - \frac{8h^3}{3!}f_i''' + \frac{16h^4}{4!}f_i^{(iv)} - \frac{32h^5}{5!}f_i^{(v)} + \cdots
\end{aligned}
$$

Zatim formiramo izraz

$$
\begin{aligned}
\alpha f_{i+2} + \beta f_{i+1} + \gamma f_i + \delta f_{i-1} + \epsilon f_{i-2} & \\
&= (\alpha + \beta + \gamma + \delta + \epsilon) f_i + h(2\alpha + \beta - \delta - 2\epsilon) f_i' \\
&+ \frac{h^2}{2}(4\alpha + \beta + \delta + 4\epsilon) f_i'' + \frac{h^3}{3!}(8\alpha + \beta - \delta - 8\epsilon) f_i''' \\
&+ \frac{h^4}{4!}(16\alpha + \beta + \delta + 16\epsilon) f_i^{(iv)} + \frac{h^5}{5!}(32\alpha + \beta - \delta - 32\epsilon) f_i^{(v)} + \cdots
\end{aligned}
$$

Da bismo postigli maksimalnu točnost formule određujemo koeficijente tako zadovoljavaju sljedeći sustav jednadžbi:

$$
\begin{aligned}
\alpha + \beta + \gamma + \delta + \epsilon &= 0 \\
2\alpha + \beta - \delta - 2\epsilon &= 1 \\
4\alpha + \beta + \delta + 4\epsilon &= 0 \\
8\alpha + \beta + \delta + 8\epsilon &= 0 \\
16\alpha + \beta + \delta + 16\epsilon &= 0.
\end{aligned}
$$

Rješenje je

$$
\alpha = -\frac{1}{12}, \quad \beta = \frac{2}{3}, \quad \gamma = 0, \quad \delta = -\frac{2}{3}, \quad \epsilon = \frac{1}{12}.
$$

Time dobivamo formulu

$$
\frac{-f_{i+2} + 8f_{i+1} - 8f_{i-1} + f_{i-2}}{12h} = f_i' - \frac{h^4}{30}f_i^{(v)} + \cdots
$$

M. Jurak 16. studenog 2005.

Zadatak. Istom metodom dokažite i ove formule:

$$\frac{2f_{i+1}+3f_i-6f_{i-1}+f_{i-2}}{6h} = f_i' + \frac{h^3}{12}f_i^{(iv)} + \cdots$$
$$\frac{-f_{i+2}+6f_{i+1}-3f_i-2f_{i-1}}{6h} = f_i' - \frac{h^3}{12}f_i^{(iv)} + \cdots$$

1.3 Richardsonova ekstrapolacija

Richardsonova ekstrapolacija je općenit postupak kojim se mogu dobiti formule višeg reda točnosti u slučaju kada računamo neku veličinu $\phi(h)$ koja ovisi o malom parametru h. Vrijednost koja nas zanima je $\lim_{h\to 0}\phi(h) = L$, no mi možemo izračunati jedino $\phi(h)$ za male vrijednosti parametra h. Jednu takvu situaciju predstavljaju formule za numeričko deriviranje; tu je $\phi(h)$ dana formula numeričkog deriviranja s korakom h, a L je stvarna vrijednost derivacije.

Da bismo mogli primijeniti Richardsonovu ekstrapolaciju greška $L-\phi(h)$ mora se dati razviti u red potencija po parametru h. Na primjer, možemo imati

$$L = \phi(h) + a_2h^2 + a_4h^4 + a_6h^6 + \cdots \tag{1.4}$$

Takva se situacija javlja u slučaju centralne diferencije ako je funkcija f koju deriviramo analitička. Naime, tada imamo

$$f(x+h) = \sum_{k=0}^{\infty}\frac{h^k}{k!}f^{(k)}(x),$$

$$f(x-h) = \sum_{k=0}^{\infty}(-1)^k\frac{h^k}{k!}f^{(k)}(x).$$

Tada je

$$f'(x) = \frac{f(x+h)-f(x-h)}{2h} - \frac{h^2}{3!}f^{(3)}(x) - \frac{h^4}{5!}f^{(5)}(x) - \cdots.$$

Ukoliko vrijedi razvoj (1.4) za $\phi(h/2)$ imamo sljedeći razvoj:

$$L = \phi(\frac{h}{2}) + a_2\frac{h^2}{2^2} + a_4\frac{h^4}{2^4} + a_6\frac{h^6}{2^6} + \cdots \tag{1.5}$$

Razvoje (1.4) i (1.5) možemo iskoristiti da skratimo član s koeficijentom a_2. Dobivamo

$$3L = 4L - L = 4\phi(\frac{h}{2}) - \phi(h) - \frac{3}{4}a_4h^4 - \frac{15}{16}a_6h^6 - \cdots$$

odnosno

$$L = \frac{4}{3}\phi(\frac{h}{2}) - \phi(h) - \frac{1}{4}a_4h^4 - \frac{5}{16}a_6h^6 - \cdots$$

Vidimo dakle da izraz

$$\phi_1(h) = \frac{4}{3}\phi(\frac{h}{2}) - \phi(h)$$

aproksimira L s četvrtim redom točnosti.

Proces Richardsonove ekstrapolacije može se nastaviti i dalje. Za funkciju ϕ_1 vrijedi

$$L = \phi_1(h) + b_4 h^4 + b_6 h^6 + \cdots$$
$$L = \phi_1(\frac{h}{2}) + b_4 \frac{h^4}{2^4} + b_6 \frac{h^6}{2^6} + \cdots$$

s nekim koeficijentima b_n koje nije potrebno poznavati. Sličnom manipulacijom kao ranije dobivamo

$$4^2 L - L = 4^2 \phi_1(\frac{h}{2}) - \phi_1(h) - \frac{3}{4} b_6 h^6 - \cdots$$

odnosno

$$L = \frac{4^2}{4^2-1}\phi_1(\frac{h}{2}) - \frac{1}{4^2-1}\phi(h) - \frac{3}{4}\frac{1}{4^2-1} b_6 h^6 - \cdots$$

Time smo došli do aproksimacije šestog reda. Postupak se može nastaviti dalje. Lako je uočiti da općenito dobivamo izraz

$$\phi_n(h) = \frac{4^n \phi_{n-1}(\frac{h}{2}) - \phi_{n-1}(h)}{4^n - 1}$$

čiji je red aproksimacije $2n+2$. Shema računanja Richardsonove aproksimacije može se predstaviti sljedećom tablicom.

$O(h^2)$	$O(h^4)$	$O(h^6)$	$O(h^8)$
$\phi(h)$			
$\phi(\frac{h}{2})$	$\phi_1(h)$		
$\phi(\frac{h}{4})$	$\phi_1(\frac{h}{2})$	$\phi_2(h)$	
$\phi(\frac{h}{8})$	$\phi_1(\frac{h}{4})$	$\phi_2(\frac{h}{2})$	$\phi_3(h)$

Tablica se izračunava po stupcima. Prvo se izračuna prvi stupac do željene dubine. Zatim se izračunava drugi. On koristi samo vrijednosti prvog stupca. Postupak se nastavlja dok se ne dođe do dubine koja je odabrana.

1.4 Metoda konačnih diferencija

Ovdje ćemo pokazati kako se primijenjuju formule za numeričko deriviranje u diskretizaciji rubnih problema za obične diferencijalne jednadžbe.

M. JURAK 16. studenog 2005.

Zadan je rubni problem za običnu diferencijalnu jednadžbu

$$-\frac{d}{dx}\left(k(x)\frac{du}{dx}\right) + w(x)\frac{du}{dx} + d(x)u = f(x) \quad x \in (0,1) \tag{1.6}$$

$$u(0) = \alpha, \quad u(1) = \beta. \tag{1.7}$$

Zadane vrijednosti su funkcije k, w, d i f te brojevi α i β. Traži se funkcija u, definirana na segmantu $[0,1]$, koja zadovoljava diferencijalnu jednadžbu u svakoj točki $x \in (0,1)$ i prima zadane vrijednosti u rubnim točkama. Pretpostavit ćemo da su sve zadane funkcije glatke, da vrijedi

$$k(x) \geq k_0 > 0, \quad d(x) \geq 0 \quad \forall x \in [0,1],$$

te da je derivacija funkcije w dovoljno mala, npr. strogo manja od $4k_0$. Uz takve se uvjete može pokazati da rubni problem (1.6), (1.7) ima jedinstveno rješenje.

Da bismo konstruirali približno rješenje problema uvedimo na segmentu $[0,1]$ ekvidistantnu mrežu

$$0 = x_0 < x_1 < x_2 < \cdots < x_n < x_{n+1} = 1,$$

pri čemu je

$$x_j = jh,\ j = 0, 1, \ldots, n+1, \quad h = \frac{1}{n+1}.$$

Time smo segment $[0,1]$ razbili na $n+1$ segmenata $[x_j, x_{j+1}]$. Označimo još sredinu svakog od tih segmenata sa

$$x_{j+1/2} = \frac{1}{2}(x_j + x_{j+1}).$$

Diferencijalne operatore koji se javljaju na lijevoj strani u (1.6) aproksimirat ćemo tako da derivacije zamijenimo diferencijskim kvocijentima računatim u točkama mreže. Da bismo pojednostavili zapis uvesti ćemo oznake

$$k_{j+1/2} = k(x_{j+1/2}),\ w_j = w(x_j),\ d_j = d(x_j),\ f_j = f(x_j).$$

Imamo:

$$\left(p(x)\frac{du}{dx}\right)\Big|_{x=x_j} \approx p_j \frac{u(x_{j+1}) - u(x_{j-1})}{h},$$

Diferencijalni operator drugog reda aproksimiramo u dva koraka:

$$\begin{aligned}
&-\frac{d}{dx}\big(k(x)\frac{du}{dx}\big)\Big|_{x=x_j} \\
&\quad \approx -\frac{1}{h}\big[k_{j+1/2}\big(\frac{du}{dx}\big)\Big|_{x=x_{j+1/2}} - k_{j-1/2}\big(\frac{du}{dx}\big)\big|_{x=x_{j-1/2}}\big] \\
&\quad \approx -\frac{1}{h}\big[k_{j+1/2}\frac{u(x_{j+1}) - u(x_j)}{h} - k_{j-1/2}\frac{u(x_j) - u(x_{j-1})}{h}\big] \\
&\quad = \frac{1}{h^2}\big[(k_{j+1/2} + k_{j-1/2})u(x_j) - k_{j+1/2}u(x_{j+1}) - k_{j-1/2}u(x_{j-1})\big].
\end{aligned}$$

Sumirajući razne doprinose vidimo da vrijednosti $u(x_j)$ zadovoljavaju približnu jednakost

$$(k_{j+1/2}+k_{j-1/2}+h^2d_j)u(x_j)-(k_{j+1/2}-\frac{1}{2}hw_j)u(x_{j+1}) \\ -(k_{j-1/2}+\frac{1}{2}hw_j)u(x_{j-1})\approx h^2f_j \quad (1.8)$$

za $j=1,2,\ldots,n$ te $u(x_0)=\alpha$ i $u(x_{n+1})=\beta$. Aproksimaciju $u_j\approx u(x_j)$ točnog rješenja definiramo tako da u (1.8) približnu jednakost zamijenimo s jednakošću. Time dobivamo diferencijsku jednadžbu

$$(k_{j+1/2}+k_{j-1/2}+h^2d_j)u_j-(k_{j+1/2}-\frac{1}{2}hw_j)u_{j+1}-(k_{j-1/2}+\frac{1}{2}hw_j)u_{j-1}=h^2f_j$$

za $j=1,2,\ldots,n$. Rubne uvjete zadovoljavamo neposredno: $u_0=\alpha$ i $u_{n+1}=\beta$.

Uvedimo radi jednostavnosti zapisa sljedeće oznake

$$a_j=k_{j+1/2}+k_{j-1/2}+h^2d_j \quad j=1,2,\ldots,n \quad (1.9)$$

$$b_j=k_{j-1/2}+\frac{1}{2}hw_j \quad j=1,2,\ldots,n \quad (1.10)$$

$$c_j=k_{j+1/2}-\frac{1}{2}hw_j \quad j=1,2,\ldots,n. \quad (1.11)$$

Prebacivanjem poznatih vrijednosti $u_0=\alpha$ i $u_{n+1}=\beta$ na desnu stranu dobivamo sljedeći sustav jednadžbi:

$$\begin{aligned} a_1u_1-c_1u_2&=h^2f_1+b_1\alpha \\ -b_ju_{j-1}+a_ju_j-c_ju_{j+1}&=h^2f_j \quad j=2,\ldots,n-1 \\ -b_nu_{n-1}+a_nu_n&=h^2f_n+c_n\beta, \end{aligned}$$

ili u matričnom zapisu

$$\begin{bmatrix} a_1 & -c_1 & & & & \\ -b_2 & a_2 & -c_2 & & & \\ & -b_3 & a_3 & -c_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -b_{n-1} & a_{n-1} & -c_{n-1} \\ 0 & & & & -b_n & a_n \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_{n-1} \\ u_n \end{bmatrix} = \begin{bmatrix} h^2f_1+b_1\alpha \\ h^2f_2 \\ h^2f_3 \\ \vdots \\ h^2f_{n-1} \\ h^2f_n+c_n\beta \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ \vdots \\ F_{n-1} \\ F_n \end{bmatrix}$$

Dobili smo sustav s trodijagonalnom matricom. Gaussove eliminacije bez pivotiranja svode se na dvije petlje. U prvoj se eliminiraju b-ovi i modificiraju a-ovi, dok c-ovi ostaju nepromijenjeni. U drugoj petlji se supstitucijama nalazi rješenje. Preciznije, u prvoj petlji računamao

$$a_i=a_i-\frac{b_i}{a_{i-1}}c_{i-1}, \quad i=2,\ldots,n,$$

M. JURAK 16. studenog 2005.

$$F_i = F_i + \frac{b_i}{a_{i-1}} F_{i-1}, \quad i = 2, \ldots, n,$$

i time dobivamo sustav oblika

$$\begin{bmatrix} a_1 & -c_1 & & & & \\ & a_2 & -c_2 & & & \\ & & a_3 & -c_3 & & \\ & & & \ddots & \ddots & \\ & & & & a_{n-1} & -c_{n-1} \\ 0 & & & & & a_n \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_{n-1} \\ u_n \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ \vdots \\ F_{n-1} \\ F_n \end{bmatrix}$$

gdje su koeficijenti a_i i F_i promijenjeni. Povratne supstitucije sada daju

$$\begin{aligned} u_n &= F_n/a_n \\ u_i &= (F_i + c_i u_{i+1})/a_i \quad i = n-1, \ldots, 1. \end{aligned}$$

`Rubni uvjeti` oblika $u(0) = \alpha$ i $u(1) = \beta$, u kojima se zadaje vrijednost rješenja u graničnim točkama nazivaju se Dirichletovi rubni uvjeti. Pored njih moguće je postaviti rubne uvjete koji uključuju vrijednost prve derivacije u graničnim točkama. Na primjer, $u'(1) = \beta$ (Neumannov rubni uvjet) ili $u'(1) + \gamma u(1) = \beta$, (Robinov rubni uvjet) gdje su β i γ zadani brojevi. Postavlja se pitanje, ako diskretiziramo rubni uvjet koji sadrži derivaciju u sebi hoće li matrica ostati trodijagonalna? Odgovor na to pitanje ovisi o načinu diskretizacije rubnog uvjeta. Uzmimo na primjer, mješoviti rubni uvjet

$$u'(1) + \gamma u(1) = \beta$$

i diskretizirajmo derivaciju diferencijom unatrag. Dobivamo

$$\frac{u_{n+1} - u_n}{h} + \gamma u_{n+1} = \beta,$$

što je dodatna jednadžba za u_{n+1}, pa tu vrijednost treba tretirati kao varijablu. Dimenzija sustava se povećala za jedan ali nova jednadžba ima isti oblik kao i prethodne te se trodijagonalna struktura matrice ne kvari.

Uočimo da ovakva diskretizacija ima i jedan nedostatak. Diskretizacija diferencijalne centralnim diferencijama ima drugi red točnosti, dok diskretizacija rubnog uvjeta diferencijom unatrag ima samo prvi red točnosti. Ako bismo željeli imati konsistentnu diskretizaciju drugog reda, morali bismo iskoristiti točniju formulu za numeričko deriviranje koja nužno uključuje vrijednost funkcije i u točki x_{n-1}. Time bi struktura matrice u zadnjem retku bila narušena.

Pokazat ćemo sada malo drugačiji pristup rješavanju dobivenog sustava, koji vodi na metodu koja je fleksibilnija u odnosu na rubne uvjete.

1.4.1 Thomasov algoritam

Želimo riješiti diferencijsku jednadžbu

$$-b_i u_{i-1} + a_i u_i - c_i u_{i+1} = h^2 f_i, \quad i = 1, \ldots, n \tag{1.12}$$

$$u_0 = \alpha, \;\; u_{n+1} = \beta. \tag{1.13}$$

Na osnovu Gaussovih eliminacija znamo da rješenje možemo tražiti u obliku

$$u_i = p_{i+1} u_{i+1} + q_{i+1}, \quad i = 0, 1, \ldots, n \tag{1.14}$$

pri čemu su nam za sad koeficijenti p_i, q_i nepoznati. Uvrštavanjem u (1.12) dobivamo

$$-b_i(p_i u_i + q_i) + a_i u_i - c_i u_{i+1} = h^2 f_i$$

što nakon sređivanja daje

$$u_i = \frac{c_i}{a_i - b_i p_i} u_{i+1} + \frac{h^2 f_i + b_i q_i}{a_i - b_i p_i} \tag{1.15}$$

za $i = 1, 2, \ldots, n$. Uspoređivanjem s (1.14) zaključujemo da mora vrijediti

$$p_{i+1} = \frac{c_i}{a_i - b_i p_i}, \quad q_{i+1} = \frac{h^2 f_i + b_i q_i}{a_i - b_i p_i}$$

za $i = 1, 2, \ldots, n$. Time smo dobili rekurzije na osnovu kojih možemo izračunati sve koeficijente ako znamo p_1 i q_1. Te koeficijente dobivamo iz rubnog uvjeta. Općenito mora vrijediti

$$u_0 = p_1 u_1 + q_1,$$

pa u slučaju Dirichletovog rubnog uvjeta $u_0 = \alpha$ treba uzeti

$$p_1 = 0, \quad q_1 = \alpha.$$

Pomoću (1.15) možemo odrediti čitavo rješenje ako znamo u_{n+1}, ali ta vrijednost je dana rubnim uvjetom: $u_{n+1} = \beta$. Time smo došli do Thomasovog algoritma.

Algoritam 1 rješava jednadžbu (1.6) uz rubne uvjete (1.7).

Izvedivost i stabilnost Thomasovog algoritma ovisi o tome da li nazivnik $a_i - b_i p_i$ može postati blizak ili jednak nuli. Algoritam nije izvediv ako je za neki indeks taj nazivnik jednak nuli. S druge strane, kada postane vrlo malen faktori p_i mogu postati veći od 1 što može voditi do gubitka stabilnosti algoritma.

Lema 1.1 Thomasov algoritam je provediv i stabilan ako je

$$|a_j| \geq |b_j| + |c_j|, \quad j = 1, \ldots, n, \tag{1.16}$$

te $c_j \neq 0$ za $j = 1, \ldots, n$.

M. JURAK 16. studenog 2005.

Algorithm 1 Thomasov algoritam

Ulaz: n, h, $(a_j)_{j=1}^n$, $(b_j)_{j=1}^n$, $(c_j)_{j=1}^n$, $(f_j)_{j=1}^n$
n = broj unutarnjih točaka mreže,
h = korak mreže,
$a_j, b_j c_j$ = koeficijenti iz formula (1.9)–(1.11),
f_j = vektor desne strane,
$p_1 = 0, q_1 = \alpha$
for $i = 1, 2, \ldots, n$ **do**
$p_{i+1} = c_i/(a_i - b_i p_i)$
$q_{i+1} = (h^2 f_i + b_i q_i)/(a_i - b_i p_i)$
end for
$u_{n+1} = \beta$
for $i = n, n-1, \ldots, 0$ **do**
$u_i = p_{i+1} u_{i+1} + q_{i+1}$
end for
Izlaz: $(u)_{i=0}^{n+1}$ = rješenje u točkama mreže.

Dokaz. Iz $p_1 = 0$ izlazi

$$|p_2| = |\frac{c_1}{a_1}| \leq 1.$$

Pokažimo indukcijom da za sve $j = 1, \ldots, n$ vrijedi

$$a_j - b_j p_j \neq 0 \quad \text{i} \quad |p_{j+1}| \leq 1.$$

Za $j = 1$ tvrdnja je provjerena. Općenito za $j > 1$ imamo

$$\begin{aligned} |a_j - b_j p_j| &\geq |a_j| - |b_j||p_j| \\ &\geq |a_j| - |b_j| \quad \text{pretpostavka indukcije } |p_j| \leq 1 \\ &\geq |c_j| > 0 \quad \text{pretpostavke leme.} \end{aligned}$$

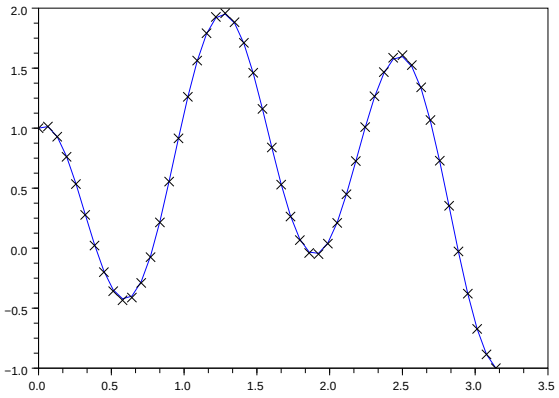
Time smo ujedno dobili

$$|p_{j+1}| = |\frac{c_j}{a_j - b_j p_j}| \leq 1. \quad \square$$

Primjer 1.1 Rubni problem

$$\begin{aligned} -u'' + u' = 2\sin(x) + 26\cos(5x) \quad &\text{na } (0, \pi) \\ u(0) = 1, \quad u(1) = -1, \end{aligned}$$

ima rješenje $u(x) = \sin(x) + \cos(5x)$. Aproksimacija s 50 točaka pokazana je na Slici 1.4.1. Točno rješenje je prikazano punom crtom, a diferencijska aproksimacija križićima. Variranjem broja točaka lako se vidi da je greška računata po formuli $err = \max_i |u_i - u(x_i)|$ proporcionalana s h^2 ($err \approx 3h^2$).



Slika 1.1:

Primjer 1.2 Rubni problem

$$\begin{aligned} -\varepsilon u'' + u' &= 1 \quad \text{na } (0,1) \\ u(0) = u(1) &= 0, \end{aligned}$$

ima jedinstveno rješenje

$$u(x) = x - (e^{-(1-x)/\varepsilon} - e^{-1/\varepsilon})/(1 - e^{-1/\varepsilon}).$$

Numerička aproksimacija rješenja ($\varepsilon = 1/70$) pokazana je an Slici 1.4.1 za 20 (lijeva) i 50 (desna) diskretizacijskih točaka. Vidimo da je stabilnost izgubljena. Matrica sustava ima oblik

$$\begin{bmatrix} a & -c & & & & \\ -b & a & -c & & & \\ & -b & a & -c & & \\ & & \ddots & \ddots & \ddots & \\ & & & -b & a & -c \\ 0 & & & & -b & a \end{bmatrix}, \qquad \begin{aligned} a &= 2\varepsilon \\ b &= \varepsilon + h/2 \\ c &= \varepsilon - h/2, \end{aligned}$$

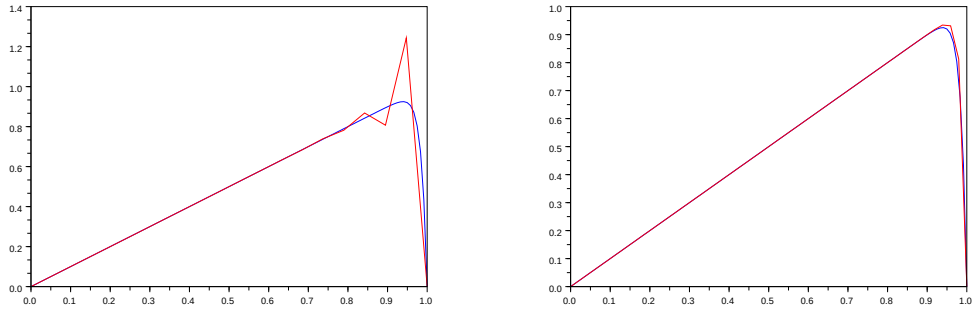
pa vidimo da je uvjet stabilnosti zadovoljen samo ako je $\varepsilon > h/2$. Drugim riječima, za zadani ε prostorni korak mora biti manji od 2ε.

1.4.2 Jednadžba $-\varepsilon u'' + w(x)u' = f(x)$

Analizirajmo nestabilnost koja se javila u Primjer 1.2 na modelnom rubnom problemu:

$$-\varepsilon u'' + w(x)u' = f(x), \quad x \in (0,1) \tag{1.17}$$

$$u(0) = u(1) = 0. \tag{1.18}$$



Slika 1.2: Rješenje za $\varepsilon = 1/70$ te 20 i 50 diskretizacijskih točaka

Diskretni sustav glasi

$$\begin{bmatrix} a & -c_1 & & & & \\ -b_2 & a & -c_2 & & & \\ & -b_3 & a & -c_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -b_{n-1} & a & -c_{n-1} \\ 0 & & & & -b_n & a \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_{n-1} \\ u_n \end{bmatrix} = \begin{bmatrix} h^2 f_1 \\ h^2 f_2 \\ h^2 f_3 \\ \vdots \\ h^2 f_{n-1} \\ h^2 f_n \end{bmatrix} \tag{1.19}$$

pri čemi je

$$a = 2\varepsilon, \quad b_j = \varepsilon + \frac{h}{2} w_j, \quad c_j = \varepsilon - \frac{h}{2} w_j.$$

Uvjet stabilnosti (1.16)

$$2\varepsilon \geq |\varepsilon + \frac{h}{2} w_j| + |\varepsilon - \frac{h}{2} w_j|$$

bit će zadovoljen ako i samo ako su članovi na desnoj strani pozitivni, što daje uvjet

$$h \leq \frac{2\varepsilon}{\max_j |w_j|}.$$

Taj uvjet postaje suviše restriktivan kada je ε mali.

Napomena. Za mali ε diferencijalna jednadžba je singularno perturbirana. Član s drugom derivacijom množi se s malim brojem, tako da je jednadžba (1.17) bliska diferencijalnoj jednadžbi prvog reda

$$w(x)u' = f(x), \quad x \in (0, 1). \tag{1.20}$$

Rješenje jednadžbe (1.20) može zadovoljiti smo jedan rubni uvjet pa rješenja jednadžbi (1.17) i (1.20) ne mogu biti bliska. U Primjeru 1.2 smo vidjeli što se događa. Rješenje jednadžbe za $\varepsilon = 0$ je u tom slučaju $u(x) = x$ i dva rješenja

su bliska svugdje osim u okolini rubne točke $x = 1$, gdje se rješenje ε-jednadžbe eksponencijalno spušta na nulu.

Kada rješenje diferencijalne jednadžbe trpi veliku promjenu u blizini granice domene onada kažemo da ono ima rubni sloj. Položaj rubnog sloja ovisi o predznaku funkcije $w(x)$. Na primjer, jednadžba $-\varepsilon u'' - u' = 1$ uz rubne uvjete $u(0) = u(1) = 0$ ima rubni sloj u $x = 0$ (izračunajte točno rješenje).

Rubni slojevi su važni s numeričkog stanovišta jer izazivaju nestabilnost u numeričkim shemama. Posebnim tehnikama diskretizacije se takve nestabilnosti izbjegavaju. □

Struktura matrice diskretnog sustava je posljedica načina na koji smo diskretizirali prvu i drugu derivaciju. Pokušajmo promijeniti način diskretizacije derivacije prvog reda (diskretizaciju člana $-\varepsilon u''$ nećemo mijenjati). Umjesto centralne diferencije možemo iskoristiti 1) diferenciju unaprjed:

$$w(x_j)u'(x_j) \approx w_j(u_{j+1} - u_j)/h$$

ili 2) diferenciju unazad:

$$w(x_j)u'(x_j) \approx w_j(u_j - u_{j-1})/h$$

U oba slučaja dobivamo diskretnu jednadžbu u obliku (1.19) u kojoj je

$$\begin{aligned} a &= 2\varepsilon - hw_j, \quad c_j = \varepsilon - hw_j, \quad b_j = \varepsilon \quad \text{diferencija unaprijed} \\ a &= 2\varepsilon + hw_j, \quad c_j = \varepsilon, \quad b_j = \varepsilon + hw_j \quad \text{diferencija unazad.} \end{aligned}$$

Koja je od te dvije sheme stabilna? Odmah se vidi da to ovisi o predznaku od w_j. Za $w_j > 0$ treba uzeti diferenciju unazad, a za $w_j < 0$ unaprijed. Tamo gdje je $w_j = 0$ izbor je nebitan.

Tako dobivena shema naziva se upwind shema. Za nju je

$$a = 2\varepsilon + h|w_j|, \quad c_j = \varepsilon - hw_j^-, \quad b_j = \varepsilon + hw_j^+$$

gdje je

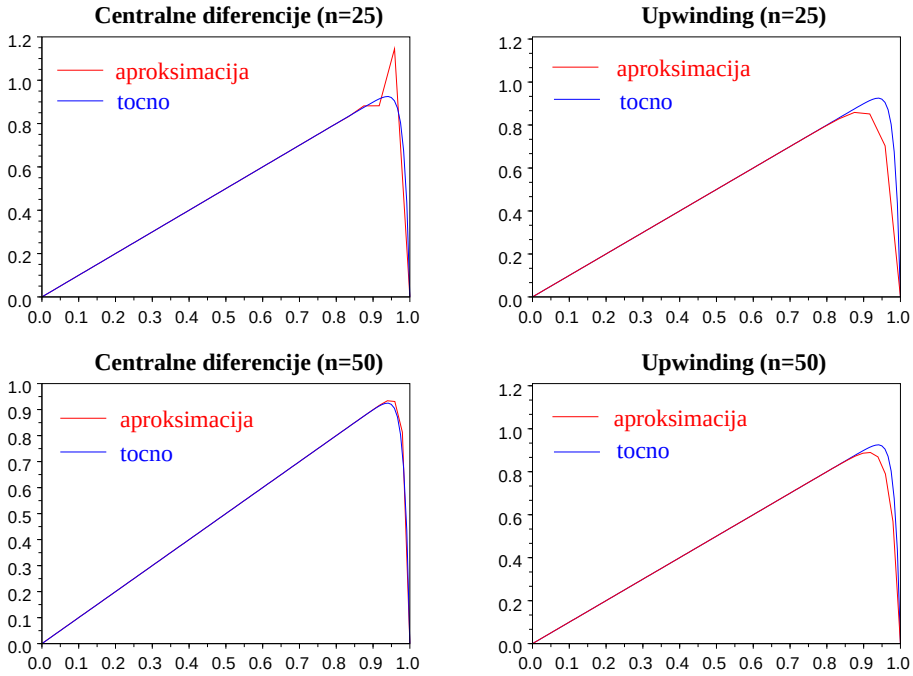
$$w^+ = \max(w, 0), \quad w^- = \min(w, 0).$$

Uočite da je $w^+ - w^- = |w|$ i $w^+ + w^- = w$. Shema se može zapisati u obliku

$$-\varepsilon(u_{j+1} - 2u_j + u_{j-1}) + hw_j^+(u_j - u_{j-1}) + hw_j^-(u_{j+1} - u_j) = h^2 f_j.$$

Zadatak. Usporedite centralne diferencije i upwind shemu na Primjeru 1.2. Pokažite da za $\varepsilon = 1/70$ i $n = 25$ te $n = 50$ dobivamo rezultate kao na Slici 1.4.2. Vidimo da je upwind shema stabilnija i da nam dozvoljava izračunati rješenje na grubljoj diskretizacijskoj mreži, bez oscilacije. S druge strane, evidentno smo upwindingom umanjili točnost sheme. To je za očekivati jer smo prvu derivaciju diskretizirali formulom prvog reda točnosti.

M. JURAK 16. studenog 2005.



Slika 1.3: Rješenje za $\varepsilon = 1/70$ te 25 i 50 diskretizacijskih točaka

1.4.3 Lokalna greška diskretizacije

1.4.4 Jednadžba $-u''(x) = f(x)$

U ovoj sekciji promatramo najjednostavniji primjer rubnog problema za običnu diferencijalnu jednadžbu. Naša motivacija nije važnost tog problema već želja da s minimalnim tehničkim komplikacijama ilustriramo analizu metode konačnih diferencija koja je primijenjiva i na daleko složenije zadaće.

Pogledajmo sada specijalan slučaj jednadžbe (1.6)

$$-u''(x) = f(x), \quad x \in (0,1), \tag{1.21}$$

$$u(0) = u(1) = 0. \tag{1.22}$$

Diskretiziran sustav glasi

$$\frac{1}{h^2}\begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_{n-1} \\ u_n \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_{n-1} \\ f_n \end{bmatrix} \tag{1.23}$$

Vidimo da diskretizacijom operatora $\mathcal{A} = -\frac{d^2}{dx^2}$ dobivamo matricu

$$A_h = \frac{1}{h^2}\begin{bmatrix} 2 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 2 & \end{bmatrix}. \tag{1.24}$$

Želja nam je proučiti svojstva te matrice i pokazati da su ona analogna svojstvima diferencijalnog operatora $\mathcal{A}$.

Diferencijalnu jednažbu možemo promatrati kao operatorsku jednadžbu na nekim funkcijskim prostorima. Uvedimo stoga oznaku $C^k([0,1])$ za linearan prostor svih k-puta neprekidno derivabilnih funkcija na segmentu $[0,1]$.[1] Umjesto $C^0([0,1])$ (prostor neprekidnih funkcija na segmentu) pisat ćemo $C([0,1])$. Uvedimo još prostor

$$C_0^2([0,1]) = \{u \in C^2([0,1]) \colon u(0) = u(1) = 0\}.$$

Evidentno je $\mathcal{A}$ linearan operator na sljedećim prostorima:

$$\mathcal{A} \colon C_0^2([0,1]) \to C([0,1]).$$

Lako je pokazati (zadatak) da je operator $\mathcal{A}$ bijekcija na tim prostorima. Riješiti postavljenu rubnu zadaću znači izračunati za zadano $f \in C([0,1])$ funkciju $\mathcal{A}^{-1}f$, gdje je $\mathcal{A}^{-1} \colon C([0,1]) \to C_0^2([0,1])$ inverz operatora $\mathcal{A}$.

U diskretnom slučaju operatoru $\mathcal{A}^{-1}$ odgovara inverzna matrica A_h^{-1}. Pitanje je što možemo zaključiti o ta dva operatora.

Korisna svojstva inverznog opertora možemo dobiti na osnovu principa maksimuma. Radi se o sljedećem: ako funkcija $u \in C_0^2([0,1])$ zadovoljava diferencijalnu nejednakost

$$-u''(x) < 0 \quad x \in (0,1),$$

onda ona ne može imati lokalni maksimum u intervalu $(0,1)$. Zaista, u točki lokalnog maksimuma $x_0 \in (0,1)$ vrijedilo bi $u'(x_0) = 0$ i $u''(x_0) \leq 0$. Ova druga nejednakost je u suprotnosti s gornjom pretpostavkom. Na temelju toga i rubnog uvjeta $u(0) = u(1) = 0$ zaključujemo da vrijedi $u(x) \leq 0$ za sve $x \in [0,1]$.

Time smo dobili zaključak:

$$\mathcal{A}u < 0 \quad \text{na } (0,1) \quad \Rightarrow \quad u \leq 0 \quad \text{na } [0,1].$$

Posve analogno se dokazuje (a slijedi i množenjem s -1)

$$\mathcal{A}u > 0 \quad \text{na } (0,1) \quad \Rightarrow \quad u \geq 0 \quad \text{na } [0,1].$$

[1] Neprekidnost na segmentu $[0,1]$ znači da je funkcija uniformno neprekidna na $(0,1)$, pa je stoga ograničena i ima jedinstveno proširenje po neprekidnosti u rubnim točkama intervala.

Dakle, princip maksimuma pokazuje da naš operator ima određeno svojstvo monotonosti.
[2]

Uzmimo sada proizvoljni $f \in C([0,1])$ i neka je $u \in C_0^2([0,1])$ rješenje problema $\mathcal{A}u = f$. Funkcija $w(x) = x(1-x)/2$ je u prostoru $C_0^2([0,1])$ i zadovoljava jednadžbu $\mathcal{A}w = 1$ (provjerite). Za proizvoljnu konstantu C imamo

$$\mathcal{A}(u - Cw) = f - C.$$

Odaberemo li $C > \|f\|_\infty$, gdje je

$$\|f\|_\infty = \max_{x\in[0,1]} |f(x)|, \tag{1.25}$$

onda dobivamo

$$\mathcal{A}(u - Cw) < 0 \quad \Rightarrow \quad u - Cw \leq 0.$$

Time smo dobili ocjenu

$$u(x) \leq Cw(x) \leq \frac{1}{8}C \quad \text{za sve } x \in [0,1],$$

koja vrijedi za svaki $C > \|f\|_\infty$ pa stoga mora vrijediti i za $C = \|f\|_\infty$. Dakle

$$u(x) \leq \frac{1}{8}\|f\|_\infty \quad \text{za sve } x \in [0,1].$$

Posve se analogno dokazuje

$$-\frac{1}{8}\|f\|_\infty \leq u(x) \quad \text{za sve } x \in [0,1], \tag{1.26}$$

tako da imamo

$$|u(x)| \leq \frac{1}{8}\|f\|_\infty \quad \text{za sve } x \in [0,1], \tag{1.27}$$

odnosno $\|u\|_\infty \leq \|f\|_\infty/8$. Ocjena tog tipa naziva se apriorna ocjena. Ona nam kaže da je operator $\mathcal{A}^{-1}$ ograničen ako se promatra kao operator s prostora $C([0,1])$ u prostor $C([0,1])$, pri čemu je norma u $C([0,1])$ odabrana s (1.25). U operatorskoj normi imamo

$$\|\mathcal{A}^{-1}\| = \sup_{f\in C([0,1])} \frac{\|\mathcal{A}^{-1}f\|_\infty}{\|f\|_\infty} \leq \frac{1}{8}. \tag{1.28}$$

Napomena. Apriorna ocjena (1.27) dokazuje korektnosti zadaće (1.21), (1.22). Pri tome kažemo da je rubna zadaća korektno postavljena ako ima jedinstveno rješenje koje neprekidno ovisi o zadanim podacima zadaće. (Jedini podatak o kome zadaća (1.27) ovisi je desna strana $f(x)$.) U operatorskoj formi to se svodi na ograničenost inverznog operatora. □

[2] Sada je vidljivo zašto držimo minus ispred druge derivacije.

Možemo li analogne zaključke izvesti o inverznoj matrici A_h^{-1}?
Uvest ćemo odgovarajuću normu na vektorima:

$$\text{Za } x \in \mathbb{R}^n, \quad \|x\|_\infty = \max_{1\le i\le n} |x_i|. \tag{1.29}$$

Pripadna matrična norma je tada

$$\text{Za } A \in \mathbb{R}^{n\times n}, \quad \|A\|_\infty = \max_{x\in\mathbb{R}^n} \frac{\|Ax\|_\infty}{\|x\|_\infty}.$$

Lako se pokazuje da je za matricu $A_h = (a_{i,j}) \in \mathbb{R}^{n\times n}$

$$\|A\|_\infty = \max_{1\le i\le n} \sum_{j=1}^{n} |a_{i,j}|. \tag{1.30}$$

Za matricu A_h iz (1.24) lako izlazi

$$\|A_h\|_\infty \le \frac{4}{h^2}.$$

Nadalje bismo htjeli pokazati da je matrica A_h regularna. To je najjednostavnije vidjeti tako da dokažemo da je A_h pozitivno definitna, odnosno da postoji konstanta $\alpha > 0$, takva da je

$$\forall x \in \mathbb{R}^n, \quad A_h x \cdot x \ge \alpha |x|^2. \tag{1.31}$$

Pri dokazivanju pozitivne definitnosti najčešće se koristi ovaj kriterij: Matrica A_h je pozitivno definitna ako i samo ako vrijedi (Zadatak 18)

$$\forall x \in \mathbb{R}^n, \quad x \ne 0 \quad \Rightarrow \quad A_h x \cdot x > 0. \tag{1.32}$$

Izravnim računom dobivamo

$$\begin{aligned} A_h x \cdot x &= (2x_1 - x_2)x_1 + \sum_{i=2}^{n-1} (-x_{i-1} + 2x_i - x_{i+1})x_i + (-x_{n-1} + 2x_n)x_n \\ &= x_1^2 + (x_1 - x_2)x_1 + \sum_{i=2}^{n-1} (x_i - x_{i-1})x_i \\ &\qquad + \sum_{i=2}^{n-1} (x_i - x_{i+1})x_i + (x_n - x_{n-1})x_n + x_n^2 \\ &= x_1^2 + \sum_{i=2}^{n} (x_i - x_{i-1})x_i + \sum_{i=1}^{n-1} (x_i - x_{i+1})x_i + x_n^2 \end{aligned}$$

Zamjenom indeksa na primjer u prvoj sumi dobivamo

$$A_h x \cdot x = x_1^2 + \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2 + x_n^2.$$

Lako je vidjeti da je taj izraz strogo pozitivan za svako $x \neq 0$ pa je time je pozitivna definitnost matrice dokazana. Sljedeća definicija je uobičajena u linearnoj algebri.

Definicija. Matrica $A = (a_{i,j}) \in \mathbb{M}^{n\times n}$ je pozitivna ako su svi njeni elementi pozitivni, tj. za sve $1 \leq i, j \leq n$ vrijedi $a_{i,j} \geq 0$. Vektor $v = (v_i) \in \mathbb{R}^n$ je pozitivan ako je za sve $1 \leq i \leq n$, $v_i \geq 0$. Tada jednostavno pišemo

$$A \geq 0, \quad v \geq 0. \ \square$$

Princip maksimuma vrijedi i u diskretnom slučaju i tada se naziva diskretni princip maksimuma. Kao i u kontinuiranom slučaju zaključujemo: ako vektor $x \in \mathbb{R}^n$ zadovoljava $A_h x \leq 0$, tada x nema strogog lokalnog maksimuma, tj. ne postoji indeks $i \in \{1, \ldots, n\}$ takav da je

$$x_i > x_{i-1} \quad \text{ili} \quad x_i > x_{i+1}.$$

U tom bi slučaju bi, naime, vrijedilo

$$-x_{i-1} + 2x_i - x_{i+1} > 0,$$

što je u suprotnosti s pretpostavkom. Zbog rubnog uvjeta, $x_0 = x_{n+1} = 0$ zaključujemo da je $x_i \leq 0$ za sve $i = 1, \ldots, n$. Drugim riječima dobivamo sljedeći zaključak:

$$\forall x \in \mathbb{R}^n, \quad A_h x \geq 0 \quad \Rightarrow \quad x \geq 0. \tag{1.33}$$

Monotonost matrice se definira na sljedeći način:

Definicija. Matrica $A \in \mathbb{M}^{n\times n}$ je monotona ako je regularna i ako je A^{-1} pozitivna. $\square$

Lema 1.2 Matrica $A_h \in \mathbb{M}^{n\times n}$ je monotona ako i samo ako vrijedi (1.33).

Dokaz. Neka je matrica A_h monotona i neka je $x \in \mathbb{R}^n$ vektor sa svojstvom $Ax \geq 0$. Tada je zbog pozitivnosti elemenata matrice A^{-1}, $x = A^{-1}Ax \geq 0$.

Obrat. Neka vrijedi (1.33). Tada je prvo natrica A regularna. Zaista:

$$Ax = 0 \quad \Rightarrow \quad A(\pm x) \geq 0 \quad \Rightarrow \quad \pm x \geq 0 \quad \Rightarrow \quad x = 0.$$

Time je dokazana injektivnost, koja povlači i surjektivnost. Nadalje, ako je $x \in \mathbb{R}^n$ rješenje sustava

$$Ax = e_i,$$

gdje je e_i i-ti vektor kanonske baze, onda (1.33) daje $x = A^{-1}e_i \geq 0$; i-ti stupac inverzne matrice je pozitivan. Kako to vrijedi za $1 \leq i \leq n$, cijela je matrica A^{-1} pozitivna. □

Napomena. Uočite da smo pomoću principa maksimuma ponovo dokazali da je matrica A_h regularna. Pokazali smo, k tome, da matrica A_h^{-1} ima pozitivne elemente. □

Sljedeći korak je dokazati korektnost diskretne zadaće. Već smo dokazali da diskretna zadaća ima jedinstveno rješenje. Ostaje još samo pokazati neprekidnu zavisnost rješenja o desnoj strani. To se svodi na ocjenu norme inverzne matrice A_h^{-1}. Kod problema koji se dobivaju diskretizacijom kontinuiranih problema uvijek je potrebno zahtijevati više od same ograničenosti inverznog operatora: Konstanta koja se pojavljuje u toj ocjeni mora biti neovisna o parametru diskretizacije h. Kada je taj uvjet zadovoljen kažemo da je diskretna zadaća stabilna. Stabilnost diskretne zadaće je pojam koji odgovara korektnosti kontinuirane zadaće.

Pojmovi stabilnosti i korektnosti bitno ovise o izboru norme. Pokazat ćemo stabilnost diskretne zadaće (1.23) u ∞-normi (1.29). Uočimo prvo da zbog pozitivnosti matrice $A_h^{-1} = (b_{i,j})$, po definiciji matrične norme (1.30), imamo

$$\|A_h^{-1}\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} b_{i,j} = \|A_h^{-1}e\|_\infty, \quad e = (1, 1, \ldots, 1)^\tau.$$

Da bismo izračunali $A_h^{-1}e$ treba nam rješenje sustava

$$A_h w = e. \tag{1.34}$$

No ovdje nam pomaže kontinuirani slučaj. U izvodu apriorne ocjene služili smo se funkcijom $w \in C_0^2([0,1])$ koja je zadovoljvala jednadžbu $-w'' = 1$. Rješenje je bilo $w(x) = x(1-x)/2$. Linearan sustav $A_h w = e$ je upravo diskretizacija te jednadžbe. To nam sugerira da rješenje sustava (1.34) mora biti

$$w_i = \frac{1}{2}x_i(1-x_i) = \frac{1}{2}i(n-i)h^2, \quad 1 \leq i \leq n. \tag{1.35}$$

To se lako provjerava (Zadatak 20). Sada je

$$\|A_h^{-1}\|_\infty = \|A_h^{-1}e\|_\infty = \max_{1\leq i\leq n} |w_i| = \frac{1}{8}.$$

Time je stabilnost zadaće dokazana.

Spomenimo još da se pored monotonih matrica u teoriji konačnih diferencija i konačnih elemenata promatra još uža klasa matrica: to su tzv. M-matrice

Definicija. Regularna matrica $A \in \mathbb{R}^{n\times n}$ naziva se M-matrica ako je $A^{-1} \geq 0$ te

$$\forall i, \quad a_{ii} > 0, \quad \forall i \neq j, \quad a_{ij} \leq 0.$$

Matrica (1.24) je primjer M-matrice.

M. JURAK 16. studenog 2005.

Ocjena greške metode

Nakon što smo dokazali stabilnost numeričke metode lako je dokazati njenu konvergenciju i štoviše, dobiti ocjenu greške. Potrebno je samo uvesti u razmatranje greške diskretizacije.

Neka je $u(x) \in C^4([0,1])$ točno rješenje zadaće (1.21), (1.22). Formiramo vektor $U = (U_i)_{i=1}^n$, $U_i = u(x_i)$, za $i = 0, 1, \ldots, n, n+1$. Koristeći Taylorov razvoj lako se pokazuje (Zadatak 21) da vektor U zadovoljava sustav

$$A_h U = f - \frac{h^2}{12} b, \tag{1.36}$$

gdje je $b_i = u^{(4)}(x_i + \theta_i h)$, za neke brojeve $\theta_i \in (0,1)$. Vektor $-\frac{h^2}{12}b$ predstavlja grešku diskretizacije. Neka je $u^h \in \mathbb{R}^n$ rješenje diskretnog problema

$$A_h u^h = f.$$

Oduzimanjem dobivamo

$$A_h(u^h - U) = \frac{h^2}{12} b,$$

a stabilnost metode tada daje

$$\|u^h - U\|_\infty \leq \frac{h^2}{12} \|A_h^{-1}\|_\infty \|b\|_\infty \leq \frac{h^2}{96} \|b\|_\infty.$$

Time dobivamo ocjenu greške metode

$$\max_{1 \leq i \leq n} |u_i - u(x_i)| \leq \frac{h^2}{96} \max_{x \in [0,1]} |u^{(4)}(x)|.$$

1.4.5 Crank-Nicolsonova shema za paraboličku PDJ

Promatramo rubnu zadaću za paraboličku jednadžbu

$$\begin{aligned} u_t &= b u_{xx} + f \quad x \in (0,1), t \in (0,T), \\ u(x,0) &= u_0(x) \quad x \in (0,1) \\ u(0,t) &= u(1,t) = 0. \end{aligned}$$

Koeficijent b je konstantan i strogo pozitivan. Funkcije f i u_0 su zadane.

Rješenje ovog problema je funkcija dvije varijable $u = u(x,t)$. Stoga moramo pored prostornog koraka h uvesti i vremenski korak Δt. Metoda konačnih diferencija daje aproksimaciu

$$u_i^n \approx u(ih, n\Delta t), \quad i = 0, 1, \ldots, N+1, \ n = 0, 1, \ldots, M$$

gdje je

$$h = \frac{1}{N+1}, \quad \Delta t = \frac{1}{M}.$$

Crank-Nicolsonova shema glasi

$$\frac{u_i^{n+1}-u_i^n}{\Delta t}=\frac{b}{2}\frac{u_{i+i}^{n+1}-2u_i^{n+1}+u_{i-1}^{n+1}}{h^2}+\frac{b}{2}\frac{u_{i+i}^{n}-2u_i^{n}+u_{i-1}^{n}}{h^2}+\frac{1}{2}(f_i^{n+1}+f_i^n)$$

Shema je dobivena tako da prvo izvršena aproksimacija vremenske derivacije

$$\frac{u_i^{n+1}-u_i^n}{\Delta t}\approx bu_{xx}(x_i,t_{n+\frac{1}{2}})\approx\frac{1}{2}b(u_{xx}(x_i,t_{n+1})+u_{xx}(x_i,t_n)),$$

gdje je $x_i=ih$ i $t_k=k\Delta t$. Zatim je druga prostorna derivacija aproksimirana centralnom diferencijom. Ovakva je aproksimacija drugog reda točnosti i ima svojstvo bezuvjetne stabilnosti.

Algoritam rješavanja je sljedeći: Vrijednosti u_i^0 su poznate iz početnog uvjeta. Za izračunati vrijednosti na sljedećem vremenskom sloju potrebno je riješiti sustav s trodijagonalnom matricom. Taj se postupak ponavlja za $n=1,2,\ldots,M$.

1.5 Zadaci

Z1.1. Dokažite da za cetralnu diferenciju drugog reda vrijedi

$$u''(x)=\Delta^2u(x)-\frac{h^2}{12}u^{(4)}(\xi)\quad\text{za}\quad\xi\in J(x-h,x+h).$$

Z1.2. Diferencijom unaprijed računati derivaciju funkcije $u(x)=\arctan x$ u točki $a=\sqrt{2}$ $(u'(a)=1/3)$. Korak h u početku uzeti jednak 1, a zatim ga u svakom sljedećem koraku poloviti. Odredite na taj način eksperimentalno h_{min} i usporedite ga s teorijskim (formula (1.2)).

Z1.3. Dokažite formulu (1.3).

Z1.4. Dokažite formulu numeričke drivacije

$$f'(x)=\frac{3f(x)-4f(x-h)+f(x-2h)}{2h}+\frac{h^2}{3}f^{(3)}(\xi),$$

za neki $\xi\in(x-2h,x)$.

Z1.5. Koristeći interpolacijski polinom drugog stupnja (u točkama x_0, x_1 i x_2) nađite formule za približno računanje druge derivacije. U slučaju ekvidistantnih točaka pokažite da su formule dobivene u točkama x_0 i x_2 prvog reda točnosti dok se u x_1 dobiva centralna diferencija drugog reda koja je drugog reda točnosti.

M. JURAK 16. studenog 2005.

Z1.6. Izvedite sljedeće formule formule drugog reda za drugu derivacuju:

$$\begin{aligned} f''(x) =& \frac{2f(x) - 5f(x+h) + 4f(x+2h) - f(x+3h)}{h^2} + \frac{11}{12}h^2 f^{(4)}(\xi) \\ =& \frac{2f(x) - 5f(x-h) + 4f(x-2h) - f(x-3h)}{h^2} + \frac{11}{12}h^2 f^{(4)}(\xi). \end{aligned}$$

Z1.7. Za funkciju $f(x) = \operatorname{arctg}(x)$ u točki $a = \sqrt{2}$ ispisati vrijednosti centralne diferencije $\Delta^0 f(a)$ i njene ekstrapolacije do nekog nivoa M. Ulazni parametri su početni korak h i nivo M. Raditi u jednostrukoj i dvostrukoj preciznosti.

Z1.8. Numerički riješiti diferencijalnu jednadžbu

$$u''(x) + 9u(x) = \cos(2x), \quad x \in (0, \pi/2),$$

uz rubne uvjete $u(0) = 1$ i $u(\pi/2) = -1$. Izračunajte točno rješenje i odredite eksperimentalno grešku metode. Pokažite da je drugog reda.

Z1.9. Riješite diferencijalnu jednadžbu

$$u''(x) + 9u(x) = \cos(2x), \quad x \in (0, \pi/2),$$

uz rubne uvjete $u(0) = u(1) = 0$. Izračunajte točno rješenje i usporedite s približnim. Eksperimentalno odredite red konvergencije.

Z1.10. Modificirajte Thomasov algoritam tako da možete riješiti rubni problem

$$u''(x) - 2u'(x) + u(x) = 2e^x, \quad x \in (0, 1),$$

uz rubne uvjete $u(0) = 0$, $u'(0) = e$.

Z1.11. Dokažite stabilnost modifikacije Thomasovog algoritma iz prethodnog zadatka.

Z1.12. Diskretizirajte jednadžbu (1.6) s periodičkim rubnim uvjezom $u(0) = u(1)$. Kakvu strukturu ima matrica diskretiziranog problema?

Z1.13. Pokažite da su skupovi funkcija $C^k([0,1])$ $k = 0, 1, 2, \ldots$ i $C_0^2([0,1])$ vektorski prostori u odnosu na uobičajene operacije zbrajanja funkcija i množenja skalarom. Pokažite da ti prostori nisu konačnodimenzionalni.

Z1.14. Nađite formulu za rješenje problema $-u''(x) = f(x)$ uz homogene Dirichletove rubne uvjete $u(0) = u(1) = 0$ i na osnovu toga uvjerite se da je operator $\mathcal{A}$ surjekcija. Dokažite i njegovu injektivnost.

Z1.15. Neka su X i Y normirani vektorski prostori. Dokažite da je linearan operator $A\colon X \to Y$ neprekidan ako i samo ako je ograničen, tj. postoji konstanta C, takva da je

$$\forall x \in X, \quad \|Ax\|_Y \leq C\|x\|_X.$$

Z1.16. Dokažite ocjenu (1.26).

Z1.17. Dokažite formulu (1.30).

Z1.18. Dokažite kriterij pozitivne definitnosti (1.32). Uputa: koristiti kompaktnost jedinične sfere i činjenicu da neprekidna funkcija dostiže svoj minimum (i maksimum) na kompaktnom skupu.

Z1.19. Dokažite da je svaka pozitivno definitna matrica regularna.

Z1.20. Dokažite da je (1.35) rješenje sustava (1.34).

Z1.21. Razvojem u Taylorov red do člana četvrtog reda dokažite formulu (1.36). Uočite da treba iskoristiti Zadatak 1.

Z1.22. Izračunajete svojstvene vektore i svojstvene vrijednosti matrice (1.24) polazeći od rješenja kontinuiranog problema svojstvenih vrijednosti: naći $u \in C_0^2([0,1])$ i $\lambda \in \mathbb{R}$ za koje vrijedi

$$-u''(x) = \lambda u(x), \quad x \in (0,1).$$

Z1.23. Crank-Nicolsonovom shemom riješiti rubnu zadaću

$$\begin{aligned} u_t &= 2u_{xx} \quad x \in (0,1), t \in (0,5), \\ u(x,0) &= \sin(8\pi x) \quad x \in (0,1) \\ u(0,t) &= u(1,t) = 0. \end{aligned}$$

Nađite točno rješenje separacijom varijabli. Izračunajete aproksimativno rješenje i grešku u normi

$$\|u\|_{2,\infty} = \max_{0 \leq n \leq M} \left(\sum_{i=0}^{n+1} |u_i^n|^2 \right)^{1/2}.$$

Pokažite eksperimentalno na ovom primjeru da metoda konvergira i odredite red konvergencije.