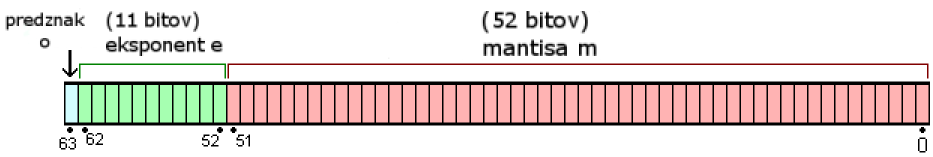
Poglavje 1

Plavajoča vejica

Naloga 1.1 (Vaje) *Pretvori število* 7.721 *v sistem* $P(2, 6, -10, 10)$.



Slika 1.1: Plavajoča vejica

Če hočemo število x zapisati v sistemu $P(b, , t, L, U)$, iščemo zapis normalizirane oblike

$$\pm 1.c_1 \dots c_t \cdot b^e,$$

kjer je $L \leq e \leq U$. V primeru, ko število ni predstavljivo, uporabimo zaokrožanje. Naj bo $x = 1.c_1c_2 \dots c_t c_{t+1} \dots \cdot b^e$ v neskončnem decimalnem zapisu. Če je $c_{t+1} < b/2$ odrežemo števke od $(t+1)$-ve naprej in je $\mathrm{fl}(x) = \pm 1.c_1c_2 \dots c_t \cdot b^e$, če pa je $c_{t+1} \geq b/2$ zaokrožimo navzgor, $\mathrm{fl}(x) = \pm \left(1.c_1c_2 \dots c_t + b^{-t}\right) \cdot b^e$.

Rešitev. Število najprej pretvorimo v dvojiški zapis. Najprej se lotimo celega dela.

$$\begin{aligned} 7 &= 3 * 2 + \mathbf{1} \\ 3 &= 1 * 2 + \mathbf{1} \\ 1 &= 0 * 2 + \mathbf{1} \uparrow \end{aligned}$$

Preberemo navzgor in dobimo $7 = 111_{(2)}$. V naslednjem koraku pretvorimo še decimalni del v dvojiški zapis.

$$\begin{aligned} 0.721 * 2 &= 0.442 + \mathbf{1} \downarrow \\ 0.442 * 2 &= 0.884 + \mathbf{0} \\ 0.884 * 2 &= 0.768 + \mathbf{1} \\ 0.768 * 2 &= 0.536 + \mathbf{1} \\ 0.536 * 2 &= 0.072 + \mathbf{1} \end{aligned}$$

Torej je

$$\mathrm{fl}\,(7.721) = \mathrm{fl}\left(111.10111_{(2)}\right) \overbrace{=}^{\text{zaokrožimo}} 111.1100.$$

Ker iščemo zapis normalizirane oblike, moramo število deliti/množiti z dva, dokler ne dobimo oblike $1.c_1c_2c_3c_4c_5c_6$. Torej $111.1100_{(2)} = 1.111100(2) * 2^3$. Mantisa je 0.111110, eksponent pa $3_{(2)} = 11$.

■

Naj bo x število in $fl(x)$ najbližje predstavljivo število. Velja

$$\mathrm{fl}(x) = x(1+\delta) \text{ in } |\delta| \leq u,$$

kjer je u osnovna zaokrožitvena napaka. Za predstavljivo število velja $\mathrm{fl}(x) = x$.

Poglavje 2

Numerična stabilnost

Algoritem je **direktno stabilen**, če vrne rezultat, ki se malo razlikuje od prave vrednosti. Ponavadi preverjamo relativno direktno stabilnost. Direktna absolutna in relativna napaka sta

$$|y - \widehat{y}| \text{ in } \frac{|y - \widehat{y}|}{|y|}.$$

Algoritem je **obratno stabilen**, če za izračunan rezultat obstajajo taki malo zmoteni podatki, da iz njih s točnim izračunom dobimo izračunano vrednost. Obratna absolutna napaka je najmanjši $|\Delta x|$, tako da velja $f(x + \Delta x) = \widehat{y}$. Obratna relativna napaka je $\frac{|\Delta x|}{|x|}$.

Pri obratni napaki je izračunana vrednost enaka $\widehat{y} = f(\widehat{x})$. Če je funkcija f *zvezno odvedljiva* v x, potem velja

$$|\widehat{y} - f(x)| = |f(\widehat{x}) - f(x)| \le |f'(x)||\widehat{x} - x|.$$

Če f ni absolutno občutljiva, bo pri majhnih vrednostih $\Delta x = |\widehat{x} - x|$ napaka absolutno obratno stabilne metode majhna in metoda bo absolutno direktno stabilna. Podobno velja za relativno stabilnost in relativne napake. Več ste povedali na predavanjih.

Naloga 2.1 *Vrednost $z = x^2 - y^2$ računamo na dva načina:*

(i). $z = x^2 - y^2$

(ii). $z = (x - y)(x + y)$

Analiziraj algoritma. Oceni relativno napako $\frac{|\widehat{z} - z|}{|z|}$. Ali je kateri od obeh algoritmov direktno/obratno stabilen?

Rešitev.

(i). Imamo $z = x * x - y * y$, $\widehat{a} = x * x(1 + \alpha)$, $\widehat{b} = y * y(1 + \beta)$, $\widehat{z} = (\widehat{a} - \widehat{b})(1 + \gamma)$, kjer je $|\alpha|, |\beta|, |\gamma| \le u$, u je relativna natančnost. Sledi, da je

$$\widehat{z} = x^2 \overbrace{(1 + \alpha)(1 + \gamma)}^{(1+\delta_1)} - y^2 \overbrace{(1 + \beta)(1 + \gamma)}^{(1+\delta_2)}.$$

Iz ocene $1 - 2u + u^2 = (1 - u)^2 \le (1 + \delta_1) \le (1 + u)^2 = 1 + 2u + u^2$, pri majhnem u dobimo $\delta_1, \delta_2 \le 2 * u$. Ocenimo izraz

$$|\widehat{z} - z| = |x^2\delta_1 - y^2\delta_2| \le x^2|\delta_1| + y^2|\delta_2| \le 2u(x^2 + y^2).$$

Torej velja ocena:

$$\frac{|\widehat{z}-z|}{|z|} \leq 2u\frac{x^2+y^2}{x^2-y^2}.$$

Ocena je smiselna. Najlažje to vidimo tako, da izberemo $\delta_1 = u$, $\delta_2 = -u$. Iz tega vidimo, da ta algoritem ni direktno stabilen, saj lahko pri x in y, ki sta si blizu, dobimo veliko napako. Algoritem je obratno stabilen. Če definiramo $\widehat{x} = x\sqrt{(1+\delta_1)}$ in $\widehat{y} = y\sqrt{(1+\delta_2)}$, ki sta blizu x in y, ter predpostavimo, da je računanje točno, dobimo zmoten izraz $\widehat{z} = \widehat{x}^2 - \widehat{y}^2$.

(ii). Oglejmo si še $z = (x-y)(x+y)$. Definirajmo $\widehat{a} = (x-y)(1+\alpha)$, $\widehat{b} = (x+y)(1+\beta)$ ter $\widehat{z} = \widehat{a}\widehat{b}(1+\gamma)$, kjer je $|\alpha|, |\beta|, |\gamma| \leq u$. Izraz je enak

$$\widehat{z} = (x-y)(x+y)\overbrace{(1+\alpha)(1+\beta)(1+\gamma)}^{(1+\delta)},$$

kjer s podobno oceno kot v prejšnji točki dobimo $|\delta| \leq 3u$. Torej velja ocena $|\widehat{z}-z| \leq 3u|z|$. Na koncu dobimo $\frac{|\widehat{z}-z|}{|z|} \leq 3u$. Algoritem je direktno stabilen. Prav tako je obratno stabilen, iskana zmotena x in y sta recimo: $\widehat{x} = x\sqrt{(1+\delta)}$, $\widehat{y} = y\sqrt{(1+\delta)}$.

■

Naloga 2.2 *Dan je polinom*

$$p(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n,$$

radi bi izračunali njegovo vrednost v točki x po Hornerjevem algoritmu. Predpostavka je, da so koeficienti $a_0, \ldots, a_n$ in argument x predstavljiva števila. Analiziraj stabilnost izračuna.

Rešitev.

Algoritem 1: Eksaktni algoritem	**Algoritem 2:** Dejanski algoritem
$p_0 = a_0$;	$\widehat{p}_0 = a_0$;
for $i = 1, \ldots, n$ **do**	**for** $i = 1, \ldots, n$ **do**
$p_i = p_{i-1} * x + a_i$;	$\widehat{p}_i = \left(\widehat{p}_{i-1}(1+\alpha_i) + a_i\right)(1+\beta_i)$;
$p = p_n$;	$p = p_n$;
end	**end**

Z analizo zaokrožitvenih napak dobimo

$$\widehat{p} = a_0x^n(1+\gamma_0) + a_1x^{n-1}(1+\gamma_1) + \cdots + a_n(1+\gamma_n),$$

kjer je

$$\begin{aligned} 1+\gamma_0 &= (1+\alpha_1)\cdots(1+\alpha_n)(1+\beta_1)\cdots(1+\beta_n), \\ 1+\gamma_i &= (1+\alpha_{i+1})\cdots(1+\alpha_n)(1+\beta_i)\cdots(1+\beta_n), \quad i = 1, \cdots, n-1, \\ 1+\gamma_n &= 1+\alpha_n. \end{aligned}$$

Napako γ_0 obravnavamo posebej, saj pri izračunu $p_0 = a_0$, ne zagrešimo napake. Naredimo samo dva koraka. Dokaz koraka indukcije je preprost in prepuščen bralcu.

$$\widehat{p}_1 = \left(\widehat{p}_0(1+\alpha_1) + a_1\right)(1+\beta_1) = a_0x(1+\alpha_1)(1+\beta_1) + a_1(1+\beta_1),$$

$$\widehat{p}_2 = \left(\widehat{p}_1(1+\alpha_2) + a_2\right)(1+\beta_1) = a_0x^2(1+\alpha_1)(1+\alpha_2)(1+\beta_1)(1+\beta_2) + a_1x(1+\alpha_2)(1+\beta_1)(1+\beta_2) + a_2(1+\beta_2).$$

Narediti je treba še korak indukcje po stopnji polinoma.

Ocenimo lahko $|\gamma_0| \leq 2nu$ in $|\gamma_i| \leq (2(n-i)+1)\, u$ za $i = 1, \ldots, n$. Tukaj smo upoštevali, da se relativne napaka seštejejo, člene višjega reda zanemarimo.

Računanje vrednosti polinoma je obratno stabilno, saj se izračunane vrednosti ujemajo z eksaktno vrednostjo bližnjega polinoma, ki ima koeficiente $a_i(1+\gamma_i)$ namesto a_i, pri nespremenjenem argumentu x.

Iz absolutne napake $\widehat{p} - p = a_0 x^n \gamma_0 + a_1 x^{n-1}\gamma_1 + \cdots a_n \gamma_n$, sledi ocena

$$|\widehat{p} - p| \leq 2nu \left(|a_0||x^n| + |a_1||x^{n-1}| + \cdots + |a_n| \right),$$

od tod pa

$$\frac{|\widehat{p} - p|}{|p|} \leq \frac{2nu \left(|a_0||x^n| + |a_1||x^{n-1} + \cdots + |a_n| \right)}{|a_0 x^n + \cdots + a_n|}.$$

Računanje vrednosti polinoma po Hornerjevem algoritmu ni direktno stabilno. Težave pa, podobno kot pri računanju skalarnega produkta, lahko pričakujemo, če je vsota blizu 0, členi v vsoti pa niso enako predznačeni.

Modelni primer je Wilkinsov zgled za polinom $(x-2)^{19}$. Oglejte si demonstracijski zgled `horner_wilkinson.m` v Matlabu. ∎

Naloga 2.3 *Podani sta dve približno enaki števili* $x = 76.54320$ *in* $y = 76.54311110$ *v sistemu* $P(10, 7, ...)$. *Izračunaj relativne napake* $\frac{x - \mathrm{fl}(x)}{x}$, $\frac{y - \mathrm{fl}(y)}{y}$ *in* $\frac{z - \mathrm{fl}(z)}{z}$, *kjer je* $z = x - y$.

Rešitev. Števila so že podana v desetiškem sistemu, zato moramo poskrbeti le, da bo mantisa dolžine 7 in število zaokroženo.

$$\mathrm{fl}(76.54320) = 0.7654320 * 10^2, \qquad \mathrm{fl}(76.54311110) = 0.7654311 * 10^2.$$

$$\frac{x - \mathrm{fl}(x)}{x} = 0, \frac{y - \mathrm{fl}(y)}{y} = 0.000000014.$$

Izračunajmo še $z = x - y = 0.0000889$ in $\mathrm{fl}(z) = \mathrm{fl}(\mathrm{fl}(x) - \mathrm{fl}(y)) = \mathrm{fl}(0.7654320 * 10^2 - 0.7654311 * 10^2) = 0.0000009 * 10^2$. Torej velja $\frac{z - \mathrm{fl}(z)}{z} = \frac{0.0000889 - 0.00009}{0.0000889} = -0.012373453$. Relativna napaka je velika.

∎

Naloga 2.4 *Pretvori naslednje izraze v stabilno obliko:*

(i). $\sqrt{1+x} - 1$ *za majhne* x,

(ii). $\sqrt{x^2 + x} - x$ *za velike* x,

(iii). $\tan(x) - \sin(x)$ *za majhne* x.

Rešitev.

(i). Znebiti se moremo odštevanja dveh približno enakih števil,

$$\sqrt{1+x}-1=\frac{(\sqrt{1+x}-1)(\sqrt{1+x}+1)}{\sqrt{1+x}+1}=\frac{x}{1+\sqrt{1+x}}.$$

Zadnji izraz je v stabilni obliki.

(ii). Probleme nam lahko povzroča overflow, ko hočemo izračunati x^2.

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

(iii). Odštevamo dve približno enaki števili, saj velja $\sin(x)=\tan(x)\approx x$.

$$\tan(x)-\sin(x)=\frac{\sin(x)}{\cos(x)}-\sin(x)=\frac{\sin(x)(1-\cos(x))}{\cos(x)}=\frac{2\sin^2(x/2)\sin(x)}{\cos(x)}.$$

Upoštevali smo $\cos(2(x/2))=\cos^2(x/2)-\sin^2(x/2)$, $1=\cos^2(x/2)+\sin^2(x/2)$ in $\sin(x)=2\sin(x/2)\cos(x/2)$.

■

Primer 2.1 *Stabilno računanje ničel kvadratnega polinoma ax^2+bx+c. Rešitev kvadratne enačbe ax^2+bx+c je enaka*

$$x_{1,2}=\frac{-b\pm\sqrt{b^2-4ac}}{2a}=\frac{2c}{-b\mp\sqrt{b^2-4ac}}.$$

Drugi izraz uporabimo, ko gre $a\to 0$, saj potem velja $x_1\to-\frac{c}{b}$ in $x_2\to\infty$. Eno ničlo lahko zmeraj izračunamo stabilno, drugo pa računamo preko Vietovih formul $x_1x_2=\frac{c}{a}$, $x_1+x_2=-\frac{b}{q}$. Naslednji izračun je stabilen:

$$q=-\frac{1}{2}\left(b+\operatorname{sign}(b)\sqrt{b^2-4ac}\right),$$

$$x_1=\frac{q}{a}\qquad \textit{in}\qquad x_2=\frac{c}{ax_1}.$$

Poglavje 3

Nelinearne enačbe

Naloga 3.1 *Za funkcijo* $f(x) = x^3 + 1$ *naredi korak bisekcije in tangentne metode. Vzemi* $a = -1.2$ *in* $b = x_0 = -0.9$.

Rešitev. Korak bisekcije je $fb = f(-0.9) = (-0.9)^3+1 = 0.271$, $fa = f(-1.2) = (-1.2)^3+1 = -0.728$, $\text{sign}(fafb) = -1$. Zato lahko izvedemo korak bisekcije. Nadaljujmo: $c = (a+b)/2 = -1.05$, $fc = (-1.05)^3 + 1 = -0.157625$, $\text{sign}(fafc) < 0$, torej vzamemo $a = c$. To ponavljamo.

Za korak tangentne metode potrebujemo odvod $f'(x) = 3x^2$. Izračunajmo

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = -0.9 - \frac{(-0.9)^3 + 1}{3(-0.9)^2} \doteq 1,0115.$$

Bisekcijo in sekantno metoda uporabljam takrat, ko ne poznamo odvoda ali pa je računanje odvoda zahtevno. ∎

Naloga 3.2 *Podani sta dve naložbi z naslednjimi podatki*

	Začetni vložek	*Končna vrednost naložbe*	*Čas dospetja*
Naložba 1	*1000€*	*3000€*	*5 let*
Naložba 2	*1000€*	*4000€*	*7 let*

Kolikšna sta donosnosti do dospetja posameznih naložb? Kolikšen je donosnost do dospetja, če vložimo v vsako naložbo po 1000€*?*

Rešitev. Donosnost do dospetja prve naložbe je $\sqrt[5]{\frac{3000}{1000}} = \sqrt[5]{3}$, druge naložbe pa $\sqrt[7]{\frac{4000}{1000}} = \sqrt[7]{4}$. Neto sedanja vrednost je vrednost, ki bi jo morala imeti naložba danes, da bi dobili končno vrednost naložbe pri obrestovalnem faktorju r. Donosnost do dospetja sestavljene naložbe izračunamo tako, da izenačimo neto sedanjo vrednost naložb in začetni vložek v naložbe.

$$3000/(1+r)^5 + 4000/(1+r)^7 = 1000 + 1000 = 2000.$$

Dobili smo nelinearno enačbo, ki jo rešimo s kakšno numerično metodo. Dobimo rešitev $r \doteq 0.235$. ∎

Naloga 3.3 *Na voljo imamo kuponsko obveznico s ceno* 950€ *in kuponsko obrestno mero* 7 %*. Čas dospetja je* 4*. leta, njena nominalna vrednost* N *je* 1000€*. Kolikšna je njena donosnost do dospetja? Ali je večja od trenutnega donosa?*

Rešitev. Kuponska obrestna mera je 7 %. Vsako leto tako dobimo izplačanih 70 €. Kuponsko obveznico si lahko predstavljamo kot sestavljeno naložbo

	Skupni začetni vložek	Končna vrednost naložbe	Čas dospetja
Naložba	950€		
Naložba 1		70€	1 leto
Naložba 2		70€	2 leti
Naložba 1		70€	3 leta
Naložba 2		1000€ + 70€	4 leta

Tako dobimo nelinearno enačbo

$$\frac{70}{(1+r)} + \frac{70}{(1+r)^2} + \frac{70}{(1+r)^3} + \frac{1070}{(1+r)^4} = 950.$$

Približna rešitev je $r = 0.085$. Trenutna donosnost naložbe je $70/950 = 0.0737 = 7.37\%$. ■

Izrek 3.1 *Naj bo α koren enačbe $x = g(x)$, naj bo g zvezno odvedljiva na $I = [\alpha - d, \alpha + d]$ in naj velja $|g'(x)| \leq m < 1$ za vsak $x \in I$. Potem za vsak $x_0 \in I$ zaporedje*

$$x_{r+1} = g(x_r),\ r = 0, 1, \cdots$$

konvergira k α in velja ocena za napako

$$|x_r - \alpha| \leq \frac{m}{1-m}|x_r - x_{r-1}|.$$

Dokaz 3.1 *Za $z_0 \in [\alpha - d, \alpha + d]$ velja*

$$|g(z_0) - \alpha| = |g(z_0) - g(\alpha)| = |(z_0 - \alpha)g'(\xi)|,$$

kjer je ξ med z_0 in α. Torej velja $\xi \in I$, posledično je $|g'(\xi)| \leq m < 1$. Dobili smo da velja $|z_1 - \alpha| \leq m|z_0 - \alpha|$. Naša indukcijska predpostavka je $|z_n - \alpha| \leq |z_0 - \alpha|m^n$. Korak indukcije je trivialen,

$$|z_{n+1} - \alpha| = |g(z_n) - \alpha| = |z_n - \alpha|g'(\xi_1) \leq m|z_n - \alpha| \leq^{IP} |z_0 - \alpha|m^{n+1}$$

Torej velja

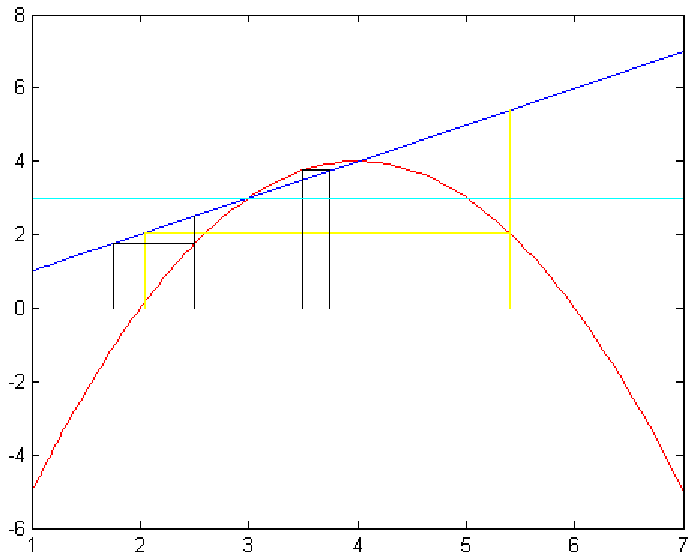
$$\lim_{n\to\infty} |z_{n+1} - \alpha| \leq \lim_{n\to\infty} |z_0 - \alpha|m^{n+1} = 0,$$

saj je m strogo manjši kot ena.

Izrek 3.2 *Imamo začetno točko x_0 in interval $I = [x_0 - d, x_0 + d]$. Če velja $|g(x) - g(y)| \leq m|x - y|$ za $m < 1$ in $|g(x_0) - x_0| \leq (1 - m)d$, potem zaporedje x_r konvergira k α in velja $g(\alpha) = \alpha$.*

Izrek 3.3 *(Banachovo skrčitveno načelo) Naj bo M poln metričen prostor in $g : M \to M$. Tedaj obstaja natanko ena negibna (fiksna) točka preslikava g, t.j. taka točka $a \in M$, da je $g(a) = a$. Če je $x_0 \in M$ poljubna točka, tedaj zaporedje $x_{r+1} = g(x_r)$ konvergira k a. Primer polnega metričnega prostora v $\mathbb{R}$ je zaprti interval.*

Naloga 3.4 *Za katere začetne približke je navadna iteracija za reševanje $x = g(x)$, kjer je $g(x) = -12 + 8x - x^2$, konvergentna? Kam konvergira zaporedje $x_{r+1} = g(x_r)$? Kakšen je red konvergence? Kje nam konvergenco zagotavlja izrek (Banachovo skrčitveno načelo) ? Namig: začetne približke išči na intervalu.*



Slika 3.1: Koraki iteracije

Rešitev. Iz slike preberemo, da je kandidat za območje konvergence interval $[3,5]$. V naslednjih točkah bomo pokazali, da je ta interval res dobra izbira.

(i). Pokažimo, da velja: $x_r < 3 \Rightarrow x_{r+1} < x_r$. Za take začetne približke bomo imeli divergenco. Oglejmo si $x_r - x_{r+1}$ in pokažimo, da je izraz pozitiven. Izračunajmo

$$x_r - x_{r+1} = x_r - g(x_r) = x_r - (-12 + 8x_r - x_r^2) = 12 - 7x_r + x_r^2 = (x_r - 4)(x_r - 3).$$

Zadnji izraz v enakosti je v primeru $x_r < 3$ strogo pozitiven.

(ii). Velja tudi: $x_0 > 5 \Rightarrow x_1 < 3$. Spet dobimo divergenco. Oglejmo si izraz $3 - x_1$, to nam da

$$3 - x_1 = 15 - 8x_0 + x_0^2 = (x_0 - 5)(x_0 - 3) > 0.$$

(iii). Poleg tega velja $|x_{r+1} - 4| \leq |x_r - 4|^2$ za $x_r \in (3,5)$. Izračunajmo

$$|x_{r+1} - 4| = |-12 + 8x_r - x_r^2 - 4| = |-16 + 8x_r - x_r^2| = (x_r - 4)^2, \tag{3.1}$$

ker je x_r na $(3,5)$, sledi $|x_r - 4| < 1$ in potem $|x_r - 4|^2 < |x_r - 4|$.Vidimo, da velja tudi $|x_r - 4| = |x_0 - 4|^{2^r}$, kar že zagotavlja konvergenco. Zvezo dobimo z zaporedno uporabo enačbe (3.1).

(iv). Za robne točke velja $g(3) = g(5) = 3$.

Naši možni rešitvi sta $x_1 = 4$ in $x_2 = 3$, x_2 dobimo samo v primeru robnih točk, kar je praktično nemogoče. Zato si oglejmo red konvergence v okolici ničle $x_1 = 4$. Spomnimo se, da velja: metoda je reda p, natanko takrat ko $g'(x_1) = \cdots = g^{(p-1)}(x_1) = 0$ in $g^{(p)}(x_1) \neq 0$. Odvod je $g'(x) = 8 - 2x$, torej je $g'(4) = 0$. Red je vsaj kvadratičen. Drugi odvod je enak $g''(x) = -2$. Torej je red kvadratičen.

Izrek nam konvergenco zagotavlja na intervalu, kjer velja $|g'(x)| < 1$. Torej mora veljati |8-2x|< 1, to pa je $3.5 < x < 4.5$. Izrek nam da torej le potreben pogoj, iterativna metoda pa lahko konvergira še na večjem intervalu.

∎

Naloga 3.5 *Iščemo rešitve enačbe* $f(x) = x^5 - 10x + 1 = 0$. *Za iteracijsko funkcijo izberemo* $g(x) = \frac{x^5+1}{10}$. *Ali nam izrek izrek* 3.2 *zagotavlja konvergenco za* $x_0 = 0$? *Oceni napako drugega približka.*

Rešitev. Za odvod mora veljati $|g'(x)| \leq m < 1$. Kar je $|\frac{x^4}{2}| < 1$, $|x| < \sqrt[4]{2} \doteq 1.1892$. Torej imamo za x_0 konvergenco. Napako drugega približka bomo ocenili s pomočjo formule $|x_r - \alpha| \leq \frac{m}{1-m}|x_r - x_{r-1}|$. Oceniti moramo še m in si izbrati primeren interval. Če hočemo uporabiti izrek 3.2, izberemo recimo interval $[-0.2, 0.2]$, kjer je $d = 0.2$. Na tem intervalu ocenimo $m \leq \frac{0.2^4}{2} \leq 0.0008$. Poleg tega velja tudi $|g(0) - 0| = 0.1 \leq (1-0.0008)0.2$. Vsi pogoji so izpolnjeni. ■

Tangentno metodo dobimo, če za naslednjo točko v iteraciji vzamemo presečišče tangente na funkcijo f z osjo x. Tako dobimo $x_{r+1} = g(x_r) = x_r - \frac{f(x_r)}{f'(x_r)}$.

Izrek 3.4 *Za tangentno metodo velja naslednji izrek. Naj velja* $f(\alpha) = 0$ *in naj bo* α m*-kratna ničla. Če je* $m = 1$, *je konvergenca vsaj kvadratična. Če velja še* $f''(\alpha) = 0$, *je konvergenca kubična. Če imamo večkratno ničlo* $m \geq 2$, *velja* $\lim_{x\to\alpha} g'(x) = 1 - 1/m$.

Rešitev. Dokazali bomo samo drugi del izrek, saj je prvi enostaven. Če je α m-kratna ničla funkcije f, potem po definiciji obstaja $l \neq 0$, da velja $\lim_{x\to\alpha} \frac{f(x)}{(x-\alpha)^m} = l$. Po L'Hopitalu dobimo $\lim_{x\to\alpha} \frac{f'(x)}{m(x-\alpha)^{m-1}} = l$, saj gresta imenovalec in števec proti 0. Najprej pokažimo, da velja $\lim_{x\to\alpha} x - \frac{f(x)}{f'(x)} = \alpha$. Zadosti bo, da dokažemo $\lim_{x\to\alpha} \frac{f(x)}{f'(x)} = 0$. Imenovalec in števec delimo z $(x-\alpha)^m$, dobimo

$$\lim_{x\to\alpha}(x-\alpha)\frac{\frac{f(x)}{(x-\alpha)^m}}{m\frac{f'(x)}{m(x-\alpha)^{m-1}}} = \lim_{x\to\alpha}(x-\alpha)\frac{l}{ml} = 0.$$

Izračunajmo še limito odvoda:

$$g'(\alpha) = \lim_{x\to\alpha}\frac{x - \frac{f(x)}{f'(x)} - g(\alpha)}{x-\alpha} = \lim_{x\to\alpha}\frac{x - \frac{f(x)}{f'(x)} - \alpha}{x-\alpha} = 1 - \lim_{x\to\alpha}\frac{f(x)}{f'(x)(x-\alpha)}.$$

Nadaljujmo

$$1 - \lim_{x\to\alpha}\frac{f(x)}{f'(x)(x-\alpha)} = 1 - \lim_{x\to\alpha}\frac{\frac{f(x)}{(x-\alpha)^m}}{m\frac{f'(x)}{m(x-\alpha)^{m-1}}} = 1 - \frac{l}{ml} = 1 - \frac{1}{m}.$$

■

Laguerrova metoda je metoda za iskanje ničel polinoma

$$f(z) = a_0(z-\alpha_1)(z-\alpha_2)\dots(z-\alpha_n).$$

Iteracije je oblike

$$z_{r+1} = z_r - \frac{nf(z_r)}{f'(z_r) \pm \sqrt{(n-1)\left[(n-1)f'^2(z_r) - nf(z_r)f''(z_r)\right]}}.$$

Velja $\epsilon_r = x_r - \alpha$, $\epsilon_{r+1} = C\epsilon_r^3 + O(\epsilon_r^4)$, kjer je $C = \frac{(n-2)f''^2(\alpha)}{8(n-1)f'^2(\alpha)} - \frac{f'''(\alpha)}{6f'(\alpha)}$. Predznak izberemo tako, da je izraz v imenovalcu po absolutni vrednosti največji. Konvergenca je pri enostavnih ničlah vsaj kubična.

Naloga 3.6 *Izpelji metodo kubičnega korena, tako da v Laguerrovi metodi pošlješ* $n \to \infty$. *Z izpeljano metodo izračunaj* $\sqrt[3]{a}$. *Naredi tri korake za* $a = 3$ *in* $x_0 = 1$.

Rešitev. Izračunajmo

$$\begin{aligned} z_{r+1} &= z_r - \lim_{n\to\infty} \frac{f(z_r)}{\frac{f'(z_r)}{n} \pm \sqrt{(1-\frac{1}{n})\left[(1-\frac{1}{n})f'^2(z_r) - f(z_r)f''(z_r)\right]}} \\ &= z_r - \frac{f(z_r)}{\sqrt{f'^2(z_r) - f(z_r)f''(z_r)}}. \end{aligned}$$

Velja $\epsilon_{r+1} = C\epsilon_r^3 + O(\epsilon_r^4)$, kjer je $C = \frac{f''^2(\alpha)}{8f'^2(\alpha)} - \frac{f'''(\alpha)}{6f'(\alpha)}$.
Funkcijo f si izberemo tako, da velja $f(\sqrt[3]{a}) = 0$. Izberemo si $f(x) = x^3 - a$, torej je $f'(x) = 3x^2$ in $f''(x) = 6x$. Dobimo metodo

$$\begin{aligned} x_{r+1} &= x_r - \frac{x_r^3 - a}{\sqrt{9x_r^4 - (x_r^3 - a)6x_r}} \\ &= x_r - \frac{x_r^3 - a}{\sqrt{3x_r^4 + 6ax_r}}. \end{aligned}$$

Za $a = 3$, $x_0 = 1$ dobimo: $x_1 = 1 - \frac{1-3}{\sqrt{3+6\cdot 3}} \doteq 1.4364358$, $x_2 \doteq 1.4422496$, $x_3 \doteq 1.4422496$. ■

Naloga 3.7 *V iteracijski formuli* $x_{r+1} = \frac{A^2}{x_r^5}\left(\alpha + \frac{\beta x_r^3}{A} + \frac{\gamma x_r^6}{A^2}\right)$ *za* $\sqrt[3]{A}$ *določi* α, β, γ, *da bo metoda vsaj kubična.*

Rešitev. Veljati mora $g(\sqrt[3]{A}) = \sqrt[3]{A}$, $g'(\sqrt[3]{A}) = 0$, $g''(\sqrt[3]{A}) = 0$. Imamo

$$\begin{aligned} g(x) &= \alpha A^2 x^{-5} + \beta A x^{-2} + \gamma x, \\ g'(x) &= -5\alpha A^2 x^{-6} - 2\beta A x^{-3} + \gamma, \\ g''(x) &= 30\alpha A^2 x^{-7} + 6\beta A x^{-4}. \end{aligned}$$

Dobimo sistem:

$$\begin{aligned} \alpha + \beta + \gamma &= 1, \\ -5\alpha - 2\beta + \gamma &= 0, \\ 30\alpha + 6\beta &= 0. \end{aligned}$$

Iz tretje enačbe dobimo $\beta = -5\alpha$. Vstavimo v drugo in dobimo $\gamma = -5\alpha$. Nazadnje iz prve enačbe dobimo $-9\alpha = 1$. Torej je $\alpha = -\frac{1}{9}$, $\beta = \gamma = \frac{5}{9}$. Za $A = 5$ in $x_0 = 2$ dobimo $g(x_r) = \frac{25}{9x_r^5}\left(-1 + x_r^3 + \frac{x_r^6}{5}\right)$. Navedimo nekaj začetnih približkov: $x_1 \doteq 1.71875$, $x_2 \doteq 1.709976326$, $x_3 \doteq 1.709975947$. ■

Naloga 3.8 *Določi vse polinome četrte stopnje z vodilnim koeficientom* 1, *pri katerih se tangentna metoda veda takole:*

- *v bližini* α *ima linearno konvergenco,*
- *v bližini* $-\alpha$ *ima kubično konvergenco.*

Določi še preostale ničle polinoma in red konvergence v njihovi bližini.

Rešitev. Spomnimo se, da je v bližini enostavnega korena konvergenca vsaj kvadratična. Če je prvi odvod enak 0, je konvergenca kubična. V primeru večkratne ničle je konvergenca linearna. Iz tega ugotovimo, da je α vsaj dvakratna ničla in $-\alpha$ enostavna ničla. Polinom je oblike $p(x) = (x+\alpha)(x-\alpha)^2(x-d)$, kjer je d neznana ničla. Uporabimo formulo

$$(f(x)g(x))^{(n)} = \sum_{i=0}^{n} \binom{n}{i} f^{(i)}(x) g^{(n-i)}(x).$$

Dobimo

$$p''(x) = 2(x-\alpha)^2 + 4(x-\alpha)(x-d) + 4(x+\alpha)(x-d) + (x-\alpha)(x+\alpha).$$

Če hočemo, da bo konvergenca kubična, mora veljati $p''(-\alpha) = 0$. Vstavimo $-\alpha$:

$$p''(-\alpha) = 2(-\alpha-\alpha)^2 + 4(-\alpha-\alpha)(-\alpha-d) + 0 + 0 = 8\alpha(\alpha+\alpha+d) = 0,$$

iz česar dobimo $d = -2\alpha$. Ničla -2α je enostavna, konvergenca je vsaj kvadratična. Izračunajmo še vrednost drugega odvoda v -2α:

$$p''(-2\alpha) = 2(-2\alpha-\alpha)^2 + 0 + 0 + (-2\alpha+\alpha)(-2\alpha-\alpha) = 30\alpha^2 \neq 0.$$

Konvergenca v -2α je kvadratična. ∎

Newtonova metoda se uporablja za iskanje rešitev sistema $F(z) = 0$. Metoda je posplošitev tangentne metode. Označimo $\Delta z_r = z_{r+1} - z_r$. Recimo, da je z_{r+1} ničla. Velja

$$0 \doteq F(z_r) + J(z_r)(z_{r+1} - z_r),$$

kjer je J Jacobijeva matrika parcialnih odvodov reda 1. Dobimo

$$J(z_r)\Delta z_r = -F(z_r).$$

Pri vsakem koraku rešimo sistem in dobimo Δz_r,

Naloga 3.9 *Nastavi Newtonovo metodo za reševanje sistema:*

$$\begin{aligned} x^2 + y^2 &= 4, \\ x^2 - y^2 &= 1. \end{aligned}$$

Naredi en korak z začetnim približkom $x_0 = 1.5$, $y_0 = 1$.

Rešitev. Velja $F(x,y) = \begin{bmatrix} f_1(x,y) \\ f_2(x,y) \end{bmatrix} = \begin{bmatrix} x^2+y^2-4 \\ x^2-y^2-1 \end{bmatrix}$. Izračunajmo

$$JF = \begin{bmatrix} f_{1x} & f_{1y} \\ f_{2x} & f_{2y} \end{bmatrix} = \begin{bmatrix} 2x & 2y \\ 2x & -2y \end{bmatrix}.$$

Na vsakem koraku rešimo:

$$\begin{bmatrix} 2x_r & 2y_r \\ 2x_r & -2y_r \end{bmatrix} \begin{bmatrix} \Delta x_r \\ \Delta y_r \end{bmatrix} = - \begin{bmatrix} x_r^2 + y_r^2 - 1 \\ x_r^2 - y_r^2 - 1 \end{bmatrix}.$$

Izračunamo $x_{r+1} = x_r + \Delta x_r$ in $y_{r+1} = y_r + \Delta y_r$. To ponavljamo za $r = 0, 1, \ldots$

$$JF(1.5, 1) = \begin{bmatrix} 3 & 2 \\ 3 & -2 \end{bmatrix}.$$

Rešimo sistem

$$\begin{bmatrix} 3 & 2 \\ 3 & -2 \end{bmatrix} \begin{bmatrix} \Delta x_r \\ \Delta y_r \end{bmatrix} = - \begin{bmatrix} -0.75 \\ 0.25 \end{bmatrix}$$

in dobimo $\Delta x_0 = \frac{1}{12}$ ter $\Delta y_0 = 0.25$. Kar pomeni $x_1 = x_0 + \Delta x_0 = 1.5 + \frac{1}{12} = \frac{19}{12}$ in $y_1 = y_0 + \Delta y_0 = 1 + \frac{1}{4} = \frac{5}{4}$. ∎

Naloga 3.10 *Poišči ničle funkcije* $f(x) = x + 4 - e^{-x^2}$ *z Matlabom in Mathematico.*

Rešitev. Oglejte si priložene datoteke. ∎

Poglavje 4

Matrične norme

Matrična norma je preslikava $||.|| : \mathbb{C}^{n\times n} \to \mathbb{R}^+$, za katero velja

(i). $||A|| \geq 0,\ ||A|| = 0 \Leftrightarrow A = 0$,

(ii). $||\alpha A|| = |\alpha| ||A||$,

(iii). $||A + B|| \leq ||A|| + ||B||$,

(iv). $||AB|| \leq ||A|| \cdot ||B||$, submultiplikativnost.

Za poljubni matriki A in B in poljuben $\alpha \in \mathbb{C}$.

Naloga 4.1 *Dokaži, da velja* $||A||_F^2 = \mathrm{sl}(A^H A) = \sum_{i=1}^n \lambda_i(A^H A)$, *kjer so* λ_i *lastne vrednosti* $A^H A$.

Rešitev. Za $B = A^H A$ velja $b_{ii} = \sum_{k=1}^n \overline{a_{ki}} a_{ki} = \sum_{k=1}^n |a_{ki}|^2$. Tako za sled B dobimo

$$\mathrm{sled}(A^H A) = \sum_{i=1}^n \left(\sum_{k=1}^n |a_{ki}|^2 \right) = ||A||_F^2.$$

Matrika $A^H A$ je simetrična, torej se da diagonalizirati. Podobna je matriki z lastnimi vrednostmi na diagonali. Sled podobne matrike je enaka sledi prvotne matrike, tako dobimo $||A||_F^2 = sl(A^H A) = \sum_{i=1}^n \lambda_i(A^H A)$. ∎

Naloga 4.2 *Pokaži, da velja*

$$||A||_1 = \max_i ||A(:, i)|| = \max_i \sum_{j=1}^n |a_{ij}.$$

Rešitev. Upoštevajmo definicijo norme $||A||_1 = \max_{||x||_1=1} ||Ax||_1$, označimo še $y = Ax$. Tako velja $y_i = \sum_{j=1}^n a_{ij} x_j$ in

$$||y|| = \sum_{i=1}^n \left| \sum_{j=1}^n a_{ij} x_j \right| \leq \sum_{j=1}^n |x_j| \sum_{i=1}^n |a_{ij}| \leq \sum_{j=1}^n |x_j| \max_j \left(\sum_{i=1}^n |a_{ij}| \right) = \max_j \left(\sum_{i=1}^n |a_{ij}| \right).$$

Enakost je dosežene za vektor e_k, kjer je k indeks stolpca z največjo prvo normo. ∎

Naloga 4.3 *Pokaži naslednje:*

a) $\frac{1}{\sqrt{n}}||A||_F \leq ||A||_2 \leq ||A||_F$

b) $\frac{1}{\sqrt{n}}||A||_\infty \leq ||A||_2 \leq \sqrt{n}||A||_\infty$

c) $\frac{1}{\sqrt{n}}||A||_1 \leq ||A||_2 \leq \sqrt{n}||A||_1$

d) $N_\infty(A) \leq ||A||_2 \leq nN_\infty(A)$

Rešitev. a)
Vemo, da velja

$$||A||_2 = \max_{i=1,\dots,n} \sqrt{\lambda_i(A^H A)} = \sigma_1(A) \text{ in } ||A||_F^2 = \sum_{i=1}^{n} \lambda_i(A^H A).$$

Lastne vrednosti $A^H A$ so nenegativne, saj velja $\langle A^H Ax, x\rangle = \langle \mu x, x\rangle = \mu = \langle Ax, Ax\rangle \geq 0$, $\lambda_i = \sigma_i^2$. Razporedimo jih v zaporedje $\sigma_1^2 \geq \sigma_2^2 \ldots \geq \sigma_n^2 \geq 0$. Pozitivni koreni teh lastnih vrednosti $\sigma_1 \geq \sigma_2 \geq \ldots \geq 0$ so singularne vrednosti matrike A. Očitno je $||A||_2 \leq ||A||_F$, saj je $||A||_2$ enaka $\sigma_1(A)$, kar je največja singularna vrednost matrike A. Poglejmo si $||A||_F^2 = \sum_{i=1}^n \sigma_i^2 \leq n\sigma_1^2 = n||A||_2$, kar pomeni $\frac{1}{\sqrt{n}}||A||_F \leq ||A||_2$. Neenačaj dobimo, ker je $\sigma_1(A)$ največja singularna vrednost.
b)
Za vektorske norme velja

$$||x||_\infty = \max_{1\leq i\leq n} |x_i| \leq \sqrt{|x_1|^2 + \ldots |x_n|^2} = ||x||_2 \leq \sqrt{n \max_i |x_i|^2} = \sqrt{n}||x||_\infty,$$

to je $||x||_\infty \leq ||x||_2 \leq \sqrt{n}||x||_\infty$. Razmisli zakaj neeneakosti držijo, vsi sklepi so enostavni. Iz tega dobimo

$$||A||_2 = \max_{x\neq 0} \frac{||Ax||_2}{||x||_2} \leq \max_{x\neq 0} \frac{\sqrt{n}||Ax||_\infty}{||x||_\infty} = \sqrt{n}||A||_\infty$$

in

$$||A||_2 = \max_{x\neq 0} \frac{||Ax||_2}{||x||_2} \geq \max_{x\neq 0} \frac{||Ax||_\infty}{\sqrt{n}||x||_\infty} = \frac{1}{\sqrt{n}}||A||_\infty.$$

c)
Velja $||A||_1 = ||A^H||_\infty$ in $||A||_2 = ||A^H||_2$. Iz tega že sledi neenakost.
d)
Iz a) vemo

$$||A||_2 \leq ||A||_F = \sqrt{\sum_{i,j} |a_{ij}|^2} \leq \sqrt{n^2 \max_{i,j} |a_{ij}|^2} = n \max_{i,j} |a_{ij}| = nN_\infty(A).$$

Za drugi del neenakosti upoštevamo $a_{ij} = e_i^T Ae_j$. Računajmo

$$|a_{ij}| = |e_i^T Ae_j| \leq ||e_i||_2||Ae_j||_2 = \frac{||Ae_j||_2}{||e_j||_2} \leq ||A||_2.$$

Prvi neenačaj dobimo po Cauchy-Schwartzovi neenakosti, zadnjega pa po definiciji norme $||A||_2$.
∎

Naloga 4.4 *Pokaži, da velja*

a) $|\lambda| \leq ||A||$

b) $||A||_2^2 \leq ||A||_\infty ||A||_1$

c) $||a_i||_2 \leq ||A||_2$

Rešitev.

a) Točka a) je očitna. Naj bo x enotski lastni vektor za lastno vrednost λ, potem velja $||Ax|| = ||\lambda x|| = |\lambda| ||x|| = \lambda$. Norma je supremum po vseh enotskih x, torej večja ali enaka.

b) Vemo, da je $||A||_2^2$ največja lastna vrednost matrike $A^H A$. Iz a) dobimo

$$||A||_2^2 \leq ||A^H A||_\infty \leq ||A^H||_\infty ||A||_\infty = ||A||_1 ||A||_\infty.$$

c) Trditev sledi za i-ti enotski vektor e_i. Kjer velja $Ae_i = a_i$.

■

Naloga 4.5 *Izračunaj* $||\ ||_1$, $||\ ||_\infty$, $||\ ||_F||$ *za* $\begin{bmatrix} 2 & -1 & 3 \\ 5 & 4 & 1 \\ -2 & -1 & 2 \end{bmatrix}$ *in oceni* $||\ ||_2$.

Rešitev. Izračunajmo

$$\begin{aligned} ||A||_1 &= \max\{|2| + |5| + |-2|,\ |-1| + |4| + |-1|,\ |3| + |1| + |2|\} = 9, \\ ||A||_\infty &= \max\{|2| + |-1| + |3|,\ |5| + |4| + |1|,\ |-2| + |-1| + |2|\} = 10, \\ ||A||_F &= \sqrt{4+1+9+25+16+1+4+4} = \sqrt{65} \doteq 8.06226. \end{aligned}$$

Ocenimo $||A||_2$

$$\begin{aligned} \frac{1}{\sqrt{3}}||A||_F \leq ||A||_2 \leq ||A||_F &\quad\Rightarrow\quad 4.65475 \leq ||A||_2 \leq 8.06226 \\ \frac{1}{\sqrt{3}}||A||_\infty \leq ||A||_2 \leq \sqrt{3}||A||_\infty &\quad\Rightarrow\quad 5.7735 \leq ||A||_2 \leq 17.3205 \\ \frac{1}{\sqrt{3}}||A||_1 \leq ||A||_2 \leq \sqrt{3}||A||_1 &\quad\Rightarrow\quad 5.19615 \leq ||A||_2 \leq 15.58846 \\ N_\infty(A) \leq ||A||_2 \leq 3N_\infty(A) &\quad\Rightarrow\quad 5 \leq ||A||_2 \leq 15. \end{aligned}$$

Skupaj dobimo oceno $5.7735 \leq ||A||_2 \leq 8.06226$. Boljšo oceno dobimo, če poskusimo oceniti spektralni radij matrike

$$B = A^H A = \begin{bmatrix} 2 & 5 & -2 \\ -1 & 4 & -1 \\ 3 & 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & -1 & 3 \\ 5 & 4 & 1 \\ -2 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 33 & 20 & 7 \\ 20 & 8 & -1 \\ 7 & -1 & 14 \end{bmatrix}.$$

Za vsako lastno vrednost in vsako normo velja $\lambda_i(B) \leq ||B|_F = 50.0899$. Vemo, da velja

$$||A||_2 \leq \sqrt{||B||_F} = 7.0774.$$

Oceno navzdol dobimo, če upoštevamo $||A||_2 = ||A^T||_2$ in $||Ae_i||_2 \leq ||A||_2 ||e_i||_2 = ||A||_2$, zadnja enakost sledi iz definicije matrične norme. Vektorji Ae_i so ravno i-ti stolpci, $A^T e_i$ so i-te vrstice. Norme stolpcev so $5.7446,\ 4.2426, 3.7417$. Norme vrstic so $3.7417,\ 6.4807,\ 3$. Torej dobimo $||A||_2 \geq 6.4807$. Končna ocena je $6.4807 \leq ||A||_2 \leq 7.0774$. Ukaz iz Matlaba vrne $\texttt{norm}(A) = 6.9044$. ■

Naloga 4.6 *Izračunaj* $||A||_1$, $||A||_\infty$, $||A||_F$ *in čim natančneje oceni* $||A||_2$ *na obe strani, če je*

$$\begin{aligned} a_{ii} &= -2i,\ i = 1,\ 2, \ldots,\ n \\ a_{i+1,i} &= a_{i,i+1} = n-i,\ i = 1,\ 2, \ldots,\ n, \end{aligned}$$

drugi elementi so enaki 0.

$$A = \begin{bmatrix} -2 & n-1 & & & \\ n-1 & -4 & n-2 & & \\ & n-2 & -6 & & \\ & & \ddots & \ddots & 1 \\ & & & 1 & -2n \end{bmatrix}.$$

Upoštevaj, da velja

$$\sum_{i=1}^{n-1} i^2 = \frac{(n-1)n(2n-1)}{6}.$$

Rešitev. Pri $i = 1$ dobimo za vsoto absolutnih vrednosti v stolpcu $n+1$, kar očitno ni maksimum. Izračunajmo

$$\begin{aligned} ||A||_1 &= \max_{i \neq 1}\{|-2i| + |n-i| + |n-i+1|\} = \max_{i \neq 1}\{2i + n - i + n - i + 1\} \\ &= \max_{i \neq 1}\{2n+1\} = 2n+1. \end{aligned}$$

Ker je matrika simetrična velja $||A||_1 = ||A^H||_\infty = ||A||_\infty = 2n+1$. Frobeniusova norma je

$$\begin{aligned} ||A||_F^2 = \sum_{i=1}^{n}(2i)^2 + 2\sum_{i=1}^{n} i^2 = 6\sum_{i=1}^{n-1} i^2 + 4n^2 = 6\frac{(n-1)n(2n-1)}{6} + 4n^2 = \\ n(n-1)(2n-1) + 4n^2 = n(2n^2 - 3n + 1) + 4n^2 = 2n^3 + n^2 + n. \end{aligned}$$

Poglejmo si ocene za drugo normo

$$\begin{aligned} \frac{1}{\sqrt{n}}||A||_F \leq ||A||_2 \leq ||A||_F \Rightarrow \\ \frac{1}{\sqrt{n}}\sqrt{2n^3 + n^2 + n} \leq ||A||_2 \leq \sqrt{2n^3 + n^2 + n} \\ \frac{1}{\sqrt{n}}||A||_\infty \leq ||A||_2 \leq \sqrt{n}||A||_\infty \Rightarrow \frac{1}{\sqrt{n}}(2n+1) \leq ||A||_2 \leq \sqrt{n}(2n+1) \\ N_\infty(A) \leq ||A||_2 \leq nN_\infty(A) \Rightarrow 2n \leq ||A||_2 \leq 2n^2. \end{aligned}$$

Upoštevamo še, da velja $||A||_2^2 \leq ||A||_1 ||A||_\infty$. Iz tega dobimo

$$||A||_2^2 \leq (2n+1)^2,$$

kar pomeni $||A||_2 \leq 2n+1$. Za drugo normo smo dobili oceno

$$2n \leq ||A||_2 \leq 2n+1.$$

■

Poglavje 5

Reševanje linearnih sistemov

Pogojenostno število matrike, $\mathrm{cond}(A) = ||A||_2||A^{-1}||_2$ nam pove kako občutljivo je reševanje sistema. Da lahko dobimo poljubno občutljive matrike v praksi, pokaže naslednja naloga.

Naloga 5.1 *Iščemo koeficiente polinoma $p(x) = a_1 + a_2x + \cdots + a_nx^{n-1}$, ki na $[0, 1]$ aproksimira zvezno funkcijo f tako, da je napaka*

$$E = \int_0^1 (f(x) - p(x))^2 dx$$

minimalna. Torej mora veljati $\frac{\partial E}{\partial a_i} = 0$ za $i = 1, \ldots, n$, kjer je

$$\frac{\partial E}{\partial a_i} = 2\int_0^1 (p(x) - f(x))x^{i-1}dx.$$

Definirajmo vektor

$$F = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{bmatrix} \qquad \textit{in;} \quad a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

Od tod dobimo

$$F_i = \int_0^1 f(x)x^{i-1}dx = \int_0^1 \left(\sum_{j=1}^n a_jx^{j-1}x^{i-1}\right) dx = \sum_{j=1}^n a_j \int_0^1 x^{i+j-2}dx = \sum_{j=1}^n a_j \frac{1}{i+j-1}.$$

Dobimo sistem s Hilbertovo matriko H_n, $H_{ij} = \frac{1}{i+j-1}$. Kar je sistem $Ha = F$. Primer za $n = 5$ je

$$H_5 = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} \\ \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} & \frac{1}{9} \end{bmatrix}.$$

Hilbertove matrike so primer zelo občutljivih matrik. Pogojenostno število H_5 je recimo približno $4.766 \cdot 10^5$, kar lahko izračunamo z ukazom `cond` *v Matlabu. Pogojenostna števila matrik H_n hitro rastejo.*

Izkaže se, da smo dobili občutljiv sistem, ker smo vzeli standardno bazo polinomov stopnje n. Za stabilno računanje moramo vzeti ortogonalno bazo polinomov. Priložena je Matlabova skripta za aproksimacijo integrala funkcije $f(x) = e^x$ s parabolo.

Naloga 5.2 *Oceni relativno neodstranljivo napako, ki nastane pri množenju vektorja x z A, če x spremenimo za δx. Matrika A je obrnljiva.*

Rešitev. Velja $y = Ax$ in $A(x + \delta x) = y + \delta y$. Zanima nas

$$\frac{||\delta y||}{||y||} = \frac{||A(x + \delta x) - Ax||}{||Ax||}.$$

Očitno velja $||\delta y|| \leq ||A||||\delta x||$ in $||x|| \leq ||A^{-1}||||y||$. Izračunamo

$$\frac{||\delta y||}{||y||} = \frac{||A\delta x|| \cdot ||x||}{||y|| \cdot ||x||}||x|| \leq \frac{||A|| \cdot ||\delta x|| \cdot ||A^{-1}|| \cdot ||y||}{||y|| \cdot ||x||} = \underbrace{||A||||A^{-1}||}_{\text{cond}(A)} \frac{||\delta x||}{||x||}.$$

∎

Naloga 5.3 *Izpelji aposteriorno oceno*

$$\frac{||\delta x||}{||x||} \leq \text{cond(A)}\frac{||r||}{||b||},$$

ki je uporabna za oceno točnosti rešitve.

Rešitev. Naj bo $\widetilde{x}$ rešitev, ki jo dobimo z numeričnim izračunom. Naj bo $\delta x = \widetilde{x} - x$ in $r = A\widetilde{x} - b$. Ocenimo

$$||\widetilde{x} - x|| = ||A^{-1}(A\widetilde{x} - b)|| = ||A^{-1}r|| \leq ||A^{-1}|| \cdot ||r||$$

in

$$||b|| = ||Ax|| \leq ||A|| \cdot ||x||.$$

Oceni združimo in dobimo željeno.

∎

Izrek 5.1 *Naslednji trditvi sta ekvivalentni*

(i). Obstaja enolični razcep $A = LU$, kjer je L spodnje trikotna matrika z enicami na diagonali, U pa nesingularna zgornje trikotna matrika.

(ii). Vse vodilne podmatrike $A(1 : k, 1 : k)$ so nesingularne.

Algoritem 3: LU razcep brez pivotiranja

---

for $j = 1, \ldots, n - 1$ **do**
  for $i = 1, \ldots, n$ **do**
    $l_{ij} = \frac{a_{ij}}{a_{jj}}$;
    for $k = j + 1, \ldots, n$ **do**
      $a_{ik} = a_{ik} - l_{ij}a_{jk}$;
    end
  end
end

Število operacij je $\frac{2}{3}n^2 + O(n^2)$.

Izrek 5.2 *Če je A nesingularna, potem obstaja taka permutacijska matrika P, da obstaja LU razcep $PA = LU$.*

Algoritem 4: LU razcep z delnim pivotiranja

```
L = 0;
P = I;
for j = 1,...,n-1 do
    Poišči |a_qj| = max_{j≤p≤n} |a_pj|;
    Zamenjaj vrstici q in j v L, A in P;
    for i = j,...,n do
        l_ij = a_ij / a_jj;
        for k = j+1,...,n do
            a_ik = a_ik - l_ij a_jk;
        end
    end
end
```

Število operacij je $\frac{2}{3}n^3 + O(n^2)$.

Naloga 5.4 *Z LU-razcepom brez pivotiranja reši sistem $Ax = b$, kjer je*

$$A = \begin{bmatrix} 2 & 3 & -4 \\ -6 & -11 & 13 \\ 4 & 0 & -8 \end{bmatrix}, \; b = \begin{bmatrix} -11 \\ 38 \\ -16 \end{bmatrix}.$$

Zapiši matriki L, U in vektor y, ki ga dobiš pri računanju.

Rešitev.
Lotimo se Gaussove eliminacije:

$$\begin{bmatrix} 2 & 3 & -4 \\ -6 & -11 & 13 \\ 4 & 0 & 8 \end{bmatrix} \to \begin{bmatrix} 2 & 3 & -4 \\ 0 & -2 & 1 \\ 0 & -6 & 0 \end{bmatrix} \to \begin{bmatrix} 2 & 3 & -4 \\ 0 & -2 & 1 \\ 0 & 0 & 3 \end{bmatrix} = U$$

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ -3 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 2 & 3 & 1 \end{bmatrix} = L.$$

Izračunamo lahko tudi v bolj kompaktni obliki. Zgornji trikotnik končne matrike vsebuje matriko U, spodnji trikotnik brez diagonale pa matriko L brez diagonale. Upoštevamo, da ima matrika L na diagonali enice.

$$\begin{bmatrix} 2 & 3 & -4 \\ -6 & -11 & 13 \\ 4 & 0 & -8 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & 3 & -4 \\ -3 & -2 & 1 \\ 2 & -6 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & 3 & -4 \\ -3 & -2 & 1 \\ 2 & 3 & -3 \end{bmatrix},$$

zgornji trikotnik je U, spodnji trikotnik brez diagonale je matrika L brez diagonale.

Rešimo sistem $Ax = b$, $L(Ux) = b$, $y = Ux$.

Algoritem 5: Prema substitucija, $Ly = b$, $n^2 - n$ operacij

---

```
for i = 1,2,...,n do
    y_i = b_i - sum_{k=1}^{i-1} l_ik y_k;
end
```

$$\begin{bmatrix} 1 & & \\ -3 & 1 & \\ 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -11 \\ 38 \\ -16 \end{bmatrix}, \qquad \begin{aligned} y_1 &= -11 \\ y_2 &= 38 + 3 \cdot (-11) = 5 \\ y_3 &= -16 - 2 \cdot (-11) - 35 = -9 \end{aligned}$$

Algoritem 6: Obratna substitucija, $Ux = y$, n^2 operacij

for $i = n, n-1, \ldots, 1$ **do**
$\quad x_i = \frac{1}{u_{ii}} \left(y_i - \sum_{k=i+1}^{n} u_{ik} x_k\right)$;
end

$$\begin{bmatrix} 2 & 3 & -4 \\ 0 & -2 & 1 \\ 0 & 0 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -11 \\ 5 \\ -9 \end{bmatrix}, \qquad \begin{aligned} x_3 &= 3 \\ x_2 &= \tfrac{-1}{2}(5-3) = -1 \\ x_1 &= \tfrac{1}{2}(-11+3+12) = 2 \end{aligned}.$$

Torej dobimo $x = \begin{bmatrix} 2 & -1 & 3 \end{bmatrix}^T$. ∎

Naloga 5.5

- *Sestavi ekonomičen algoritem za izračun inverzne matrike spodnje trikotne matrike L z enicami po diagonali in preštej število operacij.*
- *Prilagodi algoritem iz prve točke, tako da boš lahko izračunal inverz zgornje trikotne matrike U. Preštej še število operacij.*
- *S pomočjo prejšnjih dveh točk, sestavi učinkovit algoritem za izračun inverza matrike A, če imaš podan njen LU razcep. Kakšno je število operacije?*

Rešitev.

- Iz pravila za izračun inverza vidimo, da je inverzna matrika zopet spodnje trikotna matrika z enicami na diagonali. Rešujemo sistem $LY = I$, Y je oblike $Y = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}$. Dobimo več podproblemov oblike

$$\begin{bmatrix} 1 & & & & \\ * & \ddots & & & \\ l_{j,1} & l_{j,2} \cdots l_{j,j-1} & 1 & & \\ lj+1,1 & l_{j+1,2} \cdots l_{j+1,j} & & \ddots & \\ * & * & * & * & 1 \end{bmatrix} \begin{bmatrix} 0 \\ \\ y_{j,j} \\ \vdots \\ y_{n,j} \end{bmatrix} = \begin{bmatrix} 0 \\ \\ 1 \\ \\ 0 \end{bmatrix} = e_j.$$

Vemo, da je $y_{j,j} = 1$. Za $i > j$ dobimo

$$l_{i,j} y_{j,j} + l_{i,j+1} y_{j+1,j} + \cdots l_{i,i-1} y_{i-1,j} + l_{i,i} y_{i,j} = 0,$$

kar nam da $y_{ij} = -\sum_{k=j}^{i-1} l_{ik} y_{kj}$. Vseh operacij je

$$\sum_{j=1}^{n} \sum_{i=j+1}^{n} 2(i-j) = 2 \sum_{j=1}^{n} \frac{(n-j)(n-j+1)}{2} = \sum_{k=1}^{n-1} k(k+1) = \sum_{k=1}^{n-1} (k^2 + k)$$

$$= \frac{(n-1)n(2n-1)}{6} + \frac{n(n-1)}{2} = \frac{n^3}{3} + O(n^2).$$

---

for $j = 1, \ldots, n$ **do**
  $y_{j,j} = 1$;
  for $i = j+1, \ldots, n$ **do**
    $y_{i,j} = -\sum_{k=j}^{i-1} l_{ik} y_{kj}$;
  end
end

---

for $j = 1, \ldots, n$ *($UY_j = e_j$)* **do**
  $y_{jj} = 1/u_{jj}$;
  for $i = j-1, \ldots, 1$ **do**
    $y_{ij} = \frac{1}{u_{ii}} \left(-\sum_{k=j-1}^{i} u_{ik} y_{kj} \right)$;
  end
end

- Uporabimo idejo obratne substitucije in resujemo podobno kot pri prvi tocki.

  Algoritem 7: Obratna substitucija, $Ux = y$

  for $i = n, n-1, \ldots, 1$ **do**
    $x_i = \frac{1}{u_{ii}} \left(y_i - \sum_{k=i+1}^{n} u_{ik} x_k \right)$;
  end

  Število operacij je spet $\frac{n^3}{3} + O(n^2)$, v resnici imamo samo $O(n^2)$ deljenj več.

- Upoštevajmo še, da velja

$$A^{-1} = (LU)^{-1} = U^{-1} L^{-1}.$$

  Ugotovimo moramo še kakšna je zahtevnost množenja spodnje in zgornje trikotne matrike.

$$\sum_{i=1}^{n} \sum_{j=1}^{i} j + (n-i)i = \sum_{i=1}^{n} i(i+1)/2 + ni - i^2 = \frac{2n^3}{3} + O(n^2).$$

  V resnici bi bilo bolje (bolj stabilno), če bi zadnji dve točki naredli hkrati, tako da bi rešili sistem $UX = L^{-1}$ in direktno izračunali $U^{-1}L^{-1}$. Skupno število operacij ostane $2n^3 + O(n^2)$.

■

Naloga 5.6 *Sestavi učinkovit algoritem za reševanje sistema linearnih enačb*

$$\begin{bmatrix} U & -I \\ B & L \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}.$$

Matrika $B = LU$ je nesingularna matrika in ima podan LU razcep. Preštej število množenj in deljenj.

Rešitev. Zmnožimo po blokih in dobimo

$$\begin{aligned} Ux - y &= a \\ L(Ux + y) &= b. \end{aligned}$$

Rešimo drugi sistem in dobimo vektor $z = Ux + y$, kar nas stane $n^2 + O(n)$ operacij. Označimo še $w = Ux - y$ in izrazimo $y = \frac{1}{2}(z - w) = \frac{1}{2}(z - a)$. Za izračun y porabimo $O(n)$ operacij. Na koncu rešimo še sistem $Ux = a + y$, za kar porabimo $n^2 + O(n)$ operacij. Skupaj smo porabili $2n^2 + O(n)$ operacij. ■

Naloga 5.7 *Za matriko*

$$A = \begin{bmatrix} 3 & 1 & 1 \\ 0 & 0 & 1 \\ 6 & 1 & 7 \end{bmatrix}$$

naredi:

(i). LU razcep z delnim pivotiranjem,

(ii). LU razcep s kompletnim pivotiranjem.

Rešitev. (i)Algoritem poteka takole. V stolpcu i poiščemo največji element, izmed a_{ji}, $j \geq i$ po absolutni vrednostni. Vrstico največjega elementa in i-to vrstico zamenjamo v L, A in P. Normalno nadaljujemo z Gaussovo eliminacijo. Na koncu velja $PA = LU$.

$$A = \begin{bmatrix} 3 & 1 & 1 \\ 0 & 0 & 1 \\ \underline{6} & 1 & 7 \end{bmatrix} \overset{1\updownarrow 3}{\rightarrow} \begin{bmatrix} 6 & 1 & 7 \\ 0 & 0 & 1 \\ 3 & 1 & 1 \end{bmatrix} \overset{G}{\rightarrow} \begin{bmatrix} 6 & 1 & 7 \\ 0 & 0 & 1 \\ 0 & \underline{1/2} & -5/2 \end{bmatrix} \overset{2\updownarrow 3}{\rightarrow} \begin{bmatrix} 6 & 1 & 7 \\ 0 & 1/2 & -5/2 \\ 0 & 0 & 1 \end{bmatrix} = U$$

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \overset{1\updownarrow 3}{\rightarrow} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \overset{G}{\rightarrow} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1/2 & 0 & 0 \end{bmatrix} \overset{2\updownarrow 3}{\rightarrow} \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \rightrightarrows \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = L$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \overset{1\updownarrow 3}{\rightarrow} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \overset{2\updownarrow 3}{\rightarrow} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = P$$

(ii)Algoritem poteka takole. V podmatriki $A(i:n, i:n)$ poiščemo največji element po absolutni vrednostni, recimo da je to a_{kl}. Da ta element spravimo na mesto (i,i) zamenjamo i-to vrstico in k-to vrstico ter i-ti stolpec in k-ti stolpec. To naredimo tudi za matriko L. V P menjamo samo vrstice, v Q pa samo stolpce. Normalno nadaljujemo z Gaussovo eliminacijo. Na koncu velja $PAQ = LU$.

$$A = \begin{bmatrix} 3 & 1 & 1 \\ 0 & 0 & 1 \\ 6 & 1 & \underline{7} \end{bmatrix} \overset{1\updownarrow 3, 1\leftrightarrow 3}{\rightarrow} \begin{bmatrix} 7 & 1 & 6 \\ 1 & 0 & 0 \\ 1 & 1 & 3 \end{bmatrix} \overset{G}{\rightarrow} \begin{bmatrix} 7 & 1 & 6 \\ 0 & -\frac{1}{7} & -\frac{6}{7} \\ 0 & \frac{6}{7} & \underline{\frac{15}{7}} \end{bmatrix} \overset{2\updownarrow 3, 2\leftrightarrow 3}{\rightarrow} \begin{bmatrix} 7 & 6 & 1 \\ 0 & \frac{15}{7} & \frac{6}{7} \\ 0 & -\frac{6}{7} & \frac{1}{7} \end{bmatrix} \overset{G}{\rightarrow} \begin{bmatrix} 7 & 6 & 1 \\ 0 & \frac{15}{7} & \frac{6}{7} \\ 0 & 0 & \frac{1}{5} \end{bmatrix} = U$$

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \overset{1\updownarrow 3, 1\leftrightarrow 3}{\rightarrow} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \overset{G}{\rightarrow} \begin{bmatrix} 1 & 0 & 0 \\ 1/7 & 0 & 0 \\ 1/7 & 0 & 0 \end{bmatrix} \overset{2\updownarrow 3, 2\leftrightarrow 3}{\rightarrow} \begin{bmatrix} 1 & 0 & 0 \\ 1/7 & 0 & 0 \\ 1/7 & 0 & 0 \end{bmatrix} \overset{G}{\rightarrow} \begin{bmatrix} 1 & 0 & 0 \\ 1/7 & 1 & 0 \\ 1/7 & -2/5 & 1 \end{bmatrix} = L$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \overset{1\updownarrow 3}{\to} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \overset{2\updownarrow 3}{\to} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = P$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \overset{1\leftrightarrow 3}{\to} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \overset{2\leftrightarrow 3}{\to} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = Q$$

■

Naloga 5.8 *Sestavi ekonomičen (čim manj operacij) algoritem za računanje tridiagonalnega sistema preko LU razcepa brez pivotiranja.*

Rešitev. Matrika A je oblike

$$A = \begin{bmatrix} a_1 & c_1 & & & \\ b_2 & a_2 & c_2 & & \\ & b_3 & \ddots & \ddots & \\ & & \ddots & \ddots & c_{n-1} \\ & & & b_n & a_n \end{bmatrix},$$

torej bosta matriki L in U oblike

$$L = \begin{bmatrix} 1 & & & & \\ l_2 & 1 & & & \\ & l_3 & 1 & & \\ & & \ddots & \ddots & \\ & & & l_n & 1 \end{bmatrix} \text{ in } U = \begin{bmatrix} u_1 & v_1 & & & \\ & u_2 & v_2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & v_{n-1} \\ & & & & u_n \end{bmatrix}.$$

Zmnožek L in U je oblike

$$LU = \begin{bmatrix} u_1 & v_1 & & & \\ l_2u_1 & l_2v_1 + u_2 & v_2 & & \\ & l_3u_2 & l_3v_2 + u_3 & \ddots & \\ & & \ddots & \ddots & v_{n-1} \\ & & & l_nu_{n-1} & l_nv_{n-1} \end{bmatrix},$$

iz česar dobimo $v_i = c_i$, $l_{i+1}u_i = b_{i+1}$ in $l_{i+1}v_i + u_{i+1} = a_i$ za $i = 1, \ldots, n-1$. Poleg tega velja še $u_1 = a_1$.

$u_1 = a_1$;
$v_1 = c_1$;
for $i = 2, \ldots, n$ **do**
  $v_i = c_i$;
  $l_i = \frac{b_i}{u_{i-1}}$;
  $u_i = a_i - v_{i-1}l_i$;
end

Število operacij je $\sum_{i=2}^{n} 3 = 3(n-1) = 3n-3$, kar je veliko bolje kot $2/3n^2 + O(n^2)$. Iz $L(Ux) = d$, $y = Ux$ dobimo $Ly = d$, kar je

$$\begin{bmatrix} 1 & & & & \\ l_2 & 1 & & & \\ & l_3 & 1 & & \\ & & \ddots & \ddots & \\ & & & l_n & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ \vdots \\ d_n \end{bmatrix}.$$

To da $y_1 = d_1$ in $y_{i-1}l_i + y_i = d_i$. Dobimo algoritem

$y_1 = d_1$;
for $i = 2, \ldots, n$ **do**
$\quad y_i = d_i - y_{i-1}l_i$;
end

Število operacij $\sum_{i=2}^{n} 2 = 2n-2$. Podobno za $Ux = y$ dobimo

$$\begin{bmatrix} u_1 & v_1 & & & \\ & u_2 & v_2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & v_{n-1} \\ & & & & u_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{bmatrix}.$$

Dobimo $u_ix_i + v_ix_{i+1} = y_i$ ter $u_nx_n = y_n$.
Za to porabimo $1 + \sum_{i=1}^{n-1} = 3(n-1) + 1 = 3n-2$ operacij. Vse skupaj porabimo $3n - 3 + 2n -$

$x_n = \frac{y_n}{u_n}$;
for $i = n-1, \ldots, 1$ **do**
$\quad x_i = \frac{y_i - v_ix_{i+1}}{u_i}$;
end

$2 + 3n - 2 = 8n - 7$ operacij. Ali lahko kaj podobnega naredimo tudi z delnim pivotiranjem? ■

Naloga 5.9 *Dana je nesingularna matrika A in vektorja x in b_1, da velja $Ax = b_1$ ter njen LU razcep. Zapiši algoritem za reševanje sistema $Bx = b$, kjer je*

$$B = \begin{bmatrix} A & u \\ v^T & \alpha \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} in x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

kjer je $u, v \in \mathbb{R}^n, \alpha \in \mathbb{R}$. Koliko operacij porabiš za reševanje razširjenega sistema?

Rešitev. Zmnožimo sistem po blokih,

$$\begin{bmatrix} A & u \\ v^T & \alpha \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} Ax_1 + ux_2 \\ v^Tx_1 + \alpha x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Iz prve enačbe dobimo $Ax_1 + ux_2 = b_1$, oziroma $x_1 + yx_2 = x$, kjer je y rešitev sistema $Ay = u$. Za reševanje sistem porabimo $2n^2 + O(n)$ operacij. Drugo enačba je $v^Tx_1 + \alpha x_2 = b_2$. Prvo enačbo pomnožimo z v^T in dobimo $v^T x_1 + v^T y x_2 = v^T x$, le to odštejemo od druge enačbe in dobimo $x_2(\alpha - v^T y) = b_2 - v^T x$. Od tukaj izrazimo $x_2 = (b_2 - v^T x)/(\alpha - v^T y)$. Za izračun $v^T x$ in $v^T y$ porabimo $4n - 2$ operacij, dodati moramo še dve odštevanji in eno deljenje, skupaj dobimo $4n - 1$ operacij. Nato iz prve enačbe izrazimo še $x_1 = x - yx_2$, za kar porabimo $2n$ operacij. Vse skupaj smo porabili $2n^2 + 6n - 1 = 2n^2 + O(n)$ operacij.

■

Naloga 5.10 *Poišči razcep Choleskega za matriko*

$$A = \begin{bmatrix} 1 & 2 & -2 & 3 \\ 2 & 8 & -2 & 8 \\ -2 & -2 & 14 & -11 \\ 3 & 8 & -11 & 15 \end{bmatrix}.$$

Kaj to pomeni za matriko A?

Rešitev.

Algoritem 8: Razcep Choleskega – $A = V \cdot V^T,\ V = ?$

for $k = 1, \dots, n$ **do**
  $v_{kk} = \sqrt{a_{kk} - \sum_{i=1}^{k-1} v_{ki}^2}$;
  for $j = k+1, \dots, n$ **do**
    $v_{jk} = \frac{1}{v_{kk}} \left(a_{jk} - \sum_{i=1}^{k-1} v_{ji} \cdot v_{ki} \right)$;
  end
end

$$\begin{bmatrix} 1 & 2 & -2 & 3 \\ 2 & 8 & -2 & 8 \\ -2 & -2 & 14 & -11 \\ 3 & 8 & -11 & 15 \end{bmatrix} \to \begin{bmatrix} \sqrt{1} = 1 \\ 2/1 \\ -2/1 \\ 3/1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 2 & \sqrt{8 - 2^2} = 2 \\ -2 & 1/2(-2 - (-2 \cdot 2)) \\ 3 & 1/2(8 - (3 \cdot 2)) \end{bmatrix}$$

$$\to \begin{bmatrix} 1 & 0 & 0 \\ 2 & 2 & 0 \\ -2 & 1 & \sqrt{14 - (-2)^2 - 1^2} = 3 \\ 3 & 1 & 1/3(-11 - (3 \cdot -2 + 1 \cdot 1)) \end{bmatrix}$$

$$\to \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ -2 & 1 & 3 & 0 \\ 3 & 1 & -2 & \sqrt{15 - (3^2 + 1^2 + (-2)^2)} = 3 \end{bmatrix} = V.$$

Matrika A je simetrična in pozitivno definitna, saj se je razcep Choleskega izvedel. ■

Poglavje 6

Problemi najmanjših kvadratov, predoločeni sistemi

Radi bi rešili sistem $Ax = b$, kjer je $A \in \mathbb{C}^{m\times n}$ matrika polnega ranga, $x \in \mathbb{C}^n$, $b \in \mathbb{C}$ ter $m \geq n$. Sistem v splošnem ni rešljiv, zato minimiziramo normo ostanka $||Ax - b||_2$. Izkaže se, da je tak x ravno rešitev **normalnega sistema** $A^T Ax = A^T b$.

Naloga 6.1 *Merili smo tir delca, ki naj bi se gibal po paraboli.*

i	x_i	$f(x_i)$
1	-1	$\frac{11}{4}$
2	0	$\frac{7}{4}$
3	1	$\frac{1}{4}$
4	2	$\frac{13}{4}$

V bližini teh točk bi radi potegnili parabolo. Določi njene koeficiente po metodi najmanjših kvadratov. Uporabi normalni sistem in ga reši z Gaussovo eliminacijo (po metodi Choleskega).

Rešitev. Parabola naj bo $a + bx + cx^2$. Dobimo

$$\begin{array}{ccccccc} a & - & b & + & c & = & \frac{11}{4} \\ a & + & 0b & + & 0c & = & \frac{7}{4} \\ a & + & b & + & c & = & \frac{1}{4} \\ a & + & 2b & + & 4c & = & \frac{13}{4} \end{array} \qquad \overbrace{\begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}}^{A} \overbrace{\begin{bmatrix} a \\ b \\ c \end{bmatrix}}^{x} = \overbrace{\begin{bmatrix} \frac{11}{4} \\ \frac{7}{4} \\ \frac{1}{4} \\ \frac{13}{4} \end{bmatrix}}^{z},$$

kar je predoločen sistem $Ax = z$. Iz tega dobimo normalni sistem $\overbrace{A^T A}^{B} x = A^T z = v$. Izračunamo

$$B = A^T A = \begin{bmatrix} 4 & 2 & 6 \\ 2 & 6 & 8 \\ 6 & 8 & 18 \end{bmatrix} \text{ in } v = A^T z = \begin{bmatrix} 8 \\ 4 \\ 16 \end{bmatrix}.$$

Sistem rešimo z Gaussovo eleminacijo

$$\begin{bmatrix} 4 & 2 & 6 & | & 8 \\ 2 & 6 & 8 & | & 4 \\ 6 & 8 & 18 & | & 16 \end{bmatrix} \to \begin{bmatrix} 2 & 6 & 8 & | & 4 \\ 0 & -10 & -10 & | & 0 \\ 0 & -10 & -6 & | & 4 \end{bmatrix} \to \begin{bmatrix} 2 & 6 & 8 & | & 4 \\ 0 & -10 & -10 & | & 0 \\ 0 & 0 & 4 & | & 4 \end{bmatrix}$$

Lahko uporabimo razcep Choleskega. Najprej izračunamo razcep Choleskega matrike $B = VV^T$ in dobimo

$$V = \begin{bmatrix} 2 & & \\ 1 & \sqrt{5} & \\ 3 & \sqrt{5} & 2 \end{bmatrix}.$$

Dobimo sistem $V\overbrace{(V^T x)}^{y} = z$. Iz

$$\begin{bmatrix} 2 & & \\ 1 & \sqrt{5} & \\ 3 & \sqrt{5} & 2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \\ 16 \end{bmatrix}, \quad \begin{array}{ccccccc} 2y_1 & + & 0y_2 & + & 0y_3 & = & 8 \\ y_1 & + & \sqrt{5}y_2 & + & 0y_3 & = & 4 \\ 3y_1 & + & \sqrt{5}y_2 & + & 2y_3 & = & 16 \end{array}, \text{ dobimo } \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 2 \end{bmatrix}.$$

Rešiti moramo še sistem $V^T x = y$, kar je

$$\begin{bmatrix} 2 & 1 & 3 \\ 0 & \sqrt{5} & \sqrt{5} \\ 0 & 0 & 2 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 2 \end{bmatrix}.$$

Rešimo sistem in dobimo $a = 1$, $b = -1$, $c = 1$. Parabola je $1 - x + x^2$. ■

Naloga 6.2 *Podjetje ima na voljo podatke*

Območje	*Prodaja (y_i)*	*Populacija (a_i)*	*Zaslužek na prebivalca (b_i)*
1	*162*	*274*	*850*
2	*120*	*180*	*1120*
3	*223*	*375*	*740*
4	*131*	*205*	*970*
5	*67*	*86*	*1032*

.

Za napovedovanje prodaje uporabljajo model $y_i = x_i + a_i x_2 + b_i x_3$, *kjer so* x_1, x_2, x_3, *ki jih izračunajo po metodi najmanjših kvadratov. Določi* x_1, x_2, x_3, *in napovej prodajo na območju z populacijo* $a = 60$ *in zaslužkom* $b = 1050$. *Nalogo reši v Matlabu.*

Rešitev. Sistem modelnih enačb prepišemo v matrično obliko

$$\begin{bmatrix} 1 & a_1 & b_1 \\ 1 & a_2 & b_2 \\ 1 & a_3 & b_3 \\ 1 & a_4 & b_4 \\ 1 & a_5 & b_5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}.$$

Za rešitev v Matlabu glej priloženo datoteko `prodaja.m` ■

Izrek 6.1 *Naj bo* $A \in \mathbb{R}^{m\times n}$, *kjer je* $m \geq n$ *in* $rang(A) = n$. *Potem obstaja enolični QR razcep* $A = QR$, *kjer je* Q *pravokotna matrika dimenzije* $m \times n$ *z ortogonalnimi stolpci,* R *pa zgornje trikotna matrika s pozitivnimi diagonalnimi elementi.*

Pri reševanju predoločenega sistema si lahko pomagamo s QR razcepom. Boljša je različica z MGS. Dobimo, da moramo rešiti sistem $Rx = Q^T b$. Boljše je narediti razširjen QR razcep:

$$Ax - b = \begin{bmatrix} A & b \end{bmatrix}\begin{bmatrix} x \\ -1 \end{bmatrix} = \begin{bmatrix} Q & q_{n+1} \end{bmatrix}\begin{bmatrix} R & z \\ 0 & \rho \end{bmatrix}\begin{bmatrix} x \\ -1 \end{bmatrix} = Q(Rx - z) - \rho q_{n+1}.$$

Iz česar dobimo, da moramo rešiti sistem $Rx = z$, maksimum pa je enak ρ.

for $k = 1, \dots, n$ **do**
  $q_k = a_k$;
  for $i = 1, \dots, k-1$ **do**
    $r_{ik} = q_i^T a_k (CGS)$ ali $r_{ik} = q_i^T q_k (MGS)$;
    $q_k = q_k - r_{ik} q_i$;
  end
  $r_{kk} = ||q_k||_2$;
  $q_k = \frac{q_k}{r_{kk}}$;
end

Naloga 6.3 *Poišči normalni sistem za reševanje naslednjih dveh problemov najmanjših kvadratov, kjer je $A \in \mathbb{R}^{m \times n}$, $m \geq n$ in $rang(A) = n$.*

(i). Uteženi problem najmanjših kvadratov. Iščemo

$$\min_x ||D(Ax - b)||_2,$$

kjer je D nesingularna diagonalna.

(ii). Iščemo

$$\min_x ||Ax - b||_C,$$

kjer je C simetrična pozitivno definitna, ki generira normo

$$||x||_C = (x^T C x)^{\frac{1}{2}}.$$

Rešitev. (i)
Izračunajmo

$$\min_x ||D(Ax - b)|| = \min_x ||DAx - Db||,$$

matrika $F = DA$ je spet dimenzije $m \times n$. Označimo še $Db = \widehat{b}$, tako dobimo predoločen sistem $Fx = \widehat{b}$, iz česar dobimo

$$F^T F x = (DA)^T (DA) x = F^T \widehat{b} = (DA)^T (Db).$$

Razpisano je to enako

$$A^T D^2 A x = A^T D^2 b.$$

(ii)
Matrika C je simetrična pozitivno definitna, zato zanjo obstaja razcep Choleskega $C = V^T V$, kjer V zgornje trikotna. Izračunajmo

$$\min_x ||Ax - b||_C = \min_x \Big((Ax - b)^T C (Ax - b)\Big)^{\frac{1}{2}} = \min_x \left((Ax - b)^T V^T \overbrace{V(Ax - b)}^{y}\right)^{\frac{1}{2}},$$

uvedemo novo neznanko $y = U(Ax - b)$, dobimo

$$\min_x (y^T y)^{\frac{1}{2}} = \min_x < y, y >^{\frac{1}{2}} = \min_x ||y||_2 = \min_x ||V(Ax - b)||_2.$$

Torej iščemo
$$\min_x ||VAx - Vb||_2,$$
to je predoločen sistem
$$VAx = Vb \ \ /(VA)^T,$$
kar da
$$A^TV^TVAx = A^TV^TVb.$$
Z upoštevanjem $C = V^TV$ dobimo
$$A^TCAx = A^TCb.$$

■

Givensova rotacija R_{ik} je matrika enaka identiteti povsod razen v i-ti in k-ti vrstici in preslika i-to in k-to komponento vektorja x v vektorja y, ki ima k-to komponento enako 0.
$$R_{ik}^T([i,k],[i,k]) = \begin{bmatrix} \frac{x_i}{\sqrt{x_i^2+x_k^2}} & \frac{x_k}{\sqrt{x_i^2+x_k^2}} \\ -\frac{x_k}{\sqrt{x_i^2+x_k^2}} & \frac{x_i}{\sqrt{x_i^2+x_k^2}} \end{bmatrix}.$$

Householderjevo zrcaljenje je matrika oblike $H = I - \frac{2}{w^Tw}ww^T$ in predstavlja zrcaljenje prek ravnine določene z normalo w. Če hočemo preslikati vektor x v $\pm ke_1$, lahko to naredimo s Householderjevim zrcaljenjem določenim z $w = x + sign(x_1)||x||_2 e_1$.

Naloga 6.4 *Izračunaj QR razcep matrike*
$$A = \begin{bmatrix} 1 & -2 & 1 \\ 2 & 1 & 1 \\ 1 & 3 & 1 \end{bmatrix}$$
na različne načine:

- *s pomočjo MGS,*
- *s pomočjo Givensovih rotacij,*
- *s pomočjo Householderjevih zrcaljenj.*

Rešitev.

- Najprej normiramo prvi stolpec A, $||a_1|| = \sqrt{1+4+1} = \sqrt{6}$. Torej je
$$q_1 = a_1/||a_1|| = \begin{bmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$$
in $r_{11} = ||a_1|| = \sqrt{6}$. V drugem koraku ortogonaliziramo drugi stolpec A a_2 glede na q_1. Izračunamo $a_2 = a_2 - \langle q_1, a_2 \rangle q_1$. Velja $\langle q_1, a_2 \rangle = \frac{3}{\sqrt{6}}$. Tako dobimo $a_2 = \begin{bmatrix} -5/2 & 0 & 5/2 \end{bmatrix}^T$ in $r_{12} = \frac{3}{\sqrt{6}}$. Izračunamo $q_2 = a_2/||a_2|| = \begin{bmatrix} -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix}^T$ in $r_{22} = ||a_2|| = 5/\sqrt{2}$. Na koncu izračunamo še $a_3 = a_3 - \langle q_1, a_3 \rangle q_1 - \langle q_2, a_3 \rangle q_2$. Dobimo $r_{13} = \langle q_1, a_3 \rangle = 2\sqrt{\frac{2}{3}}$, $r_{23} = \langle q_2, a_3 \rangle = 0$, $r_{33} = ||a_3|| = \frac{1}{\sqrt{3}}$, $a_3 = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix}^T$.

- S pomočjo Givensove rotacije

$$R_{12}^T = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} & 0 \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

uničimo 2. element v prvem stolpcu:

$$\begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} & 0 \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & 1 \\ 2 & 1 & 1 \\ 1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{5} & 0 & 3/\sqrt{5} \\ 0 & \sqrt{5} & -1/\sqrt{5} \\ 1 & 3 & 1 \end{bmatrix}.$$

Z Givensovo rotacijo

$$R_{13}^T = \begin{bmatrix} \sqrt{5/6} & 0 & 1/\sqrt{6} \\ 0 & 1 & 0 \\ -1/\sqrt{6} & 0 & \sqrt{5/6} \end{bmatrix}$$

uničimo 3. element v dobljeni matriki. Dobimo

$$\begin{bmatrix} \sqrt{5/6} & 0 & 1/\sqrt{6} \\ 0 & 1 & 0 \\ -1/\sqrt{6} & 0 & \sqrt{5/6} \end{bmatrix} \begin{bmatrix} \sqrt{5} & 0 & 3/\sqrt{5} \\ 0 & \sqrt{5} & -1/\sqrt{5} \\ 1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{6} & \sqrt{3/2} & \sqrt{3/2}+1/\sqrt{6} \\ 0 & \sqrt{5} & -1/\sqrt{5} \\ 0 & \sqrt{15/2} & -\sqrt{3/10}+\sqrt{5/6} \end{bmatrix}.$$

Uporabimo še rotacijo

$$R_{23}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{2/5} & \sqrt{3/5} \\ 0 & -\sqrt{3/5} & \sqrt{2/5} \end{bmatrix}$$

in dobimo isto matriko kot pri MGS, matrika $Q = R_{12}R_{13}R_{23}$.

■

Naloga 6.5 *S Householderjevimi zrcaljenji in QR razcepom reši linearni sistem*

$$\begin{array}{rcrcrcr} x_1 & + & 6x_2 & + & -2x_3 & = & -7 \\ 2x_1 & + & x_2 & + & -2x_3 & = & -1 \\ 2x_1 & + & 2x_2 & + & 6x_3 & = & 6 \end{array}.$$

Rešitev. Najprej transformiramo prvi stolpec

$$w = \begin{bmatrix} 1+1\cdot(3=||A_1||_2) \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 2 \end{bmatrix},\ w^Tw = 24,\ P_1 = I - \frac{1}{12}ww^T = \frac{1}{3}\begin{bmatrix} -1 & -2 & -2 \\ -2 & 2 & -1 \\ -2 & -1 & 2 \end{bmatrix},$$

$$P_1 \begin{bmatrix} 1 & 6 & -2 & -7 \\ 2 & 1 & -2 & -1 \\ 2 & 2 & 6 & 6 \end{bmatrix} = \begin{bmatrix} -3 & -4 & -2 & -1 \\ 0 & -4 & -2 & 2 \\ 0 & -3 & 6 & 9 \end{bmatrix}.$$

Izračunajmo P_2:

$$k = ||\begin{bmatrix} -4 \\ -3 \end{bmatrix}|| = 5,\ w_1 = \begin{bmatrix} -4-5 \\ -3 \end{bmatrix} = \begin{bmatrix} -9 \\ -3 \end{bmatrix},\ w_1^Tw_1 = 90,$$

$$P_2 = I - \frac{1}{45} w_1 w_1^T = \frac{1}{5} \begin{bmatrix} -4 & -3 \\ -3 & 4 \end{bmatrix}.$$

Izračunajmo

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{-4}{5} & \frac{-3}{5} \\ 0 & \frac{-3}{5} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} -3 & -4 & -2 & -1 \\ 0 & -4 & -2 & 2 \\ 0 & -3 & 6 & 9 \end{bmatrix} = \begin{bmatrix} -3 & -4 & -2 & -1 \\ 0 & 5 & -2 & -7 \\ 0 & 0 & 6 & 6 \end{bmatrix}.$$

Iz česar dobimo $x_3 = 1$, $x_2 = \frac{-7+2}{5} = -1$, $x_1 = \frac{-1+2-4}{-3} = 1$. ∎

Naloga 6.6 *Naj bo A zgornja Hessenbergova matrika ($a_{ik} = 0,\ i > k+1$). Zapiši algoritem za QR razcep s pomočjo Givensovih rotacij in z njegovo pomočjo reši sistem $Ax = b$. Poleg tega še preštej število operacij (korenjenja, seštevanja, množenja). Matrike Q ti ni potrebno izračunati. Kakšna je oblika matrike Q?*

Rešitev. Givensova rotacija R_{ik} spremeni samo i-to in k-to vrstico. Poglejmo si primer, ko je matrika dimenzije 4:

$$A = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ & \times & \times & \times \\ & & \times & \times \end{bmatrix} \overset{R_{12}^T}{\to} \begin{bmatrix} \times & \times & \times & \times \\ & \times & \times & \times \\ & \times & \times & \times \\ & & \times & \times \end{bmatrix} \overset{R_{23}^T}{\to} \begin{bmatrix} \times & \times & \times & \times \\ & \times & \times & \times \\ & & \times & \times \\ & & \times & \times \end{bmatrix} \overset{R_{34}^T}{\to} \begin{bmatrix} \times & \times & \times & \times \\ & \times & \times & \times \\ & & \times & \times \\ & & & \times \end{bmatrix}$$

---

for $i = 1, \ldots, n-1$ **do**
  $r = \sqrt{a_{ii}^2 + a_{i+1,i}^2}$ -|- $n-1\ \surd,\ 2n-2\ *,\ n-1\ +$;
  $c = \frac{a_{ii}}{r}$ -|- $n-1\ *$;
  $s = \frac{a_{i+1,i}}{r}$ -|- $n-1\ *$;
  $a_{ii} = r$;
  $z_1 = b_i$;
  $z_2 = b_{i+1}$;
  $b_i = cz_1 + sz_2$ -|- $2n-2\ *,\ n-1\ +$;
  $b_{i+1} = -sz_1 + cz_2$ -|- $2n-2\ *,\ n-1\ +$;
  for $k = i+1, \ldots, n-1$ **do**
    $aik = a_{ik}$;
    $a_{ik} = ca_{ik} + sa_{i+1,k}$ -|- $2\sum_{i=1}^{n-1}(n-i-1)\ *,\ \sum_{i=1}^{n-1}(n-i-1)\ +$;
    $a_{i+1,k} = -s \cdot aik + ca_{i+1,k}$ -|- $2\sum_{i=1}^{n-1}(n-i-1)\ *,\ \sum_{i=1}^{n-1}(n-i-1)\ +$;
  end
end
Reši zgornje trikotni sistem $Rx = b$.

Vsota je enaka

$$6\sum_{i=1}^{n-1}(n-i-1) + 12(n-1) = 6\sum_{k=1}^{n-2} k + 12(n-1) = 6\frac{(n-2)(n-1)}{2} + 12(n-1) = 3n^2 + 3n - 6.$$

Matrika Q je zgornja Hessenbergova. ∎

Naloga 6.7 *Naj bo $A \in \mathbb{R}^{m \times n}$, $m \geq n$ in $rang(A) = n$. Podana je še matrika $C \in \mathbb{R}^{p \times n}$, $p \leq n$, $rang(C) = p$. Zapiši algoritem za reševanje predoločenega sistema $Ax = b$, pri pogoju $Cx = d$.*

Rešitev. Iščemo $\min_{Cx=d}||Ax-b||_2$, poglejmo si oblike matrik. Pomagajmo si z razširjenim

$Q \begin{array}{c} p \\ n-p \end{array} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$ razcepom matrike C^T. Rešujemo problem

$$\min_{Cx=d}||Ax-b||_2 = \min_{CQQ^Tx=d}||AQQ^Tx-b||_2.$$

Izračunajmo $CQ = (C^T)^TQ = (QR)^TQ = R^TQ^TQ = R^T = \begin{bmatrix} R_1^T & 0 \end{bmatrix}$, uvedemo še nove neznanke

$Q^Tx = \begin{array}{c} p \\ n-p \end{array} \begin{bmatrix} y \\ z \end{bmatrix}$ in $AQ = \begin{array}{c} \begin{array}{cc} p & n-p \end{array} \\ \begin{bmatrix} A_1 & A_2 \end{bmatrix} \end{array}$. Nadaljujemo z

$$\min_{\begin{bmatrix} R_1^T & 0 \end{bmatrix}\begin{bmatrix} y \\ z \end{bmatrix}=d} ||\begin{bmatrix} A_1 & A_2 \end{bmatrix}\begin{bmatrix} y \\ z \end{bmatrix} - b||_2 = \min_{R_1^Ty=d}||A_1y+A_2z-b||_2.$$

Z $R^Ty=d$ je y enolično določen. Torej nam

$$\min_{R_1^Ty=d}||A_2z-(b-\overbrace{A_1y}^{\text{znano}})||_2$$

predstavlja predoločeni sistem $A_2z = b - A_1y$.

$C^T = QR$;
$R_1^Ty = d$ (rešimo sistem, dobimo y);
$AQ = \begin{array}{c} \begin{array}{cc} p & n-p \end{array} \\ \begin{bmatrix} A_1 & A_2 \end{bmatrix} \end{array}$;
$\min_z ||A_2z-(b-A_1y)||_2$ (rešimo predoločen sistem, dobimo z) ;
$x = Q\begin{bmatrix} y \\ z \end{bmatrix}$;

■

Naloga 6.8 *Dana je funkcija* $f(x) = sin(\pi x)\exp(x/5)$. *Na intervalu* $[0,4]$ *ekvidistantno izberite točke,* $x_i = ih, i = 0,1,\ldots,16, h = \frac{1}{4}$ *in izračunajte vrednosti* $y_i = f(x_i)$. *Poiščite polinom* p *stopnje* 5, *ki zadošča*

$$p(1) = 0, \quad p(4) = 1$$

in aproksimira točke (x_i, y_i) *po metodi najmanjših kvadratov. Problem prevedite na iskanje*

$$min_{Cx=d}||Ax-b||_2,$$

in ga rešite z metodo opisano v prejšnji nalogi. Izpišite koeficiente polinoma ter narišite sliko, na kateri so označene točke, funkcija f *ter izračunan polinom* p.

Rešitev. Polinom zapišemo v obliki $a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^5 + a_5x^5$. Pogoji nam dajo sistem $Cx = d$

$$\begin{array}{ccccccccccccc} a_0 & + & a_1 & + & a_2 & + & a_3 & + & a_4 & + & a_5 & = 0 \\ a_0 & + & 4a_1 & + & 16a_2 & + & 64a_3 & + & 256a_4 & + & 1024a_5 & = 1, \end{array}$$

kjer je

$$C = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 4 & 16 & 64 & 256 & 1024 \end{bmatrix}, x = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & a_5 \end{bmatrix}^T \text{ in } d = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Zapišimo še predoločen sistem $Ax = b$. Matrika bo dimenzije 17×6, saj imamo sedemnajst interpolacijskih točk, njeni elementi so $a_{ij} = x_i^{j-1}$. Oglejte si demonstracijsko datoteko v Matlabu. ■

Naloga 6.9 *Dan imamo naslednji zvezni problem najmanjših kvadratov, kjer je* $p(x) = a_0 + a_1x + a_2x^2$ *in* $f(x) = e^x$,

$$\min_{p \text{ je parabola}} \int_{-1}^{1} (f(x) - p(x))^2 .$$

Polinom p je linearna kombinacija baznih polinomov $1, x, x^2, x^3$. *Tisti polinom p, ki predstavlja rešitev, je pravokotna projekcija funkcije f na prostor vseh poli- nomov stopnje 3, torej mora veljati*

$$\left\langle f(x) - p(x), x^k \right\rangle = 0 \text{ za } k = 0, 1, 2, 3,$$

kjer je skalarni produkt dveh funkcij definiran z

$$\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)dx.$$

Denimo, da na intervalu $[a, b]$ *iščemo najboljšo aproksimacijo po metodi najmanjših kvadratov funkcije f z linearno kombinacijo baznih funkcij* $p_1, \ldots, p_n$. *Potem je rešitev* $p(x) = \alpha_1 p_1(x) + \ldots + \alpha_n p_n(x)$, *kjer so* $\alpha_1, \ldots, \alpha_n$ *rešitve linearnega sistema*

$$\begin{bmatrix} \langle p_1, p_1 \rangle & \langle p_1, p_2 \rangle & \cdots & \langle p_1, p_n \rangle \\ \langle p_2, p_1 \rangle & \langle p_2, p_2 \rangle & \cdots & \langle p_2, p_n \rangle \\ \vdots & \vdots & & \vdots \\ \langle p_n, p_1 \rangle & \langle p_n, p_2 \rangle & \cdots & \langle p_n, p_n \rangle \end{bmatrix}$$

Matriko skalarnih produktov iz zgornjega sistema imenujemo ***Gramova matrika***. *Če so bazne funkcije ortogonalne, kar lahko npr. dosežemo z uporabo Gram-Schmidtove orto- gonalizacije, potem je Gramova matrika diagonalna.*

Rešitev. Oglejte si demonstracijsko datoteko v Matlabu. ■

Naloga 6.10 *Naj bo* $A \in \mathbb{R}^{m \times n}$, *kjer je* $m \geq n$ *in* $rang(A) = n$. *Dokaži, da ima sistem*

$$\begin{array}{cc} & \begin{array}{cc} m & n \end{array} \\ \begin{array}{c} m \\ n \end{array} & \begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} \end{array} \begin{bmatrix} r \\ x \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}$$

rešitev, ki ustreza rešitvi predoločenega sistema $Ax = b$ *po metodi najmanjših kvadratov.*

Rešitev. Poglejmo kaj mora veljati za rešitev sistema. Po blokih dobimo $r + Ax = b$ in $A^T r = 0$. Iz prve enačbe dobimo $r = b - Ax$, kar vstavimo v drugo enačbo. Iz tega sledi

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

Dobili smo ravno normalni sistem. ■

Naloga 6.11 *Dokaži, da za matriko X z najmanjšo normo $||X||_F$, ki minimizira $||XA - I_n||_F$ velja $X = A^+$. Matika A je dimenzije $m \times n$, kjer je $m \geq n$ in $rang(A) = r$.*

Rešitev. Naj bo podan singularni razcep matrike $A = U\Sigma V^T$ dimenzije $m \times n$ in ranga r. Tukaj sta U in V ortogonalni matriki dimenzij $m \times m$ in $n \times n$. Psevdoinverz matrike A je enak $A^+ = V\Sigma^+ U^T$, kjer je

$$\Sigma = \begin{array}{c c} & \begin{array}{cc} r & n-r \end{array} \\ \begin{array}{c} r \\ m-r \end{array} & \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \end{array}, \quad \Sigma^+ = \begin{array}{c c} & \begin{array}{cc} r & n-r \end{array} \\ \begin{array}{c} r \\ m-r \end{array} & \begin{bmatrix} S^{-1} & 0 \\ 0 & 0 \end{bmatrix} \end{array}$$

in $S = diag(\sigma_1, \sigma_2, \ldots, \sigma_r)$. Lotimo se reševanja. Najprej zapišemo X kot $X = VDU^T$, potem je $||X||_F = ||D||_F$ in velja

$$||XA - I||_F = ||VD\overbrace{U^T U}^{I}\Sigma V^T - I||_F \overset{\text{V ortogonalna}}{=} ||D\Sigma - I||_F.$$

Zadnji enačaj dobimo tako, da z leve pomnožimo z V^T, z desne pa z V. Matrika V je ortogonalna, tako se Frobeniusova norma ne spremeni. Matriko D predstavimo v bločni obliki

$$D = \begin{array}{c c} & \begin{array}{cc} r & m-r \end{array} \\ \begin{array}{c} r \\ n-r \end{array} & \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \end{array}.$$

Potem je $D\Sigma$ oblike

$$D\Sigma = \begin{array}{c c} & \begin{array}{cc} r & n-r \end{array} \\ \begin{array}{c} r \\ n-r \end{array} & \begin{bmatrix} D_{11}S & 0 \\ D_{21}S & 0 \end{bmatrix} \end{array}.$$

Če želimo, da bo $||D\Sigma - I||_F^2 = ||D_{11}S - I_r||_F^2 + ||D_{21}S||_F^2 + || - I_{n-r}||_F^2$ minimalna, moramo izbrati $D_{11} = S^{-1}$ in $D_{21} = 0$. Radi bi tudi, da je $||D||_F = ||X||_F$ čim manjša, torej je očitno najbolje izbrati $D_{12} = D_{22} = 0$. Velja namreč $||D||_F^2 = ||D_{11}||_F^2 + ||D_{21}||_F^2 + ||D_{12}||_F^2 + ||D_{22}||_F^2$. Dobili smo, da mora veljat $D = \Sigma^+$, kar nam da $X = V\Sigma^+ U^T = A^+$. ■

Če A ni polnega ranga, pravimo, da ima defekten rang. V primeru defektnega ranga rešitev, ki minimizira $||Ax - b||_2$ ni enolična, saj lahko x prištejemo poljuben $z \in \ker A$. V takšnem c primeru pravimo, da je izmed vseh vektorjev x, ki minimizirajo $Ax = b$, rešitev tisti, ki ima minimalno normo.

Naloga 6.12 *Če je $A \in \mathbb{R}^{m \times n}$, $m \geq n$, $rang(A) = r < n$ in $A = U\Sigma V^T$ singularni razcep, potem ima izmed vseh vektorjev x, ki minimizirajo normo $||Ax - b||_2$, najmanjšo normo $x = A^+ b$ oziroma*

$$x = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i$$

in

$$||Ax - b||_2^2 = \sum_{i=r+1}^{m} (u_i^T b)^2.$$

Rešitev. Za poljuben $x \in \mathbb{R}^n$ velja

$$||Ax - b||_2^2 = ||U^T AV(V^T x) - U^T b||_2^2 = ||\Sigma a - U^T b||_2^2 = \sum_{i=1}^{r} (\sigma_i a_i - u_i^T b)^2 + \sum_{i=r+1}^{m} (u_i^T b)^2,$$

kjer je $a = V^T x$. Minimum je očitno dosežen pri $a_i = \frac{u_i^T b}{\sigma_i}$ za $i = 1, \ldots, r$, komponente $a_{r+1}, \ldots, a_m$ pa so poljubne. Da dobimo minimalni $x = Va$ vzamemo $a_{r+1} = \cdots = a_m = 0$. ■