Aplikace matematiky

Václav Fabian

Odhad chyby zaokrouhlování při lineárních iteračních procesech, zejména při Seidelově řešení Dirichletova problému pro čtverec 10×10

Aplikace matematiky, Vol. 3 (1958), No. 1, 22–44

Persistent URL: http://dml.cz/dmlcz/102600

ODHAD CHYBY ZAOKROUHLOVÁNÍ PŘI LINEÁRNÍCH ITERAČNÍCH PROCESECH, ZEJMÉNA PŘI SEIDELOVĚ ŘEŠENÍ DIRICHLETOVA PROBLÉMU PRO ČTVEREC 10×10

VÁCLAV FABIAN

(Došlo dne 8. května 1957.) DT: 519.281

V článku se diskutuje použití statistických method k odhadu chyby při běžném a náhodném způsobu zaokrouhlování. Jako příklad aplikace obecných výsledků práce [4] se podávají některé numerické odhady velikosti chyby ve speciálním případě Seidelovy methody pro diferenční analogii Dirichletova problému pro čtverec 10×10.

1. Úvod

V práci [4] jsme studovali vliv zaokrouhlování na lineární iterační postupy v případě náhodného zaokrouhlování a zmínili jsme se o tom, že aplikace statistických method při běžném způsobu zaokrouhlování je pochybná. Nyní si této otázky všimneme podrobněji.

V druhé části tohoto článku aplikujeme obecné výsledky [4] na případ Seidelovy iterační methody pro řešení Dirichletovy úlohy pro čtverce 10×10; výsledky srovnáme s odhadem chyby zaokrouhlení pro Rietzovu methodu.

2. Aplikabilita statistických method k odhadu chyby zaokrouhlení při běžném způsobu zaokrouhlování

Buďte L_i transformace prostoru p-rozměrných vektorů do sebe, x_0 budiž p-rozměrný vektor,

$$x_i = L_i x_{i-1}\,; \tag{2.1}$$

při zaokrouhlování se dopouštíme na i-tém kroku zaokrouhlovací chyby ε_i, takže skutečně napočítané hodnoty vyhovují vztahům

$$\xi_i = L_i \xi_{i-1} + \varepsilon_i\,; \tag{2.2}$$

vektor $\delta_n = \xi_n - x_n$ nazýváme chybou zaokrouhlení.

Mluvíme o lineárním případě, je-li $L_i x = A_i x + y_i$ pro každý vektor x a každé přirozené číslo i, při čemž y_1, y_2, ... jsou p-rozměrné vektory a A_1, A_2, ... matice typu $p \times p$.

V celém článku se předpokládá, pokud není výslovně uvedeno jinak, že se zaokrouhluje na celá čísla. V lineárním případě to znamená v podstatě jen to omezení, že se v průběhu procesu (2.2) zaokrouhluje na stále stejný počet míst; i toto omezení lze však snadno odstranit.

Přístup k odhadu chyb δ_n je, pokud mi je známo, v zásadě dvojí. První t. zv. maximalistické odhady udávají takové dva vektory a a b, že platí $a \leqq \delta_n \leqq b$ (t. j., že tato nerovnost platí mezi odpovídajícími složkami), nebo takové číslo K, že platí $\|\delta_n\| \leqq K$ (kde $\| \|$ je nějaká norma vektorů). Bohužel tyto methody obvykle silně nadceňují skutečnou chybu, takže se na př. velmi málo hodí ke stanovení nejekonomičtějšího počtu míst, s nimiž se pracuje. Vzhledem k této nesnázi je snahou využít pro odhad chyb δ_n statistických method, při čemž statistickými rozumíme zde methody založené na předpokladu, že zaokrouhlovací chyby ε_i jsou náhodné vektory vzájemně nekorelované s nulovou očekávanou hodnotou. Výsledky statistických method bývají příjemné v tom smyslu, že dávají obvykle značně menší odhady chyb δ_n, než odhady maximalistické.

Z předpokladu nekorelovanosti a nulové očekávané hodnoty zaokrouhlovacích chyb vycházejí na příklad Abramov [1] v lineárním případě, Blanc, Liniger [2] a von Neumann, Goldstine [9] v případě nelineárním (formálně se často předpokládá více než nekorelovanost, předpokládá se nezávislost). Z těchto předpokladů se odvozují očekávaná hodnota a kovariační matice (nebo její odhad) chyby zaokrouhlení δ_n. Autoři obvykle poukazují na to, že v některých uměle konstruovaných příkladech nemusí ani předpoklady ani výsledky podaného matematického modelu odpovídat skutečnosti; tvrdí však, že v prakticky přicházejících úlohách k neshodám odvozených výsledků se skutečností nedochází.

Huskey [7] však uvádí příklad numerické integrace, který lze těžko považovat za uměle konstruovaný, v němž se počítá na mnoho míst (výpočty byly provedeny na elektronkovém počítači ENIAC). Huskey zjistil, že statistická methoda založená na obvyklých předpokladech silně podhodnocuje chybu. Vyslovuje závěr, že s jistotou lze použíti jen maximalistických a nikoliv statistických method.

Při konstrukci tabulek náhodných čísel nebo při tvorbě náhodných čísel pro Monte Carlo methody se ukazuje, že některé speciální iterační aritmetické postupy vedou k posloupnostem čísel, které se chovají v mnohém ohledu podobně, jako posloupnosti nezávislých pozorování nějaké náhodné proměnné. To se však týká speciálních postupů; naopak u jiných speciálních iteračních nebo jiných postupů — zdánlivě velmi rozumných — výsledná čísla nelze považovat za nezávislá pozorování jedné náhodné proměnné.

Předpokládejme na příklad, že jsme zvolili nějaké přirozené číslo k a že α_i je číslice na k-tém desetinném místě logaritmu o základě 10 čísla i. J. FRANEL ukázal v [6], že čísla α_i nelze považovat za nezávislá pozorování náhodné proměnné, neboť relativní četnost žádného z čísel 0, 1, ..., 9 mezi čísly $\alpha_1, \ldots, \alpha_N$ nekonverguje pro $N \to \infty$.

Poznamenáme, že mluví-li se v konkrétních případech o shodě matematického modelu, založeného na předpokladu nekorelovanosti a nulové očekávané hodnoty, se skutečností, míní se tím a ověřuje se shoda jen některých výsledků — obvykle odhadu kovariační matice chyby δ_n; i v těchto případech sám předpoklad nebo jiné, subtilnější jeho důsledky, shodu se skutečností sotva ukazují.

Jedním z takových důsledků v lineárním případě je na příklad vztah $P\left(\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \delta_n = \mathbf{0}\right) = 1$, který platí, existuje-li takové m, že $A_{m+i} = A_i$ a $(A_m \ldots A_1)^s \to \mathbf{0}$ (tento vztah byl odvozen autorem v práci [3] za předpokladu náhodného zaokrouhlování, plyne však, jak lze snadno zjistit nahlédnutím do důkazu, i z předpokladů nekorelovanosti, nulových očekávaných hodnot a omezenosti náhodných vektorů ε_i). Pokud se předpoklad nekorelovanosti nahradí silnějším předpokladem nezávislosti zaokrouhlovacích chyb, pak v prakticky všech netriviálních případech obsahuje matematický model sám v sobě logický spor.

Pro osvětlení uvedeme příklad, z něhož bude patrné, v jakých případech zejména lze očekávat nesouhlas modelu se skutečností. Pro jednoduchost uvedeme jednorozměrný případ; existence analogických případů vícerozměrných je zřejmá.

Má-li (2.1) tvar $x_i = 0{,}96x_{i-1}$, je-li $x_0 = 10$, je $x_n = (0{,}96)^n\, 10$ a $\xi_n = 10$, neboť $\xi_0 = 10$ a $0{,}96 \,.\, 10$ zaokrouhleno dává $\xi_1 = 10$, atd. Je $\varepsilon_i = 0{,}4$ pro každé i (což odporuje předpokladům nekorelovanosti a nulové očekávané hodnotě), $\delta_n = 10\,(1 - 0{,}96^n)$, $\delta_n \to 10$, což odporuje vztahu

$$P\left(\lim \frac{1}{N} \sum_{n=1}^{N} \delta = \mathbf{0}\right) = 1 \,.$$

Z předpokladu nekorelovanosti a nulové očekávané hodnoty bychom dále odvodili snadno (viz také [4], pozn. 3 odst. (5.9))

$$\mathbf{E}\delta_n^2 \leqq \frac{1}{4} \sum_{i=0}^{n-1} 0{,}96^{2i} \leqq \frac{1}{4} \frac{1}{1 - 0{,}96^2} = 3{,}188 \,.$$

Odtud je směrodatná odchylka $\sigma(\delta_n) \leqq 1{,}78571$, což opět není v právě dobré shodě se skutečnými hodnotami δ_n. Použijeme-li ostřejšího předpokladu, který činí na př. Abramov [1] a Blanc, Liniger [2], že totiž ε_i jsou stejnoměrně

rozloženy na intervalu $\langle -\frac{1}{2}, \frac{1}{2}\rangle$, dostáváme $\mathbf{D}\varepsilon_i = \int_{-\frac{1}{2}}^{\frac{1}{2}} x^2 \, dx = \frac{1}{12}$ a odtud $\mathbf{D}\delta_n \leq \frac{1}{12} \frac{1}{1 - 0{,}96^2} = 1{,}06292$, což tím více odporuje skutečnosti.

Shrneme-li tedy, pak odhady hodnot $\mathbf{E}\delta_n$ a $\mathbf{D}\delta_n$ odvozené z předpokladu nekorelovanosti a nulové očekávané hodnoty zaokrouhlovacích chyb ε_i v některých případech odpovídají skutečnosti a v jiných nikoliv. Některé jiné důsledky těchto předpokladů obvykle skutečnosti neodpovídají.

Příklad, který jsme uvedli, nám přes svou jednoduchost ukazuje, že nepříjemnosti spočívají nikoliv v odhadu chyby zaokrouhlování, avšak ve věci samé, tedy v zaokrouhlování běžným způsobem. To vedlo v roce 1950 G. E. Forsytha [5] k navržení jiného způsobu zaokrouhlování, jemuž věnujeme další odstavec. Toto t. zv. náhodné zaokrouhlování nemá nevýhod běžného způsobu zaokrouhlování, o nichž jsme právě mluvili, a umožňuje další methody odhadu chyb δ_n, které jsou při běžném způsobu zaokrouhlování nemožné.

3. Náhodné zaokrouhlování

Z různých způsobů náhodného zaokrouhlování ([4]) je nejdůležitější ten, při němž se číslo a z intervalu $\langle 0, 1)$ zaokrouhlí s pravděpodobností a nahoru na číslo 1, s pravděpodobností $1 - a$ dolů na 0; číslo $b = b_1 + b_2$, kde b_1 je celé a b_2 je z intervalu $\langle 0, 1)$, se zaokrouhlí zaokrouhlením čísla b_2 a přičtením b_1. K provedení náhodného zaokrouhlování je třeba a stačí provést experiment, jehož dva možné výsledky mají pravděpodobnost a a $1 - a$. To je skutečně možné, aspoň je-li a racionální číslo. Při výpočtech na kalkulačních strojích bude vhodné použít pro zaokrouhlování tabulek náhodných čísel [10], [11], [12]. Jak tyto tabulky používat, je vyloženo na př. v knize [8], str. 192. Při některých výpočtech, kdy zaokrouhlovaná čísla mohou nabývat jen určitých hodnot (na př. při řešení Dirichletova problému methodou sítí zaokrouhlujeme čísla tvaru $a + \frac{1}{4}$, $a + \frac{1}{2}$, $a + \frac{3}{4}$, kde a je celé), lze provádět náhodné zaokrouhlování skoro stejně rychle, jako běžné nenáhodné zaokrouhlování. Při výpočtech na velkých samočinných počítacích strojích (kde problém zaokrouhlování neztrácí svůj význam) by používání tabulek náhodných čísel bylo nevhodné.

Jak dodávat s dostatečnou rychlostí automatickému počítači náhodná čísla, to je otázka, která se vyskytuje a řeší v souvislosti s Monte Carlo methodami. Je možné použít zdroje náhodných impulsů založeného na nějakém fysikálním ději (na př. na radioaktivním záření; takový zdroj zkonstruoval Zdeněk Koutský v matematickém oddělení Ústavu radiotechniky a elektroniky ČSAV), nebo je možné použít náhodných čísel vytvořených nějakou matematickou operací, jejíž vhodnost byla ověřena. K poslednímu způsobu má autor značné

výhrady, avšak zatím tento způsob při aplikacích Monte Carlo method převládá.

Zaokrouhlujeme-li vektor a, zaokrouhlíme každou složku buď na základě nezávislých experimentů (v tomto případě budeme mluvit o P° — zaokrouhlování), nebo tak, že korelace mezi zaokrouhlenými složkami je záporná (budeme mluvit o P^--zaokrouhlování).

Příkladem P^--zaokrouhlení je postup, při němž vektor $\begin{pmatrix}\frac{1}{2}\\ \frac{1}{2}\end{pmatrix}$ zaokrouhlíme se stejnou pravděpodobností $\frac{1}{2}$ buď na $\begin{pmatrix}0\\1\end{pmatrix}$ nebo na $\begin{pmatrix}1\\0\end{pmatrix}$. Při P°-zaokrouhlení bychom tento vektor zaokrouhlili se stejnou pravděpodobností $\frac{1}{4}$ buď na $\begin{pmatrix}0\\0\end{pmatrix}$ nebo $\begin{pmatrix}1\\0\end{pmatrix}$ nebo $\begin{pmatrix}0\\1\end{pmatrix}$ nebo $\begin{pmatrix}1\\1\end{pmatrix}$.

Zbývá říci, jak zaokrouhlíme náhodně proces (2.1), t. j. jakým způsobem provádíme pozorování náhodných vektorů ξ_n ve vztahu (2.2). Položíme $\xi_0 = x_0$, spočteme vektor $L_1\xi_0$ a na základě náhodného experimentu $\mathscr{E}_1$ zaokrouhlíme $L_1\xi_0$; výsledek bude pozorování náhodného vektoru ξ_1, které označíme $\xi_1(\omega)$ (je třeba rozlišovat mezi náhodným vektorem a mezi jeho hodnotou, kterou napozorujeme po provedení příslušného experimentu, analogicky, jako je třeba rozlišovat mezi funkcí f a její hodnotou $f(a)$ v bodě a). Předpokládejme, že jsme již napozorovali n hodnot $\xi_0(\omega)$, $\xi_1(\omega)$, ..., $\xi_{n-1}(\omega)$ náhodných vektorů ξ_0, ξ_1, ..., ξ_{n-1}, při čemž jsme k zaokrouhlování použili náhodných experimentů $\mathscr{E}_1, \mathscr{E}_2, \ldots, \mathscr{E}_{n-1}$. Vypočteme vektor $L_n\xi_{n-1}(\omega)$ a zaokrouhlíme jej podle výsledku náhodného experimentu $\mathscr{E}_n$, *nezávislého* na experimentech $\mathscr{E}_1, \ldots, \mathscr{E}_{n-1}$: obdržená hodnota $\xi_n(\omega)$ je pozorováním náhodného vektoru ξ_n.

Při opakování celého výpočtu (vlastně experimentu) je třeba rozlišit dva případy. Jednak je možno použít při zaokrouhlování v obou případech stejných výsledků stejných experimentů $\mathscr{E}_1, \mathscr{E}_2, \ldots$ V tomto případě i napočtené hodnoty budou stejné. Za druhé můžeme při opakování výpočtu použít experimentů $\mathscr{F}_1, \mathscr{F}_2, \ldots$, nezávislých na experimentech $\mathscr{E}_1, \mathscr{E}_2, \ldots$, použitých při prvém výpočtu. V tomto případě dostáváme obecně různé hodnoty $\xi_n(\omega_1)$ a $\xi_n(\omega_2)$, které jsou nezávislými pozorováními náhodného vektoru ξ_n.

Používáme-li průběhem celého výpočtu P°-zaokrouhlování, řekneme, že proces ξ_n vznikl P°-zaokrouhlováním; analogicky pro P^--zaokrouhlování.

Příklad. Pro ilustraci a srovnání uvažujme opět jednoduchý případ $x_i = 0{,}96x_{i-1}$, $x_0 = 10$. Vektory jsou jednorozměrné, a tedy oba způsoby zaokrouhlování P° i P^- jsou totožné. Zaokrouhlovat bude třeba čísla tvaru $\frac{i}{100}$.

Použijeme tabulek náhodných čísel. Čteme dvojice číslic $\alpha\beta$; zaokrouhlíme dolů, je-li $\alpha\beta \geq i$, nahoru, je-li $\alpha\beta < i$. Výňatek z tabulek [10] je:

77603	59467	58309
73258	73452	17619
03469	27635	56293
29718	86040	02596
45938	52403	94255
37510	14663	46581
23439	53633	94365
67317	43514	34261

Umluvíme se, že postupně čteme zleva doprava dva řádky takto: 77, 73, ..., 99 (konec prvého řádku); 02, 39, ... Dostáváme tedy postupně tyto hodnoty $\xi_n(\omega)$:

$$\xi_0 = 10;\ 0{,}96 \,.\, 10 = 9{,}60\,,\quad 77 \geqq 60 \Rightarrow \xi_1(\omega) = 9\,;$$
$$0{,}96 \,.\, 9 = 8{,}64\,,\quad 73 \geqq 64 \Rightarrow \xi_2(\omega) = 8\,;\quad 0{,}96 \,.\, 8 = 7{,}68\,,$$
$$62 < 68 \Rightarrow \xi_3(\omega) = 8\,;\quad 0{,}96 \,.\, 8 = 7{,}68\,,\quad 05 < 68 \Rightarrow \xi_4(\omega) = 8$$

atd.; celá posloupnost je uvedena v tabulce 3.1.

Tab. 3.1.

n	$\xi_n(\omega)$	n	$\xi_n(\omega)$	n	$\xi_n(\omega)$	n	$\xi_n(\omega)$
0	10						
1	9	16	5	31	3	46	1
2	8	17	5	32	3	47	1
3	8	18	5	33	2	48	1
4	8	19	5	34	2	49	1
5	8	20	4	35	2	50	0
6	8	21	4	36	2	51	0
7	7	22	4	37	2	52	0
8	7	23	4	38	2	53	0
9	7	24	4	39	2	54	0
10	7	25	4	40	2	55	0
11	7	26	4	41	1	56	0
12	6	27	4	42	1	57	0
13	6	28	4	43	1	58	0
14	6	29	3	44	1	59	0
15	5	30	3	45	1	60	0

Podstatnou vlastností náhodného zaokrouhlování je, že vznikl-li proces ξ_n náhodným zaokrouhlováním, jsou zaokrouhlovací chyby ε_i nakorelované náhodné vektory s nulovou očekávanou hodnotou (viz [4]). To umožňuje přesnou aplikaci statistických method ke studiu vlivu chyby vznikající náhodným zaokrouhlováním. Kromě této formální výhody má proces ξ_n, vzniklý náhodným zaokrouhlováním, mnoho výhodných vlastností od možnosti nezávislých pozorování ξ_n až po vlastnosti konvergenční dokázané v [3], [4], které zřejmě při obvyklém způsobu zaokrouhlování splněny nejsou a které byly demonstrovány v příkladu tohoto odstavce.

4. Odhad chyby zaokrouhlování pro diferenční analogii Dirichletova problému pro čtverec 10 × 10

Při náhodném zaokrouhlování v lineárním případě lze odhadovat chybu zaokrouhlení δ_n různým způsobem ([4]): nezávislým (aspoň dvojnásobným) pozorováním náhodné proměnné ξ_n (tak dostáváme současně intervalové odhady pro x_n), odhadem kovariační matice $\mathbf{D}\delta_n$ a konečně napozorováním náhodné proměnné Δ_n majorisující δ_n ve smyslu $\mathbf{D}\Delta_n \succsim k\mathbf{D}\delta_n$, kde $k \neq 0$ a $A \succsim B$ značí, že matice $B - A$ je semidefinitně positivní. (Jestliže $A_i \geqq \mathbf{0}$ pro $i = 1, \ldots, n$, lze relaci $\succsim$ nahradit relací $\geqq$.)

Je-li (2.1) iterační proces pro řešení systému lineárních rovnic, není cílem výpočtu stanovení vektoru x_n, ale odhad řešení $x = \lim_{n\to\infty} x_n$. Je-li a libovolný vektor, lze odhadnout vektor $a - x$ z residua $a - Aa - y$, které je ovšem nezávislé na tom, jak jsme k vektoru a dospěli. Není tedy třeba znát k odhadu přesnosti přiblížného řešení ξ_n chybu zaokrouhlení δ_n, i když není vyloučeno, že by se daly najít methody, využívající znalosti struktury δ_n k přesnějším odhadům vektoru $\xi_n - x$, kde x je řešení rovnice. To by se zdálo být slibné zvlášť v těch případech, kdy δ_n je velké v porovnání s $x - x_n$. V případě Dirichletova problému se nám nepodařilo vhodnou methodu tohoto druhu konstruovat, což je pochopitelné vzhledem k malým, jak ukážeme, hodnotám δ_n.

Zdá se tedy, že při řešení systému rovnic iterační methodou je třeba odhadu chyby zaokrouhlování především pro možnost ekonomické volby počtu míst, na něž se má zaokrouhlovat. Zde přichází v úvahu především použití majorisující posloupnosti Δ_n, která umožňuje odhad chyby před zahájením výpočtu, nebo odhad matice $\mathbf{D}\delta_n$, pro niž při P^- nebo P°-zaokrouhlování (na celá čísla) platí ([4])

$$\mathbf{D}\delta_m \prec \frac{1}{4} \sum_{i=1}^{n} A_n A_{n-1} \ldots A_{i+1} M_i A'_{i+1} \ldots A'_{n-1} A'_n , \tag{4.1}$$

kde pro Rietzovu methodu $M_i = \mathbf{1}$. Pro Seidelovu methodu je element $M_i^{(ii)}$ roven 1 a ostatní elementy matice M_i jsou rovny 0 (pro $i = 1, \ldots, p$). Dále klademe $M_{i+sp} = M_i$ pro $i = 1, \ldots, p$ a $s = 1, 2, \ldots$ Je-li splněna podmínka $A_i \geqq \mathbf{0}$ pro každé $i = 1, \ldots, n$, můžeme psát ve vztahu (4.1) místo $\prec \leqq$.

Tento druhý způsob je ovšem podstatně komplikovanější, alespoň pokud se nám nepodaří zjednodušit matici D_n stojící v pravé straně vztahu (4.1). Oba způsoby dávají odhady chyb δ_n nezávislé na pravých stranách y_i a na počátečním vektoru x_0. Je možný ještě třetí způsob, spočívající v souběžném dvojnásobném nezávislém opakování výpočtu posloupnosti $\xi_1, \xi_2, \ldots$, který odhaduje chybu až při průběhu výpočtu a je závislý na pravých stranách y_i i na počátečním vektoru x_0 (tím by ovšem mohl dát někdy ještě lepší odhad).

V této práci provedeme odhad chyby δ_n pro speciální případ diferenční analogie Dirichletova problému pro čtverec 10 × 10, libovolnou pravou stranu

(t. j. libovolné okrajové podmínky) a libovolný počáteční vektor x_0; použijeme při tom prvních dvou z naznačených method; třetí methody nepoužijeme, neboť nemůže dát odhad nezávislý na x_0 a y_i[1]). Při odhadu chyby δ_n předpokládáme, že proces je náhodně zaokrouhlován způsobem P°.

V dalším budeme uvažovat 100-rozměrné vektory, které je výhodné zapisovat do čtverce 10×10.

Každé uspořádané dvojici $[\alpha\beta]$ ($\alpha = 1, \ldots, 10$; $\beta = 1, \ldots, 10$) přiřaďme číslo $\overline{\alpha\beta}$ ($\overline{\alpha\beta} = 1, \ldots, 100$) podle tohoto schematu:

$[\alpha\beta]$				$\overline{\alpha\beta}$			
11	12	. . .	1; 10	1	2	. . .	10
21	22	. . .	2; 10	11	12	. . .	20
.	.		.	.	.		.
.	.		.	.	.		.
.	.		.	.	.		.
10; 1	10; 2	. . .	10; 10	91	92	. . .	100

Je tedy $\alpha\beta = \overline{10}(\alpha - 1) + \beta$.

Systém rovnic, který uvažujeme, je dán vztahy

$$x^{(\overline{\alpha\beta})} = \frac{1}{4}\left[x^{(\overline{\alpha-1,\beta})} + x^{(\overline{\alpha+1,\beta})} + x^{(\overline{\alpha,\beta-1})} + x^{(\overline{\alpha,\beta+1})}\right] \tag{4.2}$$

pro $\alpha = 1, \ldots, 10$; $\beta = 1, \ldots, 10$.

Přitom $x^{(\overline{\alpha\beta})}$ pro $\alpha = 1, \ldots, 10$; $\beta = 1, \ldots, 10$ jsou složky hledaného vektoru x a $x^{(\overline{\alpha\beta})}$ pro ostatní α, β jsou daná čísla (pravé strany rovnic, resp. okrajové podmínky).

Systém rovnic (4.2) můžeme přepsat ve vektorovou rovnici

$$x = Ax + y\,, \tag{4.3}$$

kde A je matice soustavy (typu 100×100) a y je pravá strana; vypisovat explicite matici A a vektor y je pro naše účely zbytečné.

K řešení rovnice (4.3) je možno použít jak Rietzovy tak i Seidelovy iterační methody. Rietzově methodě odpovídá posloupnost x_n (ξ_n při zaokrouhlování) definovaná vztahy (2.1) a (2.2), kde je $L_i a = Aa + y$ pro každý vektor a. Matice A je nezáporná symetrická, $(\mathbf{1} - A)^{-1}$ existuje, $M_i = \mathbf{1}$ a lze tedy přepsat (4.1) v odhad

$$\mathbf{D}\delta_n \leqq \frac{1}{4}\,(\mathbf{1} - A^{2n})\,(\mathbf{1} - A^2)^{-1} \leqq \frac{1}{4}\,(\mathbf{1} - A^2)\,, \tag{4.4}$$

což je výraz, ke kterému dospěl Abramov [1], který ovšem ze svých předpo-

[1]) Třetí methody však použijeme v odstavci 6.

kladů odvodil o něco ostřejší závěr; Abramov též ukázal, jak lze poměrně jednoduše spočítat nebo odhadnout prvky matice $(\mathbf{1} - A^2)(\mathbf{1} - A^{2^n})$.

Při Seidelově methodě každému kroku Rietzovy methody odpovídá tolik částečných operací, jakého rozměru jsou vektory, v našem případě tedy 100. Zhruba lze Seidelovu methodu charakterisovat tím, že používáme stále posledně napočítaných souřadnic. Přesněji: Definujeme pro $i = 1, \ldots, 100$ matici A_i vztahem[2])

$$A_i^{(\alpha\beta)} = \begin{cases} A^{(i\beta)} & \text{pro } \alpha = i\,, \\ 1 & \text{pro } \alpha \neq i\,,\ \alpha = \beta\,. \\ 0 & \text{pro } \alpha \neq i,\ \alpha \neq \beta \end{cases}$$

Vektor y_i definujeme vztahy $y_i^{(\alpha)} = 0$ pro $\alpha \neq i$ a $y_i^{(\alpha)} = y^{(\alpha)}$ pro $\alpha = i$.

Pro každý vektor a má vektor $A_i a + y_i$ všechny složky až na i-tou shodné s vektorem a; i-tou složku má shodnou s vektorem $Aa + y$. Definujeme dále $A_{100s+i} = A_i$, $y_{100s+i} = y_i$ pro každé přirozené s a každé $i = 1, \ldots, 100$. Je-li $L_i a = A_i a + y$, odpovídají (2.1) a (2.2) iteračnímu procesu Seidelovu bez a se zaokrouhlováním. Výraz (4.1) by bylo sice možno poněkud zjednodušit vzhledem ke vztahu $A_{100+i} = A_i$, avšak toto zjednodušení je málo užitečné a nevede, pokud je mi známo, k tak jednoduchým výrazům, k jakým je možno dospět při methodě Rietzově.

Protože předpokládáme P°-zaokrouhlování, všimneme si blíže, jak technicky lze toto zaokrouhlování provádět při Seidelově methodě. Na každém kroku se bude zaokrouhlovat vždy jen jedna složka vektoru (ostatní nebudou změněny), a tedy oba způsoby P° a P^- jsou shodné. Protože čísla, která budeme zaokrouhlovat, budou tvaru a, $a + \frac{1}{4}$, $a + \frac{1}{2}$, $a + \frac{3}{4}$ (a je celé), stačí na příklad číst dvojice číslic cd z tabulek náhodných čísel a zaokrouhlovat takto:

	nahoru	dolů
$\frac{1}{4}$	je-li c i d liché	je-li aspoň jedno z čísel c i d sudé
$\frac{1}{2}$	je-li c liché	je-li c sudé
$\frac{3}{4}$	je-li aspoň jedno z čísel c i d liché	je-li c i d sudé

Přitom se stále používá nových dvojic c, d, což zaručuje nezávislost zaokrouhlování na n-tém kroku na předchozích zaokrouhlováních.

[2]) $A^{(\alpha\beta)}$ je element matice A v řádku α a sloupci β; podobného označení používáme i pro vektory.

5. Dirichletův problém. Majorisující posloupnost Δ_n

Příjemnou vlastností odhadu (4.1) je jeho nezávislost na $x_0, y_1, \ldots, y_n$. Ta upozorňuje na možnost, odhadnout matici

$$D_n = \frac{1}{4} \sum_{i=1}^{n} A_n \ldots A_{i+1} M_i A'_{i+1} \ldots A'_n ,$$

výpočtem procesu ξ_n odpovídajícímu zvláštní volbě vektorů y_0, y_i tak, aby byl výpočet podstatně usnadněn a abychom znali vektory x_n. Potřebovali bychom k tomu ovšem opačnou nerovnost než jaká je v (4.1). Takový postup je skutečně možný, je popsán větou (5.10) práce [4] a zde udáme konkretní návod pozorování Δ_n, splňujících podmínku

$$\frac{4}{3} \mathbf{D}\Delta_n \geqq D_n, \text{ a tedy } \frac{4}{3} \mathbf{D}\Delta_n \geqq \mathbf{D}\delta_n .$$

Pozorování náhodných vektorů Δ_n získáme náhodným zaokrouhlováním procesu $x_i = A_i x_{i-1}$ (odpovídajícímu tedy pravé straně $y = \mathbf{0}$) a s počátečním vektorem $x_0 = \mathbf{0}$. Zaokrouhlování neprovedeme ovšem způsobem P° (dostali bychom $\Delta_n = \mathbf{0}$ pro všechna n), ale takto: Číslo b tvaru $a + \frac{i}{4}$ ($i = 1, 2, 3$), kde a je celé, zaokrouhlíme jako při způsobu P° — tedy podle tabulky uvedené ke konci předchozího odstavce. Celé číslo a však také zaokrouhlíme: s pravděpodobností $\frac{1}{8}$ dolů na $a - 1$, s pravděpodobností $\frac{1}{8}$ nahoru na $a + 1$ a s pravděpodobností $\frac{6}{8} = \frac{3}{4}$ zaokrouhlujeme na číslo a.

V každém případě při popsaném způsobu zaokrouhlování bude očekávaná hodnota výsledku zaokrouhlovaného čísla a rovna 0 a rozptyl D_a bude větší nebo roven $1\frac{3}{6}$ vzhledem k hodnotám D_a udaným v tabulce:

a	0	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{3}{4}$
D_a	$\frac{1}{4}$	$1\frac{3}{6}$	$\frac{1}{4}$	$1\frac{3}{6}$

(5.1)

Technicky provedeme zaokrouhlení celého čísla a takto: Z tabulek náhodných čísel přečteme tři dosud v procesu nepoužité číslice c, d, e a zaokrouhlíme na $a + 1$, jsou-li všechna tři čísla lichá, zaokrouhlíme na $a - 1$, jsou-li všechna tři čísla sudá; zaokrouhlíme na a v ostatních případech.

Pozorování náhodných vektorů Δ_n nyní provádíme analogicky, jako by probíhalo pozorování náhodných vektorů ξ_n; použijeme ovšem právě popsaného způsobu náhodného zaokrouhlování. Položíme $\Delta_0 = \mathbf{0}$. Jestliže jsme již napozorovali hodnotu $\Delta_{i-1}(\omega)$, vypočteme vektor $d_i = A_i \Delta_{i-1}(\omega)$, lišící se od $\Delta_{i-1}(\omega)$ pouze svou i-tou složkou. Položíme $\Delta_i^{(\alpha)}(\omega) = d_i^{(\alpha)}(\omega) = [\Delta_{i-1}(\omega)]^{(\alpha)}$ pro $\alpha \neq i$, avšak i-tou složku (která může, ale nemusí být celé číslo) zaokrouhlíme (nezávisle na zaokrouhleních, která jsme prováděli v předchozím průběhu

procesu); výsledek zaokrouhlení bude zbývající i-tá složka vektoru $\Delta_i(\omega)$, hodnoty náhodného vektoru Δ_i.

Z věty (5.10) práce [4] a ze vztahu $D_a \geqq \frac{3}{16}$ plyne $D_n \leqq \frac{4}{3}\mathbf{D}\Delta_n$.

Popsaný pokus byl proveden nezávisle dvakrát na sobě a byla tak získána dvě nezávislá pozorování $\{\Delta_n(\omega_i)\}_{n=0}^{5000}$ $(i = 1, 2)$ posloupnosti náhodných vektorů $\{\Delta_n\}_{n=1}^{5000}$. Přesto, že nás zajímá jen vybraná posloupnost Δ_{100n}, neuvedeme všechny napozorované hodnoty, neboť se jedná o $2 \times 50 \times 100 = 10\,000$ čísel; uvedeme však některé charakteristiky, které snad podají dosti výstižný přehled o výsledcích.

Především uvedeme hodnoty $\|\Delta_n(\omega)\|^2 = \sum_{\alpha=1}^{100} (\Delta_n^{(\alpha)}(\omega))^2$. Očekávaná hodnota náhodné proměnné $\|\Delta_n\|^2$ je rovna číslu st $\mathbf{D}\Delta_n$, a tedy

$$\frac{4}{3}\mathbf{E}\,\|\Delta_n\|^2 \geqq \mathrm{st}\,\mathbf{D}\delta_n = \sum_{\alpha=1}^{100} \mathbf{E}[\delta_n^{(\alpha)}]^2 .$$

Tabulka 5.2

n	$\|\Delta_{100n}(\omega_1)\|^2$	$\|\Delta_{100n}(\omega_2)\|^2$	n	$\|\Delta_{100n}(\omega_1)\|^2$	$\|\Delta_{100n}(\omega_2)\|^2$
1	20	19	26	36	32
2	23	24	27	18	37
3	30	29	28	24	32
4	22	27	29	27	26
5	42	22	30	37	24
6	37	27	31	36	27
7	31	32	32	24	29
8	28	24	33	31	30
9	27	33	34	22	29
10	28	35	35	32	40
11	27	27	36	32	25
12	23	31	37	18	26
13	27	35	38	30	40
14	26	36	39	35	34
15	25	39	40	29	32
16	29	39	41	29	26
17	37	40	42	31	25
18	25	40	43	32	33
19	38	35	44	37	41
20	24	29	45	38	28
21	23	40	46	34	24
22	30	43	47	40	28
23	26	35	48	34	35
24	34	29	49	31	36
25	36	27	50	35	31

Složky vektorů $\Delta_{100n}(\omega_i)$ jsou vesměs 0, 1, -1, 2, -2; poslední dva případy (2, -2) však nastaly jen ve čtyřech případech:

$$(\omega_1, 6, \overline{58}), \quad (\omega_1, 47, \overline{48}), \quad (\omega_2, 35, \overline{46}), \quad (\omega_2, 44, \overline{63}),$$

kde symbol (ω_i, n, α) značí případ $|\Delta_{100n}^{(\alpha)}(\omega_i)| = 2$.

Pro umožnění představy o tom, jak se liší obě nezávislá pozorování, uvedeme v tab. 5.3 hodnoty $\Delta^{(\overline{66})}_{100n}(\omega_i)$, $\Delta^{(\overline{77})}_{100n}(\omega_i)$, $\Delta^{(\overline{88})}_{100n}(\omega_1)$ pro $i = 1, 2$ a $n = 1, 2, \ldots, 50$.

Tabulka 5.3

$\Delta^{(\overline{66})}_{100(5\mu+\nu)}(\omega_1)$

ν \ μ	0	1	2	3	4	5	6	7	8	9
1	1	1	0	0	0	0	0	−1	1	0
2	1	0	0	0	−1	0	0	0	0	0
3	0	0	1	0	1	0	0	1	0	1
4	0	0	0	0	1	−1	0	1	0	−1
5	1	0	0	0	0	0	0	0	0	1

$\Delta^{(\overline{77})}_{100(5\mu+\nu)}(\omega_1)$

ν \ μ	0	1	2	3	4	5	6	7	8	9
1	1	1	0	0	0	1	0	1	0	0
2	0	0	0	0	0	0	0	1	0	1
3	1	0	0	0	1	1	1	1	0	1
4	0	0	0	0	0	0	0	0	0	1
5	0	0	0	0	−1	1	0	−1	0	1

$\Delta^{(\overline{88})}_{100(5\mu+\nu)}(\omega_1)$

ν \ μ	0	1	2	3	4	5	6	7	8	9
1	0	0	0	0	0	1	1	1	0	0
2	0	0	0	−1	0	0	0	0	0	0
3	0	0	0	−1	0	0	0	−1	0	0
4	0	0	0	0	1	0	0	−1	0	1
5	0	0	0	0	0	1	1	0	0	1

Tabulka 5.3 (pokračování)

$A^{(\bar{6}\bar{6})}_{100(5\mu+\nu)}(\omega_2)$

ν \ μ	0	1	2	3	4	5	6	7	8	9
1	0	0	0	−1	0	0	1	1	0	0
2	0	1	0	0	0	−1	1	1	0	1
3	0	1	0	0	0	0	0	1	0	0
4	0	0	−1	0	0	−1	1	1	0	0
5	0	0	−1	0	1	0	1	1	0	1

$A^{(\bar{7}\bar{7})}_{100(5\mu+\nu)}(\omega_2)$

ν \ μ	0	1	2	3	4	5	6	7	8	9
1	0	1	0	−1	0	−1	1	1	0	0
2	0	0	0	−1	−1	0	1	1	1	0
3	0	0	0	−1	−1	0	0	0	1	0
4	0	0	0	−1	0	1	1	0	0	0
5	−1	0	0	0	−1	0	1	1	1	1

$A^{(\bar{8}\bar{8})}_{100(5\mu+\nu)}(\omega_2)$

ν \ μ	0	1	2	3	4	5	6	7	8	9
1	0	0	0	0	−1	0	0	0	1	0
2	0	0	0	−1	−1	0	0	0	0	1
3	0	0	0	−1	−1	1	0	0	1	0
4	0	0	0	−1	0	1	−1	0	1	0
5	1	0	−1	−1	0	1	0	0	1	0

Konečně uvedeme v tabulce 5.4 vektory $\Delta_{100}(\omega_i)$, $\Delta_{1000}(\omega_i)$ a $\Delta_{5000}(\omega_i)$ pro $i = 1, 2$.

Tabulka 5.4

$\Delta_{100}^{(\overline{\alpha\beta})}(\omega_1)$

α \ β	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	1	0	−1	0	0
3	0	0	1	0	0	0	0	0	0	−1
4	0	0	0	0	0	0	0	1	0	0
5	0	0	0	0	0	0	0	0	0	0
6	−1	0	0	1	0	1	1	0	−1	0
7	−1	−1	0	1	0	1	1	0	−1	0
8	−1	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	−1	0	0	0	0
10	0	0	0	1	0	−1	0	0	0	0

$\Delta_{1000}^{(\overline{\alpha\beta})}(\omega_1)$

α \ β	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	1	0	0	0	0
2	0	0	0	1	0	1	0	−1	−1	0
3	0	0	1	1	1	0	0	−1	−1	0
4	0	0	0	1	1	0	0	0	0	0
5	0	0	1	0	0	0	0	0	1	1
6	0	0	1	1	0	0	0	0	0	0
7	1	0	0	0	0	0	0	0	0	0
8	0	0	1	0	1	1	0	0	0	0
9	0	0	1	0	1	1	1	0	0	1
10	0	0	0	1	1	0	0	0	0	0

Tabulka 5.4 (pokračování)

$\Delta_{100}^{(\overline{\alpha\beta})}(\omega_2)$

α \ β	1	2	3	4	5	6	7	8	9	10
1	0	0	−1	0	0	0	0	−1	0	0
2	0	0	−1	0	1	0	1	0	0	0
3	−1	0	−1	0	0	1	0	0	0	0
4	0	0	0	0	0	1	0	0	0	0
5	0	0	−1	0	0	0	0	0	0	0
6	0	0	0	−1	0	0	0	0	1	0
7	0	0	0	0	0	1	0	0	0	0
8	0	−1	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	−1	−1	0	0
10	0	0	1	0	0	0	−1	0	0	1

$\Delta_{1000}^{(\overline{\alpha\beta})}(\omega_2)$

α \ β	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	1	0	0	−1	0
2	0	0	0	0	0	0	−1	0	−1	−1
3	0	0	1	0	1	0	−1	0	−1	−1
4	0	0	0	1	0	−1	0	−1	−1	−1
5	0	−1	−1	0	1	0	0	0	0	0
6	0	−1	0	0	0	0	0	1	0	0
7	0	−1	−1	−1	0	0	0	0	1	1
8	0	−1	−1	−1	0	−1	0	0	0	0
9	0	−1	−1	0	−1	0	0	0	−1	0
10	1	−1	0	0	0	0	0	0	0	0

Tabulka 5.4 (pokračování)

$\Delta_{5000}^{(\overline{\alpha\beta})}(\omega_1)$

α \ β	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	−1	0	0	1	0	1
3	0	0	0	−1	−1	1	0	0	0	0
4	0	0	−1	0	−1	0	1	0	1	0
5	0	0	0	0	0	0	1	1	1	1
6	0	0	0	1	0	1	1	1	1	0
7	−1	1	0	1	0	0	1	1	0	0
8	0	0	0	0	0	1	0	1	0	0
9	0	0	−1	−1	0	1	1	0	1	0
10	0	−1	0	0	0	1	0	0	1	1

$\Delta_{5000}^{(\overline{\alpha\beta})}(\omega_2)$

α \ β	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	1	−1	0	0	0	0
2	−1	0	0	0	0	−1	0	0	0	0
3	−1	0	0	1	0	0	−1	0	−1	0
4	0	0	0	1	0	0	0	0	0	0
5	−1	−1	0	1	1	1	1	0	−1	0
6	−1	−1	0	0	0	1	1	0	0	0
7	0	0	0	1	1	0	1	0	−1	0
8	0	0	0	1	1	0	0	0	0	0
9	−1	0	0	0	0	0	0	−1	−1	0
10	0	−1	1	0	0	0	0	0	0	0

Z uvedených výsledků a ze vztahu $\frac{4}{3}\mathbf{D}\Delta_n \succcurlyeq \mathbf{D}\delta_n$ se přirozeně tvoří představa, že rozptyly jednotlivých složek chyb $\delta_n^{(\alpha)}$ se od sebe liší poměrně velmi

málo, že také $\mathbf{D}\delta_{100n}$ s výjimkou několika málo (5 až 10) prvých indexů n, s rostoucím n již dále v podstatě neroste, a že chyba zaokrouhlení δ_{100n} (pro libovolné n) ovlivní nanejvýše poslední místo, s nímž počítáme.

Verifikace této představy formálním aparátem matematické statistiky není zcela jednoduchá. Domnívám se, že v případě, kdy se při řešení konkrétního systému rovnic jedná o stanovení ekonomicky vhodné zaokrouhlovací jednotky, stačí se rozhodnout na základě výsledků pozorování vektorů $\Delta_n(\omega)$, a to dokonce s jedním opakováním a s menším rozsahem indexů n — počínat si prostě při volbě zaokrouhlovací jednotky tak, jako by hodnoty $\frac{4}{3}\Delta_n(\omega)$ představovaly skutečné chyby, jichž se dopustíme při výpočtu hodnot x_n.

Odhady jsou ovšem možné, zejména použijeme-li zjednodušujících předpokladů. Nesnáze spočívají v tom, že máme, přes rozsáhlé množství dat jen dvě nezávislá pozorování závislých vektorů Δ_n. Sledování výrazu $\|\Delta_{100n}\|^2$ ukázalo, že pro $n \geqq 10$ již nezávisí v podstatě na n; odtud však plyne, že $\Delta_{100(n+j)}$ a Δ_{100n} jsou přibližně nekorelované vektory pro $j \geqq 10$. (Tento předpoklad také neodporoval vypočtené výběrové korelační funkci náhodných proměnných $\|\Delta_{1000n}\|^2$.) Dostáváme tak přibližně nezávislá pozorování:

$$\|\Delta_{100n}(\omega_i)\|^2 \quad \text{pro} \quad n = 10, 20, \ldots, 50; \quad i = 1, 2:$$
$$28,\ 24,\ 37,\ 29,\ 35,\ 35,\ 29,\ 24,\ 32,\ 31.$$

Chceme-li použít dalšího zjednodušujícího předpokladu přibližné normality, můžeme použít t-testu a dostáváme intervalový odhad

$$\mathbf{E}\,\|\Delta_n\|^2 \leqq 35^2) \tag{5.4}$$

při stupni spolehlivosti přibližně (vzhledem k přibližnému splnění předpokladů t-testu) 0,99.

Ze vztahu (5.4) ovšem plyne

$$\mathrm{st}\,\mathbf{D}\delta_{nm} = \mathbf{E}\,\|\delta_{nm}\|^2 \leqq 46 \tag{5.5}$$

Z hodnot $\Delta_n(\omega)$ můžeme ovšem odhadovat i jiné charakteristiky chyby δ_n podle požadavků konkrétní situace.

6. Dirichletův problém. Výpočet $\mathbf{D}_n^{(\alpha\alpha)}$

Methoda předešlého odstavce nám dala poměrně jednoduchou cestou přehled o pravděpodobnostním chování chyb zaokrouhlení δ_n.

Druhou cestou je odhad matice $\mathbf{D}\delta_n$, resp. matice D_n. Uvědomíme si, že matice D_n je typu 100×100 a je zadána poměrně dosti složitým způsobem. Nejdůležitější ovšem jsou pro nás prvky na diagonále matice D_n. Můžeme přibližně odhadnout, který z prvků $D_n^{(\alpha\alpha)}$ je největší a omezit se jen na výpočet

[2]) Výsledek je zaokrouhlen nahoru.

tohoto prvku. Tento postup bude stále pracný; dá sice přesnější, avšak také užší výsledek, než methoda majorisující posloupnosti A_n. V konkrétním případě bude třeba uvážit, zda užitek plynoucí ze stanovení prvku $D_n^{(\alpha\alpha)}$ vyváží námahu se stanovením spojenou. V našem případě počítáme $D_n^{(\alpha\alpha)}$ 1. pro ilustraci jako příklad, 2. pro závažnost a v praxi časté řešení Dirichletova problému. 3. pro možnost srovnání výsledku s analogickým výsledkem pro Rietzovu methodu; zatím co při Rietzově methodě se poměrně snadno zjistí, že chyba zaokrouhlení δ_n je velmi malá, není zřejmé (aspoň neznáme-li výsledky předešlého odstavce), zda tato chyba není při Seidelově methodě podstatně vyšší.

Nyní je třeba udat návod, jak počítat prvek $D_n^{(\alpha\alpha)}$. Pro každé přirozené číslo n tvaru $100s + i$ (s přirozené, $i = 1, 2, \ldots, 100$) pišme $[n] = i$. Označme symbolem e_α vektor, jehož α-tá složka je 1, ostatní složky jsou rovny nule. Pro každé přirozené číslo n položme $B_n = A'_{101-[n]}$. Definujme vektory z_i vztahem

$$z_0 = e_\alpha\,, \quad z_i = B_i z_{i-1}\,. \tag{6.1}$$

Pro $n = 100s$ platí

$$D_n^{(\alpha\alpha)} = \frac{1}{4}\sum_{i=1}^{n} [z_i^{(101-[i])}]^2\,. \tag{6.2}$$

Skutečně je $[BB']^{(\alpha\alpha)} = \|B'e_\alpha\|^2$ pro každou matici B a

$$D_n^{(\alpha\alpha)} = \sum_{i=1}^{n} [C_i C'_i]^{(\alpha\alpha)}\,,$$

kde $C_i = A_n A_{n-1} \ldots A_{i+1} M_i = B'_1 B'_2 \ldots B'_{n-i} M_i$, a tedy $C'_i = M_i B_{n-i} B_{n-i-1} \ldots B_1$ (speciálně pro $i = n$ je ovšem $C_n = M_n$).

Je tedy

$$D_n^{(\alpha\alpha)} = \sum_{i=1}^{n} \|C'_i e_\alpha\|^2 = \sum_{i=1}^{n} \|M_i z_{n-i}\|^2\,.$$

Vektor $M_i z$ má ovšem všechny složky kromě $[i]$-té, která je rovna $z^{[i]}$, rovny nule, odtud plyne, že $\|M_i z_{n-i}\|^2 = [z_{n-i}^{[i]}]^2$. Můžeme tedy dále psát $D_n^{(\alpha\alpha)} = \sum_{i=1}^{n} [z_{n-i}^{[i]}]^2 = \sum_{i=1}^{n} [z_i^{(101-[i])}]^2$, čímž je vztah (6.2) dokázán. Probíhá-li i čísla 1, 2, ..., probíhá $(101 - [i])$ čísla 100, 99, ..., 1, 100, 99, ..., 1, ...; zavedeme-li označení $\tilde{i} = (101 - [i])$, můžeme přepsat vztahy (6.1) a (6.2) na vztahy

$$z_0 = e_\alpha, \quad z_i = A'_{\tilde{i}} z_{i-1}\,, \tag{6.3}$$

$$D_n^{(\alpha\alpha)} = \frac{1}{4}\sum_{i=1}^{n} z_i^{(\tilde{i})}]^2\,; \tag{6.4}$$

použitím symetrie dostaneme vztahy pohodlnější pro zapisování:

$$D_n^{(\alpha\alpha)} = \frac{1}{4}\sum_{i=1}^{n} [u_i^{(i)}]^2\,, \tag{6.5}$$

kde $$u_0 = e_{\tilde{\alpha}}\,, \quad u_i = A'_i u_{i-1}\,. \tag{6.6}$$

Početní postup daný vztahy (6.5) a (6.6) lze ještě poněkud upravit. Označme pro každé $j = 0, 1, \ldots$ symbolem v_j vektor, jehož složky jsou definovány vztahy $v_j^{(\beta)} = u_{100j+\beta-1}^{(\beta)}$ pro $\beta = 1, 2, \ldots, 100$. Lze pak psáti

$$D_{100m}^{(\alpha\alpha)} = \frac{1}{4}\sum_{i=0}^{m-1} \|v_i\|^2\,; \tag{6.7}$$

toto formální vyjádření nabude zajímavosti důkazem vztahu

$$v_{i+1} = A_{100}A_{99}\ldots A_1 v_i \quad (i = 0, 1, \ldots)\,. \tag{6.8}$$

Dokážeme tento vztah. Upozorníme nejprve, že vektor $A'_{\overline{\gamma\delta}}u$ má stejné složky jako u, kromě složky $\overline{\gamma\delta}$, která je rovna nule ($[A'_{\overline{\gamma\delta}}u]^{(\overline{\gamma\delta})} = 0$) a složek „sousedních" s indexy $\overline{\gamma-1, \delta}$, $\overline{\gamma, \delta-1}$, $\overline{\gamma+1, \delta}$, $\overline{\gamma, \delta+1}$ (samozřejmě jen těch, pro něž má uvedený symbol smysl), které přejdou v součet původní složky s jednou čtvrtinou složky $u^{(\overline{\gamma\delta})}$).

K verifikaci vztahu (6.8) je třeba dokázat, že pro každé i; $\gamma, \delta = 1, \ldots, 10$ platí

$$v_{i+1}^{(\overline{\gamma\delta})} = \frac{1}{4}\left[v_{i+1}^{(\overline{\gamma-1,\delta})} + v_{i+1}^{(\overline{\gamma,\delta-1})} + v_i^{(\overline{\gamma+1,\delta})} + v_i^{(\overline{\gamma+1,\delta})}\right];$$

při tom zde i dále rozumíme složkou s indexem $\overline{\gamma', \delta'}$ non $\epsilon\,\{1, \ldots, 100\}$ prostě nulu. Označme $\overline{\gamma\delta} = \beta$, $\overline{\gamma-1, \delta} = \beta_1$, $\overline{\gamma, \delta-1} = \beta_2$, $\overline{\gamma, \delta+1} = \beta_3$, $\overline{\gamma+1, \delta} = \beta_4$ (je $\beta_1 < \beta_2 < \beta < \beta_3 < \beta_4$); hoření vztah můžeme přepsat jako

$$u_{100(i+1)+\beta-1}^{(\beta)} = \frac{1}{4}\left[u_{100(i+1)+\beta_1-1}^{(\beta_1)} + u_{100(i+1)+\beta_2-1}^{(\beta_2)} + u_{100i+\beta_3-1}^{(\beta_3)} + u_{100i+\beta_4-1}^{(\beta_4)}\right]. \tag{6.9}$$

Všimněme si nyní, jak se mění posloupnost čísel $u_{100i+\beta+j}^{(\beta)}$ pro $j = 0, \ldots$ $\ldots, 99$. Předpokládejme na okamžik, že všechna $\beta_i \epsilon\,\{1, \ldots, 100\}$. Z definice vektorů u_i a z významu operací A'_i ihned vyplývá, že $u_{100i+\beta}^{(\beta)} = (A'_\beta u_{100i+\beta-1})^{(\beta)} = 0$. Další operace A'_{β_3} změní tuto složku na $0 + \frac{1}{4}u_{100i+\beta_3-1}^{(\beta_3)}$; další změna nastává až při operaci A'_{β_4} a je $u_{100i+\beta_4}^{(\beta)} = \frac{1}{4}u_{100i+\beta_3-1}^{(\beta_3)} + \frac{1}{4}u_{100i+\beta_4-1}^{(\beta_4)}$. Operace $A'_{\beta_4+1}, A'_{\beta_4+2}, \ldots, A'_{100}, A'_1, \ldots, A'_{\beta_1-1}$ ponechávají β-tou složku nezměněnu; operace A'_{β_1} přičte k ní opět číslo $\frac{1}{4}u_{100(i+1)+\beta_1-1}^{(\beta_1)}$, takže $u_{100(i+1)+\beta_1}^{(\beta)} = \frac{1}{4}(u_{100i+\beta_3-1}^{(\beta_3)} +$ $+ u_{100i+\beta_4-1}^{(\beta_4)} + u_{100(i+1)+\beta_1-1}^{(\beta_1)})$. Konečně další změna nastává při operaci A'_{β_2} a je $u_{100(i+1)+\beta-1}^{(\beta)} = u_{100(i+1)+\beta_2}^{(\beta)} = \frac{1}{4}(u_{100i+\beta_3-1}^{(\beta_3)} + u_{100i+\beta_4-1}^{(\beta_4)} + u_{100(i+1)+\beta_1-1}^{(\beta_1)} +$ $+ u_{100(i+1)+\beta_2-1}^{(\beta_2)})$; jestliže opět $u^{(\epsilon)}$ značí 0 pro každý symbol ε, který neoznačuje žádné z čísel $1, \ldots, 100$, platí právě uvedený vztah pro všechna β a vztah (6.9), a tedy i vztah (6.8) je dokázán.

Výpočet vektorů v_i podle vztahů (6.8) je snadnější než výpočet podle vztahu (6.6). Protože však nám vztah (6.8) nebyl na počátku znám, byl výpočet vektorů $v_0, \ldots, v_7$ prováděn podle vztahů (6.6) a dále se postupovalo podle (6.8). Bylo ovšem nutno opět zaokrouhlovat; zaokrouhlování bylo prováděno ná-

hodně, pracovalo se se 3 desetinnými místy. Při postupu podle (6.6) bylo třeba na každém kroku zaokrouhlovat čtyři čísla; na př. vektor

$$\begin{Vmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{Vmatrix}$$

se zaokrouhloval se stejnou pravděpodobností na jeden z těchto vektorů

$$\begin{Vmatrix} 1 \\ 0 \\ 0 \\ 0 \end{Vmatrix}, \quad \begin{Vmatrix} 0 \\ 1 \\ 0 \\ 0 \end{Vmatrix}, \quad \begin{Vmatrix} 0 \\ 0 \\ 1 \\ 0 \end{Vmatrix}, \quad \begin{Vmatrix} 0 \\ 0 \\ 0 \\ 1 \end{Vmatrix}.$$

Při postupu podle (6.8) se zaokrouhlovalo P°-způsobem. Pro každé i byl tedy místo vektoru v_i pozorován náhodný vektor $\nu_i = v_i + \vartheta_i$ takový, že $\mathbf{E}\vartheta_i = 0$. Je

$$\mathbf{E}\sum_{i=0}^{m-1}\sum_{\beta=1}^{p}[(v_i + \vartheta_i)^{(\alpha)}]^2 = \mathbf{E}\sum_{i=0}^{m-1}\sum_{\alpha=1}^{p}[(v_i^{(\alpha)})^2 + 2\vartheta_i^{(\alpha)}v_i^{(\alpha)} + (\vartheta_i^{(\alpha)})^2] =$$

$$= \sum_{i=0}^{m-1}\sum_{\alpha=1}^{p}[(v_i^{(\alpha)})^2 + \mathbf{E}[\vartheta_i^{(\alpha)}]^2] > \sum_{0=1}^{m-1}\sum_{\alpha=1}^{p}(v_i^{(\alpha)})^2 = \sum_{0=1}^{m-1}\|v_i\|^2 .$$

Položíme-li tedy

$$\lambda_m = \frac{1}{4}\sum_{i=0}^{m-1}\|\nu_i\|^2 ,$$

je $\mathbf{E}\lambda_m > D_{100m}^{(\alpha\alpha)}$.

Provedli jsme dvě nezávislá pozorování posloupnosti vektorů $\{\lambda_m\}_{m=1}^{\infty}$ pro speciální hodnotu $\alpha = \overline{7,7}$. Pro Rietzovu methodu je $D_n^{(\alpha\alpha)}$ maximální pro $\alpha = (\overline{5, 5}), (\overline{5, 6}), (\overline{6, 5}), (\overline{6, 6})$; domníváme se, že pro Seidelovu methodu bude $D_n^{(\alpha\alpha)}$ největší opět pro α na diagonále, tedy pro α tvaru $(\overline{\gamma\ \gamma})$, avšak zdá se z výsledků předchozího paragrafu, že maxima bude dosaženo pro $\alpha \geqq 6$; odtud plyne naše volba (viz také tabulku 5.3).

Hodnoty pozorování $\lambda_m(\omega_1)$ a $\lambda_m(\omega_2)$ náhodných proměnných λ_m jsou udány v tabulce 6.1. Ověřuje se tvrzení z předchozích odstavců, že pro vyšší m (5 až 10) se již $D_{100m}^{(\alpha\alpha)}$ podstatně nemění.

(To, že se $\lambda_m(\omega)$ od jistého indexu $M(\omega)$ již nemění plyne z toho, že pro tento index (a tedy i pro další) byla napozorovaná hodnota vektoru $\nu_i(\omega)$ rovna $\mathbf{0}$.

Předpokládáme-li, že náhodný vektor $\lambda_\infty = \lim_{m\to\infty}\lambda_m$ má přibližně normální rozložení, dostáváme intervalový odhad (zvolíme-li stupeň spolehlivosti $P = 0{,}99$) $\mathbf{E}\lambda_\infty < 0{,}547$, a tedy

$$\mathbf{D}^{(\overline{77,77})}\delta_n < 0{,}547 . \tag{6.10}$$

Při volbě stupně spolehlivosti $P = 0{,}999$ bychom dostali

$$\mathbf{E}\lambda_\infty < 0{,}784 ,$$

Tabulka 6.1.

m	$\lambda_m(\omega_1)$	$\lambda_m(\omega_2)$	m	$\lambda_m(\omega_1)$	$\lambda_m(\omega_2)$
1	0,2887270	0,2886770	37	0,5194862	0,5210338
2	0,3607665	0,3611622	38	0,5195370	0,5210858
3	0,3985355	0,3992562	39	0,5195778	0,5211348
4	0,4239322	0,4250132	40	0,5196090	0,5211758
5	0,4430680	0,4444062	41	0,5196375	0,5112202
6	0,4579552	0,4592172	42	0,5196612	0,5212540
7	0,4697512	0,4708942	43	0,5196828	0,5212820
8	0,4789200	0,4801402	44	0,5197008	0,5213058
9	0,4861670	0,4874998	45	0,5197190	0,5213245
10	0,4918880	0,4933800	46	0,5197320	0,5213395
11	0,4965148	0,4982300	47	0,5197452	0,5213475
12	0,5002782	0,5021168	48	0,5197582	0,5213555
13	0,5033018	0,5052810	49	0,5197705	0,5213632
14	0,5058430	0,5079002	50	0,5197795	0,5213722
15	0,5079730	0,5100855	51	0,5197902	0,5213792
16	0,5097242	0,5119015	52	0,5198092	0,5213852
17	0,5112130	0,5134295	53	0,5198070	0,5213948
18	0,5124885	0,5146952	54	0,5198100	0,5214048
19	0,5135678	0,5157595	55	0,5198110	0,5214152
20	0,5144790	0,5166335	56	0,5198110	0,5214218
21	0,5152455	0,5173548	57		0,5214270
22	0,5159213	0,5179602	58		0,5214335
23	0,5164922	0,5184702	59		0,5214370
24	0,5169968	0,5188810	60		0,5214405
25	0,5174235	0,5192305	61		0,5214440
26	0,5177995	0,5195235	62		0,5214485
27	0,5180952	0,5197862	63		0,5214522
28	0,5183472	0,5200125	64		0,5214540
29	0,5185670	0,5202070	65		0,5214562
30	0,5187378	0,5203718	66		0,5214582
31	0,5188850	0,5205105	67		0,5214595
32	0,5190105	0,5206350	68		0,5214600
33	0,5191288	0,5207418	69		0,5214605
34	0,5192345	0,5208305	70		0,5214605
35	0,5193320	0,5209040			
36	0,5194115	0,5209720	∞	0,5198110	0,5214605

a tedy

$$\mathbf{D}^{(\overline{77},\overline{77})}\delta_n < 0{,}784\,. \tag{6.11}$$

Pro srovnání uvádíme analogickou hodnotu pro Rietzovu methodu. Maximální je zde element s indexem $\overline{66}$, $\overline{66}$ a platí

$$\mathbf{D}^{(\overline{66},\overline{66})}\delta_n < 0{,}537872\,, \tag{6.12}$$

při čemž číslo vpravo bylo napočítáno vzhledem ke vztahu (4.4) podle Abramova [1] ze vzorce

$$\frac{1}{4}\,\frac{4}{11^2}\sum_{\mu=1}^{10}\sum_{\nu=1}^{10}\frac{\sin^2\frac{6\mu\pi}{11}\sin^2\frac{6\nu\pi}{11}}{1-\left[\frac{\cos\frac{\mu\pi}{11}+\cos\frac{\nu\pi}{11}}{2}\right]^2}\,.$$

Žádáme čtenáře, aby si lask. doplnil v literatuře k článku V. Fabiana na str. 43 u citace [3]: Mathematische Nachrichten, 16 (1957), No 5–6.

LITERATURA

[1] *Абрамов:* О влиции ошибок округления при решении уравнения Лапласса, Вычислительная математика и вычислительная техника, Сб. I АН СССР, Москва 1953.

[2] *Ch. Blanc. W. Liniger:* Erreurs de chute dans la résolution de systèmes algébriques linéaires, Com. Math. Helvetici 30, Fasc. 4, 257—264.

[3] *Václav Fabian:* Zufälliges Abrunden und die Konvergenz des linearen (Seidelschen) Iterationverfahrens, Mathematische Nachrichten.

[4] *Václav Fabian:* L'influence d'arrondissement aux évaluations numériques linéaires. Czech. math. j. 8 (83), 1958, No 2.

[5] *G. E. Forsythe:* Note on rounding-off errors, Nat. Bur. Stand., Los Angeles, Calif., 3 pp. (1950).

[6] *J. Franel:* A propos des tables de logarithme, Vierteljahrschrift der Naturforschenden Gesellschaft in Zürich (1917), 62, s. 286.

[7] *Harry D. Huskey:* On the Precision of a certain Procedure of Numerical Integration, J. Research Nat. Bur. Stand. 42 (1949), 57—62.

[8] *M. G. Kendall:* The Advanced Theory of Statistics, London 1945.

[9] *J. von Neumann, H. H. Goldstine:* Numerical Inverting of Matrices of High Order, Bull. Amer. Math. Soc. 53, No 11 (1947), 1021—1099.

Tabulky náhodných čísel:

[10] Journal of the American Statistical Association, 48 (1953), str. 167 a další.

[11] *M. G. Kendall, B. Babington Smith:* Tables of Random Sampling Numbers, Tracts for computers, No 24, 1940.

[12] *L. H. C. Tipett:* Random Sampling Numbers, Tracts for computers, No 15, 1927.

Резюме

ОЦЕНКА ОШИБКИ ЗАКРУГЛЕНИЯ ПРИ ЛИНЕЙНЫХ ИТЕРАЦИОННЫХ ПРОЦЕССАХ, В ОСОБЕННОСТИ ПРИ РЕШЕНИИ ПО СЕЙДЕЛУ ПРОБЛЕМЫ ДИРИХЛЕ ДЛЯ КВАДРАТА 10×10

ВАЦЛАВ ФАБИАН (Václav Fabian)

(Поступило в редакцию 8/V 1957 г.)

В статье анализируется применение статистических методов к оценке ошибки закругления при обычном и случайном способе закругления. В качестве примера приложений общих результатов работы [4] проводятся некоторые численные оценки величины ошибки в специальном случае метода Сейдела для разностного аналога проблемы Дирихле для квадрата 10×10.

Résumé

ESTIMATION DE L'ERREUR DUE A L'ARRONDISSEMENT DANS LES PROCESSUS ITERATIFS LINEAIRES, EN PARTICULIER DANS LE PROCEDE DE SEIDEL POUR LA SOLUTION DU PROBLEME DE DIRICHLET DANS LE CARRE 10×10

VÁCLAV FABIAN

(Reçu le 8 mai 1957.)

Dans cet article, on discute l'emploi des méthodes statistiques d'estimation de l'erreur due à l'arrondissement pour les procédés d'arrondissement habituel et aléatoire. A titre d'exemple d'application des résultats généraux obtenus dans le travail [4], on a évalué ici quelques estimations numériques de l'erreur pour le cas spécial du procédé de Seidel appliqué à l'analogie en différences finies du problème de Dirichlet, dans le carré 10×10.