ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
SERIA III: MATEMATYKA STOSOWANA V (1975)

HO THUAN (Hanoi)

Pewne warunki dostateczne zbieżności metod Jacobiego i Gaussa–Seidela dla wielkich układów równań liniowych

0. Wprowadzenie. Praca zawiera trzy warunki dostateczne zbieżności metod Jacobiego i Gaussa–Seidela iteracyjnego rozwiązywania układów równań liniowych postaci $A\vec{x} = \vec{b}$. Warunki te opierają się na pewnej własności macierzy A, zdefiniowanej w pracy jako „kryterium sumy". Własność ta na ogół nie jest równoważna nieredukowalności macierzy A.

Przedstawione wyniki traktujemy przede wszystkim jako przyczynek teoretyczny do poznania metod Jacobiego i Gaussa–Seidela. Stosowalność tych metod w praktyce jest znikoma, gdyż na ogół potrafimy wskazać inne znacznie efektywniejsze metody iteracyjne.

1. Niech $A = (a_{ij})$ będzie zespoloną macierzą $n \times n$, a $\vec{b}$ danym wektorem $n \times 1$. Poszukujemy rozwiązania układu

$$A\vec{x} = \vec{b}. \tag{1.1}$$

Macierz A przedstawmy w postaci

$$A = D - E - F, \tag{1.2}$$

gdzie $D = \operatorname{diag}(a_{ii})$, E jest macierzą ściśle dolną trójkątną, F ściśle górną trójkątną. Zakładamy, że macierz D jest nieosobliwa.

Rozważmy metody iteracyjne Jacobiego i Gaussa–Seidela definiowane następująco:

$$\vec{x}_{m+1} = D^{-1}(E+F)\vec{x}_m + D^{-1}\vec{b}, \qquad m \geqslant 0, \tag{1.3}$$

(metoda Jacobiego, por. [5], str. 71 i dalsze),

$$\vec{x}_{m+1} = (D-E)^{-1}F\vec{x}_m + (D-E)^{-1}\vec{b} \tag{1.4}$$

(metoda Gaussa–Seidela, por. [5], str. 72 i dalsze).

Macierze

$$B = D^{-1}(E+F) \tag{1.5}$$

oraz

$$C = (D-E)^{-1}F \tag{1.6}$$

nazywamy macierzą Jacobiego i macierzą Gaussa–Seidela odpowiednio. Metody Jacobiego i Gaussa–Seidela są szczególnym przypadkiem metody iteracji prostych, zastosowanej do układu równań postaci

$$\vec{x} = G\vec{x} + \vec{g}.$$

Konstruowany ciąg iteracyjny $\{\vec{x}_n\}$ jest określony rekurencyjnie:

$$\vec{x}_{k+1} = G\vec{x}_k + \vec{g}.$$

Zbieżność tej metody jest uzależniona od $\rho(G)$, gdzie $\rho(G)$ jest promieniem spektralnym macierzy G $(\rho(G) = \max_{\lambda \in \mathrm{spect}(G)} |\lambda|)$.

Prawdziwe jest twierdzenie (por. [4] i [5]):

Ciąg $\{\vec{x}_n\}$ *jest zbieżny do rozwiązania przy dowolnym przybliżeniu początkowym* $\vec{x}_0$ *wtedy i tylko wtedy, gdy* $\rho(G) < 1$.

Dla uzyskania warunków dostatecznych zbieżności metod Jacobiego i Gaussa–Seidela poszukuje się na ogół takich własności macierzy A, aby odpowiadające jej macierze B i C miały promienie spektralne mniejsze od jedności. Wielu autorów (por. [5]) zakłada, że A jest macierzą nieredukowalną, a następnie, przyjmując dodatkowe założenia, dowodzi zbieżności metod Jacobiego i Gaussa–Seidela.

W tej pracy założenie nieredukowalności zastępujemy innym, na ogół nie równoważnym założeniem (tzw. „kryterium sumy"), a następnie otrzymujemy analogiczne twierdzenia o zbieżności rozważanych procesów iteracyjnych. Przypomnijmy pojęcie redukowalności macierzy (por. [5], str. 37).

DEFINICJA 1.1. Niech $n \geqslant 2$. Macierz A $(n \times n)$ jest *redukowalna* jeśli istnieje macierz permutacji P $(n \times n)$ taka, że dla pewnego $r \in N$, $N = \{1, 2, \ldots, n\}$,

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

gdzie A_{11} jest macierzą $r \times r$, A_{22} macierzą $(n-r) \times (n-r)$. Jeśli nie istnieje macierz P o powyższej własności, to A jest macierzą *nieredukowalną*.

Łatwo sprawdzić (por. [5], str. 37 i 38), że z definicji 1.1 wynika

TWIERDZENIE 1.1. 1. *Macierz A jest redukowalna wtedy i tylko wtedy, gdy istnieją dwa rozłączne i niepuste zbiory I, J takie, że*

$$I \cup J = N, \qquad n \geqslant 2$$

oraz

$$a_{ij} = 0 \qquad dla \qquad i \in I, \ j \in J.$$

Jeśli takie zbiory I, J nie istnieją, to macierz A jest nieredukowalna.

2. *Macierz A jest nieredukowalna wtedy i tylko wtedy, gdy do każdej pary* i, j $(1 \leqslant i, j \leqslant n, i \neq j)$ *istnieją wskaźniki* $i_1, i_2, \ldots, i_s$ *takie, że*

$$a_{ii_1} a_{i_1 i_2} \ldots a_{i_s j} \neq 0. \ \blacksquare$$

Z zacytowanego powyżej twierdzenia 1.1 wynikają lematy:

LEMAT 1.1. *Jeśli istnieje wskaźnik* k_0 *taki, że*

$$a_{k_0 j} \neq 0 \qquad dla \qquad j = 1, 2, \ldots, n, \quad j \neq k_0,$$

oraz

$$a_{i k_0} \neq 0 \qquad dla \qquad i = 1, 2, \ldots, n, \quad i \neq k_0,$$

to A jest macierzą nieredukowalną.

D o w ó d. Skorzystamy z punktu 2 twierdzenia 1.1. Przyjmując $i_1 = i_s = k_0$ mamy, na mocy założeń lematu 1.1,

$$a_{i k_0} \cdot a_{k_0 j} \neq 0,$$

co oznacza nieredukowalność macierzy A. ∎

LEMAT 1.2. *Jeśli liczba zerowych elementów niediagonalnych jest mniejsza niż $n-1$, to A jest macierzą nieredukowalną.*

D o w ó d. Załóżmy nie wprost, że A jest redukowalna. Na mocy punktu 1 twierdzenia 1.1 liczba zerowych elementów niediagonalnych A jest nie mniejsza niż

$$|I| * (n - |I|), \tag{1.7}$$

gdzie $|I|$ przybiera wartości całkowite z przedziału $[1, n-1]$. Iloczyn (1.7) jest zawsze nie mniejszy niż $n-1$, co jest sprzeczne z założeniem i tym samym oznacza, że A jest macierzą nieredukowalną. ∎

2. W tym rozdziale podajemy trzy warunki wystarczające zbieżności metod Jacobiego i Gaussa–Seidela. Warunki te opierają się na następującym kryterium.

DEFINICJA 2.1. Macierz A $(n \times n)$ spełnia *kryterium sumy* jeśli istnieją dwa rozłączne zbiory I, J oraz wskaźnik $k_0 \in I$, takie, że

$$I \neq \emptyset, \qquad I \cup J = N,$$

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}| \qquad \text{dla} \qquad i \in I,$$

$$|a_{ii}| = \sum_{j \neq i} |a_{ij}| \qquad \text{dla} \qquad i \in J,$$

oraz

$$a_{i k_0} \neq 0 \qquad \text{dla} \qquad i \in J. \quad ∎$$

Zauważmy przede wszystkim, że z kryterium sumy nie wynika nieredukowalność macierzy A. Świadczy o tym przykład

$$A = \begin{bmatrix} 2 & 0 & 2 & 0 \\ 0 & 2 & 1 & -1 \\ 2 & 0 & -4 & 0 \\ 1 & -2 & 0 & 4 \end{bmatrix}.$$

Macierz A spełnia kryterium sumy, gdy $I = \{3, 4\}$, $J = \{1, 2\}$ oraz $k_0 = 3$, natomiast A jest redukowalna na mocy punktu 1 twierdzenia 1.1, gdy $I = \{1, 3\}$ oraz $J = \{2, 4\}$.

TWIERDZENIE 2.1. *Jeśli A spełnia kryterium sumy, to A jest nieosobliwa.*

D o w ó d. Załóżmy nie wprost, że A jest macierzą osobliwą. Wówczas równanie jednorodne

$$A\vec{x} = \vec{0} \tag{2.1}$$

ma niezerowe rozwiązanie $\vec{x} = [x_1, \ldots, x_n]^T$.

Niech

$$|x_r| = \max_{1 \leqslant i \leqslant n} |x_i|.$$

Rozważmy r-te równanie w (2.1),

$$a_{rr} x_r = -\sum_{j \neq r} a_{rj} x_j.$$

Stąd

$$|a_{rr}| \leqslant \sum_{j \neq r} |a_{rj}| \left|\frac{x_j}{x_r}\right| \leqslant \sum_{j \neq r} |a_{rj}|. \tag{2.2}$$

Z (2.2) wynika, że $r \in J$, oraz

$$|a_{rr}| = \sum_{j \neq r} |a_{rj}| \left|\frac{x_j}{x_r}\right| = \sum_{j \neq r} |a_{rj}|,$$

co oznacza, że

$$|x_j| = |x_r| \qquad \text{dla tych } j, \text{ dla których } a_{rj} \neq 0.$$

Z własności kryterium sumy wynika, że

$$a_{rk_0} \neq 0, \qquad k_0 \in I,$$

czyli $|x_{k_0}| = |x_r|$.

Stąd i z (2.2) wynika nierówność

$$|a_{k_0 \underline{k_0}}| \leqslant \sum_{j \neq k_0} |a_{k_0 j}|,$$

co stanowi sprzeczność, gdyż $k_0 \in I$. Zatem A jest macierzą nieosobliwą. ■

TWIERDZENIE 2.2. *Jeśli A spełnia kryterium sumy, to metody Jacobiego i Gaussa–Seidela są zbieżne przy dowolnym przybliżeniu początkowym.*

D o w ó d. Macierz Jacobiego $B = (b_{ij})$ z (1.5) jest równa

$$b_{ij} = \begin{cases} 0 & \text{dla} \quad i = j, \\ -\dfrac{a_{ij}}{a_{ii}} & \text{dla} \quad i \neq j. \end{cases}$$

Zdefiniujmy macierz $\tilde{B} = (\tilde{b}_{ij})$ następująco:

$$\begin{aligned} \tilde{b}_{ik_0} &= \varepsilon \quad \text{jeśli} \quad i \in I \text{ oraz } b_{ik_0} = 0, \\ \tilde{b}_{k_0 j} &= \varepsilon \quad \text{jeśli} \quad j \neq k_0 \text{ oraz } b_{k_0 j} = 0, \\ \tilde{b}_{ij} &= b_{ij} \quad \text{dla pozostałych } i, j \quad (\text{dla } \varepsilon > 0). \end{aligned}$$

Na mocy lematu 1.1 macierz B jest nieredukowalna, a dla dostatecznie małego ε mamy

$$\sum_{j=1}^{n} |\tilde{b}_{ij}| < 1 \qquad \text{dla} \qquad i \in I.$$

Ze znanych własności macierzy o nieujemnych współczynnikach (por. [4]) wynika

$$\rho(B) \leqslant \rho(|B|) \leqslant \rho(|\tilde{B}|) < 1,$$

co oznacza, że metoda Jacobiego jest zbieżna.

Macierz $|B|$ oznacza macierz o elementach $|b_{ij}|$, $\rho(B)$ jest największym modułem wartości własnej macierzy B.

Przypomnijmy, że

$$B = D^{-1}(E + F) = L + U,$$

gdzie $L = D^{-1}E$ oraz $U = D^{-1}F$. Macierz Gaussa–Seidela C jest równa

$$C = (D - E)^{-1}F = (D - E)^{-1}DD^{-1}F = [D^{-1}(D - E)]^{-1}U = (I - L)^{-1}U.$$

Stąd też

$$|B| = |L| + |U|, \qquad |\tilde{B}| = |\tilde{L}| + |\tilde{U}|,$$

gdzie $|L| \leqslant |\tilde{L}|$ i $|U| \leqslant |\tilde{U}|$.

Oznaczmy

$$\tilde{C}_+ = (I - |\tilde{L}|)^{-1}|\tilde{U}|.$$

Wykorzystując twierdzenie Steina–Rosenberga (por. [4], [5], str. 120) mamy

$$\rho(\tilde{C}_+) \leqslant \rho(|\tilde{B}|) < 1. \tag{2.3}$$

Z drugiej strony

$$\begin{aligned} |C| &= |(I + L + \ldots + L^{n-1})U| \leqslant (I + |L| + \ldots + |L|^{n-1})|U| \leqslant \\ &\leqslant (I + |\tilde{L}| + \ldots + |\tilde{L}|^{n-1})|\tilde{U}| = (I - |\tilde{L}|)^{-1}|\tilde{U}| = \tilde{C}_+. \end{aligned}$$

Z (2.3) wynika zatem, że

$$\rho(C) \leqslant \rho(|C|) \leqslant \rho(\tilde{C}_+) < 1,$$

co oznacza zbieżność metody Gaussa–Seidela. ■

Uwagi

1. W przypadku gdy J jest zbiorem pustym (czyli $|I| = n$), macierz A jest ściśle diagonalnie zdominowana i twierdzenie 2.2 pokrywa się ze znanym rezultatem Misess, Geiringer [3] oraz Collatza [1].

2. Jeśli dodatkowo założymy, że A jest macierzą nieredukowalną, to twierdzenie 2.2 pokrywa się z rezultatem Geiringer [2], a co więcej, z dowodu twierdzenia 2.2 jest jasne, że założenie o istnieniu indeksu k_0 jest teraz niepotrzebne.

TWIERDZENIE 2.3. *Jeśli istnieją dwa rozłączne zbiory I, J takie, że*

$$I \neq \emptyset, \qquad I \cup J = N,$$

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}| \quad dla \quad i \in I,$$

$$|a_{ii}| \geqslant \sum_{j \neq i} |a_{ij}| \quad dla \quad i \in J,$$

a ponadto ilość zerowych a_{ij} dla $i \neq j$, $i \in J$, jest mniejsza niż $n-1$, to A jest macierzą nieosobliwą.

Dowód. Załóżmy, że A jest osobliwa i niech $\vec{x}$ będzie niezerowym rozwiązaniem równania $A\vec{x} = \vec{0}$. Niech

$$|x_r| = \max_{1 \leqslant i \leqslant n} |x_i|.$$

Jeśli $r \in I$, to powtarzając rozumowanie z twierdzenia 2.1 dochodzimy do sprzeczności.

Zatem $r \in J$ oraz $|a_{rr}| = \sum_{j \neq r} |a_{rj}|$. Z założeń naszego twierdzenia wynika, że co najmniej jeden element a_{rs} $(s \neq r)$ jest różny od zera.

Załóżmy, że

$$a_{rs_1}, \ldots, a_{rs_t} \neq 0 \qquad (1 \leqslant t \leqslant n-1).$$

Jeśli $t = n-1$, to $|x_1| = |x_2| = \ldots = |x_n|$, czyli $I = \emptyset$, co jest sprzeczne z założeniem. A więc $t < n-1$.

Rozważmy dwa przypadki

(a) Istnieje co najmniej jeden element a_{rs_l} z $s_l \in I$ $(1 \leqslant l \leqslant t)$. Powtarzając dowód z twierdzenia 2.1 dochodzimy i tym razem do sprzeczności.

(b) Wszystkie $s_l \in I$ dla $l = 1, 2, \ldots, t$. Wówczas wśród t wierszy o numerach $s_1, s_2, \ldots, s_t$ istnieje co najmniej jeden, który zawiera niezerowe elementy niediagonalne na mocy założenia, że liczba zerowych a_{ij} dla $i \neq j$, $i \in J$, jest mniejsza niż $n-1$. Dochodzimy zatem znów do sprzeczności, co ostatecznie dowodzi nieosobliwości macierzy A. ■

TWIERDZENIE 2.4. *Jeśli A spełnia założenia twierdzenia* 2.3, *to metody Jacobiego i Gaussa–Seidela są zbieżne przy dowolnych przybliżeniach początkowych.*

D o w ó d. Teza twierdzenia 2.4 wynika z dowodu twierdzenia 2.2 z nieznacznymi zmianami.

Jako następny przypadek zapewniający zbieżność omawianych metod rozważmy twierdzenie 2.5 i 2.6.

TWIERDZENIE 2.5. *Jeśli istnieją dwa rozłączne zbiory I, J takie, że*

$$I \neq \emptyset, \qquad I \cup J = N,$$

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}| \qquad \textit{dla} \qquad i \in I,$$

$$|a_{ii}| = \sum_{j \neq i} |a_{ij}| \qquad \textit{oraz} \qquad a_{i,i-1} \neq 0, \; a_{i,i+1} \neq 0 \qquad \textit{dla} \qquad i \in J,$$

to macierz A jest nieosobliwa.

D o w ó d. Załóżmy, że A jest osobliwa i niech $\vec{x}$ będzie niezerowym rozwiązaniem równania jednorodnego $A\vec{x} = \vec{0}$. Oznaczmy

$$|x_r| = \max_{1 \leqslant i \leqslant n} |x_i|.$$

Niech $i_0 \in I$ spełnia warunek

$$|i_0 - r| = \min_{i \in I} |i - r|.$$

Jeśli $r \in I$, tzn. $r = i_0$, to powtarzając dowód twierdzenia 2.1 dochodzimy do sprzeczności.

Zatem, $r \in J$, a stąd

$$|a_{rr}| = \sum_{j \neq r} |a_{rj}|.$$

Oznacza to, że $|x_j| = |x_r|$, gdy $a_{rj} \neq 0$.

Rozważmy dwa przypadki.

(a) $r > i_0$. Z założeń twierdzenia 2.5 mamy $a_{r,r-1} \neq 0$, czyli $|x_{r-1}| = |x_r|$. Jeśli $r - 1 = i_0$, to otrzymujemy natychmiastową sprzeczność. Jeśli nie, to $(r-1) \in J$, a stąd $a_{r-1,r-2} \neq 0$. Tak więc $|x_{r-2}| = |x_r|$, a postępując analogicznie stwierdzamy, że $|x_{i_0}| = |x_r|$. Stosując rozumowanie z dowodu twierdzenia 2.1, dochodzimy do sprzeczności.

(b) $r < i_0$. Z założenia twierdzenia 2.6 wiemy, że $a_{r,r+1} \neq 0$ i tak jak w przypadku (a) stwierdzamy, że $|x_{i_0}| = |x_r|$, co prowadzi do sprzeczności.

Udowodniliśmy zatem, że A jest macierzą nieosobliwą. ∎

TWIERDZENIE 2.6. *Jeśli A spełnia założenia twierdzenia* 2.5, *to metody Jacobiego i Gaussa–Seidela są zbieżne przy dowolnych przybliżeniach początkowych.*

D o w ó d. Macierz Jacobiego $B = (b_{ij})$ jest równa,

$$b_{ij} = \begin{cases} 0 & \text{dla} \quad i = j, \\ -\dfrac{a_{ij}}{a_{ii}} & \text{dla} \quad i \neq j. \end{cases}$$

Macierz $\widetilde{B} = (\widetilde{b}_{ij})$ jest definiowana następująco:

$$\widetilde{b}_{i,i-1} = \varepsilon \quad \text{dla} \quad i \in I \text{ oraz } b_{i,i-1} = 0,$$

$$\widetilde{b}_{i,i+1} = \varepsilon \quad \text{dla} \quad i \in I \text{ oraz } b_{i,i+1} = 0,$$

$$\widetilde{b}_{ij} = b_{ij} \quad \text{dla pozostałych } i, j,$$

gdzie $\varepsilon > 0$ jest dostatecznie małe, tak aby

$$\sum_{j=1}^{n} |\widetilde{b}_{ij}| < 1 \quad \text{dla} \quad i \in I.$$

Macierz $\widetilde{B}$ jest nieredukowalna, gdyż $\widetilde{b}_{i,i+1}$, $\widetilde{b}_{i,i-1}$ są różne od zera (por. [4]).

Dalszy dowód jest analogiczny do dowodu twierdzenia 2.2, skąd też wynika zbieżność omawianych metod. ∎

Bibliografia

[1] L. Collatz, *Fehlerabschätzung für das Iterations Verfahren zur Ausflösung linearer Gleichungs-systeme*, Z. Angew. Math. Mech. 22 (1942), str. 357–361; 74, 95.

[2] H. Geiringer, *On the solution of systems of linear equations by certain iterative methods*, Reissner Anniversary volume J. W. Edwards, Ann Arbor 1949, str. 365–393, 2, 19, 25, 57, 58, 95, 96, 160.

[3] R. Misess und H. Pollaczek-Geiringer, *Praktische Verfahren der Gleichungs Auflosung*, Z. Angew. Math. Mech. 9 (1929), str. 58–77, 152–164; 94, 95, 96.

[4] R. S. Varga, *Matrix Iterative Analysis*, Prentice Hall 1962.

[5] D. M. Young, *Iterative Solution of Large Linear Systems*, Academic Press, 1971.