BIULETYN INFORMACYJNY INSTYTUTU TECHNIKI CIEPLNEJ
POLITECHNIKI WARSZAWSKIEJ

Nr 35 1972

mgr inż. Michał Podowski

Instytut Techniki Cieplnej

METODA RÓŻNICOWA ROZWIĄZYWANIA UKŁADU RÓWNAŃ KINETYKI REAKTORA

1. Wstęp

Równania dyfuzji, określające rozkład neutronów w reaktorze (z uwzględnieniem G-grup neutronów opóźnionych), można zapisać w postaci następującej [1]:

$$\left.\begin{aligned} \left[\nabla D \nabla - A + (1-\beta)\varkappa H\right]\bar{\phi} + \sum_{i=1}^{G} \lambda_i \varkappa_i C_i &= V^{-1}\frac{\partial \bar{\phi}}{\partial t} \\ \beta_i H\bar{\phi} - \lambda_i \bar{C}_i &= \frac{\partial \bar{C}_i}{\partial t} \qquad (i = 1,2,\ldots,G), \end{aligned}\right\} \quad (1.1)$$

gdzie: $\beta_i, \beta, \varkappa, \varkappa_i, \lambda_i$ - skalary, $\beta = \sum_{i=1}^{G} \beta_i$,

$\left.\begin{aligned} \bar{\phi} &= \bar{\phi}(\bar{r},t) \\ \bar{C}_i &= \bar{C}_i(r,t) \end{aligned}\right\}$ - wektory g-wymiarowe,

$\nabla D \nabla$, A, H, V^{-1} - macierze kwadratowe [g x g].

Z uwagi na trudności, związane z dokładnym rozwiązaniem układu (1.1), stosuje się tutaj metody przybliżone. Podstawą jednej z takich metod jest założenie, że wektory $\bar{\phi}(\bar{r},t)$ oraz $\bar{C}_i(\bar{r},t)$ można przedstawić w postaci:

$$\bar{\phi}(\bar{r},t) = \sum_{k=1}^{K} \bar{\psi}_k(\bar{r})\, T_k(t),$$

$$C_i(r,t) = \sum_{k=1}^{K} \bar{\psi}_k(\bar{r})\, C_{ik}(t),$$

gdzie: $\bar{\psi}_k(\bar{r})$ - znane g-wymiarowe wektory,
$T_k(t)$, $C_{ik}(t)$ - nieznane wielkości skalarne.

Przyjmując powyższe założenie można układ (1.1) doprowadzić, po pewnych przekształceniach, do postaci układu równań różniczkowych zwyczajnych:

$$\left.\begin{aligned} \Lambda\dot{\vec{n}} &= \left[R(t) - B\right]\vec{n} + \sum_{j=1}^{G} \lambda_j \vec{C}_j , \\ \dot{\vec{C}}_j &= B_j \vec{n} - \lambda_j \vec{C}_j \qquad (j = 1,2,\ldots,G), \end{aligned}\right\} \tag{1.2}$$

gdzie: $B = \sum_{j=1}^{G} B_j$,

B_j, R, Λ - macierze wymiaru $[K \times K]$, przy czym $\det\Lambda \neq 0$, $R^T = R$,

λ_j - skalary.

Stosując podstawienie

$$\vec{\xi}_j = \Lambda^{-1} \vec{C}_j$$

można układ (1.2) przekształcić do postaci

$$\left.\begin{aligned} \dot{\vec{n}} &= \Lambda^{-1}\left[R(t) - B\right]\vec{n} + \sum_{j=1}^{G} \lambda_j \vec{\xi}_j , \\ \dot{\vec{\xi}}_j &= \Lambda^{-1} B_j \vec{n} - \lambda_j \vec{\xi}_j \qquad (j = 1,2,\ldots,G). \end{aligned}\right\} \tag{1.3}$$

Oznaczając:

$$P(t) = \begin{bmatrix} \Lambda^{-1}\left[R(t) - B\right] & \lambda_1 I & \cdots & \lambda_G I \\ \Lambda^{-1} B_1 & -\lambda_1 I & 0 & \\ \vdots & & \vdots & \\ \vdots & 0 & & \vdots \\ \Lambda^{-1} B_G & & & -\lambda_G I \end{bmatrix}, \quad x(t) = \begin{bmatrix} \vec{n} \\ \vec{\xi}_1 \\ \vdots \\ \vdots \\ \vec{\xi}_G \end{bmatrix}$$

oraz uwzględniając warunek początkowy

$$x(t_0) = x_0 ,$$

otrzymuje się poprawnie sformułowane zagadnienie początkowe:

$$\left.\begin{aligned} \dot{x}(t) &= P(t)\, x(t), \\ x(t_0) &= x_0. \end{aligned}\right\} \tag{1.4}$$

W przypadku, gdy elementami macierzy P są macierzy wymiaru $[1 \times 1]$, układ (1.4) przybiera postać:

$$\left.\begin{aligned} \dot{x}(t) &= P_1(t)\, x(t), \\ x(t_0) &= x_0, \end{aligned}\right\} \tag{1.5}$$

gdzie:

$$\begin{bmatrix} \frac{1}{\Lambda}\left[\rho(t) - \beta\right] & \lambda_1 & \cdots & \lambda_G \\ \frac{\beta_1}{\Lambda} & -\lambda_1 & & 0 \\ \vdots & & \ddots & \\ \frac{\beta_G}{\Lambda} & 0 & & -\lambda_G \end{bmatrix}$$

Równania (1.5) noszą nazwę punktowych równań kinetyki. Wartości własne macierzy $P_1(t)$ są pierwiastkami równania (ze względu na ω)

$$\omega \Lambda + \sum_{j=1}^{G} \frac{\omega \beta_j}{\omega + \lambda_j} - \rho(t) = 0, \tag{1.6}$$

przy czym można wykazać [2], że jeśli $\lambda_G < \lambda_{G-1} < \cdots < \lambda_1$, to w każdym z przedziałów $(-\infty, -\lambda_1)$, $(-\lambda_1, -\lambda_2), \ldots,$ $(-\lambda_{G-1}, -\lambda_G)$ znajduje się dokładnie jedna wartość własna macierzy $P_1(t)$ oraz:

a) jeśli $\rho(t) > 0$, wówczas

$$\omega_{\min} \in \left(-\lambda_1 - \frac{\beta}{\Lambda},\ -\lambda_1\right),$$

$$\omega_{\max} \in (0, +\infty),$$

b) jeśli $\rho(t) < 0$, to

$$\omega_{min} \in \left(-\lambda_1 + \frac{\rho(t) - \beta}{\Lambda}, \ -\lambda_1 \right),$$

$$\omega_{max} \in (-\lambda_G, \ 0).$$

Dla przypadku G = 6 przykładowe wartości liczbowe stałych są następujące: $\beta = 0{,}0075$, $\Lambda = 10^{-4}$ s, $\lambda_1 = 14\ s^{-1}$, $\lambda_G = 0{,}0124\ s^{-1}$. Przedziały, wewnątrz których zawarte są ekstremalne wartości własne (wyrażone w sekundach), będą wtedy następujące:
przy założeniu a)

$$\omega_{min} \in (-89, -14),$$

natomiast przedział $(0, +\infty)$, do którego należy ω_{max}, dla praktycznie stosowanych wartości $\rho(t)$ (które są zawsze znacznie mniejsze od β) zawęża się do odcinka (0, 1):

$$\omega_{max} \in (0, 1),$$

przy założeniu b):

$$\omega_{min} \in (-89 + \frac{\rho}{\Lambda}, \ -14),$$

$$\omega_{max} \in (-0{,}0124, \ 0).$$

Powracając do ogólnej postaci równań kinetyki (wzór (1.4)), gdy elementami macierzy P(t) są macierze $[K \times K]$ -wymiarowe, można dowieść [1], że każdy z przedziałów $(-\infty, -\lambda_1)$, $(-\lambda_1, -\lambda_2)$, ..., $(-\lambda_G, +\infty)$ zawiera dokładnie K wartości własnych macierzy P przy czym, jeśli macierz R(t) jest określona dodatnio, wówczas wartości własne z przedziału $(-\lambda_G, +\infty)$ leżą w przedziale $(0, +\infty)$, natomiast jeśli R(t) jest określona ujemnie, znajdują się one w przedziale $(-\lambda_G, 0)$.

Przy rozpatrywaniu zagadnienia numerycznego rozwiązania zadania (1.4), przy tak niesymetrycznym rozkładzie wartości własnych macierzy P(t) ($\omega_{min} < 0$, $-\omega_{min} \gg |\omega_{max}|$), nasuwa się pytanie: czy istnieje możliwość dobrania schematu różnicowego, aproksymującego zadanie (1.4), pozwalającego na optymalne oszacowanie błędu aproksymacji. Celem niniejszej pracy jest

znalezienie schematu różnicowego o pewnej, z góry założonej postaci, aproksymującego z rzędem jeden różniczkowe zagadnienie początkowe typu (1.4), przy którym błąd rozwiązania przybliżonego jest minimalny względem pewnego ciągu parametrów.

2. Analiza zbieżności rozwiązania schematu różnicowego do rozwiązania zagadnienia różniczkowego

W większości zagadnień, dotyczących numerycznego rozwiązywania równań różniczkowych, rozpatruje się możliwości doboru schematów różnicowych pod kątem uzyskania odpowiedniego rzędu zbieżności względem kroku siatki. Rzadziej natomiast występuje problem oszacowania rzeczywistej wartości błędu przy zadanym kroku. Uzyskanie takiego właśnie oszacowania (w odniesieniu do układu równań różniczkowych zwyczajnych) i znalezienie jego wartości minimalnej jest celem poniższych rozważań.

2.1. Postawienie zadania

Dane jest różniczkowe zagadnienie początkowe:

$$\left.\begin{aligned} \dot{x} &= A(t)x\,, \\ x(t_0) &= x_0, \end{aligned}\right\} \tag{2.1}$$

gdzie: $A(t)$ - macierz kwadratowa wymiaru $[K \times K]$, różniczkowalna w przedziale $\langle t_0, T\rangle$,

$x = x(t)$ - szukany wektor K-wymiarowy.

Zadanie polega na takim doborze ciągu macierzy $\{B_n\}$ oraz ciągu liczb rzeczywistych $\{\gamma_n\}$, aby rozwiązanie schematu różnicowego postaci

$$\left.\begin{aligned} &y_{n+1} = \left[I + h(1-h\gamma_n)B_n\right]y_n \quad \left(n=0,1,\ldots,N=\frac{T-t_0}{h}-1\right), \\ &y_0 \text{ - dane,} \end{aligned}\right\} \tag{2.2}$$

aproksymującego z rzędem jeden zagadnienie (2.1), było zbieżne

do rozwiązania tego zagadnienia według normy euklidesowej oraz by uzyskane oszacowanie błędu było optymalne względem $\{\tau_n\}$.

2.2. Dobór macierzy B_n

Poniżej podane będą (lematy 1-3) warunki dostateczne na to, by rozwiązanie schematu różnicowego postaci (2.2) było zbieżne do rozwiązania zagadnienia (2.1), jak również zostanie wyznaczona postać macierzy B_n.

Lemat 1

Rozwiązanie schematu różnicowego

$$\left.\begin{aligned} \xi_{n+1} &= C_n \xi_n + \theta_n \qquad (n = 0,1,\ldots,N) \\ \xi_0 &= 0 \end{aligned}\right\} \tag{2.3}$$

gdzie: C_n - macierze $[K \times K]$-wymiarowe,
θ_n - dane wektory K-wymiarowe,
ma postać

$$\xi_n = \sum_{i=0}^{n-1} \left(\prod_{k=n-1}^{i+1} C_k \right) \theta_i , \tag{2.4}$$

przy czym

$$\prod_{k=n}^{n+1} C_k \overset{df}{=} 1.$$

D o w ó d (indukcyjny).

1) $\xi_1 = \theta_0$,

2) $\xi_{n+1} = \sum_{i=0}^{n} \left(\prod_{k=n}^{i+1} C_k \right) \theta_i = \sum_{i=0}^{n-1} \left(\prod_{k=n}^{i+1} C_k \right) \theta_i + \left(\prod_{k=n}^{n+1} C_k \right) \theta_n =$

$$= \sum_{i=0}^{n-1} C_n \left(\prod_{k=n-1}^{i+1} C_k \right) \theta_i + \theta_n = C_n \xi_n + \theta_n , \quad \text{cbdo.}$$

Lemat 2

Jeśli

$$Q_n = I + hA_n + O(h^2), \tag{2.5}$$

to rozwiązanie schematu różnicowego

$$\left.\begin{aligned} v_{n+1} &= Q_n v_n, \\ v_o &= x_o, \end{aligned}\right\} \tag{2.6}$$

jest zbieżne w przedziale $\langle t_o, T\rangle$ do rozwiązania zagadnienia (2.1), przy czym rząd zbieżności jest równy jeden.

D o w ó d

Jeśli x_n jest rozwiązaniem zagadnienia (2.1) w punkcie $t_n = t_o + nh$ $(n=0,1,\ldots,N)$, to

$$x_{n+1} = x_n + h\dot{x}_n + O(h^2) = (I + hA_n)x_n + O(h^2), \tag{2.7}$$

oraz

$$x_{n+1} - v_{n+1} = Q_n(x_n - v_n) + \left[I + hA_n - Q_n + O(h^2)\right]x_n. \tag{2.8}$$

Oznaczając

$$x_n - v_n \overset{df}{=} \eta_n,$$

otrzymuje się następujący schemat różnicowy ze względu na η_n:

$$\left.\begin{aligned} \eta_{n+1} &= Q_n \eta_n + \left[I + hA_n - Q_n + O(h^2)\right]x_n, \\ \eta_o &= 0. \end{aligned}\right\} \tag{2.9}$$

Rozwiązanie jego, zgodnie z lematem 1, ma postać

$$\eta_n = \sum_{i=0}^{n-1}\left(\prod_{k=n-1}^{i-1} Q_k\right)\left[I + hA_i - Q_i + O(h^2)\right]x_i, \tag{2.10}$$

a po uwzględnieniu (2.5)

$$\eta_n = \sum_{i=0}^{n-1}\left(\prod_{k=n-1}^{i+1} Q_k\right) O(h^2), \tag{2.11}$$

skąd wynika następujące oszacowanie normy

$$\|\eta_n\| \leq \sup_{\substack{0\leq i\leq n-1 \\ i+1\leq k\leq n-1}} \prod_{k=n-1}^{i+1} \|Q_k\| \sum_{i=0}^{n-1} O(h^2) = Ch^2 n \leq Ch^2\left(\frac{T-t_o}{n} - 1\right) = C_1 h,$$

z którego wynika teza lematu.

Lemat 3

Jeżeli $B_n = A_n + O(h)$ $(n=0,1,\ldots,N)$ oraz $y_0 = x_0$, to rozwiązanie schematu (2.2) jest zbieżne do rozwiązania zagadnienia (2.1) przy dowolnych wartościach γ_n.

Dowód wynika bezpośrednio z lematu 2, ponieważ

$$I + h(1 - h\gamma_n)B_n = I + hB_n - h^2\gamma_n B_n = I + hA_n + O(h^2).$$

Wprowadzając oznaczenie

$$A_{n+\frac{1}{2}} \stackrel{df}{=} A\left[t_0 + (n+\tfrac{1}{2})h\right], \tag{2.12}$$

otrzymuje się

$$A_{n+\frac{1}{2}} = A_n + O(h). \tag{2.13}$$

Wobec tego, zgodnie z lematem 3, podstawienie

$$B_n \stackrel{df}{=} A_{n+\frac{1}{2}}$$

powoduje, że rozwiązanie schematu (2.2) jest zbieżne, przy dowolnych wartościach γ_n, do rozwiązania zagadnienia (2.1). Taki właśnie wybór macierzy B_n podyktowany jest faktem, że otrzymane wówczas wyrażenie określające błąd aproksymacji ma formę dogodną do minimalizacji (według normy $\|\ \|$) względem γ_n. Dokładna analiza błędu podana jest poniżej.

2.3. Oszacowanie błędu aproksymacji

Jeżeli x_n oznacza rozwiązanie zagadnienia (2.1) w punkcie t_n, natomiast y_n - odpowiadające mu rozwiązanie schematu (2.2), to wartość błędu wyznacza zależność

$$z_n = x_n - y_n. \tag{2.14}$$

Aby uzyskać efektywną postać z_n oraz przeprowadzić jego oszacowanie, należy udowodnić następujący lemat:

Lemat 4

Rozwiązanie zagadnienia (2.1), ograniczone do węzłów siatki, spełnia równanie

$$x_{n-1} = \left[I + hA_{n+\frac{1}{2}} + \frac{h^2}{2} A^2_{n+\frac{1}{2}} \right] x_n + O(h^3). \qquad (2.15)$$

D o w ó d

Jeśli

$$\dot{x}(t) = A(t)x(t),$$

to

$$x_{n+1} = x_n + h\dot{x}_n + \frac{h^2}{2} \ddot{x}_n + O(h^3) = x_n + hA_n x_n +$$

$$+ \frac{h^2}{2} \left(\dot{A}_n + A_n^2 \right) x_n + O(h^3). \qquad (2.16)$$

Ponieważ

$$A_{n+\frac{1}{2}} = A_n + \frac{h}{2} \dot{A}_n + O(h^2),$$

więc wyznaczając stąd A_n i podstawiając do (2.16) otrzymuje się natychmiast wyrażenie (2.15) cbdo.

W dalszym ciągu pracy macierze $A_{n+\frac{1}{2}}$ będą, celem uproszczenia zapisu, oznaczane symbolem B_n. Tak więc równanie (2.15) przybiera postać

$$x_{n+1} = (I + hB_n)x_n + \frac{h^2}{2} B_n^2 x_n + O(h^3). \qquad (2.17)$$

Wartość błędu z_n, wyznaczona wzorem (2.14), spełnia wtedy równanie

$$z_{n+1} = x_{n+1} - y_{n+1} = \left[I + h(1-h\gamma_n)B_n \right] (x_n - y_n) +$$

$$+ \left[hB_n + \frac{h^2}{2} B_n^2 - h(1-h\gamma_n)B_n \right] x_n + O(h^3), \qquad (2.18)$$

a więc

$$z_{n+1} = \left[I + h(1-h\gamma_n)B_n \right] z_n + \frac{h^2}{2} B_n \left[B_n + 2\gamma_n I \right] x_n = O(h^3). \qquad (2.19)$$

Wprowadzając oznaczenie

$$R_n = B_n + 2\gamma_n I, \qquad (2.20)$$

oraz uwzględniając, że $z_0 = x_0 - y_0 = 0$, otrzymuje się następujący schemat różnicowy ze względu na z_n

$$\left.\begin{aligned} z_{n+1} &= \left[I + h(1-h\gamma_n)B_n\right]z_n + \frac{h^2}{2} B_n R_n x_n + O(h^3), \\ z_0 &= 0. \end{aligned}\right\} \quad (2.21)$$

Lemat 5

Rozwiązanie schematu (2.21) ma postać

$$z_n = \frac{1}{2} h^2 \sum_{i=0}^{n-1} \left(I + h \sum_{k=i+1}^{n-1} B_k \right) B_i R_i x_i + O(h^2). \quad (2.22)$$

D o w ó d

Uwzględniając lemat 1 otrzymuje się natychmiast

$$z_n = \sum_{i=0}^{n-1} \left[\prod_{k=n-1}^{i+1} \left(I + h(1-h\gamma_k)B_k\right)\right]\left[\frac{1}{2} h^2 B_i R_i x_i + O(h^3)\right] =$$

$$= \sum_{i=0}^{n-1} \left[\prod_{k=n-1}^{i+1} (I + hB_k)\right]\left[\frac{1}{2}h^2 B_i R_i x_i + O(h^3)\right] =$$

$$= \sum_{i=0}^{n-1} \left[I + h \sum_{k=i+1}^{n-1} B_k \right]\left[\frac{1}{2}h^2 B_i R_i x_i + O(h^3)\right] =$$

$$= \frac{1}{2}h^2 \sum_{i=0}^{n-1}\left[I + h \sum_{k=i+1}^{n-1} B_k \right] B_i R_i x_i + O(h^2), \text{ cbdo.}$$

Z równania (2.22) można otrzymać następujące oszacowanie błędu

$$\|z_n\| \leq \frac{1}{2}h^2 \sum_{i=0}^{n-1} \left\| I + h \sum_{k=i+1}^{n-1} B_k \right\| \cdot \|B_i\| \cdot \|R_i\| \cdot \|x_i\| + O(h^2). \quad (2.23)$$

Oszacowanie to będzie stanowiło punkt wyjściowy minimalizacji oszacowania błędu względem ciągu $\{\gamma_i\}$ $(i=0,1,\ldots,n-1)$, przeprowadzonej w rozdziale następnym.

3. Obliczenie wartości minimalnej oszacowania błędu aproksymacji

Jak wynika z postaci wzoru (2.23), oszacowanie błędu aproksymacji jest zależne, poprzez wyrażenia R_i, od ciągu γ_i

i przybiera wartość minimalną dla ciągu γ_{oi}, którego wyrazy spełniają związek

$$\min_{\gamma_i \in (-\infty,+\infty)} \| R_i(\gamma_i) \| = \| R_i(\gamma_{oi}) \| = \| B_i + 2\gamma_{oi} I \|$$

$$(i = 0,1,\ldots,n-1). \qquad (3.1)$$

Zadanie sprowadza się więc do znalezienia takiej wartości γ_o, która dla danej macierzy B minimalizuje normę $\| B + 2\gamma I \|$. Zostanie ono zrealizowane, poprzez lematy 6-9 oraz twierdzenie 1, w twierdzeniu 2.

Twierdzenie 1

Dla każdej macierzy B, wymiaru $[K \times K]$, istnieje taka nieosobliwa macierz S ($\det S \neq 0$), że

$$B = S\, J\, S^{-1}, \qquad (3.2)$$

gdzie:

$$J = \sum_{i=1}^{L} J_{\lambda_i} \stackrel{df}{=} \begin{bmatrix} J_{\lambda_1} & & & 0 \\ & J_{\lambda_2} & & \\ & & \ddots & \\ 0 & & & J_{\lambda_L} \end{bmatrix}, \qquad (3.3)$$

$$J_{\lambda_i} \stackrel{df}{=} \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ & & & \lambda_i & 1 \\ 0 & \cdots & & 0 & \lambda_i \end{bmatrix} \text{ - klatka Jordana rzędu } s_i \left(\sum_{i=1}^{L} s_i = K \right), \qquad (3.4)$$

λ_i - wartość własna macierzy B o krotności s_i.

Dowód tego twierdzenia można znaleźć w pracy [3] lub [4].

Lemat 6

Jeśli J oznacza macierz Jordana, określoną wzorem (3.3), to

$$\| J \| = \sup_{1 \leq i \leq L} \| J_{\lambda_i} \| . \tag{3.5}$$

D o w ó d

Z definicji normy macierzy wynika, że

$$\| J \| = \sup_{\| x \| \leq 1} \| Jx \| , \tag{3.6}$$

gdzie x oznacza dowolny wektor K-wymiarowy. Wektor ten można przedstawić w postaci

$$x = \left\{ x_i \right\} \quad (i=1,2,\ldots,L), \tag{3.7}$$

gdzie $x_i = \left\{ x_{ik} \right\}$ $(k=1,2,\ldots,s_i)$. (3.8)

Można wtedy zapisać, że

$$Jx = \left\{ J_{\lambda_i} x_i \right\} \quad (i=1,2,\ldots,L) \tag{3.9}$$

i

$$\| Jx \|^2 = \sum_{i=1}^{L} \| J_{\lambda_i} x_i \|^2 , \tag{3.10}$$

a uwzględniając (3.6) otrzymuje się

$$\| J \|^2 = \sup_{\sum_{i=1}^{L} \| x_i \|^2 \leq 1} \sum_{i=1}^{L} \| J_{\lambda_i} x_i \|^2 . \tag{3.11}$$

Jeśli teraz założyć, że

$$\sup_{1 \leq i \leq L} \| J_{\lambda_i} \| = \| J_{\lambda_{i_0}} \| , \tag{3.12}$$

to

$$\| J \|^2 \leq \sup_{\sum_{i=1}^{L} \| x_i \|^2 \leq 1} \sum_{i=1}^{L} \| J_{\lambda_i} \|^2 \cdot \| x_i \|^2 \leq \| J_{\lambda_{i_0}} \|^2 , \tag{3.13}$$

czyli

$$\| J \| \leq \sup_{1 \leq i \leq L} \| J_{\lambda_i} \| . \tag{3.14}$$

Z drugiej strony, biorąc

$$x_o^T = \left\{ 0,\ldots,0,x_{i_o},0,\ldots,0 \right\} \tag{3.15}$$

i uwzględniając (3.11), otrzymuje się

$$\| J \| \geqslant \sup_{\| x_{i_o} \| \leq 1} \| J_{\lambda_{i_o}} x_{i_o} \| = \| J_{\lambda_{i_o}} \| \tag{3.16}$$

Z nierówności (3.14) oraz (3.16) wynika równość (3.5), czyli teza lematu.

Lemat 7

Jeżeli s_i oznacza wymiar macierzy J_{λ_i}, to:

1) jeśli $s_i=1$, wówczas

$$\| J_{\lambda_i} \| = |\lambda_i|, \tag{3.17}$$

2) jeśli $s_i \geqslant 2$, wtedy

$$\left[1 + 2|\lambda_i| \left(1 - \frac{1}{s_i - 1} \right) + |\lambda_i|^2 \right]^{\frac{1}{2}} \leq \| J_{\lambda_i} \| \leq |\lambda_i| + 1. \tag{3.18}$$

D o w ó d

Słuszność 1) wynika wprost z definicji. Dla dowodu 2) trzeba oznaczyć

$$x_i = \left\{ x_{ik} \right\} \quad (k=1,2,\ldots,s_i). \tag{3.19}$$

Wtedy

$$J_{\lambda_i} x_i = \begin{bmatrix} \lambda_i x_{i1} + x_{i2} \\ \lambda_i x_{i2} + x_{i3} \\ \cdots\cdots\cdots \\ \cdots\cdots\cdots \\ \lambda_i x_{i(s_i-1)} + x_{is_i} \\ \lambda_i x_{is_i} \end{bmatrix} \tag{3.20}$$

oraz

$$\|J_{\lambda_i} x_i\|^2 = \sum_{k=1}^{s_i-1} |\lambda_i x_{ik} + x_{i(k+1)}|^2 + |\lambda_i|^2 \cdot |x_{is_i}|^2 =$$

$$= |\lambda_i|^2 \sum_{k=1}^{s_i} |x_{ik}|^2 + 2|\lambda_i| \sum_{k=1}^{s_i-1} |x_{ik}||x_{i(k+1)}| + \sum_{k=2}^{s_i} |x_{ik}|^2. \quad (3.21)$$

Ponieważ

$$2|x_{ik}||x_{i(k+1)}| \leq |x_{ik}|^2 + |x_{i(k+1)}|^2, \quad (3.22)$$

więc

$$\|J_{\lambda_i} x_i\|^2 \leq |\lambda_i|^2 \sum_{k=1}^{s_i} |x_{ik}|^2 + |\lambda_i| \left(\sum_{k=1}^{s_i-1} |x_{ik}|^2 + \sum_{k=2}^{s_i} |x_{ik}|^2 \right) +$$

$$+ \sum_{k=2}^{s_i} |x_{ik}|^2. \quad (3.23)$$

Wobec tego

$$\|J_{\lambda_i}\|^2 = \sup_{\|x_i\| \leq 1} \|J_{\lambda_i} x_i\|^2 \leq |\lambda_i|^2 + 2|\lambda_i| + 1 = (|\lambda_i| + 1)^2, \quad (3.24)$$

czyli

$$\|J_{\lambda_i}\| \leq |\lambda_i| + 1, \quad (3.25)$$

co stanowi prawą nierówność wyrażenia (3.18). Dla dowodu lewej nierówności oznaczono

$$x_{oi}^T = \left\{ 0, \frac{1}{\sqrt{s_i - 1}}, \ldots, \frac{1}{\sqrt{s_i - 1}} \right\}, \quad \|x_{oi}\| = 1. \quad (3.26)$$

Wtedy

$$\|J_{\lambda_i} x_{oi}\|^2 = |\lambda_i| \sum_{k=2}^{s_i} \left(\frac{1}{\sqrt{s_i - 1}} \right)^2 + 2|\lambda_i| \sum_{k=2}^{s_i} \left(\frac{1}{\sqrt{s_i - 1}} \right)^2 + \sum_{k=2}^{s_i} \left(\frac{1}{\sqrt{s_i - 1}} \right)^2 =$$

$$= |\lambda_i|^2 + 2|\lambda_i| \left(\frac{s_i - 2}{s_i - 1} \right) + 1 = |\lambda_i|^2 + 2|\lambda_i| \left(1 - \frac{1}{s_i - 1} \right) + 1 \quad (3.27)$$

i ostatecznie

$$\left\| J_{\lambda_i} \right\| = \sup_{\|x_i\|=1} \left\| J_{\lambda_i} x_i \right\| \geq \left\| J_{\lambda_i} x_{oi} \right\| = \left[|\lambda_i|^2 + 2|\lambda_i| \left(1 - \frac{1}{s_i - 1} \right) + 1 \right]^{\frac{1}{2}}.$$

Z lematu 7 wynika następujący wniosek

Wniosek

Jeżeli s_i oznacza wymiar macierzy J_{λ_i}, to

$$\lim_{s_i \to \infty} \left\| J_{\lambda_i} \right\| = |\lambda_i| + 1. \tag{3.28}$$

Na podstawie lematów 6 i 7 otrzymuje się natomiast dowód kolejnego lematu:

Lemat 8

Jeżeli

$$p_i \stackrel{df}{=} \begin{cases} |\lambda_i| & \text{dla } i : s_i = 1, \\ \left[|\lambda_i|^2 + 2|\lambda_i| \left(1 - \frac{1}{s_i - 1} \right) + 1 \right]^{\frac{1}{2}} & \text{dla } i : s_i \geq 2, \end{cases} \tag{3.29}$$

oraz

$$q_i \stackrel{df}{=} \begin{cases} |\lambda_i| & \text{dla } i : s_i = 1, \\ |\lambda_i| + 1 & \text{dla } i : s_i \geq 2, \end{cases} \tag{3.30}$$

to

$$\max_{1 \leq i \leq L} p_i \leq \| J \| \leq \max_{1 \leq i \leq L} q_i. \tag{3.31}$$

Lemat 9

Jeżeli J oraz J_{λ_i} są macierzami, określonymi wzorami (3.3) i (3.4), a

$$J(\delta) = \sum_{i=1}^{L} J_{(\lambda_i + \delta)}, \tag{3.32}$$

gdzie: δ - dowolna liczba rzeczywista (lub zespolona), to

$J + \delta I = J(\delta)$.

D o w ó d

Z definicji macierzy J wynika, że

$$J + \delta I = \sum_{i=1}^{L} J_{\lambda_i} + \sum_{i=1}^{L} \delta I_i = \sum_{i=1}^{L} (J_{\lambda_i} + \delta I_i), \quad (3.34)$$

gdzie I_i $(i=1,2,\ldots,L)$ - macierze jednostkowe, i-wymiarowe.

Ponieważ

$$J_{\lambda_i} + \delta I_i = J_{(\lambda_i+\delta)}, \quad (3.35)$$

więc

$$J + \delta I = \sum_{i=1}^{L} J_{(\lambda_i+\delta)} = J(\delta), \quad \text{cbdo.}$$

Z twierdzenia 1 oraz ostatniego lematu wynika, że

$$B + 2\gamma I = SJS^{-1} + 2\gamma I = S(J + 2\gamma I)S^{-1} = S\,J(2\gamma)S^{-1}. \quad (3.36)$$

Otrzymuje się stąd następujące oszacowanie

$$\|B + 2\gamma I\| \leq \|S\| \cdot \|J(2\gamma)\| \cdot \|S^{-1}\| = \|J(2\gamma)\| \cdot \|S^{-1}\|, \quad (3.37)$$

ponieważ można przyjąć, że $\|S\| = 1$.

Z postaci wyrażenia (3.37) widoczne jest, że poszukiwane liczby realizującej $\min_{\gamma \in (-\infty,+\infty)} \|B + 2\gamma I\|$ należy w rzeczywistości zastąpić poszukiwaniem wartości realizującej $\min_{\gamma \in (-\infty,+\infty)} \|J(2\gamma)\|$. Ponieważ, zgodnie z lematem 8,

$$\|J(2\gamma)\| \leq \max_{1 \leq i \leq L} q_i(\gamma), \quad (3.38)$$

$$\text{gdzie } q_i(\gamma) = \begin{cases} |\lambda_i + 2\gamma| & \text{dla } i : s_i = 1, \\ |\lambda_i + 2\gamma + 1| & \text{dla } i : s_i \geq 2, \end{cases} \quad (3.39)$$

więc zadanie sprowadza się ostatecznie do znalezienia liczby γ_0, dla której $\min_{\gamma \in (-\infty,+\infty)} \max_{1 \leq i \leq L} q_i(\gamma) = \max_{1 \leq i \leq L} q_i(\gamma_0)$.

Twierdzenie 2

Jeżeli γ_o oznacza liczbę taką, że

$$\min_{\gamma \in (-\infty,+\infty)} \max_{1 \le i \le l} q_i(\gamma) = \max_{1 \le i \le L} q_i(\gamma_o), \tag{3.40}$$

to istnieje liczba naturalna $i_o \le L$ taka, że zachodzi następująca równość:

$$\max \left[\max_{1 \le i \le l} q_i\left(-\frac{1}{2}\mathrm{Re}\lambda_i\right), \max_{\substack{-\frac{1}{2}\mathrm{Re}\lambda_i \le \gamma_{ij} \le -\frac{1}{2}\mathrm{Re}\lambda_j \\ (i,j=1,\ldots,L)}} q_i(\gamma_{ij}) = q_{i_o}(\gamma_o) \right], \tag{3.41}$$

gdzie $\gamma_{ij} = \left\{ \gamma : q_i(\gamma) = q_j(\gamma) \right\}$. (3.42)

D o w ó d

Ponieważ

$$|\lambda_i + 2\gamma| = \left[(2\gamma + \mathrm{Re}\lambda_i)^2 + (\mathrm{Im}\lambda_i)^2 \right]^{\frac{1}{2}} \tag{3.43}$$

więc uwzględniając definicję (3.39) otrzymuje się

$$\frac{dq_i(\gamma)}{d\gamma} = \frac{2\gamma + \mathrm{Re}\lambda_i}{|\lambda_i + 2\gamma|}, \quad \text{dla } |\lambda_i + 2\gamma| \neq 0. \tag{3.44}$$

Wynika stąd, że:

1) jeśli $\gamma < -\frac{1}{2}\mathrm{Re}\lambda_i$, wówczas $\frac{dq_i}{d\gamma} < 0$, czyli $q_i(\gamma)$ - funkcja malejąca,

2) jeśli $\gamma > -\frac{1}{2}\mathrm{Re}\lambda_i$, wówczas $\frac{dq_i}{d\gamma} > 0$, czyli $q_i(\gamma)$ - funkcja rosnąca.

Wobec tego dla każdego $i=1,2,\ldots,L$ prawdziwa jest równość

$$q_i\left(-\frac{1}{2}\mathrm{Re}\lambda_i\right) = \min q_i(\gamma). \tag{3.45}$$

W dalszym ciągu dowodu rozpatrzone będą dwa możliwe, wzajemnie wyłączające się przypadki:

a) $$\max_{-\frac{1}{2}\mathrm{Re}\lambda_i \le \gamma_{ij} \le -\frac{1}{2}\mathrm{Re}\lambda_j} q_i(\gamma_{ij}) \le \max_{1 \le i \le L} q_i\left(-\frac{1}{2}\mathrm{Re}\lambda_i\right) \overset{df}{=} q_{i_o}(\gamma_o), \tag{3.46}$$

b) $$\max_{1\le i\le L} q_i\left(-\tfrac{1}{2}\mathrm{Re}\lambda_i\right) < \max_{-\frac{1}{2}\mathrm{Re}\lambda_i \le \gamma_{ij} \le -\frac{1}{2}\mathrm{Re}\lambda_j} q_i(\gamma_{ij}) =$$

$$= q_{i_o}(\gamma_{i_o j_o}) \stackrel{df}{=} q_{i_o}(\gamma_o), \tag{3.47}$$

i pokazane zostanie, że tak określona liczba γ_o spełnia równanie (3.40).

Biorąc pod uwagę dowolną liczbę $\gamma \neq \gamma_o$, otrzymuje się w przypadku a)

$$q_{i_o}(\gamma) > q_{i_o}\left(-\tfrac{1}{2}\mathrm{Re}\lambda_{i_o}\right) = q_{i_o}(\gamma_o), \tag{3.48}$$

w przypadku b)

$$\left.\begin{aligned} &q_{i_o}(\gamma) > q_{i_o}(\gamma_{i_o j_o}), \quad \text{jeśli } \gamma < \gamma_{i_o j_o}, \\ \text{lub} \\ &q_{j_o}(\gamma) > q_{i_o}(\gamma_{i_o j_o}), \quad \text{jeśli } \gamma > \gamma_{i_o j_o}. \end{aligned}\right\} \tag{3.49}$$

Dowodzi to, że

$$\min_{\gamma\in(-\infty,+\infty)} \max_{1\le i\le L} q_i(\gamma) \ge q_{i_o}(\gamma_o). \tag{3.50}$$

Dla dowodu nierówności przeciwnej wystarczy założyć istnienie takiego k (k=1,...,L), dla którego spełniona jest nierówność

$$q_k(\gamma_o) > q_{i_o}(\gamma_o). \tag{3.51}$$

Uwzględniając własności funkcji $q_i(\gamma)$, wynikające ze wzorów (3.44) i (3.45), otrzymuje się wtedy:
dla przypadku a)

$$q_k\left(-\tfrac{1}{2}\mathrm{Re}\lambda_k\right) \le q_{i_o}\left(-\tfrac{1}{2}\mathrm{Re}\lambda_{i_o}\right) = q_{i_o}(\gamma_o), \tag{3.52}$$

a więc istnieje taka liczba $\gamma_{i_o k}$, że

$$\min\left(-\tfrac{1}{2}\mathrm{Re}\lambda_k, -\tfrac{1}{2}\mathrm{Re}\lambda_{i_o}\right) \le \gamma_{i_o k} \le \left(\max\left(-\tfrac{1}{2}\mathrm{Re}\lambda_k, -\tfrac{1}{2}\mathrm{Re}\lambda_{i_o}\right), \tag{3.53}$$

oraz

$$q_k(\gamma_{i_o k}) = q_{i_o}(\gamma_{i_o k}) > q_{i_o}(\gamma_o), \tag{3.54}$$

co jest sprzeczne z założeniem (3.51),
dla przypadku b)

$$q_k\left(-\tfrac{1}{2}\mathrm{Re}\lambda_k\right) \leq q_{i_o}(\gamma_{i_o j_o}) = q_{j_o}(\gamma_{i_o j_o}), \tag{3.55}$$

oraz odpowiednio

1) jeśli $-\frac{1}{2}\mathrm{Re}\lambda_k > -\frac{1}{2}\mathrm{Re}\lambda_{i_o}$, to istnieje liczba $\gamma_{i_o k}$
$\left(-\frac{1}{2}\mathrm{Re}\lambda_{i_o} \leq \gamma_{i_o k} \leq -\frac{1}{2}\mathrm{Re}\lambda_k\right)$ taka, że

$$q_{i_o}(\gamma_{i_o k}) > q_{i_o}(\gamma_o), \tag{3.56}$$

2) jeśli $-\frac{1}{2}\mathrm{Re}\lambda_k < -\frac{1}{2}\mathrm{Re}\lambda_{i_o}$, to istnieje liczba $\gamma_{j_o k}$
$\left(-\frac{1}{2}\mathrm{Re}\lambda_k < \gamma_{j_o k} \leq -\frac{1}{2}\mathrm{Re}\lambda_{j_o}\right)$ taka, że

$$q_{j_o}(\gamma_{j_o k}) > q_{j_o}(\gamma_o) = q_{i_o}(\gamma_o), \tag{3.57}$$

co znowu jest sprzeczne z założeniem (3.51). Wobec tego dla każdego $i = 1,2,\ldots,L$ musi być spełniona nierówność

$$q_i(\gamma_o) \leq q_{i_o}(\gamma_o), \tag{3.58}$$

a więc również

$$\min_{\gamma \in (\infty,+\infty)} \max_{1 \leq i \leq L} q_i(\gamma) \leq q_{i_o}(\gamma_o). \tag{3.59}$$

Biorąc teraz pod uwagę nierówność (3.50) otrzymuje się bezpośrednio równanie (3.40), co kończy dowód twierdzenia.

Powracając do rozważanego schematu różnicowego można teraz stwierdzić, że oszacowanie (2.23) normy błędu aproksymacji będzie najlepsze, jeżeli wyrazy ciągu $\{\gamma_{on}\}$ zostaną, dla każdego n, określone zgodnie z twierdzeniem 2.

Z twierdzenia tego wynikają dwa następujące wnioski:

Wniosek 1

Jeżeli macierz B ma wartości własne rzeczywiste i jednokrotne wówczas, jak wynika z lematu 8 oraz nierówności (3.38)

$$\min_{\gamma\in(-\infty,+\infty)} \|J(2\gamma)\| = \min_{\gamma\in(-\infty,+\infty)} \max_{1\le i\le K} q_i(\gamma) = \max_{1\le i,j\le K} q_i(\gamma_{ij}), \tag{3.60}$$

gdzie γ_{ij} jest pierwiastkiem równania

$$|2\gamma + \lambda_i| = |2\gamma + \lambda_j|, \tag{3.61}$$

rozwiązanie którego ma postać

$$\gamma_{ij} = -\frac{1}{4}(\lambda_i + \lambda_j). \tag{3.62}$$

Wtedy

$$|2\gamma_{ij} + \lambda_i| = \frac{1}{2}|\lambda_i - \lambda_j|, \tag{3.63}$$

a więc

$$\max_{1\le i,j\le K} q_i(\gamma_{ij}) = \max_{1\le i,j\le K} |2\gamma_{ij} + \lambda_i| = \frac{1}{2}\left(\max_{1\le i\le K}\lambda_i - \min_{1\le i\le K}\lambda_i\right) \tag{3.64}$$

czyli poszukiwana wartość γ_0 jest wyznaczona wzorem

$$\gamma_0 = -\frac{1}{4}\left(\min_{1\le i\le K}\lambda_i + \max_{1\le i\le K}\lambda_i\right). \tag{3.65}$$

Wniosek 2

Jeśli istnieje taka liczba naturalna $k\le L$ oraz liczba dodatnia ε, że:

1) $$|\mathrm{Re}\lambda_k|^{-1} < \varepsilon \tag{3.66}$$

2) $$\frac{|\mathrm{Re}\lambda_i|}{|\mathrm{Re}\lambda_k|} < \varepsilon \quad (i=1,\ldots,L),\ (i\neq k), \tag{3.67}$$

3) $$\frac{|\mathrm{Im}\lambda_i|}{|\mathrm{Re}\lambda_k|} < \varepsilon \quad (i=1,\ldots,L), \tag{3.68}$$

to wartość γ_0, określona równaniem (3.40), wyraża się wzorem

$$\gamma_0 = -\frac{1}{4}\mathrm{Re}\lambda_k(1+r), \tag{3.69}$$

gdzie $|r| < 2\varepsilon + O(\varepsilon^2)$. (3.70)

Słuszność tego wniosku wynika z faktu, że przy przyjętych założeniach

$$\min_{\gamma\in(-\infty,+\infty)} \max_{1\le i\le L} q_i(\gamma) = \max_{i:\left[\min\left(-\frac{1}{2}Re\lambda_k, -\frac{1}{2}Re\lambda_i\right) < \gamma_{ki} < \max\left(-\frac{1}{2}Re\lambda_k, -\frac{1}{2}Re\lambda_i\right)\right]} q_k(\gamma_{ki}) \tag{3.71}$$

gdzie γ_{ki} jest pierwiastkiem równania

$$q_k(\gamma) = q_i(\gamma), \tag{3.72}$$

które przybiera jedną z następujących postaci:

a) jeśli $s_k = s_i = 1$ lub $s_k \geq 2$ i $s_i \geq 2$, to

$$\left[(2\gamma + \mathrm{Re}\lambda_k)^2 + (\mathrm{Im}\lambda_k)^2\right]^{\frac{1}{2}} = \left[(2\gamma + \mathrm{Re}\lambda_i)^2 + (\mathrm{Im}\lambda_i)^2\right]^{\frac{1}{2}} \tag{3.73}$$

b) jeśli $s_k \geq 2$ oraz $s_i = 1$, to

$$\left[(2\gamma + \mathrm{Re}\lambda_k)^2 + (\mathrm{Im}\lambda_k)^2\right]^{\frac{1}{2}} + 1 = \left[(2\gamma + \mathrm{Re}\lambda_i)^2 + (\mathrm{Im}\lambda_i)^2\right]^{\frac{1}{2}} \tag{3.74}$$

(jeśli $s_k=1$ a $s_i \geq 2$ to postać równania różni się od (3.73) przestawieniem wskaźników).

Rozwiązując równanie (3.71) otrzymuje się odpowiednio: w przypadku a)

$$\gamma_{ki} = -\frac{1}{4}\mathrm{Re}\lambda_k\left[1 + \frac{\mathrm{Re}\lambda_i}{\mathrm{Re}\lambda_k} + \frac{\left(\frac{\mathrm{Im}\lambda_k}{\mathrm{Re}\lambda_k}\right)^2 - \left(\frac{\mathrm{Im}\lambda_i}{\mathrm{Re}\lambda_k}\right)^2}{1 - \frac{\mathrm{Re}\lambda_i}{\mathrm{Re}\lambda_k}}\right] = -\frac{1}{4}\mathrm{Re}\lambda_k(1+r), \tag{3.75}$$

gdzie

$$|r| = \left|\frac{\mathrm{Re}\lambda_i}{\mathrm{Re}\lambda_k} + \frac{\left(\frac{\mathrm{Im}\lambda_k}{\mathrm{Re}\lambda_k}\right)^2 - \left(\frac{\mathrm{Im}\lambda_i}{\mathrm{Re}\lambda_k}\right)^2}{1 - \frac{\mathrm{Re}\lambda_i}{\mathrm{Re}\lambda_k}}\right| \leq \left|\frac{\left(\mathrm{Re}\lambda_i\right)}{\mathrm{Re}\lambda_k}\right| +$$

$$+\frac{\max\left(\left|\frac{\mathrm{Im}\lambda_k}{\mathrm{Re}\lambda_k}\right|^2, \left|\frac{\mathrm{Im}\lambda_i}{\mathrm{Re}\lambda_k}\right|^2\right)}{\left|1-\frac{\mathrm{Re}\lambda_i}{\mathrm{Re}\lambda_k}\right|} \leqslant \varepsilon + \frac{\varepsilon^2}{1-\varepsilon} = \varepsilon + O(\varepsilon^2), \tag{3.76}$$

w przypadku b)

$$\gamma = -\frac{1}{4}\mathrm{Re}\lambda_k\left[1 + \frac{\mathrm{Re}\lambda_i}{\mathrm{Re}\lambda_k} + \frac{\left(\frac{\mathrm{Im}\lambda_k}{\mathrm{Re}\lambda_k}\right)^2 - \left(\frac{\mathrm{Im}\lambda_i}{\mathrm{Re}\lambda_k}\right)^2 - \left(\frac{1}{\mathrm{Re}\lambda_k}\right)^2}{1-\frac{\mathrm{Re}\lambda_i}{\mathrm{Re}\lambda_k}} + \right.$$

$$\left. + \frac{2}{\mathrm{Re}\lambda_k}\frac{\left[\left(\frac{2\gamma+\mathrm{Re}\lambda_i}{\mathrm{Re}\lambda_k}\right)^2 + \left(\frac{\mathrm{Im}\lambda_i}{\mathrm{Re}\lambda_k}\right)^2\right]^{\frac{1}{2}}}{1-\frac{\mathrm{Re}\lambda_i}{\mathrm{Re}\lambda_k}}\right], \tag{3.77}$$

a oznaczając:

$$a = \frac{\mathrm{Re}\lambda_i}{\mathrm{Re}\lambda_k}, \quad |a| < \varepsilon \tag{3.78}$$

i uwzględniając założenia (3.66)÷(3.68),

$$\gamma = -\frac{1}{4}\mathrm{Re}\lambda_k\left[1 + a + 4\gamma(\mathrm{Re}\lambda_k)^{-1} + O(\varepsilon^2)\right] \tag{3.79}$$

lub

$$\gamma\left[1 + (\mathrm{Re}\lambda_k)^{-1}\right] = -\frac{1}{4}\mathrm{Re}\lambda_k\left[1 + a + O(\varepsilon^2)\right], \tag{3.80}$$

skąd

$$\gamma = \gamma_{ki} = -\frac{1}{4}\mathrm{Re}\lambda_k\left[1 + a - (\mathrm{Re}\lambda_k)^{-1} + O(\varepsilon^2)\right]. \tag{3.81}$$

Oznaczając z kolei

$$r = a - (\mathrm{Re}\lambda_k)^{-1} + O(\varepsilon^2), \tag{3.82}$$

otrzymuje się

$$\gamma_{ki} = -\frac{1}{4}\mathrm{Re}\lambda_k(1 + r), \tag{3.83}$$

gdzie $|r| \leqslant |a| + |\mathrm{Re}\lambda_k|^{-1} + O(\varepsilon^2) < 2\varepsilon + O(\varepsilon^2)$. (3.84)

4. Przybliżona metoda minimalizacji oszacowania błędu aproksymacji w przypadku, gdy nie są znane wartości własne macierzy B_n

Obecnie przedstawiona zostanie metoda przybliżonego wyznaczenia ciągu $\{\gamma_{on}\}$ w przypadku, gdy nie są znane wartości własne macierzy B_n, wiadomo natomiast, że spełniają one założenia wniosku 1. lub 2.

Lemat 10

Dla dowolnej macierzy kwadratowej $B = \{b_{ij}\}$ $(i,j = 1,2,\ldots,K)$ istnieje wskaźnik i_o taki, że każda wartość własna macierzy B spełnia nierówność

$$|\lambda - b_{i_o i_o}| \leq \sum_{i_o \neq j=1}^{K} |b_{i_o j}| . \tag{4.1}$$

Dowód można znaleźć w pracy [5].

Lemat 11

Jeżeli macierz B jest rzeczywista, to każda jej wartość własna spełnia nierówność

$$\min_i \left(b_{ii} - \sum_{i \neq j=1}^{K} |b_{ij}|\right) \leq \mathrm{Re}\,\lambda \leq \max \left(b_{ii} + \sum_{i \neq j=1}^{K} |b_{ij}|\right) \tag{4.2}$$

D o w ó d

Z nierówności (4.1) wynika, że

$$|\mathrm{Re}\lambda - b_{i_o i_o}| \leq \left[(\mathrm{Re}\lambda - b_{i_o i_o})^2 + (\mathrm{Im}\lambda)^2\right]^{\frac{1}{2}} =$$

$$= |\lambda - b_{i_o i_o}| \leq \sum_{i_o \neq j=1}^{K} |b_{i_o j}| \tag{4.3}$$

czyli

$$b_{i_o i_o} - \sum_{i_o \neq j=1}^{K} |b_{i_o j}| \leq \mathrm{Re}\,\lambda \leq b_{i_o i_o} + \sum_{i_o \neq j=1}^{K} |b_{i_o j}| \tag{4.4}$$

a więc spełniona jest nierówność (4.2), cbdo.

Z lematu 11 wynika wniosek następujący:

Wniosek 3

Jeżeli spełnione są założenia wniosku 1. lub 2. oraz zachodzi jedna z nierówności:

$$\frac{\left|\max_i\left(b_{nii}+\sum_{i\neq j=1}^{K}\left|b_{nij}\right|\right)\right|}{\left|\min_i\left(b_{nii}+\sum_{i\neq j=1}^{K}\left|b_{nij}\right|\right)\right|}<\varepsilon(n=0,1,\ldots,N), \tag{4.5}$$

lub

$$\frac{\left|\min_i\left(b_{nii}+\sum_{i\neq j=1}^{K}\left|b_{nij}\right|\right)\right|}{\left|\max_i\left(b_{nii}+\sum_{i\neq j=1}^{K}\left|b_{nij}\right|\right)\right|}<\varepsilon(n=0,1,\ldots,N), \tag{4.6}$$

gdzie jest znacznie mniejsze od jedności, wówczas celowe jest zastosowanie do aproksymacji zagadnienia (2.1) schematu różnicowego postaci (2.2) przy założeniu, że ciąg $\{\tau_n\}$ jest określony następująco

$$\tau_n = \tau_{on} = -\frac{1}{4}\left[\min_i\left(b_{nii}-\sum_{i\neq j=1}^{K}|b_{nij}|\right)+\max_i\left(b_{nii}+\sum_{i\neq j=1}^{K}|b_{nij}|\right)\right] \tag{4.7}$$

gdzie $B_n=\{b_{nij}\}$ $(i,j=1,2,\ldots,K;\ n=0,1,\ldots,N)$. (4.8)

5. Ocena przydatności opisanej metody doboru schematu różnicowego

5.1. Uwagi ogólne

Z przeprowadzonej dotychczas analizy wynika w sposób oczywisty odpowiedź na pytanie: w jakich przypadkach uzasadnione jest stosowanie (do przybliżonego rozwiązania zagadnienia (2.1)) schematu różnicowego o postaci (2.2) w miejsce schematu postaci

$$\left.\begin{aligned} y_{n+1} &= (I+hB_n)y_n,\\ y_o &= x_o. \end{aligned}\right\} \tag{5.1}$$

Widoczne jest mianowicie, że jeśli spełnione są założenia (3.66)-(3.68) wniosku 2. wówczas określając ciąg $\{\gamma_n\}$ następująco

$$\gamma_n = -\frac{1}{4}\,\mathrm{Re}\lambda_{kn} \quad (n=0,1,\ldots,N), \tag{5.2}$$

oraz uwzględniając (3.39), otrzymuje się

$$q_i\left(-\frac{1}{4}\,\mathrm{Re}\lambda_{kn}\right) = \begin{cases} \left|\lambda_{in} - \frac{1}{2}\,\mathrm{Re}\lambda_{kn}\right| & \text{jeśli } s_i=1 \\ \left|\lambda_{in} - \frac{1}{2}\,\mathrm{Re}\lambda_{kn}\right|+1 & \text{jeśli } s_i\geq 2 \end{cases} \quad \begin{matrix}(i=1,\ldots,L)\\(i\neq k)\end{matrix} \tag{5.3}$$

czyli

$$q_i\left(-\frac{1}{4}\mathrm{Re}\lambda_{kn}\right) \leq \begin{cases} \frac{1}{2}\left|\mathrm{Re}\lambda_{kn}\right|\left(1+\frac{2\left|\lambda_{in}\right|}{\left|\mathrm{Re}\lambda_{kn}\right|}\right), & \text{gdy } s_i=1 \\ & (i\neq k), \\ \frac{1}{2}\left|\mathrm{Re}\lambda_{kn}\right|\left(1+\frac{2\left|\lambda_{in}\right|}{\left|\mathrm{Re}\lambda_{kn}\right|}+\frac{2}{\left|\mathrm{Re}\lambda_{kn}\right|}\right), & \text{gdy } s_i\geq 2 \end{cases} \tag{5.4}$$

a więc, jeśli $i\neq k$, to

$$q_i\left(-\frac{1}{4}\mathrm{Re}\lambda_{kn}\right) \leq \frac{1}{2}\left|\mathrm{Re}\lambda_{kn}\right|(2\sqrt{2}\,\varepsilon + 2\varepsilon + 1) = \frac{1}{2}\left|\mathrm{Re}\lambda_{kn}\right|\left[1+2(\sqrt{2}+1)\varepsilon\right] \tag{5.5}$$

natomiast dla $i=k$

$$q_k\left(-\frac{1}{4}\mathrm{Re}\lambda_{kn}\right) \leq \frac{1}{2}\left|\mathrm{Re}\lambda_{kn}\right|\left(1+2\frac{\left|\mathrm{Im}\lambda_{kn}\right|+1}{\left|\mathrm{Re}\lambda_{kn}\right|}\right) \leq \frac{1}{2}\left|\mathrm{Re}\lambda_{kn}\right|(1+4\varepsilon). \tag{5.6}$$

Ostatecznie więc

$$\max_{1\leq i\leq L} q_i\left(-\frac{1}{4}\,\mathrm{Re}\lambda_{kn}\right) < \frac{1}{2}\,\mathrm{Re}\lambda_{kn}\left[1+2(\sqrt{2}+1)\varepsilon\right] \tag{5.7}$$

Jeżeli zaś założyć, że $\gamma_n = 0$ $(n=0,1,\ldots,N)$, co ma miejsce w przypadku zastosowania schematu (5.1), wówczas:

$$q_i(0) = \begin{cases} |\lambda_{in}| & \text{gdy } s_i=1, \\ |\lambda_{in}| + 1 & \text{gdy } s_i\geq 2, \end{cases} \tag{5.8}$$

a więc

$$\max_{1\le i\le L} q_i(0) \le |\lambda_{kn}| + 1 = |\mathrm{Re}\lambda_{kn}| \left\{ \left[1 + \left(\frac{\mathrm{Im}\lambda_{kn}}{\mathrm{Re}\lambda_{kn}} \right)^2 \right] + \frac{1}{\mathrm{Re}\lambda_{kn}} \right| \quad (5.9)$$

czyli

$$|\mathrm{Re}\lambda_{kn}| \le \max_{1\le i\le L} q_i(0) < |\mathrm{Re}\lambda_{kn}| \left[\sqrt{1+\varepsilon^2} + \varepsilon \right] \le |\mathrm{Re}\lambda_{kn}| (1+2\varepsilon). \quad (5.10)$$

Biorąc teraz pod uwagę wyrażenie (2.23) oraz nierówność (3.1), łatwo można zauważyć, że w pierwszym przypadku ($\tau_n = -\frac{1}{4}\mathrm{Re}\lambda_{kn}$) wielkość oszacowania błędu rozwiązania przybliżonego jest około dwukrotnie mniejsza niż w drugim ($\tau_n = 0$). Jedyną trudność, związaną z zastosowaniem schematu (2.2) stanowi fakt, że konieczna jest wtedy znajomość wartości własnych macierzy B_n, a właściwie tej wartości własnej, dla której zachodzi równość

$$\max_i |\mathrm{Re}\lambda_{in}| = |\mathrm{Re}\lambda_{kn}| \quad (i=1,\ldots,L;\ n=0,1,\ldots,N). \quad (5.11)$$

Trudność ta nie wystąpi, gdy ciąg τ_n zostanie określony według równania (4.8), przy spełnionych założeniach wniosku 3. Nie można jednak wtedy dokładnie określić, jak zmieni się wielkość oszacowania błędu.

5.2. Zastosowanie do równań kinetyki

Ponieważ wartości własne macierzy $P(t)$ są dla każdego t rzeczywiste, więc ciąg $\{\tau_{on}\}$ określony jest wzorem

$$\tau_{on} = -\frac{1}{4}\left(\omega_{n(\max)} + \omega_{n(\min)}\right). \quad (5.12)$$

gdzie $\omega_{n(\max)}$ i $\omega_{n(\min)}$ są ekstremalnymi wartościami własnymi macierzy $P(t_n)$. Są więc one pierwiastkami równania [1]

$$\det\left[\omega\Lambda + \sum_{j=1}^{G} \frac{\omega}{\omega+\lambda_j} B_j - R(t_n)\right] = 0. \quad (5.13)$$

W przypadku, gdy norma macierzy $R(t_n)$ jest dużo mniejsza od jedności można, celem uproszczenia rachunków, zastosować

wzór przybliżony

$$\gamma_{on} = \gamma_o - \frac{1}{4}\omega_{o(min)}, \tag{5.14}$$

gdzie $\omega_{o(min)}$ jest najmniejszym pierwiastkiem równania

$$\det\left[\omega\Lambda + \sum_{j=1}^{6} \frac{\omega}{\omega + \lambda_j} B_j\right] = 0. \tag{5.15}$$

Bibliografia

1 Porsching T.A.: On the spectrum of a matrix, arising from a problem in reactor kinetics. SIAM J.Appl. Math. 1968, T.16, nr 2.

2 Ash M.: Nuclear reactor kinetics. McGraw-Hill Book Company. New York 1965.

3 Pontriagin L.S.: Równania różniczkowe zwyczajne. PWN. Warszawa 1964.

4 Birkhoff, Garrett, McLane S.: A survey of modern algebra. The McMillan Company. New York 1953.

5 Varga R.S.: Metrix iterative analysis. Prentice-Hall, Inc. Englewood Cliffs. New Jersey 1962.

РАЗНОСТНЫЙ МЕТОД РЕШЕНИЯ КИНЕТИКИ ЯДЕРНОГО-РЕАКТОРА

К р а т к о е с о д е р ж а н и е

Определено разностную схему первого порядка для решения системы обыкновенных дифференциальных уравнений. Указано, что эта система лучше других по отношении к минимальной оценке по Эвклидовой норме. Приведено несколько примеров. Более подробно рассмотрена система уравнений кинетики ядерного реактора.

A FINITE DIFFERENCE SOLUTION OF NUCLEAR REACTOR KINETICS EQUATIONS

S u m m a r y

A finite difference first order integration formula for a set of ordinary diffirential equations has been developed. It has been shown that this formula gives better estimation of an error in the meaning of an Euclidean norm than those already known. The proposed method has been illustrated by examples. A special attention has been paid to nuclear reactor kinetics equations.

Rękopis dostarczono w listopadzie 1971 r.