

S. M. RUMP

Exakte Fehlerschranken für Eigenwerte und Eigenvektoren

1. Einleitung

Wie in der Einleitung von [2] bereits deutlich gemacht, können auf Rechenanlagen durch Rundungsfehler große Ungenauigkeiten entstehen. Dies kann sich insbesondere schwerwiegend auswirken, wenn die Ergebnisse nicht als falsch erkannt werden und mit ihnen weiter gerechnet wird. So kann es passieren, daß mittels eines Gleitkommaalgorithmus als Näherung für einen Eigenwert einer 5×5 -Matrix (!) der Wert $+0.2$ berechnet wird, wohingegen der wahre Wert -0.3 ist.

In der vorliegenden Arbeit werden Algorithmen vorgestellt, die bewiesene Fehlerschranken für Eigenwert/Eigenvektorpaare berechnen. Da die erhaltenen Ergebnisse bewiesenermaßen richtig sind, entfällt der bis jetzt notwendige Mehraufwand des Benutzers für Kontrollen im Algorithmus und am Ergebnis, durch die letztlich doch keine Sicherheit gegeben wurde. Der Zeitaufwand der vorliegenden Algorithmen liegt in der Größenordnung von Gleitkommaalgorithmen. Sie können von jedermann ohne tiefere Kenntnis der Problematik benutzt und die Ergebnisse, da sie bewiesenermaßen richtig sind, bedenkenlos weiter verwendet werden.

2. Theoretischer Hintergrund

Das Problem sei als nicht-lineares Gleichungssystem

$$\begin{aligned} (A - \lambda E) x &= 0 \\ e_k \cdot x - 1 &= 0 \end{aligned} \quad \text{mit } A \in M_n \mathbb{R}, \quad x \in V_n \mathbb{R} \quad \text{und } \lambda \in \mathbb{R} \quad (1)$$

formuliert, wobei e_k der transponierte k -te Einheitsvektor bedeutet. Wir definieren weiter die Funktion

$$f \begin{pmatrix} x \\ \lambda \end{pmatrix} := \begin{pmatrix} x \\ \lambda \end{pmatrix} - R \cdot \begin{pmatrix} (A - \lambda E)x \\ e_k x - 1 \end{pmatrix}, \quad R \in M_{n+1} \mathbb{R} \tag{2}$$

Durch TAYLOR-Entwicklung von f unter Einbeziehung der zweiten Ableitung erhält man

$$f \begin{pmatrix} x \\ \lambda \end{pmatrix} = f \begin{pmatrix} \tilde{x} \\ \tilde{\lambda} \end{pmatrix} + \left\{ E - R \cdot \begin{pmatrix} A - \tilde{\lambda} E - x \\ e_k \end{pmatrix} \right\} \cdot \begin{pmatrix} x - \tilde{x} \\ \lambda - \tilde{\lambda} \end{pmatrix} \tag{3}$$

für beliebiges $\tilde{x} \in V_n \mathbb{R}$ und $\tilde{\lambda} \in \mathbb{R}$. Die Gleichheit von (2) und (3) bestätigt man durch Ausrechnen. Für nicht-singuläres R ist offenbar jeder Fixpunkt von f ein Eigenvektor/Eigenwertpaar. Nach dem Fixpunktsatz von BROUWER liegt für

$$f \begin{pmatrix} X \\ A \end{pmatrix} \subseteq \begin{pmatrix} X \\ A \end{pmatrix} \quad X \in V_n \mathbb{R}, A \in \mathbb{R} \tag{4}$$

in X ein Eigenvektor und in A der zugehörige Eigenwert von A . Das Erfülltsein von (4) ist durch intervallmäßiges Ausrechnen von f nach (3) nachzuprüfen. Sowohl die Frage nach der (notwendigen) Nicht-Singularität von R als auch nach der Eindeutigkeit von Eigenvektor und Eigenwert muß noch beantwortet werden. Dazu geben wir folgende Definition.

Definition 1: Für $I, J \in \mathbb{R}$ heißt I *echt enthalten in* J , im Zeichen $I \subsetneq J$, wenn gilt:

$$I \subseteq J \quad \text{und} \quad I \cap \partial J = \emptyset.$$

Die entsprechende Definition für Intervallvektoren gilt komponentenweise. Der folgende Satz entscheidet das erste Problem positiv.

Satz 2: Sei $C \in M_n \mathbb{R}$, $z \in V_n \mathbb{R}$ und $V \in V_n \mathbb{R}$. Dann folgt aus

$$z + C \cdot V \subsetneq V, \tag{5}$$

daß der Spektralradius von C kleiner als 1 ist.

Beweis: Nach dem Fixpunktsatz von BROUWER existiert ein $\hat{x} \in V_n \mathbb{R}$ mit $z + C\hat{x} = \hat{x} \in V$. Aus (5) folgt dann

$$C \cdot (V - \hat{x}) = C \cdot V - C\hat{x} = C \cdot V + z - \hat{x} \subsetneq V - \hat{x}. \tag{6}$$

Für $Y := V - \hat{x}$ ist also $C \cdot Y \subsetneq Y$ und mit $W := Y + i \cdot Y$ gilt

$$C \cdot W = C(Y + i \cdot Y) = CY + i \cdot CY \subsetneq Y + iY = W. \tag{7}$$

Darüber hinaus ist $U_\varepsilon(0) \subseteq W$ für ein $\varepsilon > 0$ wegen des echten Enthaltenseins in (7).

Für einen beliebigen Eigenvektor w von C mit zugehörigem Eigenwert μ definieren wir $T := \{\gamma \in \mathbb{C} \mid \gamma w \in W\}$ und $|\gamma^*| = \max_{\gamma \in T} |\gamma|$ mit $0 \neq \gamma^* \in T$. Damit ist

$$C(\gamma^* w) = \mu \gamma^* w \in W \setminus \partial W \tag{8}$$

nach (7). Aus der Definition von γ^* folgt daher $|\mu| < 1$ und aus der beliebigen Wahl von w, μ die Behauptung.

Durch Einschränkung des Intervallvektors X auf Vektoren mit k -ter Komponente 1 folgt durch Anwendung von Satz 2 auf (3) aus

$$f \begin{pmatrix} X \\ A \end{pmatrix} \subsetneq \begin{pmatrix} X \\ A \end{pmatrix} \quad \text{außer für die } k\text{-te Komponente} \tag{9}$$

sofort insbesondere die Nicht-Singularität von R und damit die Existenz eines $(x, \lambda) \in (X, A)$ mit $Ax = \lambda x$.

Schwieriger ist es, die Eindeutigkeit von Eigenvektor und Eigenwert in (X, A) nachzuweisen. Ohne Beweis sei der folgende Satz angegeben. Für eine detaillierte Diskussion und den Beweis siehe [1].

Satz 3: Mit den obigen Bezeichnungen folgt aus (8) bereits

- a) Es gibt genau einen Eigenvektor x von A mit $x \in X$.
- b) Es gibt genau einen Eigenvektor λ von A mit $\lambda \in A$.
- c) Dieser hat die Vielfachheit 1.
- d) Es gilt $Ax = \lambda x$.
- e) $\begin{pmatrix} x \\ \lambda \end{pmatrix} \in f^k \begin{pmatrix} X \\ A \end{pmatrix}$ für $0 \leq k \in \mathbb{N}$.

Die wesentliche Aussage des Satzes 3 ist, daß die Eindeutigkeit von Eigenvektor und Eigenwert einzeln nachgewiesen wird.

3. Der Algorithmus

Mit diesen Voraussetzungen ist es bereits möglich, eine einfache Version eines Algorithmus zur Bestimmung bewiesener Fehlerschranken für Eigenvektoren und Eigenwerte anzugeben. Seien \tilde{x} und $\tilde{\lambda}$ Gleitkommnäherungen

für ein Eigenvektor/Eigenwertpaar und

$$R \approx \begin{pmatrix} A - \tilde{\lambda}E & -\tilde{x} \\ \epsilon_k & 0 \end{pmatrix}^{-1} \text{ gleitkommamäßig,}$$

wobei an die Genauigkeit der Näherungen keinerlei Voraussetzungen geknüpft sind, wie die obigen Sätze zeigen. Für

$$U := f \left(\begin{array}{c} \tilde{x} \\ \tilde{\lambda} \end{array} \right) \text{ intervallmäßig}$$

iteriere man

$$\text{repeat } Y := U; \quad U := f(U)$$

$$\text{until } U \subseteq Y. \quad (9)$$

Für das Ergebnisintervall U gelten dann die Behauptungen von Satz 3. Neben vielen Verbesserungen sei speziell die Einschließung des Defekts statt der Lösung genannt. Damit ist es möglich, bei im wesentlichen einfacher Genauigkeit der Rechnung und nur punktueller Rechnung mit höherer Genauigkeit (Berechnung eines Defekts) die Lösung beliebig genau zu berechnen (siehe [1]). Die Algorithmen sind in FORTRAN implementiert und verfügbar. Als Beispiel für die Leistungsfähigkeit sei die Anwendung auf die 8×8 -HILBERT-Matrix genannt. Für einen Eigenwert berechnete der LR-Algorithmus $-2.17_{10} - 3$. Der wahre Wert $+1.26_{10} - 3$ wurde vom neuen Algorithmus wie alle Komponenten des Eigenvektors bei $8^{1/2}$ -stelliger Genauigkeit der Rechnung auf 18 Dezimalstellen eingeschlossen.

4. Zusammenfassung

Wie die obigen Beispiele zeigten, kann die laienhafte Verwendung von Gleitkommaalgorithmen bei der Eigenwertbestimmung große Gefahren bergen. Es wurde ein Algorithmus vorgestellt, der automatisch arbeitet und den Beweis für die Richtigkeit der Lösung mitliefert. Insbesondere kann damit die Richtigkeit oder Güte von Gleitkommannäherungen überprüft werden. Stoppt die Iteration (9) nicht, ist die Rechengenauigkeit gemessen an der Kondition unzureichend und es ist mit höherer Genauigkeit zu rechnen oder die Vielfachheit des Eigenwerts ist größer als 1. In fast allen gerechneten Beispielen stoppte die Iteration (9) nach einem Schritt. An der Verallgemeinerung auf Einschließung von Eigenwerten höherer Vielfachheit wird gearbeitet. Der Algorithmus ist leicht verallgemeinerbar auf komplexe, Intervall- und komplexe Intervallmatrizen.

Literatur

- 1 RUMP, S. M., Kleine Fehlerschranken bei Matrixproblemen, Dr.-Dissertation, Universität Karlsruhe, Februar 1980.
- 2 RUMP, S. M., Kleine, exakte Fehlerschranken für die Lösung linearer Gleichungssysteme, ZAMM 61 (1981), T 313 - T 315.

Auschrift: Dr. SIEGFRIED M. RUMP, Universität Karlsruhe, Institut für Angewandte Mathematik, Kaiserstraße 12, D-7500 Karlsruhe, BRD