Kapitel 5

Berechnung von Eigenwerten und Eigenvektoren

5.1 Einführung

Bemerkung 5.1 *Aufgabenstellung.* Diese Kapitel behandelt numerische Verfahren zur Lösung des Eigenwertproblems. Gegeben sei $A \in \mathbb{R}^{n\times n}$, bestimme $\lambda \in \mathbb{C}$ und $\boldsymbol{v} \in \mathbb{C}^n$, $\boldsymbol{v} \neq \mathbf{0}$, so dass

$$A\boldsymbol{v} = \lambda\boldsymbol{v} \tag{5.1}$$

gilt. Hierbei heißt λ Eigenwert und $\boldsymbol{v}$ Eigenvektor zum Eigenwert λ.

In der Vorlesung wird vor allem der Fall betrachtet, dass A eine symmetrische Matrix ist. Dann sind alle Eigenwerte und alle Eigenvektoren reell. □

Beispiel 5.2 *Spektralnorm und Spektralkondition einer symmetrischen Matrix.* Für eine symmetrische invertierbare Matrix A gelten $\|A\|_2 = |\lambda_{\max}(A)|$ und $\|A^{-1}\|_2 = |\lambda_{\min}(A)|^{-1}$. Damit muss man für die Berechnung der Spektralnorm von A den betragsmäßig größten Eigenwert bestimmen und zur Berechnung der Spektralkondition $\kappa_2(A) = \|A\|_2 \|A^{-1}\|_2$ noch zusätzlich den betragsmäßig kleinsten Eigenwert. Es ist jedoch nicht nötig, alle Eigenwerte zu berechnen. □

Beispiel 5.3 *Modellierung von Schwingungsabläufen, Sturm[1]–Liouville[2]-Problem.* Die mathematische Modellierung zur Beschreibung der Überlagerung von Schwingungsvorgängen kann durch das sogenannte Sturm–Liouville-Problem

$$u''(x) + \lambda r(x)u(x) = 0, \quad x \in (0,1), \tag{5.2}$$

mit den Randbedingungen $u(0) = u(1) = 0$ erfolgen. In (5.2) ist die Funktion $r \in C([0,1])$ mit $r(x) > 0$ gegeben und die Funktionen $u(x)$ sowie die Zahlen λ sind gesucht. Die Funktion $r(x)$ beschreibt Eigenschaften des Materials im Punkt x. Es handelt sich um ein Eigenwertproblem für ein Randwertproblem mit gewöhnlicher Differentialgleichung. Dies ist die eindimensionale Version von Modellen, wie man sie etwa beim Brückenbau verwendet, um Resonanzen zu vermeiden, die einen Brückeneinsturz verursachen könnten. Die Randbedingungen besagen, dass die Brücke an beiden Enden fest ist.

Im Fall $r(x) \equiv 1$ rechnet man direkt nach, dass $\lambda = (k\pi)^2$ und $u(x) = \sin(k\pi x)$, $k = 0, 1, 2, \ldots$ Lösungen von (5.2) sind. Es gibt also unendliche viele Paare, welche das Eigen-Randwertproblem erfüllen.

Ist $r(x)$ nicht konstant, dann gibt es aber im Allgemeinen keine geschlossene Formel, um die Lösungen darzustellen. Man muss die Lösungen numerisch approximieren. Dazu kann man $[0,1]$ in ein äquidistantes Gitter mit n Intervallen und der Schrittweite $h = 1/n$ zerlegen. Nun werden die Funktionen $r(x)$ und $u(x)$ in (5.2) in den Gitterpunkten x_i, $i = 0, \ldots, n$, betrachtet und die zweite Ableitung in den inneren Gitterpunkten wird durch einen Differenzenquotienten approximiert, siehe später in Abschnitt 7.4,

$$u''(x_i) \approx \frac{u(x_i + h) - 2u(x_i) + u(x_i - h)}{h^2}, \quad i = 1, \ldots, n-1. \tag{5.3}$$

[1] Jacques Charles Francois Sturm (1803 – 1855)
[2] Joseph Liouville (1809 – 1882)

Mit Taylor-Entwicklung kann man nachrechnen, dass diese Approximation genau bis auf Terme der Ordnung h^2 ist, wenn $u(x)$ glatt genug ist, vergleiche Beispiel 7.47. In den Randpunkten braucht man die zweite Ableitung nicht, da dort die Lösung gegeben ist.

Bezeichne u_i die approximierte Lösung im Gitterpunkt x_i, $i = 0, \ldots, n$. Dann wird in der Praxis in (5.3) $u(x_i)$ durch u_i usw. ersetzt. Einsetzen der Approximationen in (5.2) ergibt für die Knoten

$$\begin{aligned} u_0 &= 0, \\ \frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} + \lambda r(x_i) u_i &= 0, \quad i = 1, \ldots, n-1, \\ u_n &= 0. \end{aligned}$$

Die Randwerte kann man in die Gleichung für $i = 1$ beziehungsweise $i = n-1$ einsetzen und man erhält letztlich ein Gleichungssystem mit $(n-1)$ Gleichungen für die $(n-1)$ Unbekannten $\boldsymbol{u}^T = (u_1, \ldots, u_{n-1})$ der Gestalt

$$-B\boldsymbol{u} + \lambda D\boldsymbol{u} = 0$$

mit

$$B = \frac{1}{h^2} \begin{pmatrix} 2 & -1 & & & \\ -1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & -1 \\ & & & -1 & 2 \end{pmatrix}, \quad D = \begin{pmatrix} r(x_1) & & & & \\ & r(x_2) & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & r(x_{n-1}) \end{pmatrix}.$$

Da nach Voraussetzung $r(x_i) > 0$ für alle i ist, kann man

$$D^{1/2} = \operatorname{diag}\left(\sqrt{r(x_1)}, \ldots, \sqrt{r(x_{n-1})}\right)$$

bilden. Damit erhält man

$$B\boldsymbol{u} = \lambda D\boldsymbol{u} = \lambda D^{1/2} D^{1/2} \boldsymbol{u} \quad \Longleftrightarrow \quad D^{-1/2} B\boldsymbol{u} = \lambda D^{1/2} \boldsymbol{u}.$$

Setzt man $\boldsymbol{v} = D^{1/2}\boldsymbol{u}$, so ergibt sich ein Eigenwertproblem der Gestalt (5.1)

$$A\boldsymbol{v} = D^{-1/2} B D^{-1/2} \boldsymbol{v} = \lambda \boldsymbol{v}.$$

Da A symmetrisch

$$A^T = \left(D^{-1/2} B D^{-1/2}\right)^T = D^{-T/2} B^T D^{-T/2} = D^{-1/2} B D^{-1/2} = A$$

und positiv definit ist, sind alle Eigenwerte reell und positiv.

In dieser Aufgabe sind alle Eigenwerte und alle Eigenfunktionen gesucht. Man kann jedoch nur $n-1$ der Eigenwerte und Eigenfunktionen des Randwertproblems numerisch approximieren. □

5.2 Zur Theorie des Eigenwertproblems

Bemerkung 5.4 *Inhalt.* Dieser Abschnitt stellt einige Aussagen zur Theorie des Eigenwertproblems (5.1) zusammen, die zum Teil schon aus der linearen Algebra bekannt sind. □

Bemerkung 5.5 *Das charakteristische Polynom.* Die Zahl $\lambda \in \mathbb{C}$ ist genau dann Eigenwert von $A \in \mathbb{R}^{n \times n}$, wenn

$$p(\lambda) = \det(A - \lambda I) = 0.$$

Man nennt $p(\lambda) \in P_n$ das charakteristische Polynom der Matrix A. Seine Nullstellen sind die Eigenwerte von A. Da die Nullstellenberechnung eines Polynoms ein nichtlineares Problem ist, folgt, dass auch das Eigenwertproblem nichtlinear ist. Es ist insbesondere deutlich komplexer als die Lösung eines linearen Gleichungssystems.

Die Verwendung des charakteristischen Polynoms zur Berechnung der Eigenwerte von A besitzt jedoch in der Praxis entscheidende Nachteile. Zuerst müssen die Koeffizienten des Polynoms berechnet werden. Das ist für große n aufwändig. Des Weiteren hängen die Nullstellen oft sensibel von den Koeffizienten des charakteristischen Polynoms ab. Das Problem ist also schlecht konditioniert, insbesondere wenn die Matrix mehrfache Eigenwerte besitzt. Insgesamt ist das charakteristische Polynom zur numerischen Approximation von Eigenwerten einer Matrix nicht brauchbar (Übungsaufgabe). □

Bemerkung 5.6 *Weitere bekannte Aussagen, Begriffe.*

- Das charakteristische Polynom $p(\lambda)$ besitzt nach dem Fundamentalsatz der Algebra genau n (mit entsprechender Vielfachheit gezählte) reelle oder komplexe Nullstellen $\lambda_1, \ldots, \lambda_n$.
- Der Eigenwert λ_i heißt einfacher Eigenwert, wenn die entsprechende Nullstelle des charakteristischen Polynoms einfach ist.
- Die Menge aller Eigenwerte von A
$$\sigma(A) = \{\lambda_1, \ldots, \lambda_n\}$$
heißt Spektrum von A.
- Zwei Matrizen $A, B \in \mathbb{R}^{n\times n}$ heißen ähnlich (über dem Körper der reellen Zahlen), wenn es eine invertierbare Matrix $T \in \mathbb{R}^{n\times n}$ gibt mit der Eigenschaft
$$B = T^{-1}AT.$$
Ähnliche Matrizen besitzen das gleiche Spektrum
$$\sigma(A) = \sigma(T^{-1}AT)$$
für eine beliebige invertierbare Matrix $T \in \mathbb{R}^{n\times n}$, da sie dasselbe charakteristische Polynom besitzen
$$\begin{aligned} \det(B-\lambda I) &= \det\left(T^{-1}AT - \lambda T^{-1}T\right) = \det\left(T^{-1}(A-\lambda I)T\right) \\ &= \det\left(T^{-1}\right)\det(A-\lambda I)\det(T) = \det(A-\lambda I). \end{aligned}$$
- Die Matrix $A \in \mathbb{R}^{n\times n}$ heißt diagonalisierbar, wenn A zu einer Diagonalmatrix ähnlich ist. Die Matrix A ist genau dann diagonalisierbar, wenn sie n linear unabhängige Eigenvektoren hat. Die Eigenvektoren sind gerade die Spalten der entsprechenden matrix T.
- Besitzt $A \in \mathbb{R}^{n\times n}$ n verschiedene Eigenwerte, so ist A diagonalisierbar.

□

Lemma 5.7 Reelle Schur[3]-Faktorisierungen. *Zu jeder Matrix $A \in \mathbb{R}^{n\times n}$ gibt es eine orthogonale Matrix $Q \in \mathbb{R}^{n\times n}$, so dass*

$$Q^{-1}AQ = Q^TAQ = \begin{pmatrix} R_{11} & & & \\ & \ddots & * & \\ & & \ddots & \\ & 0 & & R_{mm} \end{pmatrix} = R \in \mathbb{R}^{n\times n}. \tag{5.4}$$

Dabei ist für jedes $i \in \{1, \ldots, m\}$, $m \le n$, entweder $R_{ii} \in \mathbb{R}$ oder $R_{ii} \in \mathbb{R}^{2\times 2}$. Im letzteren Fall hat R_{ii} ein Paar von konjugiert komplexen Eigenwerten. Die Menge aller Eigenwerte der R_{ii}, $i = 1, \ldots, m$, bilden das Spektrum von A. Die Zerlegung ist nicht eindeutig.

Beweis: Für einen Beweis sei auf die Literatur verwiesen, zum Beispiel (Golub & Van Loan, 1996, S. 341). ■

[3] Issai Schur (1875 – 1941)

Folgerung 5.8 Diagonalisierbarkeit symmetrischer Matrizen. *Jede symmetrische Matrix $A \in \mathbb{R}^{n\times n}$ lässt sich mittels einer orthogonalen Matrix $Q \in \mathbb{R}^{n\times n}$ durch eine Ähnlichkeitstransformation auf Diagonalgestalt bringen*

$$R = Q^{-1}AQ = D = diag\,(\lambda_1, \ldots, \lambda_n)\,.$$

Die Matrix A besitzt somit nur reelle Eigenwerte und n linear unabhängige, zueinander orthogonale Eigenvektoren, nämlich die Spalten von Q.

Beweis: Die Symmetrie von R folgt direkt aus (5.4) und der Symmetrie von A

$$R^T = \left(Q^{-1}AQ\right)^T = Q^T A^T Q^{-T} = Q^{-1}AQ = R.$$

Auf Grund der Symmetrie von R muss der mit $*$ markierte Block in (5.4) ein Nullblock sein, so dass

$$R = \mathrm{diag}\,(R_{11}, \ldots, R_{mm})\,.$$

Es können nun noch symmetrische 2×2 Blöcke R_{ii} auftreten. Man rechnet aber direkt nach, dass symmetrische 2×2 Matrizen mit nichtverschwindenden Nebendiagonalelementen immer zwei unterschiedliche reelle Eigenwerte besitzen, da die Diskriminante des charakteristischen Polynoms positiv ist (Übungsaufgabe). Somit widerspricht das Auftreten von symmetrischen 2×2 Blöcken den Aussagen von Lemma 5.7 und R muss eine Diagonalmatrix sein. ■

5.3 Kondition des Eigenwertproblems

Bemerkung 5.9 *Inhalt.* In diesem Abschnitt wird untersucht, wie stark sich Eigenwerte und Eigenvektoren bei Störungen der Koeffizienten von A verändern. □

Satz 5.10 Einfluss von Störungen auf die Eigenwerte. *Sei $A \in \mathbb{R}^{n\times n}$ diagonalisierbar, das heißt es existiert eine Matrix $T \in \mathbb{R}^{n\times n}$ mit*

$$T^{-1}AT = diag\,(\lambda_1, \ldots, \lambda_n)\,.$$

Sei $\Delta A \in \mathbb{R}^{n\times n}$ eine Störung von A und sei μ ein Eigenwert der gestörten Matrix $A + \Delta A$. Dann gilt

$$\min_{1\le i\le n} |\lambda_i - \mu| \le \|T\|_p \left\|T^{-1}\right\|_p \|\Delta A\|_p \tag{5.5}$$

für alle $p \in [1, \infty]$.

Beweis: Auch hier sei für den Beweis wieder auf die Literatur, zum Beispiel Golub & Van Loan (1996) oder Stoer & Bulirsch (2005). ■

Bemerkung 5.11 *Interpretation von Satz 5.10.* Die absolute Kondition des Eigenwertproblems

$$\sup_{\Delta A} \frac{\min_{1\le i\le n} |\lambda_i - \mu|}{\|\Delta A\|_p}$$

hängt von $\kappa_p(T) = \|T\|_p \|T^{-1}\|_p$ und nicht von $\kappa_p(A)$ ab. Da die Spalten von T gerade die Eigenvektoren von A sind, bedeutet (5.5) gerade, dass für eine diagonalisierbare Matrix die Kondition der Eigenvektorbasis eine große Rolle bei der Empfindlichkeit der Eigenwerte von A bezüglich Störungen spielt. □

Beispiel 5.12 *Kondition eines Eigenwertproblems.* Betrachte

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

und die gestörte Matrix

$$A + \Delta A = \begin{pmatrix} 0 & 1 \\ \delta & 0 \end{pmatrix}, \quad (\Delta A)^T \Delta A = \begin{pmatrix} \delta^2 & 0 \\ 0 & 0 \end{pmatrix},$$

mit $\delta > 0$. Die Eigenwerte von A sind $\lambda_1 = \lambda_2 = 0$ und die von $A + \Delta A$ sind $\tilde{\lambda}_{1,2} = \pm\sqrt{\delta}$. Für die Spektralkondition κ_2 des Eigenwertproblems ergibt sich somit, wobei man die Spektralnorm aus den Eigenwerten von $(\Delta A)^T \Delta A$ berechnen kann,

$$\kappa_2 \geq \frac{\left|\tilde{\lambda}_1 - \lambda_1\right|}{\|A + \Delta A - A\|_2} = \frac{\left|\tilde{\lambda}_1 - \lambda_1\right|}{\|\Delta A\|_2} = \frac{\sqrt{\delta}}{\delta} = \frac{1}{\sqrt{\delta}} \to \infty$$

für $\delta \to 0$.

Offenbar kann das Eigenwertproblem für beliebige Matrizen beliebig schlecht konditioniert sein. □

Folgerung 5.13 Kondition des Eigenwertproblems für symmetrische Matrizen. *Sei $A \in \mathbb{R}^{n\times n}$ symmetrisch und sei μ ein Eigenwert der gestörten Matrix $A + \Delta A$. Dann gilt*

$$\min_{1\leq i\leq n} |\lambda_i - \mu| \leq \|\Delta A\|_2 .$$

Das Eigenwertproblem für symmetrische Matrizen ist also gut konditioniert.

Beweis: Nach Folgerung 5.8 lässt sich A mittels einer Orthogonalmatrix Q diagonalisieren. Da für Orthogonalmatrizen $\kappa_2(Q) = 1$ gilt, folgt die Behauptung direkt aus (5.5). ■

5.4 Abschätzungen für Eigenwerte

Bemerkung 5.14 *Inhalt.* In diesem Abschnitt werden Abschätzungen für Eigenwerte angegeben, welche man aus direkt zugänglichen Informationen, zum Beispiel den Einträgen der Matrix, erhält, ohne dass man die Eigenwerte explizit berechnen muss. □

Lemma 5.15 Eigenschaften von Eigenwerten und des Spektrums. *Seien $A, B \in \mathbb{R}^{n\times n}$. Dann gelten die folgenden Aussagen.*

i) Falls $\det(A) \neq 0$ *und λ ein Eigenwert von A ist, so ist λ^{-1} ein Eigenwert von A^{-1}.*

ii) Ist $\lambda \in \sigma(A)$, dann ist $\alpha\lambda \in \sigma(\alpha A)$ für beliebiges $\alpha \in \mathbb{C}$.

iii) Ist $\lambda \in \sigma(A)$, dann ist $(\lambda - \alpha) \in \sigma(A - \alpha I)$. Man nennt α Spektralverschiebung oder Shift.

iv) Ist $\lambda \in \sigma(A)$, dann ist $\overline{\lambda} \in \sigma(A)$.

v) Es gilt $\sigma(A) = \sigma(A^T)$, da beide Matrizen dasselbe charakteristische Polynom besitzen.

vi) Es gilt $\sigma(AB) = \sigma(BA)$.

Beweis: Alle Aussagen sind aus der Linearen Algebra bekannt. ■

Lemma 5.16 Abschätzung von Eigenwerten mit Matrixnormen. *Es gilt $|\lambda| \leq \|A\|$ für jedes $\lambda \in \sigma(A)$ und jede Matrixnorm $\|\cdot\|$, die mit einer Vektornorm verträglich ist.*

Beweis: Seien $\lambda \in \sigma(A)$ und $\boldsymbol{v}$ ein zugehöriger Eigenvektor mit $\|\boldsymbol{v}\| = 1$, wobei die Vektornorm zur Matrixnorm verträglich ist. Dann gilt mit einer Normeigenschaft, der Eigenwertaufgabe und der Verträglichkeit der Normen

$$|\lambda| = |\lambda| \, \|\boldsymbol{v}\| = \|\lambda \boldsymbol{v}\| = \|A\boldsymbol{v}\| \leq \|A\| \, \|\boldsymbol{v}\| = \|A\| .$$

■

Bemerkung 5.17 *Zur Abschätzung von Eigenwerten mit Matrixnormen.*

- Der Spektralradius einer Matrix $A \in \mathbb{R}^{n\times n}$ ist definiert durch

  $$\rho(A) = \max\{|\lambda| \ : \ \lambda \in \sigma(A)\}.$$

  Aus Lemma 5.16 folgt sofort, dass $\rho(A) \leq \|A\|$ für jede Matrixnorm.
- Man kann zeigen, siehe Numerik II, dass es zu jedem $\varepsilon > 0$ eine Matrixnorm $\|\cdot\|$ gibt, so dass die Ungleichung $\|A\| \leq \rho(A) + \varepsilon$ gilt.

□

Satz 5.18 Kreissatz von Gerschgorin[4]. *Sei $A \in \mathbb{R}^{n\times n}$ und seien*

$$K_i = \left\{ z \in \mathbb{C} \ : \ |z - a_{ii}| \leq \sum_{j=1, i\neq j}^{n} \left|a_{ij}\right| \right\}$$

die Gerschgorin-Kreise. Dann gilt

$$\sigma(A) \subseteq \bigcup_{i=1}^{n} K_i. \tag{5.6}$$

Das heißt, alle Eigenwerte liegen in der Vereinigung der Gerschgorin-Kreise.

Beweis: Sei λ ein beliebiger Eigenwert von A mit Eigenvektor $\boldsymbol{x}$. Für einen Index i gilt $|x_i| \geq \left|x_j\right|$ für alle $j \neq i$. Da $\boldsymbol{x}$ ein Eigenvektor ist, gilt insbesondere $|x_i| > 0$. Die i-te Gleichung des Eigenwertproblems hat die Gestalt

$$\sum_{j=1, j\neq i}^{n} a_{ij} x_j + (a_{ii} - \lambda)\, x_i = 0.$$

Mit Dreiecksungleichung folgt

$$|\lambda - a_{ii}|\, |x_i| = \left| \sum_{j=1, j\neq i}^{n} a_{ij} x_j \right| \leq \sum_{j=1, j\neq i}^{n} \left|a_{ij}\right| \left|x_j\right|.$$

Nun ergibt Division mit $|x_i| > 0$ und $|x_i| \geq \left|x_j\right|$

$$|\lambda - a_{ii}| \leq \sum_{j=1, j\neq i}^{n} \left|a_{ij}\right| \frac{\left|x_j\right|}{|x_i|} \leq \sum_{j=1, j\neq i}^{n} \left|a_{ij}\right|.$$

Also liegt λ in einem der Gerschgorin-Kreise und damit erst recht in der Vereinigung aller Gerschgorin-Kreise. ∎

Folgerung 5.19 Weitere Einschränkungen des Bereiches der Eigenwerte. *Sei $A \in \mathbb{R}^{n\times n}$ mit den Gerschgorin-Kreisen K_i und seien K_i^T die Gerschgorin-Kreise von A^T. Dann gilt*

$$\sigma(A) \subseteq \left(\bigcup_{i=1}^{n} K_i \right) \cap \left(\bigcup_{i=1}^{n} K_i^T \right).$$

Falls A symmetrisch ist, gilt

$$\sigma(A) \subseteq \bigcup_{i=1}^{n} \left(K_i \cap \mathbb{R}\right).$$

Beweis: Aus Lemma 5.15 v) und dem Kreissatz von Gerschgorin für A^T folgt

$$\sigma(A) = \sigma\left(A^T\right) \subseteq \bigcup_{i=1}^{n} K_i^T.$$

Zusammen mit (5.6) folgt damit die erste Aussage.

Die zweite Aussage folgt daraus, dass alle Eigenwerte einer symmetrischen Matrix reell sind. ∎

Beispiel 5.20 *Anwendung des Kreissatzes von Gerschgorin.* Betrachte die Matrix

$$A = \begin{pmatrix} 4 & -1 & 0 \\ 0 & -2 & -1 \\ -1 & -1 & 3 \end{pmatrix}.$$

Die Eigenwerte von A sind (mit MATLAB berechnet) $\lambda_1 = -2.2223$, $\lambda_{2,3} = 3.6111 \pm 0.0974i$.

[4] Semjon Aronowitsch Gerschgorin (1901 – 1933)

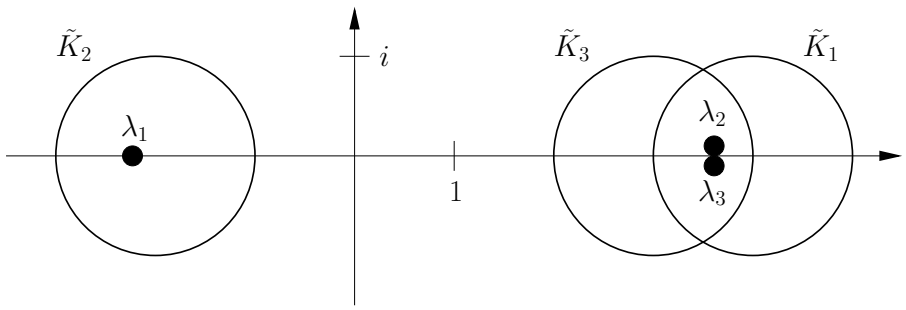

Abbildung 5.1: Beispiel 5.20. Eigenwerte und Gerschgorin-Kreise.

Zunächst kann man den Betrag der Eigenwerte mit Normen von A abschätzen. Man erhält

$$\|A\|_1 = 5, \quad \|A\|_F = 5.7446, \quad \|A\|_\infty = 5.$$

Damit ergibt sich $|\lambda_i| \leq 5$, $i = 1, 2, 3$.

Die Gerschgorin-Kreise von A und A^T sind

$$\begin{aligned} K_1 &= \{z \,:\, |z-4| \leq 1\}, & K_1^T &= \{z \,:\, |z-4| \leq 1\}, \\ K_2 &= \{z \,:\, |z+2| \leq 1\}, & K_2^T &= \{z \,:\, |z+2| \leq 2\}, \\ K_3 &= \{z \,:\, |z-3| \leq 2\}, & K_3^T &= \{z \,:\, |z-3| \leq 1\}. \end{aligned}$$

Es gilt $K_1 \cup K_2 \cup K_3 = K_2 \cup K_3$. Betrachtet man nun den Schnitt gemäß Folgerung 5.19, so erhält man, da $K_1^T \cup K_3^T \subset K_3$ ist, womit man K_3 weglassen kann,

$$\sigma(A) \subset \tilde{K}_1 \cup \tilde{K}_2 \cup \tilde{K}_3$$

mit

$$\tilde{K}_1 = \{z \,:\, |z-4| \leq 1\}, \quad \tilde{K}_2 = \{z \,:\, |z+2| \leq 1\}, \quad \tilde{K}_3 = \{z \,:\, |z-3| \leq 1\},$$

siehe Abbildung 5.1. □

5.5 Die Potenzmethode oder Vektoriteration

Bemerkung 5.21 *Grundidee.* Die Potenzmethode oder Vektoriteration ist ein Verfahren zur Berechnung des betragsgrößten Eigenwertes und eines zugehörigen Eigenvektors einer Matrix A. Dieses Verfahren geht auf von Mises[5] zurück. Es liefert das Grundkonzept für die Entwicklung weiterer Verfahren zur Berechnung von Eigenwerten und -vektoren.

Der Einfachheit halber werden für die Konvergenzanalyse einige Annahmen gemacht. Die Matrix $A \in \mathbb{R}^{n \times n}$ sei diagonalisierbar. Weiter gelte für die Eigenwerte

$$|\lambda_1| > |\lambda_2| \geq \ldots \geq |\lambda_n| \geq 0,$$

das heißt, es soll nur einen betragsgrößten Eigenwert geben und dieser soll einfach sein. Wegen der Diagonalisierbarkeit sind alle Eigenwerte reell, ihre algebraische Vielfachheit stimmt mit ihrer geometrischen Vielfachheit überein, alle Eigenvektoren $\boldsymbol{v}_j$, $j = 1, \ldots, n$, sind reell und die Eigenvektoren spannen $\mathbb{R}^n$ auf.

Für die Potenzmethode benötigt man einen Startvektor $\boldsymbol{x}^{(0)} \in \mathbb{R}^n$. Dieser lässt sich als Linearkombination der Eigenvektoren darstellen

$$\boldsymbol{x}^{(0)} = \sum_{j=1}^{n} c_j \boldsymbol{v}_j.$$

[5] Richard von Mises (1883 – 1953)

Sei $\boldsymbol{x}^{(0)}$ so gewählt, dass $c_1 \neq 0$ gilt. Multipliziert man die Darstellung von $\boldsymbol{x}^{(0)}$ mit der k-ten Potenz A^k von, so erhält man

$$A^k \boldsymbol{x}^{(0)} = \sum_{j=1}^{n} c_j A^k \boldsymbol{v}_j = \sum_{j=1}^{n} c_j \lambda_j^k \boldsymbol{v}_j.$$

Damit gilt

$$\boldsymbol{x}^{(k)} := A^k \boldsymbol{x}^{(0)} = \lambda_1^k \left(c_1 \boldsymbol{v}_1 + \sum_{j=2}^{n} c_j \left(\frac{\lambda_j}{\lambda_1} \right)^k \boldsymbol{v}_j \right) =: \lambda_1^k \left(c_1 \boldsymbol{v}_1 + \boldsymbol{r}^{(k)} \right). \tag{5.7}$$

Wegen $\left|\lambda_j / \lambda_1\right| < 1$ folgt

$$\lim_{k \to \infty} \boldsymbol{r}^{(k)} = \lim_{k \to \infty} \sum_{j=2}^{n} c_j \left(\frac{\lambda_j}{\lambda_1} \right)^k \boldsymbol{v}_j = \boldsymbol{0}.$$

Das bedeutet, für große k dominiert in (5.7) der Beitrag vom ersten Eigenwert und Eigenvektor. □

Satz 5.22 Konvergenz der Vektoriteration. *Sei $A \in \mathbb{R}^{n \times n}$ und erfülle A die Voraussetzungen aus Bemerkung 5.21. Sei $\boldsymbol{x}^{(k)} \in \mathbb{R}^n$ die k-te Iterierte der Potenzmethode und sei*

$$\lambda^{(k)} = \frac{\left(\boldsymbol{x}^{(k)}\right)^T A \boldsymbol{x}^{(k)}}{\left\|\boldsymbol{x}^{(k)}\right\|_2^2} = \frac{\left(\boldsymbol{x}^{(k)}\right)^T \boldsymbol{x}^{(k+1)}}{\left\|\boldsymbol{x}^{(k)}\right\|_2^2}.$$

Dann gilt

$$\left|\lambda_1 - \lambda^{(k)}\right| = \mathcal{O}\left(\left| \frac{\lambda_2}{\lambda_1} \right|^k \right).$$

Beweis: Dass $\lambda^{(k)}$ eine Approximation von λ_1 ist, wird im Beweis gezeigt.

Man betrachtet den Abstand zwischen dem Unterraum $S^{(k)} := \{\alpha \boldsymbol{x}^{(k)} \ : \ \alpha \in \mathbb{R}\}$ und dem Eigenvektor $\boldsymbol{v}_1$, wobei man o.B.d.A. diesen Vektor normiert, so dass $\|\boldsymbol{v}_1\|_2 = 1$,

$$d\left(S^{(k)}, \boldsymbol{v}_1\right) := \min_{\boldsymbol{x} \in S^{(k)}} \|\boldsymbol{x} - \boldsymbol{v}_1\|_2 = \min_{\alpha \in \mathbb{R}} \left\| \alpha \boldsymbol{x}^{(k)} - \boldsymbol{v}_1 \right\|_2.$$

Nun formt man (5.7) äquivalent um

$$\left(\lambda_1^k c_1\right)^{-1} \boldsymbol{x}^{(k)} = \boldsymbol{v}_1 + c_1^{-1} \boldsymbol{r}^{(k)}. \tag{5.8}$$

Eine Abschätzung für $d(S^{(k)}, \boldsymbol{v}_1)$ erhält man, wenn man den Wert

$$\alpha = \alpha_k := \left(\lambda_1^k c_1\right)^{-1}, \tag{5.9}$$

wählt. Damit, der Darstellung von $\boldsymbol{r}^{(k)}$ und der aus (5.7) folgenden Asymptotik von $\boldsymbol{r}^{(k)}$ ergibt sich

$$d\left(S^{(k)}, \boldsymbol{v}_1\right) \leq \left\| \alpha_k \boldsymbol{x}^{(k)} - \boldsymbol{v}_1 \right\|_2 = \left| c_1^{-1} \right| \left\| \boldsymbol{r}^{(k)} \right\|_2 = \mathcal{O}\left(\left| \frac{\lambda_2}{\lambda_1} \right|^k \right).$$

Da die rechte Seite für $k \to \infty$ gegen Null strebt, ist $\alpha_k \boldsymbol{x}^{(k)}$ eine Approximation an $\boldsymbol{v}_1$, also

$$A \alpha_k \boldsymbol{x}^{(k)} \approx \lambda_1 \alpha_k \boldsymbol{x}^{(k)} \quad \Longleftrightarrow \quad A \boldsymbol{x}^{(k)} \approx \lambda_1 \boldsymbol{x}^{(k)}.$$

Durch Multiplikation dieser Beziehung von links mit $(\boldsymbol{x}^{(k)})^T$ und Division durch $\|\boldsymbol{x}^{(k)}\|_2^2$ folgt, dass

$$\lambda_1 \approx \frac{\left(\boldsymbol{x}^{(k)}\right)^T A \boldsymbol{x}^{(k)}}{\left\|\boldsymbol{x}^{(k)}\right\|_2^2} = \frac{\left(\boldsymbol{x}^{(k)}\right)^T \boldsymbol{x}^{(k+1)}}{\left\|\boldsymbol{x}^{(k)}\right\|_2^2} = \lambda^{(k)}$$

eine Approximation an λ_1 ist. Genauso wie $\alpha_k \boldsymbol{x}^{(k)}$ eine Approximation von $\boldsymbol{v}_1$ ist, zeigt man, dass $\alpha_{k+1} \boldsymbol{x}^{(k+1)}$ mit $\alpha_{k+1} = (\lambda_1^{k+1} c_1)^{-1} = \alpha_k / \lambda_1$, was aus (5.9) folgt, eine Approximation von $\boldsymbol{v}_1$ ist.

Jetzt muss man noch die Güte dieser Approximation untersuchen. Sei $\mathbf{1} \in \mathbb{R}^n$ der Vektor, der in jeder Komponente Eins ist. Man erhält mit (5.8), der Definition von $\boldsymbol{r}^{(k)}$, der Asymptotik von $\boldsymbol{r}^{(k)}$ und $\|\boldsymbol{v}_1\|_2 = 1$

$$\begin{aligned}
\lambda^{(k)} &= \frac{\left(\alpha_k \boldsymbol{x}^{(k)}\right)^T \left(\alpha_k \boldsymbol{x}^{(k+1)}\right)}{\left\|\alpha_k \boldsymbol{x}^{(k)}\right\|_2^2} = \lambda_1 \frac{\left(\alpha_k \boldsymbol{x}^{(k)}\right)^T \left(\alpha_{k+1} \boldsymbol{x}^{(k+1)}\right)}{\left\|\alpha_k \boldsymbol{x}^{(k)}\right\|_2^2} \\
&= \lambda_1 \frac{\left(\boldsymbol{v}_1 + c_1^{-1} \boldsymbol{r}^{(k)}\right)^T \left(\boldsymbol{v}_1 + c_1^{-1} \boldsymbol{r}^{(k+1)}\right)}{\left\|\boldsymbol{v}_1 + c_1^{-1} \boldsymbol{r}^{(k)}\right\|_2^2} \\
&= \lambda_1 \frac{\left(\boldsymbol{v}_1 + \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)\mathbf{1}\right)^T \left(\boldsymbol{v}_1 + \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{k+1}\right)\mathbf{1}\right)}{\left\|\boldsymbol{v}_1 + \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)\mathbf{1}\right\|_2^2} \\
&= \lambda_1 \frac{1 + \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)}{1 + \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)} = \lambda_1 \left(1 + \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)\right). \qquad (5.10)
\end{aligned}$$

Mit Nutzung von $\|\boldsymbol{v}_1\|_2 = 1$ sieht man die Gültigkeit des vorletzten Schrittes aus

$$\begin{aligned}
&\left(\boldsymbol{v}_1 + \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)\mathbf{1}\right)\left(\boldsymbol{v}_1 + \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{k+1}\right)\mathbf{1}\right) \\
&= 1 + \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right) + \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{k+1}\right) + \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{2k+1}\right) \\
&= 1 + \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right),
\end{aligned}$$

vergleiche Definition des Landau[6]-Symbols. Der letzte Schritt ergibt sich aus (h.o.t. = higher order terms)

$$\begin{aligned}
\frac{1 + \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)}{1 + \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)} &= \frac{1 + C_1\left|\frac{\lambda_2}{\lambda_1}\right|^k + \text{h.o.t.}}{1 + C_2\left|\frac{\lambda_2}{\lambda_1}\right|^k + \text{h.o.t.}} = \frac{1 + C_2\left|\frac{\lambda_2}{\lambda_1}\right|^k + (C_1 - C_2)\left|\frac{\lambda_2}{\lambda_1}\right|^k + \text{h.o.t.}}{1 + C_2\left|\frac{\lambda_2}{\lambda_1}\right|^k + \text{h.o.t.}} \\
&= 1 + \left|\frac{\lambda_2}{\lambda_1}\right|^k \frac{C_1 - C_2 + \text{h.o.t.}}{1 + C_2\left|\frac{\lambda_2}{\lambda_1}\right|^k + \text{h.o.t.}} \\
&= 1 + \left|\frac{\lambda_2}{\lambda_1}\right|^k (C_1 - C_2 + \text{h.o.t.}) = 1 + \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right),
\end{aligned}$$

wobei $C_1, C_2 \geq 0$ und die Terme höherer Ordnung an unterschiedlichen Stellen verschieden sein können.

Durch Umstellen von (5.10) folgt

$$\left|\lambda_1 - \lambda^{(k)}\right| = \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right).$$

■

Bemerkung 5.23 *Symmetrische Matrizen.* Falls A eine symmetrische Matrix ist, kann man sogar

$$\left|\lambda_1 - \lambda^{(k)}\right| = \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{2k}\right)$$

zeigen. □

[6] Edmund Georg Hermann Landau (1877 – 1938)

Bemerkung 5.24 *Skalierung der Iterierten.* Wendet man das bisherige Verfahren an, so gelten, da $\boldsymbol{x}^{(k)} \approx \lambda_1^k c_1 \boldsymbol{v}_1$ wegen (5.7),

$$\begin{aligned} \left\|\boldsymbol{x}^{(k)}\right\|_2 &\to \infty \quad \text{falls} \quad |\lambda_1| > 1, \\ \left\|\boldsymbol{x}^{(k)}\right\|_2 &\to 0 \quad \text{falls} \quad |\lambda_1| < 1. \end{aligned}$$

Aus diesen Gründen ist es zweckmäßig, die Iterierten zu skalieren. Damit werden starke Änderungen in der Größenordnung und das Verlassen des Bereichs der Gleitkommazahlen vermieden. Die Konvergenzaussagen ändern sich durch Skalierung auch nicht, da weder der Unterraum $S^{(k)}$ noch die Iterierte $\lambda^{(k)}$ von einer Skalierung von $\boldsymbol{x}^{(k)}$ abhängen. □

Algorithmus 5.25 *Potenzmethode, Vektoriteration.* Seien $A \in \mathbb{R}^{n\times n}$ und $\boldsymbol{y}^{(0)} \neq \boldsymbol{0}$ mit $\|\boldsymbol{y}^{(0)}\|_2 = 1$ gegeben. Für $k = 0, 1, \ldots$ berechne man

$$\begin{aligned} \tilde{\boldsymbol{y}}^{(k+1)} &= A\boldsymbol{y}^{(k)} \\ \lambda^{(k)} &= \left(\tilde{\boldsymbol{y}}^{(k+1)}\right)^T \boldsymbol{y}^{(k)} \\ \boldsymbol{y}^{(k+1)} &= \frac{\tilde{\boldsymbol{y}}^{(k+1)}}{\left\|\tilde{\boldsymbol{y}}^{(k+1)}\right\|_2}. \end{aligned}$$

Die Bezeichnung $\boldsymbol{y}^{(k)}$ wurde gewählt, um die normierten Vektoren von den nichtnormierten Vektoren $\boldsymbol{x}^{(k)}$ zu unterscheiden. □

Bemerkung 5.26 *Zu Algorithmus 5.25.*

- Wählt man als Startiterierte $\boldsymbol{x}^{(0)}$, so weist man mit vollständiger Induktion nach, dass in exakter Arithmetik
$$\boldsymbol{y}^{(k)} = \frac{\boldsymbol{x}^{(k)}}{\left\|\boldsymbol{x}^{(k)}\right\|_2} = \frac{A^k\boldsymbol{x}^{(0)}}{\left\|A^k\boldsymbol{x}^{(0)}\right\|_2}.$$
Also liefert Algorithmus 5.25, bis auf Skalierung in $\boldsymbol{x}^{(k)}$, die oben analysierten Folgen $\{\boldsymbol{x}^{(k)}\}$ und $\{\lambda^{(k)}\}$. Insbesondere ist $\boldsymbol{y}^{(k)}$ als ein Nicht-Null-Vielfaches von $\boldsymbol{x}^{(k)}$ eine Approximation eines Eigenvektors zum Eigenwert λ_1.
- Die Konvergenzgeschwindigkeit der Potenzmethode hängt wesentlich vom Verhältnis von $|\lambda_1|$ und $|\lambda_2|$ ab.

□

Beispiel 5.27 *Sturm-Liouville-Problem.* Betrachte das in Beispiel 5.3 hergeleitete Eigenwertproblem $A\boldsymbol{v} = \lambda\boldsymbol{v}$. Sei $r(x) \equiv 1$, dann sind die Eigenwerte von A bekannt

$$\lambda_{n-j} = \frac{4}{h^2}\sin^2\left(\frac{j\pi h}{2}\right), \quad j = 1, \ldots, n-1, \ h = \frac{1}{n}.$$

Es ist $\lambda_1 > \lambda_2 > \ldots > \lambda_{n-1} > 0$. Dann erhält man mit einem Additionstheorem für die Sinusfunktion, Taylorentwicklung und Polynomdivision

$$\begin{aligned} \left|\frac{\lambda_2}{\lambda_1}\right| &= \frac{\sin^2\left(\frac{(n-2)\pi h}{2}\right)}{\sin^2\left(\frac{(n-1)\pi h}{2}\right)} = \frac{\sin^2\left(\frac{\pi}{2} - \pi h\right)}{\sin^2\left(\frac{\pi}{2} - \frac{\pi h}{2}\right)} = \frac{\cos^2(\pi h)}{\cos^2\left(\frac{\pi h}{2}\right)} \\ &\approx \frac{\left(1 - \frac{(\pi h)^2}{2}\right)^2}{\left(1 - \frac{(\pi h/2)^2}{2}\right)^2} = \frac{1 - \pi^2h^2 + \frac{\pi^4h^4}{4}}{1 - \frac{\pi^2h^2}{4} + \frac{\pi^4h^4}{64}} \approx 1 - \frac{3}{4}\pi^2h^2. \end{aligned}$$

Man erkennt, dass man im Fall $h \ll 1$ mit einer sehr langsamen Konvergenz $\lambda^{(k)} \to \lambda_1$ rechnen muss. □

Bemerkung 5.28 *Fazit.* Falls A diagonalisierbar ist, λ_1 ein einfacher Eigenwert ist und es keine weiteren Eigenwerte gibt, deren Betrag gleich $|\lambda_1|$ ist, dann konvergiert Algorithmus 5.25. Die Konvergenz kann aber sehr langsam sein. □

Bemerkung 5.29 *Berechnung anderer Eigenwerte, inverse Vektoriteration, Spektralverschiebung.* Sei $A \in \mathbb{R}^{n\times n}$ nichtsingulär und diagonalisierbar. Die Eigenwertgleichung $A\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$, $i = 1, \ldots, n$, ist äquivalent zu

$$\frac{1}{\lambda_i}\boldsymbol{v}_i = A^{-1}\boldsymbol{v}_i.$$

Damit würde die Vektoriteration angewandt mit A^{-1} unter der Annahme

$$|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_{n-1}| > |\lambda_n|$$

den betragsmäßig größten Eigenwert λ_n^{-1} von A^{-1} berechnen, das heißt den betragsmäßig kleinsten Eigenwert von A.

Nach Lemma 5.15, iii), ist λ_i ein Eigenwert von A genau dann, wenn $\lambda_i - \mu$ ein Eigenwert von $A - \mu I$ ist. Angenommen, man hätte eine Schätzung $\mu \approx \lambda_i$ eines beliebigen einfachen reellen Eigenwertes von A, so dass

$$|\lambda_i - \mu| < \left|\lambda_j - \mu\right|, \quad \text{für alle } i \neq j. \tag{5.11}$$

Dann ist $(\lambda_i - \mu)^{-1}$ der betragsmäßig größte Eigenwert von $(A - \mu I)^{-1}$. Zur Berechnung dieses Eigenwertes kann man die Vektoriteration anwenden. □

Algorithmus 5.30 *Inverse Vektoriteration mit Spektralverschiebung.* Gesucht ist der einfache reelle Eigenwert λ_i von $A \in \mathbb{R}^{n\times n}$. Wähle μ so, dass (5.11) gilt und wähle einen Startvektor $\boldsymbol{y}^{(0)} \neq \boldsymbol{0}$ mit $\|\boldsymbol{y}^{(0)}\|_2 = 1$. Für $k = 0, 1, \ldots$ berechne man

$$\begin{aligned} (A - \mu I)\,\tilde{\boldsymbol{y}}^{(k+1)} &= \boldsymbol{y}^{(k)} \\ \lambda^{(k)} &= \frac{1}{\left(\tilde{\boldsymbol{y}}^{(k+1)}\right)^T \boldsymbol{y}^{(k)}} + \mu \\ \boldsymbol{y}^{(k+1)} &= \frac{\tilde{\boldsymbol{y}}^{(k+1)}}{\left\|\tilde{\boldsymbol{y}}^{(k+1)}\right\|_2}. \end{aligned} \tag{5.12}$$

□

Bemerkung 5.31 *Zu Algorithmus 5.30.*

- In (5.12) muss man ein lineares Gleichungssystem mit der Matrix $(A - \mu I)$ lösen. Dafür berechnet man einmal eine LU- oder QR-Zerlegung von $(A - \mu I)$.
- Die Vektoriteration, Algorithmus 5.25 für $(A - \mu I)^{-1}$, strebt gegen $(\lambda_i - \mu)^{-1}$. Das bedeutet für die Iterierten aus Algorithmus 5.30, dass

  $$\lambda^{(k)} = \frac{1}{\left(\tilde{\boldsymbol{y}}^{(k+1)}\right)^T \boldsymbol{y}^{(k)}} + \mu \to \lambda_i - \mu + \mu = \lambda_i$$

  für $k \to \infty$.
- Die Konvergenzgeschwindigkeit von Algorithmus 5.30 hängt wie bei der Vektoriteration vom Verhältnis der betragsgrößten Eigenwerte ab. Das ist hier

  $$\frac{\max_{j\neq i}\left(\left|\lambda_j - \mu\right|^{-1}\right)}{|\lambda_i - \mu|^{-1}} = \frac{\left(\min_{j\neq i}\left|\lambda_j - \mu\right|\right)^{-1}}{|\lambda_i - \mu|^{-1}} = \frac{|\lambda_i - \mu|}{\min_{j\neq i}\left|\lambda_j - \mu\right|}.$$

  Hat man also eine gute Schätzung μ von λ_i, dann gilt

  $$\frac{|\lambda_i - \mu|}{\min_{j\neq i}\left|\lambda_j - \mu\right|} \ll 1$$

  und das Verfahren konvergiert sehr schnell. In der Praxis ist allerdings im Allgemeinen nicht klar, wie man μ wählen sollte.

- Die Konvergenzgeschwindigkeit kann verbessert werden, wenn man während der Iteration den Parameter μ geeignet anpasst, zum Beispiel mit der aktuellen Iterierten $\mu = \lambda^{(k)}$. Nach jeder Anpassung muss man allerdings die Matrix $(A - \mu I)$ neu faktorisieren, so dass die Kosten dieses Iterationsschrittes vergleichsweise sehr hoch sind.

□

5.6 Das QR-Verfahren

Bemerkung 5.32 *Inhalt.* Sei $A \in \mathbb{R}^{n\times n}$ eine symmetrische Matrix. Aus Folgerung 5.13 ist bekannt, dass das Eigenwertproblem für A gut konditioniert ist. Dieser Abschnitt stellt ein Verfahren zur Approximation aller Eigenwerte und Eigenvektoren von A vor. □

Bemerkung 5.33 *Transformationen von A durch Orthogonaltransformationen.* Aus Folgerung 5.8 ist bekannt, dass alle Eigenwerte von A reell sind, A diagonlisierbar ist und eine Orthonormalbasis aus Eigenvektoren $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ existiert, so dass

$$Q^{-1}AQ = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$$

mit $Q = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n]$.

Im Allgemeinen ist es jedoch nicht möglich, Q in endlich vielen Schritten zu bestimmen. Damit wäre auch ein endliches Verfahren zur Bestimmung aller Nullstellen eines Polynoms n-ten Grades gefunden. Ein solches Verfahren, basierend auf elementaren Rechenoperation und der Quadratwurzel, kann es aber nach dem Satz von Abel[7] nicht geben.

Es ist auch nicht möglich, A mit Orthogonaltransformationen, zum Beispiel mit Housholder-Spiegelungen, auf Diagonalgestalt zu bringen. Mit einer ersten Housholder-Spiegelung kann man die Elemente der ersten Spalte unterhalb der Diagonalen zu Null machen. Wendet man dann eine Housholder-Spiegelung auf die erste Zeile an, dann wird die erste Spalte wieder gefüllt

$$A = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{pmatrix} \xrightarrow[\text{v.l.}]{Q_1} \begin{pmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \end{pmatrix} \xrightarrow[\text{v.r.}]{Q_2} \begin{pmatrix} * & 0 & 0 & 0 & 0 \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{pmatrix}.$$

Man verliert auch die Symmetrie der Matrix.

Es ist jedoch möglich, A mit orthogonalen Transformationen auf Tridiagonalgestalt zu bringen. Verwendet man eine orthogonale Matrix deren erste Zeile der erste Einheitsvektor ist und deren Spalten durch eine Housholder-Spiegelung so konstruiert sind, dass die Elemente der ersten Spalte unter a_{21} verschwinden, so erhält man

$$A = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{pmatrix} \xrightarrow[\text{v.l.}]{Q_1} \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \end{pmatrix} \xrightarrow[\text{v.r.}]{Q_1^T} \begin{pmatrix} * & * & 0 & 0 & 0 \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \end{pmatrix}.$$

Die Multiplikation von rechts mit Q_1^T hat keine Auswirkungen auf die erste Spalte. Auf diese Art und Weise fährt man fort bis man eine Tridiagonalmatrix erhält. □

Lemma 5.34 Transformationen auf Tridiagonalgestalt. *Sei $A \in \mathbb{R}^{n\times n}$ symmetrisch. Dann existiert eine orthogonale Matrix $Q \in \mathbb{R}^{n\times n}$, die das Produkt von $(n-2)$ Householder-Spiegelungen ist, so dass QAQ^T eine symmetrische Tridiagonalmatrix ist.*

[7] Niels Henrik Abel (1802 – 1829)

Beweis: Die Fortsetzung des in Bemerkung 5.33 beschriebenen Prozesses und die Eigenschaften der Householder-Matrizen $Q_1, \ldots, Q_{n-2}$ liefern

$$QAQ^T = Q_{n-2} \ldots Q_1 A Q_1^T \ldots Q_{n-2}^T = \begin{pmatrix} * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ 0 & * & * & * & 0 \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}.$$

Die Symmetrie der Matrix folgt aus der linken oder mittleren Darstellung und der Symmetrie von A. ■

Bemerkung 5.35 *Reduktion des Eigenwertproblems.* Die Matrizen A und QAQ^T sind ähnlich und sie besitzen gemäß Bemerkung 5.6 dieselben Eigenwerte. Damit hat man also das Problem der Bestimmung der Eigenwerte einer symmetrischen Matrix auf das Problem der Bestimmung der Eigenwerte einer symmetrischen Tridiagonalmatrix reduziert. □

Bemerkung 5.36 *Iteration mit Tridiagonalmatrizen.* Die numerische Approximation der Eigenwerte einer Tridiagonalmatrix wird iterativ erfolgen. Die grundlegende Idee wurde in Rutishauser (1958) vorgestellt. Sei

$$B = B_1 = \begin{pmatrix} * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ 0 & * & * & * & 0 \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}.$$

Dann kann man eine Faktorisierung von B bestimmen, wobei in Rutishauser (1958) die LU-Zerlegung vorgeschlagen wurde. Für Tridiagonalmatrizen erhält man ein Produkt von zwei Bidiagonalmatrizen

$$B = B_1 = LU = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ * & 1 & 0 & 0 & 0 \\ 0 & * & 1 & 0 & 0 \\ 0 & 0 & * & 1 & 0 \\ 0 & 0 & 0 & * & 1 \end{pmatrix} \begin{pmatrix} * & * & 0 & 0 & 0 \\ 0 & * & * & 0 & 0 \\ 0 & 0 & * & * & 0 \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{pmatrix}.$$

Nun vertauscht man die Faktoren

$$B_2 = UL = \begin{pmatrix} * & * & 0 & 0 & 0 \\ 0 & * & * & 0 & 0 \\ 0 & 0 & * & * & 0 \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ * & 1 & 0 & 0 & 0 \\ 0 & * & 1 & 0 & 0 \\ 0 & 0 & * & 1 & 0 \\ 0 & 0 & 0 & * & 1 \end{pmatrix}$$

und wiederholt das Verfahren mit B_2. In den Arbeiten Francis (1962) und Kublanovskaja (1961) wurde das Verfahren modifiziert, indem statt der LU-Zerlegung die stabilere QR-Zerlegung verwendet wurde.

Mit diesem Verfahren erhält man eine Folge von Matrizen $\{B_k\}_{k\in\mathbb{N}}$. □

Lemma 5.37 Eigenschaften der Matrizen $\{B_k\}_{k\in\mathbb{N}}$. *Die Matrizen $\{B_k\}_{k\in\mathbb{N}}$ seien mit dem Verfahren aus Bemerkung 5.36 definiert, wobei die Faktorisierung mittels einer QR-Zerlegung vorgenommen wurde $B_k = Q_k R_k$, $B_{k+1} := R_k Q_k$. Dann gelten mit $B = B_1$:*

i) B_k ist ähnlich zu B, $k \geq 1$.

ii) Falls B symmetrisch ist, so ist auch B_k symmetrisch, $k \geq 1$.

iii) Falls B symmetrisch und tridiagonal ist, so ist auch B_k symmetrisch und tridiagonal, $k \geq 1$.

Beweis: *i).* Die Aussage ist bewiesen, wenn man gezeigt hat, dass B_k und B_{k+1} für beliebiges $k \geq 1$ ähnlich sind. Nach Konstruktion der Matrizen $\{B_k\}_{k\in\mathbb{N}}$ und einer Eigenschaft von Orthogonalmatrizen, Lemma 2.14, gilt

$$Q_k B_{k+1} Q_k^T = Q_k R_k Q_k Q_k^T = Q_k R_k = B_k. \tag{5.13}$$

Das ist genau die Ähnlichkeit der beiden Matrizen.

ii). Diese Aussage wird durch vollständige Induktion nach k gezeigt. Der Induktionsanfang, die Symmetrie von $B_1 = B$ ist klar. Gelte also die Symmetrie für B_k. Dann folgt mit (5.13)

$$B_{k+1}^T = \left(Q_k^T B_k Q_k\right)^T = Q_k^T B_k^T Q_k = Q_k^T B_k Q_k = B_{k+1}.$$

iii). Der Beweis wird mit vollständiger Induktion nach k erbracht. Der Induktionsanfang ist wieder klar. Sei nun B_k eine symmetrische Tridiagonalmatrix. Dann kann man mit $(n-1)$ Givens-Drehungen $G_{i,k}$ die Einträge der unteren Hauptnebendiagonalen zu Null machen und erhält eine Dreiecksmatrix R_k mit dem Besetzheitsmuster

$$B_k = Q_k R_k = G_{1,k} \dots G_{n-1,k} R_k = G_{1,k} \dots G_{n-1,k} \begin{pmatrix} * & * & * & 0 & 0 \\ 0 & * & * & * & 0 \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{pmatrix}.$$

Beispielsweise im ersten Schritt, bei der Givens-Drehung $G_{1,k}$, wird ein Vielfaches der zweiten Zeile von B_k zur ersten Zeile addiert. Da B_k tridiagonal ist, also $b_{2j} = 0$ für $j > 3$, bleiben alle Elemente der ersten Zeile der resultierenden Matrix mit einem Spaltenindex größer als Drei Null. Damit hat man eine konkrete Form von R_k. Nun ist

$$\begin{aligned} B_{k+1} &= R_k Q_k = R_k G_{1,k} \dots G_{n-1,k} = \begin{pmatrix} * & * & * & 0 & 0 \\ * & * & * & * & 0 \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{pmatrix} G_{2,k} \dots G_{n-1,k} \\ &= \begin{pmatrix} * & * & * & 0 & 0 \\ * & * & * & * & 0 \\ 0 & * & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{pmatrix} G_{3,k} \dots G_{n-1,k} = \dots = \begin{pmatrix} * & * & * & 0 & 0 \\ * & * & * & * & 0 \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}, \end{aligned}$$

da $G_{1,k}$ nur Nichtnulleinträge im Durchschnitt der ersten beiden Zeilen und Spalten erzeugt oder ändert, $G_{2,k}$ nur im Durchschnitt der zweiten und dritten Zeile und Spalte, und so weiter. Nach Teil ii) ist B_{k+1} aber auch symmetrisch. Das bedeutet, dass der ganze Fill-in im Dreieck über der oberen Hauptnebendiagonalen verschwindet. Damit ist B_{k+1} tridiagonal. ■

Satz 5.38 Iteration zur Approximation der Eigenwerte. *Sei $A \in \mathbb{R}^{n\times n}$ symmetrisch mit Eigenwerten $\lambda_1, \dots, \lambda_n$, welche die Eigenschaft*

$$|\lambda_1| > |\lambda_2| > \dots > |\lambda_n| > 0$$

haben mögen. Weiter seien die Matrizenfolgen $\{A_k\}_{k\in\mathbb{N}}$, $\{Q_k\}_{k\in\mathbb{N}}$ und $\{R_k\}_{k\in\mathbb{N}}$ durch folgenden Algorithmus definiert:

1. *(Initialisierung.) setze $A_1 = A$, $k = 1$,*
2. *(Faktorisierung von A_k.) berechne QR-Zerlegung $A_k = Q_k R_k$,*
3. *(Bildung von A_{k+1}.) bestimme $A_{k+1} = R_k Q_k$,*
4. *setze $k = k + 1$, gehe zu 2.*

Dann gibt es (Vorzeichen-) Matrizen $S_k = diag(\sigma_1^{(k)}, \dots, \sigma_n^{(k)})$ mit $|\sigma_i^{(k)}| = 1$ so dass

$$\lim_{k\to\infty} S_{k-1} Q_k S_k = I$$

und

$$\lim_{k\to\infty} S_k R_k S_{k-1} = \lim_{k\to\infty} S_{k-1} A_k S_k = diag\,(\lambda_1, \dots, \lambda_n) = D.$$

Es gilt also insbesondere, da aus der Existenz des Grenzwertes folgt, dass $S_{k-1} = S_k$ ab einem gewissen Index ist,

$$\lim_{k\to\infty} a_{jj}^{(k)} = \lambda_j, \quad j = 1, \dots, n,$$

wobei $a_{jj}^{(k)}$ das j-te Diagonalelement von A_k ist.

Beweis: Der Beweis ist recht umfangreich und deshalb wird auf die Literatur verwiesen, zum Beispiel auf Wilkinson (1965). Die Vorzeichenmatrix erscheint im Beweis, da die Givens-Drehungen nur bis auf das Vorzeichen bestimmt sind. ■

Bemerkung 5.39 *Zu Satz 5.38.*

- Satz 5.38 bietet einen Algorithmus zur Konstruktion einer Schur-Faktorisierung von A gemäß (5.4).
- Der Algorithmus aus Satz 5.38 lässt sich als Verallgemeinerung der Vektoriteration auffassen. Er entspricht der Projektion auf die Unterräume, die von den Spalten von A_k aufgespannt werden, siehe (Stoer & Bulirsch, 2005, S. 58).

□

Algorithmus 5.40 *QR-Verfahren zur Eigenwertberechnung.* Sei die symmetrische Matrix $A \in \mathbb{R}^{n \times n}$ gegeben.

1. Transformiere A mit Hilfe von Householder-Spiegelungen auf Tridiagonalgestalt $B = Q^T AQ$.
2. Wende auf B den Algorithmus aus Satz 5.38 mit Givens-Drehungen an, wobei

$$GBG^T \approx D$$

und G das Produkt aller Givens-Matrizen ist. Die Diagonale von GBG^T approximiert die Eigenwerte von A und die Spalten von GQ^T, mit Q aus Schritt 1, die zugehörigen Eigenvektoren.

□

Bemerkung 5.41 *Zum QR-Verfahren.*

- Der Aufwand des ersten Schrittes beträgt $\mathcal{O}(\frac{2}{3}n^3)$ Multiplikationen/Divisionen. Jede Iteration im 2. Schritt benötigt $\mathcal{O}(n^2)$ Multiplikationen/Divisionen.
- Man kann zeigen, dass die Konvergenzgeschwindigkeit des QR-Verfahrens von den Quotienten $\left|\lambda_{j+1}/\lambda_j\right|$ für $j = 1, \ldots, n-1$, abhängt. Liegt dieser Wert für einen oder mehrere Indizes nahe bei Eins, dann ist die Effizienz des Verfahrens schlecht. Abhilfe kann man auch hier mit einer Spektralverschiebung schaffen, vergleiche Algorithmus 5.30.
- Mit einigen Modifikationen lässt sich das Verfahren auch auf nichtsymmetrische Matrizen anwenden, siehe beispielsweise Stoer & Bulirsch (2005).

□