Kapitel 7

Verfahren zur Eigenwert- und Eigenvektorberechnung

7.1 Eigenwertprobleme

In vielen technischen und physikalischen Problemen, etwa bei der Untersuchung des Schwingungsverhaltens von mechanischen oder elektrischen Systemen, ist es von Bedeutung, die Eigenwerte und Eigenvektoren einer Matrix $A \in \mathbb{C}^{n,n}$ zu bestimmen.

Aber auch bei der PageRank-Bestimmung der Suchmaschine google werden große Eigenwertprobleme gelöst.

Beispiel: Grundschwingungen und Resonanzfrequenzen schwingender Strukturen

Wir betrachten eine mechansiche Struktur (z.B. Karosserie, Brücke, Gebäude) und interessieren uns dafür, auf welchen Frequenzen sie schwingen kann und wie die zugehörigen Schwingungen aussehen (für elektrische Schaltkreise ist die Situation ähnlich). Dies Fragestellung ist z.B. bei der Schwingungs- und Lärmbekämpfung sowie bei der Auslegung von Bauwerken, Flugzeugen, etc. von großer Relevanz.

Sei jeweils $y_i(t) \in \mathbb{R}^3$ die Verschiebung der Struktur am Punkt $x_i \in \mathbb{R}^3$ zur Zeit t, $1 \leq i \leq n$. Im Falle einer ungedämpften Schwingung, die durch eine Kraft $f(t)$ angeregt wird erfüllt $y(t) = (y_i(t))_{1 \leq i \leq n}$ das Anfangswertproblem

$$My''(t) = -Ay(t) + f(t), \;\; y(0) = y^{(0)}, \;\; y'(0) = y^{(1)}.$$

mit der invertierbaren Massenmatrix $M \in \mathbb{R}^{3n,3n}$ und der Steifigkeitsmatrix $A \in \mathbb{R}^{3n,3n}$. Die Lösung ist die Summe aus einer Lösung der inhomogenen Gleichung und einer geeigneten Lösung der homogenen Gleichung

$$My''(t) = -Ay(t),$$

die äquivalent ist zu

$$y''(t) = -M^{-1}Ay(t).$$

Man kann zeigen, dass $B := M^{-1}A$ diagonalisierbar ist mit rellen Eigenwerten $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{3n}$ und zugehörigen Eigenvektoren $v_1, \ldots, v_{3n}$. Nun ist jede der Funktionen

$$\phi_i(t) := (a_i \sin(\sqrt{\lambda_i}t) + b_i \cos(\sqrt{\lambda_i}t))v_i$$

eine Lösung der homogenen Gleichung, denn wegen $Bv_i = \lambda_i v_i$ gilt

$$\phi_i''(t) = -\sqrt{\lambda_i}^2(a_i \sin(\sqrt{\lambda_i}t) + b_i \cos(\sqrt{\lambda_i}t))v_i = -\lambda_i\phi_i(t) = -B\phi_i(t).$$

Damit sind $\phi_i(t)$ die Grundschwingungen der Struktur, die i-te Grundschwingung hat also Frequenz $\sqrt{\lambda_i}/(2\pi)$ und die zugehörige Verformung der Struktur wird durch den Eigenvektor v_i gegeben. $\Box$

Beispiel: PageRank-Algorithmus von google

Betrachte N Webseiten. Webseite i enthalte k_i Links auf andere Seiten. Die Wahrscheinlichkeit, dass ein Nutzer von Seite i nach Seite j geht, wird modelliert als

$$p_{ij} = \begin{cases} \frac{\alpha}{k_i} + \frac{1-\alpha}{N}, & \text{falls Seite } i \text{ einen Link auf Seite } j \text{ enthält,} \\ \frac{1-\alpha}{N}, & \text{falls Seite } i \text{ keinen Link auf Seite } j \text{ enthält.} \end{cases}$$

Hierbei wird in der Regel $\alpha = 0,85$ gewählt. Sei nun $P = (p_{ij})_{1 \leq i,j \leq N}$. Als Gewichte der Seiten bestimmt man nun einen Vektor $\pi \in \mathbb{R}^N$, die sogenannte stationäre Verteilung, so dass gilt

$$\pi = P^T\pi, \quad \sum_{i=1}^{N} \pi_i = 1, \quad \pi_i \geq 0.$$

Anschauliche Erklärung: Ist π_i der Anteil der Internetnutzer, die sich im Mittel auf Seite i aufhalten, dann bleibt nach dem Übergangsverhalten gemäß den Wahrscheinlichkeiten p_{ij} dieser Anteil unverändert. Also gibt π_i den Anteil der Internetnutzer an, die sich nach Einstellen eines Gleichgewichtszustandes im Mittel auf Seite i befinden. $\Box$

7.1.1 Grundlagen

Definition 7.1.1 *Eine Zahl $\lambda \in \mathbb{C}$ heißt* Eigenwert *einer Matrix $A \in \mathbb{C}^{n,n}$, wenn es einen Vektor $x \in \mathbb{C}^n$, $x \neq 0$ gibt mit*

$$Ax = \lambda x.$$

Jeder solche Vektor $x \in \mathbb{C}^n$ heißt (Rechts-)Eigenvektor *zum Eigenwert λ. Die Menge $\sigma(A)$ aller Eigenwerte von A heißt* Spektrum *von A.* $\Box$

Der Unterraum

$$\mathrm{Eig}_A(\lambda) := \{x \in \mathbb{C}^n \;:\; (A - \lambda I)x = 0\}$$

ist der *Eigenraum* von A zum Eigenwert λ. Seine Dimension

$$\gamma(\lambda) := \dim \mathrm{Eig}_A(\lambda) = n - \mathrm{Rang}(A - \lambda I)$$

ist die *geometrische Vielfachheit* von λ und gibt die Maximalzahl linear unabhängiger Eigenvektoren zu λ an.

Offensichtlich ist λ genau dann Eigenwert von A, wenn gilt

$$\chi(\lambda) := \det(A - \lambda I) = 0,$$

also wenn λ Nullstelle des *charakteristischen Polynoms* $\chi(\mu)$ von A ist. χ ist ein Polynom n-ten Grades und hat die Form

$$\chi(\mu) = (-1)^n \mu^n + (-1)^{n-1} \mu^{n-1} \mathrm{spur}(A) + \cdots + \det(A).$$

Sind $\lambda_1, \ldots, \lambda_k$ die verschiedenen Nullstellen von χ (d.h. die verschiedenen Eigenwerte von A) mit Vielfachheiten ν_i, $i = 1, \ldots, k$, so gilt $\nu_1 + \ldots + \nu_k = n$ und χ hat die Linearfaktorzerlegung

$$\chi(\mu) = (-1)^n (\mu - \lambda_1)^{\nu_1} \cdots (\mu - \lambda_k)^{\nu_k}.$$

Man nennt $\nu(\lambda_i) = \nu_i$ die *algebraische Vielfachheit* von λ_i. Es ist nicht schwer zu zeigen, dass immer gilt

$$\gamma(\lambda_i) \le \nu(\lambda_i).$$

Wir fassen einige grundlegende Eigenschaften von Eigenwerten und Eigenvektoren zusammen:

Proposition 7.1.2 *Sei $A \in \mathbb{C}^{n,n}$ ein beliebig. Dann gilt:*

a) Ist λ Eigenwert von A, so ist λ Eigenwert von A^T und $\bar{\lambda}$ Eigenwert von $A^H := \bar{A}^T$.

b) Für jede nichtsinguläre Matrix $T \in \mathbb{C}^{n,n}$ hat die zu A ähnliche Matrix *$B := T^{-1}AT$ dasselbe charakteristische Polynom und dieselben Eigenwerte wie A. Ist x Eigenvektor von A, so ist $y := T^{-1}x$ Eigenvektor von B.*

c) Ist A hermitesch, *also $A^H = A$ mit $A^H := \bar{A}^T$, dann hat A lauter reelle Eigenwerte. Ist A* unitär, *also $A^H = A^{-1}$, so gilt $|\lambda| = 1$ für jeden Eigenwert λ.*

Eine Matrix $A \in \mathbb{C}^{n,n}$ heißt *diagonalisierbar*, wenn sie n linear unabhängige Eigenvektoren $x_1, \ldots, x_n$ besitzt. Die zugehörige Matrix $T := (x_1, \ldots, x_n)$ ist dann invertierbar und mit den Eigenwerten λ_i zu x_i gilt

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

Tatsächlich haben wir

$$AT = (\lambda_1 x_1, \ldots, \lambda_n x_n) = TD.$$

Eine wichtige Rolle spielen *hermitesche* Matrizen $A \in \mathbb{C}^{n,n}$, d.h. $A^H = A$, und mithin *reelle symmetrische* Matrizen. Man kann recht einfach zeigen, dass eine hermitesche Matrix $A \in \mathbb{C}^{n,n}$ immer *diagonalisierbar* ist mit einer *unitären* Matrix $T = U$, also

$$U^{-1}AU = D, \quad U^H = U^{-1}.$$

Ist $A = A^T$ reell, dann kann $U \in \mathbb{R}^{n,n}$ orthogonal gewählt werden, also

$$U^{-1}AU = D, \quad U^T = U^{-1}.$$

Die wichtigsten Verfahren zur Berechnung von Eigenwerten und Eigenvektoren einer Matrix A nehmen zunächst eine Reihe von Ähnlichkeitstransformationen

$$A^{(0)} := A, \quad A^{(k+1)} := T_k^{-1} A^{(k)} T_k, \quad k = 0, 1, \ldots$$

vor, um A in eine Matrix einfacherer Gestalt zu transformieren, deren Eigenwerte und Eigenvektoren leichter zu bestimmen sind.

7.1.2 Grundkonzepte numerischer Verfahren

Die im folgenden besprochenen numerischen Verfahren zur Berechnung von Eigenwerten und Eigenvektoren lassen sich in zwei Klassen einordnen. Die eine beruht auf der Vektoriteration, die andere auf der Anwendung von Ähnlichkeitstransformationen.

Vektoriteration

Bei der ersten Klasse von Verfahren handelt es sich um Vektoriterationen, die allgemein von der Form sind

$$x^{(k+1)} = \frac{Bx^{(k)}}{\|Bx^{(k)}\|}, \quad k = 0, 1, \ldots$$

mit einem Startvektor $x^{(0)}$, einer Iterationsmatrix B und einer Vektornorm $\|\cdot\|$.

Ähnlichkeitstransformation auf einfachere Gestalt

Nach Proposition 7.1.2 bleiben die Eigenwerte einer Matrix A bei einer Ähnlichkeitstransformation $B = T^{-1}AT$ unverändert und aus einem Eigenvektor y von B erhält man durch $x = Ty$ einen Eigenvektor der Ausgangsmatrix A.

Es liegt daher nahe, A durch Ähnlichkeitstransformationen

$$A^{(0)} := A \to A^{(1)} \to \ldots, \quad A^{(k+1)} = T_k^{-1} A^{(k)} T_k \tag{7.1}$$

in eine einfachere Form zu überführen, für die die Bestimmung von Eigenwerten und Eigenvektoren einfacher ist. Wir betrachten hier nur das QR-Verfahren, das eines der schnellsten Verfahren zur Lösung von Eigenwertproblemen darstellt.

QR-Verfahren: Beim QR-Verfahren wird durch Anwendung unitärer Matrizen T_i erreicht, dass die Elemente von $A^{(k)}$ im strikten unteren Dreieck gegen null konvergieren. Die Diagonaleinträge von $A^{(k)}$ konvergieren wiederum gegen die Eigenwerte von A.

7.1.3 Störungstheorie für Eigenwertprobleme

Bei oberen oder unteren Dreiecksmatrizen sind die Eigenwerte nichts anderes als die Diagonalelemente. Wir haben bereits angedeutet, dass das QR-Verfahren durch Ähnlichkeitstransformationen den Außerdiagonalteil bzw. das strikte untere Dreieck reduzieren. Störungsresultate für Eigenwerte liefern unter anderem Schranken, wie gut die Diagonalelemente mit den Eigenwerten übereinstimmen.

Wir haben das folgende fundamentale Resultat.

Satz 7.1.3 *Bezeichnet $\lambda_i(A)$, $i = 1, \ldots, n$, die angeordneten Eigenwerte einer Matrix $A \in \mathbb{C}^{n,n}$ (zum Beispiel aufsteigend nach Realteil und bei gleichem Realteil aufsteigend nach Imaginärteil), dann sind die Abbildungen*

$$A \in \mathbb{C}^{n,n} \mapsto \lambda_i(A), \quad i = 1, \ldots, n$$

stetig. Eigenwerte hängen also stetig von der Matrix ab.

Beweis: Siehe zum Beispiel Werner [We92]. □

Ein wichtiges Einschließungskriterium für Eigenwerte erhält man durch die *Gershgorin-Kreise*:

Satz 7.1.4 *Es sei $A = (a_{ij}) \in \mathbb{C}^{n,n}$ beliebig.*

a) Es gilt

$$\sigma(A) \subset \bigcup_{i=1}^{n} K_i$$

mit den Gershgorin-Kreisen

$$K_i := \left\{ \mu \in \mathbb{C} \ : \ |\mu - a_{ii}| \le \sum_{\substack{j=1 \\ j \ne i}}^{n} |a_{ij}| \right\}, \quad i = 1, \ldots, n.$$

b) *Ist die Vereinigung G_1 von k Gershgorin-Kreisen disjunkt von der Vereinigung G_2 der restlichen $n-k$ Gershgorin-Kreise, dann enthält G_1 genau k Eigenwerte und G_2 genau $n-k$ Eigenwerte von A.*

Das folgende Resultat gilt für diagonalisierbare Matrizen:

Satz 7.1.5 (Bauer/Fike)
Es sei $A \in \mathbb{C}^{n,n}$ diagonalisierbar, also

$$T^{-1}AT = \operatorname{diag}(\lambda_1, \dots, \lambda_n) =: D.$$

Dann gilt für jede Matrix $\Delta A \in \mathbb{C}^{n,n}$

$$\forall\, \mu \in \sigma(A + \Delta A): \quad \min_{i=1,\dots,n} |\mu - \lambda_i| \le \operatorname{cond}_2(T) \|\Delta A\|_2.$$

Hierbei ist $\|\cdot\|_2$ die von der euklidischen Norm induzierte Matrix-Norm und $\operatorname{cond}_2(T) := \|T\|_2 \|T^{-1}\|_2$ *die zugehörige Kondition von T.*

Bemerkung: Ist A hermitesch, so kann T unitär gewählt werden und es gilt $\operatorname{cond}_2(T) = 1$.

7.2 Die Vektoriteration

7.2.1 Definition und Eigenschaften der Vektoriteration

Definition 7.2.1 *Für eine Matrix $B \in \mathbb{C}^{n,n}$ ist die zugehörige Vektoriteration gegeben durch*

$$z^{(k+1)} = \frac{1}{\|Bz^{(k)}\|} Bz^{(k)}, \quad k = 0, 1, \dots \tag{7.2}$$

mit einem Startvektor $z^{(0)} \in \mathbb{C}^n \setminus \{0\}$.

Bei geeigneter Wahl von B ergeben sich hieraus Näherungen $z^{(k)}$ für einen Eigenvektor zu einem Eigenwert λ. Eine Eigenwertnäherung für λ erhalten wir dann durch den *Rayleigh-quotienten*

$$R(z^{(k)}, B) = \frac{(z^{(k)})^H B z^{(k)}}{(z^{(k)})^H z^{(k)}}.$$

Wir untersuchen die grundlegenden Eigenschaften für eine diagonalisierbare Matrix B mit Eigenwerten $\lambda_1, \dots, \lambda_n$. Wir sagen, dass ein Vektor $x \in \mathbb{C}^n$ *einen Anteil in* $\operatorname{Eig}_B(\lambda_i)$ hat, falls in der eindeutigen Darstellung

$$x = u + v, \quad u \in \operatorname{Eig}_B(\lambda_i), \quad v \in \bigoplus_{\lambda_j \ne \lambda_i} \operatorname{Eig}_B(\lambda_j)$$

gilt $u \ne 0$. u ist der Anteil von x in $\operatorname{Eig}_B(\lambda_i)$.

Satz 7.2.2 *Es sei $B \in \mathbb{C}^{n,n}$ diagonalisierbar mit Eigenwerten $\lambda_1, \ldots, \lambda_n$,*

$$\lambda_1 = \ldots = \lambda_r, \quad |\lambda_r| > |\lambda_{r+1}| \geq \ldots \geq |\lambda_n|$$

mit $r < n$. Falls der Startvektor $z^{(0)}$ einen Anteil in $\mathrm{Eig}_B(\lambda_1)$ besitzt, gilt für die Vektoriteration (7.2)

$$R(z^{(k)}, B) = \frac{(z^{(k)})^H B z^{(k)}}{(z^{(k)})^H z^{(k)}} = \lambda_1 + O(q^k) \quad \textit{für } k \to \infty, \quad q = \frac{|\lambda_{r+1}|}{|\lambda_1|} < 1.$$

Zudem gilt

$$z^{(k)} = \frac{\lambda_1^k}{|\lambda_1|^k} \frac{x_1}{\|x_1\|} + O(q^k), \;\; k \geq 1.$$

mit einer beliebigen Vektornorm $\|\cdot\|$, wobei x_1 den Anteil von $z^{(0)}$ in $\mathrm{Eig}_B(\lambda_1)$ bezeichnet.

Beweis: Wir können genausogut die nicht normierte Folge $\tilde{z}^{(k+1)} = B\tilde{z}^{(k)}$, $\tilde{z}^{(0)} = z^{(0)}$ betrachten. Es gilt dann $z^{(k)} = \tilde{z}^{(k)} / \|\tilde{z}^{(k)}\|$, $k \geq 1$.

Es gibt eine Darstellung der Form $z^{(0)} = x_1 + \sum_{j=r+1}^n x_j$ mit $x_j \in \mathrm{Eig}_B(\lambda_j)$, $x_1 \neq 0$. Einsetzen in $\tilde{z}^{(k+1)} = B\tilde{z}^{(k)}$ ergibt

$$\tilde{z}^{(k)} = B^k z^{(0)} = \lambda_1^k x_1 + \sum_{j=r+1}^n \lambda_j^k x_j = \lambda_1^k \left(x_1 + \sum_{j=r+1}^n \left(\frac{\lambda_j}{\lambda_1} \right)^k x_j \right), \quad k \geq 0. \tag{7.3}$$

Dies liefert

$$\tilde{z}^{(k)} = \lambda_1^k (x_1 + O(q^k))$$

und somit

$$\begin{aligned}
(\tilde{z}^{(k)})^H B \tilde{z}^{(k)} &= (\tilde{z}^{(k)})^H \tilde{z}^{(k+1)} = \bar{\lambda}_1^k \lambda_1^{k+1} (x_1 + O(q^k))^H (x_1 + O(q^k)) \\
&= \lambda_1 |\lambda_1|^{2k} (\|x_1\|_2^2 + O(q^k)) \\
(\tilde{z}^{(k)})^H \tilde{z}^{(k)} &= \bar{\lambda}_1^k \lambda_1^k (x_1 + O(q^k))^H (x_1 + O(q^k)) = |\lambda_1|^{2k} (\|x_1\|_2^2 + O(q^k)).
\end{aligned}$$

Wir erhalten

$$R(z^{(k)}, B) = R(\tilde{z}^{(k)}, B) = \lambda_1 \frac{\|x_1\|_2^2 + O(q^k)}{\|x_1\|_2^2 + O(q^k)} = \lambda_1 + O(q^k).$$

Analog haben wir

$$z^{(k)} = \frac{\tilde{z}^{(k)}}{\|\tilde{z}^{(k)}\|} = \frac{\lambda_1^k (x_1 + O(q^k))}{|\lambda_1|^k (\|x_1\| + O(q^k))} = \frac{\lambda_1^k}{|\lambda_1|^k} \frac{x_1}{\|x_1\|} + O(q^k)$$

□

Bemerkung: Selbst wenn $z^{(0)}$ keinen Anteil in $\mathrm{Eig}_B(\lambda_1)$ hat, was bei "genügend allgemeiner" Wahl vom $z^{(0)}$ unwahrscheinlich ist, so stellt sich in der Praxis diese Situation durch den Einfluß von Rundungsfehlern ein. $\Box$

Im Falle hermitescher Matrizen erhält man lineare Konvergenzrate q^2 des Rayleigh-Quotienten gegen λ_1.

Satz 7.2.3 *Sei $B \in \mathbb{C}^{n,n}$ hermitesch. Dann gilt unter den Voraussetzungen von Satz 7.2.2 für den Rayleigh-Quotienten die Konvergenzaussage*

$$R(z^{(k)}, B) = \frac{(z^{(k)})^H B z^{(k)}}{(z^{(k)})^H z^{(k)}} = \lambda_1 + O(q^{2k}) \quad \text{für } k \to \infty \text{ mit} \quad q = \frac{|\lambda_{r+1}|}{|\lambda_1|} < 1.$$

7.2.2 Die Vektoriterationen nach v. Mises und Wielandt

Sei $A \in \mathbb{C}^{n,n}$ gegeben. Unterschiedliche Varianten der Vektoriteration entstehen durch die Wahl der Iterationsmatrix B.

Einfache Vektoriteration nach von Mises

Die einfache Vektoriteration erhält man durch die naheliegende Wahl $B = A$. Die Konvergenzeigenschaften können dann unmittelbar Satz 7.2.2 bzw. 7.2.3 entnommen werden.

Inverse Vektoriteration von Wielandt

Offensichtliche Nachteile der Vektoriteration sind die langsame Konvergenz bei schlechter Trennung der Eigenwerte und die Einschränkung auf die Bestimmung des betragsmäßig größten Eigenwerts. Dies kann durch die inverse Vektoriteration von Wielandt vermieden werden. Man braucht hierzu eine gute Näherung μ eines Eigenwerts λ_j, so dass

$$|\lambda_j - \mu| \ll |\lambda_i - \mu|, \quad \text{für } \lambda_i \neq \lambda_j.$$

Dann hat für $\mu \neq \lambda_j$ die Matrix $B = (A - \mu I)^{-1}$ die Eigenwerte

$$\mu_i = \frac{1}{\lambda_i - \mu},$$

wobei $|\mu_j| \gg |\mu_i|$ für alle $\mu_i \neq \mu_j$. Ferner ist x_j genau dann Eigenvektor von B zum Eigenwert μ_j, wenn x_j Eigenvektor von A zum Eigenwert λ_j ist.

Die zugehörige inverse Iteration von Wielandt lautet dann

$$z^{(k+1)} = \frac{\hat{z}^{(k+1)}}{\|\hat{z}^{(k+1)}\|} \quad \text{mit } \hat{z}^{(k+1)} = (A - \mu I)^{-1} z^{(k)}.$$

In der Praxis bestimmt man nicht $(A-\mu I)^{-1}$, sondern implementiert die Iteration in der Form

$$\text{Löse } (A-\mu I)\hat{z}^{(k+1)} = z^{(k)} \quad \text{und setze} \quad z^{(k+1)} = \frac{\hat{z}^{(k+1)}}{\|\hat{z}^{(k+1)}\|}.$$

Die inverse Iteration von Wielandt hat dann im Falle

$$q := \max_{1\le i\le n, \lambda_i\neq\lambda_j} \frac{|\lambda_j-\mu|}{|\lambda_i-\mu|} < 1$$

nach Satz 7.2.2 die Konvergenzeigenschaften

$$R(z^{(k)},(A-\mu I)^{-1}) = \frac{(z^{(k)})^H \hat{z}^{(k+1)}}{(z^{(k)})^H z^{(k)}} = \frac{1}{\lambda_j-\mu} + O(q^k),$$

$$z^{(k)} = \frac{|\lambda_j-\mu|^k}{(\lambda_j-\mu)^k}\frac{x_j}{\|x_j\|} + O(q^k),$$

wobei x_j den Anteil von $z^{(0)}$ in $\mathrm{Eig}_A(\lambda_j) = \mathrm{Eig}_{(A-\mu I)^{-1}}(1/(\lambda_j-\mu))$ bezeichnet. Ist A zudem hermitesch, so erfüllt der Rayleigh-Quotient nach Satz 7.2.3

$$R(z^{(k)},(A-\mu I)^{-1}) = \frac{(z^{(k)})^H \hat{z}^{(k+1)}}{(z^{(k)})^H z^{(k)}} = \frac{1}{\lambda_j-\mu} + O(q^{2k}).$$

7.3 Das QR-Verfahren

Das im folgenden beschriebene QR-Verfahren von Francis bildet die Basis sehr leistungsfähiger Verfahren zur Eigenwert- und Eigenvektorberechnung. Ausgehend von einer Matrix $A^{(1)} = A \in \mathbb{C}^{n,n}$ führt man beim QR-Verfahren unitäre Ähnlichkeitstransformationen folgender Form durch:

Algorithmus 6 QR-Verfahren
Sei $A \in \mathbb{C}^{n,n}$ eine gegebene Matrix.

0. Setze $A^{(1)} := A$.

1. Für $l = 1, 2, \ldots$: Berechne

$$\begin{aligned} A^{(l)} &=: Q_l R_l, \quad Q_l \in \mathbb{C}^{n,n} \text{ unitär}, \quad R_l \in \mathbb{C}^{n,n} \text{ obere Dreiecksmatrix}, \\ A^{(l+1)} &:= R_l Q_l. \end{aligned} \tag{7.4}$$

In jedem Schritt ist also die Berechnung einer QR-Zerlegung

$$A^{(l)} = Q_l R_l, \quad R_l \in \mathbb{C}^{n,n} \text{ obere Dreiecksmatrix}, \quad Q_l \in \mathbb{C}^{n,n} \text{ unitär, also } Q_l^H = Q_l^{-1}$$

erforderlich. Eine solche Zerlegung kann mit Hilfe des Householder-Verfahrens berechnet werden, das wir für Interessierte am Ende dieses Kapitels kurz beschreiben.

7.3.1 Grundlegende Eigenschaften des QR-Verfahrens

Wir beginnen mit der offensichtlichen Feststellung, dass (7.4) tatsächlich eine Folge unitär ähnlicher Matrizen $A^{(l)}$ erzeugt.

Lemma 7.3.1 *Es seien Q_l und R_l von Algorithmus 6 erzeugt. Dann gilt mit den Bezeichnungen $Q_{1\ldots l} := Q_1 Q_2 \cdots Q_l$, $R_{l\ldots 1} := R_l R_{l-1} \cdots R_1$*

$$A^{(l+1)} = Q_l^{-1} A^{(l)} Q_l = Q_{1\ldots l}^{-1} A Q_{1\ldots l}, \quad l = 1, 2, \ldots.$$

Beweis: Wegen (7.4) ist $R_l = Q_l^{-1} A^{(l)}$ und daher

$$A^{(l+1)} = R_l Q_l = Q_l^{-1} A^{(l)} Q_l.$$

Induktiv ergibt sich

$$A^{(l+1)} = Q_l^{-1} \cdots Q_1^{-1} A^{(1)} Q_1 \cdots Q_l = Q_{1\ldots l}^{-1} A Q_{1\ldots l}.$$

□

7.3.2 Konvergenz des QR-Verfahrens

Wir geben zunächst ein Resultat für Matrizen mit betragsmäßig getrennten Eigenwerten an. Unter gewissen Voraussetzungen konvergiert dann die vom QR-Verfahren generierte Folge $A^{(l)}$ nach unitärer Diagonalskalierung der Form $S_l^{-1} A^{(l)} S_l$ gegen eine obere Dreiecksmatrix U, wobei die Konvergenzgeschwindigkeit von der Trennung der Beträge der Eigenwerte abhängt.

Satz 7.3.2 *Die Matrix $A \in \mathbb{C}^{n,n}$ sei regulär mit betragsmäßig getrennten Eigenwerten $\lambda_1, \ldots, \lambda_n$,*

$$|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n|.$$

Weiter seien $v_1, \ldots, v_n$ zugehörige Eigenvektoren und die Inverse der Matrix $T = (v_1, \ldots, v_n)$ besitze ohne Zeilenvertauschung eine LR-Faktorisierung. Dann gilt für das in Algorithmus 6 angegebene QR-Verfahren

$$A^{(l)} = S_l U S_l^{-1} + O(q^{l-1}) \quad \textit{für } l \to \infty, \quad q := \max_{j=1,\ldots,n-1} \left| \frac{\lambda_{j+1}}{\lambda_j} \right|$$

mit einer oberen Dreiecksmatrix

$$U = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ & \ddots & \ddots & \vdots \\ & & \ddots & * \\ & & & \lambda_n \end{pmatrix}$$

und unitären Phasenmatrizen $S_l = \operatorname{diag}(\sigma_1^{(l)}, \dots, \sigma_n^{(l)})$, $|\sigma_i^{(l)}| = 1$. *Insbesondere gilt mit den Diagonaleinträgen* $a_{11}^{(l)}, \dots, a_{nn}^{(l)}$ *von* $A^{(l)}$

$$|a_{ii}^{(l)} - \lambda_i| = O(q^{l-1}).$$

Beweis: Siehe zum Beispiel Plato [Pl00]. $\square$

Bemerkungen:

- Die zugehörigen Eigenvektoren kann man zum Beispiel durch Inverse Vektoriteration berechnen, wobei man jeweils die Diagonalelemente von $A^{(l)}$ als Shifts μ verwendet.

- Hat T^{-1} lediglich eine LR-Faktorisierung mit Zeilenvertauschungen, dann konvergiert das QR-Verfahren nach wie vor, die Eigenwerte erscheinen in der Diagonale der Grenzmatrix U jedoch unter Umständen in anderer Reihenfolge.

- Sind nicht alle Eigenwerte betragsmäßig getrennt, also etwa

  $$|\lambda_1| > \ldots > |\lambda_r| = |\lambda_{r+1}| > \ldots > |\lambda_n|,$$

  was zum Beispiel eintritt, wenn eine reelle Matrix A konjugiert komplexe Eigenwerte hat, dann konvergiert $S_l^{-1} A^{(l)} S_l$ mit Phasenmatrizen S_l außerhalb des mit $\times$ markierten Bereichs gegen eine Matrix der Form

  $$\begin{pmatrix} \lambda_1 & \cdots & * & \times & \times & * & \cdots & \\ & \ddots & \vdots & \vdots & \vdots & \vdots & & \\ & & \lambda_{r-1} & \times & \times & * & \cdots & \\ & & & \times & \times & * & \cdots & \\ & & & \times & \times & * & \cdots & \\ & & & & & \lambda_{r+2} & & \\ & & & & & & \ddots & \\ & & & & & & & \lambda_n \end{pmatrix}.$$

  Die Eigenwerte des Blocks $\begin{pmatrix} a_{r,r}^{(l)} & a_{r,r+1}^{(l)} \\ a_{r+1,r}^{(l)} & a_{r+1,r+1}^{(l)} \end{pmatrix}$ konvergieren gegen λ_r und λ_{r+1}.

- Die Konvergenz des QR-Verfahrens ist sehr langsam, wenn die Trennung der Eigenwerte schlecht ist. Die Konvergenz der letzten Zeile gegen $(0, \dots, 0, \lambda_n)$ kann durch *Shift*-Techniken entscheidend verbessert werden, auf die wir nun kurz eingehen.

7.3.3 Shift-Techniken

Eine genauere Analyse zeigt, dass die letzte Zeile von $A^{(l)}$ die Form hat $(O(|\lambda_n/\lambda_{n-1}|^{l-1}), a_{nn}^{(l)})$. Ist also $|\lambda_n| \ll |\lambda_{n-1}|$, dann konvergiert $a_{n,j}^{(l)}$, $1 \le j < n$ sehr schnell gegen 0 und

$a_{nn}^{(l)}$ sehr schnell gegen λ_n. Nach genauer Bestimmung von λ_n kann man dann mit dem $(n-1)\times(n-1)$-Block von $A^{(l)}$ zur Bestimmung von λ_{n-1} fortfahren.

Um die Trennung von λ_n und λ_{n-1} zu verbessern, wendet man das QR-Verfahren in jedem Schritt auf $A^{(l)} - \mu_l I$ an mit $\mu_l \approx \lambda_n$ und korrigiert den Shift anschließend. Anstelle von (7.4) berechnet man also mit einem *Shift* $\mu_l \approx \lambda_n$

$$\begin{aligned} &A^{(l)} - \mu_l I =: Q_l R_l, \quad Q_l \in \mathbb{C}^{n,n} \text{ unitär}, \quad R_l \in \mathbb{C}^{n,n} \text{ obere Dreiecksmatrix}, \\ &\quad A^{(l+1)} := R_l Q_l + \mu_l I. \end{aligned}$$

Man prüft leicht nach, dass wieder gilt $A^{(l+1)} = Q_l^{-1} A^{(l)} Q_l$.

Verbreitete Shift-Strategie: Eine effizente Shift-Strategie erhält man, wenn man μ_l als denjenigen Eigenwert von $\begin{pmatrix} a_{n-1,n-1}^{(l)} & a_{n-1,n}^{(l)} \\ a_{n,n-1}^{(l)} & a_{n,n}^{(l)} \end{pmatrix}$ wählt, der am nächsten bei $a_{n,n}^{(l)}$ liegt. Im Zweifelsfall wähle den mit positivem Imaginärteil.

Das QR-Verfahren mit Shift liefert recht schnell eine Matrix $A^{(l)}$, deren letzte Zeile auf hohe Genauigkeit mit $(0,\dots,0,\lambda_n)$ übereinstimmt. Man wendet nun das QR-Verfahren mit Shift auf den oberen linken $(n-1)\times(n-1)$-Block von $A^{(l)}$ zur Bestimmung von λ_{n-1} an und so fort.

Bemerkung: Das QR-Verfahren mit Shift gilt zur Zeit als eines der besten Iterationsverfahren zur Lösung des vollständigen Eigenwertproblems.

Berechnung der Eigenvektoren: Die Eigenvektoren kann man nun zum Beispiel wieder durch Inverse Vektoriteration bestimmen, wobei man als Shifts μ die vom QR-Verfahren berechneten Eigenwerte verwendet.

7.3.4 Berechnung einer QR-Zerlegung (Ergänzung für Interessierte)

Wir geben zum Abschluss ein numerisches Verfahren an zur Berechnung einer

QR-Zerlegung:

Für $B \in \mathbb{C}^{n,n}$ bestimme eine unitäre Matrix $Q \in \mathbb{C}^{n,n}$ und eine obere Dreiecksmatrix $R \in \mathbb{C}^{n,n}$ mit

$$B = QR. \tag{7.5}$$

Householder-Verfahren zur Berechnung einer QR-Zerlegung:

Beim Householder-Verfahren berechnet man (7.5) in $n-1$ Schritten:

Initialisierung:

$$B^{(0)} := B = \left(\begin{array}{c|cc} & * & \cdots \\ b^{(0)} & \vdots & \\ & * & \cdots \end{array}\right)$$

Schritt 0: Bestimme eine unitäre Matrix T_0 (siehe (7.7), (7.8)) mit

$$B^{(1)} := T_0 B^{(0)} = \left(\begin{array}{c|ccc} * & * & * & \cdots \\ \hline 0 & * & * & \cdots \\ \vdots & \vdots & \vdots & \\ 0 & * & * & \cdots \end{array}\right) =: \left(\begin{array}{c|c|c} B_1^{(1)} & \multicolumn{2}{c}{B_2^{(1)}} \\ \hline 0 & & \\ \vdots & b^{(1)} & B_3^{(1)} \\ 0 & & \end{array}\right)$$

Schritt 1: Bestimme eine unitäre Matrix T_1 (siehe (7.7), (7.8)) mit

$$B^{(2)} := T_1 B^{(1)} = \left(\begin{array}{cc|ccc} * & * & * & * & \cdots \\ 0 & * & * & * & \cdots \\ \hline 0 & 0 & * & * & \cdots \\ \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & * & * & \cdots \end{array}\right) =: \left(\begin{array}{cc|c|c} \multicolumn{2}{c|}{B_1^{(2)}} & \multicolumn{2}{c}{B_2^{(2)}} \\ \hline 0 & 0 & & \\ \vdots & \vdots & b^{(2)} & B_3^{(2)} \\ 0 & 0 & & \end{array}\right)$$

Schritt k, $k = 2, \ldots, n-2$: Bestimme eine unitäre Matrix T_k (siehe (7.7), (7.8)) mit

$$(7.6) \qquad B^{(k+1)} := T_k B^{(k)} = \left(\begin{array}{ccc|ccc} * & \cdots & * & * & & \cdots \\ & \ddots & \vdots & & & \\ 0 & & * & * & & \cdots \\ \hline 0 & \cdots & 0 & * & & \cdots \\ \vdots & & \vdots & \vdots & & \\ 0 & \cdots & 0 & * & & \cdots \end{array}\right) \begin{array}{l} \left.\vphantom{\begin{array}{c}*\\ \ddots\\ 0\end{array}}\right\} k+1 \\ \left.\vphantom{\begin{array}{c}0\\ \vdots\\ 0\end{array}}\right\} n-(k+1) \end{array}$$

$$= \left(\begin{array}{ccc|c|c} & B_1^{(k+1)} & & \multicolumn{2}{c}{B_2^{(k+1)}} \\ \hline 0 & \cdots & 0 & & \\ \vdots & & \vdots & b^{(k+1)} & B_3^{(k+1)} \end{array}\right).$$

Ergebnis: $R := B^{(n-1)}$, $Q := (T_{n-2} \cdot \ldots \cdot T_0)^H = T_0^H \cdot \ldots \cdot T_{n-2}^H$.

Rechtfertigung des Verfahrens:

Dann gilt tatsächlich

$R = B^{(n-1)} =$ obere Dreiecksmatrix, $\quad Q = T_0^H \cdot \ldots \cdot T_{n-2}^H$ unitär als Produkt unitärer Matrizen und

$$R = B^{(n-1)} = \underbrace{T_{n-2} \cdot \ldots \cdot T_0}_{=Q^H} B = Q^H B, \quad \text{also} \quad QR = B.$$

Berechnung der Transformationen T_k:

Es bleibt, die Berechnung von T_k anzugeben. Beim Householder-Verfahren wählt man jeweils T_k von der Form

(7.7)
$$T_k = \left(\begin{array}{c|c} I_k & 0 \\ \hline 0 & H_k \end{array}\right)$$

mit

$$I_k = \text{Einheitsmatrix in } \mathbb{R}^{k,k},$$

und $H_k \in \mathbb{R}^{n-k,n-k}$ als **Householder Transformation** der Form

(7.8)
$$H_k = I - \frac{2}{w_k^H w_k} w_k w_k^H, \quad w_k = b^{(k)} + \sigma_k \|b^{(k)}\|_2 \begin{pmatrix} 1 \\ 0 \\ \vdots \end{pmatrix}, \quad \sigma_k = \begin{cases} 1 & \text{falls } b_1^{(k)} = 0, \\ \frac{b_1^{(k)}}{|b_1^{(k)}|} & \text{sonst.} \end{cases}$$

Man kann zeigen, dass mit dieser Wahl gilt

$$H_k \text{ unitär und hermitesch}, \quad H_k b^{(k)} = \begin{pmatrix} \omega_k \|b^{(k)}\|_2 \\ 0 \\ \vdots \end{pmatrix}, \quad \omega_k \in \mathbb{C}, \ |\omega_k| = 1.$$

Man sieht leicht, dass dann tatsächlich jeweils $B^{(k+1)}$ die Form (7.6) hat.