

DAS QZ-VERFAHREN

BACHELOR-ARBEIT
IM 1-FACH BACHELORSTUDIENGANG MATHEMATIK
DER MATHEMATISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT
DER CHRISTIAN-ALBRECHTS-UNIVERSITÄT ZU KIEL

VORGELEGT VON
JANINA GNUTZMANN

ERSTGUTACHTER: PROF. DR. STEFFEN BÖRM
ZWEITGUTACHTER: DIPL.-MATH. JENS BURMEISTER

KIEL IM JANUAR 2012

Inhaltsverzeichnis

1	Einleitung	3
2	Theoretische Grundlagen	5
2.1	Gewöhnliche Eigenwertprobleme	5
2.1.1	Eigenwerte und Eigenräume	5
2.1.2	Die Schur-Zerlegung	6
2.2	Das verallgemeinerte Eigenwertproblem	9
2.2.1	Die verallgemeinerte Schur-Zerlegung	12
3	Gewöhnliche Eigenwertverfahren	15
3.1	Vektoriteration	15
3.2	QR-Iteration	18
4	Das QZ-Verfahren	21
4.1	Hessenberg-Dreiecks-Form	21
4.2	Deflation	24
4.3	Der QZ-Schritt	26
4.3.1	Der QZ-Schritt mit Doppel-Shift nach Moler und Stewart	29
4.4	Das QZ-Verfahren	34
4.5	Beispiele	36
5	Ausblicke	39
5.1	Verallgemeinerte Vektoriteration	39
5.2	Selbstadjungierte positiv definite verallgemeinerte Eigenwertprobleme	39
	Literatur	41
	Erklärung	43

1 Einleitung

In vielen Bereichen der Naturwissenschaften spielen Eigenwertprobleme eine große Rolle, zum Beispiel bei der Untersuchung des Schwingungsverhaltens von Konstruktionen. Da sich diese Fragestellungen mathematisch nicht exakt lösen lassen, benötigt für die Lösungen man numerische Verfahren. Dabei geht es darum, das sogenannte (*gewöhnliche*) *Eigenwertproblem*

$$Ax = \lambda x$$

zu lösen, wobei A eine $(n \times n)$ -Matrix, x ein Vektor und λ eine Unbekannte ist. Oftmals ist allerdings auch nach Lösungen x und λ des sogenannten *verallgemeinerten Eigenwertproblems*

$$Ax = \lambda Bx,$$

mit $(n \times n)$ -Matrizen A und B , gefragt. Zum Beispiel in verschiedenen Bereichen der Physik, wie in der Mechanik oder bei Schwingungen, wenn mehrere Kräfte wirken.

Das Ziel dieser Bachelor-Arbeit ist die Präsentation eines Algorithmus zum Lösen des verallgemeinerten Eigenwertproblems, das sogenannte QZ-Verfahren. Dabei beschränkt sich diese Arbeit auf ein Verfahren mit einer Doppel-Shift-Strategie nach Moler und Stewart. In Kapitel 2 geht es vorwiegend darum, einen kurzen mathematischen Überblick über die Welt der gewöhnlichen und verallgemeinerten Eigenwertprobleme zu verschaffen. In Kapitel 3 werden kurz zwei Verfahren zur Lösung gewöhnlicher Eigenwertprobleme vorgestellt, die Vektoriteration und das QR-Verfahren, welches die Grundlage des in Kapitel 4 beschriebenen QZ-Verfahrens darstellt. Abschließend wird in Kapitel 5 eine weitere Möglichkeit zur Lösung des verallgemeinerten Eigenwertproblems vorgestellt und auf einen wichtigen Spezialfall eingegangen.

Eine Implementierung des QZ-Verfahrens ist dieser Arbeit beigelegt, die in Kapitel 4.4 vorgestellten Beispiele wurden mit diesem Programm berechnet.

2 Theoretische Grundlagen

In diesem Kapitel werden die mathematischen Grundlagen für Eigenwertprobleme behandelt. Es sei immer $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und $n, m \in \mathbb{N}$.

2.1 Gewöhnliche Eigenwertprobleme

2.1.1 Eigenwerte und Eigenräume

Wir werden hier kurz grundlegende Begriffe und Ergebnisse aus der Linearen Algebra wiederholen, die für diese Arbeit nötig sind.

Sei $A \in \mathbb{K}^{n \times n}$ eine Matrix. Ein Element $\lambda \in \mathbb{K}$ heißt genau dann *Eigenwert* von A , wenn es einen Vektor $x \in \mathbb{K}^n \setminus \{0\}$ gibt, der die Gleichung $Ax = \lambda x$ erfüllt. Den Vektor x bezeichnet man als *Eigenvektor* der Matrix A zum Eigenwert λ . Die Menge $\sigma(A) := \{\lambda \in \mathbb{K} : \lambda \text{ ist Eigenwert von } A\}$ heißt *Spektrum* von A . Weiter bezeichnet man $\mathcal{E}_\lambda(A) := \text{Kern}(A - \lambda I)$ als den *Eigenraum* von A zum Eigenwert λ , welcher von den Eigenvektoren des Eigenwerts λ aufgespannt wird.

Das Polynom $\chi_A(\lambda) := \det(A - \lambda I)$ n -ten Grades heißt *charakteristisches Polynom* von A . Mit Hilfe dieses Polynoms können wir die Eigenwerte von A gerade als die Nullstellen von χ_A charakterisieren. Betrachten wir den komplexwertigen Fall, so liefert uns der Fundamentalsatz der Algebra genau n Nullstellen von χ_A und somit n Eigenwerte. Im Fall $\mathbb{K} = \mathbb{C}$ gilt also $\sigma(A) \neq \emptyset$.

Für Blockmatrizen können wir die folgende Aussage treffen, die für die in Kapitel 3 und 4 vorgestellten Verfahren wichtig ist.

Lemma 2.1. Sei $A \in \mathbb{K}^{n \times n}$ mit

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}, \quad (2.1)$$

wobei $A_{11} \in \mathbb{K}^{p \times p}$, $A_{22} \in \mathbb{K}^{q \times q}$, $p + q = n$. Dann gilt $\sigma(A) = \sigma(A_{11}) \cup \sigma(A_{22})$.

Beweis. Es gilt

$$\det(A - \lambda I) = \det(A_{11} - \lambda I_p) \det(A_{22} - \lambda I_q)$$

und somit

$$\chi_A(\lambda) = \chi_{A_{11}}(\lambda) \chi_{A_{22}}(\lambda).$$

2 Theoretische Grundlagen

Daraus erhalten wir

$$\chi_A(\lambda) = 0 \quad \Leftrightarrow \quad \chi_{A_{11}}(\lambda) = 0 \quad \text{oder} \quad \chi_{A_{22}}(\lambda) = 0$$

und damit folgt aus

$$\lambda \in \sigma(A) \quad \Leftrightarrow \quad \lambda \in \sigma(A_{11}) \quad \text{oder} \quad \lambda \in \sigma(A_{22})$$

die Behauptung. □

2.1.2 Die Schur-Zerlegung

Die in diesem Abschnitt gegebenen Resultate basieren auf [1].

Definition 2.2 (Orthogonal und unitär). *Sei $Q \in \mathbb{K}^{n \times m}$. Falls $Q^*Q = I$ gilt, so nennen wir Q orthogonal. Gilt zudem $n = m$, so heißt Q unitär.*

Im folgenden Lemma stellen wir hilfreiche Eigenschaften für diese Klasse von Matrizen vor.

Lemma 2.3. *Sei $Q \in \mathbb{K}^{n \times m}$ orthogonal. Dann ist Q injektiv und es gilt $\|Qx\|_2 = \|x\|_2$ für alle $x \in \mathbb{K}^m$. Ist $n = m$, also Q unitär, so ist auch Q^* unitär und es ist insbesondere Q^* die Inverse von Q .*

Beweis. Sei $x \in \mathbb{K}^m$. Nach Definition der euklidischen Norm gilt

$$\|Qx\|_2^2 = \langle Qx, Qx \rangle = \langle Q^*Qx, x \rangle = \langle x, x \rangle = \|x\|_2^2.$$

Insbesondere kann $Qx = 0$ nur dann erfüllt sein, wenn bereits $x = 0$ erfüllt ist, also ist der Kern von Q trivial und die Matrix deshalb injektiv.

Sei nun Q unitär, die Matrix sei also quadratisch. Da Q injektiv ist, ist Q invertierbar und es gilt nun

$$Q^{-1}Q = I = Q^*Q.$$

Da die Inverse eindeutig bestimmt ist, folgt $Q^* = Q^{-1}$ und somit $QQ^* = I$. Damit ist Q^* ebenfalls unitär. □

Lemma 2.4 (Basisergänzung). *Sei $Q \in \mathbb{K}^{n \times m}$ orthogonal, $m < n \in \mathbb{N}$. Dann existiert eine unitäre Matrix $P \in \mathbb{K}^{n \times n}$, die $P|_{n \times m} = Q$ erfüllt.*

2.1 Gewöhnliche Eigenwertprobleme

Beweis. Seien q_1, \dots, q_m die Spaltenvektoren der Matrix Q . Nach Definition 2.2 ist dann $\{q_1, \dots, q_m\}$ eine Menge paarweise orthogonaler normierter Vektoren, die wir zu einer orthonormalen Basis $\{q_1, \dots, q_n\}$ ergänzen können. Wir definieren die Matrix P so, dass für alle $i \in \{1, \dots, n\}$ ihre i -te Spalte gerade q_i ist. \square

Folgendes Resultat besagt, dass eine Matrix durch unitäre Ähnlichkeitstransformationen auf Block-Dreiecks-Gestalt gebracht werden kann, sofern wir einen invarianten Unterraum kennen.

Lemma 2.5. *Seien $A \in \mathbb{K}^{n \times n}$, $p \leq n$, $\Lambda \in \mathbb{K}^{p \times p}$ und $X \in \mathbb{K}^{n \times p}$ orthogonal derart, dass $AX = X\Lambda$ gilt. Dann existieren eine unitäre Matrix $Q \in \mathbb{K}^{n \times n}$ sowie Matrizen $D \in \mathbb{K}^{(n-p) \times (n-p)}$ und $R \in \mathbb{K}^{p \times (n-p)}$ derart, dass die Gleichung*

$$Q^*AQ = \begin{pmatrix} \Lambda & R \\ 0 & D \end{pmatrix} \quad (2.2)$$

erfüllt ist und $Q|_{n \times p} = X$ gilt.

Beweis. Gemäß Lemma 2.4 können wir X zu einer unitären Matrix Q mit $Q|_{n \times p} = X$ ergänzen. Wir zerlegen Q dementsprechend in

$$Q = \begin{pmatrix} X & Y \end{pmatrix}$$

und stellen fest, dass auch $Y \in \mathbb{K}^{n \times (n-p)}$ orthogonal ist und aus

$$\begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} = I = Q^*Q = \begin{pmatrix} X^* \\ Y^* \end{pmatrix} \begin{pmatrix} X & Y \end{pmatrix} = \begin{pmatrix} X^*X & X^*Y \\ Y^*X & Y^*Y \end{pmatrix}$$

die Identität $Y^*X = 0$ folgt.

Ferner erhalten wir

$$\begin{aligned} Q^*AQ &= \begin{pmatrix} X^* \\ Y^* \end{pmatrix} A \begin{pmatrix} X & Y \end{pmatrix} = \begin{pmatrix} X^*AX & X^*AY \\ Y^*AX & Y^*AY \end{pmatrix} \\ &= \begin{pmatrix} X^*X\Lambda & X^*AY \\ Y^*X\Lambda & Y^*AY \end{pmatrix} = \begin{pmatrix} \Lambda & X^*AY \\ 0 & Y^*AY \end{pmatrix}. \end{aligned}$$

aus der Gleichung $AX = X\Lambda$. Die Matrizen

$$R := X^*AY, \quad D := Y^*AY$$

liefern nun das Gewünschte. \square

2 Theoretische Grundlagen

Bei wiederholter Anwendung dieses Lemmas erhalten wir die sogenannte *Schur-Zerlegung* einer Matrix.

Satz 2.6 (Schur-Zerlegung). *Sei $A \in \mathbb{C}^{n \times n}$, dann existieren eine unitäre Matrix $Q \in \mathbb{C}^{n \times n}$ und eine obere Dreiecksmatrix $D \in \mathbb{C}^{n \times n}$, sodass*

$$Q^* A Q = D \tag{2.3}$$

erfüllt ist.

Beweis. Wir führen Induktion nach der Dimension n der Matrix A . Der Induktionsanfang $n = 1$ ist klar, da A in diesem Fall eine obere Dreiecksmatrix ist und $Q = I$ das Gewünschte liefert. Angenommen, die Aussage gilt für alle Matrizen der Dimension $n - 1$ oder kleiner. Nach dem Fundamentalsatz der Algebra existiert mindestens ein Eigenwert $\lambda \in \mathbb{C}$ von A . Sei $x \in \mathbb{C}^n \setminus \{0\}$ ein zugehöriger Eigenvektor mit $\|x\|_2 = 1$. Indem wir x als Matrix $X \in \mathbb{C}^{n \times 1}$ und λ als Matrix $\Lambda \in \mathbb{C}^{1 \times 1}$ interpretieren, gilt $A X = X \Lambda$ und wir erhalten mit Lemma 2.5 eine unitäre Matrix $Q_1 \in \mathbb{C}^{n \times n}$ und Matrizen $R_1 \in \mathbb{C}^{1 \times (n-1)}$ und $A_1 \in \mathbb{C}^{(n-1) \times (n-1)}$, sodass

$$Q_1^* A Q_1 = \begin{pmatrix} \lambda & R_1 \\ 0 & A_1 \end{pmatrix}$$

gilt. Nun wenden wir die Induktionsvoraussetzung auf A_1 an und erhalten eine unitäre Matrix $Q_2 \in \mathbb{C}^{(n-1) \times (n-1)}$ und eine obere Dreiecksmatrix $R_2 \in \mathbb{C}^{(n-1) \times (n-1)}$ mit

$$Q_2^* A_1 Q_2 = R_2.$$

Definiere nun

$$Q := Q_1 \begin{pmatrix} 1 & 0 \\ 0 & Q_2 \end{pmatrix}, R := \begin{pmatrix} \lambda & R_1 Q_2 \\ 0 & R_2 \end{pmatrix}.$$

Nach Konstruktion ist Q eine unitäre Matrix, da Q_1 und Q_2 unitär sind, und R ist eine obere Dreiecksmatrix, da R_2 eine obere Dreiecksmatrix ist.

Wir erhalten somit

$$\begin{aligned} Q^* A Q &= \begin{pmatrix} 1 & 0 \\ 0 & Q_2^* \end{pmatrix} Q_1^* A Q_1 \begin{pmatrix} 1 & 0 \\ 0 & Q_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & Q_2^* \end{pmatrix} \begin{pmatrix} \lambda & R_1 \\ 0 & A_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & Q_2 \end{pmatrix} \\ &= \begin{pmatrix} \lambda & R_1 Q_2 \\ 0 & Q_2^* A_1 Q_2 \end{pmatrix} = \begin{pmatrix} \lambda & R_1 Q_2 \\ 0 & R_2 \end{pmatrix} = R. \end{aligned}$$

□

2.2 Das verallgemeinerte Eigenwertproblem

Wir wenden uns nun dem *verallgemeinerten Eigenwertproblem* $Ax = \lambda Bx$, mit $A, B \in \mathbb{K}^{n \times n}$, $x \in \mathbb{K}^n$ und $\lambda \in \mathbb{K}$, zu. Dies ist, wie der Name schon sagt, eine Verallgemeinerung des gewöhnlichen Eigenwertproblems $Ax = \lambda x$ und kann darauf zurückgeführt werden, indem man $B = I$ setzt. In diesem Abschnitt widmen wir uns der Untersuchung der mathematischen Eigenschaften des Problems. Die dabei gegebenen Definitionen und Resultate basieren auf Stewarts Abschnitt 2.4 in [5].

Definition 2.7 (Büschel). *Seien $A, B \in \mathbb{K}^{n \times n}$ sowie λ eine Unbestimmte über \mathbb{K} . Dann heißt die lineare Funktion $A - \lambda B$ Büschel und wird mit (A, B) bezeichnet.*

Wie im gewöhnlichen Fall können wir Eigenwerte und Eigenvektoren für ein Büschel definieren.

Definition 2.8 (Spektrum, Eigenwerte, Eigenvektoren). *Seien $A, B \in \mathbb{K}^{n \times n}$. Die Eigenwerte des Büschels (A, B) sind Elemente der Menge*

$$\sigma(A, B) := \{\mu \in \mathbb{K} : \det(A - \mu B) = 0\}. \quad (2.4)$$

*Dabei bezeichnet $\sigma(A, B)$ das Spektrum des Büschels (A, B) . Für $\mu \in \mathbb{K}$, $x \in \mathbb{K}^n \setminus \{0\}$ heißt das Paar (μ, x) rechtes Eigenpaar des Büschels (A, B) , falls $Ax = \mu Bx$ gilt oder linkes Eigenpaar, falls $x^*A = \mu x^*B$ gilt, der Vektor x heißt Eigenvektor.*

In Zukunft sei (A, B) stets ein Büschel mit $A, B \in \mathbb{K}^{n \times n}$. Wie bei dem normalen Eigenwertproblem meinen wir mit Eigenpaaren rechte Eigenpaare, sofern es nicht anders erwähnt wird.

Ein anschaulicher Unterschied zwischen dem gewöhnlichen und dem verallgemeinerten Eigenwertproblem ist, dass sich die geometrische Anschauung des verallgemeinerten Eigenvektors von der des normalen Eigenvektors unterscheidet. Ist nämlich (λ, x) ein Eigenpaar der Matrix A , so wird x durch A auf ein Vielfaches von x abgebildet. Dies bedeutet, dass die Richtung des Vektors invariant unter Multiplikation mit A ist. In dem Fall des verallgemeinerten Eigenwertproblems bleibt die Richtung eines Eigenvektors x unter Multiplikation mit A oder B nicht unbedingt erhalten. Stattdessen sind die Richtungen von Ax und Bx bis auf Vorzeichen dieselben, die Vektoren unterscheiden sich nur in der Länge durch den Faktor des zugehörigen Eigenwertes λ .

Ein weiterer und wichtiger Unterschied zwischen dem verallgemeinerten und dem gewöhnlichen Eigenwertproblem ist, dass nun der Fall einer singulären Matrix B auftreten kann, während es die Matrix $B = I$ in dem gewöhnlichen Problem nicht ist. Dies hat

2 Theoretische Grundlagen

mehrere Auswirkungen.

Ist B singulär, so ist es möglich, dass jede beliebige Zahl λ ein Eigenwert des Büschels (A, B) sein kann. Etwa für

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = B \quad (2.5)$$

gilt $Ae^{(2)} = 0 = Be^{(2)}$, wobei $e^{(2)}$ den zweiten Einheitsvektor bezeichnet. Damit erfüllt jedes $\lambda \in \mathbb{K}$ die Gleichung $Ae^{(2)} = \lambda Be^{(2)}$, und ist also ein Eigenwert. Allgemeiner gilt, dass (λ, x) für jedes λ ein Eigenpaar von (A, B) ist, falls $x \neq 0$ ein Nullvektor beider Matrizen A und B ist.

Ein weiteres Beispiel für unendlich viele Eigenwerte liefert das Büschel (A, B) mit

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad (2.6)$$

da für jedes $\lambda \in \mathbb{K}$ der Vektor $x = (\lambda \ 1 \ 0)^T$ ein Eigenvektor ist.

Für die beiden vorgestellten Beispiele gilt jeweils $\sigma(A, B) = \mathbb{C}$. Da wir in diesem Fall bereits alle Eigenwerte kennen, interessieren wir uns nur für Büschel mit einem endlichen Spektrum. Solche Paare können wir mit Hilfe der Determinante charakterisieren.

Lemma 2.9. *Sei (A, B) ein Büschel der Dimension n . Dann ist das Spektrum $\sigma(A, B)$ genau dann endlich, wenn die Determinante von (A, B) nicht identisch null ist.*

Beweis. Angenommen, es gelte $\det(A - \lambda B) \equiv 0$. Dann ist nach Definition 2.8 jedes $\mu \in \mathbb{K}$ ein Eigenwert des Büschels und somit $\sigma(A, B) = \mathbb{K}$.

Es sei nun $\det(A - \lambda B) \not\equiv 0$. Dann ist $p(\lambda) = \det(A - \lambda B)$ ein Polynom vom Grad $0 \leq m \leq n$ und es existieren somit maximal m Nullstellen, die nach Definition gerade die Eigenwerte des Büschels (A, B) sind. \square

Definition 2.10 (Charakteristisches Polynom). *Sei (A, B) ein Büschel der Dimension n . Dann nennen wir das Polynom*

$$p(\lambda) = \det(A - \lambda B) \quad (2.7)$$

das charakteristische Polynom von (A, B) und bezeichnen es mit $\chi_{(A, B)}$.

Es ist $\lambda \in \mathbb{K}$ also genau dann ein Eigenwert von (A, B) , wenn $\chi_{(A, B)}(\lambda) = 0$ gilt.

Betrachten wir nun wieder die beiden Beispiele 2.5 und 2.6, so sehen wir, dass die Determinante der Büschel identisch Null ist. Dies führt uns nun zu folgender Definition.

Definition 2.11 (Regularität). *Ein Matrixbüschel (A, B) heißt regulär, falls $\det(A - \lambda B)$ nicht identisch null ist. Man sagt auch, das Paar (A, B) ist regulär.*

Wir betrachten nun ein reguläres Büschel (A, B) mit einer regulären Matrix B . In diesem Fall gilt

$$\sigma(A, B) = \sigma(B^{-1}A, I) = \sigma(B^{-1}A) \quad (2.8)$$

wegen

$$Ax = \lambda Bx \Leftrightarrow B^{-1}Ax = \lambda Ix \Leftrightarrow B^{-1}Ax = \lambda x. \quad (2.9)$$

Ist $\lambda \neq 0$ ein Eigenwert von (A, B) , so ist $\mu = \lambda^{-1}$ ein Eigenwert von (B, A) , denn es gilt

$$Ax = \lambda Bx \Leftrightarrow \lambda^{-1}Ax = Bx. \quad (2.10)$$

Lassen wir $\lambda = 0$ zu und setzen $\infty = \frac{1}{0}$, so kann ∞ als Eigenwert eines Büschels auftreten. Dies geschieht, falls B singulär ist. Dann existiert nämlich ein $x \neq 0$ mit $Bx = 0$, sodass Null ein Eigenwert von (B, A) ist und wir somit ∞ als Eigenwert von (A, B) auffassen können. Es ist in dem Fall einer singulären Matrix B außerdem möglich, dass $\chi_{(A,B)} \equiv c$ für eine Konstante $c \in \mathbb{K} \setminus \{0\}$ gilt, also (A, B) keine gewöhnlichen Eigenwerte hat. Wir können dann ebenfalls ∞ als Eigenwert von (A, B) betrachten.

Wie wir gesehen haben, ist das Spektrum $\sigma(A, B)$ bei einem verallgemeinerten Eigenwertproblem nur von $\text{Rang}(B)$ abhängig, nicht aber von $\text{Rang}(A)$. Hat B keinen vollen Rang, so kann das Spektrum $\sigma(A, B)$ endlich, leer oder unendlich sein, wie das folgende Beispiel, entnommen aus [2], zeigt. Im Falle eines leeren Spektrums können wir ∞ als Eigenwert ansehen.

Beispiel 2.12. (1) Seien $A = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$ und $B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

Es ist $\chi_{(A,B)}(\lambda) = 3(1 - \lambda)$ und somit $\sigma(A, B) = \{1\}$.

(2) Seien $A = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$ und $B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$.

Es ist $\chi_{(A,B)} \equiv 3$ und somit $\sigma(A, B) = \{\infty\}$.

(3) Seien $A = \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}$ und $B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

Es ist $\chi_{(A,B)} \equiv 0$ und somit $\sigma(A, B) = \mathbb{C}$.

2.2.1 Die verallgemeinerte Schur-Zerlegung

Wir kommen nun zu einem für diese Arbeit sehr relevanten Resultat, der *verallgemeinerten Schur-Zerlegung*. Sie ist die Grundlage für die algorithmischen Überlegungen in Kapitel 4.

Definition 2.13 (Äquivalenz). *Sei (A, B) ein Büschel und seien $U, V \in \mathbb{K}^{n \times n}$ regulär. Dann heißt das Büschel (U^*AV, U^*BV) äquivalent zu (A, B) . Man sagt auch, (U^*AV, U^*BV) entsteht aus (A, B) durch Äquivalenztransformation.*

Die Auswirkungen von Äquivalenztransformationen auf Eigenwerte und Eigenvektoren eines Matrixbüschels beschreibt das folgende Lemma.

Lemma 2.14. *Seien (λ, x) und (λ, y) linkes bzw. rechtes Eigenpaar eines regulären Büschels (A, B) . Falls U und V regulär sind, so sind $(\lambda, V^{-1}x)$ und $(\lambda, U^{-1}y)$ linkes bzw. rechtes Eigenpaar von (U^*AV, U^*BV) .*

Beweis. Durch Einsetzen erhalten wir

$$\begin{aligned} (U^*AV - \lambda U^*BV)(V^{-1}x) &= (U^*A - \lambda U^*B)x = U^*(A - \lambda B)x = 0, \\ (U^{-1}y)^*(U^*AV - \lambda U^*BV) &= (y^*(U^{-1})^*)(U^*AV - \lambda U^*BV) = y^*(AV - \lambda BV) \\ &= y^*(A - \lambda B)V = 0. \end{aligned}$$

□

Äquivalenztransformationen lassen außerdem das charakteristische Polynom $\chi_{(A,B)}$ im Wesentlichen unverändert. Es gilt nämlich

$$\det(U^*AV - \lambda U^*BV) = \det(U^*) \det(V) \det(A - \lambda B). \quad (2.11)$$

Da U und V regulär sind, sorgt die Äquivalenztransformation also dafür, dass $\chi_{(A,B)}$ mit einer von Null verschiedenen Konstante multipliziert wird. Insbesondere bleiben dadurch die Nullstellen des Polynoms und deren Vielfachheiten erhalten.

Genauso, wie wir Matrizen durch Ähnlichkeitstransformationen auf aus numerischer Sicht einfachere Formen reduzieren können, können wir ein Büschel durch Äquivalenztransformationen vereinfachen. So können wir entsprechend der gewöhnlichen Schur-Zerlegung 2.6 eine *verallgemeinerte Schur-Zerlegung*, die auf Stewart [4] zurückgeht, einführen.

Satz 2.15 (Verallgemeinerte Schur-Zerlegung). *Seien $A, B \in \mathbb{C}^{n \times n}$ und (A, B) regulär. Dann existieren unitäre Matrizen $Q, Z \in \mathbb{C}^{n \times n}$ und obere Dreiecksmatrizen $S, T \in \mathbb{C}^{n \times n}$, sodass*

$$Q^*AZ = T \quad \text{und} \quad Q^*BZ = S \quad (2.12)$$

gilt. Ein Paar (T, S) , mit Matrizen T und S wie oben, heißt verallgemeinerte Schur-Form.

Beweis. Wir führen Induktion nach der Dimension n der Matrizen A, B . Der Induktionsanfang $n = 1$ ist klar, da Matrizen der Dimension $n = 1$ obere Dreiecksmatrizen sind und $U = 1 = V$ das Gewünschte liefern.

Angenommen, die Aussage gilt für alle Matrizen der Dimension $n - 1$ oder kleiner. Sei (λ, v_1) ein Eigenpaar von (A, B) mit $\|v_1\|_2 = 1$. Da (A, B) regulär ist, ist Av_1 oder Bv_1 ungleich null. Nach Lemma 2.4 existiert eine orthogonale Matrix $V_1 \in \mathbb{C}^{n \times (n-1)}$ so, dass $(v_1 \ V_1)$ unitär ist. Weiter sei $U_1 \in \mathbb{C}^{n \times (n-1)}$ orthogonal und so, dass die Spalten im orthogonalen Komplement von $\text{span}(Av_1 + Bv_1)$ liegen. Nach Lemma 2.4 finden wir $u_1 \in \mathbb{C}^n$ so, dass $(u_1 \ U_1)$ unitär ist. Mit

$$\alpha_{11} := u_1^*Av_1, \quad a_{12}^* := u_1^*AV_1 \quad \text{und} \quad A_{22} := U_1^*AV_1$$

gilt dann

$$(u_1 \ U_1)^*A(v_1 \ V_1) = \begin{pmatrix} u_1^*Av_1 & u_1^*AV_1 \\ U_1^*Av_1 & U_1^*AV_1 \end{pmatrix} = \begin{pmatrix} \alpha_{11} & a_{12}^* \\ 0 & A_{22} \end{pmatrix}.$$

Zudem erhalten wir mit $\beta_{11} := u_1^*Bv_1, b_{12}^* := u_1^*BV_1$ und $B_{22} := U_1^*BV_1$

$$(u_1 \ U_1)^*B(v_1 \ V_1) = \begin{pmatrix} u_1^*Bv_1 & u_1^*BV_1 \\ U_1^*Bv_1 & U_1^*BV_1 \end{pmatrix} = \begin{pmatrix} \beta_{11} & b_{12}^* \\ 0 & B_{22} \end{pmatrix}.$$

Indem wir die Induktionsvoraussetzung auf A_{22} und B_{22} anwenden, erhalten wir unitäre Matrizen U_2 und V_2 der Dimension $n - 1$, sodass $\hat{A}_{22} = U_2^*A_{22}V_2$ und $\hat{B}_{22} = U_2^*B_{22}V_2$ obere Dreiecksmatrizen sind. Die unitären Matrizen

$$U = (u_1 \ U_1) \begin{pmatrix} 1 & 0 \\ 0 & U_2 \end{pmatrix}, \quad V = (v_1 \ V_1) \begin{pmatrix} 1 & 0 \\ 0 & V_2 \end{pmatrix}$$

liefern nun das Gewünschte. □

Aus diesem Satz können wir nun folgende Konsequenzen ziehen, die die Bestimmung der Eigenwerte wesentlich vereinfachen.

Lemma 2.16. *Sei (A, B) ein reguläres Büschel der Dimension n und sei (\hat{A}, \hat{B}) eine verallgemeinerte Schur-Form von (A, B) . Existiert ein $k \in \{1, \dots, n\}$ mit $\hat{b}_{kk} = 0$ und $\hat{a}_{kk} \neq 0$, so ist der k -te Eigenwert unendlich. Es gilt*

$$\sigma(A, B) \setminus \{\infty\} = \left\{ \frac{\hat{a}_{kk}}{\hat{b}_{kk}} : \hat{b}_{kk} \neq 0, k \in \{1, \dots, n\} \right\}. \quad (2.13)$$

Beweis. Seien $U, V \in \mathbb{K}^{n \times n}$ die unitären Transformationen, die die verallgemeinerte Schur-Form (\hat{A}, \hat{B}) von (A, B) erzeugen. Die Aussagen über das Spektrum von (A, B) folgen aus der Gleichung

$$\begin{aligned} \det(A - \lambda B) &= \det(U\hat{A}V^* - \lambda U\hat{B}V^*) = \det(U(\hat{A} - \lambda\hat{B})V^*) \\ &= \det(U) \det(\hat{A} - \lambda\hat{B}) \det(V^*) = \det(U) \det(V^*) \prod_{i=1}^n (\hat{a}_{ii} - \lambda\hat{b}_{ii}) \\ &= \omega \prod_{\hat{b}_{ii} \neq 0} (\hat{a}_{ii} - \lambda\hat{b}_{ii}) \prod_{\hat{b}_{ii} = 0} \hat{a}_{ii}, \end{aligned}$$

wobei ω das Produkt der Determinanten der unitären Transformationen U und V ist, also $|\omega| = 1$ gilt. Aus (2.11) folgt, dass das Spektrum nicht von der Wahl von U und V abhängt. \square

Wir wollen nun eine einheitliche Schreibweise für endliche und unendliche Eigenwerte eines regulären Büschels (A, B) in verallgemeinerter Schur-Form einführen. Wie Satz 2.16 besagt, sind die Eigenwerte von (A, B) durch Paare (α_i, β_i) für $i \in \{1, \dots, n\}$ bestimmt. Ist $\beta_i \neq 0$ für $i \leq n$, so ist $\lambda_i = \alpha_i/\beta_i$ ein endlicher Eigenwert von (A, B) , andernfalls liegt ein unendlicher Eigenwert vor. Wir fassen deshalb die Eigenwerte als Paare (α_i, β_i) auf. Da diese Schreibweise bis auf Skalarmultiplikation eindeutig ist, definieren wir die *projektive Darstellung* $\langle \alpha_i, \beta_i \rangle$ eines Eigenwerts als die Menge

$$\langle \alpha_i, \beta_i \rangle = \{\tau(\alpha_i, \beta_i) : \tau \in \mathbb{C}\}. \quad (2.14)$$

So ist $\langle 0, 1 \rangle$ der Eigenwert Null und $\langle 1, 0 \rangle$ ein unendlicher Eigenwert. Einen gewöhnlichen Eigenwert λ können wir nun in der Form $\langle \lambda, 1 \rangle$ darstellen.

3 Gewöhnliche Eigenwertverfahren

Bevor wir nun zu dem Ziel dieser Arbeit, dem QZ-Verfahren, kommen, verschaffen wir uns einen kurzen Überblick über gewöhnliche Eigenwertverfahren, deren Kenntnis für das Verständnis der folgenden Kapitel erforderlich ist. Dabei wollen wir hauptsächlich die wichtigen Resultate nennen, ohne auf die Beweise einzugehen. Eine detaillierte Einführung in die Theorie der Eigenwertverfahren mit Beweisen zu den gegebenen Aussagen und algorithmischen Umsetzungen kann unter anderem den Büchern von Golub und van Loan [2] und Stewart [5] entnommen werden. Wir werden uns in diesem Kapitel überwiegend auf [1] beziehen.

3.1 Vektoriteration

Das Ziel der *Vektoriteration* ist es, Eigenwerte mit zugehörigen Eigenvektoren einer Matrix $A \in \mathbb{K}^{n \times n}$ zu berechnen.

Wir befassen uns zunächst mit Diagonalmatrizen $D = \text{diag}(\lambda_1, \dots, \lambda_n) \in \mathbb{K}^{n \times n}$ mit Eigenwerten $\lambda_1, \dots, \lambda_n \in \mathbb{K}$. Dabei seien die Eigenwerte betraglich absteigend angeordnet und λ_1 dominant und einfach. Dies bedeutet, es gilt

$$|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_n|. \quad (3.1)$$

Zu einem Startvektor $z^{(0)} \in \mathbb{K}^n$ betrachten wir nun die Folge $(z^{(m)})_{m \in \mathbb{N}}$ mit

$$z^{(m)} := D^m z^{(0)} \quad \text{für alle } m \in \mathbb{N}, \quad (3.2)$$

und somit

$$z_i^{(m)} = \lambda_i^m z_i^{(0)} \quad \text{für alle } m \in \mathbb{N}, i \in \{1, \dots, n\}. \quad (3.3)$$

Da λ_1 als dominant und einfach vorausgesetzt war, wird die Komponente $z_1^{(m)}$ des Vektors am schnellsten wachsen und sich somit durchsetzen.

Lemma 3.1. *Sei $z^{(0)} \in \mathbb{K}^n$ ein Startvektor und $(z^{(m)})_{m \in \mathbb{N}}$ die in (3.2) definierte Folge der Iterierten. Dann gilt für alle $m \in \mathbb{N}_0$*

$$\left\| \lambda_1^{-m} z^{(m)} - z_1^{(0)} e^{(1)} \right\|_2 \leq \left(\frac{|\lambda_2|}{|\lambda_1|} \right)^m \left\| z^{(0)} - z_1^{(0)} e^{(1)} \right\|_2. \quad (3.4)$$

Die Folge $(z^{(m)})_{m \in \mathbb{N}}$ der Iterierten konvergiert in Abhängigkeit des Quotienten $|\lambda_2| / |\lambda_1|$, also dem Verhältnis des zweitgrößten Eigenwerts zum größten, linear gegen ein Vielfaches

3 Gewöhnliche Eigenwertverfahren

des ersten kanonischen Einheitsvektors $e^{(1)}$, der ein Eigenvektor zum Eigenwert λ_1 ist. Wir betrachten nun den allgemeineren Fall einer diagonalisierbaren Matrix $A \in \mathbb{K}^{n \times n}$ mit einem einfachen dominanten Eigenwert $\lambda_1 \in \mathbb{K}$. Es existiert also eine reguläre Matrix $T \in \mathbb{K}^{n \times n}$ mit

$$TAT^{-1} = D = \text{diag}(\lambda_1, \dots, \lambda_n), \quad (3.5)$$

wobei die Eigenwerte $\lambda_1, \dots, \lambda_n$ wie in (3.1) angeordnet sind.

Die Folge der Iterierten definieren wir für einen Startvektor $z^{(0)} \in \mathbb{K}^n$ wie zuvor in 3.2

$$z^{(m)} := A^m z^{(0)} = Az^{(m-1)} \quad \text{für alle } m \in \mathbb{N}. \quad (3.6)$$

Sei $x^{(1)}$ ein Eigenvektor zum Eigenwert λ_1 . Wie in Lemma 3.1 kann gezeigt werden, dass die Folge der Iterierten abhängig vom Verhältnis $|\lambda_2| / |\lambda_1|$ gegen ein Vielfaches von $x^{(1)}$ konvergiert.

Eine Approximation des Eigenwertes λ_1 erhalten wir durch den *Rayleigh-Quotienten*

$$\Lambda_A(x) := \frac{\langle x, Ax \rangle_2}{\langle x, x \rangle_2} \quad \text{für } x \in \mathbb{K}^n \setminus \{0\}. \quad (3.7)$$

Ist $x \in \mathbb{K}^n \setminus \{0\}$ ein Eigenvektor der Matrix A zum Eigenwert λ_1 , so gilt $\Lambda_A(x) = \lambda_1$. Für eine Näherung $y \in \mathbb{K}^n \setminus \{0\}$ des Eigenvektors x liefert der Rayleigh-Quotient $\Lambda_A(y)$ eine gute Approximation des Eigenwertes λ_1 .

Die Vektoriteration ist ein Verfahren um den größten Eigenwert und einen zugehörigen Eigenvektor einer Matrix zu bestimmen. Da in der Praxis das Interesse allerdings meistens bei dem kleinsten Eigenwert liegt, führt uns dies zu der *inversen Vektoriteration*.

Dazu sei nun eine reguläre und diagonalisierbare Matrix $A \in \mathbb{K}^{n \times n}$ gegeben. Ist $\lambda \in \mathbb{K}$ ein Eigenwert mit Eigenvektor $x \in \mathbb{K}^n \setminus \{0\}$, so gilt

$$Ax = \lambda x \quad \Leftrightarrow \quad \lambda^{-1}x = A^{-1}x. \quad (3.8)$$

Dies bedeutet, das Inverse des betragskleinsten Eigenwerts der Matrix A ist der betragsgrößte Eigenwert von A^{-1} . Wir führen deshalb die Vektoriteration mit der Inversen A^{-1} anstelle von A durch und definieren dazu die Folge $(z^{(m)})_{m \in \mathbb{N}}$ der Iterierten für einen Startvektor $z^{(0)} \in \mathbb{K}^n \setminus \{0\}$ analog zu 3.6 mit

$$z^{(m)} := A^{-m} z^{(0)} = A^{-1} z^{(m-1)} \quad \text{für alle } m \in \mathbb{N}. \quad (3.9)$$

Da A diagonalisierbar ist, existiert also eine reguläre Matrix $T \in \mathbb{K}^{n \times n}$, die (3.5) erfüllt, wobei wir jetzt fordern, dass $1/|\lambda_1|$ dominanter Eigenwert von A^{-1} ist. Die Eigenwerte $\lambda_1, \dots, \lambda_n$ von A sind also betraglich aufsteigend geordnet:

$$|\lambda_1| < |\lambda_2| \leq \dots \leq |\lambda_n|. \quad (3.10)$$

Wir erhalten nun die Konvergenz der Folge der Iterierten gegen einen Eigenvektor $x^{(1)}$ des Eigenwertes λ_1 in Abhängigkeit von $|\lambda_1|/|\lambda_2|$, also dem Verhältnis des kleinsten Eigenwerts zum zweitkleinsten.

Führen wir nun einen Shift $\mu \in \mathbb{K}$ ein und gehen zu der Matrix $(A - \mu I)^{-1}$ über, so können wir Konvergenz gegen einen beliebigen Eigenwert der Matrix erhalten. Dieses Verfahren nennt sich *inverse Vektoriteration mit Shift μ* . Falls $\mu \notin \sigma(A)$ gilt, ist $A - \mu I$ invertierbar. Wir benötigen daher keine Regularität der Matrix A , es reicht wie zuvor die Diagonalisierbarkeit. Für $\lambda \in \mathbb{K}$ und einen zugehörigen Eigenvektor $x \neq 0$ folgt die Äquivalenz

$$Ax = \lambda x \quad \Leftrightarrow \quad (A - \mu I)^{-1}x = (\lambda - \mu)^{-1}x. \quad (3.11)$$

Wählen wir $\mu \in \mathbb{K}$ so, dass

$$|\lambda_1 - \mu| < |\lambda_2 - \mu| \leq \dots \leq |\lambda_n - \mu| \quad (3.12)$$

gilt, und definieren die Iterierte durch

$$z^{(m)} := (A - \mu I)^{-m} z^{(0)} = (A - \mu I)^{-1} z^{(m-1)} \quad \text{für alle } m \in \mathbb{N}, \quad (3.13)$$

so erhalten wir entsprechend Konvergenz gegen den Eigenwert $\lambda_1 - \mu$ von $A - \mu I$. Indem μ als gute Näherung von λ_1 gewählt wird, kann die Konvergenzrate $|\lambda_1 - \mu|/|\lambda_2 - \mu|$ deutlich verbessert werden.

Liegt kein einfacher Eigenwert vor, sondern mehrfache Eigenwerte, oder Eigenwerte gleichen Betrags, es gilt also

$$|\lambda_1| = \dots = |\lambda_k| > \dots \geq |\lambda_n| \quad \text{für ein } k \leq n, \quad (3.14)$$

so werden die Iterierten mit den bisherigen Verfahren gegen einen Vektor aus dem von den Eigenvektoren zu λ_1 bis λ_k aufgespannten invarianten Teilraum konvergieren. Um eine Basis des Teilraums zu konstruieren, starten wir die Vektoriteration mit k Vektoren. Diese werden gegen k Vektoren aus dem Teilraum konvergieren und, sofern sie linear unabhängig sind, die gewünschte Basis bilden. Wir beginnen die Iteration deswegen mit einer Startmatrix $Z^{(0)} \in \mathbb{K}^{n \times k}$ und definieren die Iterierten durch

$$Z^{(m)} := A^m Z^{(0)} \quad \text{für alle } m \in \mathbb{N}. \quad (3.15)$$

Die Konvergenzrate ist jetzt entsprechend den vorherigen Ergebnissen von dem Verhältnis $|\lambda_{k+1}|/|\lambda_k|$ abhängig, das heißt, das Verfahren kann nur konvergieren, wenn mindestens ein Eigenwert betragsmäßig echt kleiner ist, als die anderen. Um für numerische Stabilität zu sorgen, führen wir eine Orthogonalisierung der Iterierten durch, zum Beispiel

3 Gewöhnliche Eigenwertverfahren

durch die QR-Zerlegung. Dieses Verfahren wird als *orthogonale Iteration* bezeichnet. Wir bestimmen eine QR-Zerlegung $Z^{(0)} = Q^{(0)}R^{(0)}$ der Startmatrix, berechnen die Iterierten

$$Z^{(m+1)} = AQ^{(m)} \quad (3.16)$$

und bestimmen mit einer QR-Zerlegung von $Z^{(m+1)}$ anschließend $Q^{(m+1)}$ durch

$$Q^{(m+1)}R^{(m+1)} = Z^{(m+1)} \quad (3.17)$$

für alle $m \in \mathbb{N}$.

3.2 QR-Iteration

Sind wir an allen Eigenwerten und ihren zugehörigen Eigenvektoren interessiert, so ist die Idee, die orthogonale Iteration zu erweitern, indem wir $k = n$ und als Startmatrix zum Beispiel $Q^{(0)} = I$ wählen. Dieses Verfahren nennt man die *QR-Iteration* und ihr Ziel ist es, die Schur-Zerlegung 2.6 zu approximieren. Dazu definieren wir die Folge der Iterierten

$$A^{(m)} := \left(Q^{(m)}\right)^* AQ^{(m)} \quad \text{für alle } m \in \mathbb{N}, \quad (3.18)$$

Da die Matrizen $A^{(m)}$ und $Q^{(m)}$ jetzt vollen Rang haben, ist diese Vorschrift mit der Berechnung von $Q^{(m)}$ wie in (3.17) zu aufwändig. Indem wir eine QR-Zerlegung der Matrix $AQ^{(m)}$ durchführen, erhalten wir

$$Q^{(m+1)}R^{(m+1)} = AQ^{(m)} \quad \Leftrightarrow \quad Q^{(m+1)}R^{(m+1)} \left(Q^{(m)}\right)^* = A \quad (3.19)$$

und daraus folgt

$$A^{(m)} = \left(Q^{(m)}\right)^* AQ^{(m)} = \left(Q^{(m)}\right)^* Q^{(m+1)}R^{(m+1)}. \quad (3.20)$$

Setzen wir jetzt $\hat{Q}^{(m+1)} = \left(Q^{(m)}\right)^* Q^{(m+1)}$, so erhalten wir

$$\begin{aligned} A^{(m)} &= \hat{Q}^{(m+1)}R^{(m+1)} \\ A^{(m+1)} &= R^{(m+1)}\hat{Q}^{(m+1)}, \end{aligned} \quad (3.21)$$

und somit

$$A^{(m+1)} = \left(\hat{Q}^{(m+1)}\right)^* A^{(m)} \hat{Q}^{(m+1)} \quad \text{für alle } m \in \mathbb{N}_0. \quad (3.22)$$

Wir können also eine QR-Zerlegung der aktuellen Iterierten $A^{(m)}$ durchführen und aus ihr die neue Iterierte $A^{(m+1)}$ konstruieren, ohne die Ausgangsmatrix A und die Matrizen $Q^{(m)}$ zu kennen.

Die Konvergenz des Verfahrens entspricht der Konvergenz der orthogonalen Iteration, wir müssen nur fordern, dass es einen Eigenwert gibt, der betragsmäßig echt kleiner ist, als die anderen.

Um den Rechenaufwand der QR-Zerlegung, der im Allgemeinen bei $\mathcal{O}(n^3)$ liegt, zu verringern, bietet es sich an, die Matrix A vorher auf die sogenannte Hessenberg-Gestalt zu transformieren.

Definition 3.2. Eine Matrix $A \in \mathbb{K}^{n \times n}$ heißt (obere) Hessenberg-Matrix, falls $a_{ij} = 0$ für alle $i, j \in \{1, \dots, n\}$ mit $i > j + 1$ gilt.

Wird die QR-Zerlegung anschließend mit Givens-Rotationen durchgeführt, so bleibt die Hessenberg-Form erhalten und der Rechenaufwand beträgt im Allgemeinen nur noch $\mathcal{O}(n^2)$, im symmetrischen Fall sogar $\mathcal{O}(n)$.

Wir wählen $Q^{(0)}$ also so, dass $A^{(0)} = (Q^{(0)})^* A Q^{(0)}$ Hessenberg-Gestalt hat.

Um den Speicheraufwand zu verringern, wollen wir $A^{(m+1)}$ direkt aus $A^{(m)}$ berechnen, das heißt, wir führen den QR-Schritt nicht mehr explizit wie in (3.21) aus, sondern wir wollen versuchen, die Berechnung implizit durchzuführen. Da unsere iterierten Hessenberg-Gestalt haben, führen wir die QR-Zerlegung mit Givens-Rotationen durch, das heißt, $\hat{Q}^{(m)}$ hat die Form

$$\hat{Q}^{(m)} = G_1^* G_2^* \cdots G_{n-1}^*, \quad (3.23)$$

wobei G_i^* eine Givens-Rotation ist, die nur auf die i -te und $(i+1)$ -te Spalte einer Matrix wirkt. Wir können die Berechnung wie folgt durchführen

(1) Bestimme G_1 so, dass $(G_1 A^{(m)})_{21} = 0$ gilt.

(i) Berechne $G_1 A^{(m)}$.

(2) Bestimme G_2 so, dass $(G_2 G_1 A^{(m)})_{32} = 0$ gilt.

(i) Berechne $G_2 G_1 A^{(m)}$.

(ii) Berechne $G_2 G_1 A^{(m)} G_1^*$.

(3) Fahre wie in (2) fort, bis A Dreiecks-Gestalt hat.

Wir verweisen auf Satz 4.3 um zu zeigen, dass dies zum gleichen Ergebnis führt, wie die explizite Berechnung, da die erste Spalte von $\hat{Q}^{(m)}$ offensichtlich nur von G_1 abhängt.

Um die Konvergenzgeschwindigkeit zu verbessern, können wir wie bei der Vektoriteration bestimmte Shift-Strategien durchführen. Wählen wir einen Shift μ so, dass

$$|\lambda_1 - \mu| \geq \dots \geq |\lambda_k - \mu| > |\lambda_{k+1} - \mu| \geq \dots \geq |\lambda_n - \mu| \quad (3.24)$$

3 Gewöhnliche Eigenwertverfahren

und $\lambda_{k+1} = \dots = \lambda_n$ gilt, so ist die Konvergenzrate abhängig von $\frac{|\lambda_n - \mu|}{|\lambda_k - \mu|}$. Die Iterierten haben die Form

$$A^{(m)} - \mu I = \hat{Q}^{(m+1)} R^{(m+1)}, \quad A^{(m+1)} = R^{(m+1)} \hat{Q}^{(m+1)} + \mu I \quad (3.25)$$

und wir erhalten

$$A^{(m+1)} = \left(\hat{Q}^{(m+1)} \right)^* A^{(m)} \hat{Q}^{(m+1)}, \quad (3.26)$$

somit entsteht $A^{(m+1)}$ aus $A^{(m)}$ wieder nur durch Äquivalenztransformationen.

Wir können jetzt zwischen verschiedenen Shift-Strategien wählen. Eine einfache Variante ist die Wahl von $\mu = a_{nn}^{(m)}$ als Shift-Parameter, auch *Rayleigh-Quotient-Shift* genannt. Ist nämlich $q^{(m)} := Q^{(m)} e^{(n)}$ die letzte Spalte von $Q^{(m)}$, so folgt $\Lambda_A(q^{(m)}) = \mu$. Es lässt sich zeigen, dass hier unter gewissen Voraussetzungen sogar quadratische Konvergenz vorliegt. Die Hoffnung ist, dass durch diesen Shift alle Außerdiagonaleinträge der letzten Zeile schnell gegen Null konvergieren. Lemma 2.1 besagt dann, dass wir einen Eigenwert $a_{nn}^{(m)}$ erhalten, sodass wir Deflation durchführen können. Dies bedeutet, ein bereits konvergierter Diagonalblock wird abgespalten und die Iteration anschließend mit einer Matrix kleinerer Dimension fortgeführt. Mit diesem Shift erhält man in $\mathcal{O}(n)$ Iterationen eine obere Dreiecksmatrix.

Eine andere Wahl ist der sogenannte *Wilkinson-Shift*, bei dem man die Eigenwerte der untersten 2×2 -Teilmatrix von A verwendet. Da wir darauf hoffen, dass $a_{nn}^{(m)}$ gegen einen Eigenwert konvergiert, bietet es sich an denjenigen Eigenwert der Teilmatrix als Shift-Parameter zu wählen, der $a_{nn}^{(m)}$ am nächsten liegt. Demnach dürfen wir auf bessere Konvergenz hoffen, als beim Rayleigh-Quotient-Shift. Bei dieser Variante können allerdings komplexe Shift-Parameter auftreten. Um dies zu umgehen, bietet sich ein *Doppel-Shift* an, bei dem ein Paar von einfachen Shifts zu einem Iterationsschritt zusammengefasst wird. Kombinieren wir einen komplexen Shift μ mit dem komplex konjugierten $\bar{\mu}$, so erhalten wir mit

$$(A^{(m)} - \mu I)(A^{(m)} - \bar{\mu} I) = (A^{(m)})^2 - (\mu + \bar{\mu})A^{(m)} + \mu\bar{\mu}I \quad (3.27)$$

wieder eine reelle Matrix. Der Doppel-Shift führt dazu, dass die Matrix $A^{(m)}$ gegen *quasi-obere Dreiecksform* konvergiert, das heißt, auf der Diagonalen von A treten sowohl 1×1 - als auch 2×2 -Blöcke auf.

4 Das QZ-Verfahren

Wir wenden uns nun den numerischen Aspekten des verallgemeinerten Eigenwertproblems zu. Das Ziel dieses Kapitels ist die Präsentation des *QZ-Verfahrens*, ein Zugang zur numerischen Lösung des verallgemeinerten Eigenwertproblems $Ax = \lambda Bx$.

Ist die Matrix B invertierbar, so können wir zu den gewöhnlichen Eigenwertproblemen

$$B^{-1}Ax = \lambda x$$

oder

$$AB^{-1}y = \lambda y, \quad y = Bx$$

übergehen, um dann zum Beispiel das QR-Verfahren (3.2) anzuwenden mit dem wir die Eigenwerte von AB^{-1} oder $B^{-1}A$ bestimmen können. Das Problem hierbei ist, dass die Berechnung der Inversen im Allgemeinen sehr schlecht konditioniert ist, dieses Verfahren also von der numerischen Sichtweise her nicht geeignet ist. Des Weiteren ist dieser Ansatz bei einer singulären Matrix B nicht möglich. Wir suchen daher einen alternativen Zugang zum Lösen des verallgemeinerten Eigenwertproblems. Die Idee ist, wie schon bei der QR-Iteration, die Approximation der verallgemeinerten Schur-Zerlegung der Matrizen A und B , um daraus die Eigenwerte zu erhalten.

4.1 Hessenberg-Dreiecks-Form

Um einen möglichst geringen Rechenaufwand zu erhalten, werden wir, ähnlich dem QR-Verfahren, das Büschel (A, B) zunächst auf eine geeignetere Form transformieren. Hier bietet sich eine Transformation auf die sogenannte Hessenberg-Dreiecks-Form an. Wir beziehen uns dabei auf die Vorgehensweise von Golub und van Loan in [2].

Definition 4.1 (Hessenberg-Dreiecks-Form). *Sei (A, B) ein Büschel. Dann hat (A, B) Hessenberg-Dreiecks-Form, wenn A eine obere Hessenberg-Matrix und B eine obere Dreiecksmatrix ist.*

Der erste Schritt in der Berechnung der verallgemeinerten Schur-Zerlegung des Büschels (A, B) ist nun die Reduktion der Matrizen A und B auf obere Hessenberg- bzw. obere Dreiecksform durch unitäre Transformationen.

Dazu bestimmen wir zunächst eine unitäre Matrix $U \in \mathbb{R}^{n \times n}$, sodass U^*B eine obere Dreiecksmatrix ist und wenden U in gleicher Weise auf A an, damit die Eigenwerte des Büschels erhalten bleiben.

4 Das QZ-Verfahren

Wir betrachten die Transformationen skizzenhaft am Beispiel von Matrizen der Dimension $n = 4$:

$$A \leftarrow U^* A = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix}, \quad B \leftarrow U^* B = \begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{pmatrix}.$$

Im nächsten Schritt transformieren wir A auf Hessenberg-Gestalt, zum Beispiel durch geeignete Givens-Rotationen. Diese erzeugen jedoch in der Matrix B jeweils einen Nichtnulleintrag in der unteren Dreieckshälfte, der anschließend durch Multiplikation mit einer weiteren Givens-Rotation eliminiert werden muss, damit die Dreiecksform von B erhalten bleibt.

Wir bezeichnen im Folgenden mit Q_{ij} und Z_{ij} Givens-Rotationen, die auf die Zeilen bzw. Spalten i und j angewendet werden. In unserem Beispiel bestimmen wir als erstes eine Givens-Rotation Q_{34} um den Eintrag a_{41} zu eliminieren:

$$A \leftarrow Q_{34}^* A = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \end{pmatrix}, \quad B \leftarrow Q_{34}^* B = \begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{pmatrix}.$$

Der entstandene Nichtnulleintrag b_{43} in B kann durch eine Givens-Rotation Z_{34} wieder annulliert werden:

$$A \leftarrow AZ_{34} = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \end{pmatrix}, \quad B \leftarrow BZ_{34} = \begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{pmatrix}.$$

Auf dem gleichen Weg wird der Eintrag a_{31} in A eliminiert:

$$A \leftarrow Q_{23}^* A = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{pmatrix}, \quad B \leftarrow Q_{23}^* B = \begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & 0 & 0 & * \end{pmatrix},$$

$$A \leftarrow AZ_{23} = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{pmatrix}, \quad B \leftarrow BZ_{23} = \begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{pmatrix}.$$

Die Matrix A hat nun in der ersten Spalte Hessenberg-Gestalt. Nach der Elimination des Eintrages a_{42} sind wir fertig:

$$A \leftarrow \tilde{Q}_{34}^* A = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \end{pmatrix}, \quad B \leftarrow \tilde{Q}_{34}^* B = \begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{pmatrix},$$

$$A \leftarrow A \tilde{Z}_{34} = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \end{pmatrix}, \quad B \leftarrow B \tilde{Z}_{34} = \begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{pmatrix}.$$

Insgesamt erhalten wir also transformierte Matrizen $\hat{A} = \tilde{Q}_{34}^* Q_{23}^* Q_{34}^* U^* A Z_{34} Z_{23} \tilde{Z}_{34}$ in Hessenberg- sowie $\hat{B} = \tilde{Q}_{34}^* Q_{23}^* Q_{34}^* U^* B Z_{34} Z_{23} \tilde{Z}_{34}$ in Dreiecksgestalt.

Wie wir gesehen haben, benötigen wir für jeden zu eliminierenden Eintrag a_{ij} zwei unitäre Transformationen. Eine, die den Eintrag a_{ij} eliminiert, und eine, die die Dreiecksgestalt von B bewahrt. Dies kann man zum Beispiel mit Givens-Rotationen oder Householder-Spiegelungen umsetzen. Zusammenfassend ergibt sich der folgende Algorithmus.

Algorithmus 4.2 (Hessenberg-Dreiecks-Transformation). *Zu gegebenen Matrizen $A, B \in \mathbb{K}^{n \times n}$ überschreibt der folgende Algorithmus A mit einer oberen Hessenberg-Matrix $Q^* A Z$ und B mit einer oberen Dreiecksmatrix $Q^* B Z$, wobei $Q, Z \in \mathbb{K}^{n \times n}$ unitär sind. Dabei sei **givens**(a, b) die Funktion, die zu gegebenen Werten a und b eine Givens-Rotation so erstellt, dass der Eintrag b eliminiert wird.*

Löse $Q^* B = R$, wobei Q unitär und R eine obere Dreiecksmatrix ist

$A \leftarrow Q^* A$

for $j = 1 \dots n - 2$ **do**

for $i = n \dots j + 2$ **do**

$G = \mathbf{givens}(A_{i-1,j}, A_{ij})$

$A \leftarrow G^* A, B \leftarrow G^* B$

$G = \mathbf{givens}(-B_{ii}, B_{i,i-1})$

$B \leftarrow B G, A \leftarrow A G$

$i \leftarrow i - 1$

end for

$j \leftarrow j + 1$

end for

4 Das QZ-Verfahren

Dieser Algorithmus benötigt in etwa $8n^3$ Operationen, die Berechnung von Q und Z benötigt weitere $4n^3$ bzw. $3n^3$ Operationen.

Die Transformation eines Büschels (A, B) auf Hessenberg-Dreiecks-Form liefert uns die Grundlage für die verallgemeinerte QR-Iteration, auch bekannt als die QZ-Iteration, die nachfolgend beschrieben wird.

4.2 Deflation

In diesem Abschnitt geht es um die Behandlung des Falls einer singulären Matrix B , wie es von Golub und van Loan in [2] beschrieben wird. Es sei (\hat{A}, \hat{B}) eine Schur-Form des regulären Büschels (A, B) . Wir können nun ohne Einschränkung davon ausgehen, dass wir mit \hat{A} bereits eine Hessenberg-Matrix und mit \hat{B} eine obere Dreiecksmatrix vorliegen haben. Weiter können wir annehmen, dass \hat{A} *irreduzibel* ist, das heißt, dass $\hat{a}_{k+1,k} \neq 0$ für alle $k \in \{1, \dots, n-1\}$ gilt.

Wäre \hat{A} nicht irreduzibel, also $\hat{a}_{k+1,k} = 0$ für ein $k \in \{1, \dots, n-1\}$, so können wir das Problem wegen

$$\hat{A} - \lambda \hat{B} = \begin{pmatrix} \hat{A}_{11} - \lambda \hat{B}_{11} & \hat{A}_{12} - \lambda \hat{B}_{12} \\ 0 & \hat{A}_{22} - \lambda \hat{B}_{22} \end{pmatrix} \quad (4.1)$$

mit $\hat{A}_{11} - \lambda \hat{B}_{11} \in \mathbb{K}^{k \times k}$ und $\hat{A}_{22} - \lambda \hat{B}_{22} \in \mathbb{K}^{n-k \times n-k}$ durch Deflation auf zwei Probleme kleinerer Dimension zurückführen.

Bisher haben wir nur den Fall einer invertierbaren Matrix B betrachtet. Die Deflation führt nun dazu, dass wir auch den Fall einer singulären Matrix B behandeln können. Ist B nicht invertierbar, so kann auch die obere Dreiecksmatrix \hat{B} nicht invertierbar sein und dies ist genau dann der Fall, wenn ein Diagonaleintrag gleich null ist, das heißt, es gilt $\hat{b}_{kk} = 0$ für ein $k \in \{1, \dots, n\}$. Wir können diese Situation also sehr leicht erkennen. Tritt dieser Fall ein, so ist es möglich einen Nulleintrag in $\hat{a}_{n,n-1}$ zu erzeugen und dann das gesamte Problem um eine Dimension zu reduzieren, wie es das folgende Beispiel für $n = 5$ und $k = 3$ illustriert. Das Büschel (\hat{A}, \hat{B}) habe Hessenberg-Dreiecks-Form, also

$$\hat{A} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}, \quad \hat{B} = \begin{pmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{pmatrix}.$$

Der Nulleintrag in der Diagonalen von \hat{B} kann durch Givens-Rotationen mit Hilfe des sogenannten „Bulge Chasing“ an die Position $(5, 5)$ verschoben werden. Dazu erzeugen

wir zunächst einen Nulleintrag in \hat{b}_{44} mit einer Givens-Rotation Q_{34} . Diese erzeugt einen Nichtnulleintrag in \hat{a}_{42} .

$$\hat{A} \leftarrow Q_{34}^* \hat{A} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}, \quad \hat{B} \leftarrow Q_{34}^* \hat{B} = \begin{pmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & * \end{pmatrix}.$$

Damit die Hessenberg-Struktur der Matrix \hat{A} erhalten bleibt, erzeugen wir nun mittels einer Givens-Rotation Z_{23} in \hat{a}_{42} einen Nulleintrag ohne die Gestalt von \hat{B} zu beeinflussen.

$$\hat{A} \leftarrow \hat{A}Z_{23} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}, \quad \hat{B} \leftarrow \hat{B}Z_{23} = \begin{pmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & * \end{pmatrix}.$$

Der nächste Schritt ist nun die Elimination des Diagonaleintrages \hat{b}_{55} . Die dazu nötige Givens-Rotation Q_{45} erzeugt einen Nichtnulleintrag in \hat{a}_{53} .

$$\hat{A} \leftarrow Q_{45}^* \hat{A} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & * & * & * \end{pmatrix}, \quad \hat{B} \leftarrow Q_{45}^* \hat{B} = \begin{pmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Durch eine Givens-Rotation Z_{34} erzeugen wir den Nulleintrag in \hat{a}_{53} und gleichzeitig in \hat{b}_{33} einen Nichtnulleintrag.

$$\hat{A} \leftarrow \hat{A}Z_{34} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}, \quad \hat{B} \leftarrow \hat{B}Z_{34} = \begin{pmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

4 Das QZ-Verfahren

Im letzten Schritt eliminieren wir nun mittels Z_{45} den Eintrag \hat{a}_{54} und erhalten außerdem einen Nichtnulleintrag in \hat{b}_{44} .

$$\hat{A} \leftarrow \hat{A}Z_{45} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & 0 & * \end{pmatrix}, \quad \hat{B} \leftarrow \hat{B}Z_{45} = \begin{pmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Wie man sieht, haben wir nun den Nulleintrag aus \hat{b}_{33} auf \hat{b}_{55} verschoben und gleichzeitig \hat{a}_{54} eliminiert. Lemma 2.1 besagt nun, dass wir einen Eigenwert $\langle \hat{a}_{55}, 0 \rangle$ erhalten haben und ihn abspalten können, das heißt, wir können mit einem Bündel der Dimension 4 fortfahren.

Diese „Bulge Chasing“-Technik funktioniert sehr allgemein und kann dazu genutzt werden, den Eintrag $a_{n,n-1}$ zu eliminieren, unabhängig davon, wo der Nulleintrag entlang der Diagonalen von B auftritt. Wir sind jetzt also in der Lage singuläre Matrizen bei unseren Berechnungen zu berücksichtigen, wobei wir in diesem Fall die Problemgröße sogar schon besonders früh verringern können.

4.3 Der QZ-Schritt

In diesem Abschnitt wenden wir uns nun dem eigentlichen QZ-Schritt zu. Wie bei der QR-Iteration gibt es auch hier verschiedene Shift-Strategien. War unser Ziel bei der QR-Iteration aus einer Hessenberg-Matrix die gewöhnliche Schur-Form zu erhalten, so ist es jetzt unser Ziel ein Bündel in Hessenberg-Dreiecks-Gestalt auf verallgemeinerte Schur-Form zu transformieren.

Wir werden zuerst den allgemeinen impliziten QZ-Schritt beschreiben und beziehen uns dabei auf [5].

Sei (A, B) ein Bündel in Hessenberg-Dreiecks-Form und sei außerdem B regulär. Setzen wir $y = Bx$, so gilt

$$Ax = \lambda Bx \quad \Leftrightarrow \quad AB^{-1}y = \lambda y. \quad (4.2)$$

Da B^{-1} aufgrund der Dreiecksform von B ebenfalls eine obere Dreiecksmatrix ist, hat die Matrix $C^{(0)} = AB^{-1}$ Hessenberg-Gestalt und wir können ihre Schur-Zerlegung berechnen, indem wir die in der QR-Iteration auftretenden Transformationen $H^{(0)}, H^{(1)}, \dots$ anwenden:

$$C^{(0)} := AB^{-1}, \quad C^{(m+1)} := (H^{(m)})^* C^{(m)} H^{(m)} \quad \text{für alle } m \in \mathbb{N}_0. \quad (4.3)$$

Da wir aber die Berechnung der Matrix B^{-1} vermeiden wollen, um die Kondition des Problems nicht zu verschlechtern, versuchen wir den QR-Schritt so abzuändern, dass wir ihn direkt auf A und B anwenden können. Dazu definieren wir

$$\begin{aligned} A^{(0)} &:= A, & A^{(m+1)} &:= (H^{(m)}Q^{(m)})^* A^{(m)} Z^{(m)}, \\ B^{(0)} &:= B, & B^{(m+1)} &:= (H^{(m)}Q^{(m)})^* B^{(m)} Z^{(m)} \end{aligned} \quad (4.4)$$

für alle $m \in \mathbb{N}_0$, wobei $Q^{(m)}$ und $Z^{(m)}$ unitär und so gewählt sind, dass $Q^{(m)}e^{(1)} = e^{(1)}$ gilt und das Büschel

$$(A^{(m+1)}, B^{(m+1)}) = (H^{(m)}Q^{(m)})^*(A^{(m)}, B^{(m)})Z^{(m)} \quad (4.5)$$

Hessenberg-Dreiecks-Gestalt hat. Daraus erhalten wir wegen $C^{(0)} = A^{(0)}(B^{(0)})^{-1}$ nun mit Induktion

$$\begin{aligned} A^{(m+1)}(B^{(m+1)})^{-1} &= ((H^{(m)}Q^{(m)})^*(A^{(m)}Z^{(m)}))((H^{(m)}Q^{(m)})^*B^{(m)}Z^{(m)})^{-1} \\ &= ((H^{(m)}Q^{(m)})^*A^{(m)}Z^{(m)})((Z^{(m)})^*(B^{(m)})^{-1}H^{(m)}Q^{(m)}) \\ &= (H^{(m)}Q^{(m)})^*A^{(m)}(B^{(m)})^{-1}(H^{(m)}Q^{(m)}) \\ &= (H^{(m)}Q^{(m)})^*C^{(m)}(H^{(m)}Q^{(m)}) \end{aligned} \quad (4.6)$$

für alle $m \in \mathbb{N}_0$. Um einzusehen, dass $C^{(m+1)} = (H^{(m)}Q^{(m)})^*C^{(m)}(H^{(m)}Q^{(m)})$ gilt, also $C^{(m+1)}$ das Ergebnis eines impliziten QR-Schritts angewandt auf $C^{(m)} = A^{(m)}(B^{(m)})^{-1}$ ist, benötigen wir folgenden Satz.

Satz 4.3 (Implizites Q-Theorem). *Sei $A \in \mathbb{K}^{n \times n}$ und $H \in \mathbb{K}^{n \times n}$ eine irreduzible Hessenberg-Matrix. Dann ist eine unitäre Matrix $Q \in \mathbb{K}^{n \times n}$, welche $H = Q^*AQ$ erfüllt, bis auf Skalierung eindeutig durch ihre erste oder letzte Spalte bestimmt.*

Beweis. Wir fassen Q spaltenweise auf und schreiben $Q = (q_1, q_2, \dots, q_n)$. Für die erste Spalte der Gleichung $AQ = QH$ gilt damit

$$Aq_1 = h_{11}q_1 + h_{21}q_2.$$

Wegen $q_1^*q_1 = 1$ und $q_1^*q_2 = 0$ ist dies äquivalent zu

$$h_{11} = q_1^*Aq_1.$$

Weiter ist

$$h_{21}q_2 = Aq_1 - h_{11}q_1.$$

4 Das QZ-Verfahren

Da H irreduzibel ist, ist $h_{21} \neq 0$, und damit auch $Aq_1 - h_{11}q_1 \neq 0$. Wegen $\|q_2\|_2 = 1$ ist q_2 durch $|h_{21}| = \|Aq_1 - h_{11}q_1\|_2$ bis auf Vorzeichen eindeutig bestimmt.

Angenommen, $k < n$ und q_1, \dots, q_k wurden bereits bestimmt. Dann gilt für die k -te Spalte der Gleichung $AQ = QH$

$$Aq_k = h_{1k}q_1 + \dots + h_{kk}q_k + h_{k+1,k}q_{k+1}.$$

Da q_1, \dots, q_{k+1} orthonormal zueinander sind, gilt $h_{ik} = q_i^* Aq_k$ für alle $i \in \{1, \dots, k\}$. Wegen $h_{k+1,k} \neq 0$ ist auch $Aq_k - h_{1k}q_1 - \dots - h_{kk}q_k \neq 0$ und durch

$$|h_{k+1,k}| = \|Aq_k - h_{1k}q_1 - \dots - h_{kk}q_k\|_2$$

wird q_{k+1} bis auf Vorzeichen eindeutig bestimmt.

Die letzte Spalte der Gleichung $AQ = QH$ liefert

$$h_{in} = q_i^* Aq_n, \quad i = 1, \dots, n.$$

Um zu beweisen, dass Q durch die letzte Spalte eindeutig bestimmt ist, betrachten wir die Gleichung $Q^*A = HQ^*$ und wenden obige Argumente auf die Zeilen von Q^* an, beginnend mit der letzten Zeile. \square

Dieser Satz heißt implizites Q-Theorem, da er die Bestimmung von Q implizit durch die erste oder letzte Spalte erlaubt. Wir erhalten nun direkt folgendes Korollar.

Korollar 4.4. *Seien $A \in \mathbb{K}^{n \times n}$, $H_1, H_2 \in \mathbb{K}^{n \times n}$ Hessenberg-Matrizen und $Q_1, Q_2 \in \mathbb{K}^{n \times n}$ unitär mit $Q_1^* A Q_1 = H_1$ und $Q_2^* A Q_2 = H_2$. Gilt $Q_1 e^{(1)} = Q_2 e^{(1)}$, so folgt $H_1 = H_2$.*

Wir kehren nun zu unseren Überlegungen vor Satz 4.3 zurück und stellen fest, dass $H^{(m)} Q^{(m)} e^{(1)} = H^{(m)} e^{(1)}$ auf Grund der Wahl von $Q^{(m)}$ gilt und somit das implizite Q-Theorem besagt, dass $C^{(m+1)}$ das Ergebnis eines impliziten QR-Schritts angewandt auf $C^{(m)} = A^{(m)} (B^{(m)})^{-1}$ ist. Es gilt also

$$C^{(m)} = A^{(m)} (B^{(m)})^{-1} \quad \text{für alle } m \in \mathbb{N}_0. \quad (4.7)$$

Die durch (4.4) berechneten Matrizen $A^{(m)}$ und $B^{(m)}$ entsprechen denen, die bei der gewöhnlichen QR-Iteration entstehen. Wählen wir die unitären Transformationen so, dass die Matrizen $C^{(m)}$ gegen obere Dreiecksform konvergieren, erwarten wir, dass der untere Nebendiagonaleintrag $c_{n,n-1}^{(m)}$ gegen Null streben wird. Da $(A^{(m)}, B^{(m)})$ Hessenberg-Gestalt hat, folgt aus (4.7)

$$a_{n,n-1}^{(m)} = c_{n,n-1}^{(m)} b_{n-1,n-1}^{(m)}. \quad (4.8)$$

Falls also $c_{n,n-1}^{(m)} = 0$ gilt, folgt demnach $a_{n,n-1}^{(m)} = 0$. Somit haben wir einen Neben-diagonaleintrag der Matrix $A^{(m)}$ eliminiert und nach Lemma 2.1 einen Eigenwert des Büschels $(A^{(m)}, B^{(m)})$ erhalten, den wir abspalten können, um anschließend mit einem Problem kleinerer Dimension fortzufahren. Sollte $b_{n-1,n-1}^{(m)}$ gegen Null konvergieren, so ist die Matrix B singulär und wir erhalten einen unendlichen Eigenwert.

4.3.1 Der QZ-Schritt mit Doppel-Shift nach Moler und Stewart

Wir wollen nun die Strategie des Doppel-Shifts untersuchen und einen Weg zur direkten Anwendung der QR-Iteration auf A und B beschreiben. Das folgende Verfahren wurde von Moler und Stewart in [3] beschrieben.

Es sei $C^{(m)} = A^{(m)}(B^{(m)})^{-1}$ und $v^{(m)}$ die erste Spalte der Matrix $(C^{(m)} - \lambda^{(m)}I)(C^{(m)} - \mu^{(m)}I)$, wobei $\lambda^{(m)}$ und $\mu^{(m)}$ die Eigenwerte der untersten 2×2 -Teilmatrix von $C^{(m)}$ sind und demnach die Eigenwerte des Büschels $(\bar{A}^{(m)}, \bar{B}^{(m)})$ mit

$$\bar{A}^{(m)} = \begin{pmatrix} a_{n-1,n-1}^{(m)} & a_{n-1,n}^{(m)} \\ a_{n,n-1}^{(m)} & a_{nn}^{(m)} \end{pmatrix}, \quad \bar{B}^{(m)} = \begin{pmatrix} b_{n-1,n-1}^{(m)} & b_{n-1,n}^{(m)} \\ 0 & b_{nn}^{(m)} \end{pmatrix}. \quad (4.9)$$

Da wir $\lambda^{(m)}$ und $\mu^{(m)}$ nicht explizit ausrechnen wollen, können wir $v^{(m)}$, wie in [3] beschrieben, in $O(1)$ berechnen. Dazu müssen wir uns verdeutlichen, dass $v^{(m)}$ wegen

$$(C^{(m)} - \lambda^{(m)}I)(C^{(m)} - \mu^{(m)}I) = (C^{(m)})^2 - (\lambda^{(m)} + \mu^{(m)})C^{(m)} + \lambda^{(m)}\mu^{(m)}I \quad (4.10)$$

und der Hessenberg-Struktur von $C^{(m)}$ genau drei Nichtnulleinträge hat. Diese können wir wie folgt darstellen:

$$\begin{aligned} v_1 &= \left[\begin{pmatrix} a_{n-1,n-1}^{(m)} & a_{11}^{(m)} \\ b_{n-1,n-1}^{(m)} & b_{11}^{(m)} \end{pmatrix} \begin{pmatrix} a_{nn}^{(m)} & a_{11}^{(m)} \\ b_{nn}^{(m)} & b_{11}^{(m)} \end{pmatrix} - \begin{pmatrix} a_{n-1,n}^{(m)} \\ b_{nn}^{(m)} \end{pmatrix} \begin{pmatrix} a_{n,n-1}^{(m)} \\ b_{n-1,n-1}^{(m)} \end{pmatrix} \right. \\ &\quad \left. + \begin{pmatrix} a_{n,n-1}^{(m)} \\ b_{n-1,n-1}^{(m)} \end{pmatrix} \begin{pmatrix} b_{n-1,n}^{(m)} \\ b_{nn}^{(m)} \end{pmatrix} \begin{pmatrix} a_{11}^{(m)} \\ b_{11}^{(m)} \end{pmatrix} \right] \cdot \begin{pmatrix} b_{11}^{(m)} \\ a_{21}^{(m)} \end{pmatrix} + \begin{pmatrix} a_{12}^{(m)} \\ b_{22}^{(m)} \end{pmatrix} - \begin{pmatrix} a_{11}^{(m)} \\ b_{11}^{(m)} \end{pmatrix} \begin{pmatrix} b_{12}^{(m)} \\ b_{22}^{(m)} \end{pmatrix}, \\ v_2 &= \begin{pmatrix} a_{22}^{(m)} & a_{11}^{(m)} \\ b_{22}^{(m)} & b_{11}^{(m)} \end{pmatrix} - \begin{pmatrix} a_{21}^{(m)} \\ b_{11}^{(m)} \end{pmatrix} \begin{pmatrix} b_{12}^{(m)} \\ b_{22}^{(m)} \end{pmatrix} - \begin{pmatrix} a_{n-1,n-1}^{(m)} & a_{11}^{(m)} \\ b_{n-1,n-1}^{(m)} & b_{11}^{(m)} \end{pmatrix} - \begin{pmatrix} a_{nn}^{(m)} & a_{11}^{(m)} \\ b_{nn}^{(m)} & b_{11}^{(m)} \end{pmatrix} \\ &\quad + \begin{pmatrix} a_{n,n-1}^{(m)} \\ b_{n-1,n-1}^{(m)} \end{pmatrix} \begin{pmatrix} b_{n-1,n}^{(m)} \\ b_{nn}^{(m)} \end{pmatrix}, \\ v_3 &= \frac{a_{32}^{(m)}}{b_{22}^{(m)}}. \end{aligned} \quad (4.11)$$

4 Das QZ-Verfahren

Wir sehen, dass nur Differenzen von Diagonalelementen auftreten, also kleine, aber nicht vernachlässigbare Außerdiagonaleinträge durch die Berechnung des Shifts nicht verloren gehen. Weiter müssen wir fordern, dass die erste 2×2 -Hauptuntermatrix und die letzte 2×2 -Hauptuntermatrix von $B^{(m)}$ regulär sind, also $b_{11}^{(m)}, b_{22}^{(m)}, b_{n-1,n-1}^{(m)}, b_{nn}^{(m)} \neq 0$ sind. Ist dies nicht der Fall, so können wir die Regularität durch Deflation, wie in (4.2) beschrieben, erreichen.

Zu dem Vektor $v^{(m)}$ bestimmen wir dann eine 3×3 -Householder-Matrix $H^{(m)}$, die $v_2^{(m)}$ und $v_3^{(m)}$ eliminiert, und wenden sie auf das Büschel $(A^{(m)}, B^{(m)})$ an. Dies liefert folgende Form für den Fall $n = 5$:

$$A^{(m)} \leftarrow H^{(m)} A^{(m)} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}, \quad B^{(m)} \leftarrow H^{(m)} B^{(m)} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{pmatrix}.$$

Anschließend transformieren wir das Büschel durch schrittweises Eliminieren der unerwünschten Nichtnulleinträge auf Hessenberg-Dreiecks-Form zurück. Dazu bestimmen wir zuerst eine Householder-Matrix $Z_1^{(m)}$, um die Einträge $b_{31}^{(m)}$ und $b_{32}^{(m)}$ zu eliminieren und anschließend $Z_2^{(m)}$, um den Eintrag $b_{21}^{(m)}$ zu beseitigen:

$$A^{(m)} \leftarrow A^{(m)} Z_1^{(m)} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}, \quad B^{(m)} \leftarrow B^{(m)} Z_1^{(m)} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{pmatrix},$$

$$A^{(m)} \leftarrow A^{(m)} Z_2^{(m)} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}, \quad B^{(m)} \leftarrow B^{(m)} Z_2^{(m)} = \begin{pmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{pmatrix}.$$

Als nächstes bestimmen wir eine Householder-Matrix $Q_1^{(m)}$, um die Einträge $a_{31}^{(m)}$ und $a_{41}^{(m)}$ zu eliminieren:

$$A^{(m)} \leftarrow Q_1^{(m)} A^{(m)} = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}, \quad B^{(m)} \leftarrow Q_1^{(m)} B^{(m)} = \begin{pmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & * \end{pmatrix}.$$

Durch diesen Schritt werden die unerwünschten Nichtnulleinträge von ihrer ursprünglichen Position aus nach rechts unten verschoben und nachdem wir diesen Schritt insgesamt $(n - 2)$ -mal durchgeführt haben, hat das Bündel wieder Hessenberg-Dreiecks-Gestalt. Dies stellt einen typischen Schritt der QZ-Iteration dar.

Die orthogonale Matrix $H^{(m)}Q^{(m)}$, mit $Q^{(m)} = Q_1^{(m)} \cdots Q_{n-2}^{(m)}$, hat offensichtlich dieselbe erste Spalte wie $H^{(m)}$ und das implizite Q-Theorem besagt jetzt, dass

$$A^{(m+1)}(B^{(m+1)})^{-1} = (H^{(m)}Q^{(m)})^*(A^{(m)}(B^{(m)})^{-1})(H^{(m)}Q^{(m)}) \quad (4.12)$$

im Grunde gerade die Matrix ist, die wir erhalten, wenn wir den impliziten QR-Schritt mit Doppel-Shift direkt auf $A^{(m)}(B^{(m)})^{-1}$ anwenden. Insgesamt haben wir

Algorithmus 4.5 (Der QZ-Schritt mit Doppel-Shift). *Zu einer gegebenen irreduziblen oberen Hessenberg-Matrix $A \in \mathbb{K}^{n \times n}$ und einer regulären oberen Dreiecksmatrix $B \in \mathbb{K}^{n \times n}$ überschreibt der folgende Algorithmus A mit einer quasi-oberen Dreiecks-Matrix Q^*AZ und B mit einer oberen Dreiecksmatrix Q^*BZ , wobei $Q, Z \in \mathbb{K}^{n \times n}$ unitär sind.*

Sei $C = AB^{-1}$ und berechne die drei Nichtnulleinträge x, y, z der ersten Spalte wie in (4.11).

for $k = 1 \rightarrow n - 2$ **do**

Finde eine Householdermatrix Q_k , sodass $Q_k(x \ y \ z)^* = (* \ 0 \ 0)^*$.

$A \leftarrow \text{diag}(I_{k-1}, Q_k, I_{n-k-2}) A$

$B \leftarrow \text{diag}(I_{k-1}, Q_k, I_{n-k-2}) B$

Finde eine Householdermatrix Z_{k_1} , sodass $(b_{k+2,k} \ b_{k+2,k+1} \ b_{k+2,k+2}) Z_{k_1} = (0 \ 0 \ *)$

$A \leftarrow \text{Adiag}(I_{k-1}, Z_{k_1}, I_{n-k-2})$

$B \leftarrow \text{Bdiag}(I_{k-1}, Z_{k_1}, I_{n-k-2})$

Finde eine Householdermatrix Z_{k_2} , sodass $(b_{k+1,k} \ b_{k+1,k+1}) Z_{k_2} = (0 \ *)$.

$A \leftarrow \text{Adiag}(I_{k-1}, Z_{k_2}, I_{n-k-1})$

$B \leftarrow \text{Bdiag}(I_{k-1}, Z_{k_2}, I_{n-k-1})$

$x = a_{k+1,k}; y = a_{k+2,k}$

if $k < n - 2$ **then**

$z \leftarrow a_{k+3,k}$

end if

end for

Finde eine Householdermatrix Q_{n-1} so, dass $Q_{n-1}(x \ y)^* = (* \ 0)^*$.

$A \leftarrow \text{diag}(I_{n-2}, Q_{n-1}) A$

$B \leftarrow \text{diag}(I_{n-2}, Q_{n-1}) B$

4 Das QZ-Verfahren

Finde eine Householdermatrix Z_{n-1} so, dass $(b_{n,n-1} \ b_{n,n}) Z_{n-1} = (0 \ *)$.

$A \leftarrow \text{Adiag}(I_{n-2}, Z_{n-1})$

$B \leftarrow \text{Bdiag}(I_{n-2}, Z_{n-1})$

Dieser Algorithmus benötigt $22n^2$ Operationen. Die Matrizen Q und Z können für weitere $8n^2$ bzw. $13n^2$ Operationen gespeichert werden.

Der QZ-Schritt mit Doppel-Shift liefert uns im Allgemeinen noch keine richtige verallgemeinerte Schur-Form des Büschels (A, B) , sondern ein Büschel, in dem A quasi-obere Dreiecksgestalt hat. Das heißt, unser ursprüngliches Problem wurde nun in ein- und zweidimensionale Blöcke zerlegt. Da wir aber daran interessiert sind, sowohl A als auch B auf obere Dreiecksgestalt zu transformieren, werden wir hier nun noch das in [3] beschriebene Verfahren zur weiteren Reduktion auf verallgemeinerte Schur-Form vorstellen.

Es gibt mehrere Gründe, warum wir diese Transformation bevorzugen, anstelle es bei der quasi-oberen Dreiecksform der Matrix A zu belassen und die Eigenwerte durch ein quadratisches Polynom direkt zu berechnen. Zum einen ist die Berechnung der Eigenvektoren eines Büschels in strikter Schur-Form einfacher, zum anderen können wir dem Büschel so mehr Informationen über die Eigenwerte entnehmen, als bei gegebenen Eigenwerten. Sind zum Beispiel die Diagonalelemente a_{ii} und b_{ii} sehr klein, so ist der Eigenwert $\lambda_i = a_{ii}/b_{ii}$ schlecht konditioniert, egal welchen Wert er hat.

Es geht nun um die Transformation eines zweidimensionalen Büschels auf strikte verallgemeinerte Schur-Form. Dies werden wir wie gewohnt mit unitären Äquivalenztransformationen umsetzen. Es sei also (A, B) ein solches Büschel und B eine Dreiecksmatrix.

Die Fälle, in denen ein Diagonaleintrag von B Null ist, können wir leicht behandeln, indem wir jeweils das Element a_{21} eliminieren. Ist $b_{11} = 0$, so wählen wir eine unitäre Transformation Q , sodass Q^*A Dreiecksgestalt hat, und ist $b_{22} = 0$, so wählen wir entsprechend ein unitäres Z , sodass AZ Dreiecksgestalt hat. Durch die Behandlungen von links bzw. rechts bleibt die Nullspalte bzw. -zeile von B erhalten. Diesen Vorgang können wir zum Beispiel mit jeweils einer Givens-Rotation umsetzen.

Es sei B also regulär. Wir wählen nun unitäre Transformationen Q und Z so, dass

$$\hat{A} = Q^*AZ, \quad \hat{B} = Q^*BZ \quad (4.13)$$

Dreiecksgestalt haben. Dazu sei $\lambda \in \sigma(A, B)$ ein Eigenwert des Büschels und es sei

$$E := A - \lambda B. \quad (4.14)$$

Da λ ein Eigenwert des Büschels ist, sind die Zeilen von E linear abhängig. Wir wählen nun eine unitäre Matrix Z so, dass entweder e_{11} oder e_{21} eliminiert wird. Wegen der

Struktur von E folgt direkt, dass auch das jeweils andere Element annulliert wird. Als nächstes wählen wir ein unitäres Q so, dass entweder Q^*AZ oder Q^*BZ obere Dreiecksform hat. Daraus erhalten wir wegen

$$Q^*EZ = Q^*AZ - \lambda Q^*BZ, \quad (4.15)$$

dass sowohl Q^*AZ als auch Q^*BZ obere Dreiecksmatrizen sind, da Z so gewählt war, dass die erste Spalte von EZ Null ist und somit Q^*EZ eine obere Dreiecksmatrix ist. In der Praxis erhalten wir nun folgendes Verfahren.

Algorithmus 4.6. *Zu gegebenen Matrizen $A \in \mathbb{K}^{2 \times 2}$ und $B \in \mathbb{K}^{2 \times 2}$ in Dreiecksform berechnet der folgende Algorithmus unitäre Matrizen $Q, Z \in \mathbb{K}^{2 \times 2}$, sodass das Büschel (Q^*AZ, Q^*BZ) verallgemeinerte Schur-Form hat.*

```

Berechne einen Eigenwert  $\lambda$  des Büschels  $(A, B)$ 
if  $b_{11} = 0$  then
     $G = \mathbf{givens}(a_{11}, a_{21})$ 
     $A \leftarrow G^*A, B \leftarrow G^*B$ 
else if  $b_{22} = 0$  then
     $G = \mathbf{givens}(-a_{22}, a_{21})$ 
     $A \leftarrow AG, B \leftarrow BG$ 
else
    if  $\|(a_{21} - \lambda b_{21}, a_{22} - \lambda b_{22})^T\| > \|(a_{11} - \lambda b_{11}, a_{12} - \lambda b_{12})^T\|$  then
         $G = \mathbf{givens}(-(a_{22} - \lambda b_{22}), a_{21} - \lambda b_{21})$ 
         $A \leftarrow AG, B \leftarrow BG$ 
    else
         $G = \mathbf{givens}(-(a_{12} - \lambda b_{12}), a_{11} - \lambda b_{11})$ 
         $A \leftarrow AG, B \leftarrow BG$ 
    end if
    if  $\|A\| \geq |\lambda| \|B\|$  then
         $G = \mathbf{givens}(b_{11}, b_{21})$ 
         $A \leftarrow G^*A, B \leftarrow G^*B$ 
    else
         $G = \mathbf{givens}(a_{11}, a_{21})$ 
         $A \leftarrow G^*A, B \leftarrow G^*B$ 
    end if
end if

```

Dieser Algorithmus benötigt nur einen Schritt um ein 2×2 -Büschel (A, B) auf Schur-Form zu bringen. Dass die Berechnung von λ und die Wahl der Matrizen Q und Z trotz Rundungsfehlern stabil ist und uns das gewünschte Ergebnis liefert, werden wir hier nicht weiter vertiefen. Details der numerischen Analyse sind in [3] ab Seite 251 zu finden.

4.4 Das QZ-Verfahren

Mit dem QZ-Schritt haben wir nun die Grundlage des QZ-Verfahrens geschaffen. Wenden wir eine Folge von QZ-Schritten auf das Hessenberg-Dreiecks-Büschel (A, B) an, so wird die Matrix A gegen obere Dreiecksform konvergieren. Dabei ist es sinnvoll zu betrachten, ob bereits konvergierte Blöcke abgespalten werden können. Ist der Eintrag $a_{n,n-1}^{(m)}$ bereits klein genug, so können wir den Eigenwert $\langle a_{nn}^{(m)}, b_{nn}^{(m)} \rangle$ abspalten und mit einem Problem der Dimension $n - 1$ fortfahren. Wird dagegen der Eintrag $a_{n-1,n-2}^{(m)}$ klein, so können wir den untersten 2×2 -Block abspalten und ihn mit dem Algorithmus (4.6) behandeln. Auf diese Weise können wir das Problem auf Dimension $n - 2$ reduzieren.

Das vollständige Verfahren, basierend auf dem QZ-Schritt mit Doppel-Shift und der separaten Behandlung der bereits konvergierten 2×2 -Blöcke von Moler und Stewart, wie in [3] beschrieben, ist das folgende.

Algorithmus 4.7. *Zu gegebenen $A, B \in \mathbb{K}^{n \times n}$ berechnet der folgende Algorithmus unitäre Matrizen $Q, Z \in \mathbb{K}^{n \times n}$ so, dass $Q^*AZ = T$ und $Q^*BZ = S$ obere Dreiecksmatrizen sind. A wird dabei mit T und B mit S überschrieben.*

Wende Algorithmus 4.2 auf A und B an

repeat

Setze alle Unterdiagonalelemente von A gleich Null, die

$$|a_{i,i-1}| \leq \epsilon (|a_{i-1,i-1}| + |a_{ii}|) \text{ erfüllen}$$

Finde das größte nichtnegative q und das kleinste nichtnegative p , sodass

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ 0 & A_{22} & A_{23} \\ 0 & 0 & A_{33} \end{pmatrix},$$

wobei $A_{11} \in \mathbb{K}^{p \times p}$, $A_{22} \in \mathbb{K}^{(n-p-q) \times (n-p-q)}$ und $A_{33} \in \mathbb{K}^{q \times q}$, A_{33} eine Dreiecksmatrix ist und A_{22} eine irreduzible obere Hessenberg-Matrix ist.

Partitioniere B genauso:

$$B = \begin{pmatrix} B_{11} & B_{12} & B_{13} \\ 0 & B_{22} & B_{23} \\ 0 & 0 & B_{33} \end{pmatrix},$$

wobei $B_{11} \in \mathbb{K}^{p \times p}$, $B_{22} \in \mathbb{K}^{(n-p-q) \times (n-p-q)}$ und $B_{33} \in \mathbb{K}^{q \times q}$

```

if  $q < n$  then
  if  $B_{22}$  singular then
    Eliminiere  $a_{n-q, n-q-1}$ 
  else if  $n - p - q = 2$  then
    Wende Algorithmus 4.6 auf  $A_{22}$  und  $B_{22}$  an
     $A \leftarrow \text{diag}(I_p, Q, I_q) * \text{Adiag}(I_p, Z, I_q)$ 
     $B \leftarrow \text{diag}(I_p, Q, I_q) * B \text{diag}(I_p, Z, I_q)$ 
  else
    Wende Algorithmus 4.5 auf  $A_{22}$  und  $B_{22}$  an
     $A \leftarrow \text{diag}(I_p, Q, I_q) * \text{Adiag}(I_p, Z, I_q)$ 
     $B \leftarrow \text{diag}(I_p, Q, I_q) * B \text{diag}(I_p, Z, I_q)$ 
  end if
end if
until  $q = n$ 

```

Dieser Algorithmus benötigt $30n^3$ Operationen. Wenn Q benötigt wird, sind weitere $16n^3$ Operationen, falls Z benötigt wird, weitere $20n^3$ Operationen nötig. Diese Schätzungen der Laufzeit basieren auf der Erfahrung, dass etwa zwei QZ-Iterationen pro Eigenwert nötig sind. Weiter sind die Konvergenz-Eigenschaften des QZ-Verfahrens dieselben, wie bei dem QR-Verfahren. Die Geschwindigkeit des QZ-Algorithmus ist unabhängig davon, ob B vollen Rang hat oder nicht.

Da, wie wir gesehen haben, ein QZ-Schritt einem QR-Schritt entspricht, hat das QZ-Verfahren im Grunde das gleiche Konvergenzverhalten, wie die QR-Iteration. Dies bedeutet, dass das Verfahren konvergiert, sobald es einen Eigenwert des Büschels (A, B) gibt, der betragsmäßig echt kleiner ist, als die anderen. Allerdings ist die Konvergenz von den Shift-Strategien abhängig. So kann es zum Beispiel passieren, dass eine bestimmte Shift-Strategie bei einem Büschel nicht zum Ziel führt.

In der Praxis ist es daher oft so, dass man das Verfahren abbricht, wenn eine bestimmte Anzahl an Iterationen pro Eigenwert überschritten wird, und man anschließend eine andere Shift-Strategie probiert.

Sind wir an der Berechnung der Eigenvektoren eines Büschels (A, B) interessiert, so

4 Das QZ-Verfahren

berechnen wir zuerst die Eigenvektoren des auf verallgemeinerte Schur-Form transformierten Büschels (\hat{A}, \hat{B}) und wenden anschließend die unitäre Matrix Z auf die Vektoren an. Ist nämlich y ein Eigenvektor des Büschels (\hat{A}, \hat{B}) zum Eigenwert λ , so gilt

$$Q^*AZy = \lambda Q^*BZy, \quad x = Zy \quad \Leftrightarrow \quad Ax = \lambda Bx. \quad (4.16)$$

4.5 Beispiele

Mit dem vollständigen QZ-Verfahren sind wir nun in der Lage Eigenwerte von Büscheln zu bestimmen. Dazu betrachten wir zwei Beispiele, die mit dem QZ-Verfahren aus dem Anhang berechnet wurden. Das erste Beispiel ist von Moler und Stewart aus [3].

Beispiel 4.8. *Seien*

$$A := \begin{pmatrix} 50 & -60 & 50 & -27 & 6 & 6 \\ 38 & -28 & 27 & -17 & 5 & 5 \\ 27 & -17 & 27 & -17 & 5 & 5 \\ 27 & -28 & 38 & -17 & 5 & 5 \\ 27 & -28 & 27 & -17 & 16 & 5 \\ 27 & -28 & 27 & -17 & 5 & 16 \end{pmatrix} \quad \text{und} \quad B := \begin{pmatrix} 16 & 5 & 5 & 5 & -6 & 5 \\ 5 & 16 & 5 & 5 & -6 & 5 \\ 5 & 5 & 16 & 5 & -6 & 5 \\ 5 & 5 & 5 & 16 & -6 & 5 \\ 5 & 5 & 5 & 5 & -6 & 16 \\ 6 & 6 & 6 & 6 & -5 & 6 \end{pmatrix}.$$

Das Büschel (A, B) ist regulär und hat die sechs Eigenwerte

$$\begin{aligned} \lambda_1 &= \infty = \lambda_2, \\ \lambda_3 &= \frac{1}{2} + \frac{\sqrt{3}}{2}i = \lambda_6, \\ \lambda_4 &= \frac{1}{2} - \frac{\sqrt{3}}{2}i = \lambda_5. \end{aligned}$$

Das QZ-Verfahren liefert uns bei einer Genauigkeit von $\epsilon = 10^{-8}$ nach 17 Iterationsschritten folgende Näherungen der Eigenwerte $\lambda_1, \dots, \lambda_6$ in projektiver Darstellung:

$$\begin{aligned} &< -25.76867071, 0.00000000 > \\ &< 12.82184114, 0.00000000 > \\ &< 10.17806447 - 5.62671338 \cdot i, 0.21636462 - 11.62709722 \cdot i > \\ &< 10.15393396 + 5.61337338 \cdot i, 0.21585166 + 11.59953130 \cdot i > \\ &< -5.73594394 + 9.93538543 \cdot i, -11.47297631 > \\ &< -5.51035603 - 9.54463844 \cdot i, -11.02175768 > . \end{aligned}$$

Wir erhalten somit zwei unendliche und vier endliche Eigenwerte:

$$\begin{aligned}\lambda_1 &= \infty \\ \lambda_2 &= -\infty \\ \lambda_3 &\approx 0.50004743 + 0.86606929 \cdot i \\ \lambda_4 &\approx 0.50004743 - 0.86606929 \cdot i \\ \lambda_5 &\approx 0.49995257 - 0.86598152 \cdot i \\ \lambda_6 &\approx 0.49995257 + 0.86598152 \cdot i.\end{aligned}$$

Ein weiteres Beispiel von Golub und van Loan aus [2] liefert folgendes.

Beispiel 4.9. *Es seien*

$$A := \begin{pmatrix} 2 & 3 & 4 & 5 & 6 \\ 4 & 4 & 5 & 6 & 7 \\ 0 & 3 & 6 & 7 & 8 \\ 0 & 0 & 2 & 8 & 9 \\ 0 & 0 & 0 & 1 & 10 \end{pmatrix} \quad \text{und} \quad B := \begin{pmatrix} 1 & -1 & -1 & -1 & -1 \\ 0 & 1 & -1 & -1 & -1 \\ 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Wir erhalten mit einer Genauigkeit von $\epsilon = 10^{-8}$ nach 9 Iterationsschritten folgende Eigenwerte in projektiver Darstellung:

$$\begin{aligned}&< 7.95005034, 0.37418297 > \\ &< -1.92504630, -1.46583199 > \\ &< -0.30969929, 1.65302645 > \\ &< 6.72071808, 1.21357368 > \\ &< -10.98755604, -0.90883666 >\end{aligned}$$

und somit die Eigenwerte

$$\begin{aligned}\lambda_1 &\approx 21.24642472 \\ \lambda_2 &\approx 1.31327895 \\ \lambda_3 &\approx -0.18735289 \\ \lambda_4 &\approx 5.53795637 \\ \lambda_5 &\approx 12.08969285.\end{aligned}$$

4 Das QZ-Verfahren

Das Verfahren benötigt für das Beispiel 4.9 im Fall reeller Eigenwerte und einer regulären Matrix B weniger als zwei Iterationsschritte pro Eigenwert. Dagegen ist das Verfahren für das Bündel mit komplexen Eigenwerten und einer singulären Matrix B in Beispiel 4.8 etwas langsamer, hier werden fast drei Iterationsschritte je Eigenwert benötigt.

5 Ausblicke

5.1 Verallgemeinerte Vektoriteration

Wir können auch die Vektoriteration aus Kapitel 3.1 auf verallgemeinerte Eigenwertprobleme übertragen. Die inverse Iteration eignet sich zur Berechnung von gut genäherten Eigenvektoren.

Sei (A, B) ein reguläres Büschel und seien ein Startvektor $q^{(0)} \in \mathbb{K}^n$ und ein Shift $\mu \in \mathbb{K}$ gegeben.

```
while „Fehler zu groß“ do  
  Löse  $(A - \mu B)z^{(k)} = Bq^{(k-1)}$   
   $q^{(k)} \leftarrow z^{(k)} / \|z^{(k)}\|$   
end while
```

Ist B regulär, so entspricht dieses Verfahren der inversen Iteration angewandt auf die Matrix $B^{-1}A$.

Wurde mit dem QZ-Verfahren ein gut approximierter Eigenwert μ berechnet, so reicht im Normalfall, wegen der guten Konvergenzeigenschaften der inversen Iteration mit Shift, ein Schritt aus, um einen Eigenvektor x zu μ zu berechnen.

Dabei ist es sinnvoll, die inverse Iteration mit einem auf Hessenberg-Dreiecks-Gestalt transformierten Büschel (\hat{A}, \hat{B}) durchzuführen. In diesem Fall hat die Matrix $(\hat{A} - \mu B)$ ebenfalls Hessenberg-Gestalt, und eine QR-Zerlegung lässt sich, wie wir bereits gesehen haben, mit geringem Aufwand konstruieren. Somit können wir das Lösen des Gleichungssystems

$$(\hat{A} - \mu \hat{B})z^{(k)} = \hat{B}q^{(k-1)} \tag{5.1}$$

effizient durchführen. Haben wir einen Eigenvektor y zu einem Eigenwert des Büschels (\hat{A}, \hat{B}) berechnet, so erhalten wir wie in (4.16) einen zugehörigen Eigenvektor x des Büschels (A, B) .

5.2 Selbstadjungierte positiv definite verallgemeinerte Eigenwertprobleme

Bisher haben wir uns mit allgemeinen Matrizen A und B befasst. Nun soll der wichtige Spezialfall einer selbstadjungierten Matrix $A \in \mathbb{K}^{n \times n}$ und einer selbstadjungierten und positiv definiten Matrix $B \in \mathbb{K}^{n \times n}$ untersucht werden. Derartige Probleme treten in der Praxis häufig auf, zum Beispiel bei Schwingungsproblemen in der Physik.

5 Ausblicke

Unser Ziel ist es, das verallgemeinerte Eigenwertproblem

$$Ax = \lambda Bx \quad (5.2)$$

für $\lambda \in \mathbb{K}$ und $x \in \mathbb{K}^n \setminus \{0\}$ numerisch stabil auf ein gewöhnliches Eigenwertproblem zurückzuführen, ohne dabei die Selbstadjungiertheit der Matrix A zu zerstören. Da die Matrix B selbstadjungiert und positiv definit ist, existiert die Cholesky-Zerlegung, das heißt, es existiert eine untere Dreiecksmatrix $L \in \mathbb{K}^{n \times n}$ mit $B = LL^*$. Indem wir dies in unser Problem (5.2) einsetzen, die Identität $I = (L^*)^{-1}L^*$ einfügen und von links mit L^{-1} multiplizieren erhalten wir

$$Ax = \lambda Bx \quad \Leftrightarrow \quad (L^{-1}A(L^*)^{-1})(L^*x) = \lambda(L^{-1}L)(L^*x). \quad (5.3)$$

Indem wir

$$\hat{A} := L^{-1}A(L^*)^{-1}, \quad \hat{x} := L^*x \quad (5.4)$$

introduzieren, erhalten wir das gewöhnliche Eigenwertproblem

$$\hat{A}\hat{x} = \lambda\hat{x} \quad (5.5)$$

mit einer selbstadjungierten Matrix \hat{A} :

$$\hat{A}^* = (L^{-1}A(L^*)^{-1})^* = L^{-1}A^*(L^*)^{-1} = L^{-1}A(L^*)^{-1} = \hat{A}. \quad (5.6)$$

Einen Eigenvektor x zu einem Eigenwert λ des Büschels (A, B) erhalten wir aus einem Eigenvektor y des gewöhnlichen Problems (5.5) des Eigenwertes λ durch (5.4).

Wir können nun (5.5) mit gewöhnlichen Eigenwertverfahren für selbstadjungierte Probleme lösen.

Literaturverzeichnis

- [1] S. Börm. *Numerik von Eigenwertaufgaben*. Vorlesungsskript, 2011
- [2] G. H. Golub und C. F. van Loan. *Matrix Computations*. The Johns Hopkins University Press, 1996
- [3] C. B. Moler und G. W. Stewart. *An algorithm for generalized matrix eigenvalue problems*. SIAM J. Numer. Anal. Vol. 10, No.2, April 1973, pp. 241-256
- [4] G. W. Stewart. *On the sensitivity of the eigenvalue problem $Ax = \lambda Bx$* . SIAM J. Numer. Anal. Vol. 9, No. 4, December 1972, pp. 669-686
- [5] G. W. Stewart. *Matrix Algorithms Volume II: Eigensystems*. SIAM, 2001

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Weiterhin versichere ich, dass diese Arbeit noch nicht als Abschlussarbeit an anderer Stelle vorgelegen hat.

9. Januar 2012, Janina Gnutzmann.