# ÜBER JACOBI - ÄHNLICHE VERFAHREN ZUR LÖSUNG DES EIGENWERTPROBLEMS NICHT - NORMALER KOMPLEXER MATRIZEN

DISSERTATION

zur

Erlangung des Grades eines Dr. rer. nat.

des Fachbereiches Mathematik und Informatik

der FERNUNIVERSITÄT - Gesamthochschule - HAGEN

vorgelegt von

ANDREAS ERDMANN

aus Hagen

Hagen 1984

# INHALTSVERZEICHNIS


1. EINLEITUNG

1.1 Einführung und Überblick 1

1.2 Bezeichnungen, grundlegende Definitionen und Sätze 6

2. ÜBER DAS VERFAHREN VON EBERLEIN

2.1 Einführung 10

2.2 Der Verfahrens - Algorithmus 12

2.3 Betrachtung bei fast - normalen Matrizen 15

2.4 Eine Modifikation des Verfahrens von Eberlein 29

2.5 Beweis der globalen Konvergenz der Modifikation des Verfahrens von Eberlein 33

3. ÜBER EIN NEUES JACOBI - ÄHNLICHES VERFAHREN

3.1 Einführung 41

3.2 Der Verfahrens - Algorithmus 43

3.3 Beweis der asymptotisch quadratischen Konvergenz des Verfahrens im Falle getrennter Eigenwerte 49

4. NUMERISCHE RESULTATE

4.1 Numerische Resultate bei kleinen Testmatrizen 108

4.2 Parameteranalyse, Naive Methode 113

4.3 Numerische Resultate bei der Parameteranalyse großer Matrizen mittels des Vektorrechners CYBER 205 117

Literaturverzeichnis 122

# 1. EINLEITUNG

## 1.1 EINFÜHRUNG UND ÜBERBLICK

In der vorliegenden Arbeit beschäftigen wir uns mit Jacobi-ähnlichen Verfahren zur Diagonalisierung beliebiger (quadratischer) komplexer Matrizen. Solche Verfahren sind Verallgemeinerungen einer bereits 1846 von JACOBI [12] vorgeschlagenen Methode zur Berechnung aller Eigenwerte und Eigenvektoren reeller symmetrischer Matrizen. Jacobi verwendete dazu reelle zweidimensionale Drehungen, sogenannte Jacobi-Rotationen, um eine gegebene Matrix $A = A_0$ orthogonal in Matrizen $A_1$, $A_2$, $A_3$ etc. zu transformieren, so daß die Folge der $A_i$ gegen eine Matrix in Diagonalform strebt.

Aufgrund seines hohen Rechenaufwandes blieb dieses Verfahren jedoch lange Zeit unpraktikabel, bis es 1949 nach der Entwicklung erster leistungsfähiger, programmgesteuerter Rechenanlagen von GOLDSTINE, VON NEUMANN und MURRAY [9] wiederentdeckt wurde. GOLDSTINE und HORWITZ [8] entwickelten es dann derart weiter, daß mit Hilfe komplexer zweidimensionaler Drehungen beliebige normale Matrizen diagonalisiert werden konnten.

Nun ist nach einem Satz von SCHUR [19] das Quadrat der euklidischen Norm einer Matrix A nach unten beschränkt durch die Summe über die Betragsquadrate der Eigenwerte von A, und Gleichheit gilt genau für normale Matrizen. Da unitäre Ähnlichkeitstransformationen die euklidische Norm einer Matrix erhalten, ist es mit ihnen allein also nicht möglich, eine nicht-normale Matrix zu diagonalisieren.

Daher verwenden die sogenannten Jacobi-ähnlichen Verfahren für nicht-normale Matrizen zusätzlich zu komplexen Jacobi-Rotaionen noch nicht-unitäre Transformationen, mit denen die euklidische Norm der iterierten Matrix schrittweise reduziert wird.

Ein solches Verfahren wurde 1970 von EBERLEIN [4] vorgeschlagen (vgl. Kapitel 2). Obwohl die numerischen Resultate dieser Methode sehr zufriedenstellend sind und RUHE [13] für den Fall getrennter Eigenwerte asymptotisch quadratische Konvergenz für eine leicht modifizierte Version beweisen konnte, existiert bis heute kein Beweis für die globale Konvergenz des Eberlein-Verfahrens bei üblicher zyklischer Pivotwahl. Eine Schwierigkeit für einen solchen Beweis liegt darin, daß in Abhängigkeit von gewissen Bedingungen in jedem Iterationsschritt entweder das Pivotelement des hermiteschen Anteils H,

$$H := \frac{1}{2}(A + A^*) \ ,$$

oder aber des schiefhermiteschen Anteils S,

$$S := \frac{1}{2i}(A - A^*) \ ,$$

der iterierten Matrix A annihiliert wird, jedoch insgesamt keine monotone Abnahme der Gößе $\tau^2(A)$,

$$\tau^2(A) := \sum_{i \neq j} |a_{ij}|^2 \ ,$$

vorliegt.

Hier leisten wir nun in Abschnitt 2.3 einen Beitrag, indem wir folgendes zeigen (Satz 2.3.3):

Ist die Matrix A für hinreichend kleines $\varepsilon$ "fast-normal" von der Ordnung $\varepsilon^2$, d.h. gilt für die euklidische Norm des Kommutators C(A)

$$||C(A)||^2 = O(\varepsilon^4) \ ,$$

gilt weiter z.B. für den hermiteschen Anteil H der Matrix A

$$\tau^2(H) = O(\varepsilon^2),$$

und wird in dieser Situation eine Jacobi-Rotation zur Annihilierung des Pivotelementes des schiefhermiteschen Anteils S durchgeführt, so gilt anschließend

$$\tau^2(H') \leqq \tau^2(H) + O(\varepsilon^4) \ .$$

Unser Beweis beruht dabei auf einer Beweisskizze von VOEVODIN [16] für normale Matrizen.

Auf diesem Ergebnis aufbauend geben wir in Abschnitt 2.4 eine Modifikation des Verfahrens von Eberlein an, die im wesentlichen in der Einführung einer Schwellenstrategie bei Beibehaltung der zyklischen Pivotwahl besteht. Für diese modifizierte Version zeigen wir in Abschnitt 2.5, daß sie jede gegebene Matrix mit beliebiger Genauigkeit diagonalisiert.

In Kapitel 3 beschäftigen wir uns mit einem neuen Jacobi-ähnlichen Verfahren zur Diagonalisierung beliebiger komplexer Matrizen. Dieses Verfahren beruht auf einem von RUTISHAUSER [14] 1964 vorgeschlagenen Algorithmus zur näherungsweisen Normalisierung reeller Matrizen. Rutishauser verwendet hierzu pro Iterationsschritt eine reelle Jacobi-Rotation und eine nicht-orthogonale Ähnlichkeitstransformation, um die euklidische Norm der iterierten Matrix zu reduzieren. Von VESELIĆ stammt nun der Vorschlag, dieses Verfahren in jedem Iterationsschritt mit einer weiteren reellen Jacobi-Rotation zur

gleichzeitigen Diagonalisierung zu verbinden, um so eine gegen Diagonalform konvergente Methode zu erhalten. Diese Methode wurde vom Verfasser im Rahmen einer Diplomarbeit [6] untersucht.

In Abschnitt 3.2 stellen wir nun die Verallgemeinerung dieses Verfahrens für komplexe Matrizen vor. Die normreduzierende Transformation kann dabei weiter mit reellem Parameter erfolgen, der im Gegensatz zu dem Verfahren von Eberlein exakt bestimmt werden kann. Für diese neue Methode beweisen wir dann im Abschnitt 3.3 die asymptotisch quadratische Konvergenz unter zeilenzyklischer Pivotwahl im Falle getrennter Eigenwerte. Die hierzu angestellten theoretischen Untersuchungen führten dabei auch zu Änderungen des zunächst implementierten Algorithmus. So ist z.B. die in Schritt 4) unseres Verfahrens angegebene Wahl der Parameter der zweiten Jacobi-Rotation in Abhängigkeit von denen der ersten wesentlich für die quadratische Konvergenz; die entsprechenden Änderungen des Programmes erbrachten dann auch die zu erwartende praktische Konvergenzverbesserung bei Testrechnungen.

Kapitel 4 enthält schließlich die von uns mit den verschiedenen Verfahren erzielten numerischen Resultate. In Abschnitt 4.1 führen wir zunächst die Ergebnisse auf, die wir bei Anwendung unseres neuen und des Eberlein - Verfahrens auf kleine Testmatrizen ($n \leqq 40$) erhalten haben. Bei diesen mit einer IBM 3031 des Rechenzentrums der Fernuniversität - Gesamthochschule - Hagen durchgeführten Rechnungen zeigte sich unsere neue Methode für gut konditionierte Matrizen vergleichbar mit dem Eberlein - Verfahren, bei defektiven Matrizen war sie dagegen deutlich überlegen.

In Abschnitt 4.2 gehen wir dann kurz auf sogenannte Parameteranalyse - Probleme ein, bei denen eine ganze Reihe sich voneinander nur wenig unterscheidender Matrizen diagonalisiert werden muß, und beschreiben eine von K. VESELIĆ vorgeschlagene

"Naive Methode" zur Diagonalisierung bereits diagonal - dominanter Matrizen. Ausgehend von dieser Naiven Methode beziehungsweise von unserem neuen Verfahren gemäß Abschnitt 3.2 skizzieren wir schließlich noch zwei Jacobi-ähnliche Algorithmen zur praktischen Durchführung von Parameteranalyse - Aufgaben.

Diese beiden Algorithmen haben wir auf dem Vektorrechner CDC CYBER 205 der Ruhruniversität Bochum implementiert und ihre Leistungsfähigkeit mit der des bekannten komplexen QR - Verfahrens (vgl. [10]) verglichen. Die entsprechenden numerischen Resultate enthält Abschnitt 4.3 . Es stellte sich heraus, daß bei Anwendung auf Parameteranalyse - Aufgaben die Naive Methode nur um etwa den Faktor 1.8 langsamer war als das bekanntermaßen schnelle QR - Verfahren ohne Eigenwertberechnung; sollen die Eigenvektoren mitberechnet werden, so ist die Naive Methode sogar schneller.

An dieser Stelle möchte ich mich bei allen bedanken, die mich bei der vorliegenden Arbeit durch kritische, hilfreiche Diskussionen unterstützt haben. Mein besonderer Dank gilt Herrn Prof. Dr. K. Veselić, der mich zu diesen Untersuchungen angeregt und ihren Fortgang durch vielfältige Anregungen gefördert hat.

## 1.2 BEZEICHNUNGEN, GRUNDLEGENDE DEFINITIONEN UND SÄTZE

Im folgenden bezeichnen Großbuchstaben der Form A, H, S etc. quadratische komplexe Matrizen der Ordnung n. Die Elemente einer Matrix A seien bezeichnet mit $a_{ij}$, also

$$A = (a_{ij})_{1 \leq i,j \leq n} ,$$

und im Falle einer indizierten Matrix $A_\nu$ sei

$$A_\nu = (a_{ij}^{(\nu)})_{1 \leq i,j \leq n} .$$

$A^*$ sei die zu A konjugiert transponierte Matrix, I sei die Einheits- und $\Theta$ die Nullmatrix der Ordnung n.

Wir geben nun einige Definitionen sowie bekannte Sätze und Bemerkungen an, deren Beweise z.B. in [19] zu finden oder leicht nachzurechnen sind.

<u>DEFINITION 1.2.1:</u> *Der <u>Kommutator C(A)</u> einer Matrix A ist definiert durch*

$$C(A) := A^*A - AA^* .$$

*Eine Matrix A heißt <u>normal</u>, wenn gilt*

$$C(A) = \Theta .$$

<u>DEFINITION 1.2.2:</u> *Die <u>euklidische Norm</u>* $\|\cdot\|$ *einer Matrix A ist definiert durch*

$$\|A\|^2 := \sum_{i=1}^{n} |a_{ij}|^2 .$$

BEMERKUNG 1.2.3: *Die euklidische Norm einer Matrix* A *ist invariant unter unitären Ähnlichkeitstransformationen; d.h. für jede nicht-singuläre Matrix* U *mit*

$$U^* U = I$$

*gilt*

$$\|U^* A\, U\| = \|A\| \quad .$$

SATZ 1.2.4 (SCHUR) : *Die Matrix* A *habe die Eigenwerte* $\lambda_i \in \mathbb{C}$, $i=1,2,\ldots,n$. *Dann gilt*

$$\sum_{i=1}^{n} |\lambda_i|^2 \leq \|A\|^2 \quad .$$

*Gleichheit gilt genau dann, wenn* A *normal ist.*

DEFINITION 1.2.5: *Eine unitäre Ähnlichkeitstransformation*

$$\tilde{A} := U^* A\, U$$

*heißt* *(komplexe) Jacobi-Rotation der Matrix* A *, wenn für die Elemente der Transformationsmatrix* U *gilt*

$$u_{pp} = u_{qq} = \cos\Phi \ ,$$

$$u_{pq} = -\overline{u_{qp}} = -e^{i\alpha}\sin\Phi \ ,$$

$$u_{ij} = \delta_{ij} \quad \text{für} \quad (i,j) \notin \{(p,p),(p,q),(q,p),(q,q)\}.$$

*Die Parameter* $\alpha, \Phi \in \mathbb{R}$ *heißen die* *Rotationsparameter* *oder* *Rotationswinkel*, *das Paar* (p,q) *heißt das* *Pivotindexpaar* *der Jacobi-Rotation.*

BEMERKUNG 1.2.6 : *Es sei*

$$\tilde{A} := U^* A\, U$$

*eine Jacobi-Rotation der Matrix* A *mit Rotationsparametern* $\alpha, \phi$ *und Pivotindexpaar* $(p,q)$.
*Dann gilt für die Elemente der Matrix* $\tilde{A}$,

$$\tilde{A} = (\tilde{a}_{ij})_{1 \leq i,j \leq n} :$$

$$\tilde{a}_{pp} = \cos^2\phi\, a_{pp} + \cos\phi \sin\phi\, (e^{i\alpha} a_{qp} + e^{-i\alpha} a_{pq}) + \sin^2\phi\, a_{qq} ,$$

$$\tilde{a}_{qq} = \cos^2\phi\, a_{qq} - \cos\phi \sin\phi\, (e^{i\alpha} a_{qp} + e^{-i\alpha} a_{pq}) + \sin^2\phi\, a_{pp} ,$$

$$\tilde{a}_{pq} = e^{i\alpha} \{ \cos^2\phi\, e^{-i\alpha} a_{pq} + \cos\phi \sin\phi\, (a_{qq} - a_{pp}) - \sin^2\phi\, e^{i\alpha} a_{qp} \} ,$$

$$\tilde{a}_{qp} = e^{-i\alpha} \{ \cos^2\phi\, e^{i\alpha} a_{qp} + \cos\phi \sin\phi\, (a_{qq} - a_{pp}) - \sin^2\phi\, e^{-i\alpha} a_{pq} \} ,$$

$$\left.\begin{aligned} \tilde{a}_{pi} &= \cos\phi\, a_{pi} - e^{i\alpha} \sin\phi\, a_{qi} , \\ \tilde{a}_{qi} &= \cos\phi\, a_{qi} + e^{-i\alpha} \sin\phi\, a_{pi} , \\ \tilde{a}_{ip} &= \cos\phi\, a_{ip} - e^{-i\alpha} \sin\phi\, a_{iq} , \\ \tilde{a}_{iq} &= \cos\phi\, a_{iq} + e^{i\alpha} \sin\phi\, a_{ip} , \end{aligned}\right\} \quad i \neq p,q ,$$

$$\tilde{a}_{ij} = a_{ij} , \qquad i \neq p,q , \; j \neq p,q .$$

DEFINITION 1.2.7 : *Die* <u>*Betragsgröße* $\tau$</u> *der Außerdiagonalelemente einer Matrix* A *ist definiert durch*

$$\tau^2(A) := \sum_{i \neq j} |a_{ij}|^2 .$$

BEMERKUNG 1.2.8: *Sei*

$$\tilde{A} := U^* A\, U$$

*eine Jacobi-Rotation der Matrix* A *mit Pivotindexpaar* (p,q). *Dann gilt*

$$\tau^2(\tilde{A}) - |\tilde{a}_{pq}|^2 - |\tilde{a}_{qp}|^2 = \tau^2(A) - |a_{pq}|^2 - |a_{qp}|^2 \quad .$$

DEFINITION 1.2.9: *Ein Verfahren zur Diagonalisierung einer komplexen Matrix* A *heißt* *Jacobi-ähnlich*, *wenn es ausgehend von* $A = A_0$ *eine Folge* $A_k$ *konstruiert gemäß*

$$A_{k+1} := T_k^{-1} A_k T_k \quad , \quad k \in \mathbb{N}_0 \ , \tag{1.2.1}$$

*und dabei für die Elemente der Transformationsmatrix* $T_k$ *gilt*

$$t_{ij}^{(k)} = \delta_{ij} \quad , (i,j) \notin \{(p,p),(p,q),(q,p),(q,q)\}, \quad 1 \leq p,q \leq n \ . \tag{1.2.2}$$

*Die Transformation* (1.2.1) *nennt man dann einen* *Iterationsschritt* *des Verfahrens mit Pivotindexpaar* (p,q).
*Werden die Pivotindexpaare* (p,q) *in der Reihenfolge*

$$\begin{array}{ccccccccc}
(1,2)\ , & (1,3)\ , & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & , (1,n)\ , \\
 & (2,3)\ , & (2,4)\ , & \cdot & \cdot & \cdot & \cdot & \cdot & , (2,n)\ , \\
 & & \cdot & & & & & & \cdot \\
 & & & \cdot & & & & & \cdot \\
 & & & & \cdot & & & & \cdot \\
 & & & & & \cdot & & & \cdot \\
 & & & & & & \cdot & & \cdot \\
 & & & & & & & \cdot & \cdot \\
 & & & & & & & & (n-1,n)
\end{array} \tag{1.2.3}$$

*gewählt, so heißt diese Pivotwahl* (*zeilen-*)*zyklisch*, *und ein voller Pivotdurchlauf gemäß* (1.2.3) *heißt ein* *Iterationszyklus* *des Verfahrens.*

# 2. ÜBER DAS VERFAHREN VON EBERLEIN

## 2.1 EINFÜHRUNG

In [4] stellte Eberlein ein Jacobi-ähnliches Verfahren zur Diagonalisierung einer beliebigen komplexen Matrix A vor. Es erzeugt Folgen $U_k$ und $T_k$ zweidimensionaler Transformationsmatrizen, so daß mit

$$A_0 := A \quad ,$$

$$A_{k+1} := T_k^{-1} U_k^* A_k U_k T_k \quad , \quad k \in \mathbb{N}_0 \quad ,$$

die Folge der $A_k$ für $k \to \infty$ gegen eine Diagonalmatrix konvergiert. Dabei stellt die Transformation

$$A_k' := U_k^* A_k U_k$$

eine Jacobi-Rotation gemäß Definition 1.2.5 dar, wogegen die Transformation

$$A_{k+1} := T_k^{-1} A_k' T_k$$

der Normreduzierung dient. Die nicht-trivialen Elemente der Transformationsmatrix $T_k$ sind dabei für ein festes Pivotindexpaar (p,q) gegeben durch

(2.1.1) $$t_{pp}^{(k)} = t_{qq}^{(k)} = \cosh y \quad ,$$

(2.1.2) $$t_{pq}^{(k)} = \overline{t_{qp}^{(k)}} = - i\, e^{i\beta} \sinh y \; .$$

Im folgenden Abschnitt 2.2 werden wir eine leichte Modifikation dieses Verfahrens angeben, die in der Vertauschung der Reihenfolge von unitärer und nicht-unitärer Transformation besteht. Dies geschieht aus dem Grund, daß für diese Modifikation Ruhe [13] asymptotisch quadratische Konvergenz im Falle getrennter Eigenwerte bewiesen hat, wogegen das "originale" Eberlein-Verfahren aus [4] i.a. nicht quadratisch konvergent ist. Unter dem "Verfahren von Eberlein" werden wir im folgenden immer diese modifizierte Version verstehen.

## 2.2 DER VERFAHRENS - ALGORITHMUS

<u>SCHRITT 0 :</u> Die Pivotindexpaare (p,q) werden zyklisch gewählt. Für jeden Iterationsschritt führe man dann die folgenden Schritte 1 bis 4 aus, wobei A die jeweils aktuelle Matrix bezeichne.

<u>SCHRITT 1 :</u> Man berechne das Pivotelement $c_{pq}$ des Kommutators $C := C(A)$ und berechne

(2.2.1) $$e^{i\beta} := -i\frac{c_{pq}}{|c_{pq}|}$$

sowie

(2.2.2) $$G := \sum_{i \neq p,q} \{ |a_{pi}|^2 + |a_{ip}|^2 + |a_{qi}|^2 + |a_{iq}|^2 \} ,$$

(2.2.3) $$\xi := |a_{pq} e^{-i\beta} + a_{qp} e^{i\beta}| ,$$

(2.2.4) $$\eta := |a_{pp} - a_{qq}| .$$

<u>SCHRITT 2 :</u> Man führe die Transformation

(2.2.5) $$A' := T^{-1} A T$$

durch, wobei die nicht-trivialen Elemente von T durch (2.1.1), (2.1.2) gegeben sind; dabei ist der Parameter $\beta$ durch (2.2.1) und y durch

(2.2.6) $$\tanh y := \frac{|c_{pq}|}{G + 2(\xi^2 + \eta^2)}$$

definiert.

<u>SCHRITT 3 :</u> Man berechne

(2.2.7) $$te := |a'_{pq} + \overline{a'_{qp}}| + Re^2(a'_{pp} - a'_{qq}) ,$$

(2.2.8) $$tee := |a'_{pq} - \overline{a'_{qp}}| + Im^2(a'_{pp} - a'_{qq}) .$$

Man führe die komplexe Jacobi-Rotation

$$A'' := U^* A' U$$

durch, wobei die Rotationsparameter $\alpha$ und $\phi$ wie folgt bestimmt werden.

(2.2.9) i) Ist <u>$te \geqq tee$</u>, so setze man

(2.2.10) $$\operatorname{ctg} 2\phi := \frac{Re(a'_{pp} - a'_{qq})}{|a'_{pq} + \overline{a'_{qp}}|} , \quad |\phi| \leqq \frac{\pi}{4} ,$$

(2.2.11) $$e^{i\alpha} := \frac{a'_{pq} + \overline{a'_{qp}}}{|a'_{pq} + \overline{a'_{qp}}|} , \quad \cos\alpha \, Re(a'_{pq} + \overline{a'_{qp}}) \geqq 0.$$

(2.2.12) ii) Ist <u>$te < tee$</u>, so setze man

(2.2.13) $$\operatorname{ctg} 2\phi := \frac{Im(a'_{pp} - a'_{qq})}{|a'_{pq} - \overline{a'_{qp}}|} , \quad |\phi| \leqq \frac{\pi}{4} ,$$

(2.2.14) $$e^{i\alpha} := i \frac{a'_{pq} - \overline{a'_{qp}}}{|a'_{pq} - \overline{a'_{qp}}|} , \quad \cos\alpha \, Im(a'_{pq} - \overline{a'_{qp}}) \geqq 0.$$

<u>SCHRITT 4:</u> Falls $\tau^2(A)$ hinreichend klein ist, so beende man die Iteration.
Anderenfalls setze man

$$A := A''$$

und fahre fort mit Schritt 1 .

<u>ANMERKUNG:</u>: Ist die Abfrage (2.2.9) erfüllt und werden daher die Rotationsparameter $\alpha$ und $\phi$ gemäß (2.2.10) und (2.2.11) bestimmt, so gilt

$$a''_{pq} + \overline{a''_{qp}} = 0 \; ; \tag{2.2.15}$$

wir sprechen in diesem Fall im folgenden auch von einem "<u>Iterationsschritt bezüglich des hermiteschen Anteils</u>" von A.
Ist (2.2.9) dagegen nicht erfüllt und werden $\alpha$ und $\phi$ somit gemäß (2.2.13) und (2.2.14) gewählt, so gilt

$$a''_{pq} - \overline{a''_{qp}} = 0 \; , \tag{2.2.16}$$

und wir sprechen im folgenden entsprechend von einem "<u>Iterationsschritt bezüglich des schiefhermiteschen Anteils</u>" von A.

## 2.3 BETRACHTUNG BEI "FAST - NORMALEN" MATRIZEN

Zu Beginn dieses Abschnittes möchten wir noch einmal darauf eingehen, worin das Problem für einen Beweis der globalen Konvergenz des Verfahrens von Eberlein besteht. Wie bereits in der Einleitung kurz gesagt, wird bei einem Iterationsschritt bezüglich des hermiteschen Anteils H der iterierten Matrix A zwar $\tau^2(H)$ verringert, gleichzeitig kann dabei jedoch für den schiefhermiteschen Anteil S von A $\tau^2(S)$ anwachsen (und umgekehrt). Es besteht also die Gefahr des sogenannten "Ping-Pong-Effektes", so daß wegen $\tau^2(A) = \tau^2(H) + \tau^2(S)$ unmittelbar keine Aussage darüber möglich ist, inwieweit sich bei einem solchen Iterationsschritt $\tau^2(A)$ verändert oder gar verringert. Eine Abschätzung für das Verhalten von $\tau^2(A)$ zu gewinnen, ist umso interessanter, als es bekanntermaßen normale, aber nicht-diagonale Matrizen A gibt, für die man mit keiner Jacobi-Rotation $\tau^2(A)$ verkleinern kann (vgl. VOEVODIN [14]).

In diesem Abschnitt gehen wir nun aus von einer Matrix A, für die die Norm ihres Kommutators hinreichend klein und deren hermitescher Anteil H bereits bis auf $O(\varepsilon)$ diagonalisiert ist. Führt man in dieser Situation nun einen Iterationsschritt bezüglich S mit "nicht zu kleinem" Pivotelement $s_{pq}$ durch, so werden wir in Satz 2.3.3 dieses Abschnittes zeigen, daß dann $\tau^2(H)$ höchstens um $O(\varepsilon^4)$ anwachsen kann. (Dieses Resultat werden wir in Abschnitt 2.5 dazu verwenden, für eine gewisse Modifikation des Eberlein-Verfahrens nachzuweisen, daß sie jede gegebene Matrix mit beliebiger Genauigkeit diagonalisiert.)

Sei also A eine komplexe n x n - Matrix, für die wir o.B.d.A. voraussetzen

(2.3.1) $$\|A\| \leqq 1 ,$$

und sei A zerlegt in ihren hermiteschen Anteil H und in ihren schiefhermiteschen Anteil S gemäß

(2.3.2) $$A = H + iS ,$$

(2.3.3) $$H := \frac{1}{2} (A + A^*) ,$$

(2.3.4) $$S := \frac{1}{2i} (A - A^*) .$$

Dann sind H und S hermitesche Matrizen mit

(2.3.5) $$\|H\| \leqq 1, \quad \|S\| \leqq 1 ,$$

und es gilt

(2.3.6) $$HS - SH = \frac{1}{2i} C(A) .$$

Als erste Vorbereitung für den Satz 2.3.3 beweisen wir zunächst die folgende Bemerkung.

<u>BEMERKUNG 2.3.1:</u> *Gegeben sei eine Matrix* A *mit*

$$\|C(A)\|^2 \leqq \varepsilon^2 \leqq 1$$

*und*

$$\tau^2(H) \leqq \varepsilon^2 .$$

*Seien ferner die Eigenwerte von* H *bezeichnet mit* $\tilde{h}_{11}, \ldots, \tilde{h}_{nn}$ . *Dann gilt für alle* i,j *mit* $1 \leqq i,j \leqq n$:

$$|s_{ij}(\tilde{h}_{ii}-\tilde{h}_{jj})| < 3.33\,\varepsilon .$$

<u>BEWEIS :</u> Aus $\tau^2(H) \leq \varepsilon^2$ folgt zunächst nach einem Satz von Wielandt-Hoffmann [18]

$$(2.3.7) \qquad \sum_{i=1}^{n} |\tilde{h}_{ii} - h_{ii}|^2 \leq \varepsilon^2 \ .$$

Definieren wir noch

$$(2.3.8) \qquad \tilde{H} := \operatorname{diag}(\tilde{h}_{11}, \ldots , \tilde{h}_{nn}) \ ,$$

so kann H zerlegt werden gemäß

$$(2.3.9) \qquad H = \tilde{H} + E \ ,$$

wobei für die Matrix E gilt

$$(2.3.10) \qquad \begin{aligned} \|E\|^2 &= \tau^2(E) + \sum_{i=1}^{n} |h_{ii} - \tilde{h}_{ii}|^2 \\ &= \tau^2(H) + \sum_{i=1}^{n} |h_{ii} - \tilde{h}_{ii}|^2 \\ &\leq 2\varepsilon^2 \ . \end{aligned}$$

Nun gilt wegen (2.3.6) und (2.3.9)

$$(2.3.11) \qquad \tilde{H}S - S\tilde{H} = SE - ES + \frac{1}{2i} C(A) \ ,$$

und wegen

$$(\tilde{H}S - S\tilde{H})_{ij} = s_{ij} (\tilde{h}_{ii} - \tilde{h}_{jj})$$

erhalten wir mit (2.3.10) und $\|S\| \leq 1$

$$
\begin{aligned}
|s_{ij}(\tilde{h}_{ii} - \tilde{h}_{jj})| &\leq \|\tilde{H}S - S\tilde{H}\| \\
&\leq \|SE\| + \|ES\| + \frac{1}{2}\|C(A)\| \\
&\leq 2\sqrt{2}\,\varepsilon + \frac{1}{2}\varepsilon \\
&< 3.33\,\varepsilon\,.
\end{aligned}
$$

■

Wir führen nun noch eine Bezeichnung für den minimalen Abstand verschiedener Eigenwerte der Matrix $H$ ein. Es sei

$$\rho_H := \min_{\tilde{h}_{ii} \neq \tilde{h}_{jj}} |\tilde{h}_{ii} - \tilde{h}_{jj}| \,. \tag{2.3.12}$$

Damit erhalten wir aus der obigen Bemerkung sofort die

<u>FOLGERUNG 2.3.2:</u> *Es seien die Voraussetzungen der Bemerkung 2.3.1 erfüllt. Zusätzlich sei* $\tau^2(H) \neq 0$. *Dann gilt*

$$|s_{ij}| \leq 3.33\,\frac{\varepsilon}{\rho_H} \,.$$

<u>BEWEIS :</u> Die Behauptung ergibt sich sofort aus der Bemerkung 2.3.1, da wegen $\tau^2(H) \neq 0$ auch $\rho_H \neq 0$ ist. ■

Von nun an betrachten wir die folgende Situation. Gegeben sei eine Matrix A mit

(2.3.13) $$n \geqq 3 \ ,$$

(2.3.14) $$0 \neq \tau^2(H) \leqq \varepsilon^2 \ ,$$

(2.3.15) $$\tau^2(S) > \frac{1}{4}\varepsilon^2 \ ,$$

(2.3.16) $$\|C(A)\|^2 \leqq \varepsilon^4 \ ,$$

wobei für $\varepsilon$ gelte

(2.3.17) $$\varepsilon \leqq \frac{1}{64\sqrt{2}\,n^3}\,\rho_H^3 \ .$$

Damit können wir nun den folgenden Satz formulieren und beweisen.

<u>SATZ 2.3.3 :</u> *Die Matrix* A *sei gemäß (2.3.3) und (2.3.4) zerlegt in* $A = H + iS$, *und es seien die Voraussetzungen (2.3.13) bis (2.3.17) erfüllt.* H' *sei die Matrix* H *nach Ausführung einer Jacobi-Rotation zur Annihilierung des Pivotelementes* $s_{pq}$ *von* S *gemäß Schritt 3 des Eberlein-Verfahrens, und es gelte*

(2.3.18) $$|s_{pq}|^2 \geqq \frac{1}{n^2}\,\tau^2(S) \ .$$

*Dann folgt*

$$\tau^2(H') \leqq \tau^2(H) + 2\,\frac{\varepsilon^4}{\rho_H^2} \ .$$

<u>BEWEIS :</u> Zunächst einmal wissen wir aus der Folgerung 2.3.2, daß die Matrix S bis auf Permutation und Größenordnung $\frac{\varepsilon}{\rho_H}$ <u>blockdiagonal</u> ist; die Dimensionen der Diagonalblöcke sind dabei gleich den Vielfachheiten der Eigenwerte $\tilde{h}_{11}, \ldots, \tilde{h}_{nn}$ von H. Wir treffen die folgende Fallunterscheidung.

<u>FALL I :</u> Das Pivotelement $s_{pq}$ liegt <u>außerhalb</u> eines der Diagonalblöcke von S.

<u>FALL II :</u> $s_{pq}$ liegt <u>innerhalb</u> eines solchen Blockes.

<u>Zu Fall I :</u> In diesem Fall ist $\tilde{h}_{pp} \neq \tilde{h}_{qq}$. Aufgrund der Folgerung 2.3.2 und der Voraussetzung (2.3.18) gilt

$$\begin{aligned} \max_{i \neq j} \{|s_{ij}|^2\} &\leq \tau^2(S) \\ &\leq n^2 |s_{pq}|^2 \\ &\leq n^2 \left(\frac{3.33}{\rho_H}\right)^2 . \end{aligned} \tag{2.3.19}$$

Aus (2.3.6) erhalten wir

$$\begin{aligned} \frac{1}{2i} (C(A))_{pq} &= (HS - SH)_{pq} \\ &= \sum_{k \neq p,q} (h_{pk}s_{kq} - s_{pk}h_{kq}) - h_{pq}(s_{pp} - s_{qq}) + s_{pq}(h_{pp} - h_{qq}) . \end{aligned} \tag{2.3.20}$$

Wegen der Voraussetzung (2.3.16) folgt hieraus mit Hilfe der Cauchy - Schwarz'schen Ungleichung

$$|h_{pq}(s_{pp}-s_{qq}) - s_{pq}(h_{pp}-h_{qq})|^2 \leq 2\left(\left|\sum_{k\neq p,q}(h_{pk}s_{kq}-s_{pk}h_{kq})\right|^2 + \frac{\varepsilon^4}{4}\right)$$

$$\leq 4\left(\sum_{k\neq p,q}|h_{pk}|^2\sum_{k\neq p,q}|s_{kq}|^2 + \sum_{k\neq p,q}|s_{pk}|^2\sum_{k\neq p,q}|h_{kq}|^2\right) + \frac{\varepsilon^4}{2}$$

$$\leq 4\,\tau^2(S)\sum_{k\neq p,q}(|h_{pk}|^2 + |h_{kq}|^2) + \frac{\varepsilon^4}{2}\,. \tag{2.3.21}$$

Mit (2.3.19) und (2.3.14) sowie mit $\rho_H \leq \sqrt{2}$ erhalten wir

$$|h_{pq}(s_{pp}-s_{qq}) - s_{pq}(h_{pp}-h_{qq})| \leq \left(4n^2\left(\frac{3.33}{\rho_H}\right)^2\varepsilon^2 + \frac{\varepsilon^4}{2}\right)^{1/2}$$

$$\leq \left(4\left(\frac{3.33}{\rho_H}n\right)^2 + \frac{1}{2}\,\frac{n^2}{9}\,\frac{2}{\rho_H^2}\right)^{1/2}\varepsilon^2$$

$$< \frac{6.7\,n}{\rho_H}\,\varepsilon^2\,. \tag{2.3.22}$$

Wichtig ist nun, daß diese Abschätzung (2.3.22) auch für die entsprechenden Elemente der Matrizen H' und S', also nach Durchführung der Jacobi-Rotation, erfüllt ist. Denn zum einen gilt

$$\tau^2(S') \leq \tau^2(S)\,,$$

und zum anderen erhalten wir aus den Transformationsformeln

in Bemerkung 1.2.6

$$|h'_{pk}|^2 + |h'_{kq}|^2 = |h_{pk}|^2 + |h_{kq}|^2 \quad , \quad k \neq p,q.$$

Damit kann die Abschätzung (2.3.22) ganz analog ausgeführt werden, und wegen $s'_{pq} = 0$ erhalten wir

$$(2.3.23) \qquad |h'_{pq}(s'_{pp}-s'_{qq})| < \frac{6.7\,n}{\rho_H}\,\varepsilon^2 \quad .$$

Als nächstes wollen wir die Größenordnungen der einzelnen Terme auf der linken Seite von (2.3.22) bestimmen, um so zu einer unteren Schranke für das Element $h_{pq}$ zu gelangen. Zunächst gilt mit (2.3.7) und (2.3.17)

$$\begin{aligned} |h_{pp}-h_{qq}| &\geqq |\tilde{h}_{pp}-\tilde{h}_{qq}| - |h_{pp}-\tilde{h}_{pp}| - |h_{qq}-\tilde{h}_{qq}| \\ &\geqq \rho_H - \\ &\geqq \rho_H - \sqrt{2}\,\frac{1}{64\sqrt{2}\,27}\,2\rho_H \\ &> 0.9988\,\rho_H \quad . \end{aligned} \tag{2.3.24}$$

Weiter ist mit (2.3.15) und (2.3.18)

$$|s_{pq}| \geqq \left(\frac{\tau^2(S)}{n^2}\right)^{1/2} \geqq \frac{1}{2n}\,\varepsilon \quad ,$$

und wir erhalten

$$(2.3.25) \qquad |s_{pq}(h_{pp}-h_{qq})| > 0.4994\,\frac{\rho_H}{n}\,\varepsilon \quad .$$

Damit folgt aus (2.3.22) und (2.3.17)

$$
\begin{aligned}
|h_{pq}(s_{pp} - s_{qq})| &> 0.4994 \frac{\rho_H}{n} \varepsilon - \frac{6.7n}{\rho_H} \varepsilon^2 \\
&\geqq 0.4994 \frac{\rho_H}{n} \varepsilon - \frac{6.7n}{\rho_H} \varepsilon \frac{\rho_H^3}{64\sqrt{2}\, n^3} \\
&= \varepsilon \frac{\rho_H}{n} \left(0.4994 - \frac{6.7\rho_H}{64\sqrt{2}n}\right) \\
&> \varepsilon \frac{\rho_H}{n} \left(0.4994 - \frac{6.7\sqrt{2}}{64\sqrt{2}3}\right) \\
&> 0.4645 \frac{\rho_H}{n} \varepsilon \,. \qquad (2.3.26)
\end{aligned}
$$

Nun ist wegen $\tau^2(H) \leqq \varepsilon^2$

$$|h_{pq}| \leqq \varepsilon \,,$$

und wegen $\|S\| \leqq 1$ ist

$$|s_{pp} - s_{qq}| \leqq \sqrt{2} \,.$$

Damit erhalten wir aus (2.3.26) die beiden Abschätzungen

$$|s_{pp} - s_{qq}| > 0.4645 \frac{\rho_H}{n} \qquad (2.3.27)$$

und

$$
\begin{aligned}
|h_{pq}| &> 0.4645 \frac{\rho_H}{\sqrt{2}\, n} \varepsilon \qquad (2.3.28) \\
&> 0.32 \frac{\rho_H}{n} \varepsilon \,.
\end{aligned}
$$

Unser Ziel ist es nun, aus (2.3.23) eine obere Schranke für $|h'_{pq}|$ zu erhalten und diese mit (2.3.28) zu vergleichen.

Gemäß Bemerkung 1.2.6 ist zunächst

$$|s'_{pp} - s'_{qq}| \geq |\cos 2\phi\,(s_{pp} - s_{qq})| - 2|\sin 2\phi\, s_{pq}|. \tag{2.3.29}$$

Der Rotationswinkel $\phi$ ist nach (2.2.13) definiert durch

$$\operatorname{ctg} 2\phi = \frac{s_{pp}-s_{qq}}{2|s_{pq}|} ,$$

und mit (2.3.27), Folgerung 2.3.2 und (2.3.17) erhalten wir

$$\begin{aligned} |\operatorname{ctg} 2\phi| &> \frac{0.4645\,\rho_H^2}{2\cdot 3.33\,\varepsilon\cdot n} \\ &\geq \frac{0.4645\,\rho_H^2}{6.66\,n}\,\frac{64\sqrt{2}\,n^3}{\rho_H^3} \\ &\geq \frac{0.4645\cdot 64\sqrt{2}\cdot 9}{6.66\sqrt{2}} \\ &> 40 . \end{aligned} \tag{2.3.30}$$

Also gilt

$$\sin^2 2\phi < \left(\frac{1}{40}\right)^2$$

und

$$\cos^2 2\phi > 1 - \left(\frac{1}{40}\right)^2 ,$$

und wir erhalten aus (2.3.29) mit (2.3.27) und der Folgerung 2.3.2 die Abschätzung

$$|s'_{pp} - s'_{qq}| > |\cos 2\phi|\, 0.4645 \frac{\rho_H}{n} - 2|\sin 2\phi| \frac{3.33}{\rho_H} \varepsilon$$

$$> \frac{\rho_H}{n} \left[ 0.4645 |\cos 2\phi| - \frac{6.66 |\sin 2\phi|}{64\sqrt{2}\, n^2} \rho_H \right]$$

$$> \frac{\rho_H}{n} \left[ 0.4645 \sqrt{1 - \left(\frac{1}{40}\right)^2} - \frac{6.66\sqrt{2}}{64\sqrt{2}\, 9 \cdot 40} \right]$$

(2.3.31) $$> 0.464 \frac{\rho_H}{n} .$$

Somit folgt aus (2.3.23) mit (2.3.17)

$$|h'_{pq}| < \frac{6.7n}{\rho_H} \varepsilon^2 \frac{n}{0.464\rho_H}$$

$$\leq \frac{6.7\, n^2}{0.464\, \rho_H^2} \quad \frac{\rho_H^3}{64\sqrt{2}\, n^3} \quad \varepsilon$$

(2.3.32) $$< 0.16 \frac{\rho_H}{n} \varepsilon .$$

Damit erhalten wir mit (2.3.28)

$$|h'_{pq}| < 0.16 \frac{\rho_H}{n} \varepsilon$$

$$< 0.32 \frac{\rho_H}{n} \varepsilon$$

$$< |h_{pq}| .$$

Aus Bemerkung 1.2.8 folgt hieraus schließlich

$$\tau^2(H') \leq \tau^2(H)$$

und damit die Behauptung.

<u>Zu Fall II :</u> In diesem Fall liegt das Pivotelement $s_{pq}$ innerhalb eines der Diagonalblöcke der Matrix S. Somit gilt zunächst

$$\tilde{h}_{pp} = \tilde{h}_{qq} \quad . \tag{2.3.33}$$

Aus den Transformationsformeln in Bemerkung 1.2.6 erhalten wir

$$2|h'_{pq}|^2 = 2|h_{pq}|^2 + R \quad , \tag{2.3.34}$$

wobei $R \in \mathbb{R}$ definiert ist durch

$$\begin{aligned} R := & -2\sin^2\phi\cos^2\phi\{4\mathrm{Re}^2(e^{-i\alpha}h_{pq}) - (h_{pp}-h_{qq})^2\} \\ & + \frac{1}{2}\sin 2\phi\cos 2\phi\,(h_{pp}-h_{qq})2\mathrm{Re}(e^{-i\alpha}h_{pq}) \quad . \end{aligned} \tag{2.3.35}$$

Wegen

$$2\sin^2\phi\cos^2\phi = \frac{1}{2}\sin^2 2\phi = \frac{1}{2}(1-\cos^2 2\phi)$$

folgt hieraus

$$R \leq -\frac{1}{2}\sin^2 2\phi\, 4\,\mathrm{Re}^2(e^{-i\alpha}h_{pq}) + \frac{1}{2}(1-\cos^2 2\phi)|h_{pp}-h_{qq}|^2$$

$$+ 2|\sin 2\phi||\cos 2\phi||\mathrm{Re}(e^{-i\alpha}h_{pq})||h_{pp}-h_{qq}|$$

$$= -\frac{1}{2}\{4\sin^2 2\phi\ \mathrm{Re}^2(e^{-i\alpha}h_{pq}) + \cos^2 2\phi|h_{pp}-h_{qq}|^2$$

$$- 4|\sin 2\phi||\cos 2\phi||\mathrm{Re}(e^{-i\alpha}h_{pq})||h_{pp}-h_{qq}|\}$$

$$+ \frac{1}{2}|h_{pp}-h_{qq}|^2$$

$$= -\frac{1}{2}\left[2|\sin 2\phi||\mathrm{Re}(e^{-i\alpha}h_{pq})| - |\cos 2\phi||h_{pp}-h_{qq}|\right]^2$$

$$+ \frac{1}{2}|h_{pp}-h_{qq}|^2$$

$$\leq \frac{1}{2}|h_{pp}-h_{qq}|^2 \quad .$$

Nun gilt wegen (2.3.33) (vgl. [17])

$$|h_{pp}-h_{qq}| \leq |h_{pp}-\tilde{h}_{pp}| + |\tilde{h}_{pp}-\tilde{h}_{qq}| + |\tilde{h}_{qq}-h_{qq}|$$

$$\leq 2\frac{\varepsilon^2}{\rho_H} \quad ,$$

und wir erhalten aus (2.3.34) und der Bemerkung 1.2.8 schließlich

$$\begin{aligned}
\tau^2(H') &= \tau^2(H) - 2|h_{pq}|^2 + 2|h'_{pq}|^2 \\
&= \tau^2(H) + R \\
&\leqq \tau^2(H) + \frac{1}{2}|h_{pp}-h_{qq}|^2 \\
&\leqq \tau^2(H) + 2\frac{\varepsilon^4}{\rho_H^2} \quad . \qquad \blacksquare
\end{aligned}$$

<u>ANMERKUNG:</u> Vertauscht man in diesem Abschnitt die Rollen von $H$ und $S$, so läßt sich ganz analog ein Satz über den maximalen Zuwachs von $\tau^2(S)$ unter einem Iterationsschritt bezüglich $H$ beweisen.

## 2.4 EINE MODIFIKATION DES VERFAHRENS VON EBERLEIN

Die in diesem Abschnitt von uns vorgestellte Modifikation des Eberlein-Verfahrens besteht im wesentlichen in der Einführung einer Schwellenstrategie, wobei abwechselnd nur bezüglich des hermiteschen oder nur bezüglich des schiefhermiteschen Anteils der Matrix iteriert wird. Der in der folgenden Beschreibung des modifizierten Verfahrens auftretende Begriff der "Diagonalisierung" des hermiteschen oder schiefhermiteschn Matrixanteils meint dabei die Durchführung des üblichen Jacobi-Verahrens für hermitesche Matrizen bei zyklischer Pivotwahl.

Gemäß (2.3.3) und (2.3.4) sei die iterierte Matrix $A$ wieder zerlegt in $A = H + iS$. Der genaue Algorithmus des modifizierten Verfahrens von Eberlein lautet dann wie folgt.

<u>SCHRITT 0:</u> Man wähle eine gewünschte Diagonalisierungsgenauigkeit $\varepsilon \geqq 0$ sowie eine Anfangsschranke $q \geqq \varepsilon$.

<u>SCHRITT 1:</u> Man berechne $\tau^2(H)$, $\tau^2(S)$ und setze

$$F := \begin{cases} H, & \tau^2(H) \geqq \tau^2(S), \\ S, & \text{sonst}, \end{cases} \tag{2.4.1}$$

$$\tilde{F} := \begin{cases} S, & \tau^2(H) \geqq \tau^2(S), \\ H, & \text{sonst}. \end{cases} \tag{2.4.2}$$

<u>SCHRITT 2:</u> Man führe das gewöhnliche Eberlein - Verfahren gemäß Abschnitt 2.2 aus, bis nach $\nu_1$ Zyklen für den Kommutator $C(A_{\nu_1})$ der iterierten Matrix $A_{\nu_1}$ gilt

(2.4.3) $$\|C(A_{\nu_1})\|^2 \leq q^4 \ .$$

Ein Iterationsschritt des Eberlein-Verfahrens mit Pivotindexpaar $(p,q)$ werde dabei genau dann ausgeführt, wenn für die gerade aktuelle Matrix $A$ und ihren Kommutator $C := C(A)$ gilt

(2.4.4) $$4|c_{pq}|^2+(c_{pp}-c_{qq})^2 \geq \frac{2}{n(n-1)} \sum_{i<j} \{4|c_{ij}|^2+(c_{ii}-c_{jj})^2\}.$$

<u>SCHRITT 3:</u> Gilt

(2.4.5) $$\tau^2(F_{\nu_1}) \leq q^2 \ ,$$

so setze man $\nu_2 := 0$ und fahre fort mit Schritt 4; anderenfalls diagonalisiere man die Matrix $F_{\nu_1}$, bis nach $\nu_2$ unitären Jacobi-Zyklen gilt

(2.4.6) $$\tau^2(F_{\nu_1+\nu_2}) \leq q^2 \ .$$

<u>SCHRITT 4:</u> Gilt

(2.4.7) $$\tau^2(\tilde{F}_{\nu_1+\nu_2}) \leq q^2 \ ,$$

so setze man $k:=0$ und fahre fort mit Schritt 5; anderenfalls diagonalisiere man die Matrix $\tilde{F}_{\nu_1+\nu_2}$, bis nach $k$ unitären Jacobi-Schritten gilt

(2.4.8) $$\tau^2(\tilde{F}^{(k)}_{\nu_1+\nu_2}) \leq q^2 \ .$$

Eine Jacobi-Rotation mit Pivotindexpaar (p,q) werde dabei genau dann durchgeführt, wenn für die gerade aktuelle Matrix $\tilde{F}^{(i)}_{\nu_1+\nu_2}$ , $i \leq k$, gilt

(2.4.9) $$|(\tilde{F}^{(i)}_{\nu_1+\nu_2})_{pq}|^2 \geq \frac{\tau^2(\tilde{F}^{(i)}_{\nu_1+\nu_2})}{n^2} \quad .$$

<u>SCHRITT 5:</u> Falls gilt

(2.4.10) $$\tau^2(A^{(k)}_{\nu_1+\nu_2}) > \varepsilon^2 \quad ,$$

so setze man

(2.4.11) $$q := c\,q \,, \qquad 0 < c < \frac{1}{2} \quad ,$$

(2.4.12) $$A := A^{(k)}_{\nu_1+\nu_2}$$

und fahre fort mit Schritt 1;
anderenfalls beende man das Verfahren .

Zur besseren Übersicht über dieses Verfahren stellen wir in Figur 1 die Schritte 1 bis 4 schematisch in einem Programmablaufplan dar.

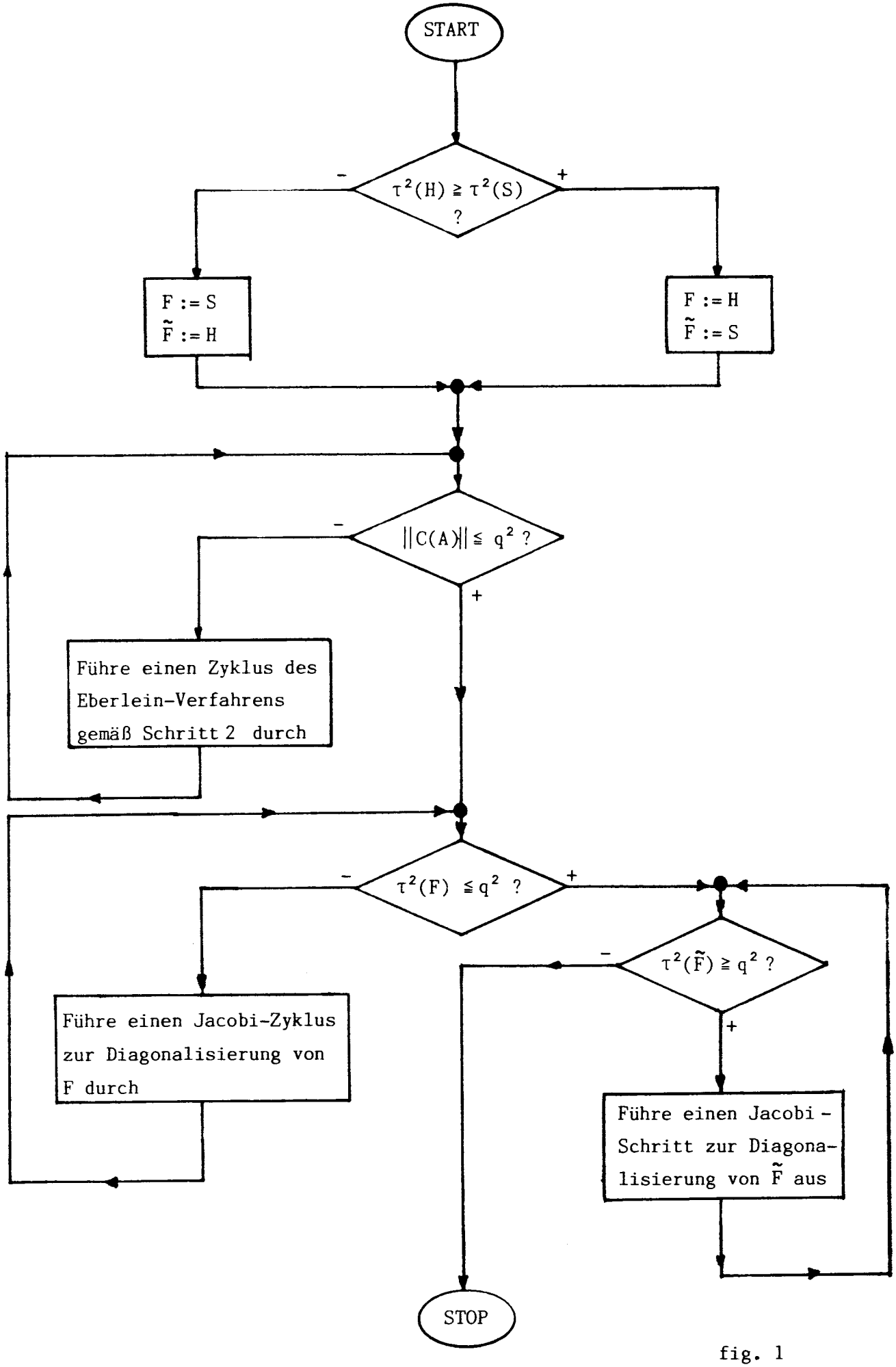
START
$\tau^2(H) \geqq \tau^2(S)$ ?
-
+
$F := S$
$\tilde{F} := H$
$F := H$
$\tilde{F} := S$
$\|C(A)\| \leqq q^2$ ?
-
+
Führe einen Zyklus des Eberlein-Verfahrens gemäß Schritt 2 durch
$\tau^2(F) \leqq q^2$ ?
-
+
$\tau^2(\tilde{F}) \geqq q^2$ ?
-
+
Führe einen Jacobi-Zyklus zur Diagonalisierung von F durch
Führe einen Jacobi-Schritt zur Diagonalisierung von $\tilde{F}$ aus
STOP

fig. 1

## 2.5 BEWEIS DER GLOBALEN KONVERGENZ DER MODIFIKATION DES VERFAHRENS VON EBERLEIN

Für die in Abschnitt 2.4 beschriebene modifizierte Version des Eberlein - Verfahrens werden wir in diesem Abschnitt den folgenden Satz beweisen.

<u>SATZ 2.5.1 :</u> *Sei* $A$ *eine beliebige Matrix mit* $\|A\| \leq 1$, *und sei* $A_\mu$ *die Matrix* $A$ *nach* $\mu$ *Durchführungen der Schritte 1 bis 5 des in Abschnitt 2.4 beschriebenen Verfahrens. Ferner sei* $A$ *gemäß (2.3.2) bis (2.3.4) zerlegt in* $A = H + iS$, *und für* $0 < q_0 \in \mathbb{R}$ *gelte*

$$(2.5.1) \qquad \frac{1}{2} n^2 \ln(1+2Kq_0)\ln\left(\frac{n^2}{(n^2-2)q_0^2}\right)\frac{1}{\ln\left(\frac{n^2}{n^2-2}\right)} \leq \ln 2 ,$$

*wobei*

$$(2.5.2) \qquad K := 2\,\frac{c^2}{32\,n^3} \quad , \; c < \frac{1}{2} \text{ aus Schritt 5.}$$

*Dann existiert zu jeder Schranke* $q$ *mit* $0 < q \leq q_0 < 1$ *und*

$$(2.5.3) \qquad q < \frac{\sqrt{2}}{4} K \rho_H^3$$

*ein Index* $\mu \in \mathbb{N}$ *mit*

$$\tau^2(A_\mu) \leq 3q^2 .$$

Bevor wir den eigentlichen Beweis dieses Satzes führen, stellen wir zunächst zwei kleine Hilfsbemerkungen bereit.

Als erstes überzeugen wir uns, daß ein Index $\nu_1$ existiert, so daß die Abfrage (2.4.3) in Schritt 1 erfüllt ist.

<u>BEMERKUNG 2.5.2 :</u> *Sei* A *eine beliebige Matrix, und sei* $A_j$ *die Matrix* A *nach* j *Iterationszyklen des gewöhnlichen Eberlein - Verfahrens gemäß Schritt 1 des in Abschnitt 2.4 beschriebenen Verfahrens.*

*Dann gilt*

$$\lim_{j\to\infty} \|C(A_j)\| = 0 \ .$$

<u>BEWEIS :</u> Siehe Eberlein [4]. ■

Als nächstes ist zu bestimmen, wieviele Jacobi-Zyklen in Schritt 4 des Verfahrens maximal durchzuführen sind. Aus Gründen der Übersichtlichkeit setzen wir dazu $\tilde{F} := \tilde{F}_{\nu_1+\nu_2}$.

<u>BEMERKUNG 2.5.3 :</u> *Es gelte*

$$q^2 < \tau^2(\tilde{F}) \leq 1 \ ,$$

*und es sei* $\tilde{k} \in \mathbb{N}$ *definiert vermöge*

$$(2.5.4) \qquad \frac{2 \ln q}{\ln(1-\frac{2}{n^2})} \leq \tilde{k} < \frac{2 \ln q}{\ln(1-\frac{2}{n^2})} + 1 \ .$$

*Sei* $\tilde{F}_{\tilde{k}}$ *die Matrix* $\tilde{F}$ *nach* $\tilde{k}$ *Iterationszyklen des Jacobi-Verfahrens gemäß Schritt 4 des in Abschnitt 2.4 beschriebenen Verfahrens. Dann gilt*

$$\tau^2(\tilde{F}_{\tilde{k}}) \leq q^2 \ .$$

<u>BEWEIS :</u> Pro Iterationsschritt existiert mindestens ein Pivotelement $\tilde{f}^{(j)}_{pq}$, das die Bedingung (2.4.9) erfüllt. Damit gilt

$$\tau^2(\tilde{F}_1) \leq \tau^2(\tilde{F}) - 2|\tilde{f}^{(1)}_{pq}|^2$$

$$\leq \tau^2(\tilde{F})\,(1 - \frac{2}{n^2}) ,$$

und mit $\tau^2(\tilde{F}) \leq 1$ und mittels vollständiger Induktion erhalten wir

$$(2.5.5) \qquad \tau^2(\tilde{F}_j) \leq (1 - \frac{2}{n^2})^j \quad , \quad j \in \mathbb{N} .$$

Aus der Voraussetzung (2.5.4) folgt aber

$$\tilde{k} \ln(1 - \frac{2}{n^2}) \leq 2 \ln q$$

und damit mit (2.5.5) die Behauptung. ■

Wir kommen nunmehr zum <u>Beweis des Satzes 2.5.1</u> über globale Konvergenz. .

Sei dazu $q$ , $0 < q \leq q_0$ beliebig. Nach Bemerkung 2.5.2 existiert ein Index $\nu_1$ mit

$$(2.5.6) \qquad \|C(A_{\nu_1})\|^2 \leq q^4 .$$

O.B.d.A. gehen wir davon aus, daß gilt

(2.5.7) $$F_{\nu_1} = H_{\nu_1} ,$$

(2.5.8) $$\tilde{F}_{\nu_1} = S_{\nu_1} .$$

Aufgrund der globalen Konvergenz des Jacobi-Verfahrens für hermitesche Matrizen (vgl. FORSYTHE - HENRICI [7]) existiert dann ein Index $\nu_2 \in \mathbb{N}_0$, so daß in Schritt 3 die Abfrage

(2.5.9) $$\tau^2(H_{\nu_1+\nu_2}) \leq q^2$$

erfüllt ist.

Gilt nun auch $\tau^2(S_{\nu_1+\nu_2}) \leq q^2$, so sind wir (wegen $q > 0$ beliebig) fertig. Sei also

(2.5.10) $$q^2 < \tau^2(S_\nu)$$

mit

(2.5.11) $$\nu := \nu_1 + \nu_2 .$$

Wir bezeichnen mit $S_\nu^{(i)}$, $H_\nu^{(i)}$ die Matrizen $S_\nu$, $H_\nu$ nach Ausführung von i Jacobi - Rotationen gemäß Schritt 4 des Verfahrens. Nach Bemerkung 2.5.3 gibt es ein $k \in \mathbb{N}_0$ mit

(2.5.12) $$k \leq \tilde{k} \frac{n(n-1)}{2} < \tilde{k} \frac{n^2}{2}$$

und

(2.5.13) $$\tau^2(S_\nu^{(k)}) \leqq q^2 ,$$

(2.5.14) $$\tau^2(S_\nu^{(j)}) \leqq q^2 .$$

Aus (2.5.12) folgt mit der Voraussetzung $0 < q \leqq q_0$ und (2.5.1), (2.5.4) zunächst einmal

(2.5.15) $$(1 + 2Kq)^i < 2 , \qquad i \leqq k .$$

Setzen wir noch

$$0 \neq \tau^2(H_\nu) := \varepsilon^2 \leqq q^2 ,$$

so sind wir in der Situation

$$0 \neq \tau^2(H_\nu) \leqq q^2 ,$$

$$\tau^2(S_\nu) > q^2 \geqq \frac{1}{4} q^4 ,$$

$$\|C(A_\nu)\|^2 \leqq q^4$$

und (wegen $c < \frac{1}{2}$ )

$$q \leqq \frac{\sqrt{2}}{4} \frac{1}{64\, n^3} \rho_H^3$$

$$< \frac{1}{64\sqrt{2}\, n^3} \rho_H^3 .$$

Unter Beachtung von (2.4.9) sind damit alle Voraussetzungen des Satzes 2.3.3 erfüllt; mit (2.5.15) und $k \geqq 1$ erhalten wir

$$\tau^2(H_\nu^{(1)}) \leqq \tau^2(H_\nu) + 2\left(\frac{\tau^2(H_\nu)}{\rho_H}\right)^2$$

$$\leqq q^2\left(1 + 2\frac{q^2}{\rho_H^2}\right)_{22}$$

$$\leqq q^2(1 + 2Kq)$$

$$\leqq 2q^2 \quad . \tag{2.5.16}$$

Damit ist

$$\tau^2(H_\nu^{(1)}) \leqq (\sqrt{2}\,q)^2 \quad ,$$

$$\tau^2(S_\nu^{(1)}) \geqq q^2 > \frac{1}{4}(\sqrt{2}\,q)^2 \quad ,$$

$$\|C(A_\nu^{(1)})\|^2 = \|C(A_\nu)\|^2 \leqq q^4 < (\sqrt{2}\,q)^4$$

sowie

$$\sqrt{2}\,q \leqq \frac{1}{2}\,\frac{1}{64\,n^3}\,\rho_H^3$$

$$< \frac{1}{64\sqrt{2}\,n^3}\,\rho_H^3 \quad ,$$

so daß wiederum die Voraussetzungen für Satz 2.3.3 erfüllt sind.

Wir führen nun vollständige Induktion durch. Sei dazu für $j \leqq k$

$$\tau^2(H_\nu^{(j-1)}) \leqq q^2(1+2Kq)^{j-1} \leqq (\sqrt{2}\,q)^2 \quad ,$$

$$\tau^2(S_\nu^{(j-1)}) \geqq \frac{1}{4}\, q^2 \quad ,$$

$$\|C(A_\nu^{(j-1)}\|^2 \leqq q^4$$

und

$$\sqrt{2}\,q \leqq \frac{1}{64\sqrt{2}\,n^3}\, \rho_H^3 \quad .$$

Dann liefert uns Satz 2.3.3

$$\tau^2(H_\nu^{(j)}) \leqq \tau^2(H_\nu^{(j-1)})\left(1+2\frac{\tau^2(H_\nu^{(j-1)})}{\rho_H^2}\right)$$

$$\leqq \tau^2(H_\nu^{(j-1)})\left(1+2\sqrt{2}\,q\,\frac{\sqrt{2}\,q}{\rho_H^2}\right)$$

(2.5.17)

$$\leqq q^2(1+2Kq)^{j-1}(1+2qK)$$

$$= q^2(1+2Kq)^j$$

und damit mit (2.5.15)

$$\tau^2(H_\nu^{(j)}) \leqq (\sqrt{2}\,q)^2 \, ,$$

$$\tau^2(S_\nu^{(j)}) \geqq \frac{1}{4}\, q^2 \, ,$$

$$\|C(A_\nu^{(j)})\|^2 \leq q^4 ,$$

$$\sqrt{2}\, q \leq \frac{1}{64\sqrt{2}\, n^3} \rho_H^3 .$$

Mit der Definition von $k$ gemäß (2.5.12), (2.5.13) und (2.5.14) und mit Bemerkung 2.5.3 erhalten wir somit nach $k$ Iterationsschritten in Schritt 4 des Verfahrens schließlich

$$\tau^2(H_\nu^{(k)}) \leq 2q^2$$

und

$$\tau^2(S_\nu^{(k)}) \leq q^2 .$$

Wegen

$$\tau^2(A_\nu^{(k)}) = \tau^2(H_\nu^{(k)}) + \tau^2(S_\nu^{(k)})$$

folgt hieraus die Behauptung. ■

# 3. ÜBER EIN NEUES JACOBI - ÄHNLICHES VERFAHREN

## 3.1 EINFÜHRUNG

Wie bereits in der Einleitung gesagt, beruht unser neues Verfahren auf einer Idee von Rutishauser [14] zur Normalisierung einer reellen Matrix $\tilde{A}$. Die Methode von Rutishauser erzeugt Folgen $\tilde{U}_k$ und $R_k$ reeller Transformationsmatrizen, so daß mit

$$\tilde{A}_0 := \tilde{A} \,,$$

(3.1.1)

$$\tilde{A}_{k+1} := R_k^{-1}\,\tilde{U}_k^T\,\tilde{A}_k\,\tilde{U}_k R_k \quad , \; k \in \mathbb{N}_0 \,,$$

die Folge der $\tilde{A}_k$ für $k \to \infty$ gegen eine normale Matrix konvergiert. Dabei stellt die Transformation

(3.1.2) $$\tilde{A}'_k := \tilde{U}_k^T \tilde{A}_k \tilde{U}_k$$

eine reelle Jacobi - Rotation zur Annihilierung des Pivotelementes des Kommutators $C(\tilde{A}_k)$ dar, wogegen die Transformation

(3.1.3) $$\tilde{A}_{k+1} := R_k^{-1}\,\tilde{A}'_k R_k$$

der Reduzierung der euklidischen Norm der iterierten Matrix dient. Die Matrix $R_k$ ist dabei eine reelle Diagonalmatrix der Form

$$(3.1.4) \qquad R_k := \begin{pmatrix} 1 & & & & & & 0 \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & x & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ 0 & & & & & & 1 \end{pmatrix} , \qquad x \in \mathbb{R}\setminus\{0\} .$$

Setzt man nun anstelle von (3.1.2) eine komplexe Jacobi-Rotation, so kann man mit dieser Methode beliebige komplexe Matrizen näherungsweise normalisieren. Verbindet man schließlich nach einem Vorschlag von VESELIĆ diesen Algorithmus in jedem Iterationsschritt mit einer zweiten komplexen Jacobi - Rotation analog zu Schritt 3 des Eberlein - Verfahrens, so gelangt man zu einem neuen Jacobi - ähnlichen Verfahren zur Diagonalisierung beliebiger komplexer Matrizen.

Mit diesem neuen Verfahren wollen wir uns in diesem Kapitel beschäftigen. Wir stellen zunächst in Abschnitt 3.2 den genauen Verfahrensalgorithmus vor und werden dann in Abschnitt 3.3 die asymptotisch quadratische Konvergenz dieser Methode im Falle getrennter Eigenwerte beweisen.

## 3.2 DER VERFAHRENS - ALGORITHMUS

<u>SCHRITT 0 :</u> Die Pivotindexpaare (p,q) werden zyklisch gewählt. Für jeden Iterationsschritt führe man die folgenden Schritte 1 bis 6 durch, wobei A die jeweils aktuelle Matrix der Iteration bezeichne.

<u>SCHRITT 1 :</u> Man berechne die Elemente $c_{pp}$, $c_{pq}$ und $c_{qq}$ des Kommutators $C := C(A)$.
Falls gilt

$$|c_{pq}| \leq 5\,\tau^2(A)\,, \tag{3.2.1}$$

so setze man $A'' := A$ und fahre fort mit Schritt 4a; anderenfalls führe man die komplexe Jacobi - Rotation

$$A' := U_1^* A\, U_1 \tag{3.2.2}$$

mit den Rotationsparametern $\alpha_1$ und $\phi_1$ aus, die wie folgt bestimmt werden :

$$e^{i\alpha_1} := \frac{c_{pq}}{|c_{pq}|}\,, \tag{3.2.3}$$

$$\operatorname{ctg} 2\phi_1 := \frac{c_{pp} - c_{qq}}{2|c_{pq}|}\,, \qquad |\phi_1| \leq \frac{\pi}{4}\,. \tag{3.2.4}$$

<u>SCHRITT 2 :</u> Man berechne die rotierten Elemente $c'_{pp}$ und $c'_{qq}$ des Kommutators $C' := C(A')$ und bestimme $k \in \{p,q\}$ vermöge

$$k := \begin{cases} p \ , & \text{falls } |c'_{pp}| \geq |c'_{qq}| \ , \\ q \ , & \text{sonst.} \end{cases} \tag{3.2.5}$$

Man berechne dann den normreduzierenden Parameter $x \in \mathbb{R}$ gemäß

$$x := \left( \frac{\sum_{i \neq k} |a'_{ki}|^2}{\sum_{i \neq k} |a'_{ik}|^2} \right)^{1/4} . \tag{3.2.6}$$

<u>SCHRITT 3 :</u> Man führe die Transformation

$$A'' := R^{-1} A' R \tag{3.2.7}$$

durch; dabei sei R reelle Diagonalmatrix mit

$$r_{kk} := x \ , \tag{3.2.8}$$

$$r_{ii} := 1 \ , \quad i \neq k \ , \quad k \text{ gemäß } (3.2.5) \ . \tag{3.2.9}$$

Man fahre fort mit Schritt 4b .

<u>SCHRITT 4a :</u> Man berechne

(3.2.10) $$te := Re^2(a''_{pp} - a''_{qq}) \quad ,$$

(3.2.11) $$tee := Im^2(a''_{pp} - a''_{qq}) \quad .$$

i) Ist <u>$te \geqq tee$</u> , so setze man

(3.2.12) $$e^{i\alpha_2} := \frac{a''_{pq} + \overline{a''_{qp}}}{|a''_{pq} + \overline{a''_{qp}}|} \quad ,$$

(3.2.13) $$\operatorname{ctg} 2\phi_2 := \frac{Re(a''_{pp} - a''_{qq})}{|a''_{pq} + \overline{a''_{qp}}|} \quad , \qquad |\phi_2| \leqq \frac{\pi}{4} \quad .$$

ii) Ist <u>$te < tee$</u> , so setze man

(3.2.14) $$e^{i\alpha_2} := i\frac{a''_{pq} - \overline{a''_{qp}}}{|a''_{pq} - \overline{a''_{qp}}|} \quad ,$$

(3.2.15) $$\operatorname{ctg} 2\phi_2 := \frac{Im(a''_{pp} - a''_{qq})}{|a''_{pq} - \overline{a''_{qp}}|} \quad , \qquad |\phi_2| \leqq \frac{\pi}{4} \quad .$$

Man fahre fort mit Schritt 5 .

<u>SCHRITT 4b :</u> Man berechne

(3.2.16) $$te := |a''_{pq} + \overline{a''_{qp}}|^2 \quad ,$$

(3.2.17) $$tee := |a''_{pq} - \overline{a''_{qp}}|^2 \quad .$$

i) Ist <u>$te \geq tee$</u>, so setze man

(3.2.18) $$e^{i\alpha_2} := s\,\frac{a''_{pq}+\overline{a''_{qp}}}{|a''_{pq}+\overline{a''_{qp}}|}\,, \qquad s:=\mathrm{sign}(\sin 2\phi_1 \mathrm{Re}(a_{qq}-a_{pp}))\,,$$

(3.2.19) $$\mathrm{ctg}\,2\phi_2 := e^{i\alpha_2}\frac{\mathrm{Re}(a''_{pp}-a''_{qq})}{a''_{pq}+\overline{a''_{qp}}}\,, \quad \mathrm{sign}\,\phi_2 = -\,\mathrm{sign}\,\phi_1\,, \quad |\phi_2| \leq \frac{\pi}{2}\,.$$

ii) Ist <u>$te < tee$</u>, so setze man

(3.2.20) $$e^{i\alpha_2} := si\,\frac{a''_{pq}-\overline{a''_{qp}}}{|a''_{pq}-\overline{a''_{qp}}|}\,, \qquad s:=\mathrm{sign}(\sin 2\phi_1 \mathrm{Im}(a_{pp}-a_{qq}))\,,$$

(3.2.21) $$\mathrm{ctg}\,2\phi_2 := ie^{i\alpha_2}\,\frac{\mathrm{Im}(a''_{pp}-a''_{qq})}{a''_{pq}-\overline{a''_{qp}}}\,, \mathrm{sign}\phi_2 = -\,\mathrm{sign}\phi_1, \quad |\phi_2| \leq \frac{\pi}{2}\,.$$

<u>SCHRITT 5 :</u> Man führe die komplexe Jacobi - Rotation

(3.2.22) $$A''' := U_2^* A'' U_2$$

mit Pivotindexpaar (p,q) und gemäß Schritt 4a bzw. Schritt 4b bestimmten Rotationsparametern $\alpha_2$ und $\phi_2$ durch.

<u>SCHRITT 6 :</u> Falls $\tau^2(A''')$ hinreichend klein ist, so beende man die Iteration; anderenfalls setze man

$$A := A'''$$

und fahre fort mit Schritt 1 .

ANMERKUNG : Im folgenden werden wir verschiedentlich die Transformation nach Schritt 1 des Verfahrens als "vorbereitende Jacobi - Rotation", den Schritt 3 als "normreduzierende Transformation" und den Schritt 5 als "zweite Jacobi - Rotation" bezeichnen.

Wir wollen an dieser Stelle noch einige Bemerkungen zu unserem soeben beschriebenen Verfahren treffen, deren Beweis man unmittelbar durch Nachrechnen erhält.

BEMERKUNG 3.2.1 : *i) Nach der vorbereitenden Jacobi-Rotation gilt mit* $C' := C(A')$

$$c'_{pq} = 0.$$

*ii) Nach der normreduzierenden Transformation gilt mit* $C'' := C(A'')$ *und gemäß (3.2.5) bestimmtem Index* $k$

$$c''_{kk} = 0 \ .$$

*iii) Nach der zweiten Jacobi - Rotation gilt*

$$a'''_{pq} + \overline{a'''_{qp}} = 0 \quad , \quad te \geq tee \ ,$$

$$a'''_{pq} - \overline{a'''_{qp}} = 0 \quad , \quad te < tee \ . \qquad \blacksquare$$

BEMERKUNG 3.2.2 : *Nach Durchführung der normreduzierenden Transformation gilt für die Elemente der Matrix* A'' ( k *gemäß (3.2.5))*

$$a''_{ik} = x\, a'_{ik} , \qquad i \neq k ,$$

$$a''_{ki} = \frac{1}{x}\, a'_{ki} , \qquad i \neq k ,$$

$$a''_{ij} = a'_{ij} , \qquad i \neq k , \; j \neq k ,$$

$$a''_{ii} = a'_{ii} . \qquad \blacksquare$$

BEMERKUNG 3.2.3 : *Es gilt*

$$\|A'\|^2 - \|A''\|^2 = \left( \left( \sum_{i \neq k} |a'_{ki}|^2 \right)^{1/2} + \left( \sum_{i \neq k} |a'_{ik}|^2 \right)^{1/2} \right)^{-2} \cdot c_{kk}^2$$

$$\geqq 0 . \qquad \blacksquare$$

## 3.3 BEWEIS DER ASYMPTOTISCH QUADRATISCHEN KONVERGENZ DES VERFAHRENS IM FALLE GETRENNTER EIGENWERTE

Wir gehen in diesem Abschnitt aus von einer komplexen, nicht - normalen Matrix $A$ mit getrennten Eigenwerten $\lambda_i$ und untersuchen zunächst, welche Auswirkungen ein einzelner Iterationsschritt unseres Verfahrens mit Pivotindexpaar $(p,q)$ auf die einzelnen Matrixelemente hat.
Es bezeichne

$A'$ die Matrix $A$ nach der vorbereitenden Jacobi - Rotation ,

$A''$ die Matrix $A'$ nach der normreduzierenden Transformation ,

$A'''$ die Matrix $A''$ nach der zweiten Jacobi - Rotation ,

$C, C', C'', C'''$ die Kommutatoren der Matrizen $A, A', A'', A'''$.

Weiter sei definiert

(3.3.1) $$\delta := \min_{i \neq j} |\lambda_i - \lambda_j| > 0 ,$$

(3.3.2) $$\varepsilon^2 := \tau^2(A) ,$$

und o.B.d.A. gelte

(3.3.3) $$\|A\| \leq 1 .$$

Schließlich gehen wir im folgenden davon aus, daß für die Matrix $A$ die folgende Generalvoraussetzung erfüllt sei :

<u>VORAUSSETZUNG 3.3.1 :</u> *Die Matrix* A *sei bereits so weit diagonalisiert, daß gilt*

$$\varepsilon \leqq \left(\frac{\delta}{10}\right)^3 .$$

Halten wir an dieser Stelle noch fest, daß aus $\|A\| \leqq 1$ für $\delta$ folgt

(3.3.4) $$\delta \leqq \sqrt{2}$$

und somit aus unserer Voraussetzung 3.3.1 die im folgenden häufig verwendete Ungleichungskette

(3.3.5) $$\varepsilon \leqq \frac{\delta^3}{1000} \leqq \frac{\sqrt{2}}{1000}\delta^2 \leqq \frac{1}{500}\delta \leqq \frac{\sqrt{2}}{500} .$$

Schließlich gilt noch die

<u>BEMERKUNG 3.3.2 :</u> *Jedes Diagonalelement* $a_{ii}$ *der Matrix* A *ist eine Näherung für einen Eigenwert* $\lambda_i$ *von* A *, und es gilt*

$$|a_{ii} - \lambda_i| \leqq \varepsilon .$$

*Daraus folgt für je zwei Diagonalelemente* $a_{ii}$, $a_{jj}$

$$|a_{ii} - a_{jj}| \leqq |\lambda_i - \lambda_j| + 2\varepsilon$$

*bzw.*

$$|a_{ii} - a_{jj}| \geqq |\lambda_i - \lambda_j| - 2\varepsilon .$$

<u>BEWEIS :</u> Eine Beweis dieser Bemerkung findet man z.B. in [1]. ■

Wir wollen nun zunächst den Fall untersuchen, in dem zu Beginn des Iterationsschrittes gilt

FALL A : $$|c_{pq}| > 5\varepsilon^2 .$$

Wir erhalten dann als erstes die

BEMERKUNG 3.3.3 : *Es gilt*

$$|\operatorname{ctg} 2\phi_1| < \frac{1}{10} .$$

BEWEIS : Gemäß (3.2.4) ist

$$\operatorname{ctg} 2\phi_1 = \frac{c_{qq} - c_{pp}}{2|c_{pq}|} ,$$

und wegen

$$|c_{qq} - c_{pp}| = \left| \sum_{i=1}^{n} |a_{iq}|^2 - |a_{qi}|^2 - |a_{ip}|^2 + |a_{pi}|^2 \right|$$

$$\leq \tau^2(A) = \varepsilon^2$$

folgt die Behauptung. ■

Aus dieser Bemerkung folgt dann sofort

KOROLLAR 3.3.4 : *Es gilt*

$$\sin^2 2\phi_1 > 0.99 ,$$

$$\cos^2 2\phi_1 < 0.01 . \qquad \blacksquare$$

Unser erstes Ziel ist es, Abschätzungen für den normreduzierenden Parameter x zu finden. Zunächst einmal halten wir fest :

BEMERKUNG 3.3.5 : *Nach Durchführung der vorbereitenden Jacobi - Rotation gilt für die Pivotelemente* $a'_{pq}$ *und* $a'_{qp}$

*i)* $\quad |a'_{pq}| > 0.48\,\delta$ ,

*ii)* $\quad |a'_{qp}| > 0.48\,\delta$ .

BEWEIS : Wir werden nur die Abschätzung i) herleiten; der Beweis von ii) verläuft dann ganz analog.
Nach Bemerkung 1.2.6 gilt

$$|a'_{pq}| = |\tfrac{1}{2}\sin 2\phi_1(a_{qq}-a_{pp})-e^{i\alpha_1}\sin^2\phi_1 a_{qp}+e^{-i\alpha_1}\cos^2\phi_1 a_{pq}|$$

$$\geq \tfrac{1}{2}|\sin 2\phi_1(a_{qq}-a_{pp})|-|e^{i\alpha_1}\sin^2\phi_1 a_{qp}-e^{-i\alpha_1}\cos^2\phi_1 a_{pq}| \tag{3.3.6}$$

Wegen Korollar 3.3.4, Bemerkung 3.3.2 und (3.3.5) erhalten wir

$$\frac{1}{2}|\sin 2\phi_1||a_{qq}-a_{pp}| > \frac{1}{2}\sqrt{0.99}\,(\delta-2\varepsilon)$$

$$\geqq \frac{1}{2}\sqrt{0.99}\,(\delta-2\frac{\delta}{500})$$

(3.3.7) $$> 0.49\,\delta\,.$$

Weiter ist

$$|e^{i\alpha_1}\sin^2\phi_1 a_{qp} - e^{-i\alpha_1}\cos^2\phi_1 a_{pq}| \leqq |a_{qp}| + |a_{pq}|$$

(3.3.8) $$\leqq \sqrt{2}\,\varepsilon\,,$$

so daß aus (3.3.6) folgt

$$|a'_{pq}| > 0.49\,\delta - \sqrt{2}\,\varepsilon$$

$$\geqq 0.49\,\delta - \frac{\sqrt{2}}{500}\,\delta$$

$$> 0.48\,\delta\,.$$

■

Zur leichteren Notation der folgenden Abschätzungen für den Parameter $x$ führen wir an dieser Stelle noch die nachstehende Bezeichnung ein.

<u>DEFINITION 3.3.6 :</u> *Es sei* $\hat{x} \in \mathbb{R}$ *definiert durch*

$$(3.3.9) \qquad \hat{x} := \begin{cases} \left(\dfrac{|a'_{pq}|^2}{|a'_{qp}|^2}\right)^{1/4} & , \text{ falls } |c'_{pp}| \geq |c'_{qq}| \ , \\ \left(\dfrac{|a'_{qp}|^2}{|a'_{pq}|^2}\right)^{1/4} & , \text{ sonst } . \end{cases}$$

<u>LEMMA 3.3.7 :</u> *Es gilt*

$$|x - \hat{x}| < \frac{1.1\,\varepsilon^2}{\delta^2}\,\hat{x} \ ,$$

$$\left|\frac{1}{x} - \frac{1}{\hat{x}}\right| < \frac{1.1\,\varepsilon^2}{\delta^2}\,\hat{x} \ .$$

<u>BEWEIS :</u> Wir beweisen die behaupteten Abschätzungen für den Fall $|c'_{pp}| \geq |c'_{qq}|$ ; im anderen Fall läßt sich der Beweis dann wörtlich übertragen.
Gemäß (3.2.5), (3.2.6) ist x definiert als

$$(3.3.10) \qquad x = \left(\frac{\sum\limits_{i \neq p} |a'_{pi}|^2}{\sum\limits_{i \neq p} |a'_{ip}|^2}\right)^{1/4} \ ,$$

und nach (3.3.9) ist

$$(3.3.11) \qquad \hat{x} = \left(\frac{|a'_{pq}|^2}{|a'_{qp}|^2}\right)^{1/4} \ .$$

Nach Bemerkung 1.2.8 gilt weiter

$$(3.3.12) \qquad \sum_{i \neq p} |a'_{pi}|^2 \leq |a'_{pq}|^2 + \varepsilon^2 \quad ,$$

$$(3.3.13) \qquad \sum_{i \neq p} |a'_{ip}|^2 \leq |a'_{qp}|^2 + \varepsilon^2 \quad .$$

Damit folgt aus (3.3.10)

$$(3.3.14) \qquad x \leq \left( \frac{|a'_{pq}|^2 + \varepsilon^2}{|a'_{qp}|^2} \right)^{1/4} = \hat{x} \left(1 + \frac{\varepsilon^2}{|a'_{pq}|^2}\right)^{1/4}$$

sowie

$$(3.3.15) \qquad \frac{1}{x} \leq \left( \frac{|a'_{qp}|^2 + \varepsilon^2}{|a'_{pq}|^2} \right)^{1/4} = \frac{1}{\hat{x}} \left(1 + \frac{\varepsilon^2}{|a'_{qp}|^2}\right)^{1/4} \quad .$$

Mit Bemerkung 3.3.5 erhalten wir hieraus

$$(3.3.16) \qquad x < \hat{x} \left(1 + \frac{\varepsilon^2}{(0.48\delta)^2}\right)^{1/4}$$

und

$$(3.3.17) \qquad \frac{1}{x} < \frac{1}{\hat{x}} \left(1 + \frac{\varepsilon^2}{(0.48\delta)^2}\right)^{1/4} \quad .$$

Wegen Voraussetzung 3.3.1 ist $\frac{\varepsilon}{0.48\delta} < 1$ , und Taylorentwicklung liefert

$$(3.3.18) \qquad \left(1 + \frac{\varepsilon^2}{(0.48\delta)^2}\right)^{1/4} < 1 + \frac{1}{4} \frac{\varepsilon^2}{(0.48\delta)^2}$$

$$< 1 + 1.1 \frac{\varepsilon^2}{\delta^2} \quad .$$

Somit folgt aus (3.3.16) , (3.3.17)

(3.3.19) $$x < \hat{x}\,(1 + 1.1\,\frac{\varepsilon^2}{\delta^2})$$

und

(3.3.20) $$\frac{1}{x} < \frac{1}{\hat{x}}\,(1 + 1.1\,\frac{\varepsilon^2}{\delta^2}) \quad .$$

Kehrwertbildung liefert dann aus (3.3.19)

$$\frac{1}{x} > \frac{1}{\hat{x}}\,\frac{1}{1 + 1.1\,\frac{\varepsilon^2}{\delta^2}}$$

$$= \frac{1}{\hat{x}}\,(\,1 - \frac{1.1\,\varepsilon^2}{\delta^2 + 1.1\varepsilon^2}\,)$$

(3.3.21) $$> \frac{1}{\hat{x}}\,(\,1 - \frac{1.1\,\varepsilon^2}{\delta^2})$$

sowie analog aus (3.3.20)

(3.3.22) $$x > \hat{x}\,(\,1 - \frac{1.1\,\varepsilon^2}{\delta^2})$$

und damit insgesamt die Behauptung. ■

<u>LEMMA 3.3.8 :</u> *i) Es gilt*

$$|\hat{x} - 1| < \frac{2.2}{\delta^2}\,\varepsilon \quad ,$$

$$|\frac{1}{\hat{x}} - 1| < \frac{2.2}{\delta^2}\,\varepsilon \quad .$$

*ii) Für* $|a'_{pp} - a'_{qq}| > \varepsilon$ *gilt*

$$|\hat{x} - 1| < 0.53 \frac{\varepsilon^2}{\delta |a'_{pp} - a'_{qq}|},$$

$$|\frac{1}{\hat{x}} - 1| < 0.53 \frac{\varepsilon^2}{\delta |a'_{pp} - a'_{qq}|}.$$

<u>BEWEIS :</u> Wir betrachten wieder nur den Fall $|c'_{pp}| \geq |c'_{qq}|$, da anderenfalls der Beweis völlig analog erfolgt.

<u>Zu i) :</u> Mit den Transformationsformeln aus Bemerkung 1.2.6 rechnet man leicht nach, daß gilt

$$|a'_{pq}|^2 - |a'_{qp}|^2 = \cos 2\phi_1 (|a_{pq}|^2 - |a_{qp}|^2) + \tag{3.3.23}$$

$$+ \sin 2\phi_1 \mathrm{Re}\{(a_{qq} - a_{pp})(e^{i\alpha_1}\overline{a_{pq}} - e^{-i\alpha_1}\overline{a_{qp}})\}$$

Hieraus folgt mit Korollar 3.3.4 und (3.3.5)

$$\left| |a'_{pq}|^2 - |a'_{qp}|^2 \right| \leq |\cos 2\phi_1| \, ( |a_{pq}|^2 + |a_{qp}|^2 )$$

$$+ |\sin 2\phi_1| \, |a_{pp} - a_{qq}| \, ( |a_{pq}| + |a_{qp}| )$$

$$\leq \sqrt{0.01}\, \varepsilon^2 + \sqrt{2}\sqrt{2}\, \varepsilon$$

$$\leq (0.1 \, (\frac{\sqrt{2}}{10})^3 + 2) \, \varepsilon$$

$$< 2.0003 \, \varepsilon. \tag{3.3.24}$$

Für $\hat{x}$ erhalten wir somit mit Bemerkung 3.3.5 die Abschätzungen

$$(3.3.25) \qquad \hat{x} = \left(\frac{|a'_{pq}|^2}{|a'_{qp}|^2}\right)^{1/4} < \left(1 + \frac{2.0003\varepsilon}{(0.48\delta)^2}\right)^{1/4}$$

sowie

$$(3.3.26) \qquad \frac{1}{\hat{x}} = \left(\frac{|a'_{qp}|^2}{|a'_{pq}|^2}\right)^{1/4} < \left(1 + \frac{2.0003\varepsilon}{(0.48\delta)^2}\right)^{1/4} .$$

Aufgrund unserer Voraussetzung 3.3.1 ist

$$1 + \frac{2.0003\,\varepsilon}{(0.48\delta)^2} < 1 ,$$

und Taylorentwicklung liefert

$$(3.3.27) \qquad \hat{x} < 1 + \frac{1}{4}\,\frac{2.0003\varepsilon}{(0.48\delta)^2}$$

$$< 1 + 2.2\,\frac{\varepsilon}{\delta^2}$$

sowie analog

$$(3.3.28) \qquad \frac{1}{\hat{x}} < 1 + 2.2\,\frac{\varepsilon}{\delta^2} .$$

Bildung des Kehrwertes in (3.3.27) ergibt schließlich

$$\frac{1}{\hat{x}} > \frac{1}{1 + 2.2\frac{\varepsilon}{\delta^2}}$$

$$= 1 - \frac{2.2\,\varepsilon}{\delta^2 + 2.2\varepsilon}$$

$$> 1 - \frac{2.2}{\delta^2}\varepsilon \quad ,$$

und analog folgt aus (3.3.28)

$$\hat{x} > 1 - \frac{2.2}{\delta^2}\varepsilon \quad .$$

<u>Zu ii) :</u> Nach Durchführung der ersten, vorbereitenden Jacobi - Rotation gilt gemäß Bemerkung 3.2.1

$$(3.3.29) \qquad 0 = c'_{pq} = a'_{pq}(\overline{a'_{pp}-a'_{qq}}) + \overline{a'_{qp}}(a'_{qq}-a'_{pp}) + K \ ,$$

wobei K definiert ist durch

$$(3.3.30) \qquad K := \sum_{i \neq p,q} (\overline{a'_{ip}}a'_{iq} - a'_{pi}\overline{a'_{qi}}) \ .$$

Also ist

$$(3.3.31) \qquad a'_{pq}(\overline{a'_{pp}-a'_{qq}}) = \overline{a'_{qp}}(a'_{pp}-a'_{qq}) - K \ ,$$

und Multiplikation dieser Gleichung mit $\overline{a'_{pq}}$ bzw. $a'_{qp}$ ergibt

$$(3.3.32) \qquad |a'_{pq}|^2(\overline{a'_{pp}-a'_{qq}}) = \overline{a'_{pq}a'_{qp}}(a'_{pp}-a'_{qq}) - K\,\overline{a'_{pq}}$$

bzw.

$$(3.3.33) \qquad |a'_{qp}|^2(a'_{pp}-a'_{qq}) = a'_{pq}a'_{qp}(\overline{a'_{pp}-a'_{qq}}) + K\,a'_{qp} \ .$$

Subtrahieren wir nun Gleichung (3.3.33) von der konjugiert komplex genommenen Gleichung (3.3.32), so erhalten wir

$$(3.3.34) \qquad (|a'_{pq}|^2-|a'_{qp}|^2)(a'_{pp}-a'_{qq}) = -\overline{K}a'_{pq} - Ka'_{qp} .$$

Hieraus wiederum folgt

$$(3.3.35) \qquad \left||a'_{pq}|^2-|a'_{qp}|^2\right| |a'_{pp}-a'_{qq}| \leqq |K| \, ( |a'_{pq}| + |a'_{qp}|)$$

und somit wegen $|a'_{pq}| + |a'_{qp}| \neq 0$

$$(3.3.36) \qquad \left||a'_{pq}|-|a'_{qp}|\right| \leqq \frac{|K|}{|a'_{pp}-a'_{qq}|} .$$

Mit Hilfe von Bemerkung 3.3.5 und wegen $|K| \leqq \frac{1}{2}\varepsilon^2$ erhalten wir

$$\hat{x} = \left(\frac{|a'_{pq}|}{|a'_{qp}|}\right)^{1/2}$$

$$\leqq (1 + \frac{\varepsilon^2}{2|a'_{qp}|\,|a'_{pp}-a'_{qq}|})^{1/2}$$

$$(3.3.37) \qquad < (1 + \frac{\varepsilon^2}{0.96\,\delta\,|a'_{pp}-a'_{qq}|})^{1/2} .$$

Wegen $|a'_{pp}-a'_{qq}| > \varepsilon$ liefert dann wieder die Taylorentwicklung

$$\hat{x} < 1 + \frac{1}{2}\,\frac{\varepsilon^2}{0.96\,\delta\,|a'_{pp}-a'_{qq}|}$$

$$(3.3.38) \qquad < 1 + 0.53\,\frac{\varepsilon^2}{\delta|a'_{pp}-a'_{qq}|} .$$

Analog leitet man her

$$\frac{1}{\hat{x}} < 1 + 0.53 \frac{\varepsilon^2}{\delta |a'_{pp} - a'_{qq}|} , \tag{3.3.39}$$

und Bildung des Kehrwertes in (3.3.38) und (3.3.39) liefert die Behauptung. ■

Wir sind nunmehr in der Lage, die gewünschten Abschätzungen für den Parameter $x$ selbst herzuleiten.

<u>LEMMA 3.3.9 :</u> *i) Es gilt*

$$|x - 1| < 2.21 \frac{\varepsilon}{\delta^2} ,$$

$$\left|\frac{1}{x} - 1\right| < 2.21 \frac{\varepsilon}{\delta^2} .$$

*ii) Für* $|a'_{pp} - a'_{qq}| > \varepsilon$ *gilt*

$$|x - 1| < 2.31 \frac{\varepsilon^2}{\delta^2 |a'_{pp} - a'_{qq}|} ,$$

$$\left|\frac{1}{x} - 1\right| < 2.31 \frac{\varepsilon^2}{\delta^2 |a'_{pp} - a'_{qq}|} .$$

<u>BEWEIS :</u> <u>Zu i) :</u> Nach Lemma 3.3.7 und Lemma 3.3.8(i) gilt

$$\begin{aligned} x &< \hat{x} \left(1 + \frac{1.1\varepsilon^2}{\delta^2}\right) \\ &< \left(1 + \frac{2.2\varepsilon}{\delta^2}\right)\left(1 + \frac{1.1\varepsilon^2}{\delta^2}\right) \\ &= 1 + \frac{\varepsilon}{\delta^2}\left(2.2 + 1.1\varepsilon + \frac{2.42\varepsilon^2}{\delta^2}\right) \end{aligned} \tag{3.3.40}$$

sowie

$$x > \hat{x}\,(1 - \frac{1.1\varepsilon^2}{\delta^2})$$

$$> (1 - \frac{2.2\varepsilon}{\delta^2})(1 - \frac{1.1\varepsilon^2}{\delta^2})$$

$$(3.3.41) \qquad > 1 - \frac{\varepsilon}{\delta^2}(2.2 + 1.1\varepsilon + \frac{2.42\varepsilon^2}{\delta^2}) \ .$$

Wegen (3.3.5) gilt

$$2.2 + 1.1\varepsilon + \frac{2.42\,\varepsilon^2}{\delta^2} \leq 2.2 + 1.1\,(\frac{\sqrt{2}}{10})^3 + \frac{2.42\,\sqrt{2}\,(\sqrt{2})^3}{10^3 \cdot 10^3}$$

$$< 2.21 \ ,$$

und somit folgt aus (3.3.40) und (3.3.41) die Behauptung. Der Beweis der zweiten behaupteten Ungleichung verläuft wieder ganz analog.

<u>Zu ii) :</u> Nach Lemma 3.3.7 und Lemma 3.3.8 (ii) gilt

$$x < \hat{x}\,(1 + \frac{1.1\varepsilon^2}{\delta^2})$$

$$< (1 + 0.53\frac{\varepsilon^2}{\delta|a'_{pp}-a'_{qq}|})(1 + \frac{1.1\varepsilon^2}{\delta^2})$$

$$(3.3.42) \qquad = 1 + \frac{\varepsilon^2}{\delta^2|a'_{pp}-a'_{qq}|}\,(0.53\delta + 1.1|a'_{pp}-a'_{qq}| + 0.53\frac{1.1\varepsilon^2}{\delta}) \ ,$$

und analog erhalten wir

$$(3.3.43) \quad x > 1 - \frac{\varepsilon^2}{\delta^2|a'_{pp}-a'_{qq}|}\,(0.53\delta + 1.1|a'_{pp}-a'_{qq}| + 0.53\frac{1.1\varepsilon^2}{\delta}) \ .$$

Nun ist wegen (3.3.5) und $|a'_{pp}-a'_{qq}| \leq \sqrt{2}$

$$0.53\delta + 1.1|a'_{pp}-a'_{qq}| + 0.53\frac{1.1\varepsilon^2}{\delta} \leq (0.53+1.1)\sqrt{2} + 0.53\frac{1.1\ 4\sqrt{2}}{10^6}$$

$$< 2.31 \ .$$

Damit folgt aus (3.3.42) und (3.3.43) die Behauptung. ■

Unser nächstes Ziel ist es nun, zu einer Abschätzung für die Pivot - Elemente $a'''_{pq}$ und $a'''_{qp}$ nach Durchführung des vollständigen Iterationsschrittes zu gelangen. Dazu benötigen wir wieder einige Vorüberlegungen.

Mit dem bisher Bewiesenen können wir zunächst eine Abschätzung für den Abstand $\big||a''_{pq}| - |a''_{qp}|\big|$ angeben.

<u>LEMMA 3.3.10 :</u> *Nach Durchführung der normreduzierenden Transformation gilt für die Pivot - Elemente* $a''_{pq}$ *und* $a''_{qp}$

$$\big||a''_{pq}| - |a''_{qp}|\big| < 1.6\frac{\varepsilon^2}{\delta^2} \ .$$

<u>BEWEIS :</u> Aus Lemma 3.3.7 und Lemma 3.3.8(i) erhalten wir mit (3.3.5)

$$|x - \hat{x}| \leq \frac{1.1\varepsilon^2}{\delta^2}\left(1 + \frac{2.2\varepsilon}{\delta^2}\right)$$

$$\leq \frac{1.1\varepsilon^2}{\delta^2}\left(1 + \frac{2.2\sqrt{2}}{10^3}\right)$$

(3.3.44) $$< 1.11\frac{\varepsilon^2}{\delta^2}$$

sowie analog

(3.3.45) $$\left|\frac{1}{x} - \frac{1}{\hat{x}}\right| < 1.11\frac{\varepsilon^2}{\delta^2} \ .$$

Sei nun wieder vorausgesetzt, daß gilt

$$|c'_{pp}| \geqq |c'_{qq}| \quad .$$

(Anderenfalls beweist man die Behauptung wieder analog.)

Mit der Definition (3.2.6) von $x$ und der Definition 3.3.6 von $\hat{x}$ gilt gemäß Schritt 3 des Verfahrens

$$\begin{aligned} |a''_{pq}| - |a''_{qp}| &= |a'_{pq}| \frac{1}{x} - |a'_{qp}|\, x \\ &= |a'_{pq}| \frac{1}{\hat{x}} + |a'_{pq}| (\frac{1}{x} - \frac{1}{\hat{x}}) - |a'_{qp}| \hat{x} - |a'_{qp}| (x - \hat{x}) \\ &= \sqrt{|a'_{pq} a'_{qp}|} + |a'_{pq}| (\frac{1}{x} - \frac{1}{\hat{x}}) - \sqrt{|a'_{pq} a'_{qp}|} - |a'_{qp}| (x - \hat{x}) \ , \end{aligned}$$

und mit (3.3.44) und (3.3.45) folgt

$$\begin{aligned} \big| |a''_{pq}| - |a''_{qp}| \big| &\leqq 1.11 \frac{\varepsilon^2}{\delta^2} (|a'_{pq}| + |a'_{qp}|) \\ &< 1.11 \sqrt{2} \frac{\varepsilon^2}{\delta^2} \\ &< 1.6 \frac{\varepsilon^2}{\delta^2} \ . \end{aligned}$$

■

Um zu den gewünschten Abschätzungen für $a'''_{pq}$ und $a'''_{qp}$ zu gelangen, benötigen wir noch Aussagen über die Abweichungen der Rotationsparameter beider Jacobi - Rotationen voneinander, die wir im folgenden herleiten wollen.

<u>BEMERKUNG 3.3.11 :</u> *i) Aus* $|a''_{pq} + \overline{a''_{qp}}| \geqq |a''_{pq} - \overline{a''_{qp}}|$ *folgt*

$$|\mathrm{Re}(a_{pp} - a_{qq})| > 0.7\delta \quad .$$

*ii) Aus* $|a''_{pq} + \overline{a''_{qp}}| < |a''_{pq} - \overline{a''_{qp}}|$ *folgt*

$$|\mathrm{Im}(a_{pp} - a_{qq})| > 0.7\delta \quad .$$

<u>BEWEIS : Zu i):</u> Es ist

$$(3.3.46) \qquad |a''_{pq} \pm \overline{a''_{qp}}|^2 = |a''_{pq}|^2 + |a''_{qp}|^2 \pm 2\mathrm{Re}(a''_{pq}a''_{qp}) \ .$$

Aus der Voraussetzung

$$|a''_{pq} + \overline{a''_{qp}}| \geqq |a''_{pq} - \overline{a''_{qp}}|$$

folgt somit

$$(3.3.47) \qquad \mathrm{Re}(a''_{pq}a''_{qp}) = \mathrm{Re}(a'_{pq}a'_{qp}) \geqq 0.$$

Weiter ist gemäß Bemerkung 1.2.6

$$(3.3.48) \qquad a'_{pq}a'_{qp} = \frac{1}{4}\sin^2 2\phi_1 (a_{qq}-a_{pp})^2 + R \ ,$$

wobei

$$(3.3.49) \qquad \begin{aligned} R := {} & (\cos^4\phi_1 + \sin^4\phi_1)a_{pq}a_{qp} \\ & - \cos^2\phi_1 \sin^2\phi_1 (e^{-2i\alpha_1}a_{pq}^2 + e^{2i\alpha_1}a_{qp}^2) \\ & + \frac{1}{2}\sin 2\phi_1 \cos 2\phi_1 (a_{qq}-a_{pp})(e^{-i\alpha_1}a_{pq} + e^{i\alpha_1}a_{qp}) \ . \end{aligned}$$

Für R gilt dann mit (3.3.5) und Korollar 3.3.4

$$\begin{aligned} R \leqq {} & (1 - \frac{1}{2}\sin^2 2\phi_1)|a_{pq}a_{qp}| + \\ & + \frac{1}{4}\sin^2 2\phi_1 (|a_{pq}|^2 + |a_{qp}|^2) + \\ & + \frac{1}{2}|\sin 2\phi_1||\cos 2\phi_1||a_{pp}-a_{qq}|(|a_{pq}| + |a_{qp}|) \\ \leqq {} & (1 - 0.495)\frac{1}{2}\varepsilon^2 + \frac{1}{4}\varepsilon^2 + \frac{1}{2}\,\frac{1}{10}\sqrt{2}\sqrt{2}\,\varepsilon \end{aligned}$$

$$\leq \varepsilon \{0.505 \frac{1}{2} \frac{\sqrt{2}}{500} + \frac{1}{4} \frac{\sqrt{2}}{500} + \frac{1}{10}\}$$

(3.3.50) $$< 0.11\,\varepsilon \ .$$

Aus (3.3.48) erhalten wir

(3.3.51) $$\mathrm{Re}(a'_{pq}a'_{qp}) = \frac{1}{4}\sin^2 2\phi_1(\mathrm{Re}^2(a_{qq}-a_{pp}) - \mathrm{Im}^2(a_{qq}-a_{pp})) + \mathrm{Re}(R) \ .$$

Wegen Bemerkung 3.3.2 und (3.3.5) gilt

$$\mathrm{Re}^2(a_{qq}-a_{pp}) + \mathrm{Im}^2(a_{qq}-a_{pp}) = |a_{pp}-a_{qq}|^2$$

$$\geq (\delta - 2\varepsilon)^2$$

$$> 0.99\delta^2$$

und somit

(3.3.52) $$\mathrm{Im}^2(a_{qq}-a_{pp}) > 0.99\delta^2 - \mathrm{Re}^2(a_{qq}-a_{pp}) \ .$$

Setzen wir dies in (3.3.51) ein, so erhalten wir zunächst

$$\mathrm{Re}(a'_{pq}a'_{qp}) < \frac{1}{4}\sin^2 2\phi_1(2\mathrm{Re}^2(a_{qq}-a_{pp}) - 0.99\delta^2) + |R|$$

und schließlich mit (3.3.47),(3.3.50) und Korollar 3.3.4

$$\mathrm{Re}^2(a_{qq}-a_{pp}) > \frac{1}{2}(0.99\delta^2 - \frac{4|R|}{\sin^2 2\phi_1})$$

$$> \frac{1}{2}(0.99\delta^2 - 0.44\,\frac{\sqrt{2}\,\delta^2}{1000\cdot 0.99})$$

$$> 0.49\delta^2 \ .$$

Zu ii) : Mit

$$\mathrm{Re}(a'_{pq}a'_{qp}) \leq 0$$

und

$$\mathrm{Re}^2(a_{qq}-a_{pp}) > 0.99\delta^2 - \mathrm{Im}^2(a_{qq}-a_{pp})$$

verläuft der Beweis analog zu i). ■

Wir können nun das folgende Lemma beweisen.

LEMMA 3.3.12 : *Für die Parameter* $\alpha_1$ *und* $\alpha_2$ *der beiden Jacobi - Rotationen gilt*

$$|e^{i\alpha_1} - e^{i\alpha_2}| < 26\frac{\varepsilon}{\delta^3} .$$

BEWEIS : Wir betrachten nur den Fall, in dem gilt

$$|a''_{pq} + \overline{a''_{qp}}| \geq |a''_{pq} - \overline{a''_{qp}}| \tag{3.3.53}$$

und beweisen die Behauptung mit Bemerkung 3.3.11 i); im anderen Fall kann der Beweis mittels Bemerkung 3.3.11 ii) analog geführt werden.

Sei also (3.3.53) erfüllt. Gemäß (3.2.18) ist dann

$$e^{i\alpha_2} := s\frac{a''_{pq} + \overline{a''_{qp}}}{|a''_{pq} + \overline{a''_{qp}}|} , \quad s \in \{-1,1\} . \tag{3.3.54}$$

Nun ist entweder ($|c'_{pp}| \geq |c'_{qq}|$)

$$a''_{pq} + \overline{a''_{qp}} = \frac{1}{x}a'_{pq} + x\,\overline{a'_{qp}} \tag{3.3.55}$$

oder $(|c'_{pp}| < |c'_{qq}|)$

(3.3.56) $$a'_{pq} + \overline{a'_{qp}} = x\, a'_{pq} + \frac{1}{x}\, \overline{a'_{qp}} \quad .$$

Setzen wir

$$K_1 := \begin{cases} a'_{pq}(\frac{1}{x} - 1) + \overline{a'_{qp}}(x - 1) \ , & |c'_{pp}| \geq |c'_{qq}|, \\ a'_{pq}(x - 1) + \overline{a'_{qp}}(\frac{1}{x} - 1) \ , & \text{sonst}, \end{cases}$$

so gilt demnach

(3.3.57) $$a'_{pq} + \overline{a'_{qp}} = a'_{pq} + \overline{a'_{qp}} + K_1 \ ,$$

und wegen Lemma 3.3.9 i) gilt für $K_1$

(3.3.58) $$|K_1| < (|a'_{pq}| + |a'_{qp}|)\, 2.21 \frac{\varepsilon}{\delta^2}$$

$$\leq 2.21\sqrt{2}\, \frac{\varepsilon}{\delta^2} \quad .$$

Nun gilt weiter (Bemerkung 1.2.6)

(3.3.59) $$a'_{pq} + \overline{a'_{qp}} = \sin 2\phi_1\, e^{i\alpha_1} \mathrm{Re}(a_{qq} - a_{pp}) + K_2 \ ,$$

wobei

(3.3.60) $$K_2 := \cos^2\phi_1 (a_{pq} + \overline{a_{qp}}) - \sin^2\phi_1 e^{2i\alpha_1} (\overline{a_{pq}} + a_{qp})$$

und

$$|K_2| \leq |a_{pq} + \overline{a_{qp}}|$$

(3.3.61) $$\leq \sqrt{2}\, \varepsilon \quad .$$

Somit erhalten wir aus (3.3.54)

$$e^{i\alpha_2} = s\,\frac{e^{i\alpha_1}\sin 2\phi_1\,\mathrm{Re}(a_{qq}-a_{pp}) + K_1 + K_2}{|e^{i\alpha_1}\sin 2\phi_1\,\mathrm{Re}(a_{qq}-a_{pp}) + K_1 + K_2|}\,, \tag{3.3.62}$$

und setzen wir noch

$$K_0 := \frac{\sin 2\phi_1\,\mathrm{Re}(a_{qq}-a_{pp}) + e^{-i\alpha_1}(K_1 + K_2)}{|e^{i\alpha_1}\sin 2\phi_1\,\mathrm{Re}(a_{qq}-a_{pp}) + K_1 + K_2\,|}\,, \tag{3.3.63}$$

so gilt

$$e^{i\alpha_2} = s\,e^{i\alpha_1}K_0\,. \tag{3.3.64}$$

Es bleibt also noch zu zeigen, daß bis auf $O(\varepsilon)$ gilt $K_0 = 1$.

Zunächst einmal ist nach Bemerkung 3.3.11 und Korollar 3.3.4

$$|e^{i\alpha_1}\sin 2\phi_1\,\mathrm{Re}(a_{qq}-a_{pp})| > \sqrt{0.99}\;0.7\delta\;; \tag{3.3.65}$$

weiter gilt wegen (3.3.58),(3.3.61) und (3.3.5)

$$|K_1 + K_2| \leq \sqrt{2}\,\varepsilon\,(2.21\frac{1}{\delta^2} + 1)$$

$$\leq \sqrt{2}\,\varepsilon\,(2.21 + 2)\frac{1}{\delta^2} \tag{3.3.66a}$$

$$\leq 4.21\sqrt{2}\,\frac{\delta}{10^3}\,. \tag{3.3.66b}$$

Damit können wir den Imaginärteil von $K_0$ abschätzen gemäß

$$|\mathrm{Im}(K_0)| \leq \frac{4.21\sqrt{2}\,\frac{\varepsilon}{\delta^2}}{\sqrt{0.99}\;0.7\delta - 4.21\sqrt{2}\,\frac{\delta}{10^3}}$$

$$< 8.63\,\frac{\varepsilon}{\delta^3}\,, \tag{3.3.67}$$

und für den Realteil von $K_0$ erhalten wir

$$|\mathrm{Re}(K_0)| \lesseqgtr \frac{|\sin 2\phi_1 \mathrm{Re}(a_{qq}-a_{pp})| \pm |K_1+K_2|}{|\sin 2\phi_1 \mathrm{Re}(a_{qq}-a_{pp})| \mp |K_1+K_2|}$$

$$= 1 \pm \frac{2|K_1+K_2|}{|\sin 2\phi_1 \mathrm{Re}(a_{pp}-a_{qq})| \mp |K_1+K_2|}$$

$$\gtreqless 1 \pm \frac{2|K_1+K_2|}{|\sin 2\phi_1 \mathrm{Re}(a_{pp}-a_{qq})| - |K_1+K_2|}$$

$$\lessgtr 1 \pm 2 \frac{\sqrt{2}\, 4.21 \frac{\varepsilon}{\delta^2}}{\sqrt{0.99}\, 0.7\delta - \sqrt{2}\, 4.21 \frac{\delta}{10^3}}$$

(3.3.68) $$\lessgtr 1 \pm 17.26 \frac{\varepsilon}{\delta^3} \quad .$$

Wegen (3.3.64) können wir schreiben

$$e^{i\alpha_2} = s\, e^{i\alpha_1}(\mathrm{Re}(K_0) + i\mathrm{Im}(K_0))$$

(3.3.69) $$= s\, e^{i\alpha_1} \mathrm{sign}(\mathrm{Re}(K_0)) + s\, e^{i\alpha_1} i\mathrm{Im}(K_0) + $$
$$+ s\, e^{i\alpha_1} \mathrm{sign}(\mathrm{Re}(K_0)) \{|\mathrm{Re}(K_0)| - 1\} \quad .$$

Aus der Definition (3.3.63) von $K_0$ folgt nun aufgrund von (3.3.65) und (3.3.66)

$$\mathrm{sign}(\mathrm{Re}(K_0)) = \mathrm{sign}(\sin 2\phi_1 \mathrm{Re}(a_{qq}-a_{pp}))$$

und damit wegen (3.2.18)

(3.3.70) $$\mathrm{sign}(\mathrm{Re}(K_0)) = s \quad .$$

Somit folgt aus (3.3.69)

$$e^{i\alpha_2} = e^{i\alpha_1} + e^{i\alpha_1}(|\mathrm{Re}(K_0)| - 1) + s e^{i\alpha_1} i\,\mathrm{Im}(K_0) ,$$

und mit (3.3.67) und (3.3.68) erhalten wir die Behauptung. ■

<u>Lemma 3.3.13 :</u> *Für die Winkel $\phi_1$ und $\phi_2$ der beiden Jacobi-Rotationen gilt*

$$|\mathrm{ctg}\, 2\phi_1 + \mathrm{ctg}\, 2\phi_2| < 6.5 \frac{\varepsilon}{\delta^3} .$$

<u>BEWEIS :</u> Wieder betrachten wir nur den Fall

(3.3.71) $$|a''_{pq} + \overline{a''_{qp}}| \geq |a''_{pq} - \overline{a''_{qp}}| ;$$

im anderen Fall läßt sich der Beweis analog übertragen.

Sei also (3.3.71) erfüllt. Dann ist nach (3.2.19)

(3.3.72) $$\mathrm{ctg}\, 2\phi_2 = \pm \frac{\mathrm{Re}(a'_{pp} - a'_{qq})}{|a''_{pq} + \overline{a''_{qp}}|} .$$

Im Beweis des Lemmas 3.3.12 haben wir gezeigt (vgl. (3.3.57) bis (3.3.61) , (3.3.66)), daß gilt

(3.3.73) $$a''_{pq} + \overline{a''_{qp}} = \sin 2\phi_1\, e^{i\alpha_1} \mathrm{Re}(a_{qq} - a_{pp}) + K_1 + K_2$$

mit

(3.3.74) $$|K_1 + K_2| \leq 4.21 \sqrt{2} \frac{\varepsilon}{\delta^2} .$$

Für $a'_{pp} - a'_{qq}$ gilt nach Bemerkung 1.2.6

(3.3.75) $$a'_{pp}-a'_{qq} = \cos 2\phi_1(a_{pp}-a_{qq}) + K_3 ,$$

wobei

(3.3.76) $$K_3 := \sin 2\phi_1(e^{i\alpha_1}a_{qp} + e^{-i\alpha_1}a_{pq})$$

und

(3.3.77) $$|K_3| \leq |a_{pq}| + |a_{qp}| \leq \sqrt{2}\,\varepsilon .$$

Damit folgt aus (3.3.72)

$$\operatorname{ctg} 2\phi_2 = \pm \frac{\cos 2\phi_1 \operatorname{Re}(a_{pp}-a_{qq}) + \operatorname{Re}(K_3)}{|e^{i\alpha_1}\sin 2\phi_1 \operatorname{Re}(a_{pp}-a_{qq}) + K_1 + K_2|}$$

(3.3.78) $$= \pm \frac{\cos 2\phi_1}{\left|e^{i\alpha_1}\sin 2\phi_1 + \frac{K_1 + K_2}{\operatorname{Re}(a_{qq}-a_{pp})}\right|} \pm K_4 ,$$

wobei

(3.3.79) $$K_4 := \frac{\operatorname{Re}(K_3)}{|e^{i\alpha_1}\sin 2\phi_1 \operatorname{Re}(a_{qq}-a_{pp}) + K_1 + K_2|} .$$

Nach (3.3.77) und (3.3.65),(3.3.66) ist

$$|K_4| \leq \frac{\sqrt{2}\,\varepsilon}{\sqrt{0.99}\,0.7\delta - 4.21\sqrt{2}\,\frac{\delta}{1000}}$$

(3.3.80) $$< 2.1\,\frac{\varepsilon}{\delta} .$$

Setzen wir nun noch

(3.3.81) $$K_5 := \frac{1}{\left|e^{i\alpha_1} + \frac{K_1 + K_2}{\sin 2\phi_1 \operatorname{Re}(a_{qq}-a_{pp})}\right|} ,$$

so können wir (3.3.78) schreiben in der Form

$$\text{(3.3.82)} \qquad \operatorname{ctg} 2\phi_2 = \pm \operatorname{ctg} 2\phi_1 \cdot K_5 \pm K_4 \quad .$$

Wir müssen nun noch $|\operatorname{ctg} 2\phi_1 + \operatorname{ctg} 2\phi_2|$ mit Hilfe von $K_5$ und $K_4$ abschätzen; dazu unterscheiden wir zwei Fälle.

<u>FALL I :</u> Es ist $|\operatorname{ctg} 2\phi_1| K_5 \leqq |K_4|$ .

Dann gilt zunächst

$$\text{(3.3.83)} \qquad |\operatorname{ctg} 2\phi_1| \leqq \frac{|K_4|}{K_5} \quad ,$$

und aus (3.3.82) folgt

$$\text{(3.3.84)} \qquad |\operatorname{ctg} 2\phi_2| \leqq 2|K_4| \quad .$$

Also ist

$$\text{(3.3.85)} \qquad |\operatorname{ctg} 2\phi_1 + \operatorname{ctg} 2\phi_2| \leqq |K_4|(2 + \frac{1}{K_5}) \quad .$$

Mit (3.3.81),(3.3.66), Korollar 3.3.5 und Bemerkung 3.3.11 gilt

$$\frac{1}{K_5} = |e^{i\alpha_1} + \frac{K_1 + K_2}{\sin 2\phi_1 \operatorname{Re}(a_{qq} - a_{pp})}|$$

$$< 1 + \frac{4.21\sqrt{2}\,\frac{\varepsilon}{\delta^2}}{\sqrt{0.99}\,0.7\delta}$$

$$\text{(3.3.86)} \qquad < 1 + 8.55\,\frac{\varepsilon}{\delta^3} \quad ,$$

so daß aus (3.3.85) mit (3.3.80) folgt

$$|\operatorname{ctg} 2\phi_1 + \operatorname{ctg} 2\phi_2| < 2.1\frac{\varepsilon}{\delta}\left(2 + 1 + 8.55\frac{\sqrt{2}}{1000\,\delta^2}\right)$$

$$< 6.5\frac{\varepsilon}{\delta^3} \quad .$$

<u>FALL II :</u> Sei nun $|\operatorname{ctg} 2\phi_1|\, K_5 > |K_4|$ .

Dann ist nach 3.3.82 wegen $K_5 > 0$ das Vorzeichen von $\operatorname{ctg} 2\phi_2$ bestimmt durch das Vorzeichen von $\operatorname{ctg} 2\phi_1$. Gemäß (3.2.19) gilt

$$\operatorname{sign}(\operatorname{ctg} 2\phi_2) = -\operatorname{sign}(\operatorname{ctg} 2\phi_1) \ ,$$

und wir erhalten aus (3.3.82)

$$\operatorname{ctg} 2\phi_2 = -\operatorname{ctg} 2\phi_1\, K_5 \pm K_4 \quad . \tag{3.3.87}$$

Für $K_5$ erhalten wir aus (3.3.86)

$$K_5 > \frac{1}{1 + 8.55\frac{\varepsilon}{\delta^3}}$$

$$= 1 - \frac{8.55\frac{\varepsilon}{\delta^3}}{1 + 8.55\frac{\varepsilon}{\delta^3}}$$

und daher

$$1 - K_5 < 8.55\frac{\varepsilon}{\delta^3} \quad . \tag{3.3.88}$$

Zusammen mit (3.3.80) und Bemerkung 3.3.3 erhalten wir aus (3.3.87)

$$|\operatorname{ctg} 2\phi_1 + \operatorname{ctg} 2\phi_2| \leq |\operatorname{ctg} 2\phi_1 (1 - K_5)| + |K_4|$$

$$< 0.1 \frac{8.55}{\delta^3}\varepsilon + 2.1 \frac{\varepsilon}{\delta}$$

$$\leq 0.855 \frac{\varepsilon}{\delta^3} + 2.1 \frac{2\varepsilon}{\delta^3}$$

$$< 6.5 \frac{\varepsilon}{\delta^3} \quad . \qquad \blacksquare$$

<u>KOROLLAR 3.3.14 :</u> *Es gilt*

$$|\sin(\phi_1+\phi_2)| < 4.6 \frac{\varepsilon}{\delta^3} \quad .$$

<u>BEWEIS :</u> Nach Lemma 3.3.13 gilt

$$|\operatorname{ctg} 2\phi_1 + \operatorname{ctg} 2\phi_2| = \left|\frac{\sin 2(\phi_1+\phi_2)}{\sin 2\phi_1 \sin 2\phi_2}\right|$$

(3.3.89) $$< 6.5 \frac{\varepsilon}{\delta^3}$$

und somit

(3.3.90) $$|\sin 2(\phi_1+\phi_2)| < 6.5 \frac{\varepsilon}{\delta^3} \quad .$$

Daraus folgt

(3.3.91) $$\frac{k\pi}{2} \leq |\phi_1+\phi_2| \leq \frac{k\pi}{2} + \frac{1}{2} \arcsin (6.5 \frac{\varepsilon}{\delta^3}) \quad , \; k \in \mathbb{N}_0 .$$

Nun ist

$$|\phi_1| \leq \frac{\pi}{4} \quad , \quad |\phi_2| \leq \frac{\pi}{2} \quad ,$$

und nach (3.2.21) gilt

$$\operatorname{sign}(\phi_1) = -\operatorname{sign}(\phi_2) \ .$$

Somit erhalten wir

$$|\phi_1+\phi_2| \leqq \frac{\pi}{2} \ . \tag{3.3.92}$$

Also kann in (3.3.91) nur $k = 0$ sein, und wir haben mit (3.3.5)

$$\begin{aligned} |\phi_1+\phi_2| &\leqq \frac{1}{2}\arcsin\left(6.5\,\frac{\varepsilon}{\delta^3}\right) \\ &\leqq \frac{1}{2}\arcsin\left(6.5\,\frac{1}{1000}\right) \\ &< \frac{\pi}{4} \ . \end{aligned} \tag{3.3.93}$$

Somit ist

$$\cos 2(\phi_1+\phi_2) > 0 \ ,$$

und es gilt

$$\begin{aligned} |\sin(\phi_1+\phi_2)| &= \sqrt{\frac{1-\cos 2(\phi_1+\phi_2)}{2}} \\ &= \sqrt{\frac{1}{2}}\,\sqrt{\frac{1-\cos^2 2(\phi_1+\phi_2)}{1+\cos 2(\phi_1+\phi_2)}} \\ &< \sqrt{\frac{1}{2}}\ |\sin 2(\phi_1+\phi_2)| \\ &< \sqrt{\frac{1}{2}}\ 6.5\,\frac{\varepsilon}{\delta^3} \\ &< 4.6\,\frac{\varepsilon}{\delta^3} \ . \end{aligned} \tag{3.3.94}$$

■

KOROLLAR 3.3.15 : *Es gilt*

$$|\cos(\phi_1+\phi_2) - 1| < 0.03\,\frac{\varepsilon}{\delta^3} \quad .$$

BEWEIS : Nach Korollar 3.3.14 gilt

$$|1-\cos^2(\phi_1+\phi_2)| = |\sin^2(\phi_1+\phi_2)| < 21.16\,\frac{\varepsilon^2}{\delta^6}$$

und damit

$$\Big|1-|\cos(\phi_1+\phi_2)|\Big| < 21.16\,\frac{\varepsilon^2}{\delta^6}\,\frac{1}{1+|\cos(\phi_1+\phi_2)|}$$

$$< 21.16\,\frac{\varepsilon^2}{\delta^6}$$

$$< 21.16\,\frac{\varepsilon}{\delta^3}\,\frac{1}{1000}$$

$$< 0.03\,\frac{\varepsilon}{\delta^3} \quad . \tag{3.3.95}$$

Wegen (3.3.93) gilt $\cos(\phi_1+\phi_2) > 0$ und somit folgt die Behauptung. ■

Bevor wir nun zu der Abschätzung für die Pivot - Elemente $a''_{pq}$ und $a''_{qp}$ kommen, benötigen wir zunächst noch eine weitere Hilfsbemerkung.

BEMERKUNG 3.3.16 : *i) Aus* $|a''_{pq} + \overline{a''_{qp}}| \geq |a''_{pq} - \overline{a''_{qp}}|$ *folgt*

$$|\mathrm{Re}(a''_{pp} - a''_{qq})| > 0.667\,\delta \quad .$$

*ii) Aus* $|a''_{pq} + \overline{a''_{qp}}| \geq |a''_{pq} - \overline{a''_{qp}}|$ *folgt*

$$|\mathrm{Im}(a''_{pp} - a''_{qq})| > 0.667\,\delta \quad .$$

<u>BEWEIS :</u> <u>Zu i) :</u> Nach Bemerkung 1.2.6 ist

$$a'''_{pp}-a'''_{qq} = \cos 2\phi_2(a''_{pp}-a''_{qq}) + \sin 2\phi_2(e^{i\alpha_2}a''_{qp}+e^{-i\alpha_2}a''_{pq})$$

$$= \cos 2\phi_2(a'_{pp}-a'_{qq}) + \sin 2\phi_2(e^{i\alpha_2}a''_{qp}+e^{-i\alpha_2}a''_{pq}) \quad . \tag{3.3.96}$$

Gemäß (3.2.18) ist $e^{i\alpha_2}$ gegeben durch

$$e^{i\alpha_2} = s\,\frac{a''_{pq}+\overline{a''_{qp}}}{|a''_{pq}+\overline{a''_{qp}}|} \quad , \quad s \in \{-1,1\} \quad ,$$

und daher gilt

$$e^{i\alpha_2}a''_{qp} + e^{-i\alpha_2}a''_{pq} = s\,\frac{|a''_{pq}|^2 + |a''_{qp}|^2 + 2a''_{pq}a''_{qp}}{|a''_{pq} + \overline{a''_{qp}}|} \quad . \tag{3.3.97}$$

Damit folgt aus (3.3.96)

$$\mathrm{Re}(a'''_{pp}-a'''_{qq}) = \cos 2\phi_2 \mathrm{Re}(a'_{pp}-a'_{qq}) + s \sin 2\phi_2 |a''_{pq}+\overline{a''_{qp}}| \quad . \tag{3.3.98}$$

Mit Bemerkung 3.3.13, Bemerkung 3.3.3 und (3.3.5) ist

$$|\mathrm{ctg}\, 2\phi_2| < |\mathrm{ctg}\, 2\phi_1| + 6.5\,\frac{\varepsilon}{\delta^3}$$

$$< 0.1 + 6.5\,\frac{1}{1000}$$

$$< 0.11 \tag{3.3.99}$$

und daher

$$\sin^2 2\phi_2 > 0.98 \quad , \tag{3.3.100}$$

$$\cos^2 2\phi_2 < 0.02 \quad . \tag{3.3.101}$$

Wie im Beweis von Lemma 3.3.13 gezeigt (vgl.(3.3.73),(3.3.74)) gilt

$$a''_{pq} + \overline{a''_{qp}} = \sin 2\phi_1 e^{i\alpha_1} \mathrm{Re}(a_{qq}-a_{pp}) + K_1 + K_2 \ , \tag{3.3.102}$$

wobei

$$|K_1 + K_2| \leq 4.21\sqrt{2}\,\frac{\varepsilon}{\delta^2} \ . \tag{3.3.103}$$

Also folgt mit (3.3.100) und Korollar 3.3.4 für den rechten Summanden aus (3.3.98)

$$\begin{aligned} |\sin 2\phi_2|\,|a''_{pq} + a''_{qp}| &> \sqrt{0.98}\,(\sqrt{0.99}\cdot|\mathrm{Re}(a_{qq}-a_{pp})| - 4.21\sqrt{2}\,\frac{\varepsilon}{\delta^2}) \\ &> 0.98\,|\mathrm{Re}(a_{qq}-a_{pp})| - 5.894\,\frac{\varepsilon}{\delta^2} \ . \end{aligned} \tag{3.3.104}$$

Für den linken Summanden aus (3.3.98) gilt gemäß Bemerkung 1.2.6 und (3.3.101) sowie Korollar 3.3.4

$$\begin{aligned} |\cos 2\phi_2|\,|\mathrm{Re}(a'_{pp}-a'_{qq})| &= |\cos 2\phi_2|\,|\cos 2\phi_1 \mathrm{Re}(a_{pp}-a_{qq}) + \sin 2\phi_1 \mathrm{Re}(e^{i\alpha_1}a_{qp} + e^{-i\alpha_1}a_{pq})| \\ &< \sqrt{0.02}\,(0.1\,|\mathrm{Re}(a_{pp}-a_{qq})| + |a_{pq}| + |a_{qp}|) \\ &\leq \frac{\sqrt{2}}{100}\,|\mathrm{Re}(a_{pp}-a_{qq})| + \frac{\sqrt{2}}{10}\sqrt{2}\,\varepsilon \ . \end{aligned} \tag{3.3.105}$$

Somit erhalten wir mit (3.3.104),(3.3.105) aus (3.3.98)

$$|\mathrm{Re}(a'''_{pp}-a'''_{qq})| > (0.98 - \frac{\sqrt{2}}{100})\,|\mathrm{Re}(a_{pp}-a_{qq})| - (5.894 + \frac{2}{10}\,2)\,\frac{\varepsilon}{\delta^2} \ ,$$

und mit (3.3.5) und Bemerkung 3.3.11 (i) folgt

$$(3.3.106) \qquad |\mathrm{Re}(a'''_{pp}-a'''_{qq})| > (0.98 - \frac{\sqrt{2}}{100})0.7\delta - (5.894 + \frac{4}{10})\frac{\sqrt{2}\,\delta}{1000}$$

$$> 0.667\,\delta \ .$$

<u>Zu ii) :</u> In diesem Fall ist wegen (3.2.20)

$$(3.3.107) \qquad e^{i\alpha_2} = si\,\frac{a''_{pq} - \overline{a''_{qp}}}{|a''_{pq} - \overline{a''_{qp}}|} \ , \qquad s \in \{-1,1\} \ .$$

Daraus folgt

$$(3.3.108) \qquad e^{i\alpha_2}a''_{qp} + e^{-i\alpha_2}a''_{pq} = -\,si\,\left(\frac{|a''_{pq}|^2 + |a''_{qp}|^2 - 2a''_{pq}a''_{qp}}{|a''_{pq} - \overline{a''_{qp}}|}\right)$$

und somit

$$\mathrm{Im}(e^{i\alpha_2}a''_{qp} + e^{-i\alpha_2}a''_{pq}) = -\,s\,\frac{|a''_{pq}|^2 + |a''_{qp}|^2 - 2\mathrm{Re}(a''_{pq}a''_{qp})}{|a''_{pq} - \overline{a''_{qp}}|}$$

$$(3.3.109) \qquad = -\,s\,|a''_{pq} - \overline{a''_{qp}}| \ .$$

Damit folgt aus (3.3.96)

$$(3.3.110) \qquad \mathrm{Im}(a'''_{pp}-a'''_{qq}) = \cos 2\phi_2\,\mathrm{Im}(a'_{pp}-a'_{qq}) - s\sin 2\phi_2|a''_{pq}-\overline{a''_{qp}}|$$

Die Behauptung folgt nun mit Bemerkung 3.3.11 (ii) analog zum Beweis im Fall I . ■

Wir sind jetzt in der Lage, das folgende Lemma über die Größenordnung der Pivot-Elemente $a_{pq}$ und $a_{qp}$ nach Ausführung eines kompletten Iterationsschrittes zu beweisen.

<u>LEMMA 3.3.17 :</u> *Es gilt*

$$|a''_{pq}| = |a''_{qp}| < 3.8 \frac{\varepsilon^2}{\delta^3} \quad .$$

<u>BEWEIS :</u> Wir definieren zunächst 2 x 2 - Matrizen $A''_{pq}$ , $A''^{+}_{pq}$, $A''^{-}_{pq}$ und $C''_{pq}$ gemäß

(3.3.111) $$A''_{pq} = \begin{vmatrix} a''_{pp} & a''_{pq} \\ a''_{qp} & a''_{qq} \end{vmatrix} ,$$

(3.3.112) $$A''^{+}_{pq} := A''_{pq} + A''^{*}_{pq} \quad ,$$

(3.3.113) $$A''^{-}_{pq} := A''_{pq} - A''^{*}_{pq} \quad ,$$

(3.3.114) $$C''_{pq} := C(A''_{pq})$$

$$= A''^{*}_{pq} A''_{pq} - A''_{pq} A''^{*}_{pq} \quad .$$

Dann gilt

(3.3.115) $$A''_{pq} = \frac{1}{2} ( A''^{+}_{pq} + A''^{-}_{pq})$$

sowie

$$(3.3.116)\qquad \|C''_{pq}\|^2 = 2\left(|a''_{qp}|^2 - |a''_{pq}|^2\right)^2 + 2\,|a''_{pq}\overline{(a''_{pp}-a''_{qq})} - \overline{a''_{qp}}(a''_{pp}-a''_{qq})|^2 .$$

Nun gilt

$$(3.3.117)\qquad a''_{pq}\overline{(a''_{pp}-a''_{qq})} - \overline{a''_{qp}}(a''_{pp}-a''_{qq}) = a'_{pq}\overline{(a'_{pp}-a'_{qq})} - \overline{a'_{qp}}(a'_{pp}-a'_{qq}) + R ,$$

wobei

$$(3.3.118)\qquad R := \begin{cases} (\frac{1}{x}-1)a'_{pq}\overline{(a'_{pp}-a'_{qq})} - (x-1)\overline{a'_{qp}}(a'_{pp}-a'_{qq}), & |c'_{pp}| \geq |c'_{qq}| \\ (x-1)a'_{pq}\overline{(a'_{pp}-a'_{qq})} - (\frac{1}{x}-1)\overline{a'_{qp}}(a'_{pp}-a'_{qq}), & |c'_{pp}| < |c'_{qq}| \end{cases}$$

Mit Hilfe von Lemma 3.3.9 erhalten wir für $|R|$ die folgenden Abschätzungen.

i) Gilt $|a'_{pp}-a'_{qq}| \leq \varepsilon$ , so folgt aus Lemma 3.3.9 (i)

$$|R| < 2.21\,\frac{\varepsilon}{\delta^2}\,\varepsilon\,(|a'_{pq}| + |a'_{qp}|)$$

$$(3.3.119)\qquad \leq 2.21\,\sqrt{2}\,\frac{\varepsilon^2}{\delta^2} .$$

ii) Gilt dagegen $|a'_{pp}-a'_{qq}| > \varepsilon$ , so erhalten wir mit Lemma 3.3.9(ii)

$$|R| < \frac{2.31\,\varepsilon^2}{\delta^2\,|a'_{pp}-a'_{qq}|}\,|a'_{pp}-a'_{qq}|\,(|a'_{pq}| + |a'_{qp}|)$$

$$(3.3.120)\qquad \leq \sqrt{2}\;2.31\,\frac{\varepsilon^2}{\delta^2} .$$

Somit können wir festhalten, daß gilt

$$|R| < 3.3\,\frac{\varepsilon^2}{\delta^2} \quad . \tag{3.3.121}$$

Zum anderen gilt nach Durchführung der vorbereitenden Jacobi-Rotation

$$0 = c'_{pq} = a'_{pq}\overline{(a'_{pp}-a'_{qq})} - \overline{a'_{qp}}(a'_{pp}-a'_{qq}) + \sum_{i\neq p,q}(\overline{a'_{ip}}a'_{iq}-a'_{pi}\overline{a'_{qi}})$$

und damit mit Bemerkung 1.2.8

$$|a'_{pq}\overline{(a'_{pp}-a'_{qq})} - \overline{a'_{qp}}(a'_{pp}-a'_{qq})| \leq \frac{1}{2}\varepsilon^2 \, . \tag{3.3.122}$$

Damit erhalten wir aus (3.3.117) und (3.3.121)

$$\begin{aligned} |a''_{pq}\overline{(a''_{pp}-a''_{qq})} - \overline{a''_{qp}}(a''_{pp}-a''_{qq})| &< \frac{1}{2}\varepsilon^2 + 3.3\,\frac{\varepsilon^2}{\delta^2} \\ &\leq (\frac{1}{2}2 + 3.3)\frac{\varepsilon^2}{\delta^2} \\ &= 4.3\frac{\varepsilon^2}{\delta^2} \quad . \end{aligned} \tag{3.3.123}$$

Für die Norm des Kommutators $C''_{pq}$ folgt dann mit Lemma 3.3.10 aus (3.3.116) die Abschätzung

$$\begin{aligned} \|C''_{pq}\|^2 &< 2(1.6\frac{\varepsilon^2}{\delta^2})^2 + 2(4.3\frac{\varepsilon^2}{\delta^2})^2 \\ &< (7\frac{\varepsilon^2}{\delta^2})^2 \, . \end{aligned} \tag{3.3.124}$$

Definieren wir die $2 \times 2$ - Matrizen $A_{pq}'''$ , $A_{pq}'''^{+}$ , $A_{pq}'''^{-}$ und $C_{pq}'''$ analog zu (3.3.111) bis (3.3.114) , so gilt dann auch

$$(3.3.125) \qquad \|C_{pq}'''\|^2 < \left(7\frac{\varepsilon^2}{\delta^2}\right)^2 ,$$

da die Norm des Kommutators $C_{pq}''' = C(A_{pq}''')$ unter der zweiten unitären Jacobi-Rotation erhalten bleibt.

Zur Abschätzung von $|a_{pq}'''|$ und $|a_{qp}'''|$ betrachten wir nun zwei Fälle.

<u>FALL I) :</u> Es sei $|a_{pq}'' + \overline{a_{qp}''}| \geq |a_{pq}'' - \overline{a_{qp}''}|$ .

Dann gilt nach Bemerkung 3.2.1

$$a_{pq}''' + \overline{a_{qp}'''} = 0 ,$$

und somit hat $A_{pq}'''^{+}$ die Gestalt

$$(3.3.126) \qquad A_{pq}'''^{+} = 2 \begin{pmatrix} \mathrm{Re}(a_{pp}''') & 0 \\ 0 & \mathrm{Re}(a_{qq}''') \end{pmatrix} .$$

Für $A_{pq}'''^{-}$ gilt

$$(3.3.127) \qquad A_{pq}'''^{-} = \begin{pmatrix} 2i\mathrm{Im}(a_{pp}''') & 2a_{pq}''' \\ -2\overline{a_{pq}'''} & 2i\mathrm{Im}(a_{qq}''') \end{pmatrix} .$$

Wegen

$$C''_{pq} = \frac{1}{2}(A''^{+}_{pq} A''^{-}_{pq} - A''^{-}_{pq} A''^{+}_{pq}) \tag{3.3.128}$$

gilt dann

$$C''_{pq} = \begin{pmatrix} 0 & 2a''_{pq}\mathrm{Re}(a''_{pp}-a''_{qq}) \\ 2\overline{a''_{pq}}\mathrm{Re}(a''_{pp}-a''_{qq}) & 0 \end{pmatrix} \tag{3.3.129}$$

und damit wegen (3.3.125)

$$\begin{aligned} \|C''_{pq}\|^2 &= 8|a''_{pq}|^2|\mathrm{Re}(a''_{pp}-a''_{qq})|^2 \\ &< (7\frac{\varepsilon^2}{\delta^2})^2 \quad . \end{aligned} \tag{3.3.130}$$

Mit Bemerkung 3.3.16 i) folgt hieraus wegen $|a''_{pq}| = |a''_{qp}|$ schließlich

$$\begin{aligned} |a''_{pq}| = |a''_{qp}| &< \sqrt{\frac{1}{8}}\, 7\, \frac{\varepsilon^2}{\delta^2|\mathrm{Re}(a''_{pp}-a''_{qq})|} \\ &< \sqrt{\frac{1}{8}}\, 7\, \frac{\varepsilon^2}{\delta^2\, 0.667\cdot\delta} \\ &< 3.8\, \frac{\varepsilon^2}{\delta^3} \quad . \end{aligned}$$

<u>FALL II) :</u> Sei nun $|a''_{pq} + \overline{a''_{qp}}| < |a''_{pq} - \overline{a''_{qp}}|$ .

Dann gilt nach Bemerkung 3.2.1

$$a'''_{pq} - \overline{a'''_{qp}} = 0 \ , \tag{3.3.131}$$

und die Matrizen $A_{pq}$ , $A'''^{-}_{pq}$ und $C'''_{pq}$ lauten

$$A'''^{+}_{pq} = 2 \begin{pmatrix} \mathrm{Re}(a'''_{pp}) & a'''_{pq} \\ \overline{a'''_{pq}} & \mathrm{Re}(a'''_{qq}) \end{pmatrix} \ , \tag{3.3.132}$$

$$A'''^{-}_{pq} = 2i \begin{pmatrix} \mathrm{Im}(a'''_{pp}) & 0 \\ 0 & \mathrm{Im}(a'''_{qq}) \end{pmatrix} \ , \tag{3.3.133}$$

$$C'''_{pq} = \begin{pmatrix} 0 & 2ia'''_{pq}\mathrm{Im}(a'''_{pp}-a'''_{qq}) \\ -2i\overline{a'''_{pq}}\mathrm{Im}(a'''_{pp}-a'''_{qq}) & 0 \end{pmatrix} \ . \tag{3.3.134}$$

Der Rest des Beweises verläuft dann mit Hilfe von Bemerkung 3.3.16 ii) analog zu Fall I) . ■

Bis hierher hatten wir nach unserer Voraussetzung für den Fall A) nur die Situation untersucht, in der zu Beginn des betrachteten Iterationsschrittes für das Pivotelement $c_{pq}$ des Kommutators galt $|c_{pq}| > 5\varepsilon^2$. Kommen wir daher nun zu

FALL B : Es gelte $|c_{pq}| \leq 5\,\varepsilon^2$ .

Nach Schritt 1 unseres Verfahrens wird in diesem Fall nur die zweite Jacobi - Rotation mit Parametern $\alpha_2$ und $\phi_2$ ausgeführt. Wir erhalten zunächst die

BEMERKUNG 3.3.18 : *Es gilt*

$$\operatorname{ctg}^2 2\phi_2 \geq 6.2\cdot 10^4 .$$

BEWEIS : Wir beweisen nur den Fall, in dem gilt

$$\mathrm{Re}^2(a_{pp}-a_{qq}) \geq \mathrm{Im}^2(a_{pp}-a_{qq}) \; ; \tag{3.3.135}$$

anderenfalls verläuft der Beweis wieder analog.
Sei also (3.3.135) erfüllt. Aufgrund von Bemerkung 3.3.2 gilt

$$\mathrm{Re}^2(a_{pp}-a_{qq}) + \mathrm{Im}^2(a_{pp}-a_{qq}) \geq (\delta - 2\varepsilon)^2 , \tag{3.3.136}$$

und mit (3.3.5) und (3.3.135) folgt

$$\begin{aligned} \mathrm{Re}^2(a_{pp}-a_{qq}) &\geq \frac{1}{2}\left(\delta - \frac{2}{500}\delta\right)^2 \\ &= 0.496\,\delta^2 . \end{aligned} \tag{3.3.137}$$

Für den durch (3.2.13) definierten Rotationswinkel $\phi_2$ gilt dann

$$\operatorname{ctg}^2 2\phi_2 = \frac{\mathrm{Re}^2(a_{pp}-a_{qq})}{|a_{pq}+\overline{a_{qp}}|^2}$$

$$\geqq \frac{0.496\,\delta^2}{(\sqrt{2}\,\varepsilon)^2} \tag{3.3.138}$$

$$= 0.248\,\frac{\delta^2}{\varepsilon^2} \quad ,$$

und mit (3.3.5) folgt

$$\operatorname{ctg}^2 2\phi_2 \geqq 0.248\,\delta^2 \frac{10^6}{4\delta^2}$$

$$= 6.2\cdot 10^4 \quad . \qquad \blacksquare$$

Aus dieser Bemerkung erhalten wir sofort das

<u>KOROLLAR 3.3.19 :</u> *Es gilt*

$$\sin^2 2\phi_2 < 2\cdot 10^{-5} \quad ,$$

$$\cos^2 2\phi_2 > 1 - 2\cdot 10^{-5} \quad . \qquad \blacksquare$$

Für die Pivotlemente $\tilde{a}_{pq}$ und $\tilde{a}_{qp}$ der durch die zweite Jacobi - Rotation transformierten Matrix $\tilde{A}$ erhalten wir nun das folgende Lemma.

<u>LEMMA 3.3.20 :</u> *Sei $\tilde{A}$ die Matrix A nach Durchführung eines kompletten Iterationszyklus mit Pivotindexpaar* (p,q) *unter der Voraussetzung*

$$|c_{pq}| \leq 5\,\tau^2(A) = 5\,\varepsilon^2 .$$

*Dann gilt*

$$|\tilde{a}_{pq}| = |\tilde{a}_{qp}| < 6\,\frac{\varepsilon^2}{\delta^3} \quad .$$

<u>BEWEIS :</u> Wir führen zunächst die folgenden Bezeichnungen ein. Es sei

(3.3.138) $$A_{pq} := \begin{pmatrix} a_{pp} & a_{pq} \\ a_{qp} & a_{qq} \end{pmatrix} \quad ,$$

(3.3.139) $$\tilde{A}_{pq} := \begin{pmatrix} \tilde{a}_{pp} & \tilde{a}_{pq} \\ \tilde{a}_{qp} & \tilde{a}_{qq} \end{pmatrix} \quad ,$$

(3.3.140) $$C_{pq} := C(A_{pq}) \ ,$$

(3.3.141) $$\tilde{C}_{pq} := C(\tilde{A}_{pq}) \quad .$$

Wegen

$$\begin{aligned} 5\,\varepsilon^2 &\geq |c_{pq}| \\ &= \Big|a_{pq}\overline{(a_{pp}-a_{qq})} - \overline{a_{qp}}(a_{pp}-a_{qq}) + \sum_{i \neq p,q} (\overline{a_{ip}}a_{iq} - a_{pi}\overline{a_{qi}})\Big| \end{aligned}$$

gilt zunächst

$$(3.3.142)\quad |a_{pq}\overline{(a_{pp}-a_{qq})} - \overline{a_{qp}}(a_{pp}-a_{qq})| \leq \frac{11}{2}\varepsilon^2,$$

so daß für den Kommutator $C_{pq}$ folgt

$$\|C_{pq}\|^2 = 2(|a_{pq}|^2 - |a_{qp}|^2)^2 + 2|a_{pq}\overline{(a_{pp}-a_{qq})} - \overline{a_{qp}}(a_{pp}-a_{qq})|^2$$

$$\leq 2\,\varepsilon^4 + 2(\frac{11}{2}\,\varepsilon^2)^2$$

$$(3.3.143)\qquad < (8\,\varepsilon^2)^2 \quad .$$

Nach Durchführung des Iterationsschrittes hat wegen $|\tilde{a}_{pq}| = |\tilde{a}_{qp}|$ der Kommutator $\tilde{C}_{pq}$ die Gestalt

$$(3.3.144)\qquad \tilde{C}_{pq} = \begin{pmatrix} 0 & \tilde{c}_{pq} \\ \overline{\tilde{c}_{pq}} & 0 \end{pmatrix} \quad ,$$

wobei

$$(3.3.145)\qquad |\tilde{c}_{pq}| = 2 \begin{cases} |\tilde{a}_{pq}\mathrm{Re}(\tilde{a}_{pp}-\tilde{a}_{qq})| \,, & \mathrm{Re}^2(a_{pp}-a_{qq}) \geq \mathrm{Im}^2(a_{pp}-a_{qq}) \,, \\ |\tilde{a}_{pq}\mathrm{Im}(\tilde{a}_{pp}-\tilde{a}_{qq})| \,, & \text{sonst.} \end{cases}$$

Da die Norm des Kommutators unter unitären Ähnlichkeitstransformationen invariant ist, folgt aus (3.3.143)

$$(3.3.146)\qquad 2|\tilde{c}_{pq}| < 8\,\varepsilon^2 \quad .$$

Betrachten wir zunächst den Fall

$$(3.3.147)\qquad \mathrm{Re}^2(a_{pp}-a_{qq}) \geq \mathrm{Im}^2(a_{pp}-a_{qq}) \quad .$$

Wegen Bemerkung 3.3.2 gilt

$$\mathrm{Re}^2(a_{pp}-a_{qq}) + \mathrm{Im}^2(a_{pp}-a_{qq}) \geq (\delta - 2\varepsilon)^2 \quad , \tag{3.3.148}$$

so daß aus (3.3.147) mit (3.3.5) folgt

$$\mathrm{Re}^2(a_{pp}-a_{qq}) \geq \frac{1}{2}\,(\delta - 2\frac{2}{1000}\delta)^2$$

$$> 0.496\,\delta^2 \,. \tag{3.3.149}$$

Mit Bemerkung 1.2.6 erhalten wir dann hiermit unter Verwendung von Korollar 3.3.19

$$|\mathrm{Re}(\tilde{a}_{qq}-\tilde{a}_{pp})| = |\mathrm{Re}\{\cos 2\phi_2\,(a_{pp}-a_{qq}) + \sin 2\phi_2(e^{i\alpha_2}a_{qp} + e^{-i\alpha_2}a_{pq})\}|$$

$$> \sqrt{1-2\cdot 10^{-5}}\,\sqrt{0.496}\,\delta - \sqrt{2\cdot 10^{-5}}\,\sqrt{2}\,\varepsilon$$

$$\geq (\sqrt{1-2\cdot 10^{-5}}\,\sqrt{0.496} - \sqrt{2\cdot 10^{-5}}\,\sqrt{2}\,\frac{1}{500})\,\delta$$

$$> 0.7\delta \,. \tag{3.3.150}$$

Damit folgt aus (3.3.145) und (3.3.146)

$$|\tilde{a}_{pq}| = |\tilde{a}_{qp}| = \frac{1}{2}\,\frac{|\tilde{c}_{pq}|}{|\mathrm{Re}(\tilde{a}_{pp}-\tilde{a}_{qq})|}$$

$$< 2\,\frac{\varepsilon^2}{0.7\,\delta}$$

$$< 2\,\frac{\varepsilon^2}{0.7\,\delta}\,\frac{2}{\delta^2}$$

$$< 6\,\frac{\varepsilon^2}{\delta^3} \,. \tag{3.3.151}$$

Nun zu dem Fall, in dem gilt

$$\text{(3.3.152)} \qquad \mathrm{Re}^2(a_{pp}-a_{qq}) < \mathrm{Im}^2(a_{pp}-a_{qq}) \ .$$

Analog zu (3.3.148) erhalten wir dann

$$\text{(3.3.153)} \qquad \mathrm{Im}^2(a_{pp}-a_{qq}) > 0.496\,\delta^2 \ ,$$

und mit Hilfe von Bemerkung 1.2.6 erhalten wir aus (3.3.145) und (3.3.146) wieder die Behauptung. ■

Wir können an dieser Stelle festhalten, daß nach Lemma 3.3.17 und Lemma 3.3.20 nach Durchführung eines Iterationsschrittes unseres Verfahrens die Pivotelemente $a_{pq}$ und $a_{qp}$ der iterierten Matrix $A$ klein sind von der Ordnung $\frac{\varepsilon^2}{\delta^3}$, wenn die Nebendiagonalelemente von $A$ vor diesem Iterationsschritt klein waren von der Ordnung $\varepsilon$.

Bevor wir nun den Satz über die asymptotisch quadratische Konvergenz unseres Verfahrens formulieren und beweisen können, benötigen wir zunächst noch eine Abschätzung über die Abweichung der Transformationsmatrix $T$ eines vollen Iterationsschrittes und ihrer Inversen $T^{-1}$ von der Einheitsmatrix. Hierzu dient das folgende Lemma.

<u>LEMMA 3.3.21 :</u> *Es sei die Voraussetzung*

$$\varepsilon \leqq \left(\frac{\delta}{10}\right)^3$$

*erfüllt. Für* $|c_{pq}| > 5\varepsilon^2$ *seien* $r, s \in \mathbb{R}$ *definiert durch*

$$r := \begin{cases} x & , \text{ falls } |c'_{pp}| \geqq |c'_{qq}| , \\ 1 & , \text{ sonst} , \end{cases}$$

$$s := \begin{cases} 1 & , \text{ falls } |c'_{pp}| \geqq |c'_{qq}| , \\ x & , \text{ sonst;} \end{cases}$$

*für* $|c_{pq}| \leqq 5\varepsilon^2$ *sei*

$$r := s := 1 \quad ,$$

$$\phi_1 := 0 \quad .$$

*Schließlich seien die* $2 \times 2$ *- Matrizen* $T_{pq}$ *und* $\tilde{T}_{pq}$ *definiert durch*

$$T_{pq} := \begin{pmatrix} t_{pp} & t_{pq} \\ t_{qp} & t_{qq} \end{pmatrix}$$

$$:= \begin{pmatrix} \cos\phi_1 & -e^{i\alpha_1}\sin\phi_1 \\ e^{-i\alpha_1}\sin\phi_1 & \cos\phi_1 \end{pmatrix} \begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix} \begin{pmatrix} \cos\phi_2 & -e^{i\alpha_2}\sin\phi_2 \\ e^{-i\alpha_2}\sin\phi_2 & \cos\phi_2 \end{pmatrix}$$

*sowie*

$$\tilde{T}_{pq} := T_{pq}^{-1} = \begin{pmatrix} \tilde{t}_{pp} & \tilde{t}_{pq} \\ \tilde{t}_{qp} & \tilde{t}_{qq} \end{pmatrix} \quad .$$

*Ist dann*

$$t_D \in \{t_{pp}, t_{qq}, \tilde{t}_{pp}, \tilde{t}_{qq}\} ,$$

$$t_N \in \{t_{pq}, t_{qp}, \tilde{t}_{pq}, \tilde{t}_{qp}\} ,$$

*so gilt*

$$|t_D - 1| < 30 \frac{\varepsilon}{\delta^3} , \qquad |t_N| < 34 \frac{\varepsilon}{\delta^3} .$$

<u>BEWEIS :</u> Wir betrachten zunächst den Fall

$$|c_{pq}| > 5 \varepsilon^2 .$$

Dann ist

$$t_{pp} = r \cos \phi_1 \cos \phi_2 - s e^{i\alpha_1} e^{-i\alpha_2} \sin \phi_1 \sin \phi_2 , \tag{3.3.154}$$

$$t_{qq} = s \cos \phi_1 \cos \phi_2 - r e^{-i\alpha_1} e^{i\alpha_2} \sin \phi_1 \sin \phi_2 , \tag{3.3.155}$$

$$t_{pq} = -r e^{i\alpha_2} \cos \phi_1 \sin \phi_2 - s e^{i\alpha_1} \sin \phi_1 \cos \phi_2 , \tag{3.3.156}$$

$$t_{qp} = r e^{-i\alpha_1} \cos \phi_2 \sin \phi_1 + s e^{-i\alpha_2} \sin \phi_2 \cos \phi_1 . \tag{3.3.157}$$

Gleichung (3.3.154) können wir nun schreiben in der Form

$$\begin{aligned} t_{pp} = {} & \cos \phi_1 \cos \phi_2 - \sin \phi_1 \sin \phi_2 + \cos \phi_1 \cos \phi_2 (r - 1) - \\ & - \sin \phi_1 \sin \phi_2 (s - 1 + s e^{i\alpha_1} (e^{-i\alpha_2} - e^{-i\alpha_1})). \end{aligned} \tag{3.3.158}$$

Mit (3.3.5), Lemma 3.3.9, Lemma 3.3.12 und Korollar 3.3.15 folgt dann hieraus

$$|t_{pp}-1| \leq |\cos(\phi_1+\phi_2)-1| + |r-1| + |s-1| + |s(e^{-i\alpha_2}-e^{-i\alpha_1})|$$

$$< 0.03\,\frac{\varepsilon}{\delta^3} + 2.21\,\frac{\varepsilon}{\delta^2} + (1+2.21\,\frac{\varepsilon}{\delta^2})\,26\frac{\varepsilon}{\delta^3}$$

$$< (0.03+2.21\,\sqrt{2} + 26\,(1+2.21\,\frac{\sqrt{2}}{1000})\,)\,\frac{\varepsilon}{\delta^3}$$

(3.3.159) $$< 30\,\frac{\varepsilon}{\delta^3} \quad .$$

Ganz analog erhält man aus (3.3.155) die Abschätzung

(3.3.160) $$|t_{qq}-1| < 30\,\frac{\varepsilon}{\delta^3} \quad .$$

Nun zu $t_{pq}$. Wir schreiben (3.3.156) in der Form

(3.3.161) $$-t_{pq} = e^{i\alpha_1}(s\sin(\phi_1+\phi_2) - \cos\phi_1\sin\phi_2(s-r)\,) - \\ - r\cos\phi_1\sin\phi_2(e^{i\alpha_1}-e^{i\alpha_2}) \quad .$$

Wegen $s=1$ oder $r=1$ ist entweder $|s-r| = |1-r|$ oder $|s-r| = |1-s|$; mit Lemma 3.3.9, Lemma 3.3.12, Korollar 3.3.14 und (3.3.5) folgt dann aus (3.3.161)

$$|t_{pq}| < (1+\,2.21\frac{\varepsilon}{\delta^2})\,4.6\frac{\varepsilon}{\delta^3} + 2.21\frac{\varepsilon}{\delta^2} + (1+2.21\frac{\varepsilon}{\delta^2})\,26\frac{\varepsilon}{\delta^3}$$

$$\leq (\,(1+2.21\frac{\sqrt{2}}{1000})\,30.6 + 2.21\,\sqrt{2}\,)\,\frac{\varepsilon}{\delta^3}$$

(3.3.162) $$< 34\,\frac{\varepsilon}{\delta^3} \quad .$$

Analog erhält man aus (3.3.157) die Abschätzung

$$(3.3.163) \qquad |t_{qp}| < 34 \frac{\varepsilon}{\delta^3} \quad .$$

Die Abschätzungen für die Elemente der Inversen $T_{pq}^{-1}$ gewinnt man ganz analog aus den (3.3.154) bis (3.3.157) entsprechenden Formeln für $\tilde{t}_{pp}$, $\tilde{t}_{pq}$, $\tilde{t}_{qp}$ und $\tilde{t}_{qq}$.

Wir haben nun noch den Fall

$$|c_{pq}| \leqq 5\,\varepsilon^2$$

zu betrachten. In diesem Fall gilt

$$(3.3.164) \qquad t_{pp} = t_{qq} = \tilde{t}_{pp} = \tilde{t}_{qq} = \cos\phi_2 \quad ,$$

$$(3.3.165) \qquad t_{pq} = -\overline{t_{qp}} = \overline{\tilde{t}_{qp}} = -\tilde{t}_{pq} = -e^{i\alpha_2}\sin\phi_2 \; .$$

Nach (3.3.138) ist

$$\operatorname{ctg}^2 2\phi_2 \geqq 0.248 \frac{\delta^2}{\varepsilon^2} > \frac{\delta^2}{5\,\varepsilon^2}$$

und somit

$$1 - \cos^2 2\phi_2 = \sin^2 2\phi_2 < \frac{5\,\varepsilon^2}{\delta^2} \; .$$

Wegen $|\phi_2| \leqq \frac{\pi}{4}$ ist $\cos 2\phi_2 \geqq 0$, und wir erhalten aus (3.3.165)

$$1 - \cos 2\phi_2 = \frac{1 - \cos^2 2\phi_2}{1 + \cos 2\phi_2}$$

$$\leqq 1 - \cos^2 2\phi_2$$

(3.3.166) $$< 5 \frac{\varepsilon^2}{\delta^2} \quad .$$

Hieraus wiederum folgt

$$\sin^2 \phi_2 = \frac{1 - \cos 2\phi_2}{2}$$

(3.3.167) $$< 2.5 \frac{\varepsilon^2}{\delta^2}$$

und wegen $\cos \phi_2 > 0$ damit auch

$$1 - \cos \phi_2 = \frac{1 - \cos^2 \phi_2}{1 + \cos \phi_2}$$

$$< \frac{\sin^2 \phi_2}{1}$$

(3.3.168) $$< 2.5 \frac{\varepsilon^2}{\delta^2} \quad .$$

Mit (3.3.167) und (3.3.168) erhalten wir nun die behaupteten Abschätzungen aus (3.3.164) und (3.3.165), da

$$2.5 \frac{\varepsilon^2}{\delta^2} < 30 \frac{\varepsilon}{\delta^3}$$

und

$$\sqrt{2.5} \frac{\varepsilon}{\delta} < 34 \frac{\varepsilon}{\delta^3} \quad .$$

■

Wir formulieren nun unseren Satz über asymptotisch quadratische Konvergenz unseres Verfahrens im Falle getrennter Eigenwerte.

<u>SATZ 3.3.22 :</u> *Gegeben sei eine* $n \times n$ *- Matrix* $A$ , $n \geqq 3$, *für deren Eigenwerte* $\lambda_i$ , $i = 1,2, \ldots , n$ *gilt*

$$(3.3.169) \qquad \delta := \min_{i \neq j} |\lambda_i - \lambda_j| > 0 \ .$$

*Weiter gelte*

$$(3.3.170) \qquad \|A\| \leqq 1$$

*sowie*

$$(3.3.171) \qquad \tau^2(A) \leqq 1.14^{-\frac{1}{N}} \left(\frac{\delta}{10}\right)^6 \ , \quad N := \frac{n(n-1)}{2} \ .$$

$\hat{A}$ *bezeichne die Matrix* $A$ *nach Durchführung eines kompletten Iterationszyklus unseres in Abschnitt 3.2 beschriebenen Verfahrens.*

*Dann existiert eine nur von* $n$ *und* $\delta$ *abhängige Konstante* $K(\delta,n) > 0$ , *so daß gilt*

$$\tau^2(\hat{A}) \leqq K(\delta,n) \ (\tau^2(A))^2 \ .$$

<u>BEWEIS*:</u> Wir bezeichnen mit $A_i$ , $i = 0,1,2, \ldots , N$ , die Matrix $A$ nach Durchführung des i-ten Iterationsschrittes unseres Verfahrens. Damit ist

$$(3.3.172) \qquad A_0 = A \ , \quad A_N = \hat{A} \ .$$

---

*) Wir führen diesen Beweis in Anlehnung an den Beweis der asymptotisch quadratischen Konvergenz des Verfahrens von Eberlein von RUHE [13].

Weiter sei $\varepsilon_i$ definiert durch

$$\varepsilon_i := \tau^2(A_i) \quad , \; i = 0,1,2, \ldots , N \; . \tag{3.3.173}$$

Wir beweisen zunächst die

<u>Zwischenbehauptung 1 :</u> Es gilt

$$\varepsilon_{i+1}^2 \leqq 1.14\, \varepsilon_i^2 \; , \; i = 0,1,2, \ldots , N-1 \; .$$

<u>Zwischen - Beweis :</u> Aufgrund der Voraussetzung (3.3.171) unseres Satzes ist für $\varepsilon_0$ sicher die Voraussetzung 3.3.1 erfüllt. Es gilt

$$\begin{aligned} \varepsilon_1^2 - \varepsilon_0^2 = & \sum_{i \neq p,q} (|a_{ip}^{(1)}|^2 + |a_{iq}^{(1)}|^2 + |a_{pi}^{(1)}|^2 + |a_{qi}^{(1)}|^2) \\ & + |a_{pq}^{(1)}|^2 + |a_{qp}^{(1)}|^2 \\ & - \sum_{i \neq p,q} (|a_{ip}|^2 + |a_{iq}|^2 + |a_{pi}|^2 + |a_{qi}|^2) \\ & - |a_{pq}|^2 + |a_{qp}|^2 \quad . \end{aligned} \tag{3.3.174}$$

Mit Bemerkung 1.2.6 und den Abschätzungen für die nicht - trivialen Elemente der Transformationsmatrix und ihrer Inversen aus Lemma 3.3.21 erhalten wir dann aus (3.3.174)

$$\begin{aligned} \varepsilon_1^2 - \varepsilon_0^2 < & \Big( \sum_{i \neq p,q} (|a_{pi}|^2 + |a_{qi}|^2 + |a_{ip}|^2 + |a_{iq}|^2) \Big) \Big( (1.03)^2 + (\tfrac{34}{1000})^2 - 1 \Big) \\ & + 2 \Big( \sum_{i \neq p,q} |a_{pi}a_{qi}| + |a_{ip}a_{iq}| \Big) \, 1.03 \, \frac{34}{1000} + |a_{pq}^{(1)}|^2 + |a_{qp}^{(1)}|^2 \end{aligned}$$

$$\leq \Big(\sum_{i \neq p,q} (|a_{pi}|^2 + |a_{qi}|^2 + |a_{ip}|^2 + |a_{ip}|^2)\Big)\Big((1.03 + \frac{34}{1000})^2 - 1\Big)$$

$$+ |a_{pq}^{(1)}|^2 + |a_{qp}^{(1)}|^2$$

(3.3.175) $$< 0.133\,\varepsilon_0^2 + |a_{pq}^{(1)}|^2 + |a_{qp}^{(1)}|^2 \quad ,$$

und mit Lemma 3.3.17, Lemma 3.3.20 und (3.3.5) folgt

$$\varepsilon_1^2 - \varepsilon_0^2 < 0.133\,\varepsilon_0^2 + 2\ 36\,\frac{\varepsilon_0^2}{\delta^6}\,\varepsilon_0^2$$

$$\leq (0.133 + \frac{72}{10^6})\,\varepsilon_0^2$$

$$< 0.14\,\varepsilon_0^2 \quad .$$

Wegen (3.3.171) ist somit auch für $\varepsilon_1$ wieder die Voraussetzung 3.3.1 erfüllt; mittels vollständiger Induktion folgt dann die Zwischenbehauptung 1.

Halten wir an dieser Stelle fest, daß wir aufgrund von Voraussetzung (3.3.171) damit gleichzeitig gezeigt haben

(3.3.176) $$\varepsilon_i \leq (\frac{\delta}{10})^3 \quad , \ i = 0,1,2,\ldots,N \ ,$$

so daß wegen (3.3.169) und (3.3.170) für jeden Iterationsschritt des betrachteten Zyklus' sämtliche in diesem Abschnitt bisher bewiesenen Bemerkungen und Lemmata anwendbar sind.

Als nächstes benötigen wir eine Abschätzung dafür, um wieviel bereits einmal als Pivot verwendete Elemente während der weiteren Transformationen des Iterationszyklus' wieder anwachsen können. Dazu dient die

<u>Zwischenbehauptung 2 :</u> Für $j \neq p,q$ und $i = 0,1,2,\ldots,N$ gilt

$$|a_{pj}^{(i+1)}| < 1.03|a_{pj}^{(i)}| + u^{(i)}\,|a_{qj}^{(i)}| \quad , \tag{3.3.177}$$

$$|a_{jp}^{(i+1)}| < 1.03|a_{jp}^{(i)}| + u^{(i)}\,|a_{jq}^{(i)}| \quad , \tag{3.3.178}$$

wobei

$$0 \leqq u^{(i)} < 34\,\frac{\varepsilon_i}{\delta^3} \quad . \tag{3.3.179}$$

Der Beweis dieser Zwischenbehauptung folgt unmittelbar aus Bemerkung 1.2.6 und dem Lemma 3.3.21 .

Die gefundenen Abschätzungen wollen wir nun dazu verwenden, Schranken für $\varepsilon_N$ , also für die Größe der Außerdiagonalelemente nach einem kompletten Iterationszyklus zu berechnen. Es sei dazu

$$s_{ij}^{(k)} := |a_{ij}^{(k)}|^2 + |a_{ji}^{(k)}|^2 \quad , \quad k = 0,1,2,\ldots,N \, , \tag{3.3.180}$$

und es sei mit der Schreibweise $k = I(p,q)$ gemeint, daß der k-te Iterationsschritt eines Zyklus' das Pivotindexpaar $(p,q)$ habe.

<u>Zwischenbehauptung 3 :</u> Es sei definiert

(3.3.181) $$\varepsilon := \max \{ \varepsilon_i \, , \, i = 0,1,2, \ldots , N \} \, .$$

und es sei

(3.3.182) $$k + 1 = I(p+1,p+2) \, .$$

Dann existieren für alle p, $p = 0,1,2, \ldots , n-1$ , reelle Konstanten $P_p$ , die jeweils nur von n abhängen, so daß gilt

$$\sum_{\substack{i<j\leq n, \\ 1\leq i\leq p}} s_{ij}^{(k)} \leq P_p \frac{\varepsilon^4}{\delta^6} \quad .$$

<u>Zwischen - Beweis :</u> Wir beweisen diese Behauptung mittels vollständiger Induktion.
Zunächst einmal gilt für $p = 0$

$$\sum_{\substack{i<j\leq n, \\ 1\leq i\leq p}} s_{ij}^{(k)} = 0 \, ,$$

so daß für $p = 0$ die Konstante $P_0 = 0$ unsere Behauptung erfüllt.

Sei nun $P_{p-1}$ bereits berechnet, d.h. es gelte

(3.3.183) $$\sum_{\substack{i<j\leq n \\ 1\leq i\leq p-1}} s_{ij}^{(k)} \leq P_{p-1} \frac{\varepsilon^4}{\delta^6} \quad .$$

Zu zeigen ist, daß dann eine nur von n abhängige Konstante $P_p$ existiert mit

$$\sum_{\substack{i<j\leq n,\\ 1\leq i\leq p}} s_{ij}^{(k+n-p)} \leq P_p \frac{\varepsilon^4}{\delta^6} \quad . \tag{3.3.184}$$

Wir teilen dazu den nichtdiagonalen Teil der Matrix A auf in die drei Teile B,D und F, wobei B aus denjenigen Außerdiagonalelementen von A besteht, die bereits als Pivot verwendet wurden, D aus denen der p-ten Zeile und Spalte und F aus dem Rest. B,D und F sind also definiert gemäß

$$b_{ij} := \begin{cases} a_{ij} & , i<p \text{ oder } j<p \text{ , } i\neq j \text{ ,} \\ 0 & , \text{sonst ,} \end{cases}$$

$$d_{ij} := \begin{cases} a_{ij} & , (i=p \text{ und } i<j) \text{ oder } (j=p \text{ und } i>j), \\ 0 & , \text{sonst ,} \end{cases}$$

$$f_{ij} := \begin{cases} a_{ij} & , i>p \text{ , } j>p \text{ , } i\neq j, \\ 0 & , \text{sonst.} \end{cases}$$

Die folgende Skizze möge diese Aufteilung verdeutlichen :

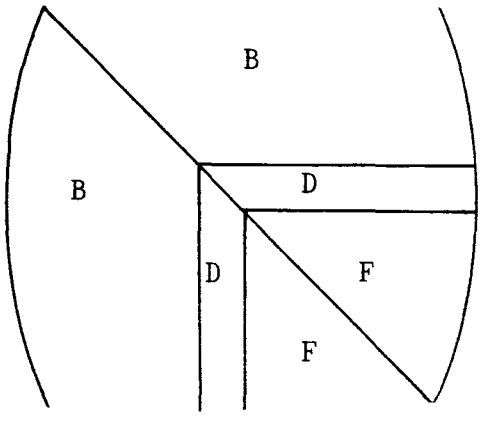


fig. 2

Mit dieser Aufteilung nun ist unsere Behauptung (3.3.184) äquivalent zu

$$(3.3.185) \qquad \|B^{(k+n-p)} + D^{(k+n-p)}\|^2 \leq P_p \frac{\varepsilon^4}{\delta^6} \quad .$$

Nach Induktionsvoraussetzung (3.3.183) ist $P_{p-1}$ bereits bekannt, und es gilt

$$(3.3.186) \qquad \|B^{(k)}\|^2 \leq P_{p-1} \frac{\varepsilon^4}{\delta^6} \quad .$$

Weiter ist

$$(3.3.187) \qquad \|F^{(k)}\|^2 \leq \tau^2(A_k) = \varepsilon_r^2 \leq \varepsilon^2 \quad .$$

Mit Hilfe unserer Zwischenbehauptung 2 können wir für jeden einzelnen Iterationsschritt die Änderungen von $a_{pj}$ und $a_{jp}$ abschätzen. Verwenden wir nacheinander alle Elemente $a_{pj}$ $(p<j)$ der p-ten Zeile als Pivotlemente, so müssen wir diese Abschätzung $(n-j)$ - mal wiederholen und erhalten

$$(3.3.188) \qquad |a_{pj}^{(k+n-p)}| \leq 1.03^{n-j}\, |a_{pj}^{(k+j-p)}| + \sum_{i=j+1}^{n} 1.03^{n-i} u^{(k+i-p-1)} |a_{ij}^{(k+i-p-1)}| \; .$$

Eine entsprechende Abschätzung gilt nach (3.3.178) auch für $a_{jp}^{(k+n-p)}$.

Wir quadrieren die Ungleichung (3.3.188) und schätzen wie folgt ab.

$$|a_{pj}^{(k+n-p)}|^2 \leq 2\cdot 1.07^{n-j}\, |a_{pj}^{(k+j-p)}|^2 + 2\,(\sum_{i=j+1}^{n} 1.03^{n-i} u^{(k+i-p-1)} |a_{ij}^{(k+i-p-1)}|\,)^2$$

$$(3.3.189) \qquad \leq 2\cdot 1.07^{n-j} \left\{ |a_{pj}^{(k+j-p)}|^2 + \sum_{i=j+1}^{n} (u^{(k+i-p-1)})^2 \sum_{i=j+1}^{n} |a_{ij}^{(k+i-p-1)}|^2 \right\} .$$

Dabei haben wir die Cauchy - Schwarz'sche Ungleichung angewandt und zugleich benutzt, daß für $i>j$ gilt

$$1.03^{2(n-i)} < 1.03^{2(n-j)} < 1.07^{n-j} \quad .$$

Wieder gilt eine (3.3.189) entsprechende Abschätzung auch für $a_{jp}^{(k+n-p)}$, und durch Addition erhalten wir

$$\sum_{j=p+1}^{n} s_{pj}^{(k+n-p)} \leq 2\left(\sum_{j=p+1}^{n} 1.07^{n-j} s_{pj}^{(k+j-p)} + \sum_{j=p+1}^{n}\left[1.07^{n-j}\sum_{i=j+1}^{n}(u^{(k+i-p-1)})^2 \sum_{i=j+1}^{n}|a_{ij}^{(k+i-p-1)}|^2\right]\right)$$

$$\text{(3.3.190)} \qquad \leq 2\left(\sum_{j=p+1}^{n} 1.07^{n-j} s_{pj}^{(k+j-p)} + \left[\sum_{i=p+2}^{n}\sum_{j=p+1}^{i-1} 1.07^{n-j} s_{ji}^{(k+i-p-1)}\right]\sum_{i=p+2}^{n}(u^{(k+i-p-1)})^2\right)$$

Um weiter abschätzen zu können, ersetzen wir wegen $p<j$ die Faktoren $1.07^{n-j}$ durch $1.07^{n-p}$ und benutzen, daß nach Lemma 3.3.17 und Lemma 3.3.20 gilt

$$\text{(3.3.191)} \qquad s_{pj}^{(k+j-p)} < 2\,36\frac{\varepsilon_{k+j-p}^4}{\delta^6} \leq 72\,\frac{\varepsilon^4}{\delta^6} \; .$$

Zusammen mit

$$\text{(3.3.192)} \qquad \sum_{j=p+1}^{i-1} s_{ji}^{(k+i-p-1)} \leq \varepsilon_{k+i-p-1}^2 \leq \varepsilon^2$$

und (3.3.179) erhalten wir

$$\sum_{j=p+1}^{n} s_{pj}^{(k+n-p)} \leq 2\cdot 1.07^{n-p}\left\{(n-p)\,72\frac{\varepsilon^4}{\delta^6} + (n-p-1)^2\,\varepsilon^2\,34\frac{\varepsilon^2}{\delta^6}\right\}$$

$$\text{(3.3.193)} \qquad =: Q_p\,\frac{\varepsilon^4}{\delta^6} \quad .$$

Nach unserer Aufteilung des Außerdiagonalteiles von A ist dieses Ergebnis äquivalent zu

$$\|D^{(k+n-p)}\|^2 \leq Q_p \frac{\varepsilon^4}{\delta^6} \quad . \tag{3.3.194}$$

Für die Elemente von B gilt, daß sie entweder unverändert bleiben oder aber mit Hilfe unserer Zwischenbehauptung 1 abgeschätzt werden können. Somit haben wir

$$\|B^{(k+n-p)}\|^2 \leq 1.14^{n-p} \|B^{(k)}\|^2 \quad . \tag{3.3.195}$$

Mittels unserer Induktionsvoraussetzung (3.3.186) erhalten wir also aus (3.3.194) und (3.3.195)

$$\|D^{(k+n-p)} + B^{(k+n-p)}\|^2 \leq Q_p \frac{\varepsilon^4}{\delta^6} + 1.14^{n-p} P_{p-1} \frac{\varepsilon^4}{\delta^6}$$

$$=: P_p \frac{\varepsilon^4}{\delta^6} \quad , \tag{3.3.196}$$

so daß wir gemäß (3.3.185) und der Tatsache, daß $Q_p$ nur von n abhängt, die Zwischenbehauptung 3 bewiesen haben.

Aus der Zwischenbehauptung 3 folgt nun insbesondere

$$\sum_{\substack{i<j\leq n \\ 1\leq i\leq n-1}} s_{ij}^{(k)} = \tau^2(A_N) \leq P_{n-1} \frac{\varepsilon^4}{\delta^6} \quad . \tag{3.3.197}$$

Aufgrund von Zwischenbehauptung 1 gilt

$$\varepsilon^2 \leq (1.14)^N \varepsilon_0^2 \quad ,$$

so daß wir aus (3.3.197) schließlich erhalten

$$\tau^2(A_N) = \tau^2(\hat{A}) \leq P_{n-1}\,(1.14)^N \frac{1}{\delta^6}\,(\tau^2(A))^2 .$$

Setzen wir noch

$$K(\delta,n) := P_{n-1}\,(1.14)^N \frac{1}{\delta^6} ,$$

so ist der Satz vollständig bewiesen. ■

# 4. NUMERISCHE RESULTATE

## 4.1 NUMERISCHE RESULTATE BEI KLEINEN TESTMATRIZEN

Dieser Abschnitt enthält die Resultate, die wir bei der Anwendung unseres neuen Verfahrens nach Abschnitt 3.2 sowie des Eberlein - Verfahrens auf kleine Testmatrizen ( $n \leq 40$ ) erhalten haben. Die entsprechenden Testrechnungen wurden am Rechenzentrum der Fernuniversität - GHS - Hagen auf der Rechenanlage IBM 3031 unter dem Betriebssystem VM/SP, Release 3, durchgeführt; die Programme wurden von uns in FORTRAN in "double precision complex" - Arithmetik formuliert und mittels des VS - FORTRAN - Compilers übersetzt.

Bei unseren zahlreichen Testrechnungen zeigte sich, daß bei gut konditionierten Testmatrizen unser Verfahren vergleichbar schnell, in einigen Fällen auch schneller als das von Eberlein war. Im allgemeinen traten bei solchen Matrizen jedoch keine signifikanten Unterschiede zwischen den beiden Methoden auf.

Dagegen besitzt unser neues Verfahren bei defektiven Matrizen deutliche Vorteile, die sich besonders dann auswirken, wenn sowohl die Anzahl der defektiven Eigenwerte als auch deren Defektivität selbst klein ist. Allgemein war weiter zu beobachten, daß unsere Methode die euklidische Norm der iterierten Matrix schneller reduziert als es das Verfahren von Eberlein tut. Eine mögliche Erklärung hierfür liegt darin, daß in unserem Verfahren der normreduzierende Parameter exakt bestimmt wird, wogegen dies bei Eberlein lediglich approximativ geschieht.

Schließlich war bei beiden Verfahren die Genauigkeit der berechneten Eigenwerte befriedigend und vergleichbar; bei gut konditionierten Matrizen wurden (bei einer Mantissenlänge von 14 Stellen) noch 12 Stellen korrekt ausgegeben.

Wir geben nun eine kleine Auswahl unserer Testmatrizen zusammen mit den von uns erhaltenen numerischen Ergebnissen an. Die Beispiele 1 bis 5 sind dabei (mit der dortigen Numerierung) entnommen aus Eberlein [4]; für die Beispiele 6 und 7 vergleiche Gregory/Karney [1], Example 5.6 bzw. 5.17.

Bei beiden betrachteten Verfahren wurde die Iteration beendet, wenn eines der folgenden Abbruchkriterien erfüllt war.

i) $\max\limits_{i \neq j} |a_{ij}| \leq 10^{-9}$ .

ii) Während eines vollen Iterationszyklus' wurde keine Transformation mehr durchgeführt.

iii) Es wurden bereits 50 Iterationszklen durchgeführt.

Darüber hinaus wurde in beiden Methoden eine Transformation mit Pivotindexpaar (p,q) übersprungen, wenn für die entsprechenden Pivotelemente $a_{pq}$ und $a_{qp}$ galt

$$|a_{pq}| + |a_{qp}| \leq \frac{2}{n^2} 10^{-10}$$

oder

$$|a_{pq}| + |a_{qp}| \leq (MAX)^2 ,$$

wobei

$$MAX := \max_{i \neq j} |\tilde{a}_{ij}|$$

und $\tilde{A}$ die iterierte Matrix nach Beendigung des unmittelbar vorhergehenden Iterationszyklus bezeichnet.

Die in den folgenden Tabellen benutzten Abkürzungen bedeuten im einzelnen:

'CPU' : die für die Iteration benötigte CPU - Zeit in Sekunden;

'Zyklen' : die Anzahl der benötigten Iterationszyklen;

'Cond' : die Konditionszahl K der berechneten Eigenvektormatrix T, wobei

$$K := \|T\|_1 \, \|T^{-1}\|_1$$

und

$$\|T\|_1 := \max_{1 \leq j \leq n} \left( \sum_{i=1}^{n} |t_{ij}| \right) ;$$

'Off' : $\max_{i \neq j} |a_{ij}|$ ;

'I' : unser neues Verfahren;

'II' : das Eberlein - Verfahren.

<u>BEISPIEL 1 :</u> Es handelt sich um eine 7 x 7 - Matrix mit den Eigenwerten $-1$, $\pm i$, $\frac{\sqrt{2}}{2}(-1 \pm i)$, $\frac{\sqrt{2}}{2}(1 \pm i)$.

| | CPU | Zyklen | Cond | Off |
|---|---|---|---|---|
| I | 0.92 | 6 | $0.12 \cdot 10^{2}$ | $0.17 \cdot 10^{-11}$ |
| II | 1.24 | 8 | $0.10 \cdot 10^{2}$ | $0.14 \cdot 10^{-11}$ |

BEISPIEL 2 : Die Eigenwerte dieser 5 x 5 - Matrix sind -1 und doppelt defektiv $\frac{1}{2}(3 \pm i\sqrt{51})$.

| | CPU | Zyklen | Cond | Off |
|---|---|---|---|---|
| I | 0.62 | 11 | $0.48 \cdot 10^{9}$ | $0.34 \cdot 10^{-12}$ |
| II | 1.89 | 38 | $0.26 \cdot 10^{9}$ | $0.38 \cdot 10^{-11}$ |

BEISPIEL 3 : Diese 13 x 13 - Matrix besitzt 7 gut und 6 schlecht konditionierte Eigenwerte.

| | CPU | Zyklen | Cond | Off |
|---|---|---|---|---|
| I | 14.16 | 33 | $0.23 \cdot 10^{10}$ | $0.68 \cdot 10^{-9}$ |
| II | 14.55 | 35 | $0.23 \cdot 10^{10}$ | $0.37 \cdot 10^{-9}$ |

BEISPIEL 4 : Hier handelt es sich um eine gut konditionierte komplexe 5 x 5 - Matrix, deren exakte Eigenwerte unbekannt sind.

| | CPU | Zyklen | Cond | Off |
|---|---|---|---|---|
| I | 0.36 | 7 | $0.24 \cdot 10^{2}$ | $0.12 \cdot 10^{-9}$ |
| II | 0.59 | 9 | $0.17 \cdot 10^{2}$ | $0.79 \cdot 10^{-9}$ |

BEISPIEL 5 : Dies ist eine schiefhermitesche 5 x 5 - Matrix mit unbekannten exakten Eigenwerten.

| | CPU | Zyklen | Cond | Off |
|---|---|---|---|---|
| I | 0.31 | 6 | $0.48 \cdot 10^{1}$ | $0.58 \cdot 10^{-12}$ |
| II | 0.30 | 6 | $0.48 \cdot 10^{1}$ | $0.54 \cdot 10^{-9}$ |

BEISPIEL 6 : Diese reelle 4 x 4 - Matrix besitzt die doppelten defektiven Eigenwerte $3 \pm \sqrt{5}$ .

| | CPU | Zyklen | Cond | Off |
|---|---|---|---|---|
| I | 0.33 | 9 | $0.44 \cdot 10^{9}$ | $0.20 \cdot 10^{-12}$ |
| II | 0.89 | 38 | $0.15 \cdot 10^{9}$ | $0.67 \cdot 10^{-9}$ |

BEISPIEL 7 : Die Eigenwerte dieser 40 x 40 - Matrix lauten 8 , 6 , 4 , 3 , -40 , -30 , -20 , -15, ±40j, ±30j, ±20j, ±15j, ±8j, ±6j, ±4j, ±3j, wobei

$$j = \frac{1}{2}(1 \pm \sqrt{3}\, i) .$$

| | CPU | Zyklen | Cond | Off |
|---|---|---|---|---|
| I | 223 | 13 | $0.69 \cdot 10^{4}$ | $0.20 \cdot 10^{-10}$ |
| II | 284 | 17 | $0.36 \cdot 10^{4}$ | $0.57 \cdot 10^{-12}$ |

## 4.2 PARAMETERANALYSE, NAIVE METHODE

In der Praxis stellt sich häufig die Aufgabe, eine gewisse Anzahl m von Matrizen $A_i$ zu diagonalisieren, die stetig von gewissen Parametern abhängen und sich voneinander nur wenig unterscheiden. Es sei also gegeben

$$A_{i+1} = A_i + E_i \,, \qquad i = 1, \ldots, m-1, \tag{4.2.1}$$

wobei

$$\|E_i\| \ll \|A_i\| \,, \qquad i = 1, \ldots, m \,. \tag{4.2.2}$$

Ein typisches Problem, das auf eine solche "Parameteranalyse" führt, ist z.B. ein schwingendes dynamisches System mit verschiedenen Dämpfern, an dessen Verhalten bei kleinen Änderungen der Dämpfungsstärken man interessiert ist.

Bei der Verwendung Jacobi - ähnlicher Verfahren zur Diagonalisierung der $A_i$ bietet sich nun folgendes Vorgehen an. Man berechnet zunächst wie gewöhnlich die Diagonalisierungsmatrix $U_1$ mit

$$U_1^{-1} A_1 U_1 = \Lambda_1 \,, \qquad \Lambda_1 \text{ Diagonalmatrix} \,, \tag{4.2.3}$$

und führt dann für $A_2$ eine Vortransformation durch gemäß

$$\begin{aligned} \tilde{A}_2 &:= U_2^{-1} A_2 U_2 \\ &= \Lambda_1 + U_1^{-1} E_1 U_1 \,. \end{aligned} \tag{4.2.4}$$

Bei kleiner Störung $E_1$ ist dann wegen Stetigkeit $\tilde{A}_2$ bereits eine diagonal-dominante Matrix, für die das verwendete Jacobi - Verfahren in relativ wenigen Iterationszyklen eine neue

Diagonalisierungsmatrix $\tilde{U}_2$ liefert. Mit $U_2$,

$$U_2 := U_2 \tilde{U}_2 \,, \tag{4.2.5}$$

wird dann entsprechend die Matrix $A_3$ vortransformiert u.s.w..

Für diagonal-dominante Matrizen, wie sie bei dieser Verfahrensweise auftreten, hat nun K. VESELIĆ eine sogenannte "Naive Methode" zur Diagonalisierung vorgeschlagen, die nach unserer Definition 1.2.9 ebenfalls Jacobi - ähnlich ist. Mit den dortigen Bezeichnungen gilt

$$\begin{vmatrix} t_{pp} & t_{pq} \\ t_{qp} & t_{qq} \end{vmatrix} = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ y & 1 \end{pmatrix} \,, \quad x,y \in \mathbb{C} \,, \tag{4.2.6}$$

wobei $x$ zur Annihilierung des Pivotelementes $a_{pq}$ und $y$ zu der von $a_{qp}$ bestimmt wird. Dies führt auf

$$x := \frac{a_{pp} - a_{qq} \pm \sqrt{4\, a_{pq} a_{qp} + (a_{qq} - a_{pp})^2}}{2\, a_{qp}} \tag{4.2.7}$$

und

$$y := \frac{a_{qp}}{a_{pp} - a_{qq} - 2x\, a_{qp}} \quad ; \tag{4.2.8}$$

dabei wird in (4.2.7) der kleinere der beiden möglichen Werte für $x$ gewählt.

Der große Vorteil dieser Naiven Methode für diagonal-dominante Matrizen liegt in dem minimalen Rechenaufwand, den sie zur Bestimmung der Transformationsparameter benötigt; allerdings garantiert sie weder eine Abnahme in der Betragsgröße $\tau$ der Außerdiagonalelemente noch die Reduzierung der euklidischen Norm der iterierten Matrix. Daher muß bei einer Implementierung dieser Methode dafür Sorge getragen werden, daß bei ihrem Versagen möglichst ohne Verlust bereits geleisteter Arbeit auf eines der üblichen Jacobi - ähnlichen Verfahren umgeschaltet werden kann.

Wir haben nun die beiden folgenden Algorithmen zur Behandlung von Parameteranalyse - Aufgaben erstellt.

I) Algorithmus 'PARAM':

1) Für die Startmatrix $A_1$ wird unser neues Verfahren gemäß Abschnitt 3.2 durchgeführt. Die erhaltene Diagonalisierungsmatrix sei $T_1$ .

2) Für $i = 2,\ldots,m$ wird zunächst die Vortransformation

$$\tilde{A}_i := T_{i-1}^{-1} A_i T_i$$

ausgeführt. Auf $\tilde{A}_i$ wird dann wieder unser Verfahren angewandt. Die Aufmultiplikation der erhaltenen Transformationsmatrix $\tilde{T}_i$ auf die alte gemäß

$$T_i := T_{i-1} \tilde{T}_i$$

geschieht dabei laufend während der Iteration.

Wächst von einem Zyklus zum nächsten die Kondition der Transformationsmatrix um mehr als den Faktor 10 an, so wird die Diagonalisierung der Matrix $A_i$ mit entsprechender Meldung abgebrochen. Damit wird verhindert, daß bei einer Matrix mit defektiven oder sehr schlecht konditionierten Eigenwerten die akkumulierte Transformationsmatrix unvernünftig stark verfälscht wird.

II) Algorithmus 'NAIV'

Dieser Algorithmus unterscheidet sich von 'PARAM' dadurch, daß auf die bereits vortransformierte Matrix jeweils die oben beschriebene Naive Methode angewandt

wird. Dabei wird nach jedem Iterationszyklus die aktuelle iterierte Matrix $\hat{A}$ als "Sicherungskopie" abgespeichert. Führt nun der nächste Zyklus der Naiven Methode zu einem Anwachsen der Außerdiagonalelemente, so wird auf die "alte" Matrix $\hat{A}$ zurückgegriffen; auf ihr werden dann 2 Zyklen unseres Verfahrens nach Abschnitt 3.2 ausgeführt. Anschließend wird dann naiv weiteriteriert.

## 4.3 NUMERISCHE RESULTATE BEI DER PARAMETERANALYSE GROSSER MATRIZEN MITTELS DES VEKTORRECHNERS CYBER 205

Die seit kurzem verfügbare neue Rechnergeneration der sogenannten Vektor- oder Parallelrechner wie z.B. CRAY 1 oder CDC - CYBER 205 ist in der Lage, Operationen auf langen Vektoren in im Vergleich zu herkömmlichen Rechenanlagen extrem kurzen Zeiten auszuführen. Da bei der Anwendung Jacobiähnlicher Verfahren auf große Matrizen (etwa $n \geq 50$) die in jedem Iterationsschritt auszuführenden Matrixtransformationen den Hauptteil des von ihnen benötigten Rechenaufwandes ausmachen, erscheint es sinnvoll, gerade solche Verfahren auf Vektorrechnern zu installieren. Wir haben daher sowohl die in Abschnitt 4.2 beschriebenen Algorithmen 'PARAM' und 'NAIV' sowie (zu Vergleichszwecken) das komplexe QR - Verfahren auf der Rechenanlage CYBER 205 des Rechenzentrums der Ruhruniversität Bochum implementiert. Die Programme hierzu wurden von uns in der FORTRAN-Erweiterung FORTRAN-200 geschrieben, die durch die Verwendung von "special calls" zur Vektorverarbeitung die Möglichkeiten der CYBER 205 optimal ausnutzt. Die Quelltexte wurden unter dem Betriebssystem VSOS - Release 2.1 mittels des FTN200 - Compilers übersetzt und ausgeführt.

Ziel unserer Untersuchungen war hauptsächlich, Aussagen über die Leistungsfähigkeit Jacobi - ähnlicher Verfahren bei ihrer Anwendung auf Probleme der Parameteranalyse unter der Verwendung eines Vektorrechners zu gewinnen. Darüber hinaus war weiter interessant zu erfahren, wieviel die einzelnen Programme von der expliziten Vektorisierung mittels der FORTRAN - 200 - Möglichkeiten profitieren konnten. Insgesamt haben wir dazu die folgenden Programme implementiert und getestet.

'QR' : Das komplexe QR - Verfahren ohne Berechnung der Eigenvektoren, unverändert übernommen aus EISPACK [5].

'QRV' : Vektorisierte Version von 'QR'. Dabei wurden genau die Spaltentransformationen und die Skalarproduktberechnungen vektorisiert (bei Abspeicherung der Matrix mittels des DIMENSION - statements); eine Vektorisierung auch der Zeilentransformationen mit den dazu nötigen Umspeicherungen erbrachte keinen weiteren Rechenzeitgewinn.

'PARAM' : Der in Abschnitt 4.2 beschriebene Algorithmus zur Parameteranalyse in nicht-vektorisierter Form.

'PARAMV' : Vektorisierte Version von 'PARAM'. Es wurden genau die Zeilen- und Spaltentransformationen sowie die Skalarproduktberechnungen vektorisiert. Um auf möglichst langen Vektoren arbeiten zu können, wurde zusätzlich die iterierte Matrix als 2n x n - Matrix abgespeichert, wobei die ersten n Zeilen den Realteil und die zweiten n Zeilen den Imaginärteil der Matrix enthalten.

'NAIV' : Der in Abschnitt 4.2 beschriebene Algorithmus zur Parameteranalyse in vektorisierter Form. Für die Startmatrix $A_1$ ist dieses Programm identisch mit 'PARAMV'; für die Durchführung der naiven Iterationszyklen wurde gegenüber 'PARAMV' lediglich der Teil zur Berechnung der Transformationsparameter ausgetauscht.

Als Ausgangsmatrizen $A_1$ für unsere Tests benutzen wir Matrizen mit von einem Zufallsgenerator gleichverteilt im Intervall [-1 , 1] erzeugten Elementen. Die weiteren Matrizen $A_i$ werden

aus $A_{i-1}$ dadurch generiert, daß jeweils n wiederum zufällig ausgewählte Elemente um einen bestimmten Prozentsatz P geändert werden. In unseren Tests haben wir dabei als realistischen Wert P = 5% gesetzt; bei kleineren Werten benötigten unsere Jacobi - ähnlichen Algorithmen lediglich noch einen Iterationszyklus zur Diagonalisierung der vortransformierten Matrizen, bei deutlich größeren Änderungen ($P \geq 10\%$) waren diese Matrizen kaum noch diagonal - dominant, so daß der Eigenwertverlauf infolge von Sprüngen der Diagonalelemente nicht mehr verfolgt werden konnte.

Im folgenden geben wir nun die von den von uns getesteten Verfahren benötigten Rechenzeiten in Sekunden für verschiedene Matrixgrößen n an. Es handelt sich hierbei um Durchschnittszeiten, da für jedes Verfahren mehrere Testreihen mit verschiedenen Startmatrizen und jeweils 10 vortransformierten Matrizen durchgeführt wurden; die bei den einzelnen Methoden aufgetretenen Abweichungen von diesen protokollierten Durchschnittszeiten waren dabei nur sehr gering.

Für das QR - Verfahren ist in der folgenden Tabelle die für eine volle, nicht vortransformierte Matrix benötigte Zeit aufgeführt. Führt man hier ebenfalls eine Vortransformation durch, so gewinnt man zwar zunächst 20% an Rechenzeit, wegen der dann aber notwendigen Berechnung der Transformationsmatrizen wird das QR - Verfahren jedoch insgesamt um etwa den Faktor 2 langsamer.

Bei den Jacobi - ähnlichen Verfahren ist schließlich unter 'I' die für die Startmatrix, unter 'II' die für die vortransformierten Matrizen benötigte Rechenzeit protokolliert. Die einzelnen Matrizen wurden dabei lediglich bis auf $10^{-4}$ diagonalisiert, da diese Genauigkeit in der Praxis i. a. völlig ausreichend ist; man gewinnt dadurch etwa einen Iterationszyklus.

Unsere Ergebnisse sind :

| n | QR | QRV | PARAM I | PARAM II | PARAMV I | PARAMV II | NAIV II |
|---|---|---|---|---|---|---|---|
| 80 | 2.05 | 1.75 | 20.7 | 6.0 | 11.9 | 3.72 | 3.0 |
| 100 | 3.92 | 2.80 | 38.0 | 15.2 | 21.0 | 8.30 | 5.3 |
| 150 | 12.7 | 8.74 | 124 | 49.0 | 66.7 | 25.0 | 15.6 |
| 180 | 22.2 | 14.4 | 239 | 83.0 | 114 | 39.2 | 25.1 |

Man erkennt zunächst, daß das QR - Verfahren durch die explizite Vektorisierung etwa um einen Faktor 1.6 schneller wird, bei unserem Algorithmus PARAM beträgt dieses Verhältnis ca. 2.1 (bei n = 180). Dies ist weniger, als von uns vor Beginn der Arbeiten erwartet worden war. Zusätzliche von uns durchgeführte Tests haben hier gezeigt, daß allein die Verwendung der Optimize - Option OPT = 1 bei der Programmübersetzung (wie wir sie bei allen oben aufgeführten Resultaten benutzt haben), bereits einen Faktor 10 (!) an Rechenzeitgewinn erreicht (Programm PARAM, n = 150). Zum Vergleich : bei der IBM 3031 der Fernuniversität - GHS - Hagen erreichten wir durch Verwendung des Optimizers bei einer 40 x 40 - Matrix lediglich eine Verbesserung um etwa 10% . Es scheint also so, daß der FTN200 - Optimizer bereits eine sehr gute Anpassung der Programme an die Möglichkeiten und Architektur der CYBER 205 vornimmt.

Betrachtet man die von uns erhaltenen absoluten Rechenzeiten, so rechtfertigen diese die Implementierung unserer Programme auf der CYBER 205. Vergleichen wir nochmals mit der IBM 3031, so erzielt man bereits bei einer 80 x 80 - Matrix eine Beschleunigung um den Faktor 80 (inclusive expliziter Vektorisierung).

Vergleichen wir schließlich die mit dem QR - Verfahren erzielten Resultate mit denen unserer Jacobi - ähnlichen Methoden. Hier zeigt sich, daß die Naive Methode für bereits vortransformierte Matrizen lediglich um einen Faktor 1.8 langsamer als QR ist. Dabei ist zu bedenken, daß das QR - Verfahren die berechneten Eigenwerte nach ihrer Größe sortiert ausgibt, so daß eine Verfolgung des Eigenwertverlaufs hier oft nur schwer möglich ist. Darüber hinaus besitzt man bei Jacobi - ähnlichen Verfahren zu jedem Zeitpunkt der Parameteranalyse zusätzlich die volle Information über die zugehörigen Eigenvektoren, so daß zum Beispiel über die Kondition der Transformationsmatrix eine Schrittweitensteuerung bei der Änderung der eingehenden Parameter möglich ist. Wird schließlich die Kontrolle über die Eigenvektoren tatsächlich benötigt, so wird wie bereits oben erwähnt das QR - Verfahren um den Faktor 2 langsamer und benötigt damit sogar mehr Zeit als die Naive Methode. In der Praxis sollte man hier also die Startmatrix einmal mittels des QR - Verfahrens unter Berechnung der Eigenvektoren diagonalisieren und anschließend mit der Naiven Methode fortfahren.

## LITERATURVERZEICHNIS

[1] BAUER, F.L., FIKE, C.T. :
Norms and exclusion theorems;
Numer. Math. 2 (1960), 137 - 141

[2] EBERLEIN, P.J. :
Solution to the eigenproblem by a norm reducing Jacobi - type method;
Numer. Math. 11 (1968), 1 - 12

[3] EBERLEIN, P.J. :
A Jacobi - like method for the automatic computation of eigenvalues and eigenvectors of an arbitrary matrix;
J. Soc. Indust. Appl. Math. Vol. 10, No. 1 (1962), 74 - 88

[4] EBERLEIN, P.J. :
Solution to the complex eigenproblem by a norm reducing Jacobi - type method;
Numer. Math. 14 (1970), 232 - 245

[5] Matrix Eigensystem Routines / EISPACK GUIDE;
Springer, Berlin-Heidelberg-New York 1976, second edition

[6] ERDMANN, A. :
Über ein neues normreduzierendes Verfahren zur numerischen Berechnung der Eigenwerte beliebiger reeller Matrizen;
Diplomarbeit, Universität Dortmund, 1980

[7] FORSYTHE, G.E., HENRICI, P. :
The cyclic Jacobi method for computing the principal values of a complex matrix;
Trans. Amer. Math. Soc. 94 (1960), 1 - 23

[8] GOLDSTINE, H.H., HORWITZ, L.P. :
A procedure for the diagonalization of normal matrices;
J. Assoc. Comp. Mach. 6 (1959), 176 - 195

[9] GOLDSTINE, H.H, MURRAY, F.J., von NEUMANN, J. :
The Jacobi method for real symmetric matrices;
J. Assoc. Comp. Mach. 6 (1959), 59 - 96

[10] GOLUB, G.H., VAN LOAN, CH.F. :
Matrix Computations ;
North Oxford Academic Publishing Co. Ltd, 1983

[11] GREGORY, R.T., KARNEY, D.L. :
A collection of matrices for testing computational algorithms;
Wiley - Interscience, New York 1969

[12] JACOBI, C.G.J. :
Über ein leichtes Verfahren, die in der Theorie der Säcularstörungen vorkommenden Gleichungen numerisch aufzulösen;
J. für die reine und angewandte Mathematik 30 (1846), 51 - 95

[13] RUHE, A. :
On the quadratic convergence of a generalisation of the Jacobi method for arbitrary matrices ;
BIT 8 (1968), 210 - 231

[14] RUTISHAUSER, H. :
Une methode pour le calcul des valeurs propres des matrices non symmétriques ;
C. R. Acad. Sc. Paris, t. 259 (1964), 2758

[15] VESELIĆ, K. :
A convergent Jacobi - method for solving the eigenproblem of arbitrary real matrices ;
Numer. Math. 25 (1976), 179 - 184

[16] VOEVODIN, V.V. :
Cislennye metody lineinaj algebry ;
Nauka, Moskva 1966

[17] WILKINSON, J.H. :
Almost diagonal matrices with multiple or close eigenvalues ;
Linear Algebra and Its Applications 1 (1968), 1 - 12

[18] WILKINSON, J.H. :
The Algebraic Eigenvalue Problem ;
Clarendon Press, Oxford 1978

[19] ZURMÜHL, R. :
Matrizen und ihre technischen Anwendungen ;
Springer, Berlin-Heidelberg-New York 1964, 4. Auflage

# L e b e n s l a u f

| | |
|---|---|
| 5.3.1955 | Geburt in Lünen<br>Eltern : Alfons Erdmann und<br>Hanna Erdmann, geb. Eppen |
| Ostern 1961 | Einschulung in die katholische Volksschule Altlünen - Alstedde |
| Ostern 1965 | Wechsel zum Freiherr vom Stein - Gymnasium Lünen |
| 16.5.1973 | Bestehen der Reifeprüfung |
| 25.9.1973 | Immatrikulation an der Universität Dortmund im Fach Mathematik, Nebenfach Informatik |
| 28.8.1980 | Bestehen der Diplom - Hauptprüfung im Fach Mathematik |
| 1.10.1980 | Eintritt in die Fernuniversität - Gesamthochschule - Hagen als wissenschaftliche Hilfskraft im Fachbereich Mathematik |
| 1.12.1980 | Weiterbeschäftigung ebendort als wissenschaftlicher Angestellter |

Hagen, den 9.8.1984