Sitzungsberichte

der

mathematisch-naturwissenschaftlichen Klasse

der

Bayerischen Akademie der Wissenschaften zu München

Jahrgang 1951

München 1952

Verlag der Bayerischen Akademie der Wissenschaften

In Kommission bei der C. H. Beck'schen Verlagsbuchhandlung München

Zur numerischen Auflösung algebraischer Gleichungen

Von Fritz Wenzl in München

Vorgelegt von Herrn Josef Lense am 6. Juli 1951

I.

Ph. Furtwängler verwendete zur numerischen Auflösung von Gleichungen 4. Grades mit zwei Paaren konjugiert komplexer Wurzeln ein offenbar wenig bekannt gewordenes Verfahren, das an folgendem Beispiel erläutert sei:[1] Gegeben sei die Gleichung

$$F(x) = x^4 + 2{,}5x^3 + 10x^2 + 4x + 1. \tag{1.1}$$

Sie hat, wie man leicht erkennt, ein Paar „großer" konjugiert komplexer Wurzeln (vom Betrag $R > 1$), zusammengefaßt in einer quadratischen Gleichung $G(x) = 0$, und ein Paar „kleiner" konjugiert komplexer Wurzeln (vom Betrag $\frac{1}{R} < 1$), zusammengefaßt in einer quadratischen Gleichung $K(x) = 0$. Die Gleichung (1.1) ist im wesentlichen gelöst, wenn die Aufspaltung

$$F(x) = G(x) \cdot K(x) \tag{1.2}$$

in die Faktoren G und K hinreichend genau bekannt ist[2]. Dazu kann man folgendermaßen vorgehen:

---

[1] Der Verfasser verdankt die Kenntnis des Verfahrens einer Vorlesung von Herrn Prof. Dr. Lense. Erst nachträglich wurde der Verfasser der vorliegenden Arbeit auf eine Abhandlung von B. Friedman aufmerksam (Comm. on Pure and Applied Math. 1949, S. 195), in der im wesentlichen derselbe Gegenstand wie hier behandelt ist. Da sich jedoch weder die Fragestellung noch der eingeschlagene Weg und die erzielten Ergebnisse decken, mag das Folgende trotzdem von Interesse sein. (Friedman beschränkt sich von vornherein auf den Fall hinreichend guter erster Näherungen und erzielt dabei auch sehr allgemeine Ergebnisse. Für Gleichungen 4. Grades mit 2 Paaren konjugiert komplexer Wurzeln werden jedoch auch unter der Voraussetzung hinreichend guter erster Näherungen die Friedmanschen Ergebnisse hier beträchtlich verschärft; vgl. dazu auch Anm. S. 88.)

[2] Der Koeffizient von x^3 sei in beiden Faktoren gleich 1.

1. Näherung K_1 für die „kleinen" Wurzeln (worunter hier wie auch weiterhin immer die Wurzeln mit dem kleineren Betrag verstanden seien; entsprechend ist die Bezeichnung „große" Wurzeln zu verstehen):

$$K_1 = x^2 + 0{,}4x + 0{,}1 \tag{1.3}$$

(entstanden durch Vernachlässigung der Glieder 3. und 4. Grades in (1.1)). Daraus 1. Näherung G_1 für die „großen" Wurzeln mittels Division unter Vernachlässigung des Restes:

$$F : K_1 \sim G_1 = x^2 + 2{,}1x + 9{,}06. \tag{1.4}$$

Daraus 2. Näherung K_2 für die „kleinen" Wurzeln mittels angenäherter Division nach steigenden Potenzen von x, so daß

$$\underbrace{1 + 4x + \cdots x^4}_{F} \sim \underbrace{(9{,}06 + 2{,}1x + x^2)}_{G_1} \cdot \underbrace{(0{,}1104 + 0{,}4159x + x^2)}_{K_2}. \tag{1.5}$$

(Übereinstimmung des konstanten Gliedes, des linearen Gliedes und des Gliedes mit x^4 auf beiden Seiten.)

Nun wieder mittels gewöhnlicher Division unter Vernachlässigung des Restes, ausgehend von K_2, eine zweite Näherung G_2 für die „großen" Wurzeln aus

$$F : K_2 \sim G_2 = x^2 + 2{,}084x + 9{,}023 \tag{1.6}$$

usw. Ein Vergleich mit $G_1 = x^2 + 2{,}10x + 9{,}06$ zeigt, daß entsprechende Koeffizienten von G_1 und G_2 nur mehr um weniger als 1% voneinander abweichen.

II.

Das in I. dargestellte Verfahren weist eine gewisse Unsymmetrie zwischen den Näherungsdivisionen $F : K_\nu$ und den Näherungsdivisionen $F : G_\nu$ auf. Während nämlich z. B. bei der Division (1.4) der Quotient G_1 wie üblich so bestimmt wird, daß F und $K_1 \cdot G_1$ in den drei höchsten auftretenden Potenzen von x (d. h. in x^4; x^3 und x^2) übereinstimmen, wird in (1.5) der Quotient K_2 so berechnet, daß F und $G_1 \cdot K_2$ in den Koeffizienten von x^0,

x^1 und x^4 übereinstimmen. Es liegt nahe, den Schritt (1.5) und die entsprechenden späteren Näherungsdivisionen $F : G_\nu$ so abzuändern, daß F und $G_\nu \cdot K_{\nu+1}$* in den drei niedrigsten Potenzen von x (also in x^0; x^1 und x^2) übereinstimmen. Dieser Forderung entspricht das folgende symmetrisch gebaute Rechenschema (die Indizes sind hier abweichend von I. so gewählt, daß K_ν; $G_{\nu+1}$; $K_{\nu+2}$ usw. aufeinander folgen):

Gegeben sei:

$$F = x^4 + Ax^3 + Bx^2 + Cx + 1 \tag{2.1}$$

und eine irgendwie gewonnene Näherung

$$K_\nu = B_\nu x^2 + C_\nu x + 1 \tag{2.2}$$

für die „kleinen" Wurzeln von $F = 0$. Dann erhält man eine Näherung $G_{\nu+1}$ für die „großen" Wurzeln aus

(2.2a)

$$\underbrace{x^4 + Ax^3 + Bx^2 + \cdots}_{F} \sim \frac{1}{B_\nu} \underbrace{(B_\nu x^2 + C_\nu x + 1)}_{K_\nu} \underbrace{(x^2 + A_{\nu+1} x + B_{\nu+1})}_{G_{\nu+1}}$$

mit

$$A_{\nu+1} = A - \frac{C_\nu}{B_\nu}; \quad B_{\nu+1} = B - A_{\nu+1} \cdot \frac{C_\nu}{B_\nu} - \frac{1}{B_\nu} \tag{2.3a}$$

und daraus eine Näherung $K_{\nu+2}$ für die „kleinen" Wurzeln mittels

(2.2b)

$$\underbrace{1 + Cx + Bx^2 + \cdots}_{F} \sim \frac{1}{B_{\nu+1}} \underbrace{(B_{\nu+1} + A_{\nu+1} x + x^2)}_{G_{\nu+1}} \underbrace{(1 + C_{\nu+2} x + B_{\nu+2} x^2)}_{K_{\nu+2}}$$

mit

$$C_{\nu+2} = C - \frac{A_{\nu+1}}{B_{\nu+1}}; \quad B_{\nu+2} = B - C_{\nu+2} \cdot \frac{A_{\nu+1}}{B_{\nu+1}} - \frac{1}{B_{\nu+1}}. \tag{2.3b}$$

Die in den Rekursionsformeln (2.3a) bzw. (2.3b) auftretenden Quotienten und Produkte lassen sich jeweils mit einer Rechenschiebereinstellung ermitteln.

---

* Bis auf einen Proportionalitätsfaktor, der für die Nullstellen nichts ausmacht.

Rechnet man das eben dargestellte Verfahren für das Beispiel (1.1) durch, so ergibt sich, ausgehend von

$$(2.4)\quad \begin{cases} K_1 = 1 + 4x \qquad + 10x^2 \\ \qquad\qquad \text{zunächst} \quad G_2 = x^2 + 2{,}1\,x + 9{,}06 \quad \text{und dann} \\ K_3 = 1 + 3{,}768\,x \qquad + 9{,}016\,x^2; \\ \qquad\qquad G_4 = x^2 + 2{,}082\,06\,x \quad + 9{,}0189; \\ K_5 = 1 + 3{,}769\,145\,x + 9{,}018\,996\,x^2; \\ \qquad\qquad G_6 = x^2 + 2{,}082\,088\,x + 9{,}018\,994. \end{cases}$$

Der Vergleich von G_4 mit G_6 läßt vermuten, daß bereits nach 3 Divisionen, d. h. bei G_4, die Koeffizienten in den ersten 4 Dezimalen festliegen. Einen Anhaltspunkt für die Güte der Näherung gibt die Übereinstimmung bzw. Differenz des Koeffizienten von x^2 in K_ν und des konstanten Gliedes in $G_{\nu+1}$ bzw. $G_{\nu-1}$ (falls die sämtlichen Näherungspolynome entsprechend den Formeln (2.2a) und (2.2b) normiert sind).

Die in (2.1) enthaltene Voraussetzung, daß F das konstante Glied 1 besitzt, stellt keine wesentliche Einschränkung der Allgemeinheit dar, denn jede Gleichung $\xi^4 + a\xi^3 + b\xi^2 + c\xi + d = 0$ mit $d > 0$ läßt sich durch $\xi = \sqrt[4]{d} \cdot x$ auf die Form (2.1) bringen (für den zunächst betrachteten Fall zweier Paare konjugiert komplexer Wurzeln ist immer $d > 0$). Natürlich kann das durch (2.3a) und (2.3b) dargestellte Verfahren sukzessiver Divisionen sinngemäß auch auf den Fall $d \neq 1$ unmittelbar (ohne vorherige Transformation der Unbekannten) angewendet werden. Für die folgenden Konvergenzbetrachtungen genügt es jedoch wegen der Möglichkeit der Transformation, wenn wir uns von vornherein auf den rechnerisch besonders bequemen Fall $d = 1$, d. h. auf den Fall der normierten Gleichung (2.1) beschränken.

III. (Konvergenzbetrachtung)

Das Polynom (2.1) enthalte die quadratischen Faktoren

$$(3.1)\quad G(x) = x^2 + \mu \cdot Rx + R^2 \quad \text{und} \quad K(x) = R^2x^2 + \varkappa Rx + 1.$$

Nach (2.1) ergeben sich dann die Koeffizienten von

$F(x) = G(x) \cdot K(x) \cdot R^{-2}$ aus

$$(3.2)\quad A = \mu R + \varkappa R^{-1};\ B = R^2 + \mu\varkappa + R^{-2};\ C = \varkappa R + \mu R^{-1}.$$

Für die Koeffizienten B_ν und C_ν der Näherung (2.2) gelte unter Verwendung zweier Korrekturgrößen β_ν und γ_ν die Darstellung

$$(3.3)\qquad B_\nu = R^2(1 + \beta_\nu);\qquad C_\nu = R(\varkappa + \gamma_\nu).$$

Dann folgt nach (2.3a) unter Berücksichtigung von (3.2) und (3.3) zunächst

$$A_{\nu+1} = \mu R + \varkappa R^{-1} - \frac{\varkappa + \gamma_\nu}{1 + \beta_\nu} \cdot R^{-1} \quad \text{oder}$$

$$(3.4a)\quad A_{\nu+1} = R(\mu + \alpha_{\nu+1}) \text{ mit } \alpha_{\nu+1} = \frac{\varkappa\beta_\nu - \gamma_\nu}{1 + \beta_\nu} \cdot R^{-2}.$$

Ebenso folgt nach (2.3a) unter Berücksichtigung von (3.2), (3.3) und (3.4a) die Gleichung

$$B_{\nu+1} = R^2 + \mu\varkappa + R^{-2} - R(\mu + \alpha_{\nu+1}) \cdot \frac{\varkappa + \gamma_\nu}{1 + \beta_\nu} \cdot R^{-1} - \frac{1}{1 + \beta_\nu} R^{-2}$$

oder

$$(3.4b)\quad B_{\nu+1} = R^2(1 + \beta_{\nu+1}) \text{ mit } \beta_{\nu+1} = \alpha_{\nu+1}\left(\mu - \frac{\varkappa + \gamma_\nu}{1 + \beta_\nu} R^{-2}\right) + \frac{\beta_\nu R^{-4}}{1 + \beta_\nu}.$$

Bei zwei Paaren konjugiert komplexer Wurzeln der Gleichung $F(x) = 0$ stellt der Faktor

$$(3.5)\qquad \rho = R^{-2},\quad |\rho| < 1 \quad \text{vorausgesetzt,}$$

wegen (3.1) das Verhältnis des Betrags einer „kleinen" Wurzel zum Betrag einer „großen" Wurzel dar. Benutzt man (3.5) zur Abkürzung, so erhält man aus (3.4a) und (3.4b) die Rekursionsformeln

$$(3.6)\quad \alpha_{\nu+1} = \rho \cdot \frac{\varkappa\beta_\nu - \gamma_\nu}{1 + \beta_\nu};\quad \beta_{\nu+1} = \alpha_{\nu+1}\left(\mu - \rho \cdot \frac{\varkappa + \gamma_\nu}{1 + \beta_\nu}\right) + \rho^2 \cdot \frac{\beta_\nu}{1 + \beta_\nu}$$

und daher für $\gamma_{\nu+2}$ entsprechend der Gleichung für $\alpha_{\nu+1}$ (Ver-

tauschung von α mit γ und $\varkappa$ mit μ; ν durch $\nu+1$ ersetzt) die Rekursionsformel

(3.7)
$$\gamma_{\nu+2}=\rho\cdot\frac{\mu\beta_{\nu+1}-\alpha_{\nu+1}}{1+\beta_{\nu+1}}=\frac{\rho\cdot\alpha_{\nu+1}\left(\mu^2-1-\rho\mu\cdot\frac{\varkappa+\gamma_\nu}{1+\beta_\nu}\right)+\rho^3\mu\frac{\beta_\nu}{1+\beta_\nu}}{1+\beta_{\nu+1}}.$$

(Bedeutung der Korrekturgröße $\gamma_{\nu+2}$ entsprechend derjenigen von γ_ν in (3.3); im Zähler der rechten Seite wurde $\beta_{\nu+1}$ mit Hilfe von (3.6) eliminiert).

Folgerungen: a) Unter Vernachlässigung höherer Potenzen der Korrekturgrößen β_ν und γ_ν folgt aus (3.6) bzw. (3.7) zunächst

(3.8)
$$\beta_{\nu+1}=\alpha_{\nu+1}(\mu-\rho\varkappa)+\rho^2\beta_\nu;\quad \gamma_{\nu+2}=\rho\alpha_{\nu+1}(\mu^2-1-\rho\mu\varkappa)+\rho^3\mu\beta_\nu$$

und ganz entsprechend

(3.9)
$$\beta_{\nu+2}=\gamma_{\nu+2}(\varkappa-\rho\mu)+\rho^2\beta_{\nu+1};\ \alpha_{\nu+3}=\rho\gamma_{\nu+2}(\varkappa^2-1-\rho\mu\varkappa)+\rho^3\varkappa\beta_{\nu+1}.$$

Durch Einsetzen von (3.8) in (3.9) könnte man nun $\beta_{\nu+2}$ und $\alpha_{\nu+3}$ unmittelbar aus β_ν und $\alpha_{\nu+1}$ erhalten in der Gestalt

(3.9a)
$$\begin{aligned}\beta_{\nu+2}&=c_{11}\beta_\nu+c_{12}\alpha_{\nu+1};\\ \alpha_{\nu+3}&=c_{21}\beta_\nu+c_{22}\alpha_{\nu+1},\end{aligned}$$

wobei die c_{ik} nur noch von μ; $\varkappa$ und ρ abhängen. Daraus folgt sicher $\beta_\nu\to 0$ und $\alpha_\nu\to 0$ mit $\nu\to\infty$, wenn sich für ein geeignetes positives λ aus $|\beta_\nu|<\varepsilon$; $|\alpha_{\nu+1}|<\lambda\varepsilon$ immer $|\beta_{\nu+2}|<q\varepsilon$ und $|\alpha_{\nu+3}|<q\cdot\lambda\varepsilon$ mit einem festen Faktor $q<1$ ergibt. Dies ist sicher der Fall, wenn zugleich

$$|c_{11}|+\lambda|c_{12}|<q,\ \text{also}\ \lambda<\frac{q-|c_{11}|}{|c_{12}|}\quad\text{und}$$
$$|c_{21}|+\lambda|c_{22}|<q\lambda,\ \text{also}\ \lambda>\frac{|c_{21}|}{q-|c_{22}|}\ \text{mit}\ |c_{22}|<q$$

für ein geeignetes positives λ gilt, d. h. für denjenigen Bereich der Größen μ; $\varkappa$ und ρ, für den die Ungleichung

(3.10)
$$\frac{1-|c_{11}|}{|c_{12}|}>\frac{|c_{21}|}{1-|c_{22}|}\ \text{mit}\ |c_{22}|<1$$

gilt (der Grenzfall, daß $c_{12} = 0$; $|c_{11}| < 1$; $|c_{22}| < 1$, sei dabei sinngemäß eingeschlossen).

In diesem Bereich konvergiert bei hinreichend kleinen Beträgen von β_1 und α_2 das angegebene Verfahren immer (vgl. hierzu auch die Bemerkung vor Satz 1). Wir beschränken uns jedoch für das Weitere auf die folgenden Abschätzungen: Es sei

$$|\varkappa| < 2; \qquad |\mu| < 2 \tag{3.11}$$

(was bei reellen $\varkappa$; μ und R nach (3.1) nichts anderes bedeutet als zwei Paare konjugiert komplexer – nicht reeller – Nullstellen von F). Ferner sei

$$|\beta_\nu| < 2\varepsilon; \quad |\alpha_{\nu+1}| < 3\rho \cdot \varepsilon. \tag{3.12}$$

Aus (3.8) folgt dann ($\varkappa$; μ und ρ weiterhin reell vorausgesetzt)

$$|\beta_{\nu+1}| < 2\varepsilon(3\rho + 4\rho^2); \quad |\gamma_{\nu+2}| < 3\rho\varepsilon \cdot \left(3\rho + \tfrac{16}{3}\rho^2\right). \tag{3.13}$$

Beide Klammern sind kleiner als 0,99 für $\rho < \frac{1}{4,3}$. Durch Vergleich mit (3.12) ergeben sich dann, indem man die entsprechende Rechnung unter Berücksichtigung von (3.9) statt (3.8) nochmal durchführt, die Ungleichungen

$$|\beta_{\nu+2}| < 2\varepsilon \cdot 0{,}99^2; \quad |\alpha_{\nu+3}| < 3\rho\varepsilon \cdot 0{,}99^2.$$

Bei Verwendung der exakten Rekursionsformeln (3.6) und (3.7) hätte man statt dessen

$$|\beta_{\nu+2}| < 2\varepsilon \cdot 0{,}99^2 + M\varepsilon^2; \quad |\alpha_{\nu+3}| < 3\rho\varepsilon \cdot 0{,}99^2 + M\rho\varepsilon^2 \tag{3.13a}$$

mit einer von ε unabhängigen Schranke M (wenn nur ε klein genug ist, daß man innerhalb des Konvergenzbereichs der Reihenentwicklungen bleibt, von denen (3.8) und (3.9) die ersten Näherungen geben). Aus (3.13a) folgt nun aber für jedes $\rho < \frac{1}{4,3}$ bei hinreichend kleinem ε sofort

$$|\beta_{\nu+2}| \leqq q \cdot 2\varepsilon \text{ und } |\alpha_{\nu+3}| \leqq q \cdot 3\rho\varepsilon \text{ mit } (0 <) q < 1. \tag{3.13b}$$

Vergleicht man dies wieder mit (3.12), so erhält man folgenden

Satz 1: Das in II. zusammengestellte Rechenschema konvergiert bei hinreichend guten Anfangswerten für die Koeffizienten des Faktorpolynoms $K(x)$ immer dann, wenn das gegebene Polynom 4. Grades $F(x)$ zwei Paare konjugiert komplexer Wurzeln besitzt, deren Beträge sich mindestens wie 4,3 : 1 verhalten.[1]

Der Vergleich von (3.13) mit (3.12) gibt besonders für kleinere ρ zugleich auch eine Abschätzung für die Güte der Konvergenz.

Unter der zusätzlichen Annahme $\mu\varkappa > 0$ (was bei reellen μ; $\varkappa$ und R nach (3.1) zwei Paare konjugiert komplexer Wurzeln bedeutet, deren Realteile dasselbe Vorzeichen haben) genügen wesentlich schwächere Voraussetzungen für ρ, wie sich folgendermaßen ergibt:

Man kann ohne Einschränkung der Allgemeinheit $|\mu| < |\varkappa|$ annehmen. Ferner läßt sich der Fall negativer μ und $\varkappa$ nach (3.8) und (3.9) auf denjenigen positiver μ und $\varkappa$ zurückführen, indem man die β_ν durch $(-\beta_\nu)$ ersetzt. Dementsprechend sei zunächst (1. Fall)

$$1{,}6 < \mu < \varkappa < 2 \quad \text{und} \tag{3.14}$$

$$(0 <)\ \rho < 0{,}5. \tag{3.14a}$$

Dann ist

$$\mu - \rho\varkappa \begin{cases} < \varkappa(1-\rho) < 2\cdot(1-\rho) \\ > 1{,}6 - 2\rho > 0{,}6 \quad \text{nach (3.14a)}, \end{cases}$$

also

$$0{,}6 < \mu - \rho\varkappa < 2\cdot(1-\rho) \tag{3.14b}$$

und daher

$$\mu^2 - 1 - \rho\varkappa\mu = \mu\cdot(\mu - \rho\varkappa) - 1 \begin{cases} < 2\cdot 2\cdot(1-\rho) - 1 = 3 - 4\rho \\ > 1{,}6\cdot 0{,}6 - 1 = -0{,}04, \end{cases}$$

[1] Dieses Ergebnis ist bereits schärfer als die entsprechende Aussage bei Friedman, wo (S. 205) $\rho < 1:5$ und dazu noch gleiches Vorzeichen des Realteils bei allen Wurzeln verlangt wird (allerdings mit der Zusatzbemerkung, daß diese hinreichende Bedingung zweifellos noch durch eine bedeutend schwächere ersetzt werden könnte).

also, weil $3 - 4\rho > 1 \ (> 0{,}04)$ nach (3.14a),

(3.14c) $$|\mu^2 - 1 - \rho\mu\varkappa| < 3 - 4\rho.$$

Aus (3.8), (3.12) und (3.14b) folgt dann

(3.15a) $$|\beta_{\nu+1}| < 3\rho\varepsilon \cdot 2(1-\rho) + \rho^2 \cdot 2\varepsilon = 2\varepsilon(3\rho - 2\rho^2) < 2\varepsilon,$$

weil $\frac{d}{d\rho}(3\rho - 2\rho^2) = 3 - 4\rho > 0$

nach (3.14a) und weil daher $3\rho - 2\rho^2$ im abgeschlossenen Bereich (3.14a)[1] seine obere Grenze für $\rho = 0{,}5$ annimmt.

Aus (3.8), (3.12), (3.14) und (3.14c) folgt ferner

(3.15b) $$\begin{aligned} |\gamma_{\nu+2}| &< 3\rho^2\varepsilon \cdot (3 - 4\rho) + 2\rho^3 \cdot 2\varepsilon \\ &= 3\rho\varepsilon\left(3\rho - \frac{8}{3}\rho^2\right) < \frac{5\rho\varepsilon}{2}, \end{aligned}$$

weil $\frac{d}{d\rho}\left(3\rho - \frac{8}{3}\rho^2\right) = 3 - \frac{16}{3}\rho > 0$

nach (3.14a) und weil daher $\left(3\rho - \frac{8}{3}\rho^2\right)$ im abgeschlossenen Bereich (3.14a) seine obere Grenze für $\rho = 0{,}5$ erreicht.

Nun folgt aber aus (3.14) und (3.14a) auch noch

(3.15c) $$\varkappa - \rho\mu \begin{cases} < 2 - 1{,}6\rho \\ > \varkappa - \rho\varkappa > 1{,}6\,(1-\rho) > 0{,}8 \end{cases}$$

und

$$\varkappa^2 - 1 - \rho\mu\varkappa \begin{cases} > \varkappa^2 - 1 - \rho\varkappa^2 = \varkappa^2(1-\rho) - 1 > 1{,}6^2 \cdot 0{,}5 - 1 > 0 \\ < \varkappa^2 - 1 - \rho \cdot 1{,}6\varkappa < 3 - 3{,}2\rho, \end{cases}$$

weil $\varkappa^2 - 1 - \rho \cdot 1{,}6\varkappa$ nach (3.14a) für $1{,}6 < \varkappa < 2$ monoton mit $\varkappa$ wächst und dabei seine obere Grenze für $\varkappa = 2$ annimmt. Also ist

(3.15d) $$0 < \varkappa^2 - 1 - \rho\mu\varkappa < 3 - 3{,}2\rho.$$

Aus (3.9), (3.15b), (3.15c) und (3.15a) folgt dann

(3.15e) $$|\beta_{\nu+2}| < \frac{5\rho\varepsilon}{2}(2 - 1{,}6\rho) + \rho^2 \cdot 2\varepsilon = 2\varepsilon\left(\frac{5\rho}{2} - \rho^2\right) < 2\varepsilon$$

[1] d. h. im Bereich $0 \leqq \rho \leqq 0{,}5$.

(nach (3.14a), ähnlich wie in der Begründung zu (3.15a) und (3.15b)); ferner ebenso unter Berücksichtigung von (3.9), (3.15a), (3.15b), (3.15d) und (3.14)

$$(3.15f) \quad |\alpha_{\nu+3}| < \frac{5\rho^2\varepsilon}{2} \cdot (3 - 3{,}2\rho) + \rho^3 \cdot 4\varepsilon = \rho\varepsilon\left(\frac{15}{2}\rho - 4\rho^2\right) < \frac{11}{4}\rho\varepsilon.$$

Der Vergleich von (3.15e) und (3.15f) mit (3.12) gibt wieder die Konvergenz (hinreichend kleine Beträge von β_1 und α_2 vorausgesetzt). Nun sei (2. Fall)

$$(3.16) \qquad 0 < \mu < 1{,}6; \quad \mu < \varkappa < 2 \quad \text{und}$$

$$(3.16a) \qquad (0 <)\ \rho < 0{,}44.$$

Dann ist (α) $\rho(\mu - \rho\varkappa) < \rho(\mu - \rho\mu) < \rho(1 - \rho) \cdot 1{,}6 < 0{,}4$, weil für $0 < \rho < 0{,}5$, also erst recht für (3.16a) immer

$$\frac{d}{d\rho}(\rho - \rho^2) = 1 - 2\rho > 0, \text{ also } \rho(1 - \rho) < \frac{1}{4}.$$

Außerdem ist dann aber

$$(\beta) \quad \rho(\rho\varkappa - \mu) < \rho^2\varkappa < \rho^2 \cdot 2 < 0{,}44^2 \cdot 2 < 0{,}4,$$

also, (α) und (β) zusammengefaßt,

$$(3.16b) \qquad \rho|\mu - \rho\varkappa| < 0{,}4.$$

Schließlich folgt aus (3.16) und (3.16a) auch noch

$$(3.16c) \quad \mu^2 - 1 - \rho\mu\varkappa < \mu^2 - 1 - \rho\mu^2 = \mu^2(1 - \rho) - 1 < 1{,}56 - 2{,}56\rho$$

und $\qquad 1 + \rho\mu\varkappa - \mu^2 < 1 + \rho\mu \cdot 2 - \mu^2.$

Aus (3.8), (3.12) und (3.16b) folgt nun

$$(3.17a) \qquad |\beta_{\nu+1}| < 3\varepsilon \cdot 0{,}4 + \rho^2 \cdot 2\varepsilon < 1{,}6\varepsilon \qquad \text{nach (3.16a).}$$

Aus (3.8), (3.12), (3.16c) und (3.16) folgt im Fall $0 < \mu^2 - 1 - \rho\mu\varkappa$

$$\begin{aligned} |\gamma_{\nu+2}| &< 3\rho^2\varepsilon \cdot (1{,}56 - 2{,}56\rho) + \rho^3 \cdot 1{,}6 \cdot 2\varepsilon \\ &< 3\rho\varepsilon \cdot (1{,}56\rho - 1{,}49\rho^2) \\ &< 3\rho\varepsilon \cdot (1{,}56 \cdot 0{,}44 - 1{,}49 \cdot 0{,}44^2)^{1} < 3\rho\varepsilon \cdot 0{,}4, \end{aligned}$$

[1] Nach (3.16a) wegen $\frac{d}{d\rho}(1{,}56\rho - 1{,}49\rho^2) > 0$ für $\rho < 0{,}44$.

dagegen im Fall $o > \mu^2 - 1 - \rho\mu\varkappa$

$$\begin{aligned}|\gamma_{\nu+2}| &< 3\rho^2\varepsilon \cdot (1 + 2\rho\mu - \mu^2) + \rho^3 \cdot 2\mu\varepsilon \\ &< 3\rho^2\varepsilon \cdot \left[1 - \mu^2 + 2\mu\left(1 + \frac{1}{3}\right) \cdot 0{,}44\right] \\ &< 3\rho^2\varepsilon \cdot \left[1 + \frac{1}{4} \cdot \left(\frac{8}{3} \cdot 0{,}44\right)^2\right],\end{aligned}$$

weil Max. $[1 - \mu^2 + c\mu] = 1 + \frac{c^2}{4}$ für festes c; damit hat man in diesem Fall $\left(\text{mit } c = \frac{8}{3} \cdot 0{,}44 < 1{,}2\right)$ die Ungleichung

(3.17b) $$|\gamma_{\nu+2}| < 3\rho \cdot 0{,}44\varepsilon \cdot \left[1 + \frac{1{,}2^2}{4}\right] = 3\rho \cdot 0{,}44\varepsilon \cdot 1{,}36 < 1{,}8\rho\varepsilon.$$

Nun folgt aus (3.16) und (3.16a) aber auch noch

(3.17c) $$o < \varkappa - \rho\mu < 2 \qquad \text{und}$$

(3.17d) $$\varkappa^2 - 1 - \rho\varkappa\mu \begin{cases} < \varkappa^2 - 1 < 3 \\ > \varkappa^2 - 1 - \rho\varkappa^2 > -1. \end{cases}$$

Aus (3.9), (3.17b), (3.17c) und (3.17a) folgt dann

(3.17e) $$|\beta_{\nu+2}| < 1{,}8\rho\varepsilon \cdot 2 + \rho^2 \cdot 1{,}6\varepsilon < 1{,}9\varepsilon \quad \text{nach (3.16a)}$$

und unter Berücksichtigung von (3.17d) und (3.16)

(3.17f) $$|\alpha_{\nu+3}| < 1{,}8\rho^2\varepsilon \cdot 3 + \rho^3 \cdot 2 \cdot 1{,}6\varepsilon < 3\rho\varepsilon \quad \text{nach (3.16a)}.$$

Der Vergleich mit (3.12) ergibt wieder (unter der Voraussetzung hinreichend kleiner Beträge von β_1 und α_2) die Konvergenz.

Auf Grund der Bemerkungen vor (3.14) lassen sich nun aber die Ergebnisse der Abschätzungen zu (3.14) und (3.16) zusammenfassen in dem folgenden

Satz 2: Das in II. zusammengestellte Rechenschema konvergiert bei hinreichend guten Anfangswerten für die Koeffizienten des Faktorpolynoms $K(x)$ immer dann, wenn das gegebene Polynom 4. Grades $F(x)$ zwei Paare konjugiert komplexer Wurzeln besitzt, deren Beträge sich mindestens wie $(1 : 0{,}44 \approx)$ $2{,}3 : 1$ verhalten und deren Realteile dasselbe Vorzeichen haben (Vgl. auch hierzu Anm. S. 88, d. h. die Friedmansche Bedingung $\frac{1}{\rho} > 1 : 0{,}2 = 5 : 1$).

Aus (3.2) und (3.5) folgt $\rho + \frac{1}{\rho} = B - \mu\varkappa$ mit $\frac{1}{\rho} > 1$; $0 < \rho < 1$ (reelles R wie bisher vorausgesetzt). Aus der Voraussetzung

$$B > 6{,}5 \tag{3.18}$$

folgt daher

$$\text{für } \mu\varkappa < 4 \quad \text{sofort} \quad \rho + \frac{1}{\rho} > 2{,}5; \quad \rho < 0{,}5 \quad (\alpha),$$

$$\text{für } \mu\varkappa < 3{,}2 \text{ dagegen } \rho + \frac{1}{\rho} > 3{,}3; \quad \rho < 0{,}4 \quad (\beta),$$

$$\text{für } \mu\varkappa < 0 \quad \text{sogar} \quad \rho + \frac{1}{\rho} > 6{,}5; \quad \rho < 0{,}2 \quad (\gamma).$$

Wir unterscheiden nun wieder die Fälle

$$1{,}6 < \mu < \varkappa < 2 \qquad \text{und} \tag{3.14}$$

$$0 < \mu < 1{,}6; \quad \mu < \varkappa < 2. \tag{3.16}$$

Aus (3.14) und (3.18) folgt dann nach (α) sofort $\rho < 0{,}5$, d. h. (3.14a), was („1. Fall") zusammen mit (3.14) für den Konvergenzbeweis genügte. Aus (3.16) und (3.18) folgt nach (β) dagegen sogar $\rho < 0{,}4$, also erst recht (3.16a), d. h. $\rho < 0{,}44$, was wieder („2. Fall") zusammen mit (3.16) die Konvergenz ergab.

Die Fälle (3.14) und (3.16) zusammengefaßt besagen aber nichts anderes als die Voraussetzung

$$0 < \mu < \varkappa < 2.$$

Dies führt also zusammen mit (3.18) in jedem Fall zur Konvergenz. Nach der Bemerkung vor (3.14), S. 88, genügt dazu nun aber auch (3.18) und

$$0 < |\mu| < 2; \quad 0 < |\varkappa| < 2, \quad \text{solange nur } \mu\varkappa > 0.^1$$

(3.18) und $\mu\varkappa < 0$ gibt jedoch nach (γ) zunächst $\rho < 0{,}2$ und daher erst recht $\rho < \frac{1}{4{,}3}$, d. h. die Voraussetzung von Satz 1, wenn man an 2 Paaren konjugiert komplexer Wurzeln festhält. Daher genügt (3.18) unter dieser Voraussetzung (2 Paare konj. kompl. Wurzeln) immer und es gilt als zusammenfassende Abschätzung der

[1] (μ sowie $\varkappa$ reell vorausgesetzt).

Satz 3: Das in II. zusammengestellte Rechenschema konvergiert bei hinreichend guten Anfangswerten und 2 Paaren konjugiert komplexer Wurzeln immer, wenn in (2.1) $B > 6{,}5$ ist.

Geht man statt von (2.1) von der allgemeinen Gleichung 4. Grades $\xi^4 + a\xi^3 + b\xi^2 + c\xi + d = 0$ aus, so lautet die entsprechende (hinreichende) Konvergenzbedingung $b : \sqrt{d} > 6{,}5$ (vgl. letzten Abschnitt vor III.).

Ebenso wie Satz 3 bloß eine hinreichende Konvergenzbedingung gibt, bedeutet z. B. auch $\rho < 0{,}44$ nur eine hinreichende, aber keine notwendige Bedingung für die Gültigkeit des Satzes 2 bei sonst unveränderten Voraussetzungen. Daß jedoch das in II. besprochene Verfahren nicht immer brauchbar ist, zeigt das Beispiel

$$\mu = 0; \quad \varkappa = 2.$$

Aus (3.8) bzw. (3.9) folgt dann nämlich

(a) $\beta_{\nu+1} = -2\rho\alpha_{\nu+1} + \rho^2\beta_\nu$; $\gamma_{\nu+2} = -\rho\alpha_{\nu+1}$ und daher

(b) $\beta_{\nu+2} = -2\rho\alpha_{\nu+1}(1+\rho^2) + \rho^4\beta_\nu$; $\alpha_{\nu+3} = -\rho^2\alpha_{\nu+1}(3+4\rho^2) + 2\rho^5\beta_\nu$.

Nun sei

(c) $$\beta_\nu\,\alpha_{\nu+1} > 0 \qquad \text{und}$$

(d) $$|\beta_\nu| < 2\rho(1+\rho^2)\cdot|\alpha_{\nu+1}|.$$

Dann ist nach (3.5), d. h. wegen $(0 <)\rho < 1$, erst recht

(e) $$\rho^4|\beta_\nu| < 2\rho(1+\rho^2)\cdot|\alpha_{\nu+1}|,$$

außerdem aber (wieder wegen (d))

(f) $$2\rho^5|\beta_\nu| < \rho^2(4\rho^4+4\rho^6)\cdot|\alpha_{\nu+1}| < \rho^2(3+4\rho^2)\cdot|\alpha_{\nu+1}|$$

für

$$\rho^2 \leqq 0{,}9$$

(es ist nämlich $4\rho^4 + 4\rho^6 < 3 + 4\rho^2$, falls $\rho^2 + \rho^4 < \frac{3}{4\rho^2} + 1$, was für $\rho^2 = 0{,}9$ und damit erst recht für $\rho^2 < 0{,}9$ erfüllt ist).

Nach (e) und (f) ist für die Vorzeichen von $\beta_{\nu+2}$ und $\alpha_{\nu+3}$ wegen (b) das Vorzeichen von $(-\alpha_{\nu+1})$ maßgebend. Es gilt also

(g) $\beta_{\nu+2}\,\alpha_{\nu+3} > 0$ sowie

(wegen (b) und (c))

(h) $|\beta_{\nu+2}| < 2\rho(1+\rho^2)\cdot|\alpha_{\nu+1}| < 2\rho(1+\rho^2)\cdot|\alpha_{\nu+3}|$,
wenn nur $|\alpha_{\nu+1}| < |\alpha_{\nu+3}|$.

Nun ist aber nach (b), (c) und (f) sicher

(i) $|\alpha_{\nu+3}| > |\alpha_{\nu+1}|\cdot[(3\rho^2+4\rho^4) - 2\rho^5(2\rho+2\rho^3)] > |\alpha_{\nu+1}|$

für $0{,}27 < \rho^2 < 0{,}8$

(es ist nämlich $3\rho^2+4\rho^4-4\rho^6-4\rho^8 > 1$, falls $3\rho^2+4\rho^4-1 > 4\rho^6(1+\rho^2)$, d. h. (Division durch $1+\rho^2$), falls $4\rho^2-1 > 4\rho^6$ (erfüllt für $0{,}27 < \rho^2 < 0{,}8$)).

Unter den für (f) und (i) benötigten (miteinander verträglichen) Voraussetzungen für ρ^2 ist also nach (i) der Betrag von $\alpha_{\nu+3}$ größer als derjenige von $\alpha_{\nu+1}$. Entsprechendes folgt durch Vergleich von (g) und (h) mit (c) und (d) mittels vollständiger Induktion dann aber auch für alle größeren ν, solange man die für Korrekturgrößen geringen Betrags gültigen Näherungen (3.8) und (3.9) benutzen kann. Dabei konnten β_ν und $\alpha_{\nu+1}$ dem Betrag nach beliebig klein sein, wenn nur die Ungleichungen (c) und (d) galten.

b) Setzt man als erste Näherung (indem man x^4 und x^3 in (2.1) vernachlässigt)

(3.19) $$K_1 = Bx^2 + Cx + 1,$$

so folgt aus (3.2) und (3.3) $\gamma_1 = \mu R^{-2}$ und $\beta_1 = \mu\varkappa R^{-2} + R^{-4}$. Nach (3.5) bedeutet dies

(3.20)

$\gamma_1 = \mu\rho$; $\beta_1 = \varkappa\mu\rho + \rho^2 = \varkappa\gamma_1 + \rho^2$, also $\alpha_2 = \frac{(\varkappa^2-1)\cdot\mu\rho^2 + \varkappa\rho^3}{1+\varkappa\mu\rho+\rho^2}$

nach (3.6).

Wegen der Symmetrie der gestellten Aufgabe und der in II. zusammengestellten Rekursionsformeln bezüglich der Faktorpolynome G und K kann man sich auch hier wieder auf die Annahme $|\mu| < |\varkappa|$ beschränken; im Fall $|\varkappa| < |\mu|$ würde man nämlich zweckmäßigerweise sowieso statt mit einer ersten Näherung für

K mit einer ersten Näherung für G beginnen, indem man $Cx + 1$ statt $x^4 + Ax^3$ in $F(x)$ wegläßt, da dann (bei unverändertem β_1) an Stelle der Korrekturgröße $\gamma_1 = \mu\rho$ die absolut genommen kleinere Korrekturgröße $\alpha_1 = \varkappa\rho$ tritt.

Darüber hinaus kann man sich auch hier wieder auf $\mu > 0$ beschränken; denn ersetzt man in (3.20) zugleich μ durch $(-\mu)$ und $\varkappa$ durch $(-\varkappa)$, so wechseln γ_1 und α_2 ihr Vorzeichen, während β_1 unverändert bleibt; andererseits bleiben aber die Rekursionsformeln (3.6) und (3.7) erhalten, wenn man bei μ und $\varkappa$ sowie bei den Korrekturgrößen γ und α das Vorzeichen wechselt, während man die β unverändert läßt.

Zunächst sei nun (mit später genauer zu bestimmenden μ_0 und ρ_0)

$$(3.21) \qquad 1 < \mu_0 < \mu < \varkappa < 2 \quad \text{und} \quad (0 <)\ \rho < \rho_0\ (< 1).$$

Nach (3.20) ist dann einerseits

$$(3.22) \qquad 2\beta_1 > \gamma_1 > 0 \quad \text{und} \quad \alpha_2 > 0.$$

Andererseits folgt aus der Annahme, für ein bestimmtes ν sei

$$(3.23) \qquad 2\beta_\nu > \gamma_\nu > 0 \quad \text{und} \quad \alpha_{\nu+1} > 0,$$

wegen $\varkappa < 2$ zunächst

$$(3.23\,a) \qquad \frac{\varkappa + \gamma_\nu}{1 + \beta_\nu} < \frac{2 + 2\beta_\nu}{1 + \beta_\nu} = 2.$$

Nach (3.6), (3.21) und (3.23) folgt daraus aber

$$(3.23\,b) \quad \beta_{\nu+1} > \alpha_{\nu+1} \cdot (\mu_0 - \rho_0 \cdot 2) > \frac{\alpha_{\nu+1}}{2}, \text{ falls nur } \mu_0 - 2\rho_0 > \frac{1}{2}.$$

Ferner ergibt sich aus (3.6) und (3.23) die Ungleichung

$$\alpha_{\nu+1} < \rho \cdot \frac{\varkappa \beta_\nu}{1 + \beta_\nu}, \qquad \text{also}$$

$$\frac{\rho \cdot \beta_\nu}{1 + \beta_\nu} > \frac{\alpha_{\nu+1}}{\varkappa} > \frac{\alpha_{\nu+1}}{2} \text{ nach (3.21)}$$

und damit

$$(3.23\,c) \qquad \frac{\rho^3 \mu \beta_\nu}{1 + \beta_\nu} > \frac{\rho^2 \cdot \mu \alpha_{\nu+1}}{2}.$$

Aus (3.23a) und (3.21) folgt andererseits

$$(3.23\,d) \qquad \mu^2 - 1 - \rho\mu \cdot \frac{\varkappa + \gamma_\nu}{1+\beta_\nu} > \mu^2 - 1 - \rho\mu \cdot 2.$$

Nach (3.23) und (3.23b) ist nun $\alpha_{\nu+1} > 0$ und $\beta_{\nu+1} > 0$.[1] Aus (3.7) ergibt sich dann wegen (3.23d) und (3.23c) schließlich

$$\gamma_{\nu+2} > \frac{1}{1+\beta_{\nu+1}} \cdot \left[\rho\,\alpha_{\nu+1}\,(\mu^2 - 1 - \rho\mu \cdot 2) + \rho^2 \cdot \mu\,\frac{\alpha_{\nu+1}}{2}\right]$$
$$= \frac{\rho\,\alpha_{\nu+1}}{1+\beta_{\nu+1}} \cdot \left(\mu^2 - 1 - \rho\mu \cdot \frac{3}{2}\right)$$

und damit nach (3.21) und (3.23) sicher

$$(3.23\,e) \qquad \gamma_{\nu+2} > 0, \text{ wenn nur } \left(\mu_0^2 - 1 - \rho_0\mu_0 \cdot \frac{3}{2}\right) > 0.^2$$

Die in (3.23b) und (3.23e) ausgesprochenen Bedingungen für μ_0 und ρ_0 sind z. B. erfüllt, wenn

$$(3.24) \qquad \mu_0 = 1{,}6; \quad \rho_0 = 0{,}5$$

oder auch, wenn

$$(3.25) \qquad \mu_0 = 1{,}281; \quad \rho_0 = \frac{1}{3}.$$

In allen Fällen, in denen ρ, $\varkappa$ und μ den Ungleichungen (3.21) mit μ_0 und ρ_0 aus (3.24) oder aus (3.25) genügen, folgt dann durch Vergleich von (3.23b) und (3.23e) mit (3.23) mittels vollständiger Induktion wegen (3.22) die Gültigkeit von (3.23) für alle ν (Daß beim nächsten Schritt α und γ sowie $\varkappa$ und μ ihre Rollen vertauschen, macht nichts aus, da bisher die in (3.21) steckende Voraussetzung $\mu < \varkappa$ nicht benutzt wurde, sondern nur daß $\mu_0 < \mu < 2$ und $\mu_0 < \varkappa < 2$). Im besonderen sind in diesen Fällen also die sämtlichen Korrekturgrößen α_ν, β_ν und γ_ν positiv.

Es gilt also dann allgemein (3.23) und, indem man ν in (3.23b) und (3.23e) durch $(\nu - 2)$ ersetzt, ebenfalls ganz allgemein (für $\nu > 2$)

$$(3.26) \qquad 2\,\beta_{\nu-1} > \alpha_{\nu-1} > 0; \quad \gamma_\nu > 0.$$

---

[1] Wenn nur $\mu_0 - 2\rho_0 > \frac{1}{2}$ wie in (3.23b).

[2] Nach (3.21) ist $\mu^2 - 1 - \rho\mu \cdot \frac{3}{2} > \mu^2 - 1 - \rho_0 \cdot \mu \cdot \frac{3}{2}$ und $\frac{d}{d\mu}(\mu^2 - 1 - \rho_0 \cdot \mu \cdot \frac{3}{2}) = 2\mu - \rho_0 \cdot \frac{3}{2} > 2 - \frac{3}{2} > 0$, also $\mu^2 - 1 - \rho \cdot \mu \cdot \frac{3}{2} > \mu_0^2 - 1 - \rho_0\mu_0 \cdot \frac{3}{2}$.

Daraus und aus (3.21) folgt zunächst entsprechend (3.23a) auch

$$\frac{\mu+\alpha_{\nu-1}}{1+\beta_{\nu-1}} < 2, \tag{3.26a}$$

weiterhin aber

$$\varkappa - \rho\cdot\frac{\mu+\alpha_{\nu-1}}{1+\beta_{\nu-1}} < 2-\rho\cdot\frac{\mu_0}{1+\beta_{\nu-1}} \quad \text{und} \tag{3.26b}$$

$$\varkappa^2 - 1 - \rho\varkappa\cdot\frac{\mu+\alpha_{\nu-1}}{1+\beta_{\nu-1}} < 3-\rho\cdot 2\cdot\frac{\mu_0}{1+\beta_{\nu-1}}, \tag{3.26c}$$

weil die Ableitung der linken Seite nach $\varkappa$ wegen (3.26a) und (3.21) sicher positiv ist.

In Anlehnung an (3.15a) und (3.15b) sei nun (für ein bestimmtes ν)

$$0<\beta_{\nu-1}<2\varepsilon;\quad 0<\gamma_\nu<\frac{5}{2}\cdot\frac{\rho\cdot\varepsilon}{1+\beta_{\nu-1}}. \tag{3.27}$$

Durch sinngemäße Anwendung[1] von (3.6) bzw. (3.7) folgen dann wegen (3.26b) bzw. (3.26c) die Ungleichungen

$$\beta_\nu < \left[\frac{5}{2}\frac{\rho\cdot\varepsilon}{1+\beta_{\nu-1}}\cdot\left(2-\frac{\rho\mu_0}{1+\beta_{\nu-1}}\right)+\frac{2\varepsilon\rho^2}{1+\beta_{\nu-1}}\right] \quad \text{bzw.} \tag{3.27a}$$

$$\alpha_{\nu+1} < \frac{1}{1+\beta_\nu}\cdot\left[\frac{5}{2}\cdot\frac{\rho^2\varepsilon}{1+\beta_{\nu-1}}\left(3-\frac{2\rho\mu_0}{1+\beta_{\nu-1}}\right)+\frac{4\varepsilon\rho^3}{1+\beta_{\nu-1}}\right]. \tag{3.27b}$$

Dabei gelte weiterhin entweder (3.24) oder (3.25). In den beiden letzten Zeilen wird dann die eckige Klammer (für feste Werte von ρ; ε; μ_0) am größten, wenn der Faktor $u=\frac{1}{1+\beta_{\nu-1}}$ am größten, d. h. wegen (3.27), wenn $u=1$ ist; denn die Ableitung von

$$(5+2\rho)\,u-\frac{5}{2}\rho\mu_0 u^2 \qquad \text{(vgl. (3.27a))}$$

und von

$$\left(\frac{15}{2}+4\rho\right)u-5\rho\mu_0 u^2 \qquad \text{(vgl. (3.27b))}$$

nach u ist für $0<u<1$ durchweg positiv, wenn (3.24), d. h. $\mu_0=1{,}6$; $\rho<0{,}5$, oder auch wenn (3.25), d. h. $\mu_0=1{,}281$; $\rho<\frac{1}{3}$

[1] μ mit $\varkappa$, α mit γ vertauscht und ν durch $(\nu-1)$ ersetzt.

vorausgesetzt ist. Aus (3.27a) wird damit (indem man $\beta_{\nu-1} = 0$ einsetzt)

$$(3.27c) \qquad (0 <) \beta_\nu < \frac{5}{2} \rho\varepsilon \cdot (2 - \rho\mu_0) + 2\rho^2\varepsilon < 2\varepsilon$$

(wie bei (3.15e)) zunächst im Fall (3.24) und erst recht im Fall (3.25). Aus (3.27b) wird auf entsprechende Weise

$$(3.27d) \qquad (0 <) \alpha_{\nu+1} < \frac{1}{1+\beta_\nu} \cdot \left[\frac{5}{2}\rho^2\varepsilon(3 - 2\rho\mu_0) + 4\rho^3\varepsilon\right] < \frac{11\rho\varepsilon}{4(1+\beta_\nu)}$$

(wie bei (3.15f)) zunächst im Fall (3.24) und erst recht im Fall (3.25). Nach (3.21) und (3.23) ist weiter

$$(3.28a) \qquad \mu - \rho \cdot \frac{\varkappa + \gamma_\nu}{1+\beta_\nu} < \varkappa - \rho \cdot \frac{\varkappa}{1+\beta_\nu} < 2 - \frac{2\rho}{1+\beta_\nu}$$

und

$$(3.28b) \qquad \mu^2 - 1 - \rho\mu \cdot \frac{\varkappa+\gamma_\nu}{1+\beta_\nu} < \varkappa^2 - 1 - \rho\varkappa \cdot \frac{\varkappa}{1+\beta_\nu} < 3 - \frac{4\rho}{1+\beta_\nu},$$

weil wegen (3.23a) und (3.21) die Ableitung der linken Seite nach μ und die Ableitung der Mitte nach $\varkappa$ immer positiv ist.

Aus (3.6) bzw. (3.7), (3.27c) und (3.27d) folgt damit

$$(3.28c) \qquad (0 <) \beta_{\nu+1} < \frac{11\rho\varepsilon}{4(1+\beta_\nu)} \cdot \left(2 - \frac{2\rho}{1+\beta_\nu}\right) + \frac{\rho^2 \cdot 2\varepsilon}{1+\beta_\nu} \quad \text{bzw.}$$

$$(3.28d) \qquad (0 <) \gamma_{\nu+2} < \frac{\rho}{1+\beta_{\nu+1}} \cdot \left[\frac{11\rho\varepsilon}{4(1+\beta_\nu)} \cdot \left(3 - \frac{4\rho}{1+\beta_\nu}\right) + \frac{\rho^2 \cdot 4\varepsilon}{1+\beta_\nu}\right].$$

Denkt man sich ρ und ε fest, so ist die rechte Seite von (3.28c) wieder am größten für $v = \frac{1}{1+\beta_\nu} = 1$, weil die Ableitung von $\rho\varepsilon\left\{\left(\frac{11}{2} + 2\rho\right)v - \frac{11}{2}\rho v^2\right\}$ nach v wegen $\beta_\nu > 0$ (d. h. $0 < v < 1$) und wegen $(0 <) \rho < \frac{1}{2}$ im Fall (3.24) und im Fall (3.25) durchweg positiv ist. Aus (3.28c) wird daher

$$(3.28e) \qquad (0 <) \beta_{\nu+1} < \frac{11\rho\varepsilon}{4}(2 - 2\rho) + \rho^2 \cdot 2\varepsilon$$
$$= \varepsilon \cdot \left(\frac{11}{2}\rho - \frac{7}{2}\rho^2\right) < \frac{15}{8}\varepsilon \quad \text{für} \quad \rho < \frac{1}{2}.$$

Die eckige Klammer in (3.28d) ist bei festem ε und β_ν proportional zu $g(\rho) = \frac{33}{4}\rho - \frac{11\rho^2}{1+\beta_\nu} + 4\rho^2$ und damit wegen

$$\frac{dg}{d\rho} = \frac{33}{4} - \left(\frac{22}{1+\beta_\nu} - 8\right)\rho > \frac{33}{4} - 14\rho \quad \left(> 0 \text{ für } \rho < \frac{1}{2}\right)$$

im Bereich $0 \leqq \rho \leqq \frac{1}{2}$ am größten für $\rho = \frac{1}{2}$. Die eckige Klammer selbst ist daher kleiner als

$$\frac{\varepsilon}{2}\left(\frac{33}{4}v - \frac{11}{2}v^2 + 2v\right) = \frac{\varepsilon}{2}\left(\frac{41}{4}v - \frac{11}{2}v^2\right) \leqq \frac{\varepsilon}{2} \cdot \frac{41^2}{8 \cdot 44} < \frac{5}{2}\varepsilon$$

$\left(\text{wobei wieder } \frac{1}{1+\beta_\nu} = v \text{ gesetzt wurde; Max. } (av - bv^2) = \frac{a^2}{4b}\right)$.
Aus (3.28d) folgt damit

$$(3.28\text{f}) \qquad (0 <)\ \gamma_{\nu+2} < \frac{5\rho\varepsilon}{2(1+\beta_{\nu+1})} \quad \left(\text{für } \rho < \frac{1}{2}\right).$$

Durch Vergleich von (3.28e) und (3.28f) mit (3.27) folgt dann unmittelbar die Konvergenz des Verfahrens, wenn man zu den Voraussetzungen (3.20) und (3.21) noch entweder (3.24) oder (3.25) hinzunimmt.

Es bleibt noch der Fall

$$(3.29) \qquad 0 < \mu < 1{,}29; \quad \mu < \varkappa < 2,$$

wobei wieder $\rho < \frac{1}{3}$ vorausgesetzt sei.

In diesem Fall folgt aus (3.20) unmittelbar bzw. durch einfache Abschätzungen

$$(3.29\text{a}) \quad 0 < \gamma_1 < 0{,}43;\ 0 < \beta_1 < 1;\ -0{,}1 < \alpha_2 < 0{,}3;\ -0{,}2 < \beta_2 < 0{,}4$$

(für β_2 unter Berücksichtigung von (3.6)).

Der Konvergenzbeweis unter den Voraussetzungen (3.29) läßt sich wegen der Möglichkeit negativer β_ν hier nicht ohne weiteres in Anlehnung an (3.16) führen. Man kann aber trotzdem zum Ziel kommen, indem man z. B. noch $\varkappa \gtrless 1{,}5$ unterscheidet. Im einzelnen sei jedoch der Konvergenzbeweis für (3.29) hier nicht mehr wiedergegeben, da er etwas mehr Rechnung als die bisherigen Abschätzungen erfordert, ohne wesentlich Neues zu enthalten.

Das Ergebnis läßt sich unter Berücksichtigung der vorausgegangenen Abschätzungen zusammenfassen in

Satz 4: Benutzt man (3.19) oder $G_1 = x^2 + Ax + B$ als erste Näherung, je nachdem $|\mu| \lesseqgtr |\varkappa|$, so konvergiert das in II. angegebene Verfahren bei zwei Paaren konjugiert komplexer Wurzeln mit einheitlichem Vorzeichen des Realteils immer, wenn das Verhältnis ρ des Betrags der „kleinen" zum Betrag der „großen" Wurzeln unter der Schranke $\rho = \frac{1}{3}$ bleibt.

Außerdem lassen sich nun aber auch die Überlegungen nach (3.18), die zu Satz 3 führten, zunächst bei einheitlichem Vorzeichen der Realteile aller 4 Wurzeln der gegebenen Gleichung auf den Fall übertragen, daß man von $K_1 = Bx^2 + Cx + 1$ oder von $G_1 = x^2 + Ax + B$ ausgeht

(Aus $B > 6{,}5$ und $0 < \mu < 1{,}6$; $\mu < \varkappa < 2$ folgt zwar zunächst nur $\rho + \frac{1}{\rho} > 3{,}3$ und damit $\rho < \frac{1}{2{,}95}$, also noch nicht genau das, was in (3.25) von ρ vorausgesetzt war; man erkennt jedoch aus den durchgeführten Abschätzungen sofort, daß für $\mu\varkappa \approx 3{,}2$ die Ungleichung $\rho < \frac{1}{2{,}95}$ völlig ausreicht).

Im folgenden soll jetzt aber gezeigt werden, daß die Voraussetzung einheitlicher Realteile hier ebenso unnötig ist wie in Satz 3. Es sei nämlich

$$(3.30) \qquad -4 < \mu\varkappa < 0 \text{ und } B = \rho + \mu\varkappa + \frac{1}{\rho} > 6{,}5.$$

Dann ist

$$(3.30a) \qquad \frac{1}{B} = \frac{\rho}{1 + \varkappa\mu\rho + \rho^2} < \frac{1}{6{,}5} \text{ und } \rho + \frac{1}{\rho} > 6{,}5 - \mu\varkappa \; (> 6{,}5),$$

also, da nach (3.5) notwendig $|\rho| < 1$, jedenfalls $\frac{1}{\rho} > 6{,}34$; $\rho < 0{,}16$ und damit

$$(3.31) \qquad \frac{1}{\rho} > 6{,}34 - \mu\varkappa; \quad \rho < \frac{1}{6{,}34 - \mu\varkappa} \left(< \frac{1}{6{,}34} \right).$$

Nach (3.20) und (3.30a) folgt

$$(3.32) \qquad |\alpha_2| < \frac{1}{6{,}5} |(\varkappa^2 - 1) \cdot \mu\rho + \varkappa\rho^2|.$$

Nun haben nach (3.30) die Glieder $(-\mu\rho)$ und $\varkappa\rho^2$ das entgegengesetzte Vorzeichen von $\varkappa^2\mu\rho$. Daher ist nach (3.32), (3.31) und (3.30), je nachdem $|\varkappa^2\mu\rho| \gtrless |-\mu\rho + \varkappa\rho^2|$, entweder

$$(3.32\,a)\quad |\alpha_2| < \frac{1}{6{,}5}\cdot|\varkappa^2\mu\rho| < \frac{2}{6{,}5}\,|\varkappa\mu\rho| < \frac{2}{6{,}5}\left|\frac{\mu\varkappa}{6{,}34-\mu\varkappa}\right| < \frac{2}{6{,}5}\cdot\frac{4^{1}}{10{,}34} < 0{,}12 \text{ oder}$$

$$(3.32\,b)\quad |\alpha_2| < \frac{1}{6{,}5}\cdot\left|\frac{2}{6{,}34}+\frac{2}{6{,}34^2}\right|,$$

also jedenfalls

$$(3.32\,c)\quad |\alpha_2| < 0{,}12.$$

Außerdem folgen aber aus (3.20), (3.30) und (3.31) die Ungleichungen

$$(3.33)\quad |\gamma_1| < \frac{2}{6{,}34} < 0{,}32 \text{ und } |\beta_1| < |\varkappa\mu\rho| < \frac{4^{1}}{10{,}34} \text{ oder } |\beta_1| < \rho^2,$$

also in beiden Fällen sicher $|\beta_1| < 0{,}39$.

Nun sei allgemein vorausgesetzt, für ein bestimmtes ν sei

$$(3.34)\quad |\gamma_\nu| < 0{,}32;\ |\beta_\nu| < 3\varepsilon \text{ und } |\alpha_{\nu+1}| < \varepsilon \text{ mit } \varepsilon < 0{,}13 \quad (\text{also } |\beta_\nu| < 0{,}39).$$

Nach (3.31) ist dann

$$\left|\rho\cdot\frac{\varkappa+\gamma_\nu}{1+\beta_\nu}\right| < \frac{1}{6{,}34-|\mu\varkappa|}\cdot\frac{|\varkappa|+0{,}32}{1-0{,}39} < \frac{2{,}32}{6{,}34+2|\mu|}\cdot\frac{1}{0{,}61},$$

weil $\dfrac{|\varkappa|+0{,}32}{6{,}34+|\mu\varkappa|}$ wegen $|\mu| < 2$ mit $|\varkappa|$ zunimmt;

daher ist

$$\left|\mu-\rho\cdot\frac{\varkappa+\gamma_\nu}{1+\beta_\nu}\right| < M(\mu) = |\mu| + \frac{2{,}32}{6{,}34+2|\mu|}\cdot\frac{1}{0{,}61} < 2 + \frac{2{,}32}{10{,}34\cdot 0{,}61},$$

weil $M(\mu)$ mit $|\mu|$ zunimmt.

Nach (3.6) wird damit

$$|\beta_{\nu+1}| < \varepsilon\left(2+\frac{2{,}32}{10{,}34}\cdot\frac{1}{0{,}61}\right)+\frac{3\varepsilon}{6{,}34^2}\cdot\frac{1}{0{,}61} < \varepsilon\cdot 2{,}5\,(<0{,}13\cdot 2{,}5<0{,}33).$$

---

[1] $\dfrac{y}{6{,}34-y}$ nimmt monoton zu, also $\left|\dfrac{y}{6{,}34-y}\right|$ für $-4 < y < 0$ monoton ab, wenn y wächst.

Entsprechend bekommt man jetzt aus (3.7)

$$|\gamma_{\nu+2}| < \frac{\varepsilon}{6{,}34 \cdot 0{,}67}\left[\left(3 + \frac{4{,}64}{10{,}34 \cdot 0{,}61}\right) + \frac{6}{6{,}34^2 \cdot 0{,}61}\right] < \varepsilon \cdot 0{,}95 \,(< 0{,}13).$$

Der Vergleich mit (3.34) gibt zusammen mit (3.33) und (3.32c) durch vollständige Induktion wieder den Konvergenzbeweis. Daher gilt allgemein der

Satz 5: Benutzt man (3.19) oder $G_1 = x^2 + Ax + B$ als erste Näherung, je nachdem $|\mu| \lesseqgtr |\varkappa|$, so konvergiert das in II. angegebene Verfahren bei zwei Paaren konjugiert komplexer Wurzeln immer, wenn nur $B > 6{,}5$, d. h. bei nicht „normierten" Gleichungen, wenn nur $b : \sqrt{d} > 6{,}5$.

IV. Bemerkungen zur Anwendbarkeit des Verfahrens

Für die praktische Verwendung des in II. angegebenen Verfahrens wird man im allgemeinen eine 1. Näherung für $K(x)$ durch Vernachlässigung von x^3 und x^4 oder eine 1. Näherung für $G(x)$ durch Vernachlässigung von x^1 und x^0 gewinnen. Für die Brauchbarkeit des Verfahrens ist nicht nur die Konvergenz selbst, sondern vor allem auch die Güte der Konvergenz maßgebend. Hierfür kann man unter Vernachlässigung höherer Potenzen von ρ die folgenden Näherungen gewinnen: Nach (3.20) gilt für die Korrekturgrößen β_ν und γ_ν zunächst

$$\beta_1 \sim \varkappa\gamma_1, \text{ also } \varkappa\beta_1 - \gamma_1 \sim (\varkappa^2 - 1)\gamma_1$$

und daher nach (3.6), (3.8) und (3.9) allgemein (für beschränktes ν)

$$(4.1) \qquad \begin{aligned} \gamma_{\nu+2} &\sim \rho^2 \cdot (\mu^2 - 1) \cdot (\varkappa^2 - 1) \cdot \gamma_\nu = q\gamma_\nu{}^{1} \text{ und} \\ \beta_{\nu+2} &\sim \rho^2 \cdot (\mu^2 - 1) \cdot (\varkappa^2 - 1) \cdot \beta_\nu = q\beta_\nu. \end{aligned}$$

Nach (3.1) und (3.5) bedeutet dabei

ρ das Verhältnis des Betrags der „kleinen" Wurzeln zu demjenigen der „großen" Wurzeln,

[1] D. h. $|\gamma_{\nu+2} - q\gamma_\nu| < M\rho^3$ mit beschränktem M für beschränktes ν.

$-\frac{1}{2}\mu$ das Verhältnis des Realteils zum Betrag bei den „großen" Wurzeln und

$-\frac{1}{2}\varkappa$ das Verhältnis des Realteils zum Betrag bei den „kleinen" Wurzeln

(2 Paare konjugiert komplexer Wurzeln vorausgesetzt).

Einen Überblick über das Konvergenzverhalten bekommt man unmittelbar aus $F(x)$, wenn man (3.2) berücksichtigt. Aus (3.2) und (3.5) folgt nämlich (unter Vernachlässigung höherer Potenzen von ρ)

$$(4.2)\quad \rho \sim \frac{1}{B\left(1-\frac{AC}{B^2}\right)};\quad \mu \sim \sqrt{\rho}\,(A-\rho C);\quad \varkappa \sim \sqrt{\rho}\,(C-\rho A).$$

Für das Beispiel (1.1) bedeutet dies

$$\rho \sim \frac{1}{10\left(1-\frac{10}{100}\right)} = \frac{1}{9};\quad \mu \sim \frac{2{,}5-0{,}4}{3} = 0{,}7;\quad \varkappa \sim \frac{4-0{,}3}{3} \sim 1{,}2$$

und daher nach (4.1) für den nach 2 Divisionen erzielten Konvergenzfaktor q jeweils näherungsweise den Faktor

$$q \sim \frac{-0{,}51 \cdot 0{,}44}{81} \sim -0{,}003,$$

wodurch die Güte der Konvergenz schon recht genau erfaßt ist, wie ein Vergleich mit (2.4) zeigt.

Die Näherungsregel (4.2) läßt sich natürlich auch übertragen auf Gleichungen, die nicht so „normiert" sind, daß der Faktor von x^4 und derjenige von x^0 gleich 1 ist. Ist

$$(4.3)\qquad \xi^4 + a\xi^3 + b\xi^2 + c\xi + d$$

gegeben, so ergibt sich mittels $\xi = x\sqrt[4]{d}$ und Division durch d sofort

$$(4.3\,a)\qquad B = \frac{b}{\sqrt{d}};\quad A = \frac{a}{\sqrt[4]{d}};\quad C = \frac{c}{\sqrt[4]{d^3}}.$$

Nach Satz 5 genügt $B > 6{,}5$ für die Konvergenz des Verfahrens.

Das geschilderte Näherungsverfahren hat den Nachteil, daß es sich nicht für alle Fälle von Gleichungen 4. Grades mit zwei Paaren konjugiert komplexer Wurzeln eignet; es hat außerdem den Nachteil, daß es nach (4.1) bloß „von der 1. Ordnung" konvergiert, d. h. daß der Fehler beim nächsten Schritt jeweils näherungsweise proportional dem Fehler der vorausgegangenen Näherung und nicht etwa (wie z. B. beim Newtonschen Verfahren) proportional dem Quadrat des Fehlers der vorausgegangenen Näherung ist. Das Verfahren hat aber den Vorteil einer sehr einfachen Rechenvorschrift und den Vorteil, daß man mit Hilfe von (4.3a), (4.2) und (4.1), wenn eine Gleichung 4. Grades vorliegt, die Brauchbarkeit des Verfahrens und die Güte der Konvergenz in vielen Fällen durch eine einfache Überschlagsrechnung sofort feststellen kann. Es wird vor allem dann dem Verfahren von Graeffe (das wie das Newtonsche im wesentlichen „quadratisches" Konvergenzverhalten zeigt) vorzuziehen sein, wenn die gewünschte Genauigkeit bereits durch 3 oder 4 Divisionen erreichbar ist (da sich der Vorteil des quadratischen Verhaltens bei den ersten Schritten noch nicht so auswirkt). Besonders im Fall $\mu\varkappa > 0$ (was z. B. bei Schwingungen um stabile Gleichgewichtslagen gilt, wenn es sich um die charakteristische Gleichung einer linearen Differentialgleichung handelt) wird das Verfahren häufig auch dann gut konvergieren, wenn die Beträge beider Wurzelpaare wenig voneinander abweichen.

Aus den Überlegungen in Abschnitt III. folgt übrigens zunächst für Gleichungen 4. Grades mit zwei Paaren konjugiert komplexer Wurzeln auch die Begründung für die folgende Rechenregel:

> Ist von der linken Seite der Gleichung $F = x^4 + Ax^3 + \cdots = 0$ näherungsweise ein quadratischer Faktor K für die Wurzeln mit dem kleineren Betrag bekannt, so ist das geeignete Verfahren zur Berechnung eines Näherungspolynoms für den anderen quadratischen Faktor (für die Wurzeln mit dem größeren Betrag) die gewöhnliche Division $F:K$; ist dagegen näherungsweise ein quadratischer Faktor G für die Wurzeln mit dem größeren Betrag bekannt, so verwendet man im Interesse größerer Genauig-

keit des anderen Faktors (für die Wurzeln mit dem kleineren Betrag) zweckmäßigerweise nicht die gewöhnliche, sondern die „umgekehrte" Division (d. h. die Division nach steigenden Potenzen von x im Sinne der Gleichung (2.2 b)).

Noch ein Wort zu dem unter I. dargelegten Verfahren: die einzelnen Schritte erfordern dort etwas weniger Rechenarbeit als bei dem unter II. behandelten Verfahren; eine einfache Konvergenzbetrachtung für das Verfahren I. (für $\rho \ll 1$ und hinreichend gute erste Näherungen) ergibt jedoch, daß diese einfachere Rechenarbeit bei den einzelnen Schritten durch den Nachteil erkauft werden muß, daß man etwa mit doppelt so vielen Schritten rechnen muß, um zu ebenso guten Näherungen wie beim Verfahren II. zu kommen.

V. Gleichungen beliebigen Grades

Das in den vorausgegangenen Abschnitten behandelte Verfahren läßt sich unter geeigneten Voraussetzungen auf Gleichungen höheren Grades verallgemeinern; um dies zu zeigen, sei die gegebene Gleichung $F = G \cdot K = 0$ (falls es nicht von vorneherein so ist) mit Hilfe einer geeigneten Transformation $\xi = \lambda \cdot x$ so normiert, daß der größte unter ihren Wurzeln auftretende Betrag gleich 1 ist. Die Gleichung läßt sich dann, falls sie m Wurzeln vom Betrag 1 (zusammengefaßt in G) und k Wurzeln von kleinerem Betrag (zusammengefaßt in K) hat, darstellen in der Gestalt

$$(5.1) \quad 0 = F = K \cdot G =$$

$$= [x^k + \varepsilon_1 x^{k-1} + \varepsilon_2 x^{k-2} + \cdots + \varepsilon_k] \cdot [x^m + c_1 x^{m-1} + c_2 x^{m-2} + \cdots + c_{m-1} x + 1]$$

$$= \underbrace{[x^k + \varepsilon_1' x^{k-1} + \varepsilon_2' x^{k-2} + \cdots + \varepsilon_k']}_{K'} \cdot \underbrace{[x^m + (c_1 + \delta_1) x^{m-1} + \cdots (1 + \delta_m)]}_{G''} + \text{Rest},$$

wobei die Größen ε_k' irgendwelche Näherungen für die Koeffizienten ε_k des Polynoms K bedeuten, während die Größen δ_μ die Korrekturglieder für die Koeffizienten des Polynoms G bedeuten, wenn man G durch $G'' = F : K'$ (unter Vernachlässigung des bei

der Division auftretenden Restes) ersetzt. Durch Koeffizientenvergleich folgt dann

$$\begin{aligned} c_1 + \varepsilon_1 &= (c_1 + \delta_1) + \varepsilon_1'; \\ c_2 + \varepsilon_1 c_1 + \varepsilon_2 &= (c_2 + \delta_2) + \varepsilon_1'(c_1 + \delta_1) + \varepsilon_2'; \\ c_3 + \varepsilon_1 c_2 + \varepsilon_2 c_1 + \varepsilon_3 &= (c_3 + \delta_3) + \varepsilon_1'(c_2 + \delta_2) + \varepsilon_2'(c_1 + \delta_1) + \varepsilon_3' \text{ usw.,} \end{aligned}$$

also unter Vernachlässigung von Gliedern der Gestalt $\varepsilon_\lambda' \delta_\mu$

$$\begin{aligned} \delta_1 &= \varepsilon_1 - \varepsilon_1'; \\ \delta_2 &= (\varepsilon_1 - \varepsilon_1') \cdot c_1 + (\varepsilon_2 - \varepsilon_2'); \\ \delta_3 &= (\varepsilon_1 - \varepsilon_1') \cdot c_2 + (\varepsilon_2 - \varepsilon_2') \cdot c_1 + (\varepsilon_3 - \varepsilon_3') \text{ und allgemein} \\ &\vdots \\ \delta_\mu &= (\varepsilon_1 - \varepsilon_1') \cdot c_{\mu-1} + (\varepsilon_2 - \varepsilon_2') \cdot c_{\mu-2} + \cdots = \sum (\varepsilon_\lambda - \varepsilon_\lambda') \cdot c_{\mu-\lambda} \end{aligned} \tag{5.2}$$

mit $c_0 = 1$.

Andrerseits ergibt sich („umgekehrte" Division durch $\frac{G''}{1+\delta_m}$) aus

(5.3)

$$\begin{aligned} F &= [1 + c_{m-1} x + c_{m-2} x^2 + \cdots c_1 x^{m-1} + x^m] \cdot [\varepsilon_k + \varepsilon_{k-1} x + \varepsilon_{k-2} x^2 + \cdots + x^k] \\ &= \left[1 + \frac{c_{m-1} + \delta_{m-1}}{1 + \delta_m} x + \frac{c_{m-2} + \delta_{m-2}}{1 + \delta_m} x^2 + \cdots\right] \cdot \\ &\cdot [\varepsilon_k + (\varepsilon_{k-1} + \varepsilon_{k-1}^*) x + (\varepsilon_{k-2} + \varepsilon_{k-2}^*) x^2 + \cdots] + \text{Rest} \end{aligned}$$

durch Koeffizientenvergleich für die Korrekturgrößen ε_λ^* zunächst die Gleichung

$$c_{m-1} \varepsilon_k + \varepsilon_{k-1} = \frac{c_{m-1} + \delta_{m-1}}{1 + \delta_m} \cdot \varepsilon_k + (\varepsilon_{k-1} + \varepsilon_{k-1}^*), \text{ also}$$

$$\varepsilon_{k-1}^* = \varepsilon_k \cdot \left[c_{m-1} - \frac{c_{m-1} + \delta_{m-1}}{1 + \delta_m}\right] = \varepsilon_k \cdot \frac{\delta_m c_{m-1} - \delta_{m-1}}{1 + \delta_m}. \tag{5.4}$$

Ferner folgt aus (5.3) (durch Vergleich der Koeffizienten von x^2)

$$\begin{aligned} &c_{m-2} \varepsilon_k + c_{m-1} \varepsilon_{k-1} + \varepsilon_{k-2} \\ &= \frac{c_{m-2} + \delta_{m-2}}{1 + \delta_m} \varepsilon_k + \frac{c_{m-1} + \delta_{m-1}}{1 + \delta_m} (\varepsilon_{k-1} + \varepsilon_{k-1}^*) + (\varepsilon_{k-2} + \varepsilon_{k-2}^*), \end{aligned}$$

also $\left(\text{wegen } c_{m-\lambda} - \frac{c_{m-\lambda}+\delta_{m-\lambda}}{1+\delta_m} = \frac{c_{m-\lambda}\,\delta_m - \delta_{m-\lambda}}{1+\delta_m}\right)$

$$\varepsilon^*_{k-2} = \varepsilon_k \cdot \frac{c_{m-2}\,\delta_m - \delta_{m-2}}{1+\delta_m} + \varepsilon_{k-1} \cdot \frac{c_{m-1}\,\delta_m - \delta_{m-1}}{1+\delta_m} - \varepsilon^*_{k-1} \cdot \frac{c_{m-1}+\delta_{m-1}}{1+\delta_m}$$

und damit wegen (5.4) unter Vernachlässigung höherer Potenzen der δ_μ schließlich

$$\varepsilon^*_{k-2} = \varepsilon_k \cdot (c_{m-2}\,\delta_m - \delta_{m-2} - c^2_{m-1}\,\delta_m + c_{m-1}\,\delta_{m-1}) + \tag{5.4a}$$
$$+ \varepsilon_{k-1}(c_{m-1}\,\delta_m - \delta_{m-1})$$

oder kurz

$$\varepsilon^*_{k-2} = \sum \varepsilon_\rho\,\delta_\sigma\,C_{\rho\sigma}, \quad \text{wobei} \quad |C_{\rho\sigma}| < M \tag{5.4b}$$

mit einer für alle Gleichungen gegebenen Grades gemeinsamen Schranke M, da die $|c_\mu|$ nach Voraussetzung (G soll lauter Wurzeln von Betrag 1 haben) höchstens gleich den Binomialkoeffizienten $\binom{m}{\mu}$ sein können. Ganz entsprechend folgt dann allgemein

$$\varepsilon^*_{k-r} = \sum \varepsilon_\rho\,\delta_\sigma\,C_{\rho\sigma r} \quad \text{mit} \quad |C_{\rho\sigma r}| < M. \tag{5.5}$$

Daraus und aus (5.2) folgt die Konvergenz des Verfahrens, falls nur die ε_ρ und die $(\varepsilon_\lambda - \varepsilon'_\lambda)$ hinreichend klein sind. Dies gilt sicher, wenn die Wurzeln von K alle hinreichend kleinen Betrag haben und wenn man für die erste Näherung K' dasjenige Polynom k. Grades verwendet, das entsteht, wenn man in F alle höheren Potenzen von x als x^k wegläßt; vgl. dazu die Bemerkungen nach (5.7). Übrigens kann man jede Gleichung $(m+k)$. Grades, die nicht lauter Wurzeln desselben Betrags hat, z. B. durch vorherige Anwendung des Graeffeschen Verfahrens[1] so transformieren, daß ihre Wurzeln die eben ausgesprochene Konvergenzbedingung erfüllen.

Aus dem Gedankengang des Konvergenzbeweises ergibt sich natürlich ohne weiteres, daß die Bedingung, G solle lauter Wurzeln gleichen Betrags haben, keineswegs notwendig ist, solange sich die Beträge der Wurzeln von G nicht allzustark voneinander unterscheiden.

[1] Und Normierung, so daß die „größten" Wurzeln den Betrag 1 haben.

Als Beispiel diene hierzu die folgende Gleichung 8. Grades (nicht im Sinn von (5.1) „normiert", was aber nichts Wesentliches ausmacht):

$$x^8 + 12x^7 + 110x^6 + 60x^5 + 400x^4 + 70x^3 + 75x^2 + 8x + 1 = 0.$$

Aus $G_0 = x^4 + 12x^3 + 110x^2 + 60x + 400$ folgt

$$K_1 = 1 + 7{,}85x + 73{,}55x^2 + 56{,}8x^3 + 371x^4$$

und dann $G_2 = x^4 + 11{,}85x^3 + 108{,}0x^2 + 41{,}1x + 372{,}05;$

$$K_3 = 1 + 7{,}89_0x + 73{,}84x^2 + 59{,}5x^3 + 371{,}7x^4;$$

$$G_4 = x^4 + 11{,}840x^3 + 107{,}91x^2 + 40{,}4x + 371{,}85;$$

$$K_5 = 1 + 7{,}891_5x + 73{,}853_5x^2 + 59{,}60x^3 + 371{,}84x^4.$$

Zu einer Abschätzung des Konvergenzbereichs und der Güte der Konvergenz kann man folgendermaßen kommen: Der größte unter den Wurzeln von K auftretende Betrag sei ρ; der Betrag der Wurzeln von G sei wie in (5.1) durchwegs 1. Wir halten m und k fest und beschränken uns auf den Fall hinreichend kleiner ρ, so daß es genügt, jeweils nur die niedrigsten Potenzen von ρ, die vorkommen, zu betrachten.[1] Dann ist $|\varepsilon_1| \leqq k\rho$ und

$$(5.6) \qquad |\varepsilon_\nu| \leqq \binom{k}{\nu} \cdot \rho^\nu \ll k\rho \quad \text{für} \quad \nu > 1, \quad \text{ferner} \quad |c_\mu| \leqq \binom{m}{\mu}.$$

Außerdem sei

$$(5.7) \qquad (\varepsilon_\nu - \varepsilon'_\nu) = \alpha_\nu \cdot \rho^{\nu+1},$$

wobei $|\alpha_\nu| \leqq M_0$ mit einer nur vom Grad $m+k$ abhängigen Schranke M_0. Dies ist sicher der Fall, wenn man die erste Näherung K' für K durch Weglassen der höheren Potenzen von x aus F gewinnt. Nach (5.1) ist nämlich

$$\begin{aligned} F = x^{m+k} &+ \cdots x^k \cdot (1 + \varepsilon_1 c_{m-1} + \varepsilon_2 c_{m-2} + \cdots) + \\ &+ x^{k-1}(\varepsilon_1 + \varepsilon_2 c_{m-1} + \varepsilon_3 c_{m-2} + \cdots) \\ &+ x^{k-2}(\varepsilon_2 + \varepsilon_3 c_{m-1} + \cdots) \\ &\vdots \\ &+ x^0 \cdot \varepsilon_k \end{aligned}$$

[1] Wir setzen dabei voraus, daß die angeschriebenen Koeffizienten $\neq 0$ sind, sonst bekäme man noch bessere Konvergenzverhältnisse, weil z. B. die ε^*_ν in (5.9) und (5.10) eine höhere Potenz von ρ als Faktor hätten.

und daher im vorliegenden Fall (mit der Abkürzung $1+\varepsilon_1 c_{m-1}+ + \varepsilon_2 c_{m-2}+\ldots = 1+\varepsilon$)

$$K' = x^k + x^{k-1}\cdot\frac{\varepsilon_1+\varepsilon_2 c_{m-1}+\cdots}{1+\varepsilon} + x^{k-2}\cdot\frac{\varepsilon_2+\varepsilon_3 c_{m-1}+\cdots}{1+\varepsilon}+\cdots$$
$$= x^k + x^{k-1}\cdot\varepsilon_1' + x^{k-2}\cdot\varepsilon_2' + \cdots \quad \text{nach (5.1).}$$

Daraus folgt

$$\varepsilon_\nu - \varepsilon_\nu' = \varepsilon_\nu - \frac{\varepsilon_\nu+\varepsilon_{\nu+1}c_{m-1}+\cdots}{1+\varepsilon} = \frac{\varepsilon_\nu\cdot\varepsilon - \varepsilon_{\nu+1}c_{m-1}+\cdots}{1+\varepsilon}.$$

Nun ist bis auf höhere Potenzen von ρ immer

$$|\varepsilon| = |\varepsilon_1 c_{m-1} + \varepsilon_2 c_{m-2} + \cdots| \leqq k\rho\cdot m \quad \text{nach (5.6)}$$

und damit wieder nach (5.6) bis auf höhere Potenzen von ρ

(5.7a)
$$|\varepsilon_\nu - \varepsilon_\nu'| \leqq \binom{k}{\nu}\rho^\nu\cdot k\rho m + \binom{k}{\nu+1}\rho^{\nu+1}\cdot m = \underbrace{m\cdot\left[\binom{k}{\nu}\cdot k + \binom{k}{\nu+1}\right]}\cdot\rho^{\nu+1}.$$

beschränkt, wenn $m+k$ gegeben.

Nach (5.2) und (5.7) folgt dann

(5.8) $\delta_\mu \sim c_{\mu-1}\cdot\alpha_1\rho^2$, speziell $\delta_m \sim c_{m-1}\alpha_1\rho^2$; $\delta_{m-1}\sim c_{m-2}\alpha_1\rho^2$.

Unter Benutzung von (5.4) folgt daraus

(5.9) $\varepsilon_{k-1}^* \sim \varepsilon_k(c_{m-1}\delta_m - \delta_{m-1}) \sim \varepsilon_k(c_{m-1}^2 - c_{m-2})\cdot\alpha_1\rho^2$.

Entsprechend folgt aus (5.4a) und (5.8)

$$\varepsilon_{k-2}^* \sim \varepsilon_{k-1}(c_{m-1}^2 - c_{m-2})\,\alpha_1\rho^2$$

und entsprechend aus (5.3) allgemein

(5.10) $\varepsilon_{k-\lambda}^* \sim \varepsilon_{k-\lambda+1}\cdot(c_{m-1}^2 - c_{m-2})\cdot\alpha_1\rho^2$,
also speziell, weil nach (5.7) $\alpha_1\rho^2 = \varepsilon_1 - \varepsilon_1'$,
$\varepsilon_1^* \sim \varepsilon_2\cdot(c_{m-1}^2 - c_{m-2})\cdot(\varepsilon_1 - \varepsilon_1')$;
$\varepsilon_0^* \sim \varepsilon_1\cdot(c_{m-1}^2 - c_{m-2})\cdot(\varepsilon_1 - \varepsilon_1')$.

Man erhält daher als zweite Näherung für $K=0$ auf Grund der in (5.3) angegebenen Bedeutung der ε_ν^* die Gleichung

(5.10a)
$$0 = (1+\varepsilon_0^*)x^k + (\varepsilon_1+\varepsilon_1^*)x^{k-1} + \cdots = (1+\varepsilon_0^*)[x^k + \varepsilon_1'' x^{k-1} + \cdots]$$

mit $\varepsilon_1'' = \frac{\varepsilon_1 + \varepsilon_1^*}{1 + \varepsilon_0^*}$, also $\varepsilon_1 - \varepsilon_1'' = \frac{\varepsilon_1 \varepsilon_0^* - \varepsilon_1^*}{1 + \varepsilon_0^*}$.

Nach (5.10) folgt daraus

$$(5.11) \qquad |\varepsilon_1 - \varepsilon_1''| \sim |\varepsilon_1^2 - \varepsilon_2| \cdot |c_{m-1}^2 - c_{m-2}| \cdot |\varepsilon_1 - \varepsilon_1'|.$$

Bei hinreichend kleinem ρ ist also nach (5.6) eine Schranke für den Konvergenzfaktor q, mit dem sich der ursprüngliche Fehler nach 2 Divisionen multipliziert hat, zunächst für die Fehler bei ε_1 näherungsweise gegeben durch

$$(5.12) \qquad q_0 = \rho^2 \cdot [k^2 + \tbinom{k}{2}] \cdot [m^2 + \tbinom{m}{2}].$$

Bei hinreichend kleinem ρ sind aber nach (5.9) und (5.10) die Fehler $|\varepsilon_\nu - \varepsilon_\nu''|$ alle entscheidend durch $\alpha_1 \rho^2 = (\varepsilon_1 - \varepsilon_1')$ bestimmt; nach (5.10a) ist nämlich

$$\varepsilon_\nu'' = \frac{\varepsilon_\nu + \varepsilon_\nu^*}{1 + \varepsilon_0^*}, \quad \text{also} \quad \varepsilon_\nu - \varepsilon_\nu'' = \frac{\varepsilon_\nu \varepsilon_0^* - \varepsilon_\nu^*}{1 + \varepsilon_0^*};$$

nach (5.10) folgt daraus

$$(5.11a) \qquad |\varepsilon_\nu - \varepsilon_\nu''| \sim |\varepsilon_\nu \varepsilon_1 - \varepsilon_{\nu+1}| \cdot |c_{m-1}^2 - c_{m-2}| \cdot |\varepsilon_1 - \varepsilon_1'| < M \cdot \rho^{\nu+1} \cdot |\varepsilon_1 - \varepsilon_1'| \quad \text{nach (5.6)}.$$

Dies entspricht der Annahme (5.7) für die $(\varepsilon_\nu - \varepsilon_\nu')$ und erlaubt damit die entsprechenden Schlüsse für weitere Näherungen, im besonderen z. B.

$$|\varepsilon_\nu - \varepsilon_\nu'''| \sim |\varepsilon_\nu \varepsilon_1 - \varepsilon_{\nu+1}| \cdot |c_{m-1}^2 - c_{m-2}| \cdot |\varepsilon_1 - \varepsilon_1''|.$$

Daher gibt q_0 ganz allgemein die gesuchte Schranke für den „Konvergenzfaktor" bei Wiederholung des Verfahrens.

Das Ergebnis läßt sich zusammenfassen im folgenden

Satz 6: Bei gegebenem Grad $k+m$ der Gleichung $F = 0$ konvergiert das Verfahren sukzessiver Divisionen im Sinn von (5.1) und (5.3) (wobei G lauter Wurzeln gleichen Betrags hat) immer, wenn nur das Verhältnis ρ des größten Wurzelbetrags von K zum Betrag der Wurzeln von G klein genug ist. Unter Vernachlässigung höherer Potenzen von ρ ist bei wieder-

holter Anwendung des Verfahrens eine obere Schranke q_0 für den „Konvergenzfaktor" q gegeben durch (5.12), d. h. durch

$$q_0 = \rho^2 \cdot [k^2 + \tbinom{k}{2}] \cdot [m^2 + \tbinom{m}{2}].$$

Dies bedeutet (vgl. die entsprechenden Überlegungen S. 87): Es ist $q \leqq q_0 \cdot (1+\varepsilon)$, wobei ε eine beliebig kleine positive Zahl bedeutet, wenn nur ρ unterhalb einer geeigneten (positiven) Schranke $\rho(\varepsilon)$ bleibt.

Im Fall reeller Polynome G vom 2. Grad mit zwei konjugiert komplexen Wurzeln ist nach (5.1) $c_{m-2} = 1$; $(4 \geqq)\ c_{m-1}^2 > 0$, also $|c_{m-1}^2 - c_{m-2}| \leqq 4 - 1 = 3$. Statt (5.12) folgt aus (5.11) dann

(5.13)
$$q_0 = 3 \cdot (k^2 + \tbinom{k}{2}) \cdot \rho^2, \text{ d. h. z. B. } q_0 = \begin{cases} 66\,\rho^2 \approx (8{,}1\,\rho)^2 \text{ für } k = 4\ (m + k = 6), \\ 153\,\rho^2 \approx (12{,}4\,\rho)^2 \text{ für } k = 6\ (m + k = 8). \end{cases}$$

Im Fall $m = 1$ ist in (5.8) sinngemäß $\delta_{m-1} = 0$ zu setzen, während $c_{m-1} = 1$. Statt (5.12) folgt dann aus (5.11) in diesem Fall

(5.14)
$$q_0 = (k^2 + \tbinom{k}{2}) \cdot \rho^2, \text{ d. h. z. B. } q_0 = \begin{cases} 22\,\rho^2 < (5\,\rho)^2 \text{ für } k = 4\ (m + k = 5), \\ 35\,\rho^2 < (6\,\rho)^2 \text{ für } k = 5\ (m + k = 6). \end{cases}$$

Von besonderem Interesse ist noch der Fall, daß F aus p quadratischen Faktoren mit je einem Paar konjugiert komplexer Wurzeln besteht, deren Beträge sich verhalten wie $1 : \rho_1 : \rho_2 : \ldots : \rho_{p-1}$. Ist dann G der quadratische Faktor mit den Wurzeln größten Betrags (1), so folgt aus (5.11) unter der Voraussetzung

$$\rho_{\nu+1} : \rho_\nu \leqq \rho \ll 1 \text{ für alle } \nu,$$

daß nicht nur $|c_{m-1}^2 - c_{m-2}| \lesssim 3$, sondern auch $|\varepsilon_1^2 - \varepsilon_2| \lesssim 3\,\rho^2$ und damit

(5.15)
$$q_0 \lesssim (3\,\rho)^2$$

für hinreichend kleine ρ in Übereinstimmung mit den Ergebnissen bei den Gleichungen 4. Grades.