

Mit den Daten für Kupfer berechnet man

$s_{11} = 1,100 \cdot 10^{-6} \text{ cm}^2/\text{kg}$, $s_{12} = -0,362 \cdot 10^{-6} \text{ cm}^2/\text{kg}$, $s_{44} = +1,790 \cdot 10^{-6} \text{ cm}^2/\text{kg}$
und erhält weiter auf dem zweiten Wege

$$E = 1146000 \text{ kg/cm}^2 \quad m = 3,256 \quad G = 438100 \text{ kg/cm}^2$$

Wie ersichtlich, weichen die beiden auf verschiedener Grundlage berechneten Werte nur verhältnismäßig wenig voneinander ab.

Die berechneten Daten gestatten uns, im Besitze entsprechender Zahlenwerte für den Einkristall, das Verhalten des quasiotropen Körpers vom Beginne des Deformationsvorganges an durch das elastische Gebiet bis zur unteren Fließgrenze numerisch zu verfolgen.

877

ZUSAMMENFASSENDE BERICHTE

Praktische Verfahren der Gleichungsauflösung.

Von R. v. MISES und H. POLLACZEK-GEIRINGER in Berlin.

In einer Vorlesung über »Praktische Analysis«, die der erstgenannte Verfasser im Sommer-Semester 1927 hielt, kamen eine Reihe von Rechnungsverfahren zur Sprache, die teils neu, teils wenig bekannt sind. Es erschien zweckmäßig, einiges davon mit kurzer Begründung und Hinweisen auf die wichtigsten Anwendungen zusammenzustellen. Dieser Arbeit hat sich der zweitgenannte Verfasser — zunächst hinsichtlich der Methoden der Gleichungsauflösung — unterzogen; von ihm rühren auch wesentliche Ergänzungen und nähere Ausführungen sowie die Zahlenbeispiele her.

Im ersten Abschnitt des vorliegenden Berichtes wird ein Verfahren zur Lösung einer beliebigen Gleichung mit einer Unbekannten entwickelt, das sich von den bisher bekannten dadurch unterscheidet, daß es stets zum Ziele führt (immer konvergiert) und in der Regel weniger Rechenarbeit verlangt. Die Aufsuchung der Wurzeln von Gleichungssystemen wird in den beiden folgenden Abschnitten nur für den Sonderfall der linearen Gleichungen behandelt. Für das »Iterationsverfahren in Gesamtschritten«, welches die unmittelbare Uebertragung des Verfahrens für Eine Unbekannte darstellt, werden in Abschnitt 2 die zum Teil bekannten, meist nicht gehörig beachteten Konvergenzbedingungen zusammengestellt sowie Fehlerabschätzungen gegeben.

Besondere Bedeutung kommt dem in Abschnitt 3 behandelten, als »Iteration in Einzelschritten« bezeichneten, von Ph. Seidel stammenden Ansatz zu. Dieser ist — nach gewissen Umformungen — auf beliebige lineare Gleichungssysteme anwendbar; andererseits liefert er in seiner ursprünglichen Form ohne jede Vorarbeit ein allgemeines, stets konvergentes Verfahren für die Lösung einer umfassenden Klasse von linearen Gleichungen, nämlich aller derer, die aus einem Minimumproblem hervorgehen. Für die Behandlung n -fach statisch unbestimmter Systeme kann man daraus einen sehr anschaulichen Satz gewinnen, der das allgemeine Problem auf die wiederholte Auflösung einfach unbestimmter Systeme zurückführt¹⁾ (Abschnitt 4).

Schließlich werden in den letzten Abschnitten²⁾ Systeme homogener linearer Gleichungen mit Parameter betrachtet, eine Aufgabe, die bekanntlich bei der Untersuchung von Eigenschwingungen elastischer Systeme und bei der Aufsuchung von Stabilitätsgrenzen (kritischen Lasten) auftritt. Die sinngemäße Uebertragung des in der Technik bekannten Verfahrens von Vianello zur Bestimmung des kleinsten Eigenwertes bei Randwertproblemen gewöhnlicher Differentialgleichungen gestattet durch einfache Iteration, ohne vorherige Lösung einer Gleichung n -ten Grades, den kleinsten Eigenwert und die zugehörige Eigenlösung zu finden. Daß man auf die gleiche Weise, durch wiederholte Anwendung des Verfahrens auch die höheren Eigenwerte und zugehörigen Eigenlösungen finden könne, ist eine oft wiederholte, unzutreffende Behauptung. Wir geben in Abschnitt 6 an, in welcher Weise man tatsächlich dieses für die Anwendungen oft wichtige Problem erledigen kann.

Soweit uns einschlägige frühere Veröffentlichungen bekannt wurden, sind sie an den betreffenden Stellen angeführt worden. Die Sätze und sonstigen Ergebnisse, die ohne Belege wiedergegeben werden, dürften im wesentlichen neu sein.

¹⁾ Vergl. den Vortragsauszug: H. Pollaczek-Geiringer, diese Zeitschr. Bd. 8 (1928), S. 446.

²⁾ Diese Abschnitte erscheinen als Fortsetzung im nächsten Heft.

1. Auflösung einer Gleichung mit einer Unbekannten. Eine Wurzel a der Gleichung $f(x)=0$ suchen [wobei $f(x)$ eine reelle stetige Funktion der reellen Variablen x ist], heißt praktisch, ein Intervall von einer genügend kleinen Länge δ bestimmen, innerhalb dessen f mindestens einmal verschwindet. Z. B. bedeutet die Angabe, $a=0,276$ sei eine Wurzel von $f(x)=0$, daß $f(x)$ innerhalb des Intervalls $0,2755$ bis $0,2765$ mindestens einmal den Wert Null annimmt. Im besonderen kann die Aufgabe gestellt sein, ausgehend von einem gegebenen Punkte $x=x_1$, die nächste, rechts von x_1 gelegene Wurzel von $f(x)$ zu suchen.

Um eine solche Wurzel zu bestimmen, muß man sich zunächst ein Bild machen von dem Anwachsen der Funktion, von ihrer »Steigung«, sei es, indem man eine rohe Skizze entwirft oder, indem man den Differenzenquotienten abschätzt. Genau genommen, ist unter Steigung der Differenzenquotient $\frac{f(x_2)-f(x_1)}{x_2-x_1}$ zu verstehen. Man wählt nun eine Zahl c , die dem Betrage nach kleiner ist als der reziproke Wert der größten Steigung in dem für die Aufsuchung der Wurzel in Frage kommenden Bereich:

$$|c| < \frac{x_2 - x_1}{|f(x_2) - f(x_1)|_{\max}} \quad (1).$$

Bei einer differenzierbaren Funktion wird man meist von der Tatsache Gebrauch machen, daß für eine solche die obere und die untere Grenze der Steigung in einem bestimmten Intervall mit der oberen und unteren Grenze der Ableitung in diesem Intervall zusammenfallen. Mit einem gemäß (1) gewählten Werte von c bildet man, ausgehend von einem Punkte x_1 , in dem $f(x_1)$ positiv sein mag, die Zahlenfolge:

$$x_2 = x_1 + c f(x_1), \quad x_3 = x_2 + c f(x_2), \dots \quad (2).$$

Wir behaupten, daß man auf diese Weise, wenn es überhaupt rechts oder links von x_1 eine Wurzel gibt, stets zu dieser gelangt, und zwar bei positivem c zu der nächsten rechts von x_1 gelegenen, bei negativem c zu der linken. Ist $f(x_1) < 0$, so vertauschen sich hier rechts und links.

Geometrisch bedeutet (2), daß man, ausgehend von dem Kurvenpunkte $y_1=f(x_1)$ eine Gerade unter dem Winkel α zieht (Abb. 1), wobei $c = \operatorname{ctg} \alpha$, eine Gerade also, die wegen (1) steiler nach rechts abwärts geht als die stärkste Sehnenneigung der Kurve im betrachteten Intervall (zwischen x_1 und der Wurzel). Aus der Figur — die für positives f gezeichnet ist — geht schon hervor, daß man auf diese Weise von links immer näher an die Wurzel herankommt, ohne über sie hinauszugelangen, und eben darin besteht die Behauptung.

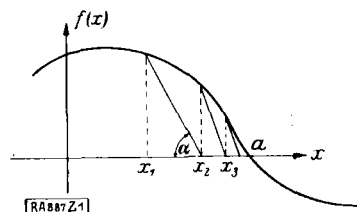


Abb. 1.

Für den Mathematiker beruht der Konvergenzbeweis in folgenden Überlegungen: Gemäß der Erklärung von c ist dieses dem Betrage nach kleiner als der reziproke Wert der Steigung für irgend zwei Punkte; wenn also die Steigung zwischen dem n -ten Näherungswert x_n und einem Wurzelwert a gebildet wird, so ist

$$|c| < \left| \frac{x_n - a}{f(x_n) - f(a)} \right| = \left| \frac{x_n - a}{f(x_n)} \right|,$$

und daher gilt wegen (2) für jedes $n=1, 2, \dots$

$$|x_{n+1} - x_n| = |c f(x_n)| < |x_n - a| \quad (3).$$

Ist $f(x_1) > 0$, c positiv gewählt und bezeichnet a den nächsten rechts von x_1 gelegenen Wurzelwert, so folgt aus (2), daß $x_2 > x_1$, und aus (3), weil a rechts von x_1 liegt, daß $x_2 - x_1 < a - x_1$, d. h. x_2 liegt zwischen x_1 und a . Demnach muß, da a die nächste rechts von x_1 gelegene Wurzel war, auch $f(x_2) > 0$ sein und man schließt wieder aus (2) und (3), daß x_3 zwischen x_2 und a liegen muß, usw. Die x_1, x_2, \dots bilden also eine beschränkte und monoton wachsende Folge, besitzen daher einen Grenzwert. Von einem gewissen, hinlänglich großen n an werden sich dann die x_n, x_{n+1}, \dots voneinander dem Betrage nach um höchstens ε unterscheiden, wobei ε eine beliebig klein vorgebbare Zahl ist:

$$|x_{v+1} - x_v| < \varepsilon, \quad \text{also wegen (2)} \quad |c f(x_v)| < \varepsilon.$$

Da c von Null verschieden gewählt war, kann somit der Grenzwert unserer Folge von einer Wurzel von $f(x)$ nicht verschieden sein; da die Folge andererseits nicht über a hin-

auswachsen kann und a die nächste rechts gelegene Wurzel ist, konvergiert die Folge $x_1, x_2, x_3 \dots$ gegen diese Zahl.

Zusammenfassend kann man sagen:

Satz 1. Bildet man mit einem Beiwert c , dessen Betrag nicht größer ist als der reziproke Wert der größten Steigung, die die Funktion $f(x)$ in dem in Betracht kommenden Intervall aufweist, die Zahlenfolge

$$x_1, \quad x_2 = x_1 + c f(x_1), \quad x_3 = x_2 + c f(x_2), \dots$$

so führt sie, je nachdem man $c f(x_1)$ positiv oder negativ gewählt hat, zu dem nächsten rechts oder links von x_1 gelegenen Wurzelwert. Die Zahlenfolge ist einsinnig wachsend oder abnehmend und führt nur dann ins Unendliche, wenn rechts bzw. links von x_1 keine Wurzel vorhanden ist.

Man sieht, daß man sich bei $c f(x_1) > 0$ mittels der sukzessive berechneten x_1, x_2, \dots einer etwa vorhandenen Wurzel a jedesfalls von links nähert. Die Konvergenz wird übrigens um so schneller sein, je größer man c unter Beachtung von (1) gewählt hat, da ja x_{n+1} aus x_n durch Hinzufügung des Zusatzes $c f(x_n)$ hervorgeht. Man kann c auch von Schritt zu Schritt geeignet verändern, also größere c wählen, wenn man in einen Bereich mit geringerer Steigung kommt (vergl. das folgende Beispiel). Wenn man so einige Schritte ausgeführt hat, wird man, um die Wurzel auch von rechts einzugrenzen, für einen Wert x'_1 , der etwas größer ist als das zuletzt berechnete x_n , den Wert von $f(x'_1)$ berechnen. Falls dieser negativ ist, kann man, wenn man das Intervall für die Wurzel noch weiter verengern will, entweder von links aus noch einige Schritte berechnen, oder von x'_1 ausgehend nun ähnlich wie in (2) eine abnehmende Folge:

$$x'_1 = x'_1 + c f(x'_1), \quad x'_2 = x'_1 + c f(x'_2), \dots \quad (2')$$

bilden, die von rechts gegen a konvergiert.

Falls $f(x)$ für x_1 negativ, für x'_1 positiv ist (x_1 kleiner als x'_1), so macht man, um von x_1 zur nächsten rechts gelegenen Wurzel zu kommen, den analogen Ansatz mit einem Beiwert $-c < 0$:

$$x_2 = x_1 - c f(x_1), \quad x_3 = x_2 - c f(x_2), \dots \quad (4),$$

und um das Intervall auch von der andern Seite abzugrenzen:

$$x'_2 = x'_1 - c f(x'_1), \quad x'_3 = x'_2 - c f(x'_2), \dots \quad (4'),$$

wobei man von einem x'_1 ausgeht, das ein wenig rechts von dem zuletzt berechneten x_n liegt und für das $f(x'_1)$ positiv ist.

Hierbei ist zunächst an einfache Wurzeln gedacht, die zugleich Zeichenwechsel der Funktion $f(x)$ sind. Besteht die Vermutung, daß die Gleichung eine mehrfache Wurzel besitzt, so wird man wohl in der Regel die Ableitung von $f(x)$ untersuchen. Das hier dargestellte Verfahren gestattet aber auch, aus dem Verhalten von $f(x)$ allein gewisse Schlüsse zu ziehen. Soll die Annahme geprüft werden, ob z. B. zwischen zwei Zahlen x_1 und x'_1 , wobei etwa $f(x_1)$ und $f(x'_1)$ positiv sei und $x_1 < x'_1$, eine Nullstelle ohne Zeichenwechsel liegt, so bilde man mit entsprechendem positiven c die Zahlenfolgen:

$$\begin{aligned} x_2 &= x_1 + c f(x_1), & x_3 &= x_2 + c f(x_2) \dots, \\ x'_2 &= x'_1 - c f(x'_1), & x'_3 &= x'_2 - c f(x'_2) \dots, \end{aligned}$$

(wobei wieder die c gegebenenfalls von Schritt zu Schritt abgeändert werden dürfen, aber stets (1) zu beachten ist). Solange sich diese Folgen nähern, ohne sich zu überschneiden, kann eine eventuelle Doppelwurzel nur zwischen ihnen liegen. Überschneiden sie sich, so ist gewiß keine Doppelwurzel vorhanden.

Unser Verfahren soll nun an einem Beispiel erläutert werden. Die Gleichung

$$\frac{1}{x} = \frac{4}{x^2} \frac{2(1 - \cos x) - x \sin x}{x \cos x - \sin x}$$

gibt für eine gewisse Klasse von Fällen die sogenannte Knickbedingung für einen eingespannten ebenen Rechtwinkelrahmen¹⁾. Dabei ist x eine Zahl, die von den Konstanten des Rahmens abhängt. Zu jedem x gibt es ein bestimmtes kleinstes α , das die Knicklast des Rahmens bestimmt. Wir wollen diesen x -Wert für $x=1$ aufsuchen und betrachten also die Funktion

$$f(x) = \frac{4}{x^2} \frac{2(1 - \cos x) - x \sin x}{x \cos x - \sin x} - 1.$$

¹⁾ Vergl. E. v. Mises und J. Ratserdorfer, diese Zeitschr., Bd. 6 (1926), S. 9.

Der Nenner des Bruches verschwindet, wenn $x = \operatorname{tg} x$. Bezeichnet man mit x_0 die kleinste (links nahe an $3\pi/2$ gelegene) Wurzel dieser transzendenten Gleichung, deren Wert 4,4934 ist¹⁾, so wird $f(x)$ für $x = x_0$ unendlich und nimmt für $x > x_0$ nach rechts ab. Für $x = 3\pi/2$ erhält man $f \sim 0,21$, für $x = 2\pi$ ist $f = -1$, da der Zähler des Bruches verschwindet. Also muß zwischen $3\pi/2$ und 2π mindestens eine Wurzel liegen; die erste rechts von $3\pi/2$ gelegene ist die von uns gesuchte. Denn für positive x -Werte links von x_0 verläuft $f(x)$, wie eine kurze Ueberschlagsrechnung lehrt, ganz im Negativen, und zwar steigt $f(x)$ ausgehend vom Werte $-\infty$ bei $x = 0$ allmählich an, erreicht zwischen $\pi/2$ und π ein (negatives) Maximum und nimmt von da an wieder bis zu der erwähnten, bei x_0 gelegenen Asymptote ab (Abb. 2).

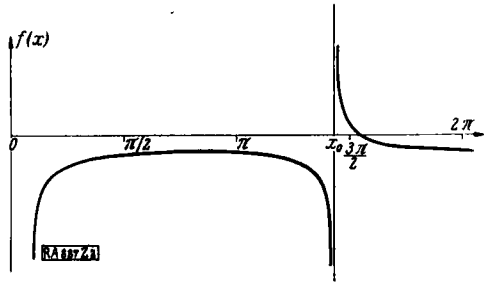


Abb. 2.

Als Ausgangspunkt wählen wir $x_1 = 3\pi/2$. Um den Beiwert c , der für unser Verfahren wesentlich ist, zu bestimmen, differenzieren wir $f(x)$ und erhalten:

$$f'(x) = -\frac{8}{x^2} \frac{2(1-\cos x) - x \sin x}{x \cos x - \sin x} - \frac{4}{x^2} + \frac{4}{x} \sin x \frac{2(1-\cos x) - x \sin x}{(x \cos x - \sin x)^2}.$$

Die Steigung nimmt von der Stelle der Asymptote her mit wachsendem x immer mehr ab. Wir rechnen daher zunächst $f'(3\pi/2)$; dieser Wert ist sicher größer als die Steigung für größere x . Man erhält (mit dem Rechenstab) $f'(3\pi/2) \sim -6,40$. Der reziproke Wert davon ist $-0,157$, also sicher absolut größer als 0,15. Diese Zahl wählen wir zunächst als Beiwert c und rechnen

$$x_2 = \frac{8\pi}{2} + 0,21 \cdot 0,15 = 4,71 + 0,0315 = 4,7415.$$

Da für 4,7415 der Wert von $f(x)$ mühsam zu rechnen ist, so rechnen wir zunächst für 4,74, was die Konvergenz im schlimmsten Fall etwas verzögert. Es ergibt sich $f(4,74) = 0,06$ und daraus:

$$x_3 = 4,74 + 0,06 \cdot 0,15 = 4,749.$$

Wir versuchen nun $x_4 = 4,75$ und rechnen diesen Wert schon genauer, da wir uns bereits in der Nähe der Wurzel befinden. Es ergibt sich $f(4,75) = 0,00422$ und daraus

$$x_5 = 4,75 + 0,15 \cdot 0,00422 = 4,7506330.$$

Nun wählen wir, um eine Annäherung von der anderen Seite zu erhalten, $x_1' = 4,751$, $f(4,751) = -0,0005125$. Um eine bessere Konvergenz zu erzielen, rechnen wir den Wert der Ableitung $f'(x)$ an der Stelle 4,751. Er ergibt sich absolut kleiner als 4,6, so daß man jetzt $c = 0,217$ nehmen darf (statt wie früher 0,15) und somit

$$x_2' = 4,751 - 0,217 \cdot 0,0005125 = 4,75088879.$$

Diese Zahl liegt oberhalb der Wurzel. Um noch eine bessere untere Schranke als das vorher errechnete x_5 zu erhalten, berücksichtigen wir den Wert der Ableitung an der Stelle 4,75, der etwas größer ist als der für 4,751. Es zeigt sich, daß man hier schon $c = 0,21$ nehmen darf; wir korrigieren daher, ohne einen neuen Funktionswert zu rechnen, das x_5 in ein x_5 :

$$x_5 = 4,75 + 0,21 \cdot 0,00422 = 4,75088620.$$

Es ist also die Wurzel zwischen 4,75088620 und 4,75088879 eingegrenzt, also innerhalb eines Intervalls von einer Länge $\delta < 3 \cdot 10^{-6}$.

Wenn man abkürzt, ergibt sich

$$x = 4,75089.$$

Es sei zum Schluß darauf hingewiesen, daß alle bekannten Verfahren der Wurzelbestimmung, wie das Newtonsche Verfahren, die regula falsi, das gewöhnliche Iterationsverfahren, als Spezialfälle des beschriebenen allgemeinen Iterationsverfahrens aufgefaßt werden können. Allgemein gesprochen haben wir ja eine Zahlenfolge der Form

$$x_{n+1} = x_n + c_n f(x_n), \quad (n = 1, 2, \dots)$$

¹⁾ Vergl. Jahnke-Emde, Funktionentafeln mit Formeln und Kurven. Leipzig 1928. S. 8.

über die ungefähre Lage der Lösung gewählten Werten $x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}$ und bestimme aus dieser neue Größen $x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)}$ nach der Vorschrift:

$$\left. \begin{aligned} x_1^{(2)} &= x_1^{(1)} + c_1 L_1(x^{(1)}) \\ x_2^{(2)} &= x_2^{(1)} + c_2 L_2(x^{(1)}) \\ &\dots \dots \dots \\ x_n^{(2)} &= x_n^{(1)} + c_n L_n(x^{(1)}) \end{aligned} \right\} \dots \dots \dots (2).$$

Man fährt fort, indem man rechts statt der $x^{(1)}$ die $x^{(2)}$ verwendet, um aus ihnen ein System $x^{(3)}$ zu bestimmen usw. Nach $\nu-1$ solchen »Gesamtschritten«, — wobei die Berechnung von n neuen Größen $x^{(\nu)}$ aus n bekannten Größen $x^{(\nu-1)}$ als ein »Gesamtschritt« bezeichnet wird — hat man ein System $x^{(\nu)}$ erhalten, aus diesem bestimmt man $x^{(\nu+1)}$ usw. Es fragt sich, ob dieser Prozeß ein konvergenter ist und ob die Folge der Wertsysteme $x^{(\nu)}$ sich der Lösung von (1) nähern; genauer gesagt, ob es unter gewissen Voraussetzungen über das vorgegebene Gleichungssystem möglich ist, eine Vorschrift für die Wahl der c_i ($i = 1, 2, \dots, n$) zu geben — analog wie dies in 1 im Fall einer Unbekannten gelungen war — derart, daß die Näherungsfolgen gegen die Lösung konvergieren¹⁾.

Zunächst ergibt sich in Verallgemeinerung des früheren Gedankenganges eine erste hinreichende Konvergenzbedingung wie folgt. Bezeichnet man mit x_1, \dots, x_n ein Lösungssystem, für das also $L_i(x) = 0$ ($i = 1, 2, \dots, n$), und führt Größen $z_i^{(\nu)}$ durch

$$x_i^{(\nu)} - x_i = z_i^{(\nu)}$$

als i te »Fehlerkomponente« der ν ten Näherung ein, so erhält man aus (2), indem man diese Gleichungen für den oberen Index ν und $\nu+1$ (statt 1 und 2) anschreibt und dann links x_i , rechts $x_i + c_i L_i(x)$ subtrahiert, das in den z homogene Gleichungssystem

$$\left. \begin{aligned} z_1^{(\nu+1)} &= (1 + c_1 a_{11}) z_1^{(\nu)} + c_1 a_{12} z_2^{(\nu)} + \dots + c_1 a_{1n} z_n^{(\nu)} \\ z_2^{(\nu+1)} &= c_2 a_{21} z_1^{(\nu)} + (1 + c_2 a_{22}) z_2^{(\nu)} + \dots + c_2 a_{2n} z_n^{(\nu)} \\ &\dots \dots \dots \\ z_n^{(\nu+1)} &= c_n a_{n1} z_1^{(\nu)} + c_n a_{n2} z_2^{(\nu)} + \dots + (1 + c_n a_{nn}) z_n^{(\nu)} \end{aligned} \right\} \dots (3).$$

Sind hier die spaltenweise genommenen Summen der absoluten Beträge kleiner als ein echter Bruch μ :

$$\left. \begin{aligned} |1 + c_1 a_{11}| + |c_1 a_{12}| + \dots + |c_1 a_{1n}| &\leq \mu < 1 \\ |c_2 a_{21}| + |1 + c_2 a_{22}| + \dots + |c_2 a_{2n}| &\leq \mu < 1 \\ &\dots \dots \dots \\ |c_n a_{n1}| + |c_n a_{n2}| + \dots + |1 + c_n a_{nn}| &\leq \mu < 1 \end{aligned} \right\} \dots \dots (4),$$

so folgt aus (3) durch Addition:

$$\begin{aligned} |z_1^{(\nu+1)}| + |z_2^{(\nu+1)}| + \dots + |z_n^{(\nu+1)}| &\leq \mu [|z_1^{(\nu)}| + |z_2^{(\nu)}| + \dots + |z_n^{(\nu)}|] \\ &\leq \dots \leq \mu^\nu [|z_1^{(1)}| + |z_2^{(1)}| + \dots + |z_n^{(1)}|] \dots (5). \end{aligned}$$

Damit ist gezeigt, daß bei jedem Gesamtschritt die Summe der Absolutwerte der Fehlerkomponenten kleiner wird und nach ν Schritten wie die ν te Potenz eines echten Bruches abnimmt, also bei unbegrenzt wachsendem ν beliebig klein wird, sobald (4) erfüllt ist. Das gleiche gilt a fortiori für jede einzelne Fehlerkomponente²⁾.

Das Aufsuchen solcher Beiwerte c_i , für die (4) erfüllt ist, läßt sich allerdings nur in Sonderfällen leicht bewerkstelligen, vor allem dann, wenn in jeder Zeile das Diagonalglied a_{ii} gegenüber den Nebengliedern a_{ik} ($i \neq k$) genug stark überwiegt. In einem solchen Falle wird man

$$c_i = -\frac{1}{a_{ii}} \quad (i = 1, \dots, n) \dots \dots \dots (6)$$

setzen, um die Diagonalglieder im Schema (3) oder (4) zum Verschwinden zu bringen. Die Iteration (2) mit der Annahme (6) erhält so die Form

$$x_i^{(\nu+1)} = -\frac{1}{a_{ii}} [a_{i1} x_1^{(\nu)} + a_{i2} x_2^{(\nu)} + \dots + a_{in} x_n^{(\nu)}] \dots \dots (7),$$

¹⁾ Das Verfahren ist bekannt. (Vergl. etwa C. Runge »Praxis der Gleichungen«, Leipzig 1921, S. 70 ff.). Doch wird als Konvergenzbedingung meist die ganz unbestimmte Forderung ausgesprochen, daß die Diagonalglieder in der Koeffizientenmatrix »überwiegend groß« sein sollen.

²⁾ Eine entsprechende Konvergenzbedingung für nicht lineare Gleichungssysteme gibt z. B. C. Runge, Praxis der Gleichungen. Leipzig 1921, S. 69/70.

so daß also die i^{te} Gleichung zur Verbesserung der i^{ten} Unbekannten benutzt (nach der i^{ten} Unbekannten aufgelöst) wird. Die Bedingung (4) nimmt die Form an:

$$\sum_i' \left| \frac{a_{ik}}{a_{ii}} \right| \leq \mu < 1 \quad (k = 1, \dots, n) \quad (8),$$

wobei \sum' bedeuten soll, daß nur über $i \neq k$ summiert wird, daß also nur die Spaltensummen aus den Nebengliedern in Betracht kommen. Im besonderen ist (8) natürlich stets erfüllt, wenn keiner der $(n-1)$ Summanden in (8) die Zahl $1/(n-1)$ übertrifft, (wobei nicht immer das Gleichheitszeichen gelten darf), d. h. wenn in jeder Zeile das Diagonalglied mindestens $(n-1)$ -mal so groß ist wie jedes der Nebenglieder. Da man jedes Gleichungssystem (1) mit nicht verschwindender Determinante so ordnen kann, daß die Diagonalglieder nicht verschwindende Koeffizienten haben, und man dann durch Dividieren jeder Gleichung durch den Koeffizienten des Diagonalgliedes ein System mit lauter Einsen in der Diagonale erhält, so lautet das bisherige Ergebnis:

Satz 2. Wenn es möglich ist, Beiwerte c_1, c_2, \dots, c_n so zu bestimmen, daß die mit ihnen gebildeten n Spaltensummen (4) sämtlich einen echten Bruch μ nicht übertreffen, so konvergiert die mit diesen c_i gebildete Iteration (2).

Insbesondere führt die Iteration mit $c_1 = c_2 = \dots = c_n = -1$ stets zum Ziel, wenn in dem »durchdividierten« System mit Diagonalkoeffizienten gleich Eins die n Summen aus den Beträgen der je $(n-1)$ untereinander stehenden Nebenkoeffizienten sämtlich kleiner als Eins sind.

Ehe wir dazu übergehen, weitere Konvergenzbedingungen anzugeben, sei noch eine Bemerkung über die Fehlerschätzung gemacht. Für die praktische Rechnung ist ja wichtiger — aber in der Regel noch schwieriger — als die Frage der Konvergenz die Frage nach einer brauchbaren Fehlerabschätzung in jedem Stadium der Rechnung. Eine nicht exakte, aber in der praktischen Analysis übliche Methode der Fehlerschätzung kann auf das Problem der Auflösung linearer Gleichungen durch Iteration folgendermaßen angewendet werden. Für zwei aufeinanderfolgende Näherungen $x^{(v)}$, $x^{(v+1)}$ und eine Lösung x gilt gemäß dem Iterationsansatz (2), da $L_i(x) = 0$,

$$x_i^{(v+1)} - x_i^{(v)} = c_i [L_i(x^{(v)}) - L_i(x)] = c_i \bar{L}_i(x^{(v)} - x) \quad (9),$$

wobei \bar{L}_i aus L_i durch Weglassen des absoluten Gliedes hervorgeht. $\bar{L}_i(x)$ ist also auch eine Bezeichnung für die i^{te} Komponente $a_{i1}x_1 + \dots + a_{in}x_n$ des transformierten Vektors $\mathcal{A}x$. Ist dann

$$x_i^{(v+1)} - x_i^{(v)} = \delta_i^{(v)} \quad (i = 1, \dots, n) \quad (10),$$

so ist nach (9), wenn man wieder für $x_i^{(v)} - x_i = z_i^{(v)}$ (Fehler der v^{ten} Näherung) setzt:

$$\bar{L}_i(z^{(v)}) = \frac{\delta_i^{(v)}}{c_i}, \quad (i = 1, \dots, n) \quad (11).$$

Denkt man sich diese Gl. (11) nach den unbekannten Fehlern $z_i^{(v)}$ aufgelöst, so ergibt sich, wenn A und A_{ix} die Determinante aus den a_{ix} bzw. die adjungierten Unterdeterminanten bedeuten,

$$z_x^{(v)} = \sum_{i=1}^n \frac{\delta_i^{(v)}}{c_i} \frac{A_{ix}}{A}, \quad |z_x^{(v)}| \leq \sum_{i=1}^n \left| \frac{\delta_i^{(v)}}{c_i} \right| \cdot \left| \frac{A_{ix}}{A} \right| < s \sum_{i=1}^n \left| \frac{A_{ix}}{A} \right| \quad (12),$$

wenn s absolut größer ist als das größte der $\left| \frac{\delta_i^{(v)}}{c_i} \right|$. Ist man dann nach einigen weiteren Schritten dahin gelangt, daß die Verbesserungen, die der q^{te} Schritt mit sich bringt, durchschnittlich nur mehr halb so groß sind, wie die eben betrachteten des v^{ten} Schrittes, d. h.

$$x_i^{(q+1)} - x_i^{(q)} = \delta_i^{(q)}, \quad |\delta_i^{(q)}| \sim \frac{1}{2} \cdot |\delta_i^{(v)}|$$

so folgt daraus

$$|z_x^{(q)}| < \frac{1}{2} s \sum_{i=1}^n \left| \frac{A_{ix}}{A} \right| \quad (12').$$

Wären (12) und (12') nicht Ungleichungen, sondern Gleichungen ohne Absolutzeichen links, so könnte man (12') von (12) subtrahieren und erhielte

$$z_x^{(v)} - z_x^{(q)} = x_x^{(v)} - x_x^{(q)} = \frac{1}{2} s \sum_{i=1}^n \left| \frac{A_{ix}}{A} \right| = |z_x^{(q)}|.$$

Die übliche, nicht beweisbare, durch das Vorstehende höchstens plausibel gemachte Behauptung geht nun dahin, daß der wirkliche Fehler $z_x^{(p)}$ der p -ten Näherung von derselben Größenordnung ist, wie die Differenz der v -ten und p -ten Näherung, wenn die Verbesserungen, die der p -te Schritt mit sich bringt, durchschnittlich halb so groß sind, wie die des v -ten Schrittes.

Eine exakte Fehlerschätzung ist möglich, wenn man das Zutreffen der Bedingung (4) bzw. (8) voraussetzt. Es sei wieder die Iteration so weit getrieben, daß die Verbesserung der i -ten Koordinate beim v -ten Schnitt

$$x_i^{(v)} - x_i^{(v+1)} = |z_i^{(v)} - z_i^{(v+1)}| \leq \delta_i \dots \dots \dots (14).$$

Dann ist

$$\sum_{i=1}^n |z_i^{(v)} - z_i^{(v+1)}| \leq \delta_1 + \delta_2 + \dots + \delta_n = \sum_{i=1}^n \delta_i$$

und erst recht

$$\sum_{i=1}^n \{|z_i^{(v)}| - |z_i^{(v+1)}|\} = \sum_{i=1}^n |z_i^{(v)}| - \sum_{i=1}^n |z_i^{(v+1)}| \leq \sum_{i=1}^n \delta_i.$$

Setzt man zur Abkürzung

$$\sum_{i=1}^n |z_i^{(v)}| = b, \quad \sum_{i=1}^n |z_i^{(v+1)}| = a, \quad \sum_{i=1}^n \delta_i = \delta,$$

so ist:

$$b - a \leq \delta \quad \text{oder} \quad \frac{b}{a} \leq 1 + \frac{\delta}{a}.$$

Andererseits besagt (5), daß $a \leq \mu b$, also ist

$$\frac{1}{\mu} \leq \frac{b}{a} \leq 1 + \frac{\delta}{a}.$$

Läßt man das Mittelglied der Ungleichheit weg, so erhält man

$$1 + \frac{\delta}{a} \geq \frac{1}{\mu}, \quad a \leq \frac{\mu}{1 - \mu} \delta$$

und wenn man dann das zweite und das dritte Glied berücksichtigt:

$$b = \sum_{i=1}^n |z_i^{(v)}| \leq a + \delta \leq \frac{1}{1 - \mu} \delta.$$

Wir erhalten somit das Ergebnis, das wir wieder der Einfachheit halber für ein System mit Einsen als Diagonalkoeffizienten aussprechen.

Satz 3. Ist die Spaltensumme der absolut genommenen Nebenkoeffizienten in einem Gleichungssystem mit den Diagonalkoeffizienten Eins kleiner als ein echter Bruch μ und ist man in Verfolgung der üblichen Iteration (7) (bei der die i -te Gleichung zur Verbesserung der i -ten Unbekannten benutzt wird) so weit gekommen, daß der Unterschied der v -ten Näherung zur folgenden in der i -ten Koordinate dem Betrage nach δ_i nicht übertrifft, so ist die Summe der Beträge der Komponenten der wirklichen Fehler der v -ten Näherung höchstens gleich

$$\frac{1}{1 - \mu} \cdot \sum_{i=1}^n \delta_i.$$

Man sieht, daß die Schranke um so besser wird, je kleiner μ ist. Unser Satz kann auch dahin ausgesprochen werden, daß in dem betrachteten Falle die Beträge der einzelnen Verbesserungen wesentlich abnehmen wie die Glieder einer geometrischen Reihe mit dem Quotienten μ .

Als ein Beispiel für diese Fehlerabschätzung betrachten wir das folgende Gleichungssystem¹⁾

$$\begin{aligned} 3x + 0,15y - 0,09z &= 6 \\ 0,08x + 4y - 0,16z &= 12 \\ 0,05x - 0,3y + 5z &= 20. \end{aligned}$$

¹⁾ Runge-König, Vorlesungen über numerisches Rechnen. Berlin 1924, S. 184.

Es seien die nachfolgenden Näherungen gerechnet:

	x	y	z
1. Näherung	2	3	4
2. Näherung	1,97	3,12	4,16
3. Näherung	1,9688	3,127	4,1675
4. Näherung	1,96868	3,12782	4,16793
5. Näherung	1,96867	3,12784	4,16795

Hier kann man $\mu = 0,11$ setzen; denn die drei absoluten Spaltensummen der Nebenglieder des »durchdividierten« Systems sind: $0,02 + 0,01 = 0,03$ bzw. $0,05 + 0,06 = 0,11$ bzw. $0,03 + 0,04 = 0,07$; dann ist $\frac{1}{1-\mu} = 1,13$. Für die $\delta_i^{(4)}$, das ist für den Unterschied der vierten und fünften Näherung, findet man in den drei Variablen $0,00001$; $0,00002$; $0,00002$. Die Summe dieser Verbesserungen ist $0,00005$ und unser Satz ergibt, daß die Summe der wirklichen Fehler der vierten Näherung die Zahl $0,00005 \cdot 1,13 = 0,00006$ nicht übertreffen kann. Für die dritte Näherung ergibt die analoge Ueberlegung $0,00098$ als obere Schranke, während tatsächlich die dritte Dezimalstelle in jeder Koordinate bei der dritten Näherung schon richtig ist.

Andererseits kann die von uns gegebene Schranke unter Umständen erreicht werden, wie folgendes Beispiel zeigt, in dem μ nicht kleiner als $0,5$ gesetzt werden kann. Geht man mit $c_1 = c_2 = -1$ in die Gleichungen

$$x + 0,5 y = 2, \quad 0,5 x + y = 2,5,$$

so ergeben die Iterationsformeln (7)

$$x^{(v+1)} = 2 - \frac{y^{(v)}}{2}, \quad y^{(v+1)} = 2,5 - \frac{x^{(v)}}{2}, \quad (v = 1, 2, 3, \dots).$$

Setzt man hier $x^{(1)} = 0$, $y^{(1)} = 2,5$, so kommt $x^{(2)} = 0,75$, $y^{(2)} = 2,5$. Hier ist $\delta_1 = 0,75$, $\delta_2 = 0$, $\delta_1 + \delta_2 = 0,75$.

$$\frac{\mu}{1-\mu}(\delta_1 + \delta_2) = 0,75 \quad \text{und} \quad \frac{\delta_1 + \delta_2}{1-\mu} = 1,50.$$

Tatsächlich beträgt, da $x = 1$, $y = 2$ die exakte Lösung ist, die absolute Fehlersumme der ersten Näherung $1 + 0,5 = 1,5$; die der zweiten $0,25 + 0,5 = 0,75$. Die angegebene Schranke ist also genau erreicht, und dies bleibt auch so bei den folgenden Schritten:

$$\begin{aligned} x^{(3)} &= 0,75, & x^{(4)} &= 0,9375, & x^{(5)} &= 0,9375, & x^{(6)} &= 0,984375, & x^{(7)} &= 0,984375, \\ y^{(3)} &= 2,125, & y^{(4)} &= 2,125, & y^{(5)} &= 2,03125, & y^{(6)} &= 2,03125, & y^{(7)} &= 2,0078125. \end{aligned}$$

usf.

Wir wenden uns jetzt der Diskussion weiterer Konvergenzfälle zu. Wenn die Nebenglieder (in dem durchdividierten System mit Einsen in der Diagonale) zwar im allgemeinen von der Größenordnung $1/n$ sind, während aber doch einzelne Spalten eine Summe aufweisen, die Eins übertrifft, so kann die Iteration (7) immer noch zum Ziel führen, wenn für die Koeffizienten die Ungleichung

$$\sum_{i \neq k} \left(\frac{a_{ik}}{a_{ii}} \right)^2 < 1 \quad (i \neq k) \quad (16)$$

besteht, die wir als Schmidtsche Bedingung bezeichnen, weil sie einer von E. Schmidt für Integralgleichungen gegebenen Konvergenzbedingung analog ist¹⁾.

In Vektorform gestaltet sich der Beweis folgendermaßen: Unser Gleichungssystem laute wie oben:

$$\mathfrak{A} \mathfrak{x} = \mathfrak{r} \quad \text{mit} \quad a_{ii} = 1 \quad (i = 1, \dots, n).$$

Dann wird die Iteration mit $c_i = -1$:

$$\mathfrak{x}^{(v+1)} = \mathfrak{x}^{(v)} - (\mathfrak{A} \mathfrak{x}^{(v)} - \mathfrak{r})$$

oder, wenn wieder die »Fehler« $\mathfrak{z}^{(v)} = \mathfrak{x}^{(v)} - \mathfrak{x}$ eingeführt werden:

$$\mathfrak{z}^{(v+1)} = \mathfrak{z}^{(v)} - \mathfrak{A} \mathfrak{z}^{(v)} = (\mathfrak{E} - \mathfrak{A}) \mathfrak{z}^{(v)} = \mathfrak{R} \mathfrak{z}^{(v)} \quad (17),$$

wobei \mathfrak{E} die Einheitsmatrix und

$$\mathfrak{R} = \mathfrak{E} - \mathfrak{A}, \quad k_{ii} = 0, \quad k_{ik} = -a_{ik} \quad (i \neq k) \quad (18).$$

¹⁾ Vergl. dazu Riemann-Weber, Die Differential- und Integralgleichungen der Mechanik und Physik. 7. Aufl., Bd. 1, Braunschweig 1925, S. 482.

den Nullpunkt führen, wenn die Eigenwerte der Matrix $\mathfrak{K} = \mathfrak{E} + c\mathfrak{A}$, das sind die Wurzeln λ' der Gleichung:

$$\begin{vmatrix} 1 + c a_{11} - \lambda' & c a_{12} & \dots & c a_{1n} \\ c a_{21} & 1 + c a_{22} - \lambda' & \dots & c a_{2n} \\ \dots & \dots & \dots & \dots \\ c a_{n1} & c a_{n2} & \dots & 1 + c a_{nn} - \lambda' \end{vmatrix} = 0 \quad (30)$$

sämtlich Beträge kleiner als Eins haben. Schreibt man, was wegen $c \neq 0$ immer möglich ist,

$$1 + c a_{ii} - \lambda' = c \left[a_{ii} - \left(\frac{\lambda'}{c} - \frac{1}{c} \right) \right] = c(a_{ii} - \lambda) \quad \text{mit} \quad \lambda' = \lambda c + 1,$$

so sieht man, daß $|\lambda'| < 1$ dasselbe bedeutet wie $|\lambda c + 1| < 1$ oder

$$\left| \lambda + \frac{1}{c} \right| < \frac{1}{c} \quad (31).$$

Das heißt aber: die Eigenwerte λ der ursprünglichen Matrix \mathfrak{A} , also die Wurzeln der Gleichung

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{vmatrix} = 0 \quad (32)$$

müssen mit dem Beiwert c in der durch (31) gegebenen Beziehung stehen. Liegen alle Wurzeln von (32) auf der linken oder alle auf der rechten Seite der imaginären Achse und wählt man c dem Zeichen nach positiv in jenem, negativ in diesem Falle und $|1/c|$ gleich dem Radius eines Kreises, der alle Eigenwerte in seinem Innern enthält und die imaginäre Achse im Nullpunkt berührt, so ist (31) stets erfüllt (vergl. Abb. 3; diese ist für den Fall gezeichnet, daß alle λ negativen Realteil haben und c daher positiv gewählt ist). Da $1/c$ mit ϱ wachsen muß, so sieht man, daß die Konvergenz des Verfahrens langsamer wird, je größer jener kleinste Kreis ist. Wir haben den:

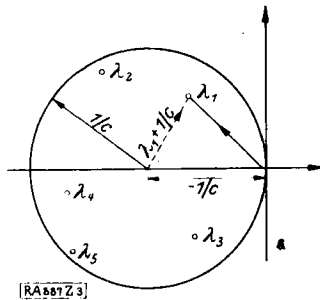


Abb. 3.

Satz 5. Das Iterationsverfahren (29) mit untereinander gleichen Beiwerten c konvergiert immer, falls alle Eigenwerte der Matrix des Systems auf ein- und derselben Seite der imaginären Achse liegen, sofern man c so wählt, daß $-1/c$ die (reelle) Mittelpunktsabszisse eines Kreises ist, der die imaginäre Achse im Null-

punkt berührt und alle Eigenwerte einschließt.

Ist z. B. die Matrix der a_{ik} positiv — oder negativ — definit, so ist die genannte Bedingung immer erfüllt, d. h. ein passendes c ist auffindbar. Ob es freilich praktisch möglich ist, dies c auch numerisch festzustellen, ist eine Frage, die sich nicht allgemein bejahen läßt. In der Regel wird es sich in den Fällen, in denen weder Satz 2 noch Satz 4 anwendbar ist, empfehlen, die im folgenden Abschnitt zur Besprechung kommenden Verfahren zu benutzen.

3. Iteration in Einzelschritten. In einer wichtigen Klasse von Fällen läßt sich ein etwas anderes Iterationsverfahren durchführen, das wir im Gegensatz zu der bisher betrachteten Iteration in »Gesamtschritten« als »Iteration in Einzelschritten« bezeichnen wollen. Zu einer einheitlichen Auffassung beider Arten von Iterationen gelangen wir, indem wir das Verfahren (2) von 2 zunächst noch verallgemeinern. Wir wollen annehmen, daß die Iterationskoeffizienten c_i (wie dies in 1 für den Fall Einer Veränderlichen auch vorgesehen war) auch von dem Index ν des Schrittes abhängen:

$$x_i^{(\nu+1)} = x_i^{(\nu)} + c_i^{(\nu)} L_i(x^{(\nu)}) \quad (1).$$

Setzt man hier $c_i^{(1)} = c_i^{(2)} = \dots = c_i^{(\nu)}$, so hat man das frühere Verfahren der Iteration in »Gesamtschritten«. Wir machen nun einen anderen speziellen Ansatz: Beim ersten Schritt $\nu = 1$ sollen alle $c_i^{(1)}$ mit Ausnahme von $c_1^{(1)}$ gleich Null sein, so daß nur

eindeutige Lösbarkeit von (1) sichergestellt ist). Ferner sind sicher die in den Diagonalen stehenden Koeffizienten a_{ii} positiv. (Denn wäre etwa $a_{11} < 0$, so würde das Glied $a_{11} x_1^2$, das bei genügender Vergrößerung von x_1 alle anderen überwiegt, den Wert von Q negativ machen; bei $a_{11} = 0$ würde aber $Q = 0$ für $x_2 = x_3 = \dots x_n = 0$ und beliebigen Wert von x_1 .) Es soll der folgende Satz bewiesen werden:

Satz 6. Sind die Koeffizienten a_{ix} des vorgelegten linearen Gleichungssystems symmetrisch, und ist die mit ihnen gebildete quadratische Form positiv definit, so führt die Iteration in Einzelschritten, bei der $c_i = -\frac{1}{a_{ii}}$ gesetzt ist, bei der also jeweils nur die i -te Gleichung gemäß (3) zur Verbesserung der i -ten Unbekannten benutzt wird, stets zum Ziel.

Zum Beweise¹⁾ betrachten wir die Funktion zweiten Grades in x_1, \dots, x_n

$$F(x_1, \dots, x_n) = \frac{1}{2} \sum_{i,x} a_{ix} x_i x_x - \sum_{i=1}^n r_i x_i = Q - \sum_{i=1}^n r_i x_i \quad (4).$$

Das Entscheidende ist, daß, wie man unmittelbar einsieht, die gegebenen linearen Gleichungen identisch sind mit

$$\frac{\partial F}{\partial x_1} = 0, \quad \frac{\partial F}{\partial x_2} = 0, \quad \dots \quad \frac{\partial F}{\partial x_n} = 0 \quad (5),$$

d. h. also, daß für ein System von x -Werten, das unsere Gleichungen befriedigt, F einen stationären Wert (Maximum, Minimum oder Sattelpunkt) annimmt. Es läßt sich weiter zeigen, daß dieser Wert ein Minimum sein muß. Denn erteilt man den Beträgen der x_i (oder auch nur einem von ihnen) hinreichend große Werte, so überwiegt der quadratische Bestandteil Q von F über den linearen und jener ist, da nach Voraussetzung die quadratische Form positiv definit ist, für alle von Null verschiedenen x_i positiv und mit wachsenden x_i proportional den Quadraten der x_i wachsend. Daher wächst die Funktion F außerhalb eines bestimmten Bereiches B mit zunehmenden Beträgen der x_i immer mehr an; eine Stelle stationären Wertes kann also nur im Innern von B liegen. Als stetige, in B beschränkte Funktion muß F einen Kleinstwert besitzen, der gewiß nicht am Rand liegt, und da unsere linearen Gleichungen nur ein Lösungssystem haben, so muß die Lösung unseres Systems die Stelle dieses Minimums liefern.

Es ist somit gezeigt, daß die Funktion F unter der Voraussetzung, daß ihr quadratischer Bestandteil Q keine negativen Werte annehmen kann, gewiß ein Minimum, und nur eines, besitzt. Wie der Gedankengang des Beweises zeigt, ist die Existenz eines Minimums erst recht sichergestellt, wenn F selbst für alle Werte der unabhängigen Veränderlichen (außer $x_i = 0$) stets positiv ist. Diese Bemerkung wird im folgenden Abschnitt herangezogen werden. (Im übrigen läßt sich leicht einsehen, daß, falls F die genannte Eigenschaft positiv definit zu sein, besitzt, auch Q sie haben muß.)

Die Lösung der Gleichungen (5) ist jedenfalls gleichbedeutend mit der Aufsuchung der Minimumstelle von F . Die Konvergenz unserer Iteration (3) ist insofern eine monotone, als sich zeigen läßt, daß jeder gemäß (3) vorgenommene Schritt den Wert von F verkleinert.

Es seien für die Unbekannten x_1, \dots, x_{i-1} die $(v+1)$ -ten, für x_{i+1}, \dots, x_n die v -ten Näherungen bekannt, und es erfolge jetzt der Schritt, der $x_i^{(v)}$ durch $x_i^{(v+1)}$ ersetzt. Der neue Wert, den F annimmt, unterscheidet sich dann von dem alten nur in den Gliedern, die x_i enthalten. Der Zuwachs ist zufolge (4):

$$\Delta F = \frac{1}{2} a_{ii} [(x_i^{(v+1)})^2 - (x_i^{(v)})^2] + \left\{ \sum_{x=1}^{i-1} a_{ix} x_x^{(v+1)} + \sum_{x=i+1}^n a_{ix} x_x^{(v)} - r_i \right\} [x_i^{(v+1)} - x_i^{(v)}].$$

Nun ist nach (3) für den Ausdruck in der geschweiften Klammer gerade $-a_{ii} x_i^{(v+1)}$ zu setzen; also wird

$$\Delta F = (x_i^{(v+1)} - x_i^{(v)}) \cdot \left[\frac{1}{2} a_{ii} (x_i^{(v+1)} + x_i^{(v)}) - a_{ii} x_i^{(v+1)} \right] = -\frac{1}{2} a_{ii} [x_i^{(v+1)} - x_i^{(v)}]^2 < 0 \quad (6),$$

da a_{ii} , wie früher bewiesen wurde, positiv ist. Bei jedem Schritt der Iteration (3) wird also der durch (4) gegebene Ausdruck F verkleinert. Da F ein endliches Minimum

¹⁾ Einen mit dem folgenden übereinstimmenden Beweis hat Hr. E. Trefftz vor einiger Zeit der Redaktion eingesandt; seine Veröffentlichung unterblieb nur mit Rücksicht auf den in Vorbereitung befindlichen Bericht.

besitzt, kann die Verkleinerung nicht unbeschränkt weiter laufen, d. h. es muß ΔF gegen 0 gehen. Nach (6) folgt daraus, daß der Betrag $|x_i^{(v+1)} - x_i^{(v)}|$ für jedes i gegen 0 gehen muß, daß also das Verfahren konvergiert.

Als Beispiel betrachten wir ein System von sechs inhomogenen Gleichungen, welche aus einem Minimumproblem — mit dem Näherungsgrade $n = 6$ — hervorgegangen sind¹⁾, deren Matrix daher gewiß positiv definit ist:

$$\begin{aligned} 651,8x - 239,3y + 94,6z - 188u + 148v + 58w &= 431,5, \\ -119,6x + 8867y - 961z - 226u - 515v + 186w &= 188,5, \\ 94,6x - 1922y + 24390z + 474u - 4820v + 592w &= 82,7, \\ -93,9x - 226y + 237z + 48100u - 2370v - 2015w &= 120,0, \\ 74,2x - 515y - 2410z - 2370u + 78600v - 6420w &= 52,5, \\ 58x + 372y + 592z - 4030u - 12840v + 158500w &= 33,5. \end{aligned}$$

Aus diesem System entsteht ein symmetrisches, wenn $x/2, z/2, w/2$ an Stelle von x, z, w als Unbekannte betrachtet werden.

Bei der Auflösung nach unserm Iterationsverfahren setzen wir zunächst für die ersten Näherungswerte $x^{(1)}, y^{(1)}, z^{(1)}$ der ersten drei Unbekannten Werte, die wir durch Ansetzen des gleichen Minimumproblems mit dem Näherungsgrad $n = 3$ erhalten haben. Diesem entspricht ein System von drei Gleichungen mit den Unbekannten x, y, z , dessen Koeffizientenmatrix in der linken obern Ecke der sechsstufigen Matrix steht und dessen rechte Seiten mit den ersten drei Zahlen der obigen rechten Seite des sechsstufigen Systems übereinstimmen. Für dieses dreigliedrige System ergibt sich zunächst die folgende Tabelle:

	x	y	z
$v = 1$	481,5	188,5	82,7
	651,8	8867	24390
2	0,6698	0,0307	0,0036
3	0,6717	0,0307	0,0033
4	0,6728	0,0307	0,0033
...

Die letztgefundenen Werte setzt man als erste Näherungen der ersten drei Unbekannten in das sechsstufige System und erhält in Verfolgung der Iteration:

	x	y	z	u	v	w
1	0,6728	0,0307	0,0033	0,0025	0,0007	0,0000
2	0,6784	0,0308	0,0033	0,0040	0,0005	0,0000
3	0,6789	0,0308	0,0032	0,0040	0,0005	0,0000
4	0,6789	0,0308	0,0032	0,0040	0,0005	0,0000
..

Die geometrische Bedeutung des angegebenen Verfahrens ist folgende: Im $(n+1)$ -dimensionalen Raum, in dem F als Funktion von x_1 bis x_n ein »Paraboloid« bildet, geht man beim ersten Schritt, der x_1 vermöge der ersten Gleichung verbessert, auf dem Paraboloid, innerhalb der zweidimensionalen Hyperebene, die durch die $(n-1)$ Gleichungen $x_2 = \text{konst.}, \dots, x_n = \text{konst.}$ bestimmt wird, bis zu dem am tiefsten gelegenen Punkt. Denn die erste Gleichung besagt ja, daß F bei Festhaltung von x_2 bis x_n zu einem Minimum gemacht wird. Das x_1 des gefundenen Punktes gibt den verbesserten Wert der ersten Unbekannten. Sodann geht man, indem man nur x_2 verändert, in einer zweidimensionalen Ebene, die den eben gefundenen Punkt enthält, in gleicher Weise weiter, usw.

Es liegt nahe, gelegentlich auch so vorzugehen, daß man statt zweidimensionaler mehrdimensionale Hyperebenen wählt, die durch Konstantsetzen von weniger als $(n-1)$ Koordinaten festgelegt werden. Dabei wird man jedenfalls auch immer zu tiefer liegenden Punkten des Paraboloids gelangen, wenn man die freien Koordinaten so wählt, daß sie den ihnen zugeordneten Gleichungen genügen. Das Verfahren gestaltet sich dann folgendermaßen: Man wählt z. B. für die dritte bis n te Unbekannte Werte $x_3^{(1)}, x_4^{(1)}, \dots, x_n^{(1)}$ und bestimmt, indem man die erste und zweite Gleichung simultan nach der ersten

¹⁾ Der Ansatz ist einer Arbeit von W. Ritz, Crelles Journal 135 (1909), S. 1 bis 61 entnommen.

usf. Da sogar die quadratische Funktion A für beliebige Werte, der Veränderlichen X , nur positiver Werte fähig ist, so ist nach der Bemerkung S. 70 die Existenz eines und nur eines Minimums von A , auf die es bei Anwendung unseres Iterationsverfahrens ankommt, gewiß sichergestellt und die Iteration in Einzelschritten somit statthaft.

Was hier von ebenen Fachwerken gezeigt wurde, gilt gleicherweise für jedes statisch unbestimmte System. Immer läßt sich die Formänderungsarbeit als Funktion zweiten Grades in n »überzähligen« ausdrücken und die Ableitungen dieser Funktion A , die als elastische Arbeit nur positive Werte annehmen kann, geben, gleich Null gesetzt, die Maxwell'schen Gleichungen des Systems. Das Ergebnis der Untersuchung fassen wir in folgendem Satze zusammen:

Satz 9. Bildet man für ein beliebiges statisch unbestimmtes System nach Wahl von n überzähligen statischen Größen $X_1 \dots X_n$ die Maxwell'schen Gleichungen, die durch Differentiation der als Funktion der X_i aufgefaßten Formänderungsarbeit nach den X_i entstehen, so konvergiert die Iteration in Einzelschritten, bei der man jedesmal die i te Gleichung zur Verbesserung der i ten Unbekannten benutzt.

Man kann einen, in bestimmter Richtung noch weitergehenden, besonders anschaulichen Satz gewinnen, wenn man den Begriff der »überzähligen« etwas einschränkt. Man wähle als die n überzähligen n »überzählige Verbindungen«, nämlich Stäbe, die zwei Punkte des Gebildes miteinander verbinden (eine Entfernung festlegen), Auflagerstäbe, die eine Verbindung zwischen einem Punkte des Systems und dem festen Boden herstellen, oder schließlich Eckverbindungen, die eine Ecke versteifen (einen Winkel fixieren). Ersetzt man die zweite bis n te dieser überzähligen Verbindungen, jede einzeln durch entsprechende Kraftgrößen (nämlich einen Stab durch zwei entgegengesetzt gleiche Kräfte, eine Eckversteifung durch zwei entgegengesetzt gleiche Momente, usf.), und fügt diese Kraftgrößen X'_1, X'_2, \dots, X'_n als eingeprägte Kräfte zu den gegebenen Kräften \mathfrak{P}_x hinzu, so hat man ein einfach statisch unbestimmtes System. Man berechnet für die Spannung in dem einen überzähligen Gliede einen Wert X_1'' . Nun bildet man wieder ein einfach statisch unbestimmtes System, indem man die erste, dritte, vierte bis n te der überzähligen Verbindungen fortläßt und durch die Kraftgrößen $X_1'', X_3'', \dots, X_n''$ in der oben erklärten Weise ersetzt, und bestimmt für die einzige noch übrig gebliebene statische Unbestimmte den Wert X_2'' . Mit $X_1'', X_2'', X_4'', X_5'', \dots, X_n''$ als gegebenen Kräften löst man wiederum ein einfach unbestimmtes System, erhält für die einzige überzählige den Wert X_3'' und fährt in dieser Weise fort, bis man die letzte der überzähligen Verbindungen als einzige beibehalten und hierfür die Spannungsgröße X_n'' ermittelt hat. Die Serie $X_1'', X_2'', \dots, X_n''$ läßt man nun an Stelle der zu Anfang willkürlich angenommenen X'_1, X'_2, \dots, X'_n treten und rechnet in dieser Weise in infinitum fort. Dabei ergeben sich, wie wir sogleich werden, im Limes die richtigen Werte sämtlicher Spannungen.

Wenn wir, um die Gedanken zu fixieren, uns wieder ein ebenes Fachwerk vorstellen, so lautet die (jetzt einzige) Maxwell'sche Gleichung für das System mit einer Unbestimmten X_1''

$$\sum_1^{2k+1} T_x^{(1)} (T_x^{(0)} + T_x^{(1)} X_1'') \left(\frac{l}{EF} \right)_x = 0 \quad \dots \quad (7).$$

Dabei bezeichnen $T_x^{(0)}$ ($x = 1, 2, \dots, 2k$) die Spannungen in den Stäben des Hauptsystems unter dem Einfluß aller jetzt wirkenden eingeprägten Kräfte und $T_{2k+1}^{(0)} = 0$. Die $T_x^{(1)}$ sind die oben ausführlich erklärten Selbstspannungen in den $(2k+1)$ Stäben. Vergleichen wir mit den Größen, die in der ersten Maxwell'schen Gleichung des Systems (5) auftreten, so sehen wir, daß

$$S_x^{(1)} = T_x^{(1)} \quad (x = 1, 2, \dots, 2k+1).$$

Gilt. Denn die $S_x^{(1)}$ sind für die genannten Indizes genau wie die $T_x^{(1)}$ definiert: Für $x = 2k+2, \dots, 2k+n$ verschwinden sowohl $S_x^{(1)}$ wie $T_x^{(1)}$; hingegen ist $T_x^{(0)}$ das Spannungsergebnis in den $2k$ Stäben des Hauptsystems zufolge der äußeren Kräfte $\mathfrak{P}_1, \dots, \mathfrak{P}_n, X'_1, \dots, X'_n$, also nach dem Superpositionsgesetz:

$$T_x^{(0)} = S_x^{(0)} + X'_1 S_x^{(2)} + \dots + X'_n S_x^{(n)} \quad (x = 1, 2, \dots, 2k) \quad \text{und} \quad T_{2k+1}^{(0)} = 0.$$

Auf der linken Seite von (7) steht somit:

$$\sum_1^{2k+1} S_x^{(1)} (S_x^{(0)} + X'_1 S_x^{(2)} + \dots + X'_n S_x^{(n)} + X_1'' S_x^{(1)}) \left(\frac{l}{EF} \right)_x$$

und das ist die linke Seite der ersten der Gleichungen (5), in der Gestalt, die sie erhält, wenn sie gemäß unserem in Satz 8 erklärten Iterationsverfahren zur Verbesserung der ersten Unbekannten X_1 benutzt wird. Ganz Analoges gilt für die zweite Maxwellsche Gleichung und den zweiten Schritt des Verfahrens usw.

Das Gesagte kann in folgenden Satz zusammengefaßt werden:

Satz 10. Die Berechnung eines statisch unbestimmten Systems mit n überzähligen Verbindungen (Stäben, Eckversteifungen, Unterstützungen), läßt sich wie folgt auf die wiederholte Berechnung von einfach unbestimmten Systemen zurückführen: Man ersetze zunächst die zweite bis n te der überzähligen Verbindungen durch entsprechende willkürlich gewählte Kraftgrößen (Kräfte, Momente, Auflager) X_2', \dots, X_n' und berechne die statisch unbestimmte Größe X_1'' , die der einzigen beibehaltenen überzähligen Verbindung entspricht. Hierauf ersetze man die erste Verbindung durch die Kraftgröße X_1' , die dritte bis n te durch X_2', \dots, X_n' und rechne die zweite jetzt allein beibehaltene Überzählige X_2'' . In dieser Weise fährt man fort, bis man zur Berechnung eines X_n'' aus bereits vorher berechneten X_1'', \dots, X_{n-1}'' kommt, und beginnt sodann von Neuem, indem man X_1'', \dots, X_n'' an Stelle der früher gewählten X_2', \dots, X_n' annimmt, daraus X_1''' bestimmt usw. Dieses Verfahren konvergiert immer.

Für die praktische Durchführung des Verfahrens ist von besonderer Wichtigkeit, daß man zur Lösung der sukzessive auftretenden einfach statisch unbestimmten Systeme nicht gerade die Maxwellsche Gleichung des betreffenden Systems heranziehen muß, sondern auch jede andere — unter Umständen einfachere — benutzen kann, da ja im Falle einer Unbestimmten jede derartige Gleichung der Maxwellschen gleichwertig sein muß.

Als Beispiel zu unserem Verfahren, bei dem auch diese letzte Bemerkung zur Anwendung gelangt, erwähnen wir kurz die von Herrn C. Biezeno¹⁾ behandelte Aufgabe, einen elastisch gestützten, statisch unbestimmten Balken zu berechnen. Hr. Biezeno gibt zur Lösung dieser Aufgabe ein Iterationsverfahren an, das nur unter ganz bestimmten, praktisch kaum nachprüfbar bedingungen konvergiert, nämlich dann, wenn die Wurzeln einer charakteristischen Gleichung alle dem Betrage nach entweder kleiner oder größer als Eins sind; eine Bedingung der Art, wie wir sie bei Besprechung der »Iteration in Gesamtschritten« in Abschnitt 2, Satz 5 kennengelernt haben.

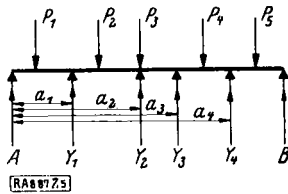


Abb. 5.

Dieses Problem läßt sich aber nach unserem Satz 9 ganz allgemein erledigen, indem man nacheinander einfach unbestimmte Systeme löst, und zwar ohne jedesmal die Maxwellsche Gleichung des Systems aufstellen zu müssen. Ist (Abb. 5) ein gerader Balken auf $n+2$ nachgiebigen Stützen gelagert, so wollen wir n Auflagerkräfte Y_v an den Abszissen $x = a_v$ ($v = 1, \dots, n$) — beispielsweise die inneren — als Überzählige betrachten. Die Größe der Senkung (positiv nach unten gerichtet), die zufolge der Nachgiebigkeit der v ten Stütze durch eine in dieser Stütze aufwärts wirkende Auflagerkraft von der Größe Y_v hervorgerufen wird, sei $c_v Y_v$, wobei c_v als die Elastizitätskonstante (Nachgiebigkeit) der Unterstützung gegeben sein muß. Faßt man den Balken als einen statisch bestimmten, mit den Auflagern bei $a_0 = 0$ und $a_{n+1} = l$ auf, indem man zu den gegebenen Kräften P_x noch die aufwärts wirkenden Y_1, Y_2, \dots, Y_n als Belastungen hinzunimmt, so läßt sich die Durchbiegung an der Stelle $x = a_v$ als lineare homogene Funktion der k Lasten P_x und der n Auflagerkräfte Y_v darstellen. Sondern wir hierbei den Bestandteil, der von Y_v selbst herrührt, ab, so hat dieser den Wert $-k_v Y_v$, wenn k_v die Größe der Durchbiegung bezeichnet, die der an beiden Enden gestützte Balken unter dem Einfluß der an der Stelle $x = a_v$ wirkenden Last von der Größe Eins an dieser Stelle erhält. Dieses k_v ist eine gegebene Funktion von a_v , beispielsweise, wenn die äußeren Auflager bei $x = 0$ und $x = l$ fest sind²⁾:

$$k_v = \frac{1}{EJ} \frac{l^3}{3} \frac{a_v^3 (l - a_v)^2}{l^2}.$$

¹⁾ C. Biezeno, Diese Zeitschr. Bd. 4, 1924, S. 98–102.

²⁾ »Hütte«, Bd. I, 25. Aufl., S. 609.

Es gilt also, wenn der restliche Teil der Durchbiegung mit y_v bezeichnet wird:

$$c_v Y_v = -k_v Y_v + y_v$$

oder, wenn man $q_v = \frac{1}{c_v + k_v}$ setzt:

$$Y_v = q_v y_v \dots \dots \dots (8).$$

Dabei ist y_v die an der Stelle $x = a_v$ auftretende Ordinate des Biegungspolygons, das für den statisch bestimmten Balken mit den sämtlichen \mathfrak{P}_x , sowie $Y_1, Y_2, \dots Y_{v-1}, Y_{v+1}, \dots Y_n$ als Belastungen konstruiert wurde. Wäre Y_v die einzige statisch Ueberzählige, so wäre mit (8) das Problem gelöst. Man kann nun im Sinne von Satz 10 wie folgt vorgehen.

Wir wählen zunächst für die zweite bis n -te der überzähligen Auflagerkräfte willkürliche Größen $Y_2', Y_3', \dots Y_n'$ und entwerfen mit diesen Kräften und den \mathfrak{P}_x für den statisch bestimmten Balken mit den Auflagern in $x = 0$ und $x = l$ das Biegungspolygon. Seine Ordinate bei $x = a_1$ sei y_1' . Daraus rechnen wir nach (8) $Y_1'' = q_1 y_1'$ und konstruieren eine neue Biegelinie des statisch bestimmten Balkens für die Kräfte $\mathfrak{P}_1, \dots \mathfrak{P}_n, Y_1'', Y_2', \dots Y_n'$. Aus der Ordinate y_2' dieser Biegelinie an der Stelle $x = a_2$ bestimmen wir $Y_2'' = q_2 y_2'$, konstruieren dann die Durchbiegung für das Kräftesystem $\mathfrak{P}_1, \mathfrak{P}_2, \dots \mathfrak{P}_n, Y_1'', Y_2'', Y_3', \dots Y_n'$ und fahren so fort. Nach Durchlaufung der n Stützen muß man natürlich von vorne anfangen und solange weitergehen, bis die Werte der Y_v sich nicht mehr merklich ändern. Bei größerem n wird man das wiederholte Aufzeichnen der Biegungspolygone durch einmalige Konstruktion der n Einflußlinien ersetzen. Jedenfalls erhalten wir auf diesem Wege sicher eine Lösung, die vielleicht gelegentlich in der Ausführung etwas umständlich sein mag, die aber einerseits gedanklich sehr einfach, andererseits an keinerlei einschränkende Konvergenzbedingung gebunden ist.

Schließlich bemerken wir, daß man gemäß Satz 7 über die Iteration in Gruppen, statt ein n -fach statisch unbestimmtes System auf die wiederholte Lösung von lauter einfach statisch unbestimmten Systemen zurückzuführen, auch die wiederholte Berechnung von zwei- oder dreifach . . . unbestimmten Systemen als Teilschritt verwenden kann. Auch in diesem Falle darf man beliebige Elastizitätsgleichungen verwenden und muß nicht gerade die Maxwell'schen Gleichungen aufstellen, da es sich nur darum handelt, jeweils ein zweifach, dreifach, unbestimmtes System richtig zu lösen und alle Gleichungssysteme, welche das leisten, untereinander äquivalent sind.

(Fortsetzung folgt.)

KLEINE MITTEILUNGEN

Ueber die allgemeine Form des Korrelationsmaßes. Bei der Ausdehnung des Korrelationsbegriffes auf ein System von mehr als zwei Variablen oder bei Zugrundelegung einer nicht linearen Funktionalbeziehung erhebt sich die Frage, wie man zu einer zweckmäßigen Verallgemeinerung des linearen Korrelationsmaßes zwischen zwei Variablen gelangen kann. Da sich der Fall nicht linearer auf den einer linearen Korrelation zwischen mehreren Variablen zurückführen läßt, wollen wir uns zunächst mit dem letzteren befassen.

Die bisherigen Untersuchungen gingen von den Regressionsgleichungen aus und gelangten zu einer Korrelation, die Pearson als »partiell« bezeichnet hat. Ihr Maß ist definiert als Produkt zweier Regressionskoeffizienten und in dieser Hinsicht allerdings eine Verallgemeinerung des für zwei Variable geltenden »totalen«. Diese partiellen Korrelationskoeffizienten sind gebrochene Funktionen der totalen, sie charakterisieren die Korrelation zwischen zwei Variablen mit Berücksichtigung der übrigen, geben jedoch keinen Aufschluß über die Gesamtkorrelation des Systems. Doch ist es gerade die

letztere, welche darüber entscheidet, mit welcher Berechtigung man von einer linearen Relation zwischen den Veränderlichen des Systems sprechen kann. Die partiellen Korrelationskoeffizienten können also nicht als geeignete Verallgemeinerung derjenigen angesehen werden, welche im speziellen Falle zweier Variabler auftreten. Um zu einer solchen Verallgemeinerung zu gelangen, müssen wir daher einen anderen Weg einschlagen.

Gegeben seien für die n Variablen $x_1, x_2, \dots x_n$ empirische Wertreihen

$$x_i^{(1)}, x_i^{(2)} \dots x_i^{(n)} \quad (i = 1, 2 \dots n),$$

wobei die zum gleichen oberen Index gehöri- gen x -Werte einander zugeordnet sind. Trifft man die Bestimmung, daß allgemein

$$\sum_{i=1}^n x_i^{(i)} = 0$$

ist, so kommt die lineare Abhängigkeit der n Größen x_i in der Form $\sum_{i=1}^n \alpha_i x_i = 0$ zum Ausdruck und die α_i berechnen sich aus einem durch die Determinante