Aplikace matematiky

Miroslav Šisler

Approximative Formeln für den Fehler bei Iterationsverfahren von höherer Ordnung

Aplikace matematiky, Vol. 12 (1967), No. 1, 1–14

Persistent URL: http://dml.cz/dmlcz/103062

Terms of use:

© Institute of Mathematics AS CR, 1967

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.

This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* http://dml.cz

SVAZEK 12 (1967) APLIKACE MATEMATIKY ČÍSLO 1

APPROXIMATIVE FORMELN FÜR DEN FEHLER BEI ITERATIONSVERFAHREN VON HÖHERER ORDNUNG

MIROSLAV ŠISLER

(Eingegangen am 12. April 1966.)

I.

Das Finden der günstigen Fehlerabschätzungen gehört immer zu den sehr wichtigen Fragen bei jedem Iterationsverfahren. Wie schon in Arbeit [1] gesagt wurde, die in der Literatur oft angeführten Fehlerabschätzungen sind meistens nicht nur sehr übertrieben, sondern auch sehr schwer realisierbar. Nimmt man z.B. das bekannte Newton-Verfahren für die Lösung des Systems von nichtlinearen algebraischen Gleichungen (das ist im allgemeinen das Verfahren zweiter Ordnung). Die theoretisch abgeleiteten Fehlerabschätzungen erfordern hier die Berechnung der Abschätzung für alle ersten und zweiten partiellen Ableitungen von oben und der Abschätzung der Jacobi-Determinanten des Systems in gewisser Umgebung der gesuchten Lösung von unten. Das ist aber die schwere und für die Rechenmaschine nicht geignete Aufgabe. Falls man wegen der Vereinfachung anstatt der Abschätzungen der partiellen Ableitungen und der Jacobi-Determinanten in der Umgebung der gesuchten Lösung nur ihre Werte in einem Punkte benutzt, bekommt man Fehlerabschätzungen, welche theoretisch im allgemeinen nicht gelten, aber die gegebene Werte für den Fehler meistens zu grob sind. Solche Fehlerabschätzungen sind allerdings vom praktischen Gesichtspunkt schwer anwendbar.

In diesem Artikel werden daher einige günstigere approximative Formeln für den Fehler bei den Iterationsverfahren von der Ordnung $N \geqq 2$ abgeleitet werden. Die approximativen Formeln für den Fehler bei linearen Iterationsverfahren (d.h. für $N = 1$) wurden in der Arbeit [1] angegeben. Absichtlich wird der Terminus „die approximative Formel für den Fehler“ anstatt des Terminus „die Fehlerabschätzung“ benutzt um anzudeuten, dass die Ungleichungen für den Fehler nicht ganz exakt abgeleitet sind, doch ihre Gültigkeit kann man nur mit Ausnahme einer kleinen Anzahl der Anfangsapproximationen erwarten. Für folgende Approximationen geben sie allerdings sehr genaue Fehlerabschätzungen. Wegen der grösseren Übersichtlichkeit und Verständlichkeit werden in der folgenden Auslegung alle

Betrachtungen für den Fall der Verfahren zweiter Ordnung (d.h. für $N = 2$) durchgeführt. Für $N > 2$ kann man ganz analog fortschreiten und darum sind nur die resultierenden Formeln angegeben (die Formeln für $N = 2$ gibt man als Spezialfall der Formeln für $N > 2$ an). Dieser Fortgang wurde auch aus folgendem Grunde gewählt: die in der Praxis benutzten Methoden sind meistens von erster oder zweiter Ordnung, wogegen die Methoden höherer Ordnungen sehr selten benutzt werden.

II.

In der Arbeit setzt man voraus, dass es sich um das durch die folgende Formel definierte Iterationsverfahren handelt:

$$\mathbf{x}_{\nu+1} = \varphi(\mathbf{x}_\nu)\,, \quad \nu = 0, 1, 2, \dots \,. \tag{1}$$

Hier $\mathbf{x}_\nu$, $\mathbf{x}_{\nu+1}$ bezeichnet die n-dimensionalen Vektoren und $\varphi = (\varphi_1, \dots, \varphi_n)$, wo $\varphi_1, \dots, \varphi_n$ die Funtionen der n Veränderlichen sind, die in der Umgebung der Lösung $\boldsymbol{a}$ (d.h. des Punktes $\boldsymbol{a}$, für welchen die Gleichung $\boldsymbol{a} = \varphi(\boldsymbol{a})$ gilt) die vollständigen Differentiale bis $N + 1$ er Ordnung ($N > 1$) haben.

Wie werden sagen, dass das durch die Formel (1) definierte Iterationsverfahren von der Ordnung N im Punkte $\boldsymbol{a}$ ist, falls alle partiellen Ableitungen der Funktionen φ_i, $i = 1, \dots, n$, bis zu der Ordnung $N - 1$ im Punkte $\boldsymbol{a}$ gleich Null sind und wenigstens eine partielle Ableitung der Ordnung N im Punkte $\boldsymbol{a}$ von Null verschiedene ist.

Nehmen wir jetzt eine beliebige Vektornorm und führen wir die Begriffe von Fehler und Korrektion der ν-ten Approximation folgenderweise ein: Als den Fehler der ν-ten Approximation wird man die Zahl

$$\delta_\nu = \|\mathbf{x}_\nu - \boldsymbol{a}\|$$

und als die Korrektion der ν-ten Approximation wird man die Zahl

$$d_\nu = \|\mathbf{x}_{\nu+1} - \mathbf{x}_\nu\|$$

nehmen. Falls $\mathbf{x}_0$ die genügend gute Anfangsapproximation ist und falls die Folge $\{\mathbf{x}_\nu\}_{\nu=0}^{\infty}$ zu dem Punkt $\boldsymbol{a}$ konvergiert, folgt sofort nach dem Taylor-Satz für das Iterationsverfahren der Ordnung N im Punkte $\boldsymbol{a}$ die Ungleichung

$$\delta_{\nu+1} \leqq P\delta_\nu^N\,, \quad \nu = 0, 1, 2, \dots\,, \tag{2}$$

wo P eine positive Konstante ist. Die folgende Auslegung gilt für alle durch die Vorschrift (1) definierten Iterationsverfahren, für welche die Ungleichung (2) (wo $N > 1$) bewiesen werden kann.

Man kann beweisen, dass für die Korrektion die Ungleichung

$$d_{\nu+1} \leqq Q d_\nu^N \tag{3}$$

gilt; dabei ist Q eine positive Konstante. Die Ungleichung (3) ist also analog zu der Ungleichung (2). Genauer gesagt, es gilt folgende Behauptung: Sei ε eine beliebige positive Zahl. Dann existiert die natürliche Zahl ν_0 so, dass für $\nu \geqq \nu_0$ die Ungleichung

$$d_{\nu+1} \leqq (P + \varepsilon)\, d_\nu^N \tag{4}$$

gilt. Aus der Dreiecksungleichung und aus der Ungleichung (2) folgt nämlich, dass

$$d_{\nu+1} \leqq \delta_{\nu+2} + \delta_{\nu+1} \leqq P(\delta_{\nu+1}^N + \delta_\nu^N)\,,$$

d.h.

$$\frac{d_{\nu+1}}{d_\nu^N} \leqq P\left[\left(\frac{\delta_{\nu+1}}{d_\nu}\right)^N + \left(\frac{\delta_\nu}{d_\nu}\right)^N\right]. \tag{5}$$

In Absatz IV (Bemerkung 1) ist folgende Behauptung bewisen: Sei c eine beliebige positive Zahl. Dann existiert die natürliche Zahl ν_2 so, dass für alle $\nu \geqq \nu_2$ die Ungleichung

$$\frac{\delta_\nu}{d_\nu} \leqq 1 + c \tag{6}$$

gilt. Mit Rücksicht auf die Gleichung $\lim\limits_{\nu\to\infty} \delta_\nu = 0$ kann man voraussetzen, dass $\delta_\nu < 1$ ist. Nach (5) und (6) folgt für $\nu \geqq \nu_2$ stufenweise

$$\frac{d_{\nu+1}}{d_\nu^N} \leqq P\left[\left(\frac{P\delta_\nu^N}{d_\nu}\right)^N + \left(\frac{\delta_\nu}{d_\nu}\right)^N\right] \leqq P\left[P^N\delta_\nu^N\left(\frac{\delta_\nu}{d_\nu}\right)^N + \left(\frac{\delta_\nu}{d_\nu}\right)^N\right] =$$

$$= P\left[(P^N\delta_\nu^N + 1)\left(\frac{\delta_\nu}{d_\nu}\right)^N\right] \leqq P[(P^N\delta_\nu^N + 1)(1 + c)^N] = \alpha\,.$$

Der Limes des Ausdruckes in der gebrochenen Klammer ist für $\nu \to \infty$ und $c \to 0$ offensichtlich gleich 1. Zu jeder beliebigen Zahl $\varepsilon > 0$ existiert dann die Zahl $c > 0$ und die natürliche Zahl ν_1 so, dass für $\nu \geqq \nu_1$

$$0 \leqq (P^N\delta_\nu^N + 1)(1 + c)^N - 1 \leqq \varepsilon$$

ist und gilt also

$$\alpha \leqq P + \varepsilon\,.$$

Für $\nu \geqq \nu_0 = \max(\nu_2, \nu_1)$ gilt also die Ungleichung (4), wo ε eine gewählte positive Zahl ist.

Zum Schluss dieses Absatzes bemerken wir, dass aus der Ungleichung (4) ferner folgt, dass die Konstanten P und Q in den Ungleichungen (2) und (3) für genügend grosse ν approximativ gleich sind, d.h. es gilt $P \doteq Q$.

III.

Die praktische Bestimmung der Konstanten P und Q in den Beziehungen (2) und (3) ist theoretisch sehr schwer. Wir versuchen darum diese Konstanten aus dem Verlauf der Werte $d_0, d_1, d_2, \ldots$ abschätzen. Setzen wir voraus, dass $N = 2$ ist. Zuerst druckt man mathematisch die sinkende Tendenz der Zahlen $d_0, d_1, d_2, \ldots$ aus. Das kann man folgenderweise machen: setzt man voraus, dass die Approximationen $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{\nu+1}$ schon berechnet sind, d.h. dass man die Werte $d_0, \ldots, d_\nu$ kennt. Man approximiert jetzt die Punkte $A_\mu = [\mu, d_\mu]$, $\mu = 0, 1, \ldots, \nu$ in der Ebene mit den Koordinaten x, y durch die Funktion in der Form

$$y_\nu(x) = Q_\nu^{2^x - 1} \Lambda_\nu^{2^x}, \tag{7}$$

wo Q_ν, Λ_ν gewisse Konstanten sind (sie hängen allerdings von der Zahl ν ab). Wir wählen die Funktion $y_\nu(x)$ in der Form (7) mit Rücksicht auf die Ungleichung (3) (für $N = 2$). Um die Methode der kleinsten Quadrate zur Berechnung der Werte Q_ν, Λ_ν leicht benutzen zu können, geht man zur logarithmischen Form über und approximiert die Punkte $B_\mu = [\mu, \log d_\mu]$, $\mu = 0, 1, \ldots, \nu$, in der Ebene mit den Koordinaten x, Y durch die Funktion in der Form

$$Y_\nu(x) = (2^x - 1)\, q_\nu + 2^x l_\nu, \tag{8}$$

wo $q_\nu = \log Q_\nu$, $l_\nu = \log \Lambda_\nu$. Jetzt sucht man solche Zahlen q_ν, l_ν, dass der Ausdruck

$$\sum_{\mu=0}^{\nu} (Y_\nu(\mu) - \log d_\mu)^2$$

d.h. der Ausdruck

$$\sum_{\mu=0}^{\nu} [(2^\mu - 1)\, q_\nu + 2^\mu l_\nu - \log d_\mu]^2$$

minimal ist. Die entsprechenden Normalgleichungen haben die Form

$$\begin{aligned} q_\nu \sum_{\mu=0}^{\nu} (2^\mu - 1)^2 + l_\nu \sum_{\mu=0}^{\nu} 2^\mu (2^\mu - 1) &= \sum_{\mu=0}^{\nu} (2^\mu - 1) \log d_\mu\,, \\ q_\nu \sum_{\mu=0}^{\nu} 2^\mu (2^\mu - 1) + l_\nu \sum_{\mu=0}^{\nu} 2^{2\mu} &= \sum_{\mu=0}^{\nu} 2^\mu \log d_\mu\,. \end{aligned}$$

Falls man $S_\nu = \sum_{\mu=0}^{\nu} \log d_\mu$ und $T_\nu = \sum_{\mu=0}^{\nu} 2^\mu \log d_\mu$ legt, bekommt man aus diesen Gleichungen für q_ν den Ausdruck

$$q_\nu = \frac{3T_\nu - (2^{\nu+1} + 1)\, S_\nu}{2^{\nu+1}(\nu - 2) + (\nu + 4)}\,. \tag{9}$$

Den gesuchten Wert Q_ν bekommt man jetzt mit Hilfe der Formel $Q_\nu = e^{q_\nu}$. Den Wert Λ_ν berechnet man nicht, dann es ist nicht in folgender Auslegung nötig.

Für eine genügend grosse natürliche Zahl ν und $1 \leqq i \leqq \nu$ kan man jetzt approximativ

(10) $$d_i \approx y_\nu(i) = Q_\nu^{2^i - 1} \Lambda_\nu^{2^i}$$

legen. Es gilt also die approximative Formel

$$\frac{d_{i+1}}{d_i^2} \approx \frac{y_\nu(i+1)}{y_\nu(i)} = \frac{Q_\nu^{2^{i+1}-1} \Lambda_\nu^{2^{i+1}}}{(Q_\nu^{2^i - 1})^2 (\Lambda_\nu^{2^i})^2} = Q_\nu$$

und damit auch

(11) $$d_{i+1} \approx Q_\nu d_i^2 .$$

Mit Rücksicht auf das Faktum, dass die Konstanten P, Q in den Ungleichungen (2) und (3) für genügend grosse ν approximativ gleich sind, kann man für genügend grosse ν approximativ

(12) $$\delta_{i+1} \approx Q_\nu \delta_i^2 , \quad 1 \leqq i \leqq \nu$$

legen.

Jetzt kann man schon die Formel für den Fehler ableiten. Nach der Dreiecksungleichung und nach der Ungleichung (2) bekommt man

$$\delta_\nu \leqq d_\nu + \delta_{\nu+1} \approx d_\nu + Q_\nu \delta_\nu^2$$

oder

$$Q_\nu \delta_\nu^2 - \delta_\nu + d_\nu \geqq 0 .$$

Diese Ungleichung ist offensichtlich erfüllt für

$$\delta_\nu \geqq \frac{1 + \sqrt{(1 - 4Q_\nu d_\nu)}}{2Q_\nu}$$

und für

(13) $$\delta_\nu \leqq \frac{1 - \sqrt{(1 - 4Q_\nu d_\nu)}}{2Q_\nu} .$$

Mit Rücksicht auf das Faktum, dass $\delta_\nu \to 0$ ist, gilt nur die Formel (13), die gewöhnlich sehr genaue Werte für den Fehler der ν-ten Approximation bietet.

Für die praktische Berechnung bemerkt man noch, dass die Werte S_ν und T_ν, die bei jedem Schritt berechnet werden müssen, man rekurrent berechnet mit Hilfe der Beziehungen

$$S_{\nu+1} = S_\nu + \log d_{\nu+1} , \quad T_{\nu+1} = T_\nu + 2^{\nu+1} \log d_{\nu+1} .$$

Bemerken wir ferner, dass mit Rücksicht auf die approximativen Ungleichungen, die zur Ableitung der Ungleichung (13) benutzt wurden, nicht exakt ihre Gültigkeit für alle v garantiert wird. In der Praxis kann man aber erwarten, dass sie nur mit Ausnahme einer kleinen Anzahl der Anfangsapproximationen gilt, solange die Konvergenz noch nicht stabilisiert wird.

IV.

Ähnlicherweise wie bei den Methoden der zweiten Ordnung kann man auch bei den Iterationsverfahren von der Ordnung $N > 2$ fortschreiten. Wie schon in Absatz I gesagt wurde, deuten wir nur der Fortgang für die Ableitung der betreffenden Formeln an und führen nur die Haupformeln ein.

Wenn man die Punkte $A_\mu = [\mu, d_\mu]$, $\mu = 0, 1, \ldots, v$ durch die Funktion $y_v(x)$ in der Form

$$y_v(x) = Q_v^{(N^v - 1)/(N-1)} A_v^{N^v} \tag{14}$$

(die Funktion (7) ist offensichtlich der Speziellfall für $N = 2$) mit Hilfe der Methode der kleinsten Quadrate approximiert, bekommt man für die Konstante Q_v die Formel

$$\log Q_v = \frac{1}{(N-1)^4} \frac{(N+1)\,T_v - (N^{v+1}+1)\,S_v}{N^{v+1}[(N-1)\,v - 2] + (N-1)\,v + 2N}, \tag{15}$$

wo

$$S_v = \sum_{\mu=0}^{v} \log d_\mu\,, \quad T_v = \sum_{\mu=0}^{v} N^\mu \log d_\mu\,.$$

Die Formel (9) ist dann der Speziellfall von der Formel (15) für $N = 2$.

Falls man jetzt für genügend grosse

$$d_i \approx y_v(i)\,, \quad 1 \leqq i \leqq v$$

legt, ist

$$\frac{d_{i+1}}{d_i^N} \approx \frac{y_v(i+1)}{y_v(i)} = \frac{Q_v^{(N^{i+1}-1)/(N-1)} A_v^{N^{i+1}}}{(Q_v^{(N^i-1)/(N-1)})^N (A_v^{N^i})^N} = Q_v$$

und gilt also

$$d_{i+1} \approx Q_v d_i^N\,, \quad 1 \leqq i \leqq v\,. \tag{16}$$

Ähnlicherweise wie in Absatz III kann man dann auch

$$\delta_{i+1} \approx Q_v \delta_i^N\,, \quad 1 \leqq i \leqq v \tag{17}$$

legen.

Auf Grund der Beziehungen (16) und (17) kann man schon die betreffenden Formeln für den Fehler ableiten. Nach der Dreiecksungleichung und nach der Beziehung (17) folgt sofort

$$\delta_v \leqq \delta_{v+1} + d_v \approx Q_v \delta_v^N + d_v$$

oder

$$Q_\nu \delta_\nu^N - \delta_\nu + d_\nu \geqq 0 .$$

Die Gleichung

$$f_\nu(x) = Q_\nu x^N - x + d_\nu = 0$$

hat offensichtlich zwei positive Wurzeln $x_1 > 0$, $x_2 > 0$ und die Funktion $f_\nu(x)$ ist konvex für $x > 0$, denn

$$f''(x) = N(N-1)\, Q_\nu x^{N-2} > 0$$

ist. Die Ungleichung $f_\nu(x) \geqq 0$ ist dann für $x \leqq x_1$ und $x_2 \leqq x$ erfüllt. Prüft man jetzt die Wurzel x_1 von oben abschätzen. Es gilt offensichtlich $d_\nu \leqq x_1$, da

$$f_\nu(d_\nu) = Q_\nu d_\nu^N - d_\nu + d_\nu = Q_\nu d_\nu^N \geqq 0$$

ist. Nun wählt man die beliebige Zahl $k > 1$. Da $\lim\limits_{\nu \to \infty} d_\nu = 0$ ist und da die Zahlen Q_ν für genügeng grosse ν approximativ konstant sind, existiert eine solches ν_0, dass für $\nu \geqq \nu_0$ die Ungleichung

$$Q_\nu k^N d_\nu^{N-1} \leqq k - 1 \tag{18}$$

und damit auch die Ungleichung $x_1 \leqq k d_\nu$ gilt.[1]) Es gilt nämlich nach (18)

$$f_\nu(kd_\nu) = Q_\nu k^N d_\nu^N - kd_\nu + d_\nu = d_\nu[Q_\nu k^N d_\nu^{N-1} - (k-1)] \leqq 0 .$$

Durch den Punkt der graphischen Darstellung der Funktion $f_\nu(x)$ mit der ersten Koordinaten kd_ν führt man jetzt die Gerade mit dem Richtungstangens $f_\nu'(x) = NQ_\nu d^{N-1} - 1$. Sie schneidet die Achse x im Punkte

$$x_0 = \frac{d_\nu - (NQ_\nu k - Q_\nu k^N)\, d_\nu^N}{1 - NQ_\nu d_\nu^{N-1}} . \tag{19}$$

Nun beweist man, dass für $\nu \geqq \nu_0$ schon die Ungleichung $x_1 \leqq x_0$ gilt, d.h. dass x_0 die gesuchte Abschätzung von oben der Wurzel x_1 ist. Wenn wir

$$\frac{1 - (NQ_\nu k - Q_\nu k^N)\, d_\nu^{N-1}}{1 - NQ_\nu d_\nu^{N-1}} = A_\nu$$

legen, ist $A_\nu \leqq k$, denn mit Rücksicht auf die Beziehung (18) ist

$$A_\nu = \frac{1 - NQ_\nu k d_\nu^{N-1} + Q_\nu k^N d_\nu^{N-1}}{1 - NQ_\nu d_\nu^{N-1}} \leqq \frac{1 - NQ_\nu k d_\nu^{N-1} + (k-1)}{1 - NQ_\nu d_\nu^{N-1}} = k .$$

---

[1]) Die Wurzeln x_1, x_2 der Gleichung $f_\nu(x) = 0$ hängen allerdings von ν ab.

Es ist also $x_0 = d_\nu A_\nu \leqq k d_\nu$. Ferner gilt

$$\begin{aligned} f_\nu(x_0) &= f_\nu(d_\nu A_\nu) = Q_\nu d_\nu^N A_\nu^N - d_\nu A_\nu + d_\nu = d_\nu(Q_\nu d_\nu^{N-1} A_\nu^N - A_\nu + 1) = \\ &= d_\nu[A_\nu(Q_\nu d_\nu^{N-1} A_\nu^{N-1} - 1) + 1] \leqq d_\nu[k(Q_\nu d_\nu^{N-1} k^{N-1} - 1) + 1] = \\ &= d_\nu[Q_\nu k^N d_\nu^{N-1} - k + 1] \leqq d_\nu[k - 1 - k + 1] = 0\,. \end{aligned}$$

Es ist also $f_\nu(x_0) \leqq 0$, sodass wirklich $x_1 \leqq x_0$ gilt. Da nun die Ungleichung $f_\nu(x) \geqq 0$ für $x \leqq x_1$ erfüllt ist und da $x_1 \leqq x_0$ ist, gilt offensichtlich nach Ungleichung (18) die Beziehung

$$\delta_\nu \leqq \frac{d_\nu - (NQ_\nu k - Q_\nu k^N)\, d_\nu^N}{1 - NQ_\nu d_\nu^{N-1}}, \tag{20}$$

solange ν so gross ist, dass die Ungleichung (18) gilt, d.h. solange die Ungleichung

$$d_\nu \leqq \sqrt[N-1]{\frac{k-1}{Q_\nu k^N}} \tag{21}$$

erfüllt ist.

Es gilt wieder dasselbe, was über die Formel (13) gesagt wurde. Mit Rücksicht auf die zur Ableitung der Formel (20) benutzten approximativen Formeln, kann man ihre Gültigkeit für alle ν mit Ausnahme einer kleinen Anzahl der Anfangsapproximationen erwarten.

Bemerkung 1. Den bereits durchgeführten Fortgang zur Ableitung der Ungleichung (20) kann man leicht zum Beweis der Formel

$$\frac{\delta_\nu}{d_\nu} \leqq 1 + c \tag{6}$$

benutzen, wo c eine beliebige positive Zahl ist und $\nu \geqq \nu_2$ (hier ist ν_2 eine genügend grosse natürliche Zahl). (Diese Beziehung wurde im Absatz II zum Beweis der Ungleichung (4) benutzt.) Nach der Ungleichung (2) und nach der Dreiecksungleichung folgt nämlich die Ungleichung

$$\delta_\nu \leqq \delta_{\nu+1} + d_\nu \leqq P\delta_\nu^N + d_\nu$$

oder

$$P\delta_\nu^N - \delta_\nu + d_\nu \geqq 0\,.$$

Die Gleichung $Px^N - x + d_\nu = 0$ hat, wie schon bekannt ist, zwei positive Wurzel. Falls x_1 die kleiner Wurzel bezeichnet, ist die vorhergehende Ungleichung erfüllt, wenn $\delta_\nu \leqq x_1$. Ganz analog, wie oben, kann man jetzt die folgende Behauptung beweisen: Sei k eine beliebige Zahl und $k > 1$. Sei ferner ν_0 eine solche natürliche Zahl, dass für alle $\nu \geqq \nu_0$ die Ungleichung

$$Pk^N d_\nu^{N-1} \leqq k - 1$$

gilt. Dann ist

$$x_1 \leqq \frac{d_\nu - (NPk - Pk^N)\, d_\nu^N}{1 - NPd_\nu^{N-1}},$$

d.h. es gilt die Abschätzung

$$\delta_\nu \leqq \frac{d_\nu - (NPk - Pk^N)\, d_\nu^N}{1 - NPd_\nu^{N-1}}.$$

Nach dieser Ungleichung folgt sofort die Ungleichung

$$\frac{\delta_\nu}{d_\nu} \leqq \frac{1 - (NPk - Pk^N)\, d_\nu^{N-1}}{1 - NPd_\nu^{N-1}}.$$

Da der Limes der rechten Seite für $\nu \to \infty$ gleich 1 ist, existiert eine solche Zahl ν_1, dass für $\nu \geqq \nu_1$

(6) $$\frac{\delta_\nu}{d_\nu} \leqq 1 + c$$

gilt (hier ist c eine beliebig gewählte positive Zahl). Die Ungleichung (6) gilt also für alle $\nu \geqq \max(\nu_0, \nu_1) = \nu_2$.

V.

Die Formel (20) für den Fehler, die bereits abgeleitet wurde, kann man vorteilhaft bei Iterationsmethoden von der zweiten Ordnung benutzen. Für $N = 2$ ist nämlich

(22) $$\delta_\nu \leqq \frac{d_\nu - (2Q_\nu k - Q_\nu k^2)\, d_\nu^2}{1 - 2Q_\nu d_\nu}$$

solange

(23) $$Q_\nu k^2 d_\nu \leqq k - 1.$$

Falls man ferner $k = 2$ legt, bekommt man nach (22) die Ungleichung

(24) $$\delta_\nu \leqq \frac{d_\nu}{1 - 2Q_\nu d_\nu},$$

solange

$$4Q_\nu d_\nu \leqq 1$$

oder

(25) $$d_\nu \leqq \frac{1}{4Q_\nu}.$$

Durch Vergleich der Ungleichungen (24) und (13) kann man sofort feststellen, dass die Ungleichung (24) für den Fehler ein wenig grösser Werte bietet (das ist klar, denn im Falle der Formel (13) benutzten wir zur Abschätzung des Fehlers δ_ν gerade die Wurzel der Gleichung $Q_\nu x^2 - x + d_\nu = 0$, wogegen bei der Ungleichung (24) eine gewisse Abschätzung dieser Wurzel von oben benutzt wurde). Die Formel (24) hat aber folgenden Vorteil: sie ist sehr einfach und enthält keine Wurzel.

VI.

Zum Schluss führen wir ein numerisches Beispiel an. Es wird das folgende Gleichungsystem mit Hilfe des Newton-Verfahren (hier ist $N = 2$) gelöst:

$$\begin{aligned} x^2 + y^2 + z^2 + v^2 - 30 &= 0\,, \\ xyzv - 24 &= 0\,, \\ 3x - 2y + 4z - v - 7 &= 0\,, \\ x^2 z + y^3 v - 2xv - 27 &= 0\,. \end{aligned}$$

Um den wirklichen Fehler mit dem durch die Formeln (13) und (24) gegebenen Abschätzungen vergleichen zu können, ist das Gleichungsystem so gewählt, dass seine genaue Lösung $x = 1$, $y = 2$, $z = 3$, $v = 4$ bekannt ist. Als die Anfangsapproximation nimmt man die Werte $x_0 = 0$, $y_0 = 3$, $z_0 = 5$, $v_0 = 7$. Die sukzessiven Approximationen, ihre wirklichen Fehler und die Fehlerabschätzungen (13) und (24) sind in der Tabelle 1 zusammengefasst. (Die Vektornorm ist dabei folgenderweise gewählt: für $\mathbf{x} = (x_i)$ ist $\|\mathbf{x}\| = \max |x_i|$.)

Literatur

[1] *M. Šisler:* Approximative Formeln für den Fehler bei Iterationsverfahren. Apl. mat. *11* (1966), 341–351.

Tabelle 1

ν	x_ν	y_ν	z_ν	v_ν
0	.000000000/+00	.299999999/+01	.500000001/+01	.700000001/+01
1	.228571429/+00	.256211741/+01	.390555429/+01	.418369661/+01
2	.662997489/+00	.215631927/+01	.329537690/+01	.385786156/+01
3	.948519671/+00	.200373374/+01	.304130061/+01	.400329397/+01
4	.998888766/+00	.200001262/+01	.300087464/+01	.400013963/+01
5	.999999495/+00	.200000000/+01	.300000038/+01	.400000006/+01
6	.100000000/+01	.200000000/+01	.299999999/+01	.400000001/+01

Výtah

PŘIBLIŽNÉ VZORCE PRO CHYBU PŘI ITERAČNÍCH METODÁCH VYŠŠÍHO ŘÁDU

MIROSLAV ŠISLER

V článku jsou podány některé přibližné vzorce pro chybu postupných aproximací $\boldsymbol{x}_\nu$ počítaných podle vzorce tvaru

$$\boldsymbol{x}_{\nu+1} = \varphi(\boldsymbol{x}_\nu), \quad \nu = 0, 1, 2, \ldots$$

($\boldsymbol{x}_{\nu+1}$, $\boldsymbol{x}_\nu$ jsou n-rozměrné vektory a $\varphi = (\varphi_1, \ldots, \varphi_n)$, kde $\varphi_1, \ldots, \varphi_n$ jsou funkce n-proměnných, mající v okolí hledaného řešení totální diferenciál až do $N + 1$-ho řádu). Je-li $\boldsymbol{a}$ hledané řešení (tj. platí-li $\boldsymbol{a} = \varphi(\boldsymbol{a})$) a je-li $\lim_{\nu\to\infty} \boldsymbol{x}_\nu = \boldsymbol{a}$, pak chybou ν-té aproximace rozumíme číslo $\delta_\nu = \|\boldsymbol{x}_\nu - \boldsymbol{a}\|$ a opravou ν-té aproximace číslo $d_\nu = \|\boldsymbol{x}_{\nu+1} - \boldsymbol{x}_\nu\|$. Předpokládáme, že se jedná o iterační metodu řádu $N \geqq 2$ v bodě $\boldsymbol{a}$ (iterační metoda je řádu N v bodě $\boldsymbol{a}$, jestliže všechny parciální derivace funkcí φ_i až do řádu $N - 1$ jsou v bodě $\boldsymbol{a}$ rovny nule a alespoň jedna parciální derivace řádu N je v bodě $\boldsymbol{a}$ od nuly různá). Pro iterační metody N-tého řádu lze pak dokázat nerovnosti $\delta_{\nu+1} \leqq P\delta_\nu^N$, $d_{\nu+1} \leqq Qd_\nu^N$, $\nu = 0, 1, 2, \ldots$, přičemž pro dostatečně velké ν je $P \doteq Q$. Pro chybu jsou pak v článku odvozeny tyto vzorce: pro $N = 2$ je

$$\delta_\nu \leqq \frac{1 - \sqrt{(1 - 4Q_\nu d_\nu)}}{2Q_\nu}$$

pro $N \geqq 2$ je

$$\delta_\nu \leqq \frac{d_\nu - (NQ_\nu k - Q_\nu k^N)\, d_\nu^N}{1 - NQ_\nu d_\nu^{N-1}}$$

d_ν	Q_ν	Abschätzung (13) für δ_ν	Abschätzung (24) für δ_ν	Wirkliche Fehler δ_ν
.281630388/+01				.300000000/+01
.610177388/+00	.769302656/−01	.641872698/+00	.67339743/+00	.905554288/+00
.285522180/+00	.242891568/+00	.308663179/+00	.33150214/+00	.337002511/+00
.503690948/−01	.356644610/+00	.513079641/−01	.52246181/−01	.514803288/−01
.111072809/−02	.396817064/+00	.111121626/−02	.11117080/−02	.111123361/−02
.506923599/−06	.408821179/+00	.508009234/−06	.50801384/−06	.504776833/−06
.283946976/−07	.186495434/+01	.284647111/−07	.28496601/−07	.149011612/−07

pokud

$$d_\nu \leqq \sqrt[N-1]{\frac{k-1}{Q_\nu k^N}}$$

(k je libovolné reálné číslo a $k > 1$). Pro $k = 2$ a $N = 2$ pak dostáváme speciálně

$$\delta_\nu \leqq \frac{d_\nu}{1 - 2Q_\nu d_\nu},$$

pokud $d_\nu \leqq 1/(4Q_\nu)$.

Číslo Q_ν, které se vyskytuje ve všech vzorcích, je dáno vzorcem

$$\log Q_\nu = \frac{1}{(N-1)^4} \frac{(N+1)\,T_\nu - (N^{\nu+1} + 1)\,S_\nu}{N^{\nu+1}[(N-1)\,\nu - 2] + (N-1)\,\nu + 2N},$$

kde

$$S_\nu = \sum_{\mu=0}^{\nu} \log d_\mu\,, \quad T_\nu = \sum_{\mu=0}^{\nu} N^\mu \log d_\mu\,.$$

Pro $N = 2$ je speciálně

$$\log Q_\nu = \frac{3T_\nu - (2^{\nu+1} + 1)\,S_\nu}{2^{\nu+1}(\nu - 2) + (\nu + 4)},$$

kde

$$S_\nu = \sum_{\mu=0}^{\nu} \log d_\mu\,, \quad T_\nu = \sum_{\mu=0}^{\nu} 2^\mu \log d_\mu\,.$$

Vzorce pro Q_ν jsou odvozeny pomocí metody nejmenších čtverců, jestliže při ν-tém kroku přibližně aproximujeme již vypočtené opravy d_i, $0 \leqq i \leqq \nu$, pomocí hodnot funkce tvaru $y_\nu(x) = Q_\nu^{(N^\nu - 1)/(N-1)} A_\nu^{N^\nu}$ (pro $N = 2$ je funkce y_ν tvaru $y_\nu(x) = Q_\nu^{2^\nu - 1} A_\nu^{2^\nu}$) v bodech $x = i$. Odvozené odhady pro chybu δ_ν platí pro všechny ν s výjimkou malého počtu počátečních aproximací. Jakost odhadů je patrna z numerického příkladu.

Резюме

ПРИБЛИЖЕННЫЕ ФОРМУЛЫ ДЛЯ ПОГРЕШНОСТИ ПРИ ИТЕРАЦИОННЫХ МЕТОДАХ ВЫСШЕГО ПОРЯДКА

МИРОСЛАВ ШИСЛЕР (MIROSLAV ŠISLER)

В работе приводятся некоторые приближенные формулы для погрешности последовательных приближений $\boldsymbol{x}_\nu$, вычисляемых при помощи формулы вида

$$\boldsymbol{x}_{\nu+1} = \varphi(\boldsymbol{x}_\nu)\,, \quad \nu = 0, 1, 2, \ldots$$

($\boldsymbol{x}_{\nu+1}$, $\boldsymbol{x}_\nu$ — n-мерные векторы и $\varphi = (\varphi_1, \ldots, \varphi_n)$, где $\varphi_1, \ldots, \varphi_n$ — функции от n переменных, имеющие в окрестности искомого решения полный дифференциал $N + 1$-го порядка). Если $\boldsymbol{a}$-точное решение (значит, если $\boldsymbol{a} = \varphi(\boldsymbol{a})$) и если $\lim\limits_{\nu \to \infty} \boldsymbol{x}_\nu = \boldsymbol{a}$, то число $\delta_\nu = \|\boldsymbol{x}_\nu - \boldsymbol{a}\|$ обозначает погрешность ν-того приближения и число $d_\nu = \|\boldsymbol{x}_{\nu+1} - \boldsymbol{x}_\nu\|$ обозначает поправку ν-того приближения. Совместно с тем мы предполагаем, что итерационный метод порядка $N \geqq 2$ в точке $\boldsymbol{a}$, если все частные производные порядка меньшего N в точке $\boldsymbol{a}$ равны нулю и хотя бы одна частная производная порядка N в точке $\boldsymbol{a}$ нулю не равняется). Для итерационных методов порядка N можно потом доказать неравенства $\delta_{\nu+1} \leqq P\delta_\nu^N$, $d_{\nu+1} \leqq Q d_\nu^N$, $\nu = 0, 1, 2, \ldots$; здесь числа P и Q приблизительно равняются друг другу для достаточно большего числа ν. Для погрешности в работе доказываются следующие формулы: для $N = 2$ имеет место формула

$$\delta_\nu \leqq \frac{1 - \sqrt{(1 - 4Q_\nu d_\nu)}}{2Q_\nu}\,;$$

для $N \geqq 2$ имеет место формула

$$\delta_\nu \leqq \frac{d_\nu - (NQ_\nu k - Q_\nu k^N)\, d_\nu^N}{1 - NQ_\nu d_\nu^{N-1}}\,,$$

если

$$d_\nu \leqq \sqrt[N-1]{\frac{k - 1}{Q_\nu k^N}}$$

(k — произвольное действительное число, $k > 1$). Для $k = 2$ и $N = 2$ получается, в частности, неравенство

$$\delta_\nu \leqq \frac{d_\nu}{1 - 2Q_\nu d_\nu}\,,$$

если $d_\nu \leqq 1/(4Q_\nu)$.

Число Q_ν получается при помощи формулы

$$\log Q_\nu = \frac{1}{(N-1)^4} \frac{(N+1)\,T_\nu - (N^{\nu+1}+1)\,S_\nu}{N^{\nu+1}[(N-1)\,\nu - 2] + (N-1)\nu + 2N}\,;$$

здесь $S_\nu = \sum_{\mu=0}^{\nu} \log d_\mu$, $T_\mu = \sum_{\mu=0}^{\nu} N^N \log d_\mu$. В частном случае для $N = 2$ имеет место формула

$$\log Q_\nu = \frac{3T_\nu - (2^{\nu+1}+1)\,S_\nu}{2^{\nu+1}(\nu-2) + (\nu+4)}\,;$$

где $S_\nu = \sum_{\mu=0}^{\nu} \log d_\mu$, $T_\nu = \sum_{\mu=0}^{\nu} 2^\mu \log d_\mu$. Формула для Q_ν получается при помощи метода наименьших квадратов, если мы при ν-том шаге приблизительно выражаем уже вычисленные поправки d_i, $0 \leqq i \leqq \nu$, при помощи значений функции вида $y_\nu(x) = Q_\nu^{(N^\nu-1)/(N-1)} \Lambda_\nu^{N^\nu}$ (для $N = 2$ функция y_ν имеет вид $y_\nu(x) = Q_\nu^{2^\nu-1} \Lambda_\nu^{2^\nu}$) в точках $x = i$. Формулы для погрешности δ_ν имеют место для всех ν за исключением некоторого небольшого числа начальных приближений. Точность введенных формул для δ_ν очевидна из пирведенного численного примера.

Adresse des Autors: Dr. *Miroslav Šisler* C.Sc., Matematický ústav ČSAV, Žitná 25, Praha 1.