ÜBER DIE APPROXIMATION DER FUNKTIONEN UND DER LÖSUNGEN EINER GLEICHUNG DURCH QUADRATISCHE INTERPOLATION

von Tiberiu Popoviciu in Cluj

1. Im folgenden bezeichnet $f = f(x)$ stets eine reellwertige Funktion, die auf einem Intervall I erklärt ist, dessen Länge von Null verschieden ist. Die Bedingungen, die diese Funktion erfüllt, werden wir im Laufe der Ausführungen angeben.

Weiterhin bezeichnet $[x_1, x_2, \ldots, x_{n+1}; f]$ die dividierte Differenz (n-ter Ordnung) und $L(x_1, x_2, \ldots, x_{n+1}; f \mid x)$ das Interpolationspolynom von Lagrange-Hermite der Funktion f bezüglich der Knotenpunkte $x_1, x_2, \ldots, x_{n+1}$. Diese Knotenpunkte können sämtlich voneinander verschieden sein oder nicht. Im letzteren Fall treten bekanntlich in der dividierten Differenz und im Interpolationspolynom Ableitungen der Funktion auf den Knotenpunkten auf.

2. Wir setzen nun voraus, daß die Funktion f den folgenden beiden Bedingungen genügt:

I. *die Gleichung*

$$f(x) = 0 \tag{1}$$

hat mindestens eine Lösung im Innern des Intervalls I.

II. *f ist eine konvexe oder konkave Funktion 0-ter, 1-ter und 2-ter Ordnung.*

Eine Funktion heißt konvex, nicht-konkav, nicht-konvex bzw. konkav von n-ter Ordnung $(n \geq -1)$, wenn ihre dividierte Differenz $(n+1)$-ter Ordnung für jedes System von $n+2$ voneinander verschiedenen Punkten des Definitionsbereiches positiv, nicht-negativ, nicht-positiv bzw. negativ ist. In allen diesen Fällen be-

sitzt dann auch die dividierte Differenz für jedes beliebige System von $n+2$ Punkten, die nicht sämtlich übereinstimmen, die gleiche Eigenschaft, vorausgesetzt, daß diese dividierte Differenz existiert.

Man kann zeigen, daß eine Funktion f, die die Bedingungen I und II erfüllt, stetig und im Innern des Intervalls I stetig differenzierbar ist, und daß die Gleichung (1) genau eine Lösung z hat.

Gilt für zwei innere Punkte a, b des Intervalls I die Ungleichung $a < z < b$, so bezeichnen wir a als unteren und b als oberen Näherungswert von z. Bekanntlich kann man dann mit der Regula falsi und dem Verfahren von Raphson-Newton bessere Näherungswerte von z bestimmen. Eine weitere Möglichkeit, bessere Näherungswerte zu erhalten, geben wir im folgenden an. O.B.d.A. kann man annehmen, daß die Funktion f steigend und konvex im üblichen Sinn ist, gemäß unserer Terminologie also konvexe Funktion 0-ter und 1-ter Ordnung. Bezeichnen z', z'' die (einzigen) Nullstellen der Polynome $L(a,b;f\,|\,x)$, $L(b,b;f\,|\,x)$, so gilt $a < z' < z < z'' < b$.

3. Weitere Näherungswerte für die Lösung z sind auch die im offenen Intervall $]a,b[$ gelegenen (einzigen) Nullstellen z_1', z_1'' der Polynome $L(a,a,b;f\,|\,x)$, $L(a,b,b;f\,|\,x)$. In der Tat, beachtet man die Gleichungen

$$f(x) - L(a,a,b;f\,|\,x) = (x-a)^2(x-b)\,[a,a,b,x;f]$$

$$f(x) - L(a,b,b;f\,|\,x) = (x-a)\,(x-b)^2[a,b,b,x;f]\,,$$

so ergibt sich für $a < x < b$ die Beziehung

$$L(a,b,b;f\,|\,x) < f(x) < L(a,a,b;f\,|\,x), \tag{2}$$

im Falle einer konvexen Funktion 2-ter Ordnung, bzw. die Beziehung

$$L(a,b,b;f\,|\,x) > f(x) > L(a,a,b;f\,|\,x), \tag{3}$$

im Falle einer konkaven Funktion 2-ter Ordnung. Daraus folgt dann $z_1' < z < z_1''$, falls f konvex, bzw. $z_1'' < z < z_1'$, falls f konkav von 2-ter Ordnung ist.

4. Es ist natürlich von Bedeutung, die erhaltenen Näherungswerte z', z'' und z_1', z_1'' zu vergleichen. Zu diesem Zweck unterscheiden wir zwei Fälle.

Ist f eine konvexe Funktion 2-ter Ordnung, so folgen für $a < x < b$ aus (2) und aus

$$L(a,a,b;f\,|\,x) - L(a,b;f\,|\,x) = (x-a)(x-b)\,[a,a,b;f]$$
$$L(a,b,b;f\,|\,x) - L(b,b;f\,|\,x) = (x-b)^2[a,b,b;f]$$

die Ungleichungen

$$L(a,a,b;f\,|\,x) < L(a,b;f\,|\,x), \quad L(b,b;f\,|\,x) < L(a,b,b;f\,|\,x).$$

Es gibt also $z' < z_1' < z < z_1'' < z''$, *d.h.* z_1', z_1'' *sind bessere Näherungswerte als* z', z''.

Ist f eine konkave Funktion 2-ter Ordnung, so folgt aus (3) und aus

$$L(a,b,b;f\,|\,x) - L(a,b;f\,|\,x) = (x-a)\,(x-b)\,[a,b,b;f]$$

die Beziehung $z' < z_1'' < z$. z_1'' *ist also ein besserer unterer Näherungswert für* z *als* z'.

5. Was die Näherungswerte z_1' und z'' von z anbelangt, so kann man diese im Falle einer konkaven Funktion 2-ter Ordnung im allgemeinen nicht vergleichen. Dies zeigen uns die folgenden Überlegungen. Aus den Gleichungen

$$(4)\quad L(a,a,b;f|x) - L(b,b;f|x) =$$
$$= (x-b)\,\{(x-a)\,[a,a,b;f] - (b-a)\,[a,b,b;f]\} =$$
$$= (x-b)\,\{(x-b)\,[a,a,b;f] - (b-a)^2[a,a,b,b;f]\}$$

folgt, daß die Differenz $L(a,a,b;f|x) - L(b,b;f\,|\,x)$ in einer Umgebung des Punktes a (rechts von a) positiv und in einer Umgebung des Punktes b (links von b) negativ ist. Demnach besitzt das Polynom (4) genau eine Nullstelle ξ im Intervall $]a,b[$. Setzt man $m = L(b,b;f\,|\,\xi)$, so gilt $f(a) < m < f(b)$. Wählt man nun eine Konstante k, so daß $f(a) < k < f(b)$ gilt, dann stellt man sofort fest, daß die Funktion $\varphi(x) = f(x) - k$ den gleichen Bedingungen I und II genügt wie die Funktion f. Weiterhin ist

$$L(a,a,b;\varphi\,|\,x) - L(b,b;\varphi\,|\,x) = L(a,a,b;f\,|\,x) - L(b,b;f\,|\,x),$$

und die der Funktion φ entsprechenden Werte z_1', z'' sind beide links von ξ, falls

(5) $$f(a) < k < m$$

ist, bzw. beide rechts von ξ, falls

(6) $$m < k < f(b)$$

ist. Beachtet man nun das Vorzeichen der Differenz (4), so folgt, daß für die Funktion φ die Beziehung $z < z_1' < z''$ gilt, falls k der Bedingung (5) genügt und $z < z'' < z_1'$, falls k der Bedingung (6) genügt. Im Falle $k = m$ ist $z_1' = z''$ für die Funktion φ.

6. Sind die Bedingungen I und II erfüllt, so hat auch das Polynom $L(a,a,b;f\,|\,x) + L(a,b,b;f\,|\,x)$ gleichfalls eine einzige Nullstelle z_1 im Intervall $]a,b[$. Die Zahl z_1 befindet sich strikt zwischen z_1', z_1'' und ist demnach ein besserer Näherungswert von z als der schlechteste der Werte z_1', z_1''.

Beachtet man, daß

$$2f(x) - L(a,a,b;f\,|\,x) - L(a,b,b;f\,|\,x) =$$
$$= (x-a)(x-b)\{(x-a)[a,a,b,x;f] + (x-b)[a,b,b,x;f]\}$$

ist, sowie die Eigenschaften der konvexen Funktionen 2-ter Ordnung, so folgt, daß die linke Seite genau eine Nullstelle η im Intervall $]a,b[$ hat. Wählt man nun eine Konstante k, so daß $f(a) < k < f(b)$ gilt, dann genügt die Funktion $\varphi(x) = f(x) - k$ den gleichen Bedingungen I und II wie die Funktion f. Ist $k \neq n = f(\eta)$, so sind die der Funktion φ entsprechenden Punkte z, z_1 beide links oder beide rechts von η. Weiterhin ist $f(a) < n < f(b)$. Setzt man also voraus, daß f eine konvexe Funktion 2-ter Ordnung ist, so gilt für die Funktion φ,

$$n < k < f(b) \Longrightarrow z_1 < z,$$
$$f(a) < k < n \Longrightarrow z < z_1.$$

Im Falle einer konkaven Funktion 2-ter Ordnung muß man auf der rechten Seite dieser beiden Formeln das Zeichen < durch > ersetzen.

7. Ähnliche Überlegungen lassen sich durchführen, wenn man voraussetzt, daß die Funktion f fallend und konkav, fallend und konvex, oder steigend und konkav 1-ter Ordnung ist. Diese Fälle kann man auf den untersuchten Fall zurückführen,

indem man die erzielten Ergebnisse auf die Funktionen $-f(x)$, $f(\frac{a+b}{2} - x)$, oder $-f(\frac{a+b}{2} - x)$ anwendet.

Bedingung II kann abgeschwächt werden. Statt der Voraussetzung, daß f konvex von 0-, 1-, und 2-ter Ordnung ist, kann man annehmen, daß f nicht-konkav oder nicht-konvex von 0-, 1- und 2-ter Ordnung sei. Einige oder alle der für z, z', z'', z_1', z_1'', z_1 bewiesenen Ungleichungen können dann in Gleichungen übergehen.

Bezüglich Bedingung II kann bemerkt werden, daß eine Funktion f mit positiver, nicht-negativer, nicht-positiver bzw. negativer $(n+1)$-ter Ableitung konvex, nicht-konkav, nicht-konvex bzw. konkav von n-ter Ordnung ist. Andererseits ist jede auf I konvexe, nicht-konkave, nicht-konvexe oder konkave Funktion n-ter Ordnung $(n > 1)$ im Innern des Intervalls I $(n-1)$-mal (stetig) differenzierbar.

Numerisches Beispiel. Gegeben sei die Funktion $f(x) = x^3 - x - 1$. Man kann sofort feststellen, daß diese Funktion in einem entsprechend gewählten Intervall, das die Punkte 1 und 2 enthält, eine konvexe Funktion 0-, 1- und 2-ter Ordnung ist. Wegen $f(1)\,f(2) < 0$ hat die Gleichung (1) zwischen den Punkten 1 $(= a)$ und 2 $(= b)$ eine Wurzel. Wendet man die Regula falsi und das Verfahren von Raphson-Newton an, so erhält man die Näherungswerte $\frac{7}{6} = 1,1\dot{6}$, $\frac{17}{11} = 1,\dot{5}\dot{4}$ für diese Wurzel. Die Polynome $L(1,1,2;f|x)$, $L(1,2,2;\, f\,|\,x)$ sind in diesem Fall $4x^2 - 6x + 1$, $5x^2 - 9x + 3$ und liefern die Näherungswerte $\frac{3+\sqrt{5}}{4} > 1,3$, $\frac{9+\sqrt{21}}{10} < 1,36$.

Demnach ist $1,3\ldots$ der Wert der positiven Wurzel der Gleichung $x^3 - x - 1 = 0$, wobei eine Dezimalstelle genau ist. Die Nullstelle $\frac{4}{3} = 1,\dot{3}$ des Polynoms $L(1,1,2;\, f\,|\,x) + L(1,2,2;\, f\,|\,x)$, die sich zwischen 1 und 2 befindet, führt uns zum gleichen Ergebnis.

8. Die Beziehungen (2) und (3) zeigen, daß die Polynome $L(a,b,b;f\,|\,x)$, $L(a,\underline{a},b;f\,|\,x)$ im Falle einer konvexen oder konkaven Funktion 2-ter Ordnung die Funktion f im Intervall $]a,b[$ sowohl von unten als auch von oben annähern. Setzen wir z.B. voraus, daß f im Intervall I konvex von 2-ter Ordnung ist, dann folgt, daß f stetig und im Innern von I (stetig) differenzierbar ist.

Sind c, a, b, d vier innere Punkte von I mit $c \leq a < b \leq d$, so ergibt sich aus

$$f(x) - L(c,a,b;f|x) = (x-c)(x-a)(x-b)[c,a,b,x;f]$$

$$f(x) - L(a,b,d;f|x) = (x-a)(x-b)(x-d)[a,b,d,x;f]$$

für $a < x < b$ die Beziehung

$$L(a,b,d;f|x) < f(x) < L(c,a,b;f|x),$$

die die Ungleichungen (2) verallgemeinert. Aus $(c < a < b < d)$

$$L(a,b,d;f|x) - L(a,b,b;f|x) = (x-a)(x-b)(d-b)[a,b,b,d;f]$$

$$L(a,a,b;f|x) - L(c,a,b;f|x) = (x-a)(x-b)(a-c)[c,a,a,b;f]$$

ergibt sich weiterhin, daß *von allen Näherungswerten* $L(a,b,d;f|x)$, $L(c,a,b;f|x)$ *mit* $c \leq a < b \leq d$, $L(a,b,b;f|x)$ *der beste untere Näherungswert von* $f(x)$ *und* $L(a,a,b;f|x)$ *der beste obere Näherungswert von* $f(x)$ *für* $a < x < b$ *ist.*

Eine ähnliche Eigenschaft kann man auch für konkave Funktionen 2-ter Ordnung beweisen. Nur ist in diesem Fall der Sinn der Ungleichungen umgekehrt.

Obige Betrachtungen lassen sich z.B. auf folgende Funktionen anwenden: $\ln x$, die auf der Menge der positiven reellen Zahlen konvex von 2-ter Ordnung, $\operatorname{arctg} x$, die auf dem Intervall $[-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}]$ konkav von 2-ter Ordnung und $\frac{1}{2}\ln(1+x^2)$, die auf dem Intervall $[0, \sqrt{3}]$ konkav von 2-ter Ordnung ist.

9. Wir setzen im folgenden voraus, daß die auf dem Intervall I erklärte Funktion f hinreichend oft differenzierbar ist, so daß alle dividierten Differenzen und Interpolationspolynome, die vorkommen werden, existieren.

Um die durch eine Tabelle dargestellte Funktion f zu interpolieren, wählen wir vier Werte c, a, b, d der Veränderlichen mit $c < a < b < d$ und approximieren die Funktion $f(x)$, wobei $a < x < b$ ist, durch das arithmetische Mittel

$$P(c,a,b,d;f|x) = \frac{1}{2}\{L(c,a,b;f|x) + L(a,b,d;f|x)\}$$

der Polynome $L(c,a,b;f|x)$, $L(a,b,d;f|x)$. Der Fehler ist dann gleich

(7) $$d(x) = f(x) - P(c,a,b,d;f|x) =$$

$$= \frac{1}{2}(x-a)(x-b)\{(x-c)(c-d)[c,a,x,b,d;f] + (2x-c-d)[a,x,b,d;f]\}.$$

Den Fehler $d_1(x)$, der sich bei der Approximation von $f(x)$ durch das arithmetische Mittel der Polynome $L(a,a,b;f|x)$, $L(a,b,b;f|x)$, also durch $P(a,a,b,b;f|x)$, ergibt, kann man erhalten, indem man in (7) c, d durch a, b ersetzt. Eine einfache Rechnung ergibt dann

(8) $$d(x) - d_1(x) = \frac{1}{2}(x-a)(x-b)\{(x-c)(c-a)[c,a,a,x,b;f] +$$

$$+(x-d)(d-b)[a,x,b,b,d;f]+(b-a)(b-d)[a,a,x,b,b;f]\}.$$

Wir setzen nun voraus, *dass die Punkte* c, a, b, d *äquidistant sind*, was bei den meisten Interpolationstabellen der Fall ist, und wollen den Wert von f im Mittelpunkt des Intervalls $[a,b]$ abschätzen, ein Fall, der ebenfalls sehr häufig eintritt. Wir setzen dann $x = \mu = \frac{a+b}{2}$ $(= \frac{c+d}{2})$, und aus (7), (8) folgt

(9)
$$d(\mu) = 9(\frac{b-a}{2})^4[c,a,\mu,b,d;f], \quad d_1(\mu) = (\frac{b-a}{2})^4[a,a,\mu,b,b;f],$$

$$d(\mu) - d_1(\mu) = (\frac{b-a}{2})^4\{3[c,a,a,\mu,b;f]+3[a,\mu,b,b,d;f]+2[a,a,\mu,b,b;f]\}.$$

Setzt man nun voraus, daß f konkav oder konvex von 3-ter Ordnung ist, so resultiert $|d(\mu)| > |d_1(\mu)|$. Es folgt also in diesem Fall, *dass* $P(a,a,b,b;f|x)$ *im Mittelpunkt des Intervalls* $[a,b]$ *einen besseren Näherungswert für* f *liefert als das Polynom* $P(c,a,b,d;f|x)$. Übrigens wird f durch beide von der gleichen Seite her angenähert.

Obige Betrachtungen lassen sich z.B. auf die schon erwähnten Funktionen: $\ln x$, die auf der Menge der positiven reellen Zahlen konkav von 3-ter Ordnung ist, $\operatorname{arctg} x$, die auf dem Intervall $[0,1]$ konvex von 3-ter Ordnung ist und $\frac{1}{2}\ln(1+x^2)$, die auf dem Intervall $[-\sqrt{2}+1, \sqrt{2}-1]$ konkav von 3-ter Ordnung ist, anwenden.

Bei der praktischen Durchführung der Rechnungen benötigt man zur Berechnung des Polynoms $P(a,a,b,b;f|x)$ außer den Werten der Funktion f, die man der Tabelle entnehmen kann, auch die Werte $f'(a)$, $f'(b)$ der Ableitung in den Punkten a, b. Die Berechnung dieser Werte wird meistens dadurch erleichtert, daß die Ableitung der Funktion f eine rationale Funktion ist, wie z.B. bei $\ln x$, $\operatorname{arctg} x$, $\frac{1}{2}\ln(1+x^2)$.

10. Im folgenden geben wir Abschätzungen für den Fehler an, der sich bei den vorhin betrachteten Approximationen ergibt. Wir werden nur einen Fall behandeln, da die übrigen ähnlich sind, und zwar die Annäherung der Funktion f im Mittelpunkt μ des Intervalls $[a, b]$ durch das Polynom $P(a, a, b, b; f|x)$.

Setzt man

$$M = \sup |[x_1, x_2, x_3, x_4, x_5; f]|,$$

wobei sich das Supremum über alle Gruppen von je 5 voneinander verschiedenen Punkten x_1, x_2, x_3, x_4, x_5 des Intervalls $[a, b]$ erstreckt, so ergibt sich aus (9) die Abschätzung

$$|d_1(\mu)| \leq \left(\frac{b-a}{2}\right)^4 M.$$

Besitzt nun f eine Ableitung 4-ter Ordnung, dann ist

$$M = \frac{1}{24} \sup_{x \in [a, b]} |f^{(4)}(x)|$$

und demnach

$$|d_1(\mu)| \leq \frac{1}{24} \left(\frac{b-a}{2}\right)^4 \sup_{x \in [a, b]} |f^{(4)}(x)|. \tag{10}$$

Natürlich ist diese Abschätzung nur dann von Interesse, wenn f beschränkt ist.

Im Falle einer nicht-konkaven oder nicht-konvexen Funktion 4-ter Ordnung kann man diese Abschätzung weiter präzisieren. In der Tat,beachtet man, daß

$$\frac{d}{dx}[a, a, x, b, b; f] = [a, a, x, x, b, b; f]$$

ist, so folgt, daß die Funktion $[a, a, x, b, b; f]$ monoton ist. Demnach liegt $d_1(\mu)$ in diesem Fall stets zwischen

$$\left(\frac{b-a}{2}\right)^4 [a, a, a, b, b; f] \quad \text{und} \quad \left(\frac{b-a}{2}\right)^4 [a, a, b, b, b; f]. \tag{11}$$

Numerisches Beispiel. Sind die Werte der Funktion $f(x) = \ln x$ für die Punkte $1, 2, 3, 4$ bekannt, dann gibt das Polynom

$$P(2, 2, 3, 3; f|x) = f(2) + (x-2)[f(3)-f(2)] + \frac{1}{2}(x-2)(x-3)[f'(3)-f'(2)]$$

im Mittelpunkt des Intervalls $[2,3]$ einen besseren Näherungswert als das Polynom

$$P(1,2,3,4;f|x) = f(2) + (x-2)[f(3)-f(2)] + \frac{1}{4}(x-2)(x-3)[f(4)-f(3)-f(2)+f(1)].$$

In der Tat, in diesem Fall $(f(x) = \ln x)$ ist

$$P(2,2,3,3;f|2,5) = \frac{1}{2}(\ln 2 + \ln 3) + \frac{1}{48}$$

$$P(1,2,3,4;f|2,5) = \frac{1}{2}(\ln 2 + \ln 3) + \frac{1}{16}(\ln 3 - \ln 2).$$

Nach Formel (10) ist der Betrag des Fehlers bei der ersten Approximation $\leq \frac{1}{1024}$. Beachtet man hingegen, daß $\ln x$ konkav von 3-ter Ordnung und konvex von 4-ter Ordnung ist, dann gibt der Betrag der ersten Zahl aus (11) die bessere Schranke

$$\frac{1}{16}\left[3(\ln 3 - \ln 2) - \frac{29}{24}\right] < \frac{1}{1600} < \frac{1}{1024}$$

sogar dann, wenn man die recht groben Abschätzungen $\ln 3 < 1,099$, $\ln 2 > 0,693$, $\frac{29}{24} > 1,208$ benutzt.

Einer Logarithmentafel mit 10 Dezimalstellen entnehmen wir die Werte

$$\ln 2 = 0,693\,147\,180\,6 \qquad \ln 3 = 1,098\,612\,288\,7$$

und erhalten dann

$$\ln 2 + \ln 3 = 1,791\,759\,469\,3$$

$$\ln 3 - \ln 2 = 0,405\,465\,108\,1.$$

Vernachlässigt man während der Rechnung die 11. Dezimalstelle, so ergibt sich

$$P(2,2,3,3;f|2,5) \approx 0,895\,879\,734\,6 + 0,020\,833\,333\,3 = 0,916\,713\,067\,9$$

$$P(1,2,3,4;f|2,5) \approx 0,895\,879\,734\,6 + 0,025\,341\,569\,2 = 0,921\,221\,303\,8.$$

In der gleichen Tafel wird der Wert von $\ln 2,5$ mit $0,916\,290\,731\,9$ angegeben. Also stimmen 3 Dezimalstellen unseres Näherungswertes mit dem Wert von $\ln 2,5$ überein.