2006–04–27

Flyttalsaritmetik, Heath 1.3 Flyttalsssytem (β, t, L, U)

IEEE double precision: $(2, 53, -1022, 1023)$

Avrundningsenhet, maskintal, maskinepsilon:

$$\mu = 0.5 \cdot \beta^{1-t} = 2^{-53}$$

`eps` i MATLAB är 2μ. `eps(1)`.

$$\frac{|\mathrm{fl}(x) - x|}{|x|} \leqslant \mu$$

Dessutom gäller i IEEE standard att om $\odot$ är någon av de aritmetiska operationerna $+$, $-$, $\times$, $\div$, så gäller

$$\frac{|\mathrm{fl}(x \odot y) - x \odot y|}{|x \odot y|} \leqslant \mu$$

Det gäller även $\sqrt{\ }$ och konvertering.

Operationer på flyttal i IEEE

EXEMPEL: $(10, 4, -9, 9)$. $x = 1.234 \cdot 10^0$, $y = 4.567 \cdot 10^{-2}$.

a) Bestäm $z = \mathrm{fl}(x + y)$:

$$1.234 \cdot 10^0 + 4.567 \cdot 10^{-2} = [\text{skift}] = 1.234 \cdot 10^0 + 0.04567 \cdot 10^0 = [\text{exakt}] = 1.27967 \cdot 10^0 =$$

$$= [\text{lagring till flyttal}] = 1.280 \cdot 10^0$$

b) Bestäm $w = \mathrm{fl}(x \cdot y)$:

$$1.234 \cdot 10^0 \times 4.567 \cdot 10^{-2} = 5.635678 \cdot 10^{-2} = [\text{lagring}] = 5.636 \cdot 10^{-2}$$

Enda felet sker vid lagring (avrundning) till flyttalssystemet.

Kancellation

Noggrannhetsförlust vid subtraktion av två nästan lika stora tal.

EXEMPEL:

$$\frac{1}{1-x} - \frac{1}{1+x}, \quad x \notin \{1, -1\}$$

ger kancellation om x ligger nära 0. $x = 0.001$ och tre siffrors precision.

$$\frac{1}{0.999} - \frac{1}{1.00} = 1.00 - 1.00 = 0.00$$

Bättre formel utan kancellation:

$$\frac{1}{1-x} - \frac{1}{1+x} = \frac{2x}{1-x^2}$$

$x = 0.001$, tre siffror ger

$$\frac{2 \cdot 0.001}{1.00} = 0.00200$$

Utskiftning, noggranhet försvinner vid skift.

I exemplet ovan räknade vi ut $1 + x$ då $x = 0.001$ med tre siffror. I datoraritmetiken blir det $x = 1.00 \cdot 10^{-3}$. Summation, skifta till största exponent: $(1.00 + 0.001) \cdot 10^0 = 1.001 \cdot 10^0 =$ [lagring] $= 1.00 \cdot 10^0$. Total utskiftning.

Tips: Att summera tal i växande storleksordning motverkar utskiftning.

Icke-linjära ekvationer, Heath 5

Problem: $f(x) = 0, \quad f: \mathbb{R} \to \mathbb{R}$.

DEFINITION: En lösning (rot) till ekvationen $f(x) = 0$ betecknas x^*. Den är en **enkelrot** om $f'(x^*) \neq 0$. Annars är den en **multipelrot** med **multiplicitet** m om $f'(x^*) = \cdots = f^{(m-1)}(x^*) = 0$ men $f^{(m)}(x^*) \neq 0$.

|||.

Iterationsmetod: x_0 start. $x = \text{formel}(x_0)$, $x_2 = \text{formel}(x_0, x_1)$... Om $x_k \to x^*$ då $k \to \infty$ så konvergens. Snabbhet: konvergensordning.

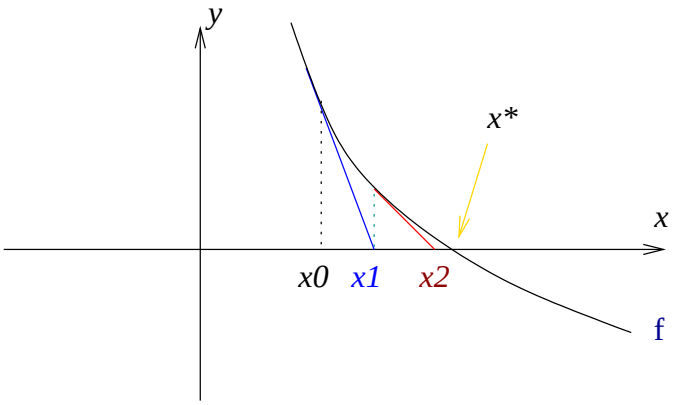
$$\lim_{k \to \infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|^q} = C < \infty$$

Största möjliga q är **konvergensordningen** och C kallas **asymptotisk felkonstant**.

Speciellt: $q = 1$: linjär konvergens. $q > 1$: superlinjär konvergens. $q = 2$: kvadratisk konvergens.

För Newtons metod gäller $q = 2$ vid enkelrot, vid multipelrot är $q = 1$.

Newtons metod, geometriskt



...analytiskt, linjärisering

x_0 är en approximation. Linjärisera f kring x_0: Taylorutveckling:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$

Linjär modell: I stället för $f(x) = 0$ löser vi $f(x_0) + f'(x_0)\ (x - x_0) = 0$. En linjär ekvation i x med lösning

$$x = x_0 - \frac{f(x_0)}{f'(x_0)}$$

Kalla denna lösning för x_1 och vi har gjort *en* iteration.

Newtons metod konvergerar om x_0 är tillräckligt nära x^*, men kan divergera.

Konvergensordning för Newtons metod. Taylorutveckling kring x_k:

$$0 = f(x^*) = f(x_k) + f'(x_k)(x^* - x_k) + \frac{1}{2} f''(\xi_k)\,(x^* - x_k)^2, \quad \xi_k \in [x^*, x_k]$$

Dividerar med $f'(x_k)$. Vi antar enkelrot.

$$0 = \frac{f(x_k)}{f'(x_k)} - x_k + x^* + \frac{1}{2}\frac{f''(\xi_k)}{f'(x_k)}\,(x^* - x_k)^2$$

Enligt Newtons metod:

$$0 = -x_{k+1} + x^* + \frac{1}{2}\frac{f''(\xi_k)}{f'(x_k)}\,(x^* - x_k)^2$$

dvs

$$x_{k+1} - x^* = \frac{1}{2} \cdot \frac{f''(\xi_k)}{f'(x_k)}\,(x^* - x_k)^2$$

$$\frac{|x_{k+1} - x^*|}{|x_k - x^*|^2} = \frac{1}{2} \cdot \frac{|f''(\xi_k)|}{|f'(x_k)|} \longrightarrow \frac{1}{2} \cdot \frac{|f''(x^*)|}{|f'(x^*)|} \quad \text{då } k \to \infty$$

$q = 2$: kvadratisk konvergens.

Övning: Visa att $q = 1$ och $C = \frac{1}{2}$ vid dubbelrot.

Anmärkning: För multiplicitet m gäller $q = 1, C = \frac{m-1}{m}$.

En variant av Newton som inte kräver derivator är **sekantmetoden**. Analytiskt är det att derivatan approximeras med en differenskvot.

$$f'(x_k) \approx \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}$$

(Bakåtdifferens.) Insatt i Newtons metod blir alltså sekantmetoden:

$$x_{k+1} = x_k - \frac{f(x_k)(x_k - x_{k-1})}{f(x_k) - f(x_{k-1})}$$

Vi måste nu ha två startvärden. Konvergensen är superlinjär med $q = 1.618$.

Låt $\delta x = \hat{x} - x^*$ vara felet i en approximation $\hat{x}$ till x^*. Taylorutveckling kring lösningen:

$$f(\hat{x}) = f(x^* + \delta x) = f(x^*) + f'(x^*)\,\delta x + \mathcal{O}(\delta x^2) = f'(x^*)\,\delta x + \mathcal{O}(\delta x^2)$$

och för små δx: $f(\hat{x}) \approx f'(x^*) \cdot \delta x$. Om enkelrot, dvs $f'(x^*) \neq 0$:

$$\delta x \approx \frac{f(\hat{x})}{f'(x^*)}$$

$$|\delta x| \lesssim \frac{|f(\hat{x})|}{|f'(\hat{x})|}$$

Detta är en metodoberoende felformel.

ANMÄRKNING: Vid multipelrot tar vi med fler termer i Taylorutvekclingen och får

$$(\delta x)^m \approx m! \frac{f(\hat{x})}{f^{(m)}(\hat{x})}$$

EXEMPEL: En teknolog gissar att en rot till ekvationen $f(x) \equiv x - \mathrm{e}^{-x} = 0$ är nära 0.55. Hur bra är gissningen?

Lösning:

$$f(0.55) = -0.0269...$$

$$f'(0.55) = 1.5769...$$

$$|x^* - 0.55| \lesssim \frac{0.0269...}{1.5769...} \leqslant \frac{0.27}{1.57} \leqslant 0.0172$$