Reliable function evaluation using continued fraction expansions

Outline

Reliable function evaluation using continued fraction expansions

- Reliable evaluation
- Steps of function evaluation
- Continued fractions vs power series
- Continued fractions: truncation error
- The ‘real life example’
- Things I did not tell you

Reliable evaluation

- Almost all computer calculations are erroneous
- ‘If I ask for t correct digits, I should get t correct digits.’
  - Correct: error at most 1 ULP
  - Example: for $t = 5$, the calculated result for

$$\sqrt{2} = 1.414\underline{2}135623731\ldots$$

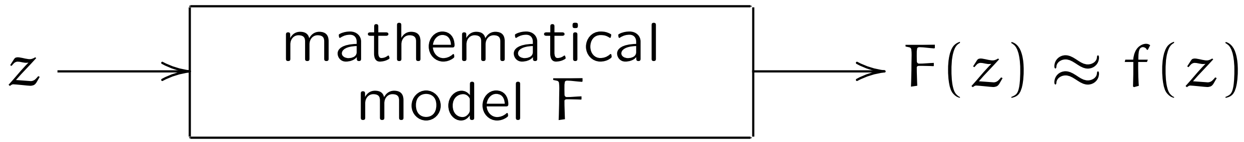
should be in

$$[1.4141135623731, 1.4143135623731]$$

Correct results:

$$y = 1.4142$$
$$y = 1.4143$$

Mathematical models

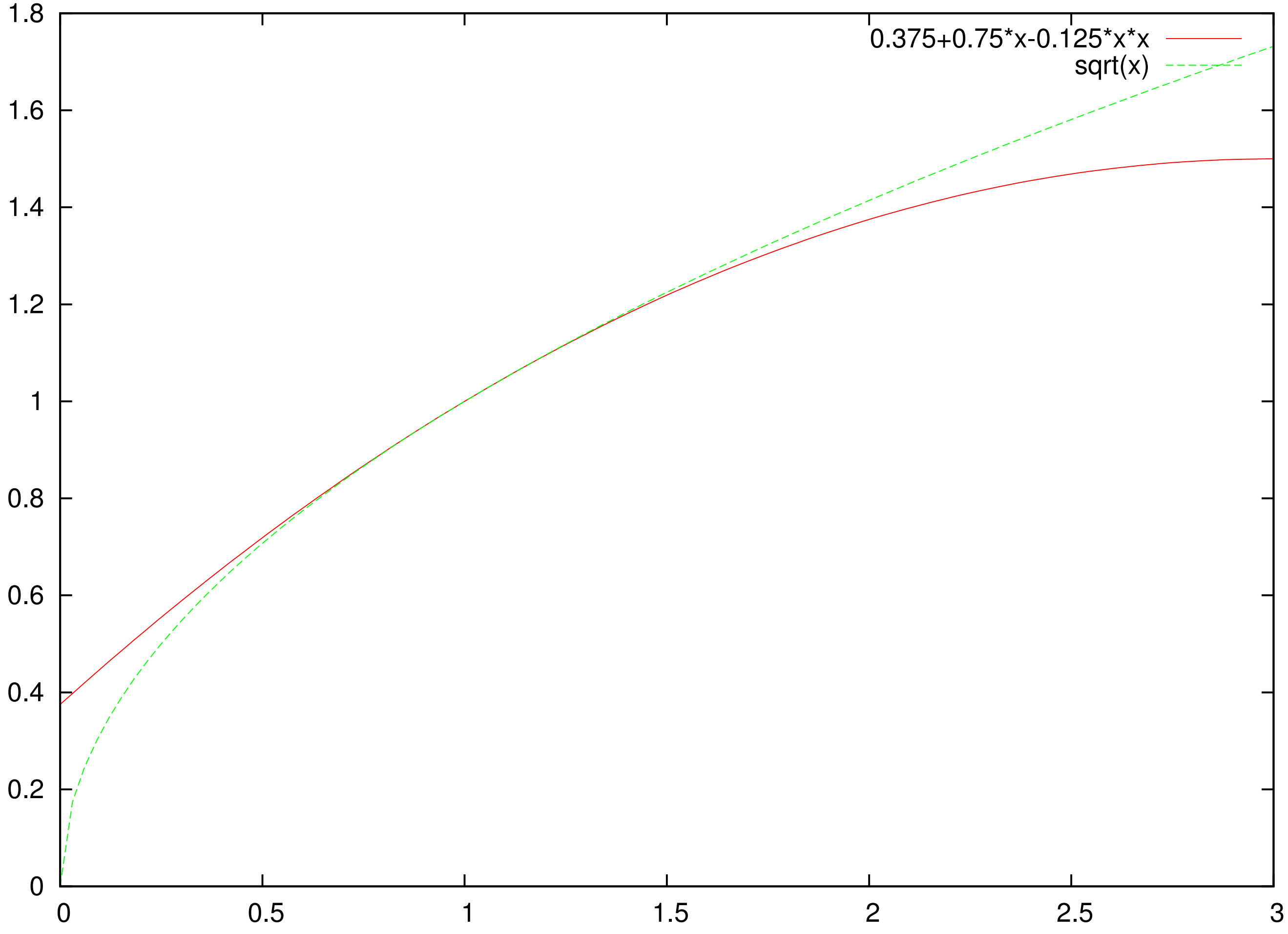


- example:

$$f(z) = \sqrt{z}$$
$$F(z) = \frac{3}{8} + \frac{3}{4}z - \frac{1}{8}z^2$$

- only an approximation of $f(z)$ (approximation error)
- only useful on a restricted domain, e.g. $[1, 1.25]$

The function $f(z) = \sqrt{z}$ and its model F



Three steps of function evaluation

$$y = \sqrt{1.25 \cdot 10^6} = 1.1180339887\ldots$$

1. Argument reduction

$$\sqrt{1.25 \cdot 10^6} = \sqrt{1.25} \cdot 10^3$$

2. Mathematical model

$$\sqrt{1.25} \approx \frac{3}{8} + \frac{3}{4}1.25 - \frac{1}{8}1.25^2$$
$$= \frac{143}{128} = 1.1171875$$

3. 'Back reduction'

$$\sqrt{1.25 \cdot 10^6} \approx \underline{1.117}1875 \cdot 10^3$$

Mathematical models

- Polynomial approximation.

$$\sqrt{z} \approx \frac{3}{8} + \frac{3}{4}z - \frac{1}{8}z^2 \qquad z \in [0.5, 1.25]$$

- Power series expansion.

$$\sqrt{z+1} = 1 + \frac{1}{2}z - \frac{1}{8}z^2 + \frac{1}{16}z^3 - \cdots \qquad |z| < 1$$

- Continued fraction expansion.

$$\sqrt{z+1} = 1 + \cfrac{z}{2 + \cfrac{z}{2 + \cfrac{z}{2 + \cdots}}} \qquad z > -1$$

Power series

A power series is an expression of the form

$$S(z) = b_0 + b_1 z + b_2 z^2 + \dots$$
$$= \sum_{k=0}^{\infty} b_k z^k$$

Example:

$$S(z) = 1 + \frac{1}{2}z - \frac{1}{8}z^2 + \frac{1}{16}z^3 - \cdots$$

Convergents of power series

The convergents $S_n(z)$ of a power series

$$S(z) = b_0 + b_1 z + b_2 z^2 + \cdots \qquad S(z) = \sum_{k=0}^{\infty} b_k z^k$$

are defined as

$$S_n(z) = b_0 + b_1 z + \cdots + b_n z^n \qquad S_n(z) = \sum_{k=0}^{n} b_k z^k$$

Examples

$$S(z) = 1 + \frac{1}{2}z - \frac{1}{8}z^2 + \frac{1}{16}z^3 - \cdots$$

$$S_0(0.25) = 1 \qquad S_0(0.25) = \underline{1}$$

$$S_1(0.25) = 1 + \frac{1}{2}0.25 \qquad S_1(0.25) = \underline{1.12}5$$

$$S_2(0.25) = 1 + \frac{1}{2}0.25 - \frac{1}{8}0.25^2 \qquad S_2(0.25) = \underline{1.117}1875$$

$$\lim_{n\to\infty} S_n(0.25) = \sqrt{1.25} \qquad S(0.25) = 1.11803398874\ldots$$

for $|z| < 1$, it is the case that $S(z) = \sqrt{1+z}$

Continued fractions (CF)

A continued fraction is an expression

$$b_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \cdots}}}$$

with all $a_j \neq 0$. The a_j and b_j are called the partial numerators and the partial denominators respectively. The continued fraction above is denoted as

$$b_0 + \mathop{\mathrm{K}}_{j=1}^{\infty} \frac{a_j}{b_j} \text{ or } b_0 + \frac{a_1}{b_1} + \frac{a_2}{b_2} + \frac{a_3}{b_3} + \cdots$$

Convergents of a continued fraction

The convergents F_n of a continued fraction

$$F = b_0 + \mathop{\mathrm{K}}_{j=1}^{\infty} \frac{a_j}{b_j} \qquad F = b_0 + \cfrac{a_1}{b_1} + \cfrac{a_2}{b_2} + \cfrac{a_3}{b_3} + \cdots$$

are defined as follows

$$F_n = b_0 + \mathop{\mathrm{K}}_{j=1}^{n} \frac{a_j}{b_j} \qquad F_n = b_0 + \cfrac{a_1}{b_1} + \cdots + \cfrac{a_n}{b_n}$$

Examples

$$F(z) = 1 + \cfrac{z}{2 + \cfrac{z}{2 + \cfrac{z}{2 \cdots}}}$$

$$F_0(0.25) = 1 \qquad F_0(0.25) = \underline{1}$$

$$F_1(0.25) = 1 + \frac{0.25}{2} \qquad F_1(0.25) = \underline{1.1}25$$

$$F_2(0.25) = 1 + \cfrac{0.25}{2 + \cfrac{0.25}{2}} \qquad F_2(0.25) = \underline{1.117}6470588235\ldots$$

$$\lim_{n\to\infty} F_n(0.25) = \sqrt{1.25} \qquad F(0.25) = 1.1180339887\ldots$$

for $z > -1$, it is the case that $S(z) = \sqrt{1+z}$

Tails of a continued fraction

The tails $F^{(m)}$ of a continued fraction

$$F = b_0 + \mathop{\mathrm{K}}_{j=1}^{\infty} \frac{a_j}{b_j} \qquad F = b_0 + \frac{a_1}{b_1} + \frac{a_2}{b_2} + \frac{a_3}{b_3} + \cdots$$

are defined as follows

$$F^{(m)} = \mathop{\mathrm{K}}_{j=m+1}^{\infty} \frac{a_j}{b_j} \qquad F^{(m)} = \frac{a_{m+1}}{b_{m+1}} + \frac{a_{m+2}}{b_{m+2}} + \cdots$$

Example

$$F(z) = 1 + \cfrac{z}{2 + \cfrac{z}{2 + \cfrac{z}{2 \cdots}}}$$

$$F(z) = \sqrt{z+1} \qquad\qquad z > -1$$

$$F^{(m)}(z) = \sqrt{z+1} - 1$$

Modified convergents

$$F(z) = b_0 + \mathop{\mathrm{K}}_{j=1}^{\infty} \frac{a_j(z)}{b_j}$$

$$F_n(z;w) = b_0 + \mathop{\mathrm{K}}_{j=1}^{n} \frac{a_j(z)}{b_j} + \frac{w}{1}$$

$$F(z) = b_0 + \frac{a_1(z)}{b_1} + \frac{a_2(z)}{b_2} + \frac{a_3(z)}{b_3} + \cdots$$

$$F_n(z;w) = b_0 + \frac{a_1(z)}{b_1} + \cdots + \frac{a_n(z)}{b_n} + \frac{w}{1}$$

Note that $F_n(z;0) = F_n(z)$ and $F_n(z;F^{(n)}(z)) = F(z)$.

Examples

$$F(z) = 1 + \frac{z}{2+} \frac{z}{2+} \frac{z}{2} \cdots$$

$$F_0(0.25) = \underline{1} \qquad F_0(0.25; 0.12) = \underline{1.12}$$

$$F_1(0.25) = \underline{1.1}25 \qquad F_1(0.25; 0.12) = \underline{1.117}92452830\ldots$$

$$F_2(0.25) = \underline{1.117}6470588235\ldots \qquad F_2(0.25; 0.12) = \underline{1.11804}0089\ldots$$

$$F(0.25) = 1.1180339887\ldots$$

CF vs PS. What's the difference?

$$F(z) = b_0 + \frac{a_1}{b_1} + \frac{a_2}{b_2} + \frac{a_3}{b_3} + \cdots$$
$$= b_0 + \underbrace{a_1/(b_1 + \underbrace{a_2/(b_2 + a_3/(b_3 + \cdots)))}_{F^{(1)}(z)}}_{F^{(0)}(z)}$$

$$S(z) = b_0 + b_1 z + b_2 z^2 + b_3 z^3 + \cdots$$
$$= b_0 + \underbrace{z \cdot (b_1 + \underbrace{z \cdot (b_2 + z \cdot (b_3 + \cdots)))}_{S^{(1)}(z)}}_{S^{(0)}(z)}$$

Tails of power series?

The tails $S^{(m)}(z)$ of a power series

$$S(z) = b_0 + b_1 z + b_2 z^2 + \cdots \qquad S(z) = \sum_{k=0}^{\infty} b_k z^k$$

are defined as

$$S^{(m)}(z) = b_{m+1} z + b_{m+2} z^2 + \cdots \quad S^{(m)}(z) = \sum_{k=m+1}^{\infty} b_k z^{k-m}$$

Note that

$$S(z) = S_n(z) + S^{(n)}(z) x^n$$

Example

$$S(z) = 1 + \frac{1}{2}z - \frac{1}{8}z^2 + \frac{1}{16}z^3 - \cdots$$
$$= \sqrt{1+z}$$

$$S^{(0)}(0.25) = 0.1180339887\ldots$$
$$S^{(1)}(0.25) = -0.02786404500042\ldots$$
$$S^{(2)}(0.25) = 0.013543819998\ldots$$
$$\lim_{m\to\infty} S^{(m)}(0.25) = 0$$

Another example

$$S(z) = 1 + z + z^2 + \cdots$$
$$= \frac{1}{1-z}$$

Since all $b_n = 1$, it follows that for all n

$$S^{(n)}(z) = z + z^2 + z^3 + \cdots$$
$$= \frac{z}{1-z}$$

Forward and backward evaluation

- Forward evaluation
  1. Calculate convergents $C_0(z), C_1(z), \ldots, C_n(z)$
  2. Use the difference between $C_{n-1}(z)$ and $C_n(z)$ to determine a stop criterium

  Numerically less stable

- Backward evaluation
  1. Find an n in advance such that $C_n(z)$ is reliable
  2. Calculate $C_1^{(n-1)}(z), C_2^{(n-2)}(z), \ldots, C_n^{(0)}(z)$
  3. $C_n(z) = b_0 + C_n^{(0)}(z)$

  Step (1) is not trivial

Truncation error of a CF's convergent

- A posteriori (forward evaluation)
  - Henrici-Pflüger
- A priori (backward evaluation)
  - Gragg-Warner (unmodified convergent)
  - Interval Sequence Theorem (modified convergent)

Henrici-Pfluger (HP)

Suppose $F(z) = b_0 + K_{j=1}^{\infty} \frac{a_j(z)}{1}$ is a converging continued fraction with $a_n(z) > 0$ for all $n \geqslant 1$. The following holds for $n \geqslant 1$:

$$|F(z) - F_n(z; 0)| \leqslant |F_n(z) - F_{n-1}(z)|$$

Example

$$F(z) = 1 + \frac{z}{2 +} \frac{z}{2 +} \frac{z}{2 +} \cdots$$
$$= 1 + \frac{\frac{1}{2}z}{1 +} \frac{\frac{1}{4}z}{1 +} \frac{\frac{1}{4}z}{1 +} \cdots$$

n	$F_n(0.25)$	t
0	1	–
1	1.125	1
2	1.117647...	3
3	1.1180555...	4
4	1.118032786885...	5

Gragg-Warner (GW)

Suppose $F(z) = b_0 + \mathrm{K}_{j=1}^{\infty} \frac{a_j(z)}{1}$ is a converging continued fraction with $a_n(z) > 0$ for all $n \geqslant 1$. The following holds for $n \geqslant 2$:

$$|F(z) - F_n(z;0)| \leqslant 2|a_1(z)| \prod_{k=2}^{n} \frac{\sqrt{1+4|a_k(z)|}-1}{\sqrt{1+4|a_k(z)|}+1}$$

Example

$$F(z) = 1 + \frac{\frac{1}{2}z}{1} + \frac{\frac{1}{4}z}{1} + \frac{\frac{1}{4}z}{1} + \cdots$$

n	$F_n(0.25)$	t (HP)	t (GW)
0	1	–	–
1	1.125	1	–
2	1.117647...	3	2
3	1.1180555...	4	4
4	1.118032786885...	5	5
5	1.1180340557...	6	6
6	1.1180339850...	8	7

Interval Sequence Theorem (IST)

Suppose $F(z) = \mathrm{K}_{j=1}^{\infty} \frac{a_j(z)}{1}$. If we can find sequences $(\ell_n)_n$ and $(r_n)_n$ such that for all n

1. $0 < \ell_n < r_n < \infty$
2. $(1 + r_n)\ell_{n-1} \leqslant a_n(z) \leqslant (1 + \ell_n)r_{n-1}$

then we can apply the ‘interval sequence theorem’:

$$|F(z) - F_n(z; w)| \leqslant (r_n - \ell_n) \frac{r_0}{1 + \ell_n} \prod_{k=1}^{n-1} \frac{r_k}{1 + r_k}$$

for $w \in [\ell_n, r_n]$.

Example

$$F(z) = 1 + \frac{\frac{1}{2}z}{1} + \frac{\frac{1}{4}z}{1} + \frac{\frac{1}{4}z}{1} + \cdots$$

$$\ell_0 = \sqrt{2} - 1 - 2^{-50}, r_0 = \sqrt{2} - 1 + 2^{-50}, \ell_n = \frac{\ell_0}{2}, r_n = \frac{r_0}{2}, n > 1$$

n	$F_n(0.25)$	t (IST)
1	1.11803398874989484517737278$4\ldots$	18
2	1.1180339887498948483732876$91\ldots$	19
3	1.11803398874989484819518545$78\ldots$	20
4	1.11803398874989484820511075$51\ldots$	21
5	1.11803398874989484820455763726...	23
6	1.11803398874989484820458846146...	24

Sufficient conditions for the IST

In general, we can find suitable ℓ_n and r_n if

- the partial numerators are non-decreasing towards a positive number.
- the partial numerators are non-increasing towards zero.
- the even partial numerators are non-decreasing towards a positive number a, and the odd partial numerators are non-increasing towards a positive number b such that $a \leqslant b$.
- the partial numerators are non-decreasing towards zero.
- the partial numerators are non-decreasing towards infinity.

Real life example: $\log(z+1)$

$$F(z) = \log(z+1) = \mathop{\mathrm{K}}_{n=1}^{\infty} \frac{a_n(z)}{1}$$

$$a_1(z) = z \tag{1}$$

$$a_n(z) = \frac{nz}{4(n-1)} \qquad n \text{ even} \tag{2}$$

$$a_n(z) = \frac{(n-1)z}{4n} \qquad n > 1,\ n \text{ odd} \tag{3}$$

$F'(z) = F^{(1)}(z)$ has even partial numerators increasing towards $\frac{1}{4}$ and odd partial numerators decreasing towards $\frac{1}{4}$.

Steps for calculating $\log(z+1)$

1. From target precision t, calculate a working precision s such that the rounding errors of the following steps will be small enough.
2. Reduce the argument to $[0, \beta^{0.5^n} - 1]$, using
$$\log(a\beta^n) = \log(a) + n\log(\beta)$$
3. Use the IST ‘the other way round’, to calculate a convergent n which ensures that the truncation error will be small enough.
4. Evaluate the continued fraction $F'(z)$.
5. Calculate $\log(z+1)$ from $F'(z)$.

Number of convergents to calculate

$$z = \sqrt{2} - 1$$

t	unm	hp	gw	mod	ist	smod	sist	ps
50	11	12	13	10	11	3	3	37
100	23	24	24	21	22	11	12	75
150	34	35	35	32	33	22	23	114
200	45	46	46	43	44	33	34	153

I did not tell you about

- Rounding error
- Multi-base arithmetic
- The C++ implementation

A final word of advice. . .

Als uw ketting gebroken is, dan moet je te voet verder.