Rounding Error Analysis of Linear Recurrences
Using Generating Series

Marc Mezzarobba

CNRS, École polytechnique

Diffferential Seminar @ RTCA6

November 30, 2023

doi:10.1553/etna_vol58s196

A Toy Example

A toy example

[Boldo 2009]

$$c_{n+1} = 2\,c_n - c_{n-1} \qquad (c_0 = \diamond(1/3),\, c_{-1} = 0)$$

$n =$	Floating-point arithmetic	Interval arithmetic
0	0.333333333333333	$[0.333333333333333 \pm 1.49e-17]$
5	2.00000000000000	$[2.00000000000000 \pm 3.78e-15]$
10	3.66666666666667	$[3.666666666667 \pm 5.74e-13]$
15	5.33333333333334	$[5.3333333333 \pm 5.29e-11]$
20	7.00000000000001	$[7.00000000 \pm 1.60e-9]$
25	8.66666666666668	$[8.666667 \pm 4.65e-7]$
30	10.3333333333333	$[10.3333 \pm 4.41e-5]$
35	12.0000000000000	$[12.000 \pm 8.82e-4]$
40	13.6666666666667	$[1.4e+1 \pm 0.406]$
45	15.3333333333334	$[\pm 21.3]$
50	17.000000000000	$[\pm 5.04e+2]$

Naïve error analysis

Model: $\diamond(x\,op\,y) = x\ op\ y + \varepsilon_{\mathrm{op}}$ with $\varepsilon_{\mathrm{op}} \in [-\mathbf{u}, \mathbf{u}]$ (∼ fixed-point arithmetic)

(multiplication by 2 is exact)

Exact rec.:

$$c_{n+1} = 2\,c_n - c_{n-1}$$

Naïve error analysis

Model: $\diamond(x\,op\,y) = x \ op \ y + \varepsilon_{\text{op}}$ with $\varepsilon_{\text{op}} \in [-\mathbf{u}, \mathbf{u}]$ ($\sim$ fixed-point arithmetic)

(multiplication by 2 is exact)

Exact rec.: $$c_{n+1} = 2\,c_n - c_{n-1}$$

Approx. rec.: $$\tilde{c}_{n+1} = \diamond\big(2\,\tilde{c}_n - \tilde{c}_{n-1}\big)$$

Naïve error analysis

Model: $\diamond(x \mathit{op}\, y) = x \mathit{op}\, y + \varepsilon_{op}$ with $\varepsilon_{op} \in [-\mathbf{u}, \mathbf{u}]$ (∼ fixed-point arithmetic)

(multiplication by 2 is exact)

Exact rec.:

$$c_{n+1} = 2\, c_n - c_{n-1}$$

Approx. rec.:

$$\begin{aligned}\tilde{c}_{n+1} &= \diamond\left(2\, \tilde{c}_n - \tilde{c}_{n-1}\right) \\ &= 2\, \tilde{c}_n - \tilde{c}_{n-1} + \varepsilon_n \quad \text{with } |\varepsilon_n| \leqslant \mathbf{u}\end{aligned}$$

Naïve error analysis

Model: $\diamond(x \, op \, y) = x \; op \; y + \varepsilon_{op}$ with $\varepsilon_{op} \in [-\mathbf{u}, \mathbf{u}]$ ($\sim$ fixed-point arithmetic)

(multiplication by 2 is exact)

Exact rec.:
$$c_{n+1} = 2\,c_n - c_{n-1}$$

Approx. rec.:
$$\begin{aligned} \tilde{c}_{n+1} &= \diamond\big(2\,\tilde{c}_n - \tilde{c}_{n-1}\big) \\ &= 2\,\tilde{c}_n - \tilde{c}_{n-1} + \varepsilon_n \quad \text{with } |\varepsilon_n| \leqslant \mathbf{u} \end{aligned}$$

$$|\tilde{c}_{n+1} - c_{n+1}| \leqslant 2\,|\tilde{c}_n - c_n| + |\tilde{c}_{n-1} - c_{n-1}| + \mathbf{u}$$

Naïve error analysis

Model: $\diamond(x\, op\, y) = x\ op\ y + \varepsilon_{op}$ with $\varepsilon_{op} \in [-\mathbf{u}, \mathbf{u}]$ (∼ fixed-point arithmetic)

(multiplication by 2 is exact)

Exact rec.: $$c_{n+1} = 2\, c_n - c_{n-1}$$

Approx. rec.: $$\begin{aligned} \tilde{c}_{n+1} &= \diamond\big(2\, \tilde{c}_n - \tilde{c}_{n-1}\big) \\ &= 2\, \tilde{c}_n - \tilde{c}_{n-1} + \varepsilon_n \quad \text{with } |\varepsilon_n| \leqslant \mathbf{u} \end{aligned}$$

$$|\tilde{c}_{n+1} - c_{n+1}| \leqslant 2\, |\tilde{c}_n - c_n| + |\tilde{c}_{n-1} - c_{n-1}| + \mathbf{u}$$

Induction: $$|\tilde{c}_n - c_n| \leqslant 3^n\, \mathbf{u}$$

Naïve error analysis

Model: $\diamond(x \,op\, y) = x \;op\; y + \varepsilon_{\mathrm{op}}$ with $\varepsilon_{\mathrm{op}} \in [-\mathbf{u}, \mathbf{u}]$ $\quad$ ($\sim$ fixed-point arithmetic)

(multiplication by 2 is exact)

Exact rec.:

$$c_{n+1} = 2\,c_n - c_{n-1}$$

Approx. rec.:

$$\begin{aligned} \tilde{c}_{n+1} &= \diamond\big(2\,\tilde{c}_n - \tilde{c}_{n-1}\big) \\ &= 2\,\tilde{c}_n - \tilde{c}_{n-1} + \varepsilon_n \quad \text{with } |\varepsilon_n| \leqslant \mathbf{u} \end{aligned}$$

$$|\tilde{c}_{n+1} - c_{n+1}| \leqslant 2\,|\tilde{c}_n - c_n| + |\tilde{c}_{n-1} - c_{n-1}| + \mathbf{u}$$

Induction:

$$|\tilde{c}_n - c_n| \leqslant 3^n\,\mathbf{u}$$

Naïve error analysis

Model: $\diamond(x\ op\ y) = x\ op\ y + \varepsilon_{op}$ with $\varepsilon_{op} \in [-\mathbf{u}, \mathbf{u}]$ (∼ fixed-point arithmetic)

(multiplication by 2 is exact)

Exact rec.: $$c_{n+1} = 2\,c_n - c_{n-1}$$

Approx. rec.: $$\begin{aligned} \tilde{c}_{n+1} &= \diamond\big(2\,\tilde{c}_n - \tilde{c}_{n-1}\big) \\ &= 2\,\tilde{c}_n - \tilde{c}_{n-1} + \varepsilon_n \quad \text{with } |\varepsilon_n| \leqslant \mathbf{u} \end{aligned}$$

$$|\tilde{c}_{n+1} - c_{n+1}| \leqslant 2\,|\tilde{c}_n - c_n| + |\tilde{c}_{n-1} - c_{n-1}| + \mathbf{u}$$

Induction: $$|\tilde{c}_n - c_n| \leqslant 3^n\,\mathbf{u}$$

Slightly better: $$\begin{aligned} |\tilde{c}_n - c_n| &\leqslant (\lambda_+\,\alpha_+^n + \lambda_-\,\alpha_-^n - 4)\,\mathbf{u} \approx 2.4^n\,\mathbf{u} \\ &\approx 2.4^n\,\mathbf{u} \end{aligned}$$

$$\alpha_\pm = 1 \pm \sqrt{2}$$
$$\lambda_\pm = 4 \pm 3\sqrt{2}$$

Naïve error analysis

Model: $\diamond(x \; op \; y) = x \; op \; y + \varepsilon_{op}$ with $\varepsilon_{op} \in [-\mathbf{u}, \mathbf{u}]$ ($\sim$ fixed-point arithmetic)

(multiplication by 2 is exact)

Exact rec.: $$c_{n+1} = 2\,c_n - c_{n-1}$$

Approx. rec.: $$\begin{aligned} \tilde{c}_{n+1} &= \diamond(2\,\tilde{c}_n - \tilde{c}_{n-1}) \\ &= 2\,\tilde{c}_n - \tilde{c}_{n-1} + \varepsilon_n \quad \text{with } |\varepsilon_n| \leqslant \mathbf{u} \end{aligned}$$

$$|\tilde{c}_{n+1} - c_{n+1}| \leqslant 2\,|\tilde{c}_n - c_n| + |\tilde{c}_{n-1} - c_{n-1}| + \mathbf{u}$$

Induction: $$|\tilde{c}_n - c_n| \leqslant 3^n\,\mathbf{u}$$

Slightly better: $$\begin{aligned} |\tilde{c}_n - c_n| &\leqslant (\lambda_+\,\alpha_+^n + \lambda_-\,\alpha_-^n - 4)\,\mathbf{u} \approx 2.4^n\,\mathbf{u} \\ &\approx 2.4^n\,\mathbf{u} \end{aligned}$$

$$\alpha_\pm = 1 \pm \sqrt{2}$$
$$\lambda_\pm = 4 \pm 3\sqrt{2}$$

This is what interval evaluation amounts to!

Errors cancel out

Exact rec.: $$c_{n+1} = 2\,c_n - c_{n-1}$$

Approx. rec.: $$\begin{aligned}\tilde{c}_{n+1} &= \diamond\big(2\,\tilde{c}_n - \tilde{c}_{n-1}\big)\\ &= 2\,\tilde{c}_n - \tilde{c}_{n-1} + \varepsilon_n \quad \text{with } |\varepsilon_n| \leqslant \mathbf{u}\end{aligned}$$

💡 The error satisfies “the same” recurrence as the computed sequence

Errors cancel out

Exact rec.: $$c_{n+1} = 2\,c_n - c_{n-1}$$

Approx. rec.: $$\begin{aligned}\tilde{c}_{n+1} &= \diamond\big(2\,\tilde{c}_n - \tilde{c}_{n-1}\big)\\ &= 2\,\tilde{c}_n - \tilde{c}_{n-1} + \varepsilon_n \quad \text{with } |\varepsilon_n| \leqslant \mathbf{u}\end{aligned}$$

Let $\delta_n = \tilde{c}_n - c_n$:

💡 The error satisfies "the same" recurrence as the computed sequence

Errors cancel out

Exact rec.: $$c_{n+1} = 2\,c_n - c_{n-1}$$

Approx. rec.: $$\begin{aligned}\tilde{c}_{n+1} &= \diamond\big(2\,\tilde{c}_n - \tilde{c}_{n-1}\big) \\ &= 2\,\tilde{c}_n - \tilde{c}_{n-1} + \varepsilon_n \quad \text{with } |\varepsilon_n| \leqslant \mathbf{u}\end{aligned}$$

Let $\delta_n = \tilde{c}_n - c_n$: $$\delta_{n+1} = 2\,\delta_n - \delta_{n-1} + \varepsilon_n \qquad (\delta_0 = \delta_1 = 0)$$

💡 The error satisfies "the same" recurrence as the computed sequence

Errors cancel out

Exact rec.: $c_{n+1} = 2\,c_n - c_{n-1}$

Approx. rec.: $\tilde{c}_{n+1} = \diamond\big(2\,\tilde{c}_n - \tilde{c}_{n-1}\big)$

$= 2\,\tilde{c}_n - \tilde{c}_{n-1} + \varepsilon_n$ with $|\varepsilon_n| \leqslant \mathbf{u}$

Let $\delta_n = \tilde{c}_n - c_n$: $\delta_{n+1} = 2\,\delta_n - \delta_{n-1} + \varepsilon_n$ $(\delta_0 = \delta_1 = 0)$

↰ **"local error"**

💡 The error satisfies "the same" recurrence as the computed sequence

Errors cancel out

Exact rec.: $$c_{n+1} = 2\,c_n - c_{n-1}$$

Approx. rec.: $$\tilde{c}_{n+1} = \diamond\big(2\,\tilde{c}_n - \tilde{c}_{n-1}\big) = 2\,\tilde{c}_n - \tilde{c}_{n-1} + \varepsilon_n \quad \text{with } |\varepsilon_n| \leqslant \mathbf{u}$$

Let $\delta_n = \tilde{c}_n - c_n$: $$\delta_{n+1} = 2\,\delta_n - \delta_{n-1} + \varepsilon_n \qquad (\delta_0 = \delta_1 = 0)$$

"global error" ↗ ↖ **"local error"**

💡 The error satisfies "the same" recurrence as the computed sequence

Errors cancel out

Exact rec.: $$c_{n+1} = 2\,c_n - c_{n-1}$$

Approx. rec.: $$\begin{aligned} \tilde{c}_{n+1} &= \diamond\big(2\,\tilde{c}_n - \tilde{c}_{n-1}\big) \\ &= 2\,\tilde{c}_n - \tilde{c}_{n-1} + \varepsilon_n \quad \text{with } |\varepsilon_n| \leqslant \mathbf{u} \end{aligned}$$

Let $\delta_n = \tilde{c}_n - c_n$: $$\delta_{n+1} = 2\,\delta_n - \delta_{n-1} + \varepsilon_n \qquad (\delta_0 = \delta_1 = 0)$$

"global error" ⤴ ⤶ **"local error"**

Then $$\begin{aligned} \delta_n &= \sum_{k=1}^{n-1} k\,\varepsilon_{n-k} \\ |\delta_n| &\leqslant \frac{n\,(n-1)}{2}\,\mathbf{u} \end{aligned}$$

Calculations can become unwieldy (nested sums, determinants…)

The error satisfies "the same" recurrence as the computed sequence

Errors cancel out

Exact rec.: $$c_{n+1} = 2\,c_n - c_{n-1}$$

Approx. rec.: $$\begin{aligned}\tilde{c}_{n+1} &= \diamond\big(2\,\tilde{c}_n - \tilde{c}_{n-1}\big) \\ &= 2\,\tilde{c}_n - \tilde{c}_{n-1} + \varepsilon_n \quad \text{with } |\varepsilon_n| \leqslant \mathbf{u}\end{aligned}$$

Let $\delta_n = \tilde{c}_n - c_n$: $$\delta_{n+1} = 2\,\delta_n - \delta_{n-1} + \varepsilon_n \qquad (\delta_0 = \delta_1 = 0)$$

"global error" ⤴ ⤶ **"local error"**

Then $$\delta_n = \sum_{k=1}^{n-1} k\,\varepsilon_{n-k}$$

$$|\delta_n| \leqslant \frac{n\,(n-1)}{2}\,\mathbf{u}$$

Calculations can become unwieldy (nested sums, determinants…)

The error satisfies "the same" recurrence as the computed sequence

Generating functions

Encode sequences by generating functions

$$\text{Sequence } (f_n)_{n \in \mathbb{Z}} \quad \longleftrightarrow \quad \text{Generating series } f(z) = \sum_{n=-\infty}^{\infty} f_n \, z^n$$

Generating functions

Encode sequences by generating functions

$$\text{Sequence } (f_n)_{n\in\mathbb{Z}} \quad \longleftrightarrow \quad \text{Generating series } f(z) = \sum_{n=-\infty}^{\infty} f_n\, z^n$$

- Formulae for product, composition, …
- Analytic methods (Cauchy integrals)
- Fast algorithms
- Method of majorants
- …

Generating functions

Encode sequences by generating functions

$$\text{Sequence } (f_n)_{n\in\mathbb{Z}} \longleftrightarrow \text{Generating series } f(z) = \sum_{n=-\infty}^{\infty} f_n z^n$$

$$\delta_{n+1} = 2\,\delta_n - \delta_{n-1} + \varepsilon_n$$

$$\Big\downarrow \sum_n \square\, z^n \qquad z\sum a_n z^n = \sum a_{n-1} z^n$$

$$z^{-1}\,\delta(z) = 2\,\delta(z) - z\,\delta(z) + \varepsilon(z)$$

- Formulae for product, composition, …
- Analytic methods (Cauchy integrals)
- Fast algorithms
- Method of majorants
- …

Generating functions

Encode sequences by generating functions

$$\text{Sequence } (f_n)_{n\in\mathbb{Z}} \longleftrightarrow \text{Generating series } f(z) = \sum_{n=-\infty}^{\infty} f_n z^n$$

Exact expression of the global error δ:

$$\delta_{n+1} = 2\,\delta_n - \delta_{n-1} + \varepsilon_n$$

$\Big\downarrow \sum_n \square\, z^n$ $\qquad z\sum a_n z^n = \sum a_{n-1} z^n$

$$z^{-1}\,\delta(z) = 2\,\delta(z) - z\,\delta(z) + \varepsilon(z)$$

$$\delta(z) = \frac{z}{(1-z)^2}\,\varepsilon(z)$$

- Formulae for product, composition, ...
- Analytic methods (Cauchy integrals)
- Fast algorithms
- Method of majorants
- ...

Generating functions

Encode sequences by generating functions

Sequence $(f_n)_{n\in\mathbb{Z}}$ $\longleftrightarrow$ Generating series $f(z) = \sum_{n=-\infty}^{\infty} f_n z^n$

Exact expression of the global error δ:

$$\delta_{n+1} = 2\,\delta_n - \delta_{n-1} + \varepsilon_n$$

$\downarrow \sum_n \square\, z^n$ $\qquad z\sum a_n z^n = \sum a_{n-1} z^n$

$$z^{-1}\,\delta(z) = 2\,\delta(z) - z\,\delta(z) + \varepsilon(z)$$

$$\delta(z) = \frac{z}{(1-z)^2}\,\varepsilon(z)$$

- Formulae for product, composition, …
- Analytic methods (Cauchy integrals)
- Fast algorithms
- Method of majorants
- …

Majorants

$$f(z) = \sum_n f_n z^n, \quad \hat{f}(z) = \sum_n \hat{f}_n z^n$$

$$f \ll \hat{f} \text{ (“}\hat{f}\text{ majorizes } f\text{”)} \quad \Leftrightarrow \quad \forall n,\ |f_n| \leqslant \hat{f}_n$$

Majorants

$$f(z) = \sum_n f_n z^n, \quad \hat{f}(z) = \sum_n \hat{f}_n z^n$$

$f \ll \hat{f}$ ("$\hat{f}$ majorizes f") $\quad\Leftrightarrow\quad \forall n,\ |f_n| \leqslant \hat{f}_n$

Bound on the global error:

($\sim$ same proof, new notation)

$$\delta(z) \quad = \quad \frac{z}{(1-z)^2}\,\varepsilon(z)$$

Majorants

$$f(z) = \sum_n f_n\, z^n, \quad \hat{f}(z) = \sum_n \hat{f}_n\, z^n$$

$f \ll \hat{f}$ (“$\hat{f}$ majorizes f”) $\Leftrightarrow$ $\forall n,\ |f_n| \leqslant \hat{f}_n$

Bound on the global error:

($\sim$ same proof, new notation)

$$\delta(z) \quad = \quad \frac{z}{(1-z)^2}\,\varepsilon(z)$$

If $f(z) \ll \hat{f}(z)$ and $g(z) \ll \hat{g}(z)$,
then $f(z)\, g(z) \ll \hat{f}(z)\, \hat{g}(z)$

Majorants

$$f(z) = \sum_n f_n z^n, \quad \hat{f}(z) = \sum_n \hat{f}_n z^n$$

$f \ll \hat{f}$ ("$\hat{f}$ majorizes f") $\Leftrightarrow$ $\forall n, |f_n| \leqslant \hat{f}_n$

Bound on the global error:

($\sim$ same proof, new notation)

$$\delta(z) = \frac{z}{(1-z)^2}\,\varepsilon(z)$$

If $f(z) \ll \hat{f}(z)$ and $g(z) \ll \hat{g}(z)$,
then $f(z)\,g(z) \ll \hat{f}(z)\,\hat{g}(z)$

Proof:
$$\Big|[z^n]\big(f(z)\,g(z)\big)\Big| = \Big|\sum_{i+j=n} f_i\,g_j\Big| \leqslant \sum_{i+j=n} \hat{f}_i\,\hat{g}_j = [z^n]\big(\hat{f}(z)\,\hat{g}(z)\big)$$

Majorants

$$f(z) = \sum_n f_n\, z^n, \quad \hat{f}(z) = \sum_n \hat{f}_n\, z^n$$

$f \ll \hat{f}$ ("$\hat{f}$ majorizes f") $\Leftrightarrow$ $\forall n,\ |f_n| \leqslant \hat{f}_n$

Bound on the global error:
($\sim$ same proof, new notation)

$$\delta(z) \quad = \quad \frac{z}{(1-z)^2}\,\varepsilon(z)$$

If $f(z) \ll \hat{f}(z)$ and $g(z) \ll \hat{g}(z)$,
then $f(z)\, g(z) \ll \hat{f}(z)\, \hat{g}(z)$

Proof:
$$\begin{aligned} \Big|[z^n]\,\big(f(z)\,g(z)\big)\Big| &= \Big|\sum_{i+j=n} f_i\, g_j\Big| \\ &\leqslant \sum_{i+j=n} \hat{f}_i\, \hat{g}_j \\ &= [z^n]\,\big(\hat{f}(z)\,\hat{g}(z)\big) \end{aligned}$$

$$\varepsilon(z) = \sum_n \varepsilon_n\, z^n \ll \frac{\mathbf{u}}{1-z}$$

Majorants

$f(z) = \sum_n f_n z^n, \quad \hat{f}(z) = \sum_n \hat{f}_n z^n$

$f \ll \hat{f}$ (“$\hat{f}$ majorizes f”) $\Leftrightarrow$ $\forall n, |f_n| \leqslant \hat{f}_n$

Bound on the global error:

($\sim$ same proof, new notation)

$$\delta(z) = \frac{z}{(1-z)^2}\,\varepsilon(z)$$

- If $f(z) \ll \hat{f}(z)$ and $g(z) \ll \hat{g}(z)$, then $f(z)\,g(z) \ll \hat{f}(z)\,\hat{g}(z)$

  Proof:
$$\left|[z^n]\left(f(z)\,g(z)\right)\right| = \left|\sum_{i+j=n} f_i\,g_j\right| \leqslant \sum_{i+j=n} \hat{f}_i\,\hat{g}_j = [z^n]\left(\hat{f}(z)\,\hat{g}(z)\right)$$

- $$\varepsilon(z) = \sum_n \varepsilon_n z^n \ll \frac{\mathbf{u}}{1-z}$$

- $$\frac{z}{(1-z)^2} \in \mathbb{R}_+[[z]]$$

Majorants

$$f(z) = \sum_n f_n\, z^n, \quad \hat{f}(z) = \sum_n \hat{f}_n\, z^n$$

$f \ll \hat{f}$ ("$\hat{f}$ majorizes f") $\Leftrightarrow$ $\forall n,\ |f_n| \leqslant \hat{f}_n$

Bound on the global error:

($\sim$ same proof, new notation)

$$\delta(z) = \frac{z}{(1-z)^2}\,\varepsilon(z) \ll \frac{\mathbf{u}\, z}{(1-z)^3}$$

- If $f(z) \ll \hat{f}(z)$ and $g(z) \ll \hat{g}(z)$, then $f(z)\, g(z) \ll \hat{f}(z)\, \hat{g}(z)$

  Proof: $$\left|[z^n]\,\big(f(z)\, g(z)\big)\right| = \left|\sum_{i+j=n} f_i\, g_j\right| \leqslant \sum_{i+j=n} \hat{f}_i\, \hat{g}_j = [z^n]\,\big(\hat{f}(z)\, \hat{g}(z)\big)$$

- $$\varepsilon(z) = \sum_n \varepsilon_n\, z^n \ll \frac{\mathbf{u}}{1-z}$$

- $$\frac{z}{(1-z)^2} \in \mathbb{R}_+[[z]]$$

Majorants

$$f(z) = \sum_n f_n z^n, \quad \hat{f}(z) = \sum_n \hat{f}_n z^n$$

$$f \ll \hat{f} \text{ (“}\hat{f}\text{ majorizes } f\text{”)} \quad \Leftrightarrow \quad \forall n,\ |f_n| \leqslant \hat{f}_n$$

Bound on the global error:

($\sim$ same proof, new notation)

$$\begin{aligned} \delta(z) &= \frac{z}{(1-z)^2}\,\varepsilon(z) \\ &\ll \frac{\mathbf{u}\, z}{(1-z)^3} \end{aligned}$$

$$|\delta_n| \leqslant \frac{n\,(n-1)}{2}\,\mathbf{u}$$

- If $f(z) \ll \hat{f}(z)$ and $g(z) \ll \hat{g}(z)$, then $f(z)\,g(z) \ll \hat{f}(z)\,\hat{g}(z)$

  Proof:
  $$\begin{aligned} \left|[z^n]\,\big(f(z)\,g(z)\big)\right| &= \left|\sum_{i+j=n} f_i\, g_j\right| \\ &\leqslant \sum_{i+j=n} \hat{f}_i\, \hat{g}_j \\ &= [z^n]\,\big(\hat{f}(z)\,\hat{g}(z)\big) \end{aligned}$$

- $$\varepsilon(z) = \sum_n \varepsilon_n z^n \ll \frac{\mathbf{u}}{1-z}$$

- $$\frac{z}{(1-z)^2} \in \mathbb{R}_+[[z]]$$

Related work

> in the backward direction. There has been less attention devoted to computation which utilizes the difference equation in the forward direction, not because a forward algorithm is more difficult to analyze, but rather for the opposite reason—that its analysis was considered straightforward. Of the above

[Wimp 1972]

'Linear' error propagation

Henrici 1962	finite difference schemes for ODE
Oliver 1967	linear recurrences

Explicit bounds (not necessarily easy to compute)

von Neumann & Goldstine 1947, Turing 1948	triangular system solving
Elliott 1968	sums of generalized Fourier series
Wimp 1972	order 2
Barrio & Melendo & Serrano 2003	order n, $O(u^2)$

Transfer functions of digital filters

Liu & Kaneko 1969	random errors
Hilaire & Lopez 2013	error bounds

Relative error propagation

Model: $\diamond(x\ op\ y) = (x\ op\ y)(1+\varepsilon_{\mathrm{op}})$ with $\varepsilon_{\mathrm{op}} \in [-\mathbf{u}, \mathbf{u}]$ ($\sim$ floating-point arithmetic)

(multiplication by 2 is exact)

Exact rec.: $c_{n+1} = 2\,c_n - c_{n-1}$ $\times(1+\varepsilon_n)$

Approx. rec.:

$$\begin{aligned} \tilde{c}_{n+1} &= \diamond(2\,\tilde{c}_n - \tilde{c}_{n-1}) \\ &= (2\,\tilde{c}_n - \tilde{c}_{n-1})\,(1+\boldsymbol{\varepsilon}_n) \quad \text{with } |\varepsilon_n| \leqslant \mathbf{u} \end{aligned}$$

With $\delta_n = \tilde{c}_n - c_n$:

Relative error propagation

Model: $\diamond(x \; op \; y) = (x \; op \; y)(1 + \varepsilon_{\mathrm{op}})$ with $\varepsilon_{\mathrm{op}} \in [-\mathbf{u}, \mathbf{u}]$ ($\sim$ floating-point arithmetic)

(multiplication by 2 is exact)

Exact rec.: $c_{n+1} = 2\,c_n - c_{n-1}$ $\times(1+\varepsilon_n)$

Approx. rec.: $\tilde{c}_{n+1} = \diamond(2\,\tilde{c}_n - \tilde{c}_{n-1})$

$= (2\,\tilde{c}_n - \tilde{c}_{n-1})\,(1 + \boldsymbol{\varepsilon_n})$ with $|\varepsilon_n| \leqslant \mathbf{u}$ $\times(-1)$

With $\delta_n = \tilde{c}_n - c_n$: $\delta_{n+1} - c_{n+1}\,\varepsilon_n = (2\,\delta_n - \delta_{n-1})\,(1 + \varepsilon_n)$

Relative error propagation

Model: $\diamond(x \; op \; y) = (x \; op \; y)(1+\varepsilon_{\mathrm{op}})$ with $\varepsilon_{\mathrm{op}} \in [-\mathbf{u}, \mathbf{u}]$ ($\sim$ floating-point arithmetic)

(multiplication by 2 is exact)

Exact rec.: $c_{n+1} = 2\,c_n - c_{n-1}$ $\quad \times(1+\varepsilon_n)$

Approx. rec.:

$$\begin{aligned} \tilde{c}_{n+1} &= \diamond(2\,\tilde{c}_n - \tilde{c}_{n-1}) \\ &= (2\,\tilde{c}_n - \tilde{c}_{n-1})\,(1+\boldsymbol{\varepsilon}_n) \quad \text{with } |\varepsilon_n| \leqslant \mathbf{u} \qquad \times(-1) \end{aligned}$$

With $\delta_n = \tilde{c}_n - c_n$:

$$\delta_{n+1} - c_{n+1}\,\varepsilon_n = (2\,\delta_n - \delta_{n-1})\,(1+\varepsilon_n)$$

$$\delta_{n+1} - 2\,\delta_n + \delta_{n-1} = \varepsilon_n\,(c_{n+1} + 2\,\delta_n - \delta_{n-1})$$

Relative error propagation

Model: $\diamond(x\ op\ y) = (x\ op\ y)(1+\varepsilon_{op})$ with $\varepsilon_{op} \in [-\mathbf{u}, \mathbf{u}]$ ($\sim$ floating-point arithmetic)

(multiplication by 2 is exact)

Exact rec.: $c_{n+1} = 2\,c_n - c_{n-1}$ $\times(1+\varepsilon_n)$

Approx. rec.:

$$\tilde{c}_{n+1} = \diamond(2\,\tilde{c}_n - \tilde{c}_{n-1})$$
$$= (2\,\tilde{c}_n - \tilde{c}_{n-1})\,(1+\boldsymbol{\varepsilon_n}) \quad \text{with } |\varepsilon_n| \leqslant \mathbf{u}$$

$\times(-1)$

With $\delta_n = \tilde{c}_n - c_n$:

$$\delta_{n+1} - c_{n+1}\,\varepsilon_n = (2\,\delta_n - \delta_{n-1})\,(1+\varepsilon_n)$$
$$\delta_{n+1} - 2\,\delta_n + \delta_{n-1} = \varepsilon_n\,(c_{n+1} + 2\,\delta_n - \delta_{n-1})$$

Translate:

$\Big\downarrow\ \sum_n \square\, z^n$

$$(z^{-1} - 2 + z)\ \delta(z) = \varepsilon(z) \odot \left(z^{-1}\,c(z) + (2-z)\ \delta(z)\right)$$

Relative error propagation

Model: $\diamond(x \; op \; y) = (x \; op \; y)(1+\varepsilon_{\text{op}})$ with $\varepsilon_{\text{op}} \in [-\mathbf{u}, \mathbf{u}]$ ($\sim$ floating-point arithmetic)

(multiplication by 2 is exact)

Exact rec.: $c_{n+1} = 2\,c_n - c_{n-1}$ $\quad\times(1+\varepsilon_n)$

Approx. rec.:
$$\tilde{c}_{n+1} = \diamond(2\,\tilde{c}_n - \tilde{c}_{n-1})$$
$$= (2\,\tilde{c}_n - \tilde{c}_{n-1})\,(1+\boldsymbol{\varepsilon_n}) \quad \text{with } |\varepsilon_n| \leqslant \mathbf{u} \qquad \times(-1)$$

With $\delta_n = \tilde{c}_n - c_n$:
$$\delta_{n+1} - c_{n+1}\,\varepsilon_n = (2\,\delta_n - \delta_{n-1})\,(1+\varepsilon_n)$$
$$\delta_{n+1} - 2\,\delta_n + \delta_{n-1} = \varepsilon_n\,(c_{n+1} + 2\,\delta_n - \delta_{n-1})$$

Translate:

$\Downarrow \; \sum_n \square\, z^n$

$$(z^{-1} - 2 + z)\;\delta(z) = \varepsilon(z) \odot \left(z^{-1}\,c(z) + (2-z)\;\delta(z)\right)$$

Solve for δ ??
$$\delta(z) = \frac{(z\,\varepsilon(z)) \odot \left(c(z) + z\,(2-z)\;\delta(z)\right)}{(1-z)^2}$$

Relative error propagation

Model: $\diamond(x \; op \; y) = (x \; op \; y)(1+\varepsilon_{op})$ with $\varepsilon_{op} \in [-\mathbf{u}, \mathbf{u}]$ $\quad$ ($\sim$ floating-point arithmetic)

(multiplication by 2 is exact)

Exact rec.:
$$c_{n+1} = 2\,c_n - c_{n-1} \qquad \times(1+\varepsilon_n)$$

Approx. rec.:
$$\begin{aligned} \tilde{c}_{n+1} &= \diamond(2\,\tilde{c}_n - \tilde{c}_{n-1}) \\ &= (2\,\tilde{c}_n - \tilde{c}_{n-1})\,(1+\varepsilon_n) \quad \text{with } |\varepsilon_n| \leqslant \mathbf{u} \qquad \times(-1) \end{aligned}$$

With $\delta_n = \tilde{c}_n - c_n$:
$$\begin{aligned} \delta_{n+1} - c_{n+1}\,\varepsilon_n &= (2\,\delta_n - \delta_{n-1})\,(1+\varepsilon_n) \\ \delta_{n+1} - 2\,\delta_n + \delta_{n-1} &= \varepsilon_n\,(c_{n+1} + 2\,\delta_n - \delta_{n-1}) \end{aligned}$$

Translate:
$$\Big\downarrow \; \textstyle\sum_n \square\, z^n$$

$$(z^{-1} - 2 + z)\;\delta(z) = \varepsilon(z) \odot \big(z^{-1}\,c(z) + (2-z)\;\delta(z)\big)$$

Solve for δ ??
$$\delta(z) = \frac{(z\,\varepsilon(z)) \odot \big(c(z) + z\,(2-z)\;\delta(z)\big)}{(1-z)^2}$$

$$^{\sharp}f(z) = \sum_n |f_n|\, z^n \qquad \delta(z) \ll \frac{z\,(2+z)\,\mathbf{u}}{(1-z)^2}\;{}^{\sharp}\delta(z) + \frac{{}^{\sharp}c(z)\;\mathbf{u}}{(1-z)^2}$$

Majorizing equations

💡 Obtain the bound as a solution of a "similar" equation

$$\delta(z) \ll \frac{z\,(2+z)\,\mathbf{u}}{(1-z)^2}\,{}^\sharp\delta(z) + \frac{{}^\sharp c(z)\,\mathbf{u}}{(1-z)^2}$$

Lemma. [∼ Cauchy]

Let $\hat{a}(z), \hat{b}(z) \in \mathbb{R}_+[[z]]$ with $\hat{a}(0) = 0$. Suppose $y \in \mathbb{R}_+[[z]]$ satisfies

$$y(z) \ll \hat{a}(z)\,y(z) + \hat{b}(z).$$

Then $y(z)$ is majorized by the solution of $\hat{y}(z) = \hat{a}(z)\,\hat{y}(z) + \hat{b}(z)$, i.e.,

$$y(z) \ll \hat{y}(z) = \frac{\hat{b}(z)}{1 - \hat{a}(z)}.$$

Majorizing equations

Obtain the bound as a solution of a “similar” equation $\delta(z) \ll \frac{z\,(2+z)\,\mathbf{u}}{(1-z)^2}\,{}^\sharp\delta(z) + \frac{{}^\sharp c(z)\,\mathbf{u}}{(1-z)^2}$

Lemma. [∼ Cauchy]

Let $\hat{a}(z), \hat{b}(z) \in \mathbb{R}_+[[z]]$ with $\hat{a}(0) = 0$. Suppose $y \in \mathbb{R}_+[[z]]$ satisfies

$$y(z) \ll \hat{a}(z)\,y(z) + \hat{b}(z).$$

Then $y(z)$ is majorized by the solution of $\hat{y}(z) = \hat{a}(z)\,\hat{y}(z) + \hat{b}(z)$, i.e.,

$$y(z) \ll \hat{y}(z) = \frac{\hat{b}(z)}{1-\hat{a}(z)}.$$

Proof.

- $y_0 \leqslant \hat{b}_0 = \hat{y}_0$
- $y_n \leqslant \sum_{i=0}^{n} \hat{a}_i\,y_{n-i} + \hat{b}_n$ □

Majorizing equations

💡 Obtain the bound as a solution of a "similar" equation $\delta(z) \ll \frac{z\,(2+z)\,\mathbf{u}}{(1-z)^2}\,{}^\sharp\delta(z) + \frac{{}^\sharp c(z)\,\mathbf{u}}{(1-z)^2}$

Lemma. [∼ Cauchy]

Let $\hat{a}(z), \hat{b}(z) \in \mathbb{R}_+[[z]]$ with $\hat{a}(0) = 0$. Suppose $y \in \mathbb{R}_+[[z]]$ satisfies

$$y(z) \ll \hat{a}(z)\,y(z) + \hat{b}(z).$$

Then $y(z)$ is majorized by the solution of $\hat{y}(z) = \hat{a}(z)\,\hat{y}(z) + \hat{b}(z)$, i.e.,

$$y(z) \ll \hat{y}(z) = \frac{\hat{b}(z)}{1 - \hat{a}(z)}.$$

Proof.

- $y_0 \leqslant \hat{b}_0 = \hat{y}_0$
- $y_n \leqslant \sum_{i=1}^{n} \hat{a}_i\,y_{n-i} + \hat{b}_n$ □

Majorizing equations

💡 Obtain the bound as a solution of a "similar" equation $\delta(z) \ll \frac{z\,(2+z)\,\mathbf{u}}{(1-z)^2}\,{}^\sharp\delta(z) + \frac{{}^\sharp c(z)\,\mathbf{u}}{(1-z)^2}$

Lemma. [∼ Cauchy]

Let $\hat{a}(z), \hat{b}(z) \in \mathbb{R}_+[[z]]$ with $\hat{a}(0) = 0$. Suppose $y \in \mathbb{R}_+[[z]]$ satisfies

$$y(z) \ll \hat{a}(z)\,y(z) + \hat{b}(z).$$

Then $y(z)$ is majorized by the solution of $\hat{y}(z) = \hat{a}(z)\,\hat{y}(z) + \hat{b}(z)$, i.e.,

$$y(z) \ll \hat{y}(z) = \frac{\hat{b}(z)}{1 - \hat{a}(z)}.$$

Proof.

- $y_0 \leqslant \hat{b}_0 = \hat{y}_0$
- $y_n \leqslant \sum_{i=1}^{n} \hat{a}_i\,y_{n-i} + \hat{b}_n \leqslant \sum_{i=1}^{n} \hat{a}_i\,\hat{y}_{n-i} + \hat{b}_n = \hat{y}_n$ □

Bound on the floating-point error

$$
{}^{\sharp}\delta(z) \ll \underbrace{\frac{z\,(2+z)\,\mathbf{u}}{(1-z)^2}}_{\hat{a}(z)}\,{}^{\sharp}\delta(z) + \underbrace{\frac{{}^{\sharp}c(z)\;\mathbf{u}}{(1-z)^2}}_{\hat{b}(z)}
$$

Bound on the floating-point error

$$\sharp\delta(z) \ll \underbrace{\frac{z\,(2+z)\,\mathbf{u}}{(1-z)^2}}_{\hat{a}(z)} \sharp\delta(z) + \underbrace{\frac{\sharp c(z)\;\mathbf{u}}{(1-z)^2}}_{\hat{b}(z)}$$

$$\begin{aligned} c(z) &= \frac{c_0}{(1-z)^2} \\ &\ll \frac{|c_0|}{(1-z)^2} =: \hat{c}(z) \end{aligned}$$

Bound on the floating-point error

$$^\sharp\delta(z) \ll \underbrace{\frac{z\,(2+z)\,\mathbf{u}}{(1-z)^2}}_{\hat{a}(z)}\, {}^\sharp\delta(z) + \underbrace{\frac{\hat{c}(z)\quad \mathbf{u}}{(1-z)^2}}_{\hat{b}(z)}$$

$$\begin{aligned} c(z) &= \frac{c_0}{(1-z)^2} \\ &\ll \frac{|c_0|}{(1-z)^2} =: \hat{c}(z) \end{aligned}$$

Bound on the floating-point error

$$\sharp\delta(z) \ll \underbrace{\frac{z\,(2+z)\,\mathbf{u}}{(1-z)^2}}_{\hat{a}(z)} \sharp\delta(z) + \underbrace{\frac{\hat{c}(z)\;\;\mathbf{u}}{(1-z)^2}}_{\hat{b}(z)}$$

$$\begin{aligned} c(z) &= \frac{c_0}{(1-z)^2} \\ &\ll \frac{|c_0|}{(1-z)^2} =: \hat{c}(z) \end{aligned}$$

By the lemma:

$$\delta(z) \ll \frac{\hat{b}(z)}{1-\hat{a}(z)}$$

Bound on the floating-point error

$$ {}^{\sharp}\delta(z) \ll \underbrace{\frac{z\,(2+z)\,\mathbf{u}}{(1-z)^2}}_{\hat{a}(z)} {}^{\sharp}\delta(z) + \underbrace{\frac{\hat{c}(z)\quad \mathbf{u}}{(1-z)^2}}_{\hat{b}(z)} $$

$$ c(z) = \frac{c_0}{(1-z)^2} \ll \frac{|c_0|}{(1-z)^2} =: \hat{c}(z) $$

By the lemma:

$$ \delta(z) \ll \frac{\hat{b}(z)}{1-\hat{a}(z)} = \frac{\hat{c}(z)\,\mathbf{u}}{1-2\,(1+\mathbf{u})\,z+(1-\mathbf{u})\,z^2} $$

Bound on the floating-point error

$$^\sharp\delta(z) \ll \underbrace{\frac{z\,(2+z)\,\mathbf{u}}{(1-z)^2}}_{\hat{a}(z)}\, ^\sharp\delta(z) + \underbrace{\frac{\hat{c}(z)\quad \mathbf{u}}{(1-z)^2}}_{\hat{b}(z)}$$

$$c(z) = \frac{c_0}{(1-z)^2} \ll \frac{|c_0|}{(1-z)^2} =: \hat{c}(z)$$

By the lemma:

$$\delta(z) \ll \frac{\hat{b}(z)}{1-\hat{a}(z)} = \frac{\hat{c}(z)\,\mathbf{u}}{1-2\,(1+\mathbf{u})\,z+(1-\mathbf{u})\,z^2} = \frac{\hat{c}(z)\,\mathbf{u}}{(1-\alpha\,z)\,(1-\beta\,z)}$$

$$\alpha = 1+2\sqrt{\mathbf{u}}+O(u)$$

Bound on the floating-point error

$$ {}^{\sharp}\delta(z) \ll \underbrace{\frac{z\,(2+z)\,\mathbf{u}}{(1-z)^2}}_{\hat{a}(z)} {}^{\sharp}\delta(z) + \underbrace{\frac{\hat{c}(z) \quad \mathbf{u}}{(1-z)^2}}_{\hat{b}(z)} $$

$$ c(z) = \frac{c_0}{(1-z)^2} \ll \frac{|c_0|}{(1-z)^2} =: \hat{c}(z) $$

By the lemma:

$$ \begin{aligned} \delta(z) &\ll \frac{\hat{b}(z)}{1-\hat{a}(z)} \\ &= \frac{\hat{c}(z)\,\mathbf{u}}{1-2\,(1+\mathbf{u})\,z+(1-\mathbf{u})\,z^2} \\ &= \frac{\hat{c}(z)\,\mathbf{u}}{(1-\alpha\,z)\,(1-\beta\,z)} \\ &\ll \frac{|c_0|\mathbf{u}}{(1-\alpha\,z)^4} \end{aligned} $$

$$ \alpha = 1 + 2\sqrt{\mathbf{u}} + O(u) $$

Bound on the floating-point error

$$^\sharp\delta(z) \ll \underbrace{\frac{z\,(2+z)\,\mathbf{u}}{(1-z)^2}}_{\hat{a}(z)}\,{}^\sharp\delta(z) + \underbrace{\frac{\hat{c}(z)\quad\mathbf{u}}{(1-z)^2}}_{\hat{b}(z)}$$

$$c(z) = \frac{c_0}{(1-z)^2} \ll \frac{|c_0|}{(1-z)^2} =: \hat{c}(z)$$

By the lemma:

$$\delta(z) \ll \frac{\hat{b}(z)}{1-\hat{a}(z)} = \frac{\hat{c}(z)\,\mathbf{u}}{1-2\,(1+\mathbf{u})\,z+(1-\mathbf{u})\,z^2} = \frac{\hat{c}(z)\,\mathbf{u}}{(1-\boldsymbol{\alpha}\,z)\,(1-\boldsymbol{\beta}\,z)} \ll \frac{|c_0|\mathbf{u}}{(1-\boldsymbol{\alpha}\,z)^4}$$

$$\boldsymbol{\alpha} = 1+2\sqrt{\mathbf{u}}+O(u)$$

Absolute error on c_n: $$|\delta_n| \leqslant \frac{|c_0|}{6}\,(n+3)^3\,\alpha^n\,\mathbf{u}$$

Bound on the floating-point error

$$^\sharp\delta(z) \ll \underbrace{\frac{z\,(2+z)\,\mathbf{u}}{(1-z)^2}}_{\hat{a}(z)} {}^\sharp\delta(z) + \underbrace{\frac{\hat{c}(z)\quad \mathbf{u}}{(1-z)^2}}_{\hat{b}(z)}$$

$$c(z) = \frac{c_0}{(1-z)^2} \ll \frac{|c_0|}{(1-z)^2} =: \hat{c}(z)$$

By the lemma:

$$\delta(z) \ll \frac{\hat{b}(z)}{1-\hat{a}(z)} = \frac{\hat{c}(z)\,\mathbf{u}}{1-2\,(1+\mathbf{u})\,z+(1-\mathbf{u})\,z^2} = \frac{\hat{c}(z)\,\mathbf{u}}{(1-\boldsymbol{\alpha}\,z)\,(1-\boldsymbol{\beta}\,z)} \ll \frac{|c_0|\mathbf{u}}{(1-\boldsymbol{\alpha}\,z)^4}$$

$$\boldsymbol{\alpha} = 1 + 2\sqrt{\mathbf{u}} + O(u)$$

Absolute error on c_n:

$$|\delta_n| \leqslant \frac{|c_0|}{6}\,(n+3)^3\,\alpha^n\,\mathbf{u}$$

Exponential for fixed $\mathbf{u}$,
but $O(n^3\,\mathbf{u})$ if $n = O(\mathbf{u}^{-1/2})$

Differential equations

Evaluation of Legendre polynomials

[Johansson & M. 2018]

Algorithm 1. Evaluation of Legendre polynomials in GMP fixed-point arithmetic.

Input: An integer x and $t \geq 0$ such that $|2^{-t}x| \leq 1$, and $n \geq 1$

Output: p, q such that $|2^{-t}p - P_{n-1}(2^{-t}x)|, |2^{-t}q - P_n(2^{-t}x)| \leq (0.75\,(n+1)(n+2)+1)\,2^{-t}$

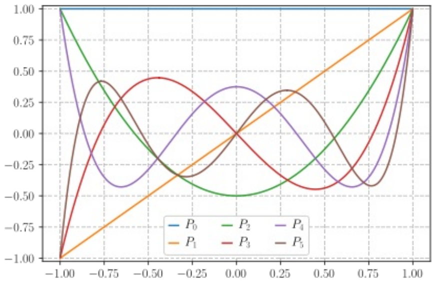
```
void legendre(mpz_t p, mpz_t q, int n, const mpz_t x, int t) {
    mpz_t tmp; int k; mpz_init(tmp);                 ▷ Comments use the notation of
    mp_limb_t denlo, den = 1;                         ▷ the proof of Corollary 6
    mpz_set_ui(p, 1); mpz_mul_2exp(p, p, t);          ▷ p0 = 2^t
    mpz_set(q, x);                                    ▷ q0 = x̂
    for (k = 1; k < n; k++) {
        mpz_mul(tmp, q, x); mpz_tdiv_q_2exp(tmp, tmp, t);  ▷ ⌊x̂ q_{k-1} 2^{-t}⌋
        mpz_mul_si(p, p, -k*k)
        mpz_addmul_ui(p, tmp, 2*k+1);                 ▷ -k^2 p_{k-1} + (2k+1) tmp
        mpz_swap(p, q);
        if (mpn_mul_1(&denlo, &den, 1, k+1)) {        ▷ If multiplication overflows
            mpz_tdiv_q_ui(p, p, den);                 ▷ ⌊p/d_{k-1}⌋
            mpz_tdiv_q_ui(q, q, den);
            den = k+1;                                ▷ d_k = k + 1
        } else den = denlo;                           ▷ d_k = (k + 1) d_{k-1}
    }
    mpz_tdiv_q_ui(p, p, den/n); mpz_tdiv_q_ui(q, q, den);
    mpz_clear(tmp);
}
```

Context: rigorous arbitrary-precision Gauss-Legendre quadrature [Johansson 2018]

$$\mathbf{P}_{n+1}(\mathbf{x}) = \frac{1}{n+1}\left[(2n+1)\,x\,\mathbf{P}_n(\mathbf{x}) - n\,\mathbf{P}_{n-1}(\mathbf{x})\right]$$

$\tilde{p}_n = P_n(x)$ evaluated using this recurrence in t-bit fixed-point arithmetic

Bound $|\tilde{p}_n - P_n(x)|$.



Legendre polynomials: error analysis

Exact rec.: $$p_{n+1} = \frac{1}{n+1}\left[(2\,n+1)\,x\,p_n - n\,p_{n-1}\right] \qquad p_n := P_n(x)$$

Approx. rec.: $$\tilde{p}_{n+1} = \frac{1}{n+1}\left[(2\,n+1)\,x\,\tilde{p}_n - n\,\tilde{p}_{n-1}\right] + \varepsilon_{n+1} \qquad \text{with}\ \varepsilon_n \leqslant 3\,\mathbf{u}$$

Legendre polynomials: error analysis

Exact rec.: $$p_{n+1} = \frac{1}{n+1}\left[(2n+1)\,x\,p_n - n\,p_{n-1}\right] \qquad p_n := P_n(x)$$

Approx. rec.: $$\tilde{p}_{n+1} = \frac{1}{n+1}\left[(2n+1)\,x\,\tilde{p}_n - n\,\tilde{p}_{n-1}\right] + \varepsilon_{n+1} \qquad \text{with } \varepsilon_n \leqslant 3\,\mathbf{u}$$

Global error:

$\delta_n = \tilde{p}_n - p_n$

$$(\mathbf{n}+1)\,\delta_{n+1} = (2\,\mathbf{n}+1)\,x\,\delta_n - \mathbf{n}\,\delta_{n-1} + (\mathbf{n}+1)\,\varepsilon_{n+1}$$

Legendre polynomials: error analysis

Exact rec.: $$p_{n+1} = \frac{1}{n+1}\left[(2n+1)\,x\,p_n - n\,p_{n-1}\right] \qquad p_n := P_n(x)$$

Approx. rec.: $$\tilde{p}_{n+1} = \frac{1}{n+1}\left[(2n+1)\,x\,\tilde{p}_n - n\,\tilde{p}_{n-1}\right] + \varepsilon_{n+1} \qquad \text{with } \varepsilon_n \leqslant 3\,\mathbf{u}$$

Global error:

$\delta_n = \tilde{p}_n - p_n$

$$(\mathbf{n}+1)\,\delta_{n+1} = (2\,\mathbf{n}+1)\,x\,\delta_n - \mathbf{n}\,\delta_{n-1} + (\mathbf{n}+1)\,\varepsilon_{n+1}$$

Translate:

$$\Big\Downarrow \ \textstyle\sum_n \Box\, z^n$$

$$(1 - 2\,x\,z + z^2)\,\boldsymbol{\delta}'(z) = z\,(x-z)\,\delta(z) + \boldsymbol{\varepsilon}'(z)$$

Legendre polynomials: error analysis

Exact rec.: $$p_{n+1} = \frac{1}{n+1}\left[(2n+1)\,x\,p_n - n\,p_{n-1}\right] \qquad p_n := P_n(x)$$

Approx. rec.: $$\tilde{p}_{n+1} = \frac{1}{n+1}\left[(2n+1)\,x\,\tilde{p}_n - n\,\tilde{p}_{n-1}\right] + \varepsilon_{n+1} \qquad \text{with } \varepsilon_n \leqslant 3\,\mathbf{u}$$

Global error:

$\delta_n = \tilde{p}_n - p_n$ $$(\mathbf{n}+1)\,\delta_{n+1} = (2\,\mathbf{n}+1)\,x\,\delta_n - \mathbf{n}\,\delta_{n-1} + (\mathbf{n}+1)\,\varepsilon_{n+1}$$

Translate:

$$\Big\downarrow \textstyle\sum_n \square\, z^n$$

$$(1-2\,x\,z+z^2)\,\boldsymbol{\delta}'(z) = z\,(x-z)\,\delta(z) + \boldsymbol{\varepsilon}'(z)$$

Solve: $$\delta(z) = \frac{1}{\sqrt{1-2\,x\,z+z^2}}\int_0^z \frac{\varepsilon'(w)}{\sqrt{1-2\,x\,w+w^2}}\,\mathrm{d}w$$

Legendre polynomials: bound

$$\delta(z) = \frac{1}{\sqrt{1-2\,x\,z+z^2}}\int_0^z \frac{\varepsilon'(w)}{\sqrt{1-2\,x\,w+w^2}}\,\mathrm{d}w \qquad \varepsilon(z) \ll \frac{3\,\mathbf{u}}{1-z}$$

Legendre polynomials: bound

$$\delta(z) = \frac{1}{\sqrt{1-2\,x\,z+z^2}} \int_0^z \frac{\varepsilon'(w)}{\sqrt{1-2\,x\,w+w^2}}\,\mathrm{d}w \qquad \varepsilon(z) \ll \frac{3\,\mathbf{u}}{1-z}$$

$$\frac{3\,\mathbf{u}}{(1-z)^2}$$

Legendre polynomials: bound

$$\delta(z) = \frac{1}{\sqrt{1-2\,x\,z+z^2}} \int_0^z \frac{\varepsilon'(w)}{\sqrt{1-2\,x\,w+w^2}}\,\mathrm{d}w \qquad \varepsilon(z) \ll \frac{3\,\mathbf{u}}{1-z}$$

$$\frac{1}{1-z} \qquad \frac{1}{1-z}\,\frac{3\,\mathbf{u}}{(1-z)^2}$$

Legendre polynomials: bound

$$\delta(z) = \frac{1}{\sqrt{1-2\,x\,z+z^2}}\int_0^z \frac{\varepsilon'(w)}{\sqrt{1-2\,x\,w+w^2}}\,\mathrm{d}w \qquad \varepsilon(z) \ll \frac{3\,\mathbf{u}}{1-z}$$

$$\ll \frac{1}{1-z}\int \frac{1}{1-z}\,\frac{3\,\mathbf{u}}{(1-z)^2}$$

Legendre polynomials: bound

$$\delta(z) = \frac{1}{\sqrt{1-2\,x\,z+z^2}} \int_0^z \frac{\varepsilon'(w)}{\sqrt{1-2\,x\,w+w^2}}\,\mathrm{d}w \qquad \varepsilon(z) \ll \frac{3\,\mathbf{u}}{1-z}$$

$$\ll \frac{1}{1-z} \int \frac{1}{1-z} \frac{3\,\mathbf{u}}{(1-z)^2}$$

$$= \frac{3}{2} \frac{1}{(1-z)^3}\,\mathbf{u}$$

Proposition. [Johansson & M.]

For all $x \in [-1, 1]$ and $n \in \mathbb{N}$, the error in the recursive fixed-point computation of Legendre polynomials satisfies

$$|\tilde{p}_n - P_n(x)| \leqslant \frac{3}{4}\,(n+1)\,(n+2)\,\mathbf{u}.$$

Legendre polynomials: bound

$$\begin{aligned}\delta(z) &= \frac{1}{\sqrt{1-2\,x\,z+z^2}} \int_0^z \frac{\varepsilon'(w)}{\sqrt{1-2\,x\,w+w^2}}\,\mathrm{d}w \\ &\ll \frac{1}{1-z} \int \frac{1}{1-z} \frac{3\,\mathbf{u}}{(1-z)^2} \\ &= \frac{3}{2} \frac{1}{(1-z)^3}\,\mathbf{u}\end{aligned}$$

$$\varepsilon(z) \ll \frac{3\,\mathbf{u}}{1-z}$$

Proposition. [Johansson & M.]

For all $x \in [-1, 1]$ and $n \in \mathbb{N}$, the error in the recursive fixed-point computation of Legendre polynomials satisfies

$$|\tilde{p}_n - P_n(x)| \leqslant \frac{3}{4}(n+1)\,(n+2)\,\mathbf{u}.$$

We were lucky that the equation could be solved explicitly

Partial sums of differentially finite series

$$L(u) = a_r\, u^{(r)} + \cdots + a_1\, u' + a_0\, u = 0, \quad a_i \in \mathbb{C}[z]$$

Given | the operator L, initial values $u_0, \ldots, u_{r-1}$, an evaluation point ζ, a truncation order N — compute an **enclosure** of $\sum_{n=0}^{N-1} u_n\, \zeta^n$.

Assumptions

ordinary point	$a_r(0) \neq 0$
"obvious" geometric convergence	$\lvert\zeta\rvert < \min\{\lvert\xi\rvert : a_r(\xi) = 0\}$

Strategy

- Compute a recurrence on the u_n
- Compute and sum the $u_n\, \zeta^n$ iteratively $\Rightarrow$ need to avoid interval blow-up

D-finite series: error propagation

Exact rec.: $$u_n = \frac{-1}{b_s(n)}\left[b_{s-1}(n)\,u_{n-1} + \cdots + b_1(n)\,u_{n-s+1} + b_0(n)\,u_{n-s}\right]$$

Approx. rec.: $$\tilde{u}_n = \frac{-1}{b_s(n)}\left[b_{s-1}(n)\,\tilde{u}_{n-1} + \cdots + b_1(n)\,\tilde{u}_{n-s+1} + b_0(n)\,\tilde{u}_{n-s}\right] + \varepsilon_n$$

local error bound $|\varepsilon_n| \leqslant \hat{\varepsilon}_n$ computed on the fly

('running' error analysis)

D-finite series: error propagation

Exact rec.: $$u_n = \frac{-1}{b_s(n)}\left[b_{s-1}(n)\, u_{n-1} + \cdots + b_1(n)\, u_{n-s+1} + b_0(n)\, u_{n-s}\right]$$

Approx. rec.: $$\tilde{u}_n = \frac{-1}{b_s(n)}\left[b_{s-1}(n)\, \tilde{u}_{n-1} + \cdots + b_1(n)\, \tilde{u}_{n-s+1} + b_0(n)\, \tilde{u}_{n-s}\right] + \varepsilon_n$$

local error bound $|\varepsilon_n| \leqslant \hat{\varepsilon}_n$ computed on the fly

('running' error analysis)

The global error $\delta_n = \tilde{u}_n - u_n$ satisfies

$$b_s(n)\, \delta_n + b_{s-1}(n)\, \delta_{n-1} + \cdots + b_0(n)\, \delta_{n-s} = b_s(n)\, \varepsilon_n$$

$$\Big\downarrow \quad \textstyle\sum_n \square\, z^n$$

$$a_r(z)\, \delta^{(r)}(z) + \cdots + a_1(z)\, \delta'(z) + a_0(z)\, \delta(z) = Q(\theta) \cdot \varepsilon(z) \qquad \theta = z\frac{d}{dz}$$

$Q(\theta) = b_s(0)\, \theta\, (\theta - 1) \cdots (\theta - s + 1)$ (ordinary point)

Compute a bound on δ_n given one on ε_n?

D-finite series: error bound

$$a_r(z)\,\delta^{(r)}(z) + \cdots + a_1(z)\,\delta'(z) + a_0(z)\,\delta(z) = Q(\theta)\cdot\varepsilon(z)$$

Lemma. [∼ Cauchy]

Let $a_0, \ldots, a_r \in \mathbb{C}[z]$. Suppose $y \in \mathbb{C}[[z]]$ satisfies

$$a_r(z)\,y^{(r)}(z) + \cdots + a_0(z)\,y(z) = Q(\theta)\cdot\varepsilon(z).$$

Suppose $\varepsilon(z) \ll \hat{\varepsilon}(z)$.

One can compute a rational series $\hat{a}(z) \in \mathbb{R}_+[[z]]$ such that $y(z)$ is majorized by any solution of

$$\hat{y}'(z) = \hat{a}(z)\,\hat{y}(z) + \hat{\varepsilon}(z)$$

such that $|y_0|, \ldots, |y_{r-1}| \leqslant \hat{y}_0$.

D-finite series: error bound

$$a_r(z)\,\delta^{(r)}(z) + \cdots + a_1(z)\,\delta'(z) + a_0(z)\,\delta(z) = Q(\theta)\cdot\varepsilon(z)$$

Lemma. [∼ Cauchy]

Let $a_0, \ldots, a_r \in \mathbb{C}[z]$. Suppose $y \in \mathbb{C}[[z]]$ satisfies

$$a_r(z)\,y^{(r)}(z) + \cdots + a_0(z)\,y(z) = Q(\theta)\cdot\varepsilon(z).$$

Suppose $\varepsilon(z) \ll \hat{\varepsilon}(z)$.

One can compute a rational series $\hat{a}(z) \in \mathbb{R}_+[[z]]$ such that $y(z)$ is majorized by any solution of

$$\hat{y}'(z) = \hat{a}(z)\,\hat{y}(z) + \hat{\varepsilon}(z)$$

such that $|y_0|, \ldots, |y_{r-1}| \leqslant \hat{y}_0$.

Solve: $$\hat{\delta}(z) = \hat{h}(z)\left(\mathrm{cst} + \int_0^z \frac{\hat{\varepsilon}(w)}{\hat{h}(w)}\,\mathrm{d}w\right), \qquad \hat{h}(z) = \exp\int_0^z \hat{a}(w)\,\mathrm{d}w$$

D-finite series: error bound

$$a_r(z)\,\delta^{(r)}(z) + \cdots + a_1(z)\,\delta'(z) + a_0(z)\,\delta(z) = Q(\theta)\cdot\varepsilon(z)$$

Lemma. [∼ Cauchy]

Let $a_0, \ldots, a_r \in \mathbb{C}[z]$. Suppose $y \in \mathbb{C}[[z]]$ satisfies

$$a_r(z)\,y^{(r)}(z) + \cdots + a_0(z)\,y(z) = Q(\theta)\cdot\varepsilon(z).$$

Suppose $\varepsilon(z) \ll \hat{\varepsilon}(z)$.

One can compute a rational series $\hat{a}(z) \in \mathbb{R}_+[[z]]$ such that $y(z)$ is majorized by any solution of

$$\hat{y}'(z) = \hat{a}(z)\,\hat{y}(z) + \hat{\varepsilon}(z)$$

such that $|y_0|, \ldots, |y_{r-1}| \leqslant \hat{y}_0$.

Solve: $$\hat{\delta}(z) = \hat{h}(z)\left(\not{c}st + \int_0^z \frac{\hat{\varepsilon}(w)}{\hat{h}(w)}\,dw\right), \qquad \hat{h}(z) = \exp\int_0^z \hat{a}(w)\,dw$$

D-finite series: error bound

$$a_r(z)\,\delta^{(r)}(z) + \cdots + a_1(z)\,\delta'(z) + a_0(z)\,\delta(z) = Q(\theta)\cdot\varepsilon(z)$$

Lemma. [$\sim$ Cauchy]

Let $a_0, \ldots, a_r \in \mathbb{C}[z]$. Suppose $y \in \mathbb{C}[[z]]$ satisfies

$$a_r(z)\,y^{(r)}(z) + \cdots + a_0(z)\,y(z) = Q(\theta)\cdot\varepsilon(z).$$

Suppose $\varepsilon(z) \ll \hat{\varepsilon}(z)$.

One can compute a rational series $\hat{a}(z) \in \mathbb{R}_+[[z]]$ such that $y(z)$ is majorized by any solution of

$$\hat{y}'(z) = \hat{a}(z)\,\hat{y}(z) + \hat{\varepsilon}(z)$$

such that $|y_0|, \ldots, |y_{r-1}| \leqslant \hat{y}_0$.

Solve: $\hat{\delta}(z) = \hat{h}(z)\left(\not{c}st + \int_0^z \frac{\hat{\varepsilon}(w)}{\hat{h}(w)}\,dw\right)$, $\qquad \hat{h}(z) = \exp\int_0^z \hat{a}(w)\,dw$

Choose $\hat{\varepsilon}_n = \bar{\boldsymbol{\varepsilon}}\,\hat{h}_n$: $\qquad = \bar{\boldsymbol{\varepsilon}}\,z\,\hat{h}(z)$

D-finite series: error bound

$$a_r(z)\ \delta^{(r)}(z)\ + \cdots + a_1(z)\ \delta'(z)\ + a_0(z)\ \delta(z)\ = Q(\theta) \cdot \varepsilon(z)$$

Lemma. [∼ Cauchy]

Let $a_0, \ldots, a_r \in \mathbb{C}[z]$. Suppose $y \in \mathbb{C}[[z]]$ satisfies

$$a_r(z)\, y^{(r)}(z) + \cdots + a_0(z)\, y(z) = Q(\theta) \cdot \varepsilon(z).$$

Suppose $\varepsilon(z) \ll \hat{\varepsilon}(z)$.

One can compute a rational series $\hat{a}(z) \in \mathbb{R}_+[[z]]$ such that $y(z)$ is majorized by any solution of

$$\hat{y}'(z) = \hat{a}(z)\, \hat{y}(z) + \hat{\varepsilon}(z)$$

such that $|y_0|, \ldots, |y_{r-1}| \leqslant \hat{y}_0$.

Solve: $$\hat{\delta}(z) = \hat{h}(z) \left(\not{c}s\not{t} + \int_0^z \frac{\hat{\varepsilon}(w)}{\hat{h}(w)}\, dw \right), \qquad \hat{h}(z) = \exp \int_0^z \hat{a}(w)\, dw$$

Choose $\hat{\varepsilon}_n = \bar{\boldsymbol{\varepsilon}}\, \hat{h}_n$: $$= \bar{\boldsymbol{\varepsilon}}\, z\, \hat{h}(z)$$

D-finite series: error bound

$$a_r(z)\,\delta^{(r)}(z) + \cdots + a_1(z)\,\delta'(z) + a_0(z)\,\delta(z) = Q(\theta)\cdot\varepsilon(z)$$

Lemma. [∼ Cauchy]

Let $a_0, \ldots, a_r \in \mathbb{C}[z]$. Suppose $y \in \mathbb{C}[[z]]$ satisfies

$$a_r(z)\,y^{(r)}(z) + \cdots + a_0(z)\,y(z) = Q(\theta)\cdot\varepsilon(z).$$

Suppose $\varepsilon(z) \ll \hat{\varepsilon}(z)$.

One can compute a rational series $\hat{a}(z) \in \mathbb{R}_+[[z]]$ such that $y(z)$ is majorized by any solution of

$$\hat{y}'(z) = \hat{a}(z)\,\hat{y}(z) + \hat{\varepsilon}(z)$$

such that $|y_0|, \ldots, |y_{r-1}| \leqslant \hat{y}_0$.

Solve: $$\hat{\delta}(z) = \hat{h}(z)\left(\mathrm{c\not{s}t} + \int_0^z \frac{\hat{\varepsilon}(w)}{\hat{h}(w)}\,dw\right), \qquad \hat{h}(z) = \exp\int_0^z \hat{a}(w)\,dw$$

Choose $\hat{\varepsilon}_n = \bar{\varepsilon}\,\hat{h}_n$: $$= \bar{\varepsilon}\,z\,\hat{h}(z)$$

Compute a bound on the truncation error at the same time

Bernoulli Numbers

Scaled Bernoulli numbers

$$B_n = 1, \frac{-1}{2}, \frac{1}{6}, 0, \frac{-1}{30}, 0, \frac{1}{42}, 0, \frac{-1}{30}, 0, \frac{5}{66}, 0, \frac{-691}{2730}, 0, \frac{7}{6}, 0, \frac{-3617}{510}, \ldots \qquad |B_{2k}| \sim \frac{2\,(2k)!}{(2\pi)^{2k}}$$

$$\mathbf{b}_k = \frac{B_{2k}}{(2k)!} \qquad b(z) = \sum_{k=0}^{\infty} \mathbf{b}_k z^k = \frac{\sqrt{z}/2}{\tanh(\sqrt{z}/2)}$$

Algorithm. [Brent 1980, based on a suggestion of Reinsch]

$$\mathbf{b}_k = \frac{1}{(2k)!\,4^k} - \sum_{j=0}^{k-1} \frac{\mathbf{b}_j}{(2k+1-2j)!\,4^{k-j}}$$

be used with sufficient guard digits, or a more stable recurrence must be used. If we multiply both sides of (30) by $\sinh(x/2)/x$ and equate coefficients, we get the recurrence

$$C_k + \frac{C_{k-1}}{3!\,4} + \cdots + \frac{C_1}{(2k-1)!\,4^{k-1}} = \frac{2k}{(2k+1)!\,4^k} \qquad (36)$$

If (36) is used to evaluate C_k, using precision n arithmetic, the error is only $O(k^2 2^{-n})$. Thus,

[Brent 1980]

Error in the floating-point computation of b_k

$$b_k = \frac{1}{(2k)!\,4^k} - \sum_{j=0}^{k-1} \frac{b_j}{(2k+1-2j)!\,4^{k-j}}, \qquad \tilde{b}_k = \text{computed values}$$

Exercise 4.35 Prove (or give a plausibility argument for) the statements made in §4.7 that: (a) if a recurrence based on (4.59) is used to evaluate the scaled Bernoulli number C_k, using precision n arithmetic, then the relative error is of order $4^k 2^{-n}$; and (b) if a recurrence based on (4.60) is used, then the relative error is $O(k^2 2^{-n})$.

[Brent & Zimmermann 2010]

Conjecture. [Brent, Zimmermann]

The computed values $\tilde{b}_k$ satisfy $\tilde{b}_k = b_k\,(1+\eta_k)$ where $\eta_k = O(k \cdot \mathbf{u})$.

$\mathbf{u}$ = unit roundoff

Remark. To be understood as $\eta_k = O(k \cdot \mathbf{u})$ when $k = O(\mathbf{u}^{-1})$
or $|\eta_k| \leqslant C_k\, u$ as $\mathbf{u} \to 0$ with $C_k = O(k)$ (resp. $O(k^2)$)

Error analysis

$$b_k = \frac{1}{(2k)!\,4^k} - \sum_{j=0}^{k-1} \frac{b_j}{(2k+1-2j)!\,4^{k-j}}, \quad \tilde{b}_k = \text{computed values}$$

Local error analysis.

$$\tilde{b}_k = \frac{1+\mathbf{s_k}}{(2k)!\,4^k} - \sum_{j=0}^{k-1} \frac{\tilde{b}_j\,(1+\mathbf{t_{k,j}})}{(2k+1-2j)!\,4^{k-j}}$$

$$|\mathbf{s_k}| \leqslant \hat{\theta}_{2k}$$
$$|\mathbf{t_{k,j}}| \leqslant \hat{\theta}_{3(k-j)+2}$$
$$\text{where } \hat{\theta}_n = (1+\mathbf{u})^n - 1$$

Error analysis

$$b_k = \frac{1}{(2k)!\,4^k} - \sum_{j=0}^{k-1} \frac{b_j}{(2k+1-2j)!\,4^{k-j}}, \quad \tilde{b}_k = \text{computed values}$$

Local error analysis.

$$\tilde{b}_k = \frac{1+\mathbf{s_k}}{(2k)!\,4^k} - \sum_{j=0}^{k-1} \frac{\tilde{b}_j\,(1+\mathbf{t_{k,j}})}{(2k+1-2j)!\,4^{k-j}}$$

$$|\mathbf{s_k}| \leqslant \hat{\theta}_{2k}$$
$$|\mathbf{t_{k,j}}| \leqslant \hat{\theta}_{3(k-j)+2}$$
$$\text{where } \hat{\theta}_n = (1+\mathbf{u})^n - 1$$

Linearity. $\delta_k := \tilde{b}_k - b_k = \text{global error}$

$$\delta_k = \frac{\mathbf{s_k}}{(2k)!\,4^k} - \sum_{j=0}^{k-1} \frac{\delta_j + (b_j + \delta_j)\,\mathbf{t_{k,j}}}{(2k+1-2j)!\,4^{k-j}}$$

Error analysis

$$b_k = \frac{1}{(2k)!\,4^k} - \sum_{j=0}^{k-1} \frac{b_j}{(2k+1-2j)!\,4^{k-j}}, \qquad \tilde{b}_k = \text{computed values}$$

Local error analysis.

$$\tilde{b}_k = \frac{1+\mathbf{s_k}}{(2k)!\,4^k} - \sum_{j=0}^{k-1} \frac{\tilde{b}_j\,(1+\mathbf{t_{k,j}})}{(2k+1-2j)!\,4^{k-j}}$$

$$|\mathbf{s_k}| \leqslant \hat{\theta}_{2k}$$
$$|\mathbf{t_{k,j}}| \leqslant \hat{\theta}_{3(k-j)+2}$$
$$\text{where } \hat{\theta}_n = (1+\mathbf{u})^n - 1$$

Linearity. $\delta_k := \tilde{b}_k - b_k = \text{global error}$

$$\delta_k = \frac{\mathbf{s_k}}{(2k)!\,4^k} - \sum_{j=0}^{k-1} \frac{\delta_j + (b_j + \delta_j)\,\mathbf{t_{k,j}}}{(2k+1-2j)!\,4^{k-j}}$$

Inequation on the global error.

$$\delta(z) \ll \check{S}(z)\,\tilde{C}(z) + \check{S}(z)\,\tilde{S}(z)\,{}^\sharp\delta(z) + \check{S}(z)\,\tilde{S}(z)\,{}^\sharp b(z) \qquad {}^\sharp f(z) = \sum_k |f_k|\,z^k$$

where

$$C(z) = \cosh(\sqrt{z}/2), \qquad S(z) = (\sqrt{z}/2)^{-1}\sinh(\sqrt{z}/2), \qquad \check{S}(z) = \frac{(\sqrt{z}/2)}{\sin(\sqrt{z}/2)},$$

$$\tilde{C}(z) = C(a^2 z) - C(z), \qquad \tilde{S}(z) = S(a^4 z) - S(z) - (a^2 - 1) \qquad \text{with } a = 1 + \mathbf{u}$$

Error analysis

$$b_k = \frac{1}{(2\,k)!\,4^k} - \sum_{j=0}^{k-1} \frac{b_j}{(2\,k+1-2\,j)!\,4^{k-j}}, \qquad \tilde{b}_k = \text{computed values}$$

Local error analysis.

$$\tilde{b}_k = \frac{1+\mathbf{s_k}}{(2\,k)!\,4^k} - \sum_{j=0}^{k-1} \frac{\tilde{b}_j\,(1+\mathbf{t_{k,j}})}{(2\,k+1-2\,j)!\,4^{k-j}}$$

$$|\mathbf{s_k}| \leqslant \hat{\theta}_{2k}$$

$$|\mathbf{t_{k,j}}| \leqslant \hat{\theta}_{3(k-j)+2}$$

where $\hat{\theta}_n = (1+\mathbf{u})^n - 1$

Linearity. $\delta_k := \tilde{b}_k - b_k = \text{global error}$

$$\delta_k = \frac{\mathbf{s_k}}{(2\,k)!\,4^k} - \sum_{j=0}^{k-1} \frac{\delta_j + (\,b_j + \delta_j\,)\,\mathbf{t_{k,j}}}{(2\,k+1-2\,j)!\,4^{k-j}}$$

Inequation on the global error.

$$\delta(z) \ll \check{S}(z)\,\tilde{C}(z) + \check{S}(z)\,\tilde{S}(z)\,{}^\sharp\delta(z) + \check{S}(z)\,\tilde{S}(z)\,{}^\sharp b(z) \qquad {}^\sharp f(z) = \sum_k |f_k|\,z^k$$

where

$$C(z) = \cosh(\sqrt{z}/2), \qquad S(z) = (\sqrt{z}/2)^{-1}\sinh(\sqrt{z}/2), \qquad \check{S}(z) = \frac{(\sqrt{z}/2)}{\sin(\sqrt{z}/2)},$$

$$\tilde{C}(z) = C(a^2\,z) - C(z), \qquad \tilde{S}(z) = S(a^4\,z) - S(z) - (a^2 - 1) \qquad \text{with } a = 1 + \mathbf{u}$$

First-order bound

$$\delta(z) \ll \check{S}(z)\,\tilde{C}(z) + \check{S}(z)\,\tilde{S}(z)\,{}^{\sharp}\mathfrak{b}(z) + \check{S}(z)\,\tilde{S}(z)\,{}^{\sharp}\delta(z)$$

'Explicit' majorant. By the first lemma on majorizing equations

$$\delta(z) \ll \frac{\check{S}(z)\,\tilde{C}(z) + \check{S}(z)\,\tilde{S}(z)\,{}^{\sharp}\mathfrak{b}(z)}{1 - \check{S}(z)\,\tilde{S}(z)} \quad =: \hat{\delta}(z)$$

First-order bound

$$\delta(z) \ll \check{S}(z)\,\tilde{C}(z) + \check{S}(z)\,\tilde{S}(z)\,{}^\sharp b(z) + \check{S}(z)\,\tilde{S}(z)\,{}^\sharp\delta(z)$$

'Explicit' majorant. By the first lemma on majorizing equations

$$\delta(z) \ll \frac{\check{S}(z)\,\tilde{C}(z) + \check{S}(z)\,\tilde{S}(z)\,{}^\sharp b(z)}{1 - \check{S}(z)\,\tilde{S}(z)} \quad =: \hat{\delta}(z)$$

Asymptotic behavior. Series notation → computer algebra

$$\hat{\delta}(z) = \left(\frac{2\,(1 - \cosh w)\cos(w)}{w^{-2}\sin(w)^2} + \frac{4\,(\cosh w - 1) + w \sinh w}{w^{-1}\sin w} \right) \mathbf{u} + O(\mathbf{u}^2)$$

$$w = \sqrt{z}/2$$

Unique dominant pole at $z = 4\,\pi^2$,
multiplicity (w.r.t. z) $= 2$

$\Rightarrow \hat{\delta}_k = O(k\,(2\,\pi)^{-2k}) \cdot u + O(\mathbf{u}^2)$

$\Rightarrow \eta_k =$ "$O(k \cdot \mathbf{u})$"

A 'hard' bound

$$\delta(z) = \hat{\delta}(z) \ll \frac{\check{S}(z)\,\tilde{C}(z) + \check{S}(z)\,\tilde{S}(z)\,{}^{\sharp}b(z)}{1 - \check{S}(z)\,\tilde{S}(z)}$$

Controlling the dominant pole.

Suppose $\mathbf{u} \leqslant 2^{-16}$.

Then $\hat{\delta}(z)$ has a pole at $\gamma = \left(\frac{2\,\pi}{1 + \varphi(\mathbf{u})}\right)^2$ where $0 \leqslant \varphi(\mathbf{u}) \leqslant 2\,(\cosh \pi - 1)\,\mathbf{u}$.

This is the only pole with $|z| < 153.7 \approx (3.9\,\pi)^2$.

(A little analysis + comparison with the limiting case using Rouché's theorem.)

A 'hard' bound

$$\delta(z) = \hat{\delta}(z) \ll \frac{\check{S}(z)\,\tilde{C}(z) + \check{S}(z)\,\tilde{S}(z)\,{}^{\sharp}b(z)}{1 - \check{S}(z)\,\tilde{S}(z)}$$

Controlling the dominant pole.

Suppose $\mathbf{u} \leqslant 2^{-16}$.

Then $\hat{\delta}(z)$ has a pole at $\gamma = \left(\frac{2\pi}{1 + \varphi(\mathbf{u})}\right)^2$ where $0 \leqslant \varphi(\mathbf{u}) \leqslant 2\,(\cosh \pi - 1)\,\mathbf{u}$.

This is the only pole with $|z| < 153.7 \approx (3.9\,\pi)^2$.

(A little analysis + comparison with the limiting case using Rouché's theorem.)

Symbolic-numeric estimate.

$$\hat{\delta}(z) = \frac{2 \text{ explicit } \mathbf{R}(\mathbf{u})}{1 - z/\gamma} - \frac{2}{1 - z/(2\pi)^2} + \text{analytic for } |z| < 153.7$$

$$\hat{\delta}(z) \ll \frac{2\,|\,\mathbf{R}(\mathbf{u}) - 1|}{1 - z/\gamma} + \frac{\text{explicit and } O(\mathbf{u})}{(1 - z/\gamma)^2} + \frac{\sup_{|z=\lambda\gamma|} \text{analytic}}{1 - z/(\lambda\gamma)}$$

Cauchy's formula + interval arithmetic.

A 'hard' bound

$$\delta(z) = \hat{\delta}(z) \ll \frac{\check{S}(z)\,\tilde{C}(z) + \check{S}(z)\,\tilde{S}(z)\,{}^{\sharp}b(z)}{1 - \check{S}(z)\,\tilde{S}(z)}$$

Controlling the dominant pole.

Suppose $\mathbf{u} \leqslant 2^{-16}$.

Then $\hat{\delta}(z)$ has a pole at $\gamma = \left(\frac{2\pi}{1 + \varphi(\mathbf{u})}\right)^2$ where $0 \leqslant \varphi(\mathbf{u}) \leqslant 2\,(\cosh \pi - 1)\,\mathbf{u}$.

This is the only pole with $|z| < 153.7 \approx (3.9\,\pi)^2$.

(A little analysis + comparison with the limiting case using Rouché's theorem.)

Symbolic-numeric estimate.

$$\hat{\delta}(z) = \frac{2\ \text{explicit}\ \mathbf{R}(\mathbf{u})}{1 - z/\gamma} - \frac{2}{1 - z/(2\pi)^2} + \text{analytic for } |z| < 153.7$$

$$\hat{\delta}(z) \ll \frac{2\,|\ \mathbf{R}(\mathbf{u})\ - 1|}{1 - z/\gamma} + \frac{\text{explicit and } O(\mathbf{u})}{(1 - z/\gamma)^2} + \frac{\sup_{|z = \lambda\gamma|}\ \text{analytic}}{1 - z/(\lambda\gamma)}$$

Cauchy's formula + interval arithmetic.

Scaled Bernoulli numbers: conclusion

$$b(z) = \frac{\sqrt{z}/2}{\tanh(\sqrt{z}/2)} \qquad\qquad b_k = \frac{1}{(2k)!\,4^k} - \sum_{j=0}^{k-1} \frac{b_j}{(2k+1-2j)!\,4^{k-j}}$$

$$\tilde{b}_k = b_k\,(1+\eta_k)$$

Theorem. The total relative error satisfies

$$|\eta_k| \leqslant (1+21.2\,\mathbf{u})^k\,(1.1\,k+446)\,\mathbf{u}$$

Corollary. Assuming $u < 2^{-16}$ and $43\,k\,u \leqslant 1$, one has $|\eta_k| \leqslant (3\,k+1213)\,\mathbf{u}$.

Conclusion

Error analyses of linear recurrences can (should!) use generating series

- Local errors → global errors via exact expressions or equations
- Cauchy majorants
- Analytic methods

- Legendre polynomials
- General D-finite functions
- Bernoulli numbers

- Other algorithms for D-finite functions, e.g., $O(n\,M(d)/d)$
- Tighter bounds in practice
- Backward recurrence schemes
- Orthogonal polynomials, numerical integration schemes, …

Conclusion

Error analyses of linear recurrences can (should!) use generating series

- Local errors → global errors via exact expressions or equations
- Cauchy majorants
- Analytic methods

- Legendre polynomials
- General D-finite functions
- Bernoulli numbers

- Other algorithms for D-finite functions, e.g., $O(n\,M(d)/d)$
- Tighter bounds in practice
- Backward recurrence schemes
- Orthogonal polynomials, numerical integration schemes, …

Thank you!

Credits