Ball arithmetic

Joris van der Hoeven

CNRS, École polytechnique

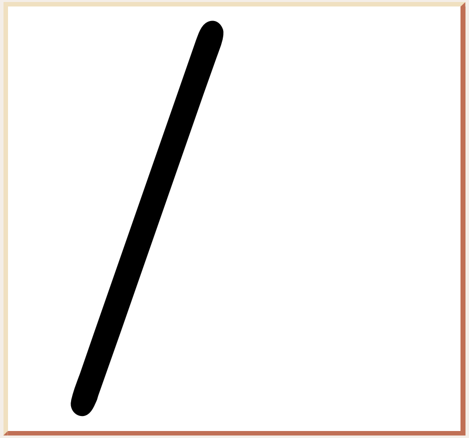
Luminy, 2010

http://www.TeXmacs.jp

Ball arithmetic

Joris van der Hoeven

CNRS, École polytechnique



Luminy, 2010

`http://www.TeXmacs.jp`

Objectives

Long term program:

- Understand the practical complexity of computable analysis.
- Design fast algorithms for common analytic computations.
- Mathemagix: a free “computer analysis” system.
- Make computable analysis “competitive” with numerical analysis.

Existing software:

- Computable real numbers. Mostly restricted to numbers
- Interval arithmetic. Mostly double precision
- Fast arithmetic. Applications mostly algebraic or discrete
- Other libraries. Mostly specialized, e.g. find roots of $P \in \mathbb{C}[z]$

The numerical hierarchy

Mathematical level

$x \in \mathbb{R}^{\text{com}} \leftrightsquigarrow \exists \check{x} \colon \varepsilon \in \mathbb{Q}^{>} \overset{\text{comp}}{\rightarrow} \tilde{x} \in \mathbb{Q}$, $|\tilde{x} - x| \leqslant \varepsilon$

Reliable level

$x^{\vdash\!\dashv} = [x_l, x_r] \in \mathcal{I}(\mathbb{D})$

Numerical level

$x \in \mathbb{D}_{p,e} = \{-2^{p-1}, \ldots, 0, \ldots, 2^{p-1} - 1\}\, 2^{\{-2^{e-1}, \ldots, 0, \ldots, 2^{e-1} - 1\}}$

$\mathbb{D}_{52,12} \leftrightsquigarrow$ “double precision IEEE 784 number”

↯ Correct rounding

Arithmetic level

$x \in \mathbb{Z}$ or $x \in \{-2^{p-1}, \ldots, 0, \ldots, 2^{p-1} - 1\}$

The numerical hierarchy

Mathematical level

$x \in \mathbb{R}^{\text{com}} \leftrightsquigarrow \exists \check{x}\colon \varepsilon \in \mathbb{D}^{>} \xmapsto{\text{comp}} \tilde{x} \in \mathbb{D}, \ |\tilde{x} - x| \leqslant \varepsilon$ $(\mathbb{D} = \mathbb{Z}\, 2^{\mathbb{Z}})$

Reliable level

$x^{\vdash\dashv} = [x_l, x_r] \in \mathcal{I}(\mathbb{D})$

Numerical level

$x \in \mathbb{D}_{p,e} = \{-2^{p-1}, \ldots, 0, \ldots, 2^{p-1} - 1\}\, 2^{\{-2^{e-1}, \ldots, 0, \ldots, 2^{e-1} - 1\}}$

$\mathbb{D}_{52,12} \leftrightsquigarrow$ "double precision IEEE 784 number"

↯ Correct rounding

Arithmetic level

$x \in \mathbb{Z}$ or $x \in \{-2^{p-1}, \ldots, 0, \ldots, 2^{p-1} - 1\}$

The numerical hierarchy

Mathematical level

$x \in \mathbb{R}^{\text{com}} \leftrightsquigarrow \exists \check{x} \colon \varepsilon \in \mathbb{D}^{>} \overset{\text{comp}}{\longmapsto} \tilde{x} \in \mathbb{D}$, $|\tilde{x} - x| \leqslant \varepsilon$ $(\mathbb{D} = \mathbb{Z}\, 2^{\mathbb{Z}})$

Reliable level

$x^{\circ} = \{\tilde{x} \in \mathbb{R} \colon |\tilde{x} - x_c| \leqslant x_r\} \in \mathcal{B}(\mathbb{D}, \mathbb{D})$

Numerical level

$x \in \mathbb{D}_{p,e} = \{-2^{p-1}, \ldots, 0, \ldots, 2^{p-1} - 1\}\, 2^{\{-2^{e-1}, \ldots, 0, \ldots, 2^{e-1} - 1\}}$

$\mathbb{D}_{52,12}$ $\leftrightsquigarrow$ “double precision IEEE 784 number”

↯ Correct rounding

Arithmetic level

$x \in \mathbb{Z}$ or $x \in \{-2^{p-1}, \ldots, 0, \ldots, 2^{p-1} - 1\}$

The numerical hierarchy

Mathematical level

$x\in\mathbb{R}^{\mathrm{com}} \leftrightsquigarrow \exists\check{x}\colon\varepsilon\in\mathbb{D}^{>}\overset{\mathrm{comp}}{\longmapsto}\tilde{x}\in\mathbb{D}$, $|\tilde{x}-x|\leqslant\varepsilon$ $(\mathbb{D}=\mathbb{Z}\,2^{\mathbb{Z}})$

Reliable level

$x^{\circ}=\{\tilde{x}\in\mathbb{R}:|\tilde{x}-x_c|\leqslant x_r\}\in\mathcal{B}(\mathbb{D},\mathbb{D})$

$f\colon\mathbb{R}\rightarrow\mathbb{R} \rightsquigarrow f^{\circ}\colon\mathcal{B}(\mathbb{R},\mathbb{R})\rightarrow\mathcal{B}(\mathbb{R},\mathbb{R})$, $\forall\tilde{x}\in x^{\circ}$, $f(\tilde{x})\in f^{\circ}(x^{\circ})$.

Numerical level

$x\in\mathbb{D}_{p,e}=\{-2^{p-1},\ldots,0,\ldots,2^{p-1}-1\}\,2^{\{-2^{e-1},\ldots,0,\ldots,2^{e-1}-1\}}$

$\mathbb{D}_{52,12} \leftrightsquigarrow$ “double precision IEEE 784 number”

↯ Correct rounding

Arithmetic level

$x\in\mathbb{Z}$ or $x\in\{-2^{p-1},\ldots,0,\ldots,2^{p-1}-1\}$

The numerical hierarchy

Mathematical level

$x \in \mathbb{R}^{\text{com}} \leftrightsquigarrow \exists \check{x} \colon \varepsilon \in \mathbb{D}^{>} \xmapsto{\text{comp}} \tilde{x} \in \mathbb{D}, \ |\tilde{x} - x| \leqslant \varepsilon$ $(\mathbb{D} = \mathbb{Z}\, 2^{\mathbb{Z}})$

Reliable level

$x^{\circ} = \mathcal{B}(x_c, x_r) = \{\tilde{x} \in \mathbb{R} \colon |\tilde{x} - x_c| \leqslant x_r\} \in \mathcal{B}(\mathbb{D}, \mathbb{D})$

$f \colon \mathbb{R}^{\text{com}} \to \mathbb{R}^{\text{com}} \rightsquigarrow f^{\circ} \colon \mathcal{B}(\mathbb{D}, \mathbb{D}) \to \mathcal{B}(\mathbb{D}, \mathbb{D}), \ \forall \tilde{x} \in x^{\circ}, \ f(\tilde{x}) \in f^{\circ}(x^{\circ}).$

$x \in \mathbb{R}^{\text{com}} \leftrightsquigarrow \exists \check{x} \colon n \in \mathbb{N} \xmapsto{\text{comp}} \check{x}_n \in \mathcal{B}(\mathbb{D}, \mathbb{D}), \ x \in \check{x}_n, \ \lim_{n \to \infty} (\check{x}_n)_r = 0$

Numerical level

$x \in \mathbb{D}_{p,e} = \{-2^{p-1}, \ldots, 0, \ldots, 2^{p-1} - 1\}\, 2^{\{-2^{e-1}, \ldots, 0, \ldots, 2^{e-1} - 1\}}$

$\mathbb{D}_{52,12} \leftrightsquigarrow$ “double precision IEEE 784 number”

↯ Correct rounding

Arithmetic level

$x \in \mathbb{Z}$ or $x \in \{-2^{p-1}, \ldots, 0, \ldots, 2^{p-1} - 1\}$

Matrix multiplication

Mathematical level

$M, N \in (\mathbb{R}^{\mathrm{com}})^{n \times n}$, $M N$?

Reliable level

$M^{\circ}, N^{\circ} \in \mathcal{B}(\mathbb{D}, \mathbb{D})^{n \times n}$, $M N$?

Numerical level

$M, N \in \mathbb{D}_{p,e}^{n \times n}$, $M N$?

- $\mathbb{D}_{p,e} = \mathbb{D}_{52,12}$ $\rightsquigarrow$ Blas
- $p > 52$, $\mathbb{D}^{n \times n} \cong \mathbb{Z}^{n \times n} \, 2^{\mathbb{Z}}$

Arithmetic level

$M, N \in \mathbb{Z}^{n \times n}$, $M N$

FFT or Chinese remaindering depending on $\frac{n}{p}$, with p = bit precision

Matrix multiplication

Mathematical level

$M, N \in (\mathbb{R}^{\text{com}})^{n \times n} \cong (\mathbb{R}^{n \times n})^{\text{com}}$

$M \in (\mathbb{R}^{\text{com}})^{n \times n} \leftrightsquigarrow \exists \check{M} \colon \varepsilon \in \mathbb{D}^{>} \xmapsto{\text{comp}} \tilde{M} \in \mathbb{D}^{n \times n}$, $|\tilde{M} - M| \leqslant \varepsilon$, sup-norm on $\mathbb{R}^{n \times n}$

Reliable level

$M^{\circ}, N^{\circ} \in \mathcal{B}(\mathbb{D}, \mathbb{D})^{n \times n}$, $M N$?

Numerical level

$M, N \in \mathbb{D}_{p,e}^{n \times n}$, $M N$?

- $\mathbb{D}_{p,e} = \mathbb{D}_{52,12} \rightsquigarrow$ Blas
- $p > 52$, $\mathbb{D}^{n \times n} \cong \mathbb{Z}^{n \times n} \, 2^{\mathbb{Z}}$

Arithmetic level

$M, N \in \mathbb{Z}^{n \times n}$, $M N$

FFT or Chinese remaindering depending on $\frac{n}{p}$, with p = bit precision

Matrix multiplication

Mathematical level

$M, N \in (\mathbb{R}^{\text{com}})^{n \times n} \cong (\mathbb{R}^{n \times n})^{\text{com}}$

$M \in (\mathbb{R}^{\text{com}})^{n \times n} \leftrightsquigarrow \exists \check{M} : n \in \mathbb{N} \xmapsto{\text{comp}} \check{M}_n \in \mathcal{B}(\mathbb{D}, \mathbb{D})^{n \times n}$

Reliable level

$M^\circ, N^\circ \in \mathcal{B}(\mathbb{D}, \mathbb{D})^{n \times n}$, $M N$?

Numerical level

$M, N \in \mathbb{D}_{p,e}^{n \times n}$, $M N$?

- $\mathbb{D}_{p,e} = \mathbb{D}_{52,12} \rightsquigarrow$ Blas
- $p > 52$, $\mathbb{D}^{n \times n} \cong \mathbb{Z}^{n \times n} 2^{\mathbb{Z}}$

Arithmetic level

$M, N \in \mathbb{Z}^{n \times n}$, $M N$

FFT or Chinese remaindering depending on $\frac{n}{p}$, with p = bit precision

Matrix multiplication

Mathematical level

$M, N \in (\mathbb{R}^{\text{com}})^{n \times n} \cong (\mathbb{R}^{n \times n})^{\text{com}}$

$M \in (\mathbb{R}^{\text{com}})^{n \times n} \leftrightsquigarrow \exists \check{M} \colon n \in \mathbb{N} \xmapsto{\text{comp}} \check{M}_n \in \mathcal{B}(\mathbb{D}, \mathbb{D})^{n \times n}$

Reliable level

$M^\circ, N^\circ \in \mathcal{B}(\mathbb{D}, \mathbb{D})^{n \times n} \cong \mathcal{B}(\mathbb{D}^{n \times n}, \mathbb{D}^{n \times n}) \cong \mathcal{B}(\mathbb{D}^{n \times n}, \mathbb{D})$

Numerical level

$M, N \in \mathbb{D}_{p,e}^{n \times n}$, $M N$?

- $\mathbb{D}_{p,e} = \mathbb{D}_{52,12} \rightsquigarrow$ Blas
- $p > 52$, $\mathbb{D}^{n \times n} \cong \mathbb{Z}^{n \times n} 2^{\mathbb{Z}}$

Arithmetic level

$M, N \in \mathbb{Z}^{n \times n}$, $M N$

FFT or Chinese remaindering depending on $\frac{n}{p}$, with $p =$ bit precision

Intervals versus balls

	Intervals	**Balls**
Representation	$[2, \infty]$	$\{z \in \mathbb{C} : \lvert z \rvert \leqslant 1\}$
Hardware support	IEEE 754 $x +^{\downarrow} y = -((-x) +^{\uparrow} (-y))$	not yet $\Delta(y) = (\lvert y \rvert +^{\uparrow} 2^{-2^{e-1}}) \cdot^{\uparrow} 2^{-p}$
Efficiency	end-points at full precision multiplication $\rightsquigarrow$ branching	radius at single precision completely vectorial
Standardization	correct rounding	explicit formulas for operations
Standardness	computer science	mathematics
Recommendation	algorithms that require subdivision of space	approximation of numbers
Predicates	values in $\mathcal{I}_{\{0,1\}}$	values in $\{0, 1\}$ also, $=: (\mathbb{R}^{\mathrm{com}})^2 \to \{0, 1\}^{\mathrm{rcom}}$

The wrapping effect and the radius type

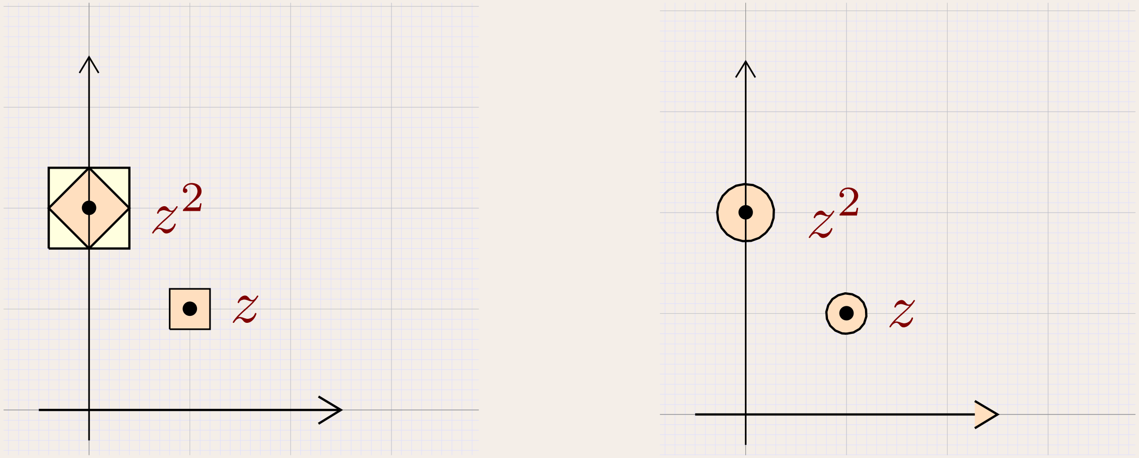


Figure. Illustration of the computation of z^2 using interval and ball arithmetic, for $z = 1 + \mathrm{i}$.

⇑ Intervals

```
Mmx] use "analyziz";
Mmx] z1 == complex (polynomial ball (1.0), polynomial ball (1.0))
Mmx] pow (u, n) == if n=1 then u else u * pow (u, n-1);
Mmx] [ pow (z1, 8*i) || i in 1 to 10 ]
Mmx]
```

⇑ Balls

```
Mmx] z2 == ball complex (1.0, 1.0)
Mmx] [ pow (z2, 8*i) || i in 1 to 10 ]
Mmx]
```

Reducing the wrapping effect

Use divide and conquer algorithms

I.e. keep depth of the arithmetic circuit small

$$z^n = z^{\lfloor \frac{n}{2} \rfloor} z^{\lceil \frac{n}{2} \rceil}$$

```
Mmx] binpow (u, n) ==
       if n = 1 then u else binpow (u, n div 2) * binpow (u, (n+1) div 2);
Mmx] [ pow (z1, 8*i), binpow (z1, 8*i) || i in 1 to 10 ]
Mmx]
```

Reducing the wrapping effect

Use divide and conquer algorithms

I.e. keep depth of the arithmetic circuit small

$$z^n = z^{\lfloor \frac{n}{2} \rfloor} z^{\lceil \frac{n}{2} \rceil}$$

Changes of coordinates

Enclose vectors by products $M V^{\vdash\!\dashv}$, $M \in \mathbb{D}^{n \times n}$, $V \in \mathcal{I}_{\mathbb{D}}^{n}$

Useful for differential equations, e.g.

$$f' = \begin{pmatrix} \cos' \\ \sin' \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \cos \\ \sin \end{pmatrix} = A f$$

Works because rounding errors all accumulate at the same side:

$$\begin{aligned} f(t_0) &= M_0 V_0^{\vdash\!\dashv} \\ f(t_1) &= \Delta_{t_0,t_1} (M_0 V_0^{\vdash\!\dashv}) = \widetilde{\Delta_{t_0,t_1} M_0} (V_0^{\vdash\!\dashv} + \varepsilon^{\vdash\!\dashv}) \end{aligned}$$

Not suited for general purpose complex ball arithmetic, because of sums

Quality and condition number

Condition number

$$\begin{aligned} f\colon \mathbb{R}^n &\to \mathbb{R}^m \\ \kappa_f(x) &= \lim_{\varepsilon\to 0}\sup_{\|\delta\|=\varepsilon} \frac{\|f(x+\delta)-f(x)\|}{\|f(x)\|}\Bigg/\frac{\|\delta\|}{\|x\|} \end{aligned}$$

Linear algebra

$$\begin{aligned} \kappa(M) &= \kappa_{M^{-1}.}(x) \\ &= \|M\|\|M^{-1}\|. \end{aligned}$$

operator norms

Integration of a dynamical system

$$\begin{aligned} f' &= \Phi(f) \\ \kappa(\Phi, f(0), 0, t) &= \max_{0\leqslant u\leqslant t} \kappa(\Delta_{0,u}(f(0))) \end{aligned}$$

Quality and condition number

Condition number

$$\begin{aligned} f\colon \mathbb{R}^n &\to \mathbb{R}^m \\ \kappa_f(x) &= \lim_{\varepsilon\to 0} \sup_{\|\delta\|=\varepsilon} \frac{\|f(x+\delta)-f(x)\|}{\|f(x)\|} \Big/ \frac{\|\delta\|}{\|x\|} \end{aligned}$$

Linear algebra

$$\begin{aligned} \kappa(M) &= \kappa_{M^{-1}.}(x) \\ &= \|M\|\|M^{-1}\|. \end{aligned}$$

operator norms

Integration of a dynamical system

$$\begin{aligned} f' &= \Phi(f) \\ \kappa(\Phi, f(0), 0, t) &= \max_{0\leqslant u\leqslant v\leqslant t} \kappa(\Delta_{u,v}(f(u)) \end{aligned}$$

Efficient ball arithmetic

Real numbers

$$\mathcal{B}(x_c, x_r) \cdot \mathcal{B}(y_c, y_r) \;=\; \mathcal{B}(x_c \cdot^{\updownarrow} y_c, (|x_c| +^{\uparrow} x_r) \cdot^{\uparrow} y_r +^{\uparrow} x_r \cdot^{\uparrow} |y_c| +^{\uparrow} \Delta(x_c \cdot^{\updownarrow} y_c))$$

Matrices

$$M^{\circ} N^{\circ} \;=\; \mathcal{B}(M_c N_c, (|M_c| + M_r) N_r + M_r |N_c| + \Delta)$$

Preconditioning

$$\begin{pmatrix} 1.0000e10 & 2.1000e5 & 6.3333e8 \\ 2.1111e16 & 1.1428e10 & 3.9876e12 \\ 2.2187e7 & 1.2134e2 & 9.8765e5 \end{pmatrix} \begin{pmatrix} 1.2222e10 & 4.3245e6 \\ 1.2345e13 & 2.3456e8 \\ 2.4325e12 & 5.3235e8 \end{pmatrix}$$

Subdivision

$$\begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} \\ a_{10} & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{20} & a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{30} & a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{40} & a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{50} & a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{pmatrix}$$

Efficient ball arithmetic

Real numbers

$$\mathcal{B}(x_c, x_r) \cdot \mathcal{B}(y_c, y_r) \; = \; \mathcal{B}(x_c \cdot^{\updownarrow} y_c, (|x_c| +^{\uparrow} x_r) \cdot^{\uparrow} y_r +^{\uparrow} x_r \cdot^{\uparrow} |y_c| +^{\uparrow} \Delta(x_c \cdot^{\updownarrow} y_c))$$

Matrices

$$\begin{aligned} M^{\circ} N^{\circ} &= \mathcal{B}(M_c N_c, [(\|M_c\| + \|M_r\|) \|N_r\| + \|M_r\| \|N_c\| + \Delta] \, \Omega) \\ \Omega &= \begin{pmatrix} \mathcal{B}(0,1) & \cdots & \mathcal{B}(0,1) \\ \vdots & & \vdots \\ \mathcal{B}(0,1) & \cdots & \mathcal{B}(0,1) \end{pmatrix} \end{aligned}$$

Preconditioning

$$\begin{pmatrix} 1.0000e10 & 2.1000e5 & 6.3333e8 \\ 2.1111e16 & 1.1428e10 & 3.9876e12 \\ 2.2187e7 & 1.2134e2 & 9.8765e5 \end{pmatrix} \begin{pmatrix} 1.2222e10 & 4.3245e6 \\ 1.2345e13 & 2.3456e8 \\ 2.4325e12 & 5.3235e8 \end{pmatrix}$$

Subdivision

$$\begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} \\ a_{10} & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{20} & a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{30} & a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{40} & a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{50} & a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{pmatrix}$$

Efficient ball arithmetic

Real numbers

$$\mathcal{B}(x_c, x_r) \cdot \mathcal{B}(y_c, y_r) \quad = \quad \mathcal{B}(x_c \cdot^{\updownarrow} y_c, (|x_c| +^{\uparrow} x_r) \cdot^{\uparrow} y_r +^{\uparrow} x_r \cdot^{\uparrow} |y_c| +^{\uparrow} \Delta(x_c \cdot^{\updownarrow} y_c))$$

Matrices

$$(M^{\circ} N^{\circ})_{ij} \quad = \quad \mathcal{B}((M_c N_c)_{ij}, (\|M_c\| + \|M_r\|) \|N_r\| + \|M_r\| \|N_c\| + \Delta)$$

Preconditioning

$$\begin{pmatrix} 1.0000e10 & 2.1000e5 & 6.3333e8 \\ 2.1111e16 & 1.1428e10 & 3.9876e12 \\ 2.2187e7 & 1.2134e2 & 9.8765e5 \end{pmatrix} \begin{pmatrix} 1.2222e10 & 4.3245e6 \\ 1.2345e13 & 2.3456e8 \\ 2.4325e12 & 5.3235e8 \end{pmatrix}$$

Subdivision

$$\begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} \\ a_{10} & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{20} & a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{30} & a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{40} & a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{50} & a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{pmatrix}$$

Efficient ball arithmetic

Real numbers

$$\mathcal{B}(x_c, x_r) \cdot \mathcal{B}(y_c, y_r) \;=\; \mathcal{B}(x_c \cdot^{\updownarrow} y_c, (|x_c| +^{\uparrow} x_r) \cdot^{\uparrow} y_r +^{\uparrow} x_r \cdot^{\uparrow} |y_c| +^{\uparrow} \Delta(x_c \cdot^{\updownarrow} y_c))$$

Matrices

$$(M^{\circ} N^{\circ})_{ij} \;=\; \mathcal{B}((M_c N_c)_{ij}, (\|(M_c)_{i\cdot}\| + \|(M_r)_{i\cdot}\|) \|(N_r)_{\cdot j}\| + \|(M_r)_{i\cdot}\| \|(N_c)_{\cdot j}\| + \Delta)$$

Preconditioning

$$\begin{pmatrix} 1.0000e10 & 2.1000e5 & 6.3333e8 \\ 2.1111e16 & 1.1428e10 & 3.9876e12 \\ 2.2187e7 & 1.2134e2 & 9.8765e5 \end{pmatrix} \begin{pmatrix} 1.2222e10 & 4.3245e6 \\ 1.2345e13 & 2.3456e8 \\ 2.4325e12 & 5.3235e8 \end{pmatrix}$$

Subdivision

$$\begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} \\ a_{10} & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{20} & a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{30} & a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{40} & a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{50} & a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{pmatrix}$$

Efficient ball arithmetic

Real numbers

$$\mathcal{B}(x_c, x_r) \cdot \mathcal{B}(y_c, y_r) \;=\; \mathcal{B}(x_c \cdot^{\updownarrow} y_c, (|x_c| +^{\uparrow} x_r) \cdot^{\uparrow} y_r +^{\uparrow} x_r \cdot^{\uparrow} |y_c| +^{\uparrow} \Delta(x_c \cdot^{\updownarrow} y_c))$$

Matrices

$$(M^\circ N^\circ)_{ij} \;=\; \mathcal{B}((M_c N_c)_{ij}, (\|(M_c)_{i\cdot}\| + \|(M_r)_{i\cdot}\|)\, \|(N_r)_{\cdot j}\| + \|(M_r)_{i\cdot}\| \|(N_c)_{\cdot j}\| + \Delta)$$

Preconditioning

$$\begin{pmatrix} 1.0000e10 & 2.1000e9 & 6.3333e10 \\ 2.1111e16 & 1.1428e13 & 3.9876e14 \\ 2.2187e7 & 1.2134e6 & 9.8765e7 \end{pmatrix} \begin{pmatrix} 1.2222e10 & 4.3245e6 \\ 1.2345e9 & 2.3456e4 \\ 2.4325e10 & 5.3235e6 \end{pmatrix}$$

Subdivision

$$\begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} \\ a_{10} & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{20} & a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{30} & a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{40} & a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{50} & a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{pmatrix}$$

Efficient ball arithmetic

Real numbers

$$\mathcal{B}(x_c, x_r) \cdot \mathcal{B}(y_c, y_r) \;=\; \mathcal{B}(x_c \cdot^{\updownarrow} y_c, (|x_c| +^{\uparrow} x_r) \cdot^{\uparrow} y_r +^{\uparrow} x_r \cdot^{\uparrow} |y_c| +^{\uparrow} \Delta(x_c \cdot^{\updownarrow} y_c))$$

Matrices

$$(M^{\circ} N^{\circ})_{ij} \;=\; \mathcal{B}((M_c N_c)_{ij}, (\|(M_c)_{i\cdot}\| + \|(M_r)_{i\cdot}\|) \|(N_r)_{\cdot j}\| + \|(M_r)_{i\cdot}\| \|(N_c)_{\cdot j}\| + \Delta)$$

Preconditioning

$$\begin{pmatrix} 1.0000e10 & 2.1000e9 & 6.3333e10 \\ 2.1111e16 & 1.1428e13 & 3.9876e14 \\ 2.2187e7 & 1.2134e6 & 9.8765e7 \end{pmatrix} \begin{pmatrix} 1.2222e10 & 4.3245e6 \\ 1.2345e9 & 2.3456e4 \\ 2.4325e10 & 5.3235e6 \end{pmatrix}$$

Subdivision

$$\left(\begin{array}{ccc|ccc} a_{00} & a_{01} & a_{02} & a_{03} & a_{04} & a_{05} \\ a_{10} & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{20} & a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ \hline a_{30} & a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{40} & a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{50} & a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{array}\right)$$

Hansen's method

Inversion of a matrix

$$\begin{aligned}
M^\circ &\in \mathcal{B}(\mathbb{D},\mathbb{D})^{n\times n}\\
N_c &:= \mathrm{fl}_p(M_c^{-1})\\
E^\circ &:= 1 - M^\circ\, N_c\\
(1-E^\circ)^{-1} &:= 1 + E^\circ + \frac{\|E^\circ\|^2}{1-\|E^\circ\|}\begin{pmatrix} \mathcal{B}(0,1) & \cdots & \mathcal{B}(0,1)\\ \vdots & & \vdots\\ \mathcal{B}(0,1) & \cdots & \mathcal{B}(0,1)\end{pmatrix}\\
(M^\circ)^{-1} &:= N_c\,(1-E^\circ)^{-1}
\end{aligned}$$

Large p

$$\|E^\circ\| \approx \kappa(M)\,2^{-p}$$

Increased quality

$$M = \begin{pmatrix} 1 & K & & & \\ & 1 & K & & \\ & & \ddots & \ddots & \\ & & & 1 & K\\ & & & & 1\end{pmatrix}$$

Hansen's method

Inversion of a matrix

$$\begin{aligned} M^\circ &\in \mathcal{B}(\mathbb{D},\mathbb{D})^{n\times n} \\ N_c &:= \mathrm{fl}_p(M_c^{-1}) \\ E^\circ &:= 1 - M^\circ N_c \\ (1-E^\circ)^{-1} &:= 1 + E^\circ + \frac{\|E^\circ\|^2}{1-\|E^\circ\|} \begin{pmatrix} \mathcal{B}(0,1) & \cdots & \mathcal{B}(0,1) \\ \vdots & & \vdots \\ \mathcal{B}(0,1) & \cdots & \mathcal{B}(0,1) \end{pmatrix} \\ (M^\circ)^{-1} &:= N_c\,(1-E^\circ)^{-1} \end{aligned}$$

Large p

$$\|E^\circ\| \;\approx\; \kappa(M)\,2^{-p}$$

Increased quality

$$(1-E^\circ)^{-1} \;:=\; (1+E^\circ)\,(1+(E^\circ)^2)\cdots(1+(E^\circ)^{2^{p-1}}) + O(\|E^\circ\|^{2^p})$$

Efficient arithmetic in $\mathcal{B}(\mathbb{D},\mathbb{D})[[z]]^{\mathrm{com}}$

⇑ Floating point coefficients at precision $p=128$

```
Mmx] use "analyziz";
Mmx] bit_precision := 128;
Mmx] time_mode? := true;
Mmx] z1 == series (0.0, 1.0);
Mmx] B1 == exp (exp z1 - 1)
Mmx] B1[5000]
```

⇑ Ball coefficients at precision $p=128$

```
Mmx] z2 == series (ball 0.0, ball 1.0);
Mmx] B2 == exp (exp z2 - 1);
Mmx] B2[5000]
```

⇑ Floating point coefficients at precision $p=256$

```
Mmx] bit_precision := 256;
Mmx] z3 == series (0.0, 1.0);
Mmx] B3 == exp (exp z3 - 1);
Mmx] B3[5000]
```

⇑ Ball coefficients at precision $p=256$

```
Mmx] z4 == series (ball 0.0, ball 1.0);
Mmx] B4 == exp (exp z4 - 1);
Mmx] B4[5000]
```

Conclusion

- Tradeoff between efficiency and quality; condition number.
- At high precision, certification should be of neglectible cost.
- High precision $\leftrightarrow$ Algebraic complexity

  Low precision $\leftrightarrow$ Geometry, condition number

Thank you !

Starring...