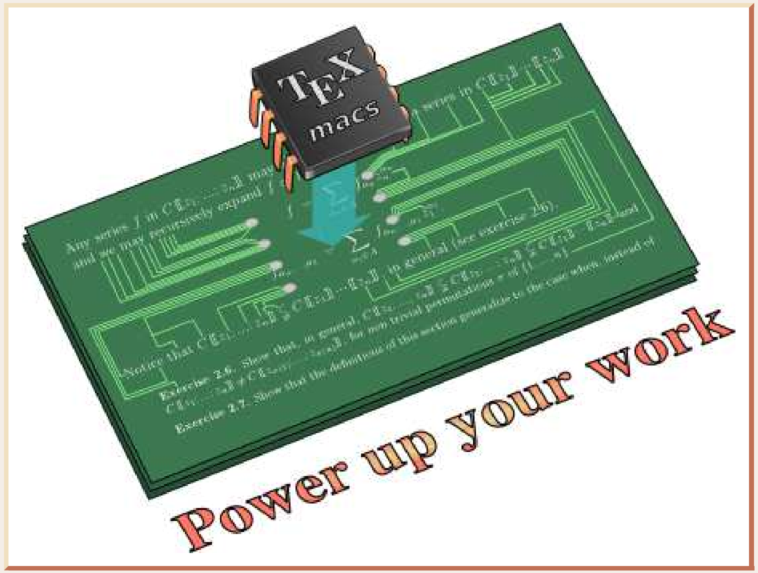
Calcul analytique I

Joris van der Hoeven

CNRS, École polytechnique



JNCF, Luminy, 2011

http://www.TeXmacs.org

Système de calcul formel/analytique

Calcul formel

Calcul mathématiquement exact avec des objets de nature algébrique, arithmétique ou combinatoire.

Analyse numérique

Calcul numérique approximatif avec estimations d'erreur généralement manuelles.

```
maple> 1.000000000000000000000001 - 1;
                                 0.
```

Calcul analytique

Calcul mathématiquement exact avec des objets de nature analytique.

Exemple 1 : intégration certifiée

Pendule

$$\ddot{y} = -\alpha \sin y,$$

Pour $\alpha = 1$, $y(0) = 0$, $\dot{y}(0) = 1$, calculer $\tilde{y} \in \mathbb{Q}$ (ou $\tilde{y} \in \mathbb{D} = \mathbb{Z}\, 2^{\mathbb{Z}}$) avec

$$|\tilde{y} - y(2011)| < 2^{-2011}.$$

Exemple 1 : intégration certifiée

Systèmes dynamiques

Soit $P = (P_1, \ldots, P_d) \in \mathbb{Q}[Y]^d = \mathbb{Q}[Y_1, \ldots, Y_d]^d$ et considérons

$$\dot{Y} = P(Y).$$

Pour $Y(0) \in \mathbb{Q}^d$, $t \in \mathbb{Q}^{>}$ et $\varepsilon \in \mathbb{Q}^{>}$, calculer (si possible) $\tilde{Y} \in \mathbb{Q}^d$ avec

$$\|\tilde{Y} - Y(t)\| < \varepsilon.$$

Fonctions spéciales

- Fonctions élémentaires $\exp, \log, \cos, \sin, \tan$, etc.
- Fonctions holonomes : $L(f) = L_r f^{(r)} + \cdots + L_0 f = 0$, $L \in \mathbb{Q}[z][\partial]$.
- Autres : $\wp$, fonctions thèta, fonctions de Matthieu, etc.

Exemple 2 : résolution polynomiale

Étant donné un système zéro-dimensionnel

$$\begin{cases} P_1(z_1, \ldots, z_n) = 0 \\ \vdots \\ P_n(z_1, \ldots, z_n) = 0 \end{cases}$$

avec $P_1, \ldots, P_n \in \mathbb{Q}[z_1, \ldots, z_n]$, trouver les zéros "dans $\mathbb{C}^n$".

Exemple 2 : résolution polynomiale

Étant donné un système zéro-dimensionnel

$$\begin{cases} P_1(z_1, \ldots, z_n) = 0 \\ \vdots \\ P_n(z_1, \ldots, z_n) = 0 \end{cases}$$

avec $P_1, \ldots, P_n \in \mathbb{E}[z_1, \ldots, z_n]$, trouver les zéros "dans $\mathbb{C}^n$".

Ici, $\mathbb{E}$ désigne l'ensemble des "nombres exp-log", construits à partir de $\mathbb{Q}$ en utilisant les opérations $+, -, \times, \exp$ et $\log$.

Exemple 2 : résolution polynomiale

Étant donné un système zéro-dimensionnel

$$\begin{cases} P_1(z_1, \ldots, z_n) = 0 \\ \vdots \\ P_n(z_1, \ldots, z_n) = 0 \end{cases}$$

avec $P_1, \ldots, P_n \in \mathbb{E}[z_1, \ldots, z_n]$, trouver les zéros "dans $\mathbb{R}^n$".

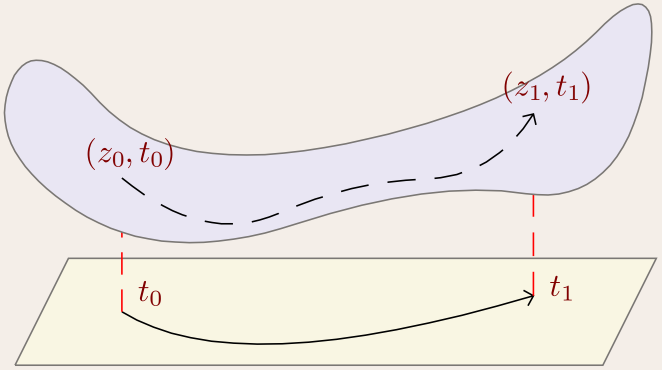
Ici, $\mathbb{E}$ désigne l'ensemble des "nombres exp-log", construits à partir de $\mathbb{Q}$ en utilisant les opérations $+, -, \times, \exp$ et $\log$.

Exemple 3 : suivi de chemin

$$H(z,t) = 0 \qquad \left\{ \begin{array}{l} H_1(z_1, \ldots, z_n, t) = 0 \\ \vdots \\ H_n(z_1, \ldots, z_n, t) = 0 \end{array} \right. ,$$

avec $H(z,t) \in \mathbb{Q}[z,t]^n$ zéro-dimensionnel en z pour $t \in \mathbb{C} \backslash \Sigma$ et Σ fini.

Étant donné (z_0, t_0) avec $H(z_0, t_0)$ et un chemin $t_0 \rightsquigarrow t_1$ qui évite Σ, calculer le chemin $z_0 \rightsquigarrow z_1$ avec $H(z_\lambda, t_\lambda) = 0$ pour tout $\lambda \in [0, 1]$.



Exemple 4 : nombres de Bell

$$B(z) = \sum_{n=0}^{\infty} \frac{B_n}{n!} z^n = \mathrm{e}^{\mathrm{e}^z - 1}.$$

Problème 1 : calculer B_{10000} avec une erreur relative de 10^{-10}

Problème 2 : trouver *automatiquement* le développement asymptotique

$$\log\left(\frac{B_n}{n!}\right) = n\left(-\log\log n + \frac{1}{\log n} + \frac{\log\log n}{\log n} + \mathcal{O}\left(\frac{1}{\log^2 n}\right)\right)$$

Problème 3 : pour $n \geqslant 10^{10}$, trouver une constante explicite pour le $\mathcal{O}\left(\frac{1}{\log^2 n}\right)$

Résolution de systèmes polynomiaux par homotopie

$$P(x,y) \quad \begin{cases} x^2+2\,x\,y-y^2-3\,x+5\,y+8=0 \\ 3\,x^2-x\,y+y^2+8\,x-2\,y+7=0 \end{cases}$$

$$\mathrm{Fastoche}(x,y) \quad \begin{cases} x^2-1=0 \\ y^2-1=0 \end{cases}$$

$$H(x,y,t) \quad \begin{cases} (x^2-1)\,t+(x^2+2\,x\,y-y^2-3\,x+5\,y+8)\,(1-t)=0 \\ (y^2-1)\,t+(3\,x^2-x\,y+y^2+8\,x-2\,y+7)\,(1-t)=0 \end{cases}$$

Calcul analytique → Calcul formel

Groupe de Galois d'une fonction algébrique

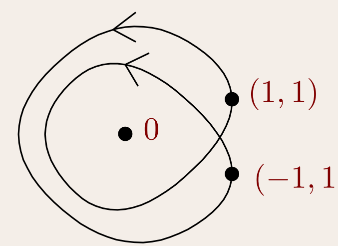
- Soit $P \in \mathbb{Q}[z,t]$
- Soit Σ l'ensemble des racines de $\operatorname{disc}_t P$
- Soient $z_{0,1}, \ldots, z_{0,d}$ les racines de P en $t_0 \in \mathbb{C} \backslash \Sigma$

Pour tout $\sigma \in \Sigma$:

- Construire un chemin t_λ^σ avec $t_0^\sigma = t_1^\sigma = t_0$ tournant autour de σ
- Remonter en un chemin z_λ^σ avec $P(z_\lambda^\sigma, t_\lambda^\sigma) = 0$ pour tout $\lambda \in [0,1]$
- Calculer la permutation π^σ de $\{1, \ldots, d\}$ avec $z_{1,i}^\sigma = z_{0,\pi^\sigma(i)}^\sigma$

Les π^σ génèrent $\operatorname{Gal}_{\mathbb{Q}(t)} \mathbb{Q}(z,t)/P$

Exemple $P = z^2 - t$



Prendre $t_\lambda = \mathrm{e}^{2\pi \mathrm{i} \lambda}$, $z_{0,1} = 1$, $z_{0,2} = -1$

$\pi^0\colon \begin{smallmatrix} 1 & \mapsto & 2 \\ 2 & \mapsto & 1 \end{smallmatrix}$, $\operatorname{Gal}_{\mathbb{Q}(t)} \mathbb{Q}\left(\sqrt{t}\right) = \mathfrak{S}_2$

Calcul analytique → Calcul formel

Monodromie d'un opérateur différentiel

$$L \;=\; z\,\partial^2 + \partial$$

Base de solutions pour $L(f) = z\,f'' + f' = 0$ en $z = 1$:

$$F(z) \;=\; \begin{pmatrix} 1 \\ \log z \end{pmatrix}$$

Prolongement analytique de la solution F autour de 0:

$$F(z\,\mathrm{e}^{2\pi\mathrm{i}}) \;=\; \begin{pmatrix} 1 \\ \log z + 2\,\pi\,\mathrm{i} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2\,\pi\,\mathrm{i} & 1 \end{pmatrix} F(z).$$

Groupe de Galois différentiel

$\mathrm{Gal}\,L$ généré par les matrices de monodromie, de Stokes, et exponentielles

Application: algorithme de factorisation de L

Calcul formel $\to$ Calcul analytique

Prétraitement symbolique $\to$ meilleure efficacité

$$e^{2^{-20112011}} - 1 \;\approx\; 1.0000000000 \cdot 2^{-20112011}$$

Plus généralement : asymptotique formelle $\to$ évaluation rapide de fonctions

Base de Groebner avant résolution numérique ?

Calcul de schémas numériques rapides

- Runge-Kutta à des ordres élevés
- Bonnes bases de fonctions pour éléments finis
- Bonnes bases d'ondelettes

Calcul formel ∩ Calcul analytique

Arithmétique rapide

- Calcul rapide en multi-précision
- Algorithmes denses rapides sur les polynômes, les matrices et les séries
- Traitement efficace de structures creuses et en évaluation

Besoin d'un bon langage

- Mathématiquement expressif
- Sémantiquement propre
- Interface conviviale
- Compilé (au moins pour le calcul analytique)

Nombre réel calculable

Définition. *Un nombre réel $x \in \mathbb{R}$ est **calculable** s'il existe un algorithme qui prend $\varepsilon \in \mathbb{D}^{>}$ en entrée et qui produit une ε-approximation $\tilde{x} \in \mathbb{D}$ de x avec $|\tilde{x} - x| \leqslant \varepsilon$. On note $\mathbb{R}^{\mathrm{cal}}$ l'ensemble des nombres réels calculables.*

Représentations alternatives

- Remplacer $\mathbb{D}$ par $\mathbb{Q}$
- $\exists$ algorithme qui prend n en entrée et calcule un 2^{-n}-approximation
- $\exists$ algorithme pour calculer $n_x \in \mathbb{Z}$, $x_n \in \{-1, 0, 1\}$ avec $x = \sum_{n \geqslant n_x} x_n 2^{-n}$
- $\exists$ suites calculables $x_0^{\mathrm{lo}} \leqslant x_1^{\mathrm{lo}} \leqslant \cdots \leqslant x$, $x_0^{\mathrm{hi}} \geqslant x_1^{\mathrm{hi}} \geqslant \cdots \geqslant x$ avec $\lim_n x_n^{\mathrm{hi}} - x_n^{\mathrm{lo}} = 0$

Références

Weihrauch : computable analysis
Bishop & Bridges : foundations of constructive analysis

Nombre réel calculable

Définition. *Un nombre réel $x \in \mathbb{R}$ est* ***calculable*** *s'il existe un algorithme qui prend $\varepsilon \in \mathbb{D}^{>}$ en entrée et qui produit une ε-approximation $\tilde{x} \in \mathbb{D}$ de x avec $|\tilde{x} - x| \leqslant \varepsilon$. On note $\mathbb{R}^{\mathrm{cal}}$ l'ensemble des nombres réels calculables.*

```
Mmx] use "asymptotix"; type_mode? := true;
```

```
Mmx] one: Approximator (Floating, Floating) == 1;
```

```
Mmx] e == exp one
```

$2.7182818284590452: \quad \mathrm{Approximator}(\mathrm{Floating}, \mathrm{Floating})$

```
Mmx] approximate (e, 1.0e-1000)
```

2.71828182845904523536028747135266249775724709369995957496696762772407663035354759457 13\
821785251664274274663919320030599218174135966290435729003342952605956307381323286279434\
907632338298807531952510190115738341879307021540891499348841675092447614606680822648001\
684774118537423454424371075390777449920695517027618386062613313845830007520449338265602\
976067371132007093287091274437470472306969772093101416928368190255151086574637721112523\
897844250569536967707854499699679468644549059879316368892300987931277361782154249992295\
763514822082698951936680331825288693984964651058209392398294887933203625094431173012381\
970684161403970198376793206832823764648042953118023287825098194558153017567173613320698\
112509961818815930416903515988885193458072738667385894228792284998920868058257492796104\
841984443634632449684875602336248270419786232090021609902353043699418491463140934317381\
436405462531520961836908887070167683964243781405927145635490613031072085103837505101157\
47704171898610687396965521267154688957035035402123407848: Floating

```
Mmx]
```

Théorème de Turing

Théorème. *Il n'existe pas de test de nullité pour* $\mathbb{R}^{\mathrm{cal}}$.

Asymmétrie. Soit $x \in \mathbb{R}^{\mathrm{cal}}$

- Si $x = 0$, alors on ne peut pas forcément le démontrer en temps fini
- Si $x \neq 0$, alors on peut le certifier en temps fini

Réstriction à des sous classes.

- Il existe un test de nullité pour $\mathbb{Q}^{\mathrm{clr}}$ (nombres algébriques réels)
- Il existe un test de nullité pour $\mathbb{E}$, modulo la conjecture de Schanuel
- **Problème ouvert difficile :** test de nullité pour les constantes holonomes

Théorème de Grzegorczyk

Théorème. *Toute fonction calculable $f\colon \mathbb{R}^{\mathrm{cal}} \to \mathbb{R}^{\mathrm{cal}}$ est continue.*

Démonstration. Considérons une machine de Turing M pour calculer f.

Soit $x = \sum_{n \geqslant n_x} x_n 2^{-n} \in \mathbb{R}^{\mathrm{cal}}$.

Sur l'entrée (n_x, x_n), M retourne (n_y, y_n) avec $y = f(x) = \sum_{n \geqslant n_y} y_n 2^{-n}$.

Soit $\varepsilon > 0$ et $k \in \mathbb{Z}$ avec $2^{-k} < \varepsilon$, considérons le calcul **fini** de $y_{n_y}, \ldots, y_k$.

Soit l maximal, tel que x_l intervient dans ce calcul.

Alors $|f(x') - f(x)| < \varepsilon$ pour tout $x' \in \mathbb{R}^{\mathrm{cal}}$ avec $|x' - x| < 2^{-l}$. $\square$

Théorème de Denef et Lipschitz

Théorème. *Considérons un système dynamique avec conditions initiales*

$$\begin{aligned} Y' &= P(Y) \\ Y(0) &\in \mathbb{Q}^d \end{aligned}$$

où $Y = (Y_1, \ldots, Y_d)$ *et* $P = (P_1, \ldots, P_d) \in \mathbb{Q}[Y]^d$. *D'après Cauchy, ce système admet une unique solution* $Y \in \mathbb{Q}[[z]]^d$. *Cette solution est convergente en* z.

Il n'existe pas d'algorithme pour calculer le rayon de convergence de Y_1.

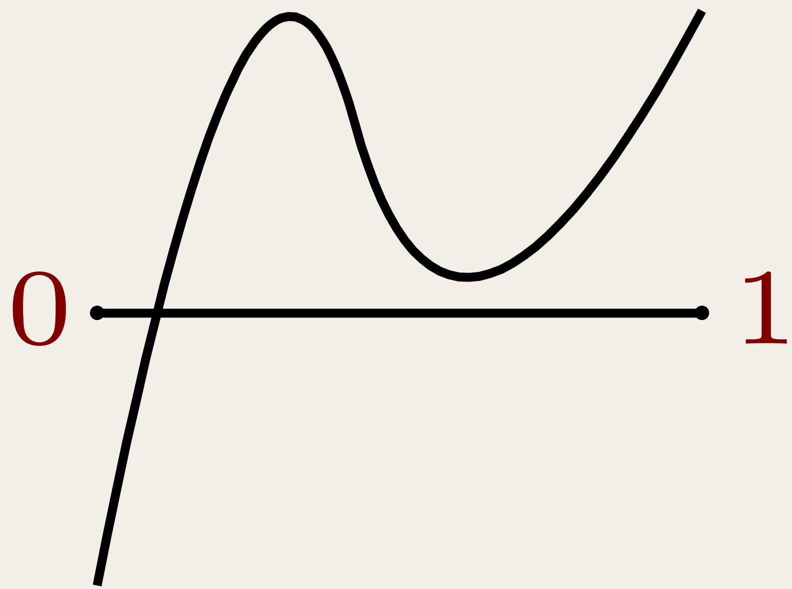
Questionnaire

Quels problèmes sont calculables ?

- Racines d'un polynôme (unitaire) à coefficients dans $\mathbb{C}^{\mathrm{cal}} = \mathbb{R}^{\mathrm{cal}}[\mathrm{i}]$
- Racines réelles d'un polynôme unitaire à coefficients dans $\mathbb{R}^{\mathrm{cal}}$
- Racines complexes d'un système polynomial zéro-dimensionnel sur $\mathbb{C}^{\mathrm{cal}}$
- Vecteur propres d'une matrice à coefficients dans $\mathbb{C}^{\mathrm{cal}}$
- Calcul d'une primitive d'une fonction calculable sur $\mathbb{R}^{\mathrm{cal}}$
- Calcul de la dérivée d'une fonction $\mathcal{C}^{\infty}$ calculable sur $\mathbb{R}^{\mathrm{cal}}$
- Calcul de la dérivée d'une fonction analytique calculable sur $\mathbb{C}^{\mathrm{cal}}$
- Calcul d'une racine de $f: [0, 1]^{\mathrm{cal}} \to \mathbb{R}^{\mathrm{cal}}$ calculable avec $f(0)\, f(1) < 0$

Quels problèmes sont calculables ?

- Calcul d'une racine de $f\colon [0,1]^{\mathrm{cal}} \to \mathbb{R}^{\mathrm{cal}}$ calculable avec $f(0)\,f(1) < 0$



Autres types de calculabilité

Calculabilité à gauche

$x \in \mathbb{R}^{\mathrm{calg}}$ s'il existe $\check{x}_0 \leqslant \check{x}_1 \leqslant \cdots$ calculable avec $\check{x}_n \in \mathbb{D}$ et $\lim_{n\to\infty} x_n = x$

Calculabilité à droite

$x \in \mathbb{R}^{\mathrm{cald}}$ si $-x \in \mathbb{R}^{\mathrm{calg}}$. On a $\mathbb{R}^{\mathrm{cal}} = \mathbb{R}^{\mathrm{calg}} \cap \mathbb{R}^{\mathrm{cald}}$

Inférieurement calculable

$X \in \mathscr{P}^{\mathrm{calinf}}(\mathbb{R}^d)$ s'il existe $\check{X}_0 \subseteq \check{X}_1 \subseteq \cdots$ calculable avec $\bigcup_{n\in\mathbb{N}} \check{X}_n = X$, où chaque $\check{X}_n$ est une réunion finie de « blocks » fermés à extrémités dans $\mathbb{D}^d$

Supérieurement calculable

$X \in \mathscr{P}^{\mathrm{calsup}}(\mathbb{R}^d)$ si $\mathbb{R}^d \setminus X \in \mathscr{P}^{\mathrm{calinf}}(\mathbb{R}^d)$

Répondaire

- $=: \mathbb{R}^{\mathrm{cal}} \times \mathbb{R}^{\mathrm{cal}} \to \{0,1\}^{\mathrm{cald}}$ calculable
- $f: \mathbb{R}^{\mathrm{cal}} \to \mathbb{R}^{\mathrm{cal}}$ calculable implique $\mathrm{graphe}(f) \in \mathscr{P}^{\mathrm{calsup}}(\mathbb{R}^2)$
- $\mathrm{rayonsol}: \mathbb{Q}[Y]^d \times \mathbb{Q}^d \to \mathbb{R}^{\mathrm{calg}}$ calculable
- $\mathrm{racines}: \mathbb{C}^{\mathrm{cal}}[z]_{\mathrm{unit}} \to \mathscr{M}(\mathbb{C})$ calculable
- $\mathrm{racines}: \mathbb{R}^{\mathrm{cal}}[z]_{\mathrm{unit}} \to \mathscr{P}^{\mathrm{calsup}}(\mathbb{R})$ calculable
- $\mathrm{racines}: \mathscr{F}([0,1]^{\mathrm{cal}}, \mathbb{R}^{\mathrm{cal}}) \to \mathscr{P}^{\mathrm{calsup}}(\mathbb{R})$ calculable
- $=: \mathbb{E} \times \mathbb{E} \to \{0,1\}^{\mathrm{Schanuel}}$ calculable (égalité modulo conjecture de Schanuel)

Formalisation générale

Structures effectives

$$E \xleftarrow{\text{représentation}} \check{E} \xrightarrow{\text{encodage}} \mathbb{N}$$

Approximateurs

$\tilde{E}$: ensemble abstrait d'approximations d'élements dans E

$x = \lim_{n\to\infty} \tilde{x}_n$: $x \in E$ est la limite des approximations $\tilde{x}_n$

$$\check{E} = \left\{ (\check{x}_n) \in \mathscr{F}^{\text{cal}}\left(\mathbb{N}, \tilde{E}\right) : (\check{x}_n) \text{ admet une limite } x \in E \right\}$$

Fonctorialité

$\widetilde{E \times F} = \tilde{E} \times \tilde{F}$

$\tilde{x} \subseteq E, \tilde{y} \subseteq F \Rightarrow \tilde{x} \times \tilde{y} \subseteq E \times F$

$\widetilde{\mathscr{F}^{\text{cal}}(E,F)} = \mathscr{F}^{\text{cal}}\left(\tilde{E}, \tilde{F}\right)$

$f : E \to F \in \tilde{f} : \tilde{E} \to \tilde{F}$ si
$x \in \tilde{x} \Rightarrow f(x) \in \tilde{f}(\tilde{x})$

Formalisation générale

Structures effectives

$$\mathbb{R}^{\mathrm{calg}} \xleftarrow{\text{représentation}} \widecheck{\mathbb{R}^{\mathrm{calg}}} \xrightarrow{\text{encodage}} \mathbb{N}$$

$$x = \lim_{n\to\infty} \check{x}_n \qquad (\check{x}_n) \in \mathscr{F}^{\mathrm{cal}}(\mathbb{N}, \mathbb{D})\Uparrow$$

Approximateurs

$\tilde{E}$: ensemble abstrait d'approximations d'élements dans E

$x = \lim_{n\to\infty} \tilde{x}_n$: $x \in E$ est la limite des approximations $\tilde{x}_n$

$\check{E} = \left\{(\check{x}_n) \in \mathscr{F}^{\mathrm{cal}}\big(\mathbb{N}, \tilde{E}\big) : (\check{x}_n) \text{ admet une limite } x \in E\right\}$

Fonctorialité

$\widetilde{E \times F} = \tilde{E} \times \tilde{F}$

$\tilde{x} \subseteq E, \tilde{y} \subseteq F \Rightarrow \tilde{x} \times \tilde{y} \subseteq E \times F$

$\widetilde{\mathscr{F}^{\mathrm{cal}}(E, F)} = \mathscr{F}^{\mathrm{cal}}\big(\tilde{E}, \tilde{F}\big)$

$f : E \to F \in \tilde{f} : \tilde{E} \to \tilde{F}$ si

$x \in \tilde{x} \Rightarrow f(x) \in \tilde{f}(\tilde{x})$

Formalisation générale

Structures effectives

$$E \xleftarrow{\text{représentation}} \check{E} \xrightarrow{\text{encodage}} \mathbb{N}$$

Approximateurs

$\tilde{E}$: ensemble abstrait d'approximations d'élements dans E

$x = \lim_{n\to\infty} \tilde{x}_n$: $x \in E$ est la limite des approximations $\tilde{x}_n$

$$\check{E} = \left\{ (\check{x}_n) \in \mathscr{F}^{\mathrm{cal}}\big(\mathbb{N}, \tilde{E}\big) : (\check{x}_n) \text{ admet une limite } x \in E \right\}$$

Fonctorialité

$\widetilde{E \times F} = \tilde{E} \times \tilde{F}$

$\tilde{x} \subseteq E, \tilde{y} \subseteq F \Rightarrow \tilde{x} \times \tilde{y} \subseteq E \times F$

$\widetilde{\mathscr{F}^{\mathrm{cal}}(E, F)} = \mathscr{F}^{\mathrm{cal}}\big(\tilde{E}, \tilde{F}\big)$

$f : E \to F \in \tilde{f} : \tilde{E} \to \tilde{F}$ si
$x \in \tilde{x} \Rightarrow f(x) \in \tilde{f}(\tilde{x})$

Formalisation générale

Structures effectives

$$E \xleftarrow{\text{représentation}} \check{E} \xrightarrow{\text{encodage}} \mathbb{N}$$

Approximateurs

$\tilde{E} \subseteq \mathscr{P}(E)$ et $\tilde{x} \in \tilde{E}$ approxime $x \in E$ si $x \in \tilde{x}$

$x = \lim_{n\to\infty} \tilde{x}_n$ si $\{x\} = \bigcap_{n\in\mathbb{N}} \tilde{x}_n$

$\check{E} = \left\{(\check{x}_n) \in \mathscr{F}^{\mathrm{cal}}(\mathbb{N}, \tilde{E}) \colon \exists x \in \mathbb{E}, \{x\} = \bigcap_{n\in\mathbb{N}} \tilde{x}_n\right\}$

Fonctorialité

$\widetilde{E \times F} = \tilde{E} \times \tilde{F}$

$\tilde{x} \subseteq E, \tilde{y} \subseteq F \Rightarrow \tilde{x} \times \tilde{y} \subseteq E \times F$

$\widetilde{\mathscr{F}^{\mathrm{cal}}(E, F)} = \mathscr{F}^{\mathrm{cal}}(\tilde{E}, \tilde{F})$

$f\colon E \to F \in \tilde{f}\colon \tilde{E} \to \tilde{F}$ si
$x \in \tilde{x} \Rightarrow f(x) \in \tilde{f}(\tilde{x})$

Formalisation générale

Structures effectives

$$E \xleftarrow{\text{représentation}} \check{E} \xrightarrow{\text{encodage}} \mathbb{N}$$

Approximateurs

$$\widetilde{\mathbb{R}^{\mathrm{cal}}} = \{[a,b] : a, b \in \mathbb{D}, a \leqslant b\}$$

$x = \lim_{n\to\infty} \tilde{x}_n$ si $\{x\} = \bigcap_{n\in\mathbb{N}} \tilde{x}_n$

$$\widecheck{\mathbb{R}^{\mathrm{cal}}} = \left\{ (\check{x}_n) \in \mathscr{F}^{\mathrm{cal}}\left(\mathbb{N}, \widetilde{\mathbb{R}^{\mathrm{cal}}}\right) : \exists x \in \mathbb{R}, \{x\} = \bigcap_{n\in\mathbb{N}} \tilde{x}_n \right\}$$

Fonctorialité

$$\widetilde{E \times F} = \tilde{E} \times \tilde{F}$$

$$\tilde{x} \subseteq E, \tilde{y} \subseteq F \Rightarrow \tilde{x} \times \tilde{y} \subseteq E \times F$$

$$\widetilde{\mathscr{F}^{\mathrm{cal}}(E,F)} = \mathscr{F}^{\mathrm{cal}}\left(\tilde{E}, \tilde{F}\right)$$

$f : E \to F \in \tilde{f} : \tilde{E} \to \tilde{F}$ si

$$x \in \tilde{x} \Rightarrow f(x) \in \tilde{f}(\tilde{x})$$

Formalisation générale

Structures effectives

$$E \xleftarrow{\text{représentation}} \check{E} \xrightarrow{\text{encodage}} \mathbb{N}$$

Approximateurs

$$\widetilde{\mathbb{R}^{\mathrm{calg}}} = \{[a, \infty[: a \in \mathbb{D}\}$$

$x = \lim_{n\to\infty} \tilde{x}_n$ si $x = \inf \bigcap_{n\in\mathbb{N}} \tilde{x}_n$

$$\widecheck{\mathbb{R}^{\mathrm{calg}}} = \left\{ (\check{x}_n) \in \mathscr{F}^{\mathrm{cal}}\left(\mathbb{N}, \widetilde{\mathbb{R}^{\mathrm{calg}}} \right) : \bigcap_{n\in\mathbb{N}} \tilde{x}_n \neq \varnothing \right\}$$

Fonctorialité

$\widetilde{E \times F} = \tilde{E} \times \tilde{F}$

$\tilde{x} \subseteq E, \tilde{y} \subseteq F \Rightarrow \tilde{x} \times \tilde{y} \subseteq E \times F$

$\widetilde{\mathscr{F}^{\mathrm{cal}}(E, F)} = \mathscr{F}^{\mathrm{cal}}(\tilde{E}, \tilde{F})$

$f : E \to F \in \tilde{f} : \tilde{E} \to \tilde{F}$ si
$x \in \tilde{x} \Rightarrow f(x) \in \tilde{f}(\tilde{x})$

Formalisation générale

Structures effectives

$$E \xleftarrow{\text{représentation}} \check{E} \xrightarrow{\text{encodage}} \mathbb{N}$$

Approximateurs

$\tilde{E}$: ensemble abstrait d'approximations d'élements dans E

$x = \lim_{n\to\infty} \tilde{x}_n$: $x \in E$ est la limite des approximations $\tilde{x}_n$

$$\check{E} = \left\{(\check{x}_n) \in \mathscr{F}^{\text{cal}}\big(\mathbb{N}, \tilde{E}\big) : (\check{x}_n) \text{ admet une limite } x \in E\right\}$$

Fonctorialité

$\widetilde{E \times F} = \tilde{E} \times \tilde{F}$

$\tilde{x} \subseteq E, \tilde{y} \subseteq F \Rightarrow \tilde{x} \times \tilde{y} \subseteq E \times F$

$\widetilde{\mathscr{F}^{\text{cal}}(E, F)} = \mathscr{F}^{\text{cal}}\big(\tilde{E}, \tilde{F}\big)$

$f : E \to F \in \tilde{f} : \tilde{E} \to \tilde{F}$ si
$x \in \tilde{x} \Rightarrow f(x) \in \tilde{f}(\tilde{x})$

Arithmétique de boules

Principe

Approximation systématique des nombres $x \in \mathbb{R}$ par des **boules** fermées

$$x^\bullet = \mathcal{B}(c, r) = c + \mathcal{B}(r) \;\; = \;\; \{x \in \mathbb{R} \colon |x - c| \leqslant r\},$$

avec $c \in \mathbb{R}$ et $r \in \mathbb{R}^{\geqslant}$ (ou $c \in \mathbb{D} \cup \{\pm\infty\}$ et $r \in \mathbb{D}^{\geqslant} \cup \{\infty\}$)

$\longrightarrow$ on garantit que le « vrai résultat » x est dans $x^\bullet$

Opérations

$f^\bullet \colon \mathcal{B}(\mathbb{R}, \mathbb{R}^{>})^d \to \mathcal{B}(\mathbb{R}, \mathbb{R}^{>})$ **extension** de $f \colon \mathbb{R}^d \to \mathbb{R}$ si

$$f(x_1^\bullet, \ldots, x_d^\bullet) = \{f(x_1, \ldots, x_d) \colon x_1 \in x_1^\bullet, \ldots, x_n \in x_n^\bullet\} \;\; \subseteq \;\; f^\bullet(x_1^\bullet, \ldots, x_n^\bullet),$$

pour tous les $x_1^\bullet, \ldots, x_d^\bullet \in \mathcal{B}(\mathbb{R}, \mathbb{R}^{>})$

Arithmétique de boules

Principe

Approximation systématique des nombres $x \in \mathbb{R}$ par des **boules** fermées

$$x^{\bullet} = \mathcal{B}(c, r) = c + \mathcal{B}(r) \;\; = \;\; \{x \in \mathbb{R} : |x - c| \leqslant r\},$$

avec $c \in \mathbb{R}$ et $r \in \mathbb{R}^{\geqslant}$ (ou $c \in \mathbb{D} \cup \{\pm\infty\}$ et $r \in \mathbb{D}^{\geqslant} \cup \{\infty\}$)

$\longrightarrow$ on garantit que le « vrai résultat » x est dans $x^{\bullet}$

$\longrightarrow$ analogie notation de Landau $f(z) = 1 + z + \frac{1}{2} z^2 + \mathcal{O}(z^3)$

Opérations

$f^{\bullet} : \mathcal{B}(\mathbb{R}, \mathbb{R}^{>})^d \to \mathcal{B}(\mathbb{R}, \mathbb{R}^{>})$ **extension** de $f : \mathbb{R}^d \to \mathbb{R}$ si

$$f(x_1^{\bullet}, \ldots, x_d^{\bullet}) = \{f(x_1, \ldots, x_d) : x_1 \in x_1^{\bullet}, \ldots, x_n \in x_n^{\bullet}\} \;\; \subseteq \;\; f^{\bullet}(x_1^{\bullet}, \ldots, x_n^{\bullet}),$$

pour tous les $x_1^{\bullet}, \ldots, x_d^{\bullet} \in \mathcal{B}(\mathbb{R}, \mathbb{R}^{>})$

Arithmétique de boules

Principe

Approximation systématique des nombres $x \in \mathbb{R}$ par des **boules** fermées

$$x^{\bullet} = \mathcal{B}(c, r) = c + \mathcal{B}(r) \;\; = \;\; \{x \in \mathbb{R} : |x - c| \leqslant r\},$$

avec $c \in \mathbb{R}$ et $r \in \mathbb{R}^{\geqslant}$ (ou $c \in \mathbb{D} \cup \{\pm\infty\}$ et $r \in \mathbb{D}^{\geqslant} \cup \{\infty\}$)

$\longrightarrow$ on garantit que le « vrai résultat » x est dans $x^{\bullet}$

$\longrightarrow$ analogie notation de Landau $f(z) = 1 + z + \frac{1}{2} z^2 + \mathcal{O}(z^2)$

Opérations

$f^{\bullet} : \mathcal{B}(\mathbb{R}, \mathbb{R}^{>})^d \to \mathcal{B}(\mathbb{R}, \mathbb{R}^{>})$ **extension** de $f : \mathbb{R}^d \to \mathbb{R}$ si

$$f(x_1^{\bullet}, \ldots, x_d^{\bullet}) = \{f(x_1, \ldots, x_d) : x_1 \in x_1^{\bullet}, \ldots, x_n \in x_n^{\bullet}\} \;\; \subseteq \;\; f^{\bullet}(x_1^{\bullet}, \ldots, x_n^{\bullet}),$$

pour tous les $x_1^{\bullet}, \ldots, x_d^{\bullet} \in \mathcal{B}(\mathbb{R}, \mathbb{R}^{>})$

Opérations élémentaires

Formules exactes

$$\begin{aligned}
\mathcal{B}(x,r) +^{\bullet} \mathcal{B}(y,s) &= \mathcal{B}(x+y, r+s)\\
\mathcal{B}(x,r) -^{\bullet} \mathcal{B}(y,s) &= \mathcal{B}(x-y, r+s)\\
\mathcal{B}(x,r) \times^{\bullet} \mathcal{B}(y,s) &= \mathcal{B}(x\times y, r\times(|y|+s)+|x|\times s)
\end{aligned}$$

Formules approchées

$x, y, r, s \in \mathbb{D}_{p,e} = \pm\{0,\ldots,2^p-1\}\,2^{\pm\{0,\ldots,2^e-1\}} \cup \{\pm\infty\}$

$$\begin{aligned}
\mathcal{B}(x,r) +^{\bullet} \mathcal{B}(y,s) &= \mathcal{B}(x +^{\mathrm{near}} y, r +^{\uparrow} s +^{\uparrow} \mathrm{erreur}_{+}(x,y))\\
\mathcal{B}(x,r) -^{\bullet} \mathcal{B}(y,s) &= \mathcal{B}(x -^{\mathrm{near}} y, r +^{\uparrow} s +^{\uparrow} \mathrm{erreur}_{-}(x,y))\\
\mathcal{B}(x,r) \times^{\bullet} \mathcal{B}(y,s) &= \mathcal{B}(x \times^{\mathrm{near}} y, r \times^{\uparrow} (|y| +^{\uparrow} s) +^{\uparrow} |x| \times^{\uparrow} s +^{\uparrow} \mathrm{erreur}_{\times}(x,y))
\end{aligned}$$

Opérations élémentaires

Formules exactes

$$\begin{aligned}
\mathcal{B}(x,r) +^{\bullet} \mathcal{B}(y,s) &= \mathcal{B}(x+y, r+s)\\
\mathcal{B}(x,r) -^{\bullet} \mathcal{B}(y,s) &= \mathcal{B}(x-y, r+s)\\
\mathcal{B}(x,r) \times^{\bullet} \mathcal{B}(y,s) &= \mathcal{B}(x\times y, r\times(|y|+s)+|x|\times s)
\end{aligned}$$

Formules approchées

$x, y, r, s \in \mathbb{D}_{p,e} = \pm\{0,\ldots,2^p-1\}\,2^{\pm\{0,\ldots,2^e-1\}} \cup \{\pm\infty\}$

$$\begin{aligned}
\mathcal{B}(x,r) +^{\bullet} \mathcal{B}(y,s) &= \mathcal{B}(x +^{\mathrm{ieee}} y, r +^{\uparrow} s +^{\uparrow} \mathrm{errarr}(x +^{\mathrm{ieee}} y))\\
\mathcal{B}(x,r) -^{\bullet} \mathcal{B}(y,s) &= \mathcal{B}(x -^{\mathrm{ieee}} y, r +^{\uparrow} s +^{\uparrow} \mathrm{errarr}(x -^{\mathrm{ieee}} y))\\
\mathcal{B}(x,r) \times^{\bullet} \mathcal{B}(y,s) &= \mathcal{B}(x \times^{\mathrm{ieee}} y, r \times^{\uparrow} (|y| +^{\uparrow} s) +^{\uparrow} |x| \times^{\uparrow} s +^{\uparrow} \mathrm{errarr}(x \times^{\mathrm{ieee}} y))
\end{aligned}$$

Arithméthique d'intervalles

Arithmétique d'intervalles	Arithmétique de boules
Bon support matériel	Peu de support matériel
Standardisation naturelle	Standardisation non canonique
Subdivision naturelle	Subdivision pas naturelle
Perd facteur 2 en précision multiple	Efficace en précision multiple
Se vectorise moins bien	Se vectorise bien
Basé sur ordre total	Extension naturelle à $\mathbb{C}$ et autres
Perturbation non naturelle	Perturbation naturelle
Théorie *ad hoc* autour de la norme IEEE	Calcul ε-δ classique et bien compris

Références

Moore (1965) : Interval analysis
Alefeld & Herzberger (1983) : Introduction to interval analysis
Rump (2010) : Verification methods : rigourous results using floating-point arithmetic

Estimations d'erreur *a priori*

Exemple : calcul d'une ε-approximation de $z = x + y$

Calculer des $\frac{\varepsilon}{2}$-approximations $\tilde{x}$ et $\tilde{y}$ de x et y
Retourner $\tilde{x} + \tilde{y}$

Avantage : précision de calcul adaptatif dans $1 + 10^{-100}$

Problème : évaluation de $P_d x^d + \cdots + P_0$ par Horner en $x = 1$

Calcul d'une $\frac{\varepsilon}{2^{d-k}}$-approximation de P_k pour $k = 0, \ldots, d - 1$

$\longrightarrow$ tolérance trop petite

Problème : nécessite « dags » pour tous les résultats intermédiaires

Estimations d'erreur *a priori*

Exemple : calcul d'une ε-approximation de $z = x + y$

Calculer des $\frac{\varepsilon}{2}$-approximations $\tilde{x}$ et $\tilde{y}$ de x et y
Retourner $\tilde{x} + \tilde{y}$

Avantage : précision de calcul adaptatif dans $1 + 10^{-100}$

Problème : évaluation de $P_d x^d + \cdots + P_0$ par Horner en $x = 1$

Calcul d'une $\frac{\varepsilon}{2^{d-k}}$-approximation de P_k pour $k = 0, \ldots, d - 1$

Équilibrage des tolérances : calcul d'une $\frac{\varepsilon}{d+1}$-approximation de P_k pour tout k

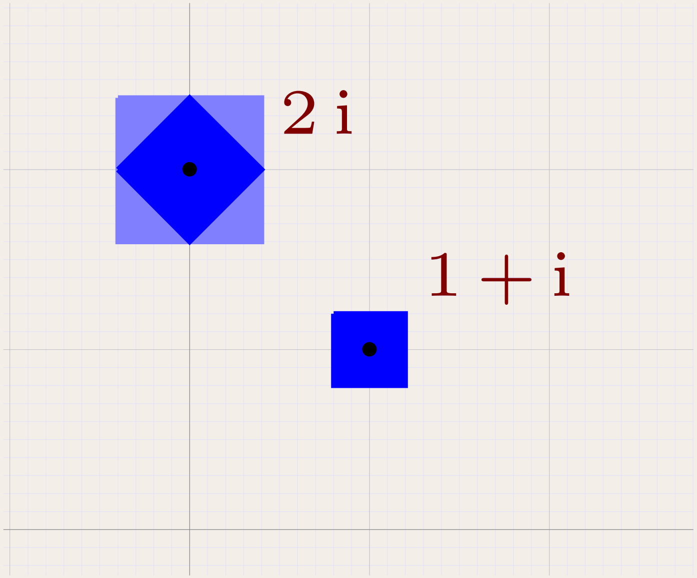
Problème : nécessite « dags » pour tous les résultats intermédiaires

Surestimation de l'erreur

$$\mathcal{B}(0,1) - \mathcal{B}(0,1) \;=\; \mathcal{B}(0,2)$$

Effet d'enveloppement

$$([1-\varepsilon, 1+\varepsilon] + [1-\varepsilon, 1+\varepsilon]\,\mathrm{i})^2 \;=\; [-2\,\varepsilon, 2\,\varepsilon] + [2-2\,\varepsilon, 2+2\,\varepsilon]\,\mathrm{i}$$



Exemple

```
Mmx] use "algebramix";
Mmx] puissance (z, n) == if n = 1 then z else z * puissance (z, n-1);
Mmx] z == complex (interval (1.0, 1.0000000001), interval (1.0, 1.0000000001))
```

$1.000000000 + 1.000000000\,\mathrm{i}$

```
Mmx] [ puissance (z, 4*n) || n in 1 to 10 ]
```

$$\begin{bmatrix} -4.00000000 \\ 16.0000000 \\ -64.000000 \\ 256.0000 \\ -1024.000 \\ 4096.00 \\ -16384.0 \\ 6.5536e4 \\ -2.62e5 \\ 1.05e6 \end{bmatrix}$$

```
Mmx]
```

Exemple

```
Mmx] use "algebramix";
Mmx] puissance (z, n) == if n = 1 then z else z * puissance (z, n-1);
Mmx] z == ball (complex (1.0, 1.0), 0.0000000001)
```

$1.000000000 + 1.000000000\,\mathrm{i}$

```
Mmx] [ puissance (z, 4*n) || n in 1 to 10 ]
```

$$\begin{bmatrix} -4.00000000 \\ 16.0000000 \\ -64.000000 \\ 256.000000 \\ -1024.00000 \\ 4096.0000 \\ -16384.0000 \\ 65536.000 \\ -262144.00 \\ 1.04857600e6 \end{bmatrix}$$

```
Mmx]
```

Minimiser la profondeur du calcul

```
Mmx] use "algebramix";
Mmx] puissance (z, n) ==
       if n = 1 then z
       else puissance (z, n quo 2) * puissance (z, n - (n quo 2));
Mmx] z == complex (interval (1.0, 1.0000000001), interval (1.0, 1.0000000001))
```

$1.000000000 + 1.000000000\,\mathrm{i}$

```
Mmx] [ puissance (z, 4*n) || n in 1 to 10 ]
```

$$\left[\begin{array}{c} -4.00000000 \\ 16.0000000 \\ -64.000000 \\ 256.000000 \\ -1024.00000 \\ 4096.0000 \\ -16384.000 \\ 65536.000 \\ -262144.00 \\ 1.0485760e6 \end{array}\right]$$

```
Mmx]
```

Minimiser la profondeur du calcul

```
Mmx] use "algebramix";
```

```
Mmx] puissance (z, n) ==
       if n = 1 then z
       else puissance (z, n quo 2) * puissance (z, n - (n quo 2));
```

```
Mmx] z == complex (interval (1.0, 1.0000000001), interval (1.0, 1.0000000001))
```

$1.000000000 + 1.000000000\,\mathrm{i}$

```
Mmx] [ puissance (z, 4*n) || n in 1 to 10 ]
```

$$\begin{bmatrix} -4.00000000 \\ 16.0000000 \\ -64.000000 \\ 256.000000 \\ -1024.00000 \\ 4096.0000 \\ -16384.000 \\ 65536.000 \\ -262144.00 \\ 1.0485760e6 \end{bmatrix}$$

```
Mmx]
```

Méthode perturbative (Hansen)

Algorithme : inversion d'une matrice $M \in \mathcal{B}(\mathbb{D}, \mathbb{D})^{n \times n}$

- Écrire $M = \mathcal{B}(C, R)$, avec $C \in \mathbb{D}^{n \times n}$ et $R \in (\mathbb{D}^{>})^{n \times n}$
- Inverser numériquement $U \approx C^{-1}$
- Calculer $E = C\,\mathcal{B}(U, 0) - \mathcal{B}(1, 0)$
- Calculer $\varepsilon = \|E\|_\infty = \max_{1 \leqslant i \leqslant n} (|E_{i1}| + \cdots + |E_{in}|)$. On a :

$$\|(1+E)^{-1} - 1\|_\infty = \|-E + E^2 - E^3 + \cdots\|_\infty \leqslant \frac{\varepsilon}{1-\varepsilon}$$

- Soit $\mathbb{1}$ la matrice avec entrées $\mathbb{1}_{ij} = 1$
- Retourner $U\,\mathcal{B}(1, \frac{\varepsilon}{1-\varepsilon}\,\mathbb{1})$

Plus généralement

Phase de **divination** ou **prospection**, puis phase de **validation**

Exemple

```
Mmx] use "analyziz"
Mmx] rnd () == {
       x == uniform_deviate (0.0, 1.0);
       return ball (x, 0.00000001);
     };
Mmx] M == [ rnd () | j in 1 to 7 || i in 1 to 7 ]
```

$$\begin{bmatrix} 0.6003643 & 0.3710189 & 0.4791752 & 0.1559513 & 0.4159951 & 0.1407053 & 0.0044911 \\ 0.5120348 & 0.7247114 & 0.1947940 & 0.655922 & 0.9355430 & 0.251486 & 0.2468534 \\ 0.462639 & 0.5648943 & 0.744350 & 0.4275757 & 0.7804102 & 0.9920186 & 0.7571225 \\ 0.2508314 & 0.0131642 & 0.8669620 & 0.6802075 & 0.7552889 & 0.3166660 & 0.3160914 \\ 0.2876469 & 0.988199 & 0.6639689 & 0.303261 & 0.498002 & 0.0277941 & 0.9222224 \\ 0.1304958 & 0.7945297 & 0.7988684 & 0.5890020 & 0.3421886 & 0.0285099 & 0.9893513 \\ 0.3882938 & 0.7500550 & 0.9237840 & 0.594590 & 0.8730709 & 0.5362681 & 0.134547 \end{bmatrix}$$

```
Mmx] row_echelon M
```

$$\begin{bmatrix} 0.6003643 & 0.3710189 & 0.4791752 & 0.1559513 & 0.4159951 & 0.1407053 & 0.0044911 \\ 0e-7 & 0.810436 & 0.4343862 & 0.2285411 & 0.298690 & -0.0396207 & 0.9200706 \\ 0e-7 & 0e-7 & 0.7427919 & 0.655052 & 0.633765 & 0.2509449 & 0.475251 \\ 0e-7 & 0e-7 & 0e-7 & 0.7893843 & 0.799480 & 0.2976308 & 0.0563714 \\ 0e-7 & 0e-7 & 0e-7 & 0e-7 & -0.3571903 & -0.102232 & -0.027367 \\ 0e-7 & 0e-7 & 0e-7 & 0e-7 & 0e-7 & 0.771327 & 0.280188 \\ 0e-7 & 0e-7 & 0e-7 & 0e-7 & 0e-7 & 0e-7 & -0.7889997 \end{bmatrix}$$

```
Mmx]
```

Nombre de conditionnement et perte de précision

Nombre de conditionnement (intrinsèque)

$$
\begin{aligned}
f\colon \mathbb{R}^n &\to \mathbb{R}^k \\
\kappa_f(x) &= \limsup_{\varepsilon\to 0} \frac{\|f(x+\varepsilon)-f(x)\|}{\|f(x)\|} \Big/ \frac{\|\varepsilon\|}{\|x\|} \\
\kappa(M) &= \|M\|\,\|M^{-1}\|
\end{aligned}
$$

Perte de précision (d'un algorithme)

$$
\begin{aligned}
f_p\colon \mathbb{D}_p^n &\to \mathbb{D}_p^k \\
\chi_{f_p}(x) &= \frac{\|f_p(x)-f(x)\|}{\|f(x)\|} \Big/ \left(2^p\,\kappa_f(x)\right)
\end{aligned}
$$

Extensions optimales et surestimation

Extension optimale (intrinsèque)

$$
\begin{array}{rcl}
f\colon \mathbb{R}^n & \to & \mathbb{R}^k \\
f^*(\mathcal{B}(x,r)) & = & \mathcal{B}\left(f(x), \sup_{x' \in \mathcal{B}(x,r)} |f(x') - f(x)| \right)
\end{array}
$$

Surestimation et surestimation ponctuelle (d'un algorithme)

$$
\begin{array}{rcl}
f^\bullet\colon \mathcal{B}(\mathbb{R},\mathbb{R})^n & \to & \mathcal{B}(\mathbb{R},\mathbb{R})^k \\
\chi_{f^\bullet}(\mathcal{B}(x,r)) & = & \inf_\lambda \{ f^\bullet(\mathcal{B}(x,r)) - f(x) \subseteq \lambda\,(f^*(\mathcal{B}(x,r)) - f(x)) \} \\
\chi_{f^\bullet}(x) & = & \limsup_{r \to 0} \chi_{f^\bullet}(\mathcal{B}(x,r))
\end{array}
$$

Question ouverte abordable : pour $f^\bullet$ construite à partir de $+^\bullet$, $-^\bullet$, $\times^\bullet$, est-ce que

$$
\chi_{f^\bullet}(\mathcal{B}(x,r)) \leqslant \sup_{x \in \mathcal{B}(x,r)} \chi_{f^\bullet}(x) \;?
$$

Surestimation de l'arithmétique standard

Exemple

$$\begin{aligned} f(x) &= (x-1)^2 \\ f^\bullet(x^\bullet) &= (x^\bullet)^2 -^\bullet 2\,x^\bullet +^\bullet 1 \end{aligned}$$

$$\begin{aligned} f^*(\mathcal{B}(x,\varepsilon)) &= \mathcal{B}(f(x), 2\,|x-1|\,\varepsilon + \mathcal{O}(\varepsilon^2)) \\ f^\bullet(\mathcal{B}(x,\varepsilon)) &= \mathcal{B}(f(x), 2\,|x|\,\varepsilon + 2\,\varepsilon + \mathcal{O}(\varepsilon^2)) \\ \chi_{f^\bullet}(x) &= \frac{|x|+1}{|x-1|} \end{aligned}$$

Généralement

$$\chi_{f^\bullet}(x) = \limsup_{\varepsilon \neq 0} \frac{(\bar{\nabla} f)(x) \cdot |\varepsilon|}{|(\nabla f)(x)| \cdot |\varepsilon|} \qquad \text{ou } (d=1) \qquad \frac{(\bar{\nabla} f)(x)}{|f'(x)|}$$

$$\begin{aligned} \bar{\nabla}\, c &= (0, \ldots, 0) && (c \in \mathbb{R}) \\ \bar{\nabla}\, X_k &= (0, \overset{k-1}{\ldots}, 0, 1, 0, \ldots, 0) && (k \in \{1, \ldots, r\}) \\ \bar{\nabla}(f \pm g) &= \bar{\nabla}\, f + \bar{\nabla}\, g \\ \bar{\nabla}(f\,g) &= (\bar{\nabla}\, f)\,|g| + |f|\,(\bar{\nabla}\, g), \end{aligned}$$

Le Graal : estimations efficaces et de qualité

Plus d'efficacité

$\mathcal{O}(1)$	Majoration triviale par $\mathcal{B}(0, \infty)$
Pénalité constante	Qualité généralement acceptable
Pénalité généralement non linéaire	Proche du nombre de condition
Généralement NP complet	Borne optimale

Plus de qualité

Terminaison précoce : commencer avec stratégie rapide et *a priori* de faible qualité
Utilisation d'algorithmes plus lents seulement si cette qualité ne convient pas

Multiplication de matrices

Majoration $|(M\,N)_{ij}| \leqslant [\max|M_{i\cdot}|]\,[\max|N_{\cdot j}|]$:

1 multiplication dans $\mathbb{D}^{n\times n}$ et majorations en $O(n^2)$

Compromis: couper M et N en matrices $k\times k$ avec k fixe:

α_k multiplications dans $\mathbb{D}^{n\times n}$ avec $1 \leqslant \alpha_k \leqslant 3$

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

3 multiplications dans $\mathbb{D}^{n\times n}$ (dont 2 multiplications en simple précision)

Produit naif M, N avec $M, N \in \mathcal{B}(\mathbb{D}, \mathbb{D})^{n\times n}$:

n^3 opérations dans $\mathcal{B}(\mathbb{D}, \mathbb{D})$

Principe :

Coût de validation asymptotiquement négligeable pour opérations bien conditionnées

Multiplication de matrices

Majoration $|(MN)_{ij}| \leqslant [\max |M_{i\cdot}|] \, [\max |N_{\cdot j}|]$:
1 multiplication dans $\mathbb{D}^{n \times n}$ et majorations en $O(n^2)$

Compromis: couper M et N en matrices $k \times k$ avec k fixe:
α_k multiplications dans $\mathbb{D}^{n \times n}$ avec $1 \leqslant \alpha_k \leqslant 3$

Réécriture $M, N \in \mathcal{B}(\mathbb{D}^{n \times n}, \mathbb{D}^{n \times n})$:
3 multiplications dans $\mathbb{D}^{n \times n}$ (dont 2 multiplications en simple précision)

Produit naif M, N avec $M, N \in \mathcal{B}(\mathbb{D}, \mathbb{D})^{n \times n}$:
n^3 opérations dans $\mathcal{B}(\mathbb{D}, \mathbb{D})$

Principe :

Coût de validation asymptotiquement négligeable pour opérations bien conditionnées

Multiplication de matrices

Majoration $|(MN)_{ij}| \leqslant [\max |M_{i\cdot}|]\,[\max |N_{\cdot j}|]$:

1 multiplication dans $\mathbb{D}^{n \times n}$ et majorations en $O(n^2)$

Compromis: couper M et N en matrices $k \times k$ avec k fixe:

α_k multiplications dans $\mathbb{D}^{n \times n}$ avec $1 \leqslant \alpha_k \leqslant 3$

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

3 multiplications dans $\mathbb{D}^{n \times n}$ (dont 2 multiplications en simple précision)

Produit naif M, N avec $M, N \in \mathcal{B}(\mathbb{D}, \mathbb{D})^{n \times n}$:

n^3 opérations dans $\mathcal{B}(\mathbb{D}, \mathbb{D})$

Principe :

Coût de validation asymptotiquement négligeable pour opérations bien conditionnées

Multiplication de matrices

Majoration $|(M\,N)_{ij}| \leqslant [\max |M_{i\cdot}|]\,[\max |N_{\cdot j}|]$:

1 multiplication dans $\mathbb{D}^{n\times n}$ et $\mathcal{O}(n^2)$ majorations

Compromis: couper M et N en matrices $k\times k$ avec k fixe:

α_k multiplications dans $\mathbb{D}^{n\times n}$ avec $1 \leqslant \alpha_k \leqslant 3$

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

3 multiplications dans $\mathbb{D}^{n\times n}$ (dont 2 multiplications en simple précision)

Produit naif M, N avec $M, N \in \mathcal{B}(\mathbb{D}, \mathbb{D})^{n\times n}$:

$\mathcal{O}(n^3)$ opérations dans $\mathcal{B}(\mathbb{D}, \mathbb{D})$

Principe :

Coût de validation asymptotiquement négligeable pour opérations bien conditionnées

Hiérarchie numérique

Niveau mathématique

↓

Niveau certifié

↓

Niveau numérique

↓

Niveau arithmétique

Multiplication de matrices

Niveau mathématique

↓

Niveau certifié

↓

Niveau numérique

↓

Niveau arithmétique

Multiplication de matrices

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

Niveau certifié

Niveau numérique

Niveau arithmétique

Multiplication de matrices

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

↓

Niveau certifié

↓

Niveau numérique

↓

Niveau arithmétique

Multiplication de matrices

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

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

Niveau numérique

Niveau arithmétique

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

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

Niveau numérique

Niveau arithmétique

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

$\downarrow$

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

$\downarrow$

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

$\downarrow$

Niveau arithmétique

Multiplication de matrices

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

$$\downarrow$$

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

$$\downarrow$$

$$\tilde{M}, \tilde{N} \in \mathbb{D}_{52,12}^{n\times n}$$

$$\downarrow$$

Niveau arithmétique

Multiplication de matrices

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

$\downarrow$

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

$\downarrow$

$\tilde{M}, \tilde{N} \in \mathbb{Z}^{n\times n}\, 2^{\mathbb{Z}}$

$\downarrow$

Niveau arithmétique

Multiplication de matrices

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

$$\downarrow$$

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

$$\downarrow$$

$$\tilde{M}, \tilde{N} \in \mathbb{Z}^{n\times n}\, 2^{\mathbb{Z}}$$

$$\downarrow$$

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