Towards accurate polynomial evaluation in rounded arithmetic

Olga Holtz

Department of Mathematics

University of California-Berkeley

`holtz@math.berkeley.edu`

joint work with James Demmel and Ioana Dumitriu

Outline

- Motivation and goal(s).

Outline

- Motivation and goal(s).
- Model of arithmetic and setting.

Outline

- Motivation and goal(s).
- Model of arithmetic and setting.
- What is allowable in classical arithmetic.

Outline

- Motivation and goal(s).
- Model of arithmetic and setting.
- What is allowable in classical arithmetic.
- Results for classical arithmetic, real and complex.

Outline

- Motivation and goal(s).
- Model of arithmetic and setting.
- What is allowable in classical arithmetic.
- Results for classical arithmetic, real and complex.
- What is allowable in black-box arithmetic.

Outline

- Motivation and goal(s).
- Model of arithmetic and setting.
- What is allowable in classical arithmetic.
- Results for classical arithmetic, real and complex.
- What is allowable in black-box arithmetic.
- Results for black-box arithmetic, real and complex.

Outline

- Motivation and goal(s).
- Model of arithmetic and setting.
- What is allowable in classical arithmetic.
- Results for classical arithmetic, real and complex.
- What is allowable in black-box arithmetic.
- Results for black-box arithmetic, real and complex.
- Open problems / Future work.

Goal and problem

Given a polynomial (or a family of polynomials) p, either produce an **accurate** algorithm to compute $y = p(x)$, or prove that none exists.

Accuracy means relative error $\eta < 1$, i.e.

- $|y_{\text{computed}} - y| \le \eta\, |y|$,
- $\eta = 10^{-2}$ yields two digits of accuracy,
- $y_{\text{computed}} = 0 \iff y = 0$.

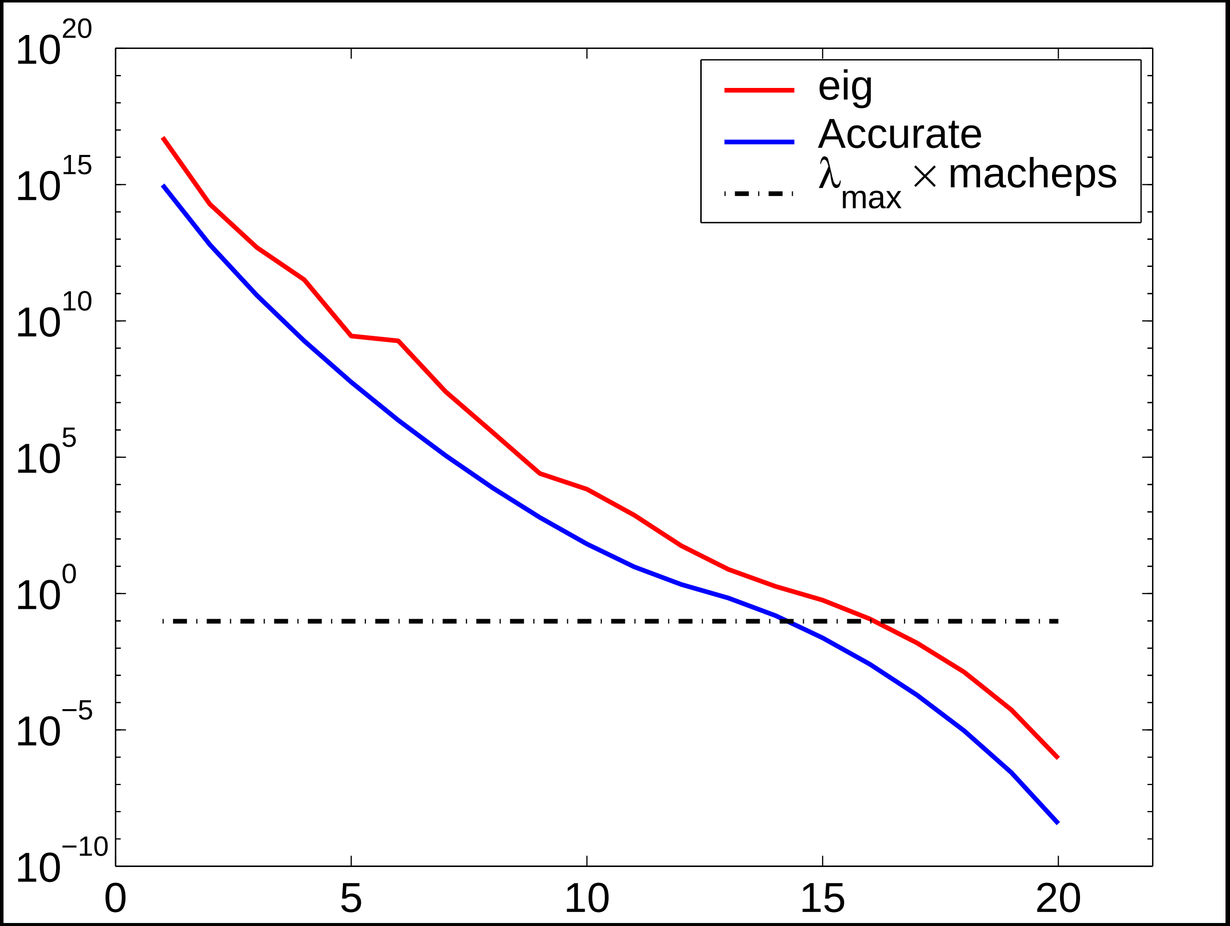
Why should we care?

- High-sensitivity computations of polynomials, like in **Hales**’ approach to prove Kepler’s conjecture.

Why should we care?

- High-sensitivity computations of polynomials, like in **Hales**’ approach to prove Kepler’s conjecture.
- In NLA of structured matrices, computing the determinant accurately is a necessary condition for obtaining accurate minors, inverse, Gaussian elimination, eigenvalues, singular values (**Demmel & Koev**).

Example



Schur complements of Vandermonde matrices

Why should we care?

- In Computational geometry and mesh generation, decisions about point inclusions are based on accurate evaluation of multivariate polynomials (**Shewchuk**).

Why should we care?

- In Computational geometry and mesh generation, decisions about point inclusions are based on accurate evaluation of multivariate polynomials (**Shewchuk**).
- Interesting mathematically! Extremely rich body of work (e.g. **von Neumann, Turing, Wilkinson, Blum, Smale, Shub, Cucker**) dealing with alternative models of arithmetic.

Traditional Model of Arithmetic

- ◇ $fl(a \otimes b) = (a \otimes b)(1 + \delta)$, with arbitrary roundoff error $|\delta| < \varepsilon \ll 1$
- ◇ Operations?

Traditional Model of Arithmetic

- ◇ $fl(a \otimes b) = (a \otimes b)(1+\delta)$, with arbitrary roundoff error $|\delta| < \varepsilon \ll 1$
- ◇ Operations?
  - ◇ in classical arithmetic, $+$, $-$, $\times$; also exact negation;
  - ◇ in black-box arithmetic, in addition to the above, polynomial expressions, like $x + y + z$, $x - yz$, etc.

Availability of Constants

Constants may be available multiplicatively as $x \mapsto cx$ or additively as $x \mapsto x - c$. Without, e.g., $\sqrt{2}$, we cannot compute $x^2 - 2$ accurately.

Availability of Constants

Constants may be available multiplicatively as $x \mapsto cx$ or additively as $x \mapsto x - c$. Without, e.g., $\sqrt{2}$, we cannot compute $x^2 - 2$ accurately.

Classical case: no constants (consider integer-coefficient polynomials without constant term).

Black-box case: any constants (which is accommodated under the black-box umbrella).

Algorithms

- give answer in finite # of steps (no iterative algorithms)

Algorithms

- give answer in finite # of steps (no iterative algorithms)
- exact answer in absence of roundoff error

Algorithms

- give answer in finite # of steps (no iterative algorithms)
- exact answer in absence of roundoff error
- branching based on exact comparisons

Algorithms

- give answer in finite # of steps (no iterative algorithms)
- exact answer in absence of roundoff error
- branching based on exact comparisons
- non-determinism (because determinism can be simulated)

Algorithms

- give answer in finite # of steps (no iterative algorithms)
- exact answer in absence of roundoff error
- branching based on exact comparisons
- non-determinism (because determinism can be simulated)
- inputs are from domains $\mathbb{C}^n$ or $\mathbb{R}^n$ (also some smaller domains)

Problem

◇ Notation:

- $p(x)$ is the multivariate polynomial to be evaluated, $x = (x_1, \ldots, x_k)$.
- $\delta = (\delta_1, \ldots, \delta_m)$ is the vector of error (rounding) variables.
- $p_{comp}(x, \delta)$ is the result of algorithm to compute p at x with errors δ.

Problem

- ◇ **Goal:** Decide if ∃ algorithm $p_{comp}(x,\delta)$ to accurately evaluate $p(x)$ on $\mathcal{D}$:

$\forall\ 0 < \eta < 1$ ... for any η = desired relative error

$\exists\ 0 < \varepsilon < 1$ ... there is an ε = maximum rounding error

$\forall\ x \in \mathcal{D}$ ... so that for all x in the domain

$\forall\ |\delta_i| \le \varepsilon$ ... and for all rounding errors bounded by ε

$$|p_{comp}(x,\delta) - p(x)| \le \eta \cdot |p(x)|$$

... relative error is at most η

- ◇ Given p and $\mathcal{D}$, seek effective procedure ("compiler") to exhibit such an algorithm, or show none exists

Examples in classical arithmetic

- $M_2(x, y, z) = z^6 + x^2 \cdot y^2 \cdot (x^2 + y^2 - 2 \cdot z^2)$
  - Positive definite and homogeneous, easy to evaluate accurately

Examples in classical arithmetic

- ◇ $M_3(x, y, z) = z^6 + x^2 \cdot y^2 \cdot (x^2 + y^2 - 3 \cdot z^2)$
  - Motzkin polynomial, nonnegative, zero at $|x| = |y| = |z|$

Examples in classical arithmetic

- $M_3(x, y, z) = z^6 + x^2 \cdot y^2 \cdot (x^2 + y^2 - 3 \cdot z^2)$

$$\begin{array}{rl}
\text{if} & |x - z| \le |x + z| \wedge |y - z| \le |y + z| \\
p = & z^4 \cdot [4((x - z)^2 + (y - z)^2 + (x - z)(y - z))] + \\
 & +z^3 \cdot [2(2(x - z)^3 + 5(y - z)(x - z)^2 + 5(y - z)^2(x - z) + \\
 & \qquad\qquad 2(y - z)^3)] + \\
 & +z^2 \cdot [(x - z)^4 + 8(y - z)(x - z)^3 + 9(y - z)^2(x - z)^2 + \\
 & \qquad\qquad 8(y - z)^3(x - z) + (y - z)^4] + \\
 & +z \cdot [2(y - z)(x - z)((x - z)^3 + 2(y - z)(x - z)^2 + \\
 & \qquad\qquad 2(y - z)^2(x - z) + (y - z)^3] + \\
 & +(y - z)^2(x - z)^2((x - z)^2 + (y - z)^2) \\
\text{else} & \ldots\ 2^{\#\text{vars}} - 1 \ \text{ more analogous cases}
\end{array}$$

Examples in classical arithmetic

- $M_4(x,y,z) = z^6 + x^2 \cdot y^2 \cdot (x^2 + y^2 - 4 \cdot z^2)$
  - Impossible to evaluate accurately

Sneak peak

The variety,

$$V(p) = \{x : p(x) = 0\},$$

plays an important role.

Allowable varieties

Basic allowable sets:

- $Z_i = \{x : x_i = 0\}$,
- $S_{ij} = \{x : x_i + x_j = 0\}$,
- $D_{ij} = \{x : x_i - x_j = 0\}$.

Allowable varieties

Basic allowable sets:

- $Z_i = \{x : x_i = 0\}$,
- $S_{ij} = \{x : x_i + x_j = 0\}$,
- $D_{ij} = \{x : x_i - x_j = 0\}$.

A variety $V(p)$ is allowable if it can be written as a finite union of intersections of basic allowable sets.

Allowable varieties

Basic allowable sets:

- $Z_i = \{x : x_i = 0\}$,
- $S_{ij} = \{x : x_i + x_j = 0\}$,
- $D_{ij} = \{x : x_i - x_j = 0\}$.

Denote by

$$\mathbf{G(p)} = \mathbf{V(p)} - \cup_{\textbf{allowable } \mathbf{A} \subset \mathbf{V(p)}} \mathbf{A}$$

the set of points in general position.

$V(p)$ unallowable $\quad \Rightarrow \quad G(p) \neq \emptyset$.

Necessity condition

Theorem 1. $V(p)$ unallowable $\Rightarrow p$ cannot be evaluated accurately on $\mathbb{R}^n$ or on $\mathbb{C}^n$.

Theorem 2. On a domain $\mathcal{D}$, if $\mathrm{Int}(\mathcal{D}) \cap G(p) \neq \emptyset$, p cannot be evaluated accurately.

Examples on $\mathbb{R}^n$ *revisited*

- $p(x, y, z) = x + y + z$

  UNALLOWABLE

- $M_2(x, y, z) = z^6 + x^2 \cdot y^2 \cdot (x^2 + y^2 - 2 \cdot z^2)$

  ALLOWABLE, $V(p) = \{0\}$.

- $M_3(x, y, z) = z^6 + x^2 \cdot y^2 \cdot (x^2 + y^2 - 3 \cdot z^2)$

  ALLOWABLE, $V(p) = \{|x| = |y| = |z|\}$

- $M_4(x, y, z) = z^6 + x^2 \cdot y^2 \cdot (x^2 + y^2 - 4 \cdot z^2)$

  UNALLOWABLE

Sketch of proof

Simplest case: non-branching, no data reuse (except for inputs), nondeterminism.

Algorithm can be represented as a tree with extra edges from the sources, each node corresponds to an operation $(+, -, \times)$, each node has a specific δ, each node has two inputs, one output.

Let $x \in G(p)$ and define $Allow(x)$ as the smallest allowable set containing x.

Sketch of proof

Key fact: for a positive measure set of δs in δ-space, a zero output can be "traced back" down the tree to "allowable" conditions $(x_i = 0$ or $x_i + x_j = 0)$, or trivial one $(x_i - x_i = 0)$.

So for a positive measure set of δs, either

- $p_{comp}(x, \delta)$ is not 0 (though $p(x) = 0$), or
- for all $y \in Allow(x) \setminus V(p)$, $p_{comp}(y, \delta) = 0$ (though $p(y) \neq 0$).

Thus p is not accurately evaluable close to x. $\square$

Sufficiency, complex case

Theorem. Let p be a polynomial over $\mathbb{C}^n$ with integer coefficients. If $V(p)$ is **allowable**, then p is **accurately evaluable**.

Sketch of proof. Can write $p(x) = c \prod_i p_i(x)$, where $p_i(x)$ is a power of some x_j or $x_j \pm x_k$, and c is an integer; all operations are accurate.

Corollary. If p is a complex multivariate polynomial, p is accurately evaluable iff p has integer coefficients and $V(p)$ is allowable.

Sufficiency, real case

Trickier... Allowability *not* sufficient:

- $q = (u^4 + v^4) + (u^2 + v^2)(x^2 + y^2 + z^2)$, $V(q) = \{u = v = 0\}$: allowable and accurately evaluable

Sufficiency, real case

Trickier... Allowability *not* sufficient:

- $q = (u^4 + v^4) + (u^2 + v^2)(x^2 + y^2 + z^2)$, $V(q) = \{u = v = 0\}$: allowable and accurately evaluable
- $p = (u^4 + v^4) + (u^2 + v^2)(x + y + z)^2$, $V(p) = \{u = v = 0\}$: allowable but NOT accurately evaluable!

Sufficiency, real case

Trickier... Allowability *not* sufficient:

- $q = (u^4 + v^4) + (u^2 + v^2)(x^2 + y^2 + z^2)$, $V(q) = \{u = v = 0\}$: allowable and accurately evaluable
- $p = (u^4 + v^4) + (u^2 + v^2)(x + y + z)^2$, $V(p) = \{u = v = 0\}$: allowable but NOT accurately evaluable!
- Has to do with locally dominant behavior (in this case, near the set $\{u = v = 0\}$).

What is dominance?

Example.

$$
\begin{aligned}
&p(x_1, x_2, x_3) \\
&= x_2^8 x_3^{12} \\
&+ x_1^2 x_2^2 x_3^{16} \\
&+ x_1^8 x_3^{12} \\
&+ x_1^6 x_2^{14} \\
&+ x_1^{10} x_2^6 x_3^4
\end{aligned}
$$

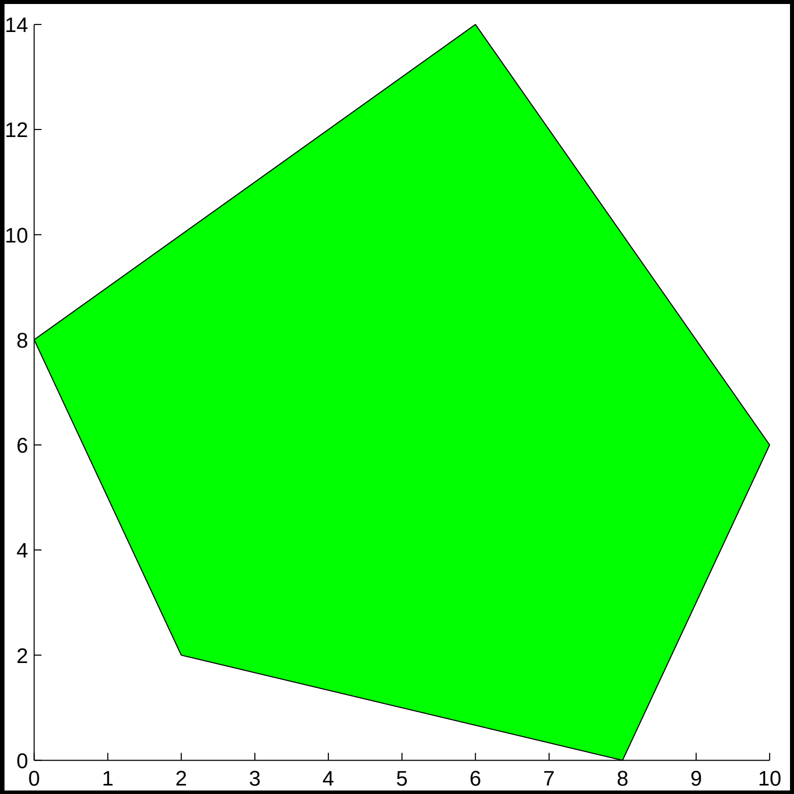


Figure 1. A Newton polytope P.

What is dominance?

Example.

$$p(x_1, x_2, x_3)$$
$$= x_2^8 x_3^{12}$$
$$+x_1^2 x_2^2 x_3^{16}$$
$$+x_1^8 x_3^{12}$$
$$+x_1^6 x_2^{14}$$
$$+x_1^{10} x_2^6 x_3^4$$

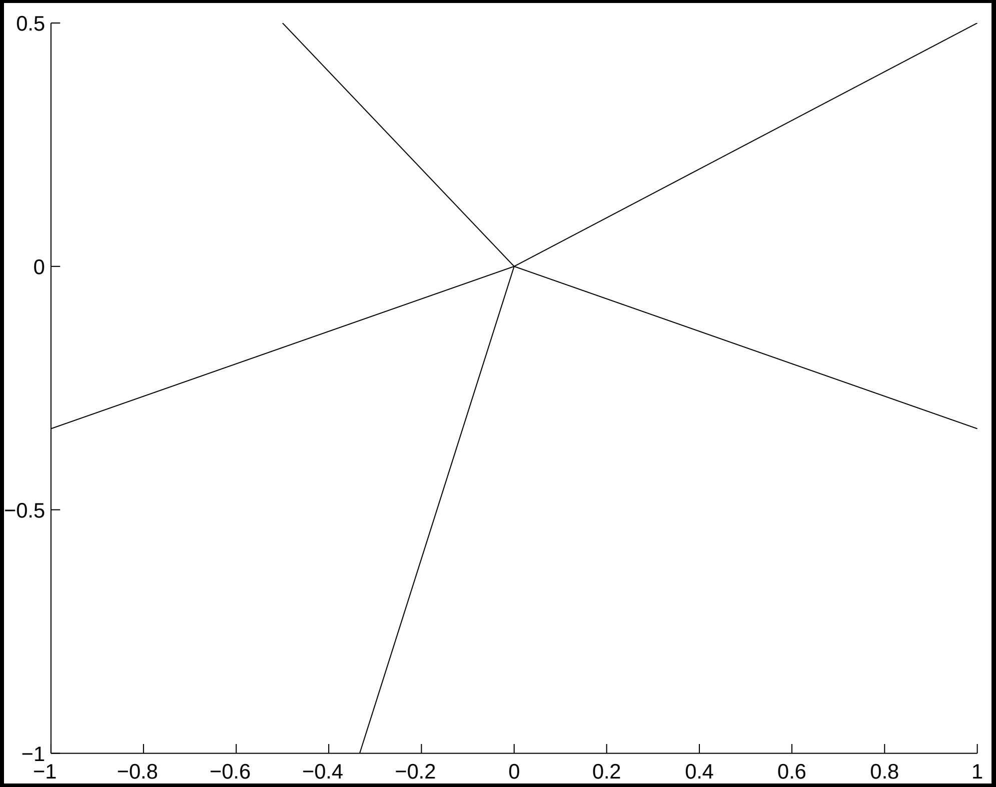


Figure 2. Its normal fan $N(P)$.

What is dominance?

Example.

$$
\begin{aligned}
&p(x_1, x_2, x_3) \\
&= x_2^8 x_3^{12} \\
&+ x_1^2 x_2^2 x_3^{16} \\
&+ x_1^8 x_3^{12} \\
&+ x_1^6 x_2^{14} \\
&+ x_1^{10} x_2^6 x_3^4
\end{aligned}
$$

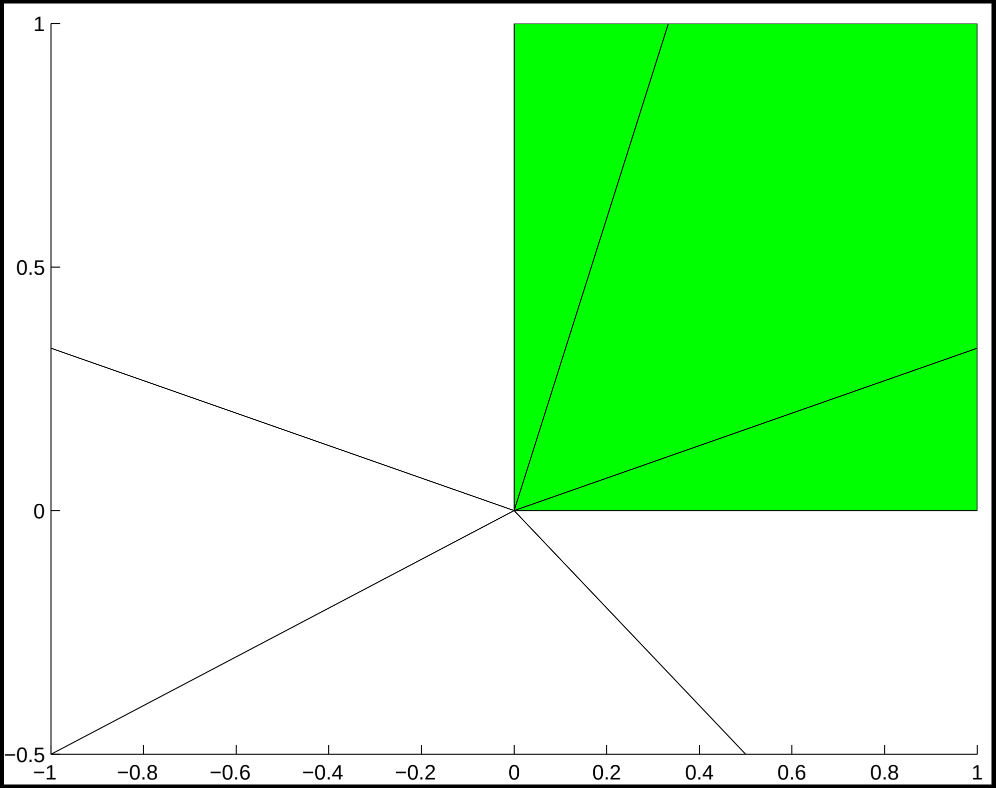


Figure 3. Positive part of $-N(P)$.

What is dominance?

Example.

$$
\begin{aligned}
& p(x_1, x_2, x_3) \\
& = x_2^8 x_3^{12} \\
& + x_1^2 x_2^2 x_3^{16} \\
& + x_1^8 x_3^{12} \\
& + x_1^6 x_2^{14} \\
& + x_1^{10} x_2^6 x_3^4
\end{aligned}
$$

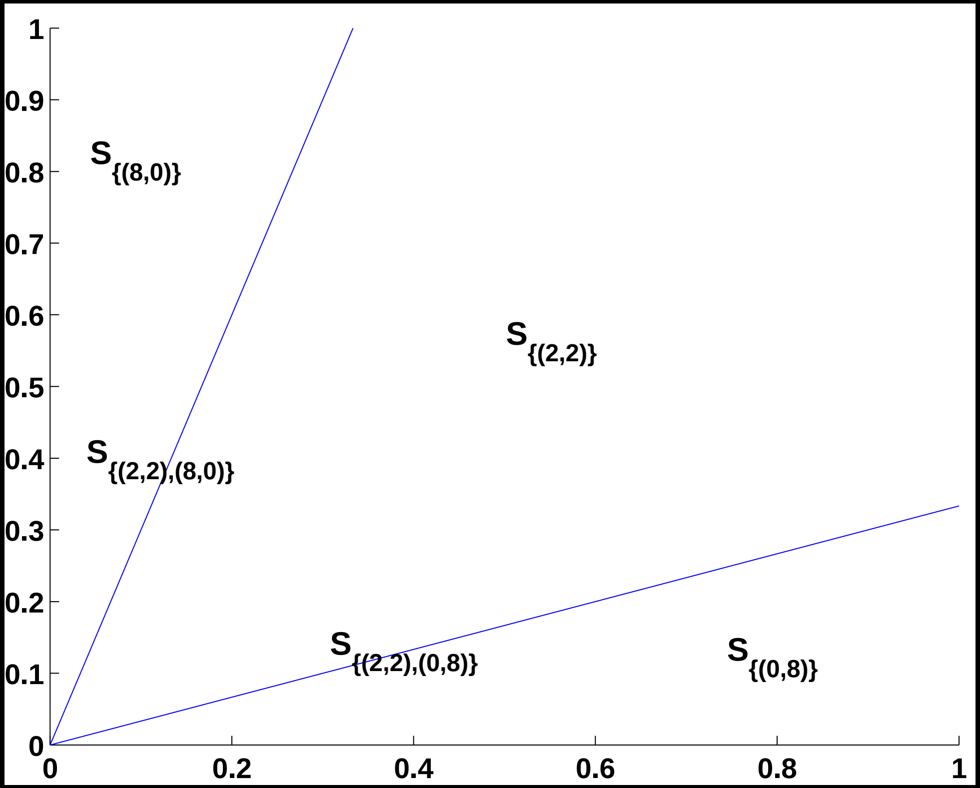


Figure 4. Labelling positive cones.

What is dominance?

Example.

$$p(x_1, x_2, x_3)$$
$$= x_2^8 x_3^{12}$$
$$+ x_1^2 x_2^2 x_3^{16}$$
$$+ x_1^8 x_3^{12}$$
$$+ x_1^6 x_2^{14}$$
$$+ x_1^{10} x_2^6 x_3^4$$

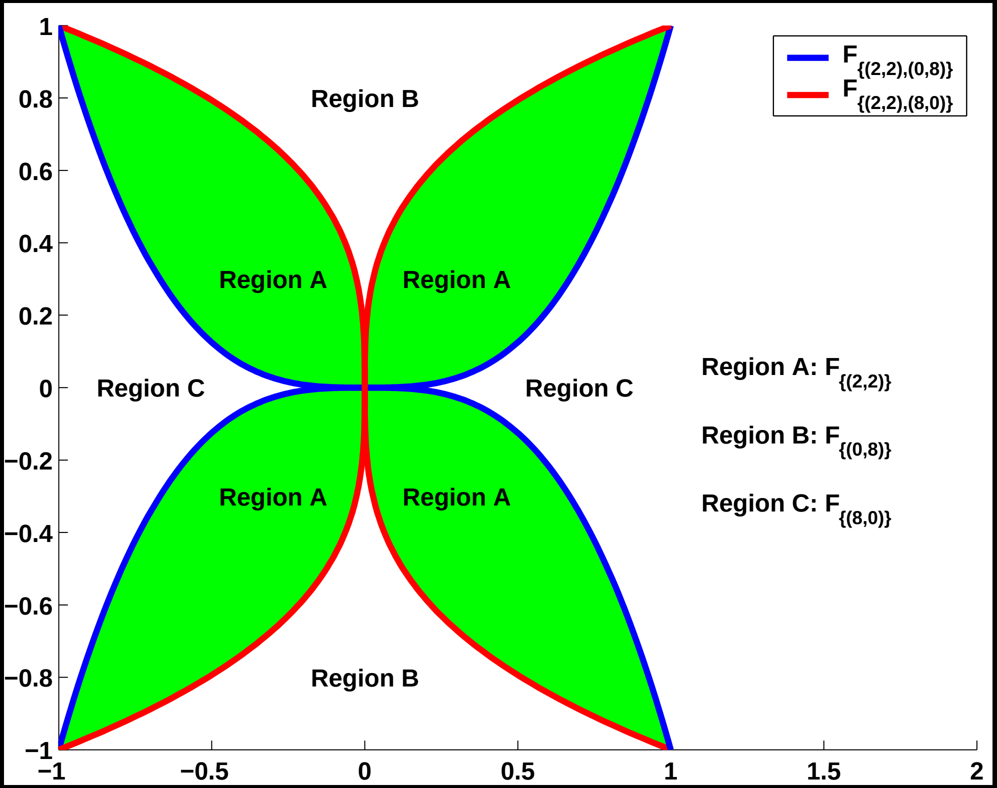


Figure 5. Exponentiating cones.

Sufficiency, real case

Theorem. If all dominant terms are accurately evaluable on $\mathbb{R}^n$, then p is accurately evaluable. In non-branching case, if p is accurately evaluable on $\mathbb{R}^n$, then so are all dominant terms.

Proof: one direction based on pruning an algorithm for p to obtain algorithms for its dominant terms; the other direction based on combining algorithms for each dominant term into one algorithm that evaluates p.

Pruning

Example: Pruning an algorithm for

$$p(x) = x_1^2 x_2^2 + (x_2 - x_3)^4 + (x_3 - x_4)^2 x_5^2.$$

Show pruning →

Induction?

Need inductive procedure of testing accurate evaluability going through all dominant terms, but so far there is no clear induction parameter.

Allowable black-box varieties

Define black-boxes $q_1, q_2, \ldots, q_k$ polynomial operations with various inputs, and for any j,

$$\mathcal{V}_j = \{V \neq \mathbb{R}^n : V \text{ can be obtained from } q_j \text{ through Process A } \}$$

Allowable black-box varieties

Define black-boxes $q_1, q_2, \ldots, q_k$ polynomial operations with various inputs, and for any j,

$$\mathcal{V}_j = \{ V \neq \mathbb{R}^n : V \text{ can be obtained from } q_j \text{ through Process A } \}$$

Process A:

Step 1. repeat and/or negate, or zero out some of the inputs,

Step 2. find the resulting variety of each black-box.

Examples of Process A

$q_1(x, y) = x - y$ has (up to symmetry)

$$\mathcal{V}_1 = \{\{x = 0\}, \{x - y = 0\}, \{x + y = 0\}\},$$

$q_2(x, y, z) = x - y \cdot z$ has (up to symmetry)

$$\begin{aligned}\mathcal{V}_2 \;=\;& \{\{x = 0\},\ \{y = 0\} \cup \{z = 0\},\\ & \{x = 0\} \cup \{x = 1\},\ \{x = 0\} \cup \{x = -1\},\\ & \{x = 0\} \cup \{y = 1\},\ \{x = 0\} \cup \{y = -1\},\\ & \{x - y^2 = 0\},\ \{x + y^2 = 0\},\\ & \{x - yz = 0\},\ \{x + yz = 0\}\}.\end{aligned}$$

Allowable black-box varieties

Define black-boxes $q_1, q_2, \ldots, q_k$ polynomial operations with various inputs, and for any j,

$$\mathcal{V}_j = \{V \neq \mathbb{R}^n : V \text{ can be obtained from } q_j \text{ through Process A}\}$$

Allowable black-box varieties

Define black-boxes $q_1, q_2, \ldots, q_k$ polynomial operations with various inputs, and for any j,

$$\mathcal{V}_j = \{V \neq \mathbb{R}^n : V \text{ can be obtained from } q_j \text{ through Process A}\}$$

Basic allowable sets:

- $Z_i = \{x : x_i = 0\}$,
- $S_{ij} = \{x : x_i + x_j = 0\}$,
- $D_{ij} = \{x : x_i - x_j = 0\}$,
- any V for which there is a j such that $V \in \mathcal{V}_j$.

Allowable black-box varieties

A variety $V(p)$ is allowable if it is a union of irreducible parts of finite intersections of basic allowable sets.

Allowable black-box varieties

A variety $V(p)$ is allowable if it is a union of irreducible parts of finite intersections of basic allowable sets.

Denote by

$$\mathbf{G}(\mathbf{p}) = \mathbf{V}(\mathbf{p}) - \cup_{\text{allowable } \mathbf{A} \subset \mathbf{V}(\mathbf{p})} \mathbf{A}$$

the set of points in general position.

Allowable black-box varieties

A variety $V(p)$ is allowable if it is a union of irreducible parts of finite intersections of basic allowable sets.

Denote by

$$\mathbf{G}(\mathbf{p}) = \mathbf{V}(\mathbf{p}) - \cup_{\text{allowable } \mathbf{A} \subset \mathbf{V}(\mathbf{p})} \mathbf{A}$$

the set of points in general position.

$V(p)$ unallowable $\quad\Rightarrow\quad G(p) \neq \emptyset$.

Black-box necessary condition

Theorem 1. $V(p)$ unallowable $\Rightarrow p$ cannot be evaluated accurately on $\mathbb{R}^n$ or on $\mathbb{C}^n$.

Theorem 2. On a domain $\mathcal{D}$, if $\text{Int}(\mathcal{D}) \cap G(p) \neq \emptyset$, p cannot be evaluated accurately.

Black-box sufficient conditions

All q_j are assumed to be irreducible.

Theorem. If $V(p)$ is a union of intersections of sets Z_i, S_{ij}, D_{ij}, and $V(q_j)$, then p is accurately evaluable over $\mathbb{C}^n$.

Corollary. If all q_j are affine, then p is accurately evaluable over $\mathbb{C}^n$ iff $V(p)$ is allowable.

NLA consequences

- $V(\det(\text{Toeplitz}))$ contains irreducible factors of arbitrarily large degree $\Rightarrow$ no set of black-boxes of bounded degree will be sufficient for accurate evaluation $\Rightarrow$ need arbitrary precision arithmetic to do NLA accurately on Toeplitz matrices.

NLA consequences

- $V(\det(\text{Toeplitz}))$ contains irreducible factors of arbitrarily large degree ⇒ no set of black-boxes of bounded degree will be sufficient for accurate evaluation ⇒ need arbitrary precision arithmetic to do NLA accurately on Toeplitz matrices.
- Same argument works for some other classes of structured matrices

NLA consequences

- $V(\det(\text{Toeplitz}))$ contains irreducible factors of arbitrarily large degree $\Rightarrow$ no set of black-boxes of bounded degree will be sufficient for accurate evaluation $\Rightarrow$ need arbitrary precision arithmetic to do NLA accurately on Toeplitz matrices.
- Same argument works for some other classes of structured matrices
- **Conjecture.** If the class of structured matrices has displacement rank ≥ 2, then accurate evaluation will not always be possible.

Type of matrix		$\det A$	A^{-1}	Any minor	LDU	SVD	Sym EVD
Acyclic (bidiagonal and other)		n	n^2	n	$\leq n^2$	n^3	N/A
Total Sign Compound (TSC)		n	n^3	n	n^4	n^4	n^4
Diagonally Scaled Totally Unimodular (DSTU)		n^3	n^5	n^3	n^3	n^3	n^3
Weakly diagonally dominant M-matrix		n^3	n^3	**No**	n^3	n^3	n^3
Displacement Rank One	Cauchy	n^2	n^2	n^2	$\leq n^3$	n^3	n^3
	Vandermonde	n^2	**No**	**No**	**No**	n^3	n^3
	Polynomial Vandermonde	n^2	**No**	**No**	**No**	$*$	$*$

Open problems / Future work

- **Complete** the decision procedure (analyze the dominant terms) when the domain is $\mathbb{R}^n$ and $V(p)$ allowable.

Open problems / Future work

- **Complete** the decision procedure (analyze the dominant terms) when the domain is $\mathbb{R}^n$ and $V(p)$ allowable.
- **Narrow** the necessity and sufficiency conditions for the black-box case

Open problems / Future work

- **Complete** the decision procedure (analyze the dominant terms) when the domain is $\mathbb{R}^n$ and $V(p)$ allowable.
- **Narrow** the necessity and sufficiency conditions for the black-box case
- **Extend** to semi-algebraic domains $\mathcal{D}$.

Open problems / Future work

- **Complete** the decision procedure (analyze the dominant terms) when the domain is $\mathbb{R}^n$ and $V(p)$ allowable.
- **Narrow** the necessity and sufficiency conditions for the black-box case
- **Extend** to semi-algebraic domains $\mathcal{D}$.
- **Incorporate** division, rational functions, perturbation theory.

Open problems / Future work

- **Extend** to interval arithmetic.

Open problems / Future work

- **Extend** to interval arithmetic.
- **Implement** decision procedure to “compile" an accurate evaluation program given $p(x)$, $\mathcal{D}$, and minimal set of “black boxes" (which could be constructed ad-hoc, with techniques *à la* **Priest**, **Shewchuk**, **Demmel**, **Hida**, etc.)

Post Scriptum

James Demmel, Ioana Dumitriu and O. H.

Toward accurate evaluation in rounded arithmetic,

arXiv.org math.NA/0508350; also

`http://www.cs.berkeley.edu/˜demmel,`

`http://www.cs.berkeley.edu/˜oholtz.`

Post Scriptum

James Demmel, Ioana Dumitriu and O. H.

Toward accurate evaluation in rounded arithmetic,

arXiv.org math.NA/0508350; also

`http://www.cs.berkeley.edu/˜demmel`,

`http://www.cs.berkeley.edu/˜oholtz`.

Thanks

for

your

attention!