Towards accurate polynomial evaluation

or

When can Numerical Linear Algebra be done accurately?

James Demmel
UC Berkeley
Math and EECS Depts.

Joint work with
Ioana Dumitriu and Olga Holtz (UC Berkeley)

Outline

1. Motivation and goal(s).
2. Model of arithmetic and setting.
3. What is *allowable* in classical arithmetic.
4. Results for classical arithmetic, real and complex.
5. What is *allowable* in black-box arithmetic.
6. Results for black-box arithmetic, real and complex.
7. Other models of arithmetic.
8. Open problems / Future work.

Outline

1. Motivation and goal(s).
2. Model of arithmetic and setting.
3. What is *allowable* in classical arithmetic.
4. Results for classical arithmetic, real and complex.
5. What is *allowable* in black-box arithmetic.
6. Results for black-box arithmetic, real and complex.
7. Other models of arithmetic.
8. Open problems / Future work.

Goal

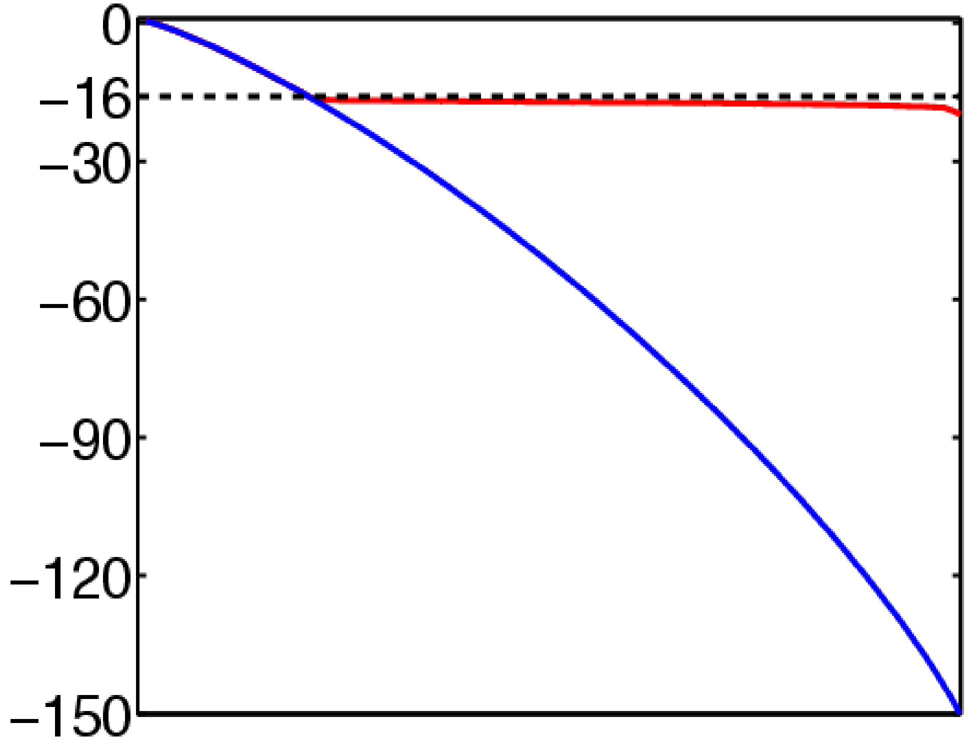
Given a family of structured matrices $M(x)$,
find **accurate** and **efficient** algorithms
to solve linear algebra problems (eg $y = \det M(x)$ or $y = \text{eig}(M(x))$),
or prove that none exist

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

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

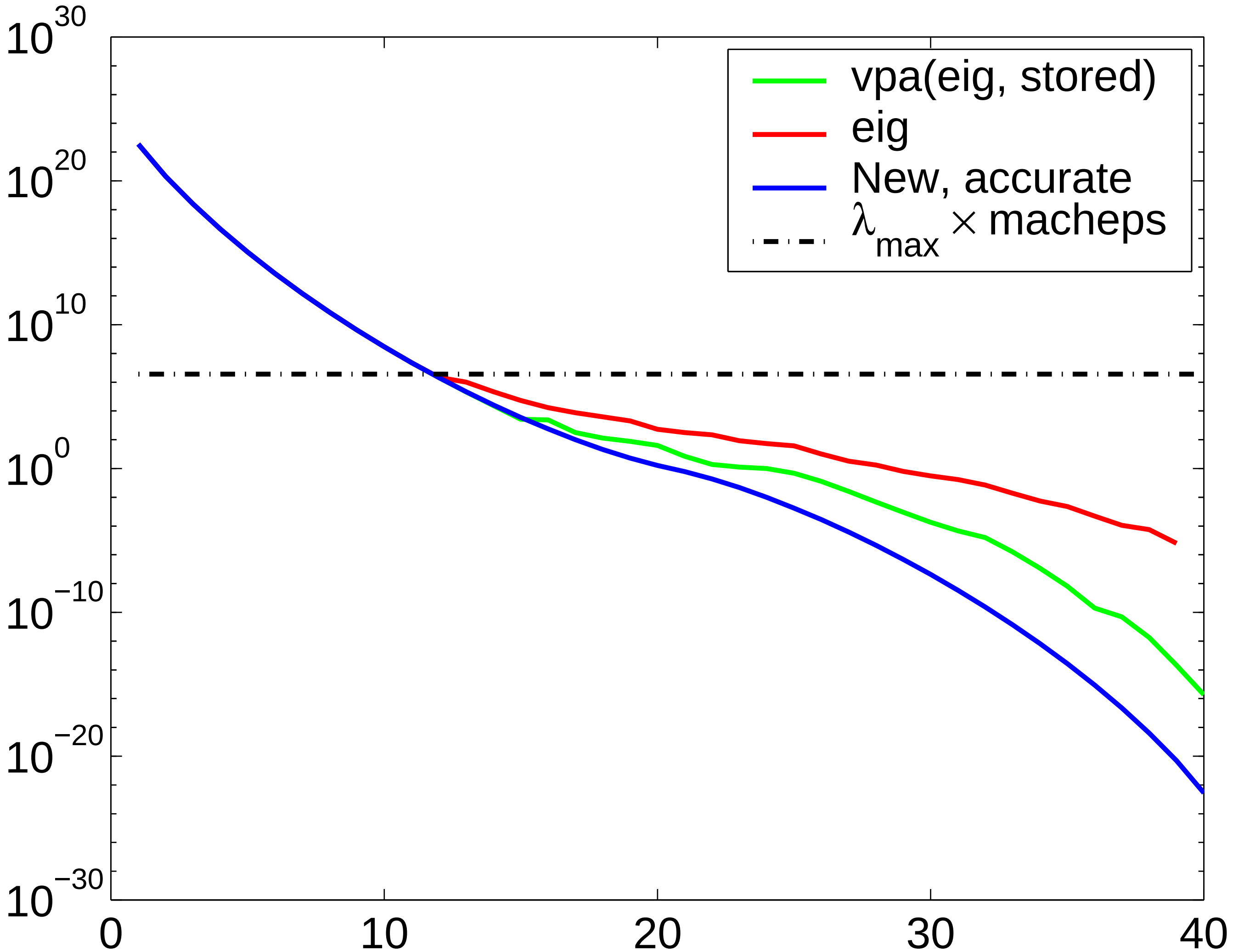
Efficiently means in polynomial time

$\mathbf{Log_{10}}$(Eigenvalues) of 50x50 Hilbert Matrix

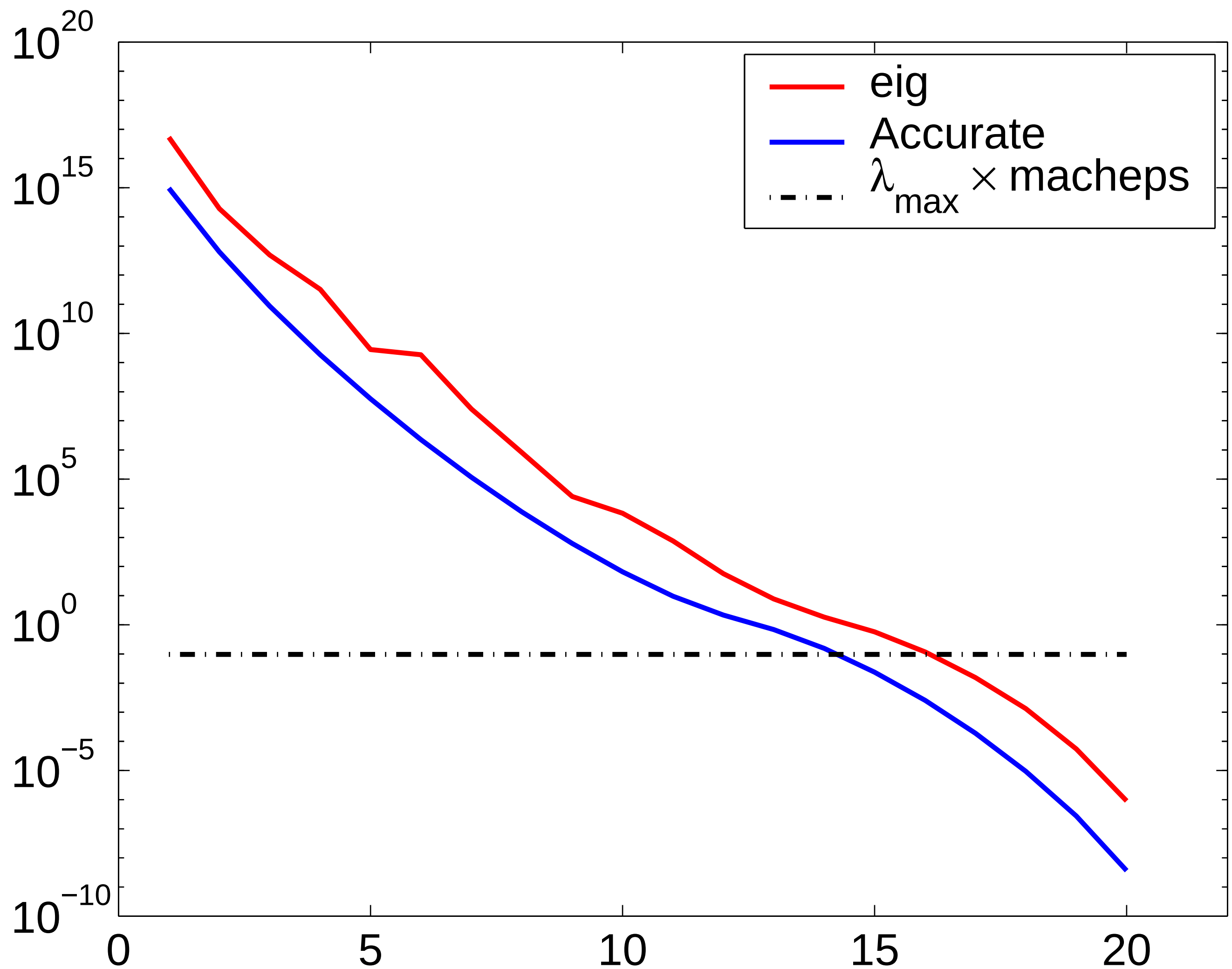


red line shows eigenvalues from conventional algorithm in 16 digits
blue line shows eigenvalues from new algorithm in 16 digits
Cost of guaranteed accuracy: $O(n^3 \log \kappa)$ vs $O(n^3 \log \log \kappa)$ where
κ = condition number

Eigenvalues of 40x40 Pascal Matrix



Eigenvalues of 20x20 Schur complement of 40x40 Vandermonde Matrix



None correct vs All correct!

General Structured Matrices

Type of matrix		det A	A^{-1}	Any minor	LDU	SVD	Sym EVD
Acyclic (bidiagonal and other)							
Total Sign Compound (TSC)							
Diagonally Scaled Totally Unimodular (DSTU)							
Weakly diagonally dominant M-matrix							
Displacement Rank One	Cauchy						
	Vandermonde						
	Polynomial Vandermonde						
Toeplitz							

General Structured Matrices

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	?	n^3	n^3	n^3
Displace-ment Rank One	Cauchy	n^2	n^2	n^2	$\leq n^3$	n^3	n^3
	Vandermonde	n^2	?	?	?	n^3	n^3
	Polynomial Vandermonde	n^2	?	?	?	?	?
Toeplitz		?	?	?	?	?	?

Totally Nonnegative Matrices

Type of Matrix	det A	A^{-1}	Any minor	Gauss. elim. NP	PP	CP	NE NP	Ax=b	SVD	Eig. Val.
Cauchy										
Vandermonde										
Generalized Vandermonde										
Any TN in Neville form										

Totally Nonnegative Matrices

Type of Matrix	det A	A^{-1}	Any minor	Gauss. elim. NP	Gauss. elim. PP	Gauss. elim. CP	NE NP	Ax=b	SVD	Eig. Val.
Cauchy	n^2	n^2	n^2	n^2	n^3	n^3	n^2	n^2	n^3	n^3
Vandermonde	n^2	n^3	n^3	n^2	n^2	poly	n^2	n^2	n^3	n^3
Generalized Vandermonde	n^2	n^3	poly	n^2	n^2	poly	n^2	n^2	n^3	n^3
Any TN in Neville form	n	n^3	n^3	n^3	n^3	n^3	0	n^2	n^3	n^3

poly = poly(n, λ), where λ = partition

(see talk by Plamen Koev, Tuesday 4pm)

Reduce Matrix problem to Polynomial problem

Theorem: Being able to compute $\det(M)$ accurately is *necessary* to be able to compute LDU, eig, SVD, ... accurately

Theorem: Being able to compute all minors of M accurately is *sufficient* for computing M^{-1}, LDU, SVD, ... accurately

(Sufficient conditions for computing eig(M) accurately unknown in nonsymmetric, non-totally positive case)

Goal - restated

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

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

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

Outline

1. Motivation and goal(s).
2. Model of arithmetic and setting.
3. What is *allowable* in classical arithmetic.
4. Results for classical arithmetic, real and complex.
5. What is *allowable* in black-box arithmetic.
6. Results for black-box arithmetic, real and complex.
7. Other models of arithmetic.
8. Open problems / Future work.

Traditional Model of Arithmetic

- $fl(a \otimes b) = (a \otimes b)(1+\delta)$, with arbitrary roundoff error $|\delta| < \epsilon \ll 1$
  - a, b and δ all real, or all complex
- Operations?

Traditional Model of Arithmetic

- $fl(a \otimes b) = (a \otimes b)(1+\delta)$, with arbitrary roundoff error $|\delta| < \epsilon \ll 1$
- Operations?
  - in classical arithmetic, $+$, $-$, $\times$; also exact negation;

Traditional Model of Arithmetic

- $fl(a \otimes b) = (a \otimes b)(1 + \delta)$, with arbitrary roundoff error $|\delta| < \epsilon \ll 1$
- Operations?
  - in classical arithmetic, $+$, $-$, $\times$; also exact negation;
  - How can we lose accuracy in this model?
    * OK to multiply or add positive numbers
    * OK to subtract exact numbers (initial data)
    * Accuracy may only be lost when subtracting approximate results:

$$
\begin{array}{rl}
 & .12345xxx \\
- & .12345yyy \\
\hline
 & .00000zzz
\end{array}
$$

Recognizing Accuracy

- **Ex: Compute** $p(x) = x_1 + x_2 + x_3$
  - **Try** $alg(x,\delta) = ((x_1 + x_2)(1+\delta_1) + x_3)(1+\delta_2)$

$$
\begin{aligned}
rel\_err(x,\delta) &= \frac{alg(x,\delta) - p(x)}{p(x)} \\
&= \frac{x_1 + x_2}{x_1 + x_2 + x_3}(\delta_1 + \delta_2 + \delta_1 \cdot \delta_2) + \frac{x_3}{x_1 + x_2 + x_3}(\delta_2)
\end{aligned}
$$

  - $\forall \epsilon > 0$, $rel\_err(x,\delta)$ **unbounded on an open subset of** (x,δ) **with** $|\delta_i| < \epsilon$
- **Generally:** $rel\_err(x,\delta) = \sum_r \frac{p_r(x)}{p(x)} \cdot q_r(\delta)$
  - **Each** $\frac{p_r(x)}{p(x)}$ **must be bounded near** $p(x) = 0$
- **Ex:** $p(x)$ **positive definite and homogeneous, degree** d
  - **If** $p_r(x)$ **also homogeneous, degree** d**, then** $\frac{p_r(x)}{p(x)}$ **bounded**

Traditional Model of Arithmetic

- $fl(a \otimes b) = (a \otimes b)(1+\delta)$, with arbitrary roundoff error $|\delta| < \epsilon \ll 1$
- Operations?
  - in classical arithmetic, $+$, $-$, $\times$; also exact negation;
  - in black-box arithmetic, above plus selected polynomial expressions
    - Ex: $x - yz$ (IBM's fused-multiply-add)
    - Ex: $wx - yz$ (using double-double)
    - Ex: small determinants (Shewchuk's Triangle)
    - Ex: dot products (using Priest or Demmel/Hida algs)

Traditional Model of Arithmetic

- $fl(a \otimes b) = (a \otimes b)(1+\delta)$, with arbitrary roundoff error $|\delta| < \epsilon \ll 1$
- Operations?
  - in classical arithmetic, $+$, $-$, $\times$; also exact negation;
  - in black-box arithmetic, above plus selected polynomial expressions
- Constants?

Availability of constants?

Example.

- Classical case:
  - without $\sqrt{2}$, we cannot compute

$$x^2 - 2 = (x - \sqrt{2})(x + \sqrt{2})$$

    accurately.
  – having no explicit constants no loss of generality for homogeneous, integer-coefficient polynomials.
- Black-box case:
  - any constants we choose can be accommodated via black-boxes

Traditional Model of Arithmetic.

- $fl(a \otimes b) = (a \otimes b)(1+\delta)$, with arbitrary roundoff error $|\delta| < \epsilon \ll 1$
- Operations?
  - in classical arithmetic, $+$, $-$, $\times$; also exact negation;
  - in black-box arithmetic, above plus selected polynomial expressions
- Constants? none in classical case, anything in black-box case.

Traditional Model of Arithmetic

- $fl(a \otimes b) = (a \otimes b)(1+\delta)$, with arbitrary roundoff error $|\delta| < \epsilon \ll 1$
- Operations?
  - in classical arithmetic, $+$, $-$, $\times$; also exact negation;
  - in black-box arithmetic, above plus selected polynomial expressions
- Constants? none in classical case, anything in black-box case.
- Algorithms?
  - exact answer in finite # of steps in absence of roundoff error

Traditional Model of Arithmetic

- $fl(a \otimes b) = (a \otimes b)(1+\delta)$, with arbitrary roundoff error $|\delta| < \epsilon \ll 1$
- Operations?
  - in classical arithmetic, $+$, $-$, $\times$; also exact negation;
  - in black-box arithmetic, above plus selected polynomial expressions
- Constants? none in classical case, anything in black-box case.
- Algorithms?
  - exact answer in finite # of steps in absence of roundoff error
  - branching based on comparisons

Traditional Model of Arithmetic

- $fl(a \otimes b) = (a \otimes b)(1 + \delta)$, with arbitrary roundoff error $|\delta| < \epsilon \ll 1$
- Operations?
  - in classical arithmetic, $+$, $-$, $\times$; also exact negation;
  - in black-box arithmetic, above plus selected polynomial expressions
- Constants? none in classical case, anything in black-box case.
- Algorithms?
  - exact answer in finite # of steps in absence of roundoff error
  - branching based on comparisons
  - non-determinism (because determinism is simulable)

Traditional Model of Arithmetic

- $fl(a \otimes b) = (a \otimes b)(1+\delta)$, with arbitrary roundoff error $|\delta| < \epsilon \ll 1$
- Operations?
  - in classical arithmetic, $+$, $-$, $\times$; also exact negation;
  - in black-box arithmetic, above plus selected polynomial expressions
- Constants? none in classical case, anything in black-box case.
- Algorithms?
  - exact answer in finite # of steps in absence of roundoff error
  - branching based on comparisons
  - non-determinism (because determinism is simulable)
  - domains to be $\mathbb{C}^n$ or $\mathbb{R}^n$ (but some domain-specific results).

Problem Restatement

◇ Notation:

– $p(x)$ 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.

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

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

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

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

$\forall\ |\delta_i| \leq \epsilon$ ... and for all rounding errors bounded by ϵ

$|p_{comp}(x, \delta) - p(x)| \leq \eta \cdot |p(x)|$ ... relative error is at most η

◇ Given $p(x)$ and $\mathcal{D}$, seek effective procedure ("compiler") to exhibit algorithm, or show one does not exist

Examples in classical arithmetic over $\mathbb{R}^n$ (none work over $\mathbb{C}^n$).

- $M_2(x,y,z) = z^6 + x^2 \cdot y^2 \cdot (x^2 + y^2 - 2 \cdot z^2)$
  - Positive definite and homogeneous, easy to evaluate accurately
- $M_3(x,y,z) = z^6 + x^2 \cdot y^2 \cdot (x^2 + y^2 - 3 \cdot z^2)$
  - Motzkin polynomial, nonnegative, zero at $|x| = |y| = |z|$

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

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

Sneak Peak.

The variety,

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

plays a necessary role.

Outline

1. Motivation and goal(s).
2. Model of arithmetic and setting.
3. What is *allowable* in classical arithmetic.
4. Results for classical arithmetic, real and complex.
5. What is *allowable* in black-box arithmetic.
6. Results for black-box arithmetic, real and complex.
7. Other models of arithmetic.
8. Open problems / Future work.

Allowable varieties in classical arithmetic

Define *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.

Denote by

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

the set of points *in general position.*

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

Outline

1. Motivation and goal(s).
2. Model of arithmetic and setting.
3. What is *allowable* in classical arithmetic.
4. Results for classical arithmetic, real and complex.
5. What is *allowable* in black-box arithmetic.
6. Results for black-box arithmetic, real and complex.
7. Other models of arithmetic.
8. Open problems / Future work.

Necessary condition on $V(p)$ for accurate evaluation of p

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

- V(det(Toeplitz)), UNALLOWABLE $\Rightarrow$ no accurate linear algebra for Toeplitz in classical arithmetic
- V(minor(Vandermonde)), UNALLOWABLE, but ok on positive orthant (TP matrices)

Necessary condition on $V(p)$, real and complex

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.

Sketch of proof.

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

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.

Necessary condition on $V(p)$, real and complex.

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.

Sketch of proof, cont'd.

Key fact: for a positive measure set of δs in δ-space, a zero output can be "traced back" down the tree to "allowable" condition ($x_i = 0$ 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 $p(y) \neq 0$).

In either case, the polynomial is not accurately evaluable arbitrarily close to x, q.e.d.

Sufficient condition on $V(p)$ for accurate evaluation of p, 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.

Sufficient Condition on $V(p)$ for accurate evaluation of p, 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.

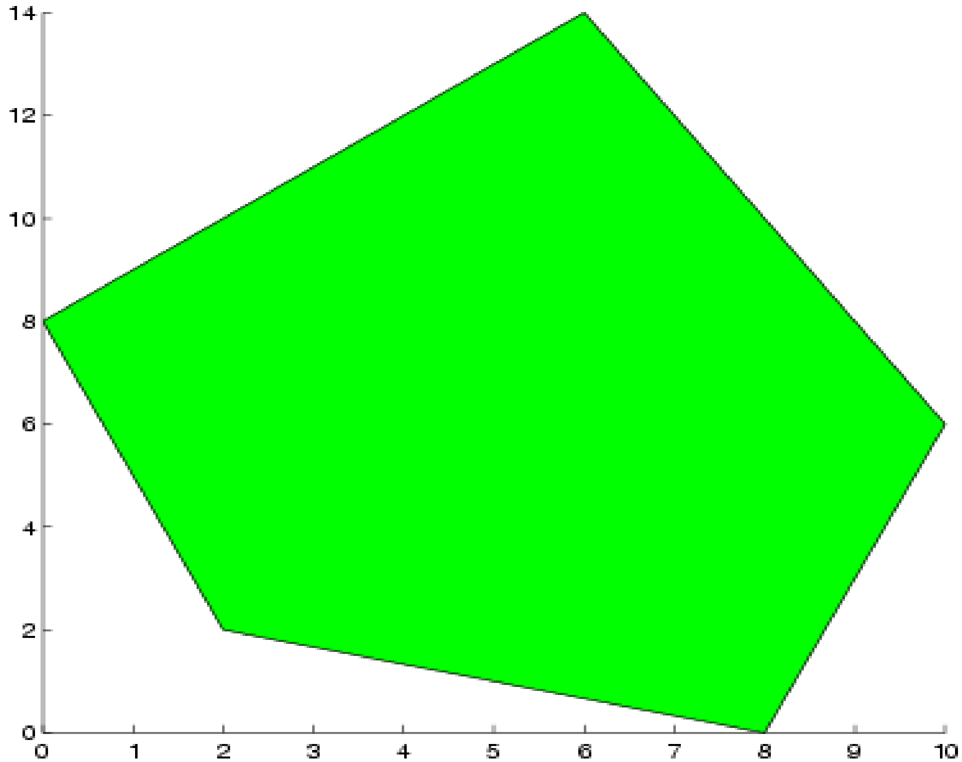
Corollary. If p is a complex multivariate polynomial, p is accurately evaluable iff p has integer coefficients and $V(p)$ is allowable.

Sufficient condition for accurate evaluation, real case.

Trickier... Allowability (or any condition) of $V(p)$ *not* sufficient:

- $q = (u^4 + v^4) + (u^2 + v^2)(x^2 + y^2 + z^2)$, $V(p) = \{u = v = 0\}$: 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!
- Say $p = (u^4 + v^4) + (u^2 + v^2)\hat{p}$ is "locally dominated" by $\hat{p}$ near $V(p)$
  - Accurate evaluabilty of p depends on that of $\hat{p}$
  - Leads to induction on hierarchy of varieties and polynomials defined by "dominance"
  - Need to formally define dominance
  - Induction is work in progress

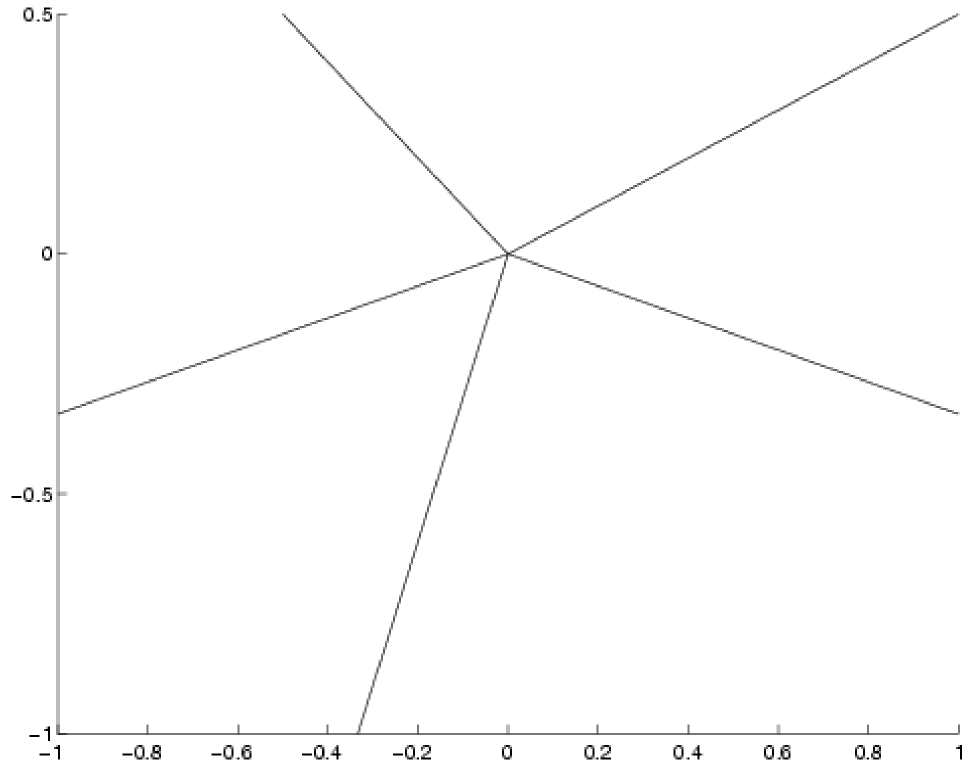
What is Dominance? Newton Polytope



$$p(x,y,z) = y^8z^{12} + x^2y^2z^{16} + x^8z^{12} + x^6y^{14} + x^{10}y^6z^4$$

Component of $V(p)$ where $\{x = y = 0\}$

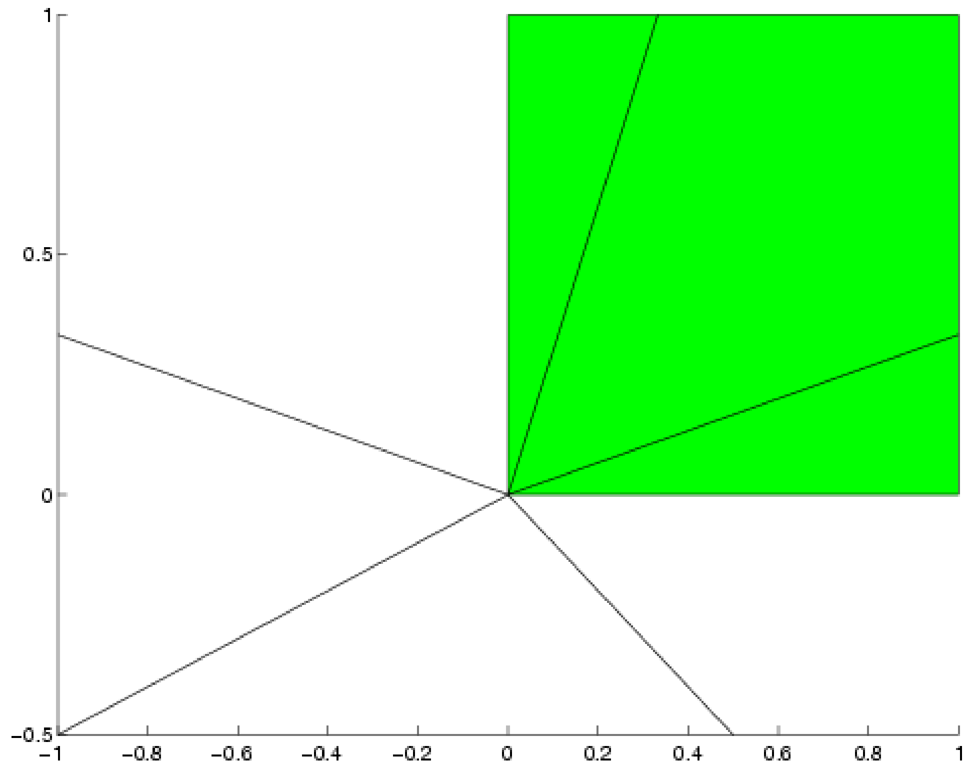
What is Dominance? Normal Fan



$$p(x,y,z) = y^8z^{12} + x^2y^2z^{16} + x^8z^{12} + x^6y^{14} + x^{10}y^6z^4$$

Component of $V(p)$ where $\{x = y = 0\}$

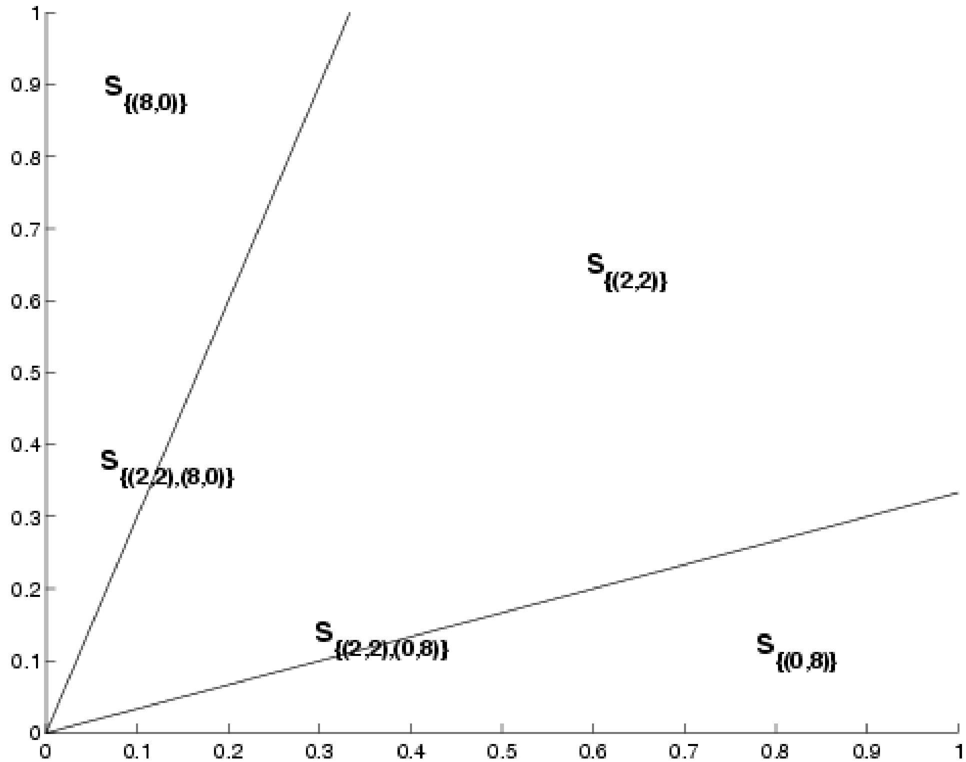
What is Dominance? First orthant of -(Normal Fan)



$$p(x, y, z) = y^8 z^{12} + x^2 y^2 z^{16} + x^8 z^{12} + x^6 y^{14} + x^{10} y^6 z^4$$

Component of $V(p)$ where $\{x = y = 0\}$

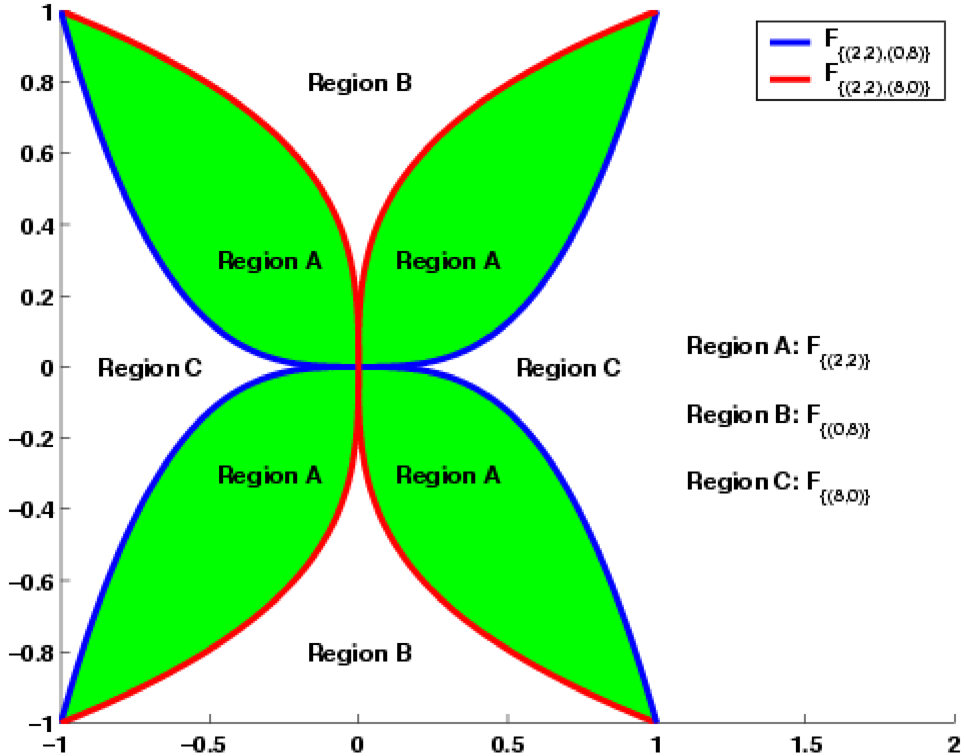
What is Dominance? Labeling cones by dominant terms



$$p(x,y,z) = y^8 z^{12} + x^2 y^2 z^{16} + x^8 z^{12} + x^6 y^{14} + x^{10} y^6 z^4$$

Component of $V(p)$ where $\{x = y = 0\}$

What is Dominance? (x, y) regions where different terms dominate - by exponentiating cones



$$p(x, y, z) = y^8z^{12} + x^2y^2z^{16} + x^8z^{12} + x^6y^{14} + x^{10}y^6z^4$$

Component of $V(p)$ where $\{x = y = 0\}$

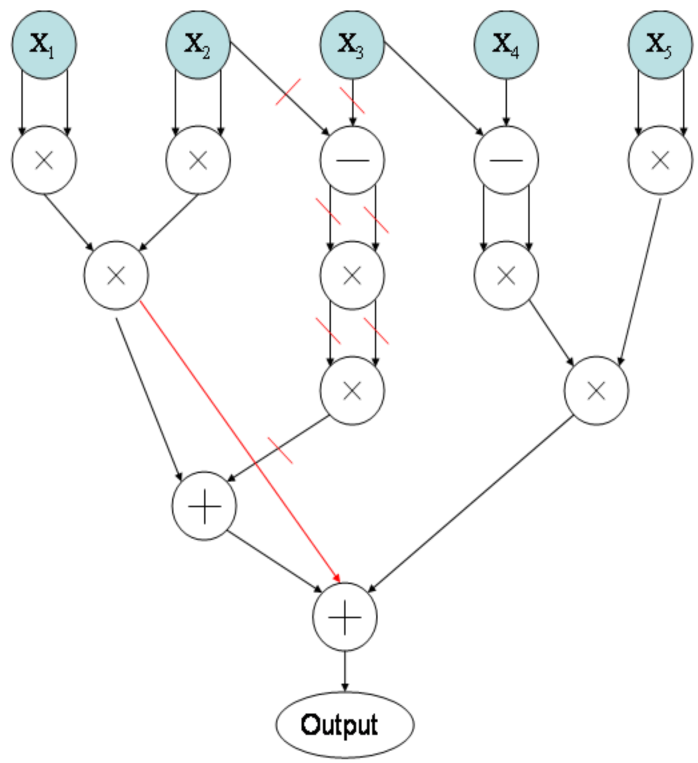
Sufficient condition for accurate evaluation, real case.

Trickier... Allowability *not* sufficient:

- $q = (u^4 + v^4) + (u^2 + v^2)(x^2 + y^2 + z^2)$, $V(p) = \{u = v = 0\}$: 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!
- Say $p = (u^4 + v^4) + (u^2 + v^2)\hat{p}$ is "locally dominated" by $\hat{p}$ near $V(p)$

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".

Sketch of showing that accurate evaluation of dominant terms is necessary for accurate evalution of p



Pruning is used to create accurate algorithm for any dominant term from accurate algorithm for p

Outline

1. Motivation and goal(s).
2. Model of arithmetic and setting.
3. What is *allowable* in classical arithmetic.
4. Results for classical arithmetic, real and complex.
5. What is *allowable* in black-box arithmetic.
6. Results for black-box arithmetic, real and complex.
7. Other models of arithmetic.
8. Open problems / Future work.

Allowable varieties in black-box arithmetic

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

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

Process A:

Step 1. repeat and/or negate, or 0 out some of the inputs,

Step 2. of the remaining variables, keep some symbolic, and find the variety in terms of the others.

Example: $q_1(x, y) = x - y$ 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 varieties in black-box arithmetic

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

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

Define *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 varieties in black-box arithmetic

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

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

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

Outline

1. Motivation and goal(s).
2. Model of arithmetic and setting.
3. What is *allowable* in classical arithmetic.
4. Results for classical arithmetic, real and complex.
5. What is *allowable* in black-box arithmetic.
6. Results for black-box arithmetic, real and complex.
7. Other models of arithmetic.
8. Open problems / Future work.

Necessary condition on $V(p)$ for accurate evaluation of p, real and complex

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.

Sufficiency condition, complex, for all q_j 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.

Corollary: If all q_j are affine, then p is accurately evaluable iff $V(p)$ is allowable.

General Structured Matrices

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
Displace-ment 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	$*$	$*$
Toeplitz		No	No	No	No	No	No

$*$ = it depends on polynomial (eg orthogonal ok)

Other linear algebra consequences

- Let $M_n(x)$ be a family of n-by-n structured matrices
- Thm: If $\det(M_n(x))$ has an irreducible factor $p_n(x)$ over $\mathbb{C}$ whose degree grows with n, then no set of "black-boxes" of bounded degree can accurately evaluate all $\det(M_n(x))$ over $\mathbb{C}$.
- Cor: $\det(\mathrm{Toeplitz}_n(x))$ cannot be evaluated accurately by any set of "black-boxes" of bounded degree over $\mathbb{C}$.
- Thm: If $V_{\mathbb{R}}(\det(M_n(x)))$ has a regular point at which the tangent depends on a growing number of coordinates, then no set of "black-boxes" of bounded degree can accurately evaluate all $\det(M_n(x))$ over $\mathbb{R}$.
- Cor: $\det(\mathrm{Toeplitz}_n(x))$ cannot be evaluated accurately by any set of "black-boxes" of bounded degree over $\mathbb{R}$.
- Accurate Toeplitz matrix computations need "infinite precision"
- What other $M_n(x)$ share these properties?

Outline

1. Motivation and goal(s).
2. Model of arithmetic and setting.
3. What is *allowable* in classical arithmetic.
4. Results for classical arithmetic, real and complex.
5. What is *allowable* in black-box arithmetic.
6. Results for black-box arithmetic, real and complex.
7. Other models of arithmetic
8. Open problems / Future work.

Other Models of arithmetic

- Other models of real arithmetic
  - Blum/Shub/Smale, Cucker/Smale, Pour-El/Richards
- Comparing Reals and Integers
  - Reals, with rounded arithmetic as described
    - Some (most) $p(x)$ impossible to evaluate accurately
  - Integers, with bit operations (usual Turing machine)
    - All $p(x)$ evaluable exactly, only question is cost
    - $\det(M)$ evaluable in polynomial time
    - Not a good bit model for real arithmetic

A bit model for Reals

- $x = m \cdot 2^e$, m and e integers, with bit operations
- Still a Turing machine, but inputs better capture reals
- Models floating point arithmetic
- All $p(x)$ evaluable exactly, but cost can be much higher
- Cost of arbitrary bit of $\prod_i (1 + 2^{e_i})$ same as permanent
- Cost of $x + y + z$ exponential unless done carefully (next slide)
- Cost of $\det(M)$ unknown, even for tridiagonal
- Cost of new matrix algorithms exponentially lower than conventional algorithms to guarantee same accuracy
  - $\log \log \kappa$ vs $\log \kappa$
  - $\log \log \kappa$ is polynomial in size of input

Adding Numbers in Bit Model of Arithmetic

- $x = m \cdot 2^e$ **where** m **and** e **are integers**
- **Cancellation is obstable to accuracy:**
  - $(2^e + 1) - 2^e$ **requires** e **bits of intermediate precision**
  - **Not polynomial time in size of input** $\log_2 e$
- **"Sort and Sum" Algorithm for** $S = \sum_{i=1}^{n} x_i$

  Sort so $|e_1| \geq |e_2| \geq \cdots \geq |e_n|$ **...** $|x_1| \geq \cdots \geq |x_n|$ **more than enough**
  $S = 0$ **...** $B > b$ **bits**
  for $i = 1$ **to** n
  $\quad S = S + x_i$

- **Thm: Let** $N = 1 + 2^{B-b} + 2^{B-2b} + \cdots 2^{B \bmod b} = 1 + \lceil \frac{2^{B-b}}{1-2^{-b}} \rceil$**. Then**
  - **If** $n \leq N$**, then** S **accurate to nearly** b **bits, despite any cancellation**
  - **If** $n \geq N + 2$**, then** S **may be completely wrong (wrong sign)**
  - **If** $n = N + 1$**, in between these cases, depending on underflow**
- **Ex:** x_i **double** ($b = 53$)**,** S **extended** ($B = 64$) $\Rightarrow N = 2049$

Outline.

1. Motivation and goal(s).
2. Model of arithmetic and setting.
3. What is *allowable* in classical arithmetic.
4. Results for classical arithmetic, real and complex.
5. What is *allowable* in black-box arithmetic.
6. Results for black-box arithmetic, real and complex.
7. Other models of arithmetic.
8. Open problems / Future work.

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 $\mathcal{D}$.
- **Apply** to more structured matrix classes
- **Incorporate** division, rational functions, perturbation theory.
  - **Conjecture** (Demmel, '04): Accurate evaluation is possible iff condition number has only certain simple singularities (depend on reciprocal distance to set of ill-posed problems).
- **Extend** to interval arithmetic.
- **Implement** decision procedure to "compile" an accurate evaluation program given $p(x)$, $\mathcal{D}$, and minimal set of "black boxes"