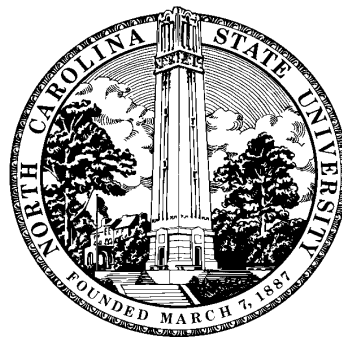


*The Seven Dwarfs of Symbolic Computation
and the Discovery of Reduced Symbolic Models*

Erich Kaltofen

North Carolina State University

google->kaltofen



Outline

The 7 Dwarfs of Symbolic Computation

Discovery of sparse rational function models

Solving overdetermined linear systems optimally

Berkeley 13 Dwarfs

http://view.eecs.berkeley.edu/wiki/Dwarf_Mine

“A dwarf is an **algorithmic method** that captures a **pattern** of computation and communication”

1. Dense Linear Algebra
2. Sparse Linear Algebra
3. Spectral Methods
4. N-Body Methods
5. Structured Grids
6. Unstructured Grids
7. MapReduce
8. Combinational Logic
9. Graph Traversal
10. Dynamic Programming
11. Backtrack and Branch-and-Bound
12. Graphical Models
13. Finite State Machines

How about Logic Programming, Symbolic Computation?

My 7 Dwarfs of Symbolic Computation

1. Exact linear algebra including algorithms for integer lattices

My 7 Dwarfs of Symbolic Computation

1. Exact linear algebra including algorithms for integer lattices
2. Exact polynomial and differential algebra, including polynomial arithmetic and computation of canonical forms such as Gröbner bases

My 7 Dwarfs of Symbolic Computation

1. Exact linear algebra including algorithms for integer lattices
2. Exact polynomial and differential algebra, including polynomial arithmetic and computation of canonical forms such as Gröbner bases
3. Inverse symbolic problems such as sparse interpolation and curve and surface parameterization

My 7 Dwarfs of Symbolic Computation

1. Exact linear algebra including algorithms for integer lattices
2. Exact polynomial and differential algebra, including polynomial arithmetic and computation of canonical forms such as Gröbner bases
3. Inverse symbolic problems such as sparse interpolation and curve and surface parameterization
4. Hybrid symbolic-numeric computation

My 7 Dwarfs of Symbolic Computation

1. Exact linear algebra including algorithms for integer lattices
2. Exact polynomial and differential algebra, including polynomial arithmetic and computation of canonical forms such as Gröbner bases
3. Inverse symbolic problems such as sparse interpolation and curve and surface parameterization
4. Hybrid symbolic-numeric computation
5. Tarski's algebraic theory of real geometry

My 7 Dwarfs of Symbolic Computation

1. Exact linear algebra including algorithms for integer lattices
2. Exact polynomial and differential algebra, including polynomial arithmetic and computation of canonical forms such as Gröbner bases
3. Inverse symbolic problems such as sparse interpolation and curve and surface parameterization
4. Hybrid symbolic-numeric computation
5. Tarski's algebraic theory of real geometry
6. Computation of closed form solutions to, e.g., sums, integrals, differential equations

My 7 Dwarfs of Symbolic Computation

1. Exact linear algebra including algorithms for integer lattices
2. Exact polynomial and differential algebra, including polynomial arithmetic and computation of canonical forms such as Gröbner bases
3. Inverse symbolic problems such as sparse interpolation and curve and surface parameterization
4. Hybrid symbolic-numeric computation
5. Tarski's algebraic theory of real geometry
6. Computation of closed form solutions to, e.g., sums, integrals, differential equations
7. Rewrite rule systems (simplification and theorem proving) and computational group theory

My 7 Dwarfs of Symbolic Computation

1. Exact linear algebra including algorithms for integer lattices
2. Exact polynomial and differential algebra, including polynomial arithmetic and computation of canonical forms such as Gröbner bases
3. Inverse symbolic problems such as sparse interpolation and curve and surface parameterization
4. Hybrid symbolic-numeric computation
5. Tarski's algebraic theory of real geometry
6. Computation of closed form solutions to, e.g., sums, integrals, differential equations
7. Rewrite rule systems (simplification and theorem proving) and computational group theory

Model Discovery and Verification: $1+2+3+4+5$

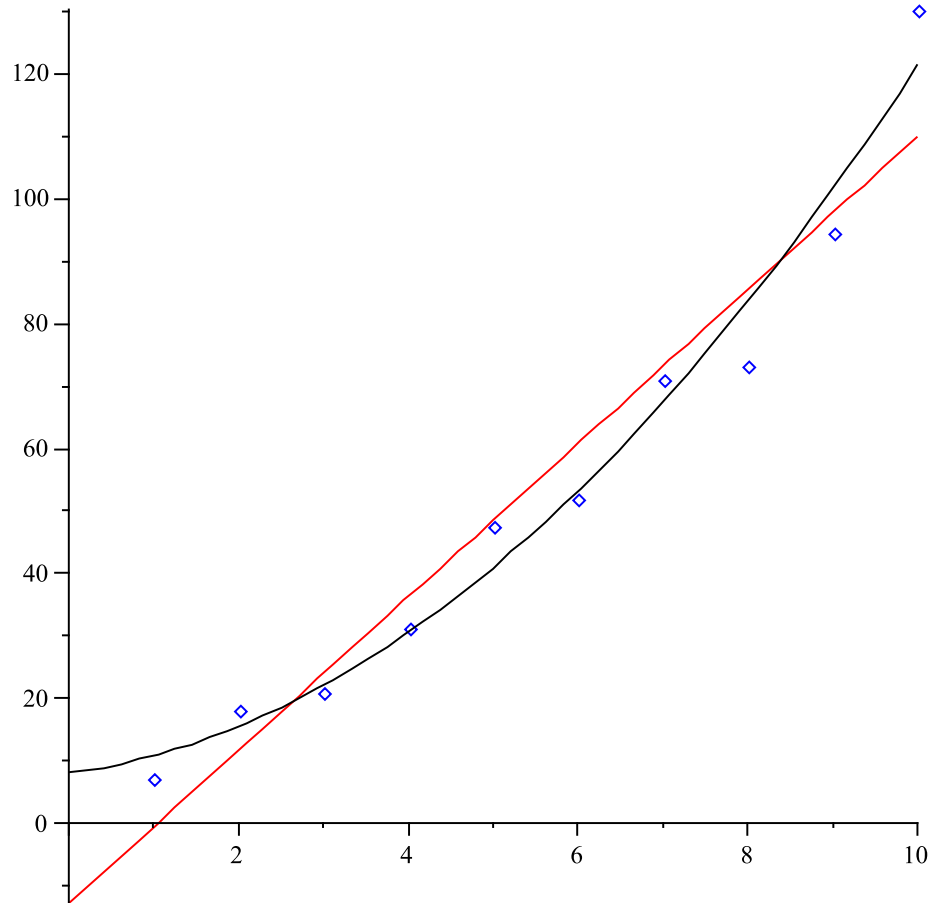
Outline

The 7 Dwarfs of Symbolic Computation

Discovery of sparse rational function models

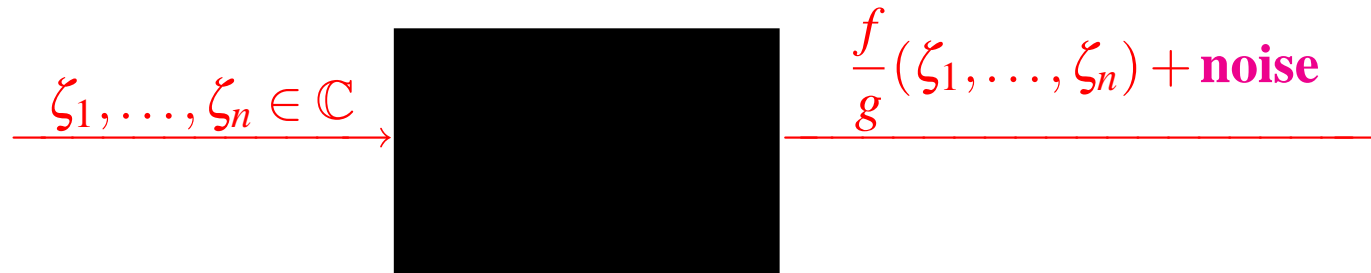
Solving overdetermined linear systems optimally

Model Discovery Example



Linear or quadratic best fit? What if best model is $\frac{2.5x^7y^{10} + 1.3}{x^2 - y^9}$?

Sparse rational function models [Kaltofen, Yang, Zhi '07]



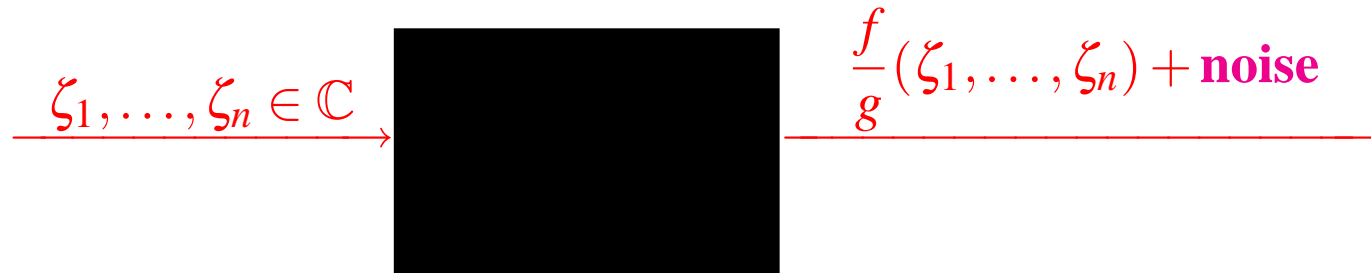
$$f, g \in \mathbb{C}[x_1, \dots, x_n], \text{GCD}(f, g) = 1$$

By sampling black box, compute sparse representation

$$\frac{\sum_{j=1}^{t_f} \tilde{a}_j x_1^{d_{j,1}} \cdots x_n^{d_{j,n}}}{\sum_{k=1}^{t_g} \tilde{b}_k x_1^{e_{k,1}} \cdots x_n^{e_{k,n}}} = \frac{\tilde{f}}{\tilde{g}}, \quad \tilde{a}_j \neq 0, \tilde{b}_k \neq 0$$

Note: Terms are **not** known.

Sparse rational function models [Kaltofen, Yang, Zhi '07]



$$f, g \in \mathbb{C}[x_1, \dots, x_n], \text{GCD}(f, g) = 1$$

By sampling black box, compute sparse representation

$$\frac{\sum_{j=1}^{t_f} \tilde{a}_j x_1^{d_{j,1}} \cdots x_n^{d_{j,n}}}{\sum_{k=1}^{t_g} \tilde{b}_k x_1^{e_{k,1}} \cdots x_n^{e_{k,n}}} = \frac{\tilde{f}}{\tilde{g}}, \quad \tilde{a}_j \neq 0, \tilde{b}_k \neq 0$$

Note: Terms are **not** known.

ZNIPR algorithm: Sparsity by a **numeric** Zippel-Schwartz lemma

Numeric noise in values by condition

numbers in **random structured matrices**

Zippel Exact and Numerical Interpolation of Rational Functions

Idea:

- Variable by variable interpolation
Term supports like [Kaltofen 1986], instead of [Zippel 1979]
- Early termination [cf. Kaltofen and Lee 2003]
- Structured Total Least Norm (STLN) method employed in numerical case [cf. Kaltofen, Yang, Zhi 2005, 2006]
 - STLN method to decide the support of the numerator and denominator.
 - STLN method to compute the coefficients corresponding to the support.

Ingredients of ZNIPR algorithm

Probabilistic analysis of exact method

Numerical Zippel/Schwartz lemma

Condition numbers of randomized matrices

Example (Exact Case)

Given the black box of the rational function f/g

$$f = x_1^3 + 3x_1x_2^2, \quad g = 2x_1^3 + 3x_2,$$

and the degree bounds $\bar{d} = 4, \bar{e} = 4$. Suppose

$$f_1 = f(x_1, a) = b_1x_1^3 + b_2x_1, \quad g_1 = g(x_1, a) = b_3x_1^3 + b_4.$$

and $\bar{D}_2 = \{1, x_2^2\}, \bar{E}_2 = \{1, x_2\}$, where $b_1, b_2, b_3, b_4 \in K \setminus \{0\}$.

Example (Exact Case)

Given the black box of the rational function f/g

$$f = x_1^3 + 3x_1x_2^2, \quad g = 2x_1^3 + 3x_2,$$

and the degree bounds $\bar{d} = 4, \bar{e} = 4$. Suppose

$$f_1 = f(x_1, a) = b_1x_1^3 + b_2x_1, \quad g_1 = g(x_1, a) = b_3x_1^3 + b_4.$$

and $\bar{D}_2 = \{1, x_2^2\}, \bar{E}_2 = \{1, x_2\}$, where $b_1, b_2, b_3, b_4 \in K \setminus \{0\}$.

The sets of the possible terms of f and g are

$$D_2 = \{x_1, x_1^3, x_1x_2^2\}, \quad E_2 = \{1, x_1^3, x_1^3x_2, x_2\}.$$

Example (Exact Case)

f and g can be represented as

$$f = y_1 x_1 + y_2 x_1^3 + y_3 x_1 x_2^2, \quad g = z_1 + z_2 x_1^3 + z_3 x_1^3 x_2 + z_4 x_2.$$

Pick random points $p_1, p_2 \in \mathbb{K}$ and compute the values:

$$\gamma_l = \frac{f(p_1^l, p_2^l)}{g(p_1^l, p_2^l)} \in \mathbb{K} \setminus \{0, \infty\}, \quad l = 0, 1, \dots, L-1.$$

⇓

$$\begin{bmatrix} 1 & 1 & 1 & \gamma_0 & \gamma_0 & \gamma_0 & \gamma_0 \\ p_1 & p_1^3 & p_1 p_2^2 & \gamma_1 & \gamma_1 p_1^3 & \gamma_1 p_1^3 p_2 & \gamma_1 p_2 \\ p_1^2 & (p_1^3)^2 & (p_1 p_2^2)^2 & \gamma_2 & \gamma_2 (p_1^3)^2 & \gamma_2 (p_1^3 p_2)^2 & \gamma_2 p_2^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_1^{L-1} & (p_1^3)^{L-1} & (p_1 p_2^2)^{L-1} & \gamma_{L-1} & \gamma_{L-1} (p_1^3)^{L-1} & \gamma_{L-1} (p_1^3 p_2)^{L-1} & \gamma_{L-1} p_2^{L-1} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Exact Probabilistic Analysis

- Rank deficiency of G is 1 if $L \geq |D_i| \cdot |E_i|$ w.h.p.
- Using smaller L : Increment L by 1 until rank deficiency of G is 1.

Table of Exact Interpolation

<i>Ex.</i>	<i>Coeff. Range</i>	d_f, d_g	n	t_f, t_g	mod	$N(\text{ZEIPR})$	$N(\text{KY}'07)$
1	$[-10,10]$	3,3	2	6,6	503	50	221
2	$[-10,10]$	5,2	4	6,3	1009	97	339
3	$[-20,20]$	2,4	6	2,5	120011	126	357
4	$[-20,20]$	1,6	8	4,8	8009	270	777
5	$[-30,30]$	10,5	10	7,4	4001	526	2246
6	$[-10,10]$	15,15	15	15,15	50021	2164	17120
7	$[-10,10]$	20,20	20	20,20	50021	3842	38682
8	$[-30,30]$	30,15	5	20,10	10007	1183	12896
9	$[-50,50]$	50,50	50	50,50	1000003	30405	603638
10	$[-10,10]$	2,8	90	10,50	1000003	10101	75082

Numerical Zippel/Schwartz Lemma

Let

$$0 \neq \Delta(\alpha_1, \dots, \alpha_s) \in \mathbb{Z}[\mathbf{i}][\alpha_1, \dots, \alpha_s], \quad \mathbf{i} = \sqrt{-1},$$

$\zeta_j = \exp\left(\frac{2\pi\mathbf{i}}{P_j}\right) \in \mathbb{C}$, $P_j \in \mathbb{Z}_{\geq 3}$ distinct prime numbers $\forall 1 \leq j \leq s$
 [cf. Giesbrecht, Labahn, Lee 2007]

Suppose $\Delta(\zeta_1, \dots, \zeta_s) \neq 0$ (use algebraic lemma to enforce)

Then for random integers R_j with $1 \leq R_j < P_j$

$$\text{Expected value} \left\{ \left| \Delta(\zeta_1^{R_1}, \dots, \zeta_s^{R_s}) \right| \right\} \geq 1.$$

Numerical Zippel/Schwartz Lemma

Let

$$0 \neq \Delta(\alpha_1, \dots, \alpha_s) \in \mathbb{Z}[\mathbf{i}][\alpha_1, \dots, \alpha_s], \quad \mathbf{i} = \sqrt{-1},$$

$\zeta_j = \exp\left(\frac{2\pi\mathbf{i}}{P_j}\right) \in \mathbb{C}$, $P_j \in \mathbb{Z}_{\geq 3}$ distinct prime numbers $\forall 1 \leq j \leq s$
 [cf. Giesbrecht, Labahn, Lee 2007]

Suppose $\Delta(\zeta_1, \dots, \zeta_s) \neq 0$ (use algebraic lemma to enforce)

Then for random integers R_j with $1 \leq R_j < P_j$

$$\text{Expected value} \left\{ \left| \Delta(\zeta_1^{R_1}, \dots, \zeta_s^{R_s}) \right| \right\} \geq 1.$$

Useful for support pruning.

Also need estimates on matrix condition numbers?

Condition numbers of random $n \times n$ matrices

Entries from standard Gaussian distribution:

$$\text{Expected value} \{ \log \kappa_2(G_n) \} < \log(n) + 2.258$$

[Edelman 1988; Chen and Dongarra 2005]

Condition numbers of random $n \times n$ matrices

Entries from standard Gaussian distribution:

$$\text{Expected value} \{ \log \kappa_2(G_n) \} < \log(n) + 2.258$$

[Edelman 1988; Chen and Dongarra 2005]

Random discrete noise:

$$\kappa_2(A + R_n) = n^{O(1)} \text{ with high probability}$$

[Spielman and Teng 2004; Tao and Vu 2007]

Condition numbers of random $n \times n$ matrices

Entries from standard Gaussian distribution:

$$\text{Expected value} \{ \log \kappa_2(G_n) \} < \log(n) + 2.258$$

[Edelman 1988; Chen and Dongarra 2005]

Random discrete noise:

$$\kappa_2(A + R_n) = n^{O(1)} \text{ with high probability}$$

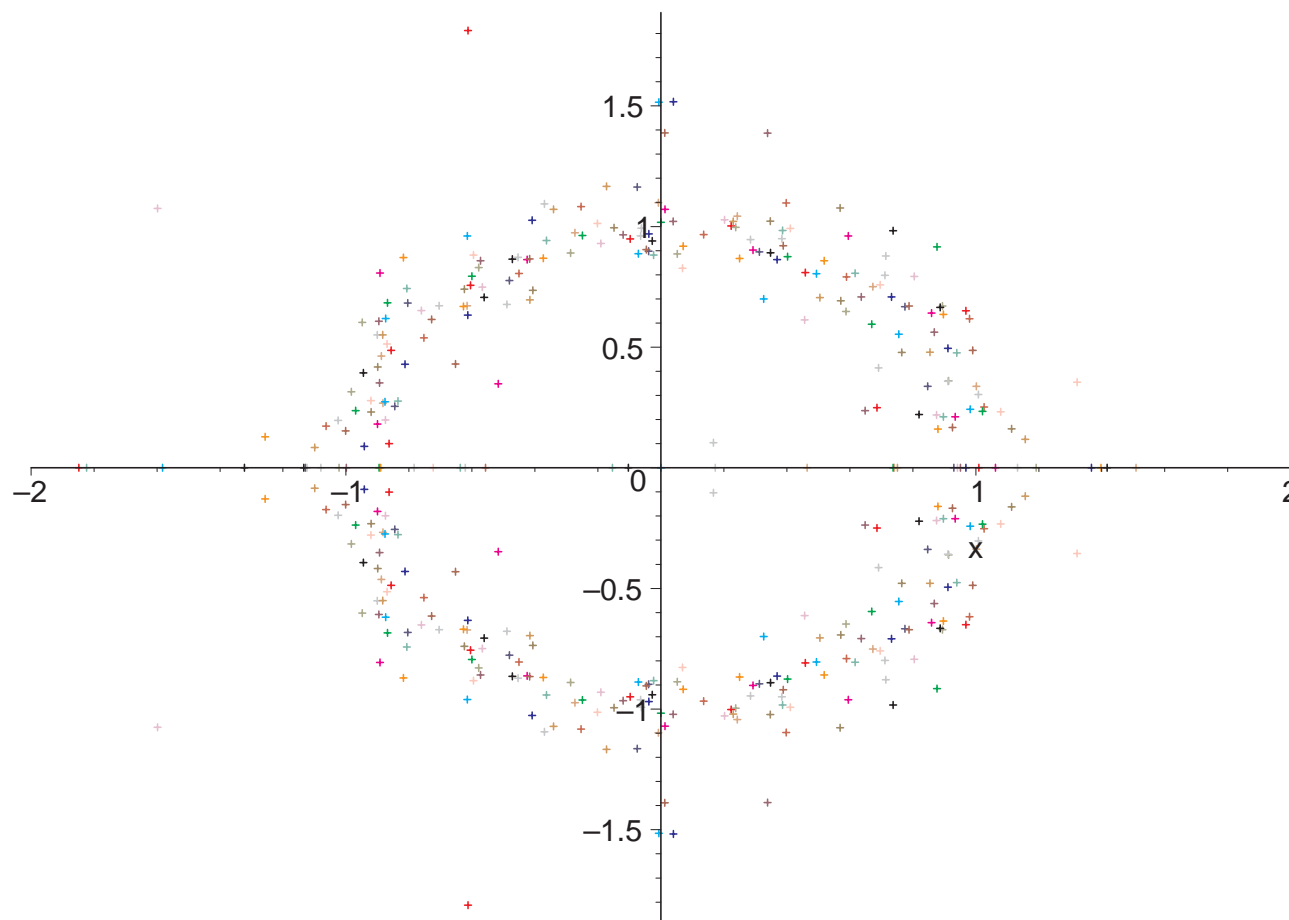
[Spielman and Teng 2004; Tao and Vu 2007]

“... one rarely encounters ill-conditioned matrices in practice”

Root distribution 20 polynomials of degree 20

Uniformly distributed coefficients

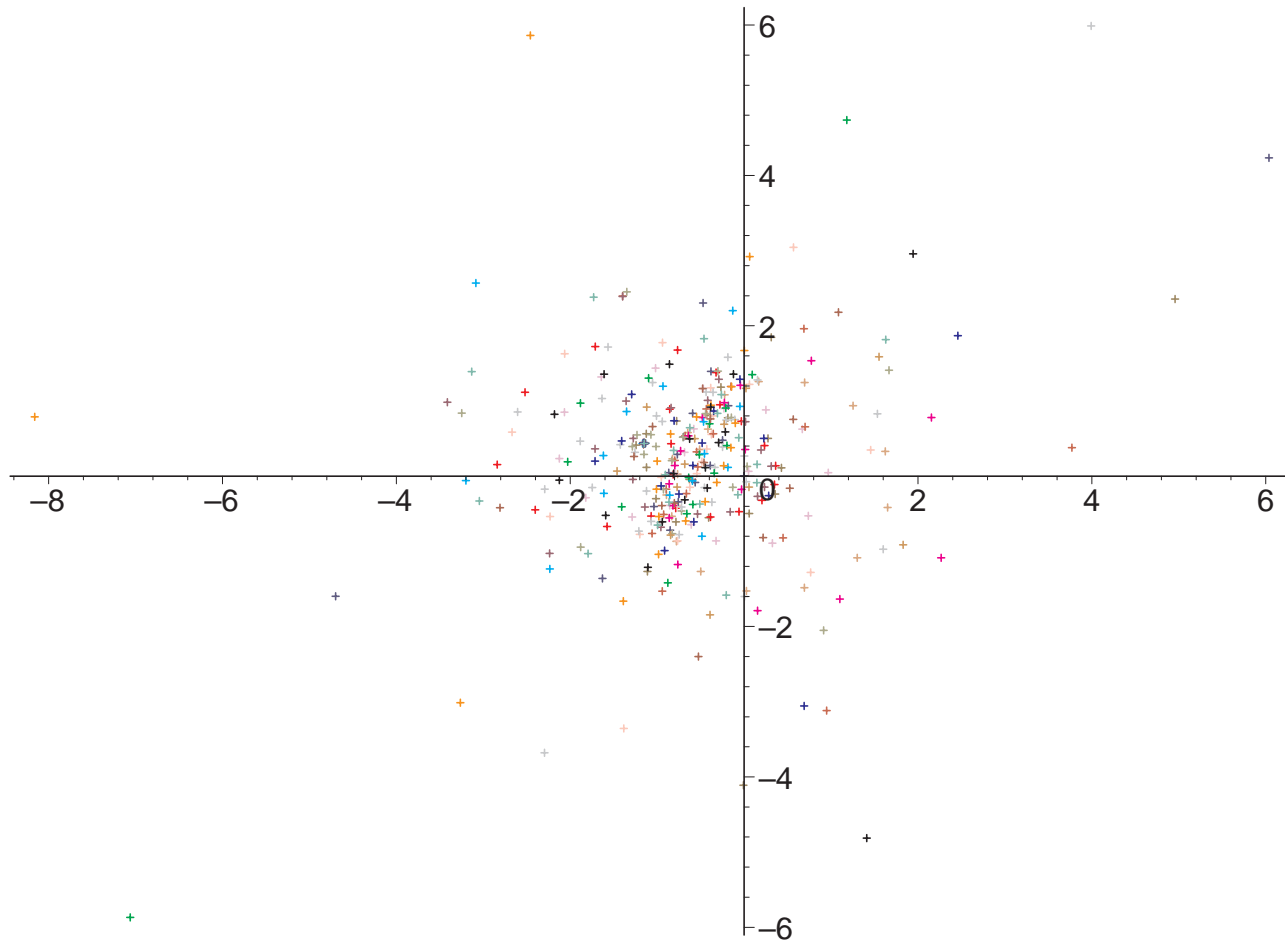
\implies small structured condition number



Root distribution 20 polynomials of degree 20

Binomially distributed coefficients

\implies large structured condition number



Conditioning of random projections

[Kaltofen and Trager 1988; Kaltofen and Yang 2007]

$$\frac{\sum_{j=1}^{t_f} a_j x_1^{d_{j,1}} (\xi_2 x_1 + \eta_2)^{d_{j,2}} \cdots (\xi_n x_1 + \eta_n)^{d_{j,n}}}{\sum_{k=1}^{t_g} b_k x_1^{e_{k,1}} (\xi_2 x_1 + \eta_2)^{e_{k,2}} \cdots (\xi_n x_1 + \eta_n)^{e_{k,n}}}, \quad \xi_i, \eta_i \in S \text{ random}$$

large structured condition numbers of Sylvester matrices

[Zippel 1979; ours here]

$$\frac{\sum_{j=1}^{t_f} a_j x_1^{d_{j,1}} \xi_2^{d_{j,2}} \cdots \xi_n^{d_{j,n}}}{\sum_{k=1}^{t_g} b_k x_1^{e_{k,1}} \xi_2^{e_{k,2}} \cdots \xi_n^{e_{k,n}}}, \quad \xi_i \in S \text{ random}$$

small structured condition numbers of Sylvester matrices

Well-conditioning of arising Fourier matrices to be established.

Table of Numerical Interpolation

<i>Ex.</i>	<i>Random Noise</i>	d_f, d_g	t_f, t_g	n	N	<i>error</i> (<i>ZNIPR</i>)
1	$10^{-5} \sim 10^{-3}$	1, 1	2, 2	2	136	9.46659e-7
2	$10^{-5} \sim 10^{-3}$	2, 2	3, 3	2	140	9.98831e-7
3	$10^{-5} \sim 10^{-3}$	1, 4	2, 4	3	233	3.70021e-7
4	$10^{-6} \sim 10^{-4}$	5, 2	10, 6	3	308	2.38743e-8
5	$10^{-7} \sim 10^{-5}$	7, 7	25, 25	5	1096	1.66383e-8
6	$10^{-7} \sim 10^{-5}$	10, 3	15, 5	8	2553	2.5896 e-5
8	$10^{-7} \sim 10^{-5}$	20, 20	7, 7	15	3107	1.90755e-9
9	$10^{-8} \sim 10^{-6}$	30, 30	6, 6	20	5213	3.77521e-12
11	$10^{-8} \sim 10^{-6}$	60, 60	7, 7	4	1987	5.31254e-12
12	$10^{-8} \sim 10^{-6}$	80, 80	6, 6	10	5127	5.10492e-11
13	$10^{-8} \sim 10^{-6}$	60, 0	6, 1	20	2862	2.00141e-12

Outline

The 7 Dwarfs of Symbolic Computation

Discovery of sparse rational function models

Solving overdetermined linear systems optimally

Fast matrix multiplication

Strassen's [1969] $O(n^{2.81})$ matrix multiplication algorithm

$$m_1 \leftarrow (a_{1,2} - a_{2,2})(b_{2,1} - b_{2,2})$$

$$m_2 \leftarrow (a_{1,1} + a_{2,2})(b_{1,1} + b_{2,2})$$

$$m_3 \leftarrow (a_{1,1} - a_{2,1})(b_{1,1} + b_{1,2})$$

$$m_4 \leftarrow (a_{1,1} + a_{1,2})b_{2,2} \quad \left| \quad a_{1,1}b_{1,1} + a_{1,2}b_{2,1} = m_1 + m_2 - m_4 + m_6$$

$$m_5 \leftarrow a_{1,1}(b_{1,2} - b_{2,2}) \quad \left| \quad a_{1,1}b_{1,2} + a_{1,2}b_{2,2} = m_4 + m_5$$

$$m_6 \leftarrow a_{2,2}(b_{2,1} - b_{1,1}) \quad \left| \quad a_{2,1}b_{1,1} + a_{2,2}b_{2,1} = m_6 + m_7$$

$$m_7 \leftarrow (a_{2,1} + a_{2,2})b_{1,1} \quad \left| \quad a_{2,1}b_{1,2} + a_{2,2}b_{2,2} = m_2 - m_3 + m_5 - m_7$$

Coppersmith and Winograd [1990]: $O(n^{2.38})$

Problems reducible to matrix multiplication:

- linear system solving, determinant [Bunch and Hopcroft 1974],
- characteristic polyn. [Keller-Gehrig'85, Pernet&Storjohann'07],
- rational canonical form [Giesbrecht'92],
- factoring in $\mathbb{Z}_2[x]$ [Berlekamp'69, Kaltofen and Shoup'95]

Fast matrix multiplication

Strassen's [1969] $O(n^{2.81})$ matrix multiplication algorithm

$$m_1 \leftarrow (a_{1,2} - a_{2,2})(b_{2,1} - b_{2,2})$$

$$m_2 \leftarrow (a_{1,1} + a_{2,2})(b_{1,1} + b_{2,2})$$

$$m_3 \leftarrow (a_{1,1} - a_{2,1})(b_{1,1} + b_{1,2})$$

$$m_4 \leftarrow (a_{1,1} + a_{1,2})b_{2,2} \quad \left| \quad a_{1,1}b_{1,1} + a_{1,2}b_{2,1} = m_1 + m_2 - m_4 + m_6$$

$$m_5 \leftarrow a_{1,1}(b_{1,2} - b_{2,2}) \quad \left| \quad a_{1,1}b_{1,2} + a_{1,2}b_{2,2} = m_4 + m_5$$

$$m_6 \leftarrow a_{2,2}(b_{2,1} - b_{1,1}) \quad \left| \quad a_{2,1}b_{1,1} + a_{2,2}b_{2,1} = m_6 + m_7$$

$$m_7 \leftarrow (a_{2,1} + a_{2,2})b_{1,1} \quad \left| \quad a_{2,1}b_{1,2} + a_{2,2}b_{2,2} = m_2 - m_3 + m_5 - m_7$$

Coppersmith and Winograd [1990]: $O(n^{2.38})$

Problems reducible to matrix multiplication:

- linear system solving, determinant [Bunch and Hopcroft 1974],
- characteristic polyn. [Keller-Gehrig'85, Pernet&Storjohann'07],
- rational canonical form [Giesbrecht'92],
- factoring in $\mathbb{Z}_2[x]$: $n^{1.5+o(1)}$ [Umans'07, w/o FMM]

Matrix preconditioners

Theorem [Kaltofen and Saunders '91]

Let $A \in K^{n \times n}$ with $r = \text{rank}(A)$, $S \subseteq K$ with $|S| < \infty$.

Probab. (the first r rows of

$$\begin{bmatrix} 1 & u_2 & u_3 & \dots & u_n \\ 0 & 1 & u_2 & \dots & u_{n-1} \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & u_2 \\ 0 & \dots & & 0 & 1 \end{bmatrix} \cdot A$$

are lin. indep. | $u_i \in S$ uniformly randomly)

$$\geq 1 - \frac{r}{|S|}$$

Matrix preconditioners

Theorem [Kaltofen and Saunders '91]

Let $A \in K^{n \times n}$ with $r = \text{rank}(A)$, $S \subseteq K$ with $|S| < \infty$.

Probab. (the first r rows of

$$\begin{bmatrix} 1 & u_2 & u_3 & \dots & u_n \\ 0 & 1 & u_2 & \dots & u_{n-1} \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & u_2 \\ 0 & \dots & & 0 & 1 \end{bmatrix} \cdot A$$

are lin. indep. | $u_i \in S$ uniformly randomly)

$$\geq 1 - \frac{r}{|S|}$$

Note: Toeplitz times vector costs $O(n \log n)$ ops. (w. roots of unity)

Application: solving an overdetermined system $Ax = b$

$$\begin{array}{c} n \\ \boxed{} \\ n \end{array} \cdot \begin{array}{c} \boxed{A} \\ \\ \mathbf{0} \end{array} = \begin{array}{c} \boxed{B} \\ \\ \mathbf{0} \end{array}$$

n p

With high probability, B has rank of A , so solve $Bx = c = \begin{bmatrix} (Tb)_1 \\ \vdots \\ (Tb)_p \end{bmatrix}$

Arithmetic cost: $TA, Tb: O(pn \log(n))$
 solve $Bx = c: O(p^3)$ or $O(p^{2.38})$
 check $Ax = b: O(pn)$

For $p = O(\sqrt{n \log(n)})$ or $O((n \log(n))^{0.72})$ the cost is $O(pn \log(n))$

Application: solving an overdetermined system $Ax = b$

$$\begin{array}{c} n \\ \square \\ T \\ \square \\ n \end{array} \cdot \begin{array}{c} \square \\ A \\ \square \\ \mathbf{0} \\ \square \\ p \end{array} = \begin{array}{c} \square \\ B \\ \square \\ \mathbf{0} \\ \square \end{array}$$

With high probability, B has rank of A , so solve $Bx = c = \begin{bmatrix} (Tb)_1 \\ \vdots \\ (Tb)_p \end{bmatrix}$

Arithmetic cost: $TA, Tb: O(pn \log(n))$
 solve $Bx = c: O(p^3)$ or $O(p^{2.38})$
 check $Ax = b: O(pn)$

For $p = O(\sqrt{n \log(n)})$ or $O((n \log(n))^{0.72})$ the cost is $O(pn \log(n))$
 (Previously: $O(p^2n)$ or $O(p^{1.38}n)$)

Application: solving an overdetermined system $Ax = b$

$$\begin{array}{c} n \\ \boxed{} \\ n \end{array} \cdot \begin{array}{c} \boxed{A} \\ \phantom{\boxed{A}} \\ \mathbf{0} \end{array} = \begin{array}{c} \boxed{B} \\ \phantom{\boxed{B}} \\ \mathbf{0} \end{array}$$

n p

With high probability, B has rank of A , so solve $Bx = c = \begin{bmatrix} (Tb)_1 \\ \vdots \\ (Tb)_p \end{bmatrix}$

Arithmetic cost: $TA, Tb: O(pn \log(n))$
 solve $Bx = c: O(p^3)$ or $O(p^{2.38})$
 check $Ax = b: O(pn)$

For $p = O(\sqrt{n \log(n)})$ or $O((n \log(n))^{0.72})$ the cost is $O(pn \log(n))$
 Highly over/underdet. systems can be solved **essentially optimally**.

Cond. Number Distr. of Precond.

Bryc, Włodzimierz, Dembo, Amir, and Jiang, Tiefeng. Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Ann. Probab.*, 34(1):1–38, 2006.

Kaltofen, Zhi \rightarrow Bryc \rightarrow Zhidong Bai \rightarrow Bingyu Li,
Jack Silverstein

The distributions are tricky (we so far have no separation from zero theorem)!

Danke schön!

Thank you!