Efficient linear algebra algorithms in symbolic computation

Erich Kaltofen 寒爐
North Carolina State University
google->kaltofen
google->han lu

Symbolic computation: my 1987 definition

– Results are symbolic.
  entries can be parametric: multivariate polynomials

– Results are exact.
  number of bits in results depends on dimension,
  rank is well-posed, no ill-conditionedness

Symbolic computation: my 1987 definition

– Results are symbolic.
entries can be parametric: multivariate polynomials

– Results are exact.
number of bits in results depends on dimension,
rank is well-posed, no ill-conditionedness

Arithmetic cost: Count number of scalar operations
faithful complexity for modular arithmetic in $\mathbb{Z}_p$
convert to rational numbers via Chinese remaindering, lifting

Show Maple worksheet

pascalmatrix.pdf

Outline of talk

Exact efficient linear algebra algorithms

Black box linear algebra and the LinBox library

Hybrid symbolic-numeric applications
[joint with Zengfeng Yang and Lihong Zhi]

Fast matrix multiplication

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

$$
\begin{array}{l|l}
m_1 \leftarrow (a_{1,2}-a_{2,2})(b_{2,1}-b_{2,2}) & \\
m_2 \leftarrow (a_{1,1}+a_{2,2})(b_{1,1}+b_{2,2}) & \\
m_3 \leftarrow (a_{1,1}-a_{2,1})(b_{1,1}+b_{1,2}) & \\
m_4 \leftarrow (a_{1,1}+a_{1,2})b_{2,2}) & a_{1,1}b_{1,1}+a_{1,2}b_{2,1} = m_1+m_2-m_4+m_6 \\
m_5 \leftarrow a_{1,1}(b_{1,2}-b_{2,2}) & a_{1,1}b_{1,2}+a_{1,2}b_{2,2} = m_4+m_5 \\
m_6 \leftarrow a_{2,2}(b_{2,1}-b_{1,1}) & a_{2,1}b_{1,1}+a_{2,2}b_{2,1} = m_6+m_7 \\
m_7 \leftarrow (a_{2,1}+a_{2,2})b_{1,1}) & a_{2,1}b_{1,2}+a_{2,2}b_{2,2} = m_2-m_3+m_5-m_7
\end{array}
$$

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

Problems reducible to matrix multiplication:

- linear system solving, determinant [Bunch and Hopcroft 1974],
- characteristic polynomial [Keller-Gehrig'85, Pernet and 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

$$
\begin{array}{l|l}
m_1 \leftarrow (a_{1,2}-a_{2,2})(b_{2,1}-b_{2,2}) & \\
m_2 \leftarrow (a_{1,1}+a_{2,2})(b_{1,1}+b_{2,2}) & \\
m_3 \leftarrow (a_{1,1}-a_{2,1})(b_{1,1}+b_{1,2}) & \\
m_4 \leftarrow (a_{1,1}+a_{1,2})b_{2,2}) & a_{1,1}b_{1,1}+a_{1,2}b_{2,1} = m_1+m_2-m_4+m_6 \\
m_5 \leftarrow a_{1,1}(b_{1,2}-b_{2,2}) & a_{1,1}b_{1,2}+a_{1,2}b_{2,2} = m_4+m_5 \\
m_6 \leftarrow a_{2,2}(b_{2,1}-b_{1,1}) & a_{2,1}b_{1,1}+a_{2,2}b_{2,1} = m_6+m_7 \\
m_7 \leftarrow (a_{2,1}+a_{2,2})b_{1,1}) & a_{2,1}b_{1,2}+a_{2,2}b_{2,2} = m_2-m_3+m_5-m_7
\end{array}
$$

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

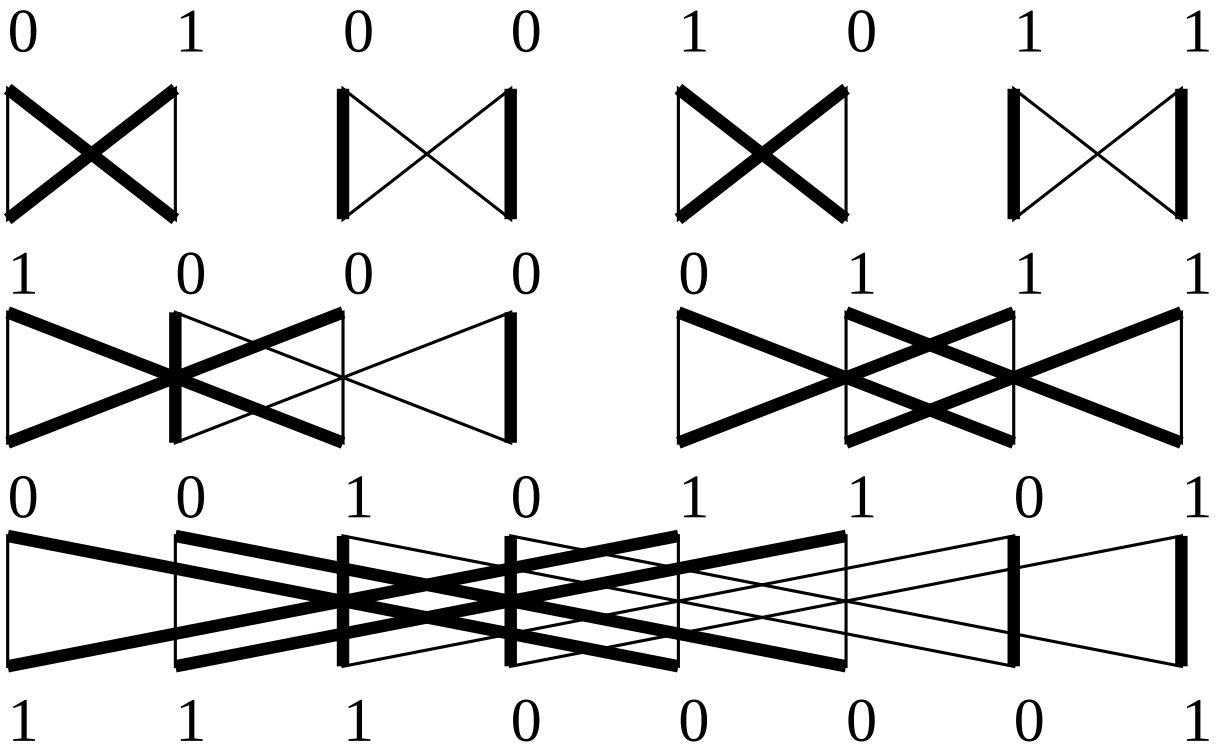
Problems reducible to matrix multiplication:

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

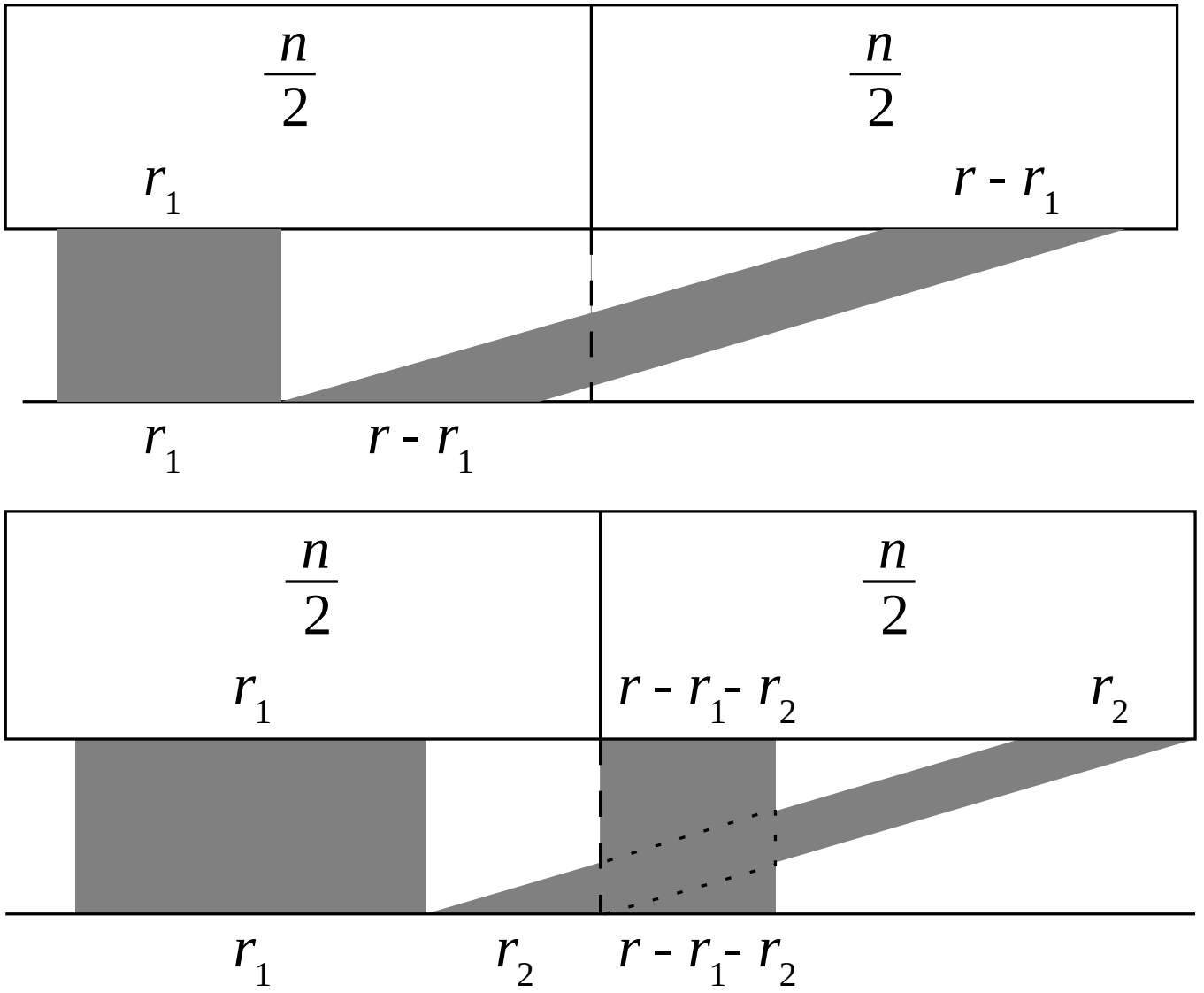
Butterfly preconditioners [Wiedemann'86, Parker'95, Chen et al.'02]

Lemma: *By row/column exchanges in a butterfly network manner, linearly independent rows/columns can be moved to any contiguous block of rows/columns including wrap-around.*

Case $n = 2^k$. Example:



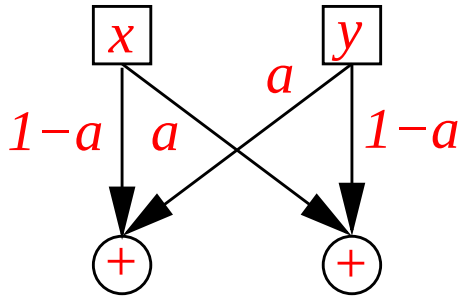
Induction argument: any contiguous position



Wrap-around: complementary rows/columns

General n: depth $= \lceil \log_2(n) \rceil$, number of switches $s \leq n \lceil \log_2(n) \rceil / 2$.

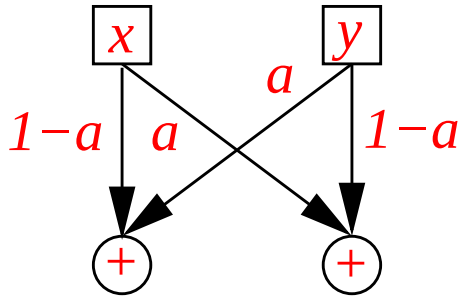
Symbolic exchange matrices + Zippel/Schwartz lemma



$$\underbrace{\begin{bmatrix} 1-a & a \\ a & 1-a \end{bmatrix}}_{\mathscr{E}(a)} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} (1-a)x + ay \\ ax + (1-a)y \end{bmatrix}$$

Unimodular alternate: $\begin{bmatrix} 1 & a \\ 1 & 1+a \end{bmatrix}$

Symbolic exchange matrices + Zippel/Schwartz lemma



$$\underbrace{\begin{bmatrix} 1-a & a \\ a & 1-a \end{bmatrix}}_{\mathscr{E}(a)} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} (1-a)x + ay \\ ax + (1-a)y \end{bmatrix}$$

Let $\mathscr{L}(\alpha_1, \ldots, \alpha_s) = \prod_{k=1}^{s} \mathscr{E}_k(\alpha_k)$, $\quad \mathscr{E}_k$: k-th butterfly switch

Theorem: Let $A \in \mathsf{K}^{n \times n}$ with $r = \operatorname{rank}(A)$.
The first r rows of $\mathscr{L}(\alpha_1, \ldots, \alpha_s) \cdot A$ are linearly independent over $\mathsf{K}(\alpha_1, \ldots, \alpha_s)$.

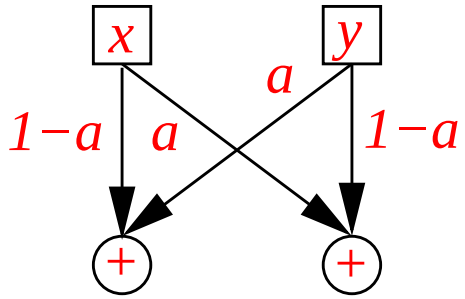
Proof. There exists a setting $\alpha_k \leftarrow \pm 1$ such that linearly independence, hence non-vanishing of a certain determinant, is true.

Symbolic exchange matrices + Zippel/Schwartz lemma

Lemma: [DeMillo-Lipton'78, Schwartz'79, Zippel'79]
Let $0 \neq \Delta \in \mathsf{K}[\alpha_1, \ldots, \alpha_s]$, $S \subseteq \mathsf{K}$ with $|S| < \infty$ elements. Then

$$\text{Probab.}(\Delta(a_1, \ldots, a_s) \neq 0 \mid a_i \in S \text{ uniformly randomly}) \geq 1 - \frac{\text{total deg}(\Delta)}{|S|}$$

Symbolic exchange matrices + Zippel/Schwartz lemma



$$\underbrace{\begin{bmatrix} 1-a & a \\ a & 1-a \end{bmatrix}}_{\mathscr{E}(a)} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} (1-a)x + ay \\ ax + (1-a)y \end{bmatrix}$$

Let $\mathscr{L}(\alpha_1, \ldots, \alpha_s) = \prod_{k=1}^{s} \mathscr{E}_k(\alpha_k)$, $\quad \mathscr{E}_k$: k-th butterfly switch

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

$$\text{Probab.}(\text{the first } r \text{ rows of } \mathscr{L}(a_1, \ldots, a_s) \cdot A \text{ are lin. indep.} \mid a_i \in S \text{ uniformly randomly}) \geq 1 - \frac{r\lceil \log_2(n) \rceil}{|S|}$$

Other preconditioner matrices

Triangular Toeplitz matrices [Kaltofen and Saunders'91]

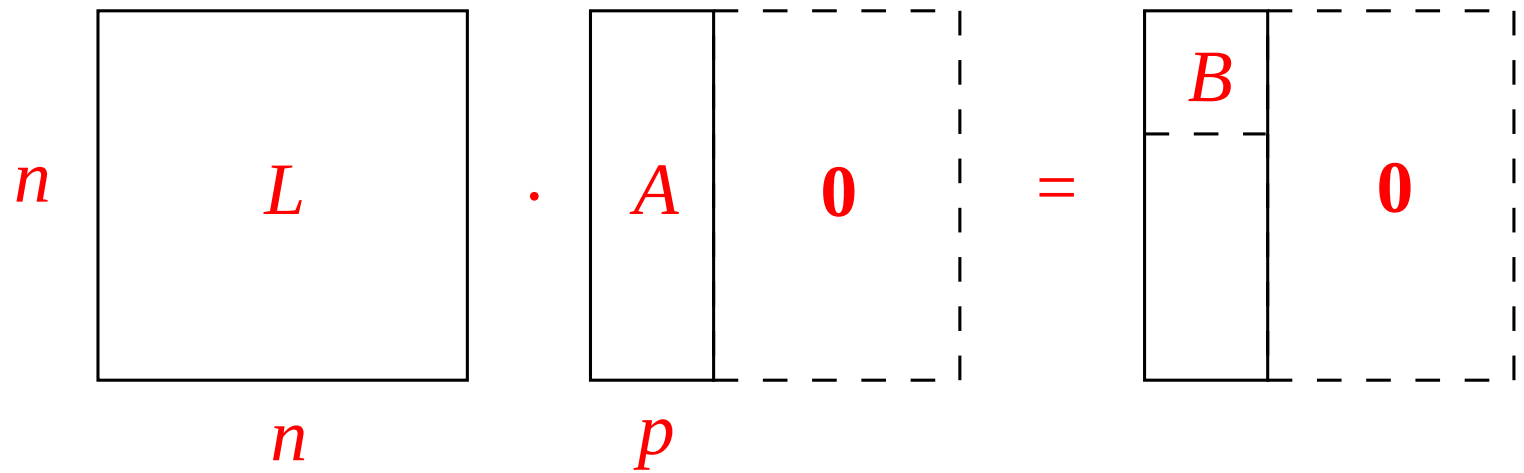
$$\begin{bmatrix} u_1 & u_2 & u_3 & \dots & u_n \\ 0 & u_1 & u_2 & \dots & u_{n-1} \\ 0 & 0 & u_1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & u_2 \\ 0 & \dots & & 0 & u_1 \end{bmatrix}$$

Keep displacement rank low

Sparse matrices with $\{0,1\}$ entries [Wiedemann'86]: work over $\mathbb{Z}_2$.

DAD, where A is symmetric, D random diagonal [Eberly and Kaltofen'97]: most efficient

Application: solving an overdetermined system $Ax = b$



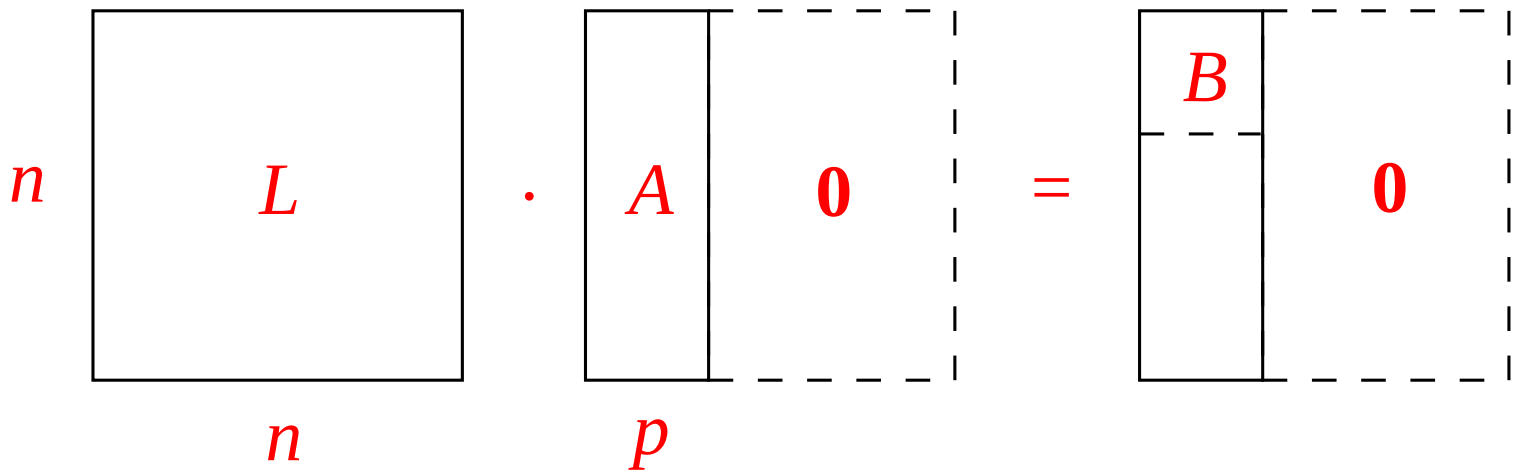
With high probability, B has rank of A, so solve $Bx = c = \begin{bmatrix} (Lb)_1 \\ \vdots \\ (Lb)_p \end{bmatrix}$

Arithmetic cost: LA, Lb: $O(pn\log(n))$
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: $O(p^2 n)$ or $O(p^{1.38} n)$)

Application: solving an overdetermined system $Ax = b$



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

Arithmetic cost: LA, Lb: $O(pn\log(n))$
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.**

Other essentially optimal linear algebra algorithms

– Toeplitz/Hankel systems [Brent et al.'81]

– Toeplitz-like matrices with displacement rank $\alpha = O(1)$
$O(\alpha^2 n^{1+o(1)})$ [Kaltofen'94] $O(\alpha^{1.38} n^{1+o(1)})$ [Bostan et al.'06]

– Multivariable realization (matrix Padé approximation)
via σ-bases [Beckermann and Labahn'94]

– A^{-1} for $A \in \mathsf{K}[x]^{n \times n}$ [Villard and Jeannerod'02]
output has size $\sim n^3 \deg(A)$
algorithm works only for $n = 2^k$ and generic input matrices

Outline of talk

Exact efficient linear algebra algorithms

Black box linear algebra and the LinBox library

Hybrid symbolic-numeric applications

[joint with Zengfeng Yang and Lihong Zhi]

Black box matrix concept

K an arbitrary, e.g., finite field

Perform linear algebra computations, e.g., $A^{-1}b$
within the simultaneous resource bounds:

$O(n)$	black box calls and
$n^{2+o(1)}$	arithmetic operations in K and
$O(n)$	intermediate storage for field elements

Sparse matrices with $O(n)$-time black box: $n^{2+o(1)}$ algorithms

Black box model is useful for dense, structured matrices

$$\begin{bmatrix} 1 & \cdots & & \cdots & \frac{1}{n} \\ & & \vdots & & \\ & & \frac{1}{i+j-1} & & \\ & & \vdots & & \\ \frac{1}{n} & \cdots & & \cdots & \frac{1}{2n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ \\ \vdots \\ \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \\ \vdots \\ \\ b_n \end{bmatrix}$$

(Hilbert matrix)

Savings may be in space, not time: $O(1)$ vs. $O(n^2)$.

Selected black box algorithms

linear system solution, determinant, minimal polynomial	Wiedemann'86
matrix rank	Kaltofen and Saunders'91
block/parallel algorithms	Coppersmith'93,'94 Montgomery'95 Kaltofen'95, Villard'97 Kaltofen and Lobo'99 Eberly and Hovinen'05
diophantine linear system solution	Giesbrecht'97 Mulders and Storjohann'99
integer Smith form	Dumas et al.'01, Giesbrecht'01
rational linear system solution	Eberly et al.'06, '07

The LinBox project www.linalg.org

Objective a generic library for exact linear algebra
("Symbolic MatLab")

New abstraction mechanism black box matrix

Programming languages C++, Maple, GAP, C (Saclib)

Design principle
genericity through template parameter types (matrix entries)
and black box matrix model (sparseness and structuredness)

Participants 22 current researchers and students in USA, Canada and France

Outline of talk

Exact efficient linear algebra algorithms

Black box linear algebra and the LinBox library

Hybrid symbolic-numeric applications

[joint with Zengfeng Yang and Lihong Zhi]

Sparse rational function models

$$f, g \in \mathbb{C}[x_1, \ldots, x_n], \gcd(f, g) = 1$$

By sampling black box, compute sparse representation

$$\frac{\sum_{j=1}^{t_f} a_j x_1^{d_{j,1}} \cdots x_n^{d_{j,n}}}{\sum_{k=1}^{t_g} b_k x_1^{e_{k,1}} \cdots x_n^{e_{k,n}}} = \frac{f}{g}, \quad a_j \neq 0, b_k \neq 0$$

Note: Terms are not known.

Sparse rational function models

$$f, g \in \mathbb{C}[x_1, \ldots, x_n], \gcd(f, g) = 1$$

By sampling black box, compute sparse representation

$$\frac{\sum_{j=1}^{t_f} a_j x_1^{d_{j,1}} \cdots x_n^{d_{j,n}}}{\sum_{k=1}^{t_g} b_k x_1^{e_{k,1}} \cdots x_n^{e_{k,n}}} = \frac{f}{g}, \quad a_j \neq 0, b_k \neq 0$$

Note: Terms are not known.

Deal with: Sparsity by Zippel interpolation + lemma
Numeric noise in values by condition numbers in random structured matrices

Numerical Zippel/Schwartz Lemma: *Let*

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

$\zeta_j = \exp(\frac{2\pi \boldsymbol{i}}{P_j}) \in \mathbb{C}$, $P_j \in \mathbb{Z}_{\geq 3}$ *distinct prime numbers* $\forall 1 \leq j \leq s$
[cf. Giesbrecht, Labahn, Lee 2006]

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

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

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

Numerical Zippel/Schwartz Lemma: *Let*

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

$\zeta_j = \exp(\frac{2\pi \boldsymbol{i}}{P_j}) \in \mathbb{C}$, $P_j \in \mathbb{Z}_{\geq 3}$ *distinct prime numbers* $\forall 1 \leq j \leq s$
[cf. Giesbrecht, Labahn, Lee 2006]

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

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

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

How about matrix condition numbers?

Condition numbers of random $n \times n$ matrices

Entries from standard Gaussian distribution:

$$\text{Expected value}\{\ \kappa_2(G_n)\ \} < 10n$$

[Edelman 1988; Chen and Dongarra 2005]

Condition numbers of random $n \times n$ matrices

Entries from standard Gaussian distribution:

$$\text{Expected value}\{\ \kappa_2(G_n)\ \} < 10n$$

[Edelman 1988; Chen and Dongarra 2005]

Random discrete noise:

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

[Spielman and Teng; Tao and Vu 2007]

Condition numbers of random $n \times n$ matrices

Entries from standard Gaussian distribution:

$$\text{Expected value}\{\ \kappa_2(G_n)\ \} < 10n$$

[Edelman 1988; Chen and Dongarra 2005]

Random discrete noise:

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

[Spielman and Teng; Tao and Vu 2007]

"... one rarely encounters ill-conditioned matrices in practice"

Approximate Polynomial GCD: Sylvester matrices

$$\begin{bmatrix}
a_m & a_{m-1} & \cdots & a_0 & & & \\
 & a_m & \cdots & a_1 & a_0 & \mathbf{0} & \\
 & & \ddots & & \ddots & \ddots & \\
 & \mathbf{0} & & a_m & a_{m-1} & \cdots & a_0 \\
b_n & b_{n-1} & \cdots & b_0 & & & \\
 & b_n & \cdots & b_1 & b_0 & \mathbf{0} & \\
 & & \ddots & & \ddots & \ddots & \\
 & \mathbf{0} & & b_n & \cdots & & b_0
\end{bmatrix}$$

approximate polynomial GCD (common root)
for $a_m x^m + \cdots + a_0$ and $b_n x^n + \cdots + b_0$

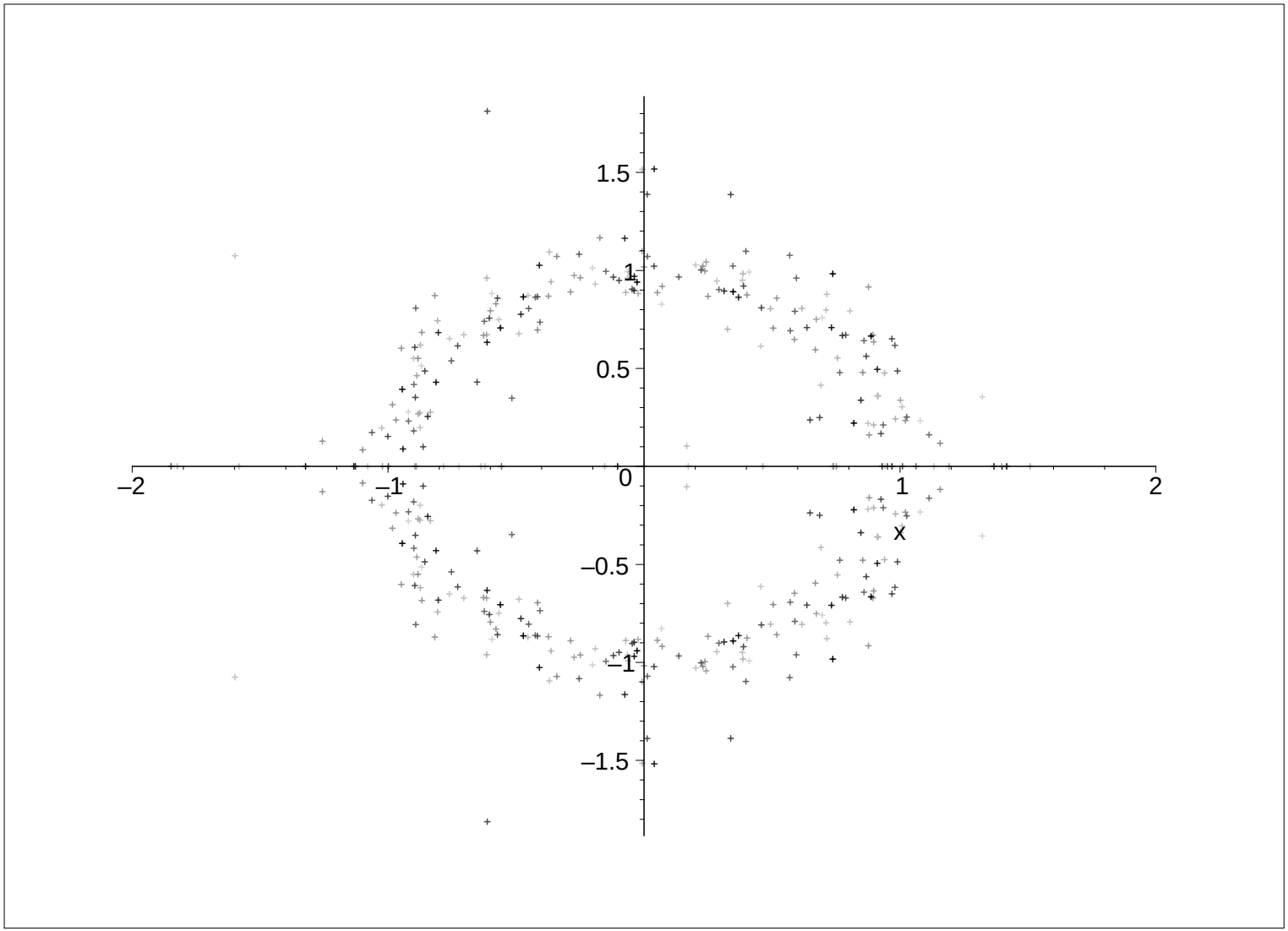
$\Updownarrow$

Sylvester-structure preserving deformation to singularity

Root distribution 20 polynomials of degree 20

Uniformly distributed coefficients

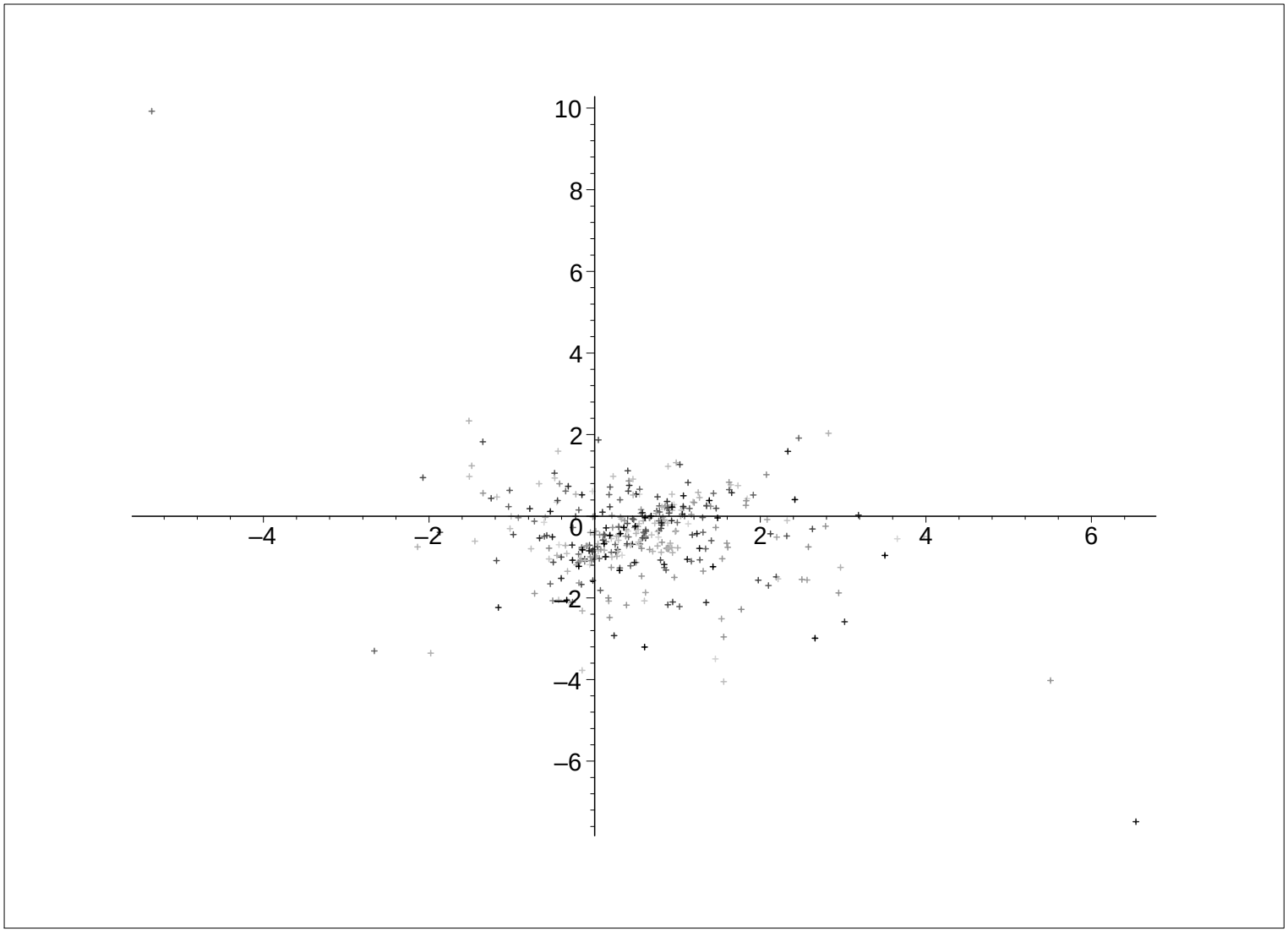
$\Longrightarrow$ small structured condition number



Root distribution 20 polynomials of degree 20

Binomially distributed coefficients

$\Longrightarrow$ 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; Kaltofen, Yang and Zhi 2007

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

small structured condition numbers of Sylvester matrices

Well-conditioning of arising Fourier matrices to be established.

Symbolic linear algebra algorithms = lots of fireworks

谢

谢

THANK YOU!