Asymptotically fast solution of Toeplitz-like singular linear systems

Erich Kaltofen

Rensselaer Polytechnic Institute
Department of Computer Science
Troy, New York, USA

Outline

- **examples of Toeplitz-like matrices**
  - in Coppersmith's block Wiedemann method
  - in Euclidean schemes
- **Toeplitz-like matrices**
  - definition by displacement operators
  - products and inverses
- **divide-and-conquer inversion**
- **randomizations**
  - attaining generic rank profile
  - minimizing the generator length
  - singular systems
- **loose ends**

Wiedmann's coordinate recurrence method (1986)

For $u, v \in \mathbb{K}^N$ and $A \in \mathbb{K}^{N\times N}$ consider the sequence of field elements

$$a_i = u^{\mathrm{tr}} A^i v, \quad i = 0, 1, 2, \ldots$$

Let $f^{(A)}(\lambda) = \sum_{k=0}^{M} f_k^{(A)} \lambda^k \in \mathbb{K}[\lambda]$ with $f^{(A)}(A) = 0$.
Since $u^{\mathrm{tr}} A^j f^{(A)}(A) v = 0$, we have

$$\forall\, j \geq 0\colon \quad \sum_{k=0}^{M} f_k^{(A)} a_{k+j} = 0$$

that is, $\{a_i\}_{i=0,1,\ldots}$ satisfies a **linear recurrence**.

Randomly precondition A and choose **random** u and v; then

$$\mathrm{Det}(\lambda I - A) = \text{minimum recurrence polynomial of } \{a_i\}_{i=0,1,\ldots}.$$

The associated Toeplitz system

Coefficients $f_0^{(A)}, \ldots, f_{M'}^{(A)}$ of a multiple of $f^{(A)}$ can be found by computing a **non-zero** solution to the Toeplitz system

$$\begin{bmatrix} a_N & a_{N-1} & \cdots & a_1 & a_0 \\ a_{N+1} & a_N & \cdots & a_2 & a_1 \\ \vdots & a_{N+1} & \ddots & \vdots & a_2 \\ & \vdots & & & \vdots \\ a_{2N-2} & & & a_{N-1} & \\ a_{2N-1} & a_{2N-2} & \cdots & a_N & a_{N-1} \end{bmatrix} \cdot \begin{bmatrix} f_N \\ f_{N-1} \\ f_{N-2} \\ \vdots \\ \\ f_0 \end{bmatrix} = \mathbf{0}.$$

Achievable in $O(N(\log N)^2 \operatorname{loglog} N)$ arithmetic steps by the Brent-Gustavson-Yun half-GCD Toeplitz solver (1980).

Coppersmith's (1992) parallelization (modified)

Use of the block vectors $\boldsymbol{x} \in \mathbb{K}^{N\times m}$ in place of u

$\boldsymbol{z} \in \mathbb{K}^{N\times n}$ in place of v

$$\boldsymbol{a}_i = \boldsymbol{x}^{\mathrm{tr}} B^{i+1} \boldsymbol{z} \in \mathbb{K}^{m\times n}$$

Find a vector polynomial $c_L\lambda^L + c_{L+1}\lambda^{L+1} + \cdots + c_D\lambda^D \in \mathbb{K}^n[\lambda]$, such that

$$\forall\, j \geq 0\colon\ \sum_{i=L}^{D} \boldsymbol{a}_{j+i} c_i = \sum_{i=L}^{D} \boldsymbol{x}^{\mathrm{tr}} B^{i+j}\, B \boldsymbol{z} c_i = \boldsymbol{0} \in \mathbb{K}^{m\times n}$$

The associated block-Toeplitz system

Let $D = \lceil N/n \rceil$, $S = n(D+1)$, $E = \lceil S/m \rceil$, and let $R = mE$. Compute a non-zero solution to the linear homogeneous $R \times S$ system

$$\left[\begin{array}{c|c|c|c|c} \boldsymbol{a}_D & \ldots & & \boldsymbol{a}_1 & \boldsymbol{a}_0 \\ \hline \boldsymbol{a}_{D+1} & \boldsymbol{a}_D & & \boldsymbol{a}_2 & \boldsymbol{a}_1 \\ \hline \vdots & & \ddots & & \vdots \\ \hline \boldsymbol{a}_{D+E-1} & \ldots & & & \boldsymbol{a}_{E-1} \end{array}\right] \left[\begin{array}{c} c_D \\ \hline c_{D-1} \\ \hline \vdots \\ \hline c_0 \end{array}\right] = \mathbf{0},$$

where $c_i \in \mathbb{K}^n$.

Achievable in $O((m+n)^2 N (\log N)^2 \operatorname{loglog} N)$ arithmetic steps by a **generalization/randomization** of the Bitmead-Anderson/Morf (1980) fast inversion algorithm for Toeplitz-like matrices.

Parallel coarse-grain realization

The ν^{th} processor computes the ν^{th} column of $\boldsymbol{a}_i$, $i \lessapprox \dfrac{N}{m} + \dfrac{N}{n}$

Implementation: sparse random matrices over GF(32 749)

N	Task	Blocking Factor 2	4	8
10,000†	(1) $\langle a^{(i)} \rangle$	$7^h 29'$	$3^h 54'$	$2^h 09'$
	(2) b-massey	$2^h 25'$	$4^h 08'$	$8^h 00'$
	(3) evaluation	$3^h 47'$	$1^h 59'$	$1^h 05'$
	total	$13^h 41'$	$10^h 06'$	$11^h 14'$
20,000‡	(1) $\langle a^{(i)} \rangle$	$57^h 17'$	$28^h 43'$	$15^h 21'$
	(2) b-massey	$9^h 48'$	$16^h 36'$	$33^h 39'$
	(3) evaluation	$29^h 42'$	$14^h 44'$	$7^h 53'$
	total	$96^h 47'$	$60^h 02'$	$56^h 53'$

Distributed on our DSC system
Each processor rated at 28.5 MIPS
† ≈ 350 000 non-zero entries
‡ ≈ 1 300 000 non-zero entries

Example: Euclidean scheme

Given $L \leq \min\{M, N\}$ and

$$f_{-1}(x) = a_M x^M + a_{M-1} x^{M-1} + \cdots + a_0 \in \mathbb{K}[x]$$

and

$$f_0(x) = b_N x^N + b_{N-1} x^{N-1} + \cdots + b_0 \in \mathbb{K}[x]$$

compute the remainder f_i in the Euclidean chain with

$$\deg(f_i) \leq L < \deg(f_{i-1})$$

and the multipliers s_i and t_i with

$$s_i f_{-1} + t_i f_0 = f_i.$$

Solve for the coefficients of $S(x)$, $T(x)$, and $F(x)$:

$$S f_{-1} + T f_0 = F, \quad \begin{cases} \deg(F) \le L, \\ \deg(S) \le N - L - 1, \\ \deg(T) \le M - L - 1. \end{cases}$$

$\iff$ compute right null space of dimension $M + N - (L - \deg f_i)$ of

$$\left[\begin{array}{cccc|ccc|ccc}
a_0 & & & \mathbf{0} & b_0 & & \mathbf{0} & -1 & & \mathbf{0} \\
a_1 & a_0 & & & b_1 & \ddots & & & \ddots & \\
\vdots & a_1 & \ddots & & \vdots & \ddots & b_0 & \mathbf{0} & & -1 \\
a_M & \vdots & & a_0 & & & & & & \\
0 & a_M & & & b_N & & & & & \\
 & 0 & \ddots & \vdots & 0 & \ddots & \vdots & & \mathbf{0} & \\
 & & \ddots & a_M & & \ddots & b_N & & & \\
\mathbf{0} & & & 0 & \mathbf{0} & & 0 & & &
\end{array}\right].$$

$$\underbrace{\qquad\qquad\qquad}_{N-L} \qquad \underbrace{\qquad\qquad}_{M-L} \qquad \underbrace{\qquad\qquad}_{L+1}$$

Toeplitz-like matrices

Kailath et al. 1979 consider the **matrix displacement operators**

$\phi_+(A) = A - \downarrow(\Rsh A) =$

$$\begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,N} \\ a_{2,1} & a_{2,2} & \dots & a_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N,1} & a_{N,2} & \dots & a_{N,N} \end{bmatrix} - \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & a_{1,1} & \dots & a_{1,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{N-1,1} & \dots & a_{N-1,N-1} \end{bmatrix}$$

and $\phi_-(A) = A - \uparrow(\Lsh A)$.

A matrix is **Toeplitz-like** if the matrix ranks $\alpha_+(A) = \mathrm{rank}(\phi_+(A))$ or $\alpha_-(A) = \mathrm{rank}(\phi_-(A))$ are small.

If A is an $m \times n$ block matrix with Toeplitz blocks, then $\alpha_+(A) \leq m + n$.

Displacement rank formulas

$$(1a)\ \phi_+(A) = \sum_{j=1}^{\alpha_+(A)} y_j z_j^{\mathrm{tr}} \iff A = \sum_{j=1}^{\alpha_+(A)} L[\![y_j]\!]\, U[\![z_j^{\mathrm{tr}}]\!] \qquad (\Sigma\text{LU-rep.})$$

$$(1b)\ \phi_-(A) = \sum_{k=1}^{\alpha_-(A)} \bar{y}_k \bar{z}_k^{\mathrm{tr}} \iff A = \sum_{k=1}^{\alpha_-(A)} U[\![(\bar{y}_k^{\mathrm{rev}})^{\mathrm{tr}}]\!] L[\![\bar{z}_k^{\mathrm{rev}}]\!] \quad (\Sigma\text{UL-rep.})$$

$$(2)\quad -2 \leq \alpha_+(A) - \alpha_-(A) \leq 2$$

$$(3)\quad \alpha_+(A) = \alpha_-(A^{-1}) \quad \text{and} \quad \alpha_-(A) = \alpha_+(A^{-1})$$

$$(4)\quad \alpha_+(AB) \leq \alpha_+(A) + \alpha_+(B) + 1$$

y_j, z_j, $\bar{y}_k$, $\bar{z}_k$ are N-dimensional vectors
$\bar{y}_k^{\mathrm{rev}}$, $\bar{z}_k^{\mathrm{rev}}$ are the mirror images of $\bar{y}_k$, $\bar{z}_k$
$L[\![y]\!]$ is a **lower-triangular Toeplitz** matrix whose first column is y
$U[\![z^{\mathrm{tr}}]\!]$ is an **upper triangular Toeplitz** matrix whose first row is z^{tr}

Main algorithmic problems

Given the ΣLU representation for an $N \times N$ non-singular matrix A of displacement rank α, compute the ΣUL representation for A^{-1}.
Note: input and output occupies $O(\alpha N)$ elements.

Given the ΣLU representation for an $N \times N$ singular matrix A, compute rank(A) and a vector w such that $Aw = \mathbf{0}$ and $w \neq \mathbf{0}$.

By use of randomization we can solve both problems in

$$O(\alpha^2 N (\log N)^2 \operatorname{loglog} N)$$

arithmetic operations.

Divide-and-conquer strategy á la Strassen

Suppose all possible leading principal submatrices are non-singular (**“generic rank profile”**): for

$$A = \left[\begin{array}{c|c} A_{1,1} & A_{1,2} \\ \hline A_{2,1} & A_{2,2} \end{array}\right]$$

we have

$$A^{-1} = \left[\begin{array}{c|c} A_{1,1}^{-1} + A_{1,1}^{-1} A_{1,2} \Delta^{-1} A_{2,1} A_{1,1}^{-1} & -A_{1,1}^{-1} A_{1,2} \Delta^{-1} \\ \hline -\Delta^{-1} A_{2,1} A_{1,1}^{-1} & \Delta^{-1} \end{array}\right].$$

where $\Delta = A_{2,2} - A_{2,1} A_{1,1}^{-1} A_{1,2}$ is the **Schur complement**.

Lemma (cf. Bitmead-Anderson/Morf 1980): If $A_{1,1}$ is non-singular and if $A[1,1] \neq 0$ then $\alpha_+(\Delta) \leq \alpha_+(A)$.

Generic rank profile by randomization

Theorem (Kaltofen and Saunders 1991): Let v and w be vectors whose entries are randomly selected from a subset S of the field of entries. Then

$$\widetilde{A} = \underbrace{U[\![v^{\mathrm{tr}}]\!]}_{V} \cdot A \cdot \underbrace{L[\![w]\!]}_{W}$$

has generic rank profile with probability $1 - \frac{N(N+1)}{\mathrm{cardinality}(S)}$.

<u>Note:</u> $\alpha_+(\widetilde{A}) \le \alpha_+(A) + 4$.

Minimal-length generators by randomization

Suppose we are given a **non-minimal** ΣLU representation

$$A = \sum_{k=1}^{\beta} L[\![\hat{y}_k]\!]\, U[\![\hat{z}_k^{\mathrm{tr}}]\!], \quad \beta > \alpha_+(A).$$

Then we may probabilistically find a **minimal** ΣLU representation

$$A = \sum_{j=1}^{\alpha} L[\![y_j]\!]\, U[\![z_j^{\mathrm{tr}}]\!], \quad \alpha = \alpha_+(A),$$

in $O(\alpha\beta N + \beta N \log N \operatorname{loglog} N)$ arithmetic operations.

Uses randomizations for generic rank profile:

$$V \cdot \phi_+(A) \cdot W = \tilde{\boldsymbol{y}} \cdot \tilde{\boldsymbol{z}}^{\mathrm{tr}} \Longrightarrow \phi_+(A) = (V^{-1}\tilde{\boldsymbol{y}}) \cdot (\tilde{\boldsymbol{z}}^{\mathrm{tr}} W^{-1})$$

Picking a random solution of a singular system

Let $\widetilde{A} \in \mathbb{K}^{n\times n}$ be of rank r and generic rank profile. Then for

$$\widetilde{A} \cdot \underbrace{\begin{bmatrix} y' \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{y} \left.\vphantom{\begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}}\right\} n-r = \widetilde{A}v, \quad v \text{ random},$$

$y - v$ uniformly samples the right null space of $\widetilde{A}$.

Loose ends

- avoid randomization

- can complexity be reduced to $\alpha^\eta N(\log N)^{O(1)}$ with $\eta < 2$ by fast matrix multiplication?

- give efficient parallel algorithm; that is, algorithm with $(\log N)^{O(1)}$ parallel time and $\alpha^2 N$ processors
  Best-known solution takes αN^2 processors

- generalize shift operators to Macaulay matrices

- prove fast method practical in comparision to the $O(\alpha N^2)$ Levinson/Durbin method