Somewhat Cheaper Parallel Linear Algebra

Wayne Eberly

Including work by Xuming Chen

Department of Computer Science

University of Calgary

Presented at ACA 2002

Slides at

http://www.cpsc.ucalgary.ca/~eberly/

Fundamental Problems

- Solving a Nonsingular System
- Computing Matrix Rank, Nullspace
- Finding One, or All, Solution(s)
  for a Linear System

Related Problems

- Matrix Inverse
- Structured Matrix Computations
- Linearly Independent Subsets of Vectors
- Matrix Normal Forms

A General Goal

(Randomized) Parallel Algorithms

- using polylogarithmic time
- with a reasonable number of processors
- working over all fields
- and having a small error probability

Goals for Much of This Work

Algorithms for computations over small fields,

- with error bounded by a constant, and
- with a smaller time-processor product than previously known algorithms

Parallel time for computations over arbitrary fields will also be considered.

Past Work

Kaltofen & Pan (1991, 1992) presented algorithms

- to solve an $n \times n$ system of linear equations
  - over a field of characteristic zero, using
    - time $O((\log n)^2)$ and
    - work $O(\mathcal{MM}(n) \log n)$
  - with slightly higher time over large fields of positive characteristic,
  - with slightly higher time and work over small finite fields; and
- for matrix rank and nullspace, and inverse, with similar costs

Other authors have subsequently applied these results to solve related problems.

An Extension and Two Applications

Eberly

- reduced the **work** for nonsingular systems over small fields (1997),
  - to match the bound given by Kaltofen and Pan for the large field case,
  - but **increasing the error probability** to an arbitrary positive constant,

  and

- gave processor-efficient parallel algorithms
  - for independent subsets of vectors, and
  - for an L-U-P decomposition of a non-singular matrix,

  using Kaltofen and Pan's algorithm as a subroutine (1991)

This Talk

- A similar improvement is described for matrix rank and nullspace (Chen, 2002).

- Algorithm for independent subsets is improved to replace a log-factor in **time** with a log-log factor, in the large field case.

- The improvement is modified, to reduce the time and work required in the small field case for this problems, too.

- Additional problems are briefly discussed.

I. Rank and Nullspace

Definitions: Minimal Polynomials

1. A sequence $a_0, a_1, a_2, \ldots$ of elements of a field F is *linearly recurrent* if there exist values $c_0, c_1, \ldots, c_n \in \mathsf{F}$, not all zero, such that

$$\sum_{j=0}^{n} c_j a_{h+j} = 0 \qquad \text{for all } j \geq 0. \qquad (1)$$

2. The *minimal polynomial* of such a sequence is the monic polynomial

$$f = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0 \in \mathsf{F}[x]$$

with least degree, such that equation (1) is satisfied (if we set $c_n = 1$).

3. If $u \in \mathsf{F}^{1\times n}$, $A \in \mathsf{F}^{n\times n}$, and $v \in \mathsf{F}^{n\times 1}$, then minpol$(u, A, v)$ is the minimal polynomial of the linearly recurrent sequence $a_0, a_1, a_2, \ldots$, where

$$a_i = uA^iv \qquad \text{for } i \geq 0.$$

4. If $A \in \mathsf{F}^{n\times n}$ and $v \in \mathsf{F}^{n\times 1}$ then minpol(A, v) is the monic polynomial $f \in \mathsf{F}[x]$ with least degree such that

$$f(A)v = 0.$$

5. If $A \in \mathsf{F}^{n\times n}$ then minpol(A) is the monic polynomial $f \in \mathsf{F}[x]$ with least degree such that

$$f(A) = 0.$$

Each polynomial divides the next, and minpol(A) divides the characteristic polynomial of A.

Useful Properties

Suppose $A \in \mathsf{F}^{n \times n}$, $u \in \mathsf{F}^{1 \times n}$, and $v \in \mathsf{F}^{n \times 1}$.

1. If A is nonsingular, and minpol(u, A, v) has degree n, then this is the characteristic polynomial of A and it can be used to solve a linear system with coefficient matrix A.

2. If A has rank r and minpol(u, A, v) has rank $r+1$, then this polynomial is minpol(A). It can be used to solve linear systems with coefficient matrix A.

3. minpol(u, A, v) can be computed efficiently in parallel from u, A, and v (see Pan, 1996, for an extensive discussion)

Useful Technique: Conditioning

A system $Ax = b$ can be solved using the solution for a "conditioned" system

$$XAYz = Xb.$$

The original system has solution $x = Yz$.

Kaltofen & Pan apply structured "conditioners" X and Y:

For random X, Y, and vectors u and v,

minpol(u, XAY, v) has maximal degree

with high probability, if the ground field F is sufficiently large.

One works over a field extension if F is small.

If F is finite and small ...

Eberly (1997): Select X uniformly from $\mathsf{F}^{n\times n}$ and let Y be the identity matrix.

This eliminates the need for field extensions and reduces the work required.

Chen (2002) has proved Eberly's analysis:

Suppose $A \in \mathbb{F}_q^{n\times n}$ is nonsingular, and that $X \in \mathbb{F}_q^{n\times n}$, $u \in \mathbb{F}_q^{1\times n}$, and $v \in \mathbb{F}_q^{n\times 1}$ are uniformly and independently chosen.

Theorem (Chen, 2002): X is nonsingular and minpol(u, AX, v) has degree n with probability

$$\frac{(q-1)}{(q+1)}\left(1-\frac{1}{q^{2n}}\right)\tau_q(n)$$

where $\tau_q(n)$ is the probability that a randomly chosen matrix from $\mathbb{F}_q^{n\times n}$ is nonsingular.

Extension: Computing Matrix Rank and Nullspace

Suppose $A \in \mathbb{F}_q^{n \times n}$ has rank r, where $0 < r < n$.

minpol(u, AX, v) has degree at most $r + 1$, for all $u \in \mathbb{F}_q^{1 \times n}$, $X \in \mathbb{F}_q^{n \times n}$, and $v \in \mathsf{F}^{n \times 1}$.

Suppose $u \in \mathbb{F}_q^{1 \times n}$, $X \in \mathbb{F}_q^{n \times n}$, and $v \in \mathsf{F}^{n \times 1}$ are uniformly and independently selected.

Theorem (Chen, 2002): X is nonsingular and minpol(u, AX, v) has degree $r + 1$ with probability

$$\frac{(q-1)}{(q+1)}\left(1 - \frac{1}{q}\right)\left(1 - \frac{1}{q^{2r}}\right) \cdot \left(1 - \frac{1}{q^{n-r}}\right)\tau_q(r)\tau_q(n-r).$$

The rank is easily certified, and a basis for nullspace of A easily obtained, once r has been guessed correctly.

Consequence

Work for small field computations can be reduced to match bounds for the large field case, with error probability bounded by any positive constant, for each of the following problems.

- Matrix Rank
- Computation of the Basis for the Nullspace
- Solving a System of Linear Equations

Time is dominated by that needed to compute the minimal polynomial of a linear recurrence (when the degree is bounded, but not known in advance)

II. Independent Subsets of Vectors

Suppose $v_1, v_2, \ldots, v_s, v_{s+1}, v_{s+t} \in \mathsf{F}^{n\times 1}$, and $N \in \mathsf{F}^{n\times n}$ is a matrix of **maximal possible** rank such that

$$Nv_i = 0 \quad \text{for } 1 \leq i \leq s,$$

and suppose $v_{h_1}, v_{h_2}, \ldots, v_{h_k}$ is a maximal linearly independent subset of $v_1, v_2, \ldots, v_s$.

Then $Nv_{j_1}, Nv_{j_2}, \ldots, Nv_{j_\ell}$ is a maximal linearly independent subset of $Nv_{s+1}, Nv_{s+2}, \ldots, Nv_{s+t}$ if and only if

$$v_{h_1}, v_{h_2}, \ldots, v_{h_k}, v_{j_1}, v_{j_2}, \ldots, v_{j_\ell}$$

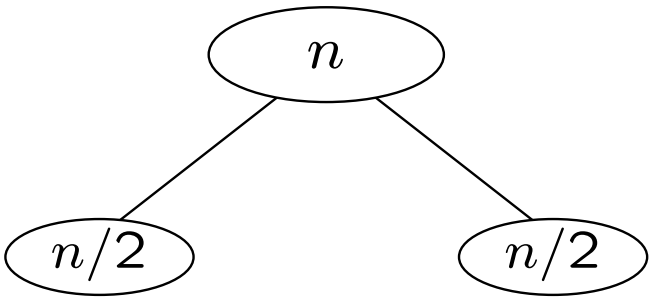
is a maximal linearly independent subset of

$$v_1, v_2, \ldots, v_s, v_{s+1}, \ldots, v_{s+t}.$$

Eberly (1991) used this to obtain a processor-efficient parallel algorithm for linearly independent subsets.

Cheriyan and Reif (1993) subsequently reduced the time required by a logarithmic factor.

Result:



Time:

$$T_{IS}(n) \leq T_{IS}(n/2) + T_{\mathit{Nullspace}}(n)$$

Work:

$$W_{IS}(n) \leq 2W_{IS}(n/2) + W_{\mathit{Nullspace}}(n).$$

Acceleration of the Process

Suppose we split input into k subsets, each of approximately the same size.

S_i: A matrix whose columns are formed from the i^{th} of these subsets, for $1 \leq i \leq k$

For $1 \leq i \leq k-1$, let N_i be a matrix of **maximal rank** such that

$$N_i S_j = 0 \qquad \text{for } 1 \leq j \leq i-1.$$

We may now consider instances consisting of the columns of each of

$$S_1, \quad N_1 S_2, \quad N_2 S_3, \quad \ldots, \quad N_{k-1} S_k.$$

in parallel.

Cost of Resulting Algorithm:

Time:

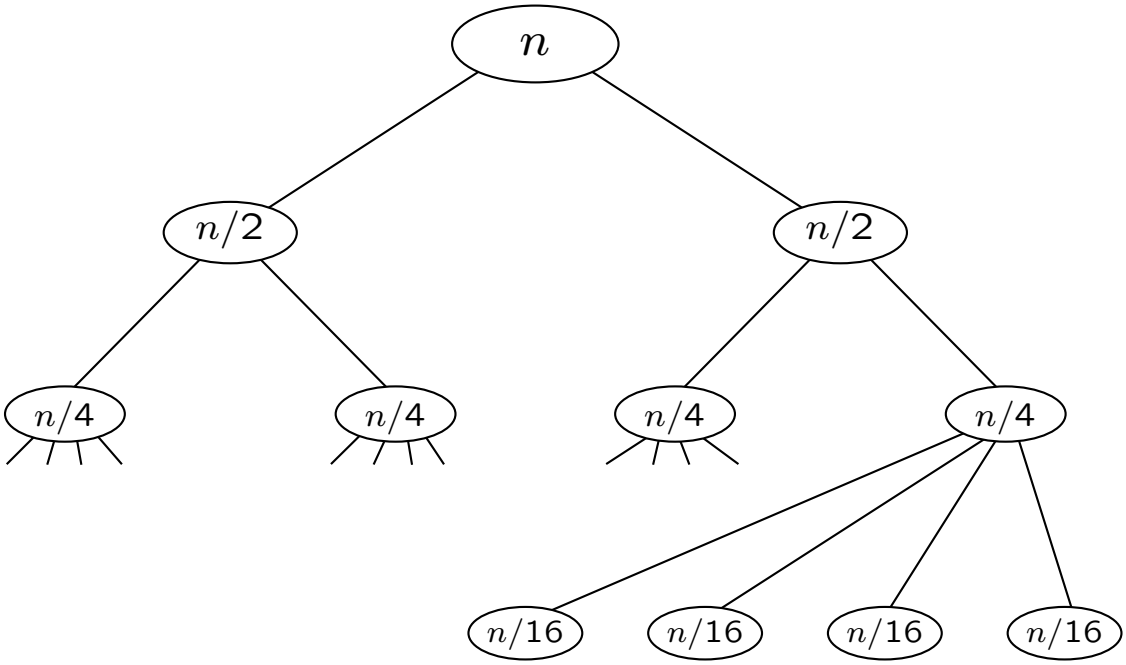
$$T_{IS}(n) \leq T_{IS}(n/k) + T_{\mathit{Nullspace}}(n)$$

Work:

$$W_{IS}(n) \leq kW_{IS}(n/k) + (k-1)W_{\mathit{Nullspace}}(n).$$

Next Trick: Allow k to vary.

Squaring k at each level ...



Work to go from $2^{2^{h-1}}$ subproblems to 2^{2^h}:

$$2^{2^{h-1}} \cdot 2^{2^{h-1}} \cdot W_{\mathit{Nullspace}}\left(n/\left(2^{2^{h-1}}\right)\right)$$

Since $W_{\mathit{Nullspace}}(\ell) \in \Omega(\ell^2)$, this is in

$$O(W_{\mathit{Nullspace}}(n))$$

Assuming $W_{Nullspace}(n) \in \Omega\left(n^{2+\varepsilon}\right)$ for some **positive constant** ε,

$$T_{IS}(n) \in O(T_{Nullspace}(n) \log \log n),$$

and

$$W_{IS}(n) \in O(W_{Nullspace}(n))$$

(if $W_{Nullspace}(n) \in \Omega(\mathcal{MM}(n))$ as well).

However, we do not know that

$$W_{Nullspace}(n) \in \Omega\left(n^{2+\varepsilon}\right) \ldots$$

Slowing Things Down, a Bit

Suppose we increase the total number of instances from k to

- $2k$ (as in the original algorithm), if $k \neq 4^\ell$ for some positive integer ℓ,
- $k^{3/2} = 8^\ell$, if $k = 4^\ell$ for a positive integer ℓ.

We increase from 4^ℓ to 8^ℓ at (at least) every second stage.

In the "rapid acceleration" case, the total work for this level is at most

$$8^\ell \; W_{\mathit{Nullspace}}\left(n/\left(4^\ell\right)\right) \leq 2^{-\ell} W_{\mathit{Nullspace}}(n),$$

... so the total work is in $O(W_{\mathit{Nullspace}}(n))$.

If F is sufficiently large ...

The total number of instances to be considered is bounded by a polynomial function of n.

The Schwartz-Zippel lemma can be used to design and analyze a large-field algorithm whose failure probability is inverse-polynomial in n.

III. Rapid Acceleration in the Small Field Case

Slowing things down, again ...

Increase k to

- $2k$, if $k \neq 8^\ell$ for a positive integer ℓ,
- $16^\ell = k^{4/3}$, if $k = 8^\ell$ for a positive integer ℓ.

There is a "rapid acceleration" phase at (at least) every third level.

To Reduce Error: Use $2^\ell = k^{1/3}$ independent trials of a Las Vegas algorithm to solve each instance of the *Nullspace* problem at this level.

Examination of recurrences confirms that

- The asymptotic time for the computation has not changed.

- The work for each level decreases geometrically, with total linear in $W_{Nullspace}(n)$.

- The probability of failure is now dropping geometrically at each level, too.

  The total probability of failure can therefore be bounded by any positive constant that we care to use.

IV. Additional Problems

Back to Matrix Rank and Nullspace

Suppose we begin an attempt to compute the rank of $A \in \mathsf{F}^{n \times n}$ as follows.

1. Condition, so that the principal $n/2 \times n/2$ submatrix is nonsingular if the rank is at least $n/2$.

2. Check whether the submatrix is invertible, and then examine either this submatrix or its Schur complement.

This "Divide and Conquer" algorithm can be accelerated, in the manner just described for Independent Subset.

Time: $O(T_{Det}(n) \log \log n)$

Work: $O(W_{Det}(n)) \subseteq O(\mathcal{MM}(n) \log n)$

Additional Problems

L-U-P factorization of a nonsingular matrix can also improved in this way.

Question: Is this true for other problems as well?

Plausible Targets:

Time:

- $O\left((\log n)^2(\log\log n)^{O(1)}\right)$
  if $p = \text{char F} = 0$ or $> n$

- $O\left((\log n)^2 \log_p n(\log\log n)^{O(1)}\right)$
  if $0 < p < n$

Work: Matching current bounds for the large field case, even for small fields if constant failure probability is allowed.