Processor-Efficient Parallel Matrix Inversion over Abstract Fields: Two Extensions

W. Eberly
Department of Computer Science
University of Calgary
Calgary, Alberta, Canada, T2N 1N4
eberly@cpsc.ucalgary.ca
http://www.cpsc.ucalgary.ca/~eberly *

Abstract

Kaltofen and Pan's processor-efficient parallel algorithm for the solution of a general $n \times n$ system of linear equations over an abstract field is extended in two ways. First, it is shown that dense, unstructured systems of linear equations over small fields can be solved using the time established by Kaltofen and Pan for this case, and with the time-processor product established by Kaltofen and Pan for computations over *large* fields — reducing the work required for the small field case by slightly more than a logarithmic factor. Second, a processor-efficient parallel algorithm is given for computation of the inverse of a matrix over an abstract field. This algorithm has essentially the same complexity as Kaltofen and Pan's algorithm for this problem, but it does not rely on any program transformation of the type given by Baur and Strassen, and used by Kaltofen and Pan to obtain an algorithm for matrix inversion. Thus, this is the first "explicitly given" processor-efficient parallel algorithm for matrix inversion over an abstract field.

1 Introduction

In the early 1990's, Kaltofen and Pan presented the first known processor-efficient parallel algorithm for the solution of systems of linear equations over abstract fields ([6], [7]). This paper includes two extensions of this work. We assume the same model of computation as Kaltofen and Pan (as described in [6]).

Henceforth we will consider the "work" used by a parallel algorithm to be its time-processor product, and we will define $\mathcal{M}(n)$ to be a function of n such that it is possible to compute the product of two $n \times n$ matrices with entries in a field F using time $O(\log n)$ and work $O(\mathcal{M}(n))$. Clearly, one can choose $\mathcal{M}(n) \in O(n^3)$; we will assume henceforth that $\mathcal{M}(n) \geq n^2 \log^3 n$, as well.

Kaltofen and Pan show (in [6]) that one can solve a nonsingular $n \times n$ system reliably in time $O(\log^2 n)$ using work $O(\mathcal{M}(n) \log n)$, provided that the size of F is greater than cn^2 for some constant c greater than three, and provided that the characteristic of F is greater than n. In subsequent work ([7], [8]) Kaltofen and Pan have eliminated these conditions on F, at the cost of increasing the time and work bounds for this computation by polylogarithmic factors.

In this paper, it is shown that a first stage of Kaltofen and Pan's algorithm can be modified slightly, so that it can be performed reliably over small fields within the time and work bounds that have been established for Kaltofen and Pan's algorithm in the large field case. Later stages of Kaltofen and Pan's algorithm require greater time, but the same or less work, in the small field (and small characteristic) case. Thus, this change does not improve the time required to solve nonsingular linear systems in the small field case. However, it does imply that the work bound is improved, so that nonsingular systems over arbitrary fields can be solved using polylogarithmic time and work $O(\mathcal{M}(n) \log n)$.

Kaltofen and Pan also consider several related problems, including the problem of computing the inverse A^{-1} of a given nonsingular matrix $A \in \mathsf{F}^{n \times n}$. They establish (in [6]) that this problem can be solved with at most the same asymptotic cost as is needed to compute the determinant — and, hence, within the bounds on time and work that they have established for solving linear systems. Kaltofen and Pan obtain this reduction by modifying a construction of Baur and Strassen [1], which transforms a straightline program computing a function f into one which also computes the partial derivatives of f; the modified construction of Kaltofen and Pan preserves circuit depth as well as size. A more "explicit" algorithm — one which does not rely on Baur and Strassen's construction or Kaltofen and Pan's extension — might arguably be preferable, since it could be easier to implement, and it might more easily prove the existence of a "uniform" family of circuits for matrix inversion of small depth and size.

An explicit processor-efficient parallel algorithm for matrix inversion over abstract fields, with essentially the same cost as Kaltofen and Pan's algorithm for solving linear systems, is given in this paper. Several problems, including the *LU* factorization of matrices ([9]), computation of a maximal nonsingular minor and *PLU* factorization ([3]), and (Las Vegas) computation of the Frobenius normal form, mini-

*Research was supported in part by Natural Sciences and Engineering Research Council of Canada research grant OGP0155376, and was conducted in the Department of Computer Science, Rensselaer Polytechnic Institute, Troy, New York, USA.

mum polynomial, and characteristic polynomial of a matrix ([4]), have previously been reduced to matrix inversion. The new algorithm can therefore be used as a subroutine to provide somewhat more explicit processor-efficient parallel algorithms, than have previously been available, for all these problems as well.

A more detailed description of Kaltofen and Pan's method will be given in Section 2 of this paper. The improvement of Kaltofen and Pan's algorithm for the small field case appears in Section 3, while the new algorithm for matrix inversion is given in Section 4. These last two sections can be read independently (but both depend on Section 2).

2 Kaltofen and Pan's Method

Kaltofen and Pan's algorithm can be regarded as a two part process. These parts are discussed in detail in the next two subsections.

2.1 Reduction to Solving a Toeplitz Linear System

Recall that a matrix $T \in \mathsf{F}^{n\times n}$ is a *Toeplitz matrix* if

$$T = \begin{bmatrix} a_{n-1} & a_n & \cdots & a_{2n-3} & a_{2n-2} \\ a_{n-2} & a_{n-1} & \cdots & a_{2n-4} & a_{2n-3} \\ \vdots & \vdots & \ddots & \vdots & \\ a_1 & a_2 & \cdots & a_{n-1} & a_n \\ a_0 & a_1 & \cdots & a_{n-2} & a_{n-1} \end{bmatrix} \qquad (2.1)$$

for elements $a_0, a_1, \ldots, a_{2n-2} \in \mathsf{F}$, so that each band parallel to the diagonal includes copies of the same element of F. The first part of Kaltofen and Pan's algorithm is a (Las Vegas) reduction from the problem of solving an arbitrary nonsingular system of linear equations to that of solving a nonsingular Toeplitz linear system.

Suppose, now, that $A \in \mathsf{F}^{n\times n}$, $b \in \mathsf{F}^{n\times 1}$, and that we wish to find the vector $x \in \mathsf{F}^{n\times 1}$ such that $Ax = b$. In order to apply Kaltofen and Pan's method, it is necessary to find "conditioners" $X, Y \in \mathsf{F}^{n\times n}$ such that the matrix $\tilde{A} = XAY$ is nonsingular and has a minimum polynomial $f^{\tilde{A}} \in \mathsf{F}[x]$ with degree n. If $\tilde{b} = Xb \in \mathsf{F}^{n\times 1}$ and $y \in \mathsf{F}^{n\times 1}$ is a solution for the system $\tilde{A}y = \tilde{b}$, then $x = Yy$ is a solution for the original system $Ax = b$, so that X and Y serve to reduce the original problem to one such that the coefficient matrix has a minimum polynomial with full degree.

Kaltofen and Pan present two randomized constructions of appropriate conditioners X and Y over sufficiently large ground fields. In particular, they show in [6] that one can set X to be the identity matrix, and $Y = HD$, a product of a Hankel matrix H and a diagonal matrix D. The results of [7] imply that one can also set X to be an upper triangular Toeplitz matrix and Y to be a lower triangular Toeplitz matrix, both with ones on the diagonal. If the entries of the above structured matrices (H and D in the first construction, X and Y in the second) are selected uniformly and independently from a finite subset S of the ground field F, then the probability that either construction fails — that is, $\tilde{A} = XAY$ is singular, or has a minimum polynomial with degree less than n — is in $O(n^2/|S|)$.

Now, recall that if $\alpha_0, \alpha_1, \alpha_2, \ldots$ is an infinite sequence of elements of F, then this sequence is *linearly generated* if there exist elements $c_0, c_1, \ldots, c_m \in \mathsf{F}$, not all zero, such that

$$c_0\alpha_i + c_1\alpha_{i+1} + \cdots + c_m\alpha_{i+m} = 0$$

for all $i \geq 0$. A polynomial $c_m x^m + c_{m-1}x^{m-1} + \cdots + c_1 x + c_0 \in \mathsf{F}[x]$ is a *generating polynomial* for the sequence $\alpha_0, \alpha_1, \alpha_2, \ldots$ if the above condition is satisfied (using the coefficients of the polynomial, as above), while the *minimum polynomial* of the sequence is the (unique) monic generating polynomial of the sequence of least degree.

Note that if $u, v \in \mathsf{F}^{n\times 1}$ then the sequence

$$\alpha_0 = u^T v, \alpha_1 = u^T \tilde{A} v, \ldots, \alpha_i = u^T \tilde{A}^i v, \ldots \qquad (2.2)$$

is linearly generated; indeed, the minimum polynomial $f^{\tilde{A}}$ of $\tilde{A}$ is a generating polynomial for this sequence. In order to complete the first stage of Kaltofen and Pan's method, it is necessary to find a pair of "projection vectors" $u, v \in \mathsf{F}^{n\times 1}$ such that the minimum polynomial $f_{u^T}^{\tilde{A},v}$ for the above sequence has full degree n; in this case, $f_{u^T}^{\tilde{A},v} = f^{\tilde{A}}$.

Kaltofen and Pan show that if $\tilde{A}$ is nonsingular and has a minimum polynomial with degree n, and if the entries of vectors u and v are selected uniformly and independently from S, then the probability that the sequence (2.2) *does not* have a minimum polynomial with degree n is in $O(n/|S|)$. Alternatively, if the ground field F is finite, one can observe that the minimum polynomial of the sequence has full degree if the Toeplitz matrix

$$T = \begin{bmatrix} \alpha_{n-1} & \alpha_n & \cdots & \alpha_{2n-3} & \alpha_{2n-2} \\ \alpha_{n-2} & \alpha_{n-1} & \cdots & \alpha_{2n-4} & \alpha_{2n-3} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha_1 & \alpha_2 & \cdots & \alpha_{n-1} & \alpha_n \\ \alpha_0 & \alpha_1 & \cdots & \alpha_{n-2} & \alpha_{n-1} \end{bmatrix} \qquad (2.3)$$

is nonsingular, and that $T = \tilde{K}_u K_v$, where $\tilde{K}_u$ and K_v are the "Krylov matrices" such that $\tilde{K}_u$ has rows $u^T\tilde{A}^{n-1}, u^T\tilde{A}^{n-2}, \ldots, u^T\tilde{A}, u^T$ (from top to bottom), and K_v has columns $v, \tilde{A}v, \ldots, \tilde{A}^{n-1}v$ (from left to right). A result of Wiedemann [10] can be used to establish that if $\tilde{A}$ is nonsingular and has a minimum polynomial with degree n, and the entries of u and v are selected uniformly and independently from the finite field F, then the events "$\tilde{K}_u$ is nonsingular" and "K_v is nonsingular" are independent, and each has probability at least $1/(6\log_q n)$, where $q = |\mathsf{F}|$. Thus, for randomly chosen u and v, the probability that T is nonsingular is at least $1/(36\log_q^2 n)$.

In fact, Wiedemann states a slightly more precise bound (replacing $\log_q n$ with the smallest positive integer j such that $n \leq q + q^2 + q^3 + \cdots + q^j$). His bound is optimal in the sense that there exist nonsingular matrices $\tilde{A} \in \mathsf{F}^{n\times n}$ whose minimum polynomials have full degree for which this bound, on the probability of choosing u and v such that the corresponding matrix T is nonsingular, is tight.

Suppose now that conditioners $X, Y \in \mathsf{F}^{n\times n}$ and projection vectors $u, v \in \mathsf{F}^{n\times 1}$ that satisfy the above conditions are available. As shown by Kaltofen and Pan, the matrix $\tilde{A} = XAY \in \mathsf{F}^{n\times n}$, values $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_{2n-1} \in \mathsf{F}$ (as given in equation (2.2)) , and vectors $\tilde{b} = XB, \tilde{A}\tilde{b}, \tilde{A}^2\tilde{b}, \ldots, \tilde{A}^{n-1}\tilde{b} \in \mathsf{F}^{n\times 1}$ can all be computed using time $O(\log^2 n)$ with work $O(\mathcal{M}(n)\log n)$. The minimum polynomial of $\tilde{A}$ can then be computed by solving a nonsingular system of linear equations whose coefficient matrix is the Toeplitz matrix given in equation (2.3). Finally, given this minimum polynomial, the solution y of the system $\tilde{A}y = \tilde{b}$ can be computed as a linear combination of the vectors $\tilde{b}, \tilde{A}\tilde{b}, \ldots, \tilde{A}^{n-1}\tilde{b}$, and the solution $x = Yy$ of the original system can be computed using matrix-vector multiplication, using additional time $O(\log n)$ and work $O(n^2)$.

When applying this process in the small field case it is necessary to perform computations over a field extension E whose degree over the ground field F is at most logarithmic in n, in order to ensure that suitable conditioners $X, Y \in \mathsf{E}^{n\times n}$ and projection vectors $u, v \in \mathsf{E}^{n\times 1}$ can be found with high probability. The time and work required are increased by factors that are polynomial in $\log\log n$ and in $\log n$, respectively, if arithmetic over the ground field is considered to be at unit cost.

2.2 Solution of a Toeplitz Linear System

Kaltofen and Pan have now contributed (at least) three algorithms for this remaining part of the computation, which is used to compute the minimum polynomial of the above matrix $\tilde{A}$. The details of these algorithms will not be given here, but it will be observed that the analysis of these algorithms provided by Kaltofen and Pan imply that the improvement described in Section 3, for the *first* part of the computation, results in the improved work bound that has been claimed for the small field case.

The algorithm for solving nonsingular Toeplitz systems given (first) for this problem (in [6]) also uses time $O(\log^2 n)$ and work $O(\mathcal{M}(n)\log n)$, provided that the ground field is sufficiently large and the field characteristic is either zero or greater than n. It is clear, though, that the condition on field size can be discarded if the field characteristic is in this range, without increasing the time or work needed by more than a linear factor — for either the field is infinite or (when the characteristic is positive) the size of the field is at least as large as the characteristic, so that Kaltofen and Pan's method can certainly be used reliably by working over an extension E of F whose degree over the ground field is less than or equal to three. In this case every operation in E can be implemented using a constant number of operations over the ground field, so that the overhead required to work in a field extension does not affect the asymptotic cost of the method.

Kaltofen and Pan presented a second, more general, method in [7] that uses $O(\log^2 n)$ time and $O(\mathcal{M}(n))$ work to reduce the problem of solving a nonsingular $n\times n$ Toeplitz linear system to that of solving a dense, unstructured system whose order is $\frac{n}{p}$, where p is the characteristic of the ground field. If p is sufficiently large — in particular, $\sqrt{n} < p \le n$ — then this establishes once again that the entire computation can be performed using time $O(\log^2 n)$ and $O(\mathcal{M}(n)\log n)$ work, without requiring that the ground field F is large; for if the characteristic of F is positive and greater than $\sqrt{n}$, then so is the size of F. In this case computations can be performed reliably, and with low overhead, using a field extension whose degree is at most six. The first stage of Kaltofen and Pan's algorithm in [6] can be used to reduce the problem to that of solving a nonsingular $n\times n$ Toeplitz linear system. A single iteration of the process described in [7] reduces this to the problem of solving a dense, unstructured linear system whose order is less than $\sqrt{n}$ — and, since this is less than the characteristic of the ground field, the algorithm of [6] can be used to complete the computation.

The algorithm of [7] can be used instead if the characteristic p is positive and less than or equal to $\sqrt{n}$. However, the time required increases by an $O(\log_p n)$ factor, even when the ground field is sufficiently large, and it may be necessary to work in an extension with logarithmic degree over F in order to keep the probability of failure small. In this case the additional overhead introduced by the use of field extensions is also nontrivial, and increases both the time and work that must be used.

A more recent algorithm of Kaltofen and Pan [8] can be used to solve a nonsingular $n\times n$ Toeplitz-like system of linear equations over a sufficiently large field of positive characteristic $p \le \sqrt{n}$ in $O(\log^3 n)$ time using $O(n^2 \log n \log\log n)$ work. This can be used for computations over small finite fields, by working over a field extension whose degree is at most logarithmic in n. If one counts operations over the ground field rather than over the extension, then the time used increases to $O((\log n)^{3+o(1)})$ and the work increases by slightly more than a logarithmic factor. Thus the work for this step is still in $O(\mathcal{M}(n)\log n)$ assuming, again, that $\mathcal{M}(n) \ge n^2 \log^3 n$.

3 Matrix Conditioning and Projections over Small Fields

In this section, it will be shown that if F is a finite field of size $q \ge 2$, and one chooses conditioners X and Y to be the identity matrix and a (dense, unstructured) matrix whose entries are selected uniformly and independently from F, respectively, then the probability that $\tilde{A} = XAY$ is nonsingular and has a minimum polynomial $f^{\tilde{A}}$ with degree n can be bounded away from zero. Furthermore, if Y is chosen uniformly from $\mathsf{F}^{n\times n}$ and projection vectors u and v are chosen uniformly and independently from $\mathsf{F}^{n\times 1}$, then the corresponding matrix T (given in equation (2.3)) is nonsingular with probability bounded away from zero as well. In particular, for $q \ge 2$, let

$$\rho_q = \begin{cases} \frac{1}{10}\sqrt{2}, & \text{if } q = 2, \\ \sqrt{\frac{2}{15}}, & \text{if } q = 3, \\ \frac{6}{\sqrt{135}}, & \text{if } q = 4, \\ \left(1-\frac{1}{q}\right)\left(1-\frac{1}{q-1}\right), & \text{if } q \ge 5. \end{cases} \tag{3.1}$$

Theorem 3.1. *Suppose* F *is a finite field of size* q *and* $A \in \mathsf{F}^{n\times n}$ *is nonsingular. If the entries of a matrix* $Y \in \mathsf{F}^{n\times n}$ *and vector* $v \in \mathsf{F}^{n\times 1}$ *are chosen uniformly and independently from* F*, then the probability that the matrix* $\tilde{A} = AY$ *is nonsingular and has a minimum polynomial with degree* n*, and the* n *vectors* $v, \tilde{A}v, \tilde{A}^2 v, \ldots, \tilde{A}^{n-1}v$ *are linearly independent, is equal to the product of* $\left(1-\frac{1}{q}\right)$ *and the probability that a randomly chosen matrix in* $\mathsf{F}^{n\times n}$ *is nonsingular, and is therefore greater than or equal to* ρ_q*.*

The next result is an immediate corollary, since it refers to a condition that is implied by the one described in the theorem.

Corollary 3.2. *Suppose* F *is a finite field of size* q *and* $A \in \mathsf{F}^{n\times n}$ *is nonsingular. If the entries of a matrix* $Y \in \mathsf{F}^{n\times n}$ *are chosen uniformly and independently from* F *then the probability that the matrix* $\tilde{A} = AY$ *is nonsingular, and has a* *minimum polynomial with degree n, is greater than or equal to ρ_q.*

It will also be shown that, since a suitable conditioner and one projection can be found with positive probability, a suitable conditioner and *two* projections can also be chosen with positive probability.

Theorem 3.3. *Suppose F is a finite field of size q and $A \in \mathsf{F}^{n\times n}$ is nonsingular. If the entries of a matrix $Y \in \mathsf{F}^{n\times n}$ and vectors $u, v \in \mathsf{F}^{n\times 1}$ are chosen uniformly and independently from F, then the probability that the matrix $\tilde{A} = AY$ is nonsingular, and the minimum polynomial for the sequence*

$$u^T v, u^T \tilde{A} v, u^T \tilde{A}^2 v, \ldots, u^T \tilde{A}^i v, \ldots$$

has degree n, is greater than or equal to $\max(2\rho_q - 1, \frac{1}{4}\rho_q^2)$.

It is easily checked that $\frac{1}{4}\rho_q^2 \geq 2\rho_q - 1$ if $q \leq 4$, and $\frac{1}{4}\rho_q^2 \leq 2\rho_q - 1 = 1 - \frac{4}{q}$ if $q \geq 5$. Hence the following is equivalent to the above theorem, and simply provides a more explicit statement of the same bounds.

Corollary 3.4. *Suppose F is a finite field of size q and $A \in \mathsf{F}^{n\times n}$ is nonsingular. If the entries of a matrix $Y \in \mathsf{F}^{n\times n}$ and vectors $u, v \in \mathsf{F}^{n\times 1}$ are chosen uniformly and independently from F, then the matrix $\tilde{A} = AY$ is nonsingular and the minimum polynomial for the sequence*

$$u^T v, u^T \tilde{A} v, u^T \tilde{A}^2 v, \ldots, u^T \tilde{A}^i v, \ldots$$

has degree n with probability greater than or equal to σ_q, for

$$\sigma_q = \begin{cases} \frac{1}{200}, & \textit{if } q = 2, \\ \frac{1}{30}, & \textit{if } q = 3, \\ \frac{1}{15}, & \textit{if } q = 4, \\ 1 - \frac{4}{q}, & \textit{if } q \geq 5. \end{cases} \tag{3.2}$$

In particular, this probability is greater than or equal to one-half if $q \geq 8$.

Theorem 3.3 implies that the first stage of Kaltofen and Pan's algorithm can be performed reliably, over all fields, without taking field extensions — provided that one is willing to use a dense, "unstructured" matrix as a conditioner. This is unacceptable for a sparse- or structured-linear system solver, and it would be desirable (and much more useful, for the solution of sparse and structured systems over small fields) if results similar to the above could be established when randomly chosen *sparse* or *structured* conditioners X and Y are used. However, the above results do show that the first stage of Kaltofen and Pan's parallel algorithm for the solution of a nonsingular $n \times n$ linear system can be implemented using time $O(\log^2 n)$ with $O(\mathcal{M}(n) \log n)$ work without taking field extensions. Since the work required for the second part of this computation is already known to be in $O(\mathcal{M}(n) \log n)$, the work required for the entire computation over small fields can be reduced to $O(\mathcal{M}(n) \log n)$ without changing the required parallel time, as has been claimed.

Theorem 3.1 will be proved using a bound on the density of nonsingular matrices in $\mathsf{F}^{n\times n}$ that is slightly better (for small q) than the bound that is commonly used. This bound will be established in the next subsection, using a slightly more elaborate application of techniques that have been used already to establish prior bounds. The improved bound will be used to prove Theorem 3.1 in the subsection after that. Theorem 3.3 will proved at the end of Section 3.

3.1 The Density of Nonsingular Matrices over Small Fields

Let $\tau_q(n)$ be the probability that a uniformly chosen matrix in $\mathsf{F}^{n\times n}$ is nonsingular. It is well known that $\tau_q(n) \geq \frac{1}{4}$ for all n if $q = 2$, and that $\tau_q(n) \geq 1 - \frac{1}{q-1}$ if $q \geq 3$. It is easy to adapt a proof of these bounds to show, as well, that

$$\tau_q(n) \geq \begin{cases} \frac{1}{5}\sqrt{2}, & \text{if } q = 2, \\ \sqrt{\frac{3}{10}}, & \text{if } q = 3, \\ \frac{8}{\sqrt{135}}, & \text{if } q = 4. \end{cases} \tag{3.3}$$

Lemma 3.5. *If $x_i \in \mathbf{R}$ and $0 < x_i < 1$ for $1 \leq i \leq n$, then*

$$\prod_{i=1}^{n}(1 - x_i) \geq 1 - \sum_{i=1}^{n} x_i.$$

Proof. If $\sum_{i=1}^{n} x_i \geq 1$ then the lemma is trivial, since the left hand size is positive and the right hand size less than or equal to zero. Otherwise both quantities are positive, so it is sufficient to show that

$$\ln \prod_{i=1}^{n}(1 - x_i) \geq \ln\left(1 - \sum_{i=1}^{n} x_i\right).$$

If $0 < z < 1$ then

$$\ln(1 - z) = -\sum_{j\geq 1} \frac{1}{j!} z^j$$

This can be used to express each of the above logarithms as an infinite sum of symmetric polynomials in $x_1, x_2, \ldots, x_n$. The above inequality is then a consequence of the fact that $(\sum_{i=1}^{n} x_i)^j \geq \sum_{i=1}^{n} x_i^j$ for $j \geq 1$, whenever $x_i > 0$ for all i. □

It is well known (and easy to argue) that

$$\tau_q(n) = \prod_{i=1}^{n}\left(1 - \frac{1}{q^i}\right) \tag{3.4}$$

for every prime power q and every positive integer n. Now it follows by the lemma that, for every integer m between 0 and n,

$$\begin{aligned} \tau_q(n) &= \prod_{i=1}^{m}\left(1 - \frac{1}{q^i}\right) \cdot \prod_{j=m+1}^{n}\left(1 - \frac{1}{q^j}\right) \\ &\geq \prod_{i=1}^{m}\left(1 - \frac{1}{q^i}\right) \cdot \left(1 - \sum_{j=m+1}^{n} q^j\right) \\ &\geq \prod_{i=1}^{m}\left(1 - \frac{1}{q^i}\right) \cdot \left(1 - \sum_{j\geq m+1} q^j\right) \\ &= \prod_{i=1}^{m}\left(1 - \frac{1}{q^i}\right) \cdot \left(1 - \frac{1}{q^m(q-1)}\right). \end{aligned}$$

Set $b_q(m) = (\prod_{j=1}^{m}(1 - q^{-i}))(1 - 1/(q^m(q-1)))$; it has been shown that $\tau_q(n) \geq b_q(m)$ whenever $n \geq m \geq 0$. Since $\tau_q(n) \geq \tau_q(n+1)$ for all q and n, it is easy to see that $\tau_q(n) \geq b_q(m)$ for $n < m$ as well. Now, it is easy to check that $b_2(3) \geq \frac{1}{5}\sqrt{2}$, $b_3(1) \geq \sqrt{\frac{3}{10}}$, and $b_4(3) \geq \frac{8}{\sqrt{135}}$, so that $\tau_2(n) \geq \frac{1}{5}\sqrt{2}$, $\tau_3(n) \geq \sqrt{\frac{3}{10}}$, and $\tau_4(n) \geq \frac{8}{\sqrt{135}}$ for all $n \geq 1$, as claimed. The better known bounds, $\tau_2(n) \geq \frac{1}{4}$ and $\tau_q(n) \geq 1 - \frac{1}{q-1}$ for $q \geq 3$, are obtained by considering $b_2(1)$ and $b_q(0)$ for $q \geq 3$.

3.2 Proof of Theorem 3.1

Since the matrix A is nonsingular, there is exactly one matrix $Y \in \mathsf{F}^{n\times n}$ such that $Z = AY$, for any given matrix $Z \in \mathsf{F}^{n\times n}$. Thus it is sufficient (indeed, equivalent) to prove that if the entries of a matrix $Z \in \mathsf{F}^{n\times n}$ and vector $v \in \mathsf{F}^{n\times 1}$ are uniformly and independently chosen from F, then the probability that Z is nonsingular and the n vectors $v, Zv, Z^2v, \ldots, Z^{n-1}v$ are linearly independent is at least ρ_q when $|\mathsf{F}| = q$.

Once again, vectors $v, Zv, Z^2v, \ldots, Z^{n-1}v$ can only be linearly independent if the minimum polynomial f^Z of Z has degree n. A matrix Z satisfies this condition, and is nonsingular, if and only if f^Z has degree n and $f^Z(0) \neq 0$.

Fix one polynomial $f = x^n + f_{n-1}x^{n-1} + \cdots + f_1x + f_0 \in \mathsf{F}[x]$ such that $f(0) \neq 0$. Following the notation of Wiedemann [10], let $\Phi(f)$ be the proportion of polynomials in $\mathsf{F}[x]$ with degree less than n that are relatively prime to f, so that there are exactly $q^n\Phi(f)$ polynomials $g \in \mathsf{F}[x]$ with degree less than n such that $\gcd(f, g) = 1$. Since f has degree n, a matrix $Z \in \mathsf{F}^{n\times n}$ has minimum polynomial f if and only if Z is similar to the *companion matrix* C_f for f,

$$C_f = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & -f_0 \\ 1 & 0 & \cdots & 0 & 0 & -f_1 \\ 0 & 1 & \cdots & 0 & 0 & -f_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & -f_{n-2} \\ 0 & 0 & \cdots & 0 & 1 & -f_{n-1} \end{bmatrix} \in \mathsf{F}^{n\times n}. \qquad (3.5)$$

It is well known that the number of nonsingular matrices similar to any given nonsingular matrix W equals the number of nonsingular matrices in $\mathsf{F}^{n\times n}$ divided by the number of nonsingular matrices that commute with W. Indeed, if U and V are nonsingular matrices in $\mathsf{F}^{n\times n}$ then $U^{-1}WU = V^{-1}WV$ if and only if $U = CV$ for some nonsingular matrix C such that $CW = WC$.

It is also known that the only matrices in $\mathsf{F}^{n\times n}$ that commute with a companion matrix C_f belong to the algebra $\mathsf{F}[C_f] \subseteq \mathsf{F}^{n\times n}$. This algebra is isomorphic to $\mathsf{F}[x]/(f)$ — so it has size q^n and includes exactly $q^n\Phi(f)$ invertible elements.

Thus there are exactly $q^n\Phi(f)$ nonsingular matrices in $\mathsf{F}^{n\times n}$ that commute with C_f, and there are exactly $\tau_q(n)q^{n^2}/(q^n\Phi(f)) = q^{n^2-n}\frac{\tau_q(n)}{\Phi(f)}$ matrices in $\mathsf{F}^{n\times n}$ with minimum polynomial f.

Now, fix any one matrix Z with minimum polynomial f. As argued by Wiedemann [10], the number of vectors $v \in \mathsf{F}^{n\times 1}$ such that $v, Zv, Z^2v, \ldots, Z^{n-1}v$ are linearly independent is exactly $q^n\Phi(f)$. Thus, the number of pairs (Z, v) of matrices $Z \in \mathsf{F}^{n\times n}$ and vectors $v \in \mathsf{F}^{n\times 1}$ such that Z has minimum polynomial f, and $v, Zv, Z^2v, \ldots, Z^{n-1}v$ are linearly independent, is exactly $\left(q^{n^2-n}\frac{\tau_q(n)}{\Phi(f)}\right) \cdot q^n\Phi(f) = q^{n^2}\tau_q(n)$.

Since this quantity is independent of the choice of f, the number of pairs (Z, v) of matrices $Z \in \mathsf{F}^{n\times n}$ and vectors $v \in \mathsf{F}^{n\times 1}$, such that Z is nonsingular, and $v, Zv, Z^2v, \ldots, Z^{n-1}v$ are linearly independent, is the product of $q^{n^2}\tau_q(n)$ and the number of monic polynomials $f \in \mathsf{F}[x]$ with degree n such that $f(0) \neq 0$ — that is, $\left(q^{n^2}\tau_q(n)\right)q^{n-1}(q-1) = q^{n^2+n-1}(q-1)\tau_q(n)$. Since there are q^{n^2+n} choices of Z and v, the probability of randomly selecting a matrix $Z \in \mathsf{F}^{n\times n}$ and $v \in \mathsf{F}^{n\times 1}$ that satisfy the above conditions is exactly $\left(1 - \frac{1}{q}\right)\tau_q(n)$.

It now follows that $\rho_2 \geq \frac{1}{10}\sqrt{2}$ since $\tau_2(n) \geq \frac{1}{5}\sqrt{2}$; $\rho_3 \geq \sqrt{\frac{2}{15}}$ since $\tau_3(n) \geq \sqrt{\frac{3}{10}}$; $\rho_4 \geq \frac{6}{\sqrt{135}}$ since $\tau_4(n) \geq \frac{8}{\sqrt{135}}$; and $\rho_q \geq (1 - \frac{1}{q})(1 - \frac{1}{q-1})$ since $\tau_q(n) \geq (1 - \frac{1}{q-1})$ for $q \geq 5$.

3.3 Proof of Theorem 3.3

Theorem 3.1 implies that if a matrix Y is uniformly chosen from $\mathsf{F}^{n\times n}$, and a vector v is uniformly and independently chosen from $\mathsf{F}^{n\times 1}$, then the probability that $\tilde{A} = AY$ is nonsingular and vectors $v, \tilde{A}v, \tilde{A}^2v, \ldots, \tilde{A}^{n-1}v$ are linearly independent is $(1 - \frac{1}{q})\tau_q(n)$. Clearly, this is also the probability that $\tilde{A}$ is nonsingular and column vectors $u, \tilde{A}^Tu, (\tilde{A}^T)^2u, \ldots, (\tilde{A}^T)^{n-1}u$ are linearly independent — or, equivalently, that the matrix $\tilde{A}$ is nonsingular and row vectors $u^T, u^T\tilde{A}, u^T\tilde{A}^2, \ldots, u^T\tilde{A}^{n-1}$ are linearly independent, if the matrix Y is chosen uniformly and randomly from $\mathsf{F}^{n\times n}$ and the vector u is chosen uniformly, independently, and randomly from $\mathsf{F}^{n\times 1}$.

Now suppose that Y is uniformly and randomly chosen from $\mathsf{F}^{n\times n}$, that $\tilde{A} = AY$, and that a pair of vectors u and v are chosen uniformly and independently from $\mathsf{F}^{n\times 1}$. The probability that either $\tilde{A}$ is singular or vectors $v, \tilde{A}v, \tilde{A}^2v, \ldots, \tilde{A}^{n-1}v$ are linearly dependent (or both) is at most $1 - \rho_q$. Recall that the matrix K_v is defined to have columns $v, \tilde{A}v, \tilde{A}^2v, \ldots, \tilde{A}^{n-1}v$ — so this bounds the probability that at least one of $\tilde{A}$ and K_v is singular. The probability that $\tilde{A}$ is singular or row vectors $u^T, u^T\tilde{A}, u^T\tilde{A}^2, \ldots, u^T\tilde{A}^{n-1}$ are linearly dependent is also at most $1 - \rho_q$. The matrix $\tilde{K}_u$ has been defined to have rows $u^T, u^T\tilde{A}, \ldots, u^T\tilde{A}^{n-1}$ (from bottom to top), so this also bounds the probability that at least one of $\tilde{A}$ and $\tilde{K}_u$ is singular. Clearly, the probability that one or more of $\tilde{A}$, K_v, and $\tilde{K}_u$ is singular is at most $2(1 - \rho_q)$, since this condition would imply that either "at least one of $\tilde{A}$ and K_v is singular," or "at least one of $\tilde{A}$ and $\tilde{K}_u$ is singular," or both. It follows that the probability that all three of $\tilde{A}$, K_v, and $\tilde{K}_u$ are nonsingular is at least $1 - 2(1 - \rho_q) = 2\rho_q - 1$.

It has already been noted (in Section 2) that if $\tilde{K}_u$ and K_v are both nonsingular, then the sequence

$$u^Tv, u^T\tilde{A}v, u^T\tilde{A}^2v, \ldots, u^T\tilde{A}^iv, \ldots$$

has a minimum polynomial with degree n. Thus, $\tilde{A}$ is nonsingular and this sequence has a minimum polynomial with degree n with probability at least $2\rho_q - 1$, as claimed in Theorem 3.3.

It remains only to show that this probability is also at least $\frac{1}{4}\rho_q^2$. In order to prove this, we will consider the following events, assuming as usual that A is nonsingular, and $\tilde{A} = AY$, where the entries of the matrix $Y \in \mathsf{F}^{n\times n}$ and vectors $u, v \in \mathsf{F}^{n\times 1}$ are selected uniformly and independently from F (Recall that $f^{\tilde{A}}$ is the minimum polynomial of the matrix $\tilde{A}$, and that $\Phi(f^{\tilde{A}})$ is the proportion of polynomials with degree less than the degree of $f^{\tilde{A}}$ in $\mathsf{F}[x]$ that are relatively prime to $f^{\tilde{A}}$):

E0: $\tilde{A}$ and K_v are nonsingular.

E1: $\tilde{A}$ and K_v are nonsingular, and $\Phi(f^{\tilde{A}}) < \frac{1}{2}\rho_q$.

E2: $\tilde{A}$ is nonsingular, $f^{\tilde{A}}$ has degree n, and $\Phi(f^{\tilde{A}}) < \frac{1}{2}\rho_q$.

E3: $\tilde{A}$ and K_v are nonsingular, and $\Phi(f^{\tilde{A}}) \geq \frac{1}{2}\rho_q$.

E4: $\tilde{A}$, K_v, and $\tilde{K}_u$ are all nonsingular, and $\Phi(f^{\tilde{A}}) \geq \frac{1}{2}\rho_q$.

It has already been shown that the probability Pr(E0) of event E0 is greater than or equal to ρ_q. Since event E0 is the disjoint union of events E1 and E3, $\Pr(\mathsf{E1}) + \Pr(\mathsf{E3}) = \Pr(\mathsf{E0}) \geq \rho_q$.

Since event E1 can only occur if the minimum polynomial of $\tilde{A}$ has degree n, E1 implies E2, so that $\mathsf{E1} = \mathsf{E1} \wedge \mathsf{E2}$ and

$$\begin{aligned}\Pr(\mathsf{E1}) &= \Pr(\mathsf{E1} \wedge \mathsf{E2}) \\ &= \Pr(\mathsf{E1} \mid \mathsf{E2}) \cdot \Pr(\mathsf{E2}) \\ &\leq \Pr(\mathsf{E1} \mid \mathsf{E2}).\end{aligned}$$

The conditional event "E1 | E2" is the event that for a (randomly chosen) vector v, the vectors $v, \tilde{A}v, \tilde{A}^2v, \ldots, \tilde{A}^{n-2}v$ are linearly independent, *provided that* $f^{\tilde{A}}$ has degree n and $\Phi(f^{\tilde{A}}) < \frac{1}{2}\rho_q$. This clearly has probability less than or equal to $\frac{1}{2}\rho_q$, since the probability that such a vector v can be found for any fixed matrix $\tilde{A}$, whose minimum polynomial $f^{\tilde{A}}$ has degree n, is $\Phi(f^{\tilde{A}})$. Thus $\Pr(\mathsf{E1}) \leq \frac{1}{2}\rho_q$, and $\Pr(\mathsf{E3}) = \Pr(\mathsf{E0}) - \Pr(\mathsf{E1}) \geq \rho_q - \frac{1}{2}\rho_q = \frac{1}{2}\rho_q$.

By essentially the same argument, $\Pr(\mathsf{E4} \mid \mathsf{E3}) \geq \frac{1}{2}\rho_q$, for this is the probability that a vector u can be found such that $\tilde{K}_u$ is nonsingular, provided that $\tilde{A}$ is nonsingular, has a minimum polynomial $f^{\tilde{A}}$ with degree n, and $\Phi(f^{\tilde{A}}) \geq \frac{1}{2}\rho_q$. Thus (since event E4 implies E3),

$$\begin{aligned}\Pr(\mathsf{E4}) &= \Pr(\mathsf{E4} \wedge \mathsf{E3}) \\ &= \Pr(\mathsf{E4} \mid \mathsf{E3}) \cdot \Pr(\mathsf{E3}) \\ &\geq \frac{1}{2}\rho_q \cdot \frac{1}{2}\rho_q = \frac{1}{4}\rho_q^2.\end{aligned}$$

Since event E4 clearly implies that $\tilde{A}$, K_v, and $\tilde{K}_u$ are all nonsingular, it follows that the probability that $\tilde{A}$ is nonsingular and the minimum polynomial of the sequence

$$u^Tv, u^T\tilde{A}v, u^T\tilde{A}^2v, \ldots, u^T\tilde{A}^iv, \ldots$$

has degree n is also at least $\frac{1}{4}\rho_q^2$, as desired.

4 An Explicit Parallel Algorithm for Matrix Inversion

Suppose again that $A \in \mathsf{F}^{n\times n}$ is nonsingular. It is easy to find matrices $X, Y \in \mathsf{F}^{n\times n}$ such that the matrix $\tilde{A} = XAY$ is nonsingular and has a minimum polynomial with degree n, and vectors $u, v \in \mathsf{F}^{n\times 1}$ such that the minimum polynomial of the sequence

$$u^Tv, u^T\tilde{A}v, u^T\tilde{A}^2v, \ldots, u^T\tilde{A}^iv, \ldots$$

also has degree n. In particular, as noted in Section 2, one can choose X and Y using either of the constructions of Kaltofen and Pan [6], [7], and then select u and v uniformly and randomly (and independently) from $\mathsf{F}^{n\times 1}$, if $|\mathsf{F}| \in \Omega(n^2)$; one can use the construction described in Section 3 otherwise.

Suppose, then, that $\tilde{A}$ has a minimum polynomial $f^{\tilde{A}} = x^n + f_{n-1}x^{n-1} + \cdots + f_1x + f_0$, that $f(0) \neq 0$ (so that $\tilde{A}$ is nonsingular), and that u, v are vectors in $\mathsf{F}^{n\times 1}$ such that the minimum polynomial of the above sequence has degree n as well. Consider the "Krylov" matrices $\tilde{K}_u, K_v \in \mathsf{F}^{n\times n}$ introduced in Section 2, so that $\tilde{K}_u$ has rows $u^T\tilde{A}^{n-1}, u^T\tilde{A}^{n-2}, \ldots, u^T\tilde{A}, u^T$ (from top to bottom), and K_v has columns $v, \tilde{A}v, \tilde{A}^2v, \ldots, \tilde{A}^{n-1}v$ (from left to right). Then $\tilde{K}_u$ and K_v are both nonsingular and have the Toeplitz matrix T given in (2.3) as a product:

$$\tilde{K}_uK_v = T. \tag{4.1}$$

Let $C_{f^{\tilde{A}}}$ be the companion matrix for $f^{\tilde{A}}$ (as shown in equation (3.5)). Since the matrix K_v is nonsingular, the following matrix identity is correct, and easily verified by comparing the actions of the matrices shown on the vectors $v, \tilde{A}v, \tilde{A}^2v, \ldots, \tilde{A}^{n-1}v$.

$$\tilde{A} = K_vC_{f^{\tilde{A}}}K_v^{-1}. \tag{4.2}$$

Now, equations (4.1) and (4.2) provide a way to compute the companion matrix $C_{f^{\tilde{A}}}$ from $\tilde{A}$, $\tilde{K}_u$, K_v, and the inverse of the Toeplitz matrix T.

$$\begin{aligned}C_{f^{\tilde{A}}} &= K_v^{-1}\tilde{A}K_v && \text{(by (4.2))} \\ &= T^{-1}\tilde{K}_u\tilde{A}K_v. && \text{(by (4.1))}\end{aligned} \tag{4.3}$$

This clearly implies that $\tilde{A} = \tilde{K}_u^{-1}TC_{f^{\tilde{A}}}K_v^{-1}$, and this allows us to express the inverse of $\tilde{A}$ as a product of $\tilde{K}_u$, K_v, and the inverses of the Toeplitz matrix T and companion matrix $C_{f^{\tilde{A}}}$:

$$\tilde{A}^{-1} = K_vC_{f^{\tilde{A}}}^{-1}T^{-1}\tilde{K}_u. \tag{4.4}$$

Finally, since $\tilde{A} = XAY$, it is easy to recover the inverse of A from the inverse of $\tilde{A}$:

$$A^{-1} = Y\tilde{A}^{-1}X = YK_vC_{f^{\tilde{A}}}^{-1}T^{-1}\tilde{K}_uX. \tag{4.5}$$

Thus, A^{-1} can be computed from A as follows.

1. Choose conditioners $X, Y \in \mathsf{F}^{n\times n}$ and vectors $u, v \in \mathsf{F}^{n\times n}$ using one of the constructions described by Kaltofen and Pan ([6], [7]) if $|\mathsf{F}| \in \Omega(n^2)$, or as described in Section 3 otherwise.

2. $\tilde{A} := XAY$

3. Compute the vectors $\tilde{A}^iv$ and $(\tilde{A}^T)^iu$ for $0 \leq i \leq n-1$. Use these to construct the matrices $\tilde{K}_u, K_v \in \mathsf{F}^{n\times n}$ described above, and the Toeplitz matrix $T = \tilde{K}_uK_v \in \mathsf{F}^{n\times n}$.

4. Attempt to compute T^{-1}; report failure, and stop, unless this step succeeds.

5. $C := T^{-1}\tilde{K}_u\tilde{A}K_v$; this is the companion matrix for the minimum polynomial of $\tilde{A}$.

6. Compute the inverse of the companion matrix C.

7. Return $A^{-1} = YK_vC^{-1}T^{-1}\tilde{K}_uX$.

Steps 1, 2, 5 and 7 can each be performed using time $O(\log n)$ and work $O(\mathcal{M}(n))$. Step 3 requires two "Krylov matrix" computations, and can be performed in time $O(\log^2 n)$ and using work $O(\mathcal{M}(n)\log n)$. It will be shown in the next subsection that step 4 can be reduced to the problem of solving a constant number of Toeplitz linear systems, so that this step can be performed in time $O((\log n)^{3+o(1)})$ and using work $O(\mathcal{M}(n)\log n)$ if the characteristic of F is positive and less than or equal to $\sqrt{n}$, or time $O(\log^2 n)$ with work $O(\mathcal{M}(n)\log n)$ otherwise. Finally, if the companion matrix $C_{f^{\tilde{A}}}$ is nonsingular then $f^{\tilde{A}}(0)\neq 0$ and the inverse of $C_{f^{\tilde{A}}}$ is easily generated: In particular, if $f^{\tilde{A}} = x^n + f_{n-1}x^{n-1} + \cdots + f_1x + f_0$, so that $f_0 = f^{\tilde{A}}(0) \neq 0$, then it easily checked that

$$C_{f^{\tilde{A}}}^{-1} = \begin{bmatrix} -f_1/f_0 & 1 & 0 & \cdots & 0 & 0 \\ -f_2/f_0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -f_{n-2}/f_0 & 0 & 0 & \cdots & 1 & 0 \\ -f_{n-1}/f_0 & 0 & 0 & \cdots & 0 & 1 \\ -1/f_0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Thus the entries of $C_{f^{\tilde{A}}}^{-1}$ can computed from the coefficients of $f^{\tilde{A}}$ using constant time and work $O(n^2)$.

It follows that the above "explicit" algorithm for matrix inversion can be implemented at essentially the cost required by the best currently known algorithm for solving dense linear systems — namely, time $O((\log n)^{3+o(1)})$ (or $O(\log^2 n)$ if the characteristic of F is at least $\sqrt{n}$) and work $O(\mathcal{M}(n)\log n)$.

4.1 Explicit Parallel Inversion of a Nonsingular Toeplitz Matrix

A formula of Gohberg and Semencul can be used to compute the inverse of an $n \times n$ Toeplitz matrix T from its first and last columns, provided that the top left entry of this matrix is nonzero — see Gohberg and Semencul [5] (in Russian), or, for example, Brent, Gustavson and Yun [2] or Kaltofen and Pan [6] for a statement of this version of the "Gohberg-Semencul formula." Unfortunately, an algorithm based on this formula would include a division by zero if the top left entry of T^{-1} is zero.

Fortunately, Gohberg and Semencul state an additional formula which can be used to compute T^{-1} from the first and last column of the inverse of an $(n+1)\times(n+1)$ matrix $\hat{T}$ that has T as its principal $n \times n$ minor — provided that the top left entry of the inverse of $\hat{T}$ is nonzero.

In particular, suppose the matrix T is as shown in equation (2.1), let $a_{-1}, a_{2n-1} \in \mathsf{F}$, and suppose

$$\hat{T} = \begin{bmatrix} a_{n-1} & a_{n-2} & \cdots & a_0 & a_{-1} \\ a_n & a_{n-1} & \cdots & a_1 & a_0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{2n-2} & a_{2n-3} & \cdots & a_{n-1} & a_{n-2} \\ a_{2n-1} & a_{2n-2} & \cdots & a_n & a_{n-1} \end{bmatrix} = \begin{bmatrix} T & c \\ r^T & a_{n-1} \end{bmatrix} \in \mathsf{F}^{(n+1)\times(n+1)},$$

for

$$r = \begin{bmatrix} a_{2n-1} \\ a_{2n-2} \\ \vdots \\ a_n \end{bmatrix} \quad \text{and} \quad c = \begin{bmatrix} a_{-1} \\ a_0 \\ \vdots \\ a_{n-2} \end{bmatrix}.$$

Theorem 4.1 (Gohberg and Semencul). *If solutions*

$$x = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \textit{and} \quad y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}$$

exist for the systems of equations $\hat{T}x = e_0$ and $\hat{T}y = e_n$, where e_0 and e_n are the first and last columns of the identity matrix of order $n+1$, and if $x_0 \neq 0$, then

$$T^{-1} = \frac{1}{x_0}\left(L_x \cdot U_y - L_y \cdot U_x\right),$$

where

$$L_x = \begin{bmatrix} x_0 & & & & 0 \\ x_1 & x_0 & & & \\ \vdots & \vdots & \ddots & & \\ x_{n-2} & x_{n-3} & \cdots & x_0 & \\ x_{n-1} & x_{n-2} & \cdots & x_1 & x_0 \end{bmatrix},$$

$$U_y = \begin{bmatrix} y_n & y_{n-1} & \cdots & y_2 & y_1 \\ & y_n & \cdots & y_3 & y_2 \\ & & \ddots & \vdots & \vdots \\ & & & y_n & y_{n-1} \\ 0 & & & & y_n \end{bmatrix},$$

$$L_y = \begin{bmatrix} y_0 & & & & 0 \\ y_1 & y_0 & & & \\ \vdots & \vdots & \ddots & & \\ y_{n-2} & y_{n-3} & \cdots & y_0 & \\ y_{n-1} & y_{n-2} & \cdots & y_1 & y_0 \end{bmatrix},$$

and

$$U_x = \begin{bmatrix} x_n & x_{n-1} & \cdots & x_2 & x_1 \\ & x_n & \cdots & x_3 & x_2 \\ & & \ddots & \vdots & \vdots \\ & & & x_n & x_{n-1} \\ 0 & & & & x_n \end{bmatrix}.$$

This theorem is also stated[1] and used by Brent, Gustavson, and Yun [2].

Now, in order to reduce the problem of computing T^{-1} to the problem of solving a small number of nonsingular Toeplitz linear systems, it suffices to argue that it is easy to choose elements a_{-1} and a_{2n-1} from F such that the corresponding matrix $\hat{T}$ satisfies all the conditions of the above theorem. Brent, Gustavson, and Yun have done this already. In particular, they consider the matrix $T_\beta \in \mathsf{F}^{(n+1)\times(n+1)}$ resulting from the selection $a_{2n-1} = \beta$ and $a_0 = 0$,

$$T_\beta = \begin{bmatrix} T & c_0 \\ r_\beta^T & a_{n-1} \end{bmatrix} = \begin{bmatrix} T & 0 \\ r_\beta^T & \gamma \end{bmatrix} \begin{bmatrix} I_n & s \\ 0 & 1 \end{bmatrix},$$

where

$$r_\beta = \begin{bmatrix} \beta \\ a_{2n-2} \\ a_{2n-3} \\ \vdots \\ a_n \end{bmatrix}, \quad c_0 = \begin{bmatrix} 0 \\ a_0 \\ a_1 \\ \vdots \\ a_{n-2} \end{bmatrix}, \quad s = \begin{bmatrix} s_0 \\ s_1 \\ s_2 \\ \vdots \\ s_{n-1} \end{bmatrix} = T^{-1}c_0,$$

and $\gamma = a_{n-1} - r_\beta^T T^{-1} c_0 = a_{n-1} - r_\beta^T s$.

Clearly, $\det T_\beta = \gamma \det T$ and (since T_β is Toeplitz) Cramer's rule implies that the top left entry of T_β^{-1} is $\frac{\det T}{\det T_\beta} = \frac{1}{\gamma}$ whenever $\gamma \neq 0$.

Now,

$$\gamma = a_{n-1} - r_\beta^T s = -s_0\beta - \sum_{j=1}^{n-1} s_j a_{2n-1-j} + a_{n-1}.$$

Since T and c_0 do not depend on β, and since $s = T^{-1}c_0$, the entries $s_0, s_1, \ldots, s_{n-1}$ are independent of β as well. Thus, γ is either linear in β or independent of β, and is only independent of β if $s_0 = 0$.

The following result follows from Lemma 4 in [2].

Lemma 4.2 (Brent, Gustavson, and Yun). *If T and s are as above and T is nonsingular then $s_0 \neq 0$.*

It follows that $\hat{T}$ is singular for at most one choice of β in F. Otherwise, $\hat{T}$ is nonsingular and has an inverse whose top left entry is nonzero, so that $\hat{T}$ can be used to compute the inverse of T. Thus, the problem of inverting T can be reduced to the solution of Toeplitz linear systems, in a deterministic manner, by attempting to solve four systems,

$$T_0 x = e_0, \quad T_0 y = e_n, \quad T_1 x = e_0, \quad T_1 y = e_n.$$

If the first two systems have solutions then (by the more commonly used version of the Gohberg-Semencul formula) T_0 is nonsingular and, as argued above, the top left entry of T_0^{-1} must be nonzero. Otherwise (if one or both of the first two equations do not have a solution), it is guaranteed that T_1 is nonsingular and has an inverse whose top left entry is nonzero. Thus, one can always either set $\hat{T} = T_0$ or $\hat{T} = T_1$ and then apply Theorem 4.1 in order to invert T.

Alternatively, if $|\mathsf{F}| = q \geq 2$, one might simply choose β uniformly from F. Then it suffices to solve two Toeplitz systems, provided that one is willing to accept a probability of $\frac{1}{q}$ of failure.

Acknowledgements

Erich Kaltofen was my host during the visit to Rensselaer when this research was conducted, asked the questions that motivated the work, and made several suggestions that helped along the way.

Mark Giesbrecht informed me of the relationship between the number of matrices similar to a given nonsingular matrix, and the number of nonsingular matrices that commute with it, which is the basis for the proof of Theorem 3.1. Nader Bshouty suggested Lemma 3.5, and its application to the problem of estimating the probability that a randomly chosen matrix over a finite field is nonsingular.

References

[1] Baur, W., and Strassen, V. The complexity of partial derivatives. *Theoretical Computer Science 22* (1982), 317–330.

[2] Brent, R. P., Gustavson, F. G., and Yun, D. Y. Y. Fast solution of Toeplitz systems of equations and computation of Padé approximants. *Journal of Algorithms 1* (1980), 259–295.

[3] Eberly, W. Efficient parallel independent subsets and matrix factorizations. In *Proceedings, 3rd IEEE Symposium on Parallel and Distributed Processing* (Dallas, USA, 1991), pp. 204–211.

[4] Giesbrecht, M. Nearly optimal algorithms for canonical matrix forms. *SIAM Journal of Computing 24* (1995), 948–969.

[5] Gohberg, I. C., and Semencul, A. A. On the inversion of finite Toeplitz matrices and their continuous analogs. *Mat. Issled. 2* (1972), 201–233. In Russian.

[6] Kaltofen, E., and Pan, V. Processor-efficient parallel solution of linear systems over an abstract field. In *Proceedings, 3rd Annual ACM Symposium on Parallel Algorithms and Architectures* (1991), ACM Press, pp. 180–191.

[7] Kaltofen, E., and Pan, V. Processor-efficient parallel solution of linear systems II: The general case. In *Proceedings, 33rd IEEE Symposium on Foundations of Computer Science* (Pittsburgh, USA, 1992), pp. 714–723.

[8] Kaltofen, E., and Pan, V. Parallel solution of Toeplitz and Toeplitz-like linear systems over fields of small positive characteristic. In *Proceedings, PASCO '94: First International Symposium on Parallel Symbolic Computation* (1994), World Scientific Publishing, pp. 225–233.

[9] Pan, V. Y. Complexity of parallel matrix computations. *Theoretical Computer Science 54* (1987), 65–85.

[10] Wiedemann, D. H. Solving sparse linear equations over finite fields. *IEEE Transactions on Information Theory IT-32* (1986), 54–62.

[1] As stated by Brent, Gustavson, and Yun, there appears to be an additional requirement that "$x_0 = y_n$." However, this is guaranteed *for any Toeplitz matrix $\hat{T}$, since the inverse of $\hat{T}$* is always *"persymmetric"*: $(\hat{T}^{-1})^T = J\hat{T}^{-1}J$, if J is the reflection matrix of order $n+1$, with ones on the antidiagonal and zeroes everywhere else.