# On Wiedemann's Method of Solving Sparse Linear Systems*

*Erich Kaltofen*

Department of Computer Science, Rensselaer Polytechnic Institute
Troy, New York 12180-3590; Inter-Net: kaltofen@cs.rpi.edu

*B. David Saunders*

Department of Computer and Information Sciences, University of Delaware
Newark, Delaware 19716; Inter-Net: saunders@dewey.udel.edu

Preliminary Report

## 1. Introduction

Douglas Wiedemann's (1986) landmark approach to solving sparse linear systems over finite fields provides the symbolic counterpart to non-combinatorial numerical methods for solving sparse linear systems, such as the Lanczos or conjugate gradient method (see Golub and van Loan (1983)). The problem is to solve a sparse linear system, when the individual entries lie in a generic field, and the only operations possible are field arithmetic; the solution is to be exact. Such is the situation, for instance, if one works in a finite field. Wiedemann bases his approach on Krylov subspaces, but projects further to a sequence of individual field elements. By making a link to the Berlekamp/Massey problem from coding theory — the coordinate recurrences — and by using randomization an algorithm is obtained with the following property. On input of an $n \times n$ coefficient matrix $A$ given by a so-called black box, which is a program that can multiply the matrix by a vector (see Figure 1), and of a vector $b$, the algorithm finds, with high probability in case the system is solvable, a random solution vector $x$ with $Ax = b$. It is assumed that the field has sufficiently many elements, say no less than $50n^2 \log(n)$, otherwise one goes to a finite algebraic extension. The complexity of the method is in the general singular case $O(n \log(n))$ calls to the black box for $A$ and an additional $O(n^2 \log(n)^2)$ field arithmetic operations.

Note that the black box model for matrix sparsity is a significant abstraction. For a matrix that has an abundance of zero entries, multiplying the matrix by a vector may cost no more than $O(n)$ field operations, in which case the algorithm becomes almost quadratic. However, the model also applies to structured matrices with few or no zero entries, such as Toeplitz- and Vandermonde-like matrices, or matrices that correspond to resultants (Canny

* This material is based on work supported in part by the National Science Foundation under Grant No. CCR-90-06077 and under Grant No. CDA-88-05910 (first author). This paper appears in Proc. Ninth Internat. Symp. Applied Algebra, Algebraic Algor., Error-Correcting Codes, Springer Lect. Notes Comput. Sci. **539**, pp. 29–38 (1991).


$A \in \mathsf{K}^{n \times n}$

$\mathsf{K}$ a field of sufficiently large cardinality

Figure 1: Black box representation of a matrix.

et al. 1989). Most importantly, the algorithm makes no assumptions on the systems, such as symmetricity or positive definiteness.

We contribute to Wiedemann's approach in several ways. For singular systems, we show how to randomly sample from the solution manifold by randomly perturbing the entire system and then solving a non-singular one. Our method is a purely algebraic one, while Wiedemann uses padding with random sparse rows to enforce non-singularity. When computing the rank or determinant of a matrix, one requires left and right multiplier matrices such that the product with the coefficient matrix has a maximal non-zero minor in the left upper corner. We present an alternate to Wiedemann's perturbation that requires asymptotically fewer field elements and is again based on algebraic rather than combinatorical properties. As it turns out, the multipliers can be chosen unit triangular Toeplitz matrices. We also present a new method for finding the rank of a matrix that is asymptotically a factor $\log(n)$ faster than the previous ones. Furthermore, we present in greater detail a method based on p-adic lifting for solving a sparse system over the rationals.

LaMacchia and Odlyzko (1991) have explored the use of conjugate gradients for solving sparse systems over finite fields. While that approach appears, in practice, to be competitive with ones based on coordinate recurrences, the probability of success for their randomizations seems difficult to analyze. Of course, for particularly structured matrices one may also proceed by nested dissection (Lipton et al. 1979) or block elimination (Abdali and Wise 1988) and (Wise and Franco 1990).

In his concluding remarks, Wiedemann raises the question whether it may be possible to solve a transposed problem $x^{\text{tr}}A = b$ from a black box for $A$. We wish to add that if the black box is an algebraic circuit, it is possible to construct a black box for $A^{\text{tr}}$ with the same asymptotic complexity (Kaminski et al. 1988) and (Kaltofen and Pan 1991, §4).

## 2. Wiedemann's Method for Non-Singular Matrices

Wiedemann (1986) presents a randomized Las Vegas algorithm for solving a sparse linear system over a finite field. As it turns out, his method constitutes an algorithm based on field arithmetic alone that can solve a non-singular system given as an $n$-dimensional black box matrix. It requires linear space and quadratic time, while applying the black box for the coefficient matrix no more than $3n$ times. In the following we present Wiedemann's argument with the change in the probabilistic analysis taken from (Kaltofen and Pan 1991), which is warranted because we work over an abstract field.

Let $V$ be a vector space over the field $\mathsf{K}$, and let $\{a_i\}_{i=0}^{\infty}$ be an infinite sequence with elements $a_i \in V$. The sequence $\{a_i\}_{i=0}^{\infty}$ is *linearly generated* over $\mathsf{K}$ if there exist $c_0, c_1, \ldots, c_n \in \mathsf{K}$, $n \geq 0$, $c_k \neq 0$ for some $k$ with $0 \leq k \leq n$, such that

$$\forall j \geq 0\colon c_0 a_j + \cdots + c_n a_{j+n} = 0.$$

The polynomial $c_0 + c_1\lambda + \cdots + c_n\lambda^n$ is called a *generating polynomial* for $\{a_i\}_{i=0}^{\infty}$. The set of all generating polynomials for $\{a_i\}_{i=0}^{\infty}$ together with the zero polynomial forms an ideal in $\mathsf{K}[\lambda]$. The unique polynomial generating that ideal, normalized to have leading coefficient 1, is called the *minimum polynomial* of a linearly generated sequence $\{a_i\}_{i=0}^{\infty}$. Every generating polynomial is a multiple of the minimum polynomial.

Let $A \in \mathsf{K}^{n\times n}$ be a square matrix over a field. The sequence $\{A^i\}_{i=0}^{\infty} \in (\mathsf{K}^{n\times n})^{\mathbb{N}}$ is linearly generated, and its minimum polynomial is the minimum polynomial of $A$, which will be denoted by $f^A$. For any column vector $b \in \mathsf{K}^n$, the sequence $\{A^i b\}_{i=0}^{\infty} \in (\mathsf{K}^n)^{\mathbb{N}}$ is also linearly generated by $f^A$. However, its minimum polynomial, denoted by $f^{A,b}$, can be a proper divisor of $f^A$. For any row vector $u \in \mathsf{K}^{1\times n}$, the sequence $\{uA^i b\}_{i=0}^{\infty} \in \mathsf{K}^{\mathbb{N}}$ is linearly generated as well, and its minimum polynomial, denoted by $f_u^{A,b}$, is again a divisor of $f^{A,b}$. Wiedemann proves the following fact (loc. cit., §VI).

**Theorem 1.** *Let $m = \deg(f^{A,b})$, and let $W$ be the linear space of polynomials of degree less than $m$ in $\mathsf{K}[\lambda]$. There exists a surjective linear map $\ell\colon \mathsf{K}^{1\times n} \longrightarrow W$ such that*

$$\forall u \in \mathsf{K}^{1\times n}\colon f_u^{A,b} = f^{A,b} \iff \mathrm{GCD}(f^{A,b}, \ell(u)) = 1.$$

Thus, the probability that $f_u^{A,b} = f^{A,b}$ for a randomly selected row vector $u$ is essentially the probability of randomly selecting a polynomial of degree less than $n$ that is relatively prime to $f^{A,b}$. For a finite field with $q$ elements, Wiedemann (loc. cit., Proposition 3) proves that the probability is no less than

$$\frac{1}{6\max\{\lceil \log_q(\deg\ f^A)\rceil, 1\}}. \tag{1}$$

In (Kaltofen and Pan 1991, §2) we establish the following alternate approach.

**Lemma 1.** *Let $A \in \mathsf{K}^{n\times n}$, $b \in \mathsf{K}^n$, and let $S \subset \mathsf{K}$. Randomly and uniformly select a row vector $u \in S^{1\times n}$. Then the probability*

$$\mathrm{Prob}(f_u^{A,b} = f^{A,b}) \geq 1 - \frac{\deg(f^{A,b})}{\mathrm{card}(S)}.$$

If $A$ is non-singular, we may compute $x = A^{-1}b$ from

$$f^{A,b}(\lambda) =: c_0 + c_1\lambda + \cdots + c_{m-1}\lambda^{m-1} + \lambda^m$$

by $m-2$ applications of $A$ as

$$x \leftarrow -\frac{1}{c_0}(A^{m-1}b + c_{m-1}A^{m-2}b \cdots + c_1 b), \tag{2}$$

since $f^{A,b}(A)b = 0$. The polynomial $f^{A,b}$ is computed by picking a random row vector $u$ and computing $f_u^{A,b}$. That is accomplished by first construction the sequence of field elements

$$\{a_0, a_1, \ldots, a_{2m-1}\}, \quad a_i := uA^i b.$$

and finding its minimum degree linear generating polynomial. By the theory of linearly generated sequences, this polynomial is equal to $f_u^{A,b}$, and it can be determined by the Berlekamp/Massey algorithm in $O(m \deg(f_u^{A,b}))$ field operations. Wiedemann shows further that $f_u^{A,b}$ for an unlucky choice of $u$ can be used with the next trial.

**Algorithm** *Minimum Polynomial*

*Input:* $A \in \mathsf{K}^{n\times n}$, $b \in \mathsf{K}^n$, and $d \geq \deg(f^{A,b})$.
*Output:* $f^{A,b} \in \mathsf{K}[\lambda]$.

**Step 1:** *Pick a random row vector* $u \in S^{1\times n}$, $S \subset \mathsf{K}$, *and compute*

$$a_0 \leftarrow ub, a_1 \leftarrow uAb, \ldots, a_i \leftarrow uA^ib, \ldots, a_{2d-1} \leftarrow uA^{2d-1}b.$$

**Step 2:** Here we determine $f_u^{A,b}$ by the Berlekamp/Massey algorithm (Massey 1969). For completeness, we give the entire method.
$\Lambda_0(\lambda) \leftarrow 1; \Sigma_0(\lambda) \leftarrow 0; l_0 \leftarrow 0; \delta \leftarrow 1;$
*For* $r = 1, 2, \ldots, 2d$ *Do* {
  *With* $\Lambda_r(\lambda) = c_0\lambda^{n_r} + c_1\lambda^{n_r-1} + \cdots + c_{n_r}$, $c_0 \neq 0$, *find the* $r$*-th discrepancy*
  $\delta_r \leftarrow c_{n_r}a_{r-1} + c_{n_r-1}a_{r-2} + \cdots + c_0a_{r-n_r-1};$
  *If* $\delta_r = 0$ *Then* {$\Lambda_r(\lambda) \leftarrow \Lambda_{r-1}(\lambda)$; $\Sigma_r(\lambda) \leftarrow \lambda\Sigma_{r-1}(\lambda)$; $l_r \leftarrow l_{r-1}$; }
    *Else* {$\Lambda_r(\lambda) \leftarrow \Lambda_{r-1}(\lambda) - \frac{\delta_r}{\delta}\lambda\Sigma_{r-1}(\lambda)$;
      *If* $2l_r < r$ *Then* {$\Sigma_r(\lambda) \leftarrow \Lambda_{r-1}(\lambda)$; $l_r \leftarrow r - l_{r-1}$; $\delta \leftarrow \delta_r$; }
        *Else* {$\Sigma_r(\lambda) \leftarrow \lambda\Sigma_{r-1}(\lambda)$; $l_r \leftarrow l_{r-1}$; }}}
$f_u^{A,b}(\lambda) \leftarrow \lambda^{l_{2d}}\Lambda_{2d}(1/\lambda).$

**Step 3:** Now we check if $f_u^{A,b}$ is a proper divisor of $f^{A,b}$.
*If* $d = \deg(f_u^{A,b})$ *Then Return* $f^{A,b} \leftarrow f_u^{A,b}$;
*Else* {$b' \leftarrow f_u^{A,b}(A)b$. Clearly, $b'$ can be determined by $\deg(f_u^{A,b}) - 1$ multiplications of $A$ by vectors, or from the vectors $A^ib$ if they have been saved in Step 1.
  *If* $b' = 0$ *Then Return* $f^{A,b} \leftarrow f_u^{A,b}$;
  *Else* {*Call the algorithm recursively with* $A$, $b'$ *and* $d - \deg(f_u^{A,b})$ *to determine* $f^{A,b'}$.
    *Finally, Return* $f^{A,b} \leftarrow f_u^{A,b} \times f^{A,b'}$.}} □

Several observations can be made about this algorithm. First, for randomly chosen $b$, the probability that $f^{A,b} = f^A$, the minimum polynomial of $A$, can be also shown to be bounded by (1) and as in Lemma 1. That observation gives a Las Vegas randomized algorithm to determine that a matrix is singular by establishing that $f^{A,b}(0) = 0$ for a random $b$. For non-singular matrices, the determinant can be found with a Las Vegas randomized method as well (Wiedemann, loc. cit., §V), but we will not need that algorithm in this paper. Second, the probability that the algorithm determines $f^{A,b}$ after $k$ invocations is much higher than if one were to try to obtain $f_u^{A,b}(A)b = 0$ for one of $k$ different $u$'s. In fact, Wiedemann (loc. cit., Eq. (12)) proves that for $k \geq 2$ and $\mathsf{K}$ a field with $q$ elements, the probability is no less than

$$1 - \log\left(\frac{q^{k-1}}{q^{k-1}-1}\right) \geq 1 - \frac{1}{q^{k-1}-1}.$$

Third, the algorithm can be implemented storing only $O(n)$ many field elements, using, for a sufficiently large field, with high probability no more than $3n$ multiplications of $A$ by a vector, and $O(n^2)$ additional arithmetic operations in $\mathsf{K}$. Clearly, from (2) and Step 3, the same complexity bounds hold for computing $x$ with $Ax = b$, provided $f^{A,b}(0) \neq 0$.

## 3. The Rational Non-Singular Case

Although the algorithm presented in §2 for abstract fields is applicable to the rational numbers, the bit size of the intermediately computed rational numbers requires analysis. Alternately, one can lift a modular solution $p$-adically and then convert to a rational one. Here we shall discuss that method further (cf. Wiedemann, loc. cit., §7). We shall assume that the system to be solved is square, non-singular, with integer entries:

$$Ax = b, \quad A \in \mathbb{Z}^{n\times n},\ b \in \mathbb{Z}^n.$$

Furthermore, we suppose that the black box for $A$ can be supplied with both a modulus $q \in \mathbb{Z}$ and a vector $y \in \mathbb{Z}/(q)$, and then computes $(A \bmod q)\, y \in (\mathbb{Z}/(q))^n$.

**Algorithm** *Rational Non-Singular System Solver*

**Step 1:** *Find a prime $p$ that does not divide* $\mathrm{Det}(A)$ *by probabilistically testing if* $\mathrm{Det}(A \bmod p) \neq 0$ *using the method described in §2.* As a by-product, we will have a polynomial $f_u^{A \bmod p, v} \in \mathbb{Z}/(p)[\lambda]$. *Initialize the estimate for* $f^{A \bmod p}$, $f(\lambda) \leftarrow f_u^{A \bmod p, v}(\lambda)$.

**Step 2:** Now we determine how far the $p$-adic solution must be lifted in order to recover the rational solution from the $p$-adic one. Let

$$\|A\|_2 := \max_{1\le j\le n} \left\{ \sqrt{A[1,j]^2 + \cdots + A[n,j]^2} \right\}, \quad \|b\|_2 := \sqrt{b_1^2 + \cdots + b_n^2},$$

where $A[i,j]$ denotes the entry in row $i$ and column $j$ of the matrix $A$. By Hadamard's determinant inequality, $|\mathrm{Det}(A)| \le \|A\|_2^n =: B_1$, and by Cramer's rule the numerator of $(A^{-1}b)_j$ is bounded by $\|A\|_2^{n-1}\|b\|_2 =: B_2$. By the well-known continued fraction recovery procedure (see, e.g., Kaltofen and Rolletschek (1989, §5)) the necessary modulus is twice the product of the numerator and denominator bound, hence

$$p^k \ge 2\|A\|_2^{2n-1}\|b\|_2 =: B_0.$$

*For* $i \leftarrow 0, \ldots, k$ *Do Steps 3 and 4.*

**Step 3:** The $p$-adic expansions of $A$, $b$, and $x$, are denoted by

$$\begin{aligned} A &\equiv \bar{A}^{(i-1)} + p^i A^{(i)} \pmod{p^{i+1}}, & \bar{A}^{(i-1)} &\in (\mathbb{Z}/(p^i))^{n\times n}, & A^{(i)} &\in (\mathbb{Z}/(p))^{n\times n},\\ b &\equiv \bar{b}^{(i-1)} + p^i b^{(i)} \pmod{p^{i+1}}, & \bar{b}^{(i-1)} &\in (\mathbb{Z}/(p^i))^{n}, & b^{(i)} &\in (\mathbb{Z}/(p))^{n},\\ x &\equiv \bar{x}^{(i-1)} + p^i x^{(i)} \pmod{p^{i+1}}, & \bar{x}^{(i-1)} &\in (\mathbb{Z}/(p^i))^{n}, & x^{(i)} &\in (\mathbb{Z}/(p))^{n}. \end{aligned}$$

At this point we have $\bar{x}^{(i-1)}$ and we find $x^{(i)}$ from

$$A(\bar{x}^{(i-1)} + p^i x^{(i)}) \equiv b \pmod{p^{i+1}}.$$

*Compute over the integers*

$$\widehat{b}^{(i)} \leftarrow \frac{\bar{b}^{(i)} - \bar{A}^{(i)}\bar{x}^{(i-1)}}{p^i} \in \mathbb{Z}^n,$$

*and set* $\widetilde{b}^{(i)} \leftarrow \widehat{b}^{(i)} \bmod p$.

**Step 4:** We have $x^{(i)} = (A \bmod p)^{-1}\widetilde{b}^{(i)}$ over $\mathbb{Z}/(p)$. The mod $p$ solution is determined from the current estimate $f$ for $f^{A \bmod p}$, which is normalized as

$$f(\lambda) =: 1 - c_1\lambda - \cdots - c_{m-1}\lambda^{m-1} - c_m\lambda^m \in \mathbb{Z}/(p)[\lambda], \quad c_m \not\equiv 0 \pmod p.$$

*Compute*

$$x^{(i)} \leftarrow c_1\widetilde{b}^{(i)} + c_2(A \bmod p)\,\widetilde{b}^{(i)} + \cdots + c_m(A \bmod p)^m\,\widetilde{b}^{(i)}.$$

*If* $b'^{(i)} \leftarrow (A \bmod p)x^{(i)} \neq \widetilde{b}^{(i)}$, *call algorithm Minimum Polynomial with* $A \bmod p$, $b'^{(i)}$, $n-m$, *over the field* $\mathsf{K} = \mathbb{Z}/(p)$. *Set* $f(\lambda) \leftarrow f^{A \bmod p, b'^{(i)}}(\lambda)f(\lambda)$ *and repeat Step 4, unless* $f(0) = 0$, *in which case the prime* $p$ *divides the determinant of* $A$ *and must be changed.*

**Step 5:** We now convert $\bar{x}^{(k)} \in (\mathbb{Z}/(p^k))^n$ to a rational vector $x$. Since the least common denominator of all components $x_j$ is a divisor of $\mathrm{Det}(A)$, we may incrementally find that denominator from the denominators of initially converted $x_j$'s. *Set* $\Delta \leftarrow 1$.

*For* $j \leftarrow 1, \ldots, n$ *Do* {

First, we divide out the current common denominator and adjust the modulus bound. *Determine the smallest* $k'$ *such that* $p^{k'} \geq B_0/\Delta$ *(see Step 2). Set* $\bar{x}_j^{(k')} \leftarrow \Delta^{-1}\bar{x}_j^{(k)} \pmod{p^{k'}}$. *If* $\bar{x}_j^{(k')}$, *represented as an integer between* $-p^{k'}/2$ *and* $p^{k'}/2$, *is in absolute value no larger than* $B_2$, *then* $x_j \leftarrow \bar{x}_j^{(k')}/\Delta \in \mathbb{Q}$. *Otherwise, compute that convergent* $u_l/v_l$ *of the continued fraction approximations (Hardy and Wright 1979, §10) of* $\bar{x}_j^{(k')}/p^{k'}$ *that satisfies* $v_l \leq B_1/\Delta < v_{l-1}$; $x_j \leftarrow (\bar{x}_j^{(k')}v_l - p^{k'}u_l)/v_l$; $\Delta \leftarrow \Delta v_l$.} □

The advantage of this algorithm lies in the fact that one only needs the minimum polynomial of $A \bmod p$. The number of applications of the black box for $A$ now depends also on the length of the entries in $A$ and $b$, but those multiplications are modular ones and therefore computationally more efficient than the ones arising in the method of §2.

## 4. The Singular Case

The problem at hand is to solve $Ax = b$ in case where the matrix $A$ is singular. We give a randomized algorithm that returns a random vector in the solution manifold, provided one exists. First, we give a perturbation scheme that makes the leading principal submatrices of dimension up to the rank of $A$ non-singular.

**Theorem 2.** *Let* $A \in \mathsf{K}^{n\times n}$, *and let* $S \subset \mathsf{K}$. *Consider the matrix*

$$\widetilde{A} := U A L, \quad U := \begin{pmatrix} 1 & u_2 & u_3 & \ldots & u_n \\ & 1 & u_2 & \ldots & u_{n-1} \\ & & 1 & \ddots & \vdots \\ & & & \ddots & u_2 \\ & & & & 1 \end{pmatrix}, \quad L := \begin{pmatrix} 1 & & & & \\ w_2 & 1 & & & \\ w_3 & w_2 & 1 & & \\ \vdots & & \ddots & \ddots & \\ w_n & w_{n-1} & \ldots & w_2 & 1 \end{pmatrix},$$

*where the elements of the unit upper triangular Toeplitz matrix* $U$ *and the elements of the unit lower triangular Toeplitz matrix* $L$ *are randomly and uniformly selected from the set* $S$. *Let* $r$ *be the rank of* $A$, *and let* $\widetilde{A}_i$ *denote the leading principal* $i \times i$ *submatrix of* $\widetilde{A}$. *Then*

$$\mathrm{Prob}(\mathrm{Det}(\widetilde{A}_i) \neq 0 \text{ for all } 1 \leq i \leq r) \geq 1 - \frac{r(r+1)}{\mathrm{card}(S)}.$$

*Proof.* For an $n \times n$ matrix $B$, denote by $B_{I,J}$ the determinant of the submatrix of $B$ that is formed by removing from $B$ all rows not contained in the set $I$ and all columns not contained in the set $J$. First, assume that $\mathcal{U}$ is a generic unit upper triangular Toeplitz matrix whose entries are new variables $\upsilon_2, \ldots, \upsilon_n$ replacing $u_2, \ldots, u_n$, and assume that $\mathcal{L}$ is a generic unit lower triangular Toeplitz matrix whose entries are new variables $\omega_2, \ldots, \omega_n$. Let $\widetilde{\mathcal{A}} = \mathcal{U} A \mathcal{L} \in \mathsf{L}^{n\times n}$, where $\mathsf{L} := \mathsf{K}(\upsilon_2, \ldots, \omega_n)$. For $K = \{1, \ldots, k\}$ the Cauchy-Binet formula yields

$$\widetilde{\mathcal{A}}_{K,K} = \sum_{\substack{I=\{i_1,\ldots,i_k\}\\ 1\le i_1<\cdots<i_k\le n}} \sum_{\substack{J=\{j_1,\ldots,j_k\}\\ 1\le j_1<\cdots<j_k\le n}} \mathcal{U}_{K,I} A_{I,J} \mathcal{L}_{J,K}. \tag{3}$$

We claim that for $k \le r$ the determinant $\widetilde{\mathcal{A}}_{K,K}$, which is the $k$th principal minor of $\widetilde{\mathcal{A}}$, is non-zero in $\mathsf{L}$. To prove this we consider the minor expansions of $\mathcal{U}_{K,I}$ and $\mathcal{L}_{J,K}$. For a given set $I = \{i_1, \ldots, i_k\}$ with $1 \le i_1 < \cdots < i_k \le n$, consider all terms in the minor expansion of $\mathcal{U}_{K,I}$. The matrix $\mathcal{U}$ restricted to rows $1, \ldots, k$ and columns in $I$ has the form, for instance,

$$\begin{array}{c} \\ 1 \\ 2 \\ \\ \\ k \end{array}\begin{array}{c} \begin{array}{cccc} i_1 & i_2 & & i_k \end{array} \\ \begin{pmatrix} \upsilon_{i_1} & \upsilon_{i_2} & \ldots & \upsilon_{i_k} \\ \upsilon_{i_1-1} & \upsilon_{i_2-1} & & \upsilon_{i_k-1} \\ \vdots & \vdots & & \vdots \\ 1 & & & \\ 0 & 1 & \ldots & \upsilon_{i_k-k+1} \end{pmatrix} \end{array}.$$

If we write the terms in the minor expansion in *descending* order of the variables, using the variable order $\upsilon_2 < \upsilon_3 < \cdots < \upsilon_n$, then the diagonal term,

$$\upsilon_{i_k-k+1}\upsilon_{i_{k-1}-k+2}\cdots\upsilon_{i_1}, \quad \text{with } \upsilon_1 = 1, \tag{4}$$

is the lexicographically *smallest* term of all non-zero monomials in that minor expansion. Moreover, this term uniquely identifies each minor of $\mathcal{U}$ under the sum (3). The latter is most easily seen from the fact that the set $I$ can be reconstructed from (4) by observing that $\upsilon_{i_\kappa-\kappa+1} = 1$ forces $i_\kappa = \kappa$ and all lower indices $i_\mu = \mu$, $\mu < \kappa$. Similarly, the diagonal terms in the minor expansions of $\mathcal{L}_{J,K}$,

$$\omega_{j_k-k+1}\omega_{j_{k-1}-k+2}\cdots\omega_{j_1}, \quad \text{with } \omega_1 = 1, \tag{5}$$

are the lexicographically smallest among all terms of non-zero monomials in the expansions, and uniquely correspond to an index set $J$. Therefore, the polynomials $\mathcal{U}_{K,I}\mathcal{L}_{J,K} \in \mathsf{K}[\upsilon_2, \ldots, \omega_n]$ have unique lexicographically lowest terms, namely the product of (4) and (5) with the prescribed variable ordering, hence are linearly independent over $\mathsf{K}$. Moreover, since for $k \le r$ there exist sets $I_0$ and $J_0$ with $A_{I_0,J_0} \neq 0$, the linear sum (3) of the linearly independent polynomials cannot be zero, which establishes the claim.

Set

$$0 \neq \sigma(\mathcal{U}, \mathcal{L}) := \prod_{k=1}^{r} \widetilde{\mathcal{A}}_{K,K} \in \mathsf{K}[\upsilon_2, \ldots, \omega_n].$$

It is clear that all those $U$ and $L$, for which $\sigma(U, L) \neq 0$, satisfy the lemma. By the Schwartz (1980) / Zippel (1979) lemma, that probability is no less than $1 - \deg(\sigma)/\mathrm{card}(S)$. The probability estimate follows from $\deg(\sigma) \le \sum_{k=1}^{r} 2k$. ⊠

We remark that applying $\widetilde{A}$ to a vector costs one application of $A$ to a vector and two polynomial multiplications, since the Toeplitz matrix times vector products can be accomplished by polynomial multiplication. For example, for

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} := \begin{pmatrix} w_1 & & & & \\ w_2 & w_1 & & & \\ w_3 & w_2 & w_1 & & \\ \vdots & & \ddots & \ddots & \\ w_n & w_{n-1} & \dots & w_2 & w_1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{pmatrix}$$

we have

$$(w_1 + \cdots + w_n z^{n-1})(v_1 + \cdots + v_n z^{n-1}) \equiv y_1 + y_2 z + \cdots + y_n z^{n-1} \pmod{z^n}.$$

Thus applying $\widetilde{A}$ to a vector consumes an additional $O(n \log(n) \operatorname{loglog}(n))$ arithmetic operations in $\mathsf{K}$ (Cantor and Kaltofen 1987). Note that Wiedemann (1986, §V) proposes a different perturbation scheme with the same effect, which is based on rearrangable permutation networks (Beneš 1964). That scheme requires $O(n \log(n))$ random field elements, but only costs an additional $O(n \log(n))$ arithmetic operations.

We may determine the rank of $A$ by performing a binary search for the largest nonsingular leading principal submatrix of $\widetilde{A}$. However, that strategy adds a $\log(n)$ factor to all timings, and we have found the following alternate way to determine the rank without that problem.

**Lemma 2.** *Let $A \in \mathsf{K}^{n \times n}$ have leading principal minors nonzero up to $A_r$, where $r$ is the (unknown) rank of $A$, and suppose that $r < n$. Let $S \subset \mathsf{K}$ and let $X = \operatorname{diag}(x_1, \ldots, x_n)$, $x_i$ chosen uniformly from $S$. Then $r = \deg(f^{AX}) - 1$ with probability at least*

$$1 - n(n-1)/(2 \operatorname{card}(S)).$$

*Proof.* Consider the conformal partitioning

$$AX = \begin{pmatrix} A_r & B \\ C & D \end{pmatrix} \begin{pmatrix} Y & 0 \\ 0 & Z \end{pmatrix} = \begin{pmatrix} A_r Y & BZ \\ CY & DZ \end{pmatrix}.$$

Provided the $x_i$'s are nonzero, this matrix has the same rank as $A$ and also has leading principal minors nonzero up to the $r$-th. Now consider the following matrix similar to $AX$,

$$\begin{aligned} M &:= \begin{pmatrix} I & Y^{-1} A_r^{-1} BZ \\ 0 & I \end{pmatrix} \begin{pmatrix} A_r Y & BZ \\ CY & DZ \end{pmatrix} \begin{pmatrix} I & -Y^{-1} A_r^{-1} BZ \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} A_r Y + Y^{-1} A_r^{-1} BZCY & 0 \\ CY & 0 \end{pmatrix}. \end{aligned}$$

Of course this matrix has the same minimal polynomial and characteristic polynomial as $AX$. Let $c^M(\lambda)$ denote the characteristic polynomial of matrix $M$. We have $c^{AX}(\lambda) = \lambda^{n-r} c^{A'Y}(\lambda)$, where $A'Y$ is the upper left corner, i.e., $A'$ denotes $A_r + Y^{-1} A_r^{-1} BZC$. If $A'$ has leading principal minors nonsingular, then Wiedemann's lemma applies (loc. cit., Section V), and

$f^{A'Y}(\lambda) = c^{A'Y}(\lambda)$. It follows that $f^{AX}(\lambda) = f^M(\lambda) = \lambda c^{A'Y}(\lambda)$ is a polynomial of degree $r+1$. To see this, note that on the one hand the right hand polynomial is clearly a factor of $f^M(\lambda)$, while on the other hand there is a relation among $M^i$, $i = 1, \ldots, r+1$, given the relation among $(A'Y)^i$, $i = 1, \ldots, r$ and that

$$M^i = \begin{pmatrix} (A'Y)^i & 0 \\ CY(A'Y)^{i-1} & 0 \end{pmatrix}.$$

It remains to show $A'$ has leading principal minors nonsingular. This is so if $x_{r+1}, \ldots, x_n$ (the entries of $Z$) are indeterminates. For then the $i$th leading principal minor of $A'$ is a polynomial of degree no more than $i$ in these variables with a constant term which is the $i$th leading principal minor of $A_r$. The product of these minors is a polynomial of degree $r(r+1)/2$, hence if the entries of $Z$ are randomly chosen, the probability of a leading principal minor being zero is, by the Schwartz/Zippel lemma, bounded by $r(r+1)/(2\,\mathrm{card}(S))$. ⊠

We may compute $f^{\widetilde{A}X} = f_u^{\widetilde{A}X,b}$, for random $u, b$ as in §2. Thus we have the following result.

**Theorem 3.** *Let $A \in \mathsf{K}^{n\times n}$, and let $S \subset \mathsf{K}$. Using $5n-2$ random elements from $S$ (the entries of $U$, $L$, $X$, $u$, and $b$), we may probabilistically determine the rank of $A$ by $O(n)$ multiplications of $A$ by vectors and $O(n^2 \log(n) \operatorname{loglog}(n))$ arithmetic operations in $\mathsf{K}$. The algorithm returns an integer that is with probability no less than*

$$1 - \frac{3}{2}\,\frac{n(n+1)}{\mathrm{card}(S)}$$

*the rank of $A$.*

Once we have determined the rank of $A$, it is relatively easy to compute a random solution to $Ax = b$. Clearly, it suffices to compute a random solution to $\widetilde{A}\,\widetilde{x} = Ub$, since then $x = L\widetilde{x}$ solves $Ax = b$. Hence we may restrict ourselves to the case where the coefficient matrix has the properties of $\widetilde{A}$, namely $A = \begin{pmatrix} A_r & B \\ C & D \end{pmatrix}$ where $r$ is the rank of $A$ and $A_r$ is nonsingular. The equivalent matrix

$$\begin{pmatrix} A_r & B \\ C & D \end{pmatrix}\begin{pmatrix} I & -A_r^{-1}B \\ 0 & I \end{pmatrix} = \begin{pmatrix} A_r & 0 \\ C & D - CA_r^{-1}B \end{pmatrix}$$

has the same rank, hence $D = CA_r^{-1}B$.

Now for any $x_2$,

$$\begin{pmatrix} A_r & B \\ C & D \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} A_r x_1 + Bx_2 \\ Cx_1 + CA_r^{-1}Bx_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

if and only if $A_r x_1 = b_1 - Bx_2$ and $b_2 = CA_r^{-1}b_1$. The latter condition, independent of $x_2$ is a necessary and sufficient condition for the existence of a solution to $Ax = b$ and the first equation describes the solution space. For the case $x_2 = 0$ we solve for $x_1$ as described above. Finally, we show how to find a random element of the solution manifold.

**Lemma 3.** *Let $A \in \mathsf{K}^{n\times n}$ be of rank $r$, and suppose that $A_r$, the $r \times r$ leading principal submatrix of $A$, is non-singular. For any column vector $w \in \mathsf{K}^n$, there exists a unique vector $y_w \in \mathsf{K}^n$ such that*

$$A \times \underbrace{\begin{pmatrix} y'_w \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{=:\, y_w} \left.\vphantom{\begin{matrix}0\\ \vdots\\ 0\end{matrix}}\right\} n - r = Aw, \quad y'_w \in \mathsf{K}^r.$$

*Furthermore the map $\ell\colon \mathsf{K}^n \longrightarrow \mathsf{K}^n$ defined by $\ell(w) := w - y_w$ is linear with range the right null space of $A$.*

*Proof.* The existence of $y_w$ with the bottom $n - r$ entries equal to 0 easily follows from the fact that $Aw$ is a linear combination of the column vectors of $A$, which by the assumption on $A$ can be also expressed as a linear combination of the first $r$ column vectors. Since $y'_w = A_r^{-1} A' w$, where $A' \in \mathsf{K}^{r\times n}$ is the matrix formed by the first $r$ rows of $A$, $\ell$ is a linear map. Clearly, $A \times \ell(w) = 0$, so range$(\ell)$ is a subspace of the right nullspace of $A$. Now $\ell(w) = 0$ iff $w - y_w = 0$, which means that

$$w = \begin{pmatrix} w' \\ 0 \\ \vdots \\ 0 \end{pmatrix} \left.\vphantom{\begin{matrix}0\\ \vdots\\ 0\end{matrix}}\right\} n - r \quad \text{and thus } y'_w = w'.$$

Hence the kernel of the map $\ell$ has dimension $r$, which implies that range$(\ell)$ has the dimension $n - r$ and is therefore the full right nullspace of $A$. ⊠

**Theorem 4.** *Let $A \in \mathsf{K}^{n\times n}$ be of rank $r$ with the leading principal $r \times r$ submatrix non-singular, and let $b \in \mathsf{K}^n$ be such that $Ax = b$ is solvable in $x \in \mathsf{K}^n$. Then for a random column vector $v \in \mathsf{K}^n$ there exists a unique vector $y_{b,v} \in \mathsf{K}^n$ such that*

$$A \times \underbrace{\begin{pmatrix} y'_{b,v} \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{=:\, y_{b,v}} \left.\vphantom{\begin{matrix}0\\ \vdots\\ 0\end{matrix}}\right\} n - r = b + Av.$$

*Furthermore, $y_{b,v} - v$ uniformly samples the solution manifold of $Ax = b$.*

*Proof.* The existence of the special vector $y_{b,v}$ follows as in Lemma 3. Let $x_0 \in \mathsf{K}^n$ be a solution $Ax_0 = b$ and let $w := v + x_0$. Then $Ay_{b,v} = Aw$, which by Lemma 2 means that $y_{b,v} - w$ samples the right nullspace of $A$. Thus $y_{b,v} - v$ samples the solution manifold of $Ax = b$. ⊠

Therefore, once we know the rank $r$ of $A$, a random solution can be obtained with an additional $n$ random field elements. Note that $y'_{b,v}$ can be obtained from a nonsingular subsystem, which has a black box coefficient matrix by setting the bottom $n-r$ components of the input vector to full black box matrix to 0. Taking the randomization of Theorem 2 into account, we need $O(r)$ applications of $A$ and an additional $O(rn \log(n) \operatorname{loglog}(n))$ arithmetic operations.

## Literature Cited


Abdali, S. K. and Wise, D. S., “Experiments with quadtree representation of matrices,” *Proc. ISSAC ’88, Springer Lect. Notes Comput. Sci.* **358**, pp. 467–474 (1988).

Beneš, V. E., “Optimal rearrangeable multistage connecting networks,” *Bell System Tech. J.* **43**, pp. 1641–1656 (1964).

Canny, J., Kaltofen, E., and Lakshman Yagati, “Solving systems of non-linear polynomial equations faster,” *Proc. ACM-SIGSAM 1989 Internat. Symp. Symbolic Algebraic Comput.*, pp. 121–128 (1989).

Cantor, D. G. and Kaltofen, E., “On fast multiplication of polynomials over arbitrary rings,” *Acta Inform.* **28**/7, pp. 693–701 (1991).

Golub, G. H. and van Loan, C. F., *Matrix Computations*; Johns Hopkins University Press, Baltimore, Maryland, 1987.

Hardy, G. H. and Wright, E. M., *An Introduction to the Theory of Numbers*; Oxford Univ. Press, Oxford, 1979.

Kaltofen, E. and Pan, V., “Processor efficient parallel solution of linear systems over an abstract field,” in *Proc. 3rd Ann. ACM Symp. Parallel Algor. Architecture*; ACM Press, pp. 180–191, 1991.

Kaltofen, E. and Rolletschek, H., “Computing greatest common divisors and factorizations in quadratic number fields,” *Math. Comp.* **53**/188, pp. 697–720 (1989).

Kaminski, M., Kirkpatrick, D. G., and Bshouty, N. H., “Addition requirements for matrix and transposed matrix products,” *J. Algorithms* **9**, pp. 354–364 (1988).

LaMacchia, B. A. and Odlyzko, A. M., “Solving large sparse linear systems over finite fields,” in *Advances in Cryptology: CRYPTO ’90*, Lect. Notes Comput. Sci. **537**, edited by A. J. Menezes and S. Vanstone; Springer Verlag, 1991.

Lipton, R., Rose, D., and Tarjan, R. E., “Generalized nested dissection,” *SIAM J. Numer. Anal.* **16**, pp. 346–358 (1979).

Massey, J. L., “Shift-register synthesis and BCH decoding,” *IEEE Trans. Inf. Theory* **IT-15**, pp. 122–127 (1969).

Schwartz, J. T., “Fast probabilistic algorithms for verification of polynomial identities,” *J. ACM* **27**, pp. 701–717 (1980).

Wiedemann, D., “Solving sparse linear equations over finite fields,” *IEEE Trans. Inf. Theory* **IT-32**, pp. 54–62 (1986).

Wise, D. S. and Franco, J., “Costs of quadtree representation of non-dense matrices,” *J. Parallel Distributed Comput.* **9**, pp. 282–296 (1990).

Zippel, R., “Probabilistic algorithms for sparse polynomials,” *Proc. EUROSAM ’79, Springer Lec. Notes Comp. Sci.* **72**, pp. 216–226 (1979).