Early Termination over Small Fields

Extended Abstract

Wayne Eberly *
Department of Computer Science
University of Calgary
Calgary, Alberta, Canada
eberly@cpsc.ucalgary.ca

ABSTRACT

Krylov-based algorithms have recently been used (alone, or in combination with other methods) in order to solve systems of linear equations that arise during integer factorization and discrete logarithm computations. Since these include systems over small finite fields, the behaviour of these algorithms in this setting is of interest.

Unfortunately, the application of these methods is complicated by the possibility of several kinds of breakdown. Orthogonal vectors can arise when a variant of the Lanczos algorithm is used to generate a basis, and zero-discrepancies can arise during the computation of minimal polynomials of linearly recurrent sequences when Wiedemann's algorithm is applied.

Several years ago, Austin Lobo reported experimental evidence that zero-discrepancies are extremely unlikely when a randomized version of Wiedemann's algorithm is applied to solve systems over large fields. With high probability, results are correct if a computation is terminated as soon as such a sequence is detected. "Early termination" has consequently been included in recent implementations.

In this paper, we analyze the probability of long sequences of zero-discrepancies during computations of minimal polynomials of the linearly recurrent sequences that arise when simple Krylov-based algorithms are used to solve systems over very small finite fields. Variations of these algorithms that incorporate early termination are briefly presented and analyzed in the small field case.

Categories and Subject Descriptors

I.1.2 [**Symbolic and Algebraic Manipulation**]: Algorithms—*algebraic algorithms, analysis of algorithms*; F.2.1 [**Analysis of Algorithms and Problem Complexity**]: Numerical Algorithms and Problems—*computations in finite fields, computations on matrices*

General Terms

Algorithms, Performance, Reliability, Theory

Keywords

Berlekamp-Massey algorithm, black box matrix, early termination, finite field, Lanczos algorithm, linear system solution, randomized algorithm

1. INTRODUCTION

Consider the problem of solving a system of linear equations

$$Ax = b$$

where $A \in \mathsf{F}^{N \times N}$ is an $N \times N$ matrix and $b \in \mathsf{F}^{N \times 1}$ is a vector with dimension N over the finite field $\mathsf{F} = \mathsf{F}_q$ with q elements. A related problem which is also of interest is the computation of an element of the nullspace of such a matrix A. Indeed, instances of these problems are formed and solved in modern algorithms for integer factorization and discrete logarithm computations. In particular, the latter problem arises with $\mathsf{F} = \mathsf{F}_2$ when the number field sieve is applied ([2]), while computations over F_q arise for larger q during discrete logarithm computations.

Several different "Krylov-based" methods for these problems have been proposed, implemented and analyzed in recent years. In contrast with elimination-based methods, these do not manipulate the entries of the coefficient matrix A. Instead, these algorithms work over the vector subspace generated by the vectors

$$b, Ab, A^2b, A^3b, \ldots$$

for some vector b.

In particular, one version of Wiedemann's algorithm [19] considers the linearly recurrent sequences

$$c_0, c_1, c_2, \ldots, \qquad \text{where } c_i = u^T A^i b$$

that are formed using uniformly and independently selected vectors $u \in \mathsf{F}^{n \times 1}$. The minimal polynomials of these sequences are computed using the Berlekamp-Massey algorithm [1], [14], and these polynomials are combined to obtain the minimal polynomial of the matrix A and vector b — that is, the monic polynomial $f \in \mathsf{F}[x]$ with least degree

*Research was supported in part by Natural Sciences and Engineering Research Council of Canada research grant OGP0089756.

such that $f(A)b = 0$. If A is nonsingular or, more generally, if $f(0) \neq 0$, then a solution for the given system is easily recovered after that.

Similarly, a version of the Lanczos algorithm [11] works with one or more uniformly and independently selected vectors u as well. In this case, an orthogonalization process is used to try to construct dual orthogonal bases for the pair of subspaces that are generated by the vectors

$$b, Ab, A^2b, A^3b, \ldots$$

and

$$u, A^Tu, (A^T)^2u, (A^T)^3u, \ldots .$$

While these algorithms are not identical, they are closely related; Lambert [10] provides a unified treatment of these and several other variants.

Unfortunately, these computations are complicated by the possibility of various kinds of breakdown. A long sequence of zero-discrepancies might arise when the Berlekamp-Massey algorithm is applied during an execution of Wiedemann's algorithm, while one might obtain a long sequence of orthogonal vectors when the Lanczos algorithm is applied.

Several years ago, Austin Lobo [12] reported experimental evidence that zero-discrepancies are extremely unlikely when the Berlekamp-Massey algorithm is used in an application of Wiedemann's algorithm to solve a system of linear equations over a large field. If "random" field elements are chosen uniformly and independently from a sufficiently large subset of the ground field, and the computation is terminated as soon as a short sequence of zero-discrepancies has been encountered, then the probability that the resulting values are correct appears to be high. Lobo reported that a window of twenty zero-discrepancies was a good early-termination threshold, and he has subsequently conjectured that it is also safe to terminate the algorithm after a single zero-discrepancy has been encountered in the large field case.

"Early termination" has consequently been included in recent implementations of Krylov-based algorithms.

Kaltofen, Lee, and Lobo [8] describe an application of this "early termination" idea in a different setting. They have also reported experimental evidence that the current analysis of this may be pessimistic. In particular, their work provides additional evidence that early termination might also be reliable for computations over small fields.

As noted by Dornstetter [5], the Berlekamp-Massey algorithm and the Euclidean algorithm are closely related. A study involving zero-discrepancies has consequently been a part of the analysis of the Euclidean algorithm. Ma and von zur Gathen [13] present relevant results and can be consulted for additional references. However, the work mentioned there concerns a different situation. Furthermore, it would seem to be more relevant to an average case analysis than a worst case analysis of the algorithms considered here, since it seems to require that the polynomial f (corresponding to the minimal polynomial of A and b in the above discussion) is also randomly selected. Nevertheless it also suggests that zero-discrepancies are infrequent in the small field case.

In this paper, it is established that a version of early termination is, indeed, reliable when the randomized Krylov-based algorithms, mentioned above, are used to solve systems over small fields. If a vector u is uniformly selected and used to form a linearly recurrent sequence

$$c_0, c_1, c_2, \ldots, \qquad \text{where } c_i = u^T A^i b,$$

and if a sequence of more than a logarithmic number of zero-discrepancies is detected during an application of the Berlekamp-Massey algorithm to compute the minimal polynomial of the above sequence $c_0, c_1, c_2, \ldots$, then one can reliably terminate the computation with high probability, regardless of the choice of the matrix A or vector b, and for a computation over any field. On the other hand, early termination is provably unreliable if it is performed before a sequence of $\Theta(\log_q N)$ zero-discrepancies has been seen. The results of this paper also establish a bound on the expected amount of "lookahead" that is required when a randomized Lanczos algorithm of the type described above is used to solve an arbitrary nonsingular system of linear equations or to sample from the nullspace of a given matrix.

This paper also includes a brief presentation of Krylov-based algorithms that incorporate early termination and that can be used to solve nonsingular systems of linear equations over finite fields. As noted above, algorithms that sample from the nullspace of a singular matrix are also of considerable interest. While some conclusions about these algorithms can be reached, on the basis of this work, these algorithms are not considered in any detail here. A more complete analysis of such algorithms requires additional results and will be considered in future work.

Finally, it should be noted that this work is part of an ongoing study of "black box linear algebra." The report of Chen, et al [3] includes a discussion of the application of Krylov-based algorithms to solve related problems as well as additional techniques that should be considered.

Linearly recurrent sequences and their properties are considered below, in Section 2. Additional details concerning the Berlekamp-Massey algorithm are presented in Section 3. Properties of "randomly chosen" linearly recurrent sequences, and the main technical results in this report, are found in Section 4. These technical results are applied, to consider Krylov-based algorithms to solve nonsingular systems of linear equations, in Section 5. Finally, related problems that should be considered in future work are described in Section 6.

This extended abstract does not include proofs of the technical results that are presented. A more complete version [6] that includes these arguments is now available.

2. LINEARLY RECURRENT SEQUENCES

2.1 Characteristic and Minimal Polynomials

Once again, let

$$c_0, c_1, c_2, \ldots$$

be a sequence of values in a field F, and let g be a nonzero polynomial

$$g = \beta_0 + \beta_1 x + \cdots + \beta_{m-1}x^{m-1} + \beta_m x^m \in \mathsf{F}[x]$$

with degree $m \geq 0$. We say that g is a *characteristic polynomial* of the sequence $c_0, c_1, c_2, \ldots$ if

$$\beta_0 c_j + \beta_1 c_{j+1} + \cdots + \beta_{m-1} c_{j+m-1} + \beta_m c_{j+m} = 0 \qquad (1)$$

for every integer $j \geq 0$.

It is not necessarily the case that a given sequence has a characteristic polynomial at all. A sequence $c_0, c_1, c_2, \ldots$

is *linearly recurrent* if it does have a nonzero characteristic polynomial.

Suppose now that a given sequence is linearly recurrent. Such a sequence has more than one characteristic polynomial. Indeed, it can be shown that the set of polynomials that are characteristic polynomials of this sequence (together with the zero polynomial) forms an ideal in $\mathsf{F}[x]$. Since $\mathsf{F}[x]$ is a principal ideal domain, this ideal has a generator. There is, consequently, a unique monic polynomial $g \in \mathsf{F}[x]$ such that g is a characteristic polynomial of the sequence and such that $h \in \mathsf{F}[x]$ is a characteristic polynomial of the sequence if and only if h is a nonzero multiple of g, for every polynomial $h \in \mathsf{F}[x]$. This polynomial, g, is called the *minimal polynomial* of the above linearly recurrent sequence.

Henceforth, let us consider a fixed linearly recurrent sequence $c_0, c_1, c_2, \ldots$. We will use the expression "$\mathsf{CharPol}[g]$" to denote the property that a polynomial g is a characteristic polynomial of the sequence $c_0, c_1, c_2, \ldots$. The expression "$\mathsf{MinPol}[g]$" will denote the property that g is the minimal polynomial.

2.2 Annihilators

It will also be useful to consider initial sequences of finite length. For a positive integer i and a polynomial

$$g = \beta_0 + \beta_1 x + \cdots + \beta_{m-1} x^{m-1} + \beta_m x^m \in \mathsf{F}[x],$$

with degree m, we will say that g is an *annihilator* of the initial sequence $c_0, c_1, \ldots, c_{i-1}$ if

$$\beta_0 c_j + \beta_1 c_{j+1} + \cdots + \beta_{m-1} c_{j+m-1} + \beta_m c_{j+m} = 0$$

for every integer j such that $0 \leq j \leq i-1-m$, and we will use the expression "$\mathsf{Ann}[g, i]$" to denote this property. Note that the property is trivial if $i \leq m$. On the other hand, g is a characteristic polynomial for the sequence $c_0, c_1, c_2, \ldots$ if and only if $\mathsf{Ann}[g, i]$ for every integer i.

It will be useful to have a notion of a "minimal annihilator" of an initial sequence as well. We will say that a polynomial g is a *minimal annihilator* of the initial sequence $c_0, c_1, \ldots, c_{i-1}$ if g is monic, g is an annihilator of this initial sequence, and if no nonzero polynomial whose degree is less than that of g is an annihilator of this initial sequence as well. We will use the expression "$\mathsf{MinAnn}[g, i]$" to denote this property.

Unfortunately, these "minimal annihilators" are not generally unique. They are, however, unique under some additional conditions, and this will be sufficient for our purposes. Consider, therefore, the Hankel matrix

$$H = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{n-1} \\ c_1 & c_2 & c_3 & \cdots & c_n \\ c_2 & c_3 & c_4 & \cdots & c_{n+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n-1} & c_n & c_{n-1} & \cdots & c_{2n-2} \end{bmatrix} \tag{2}$$

whose entry in row i and column j is c_{i+j-2} for $1 \leq i, j \leq n$, where n is a given upper bound on the degree of the minimal polynomial of the sequence $c_0, c_1, c_2, \ldots$. Consider the m^{th} principal minor of this matrix

$$H_m = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{m-1} \\ c_1 & c_2 & c_3 & \cdots & c_m \\ c_2 & c_3 & c_4 & \cdots & c_{m+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{m-1} & c_m & c_{m-1} & \cdots & c_{2m-2} \end{bmatrix} \tag{3}$$

for $1 \leq m \leq n$.

LEMMA 1. *Let m be an integer such that $1 \leq m \leq n$ and consider the initial sequence*

$$c_0, c_1, \ldots, c_{2m-1}$$

with length $2m$. This sequence has a minimal annihilator with degree m if and only if the matrix H_m is nonsingular.

Furthermore, if H_m is nonsingular, then the minimal annihilator of the above initial sequence is unique.

LEMMA 2. *Suppose that*

$$c_0, c_1, c_2, \ldots$$

is a linearly recurrent sequence whose entries are not all zero. Suppose as well that g is the minimal polynomial of this sequence and let m be the degree of g.

Then the Hankel matrix H_m is nonsingular, and g is also the unique minimal annihilator of the initial sequence

$$c_0, c_1, \ldots, c_{2m-1}.$$

3. THE BERLEKAMP-MASSEY ALGORITHM

The properties that were presented in the previous section are exploited by the *Berlekamp-Massey algorithm*. This algorithm uses an upper bound n on the degree of the minimal polynomial of a linearly recurrent sequence, and the first $2n$ entries

$$c_0, c_1, \ldots, c_{2n-1}$$

of the sequence, to compute the minimal polynomial. The algorithm generates a sequence

$$g_1, g_2, \ldots, g_{2n}$$

of monic polynomials such that g_i is a minimal annihilator of the initial sequence $c_0, c_1, \ldots, c_{i-1}$ for $1 \leq i \leq 2n$.

Now suppose that the entire sequence has minimal polynomial g and let m be the degree of g. Then $m \leq n$ and it follows by Lemma 2, above, that $g = g_{2m}$. Furthermore, since g is both the only monic annihilator for the initial sequence

$$c_0, c_1, \ldots, c_{2m-1}$$

with degree at most m, and the minimal polynomial of the entire sequence, it must the only monic annihilator for each sequence

$$c_0, c_1, \ldots, c_i$$

with degree at most m, for every integer $i \geq 2m-1$ as well. Thus

$$g = g_{2m} = g_{2m+1} = \cdots = g_{2n}.$$

The algorithm therefore returns the final polynomial, g_{2n}, that it generates, as the minimal polynomial of the entire sequence.

As this description may suggest, time can be saved when $m < n$, if one can determine (reliably) that the minimal polynomial of the entire sequence has been generated before all of the initial $2n$ entries of the sequence have been considered.

Let us now consider a fixed sequence $c_0, c_1, c_2, \ldots$, along with an upper bound n on the degree of its minimal polynomial. Let $g_1, g_2, \ldots, g_{2n}$ be the sequence of polynomials generated by the Berlekamp-Massey algorithm when it is given $c_0, c_1, \ldots, c_{2n-1}$ and n as input. For the purposes of the following definitions, set $g_{-1} = 0$ and $g_0 = 1$.

We will say that the given sequence $c_0, c_1, c_2, \ldots$ has a *zero-discrepancy at position i* (for $i \leq 2n-1$) if $g_i = g_{i+1}$, and that the sequence has a *sequence of zero-discrepancies of length j beginning at position i* (for $i + j \leq 2n$) if

$$g_{i-1} \neq g_i = g_{i+1} = \cdots = g_{i+j}.$$

We will call this a *harmful* sequence of zero-discrepancies if g_i is not equal to the minimal polynomial of the linearly recurrent sequence $c_0, c_1, c_2, \ldots$. We will say that the sequence $c_0, c_1, c_2, \ldots$ has a *harmful sequence of zero-discrepancies of length j* if it has a harmful sequence of zero-discrepancies of length j beginning at position i, for some integer i such that $0 \leq i < 2n - j$.

Consider again the matrix H shown in Equation (2). Let Δ_i be the determinant of the i^{th} principal minor H_i of this matrix if $1 \leq i \leq n$ and define Δ_0 to be 1. We will say that H has a *gap of length j beginning at position i* if

$$\Delta_i \neq 0 = \Delta_{i+1} = \Delta_{i+2} = \cdots = \Delta_{i+j}.$$

We will call this a *harmful* gap if $\Delta_\ell \neq 0$ for some integer $\ell > i + j$. We will say that H has a *harmful gap of length j* if H has a harmful gap of length j beginning at position i for some integer i such that $0 \leq i < n - j$.

4. RANDOM SEQUENCES

One objective of this work is to show that long sequences of zero-discrepancies, that are harmful, are unlikely when the Berlekamp-Massey algorithm is applied as part of a randomized algorithm to solve a system of linear equations over a finite field. We will show that harmful gaps that are long are unlikely, as well.

Let us therefore return attention to the original problem, namely, the consideration of a system

$$Ax = b,$$

where $A \in \mathsf{F}^{N \times N}$, $b \in \mathsf{F}^{N \times 1}$, and $F = \mathsf{F}_q$ is a finite field with q elements. The set of polynomials $g \in \mathsf{F}[x]$, such that $g(A)b = 0$, forms an ideal in $\mathsf{F}[x]$. Since the characteristic polynomial of A is an element of this ideal, the ideal is nonzero. There is, therefore, a monic polynomial $f \in \mathsf{F}[x]$ that generates this ideal. This polynomial is called the *minimal polynomial* of A and b, and it follows by its definition that $g(A)b = 0$ if and only if g is a multiple of f, for every polynomial $g \in \mathsf{F}[x]$. As suggested in the introduction, this is also the (unique) monic polynomial f of least degree such that $f(A)b = 0$.

Suppose, now, that f has degree n and that

$$f = \alpha_0 + \alpha_1 x + \cdots + \alpha_{n-1} x^{n-1} + x^n.$$

Then $0 \leq n \leq N$, since the characteristic polynomial of A has degree N and is a multiple of f.

As mentioned in the introduction, the algorithms to be studied select a random vector $u \in \mathsf{F}^{N \times 1}$ and consider the sequence $c_0, c_1, c_2, \ldots$, where

$$c_i = u^T A^i b \qquad \text{for } i \geq 0. \tag{4}$$

Note that if $j \geq 0$ then

$$\alpha_0 c_j + \alpha_1 c_{j+1} + \cdots + \alpha_{n-1} c_{j+n-1} + c_{j+n} = u^T A^j f(A) b = 0.$$

Therefore any sequence $c_0, c_1, c_2, \ldots$ that is generated in this way is linearly recurrent, and f is a characteristic polynomial (although, not necessarily the minimal polynomial) of this sequence.

It will be necessary to identify the probability that a given linearly recurrent sequence is generated, by the above process, in order to analyze the algorithms that are of interest. It follows by the definition of a "characteristic polynomial" of a linearly recurrent sequence that if the initial n entries

$$c_0, c_1, \ldots, c_{n-1}$$

are given, along with the information that f is a characteristic polynomial of the sequence to be studied, then the remaining entries

$$c_n, c_{n+1}, c_{n+2}, \ldots$$

are fixed. On the other hand, the condition that f is a characteristic polynomial of the sequence does not constrain the choice of the initial n entries of the sequence at all. It follows that (since $\mathsf{F} = \mathsf{F}_q$ is a finite field of size q) there are exactly q^n linearly recurrent sequences with entries in F that have f as a characteristic polynomial.

LEMMA 3. *Let $A \in \mathsf{F}^{N \times N}$, let $b \in \mathsf{F}^{N \times 1}$, and suppose that $f \in \mathsf{F}[x]$ is the minimal polynomial of A and b. Let n be the degree of f.*

Let $s_0, s_1, s_2, \ldots$ be any linearly recurrent sequence with entries in F and with characteristic polynomial f.

Finally, suppose that a vector u is chosen uniformly and randomly from $\mathsf{F}^{N \times 1}$, and let $c_j = u^T A^j b$ for $j \geq 0$. Then

$$c_i = s_i \quad \textit{for every integer } i \geq 0$$

with probability q^{-n}.

In other words, the randomized algorithms that are to be studied generate the linearly recurrent sequences with characteristic polynomial f *uniformly*.

The next result follows by an application of the theory of subresultants. The text of von zur Gathen and Gerhard [7] includes a readable introduction to this theory, as well as additional references.

LEMMA 4. *Let $f \in \mathsf{F}[z]$ be a monic polynomial with degree n and let $g \in \mathsf{F}[z]$ be a monic polynomial with degree m, where $m \leq n$ and where $\mathsf{F} = \mathsf{F}_q$ is the finite field with q elements. Suppose that the greatest common divisor h of f and g has degree k. Finally, let s be an integer such that $m \leq s \leq 2n$.*

Let $c_0, c_1, c_2, \ldots$ be uniformly chosen from the set of sequences with characteristic polynomial f.

If $s < n + m - k$ then the above sequence satisfies the condition

$$\mathsf{Ann}[g, s]$$

with probability q^{m-s}.

If $s \geq n + m - k$ *then the above sequence satisfies the condition*

$$\mathsf{Ann}[g, s]$$

with probability q^{k-n}*, and the conditions*

$$\mathsf{Ann}[g, s] \qquad \textit{and} \qquad \mathsf{CharPol}[h]$$

are equivalent.

This lemma is used to establish the one that follows.

LEMMA 5. *Let* f*,* g*,* h*,* n*,* m*, and* k *be as in the statement of the previous lemma. Suppose, once again, that*

$$c_0, c_1, c_2, \ldots$$

is uniformly chosen from the set of sequences with with characteristic polynomial f*. Let* s *and* t *be nonnegative integers such that* $m \leq s \leq s + t \leq 2n$*, and let* ϵ *be a positive real number. Then either*

$$\text{Prob}\left(\mathsf{MinAnn}[g, s]\right) \leq \epsilon \tag{5}$$

or

$$\frac{\text{Prob}\left(\mathsf{MinAnn}[g, s + t] \wedge \neg\mathsf{MinPol}[g]\right)}{\text{Prob}\left(\mathsf{MinAnn}[g, s]\right)} \leq q^{m-s-t}/\epsilon. \tag{6}$$

These results can be used to prove the following theorem.

THEOREM 6. *Let* $f \in \mathsf{F}[z]$ *be a monic polynomial with degree* n *over the finite field* $\mathsf{F} = \mathsf{F}_q$ *and suppose that the linearly recurrent sequence*

$$c_0, c_1, c_2, \ldots$$

is uniformly chosen from the set of sequences with characteristic polynomial f*. Let* m *and* t *be integers such that* $0 \leq m \leq m + t \leq n - 1$*. Then the matrix* H *corresponding to the above sequence has a harmful gap of length* t*, beginning at position* m*, with probability at most* $2q^{-t/2}$.

COROLLARY 7. *Let* $f \in \mathsf{F}[z]$ *be a monic polynomial with degree* n *over the finite field* $\mathsf{F} = \mathsf{F}_q$ *and suppose that the linearly recurrent sequence*

$$c_0, c_1, c_2, \ldots$$

is uniformly chosen from the set of sequences with characteristic polynomial f*. Then the probability that the corresponding Hankel matrix* H *has a harmful gap of length* t *is at most* $2(n - t)q^{-t/2}$.

COROLLARY 8. *Let* $f \in \mathsf{F}[z]$ *be a monic polynomial with degree* n *over the finite field* F_q *and suppose that the linearly recurrent sequence*

$$c_0, c_1, c_2, \ldots$$

is uniformly chosen from the set of sequences with characteristic polynomial f*. Then the probability that this linearly recurrent sequence has a harmful sequence of zero discrepancies of length* t *is at most* $2(n - t/2)q^{-t/4}$.

It is unlikely that the bounds in the above corollaries are tight. However, the next results suggest that improvements to these bounds will not lead to significant improvements of results concerning the reliability of algorithms.

THEOREM 9. *Let* f *be a monic polynomial with degree* n *over the finite field* F_q *and suppose that the linearly recurrent sequence*

$$c_0, c_1, c_2, \ldots$$

is uniformly chosen from the set of sequences with characteristic polynomial f*. Let* t *be a positive integer such that* $n \geq 2t$.

Then the probability that the above sequence does not have a harmful sequence of zero-discrepancies, of length $t - 1$*, is at most* $e^{-n/(2tq^t)}$.

COROLLARY 10. *Let* f *be a monic polynomial with degree* n *over the finite field* F_q *and suppose that the linearly recurrent sequence*

$$c_0, c_1, c_2, \ldots$$

is uniformly chosen from the set of sequences with characteristic polynomial f*. Let* t *be a positive integer such that* $n \geq 4t$.

Then the probability that the Hankel matrix H *that corresponds to the above sequence does not have harmful gap, of length* $t - 1$*, is at most* $e^{-n/(4tq^{2t})}$.

5. SOLVING SYSTEMS OF LINEAR EQUATIONS

Lemma 3 implies that the bounds on probabilities given in Theorems 6 and 9 and Corollaries 7, 8, and 10 are correct when one attempts to solve a system of linear equations

$$Ax = b$$

for a given nonsingular matrix $A \in \mathsf{F}^{N \times N}$ and vector $b \in \mathsf{F}^{N \times 1}$, over a finite field $\mathsf{F} = \mathsf{F}_q$, by choosing a vector u uniformly from $\mathsf{F}^{N \times 1}$ and considering the resulting linearly recurrent sequence

$$c_0, c_1, c_2, \ldots$$

where $c_i = u^T A^i b$ for every integer $i \geq 0$.

There are several different (closely related) algorithms that make use of this sequence, in some way. These include Wiedemann's algorithm [19], a modification of the algorithm of Lanczos [11] that can be applied to systems whose coefficient matrix is not symmetric and that incorporates a "lookahead" process to continue computation when orthogonal vectors are encountered, and, finally, a hybrid algorithm that computes both the sequence of polynomials generated using the Berlekamp-Massey process, and the vectors generated by the Lanczos computation, such as the algorithm described in Section 3.4 of the thesis of Lambert [10]. Each of these is discussed below.

5.1 Wiedemann's Algorithm

Once again, consider the given system of linear equations, $Ax = b$. Let $b_1 = b$ and let f_1 be the minimal polynomial of the matrix A and vector b_1. Let d_1 be the degree of f_1.

When Wiedemann's algorithm is applied, the Berlekamp-Massey algorithm is used to recover the minimal polynomial g_1 of the above linearly recurrent sequence,

$$c_0, c_1, c_2, \ldots \qquad \text{where } c_i = u_1^T A^i b_1,$$

for a randomly selected vector u_1, along with an estimate of the solution for the given system. If $g_1 = f_1$ then the estimate is, in fact, the solution for this system of equations.

On the other hand, if the estimate is not the solution, so that g_1 is a divisor of f_1 and $g_1 \neq f_1$, then the information that has been generated is applied to reduce the originally given problem to that of solving a system

$$Ax = b_2,$$

where b_2 is a vector such the minimal polynomial of A and b_2 is $f_2 = f_1/g_1$.

Continuing as needed, one obtains an iterative process, in which one wishes to solve a system $Ax = b_i$ at the beginning of the i^{th} iteration, and in which one is either successful (so that the process terminates) or a vector b_{i+1} is formed for use in the $i + 1^{\text{st}}$ iteration of the process. If f_i is the minimal polynomial of A and b_i, and d_i is the degree of f_i, then $d_{i+1} \leq d_i$ if an $i + 1^{\text{st}}$ iteration is required.

Wiedemann establishes that if the resulting iterative process is applied, and the vectors $u_1, u_2, \ldots$ that are required for each iteration are independently chosen, then a solution for the original solution is obtained, with high probability, after a constant number of iterations.

Wiedemann also analyzes the cost of each iteration. Suppose that n is an upper bound on the degree of the unknown minimal polynomial of the matrix A and vector b that is being considered during a given iteration. Then one iteration of Wiedemann's process can either be implemented to use up to $3n$ multiplications of the given matrix A by vectors, $O(nN)$ additional arithmetic operations over F, and while storing $O(N)$ elements of F, or it can be implemented to use up to $2n$ multiplications of the given matrix A by vectors, $O(nN)$ additional arithmetic operations over F, and while storing $O(nN)$ field elements.

The time required for this process is generally dominated by the cost of multiplications of the given matrix A by vectors. Consequently the time used by the second implementation can be considerably lower than that of the first. However, the storage requirements for the second implementation frequently prohibit its use.

One can obtain a rather naive (and, probably, pessimistic) upper bound on the expected cost of the entire process by multiplying the expected number of iterations that are required by the worst-case cost of a single iteration.

In contrast, a Las Vegas algorithm whose worst case expected running time closely matches that of a single iteration of the Wiedemann process can be obtained by incorporating early termination. Consider, once again, a system $Ax = b$ that is to be solved during a given iteration. Once again, let n be an upper bound on the degree of the minimal polynomial of A and b; one can certainly use N as this upper bound for the initial iteration of the process. Suppose, furthermore, that the Berlekamp-Massey process is terminated, either after $2n$ terms of the corresponding linearly recurrent sequence have been processed, or after a sequence of zero-discrepancies with length $\lceil 8 \log_q N \rceil + 1$ has been encountered. Suppose, as well, that the minimal polynomial of the linearly recurrent sequence that is currently being processed (using the Berlekamp-Massey algorithm) has degree d. Then the number of multiplications of A by vectors, required for this iteration, can be bounded by either $2d + O(\log_q N)$, if the space-inefficient implementation of Wiedemann's process is used, or $3d + O(\log_q N)$, if the space-efficient one is used instead. Each iteration is correct (that is, early termination does not introduce an error) with probability at least $1 - 1/N$.

Wiedemann's bound on the expected number of iterations can now be applied to conclude that the worst case expected number of multiplications of A by vectors, required for the entire process, is either $2n + O(\log_q N)$ for the space-inefficient implementation, or $3n + O(\log_q N)$ for the space-efficient implementation, where n is used here to denote the degree of the minimal polynomial of A and b, where b is the originally given vector, and where $A \in \mathsf{F}^{N \times N}$ as above. The worst case expected number of additional operations over F changes by at most a small constant factor, and the storage requirements are unchanged.

Theorem 9 indicates that one should not expect to be able to do significantly better than this in all cases. Suppose, once again, that one is processing a linearly recurrent sequence that is derived from a matrix A and vector b, such that the minimal polynomial of A and b has degree n. Suppose, as well, that the Berlekamp-Massey algorithm is terminated as soon as a sequence of $\frac{1}{3} \log_q n$ zero-discrepancies is encountered. Then the probability that the result is correct is provably low. For example, an upper bound on the probability of correctness of $n^{-1/2}$ is easily established, for sufficiently large n.

5.2 The Lanczos Process

There are several different ways in which one might modify the Lanczos algorithm in order to solve systems of linear equations over finite fields. The discussion of the cost of this approach is based on the work of Lambert [10], who contributes a detailed analysis along with additional references.

In general, when applying a version of the Lanczos algorithm that does not require the given coefficient matrix A to be symmetric, one attempts to construct a dual orthogonal basis for a pair of vector spaces, namely, the spaces generated by the sequences of vectors

$$b, Ab, A^2b, A^3b, \ldots$$

and

$$u, A^T u, (A^T)^2 u, (A^T)^3 u, \ldots .$$

Difficulties arise when a sequence of vectors from the former space, that are all orthogonal to a given vector in the latter space, are encountered. A "lookahead" process is included to handle these difficulties.

As noted by Lambert, one can implement a lookahead process in more than one way; space-efficient and space-inefficient implementations can be considered once again. The worst-case number of multiplications of A or A^T by vectors, for the space-efficient implementation, does not appear to be very different from the number given above, for the space-efficient implementation of a single iteration of Wiedemann's process. The worst case number of multiplications of A or A^T by vectors, for the space-inefficient implementations of (a single iteration of) the Wiedemann process and a Lanczos process, appear to agree as well. However, the space requirements for the "space-inefficient" implementation are much better: The number of elements of F that must be stored (at one time) can now be bounded by $O(NL)$ where L is the maximum "size of a lookahead block" (as defined by Lambert). One can see by Lambert's analysis that this is the same as the maximum length of a harmful gap for the Hankel matrix H that corresponds to the linearly current sequence that is being processed. Thus the expected amount

of space required, in order to match the time requirements given for the space-inefficient version of Wiedemann's algorithm, is in $O(N \log_q N)$ — rather than $\Theta(N^2)$, as is the case for the Wiedemann process.

Unfortunately, if the only modifications to the Lanczos process are the ones mentioned above, then one should not expect the process to result in a solution for the given system unless the minimal polynomial of the linearly recurrent sequence

$$c_0, c_1, c_2, \ldots \qquad \text{where } c_i = u^T A^i b$$

is the same as the minimal polynomial f of A and b. Early termination can be incorporated to determine whether this is the case, somewhat sooner than would otherwise be possible. However, this version of the Lanczos process does not provide a way to use the information gained, when the two "minimal polynomials" mentioned above are different, in order to reduce the cost of later attempts.

One naive approach that can be used to overcome this difficulty is to use independent trials of the Lanczos process, in hopes that one of these trials will succeed (that is, in hopes that the two "minimal polynomials" mentioned above are, in fact, the same). A part of the probabilistic analysis of Wiedemann's algorithm (specifically, Proposition 3 in Section VI of Wiedemann's paper [19]) can be used to establish that this approach will succeed, with high probability, if $\Theta(\log_q N)$ trials are used. However, the time required for this process is considerably higher than that needed with Wiedemann's approach, when this number of independent trials is used.

5.3 Lambert's Combined Approach

Lambert's work provides a unification of the Wiedemann and Lanczos approaches. As part of this work, a hybrid algorithm that produces both the sequence of vectors one would obtain from the Lanczos process, and the polynomials generated by the Berlekamp-Massey algorithm, is described in Chapter 3 of Lambert's thesis [10].

Lambert's thesis should be consulted for a detailed description of this algorithm. A combination of a brief analysis of the algorithm that is presented at the end of Chapter 3 of the thesis, the results of this paper (which eliminate an assumption that is used in Lambert's analysis), and results from Wiedemann's analysis of his own algorithm, provides an analysis of a Wiedemann-style iterative algorithm in which applications of the Berlekamp-Massey algorithm are replaced with applications of Lambert's. The expected number of multiplications of the matrix A by vectors is $2n + O(\log_q n)$, the expected number of additional operations in F is in $O(nN)$, and the expected amount of storage space required is in $O(N \log_q N)$. Thus this algorithm comes close to combining the advantages of both implementations of Wiedemann's algorithm.

6. RELATED PROBLEMS

The results presented above require the assumption that elements of the ground field F are selected uniformly and independently from F when vectors are formed. One might also consider the case that these elements are selected uniformly and independently from a smaller subset S of F. In an extreme case, F is infinite and $S = \{0, 1\}$.

The Schwartz-Zippel lemma [17], [20] has been applied to closely related problems. For example, the results of Kaltofen and Pan [9] can be used to bound the probability that a sequence $c_0, c_1, c_2, \ldots$, whose elements are randomly selected as discussed here, has a harmful zero-discrepancy with length at least two. The resulting probability bound is nontrivial (that is, less than 1) when $|S| > N$, and it decreases as $|S|$ increases. Unfortunately, there is no apparent way to obtain improved bounds for longer sequences of zero-discrepancies, or to obtain bounds that are of much use at all for the case $|S| < N$. There is no obvious way to modify the results presented in Section 4, above, in order to obtain a probability analysis for this version of the problem, either. Since one might wish to choose values from a very small set S, in order to reduce the precision needed for computations, this version of the problem is of potential interest.

The work presented in this paper does not address the behaviour of some additional Krylov-based algorithms that are in use. In particular, it is not directly relevant to versions of either Wiedemann's algorithm or a Lanczos algorithm that require the coefficient matrix A to be symmetric and that perform computations involving a linearly recurrent sequence

$$c_0, c_1, c_2, \ldots$$

where $c_i = b^T A^i b$ for a single randomly chosen vector b. There is work to be done to analyze the reliability of these algorithms when they are used to solve symmetric linear systems of equations over small finite fields.

There is also work remaining in order to analyze the reliability of algorithms that process blocks of vectors. While block Wiedemann algorithms are now well understood in the small field case (see, in particular, the work of Villard [18] and the references therein), the same cannot be said for Lanczos-style algorithms that process blocks of vectors. Such "block-Lanczos" algorithms have been considered by several authors, including Coppersmith [4] and Montgomery [15], [16]; Montgomery's algorithm includes a form of early termination and has not been completely analyzed. In addition, Austin Lobo [12] reports experimental results concerning the use of early termination, for block algorithms in the small field case, providing questions for additional study.

Therefore, regardless of whether (or how) the results of the current paper can be applied, it is clear that interesting work in this area remains to be done.

7. ACKNOWLEDGEMENTS

Austin Lobo and other members of the LinBox project have made numerous helpful comments, in the course of this work, and have my thanks.

8. REFERENCES

[1] E. R. Berlekamp. *Algebraic Coding Theory.* McGraw-Hill, 1968.

[2] J. P. Buhler, H. W. Lenstra, Jr., and C. Pomerance. Factoring integers with the number field sieve. In *The Development of the Number Field Sieve*, volume 1554 of *Lecture Notes in Mathematics*, pages 50–94. Springer-Verlag, 1993.

[3] L. Chen, W. Eberly, E. Kaltofen, B. D. Saunders, W. J. Turner, and G. Villard. Efficient matrix preconditioners for black box linear algebra. *Linear Algebra and its Applications*, 343–344:119–146, 2002.

[4] D. Coppersmith. Solving linear equations over GF(2); block Lanczos algorithm. *Linear Algebra and its Applications*, 192:33–60, 1993.

[5] J. L. Dornstetter. On the equivalence between Berlekamp's and Euclid's algorithms. *IEEE Transactions on Information Theory*, IT-33:428–431, 1987.

[6] W. Eberly. Early termination over small fields. Technical Report 2003-723-26, Department of Computer Science, University of Calgary, May 2003.

[7] J. von zur Gathen and J. Gerhard. *Modern Computer Algebra*. Cambridge University Press, 1999.

[8] E. Kaltofen, W.-S. Lee, and A. Lobo. Early termination in Ben-Or/Tiwari sparse interpolation and a hybrid of Zippel's algorithm. In *Proceedings, 2000 International Symposium on Symbolic and Algebraic Computation (ISSAC '00)*, pages 192–201, 2000.

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

[10] R. Lambert. *Computational Aspects of Discrete Logarithms*. PhD thesis, University of Waterloo, Waterloo, Ontario, Canada, 1996.

[11] C. Lanczos. Solution of systems of linear equations by minimized iterations. *J. Res. Nat. Bur. Standards*, 49:33–53, 1952.

[12] A. Lobo. *Matrix-Free Linear System Solving and Applications to Symbolic Computation*. PhD thesis, Rensselaer Polytechnic Institute, Troy, New York, 1995.

[13] K. Ma and J. von zur Gathen. Analysis of Euclidean algorithms for polynomials over finite fields. *Journal of Symbolic Computation*, 9:429–455, 1990.

[14] J. L. Massey. Shift-register synthesis and BCH decoding. *IEEE Transactions on Information Theory*, IT-15:122–127, 1969.

[15] P. Montgomery. A block Lanczos algorithm for finding dependencies over GF(2). In *Advances in Cryptology — EUROCRYPT '95*, volume 921 of *Lecture Notes in Computer Science*, pages 106–120. Springer, 1995.

[16] P. Montgomery. Distributed linear algebra. In *Proceedings, 4th Workshop on Elliptic Curve Cryptography*, 2000.

[17] J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *Journal of the Association of Computing Machinery*, 27:701–717, 1980.

[18] G. Villard. Further analysis of Coppersmith's block Wiedemann algorithm using matrix polynomials. Rapport de Recherche 975 IM, Institut d'Informatique et de Mathematiques Applicees de Grenoble, 1997.

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

[20] R. Zippel. Probabilistic algorithms for sparse polynomials. In *EUROSAM '79*, volume 72 of *Lecture Notes in Computer Science*, pages 216–226. Springer-Verlag, 1979.