Early Termination over Small Fields *

Wayne Eberly†
Department of Computer Science
University of Calgary

May, 2003

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

---

*An extended abstract of this work was presented at the International Symposium on Symbolic and Algebraic Computation [6].

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

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 $f \in \mathsf{F}[z]$ with least degree 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. This paper also provides upper and lower bounds on the expected amount of "lookahead" that is required, in the worst case, 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 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.

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.

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.

Definition 2.1. Let f be a nonzero polynomial

$$f = \alpha_0 + \alpha_1 z + \cdots + \alpha_{n-1} z^{n-1} + \alpha_n z^n \in \mathsf{F}[z]$$

with degree $n \geq 0$, where z is an indeterminate over the field F. Then f is a *characteristic polynomial* of the sequence $c_0, c_1, c_2, \ldots$ if

$$\alpha_0 c_i + \alpha_1 c_{i+1} + \cdots + \alpha_{n-1} c_{i+n-1} + \alpha_n c_{i+n} = 0 \tag{1}$$

for every integer $i \geq 0$.

It is not necessarily the case that a given sequence has a characteristic polynomial at all.

Definition 2.2. A sequence $c_0, c_1, c_2, \ldots$ is *linearly recurrent* if it has 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 $\mathsf{F}[z]$. Since $\mathsf{F}[z]$ is a principal ideal domain, this ideal has a generator. It follows that a linearly recurrent sequence has a unique "minimal polynomial," where this is defined as follows.

Definition 2.3. A polynomial $f \in \mathsf{F}[z]$ is the *minimal polynomial* of the linearly recurrent sequence $c_0, c_1, c_2, \ldots$ if f is monic, f is a characteristic polynomial of the sequence $c_0, c_2, c_2, \ldots$, and if $g \in \mathsf{F}[z]$ is a characteristic polynomial of this sequence if and only if g is a nonzero multiple of f, for every polynomial $g \in \mathsf{F}[z]$.

In other words, f is the *minimal polynomial* of the linearly recurrent sequence $c_0, c_1, c_2, \ldots$ if f is the unique monic generator of the ideal that consists of the zero polynomial and the set of characteristic polynomials of the given sequence.

Henceforth, let us consider a fixed linearly recurrent sequence $c_0, c_1, c_2, \ldots$. We will use the expression

$$\text{“}\mathsf{CharPol}[f]\text{”}$$

to denote the property that a polynomial f is a characteristic polynomial of the given sequence. The expression

$$\text{“}\mathsf{MinPol}[f]\text{”}$$

will denote the property that f is the minimal polynomial of this sequence.

2.2 Annihilators

It will also be useful to consider initial sequences of finite length. As above, we will consider a fixed linearly recurrent sequence $c_0, c_1, c_2, \ldots$.

Definition 2.4. Let i be a positive integer and let f be a nonzero polynomial

$$f = \alpha_0 + \alpha_1 z + \cdots + \alpha_{n-1} z^{n-1} + \alpha_n z^n \in \mathsf{F}[z],$$

with degree $n \geq 0$. Then f is an *annihilator* of the initial sequence $c_0, c_1, \ldots, c_{i-1}$ (or, "*f annihilates* this sequence") if

$$\alpha_0 c_j + \alpha_1 c_{j+1} + \cdots + \alpha_{n-1} c_{j+n-1} + \alpha_n c_{j+n} = 0 \tag{2}$$

for every integer j such that $0 \leq j \leq i - 1 - n$.

We will use the expression

$$\text{“}\mathsf{Ann}[f, i]\text{”}$$

to denote this property. Note that the property is trivial if $i \leq n$. On the other hand, f is a characteristic polynomial for the sequence $c_0, c_1, c_2, \ldots$ if and only if $\mathsf{Ann}[f, i]$ for every integer i.

The next four properties are easily established using the above definition of an "annihilator."

Lemma 2.5. *Let $f \in \mathsf{F}[z]$ be a nonzero polynomial such that f is an annihilator of the initial sequence*

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

for an integer $i \geq 1$. Let α be a nonzero element of F. Then the polynomial αf is an annihilator of the above initial sequence, as well.

Lemma 2.6. *Let $f, g \in \mathsf{F}[z]$ be nonzero polynomials such that the degrees of f and g are not the same, and suppose that f and g are both annihilators of the initial sequence*

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

for an integer $i \geq 1$. Then the sum $f + g$ of these polynomials is an annihilator of the above initial sequence as well.

Lemma 2.7. *Let $f, g \in \mathsf{F}[z]$ be nonzero polynomials such that f and g both have degree n, and suppose that f and g are both annihilators of the initial sequence*

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

for an integer $i \geq 1$. Suppose that the sum $f + g$ has degree $m \leq n$.

Then either $i \leq n - m$, or $f + g$ is an annihilator of the (shorter) initial sequence

$$c_0, c_1, \ldots, c_{i-(n-m)-1}.$$

Lemma 2.8. *Let $f \in \mathsf{F}[z]$ be a nonzero polynomial such that f is an annihilator of the initial sequence*

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

for an integer $i \geq 1$. Then the polynomial zf is an annihilator of the above initial sequence as well.

The next property is easily established using Lemmas 2.5, 2.6 and 2.8.

Lemma 2.9. *Let $f \in \mathsf{F}[z]$ be a nonzero polynomial such that f is an annihilator of the initial sequence*

$$c_0, c_1, \dots, c_{i-1}$$

for an integer $i \geq 1$. Then any nonzero multiple of f in $\mathsf{F}[z]$ is an annihilator of the above initial sequence as well.

It will be useful to have a notion of a "minimal" annihilator of an initial sequence.

Definition 2.10. Let i be a positive integer and let f be a polynomial in $\mathsf{F}[z]$. Then f is a *minimal annihilator* of the initial sequence $c_0, c_1, \dots, c_{i-1}$ if f is monic, f is an annihilator of this initial sequence, and if no factor of f (other than f, itself) is an annihilator of the initial sequence $c_0, c_1, \dots, c_{i-1}$ as well.

We will use the expression

$$\text{“}\mathsf{MinAnn}[f, i]\text{”}$$

to denote this property.

Unfortunately, these minimal annihilators are not generally unique. However, uniqueness can be proved under some additional conditions, and this will be sufficient for our purposes. Consider, therefore, the Hankel matrix

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

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 $c_0, c_1, c_2, \dots$. Consider the i^{th} principal minor of this matrix

$$H_i = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{i-1} \\ c_1 & c_2 & c_3 & \cdots & c_i \\ c_2 & c_3 & c_4 & \cdots & c_{i+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{i-1} & c_i & c_{i-1} & \cdots & c_{2i-2} \end{bmatrix} \tag{4}$$

for $1 \leq 1 \leq n$.

In the remainder of this section we will state and prove a sequence of lemmas that establish a situation in which the minimal annihilator of a given initial sequence is unique.

Lemma 2.11. *Suppose i is an integer such that $1 \leq i \leq n$ and H_i is nonsingular. Then the initial sequence*

$$c_0, c_1, \dots, c_{2i-1}$$

has a unique minimal annihilator, $f \in \mathsf{F}[z]$, and the degree of f is equal to i.

Proof. Since the principal minor H_i is nonsingular, the system

$$H_i \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{i-1} \end{bmatrix} = \begin{bmatrix} c_i \\ c_{i+1} \\ \vdots \\ c_{2i-1} \end{bmatrix} \tag{5}$$

has a unique solution, and this defines the only monic polynomial

$$f = -\alpha_0 - \alpha_1 z - \ldots - \alpha_{i-1} z^{i-1} + z^i$$

in $\mathsf{F}[z]$ with degree i that annihilates the initial sequence

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

Suppose there exists another monic polynomial

$$g = \beta_0 + \beta_1 z + \cdots + \beta_{i-2} z^{i-2} + \beta_{i-1} z^{i-1},$$

in $\mathsf{F}[z]$, with degree less than i, that is also an annihilator of the above initial sequence (so that $\beta_{i-1} = \beta_{i-2} = \cdots = \beta_{j+1} = 0$ and $\beta_j = 1$, if $j \le i-1$ is the degree of g). Then the vector

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{i-1} \end{bmatrix}$$

is in the right nullspace of H_i, and is nonzero, contradicting the fact that H_i is nonsingular.

Thus the above polynomial f, with degree i, is the only minimal annihilator of the initial sequence $c_0, c_1, \ldots, c_{2i-1}$, as claimed in the lemma. ◻

Lemma 2.12. *Suppose i is an integer such that $1 \le i \le n$ and H_i is nonsingular. Let $f \in \mathsf{F}[z]$ be the unique minimal annihilator of the initial sequence*

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

whose existence has been established in Lemma 2.11, above.

Suppose that $j \ge 0$ and that f is also an annihilator of the initial sequence

$$c_0, c_1, \ldots, c_{2i-1+j}.$$

Let $g \in \mathsf{F}[z]$ be another annihilator of the sequence $c_0, c_1, \ldots, c_{2i-1+j}$.

If the degree of g is less than or equal to $i + j$, then g is divisible by f.

Proof. Let $f \in \mathsf{F}[z]$ be as given in the statement of the lemma. It follows by Lemma 2.11 that f has degree i.

This lemma can now be established by induction on j. The strong form of mathematical induction will be used.

Basis: If $j = 0$ then $H_{i+j} = H_i$, and this matrix is nonsingular. Either there is nothing to be established, or the degree of the given polynomial g is at most $i + j = i$. It follows (in the latter case) that g is a scalar multiple of f — for, otherwise, one could divide g by its leading coefficient in order to produce a monic polynomial with degree at most i that is different from f and that is also an annihilator of the initial sequence $c_0, c_1, \ldots, c_{2i-1}$, contradicting the choice of f. It follows that g is divisible by f, if the degree of g is at most i, as required.

Inductive Step: Suppose that the claim is correct for a given value of j (and for all smaller nonnegative values). Suppose, as well, that f is an annihilator of the initial sequence

$$c_0, c_1, \ldots, c_{2i-1+(j+1)},$$

and that g is an annihilator of this initial sequence with degree at most $i + j + 1$.

If the degree of g is less than or equal to $i + j$ as well, then the desired result follows by the inductive hypothesis, since g is also an annihilator of the shorter sequence

$$c_0, c_1, \ldots, c_{2i-1+j}.$$

It is therefore sufficient to consider the case that the degree of g is exactly $i + j + 1$.

Let γ be the leading coefficient of g and consider the polynomial $g - \gamma z^{j+1} f$. Since f is monic with degree i, this polynomial has degree at most $i + j$.

If this polynomial is equal to zero then we are done, since this implies that g is divisible by f.

If this polynomial is nonzero and its degree is less than i, then the vector (with dimension i) whose entries consist of the coefficients of this polynomial is in the right nullspace of H_i — see, again, the last half of the proof of Lemma 2.11 for an application of this argument. This contradicts the fact that H_i is nonsingular.

The only remaining case is that this polynomial has degree $i+k$ for some nonnegative integer $k \leq j$. Since f and g are both annihilators of

$$c_0, c_1, \ldots, c_{2i+j},$$

it can be shown by an application of Lemma 2.9 that the polynomial $-\gamma z^{j+1} f$ is an annihilator of the initial sequence

$$c_0, c_1, \ldots, c_{2i+j},$$

as well, and it can then be shown by an application of Lemma 2.7 that the polynomial $g - \gamma z^{j+1} f$ is an annihilator of the initial sequence

$$c_0, c_1, \ldots, c_{2i-1+k}.$$

Since the polynomial $g - \gamma z^{j+1} f$ has degree $i + k$, and $k \leq j$, it now follows by the inductive hypothesis that $g - \gamma z^{j+1} f$ is divisible by f. This clearly implies that g is divisible by f as well, as is required to complete the proof. $\square$

Lemma 2.13. *Once again, let i be an integer such that $1 \leq i \leq n$ and H_i is nonsingular, and let $f \in \mathsf{F}[z]$ be the unique minimal annihilator of the initial sequence*

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

whose existence is established by Lemma 2.11.

Suppose that j is a positive integer such that f is an annihilator of the initial sequence

$$c_0, c_1, \ldots, c_{2i-2+j}$$

but f is not an annihilator of the initial sequence

$$c_0, c_1, \ldots, c_{2i-1+j}.$$

Then there are no nonzero polynomials in $\mathsf{F}[z]$ that have degree less than $i+j$ and that annihilate the sequence

$$c_0, c_1, \ldots, c_{2i-1+j}$$

at all, and the matrix H_{i+j} is nonsingular.

Proof. Let f be as in the statement of the lemma. It follows by Lemma 2.11 that f has degree i.

Suppose, in order to obtain a contradiction, that there does exist a nonzero polynomial $g \in \mathsf{F}[z]$ with degree less than $i + j$ such that g annihilates the sequence

$$c_0, c_1, \ldots, c_{2i-1+j}.$$

Then g also annihilates the shorter sequence

$$c_0, c_1, \ldots, c_{2i-2+j}$$

and it follows by Lemma 2.12 that g is divisible by f.

Let d be the degree of g, and let γ be the leading coefficient of g, so that $\gamma \neq 0$ and

$$g = \gamma z^{d-i} f + h,$$

where h is a polynomial with degree less than d that is also divisible by f. Since f is an annihilator of the initial sequence

$$c_0, c_1, \ldots, c_{2i-2+j}$$

and since h is a multiple of f, it follows by Lemma 2.9 that h and $-h$ are both annihilators of the initial sequence

$$c_0, c_1, \ldots, c_{2i-2+j}$$

as well. Since the degree of g is strictly greater than that of $-h$, it follows by Lemma 2.6 that the polynomial

$$g - h = \gamma z^{d-i} f$$

is an annihilator of the initial sequence

$$c_0, c_1, \ldots, c_{2i-1+j}$$

if g is. However, since f is an annihilator of the initial sequence

$$c_0, c_1, \ldots, c_{2i-2+j},$$

and since $\gamma \neq 0$ and the degree of the polynomial $z^{d-i} f$ is $d \leq i + j - 1 < 2i - 1 + j$, it now follows by a straightforward application of Definition 2.4 that f is an annihilator of the initial sequence

$$c_0, c_1, \ldots, c_{2i-1+j}$$

as well. This contradicts the choice of j.

There is, therefore, no nonzero polynomial $g \in \mathsf{F}[z]$ with degree less than $i + j$ that annihilates the sequence

$$c_0, c_1, \ldots, c_{2i-1+j}.$$

In order to see that H_{i+j} is nonsingular, suppose that

$$f = \alpha_0 + \alpha_1 z + \cdots + \alpha_{i-1} z^{i-1} + z^i$$

and consider the product

$$H_{i+j} \cdot \begin{bmatrix} 1 & & & 0 & \alpha_0 & 0 & \cdots & 0 \\ & 1 & & & \alpha_1 & \alpha_0 & \cdots & 0 \\ & & \ddots & & \vdots & \vdots & \ddots & \vdots \\ 0 & & & 1 & \alpha_{i-1} & \alpha_{i-2} & \cdots & \alpha_{i-j-1} \\ 0 & 0 & \cdots & 0 & 1 & \alpha_{i-1} & \cdots & \alpha_{i-j} \\ 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & \alpha_{i-j+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}, \tag{6}$$

where $\alpha_i = 1$ and $\alpha_j = 0$ if $j < 0$ or if $j > i$. The matrix on the right is a block matrix whose top left $i \times i$ block is the identity matrix, whose bottom left $j \times i$ block is a zero matrix, and whose right $(i + j) \times j$ block is a Toeplitz matrix whose columns are filled with the coefficients of f.

The product of these matrices can be considered as a block matrix as well. Its top left $i \times i$ block is the nonsingular matrix H_i. Its top right $i \times j$ block is a zero matrix, because f is an annihilator of the initial sequence

$$c_0, c_1, \ldots, c_{2i-2+j},$$

and its bottom $j \times j$ block is a Hankel matrix whose first column is

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \zeta \end{bmatrix}$$

where

$$\zeta = \alpha_0 c_{i+j-1} + \alpha_1 c_{i+j} + \cdots + \alpha_{i-1} c_{2i+j-2} + c_{2i+j-1}.$$

Now $\zeta \neq 0$, because f is an annihilator of the initial sequence

$$c_0, c_1, \ldots, c_{2i-2+j}$$

but f is not an annihilator of the initial sequence

$$c_0, c_1, \ldots, c_{2i-1+j}.$$

It follows that the bottom $j \times j$ block of the above product is a nonsingular Hankel matrix, because a lower triangular Toeplitz matrix with nonzero entries on the diagonal would be obtained if the order of its columns was reversed. Therefore the entire (block upper triangular) matrix product shown in Equation (6) is nonsingular. It follows that H_{i+j} is nonsingular, as well. $\square$

Lemma 2.14. *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 initial sequence has a minimal annihilator with degree m if and only if H_m is nonsingular.

Proof. Suppose first that H_m is nonsingular. Then it follows by Lemma 2.11 that the initial sequence

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

has a unique minimal annihilator, $f \in \mathsf{F}[z]$, and that the degree of f is m, as required.

Suppose, instead, that H_m is singular. It is necessary in this case to prove that the initial sequence

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

does not have a minimal annihilator with degree m at all.

Consider the case that H_i is singular for every integer i between 1 and m. Then, since the matrices $H_1, H_2, \ldots, H_m$ are all Hankel matrices, one can see by inspection of these matrices that

$$c_0 = c_1 = \cdots = c_{m-1} = 0,$$

so that

$$H_m = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & c_m \\ 0 & 0 & 0 & \cdots & c_{m+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & c_m & c_{m+1} & \cdots & c_{2m-2} \end{bmatrix}.$$

Now if

$$c_m = c_{m+1} = \cdots = c_{2m-1} = 0$$

as well, then the monic polynomial 1 with degree zero is an annihilator of the sequence

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

Since $0 < m$, this initial sequence does not have a minimal annihilator whose degree is m in this case.

Otherwise, c_i is nonzero for some integer i such that $m \le i \le 2m-1$. Consider the smallest integer i such that this is the case, so that $c_m = c_{m+1} = \cdots = c_{i-1} = 0$. An inspection of the linear system

$$H_m x = \begin{bmatrix} c_m \\ c_{m+1} \\ \vdots \\ c_{2m-1} \end{bmatrix}$$

reveals that this system is inconsistent, since the $i+1-m^{\text{th}}$ row of the matrix H_m is filled with zeroes and the corresponding entry in the vector on the right is nonzero. On the other hand, if the initial sequence

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

has a nonzero annihilator whose degree is at most m, then this annihilator can be used to form a solution for the above system. It follows that the initial sequence

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

does not have a minimal annihilator with degree m, in this case, because it does not have an annihilator with degree m at all.

In the remaining case there exists an integer i such that $1 \leq i < m$, H_i is nonsingular, and such that all of the minors $H_{i+1}, H_{i+2}, \ldots, H_m$ are singular. Since H_i is nonsingular it follows by Lemma 2.11, above, that the initial sequence

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

has a unique minimal annihilator $f \in \mathsf{F}[z]$, and the degree of f is i.

Suppose that f is also an annihilator of the sequence

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

Then, since $i < m$, it is clear that the above initial sequence does not have a minimal annihilator with degree m, once again.

It remains only to consider the case that f is not an annihilator of the initial sequence

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

Consider the smallest positive integer j such that f is not an annihilator of the initial sequence

$$c_0, c_1, \ldots, c_{2i-1+j}.$$

Then, of course, f is an annihilator of the shorter sequence

$$c_0, c_1, \ldots, c_{2i-2+j}.$$

Clearly $2i - 1 + j \leq 2m - 1$. On the other hand, Lemma 2.13 implies that the matrix H_{i+j} is nonsingular. It therefore follows by the choice of i that $m < i + j$.

Consider any nonzero polynomial $g \in \mathsf{F}[z]$ such that g is an annihilator of the initial sequence

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

As noted above, $2i - 1 + j \leq 2m - 1$, so that g is also an annihilator of the sequence

$$c_0, c_1, \ldots, c_{2i-1+j}.$$

It follows by Lemma 2.13 that the degree of g is at least $i + j$. Thus the degree of g is at least $m + 1$, since $m < i + j$ as noted above. Therefore the sequence

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

does not have a minimal annihilator with degree m in this remaining case, because the sequence does not have an annihilator with degree m, at all. $\square$

The above lemma will be used in the proof of Theorem 4.5, the main technical result of this report. The next result is needed to establish the correctness of the Berlekamp-Massey algorithm, which is discussed in the following section.

Lemma 2.15. *Suppose that*

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

is a linearly recurrent sequence whose entries are not all zero. Let $g \in \mathsf{F}[z]$ be 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 sequence.

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

Proof. Suppose that

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

is a linearly recurrent sequence whose entries are not all zero, as given in the lemma, and that the minimal polynomial g of this sequence has degree m. Consider the Hankel matrices $H_1, H_2, \ldots$ that are formed using the entries of the above sequence.

In order to obtain a contradiction, let us assume that H_m is singular.

If H_i is nonsingular for any integer $i > m$, then it follows by Lemma 2.11 that every monic annihilator of the initial sequence

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

has degree at least i. However, g is a monic annihilator of this initial sequence with degree $m < i$. We may conclude as a result of this contradiction that H_i is singular for every integer $i > m$.

Consider the case that H_i is singular for every positive integer $i < m$ as well. Then H_i is singular for every integer $i \geq 0$ and one can see, by an inspection of these matrices, that $c_i = 0$ for every integer $i \geq 0$ as well. Since it is given that that the entries of the sequence $c_0, c_1, c_2, \ldots$ are not all zero, we may conclude that at least one matrix H_i is nonsingular.

Considering the largest such matrix, one finds that there exists a positive integer $i < m$ such that H_i is nonsingular and such that H_{i+j} is singular for every integer $j > 0$.

It follows by Lemma 2.11 that the initial sequence

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

has a unique minimal annihilator $f \in \mathsf{F}[z]$, and the degree of f is i.

Now, since the minimal polynomial of the entire sequence has degree $m > i$ there must exist a positive integer j such that f is not an annihilator of the initial sequence

$$c_0, c_1, \ldots, c_{2i-1+j}.$$

If one considers the smallest positive integer j for which this is the case then it follows by an application of Lemma 2.13 that the matrix H_{i+j} is nonsingular. Since it has been established already that H_k is singular, for every integer $k \geq m$, it follows that j a positive integer such that $i + j \leq m$ and H_{i+j} is nonsingular. However, this contradicts the above choice of i.

A contradiction has now been obtained in every possible case. We may now conclude that H_m is nonsingular. It follows by Lemma 2.11 that the initial sequence

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

has a unique minimal annihilator, $h \in \mathsf{F}[z]$, and that h has degree m.

Since the minimal polynomial g of the entire sequence is also a monic polynomial in $\mathsf{F}[z]$ with degree m that annihilates the above initial sequence, the uniqueness of the minimal annihilator implies that $h = g$, as required to complete the proof. $\square$

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 for 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

$$g_1, g_2, \dots, g_{2n} \in \mathsf{F}[z]$$

of monic polynomials such that g_i is a minimal annihilator of the initial sequence

$$c_0, c_1, \dots, c_{i-1},$$

for $1 \leq i \leq 2n$.

Now suppose that the entire sequence has minimal polynomial $g \in \mathsf{F}[z]$ and let m be the degree of g. Then $m \leq n$ and it follows by Lemma 2.15, above, that $g = g_{2m}$. Furthermore, since g is both the only monic annihilator of the initial sequence

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

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

$$c_0, c_1, \dots, c_i$$

with degree at most m, for every integer $i \geq 2m-1$ as well. Thus g is the only minimal annihilator for the above initial sequence (again, for $i \geq 2m-1$), and it follows that

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

The final polynomial, g_{2n} that is generated can therefore be returned 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 $c_0, c_1, c_2, \dots$ has been generated before all of the initial $2n$ entries of this sequence have been considered.

Once again, let us consider a fixed sequence $c_0, c_1, c_2, \dots$, and suppose that a given value n is greater than or equal to the degree of the minimal polynomial of this sequence. Let $g_1, g_2, \dots, g_{2n}$ be the sequence of polynomials generated by the Berlekamp-Massey algorithm when it is given $c_0, c_1, \dots, c_{2n-1}$ and n as input. For the purposes of the following definitions, set $g_{-1} = 0$ and $g_0 = 1$.

Definition 3.1. Let i be an integer such that $0 \leq i \leq 2n-1$. Then the given sequence $c_0, c_1, c_2, \dots$ has a *zero-discrepancy at position* i if $g_i = g_{i+1}$.

Definition 3.2. Let i and j be integers such that $i \geq 0$, $j \geq 1$, and $i + j < 2n$.

Then the given sequence $c_0, c_1, c_2, \dots$ has a *harmful sequence of zero-discrepancies of length* j *beginning at position* i if

$$g_{i-1} \neq g_i = g_{i+1} = \dots = g_{i+j} \tag{7}$$

and g_i is not equal to the minimal polynomial of the linearly recurrent sequence $c_0, c_1, c_2, \dots$.

We will say that the given sequence 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 $i \geq 0$ and $i + j < 2n$.

Consider again the matrix H shown in Equation (3). Henceforth, we will let Δ_i denote the determinant of the i^{th} principal minor H_i of this matrix if $1 \leq i \leq n$, and we will set $\Delta_0 = 1$.

Definition 3.3. Let i and j be integers such that $i \geq 0$, $j \geq 1$, and such that $i + j < n$. Then the above matrix H has a *harmful gap of length j beginning at position i* if

$$\Delta_i \neq 0 = \Delta_{i+1} = \Delta_{i+2} = \cdots = \Delta_{i+j}, \tag{8}$$

but $\Delta_k \neq 0$ for some integer $k > i + j$.

We will say that the matrix H has a *harmful gap of length j* if it has a harmful gap of length j beginning at position i, for some integer i such that $i \geq 0$ and $i + j < n$.

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 $f \in \mathsf{F}[z]$, such that $f(A)b = 0$, forms an ideal in $\mathsf{F}[z]$. Since the characteristic polynomial of A is an element of this ideal, the ideal is nonzero. There is, therefore, a monic polynomial $f \in \mathsf{F}[z]$ that generates this ideal. This polynomial is also the "minimal polynomial" of A, and b, as defined below.

Definition 4.1. A polynomial $f \in \mathsf{F}[z]$ is the *minimal polynomial* of a matrix $A \in \mathsf{F}^{N \times N}$ and vector $b \in \mathsf{F}^{N \times 1}$ if f is monic, $f(A)b = 0$, and if $g(A)b = 0$ if and only if g is divisible by f, for every polynomial $g \in \mathsf{F}[z]$.

As suggested in the introduction, the minimal polynomial of A and b is also the monic polynomial f with least degree such that $f(A)b = 0$.

Suppose, now, that the minimal polynomial f of A and b has degree n and that

$$f = \alpha_0 + \alpha_1 z + \cdots + \alpha_{n-1} z^{n-1} + z^n. \tag{9}$$

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

$$c_j = u^T A^j b \qquad \text{for } j \geq 0. \tag{10}$$

Note that if $i \geq 0$ then

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

Thus the condition given in Definition 2.1 is satisfied, so that f is a characteristic polynomial (although, not necessarily the minimal polynomial) of the sequence $c_0, c_1, c_2, \ldots$, and this sequence is linearly recurrent.

4.1 Identification of the Probability Distribution

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. Let $f \in \mathsf{F}[z]$ be as given in Equation (9), above. Then it follows by Definition 2.1 that if the initial n entries

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

of a linearly recurrent sequence are given, along with the information that f is a characteristic polynomial of this sequence, then the remaining entries

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

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. This implies 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 4.2. *Let $A \in \mathsf{F}^{N\times N}$, let $b \in \mathsf{F}^{N\times 1}$, and suppose that $f \in \mathsf{F}[z]$ is the minimal polynomial of A and b. Let n be the degree of f.*

Let $s_0, s_1, s_2, \dots$ be any linearly recurrent sequence with entries in F 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.*

Proof of Lemma 4.2. Let A, b, f, and n be as given in the statement of the lemma. Then, since f is the minimal polynomial of A and b, the subspace of $\mathsf{F}^{N\times 1}$ that is spanned by the vectors

$$b, Ab, A^2b, \dots$$

has dimension n, and the vectors

$$b, Ab, A^2b, \dots, A^{n-1}b$$

form a basis for this subspace.

Let us add vectors $y_1, y_2, \dots, y_{N-n} \in \mathsf{F}^{N\times 1}$ to complete a basis for $\mathsf{F}^{N\times 1}$.

This basis has a dual orthogonal basis. That is, there exists another basis for $\mathsf{F}^{N\times 1}$ consisting of vectors

$$v_0, v_1, \dots, v_{n-1}, w_1, w_2, \dots, w_{N-n} \in \mathsf{F}^{N\times 1}$$

such that

$$v_i^T A^j b = \begin{cases} 1 & \text{if } 0 \leq i,j \leq n-1 \text{ and } i = j, \\ 0 & \text{if } 0 \leq i,j \leq n-1 \text{ and } i \neq j, \end{cases}$$

$$w_i^T y_j = \begin{cases} 1 & \text{if } 1 \leq i,j \leq N-n \text{ and } i = j, \\ 0 & \text{if } 1 \leq i,j \leq N-n \text{ and } i \neq j, \end{cases}$$

and such that $v_i^T y_j = 0 = w_j^T A^i b$ for all integers i and j such that $0 \le i \le n-1$ and $1 \le j \le N-n$.

Now let u be a uniformly selected vector in $\mathsf{F}^{N\times 1}$, as given in the statement of the lemma. This vector can be chosen as a random linear combination of the elements of any basis that we wish to choose. In particular, u can be chosen as

$$u = \gamma_0 v_0 + \gamma_1 v_1 + \cdots + \gamma_{n-1} v_{n-1} + \delta_1 w_1 + \delta_2 w_2 + \cdots + \delta_{N-n} w_{N-n} \tag{11}$$

where the values $\gamma_0, \gamma_1, \ldots, \gamma_{n-1}$ and $\delta_1, \delta_2, \ldots, \delta_{N-n}$ are chosen uniformly and independently from the finite field F.

Now it suffices to notice that if $0 \le j \le n-1$ then

$$c_j = u^T A^j b = \gamma_j v_j^T A^j b = \gamma_j,$$

by the choice of u (as given in Equation (11)) and the fact that $v_0, v_1, \ldots, v_{n-1}, w_1, w_2, \ldots, w_{N-n}$ was chosen as a dual orthogonal basis for the original basis

$$b, Ab, A^2 b, \ldots, A^{n-1} b, y_1, y_2, \ldots, y_{N-n}.$$

Consequently, if $s_0, s_1, s_2, \ldots$ is any given linearly recurrent sequence with characteristic polynomial f, then the probability that

$$c_j = s_j \qquad \text{for } 0 \le j \le n-1$$

is the same as the probability that

$$\gamma_j = s_j \qquad \text{for } 0 \le j \le n-1.$$

Since the values $\gamma_0, \gamma_1, \ldots, \gamma_{n-1}$ are uniformly and independently selected from $\mathsf{F} = \mathsf{F}_q$, this probability is q^{-n}.

Recall that the sequences

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

and

$$s_0, s_1, s_2, \ldots$$

are linearly recurrent sequences with characteristic polynomial f. Since the remaining terms of each sequence are determined by the initial n entries of the sequence, it follows that if $c_j = s_j$ for $0 \le j \le n-1$, then $c_j = s_j$ for every integer $j \ge n$ as well. This completes the proof. $\square$

4.2 Bounding the Probability that Harmful Sequences are Long

The next result follows by an application of the theory of subresultants.

Lemma 4.3. *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 $m \le n$ 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 $m \le s \le 2n$.*

Let $c_0, c_1, c_2, \ldots$ be a uniformly chosen linearly recurrent sequence 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.

Proof. Suppose, once again, that

$$f = \alpha_0 + \alpha_1 z + \cdots + \alpha_{n-1} z^{n-1} + \alpha_n z^n \in \mathsf{F}[z]$$

where $\alpha_n = 1$, so that f is a monic polynomial with degree n.

Suppose also that $m \leq n$ and that

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

where $\beta_m = 1$, so that g is a monic polynomial with degree m.

Suppose as well that the greatest common divisor h of f and g has degree k, as in the statement of the lemma.

Before establishing the claim in the lemma it will be useful to consider a related result. We will therefore initially consider sequences of values

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

that are uniformly and independently selected from F.

Let

$$\vec{c} = \begin{bmatrix} c_{2n-1} \\ \vdots \\ c_1 \\ c_0 \end{bmatrix} \in \mathsf{F}^{2n \times 1}$$

be the vector whose elements are the given values $c_0, c_1, \ldots, c_{2n-1}$, listed in reverse order.

Now let s be an integer such that $m \leq s \leq 2n$, and consider the system of linear equations

$$M_{f,g,s} \vec{c} = 0$$

that expresses the condition that $c_0, c_1, \ldots, c_{2n-1}$ are the initial entries of a linearly recurrent sequence satisfying the condition

$$\mathsf{Ann}[g, s] \wedge \mathsf{CharPol}[f].$$

The matrix $M_{f,g,s}$ has $n+s-m$ rows and $2n$ columns. Its top n rows correspond to the condition $\mathsf{CharPol}[f]$ and consist of the Toeplitz matrix

$$\begin{bmatrix} \alpha_n & \alpha_{n-1} & \cdots & \alpha_0 & & & \\ & \alpha_n & \cdots & \alpha_1 & \alpha_0 & & \\ & & \ddots & & & \ddots & \\ & & & \alpha_n & \alpha_{n-1} & \cdots & \alpha_0 \end{bmatrix},$$

and its bottom $s-m$ rows form the Toeplitz matrix

$$\begin{bmatrix} 0 & \cdots & 0 & \beta_m & \beta_{m-1} & \cdots & \beta_0 & & & \\ 0 & \cdots & 0 & & \beta_m & \cdots & \beta_1 & \beta_0 & & \\ & \vdots & & & & \ddots & & & \ddots & \\ 0 & \cdots & 0 & & & & \beta_m & \beta_{m-1} & \cdots & \beta_0 \end{bmatrix}$$

whose leftmost $2n-s$ columns are filled with zeroes. This bottom part of the matrix corresponds to the condition $\mathsf{Ann}[g, s]$.

Notice, by the way, that if

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

are the initial entries of a linearly recurrent sequence satisfying the condition

$$\mathsf{Ann}[g, s] \wedge \mathsf{CharPol}[f]$$

then there is *exactly one* linearly recurrent sequence that satisfies the above condition and that begins with the above initial sequence of values. After all, since $s \leq 2n$, the condition $\mathsf{Ann}[g, s]$ is independent of the choice of any additional values $c_{2n}, c_{2n+1}, c_{2n+2}, \ldots$ that might be used to extend the sequence. On the other hand, since f has degree n, there is exactly one way to choose these additional values in order to produce a sequence with characteristic polynomial f.

Consider once again the matrix $M_{f,g,s}$ that has been described above, and consider the special case

$$s = n + m - k.$$

In this case, $M_{f,g,s}$ can be considered to be a block upper triangular matrix,

$$M_{f,g,s} = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}.$$

The top left block, A, is an upper triangular $(2n-s) \times (2n-s)$ matrix with ones on the diagonal (since $\alpha_n = 1$), so that the rank of the entire matrix $M_{f,g,s}$ is the sum of $2n-s$ and the rank of the bottom right block, C.

The rank of C can be discovered by an application of the theory of subresultants. The text of von zur Gathen and Gerhard [7] includes a readable introduction to this theory and additional references. It will be used in order to continue this argument.

In particular, a comparison of the dimensions of the matrices given above confirms that the matrix C has $(n+s-m) - (2n-s) = 2s-n-m$ rows and $2n - (2n-s) = s$ columns. That is, since $s = n+m-k$, it has $n+m-2k$ rows and $n+m-k$ columns. The square matrix obtained by including all the rows of C, and the leftmost $n+m-2k$ columns, is the transpose of the matrix shown at the bottom of page 169 of von zur Gathen and Gerhard's text and whose determinant is denoted there as S_k.

Since k is the degree of the greatest common divisor of f and g, S_k is nonzero (see Corollary 6.49 on page 170 of von zur Gathen and Gerhard's text). Consequently both C, and the entire matrix $M_{f,g,s}$, have full rank in this case. In particular, the rows of the matrix $M_{f,g,s}$ are linearly independent so that the rank of $M_{f,g,s}$ is equal to the number of rows, $n+s-m$.

Consider the right nullspace of this matrix. Since this $(n+s-m) \times 2n$ matrix has full rank, the dimension of this nullspace is

$$2n - (n+s-m) = n+m-s = n+m-(n+m-k) = k.$$

It follows that there are exactly q^k initial sequences

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

of linearly recurrent sequences that satisfy the condition

$$\mathsf{Ann}[g, s] \wedge \mathsf{CharPol}[f].$$

As previously noted, no two distinct linearly recurrent sequences with characteristic polynomial f have the same initial sequence with length $2n$. It follows that there are exactly q^k linearly recurrent sequences that satisfy the condition

$$\mathsf{Ann}[g, s] \wedge \mathsf{CharPol}[f]$$

as well.

Since the polynomial h has degree k, there are also exactly q^k linearly recurrent sequences that satisfy the condition

$$\mathsf{CharPol}[h].$$

Every such sequence satisfies the condition

$$\mathsf{Ann}[g, s] \wedge \mathsf{CharPol}[f]$$

as well, since h divides both g and f. It follows that if $s = n + m - k$ then the conditions

$$\mathsf{Ann}[g, s] \wedge \mathsf{CharPol}[f] \qquad \text{and} \qquad \mathsf{CharPol}[h]$$

are equivalent.

Suppose next that $s < n + m - k$, and compare the matrix $M_{f,g,s}$ to the matrix $M_{f,g,n+m-k}$ whose properties have been investigated above. The set of rows of $M_{f,g,s}$ is a subset of the set of rows of $M_{f,g,n+m-k}$, so that the rows of $M_{f,g,s}$ are linearly independent because the rows of $M_{f,g,n+m-k}$ are. The rank of $M_{f,g,s}$ is therefore the same as the number of rows in this matrix, namely,

$$n + s - m.$$

Consequently the dimension of the right nullspace of this matrix is $2n - (n + s - m) = n + m - s$. Using an argument similar to the one given above, for the case $s = n+m-k$, we may now conclude that there are exactly q^{n+m-s} linearly recurrent sequences satisfying the condition

$$\mathsf{Ann}[g, s] \wedge \mathsf{CharPol}[f]$$

whenever $s < n + m - k$.

Finally, consider the case that $s > n + m - k$. Notice that the condition

$$\mathsf{Ann}[g, s] \wedge \mathsf{CharPol}[f]$$

implies the condition

$$\mathsf{Ann}[g, n + m - k] \wedge \mathsf{CharPol}[f]$$

and it has been established that this is equivalent to the condition $\mathsf{CharPol}[h]$. Thus

$$(\mathsf{Ann}[g, s] \wedge \mathsf{CharPol}[f]) \Rightarrow \mathsf{CharPol}[h]$$

in this case. On the other hand, if h is a characteristic polynomial of a given sequence then the sequence must also satisfy the condition

$$\mathsf{Ann}[g, s] \wedge \mathsf{CharPol}[f],$$

since h divides both g and f. That is,

$$\mathsf{CharPol}[h] \Rightarrow (\mathsf{Ann}[g, s] \wedge \mathsf{CharPol}[f])$$

as well, so that the conditions

$$\mathsf{Ann}[g, s] \wedge \mathsf{CharPol}[f] \qquad \text{and} \qquad \mathsf{CharPol}[h]$$

are equivalent, and there are exactly q^k linearly recurrent sequences that satisfy the condition

$$\mathsf{Ann}[g, s] \wedge \mathsf{CharPol}[f]$$

in this case.

In summary, it has been shown that the number of linearly recurrent sequences that satisfy the condition

$$\mathsf{Ann}[g, s] \wedge \mathsf{CharPol}[f]$$

is q^{n+m-s} if $s < n+m-k$. It has also been shown that the number of linearly recurrent sequences that satisfy the above condition is q^k if $s \geq n+m-k$, and that the conditions

$$\mathsf{Ann}[g, s] \wedge \mathsf{CharPol}[f] \qquad \text{and} \qquad \mathsf{CharPol}[h]$$

are equivalent when $s \geq n+m-k$, as well.

Now let us consider the probability distribution that is discussed in the lemma. That is, consider a uniformly selected linearly recurrent sequence

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

with characteristic polynomial f. Since there are exactly q^n such sequences, the probability that the condition

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

is satisfied, under this distribution, is equal to the quotient of the number of linearly recurrent sequences that satisfy the condition

$$\mathsf{Ann}[g, s] \wedge \mathsf{CharPol}[f]$$

and the number, q^n, of all linearly recurrent sequences with characteristic polynomial f. The probabilities given in the statement of the lemma now follow.

It also follows that if $s \geq n+m-k$ and one considers linearly recurrent sequences with characteristic polynomial f, then the conditions

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

are equivalent — for the latter condition implies the former, and each occurs with the same probability. $\square$

The above lemma will be used to prove Lemma 4.4.

Lemma 4.4. *Let f, g, h, n, m, and k be as in the statement of the previous lemma. Suppose, once again, that*

$$c_0, c_1, c_2, \dots$$

is a uniformly chosen linearly recurrent sequence with characteristic polynomial f. Let s and t be nonnegative integers such that $m \le s \le s+t \le 2n$, and let ϵ be a positive real number. Then either

$$\text{Prob}\,(\mathsf{MinAnn}[g, s]) \le \epsilon \tag{12}$$

or

$$\frac{\text{Prob}\,(\mathsf{MinAnn}[g, s+t] \wedge \neg\mathsf{MinPol}[g])}{\text{Prob}\,(\mathsf{MinAnn}[g, s])} \le q^{m-s-t}/\epsilon. \tag{13}$$

Proof. Suppose, first, that $s + t \ge n + m - k$. We will show that

$$\text{Prob}\,(\mathsf{MinAnn}[g, s+t] \wedge \neg\mathsf{MinPol}[g]) = 0$$

in this case.

First suppose that $g \ne h$. Then, since $s + t \ge n + m - k$,

$$\begin{aligned} \mathsf{MinAnn}[g, s+t] &\Rightarrow \mathsf{Ann}[g, s+t] \\ &\Rightarrow \mathsf{CharPol}[h] && \text{(by Lemma 4.3)} \\ &\Rightarrow \mathsf{Ann}[h, s+t] \\ &\Rightarrow \neg\mathsf{MinAnn}[g, s+t], \end{aligned}$$

since the degree of h is less than that of g if $h \ne g$. Thus

$$\text{Prob}\,(\mathsf{MinAnn}[g, s+t] \wedge \neg\mathsf{MinPol}[g]) = 0$$

in this case.

On the other hand, if $s + t \ge n + m - k$ and $g = h$, then it follows by Lemma 4.3 that

$$\mathsf{MinAnn}[g, s+t] \Rightarrow \mathsf{Ann}[g, s+t] \Rightarrow \mathsf{CharPol}[h]$$

once again. Suppose that the condition $\mathsf{MinAnn}[g, s+t]$ is satisfied, and that h is not the minimal polynomial of the given linearly recurrent sequence. Then, since h is a characteristic polynomial of the sequence, some divisor $\widehat{h}$ of h, that is not equal to h, must be the minimal polynomial of the sequence instead. However, this implies that $\widehat{h}$ is also an annihilator of the initial sequence

$$c_0, c_1, c_2, \dots, c_{s+t-1}$$

so that the condition $\mathsf{MinAnn}[g, s+t]$ is not satisfied, after all. Consequently (since $g = h$)

$$\mathsf{MinAnn}[g, s+t] \Rightarrow \mathsf{MinPol}[g]$$

in this case, and

$$\text{Prob}\,(\mathsf{MinAnn}[g, s+t] \wedge \neg\mathsf{MinPol}[g]) = 0$$

once again.

It follows that if $s + t \ge n + m - k$ then one of conditions (12) or (13) must be satisfied — for either

$$\text{Prob}\,(\mathsf{MinAnn}[g, s]) = 0,$$

implying condition (12), or

$$\text{Prob}\left(\mathsf{MinAnn}[g, s]\right) > 0,$$

in which case the ratio shown in condition (13) is equal to zero and condition (13) is satisfied instead.

It remains only to consider the case that $s + t < n + m - k$. If this is the case, and

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

then condition (12) is satisfied. It is therefore sufficient to consider the case that $s + t < n + m - k$ and

$$\text{Prob}\left(\mathsf{MinAnn}[g, s]\right) > \epsilon.$$

However, this implies that

$$\begin{aligned}
\frac{\text{Prob}\left(\mathsf{MinAnn}[g, s+t] \wedge \neg\mathsf{MinPol}[g]\right)}{\text{Prob}\left(\mathsf{MinAnn}[g, s]\right)} &\leq \frac{\text{Prob}\left(\mathsf{MinAnn}[g, s+t] \wedge \neg\mathsf{MinPol}[g]\right)}{\epsilon} \\
&\qquad\qquad \text{(since Prob}\left(\mathsf{MinAnn}[g, s]\right) > \epsilon) \\
&\leq \frac{\text{Prob}\left(\mathsf{MinAnn}[g, s+t]\right)}{\epsilon} \\
&\leq \frac{\text{Prob}\left(\mathsf{Ann}[g, s+t]\right)}{\epsilon} \\
&\leq q^{m-s-t}/\epsilon \qquad \text{(by Lemma 4.3, once again).}
\end{aligned}$$

Thus one or the other of conditions (12) and (13) is satisfied in every case. ☐

These results can be used to prove the following theorem.

Theorem 4.5. *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 linearly recurrent 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 $2q^{-t/2}$.

Proof. To begin, recall that a harmful gap of length t can only begin at position m if

$$\Delta_m \neq 0 = \Delta_{m+1} = \Delta_{m+2} = \cdots = \Delta_{m+t}. \tag{14}$$

Since Δ_m is the determinant of H_m, the event whose probability we wish to bound can only occur if the matrix H_m is nonsingular.

It follows by Lemma 2.11 that if $\Delta_m \neq 0$ then there is exactly one monic polynomial $g \in \mathsf{F}[x]$ with degree m such that g is a minimal annihilator of the initial sequence

$$c_0, c_1, c_2, \ldots, c_{2m-1}.$$

On the other hand, if $\Delta_m = 0$ then Lemma 2.14 implies that there is no minimal annihilator for the above initial sequence with degree m at all. Therefore

$$\text{Prob}\left(\Delta_m \neq 0\right) = \sum_{\substack{g \in \mathsf{F}[x] \\ g \text{ is monic} \\ \deg(g) = m}} \text{Prob}\left(\mathsf{MinAnn}[g, 2m]\right). \tag{15}$$

Notice that if $\Delta_m \neq 0$ and $g \in \mathsf{F}[x]$ is the minimal annihilator of the initial sequence

$$c_0, c_1, c_2, \ldots, c_{2m-1},$$

then Lemma 2.13 can be used to show that the given linearly recurrent sequence has a harmful gap of length t, beginning at position m, if and only if the condition

$$\mathsf{MinAnn}[g, 2m+t] \wedge \neg\mathsf{MinPol}[g]$$

is satisfied. It follows that the probability that the given sequence has a harmful gap of length t, beginning at position m, is $\alpha(m)$, where

$$\alpha(m) = \sum_{\substack{g \in \mathsf{F}[x] \\ g \text{ is monic} \\ \deg(g)=m}} \mathrm{Prob}\left(\mathsf{MinAnn}[g, 2m+t] \wedge \neg\mathsf{MinPol}[g]\right).$$

Now let ϵ be a positive real number. Break the above sum into two pieces,

$$\alpha(m) = \beta(m) + \gamma(m),$$

where

$$\beta(m) = \sum_{\substack{g \in \mathsf{F}[x] \\ g \text{ is monic} \\ \deg(g)=m \\ \mathrm{Prob}(\mathsf{MinAnn}[g,2m])>\epsilon}} \mathrm{Prob}\left(\mathsf{MinAnn}[g, 2m+t] \wedge \neg\mathsf{MinPol}[g]\right)$$

and

$$\gamma(m) = \sum_{\substack{g \in \mathsf{F}[x] \\ g \text{ is monic} \\ \deg(g)=m \\ \mathrm{Prob}(\mathsf{MinAnn}[g,2m])\leq\epsilon}} \mathrm{Prob}\left(\mathsf{MinAnn}[g, 2m+t] \wedge \neg\mathsf{MinPol}[g]\right)$$

It follows by Lemma 4.4 (with $s = 2m$) that

$$\begin{aligned}
\beta(m) &= \sum_{\substack{g \in \mathsf{F}[x] \\ g \text{ is monic} \\ \deg(g)=m \\ \mathrm{Prob}(\mathsf{MinAnn}[g,2m])>\epsilon}} \mathrm{Prob}\left(\mathsf{MinAnn}[g, 2m+t] \wedge \neg\mathsf{MinPol}[g]\right) \\
&\leq \sum_{\substack{g \in \mathsf{F}[x] \\ g \text{ is monic} \\ \deg(g)=m \\ \mathrm{Prob}(\mathsf{MinAnn}[g,2m])>\epsilon}} \left(q^{m-2m-t}/\epsilon\right) \mathrm{Prob}\left(\mathsf{MinAnn}[g, 2m]\right) \\
&\leq \sum_{\substack{g \in \mathsf{F}[x] \\ g \text{ is monic} \\ \deg(g)=m}} \left(q^{m-2m-t}/\epsilon\right) \mathrm{Prob}\left(\mathsf{MinAnn}[g, 2m]\right) \\
&= \left(q^{-m-t}/\epsilon\right) \sum_{\substack{g \in \mathsf{F}[x] \\ g \text{ is monic} \\ \deg(g)=m}} \mathrm{Prob}\left(\mathsf{MinAnn}[g, 2m]\right) \\
&= \left(q^{-m-t}/\epsilon\right) \mathrm{Prob}\left(\Delta_m \neq 0\right) && \text{(by Equation (15), above)}
\end{aligned}$$

$$\leq q^{-m-t}/\epsilon.$$

On the other hand,

$$\begin{aligned}
\gamma(m) &= \sum_{\substack{g\in \mathsf{F}[x] \\ g \text{ is monic} \\ \deg(g)=m \\ \mathrm{Prob}(\mathsf{MinAnn}[g,2m])\leq\epsilon}} \mathrm{Prob}\left(\mathsf{MinAnn}[g, 2m+t] \wedge \neg \mathsf{MinPol}[g]\right) \\
&\leq \sum_{\substack{g\in \mathsf{F}[x] \\ g \text{ is monic} \\ \deg(g)=m \\ \mathrm{Prob}(\mathsf{MinAnn}[g,2m])\leq\epsilon}} \mathrm{Prob}\left(\mathsf{MinAnn}[g, 2m]\right) \\
&\leq \sum_{\substack{g\in \mathsf{F}[x] \\ g \text{ is monic} \\ \deg(g)=m \\ \mathrm{Prob}(\mathsf{MinAnn}[g,2m])\leq\epsilon}} \epsilon \\
&= \epsilon \sum_{\substack{g\in \mathsf{F}[x] \\ g \text{ is monic} \\ \deg(g)=m \\ \mathrm{Prob}(\mathsf{MinAnn}[g,2m])\leq\epsilon}} 1 \\
&\leq \epsilon \sum_{\substack{g\in \mathsf{F}[x] \\ g \text{ is monic} \\ \deg(g)=m}} 1 \\
&= q^m \epsilon,
\end{aligned}$$

since there are exactly q^m monic polynomials $g \in \mathsf{F}[x]$ with degree m.

Thus

$$\alpha(m) = \beta(m) + \gamma(m) \leq q^{-m-t}/\epsilon + q^m \epsilon.$$

Finally, let $\epsilon = q^{-m-t/2}$. Then

$$q^{-m-t}/\epsilon = q^m \epsilon = q^{-t/2},$$

so that

$$\alpha(m) \leq q^{-m-t}/\epsilon + q^m \epsilon = 2q^{-t/2}$$

as claimed. □

Corollary 4.6. *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 linearly recurrent sequences with characteristic polynomial f. Then the probability that the corresponding Hankel matrix H has a harmful gap of length t is at most $2(n-t)q^{-t/2}$.

Proof. If a linearly recurrent sequence $c_0, c_1, \ldots$ has a characteristic polynomial f with degree n (so that H_i is singular for every integer $i > n$, by Lemmas 2.14 and 2.15) then the Hankel matrix H that corresponds to this sequence can only have a harmful gap of length t if this gap begins at position i, for $0 \leq i < n - t$.

The event considered in the above corollary is therefore the union of $n-t$ of the events considered in Theorem 4.5. The bound given in this corollary now follows. □

The next result will be used to bound the probability of a long sequence of zero-discrepancies.

Lemma 4.7. *Suppose that a linearly recurrent sequence*

$$c_0, c_1, c_2$$

has a harmful sequence of zero-discrepancies of length t. Then the Hankel matrix corresponding to this linearly recurrent sequence has a harmful gap of length $\lceil t/2 \rceil$.

Proof. Consider the sequence of polynomials

$$g_1, g_2, g_3, \ldots$$

that are generated by the Berlekamp-Massey algorithm using the given sequence $c_0, c_1, c_2, \ldots$ as input and, as before, set $g_{-1} = 0$ and $g_0 = 1$. Suppose that this sequence has a harmful sequence of zero-discrepancies of length t. Then this must begin at position s, for some integer $s \geq 0$, and it follows by Definition 3.2 that

$$g_{s-1} \neq g_s = g_{s+1} = \cdots = g_{s+t} \tag{16}$$

but that g_s is not the minimal polynomial of the entire sequence $c_0, c_1, c_2, \ldots$.

Suppose that the minimal polynomial of the sequence has degree $\widehat{u}$. Then it follows by Lemma 2.15 that $H_{2\widehat{u}}$ is nonsingular and that the minimal polynomial of the sequence is the unique monic polynomial with minimal degree that annihilates the initial sequence

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

In this case, it must also be the unique monic polynomial with minimal degree that annihilates the initial sequence

$$c_0, c_1, \ldots, c_k$$

for every integer $k \geq 2\widehat{u} - 1$, as well. It follows that $s + t < 2\widehat{u}$ — for, otherwise, it would be true that $g_s = g_{s+t}$ is the minimal polynomial of the entire sequence, after all.

If we set $\widehat{\ell} = 0$ and recall that Δ_0 has been defined to be 1 then, since $s \geq 0$, it follows that $\widehat{\ell}$ and $\widehat{u}$ are nonnegative integers such that $2\widehat{\ell} \leq s$, $s + t < 2\widehat{u}$, and such that $\Delta_{\widehat{\ell}}$ and $\Delta_{\widehat{u}}$ are both nonzero.

Suppose, now, that $\Delta_i = 0$ for every integer i such that $s < 2i \leq s + t$. Set ℓ to be the largest integer such that $s \geq 2\ell$ and $\Delta_\ell \neq 0$, and set u to be the smallest integer such that $s + t < 2u$ and such that $\Delta_u \neq 0$. Then, one can apply Lemmas 2.12 and 2.13 to see that

$$g_{2\ell} = g_{2\ell+1} = \cdots = g_{\ell+u-1},$$

and that $g_{2\ell}$ has degree ℓ, while the polynomials

$$g_{\ell+u}, g_{\ell+u+1}, \ldots, g_{2u}$$

all have degree u. In particular, one can see that this is the case by an application of the above lemmas for the case that $j = u - \ell$. It now follows, by an inspection of Equation (16) above, that either $2\ell \leq s \leq s+t \leq \ell + u - 1$, or that $\ell + u \leq s \leq s + t \leq 2u$. It follows in either case that

$$t = (s+t) - s \leq u - \ell,$$

so that the given sequence has a harmful gap of length t. This clearly implies that it has a harmful gap of length $\lceil t/2 \rceil$.

Suppose next that there is exactly one integer i such that $s < 2i \leq s + t$ and such that $\Delta_i \neq 0$. As above, let ℓ the the largest integer such that $s \geq 2\ell$ and $\Delta_\ell \neq 0$, and let u be the smallest integer such that $s + t < u$ and such that $\Delta_u \neq 0$. Now Lemmas 2.12 and 2.13 can be applied, once again, in order to see that

$$g_{2\ell} = g_{2\ell+1} = \cdots = g_{\ell+i-1},$$

and that $g_{2\ell}$ has degree ℓ. Furthermore, the polynomials

$$g_{\ell+i}, g_{\ell+i+1} + \cdots + g_{2i}$$

all have degree i;

$$g_{2i} = g_{2i+1} = \cdots = g_{i+u-1};$$

and the polynomials

$$g_{i+u}, g_{i+u+1}, \ldots, g_{2u}$$

all have degree u. One can conclude from these relationships and Equation (16) that $\ell + i \leq s \leq s + t \leq i + u - 1$. Thus

$$t = (s+t) - s \leq (i+u) - (\ell - i) = (i - \ell) + (u - i).$$

Therefore at least one of $\ell - i$ or $i - \ell$ must be greater than or equal to $\lceil t/2 \rceil$, implying that the given sequence has a harmful gap of length $\lceil t/2 \rceil$, once again.

In the only remaining case, there must be at least two distinct integers, i_1 and i_2, such that $s < 2i_1 < 2i_2 \leq s + t$ and such that Δ_{i_1} and Δ_{i_2} are nonzero. However, it follows by Lemma 2.14 that g_{i_1} has degree i_1 and that g_{i_2} has degree i_2. Thus $g_{i_1} \neq g_{i_2}$, contracting Equation (16), above. Therefore, this case cannot arise.

The given sequence therefore has a harmful gap of length $\lceil t/2 \rceil$ in every case that can possibly arise, as is required to establish the lemma. $\square$

The next result is a consequence of Corollary 4.6 and Lemma 4.7.

Corollary 4.8. *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 linearly recurrent 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 $2(n - t/2)q^{-t/4}$.

4.3 Bounding the Probability that Harmful Sequences are Short

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 4.9. *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*

$$c_0, c_1, c_2, \dots$$

is uniformly chosen from the set of linearly recurrent 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 $e^{-n/(2tq^t)}$.

Proof. Let f and n be as given in the statement of the theorem.

Recall that there are q^n linearly recurrent sequences

$$c_0, c_1, c_2, \dots$$

with entries in the field $\mathsf{F} = \mathsf{F}_q$ with minimal polynomial f, and that there is exactly one such sequence that begins with the initial sequence

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

for any choice of values $c_0, c_1, \dots, c_{n-1} \in \mathsf{F}$. It follows that if a linearly recurrent sequence is uniformly selected from the set of sequences with characteristic polynomial f, then the first n entries of this sequence are uniformly and independently chosen from F. That is, the probability distributions are the same.

Suppose now that t is a positive integer such that $n \geq 2t$. Partition the first $\lfloor n/t \rfloor \cdot t \leq n$ entries of a given sequence into $\lfloor n/t \rfloor$ blocks that each include t consecutive coefficients of the sequence, so that the i^{th} block consists of the coefficients

$$c_{(i-1)t}, c_{(i-1)t+1}, \dots, c_{it-1}$$

for $1 \leq i \leq \lfloor n/t \rfloor$.

Consider the event, E_i, that

$$g_{(i-1)t} = g_{(i-1)t+1} = \cdots = g_{it-1} \neq g_{it}.$$

The polynomial $g_{(i-1)t}$ certainly does depend on coefficients c_j for $j < (i-1)t$. However, the above event E_i does not — for every possible choice of $g_{(i-1)t}$, there is exactly one way to choose the entries

$$c_{(i-1)t}, c_{(i-1)t+1}, \dots, c_{it-2}$$

so that

$$g_{(i-1)t} = g_{(i-1)t+1} = \cdots = g_{it-1},$$

and there are then exactly $q-1$ ways to choose the next entry, c_{it-1}, in order to achieve the condition that $g_{it-1} \neq g_{it}$.

Using this observation, one can establish that the events

$$E_1, E_2, \dots, E_{\lfloor n/t \rfloor}$$

are mutually independent, so that their negations

$$\neg E_1, \neg E_2, \ldots, \neg E_{\lfloor n/t \rfloor}$$

are mutually independent as well. It also follows that

$$\mathrm{Prob}\,(E_i) = \frac{q-1}{q^t} = q^{1-t} - q^{-t}$$

for $1 \le i \le \lfloor n/t \rfloor$.

Consider the condition discussed in the statement of the theorem, namely, the condition that the sequence

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

does not have a harmful sequence of zero-discrepancies of length $t-1$. If this condition is satisfied then none of the events $E_1, E_2, \ldots, E_{\lfloor n/t \rfloor}$, is satisfied. The probability of the condition that is mentioned in the theorem is therefore at most

$$\begin{aligned}
&\mathrm{Prob}\left(\neg E_1 \wedge \neg E_2 \wedge \cdots \wedge \neg E_{\lfloor n/t \rfloor}\right) \\
&\quad = \prod_{i=1}^{\lfloor n/t \rfloor} \mathrm{Prob}\,(\neg E_i) && \text{(by mutual independence)} \\
&\quad = \prod_{i=1}^{\lfloor n/t \rfloor} \left(1 - q^{1-t} + q^{-t}\right) \\
&\quad \le \prod_{i=1}^{\lfloor n/t \rfloor} \left(1 - q^{-t}\right) && \text{(since } q \ge 2\text{)} \\
&\quad = \left(1 - q^{-t}\right)^{\lfloor n/t \rfloor} \\
&\quad \le \left(1 - q^{-t}\right)^{n/2t} && \text{(since } \tfrac{n}{2t} \le \tfrac{n}{t} - 1 \le \lfloor \tfrac{n}{t} \rfloor \text{ when } n \ge 2t\text{)} \\
&\quad = \left(\left(1 - q^{-t}\right)^{q^t}\right)^{\frac{n}{2tq^t}} \\
&\quad \le e^{-\frac{n}{2tq^t}} && \text{(since } (1-1/x)^x \le e^{-1} \text{ for any positive real number } x\text{).}
\end{aligned}$$

This desired bound has now been established. □

Corollary 4.10. *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 linearly recurrent sequences with characteristic polynomial f. Let t be a positive integer such that $n \ge 4t$.

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 $e^{-n/(4tq^{2t})}$.

Proof. Recall that, by Lemma 4.7, if a given linearly recurrent sequence

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

has a harmful sequence of zero-discrepancies of length $2t-2$, then the corresponding Hankel matrix H has a harmful gap of length $(2t-2)/2 = t-1$, as well.

It follows that if the given Hankel matrix *does not* have a harmful gap of length $t-1$ then the linearly recurrent sequence does not have a harmful sequence with zero-discrepancies of length $2t-2$. This certainly implies this linearly recurrent sequence does not have a harmful sequence of zero-discrepancies of length $2t-1$, either.

Therefore, the probability that the Hankel matrix does not have a harmful gap of length $t-1$ is less than or equal to the probability that the linearly recurrent sequence does not have harmful sequence of zero-discrepancies of length $2t-1$.

The result now follows by a straightforward application of Theorem 4.9 (with $2t$ replacing t in the statement of the theorem). $\square$

5 Solving Systems of Linear Equations

Lemma 4.2 implies that the bounds on probabilities given in Theorems 4.5 and 4.9 and Corollaries 4.6, 4.8, and 4.10 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 $u_1 \in \mathsf{F}^{N\times 1}$. An estimate of the solution for the given system is also produced. 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 proves that if the resulting iterative process is applied, and the vectors $u_1, u_2, \ldots$ that are required for each iteration are uniformly and independently chosen from $\mathsf{F}^{N\times 1}$, 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 this 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 4.9 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 Acknowledgments

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

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. In *Proceedings, 2003 International Symposium on Symbolic and Algebraic Computation (ISSAC '03)*, 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.