Reliable Krylov-Based Algorithms for Matrix Null Space and Rank *

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

February, 2004

Abstract

Krylov-based algorithms have recently been used, in combination with other methods, to solve systems of linear equations and to perform related matrix computations over finite fields. For example, large and sparse systems of linear equations over F_2 are formed during the use of the number field sieve for integer factorization, and elements of the null space of these systems are sampled. Block Lanczos algorithms have been used to perform this computation with considerable success. However, the algorithms that are currently in use do not appear to be reliable in the worst case.

This report presents a block Lanczos algorithm that is somewhat simpler than block algorithms that are presently in use and provably reliable for computations over large fields. This can be implemented, using a field extension, in order to produce several uniformly and independently selected elements from the null space at once. The amortized cost to produce each vector closely matches the cost to generate such a vector with the methods currently in use.

An algorithm is also given to compute the rank of a matrix $A \in \mathsf{F}^{m \times n}$ over a small finite field F. The expected number of matrix-vector products by A or A^t used by this algorithm is in $O(r)$, where r is the rank of A. The expected number of additional field operations used by this algorithm is within a polylog factor of $r(n+m)$, and the expected storage space is within a polylog factor of $n+m$. This is asymptotically more efficient than existing black box algorithms to compute the rank of a matrix over a small field, assuming that the cost of matrix-vector products dominates the cost of other operations.

1 Introduction

Consider the problem of selecting a vector uniformly and randomly from the null space of a given matrix. As discussed in the report of Buhler, Lenstra, and Pomerance [1], this problem arises for large, sparse matrices over the finite field $\mathsf{F} = \mathsf{F}_2$ when the number field sieve is applied.

*An extended abstract has been submitted for publication.

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

Structured Gaussian Elimination has been used for this computation [11]. However, storage requirements may be prohibitive for large problems when this technique is applied. Krylov-based algorithms, such as the algorithm of Lanczos [12], are reliable for computations over the real numbers, but require modification if they are applied for computations over small finite fields. A block-Lanczos algorithm was proposed for this purpose by Coppersmith [3] in the early 1990's, with the objectives of improving both reliability and coarse-grain parallelism. Variants of this algorithm, including a simpler algorithm of Montgomery [13], have been used (frequently in combination with elimination-based methods) with considerable success. Unfortunately, these algorithms have not been adequately analyzed, and there is reason to believe that they are not reliable, in the worst case, for computations over small finite fields: Krylov-based algorithms for singular matrix computations perform poorly if they are applied to matrices whose minimal polynomials (in $\mathsf{F}[z]$) are divisible by z^2 and that have a large number of invariant factors, and existing heuristics do not appear to address this problem. For example, they are ineffective for computations over F_2 when applied to block-diagonal matrices that include a large number of diagonal blocks

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

along with a large identity matrix as a final block. Heuristics that use symmetrization to condition the input — replacing A by A^tA or by AA^t – are defeated by block diagonal matrices with a form similar to the above, provided that copies of the blocks

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

are also used.

A related Krylov-based algorithm — namely, that of Wiedemann [15] — has subsequently been developed and fully analyzed. Furthermore, a block variant (with improved parallelism, once again) has also been shown to be reliable — see Kaltofen [10] for the analysis in the large field case and Villard [14] for the analysis over small finite fields. Indeed, the block Wiedemann algorithm allows the use of rectangular matrices as blocks and is asymptotically faster than existing block Lanczos algorithms. Yet, variants of the Lanczos algorithm continue to be used instead. We are therefore lead to ask whether algorithms that resemble the currently used heuristics are provably reliable.

A part of the answer to this question is provided in this report. In particular, a block Lanczos algorithm that is provably reliable for computations over large fields is described in Section 2. This appears to be both simpler than and at least as efficient as any Lanczos-based heuristic now in use. If implemented over a field extension, this provides an algorithm that returns several elements of the null space of a sparse matrix over a small finite field at once; the amortized cost to compute each vector is comparable to the cost of current Lanczos-based heuristics.

A rather different algorithm is described in Section 3 for computation of the rank of a matrix $A \in \mathsf{F}^{m \times n}$ over a small field F. The number of matrix-vector products by A or A^t used by this algorithm is linear in r, where r is the rank of A. The expected number of additional field operations used by the algorithm is within a polylog factor of $(n + m)r$, and the expected amount of storage space used is within a polylog factor of $n + m$. Previously available black-box algorithms either require computations over a field extension or the use of binary search to find the rank, increasing the number of matrix-vector products required by a logarithmic factor in each case. Consequently, the new algorithm is asymptotically more efficient than existing black box algorithms, when applied to compute the rank of a matrix over a small finite field, if (as usual) the cost of matrix-vector products dominates the cost of other operations.

2 A Block Lanczos Algorithm

Eberly and Kaltofen [8] present a simple scalar Lanczos algorithm and show that it is reliable over arbitrary large fields. In this section, this algorithm is modified to produce a simple block algorithm that is provably reliable for computations over arbitrary large fields, as well, and that can be used to sample from the null space of a given matrix A.

2.1 A Matrix Conditioner

We begin with a diagonal matrix preconditioner described in the above paper. Additional information about this preconditioner can be found in the report of Chen et. al. [2].

Lemma 2.1 (Eberly and Kaltofen [8]). *Suppose* F *is a field and let* $A \in \mathsf{F}^{m\times n}$ *be a matrix with rank* r*. Let* S *be a finite subset of* $\mathsf{F} \setminus \{0\}$*, and suppose*

$$\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_1, \beta_2, \ldots, \beta_m$$

are chosen uniformly and independently from S*. Let*

$$D_{\vec{\alpha}} = \begin{bmatrix} \alpha_1 & & & \\ & \alpha_2 & & \\ & & \ddots & \\ & & & \alpha_n \end{bmatrix} \in \mathsf{F}^{n\times n} \qquad \text{and} \qquad D_{\vec{\beta}} = \begin{bmatrix} \beta_m & & & \\ & \beta_2 & & \\ & & \ddots & \\ & & & \beta_m \end{bmatrix} \in \mathsf{F}^{m\times m}.$$

Then, with probability at most $\frac{11n^2-n}{2|S|}$*, the matrix*

$$\widetilde{A} = D_{\vec{\alpha}} A^T D_{\vec{\beta}} A D_{\vec{\alpha}} \in \mathsf{F}^{n\times n}$$

is a matrix with rank r*, whose characteristic polynomial is* $z^{n-r} f$ *for some squarefree polynomial* $f \in \mathsf{F}[z]$ *with degree* r *such that* $f(0) \neq 0$*.*

A consideration of the rank of A confirms that if the above-mentioned matrix $\widetilde{A}$ has the properties described in the lemma, then the minimal polynomial of $\widetilde{A}$ is zf and $\widetilde{A}$ is similar to a diagonal matrix over a suitable extension of F (namely, a splitting field of f).

Eberly and Kaltofen observe that if $\widetilde{A}$ is as described above, and if a system of linear equations

$$\widetilde{A}x = b$$

is consistent, then a solution for the system can be found within the Krylov space of b. That is, there exists a linear combination x of the vectors

$$b, \widetilde{A}b, \widetilde{A}^2 b, \ldots,$$

that satisfies the above system of equations. We may therefore select an element from the null space of $\widetilde{A}$ by randomly selecting a vector z, choosing a vector x such that $\widetilde{A}x = b$, for $b = \widetilde{A}z$, and returning the vector $z - x$.

Since A and $\widetilde{A}$ have the same rank, $\widetilde{A} = D_{\vec{\alpha}} A^T D_{\vec{\beta}} A D_{\vec{\alpha}}$, and the diagonal matrix $D_{\vec{\alpha}}$ is nonsingular, a vector y is in the null space of $\widetilde{A}$ if and only if $D_{\vec{\alpha}} y$ is in the null space of A. Therefore, we may also return the vector $D_{\vec{\alpha}}(z - x)$ as an element of the null space of A.

Suppose now that $k \geq 1$, and that k vectors $z_1, z_2, \ldots, z_k$ have been randomly selected from $\mathsf{F}^{n\times 1}$. Let $\vec{z}$ be the matrix in $\mathsf{F}^{n\times k}$ whose i^{th} column is z_i, for $1 \leq i \leq k$. It follows by a straightforward generalization of the above process that a sequence of k vectors can be sampled from the null space of $\widetilde{A}$ by finding a solution $\vec{x} \in \mathsf{F}^{n\times k}$ for the system

$$\widetilde{A}\vec{x} = \vec{b} \qquad \text{for } \vec{b} = \widetilde{A}\vec{z},$$

and returning the columns of the matrix $D_{\vec{\alpha}}(\vec{z} - \vec{x})$.

A block Lanczos algorithm that can be used to solve consistent systems of equations will next be described. This can be applied to perform the middle step of the above process.

2.2 A Block Lanczos Algorithm

Consider the algorithm that is shown in Figure 1 on page 5. This is a straightforward generalization of the "standard Lanczos algorithm" shown in Figure 1 of the paper of Eberly and Kaltofen [8].

Suppose that $\widetilde{A} \in \mathsf{F}^{n\times n}$ is a symmetric matrix with rank r. As suggested in the previous section, we are interested in the behaviour of the given algorithm when $\widetilde{A}$ has a minimal polynomial zf for some squarefree polynomial $f \in \mathsf{F}[z]$ with degree r such that $f(0) \neq 0$, so that $\widetilde{A}$ is similar to a diagonal matrix over an extension of F.

Let $\ell = \lceil r/k \rceil - 1$, where k is the "blocking factor" used in the algorithm.

If failure is not reported, then the algorithm generates a sequence of matrices

$$\vec{w}_0, \vec{w}_1, \vec{w}_2, \ldots, \vec{w}_\ell$$

such that $\vec{w}_i \in \mathsf{F}^{n\times k}$ for $0 \leq i \leq \ell - 1$ and such that $\vec{w}_\ell \in \mathsf{F}^{n\times h}$ for some integer h such that $1 \leq h \leq k$. As noted in the next section, it will frequently be the case that $h = k$ if r is divisible by k, and that $h = r - k\ell$ if m is not divisible by k.

The columns of the matrices

$$\vec{w}_0, \vec{w}_1, \vec{w}_2, \ldots, \vec{w}_i$$

are linearly independent and form a basis for the vector space spanned by the columns of the matrices

$$\vec{b}, \widetilde{A}\vec{b}, \widetilde{A}^2\vec{b}, \ldots, \widetilde{A}^i\vec{b}$$

for each integer i such that $0 \leq i \leq \ell$. Consequently, if h has its usual value (as given above), then the columns of the matrices

$$\vec{w}_0, \vec{w}_1, \ldots, \vec{w}_\ell$$

form a basis for the column space of $\widetilde{A}$, and the number of these columns is equal to the rank of $\widetilde{A}$.

A useful *orthogonality condition* is achieved:

$$\vec{w}_i^t \widetilde{A} \vec{w}_j = 0 \tag{2.1}$$

for all integers i and j such that $0 \leq i, j \leq \ell$ and $i \neq j$, and

$$\det \vec{w}_i^t \widetilde{A} \vec{w}_i \neq 0 \tag{2.2}$$

for $0 \leq i \leq \ell$.

Input: A symmetric matrix $\widetilde{A} \in \mathsf{F}^{n \times n}$ and a matrix $\vec{b} \in \mathsf{F}^{n \times k}$
Output: A matrix $\vec{x} \in \mathsf{F}^{n \times k}$ such that $\widetilde{A}\vec{x} = \vec{b}$, or failure

1. $\vec{w}_{-1} := 0_{n \times k}$; $\vec{v}_0 := 0_{n \times k}$; $\vec{x}_{-1} := 0_{n \times k}$; $\vec{t}_{-1} := I_k$
2. $\vec{w}_0 := \vec{b}$; $\vec{v}_1 := \widetilde{A}\vec{w}_0$; $\vec{t}_0 := \vec{v}_1^t \vec{w}_0$
3. $i := 0$
4. **while** $\det \vec{t}_i \neq 0$ **do**
5. $\quad \vec{x}_i := \vec{x}_{i-1} + \vec{w}_i \vec{t}_i^{-1} \vec{w}_i^t \vec{b}$
6. $\quad \vec{w}_{i+1} := \vec{v}_{i+1} - \vec{w}_i \vec{t}_i^{-1} \vec{v}_i^t \vec{v}_{i+1} - \vec{w}_{i-1} \vec{t}_{i-1}^{-1} \vec{v}_{i-1}^t \vec{v}_{i+1}$
7. $\quad \vec{v}_{i+2} := \widetilde{A}\vec{w}_{i+1}$
8. $\quad \vec{t}_{i+1} := \vec{v}_{i+2}^t \vec{w}_{i+1}$
9. $\quad i := i + 1$

end while

10. **if** $\vec{w}_i \neq 0_{n \times k}$ **then**
11. $\quad$ Set h to the the largest integer such that the leftmost h columns of $\vec{w}_i$ are linearly independent.
12. $\quad$ Set $\vec{w}_i$ to be the matrix in $\mathsf{F}^{n \times h}$ that includes the leftmost h columns of the current $\vec{w}_i$.
13. $\quad$ Set $\vec{t}_i$ to be the top left $h \times h$ submatrix of the current $\vec{t}_i$, so that $\vec{t}_i \in \mathsf{F}^{h \times h}$.
14. $\quad$ **if** $h = 0$ **or** $\det \vec{t}_i = 0$ **then**
15. $\quad\quad$ **report** failure

$\quad$ **else**

16. $\quad\quad \vec{x} := \vec{x}_{i-1} + \vec{w}_i \vec{t}_i^{-1} \vec{w}_i^t \vec{b}$

$\quad$ **end if**

else

17. $\quad \vec{x} := \vec{x}_{i-1}$

end if

18. **if** $\widetilde{A}\vec{x} = \vec{b}$ **then**
19. $\quad$ **return** $\vec{x}$

$\quad$ **else**

20. $\quad$ **return** failure

end if

Figure 1: A Block Lanczos Algorithm

Two other sequences of matrices are computed along the way, in order to minimize the number of multiplications by $\widetilde{A}$ that are used: $\vec{v}_0, \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_\ell$ are matrices such that

$$\vec{v}_{i+1} = \widetilde{A}\vec{w}_i \qquad \text{for } 0 \leq i \leq \ell - 1, \tag{2.3}$$

and $\vec{t}_0, \vec{t}_1, \ldots, \vec{t}_\ell$ are square matrices such that

$$\vec{t}_i = \vec{w}_i^t \widetilde{A} \vec{w}_i \qquad \text{for } 0 \leq i \leq \ell. \tag{2.4}$$

The algorithm maintains one more sequence of matrices, in order to produce a solution for the given system:

$$\vec{x}_0, \vec{x}_1, \ldots, \vec{x}_{\ell-1}$$

are matrices in $\mathsf{F}^{n\times k}$ such that

$$\vec{w}_j^T (A\vec{x}_i - b) = 0 \tag{2.5}$$

for all integers i and j such that $0 \le j \le i \le \ell - 1$; this is used at the end of the algorithm to generate a matrix $\vec{x}$ such that

$$\vec{w}_j^T (A\vec{x} - b) = 0 \tag{2.6}$$

for all j such that $0 \le j \le \ell$.

A comparison of this algorithm with the scalar algorithm will confirm that this is, indeed, a straightforward generalization: The two algorithms maintain the same sequences of matrices when $k = 1$, using virtually the same sets of operations. It is somewhat simpler than block Lanczos algorithms of Coppersmith [3] or Montgomery [13], due to the omission of any kind of lookahead mechanism. There is good reason to include such mechanisms for computations over small fields. However, as argued in the next section, these are not required for computations over large fields, when the coefficient matrix $\widetilde{A}$ has the properties that have been described here and the columns of $\vec{b}$ are randomly chosen from the column space of $\widetilde{A}$.

2.3 Analysis of Reliability

The following proof of reliability of the block Lanczos algorithm is, again, a modification of that of the reliability of the algorithm of Eberly and Kaltofen [8]. Suppose, once again, that $\widetilde{A} \in \mathsf{F}^{n\times n}$ is a symmetric matrix with rank r, and that $\vec{b} \in \mathsf{F}^{n\times k}$ for an integer $k \ge 1$. Let us consider the following block-Hankel matrices. For $1 \le i \le \lfloor r/k \rfloor$, let

$$H_i(\widetilde{A}, \vec{b}) = \begin{bmatrix} \vec{b}^t\widetilde{A}\vec{b} & \vec{b}^t\widetilde{A}^2\vec{b} & \cdots & \vec{b}^t\widetilde{A}^i\vec{b} \\ \vec{b}^t\widetilde{A}^2\vec{b} & \vec{b}^t\widetilde{A}^3\vec{b} & \cdots & \vec{b}^t\widetilde{A}^{i+1}\vec{b} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{b}^t\widetilde{A}^i\vec{b} & \vec{b}^t\widetilde{A}^{i+1}\vec{b} & \cdots & \vec{b}^t\widetilde{A}^{2i-1}\vec{b} \end{bmatrix}. \tag{2.7}$$

Let $H(\widetilde{A}, \vec{b}) \in \mathsf{F}^{r\times r}$ be the matrix $H_{r/k}(\widetilde{A}, \vec{b})$ if r is divisible by k, and let $H(\widetilde{A}, \vec{b})$ be the top left $r \times r$ submatrix of $H_{\lceil r/k \rceil}(\widetilde{A}, \vec{b})$, otherwise.

Lemma 2.2. *Suppose that $\widetilde{A} \in \mathsf{F}^{n\times n}$ is a symmetric matrix with rank r, whose minimal polynomial has the form zf, where $f \in \mathsf{F}[z]$ is a squarefree polynomial with degree r such that $f(0) \ne 0$.*

Let $\vec{b} \in \mathsf{F}^{n\times k}$ be a matrix such that the system

$$\widetilde{A}\vec{x} = \vec{b}$$

is consistent — that is, each of the columns of $\vec{b}$ belongs to the column space of $\widetilde{A}$.

Finally, suppose that $\det H_i(\widetilde{A}, \vec{b}) \ne 0$ for $1 \le i \le \lfloor r/k \rfloor$ and that $\det H(\widetilde{A}, \vec{b}) \ne 0$ as well.

Then the algorithm shown in Figure 1 succeeds. In particular, it generates a sequence of matrices

$$w_0, w_1, \ldots, w_\ell$$

for $\ell = \lceil r/k \rceil - 1$ whose columns are linearly independent and form a basis for the column space of $\widetilde{A}$, and it returns a matrix $\vec{x} \in \mathsf{F}^{n\times k}$ such that $\widetilde{A}\vec{x} = \vec{b}$.

Proof. Let i be an integer such that $0 \leq i \leq \lfloor r/k \rfloor$ and suppose that matrices

$$\vec{w}_0, \vec{w}_1, \ldots, \vec{w}_{i-1}$$

have been computed as in the algorithm, and that the matrix

$$\vec{t}_j = \vec{w}_j^t \widetilde{A} \vec{w}_j$$

is nonsingular, for $0 \leq j \leq i-1$. Then a consideration of elementary row and column operations can be applied to $H_i(\widetilde{A}, \vec{b})$ to produce the matrix

$$\begin{bmatrix} \vec{t}_0 & & & 0 \\ & \vec{t}_1 & & \\ & & \ddots & \\ 0 & & & \vec{t}_i \end{bmatrix},$$

so that this matrix is similar to $H_i(\widetilde{A}, \vec{b})$. Since $H_i(\widetilde{A}, \vec{b})$ is nonsingular, the matrix $t_i = \vec{w}_i^t \widetilde{A} \vec{w}_i$ must be nonsingular as well.

A similar argument can be used to establish that the final matrix $\vec{t}_\ell$ is nonsingular as well, if $\det H(\widetilde{A}, \vec{b})$ is also nonzero. Thus condition (2.2) is satisfied. A consideration of the computation of $\widetilde{w}_i$ (at step 6 of the algorithm) establishes that condition (2.1) is satisfied in this case as well.

Taken together, these can be used to establish that the columns of the matrices

$$\vec{w}_0, \vec{w}_1, \ldots, \vec{w}_\ell$$

are all linearly independent, so that they form the basis of an r-dimensional subspace of $\mathsf{F}^{n \times 1}$.

If the system $\widetilde{A}\vec{x} = \vec{b}$ is consistent, then these columns all belong to the column space of $\widetilde{A}$. Since $\widetilde{A}$ has rank r, it follows that they form a basis for the column space of $\widetilde{A}$, as claimed.

Consider the vector $\vec{x}$; it follows by an inspection of the algorithm (noting, in particular, lines 5 and 16) that

$$\vec{x} = \sum_{i=0}^{\ell} \vec{w}_i^t \vec{t}_i^{-1} \vec{w}_i^t b.$$

The above-mentioned orthogonality conditions can be used to establish that

$$\vec{w}_i^t (A\vec{x} - \vec{b}) = 0$$

for $0 \leq i \leq \ell$, so that matrix $A\vec{x} - \vec{b}$ is orthogonal to the column space of $\widetilde{A}$. On the other hand, if the columns of $\vec{b}$ belong to the column space of $\widetilde{A}$ (as given in the lemma), then the columns of the vector $A\vec{x} - \vec{b}$ clearly do as well. If the minimal polynomial of $\widetilde{A}$ is as described in the lemma (so that $\widetilde{A}$ is similar to a diagonal matrix over an extension of F) then it follows that $\widetilde{A}\vec{x} - \vec{b} = 0$, as required. □

Suppose now that $z_{i,j}$ are distinct indeterminates over F, for $1 \leq i \leq n$ and $1 \leq j \leq k$. Let

$$\vec{z} = \begin{bmatrix} z_{1,1} & z_{1,2} & \cdots & z_{1,k} \\ z_{2,1} & z_{2,2} & \cdots & z_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n,1} & z_{n,2} & \cdots & z_{n,k} \end{bmatrix} \in \mathsf{F}[z_{1,1} \ldots, z_{n,k}]^{n \times k}, \tag{2.8}$$

and suppose that

$$\vec{b} = \widetilde{A}\vec{z}. \tag{2.9}$$

The determinants of the corresponding matrices $H_i(\widetilde{A}, \vec{b})$ and $H(\widetilde{A}, \vec{b})$ are polynomials in the ring $\mathsf{F}[z_{1,1}, \ldots, z_{n,k}]$ with total degrees at most $2ki$ and $2r$, respectively.

Lemma 2.3. *Suppose once again that $A \in \mathsf{F}^{n\times n}$ is a symmetric matrix with rank r whose minimal polynomial has the form zf, where $f \in \mathsf{F}[z]$ is a squarefree polynomial with degree r such that $f(0) \neq 0$.*

Suppose as well that k is odd and k is not divisible by the characteristic of the field F.

Then if $\vec{z}$ and $\vec{b}$ are as given in Equations (2.8) and (2.9), then the polynomials $H_i(\widetilde{A}, b)$ are nonzero for $1 \leq i \leq \lfloor r/k \rfloor$, and the polynomial $H(\widetilde{A}, \vec{b})$ is nonzero as well.

Proof. If the minimal polynomial of $\widetilde{A}$ is as described in the statement of the lemma, then $\widetilde{A}$ is similar to a diagonal matrix over some extension of F.

It follows that if k is not divisible by the characteristic of F, then then there exists a symmetric matrix $B \in \mathsf{E}^{n\times n}$ in some extension E of F such that

$$\widetilde{A} = B^k.$$

In particular, E may be chosen to be an extension of F that includes k^{th} roots of each of the roots of f.

In this case (since B is also similar to a diagonal matrix) B also has rank r, and the column spaces of $\widetilde{A}$ and B are the same, when these are considered as matrices in $\mathsf{E}^{n\times n}$.

Now a result of Eberly and Kaltofen can be applied: there exists a vector $\gamma \in \mathsf{E}^{n\times 1}$ such that γ is in the column space of B and the Hankel matrix

$$\begin{bmatrix} \gamma^t B\gamma & \gamma_t B^2\gamma & \cdots & \gamma^t B^r\gamma \\ \gamma^t B^2\gamma & \gamma_t B^3\gamma & \cdots & \gamma^t B^{r+1}\gamma \\ \vdots & \vdots & \ddots & \vdots \\ \gamma^t B^r\gamma & \gamma^t B^{r+1}\gamma & \cdots & \gamma^t B^{2r-1}\gamma \end{bmatrix} \tag{2.10}$$

is in generic profile — each of its top left submatrices is nonsingular — see the full version of Eberly and Kaltofen [8] for details.

Since k is odd, and B is similar to a diagonal matrix, there is a vector $\kappa \in \mathsf{E}^{n\times 1}$ such that

$$B^{(3k-1)/2}\kappa = \gamma.$$

Now consider the matrix

$$\vec{\zeta} = [\kappa \; B\kappa \; \cdots \; B^{k-1}\kappa] \in E^{n\times k}$$

along with the matrix

$$\vec{\beta} = \widetilde{A}\vec{\zeta}.$$

Since $\widetilde{A} = B^k$, and $B^{(3k-1)/2}\kappa = \gamma$, it is easily checked that

$$\vec{\beta}^t\widetilde{A}\vec{\beta} = \begin{bmatrix} \gamma^t B\gamma & \gamma^t B^2\gamma & \cdots & \gamma^t B^k\gamma \\ \gamma^t B^2\gamma & \gamma^t B^3\gamma & \cdots & \gamma^t B^{k+1}\gamma \\ \vdots & \vdots & \ddots & \vdots \\ \gamma^t B^k\gamma & \gamma^t B^{k+1}\gamma & \cdots & \gamma^t B^{2k-1}\gamma \end{bmatrix}.$$

Furthermore, $H(\widetilde{A}, \vec{\beta}) = H(\widetilde{A}, \widetilde{A}\vec{\zeta})$ is equal to the matrix shown in Equation (2.10), above.

It follows that the polynomials mentioned in the lemma are all nonzero — for their values are nonzero, when evaluated over E by setting $\vec{z}$ to be $\vec{\zeta}$. □

The following can be deduced using Lemmas 2.2 and 2.3, along with an application of the Schwartz-Zippel lemma.

Theorem 2.4. *Suppose that $\widetilde{A} \in \mathsf{F}^{n \times n}$ is a symmetric matrix with rank r, whose minimal polynomial has the form zf, where $f \in \mathsf{F}[z]$ is a squarefree polynomial with degree r such that $f(0) \neq 0$.*

Let $k \geq 1$ such that k is odd and k is not divisible by the characteristic of F.

Finally, suppose that the algorithm shown in Figure 1 is applied with inputs $\widetilde{A}$ and a matrix $\vec{b} \in \mathsf{F}^{n \times k}$, where

$$\vec{b} = \widetilde{A}\vec{z}$$

and the entries of the matrix $\vec{z} \in \mathsf{F}^{n \times k}$ are chosen uniformly and independently from a finite subset S of F.

Then the algorithm succeeds, and returns a matrix $\vec{x} \in \mathsf{F}^{n \times k}$ such that

$$\widetilde{A}\vec{x} = \vec{b},$$

with probability at least $1 - r(r+1)/|S|$.

Furthermore, if F is a finite field and $S = \mathsf{F}$, then the resulting matrix $\vec{x}$ is uniformly chosen from the set of solutions for the above system of equations.

It follows that the process described in this section can be used to produce a set of k elements of the null space of a given matrix A: It suffices to sample elements uniformly and independently from a finite subset S of F with size in $O(n^2/\epsilon)$ in order to bound the probability of failure by ϵ, for any given error tolerance $\epsilon > 0$.

If F is a small finite field — as is the case in notable applications — then it is sufficient to implement the given algorithm over an extension E of F whose degree d over F is logarithmic in n in order to ensure that the process is reliable.

Note that — using the coordinate basis — a vector $z \in \mathsf{E}^{n \times 1}$ in the null space of A can be used to generate a set of d vectors in $\mathsf{F}^{n \times 1}$ that are in the null space of A as well. Consequently the above process can be used to produce a set of kd uniformly and independently selected elements of the null space of A, if it is implemented using an extension E with degree d over F, and the block Lanczos algorithm applied with blocking factor k.

The amortized cost of the computation — that is, the ratio of the total cost of this computation to the number of elements of the null space that are produced — is competitive with that of any heuristic for this computation that is presently in use.

3 Estimation of the Rank

Consider the problem of computing the rank of a sparse or structured matrix $A \in \mathsf{F}^{m \times n}$ over a small finite field F. The algorithm described in the previous section can be applied to solve this problem if it is

implemented over a field extension whose degree over F is at most linear in $\log_{|\mathsf{F}|} n$ — one simply needs to keep track of the sums of the orders of the nonsingular matrices t_i that the algorithm maintains. However, both the time and storage requirements of the computation are increased by logarithmic factors when computations over field extensions are used.

In this section, we present a rather different algorithm that can be used for this computation over small fields. The algorithm is more complicated, and less amenable to parallelization, than the algorithm of the previous section. The expected amount of storage space required is also greater, by a logarithmic factor, than that required for the previous algorithm. However, the algorithm that is presented in this section avoids computations over field extensions and is asymptotically more efficient than the first algorithm when applied to solve the above problem over small finite fields, assuming (as usual) that the cost of matrix-vector products dominates the cost of other operations.

3.1 The Frobenius Form

Consider a square matrix $\widehat{A} \in \mathsf{F}^{\ell \times \ell}$ for a positive integer ℓ. It is well known (see, for example, Gantmacher [9]) that $\widehat{A}$ is similar to a unique block diagonal matrix with companion matrices of monic polynomials $f_1, f_2, \ldots, f_k$ on the diagonal, for some integer $k \leq \ell$, where f_i is divisible by f_{i+1} for $1 \leq i \leq k-1$. That is, there exists a nonsingular matrix $V \in \mathsf{F}^{\ell \times \ell}$ such that

$$V \widehat{A} V^{-1} = \begin{bmatrix} C_{f_1} & & & 0 \\ & C_{f_2} & & \\ & & \ddots & \\ 0 & & & C_{f_k} \end{bmatrix} \tag{3.1}$$

and where

$$C_g = \begin{bmatrix} 0 & \cdots & 0 & -g_0 \\ 1 & & 0 & -g_1 \\ & \ddots & \vdots & \vdots \\ 0 & & 1 & -g_{d-1} \end{bmatrix} \in \mathsf{F}^{d \times d} \tag{3.2}$$

is the companion matrix of a monic polynomial

$$g = x^d + g_{d-1}x^{d-1} + g_{d-2}x^{d-2} + \cdots + g_1 x + g_0 \in \mathsf{F}[x].$$

The block diagonal matrix shown on the right hand side of Equation (3.1) is commonly called the *Frobenius form* of $\widehat{A}$, and the polynomials $f_1, f_2, \ldots, f_k$ are called the *invariant factors* of $\widehat{A}$.

If the matrix $\widehat{A}$ is singular then one or more of the invariant factors of $\widehat{A}$ may be equal to the polynomial x; we will say that an invariant factor f_i is a *nontrivial invariant factor* if $f_i \neq x$.

A considerable number of algorithms for the computation of the Frobenius form of a matrix are known. In the rest of this section, we will compute the rank of a given matrix A by modifying a black box algorithm for computation of the Frobenius form of Eberly [5], [6].

Unfortunately, this algorithm is not space-efficient: It also computes and stores the matrix V shown in Equation (3.1). A modified algorithm is presented in the next section. As described below, the new algorithm is more space-efficient if the number of nontrivial invariant factors of the given matrix is small, and a reliable upper bound on this number is available.

3.2 A Space-Efficient Algorithm

Let $\widehat{A} \in \mathsf{F}^{\ell \times \ell}$ and suppose that we are given A along with a positive integer, h, which will be used as an upper bound on the number of nontrivial invariant factors of $\widehat{A}$.

Suppose that d is the sum of the degrees of the nontrivial invariant factors of $\widehat{A}$.

In this section, we will modify the algorithm of Eberly [5] in order to produce a Monte Carlo algorithm that satisfies the following properties when run on input $\widehat{A} \in \mathsf{F}^{\ell \times \ell}$ and h.

- If the number of nontrivial invariant factors is, indeed, less than or equal to h, then the algorithm will return the nontrivial invariant factors of $\widehat{A}$ with high probability.
- If the bound h is incorrect — that is, $\widehat{A}$ includes more than h nontrivial invariant factors — then the algorithm will report failure with high probability, instead.
- If the algorithm is successful then expected number of matrix-vector products used by the algorithm is in $O(d)$. The expected number of additional operations required over F is in $O(\ell h d)$, and the amount of storage space used by the algorithm is in $O(\ell h^2 + \ell \log \ell)$.

The algorithm of Eberly [5] makes repeated use of a procedure minpolspace that is presented and analyzed in Section 3.1 of the above paper.

On its initial application, the procedure uses a sequence of uniformly and independently selected vectors from $\mathsf{F}^{\ell \times 1}$ in order to generate a pair of vectors u_1 and v_1 in $\mathsf{F}^{\ell \times 1}$, and a monic polynomial $f_1 \in \mathsf{F}[x]$, such that the following properties hold.

- f_1 is the monic polynomial of least degree such that $f_1(\widehat{A})v_1 = 0$.
- f_1 is also the monic polynomial of least degree such that $f_1(\widehat{A}^t)u_1 = 0$.
- Finally, f_1 is the minimal polynomial of the linearly recurrent sequence

$$u_1^t v_1, u_1^t \widehat{A} v_1, u_1 \widehat{A}^2 v_1 \ldots$$

- The expected number of vectors that must be selected from $\mathsf{F}^{\ell \times 1}$ to perform this computation is in $O(1)$. The expected number of matrix-vector products by $\widehat{A}$ or $\widehat{A}^t$ that is used is linear in the degree of f_1. Finally, the expected number of additional operations over the field F that are used by this procedure is linear in the product of ℓ and the degree of f_1.
- The polynomial f_1 is always a divisor of the minimal polynomial of $\widehat{A}$; it is equal to the minimal polynomial of $\widehat{A}$ with probability at least one-half.

Suppose the above polynomial f_1 has degree d_1. If the above conditions are satisfied then the Hankel matrix

$$\begin{bmatrix} u_1^t v_1 & u_1^t \widehat{A} v_1 & \cdots & u_1^t \widehat{A}^{d_1 - 1} v_1 \\ u_1^t \widehat{A} v_1 & u_1^t \widehat{A}^2 v_1 & \cdots & u_1^t \widehat{A}^{d_1} \\ \vdots & \vdots & \ddots & \vdots \\ u_1^t \widehat{A}^{d_1 - 1} v_1 & u_1^t \widehat{A}_{d_1} v_1 & \cdots & u_1^t \widehat{A}^{2d_1 - 2} v_1 \end{bmatrix}$$

is nonsingular. However, it is desirable to ensure that leading submatrices are likely to be nonsingular as well.

A **first modification** that will be made to the algorithm will therefore be a randomization: The vector v_1 will be replaced by $g_1(A)v_1$, where g_1 is a randomly chosen polynomial in $\mathsf{F}[x]$ that is relatively prime to f_1. Then the above conditions are still satisfied, and the above Hankel matrix is still nonsingular. Furthermore, it follows by a straightforward modification of a result of Eberly [7] that a scalar Lanczos algorithm can be used, with u_1 and v_1, in order to orthogonalize a pair of set of k vectors with respect to

$$u_1, \widehat{A}^t u_1, \ldots, (\widehat{A}^t)^{d_1-1} v_1 \quad \text{and} \quad v_1, \widehat{A} v_1, \ldots, \widehat{A}^{d_1-1} v_1$$

respectively. In particular, this computation can be performed using the vectors u_1, v_1, and the vectors to be orthogonalized, while using storage space for $O(\ell \log \ell + k)$ field elements in the worst case.

A **second modification** can now be made: Rather than storing all of

$$u_1, \widehat{A}^t u_1, \ldots, (\widehat{A}^t)^{d_1-1} u_1$$

and

$$v_1, \widehat{A} v_1, \ldots, \widehat{A}^{d_1-1} v_1,$$

— or a dual basis for the Krylov spaces that are generated by u_1 and v_1 — the algorithm will store u_1 and v_1 alone.

The algorithm of Eberly [5] requires a supply of vectors that have been generated by selecting $O(h)$ vectors uniformly and independently from $\mathsf{F}^{\ell \times 1}$, and orthogonalizing these vectors with respect to Krylov spaces corresponding to the invariant factors that have currently been generated.

A **third modification** concerns the way that these vectors are produced. The first application of the revised procedure minpolspace ends with the uniform and independent selection of $2ch$ vectors from $\mathsf{F}^{\ell \times 1}$, for a suitable constant c. A scalar Lanczos algorithm is applied to u_1 and v_1 once again, in order to orthogonalize these vectors, resulting in vectors

$$\alpha_{1,1}, \ldots, \alpha_{1,s_1}, \beta_{1,1}, \ldots, \beta_{1,s_1} \in \mathsf{F}^{\ell \times 1},$$

where $s_1 = ch$, such that

$$\alpha_{1,i}^t \widehat{A}^j v_1 = u_1^t \widehat{A}^j \beta_{1,i} = 0$$

for $1 \leq i \leq s_1$ and $0 \leq j \leq d_1 - 1$.

The amount of storage space needed to perform this computation is in $O(\ell \log \ell + \ell h)$. It will be useful to use the orthogonalized vectors in later steps, so these will be stored. The total amount of storage space needed for all these vectors is linear in the product of ℓh and the total number of applications of minpolspace that must be used. Since this number of applications is linear in h, the amount of storage space required for all of these orthogonalized vectors is in $O(\ell h^2)$.

Each subsequent application of minpolspace will take place after a sequence of vectors and polynomials

$$(u_1, v_1, f_1), (u_2, v_2, f_2), \ldots, (u_i, v_i f_i)$$

have been generated. A set of $2s_j$ vectors

$$\alpha_{j,1}, \ldots, \alpha_{j,s_j}, \beta_{j,1}, \ldots, \beta_{j,s_j} \in \mathsf{F}^{\ell \times 1}$$

will be available as well, for some integer s_j such that $1 \leq s_j \leq ch$ and for $1 \leq j \leq i$. These vectors will have been orthogonalized with respect to previous Krylov spaces — that is,

$$\alpha_{j,k}^t \widehat{A}^a v_b = u_b^t \widehat{A}^a \beta_{j,k} = 0$$

for all integers j, k, a, and b such that $1 \leq b \leq j$, $1 \leq k \leq s_j$, and $0 \leq a \leq d_b$, where d_b is the degree of f_b.

In order to ensure that the vectors u_{i+1} and v_{i+1} to be generated during the current application of minpolspace are orthogonal to the Krylov spaces that have been generated already, vectors from the sequences

$$\alpha_{i,1}, \ldots, \alpha_{i,s_i} \qquad \text{and} \qquad \beta_{i,1}, \ldots, \beta_{i,s_i} \tag{3.3}$$

will be used instead of randomly selected vectors from $\mathsf{F}^{\ell \times 1}$. The vectors that are used will then be discarded (decreasing the value of s_i). A scalar Lanczos algorithm will be applied, using u_{i+1} and v_{i+1}, to orthogonalize the vectors shown at line (3.3) with respect to the $i + 1^{\text{st}}$ Krylov spaces, in order to produce the next set of vectors

$$\alpha_{i+1,1}, \ldots, \alpha_{i+1,s_{i+1}} \qquad \text{and} \qquad \beta_{i+1,1}, \ldots, \beta_{i,s_{i+1}}$$

at the end of this application of minpolspace.

The algorithm will make repeated use of the modified procedure minpolspace, generating estimates of the invariant factors (and discarding polynomials and Krylov spaces, when estimates are discovered to incorrect) as in Eberly [5].

A **fourth modification** should now be made: The computation should be terminated as soon as it has been established, with high probability, either that $\widehat{A}$ includes at most h invariant factors, or that the $h + 1^{\text{st}}$ invariant factor is different from x. The algorithm reports failure in the latter case.

Unfortunately, the result is a Monte Carlo algorithm instead of a Las Vegas one: Since a complete set of invariant factors (including all trivial factors, along with corresponding Krylov spaces) is not generated, when $\widehat{A}$ has more than $h+1$ invariant factors, there is a small possibility that the polynomials returned by this algorithm are not the invariant factors of $\widehat{A}$ in this case.

The analysis of Eberly [5] can now be modified to establish that the above algorithm computes the desired values at the costs given at the beginning of this section.

3.3 Computation of the Rank

Suppose that the nontrivial invariant factors

$$f_1, f_2, \ldots, f_k$$

have been computed, as described above. Let

$$e_i = \begin{cases} \deg f_i & \text{if } f_i \text{ is not divisible by } x, \\ \deg f_i - 1 & \text{otherwise.} \end{cases}$$

Then the rank of $\widehat{A}$ is $e_1 + e_2 + \cdots + e_k$. Consequently, the rank of $\widehat{A}$ can be computed by a Monte Carlo algorithm with the (expected) cost stated at the beginning of Section 3.2. In particular, the rank can be computed efficiently if the number of nontrivial invariant factors of $\widehat{A}$ is small.

3.4 A Sparse Matrix Conditioner

3.4.1 Definition of Conditioner

Suppose, once again, that $A \in \mathsf{F}^{n\times m}$. Let r be the (unknown) rank of A, let $q = |\mathsf{F}|$, and let

$$\ell = \min(n, m) + c$$

where c is a parameter whose value will be given later.

Consider another constant $\widehat{c}$, as well, such that

$$\widehat{c} \geq \frac{(q-1)c}{q \log_q N}, \qquad \text{so that} \qquad c \leq \frac{\widehat{c} q \log_q N}{(q-1)}. \tag{3.4}$$

Consider matrices $L \in \mathsf{F}^{\ell\times n}$ and $R \in \mathsf{F}^{m\times\ell}$ whose entries are randomly selected according to the following distribution.

- If $1 \leq i \leq \min(n, m)$ then each entry in row i of L or column i of R is set to be zero with probability
$$\max\left(1 - \frac{\widehat{c}\log_q N}{i}, \frac{1}{q}\right) \qquad \text{for } N = \max(n, m).$$
- If $1 \leq i \leq \min(n, m)$ then each entry in row i of L or column i of R that has not been set to be 0, above, is chosen uniformly and independently from $\mathsf{F} \setminus \{0\}$.
- Finally, if $\min(n, m) < i \leq \ell$ then each entry of row i of L or column i of R is chosen uniformly and independently from F.

Notice that if $1 \leq i \leq c$ then

$$\begin{aligned} 1 - \frac{\widehat{c}\log_q N}{i} &\leq 1 - \frac{\widehat{c}\log_q N}{c} \\ &\leq 1 - \frac{\widehat{c}\log_q N}{\widehat{c} q \log_q N/(q-1)} && \text{(by condition (3.4))} \\ &= 1 - \frac{q-1}{q} \\ &= \frac{1}{q}. \end{aligned}$$

Thus the entries in the top c rows of L and the leftmost c columns of R are chosen uniformly and independently from F if L and R are randomly chosen as described above.

Let

$$\widehat{A} = LAR \in \mathsf{F}^{\ell\times\ell}. \tag{3.5}$$

Sparse matrices with the a similar structure have been investigated by Wiedemann [15]; additional useful properties are discussed in the report of Chen et. al. [2]. In the rest of this section we will establish another useful property, namely, that conditioning a matrix by pre- and post-multiplying by these matrices ensures that the expected number of nontrivial invariant factors of a matrix is small.

3.4.2 Useful Lemmas

Lemma 3.1. *Let i be a positive integer such that $\widehat{c}\log_q N \le i$. Then*

$$\left(1 - \frac{\widehat{c}\log_q N}{i}\right)^i \le N^{-\widehat{c}/\ln q}.$$

Proof. It is well known that if x is a real number such that $|x| \le 1$ then

$$1 + x \le e^x \le 1 + x + x^2$$

— see, for example, page 53 of the text of Cormen, Leiserson, Rivest and Stein [4]. Since i is a positive integer such that $\widehat{c}\log_q N \le i$, this implies that

$$\left(1 - \frac{\widehat{c}\log_q N}{i}\right) \le e^{-\widehat{c}\log_q N/i}.$$

Therefore

$$\begin{aligned}\left(1 - \frac{\widehat{c}\log_q N}{i}\right)^i &\le e^{-\widehat{c}\log_q N} \\ &= e^{-\widehat{c}\ln N/\ln q} \\ &= N^{-\widehat{c}/\ln q}.\end{aligned}$$

□

It will be useful to consider two other pairs of matrices that are chosen using a distribution similar to the above.

Suppose that $L_0, R_0 \in \mathsf{F}^{r\times r}$ are randomly chosen as follows.

- If $1 \le i \le r$ then each entry of row i or L_0 or column i of R_0 is set to be zero with probability

$$\max\left(1 - \frac{\widehat{c}\log_q N}{i}, \frac{1}{q}\right).$$

- Each entry in row i of L_0 or column i of R_0 that has not been set to 0, above, is chosen uniformly and independently from $\mathsf{F} \setminus \{0\}$.

Once again, this implies that the entries of the top c rows of L_0 and the leftmost c columns of R_0 are chosen uniformly and independently from F.

Lemma 3.2. *Let $\widetilde{A} \in \mathsf{F}^{r\times r}$ be a nonsingular matrix and let $B \in \mathsf{F}^{r\times r}$ as well. Suppose that matrix $L_0 \in \mathsf{F}^{r\times r}$ is randomly chosen as described above. Let*

$$C = L_0\widetilde{A} + B \in \mathsf{F}^{r\times r}.$$

Suppose i is an integer such that $1 \le i \le r$.

(a) The probability that row i of C is a linear combination of rows $i+1, i+2, \ldots, r$ is at most

$$q^{-i} + N^{-\widehat{c}/\ln q}.$$

(b) The probability that rows $i, i+1, \ldots, r$ of C are linearly dependent is at most

$$\frac{q^{1-i}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}.$$

(c) The probability that the nullity of C is greater than or equal to i is at most

$$\frac{q^{1-i}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}.$$

(d) The expected value of the nullity of C is at most

$$\frac{q}{(q-1)^2} + N^{2-\frac{\widehat{c}}{\ln q}}.$$

Proof. To begin, let us consider part (a) in the special case that $\widetilde{A} = I_r$, the $r \times r$ identity matrix, so that

$$C = L_0 + B.$$

Suppose that $1 \leq i \leq r$, and consider the probability that the i^{th} row of C is a linear combination of rows

$$i+1, i+2, \ldots, r.$$

Let s_i be the rank of the submatrix $\widehat{C}_i$ of C that consists of the above rows $i+1, i+2, \ldots, r$. Clearly $s_i \leq r - i$, so that $r - s_i \geq i$.

The matrix $\widehat{C}_i$ has a set S_i of s_i linearly independent columns, and each of the remaining columns is a linear combination of those in this set. Consider any assignment of values to the entries in row i of L_0, in the columns of S_i. There is exactly one assignment of values, that can be made to the remaining $r - s_i$ columns in row i, in order for row i of C to be a linear combination of rows $i+1, i+2, \ldots, r$.

Either $1 - \widehat{c}\log_q N/i \geq 1/q$ or $1 - \widehat{c}\log_q N/i < 1/q$. These cases will be considered separately.

Case: $1 - \widehat{c}\log_q N/i \geq 1/q$.

In this case, it follows by the above analysis, and the choice of L_0, that row i of C is a linear combination of rows $i+1, i+2, \ldots, r$, with probability at most

$$\begin{aligned}\left(1 - \frac{\widehat{c}\log_q N}{i}\right)^{r-s_i} &\leq \left(1 - \frac{\widehat{c}\log_q N}{i}\right)^{i} && \text{(since } r - s_i \geq i\text{)} \\ &\leq N^{-\widehat{c}/\ln q} && \text{(by Lemma 3.1).}\end{aligned}$$

Case: $1 - \widehat{c}\log_q N/i < 1/q$.

In this case, the entries of row i of L_0 are chosen uniformly and independently from F. It follows by the above analysis that row i of C is a linear combination of rows $i+1, i+2, \ldots, r$ with probability at most

$$\left(\frac{1}{q}\right)^{r-s_i} \leq \left(\frac{1}{q}\right)^{i} = q^{-i}.$$

Over-approximating, we see that if $1 \leq i \leq r$, then row i of C is a linear combination of rows $i+1, i+2, \ldots, r$ with probability at most

$$q^{-i} + N^{-\widehat{c}/\ln q},$$

as required to establish part (a) of the claim when $\widetilde{A}$ is the identity matrix.

Now suppose that $\widetilde{A} \in \mathsf{F}^{r \times r}$ is a nonsingular matrix. Then

$$C = L_0 \widetilde{A} + B = C' \cdot \widetilde{A}$$

where

$$C' = L_0 + B' \qquad \text{and} \qquad B' = B \cdot \widetilde{A}^{-1}.$$

It follows by the above analysis that row i of C' is a linear combination of rows $i+1, i+2, \ldots, r$ with probability at most

$$q^{-i} + N^{-\widehat{c}/\ln q}.$$

However, since $\widetilde{A}$ is nonsingular, row i of C' is a linear combination of rows $i+1, i+2, \ldots, r$ of C' if and only if row i of C is a linear combination of rows $i+1, i+2, \ldots, r$ of C, and this implies that part (a) holds in the general case.

Now consider the probability that rows

$$i, i+1, i+2, \ldots, r$$

are linearly dependent. In this case, there must exist at least one integer j such that $i \leq j \leq r$ and row j is a linear combination of rows $j+1, j+2, \ldots, r$. The probability that this is the case is at most

$$\begin{aligned}
\sum_{j=i}^{r} \left(q^{-j} + N^{-\frac{\widehat{c}}{\ln q}} \right) &= \left(\sum_{j=i}^{r} q^{-j} \right) + \left(\sum_{j=i}^{r} N^{-\frac{\widehat{c}}{\ln q}} \right) \\
&= \left(\sum_{j=i}^{r} q^{-j} \right) + (r - i + 1) \cdot N^{-\frac{\widehat{c}}{\ln q}} \\
&\leq \left(\sum_{j=i}^{r} q^{-j} \right) + N^{1-\frac{\widehat{c}}{\ln q}} \qquad \text{(since } r \leq \min(n,m) \leq \max(n,m) = N\text{)} \\
&\leq \left(\sum_{j \geq i} q^{-j} \right) + N^{1-\frac{\widehat{c}}{\ln q}} \\
&= \frac{q^{1-i}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}},
\end{aligned}$$

establishing part (b).

In order to establish part (c) note that if the nullity of C is greater than or equal to i, then the rank of C must be less than or equal to $r - i$, and rows $i, i+1, \ldots, r$ of C must therefore be linearly dependent. Consequently, part (c) of the claim is implied by part (b).

Finally, in order to establish part (d), note that, since the nullity of any $r \times r$ matrix is a nonnegative integer that is less than or equal to r, the expected value of the nullity of C is

$$\begin{aligned}
\sum_{i=1}^{r} \left(\text{Prob}\left(\text{nullity}(C) \geq i \right) \right) &\leq \sum_{i=1}^{r} \left(\frac{q^{1-i}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}} \right) \qquad \text{(by part (c))} \\
&= \sum_{i=1}^{r} \frac{q^{1-i}}{q-1} + \sum_{i=1}^{r} N^{1-\frac{\widehat{c}}{\ln q}}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{q}{q-1} \cdot \sum_{i=1}^{r} q^{-i} + r \cdot N^{1-\frac{\widehat{c}}{\ln q}} \\
&\leq \frac{q}{q-1} \cdot \sum_{i=1}^{r} q^{-i} + N^{2-\frac{\widehat{c}}{\ln q}} \qquad (\text{since } r \leq \min(n,m) \leq \max(n,m) = N) \\
&\leq \frac{q}{q-1} \cdot \sum_{i \geq 1} q^{-i} + N^{2-\frac{\widehat{c}}{\ln q}} \\
&= \frac{q}{(q-1)^2} + N^{2-\frac{\widehat{c}}{\ln q}}. \qquad \square
\end{aligned}
$$

Notice that the transpose of the matrix R_0 is chosen using the same probability distribution as described for L_0. The next result can therefore be obtained as a consequence of the previous one, by considering the transpose of the matrix D that is mentioned below.

Corollary 3.3. *Let $\widetilde{A} \in \mathsf{F}^{r \times r}$ be a nonsingular matrix and let $B \in \mathsf{F}^{r \times r}$ as well. Suppose that the matrix $R_0 \in \mathsf{F}^{r \times r}$ is randomly chosen as described on page 15, above. Let*

$$D = \widetilde{A} R_0 + B \in \mathsf{F}^{r \times r}.$$

Suppose that i is an integer such that $1 \leq i \leq r$.

(a) The probability that column i of D is a linear combination of columns $i+1, i+2, \ldots, r$ is at most

$$q^{-i} + N^{-\widehat{c}/\ln q}.$$

(b) The probability that columns $i, i+1, \ldots, r$ of D are linearly dependent is at most

$$\frac{q^{1-i}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}.$$

(c) The probability that the nullity of D is greater than or equal to i is at most

$$\frac{q^{1-i}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}.$$

(d) The expected value of the nullity of D is at most

$$\frac{q}{(q-1)^2} + N^{2-\frac{\widehat{c}}{\ln q}}.$$

Lastly, consider matrices $L_1 \in \mathsf{F}^{(r+c) \times r}$ and $R_1 \in \mathsf{F}^{r \times (r+c)}$ that are randomly chosen as follows.

- If $1 \leq i \leq r$ then each entry in row i of L_1 or column i of R_1 is set to be zero with probability

$$\max\left(1 - \frac{\widehat{c} \log_q N}{i}, \frac{1}{q}\right).$$

- If $1 \leq i \leq r$ then each entry in row i of L_1 or column i of R_1 that has not been set to 0, above, is chosen uniformly and independently from $\mathsf{F} \setminus \{0\}$.

- Finally, if $r < i \leq r + c$ then each entry in row i of L_1 or column i of R_1 is chosen uniformly and independently from F.

Note that the top $r \times r$ submatrix of L_1 and the left $r \times r$ submatrix of R_1 are chosen using the distributions described for L_0 and R_0, respectively (as described on page 15). Consequently, the entries in the top c rows of L_1 and the leftmost c columns of R_1 are chosen uniformly and independently from F.

Lemma 3.4. *Let $\widetilde{A} \in \mathsf{F}^{r\times r}$ be a nonsingular matrix. Let $B \in \mathsf{F}^{(r+c)\times r}$, and suppose L_1 is chosen as described above. Let*

$$C = L_1\widetilde{A} + B \in \mathsf{F}^{(r+c)\times r}.$$

Then the probability that rows $c, c+1, \ldots, r$ of C are linearly dependent or C has rank less than r (or both) is at most

$$\frac{2q^{-c}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}.$$

Proof. To begin let us suppose that $\widetilde{A} = I_r$, the $r \times r$ identity matrix.

Let $L_0, B_0 \in \mathsf{F}^{r\times r}$ be the top $r \times r$ submatrices of L_1 and B, respectively. Then the top $r \times r$ submatrix of C is the matrix

$$C_0 = L_0 + B_0 \in \mathsf{F}^{r\times r}.$$

Since L_0 is selected using the distribution described on page 15, it follows by part (b) of Lemma 3.2 that rows $c+1, c+2, \ldots, r$ of C_0 are linearly dependent with probability at most

$$\frac{q^{-c}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}.$$

The corresponding rows of C are linearly dependent with the same probability, since these are the same rows.

Recall that the entries of L_1 in rows $1, 2, \ldots, c$, and $r+1, r+2, \ldots, c$, are chosen uniformly and independently from F. It follows that the entries of C in rows $1, 2, \ldots, c$ and $r+1, r+2, \ldots, r+c$ are chosen uniformly and independently from F, as well.

Suppose that rows $c+1, c+2, \ldots, r$ of C are linearly independent. Then there exists a set S of $r-c$ columns of C such that the submatrix of C including the entries in rows $c+1, c+2, \ldots, r$ and in the columns in set S is nonsingular.

In order the complete the analysis, in the case that $\widetilde{A} = I_r$, suppose that we begin with an assignment of values for the entries of L_1 in rows $c+1, c+2, \ldots, r$ and the columns in S, and that we fill in entries in each of the remaining columns, one at a time. Since the remaining entries are uniformly and independently selected from F, we can see that the probability that the resulting columns of C are *not* linearly independent is at most

$$\sum_{c+1\leq i\leq 2c} q^{-i} \leq \sum_{i\geq c+1} q^{-i} = \frac{q^{-c}}{q-1}.$$

Thus the probability that C has rank less than r, when $\overline{A} = I_r$, is at most

$$\left(\frac{q^{-c}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}\right) + \frac{q^{-c}}{q-1} = \frac{2q^{-c}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}.$$

Suppose, next, that $\widetilde{A}$ is an arbitrary nonsingular matrix in $\mathsf{F}^{r\times r}$. Then

$$C = L_1\widetilde{A} + B = C' \cdot \widetilde{A},$$

for $C' = L_1 + B'$ and $B' = B \cdot \widetilde{A}^{-1}$.

It follows by the above analysis, using B' in place of B, that the matrix C' has rank less than r with probability at most

$$\frac{2q^{-c}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}.$$

Since $\widetilde{A}$ is nonsingular, the matrices C and C' have the same rank. Thus C has rank less than r with probability at most $2q^{-c} + N^{1-\frac{\widehat{c}}{\ln q}}$ as well. □

Notice that the transpose of the matrix R_1 is chosen using the same probability distribution as described for L_1. The next result can be obtained as a consequence of the previous one by considering the transpose of the matrix D that is mentioned below.

Corollary 3.5. *Let $\widetilde{A} \in \mathsf{F}^{r\times r}$ be a nonsingular matrix. Let $B \in \mathsf{F}^{r\times(r+c)}$, and suppose R_1 is chosen as described on page 18. Let*

$$D = \widetilde{A}R_1 + B \in \mathsf{F}^{r\times(r+c)}.$$

Then the probability that columns $c, c+1, \ldots, r$ of D are linearly dependent or D has rank less than r is at most

$$\frac{2q^{-c}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}.$$

3.4.3 Preservation of Rank

If the matrix $\widehat{A}$ is as given in equation (3.5), then the rank of $\widehat{A}$ is at most that of A. The following result therefore implies that the ranks of A and $\widehat{A}$ are the same, with high probability.

Theorem 3.6. *Let $A \in \mathsf{F}^{n\times m}$ be a matrix with rank r.*

Suppose the matrices $L \in \mathsf{F}^{\ell\times n}$ and $R \in \mathsf{F}^{m\times\ell}$ are randomly chosen as described in Section 3.4.1, and that $\widehat{A} = LAR$.

(a) The probability that either rows

$$c+1, c+2, \ldots, r$$

of the matrix LA are linearly dependent, or that the $(r+c)\times m$ submatrix of LA that includes rows with indices

$$1, 2, \ldots, r \qquad \textit{and} \qquad \min(n,m)+1, \min(n,m)+2, \ldots, \min(n,m)+c$$

has rank less than r, is at most

$$\frac{2q^{-c}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}.$$

(b) The probability that $\widehat{A}$ has rank less than r is at most

$$\frac{4q^{-c}}{q-1} + 2N^{1-\frac{\widehat{c}}{\ln q}}.$$

Proof. Let us begin by considering the matrix LA, noting that this does not depend in any way on the choice of the matrix R.

Since A has rank r, there exist permutations $P \in \mathsf{F}^{n\times n}$ and $Q \in \mathsf{F}^{m\times m}$ such that the leading $r \times r$ submatrix of PAQ is nonsingular.

Notice that, since P^{-1} is also a permutation matrix, the matrices L and $L \cdot P^{-1}$ are chosen according to the same probability distribution. Now

$$LA = (L \cdot P^{-1}) \cdot (PAQ) \cdot Q^{-1},$$

and this matrix has the same rank as that of the matrix $LA \cdot Q = (L \cdot P^{-1}) \cdot (PAQ)$. We may therefore assume without loss of generality that the leading $r \times r$ submatrix of A is nonsingular in the rest of this proof.

Using this simplifying assumption, let us write A as

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}$$

where $A_{1,1}$ is now a nonsingular matrix in $\mathsf{F}^{r\times r}$, and where $A_{1,2} \in \mathsf{F}^{r\times(m-r)}$, $A_{2,1} \in \mathsf{F}^{(n-r)\times r}$, and $A_{2,2} \in \mathsf{F}^{(n-r)\times(m-r)}$.

Consider the $(r + c) \times m$ submatrix of LA that includes rows with indices

$$1, 2, \ldots, r \qquad \text{and} \qquad \min(n, m) + 1, \min(n, m) + 2, \ldots, \min(n, m) + c.$$

This can be written as

$$\begin{bmatrix} L_1 & L_2 \end{bmatrix} \cdot \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}$$

where $L_1 \in \mathsf{F}^{(r+c)\times r}$, $L_2 \in \mathsf{F}^{(r+c)\times(n-r)}$, and where L_1 is chosen using the probability distribution described on page 18.

The submatrix consisting of the leftmost r columns of this matrix is

$$\begin{bmatrix} L_1 & L_2 \end{bmatrix} \cdot \begin{bmatrix} A_{1,1} \\ A_{2,1} \end{bmatrix} = L_1 \cdot A_{1,1} + L_2 \cdot A_{2,1} \in \mathsf{F}^{(r+c)\times r}.$$

If we choose (and fix) the values in the matrix L_2 then, since $A_{1,1} \in \mathsf{F}^{r\times r}$ is nonsingular and $L_2 \cdot A_{2,1} \in \mathsf{F}^{(r+c)\times r}$, it follows by Lemma 3.4 that (when the entries of L_1 are selected) the likelihood that rows

$$c + 1, c + 2, \ldots, r$$

of this matrix are linearly dependent or that this matrix has rank less than r is at most

$$\frac{2q^{-c}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}.$$

Since this is true for any choice of values for L_2 it follows — as claimed in part (a) above — that this is also an upper bound on the probability that rows $c + 1, c + 2, \ldots, r$ of LA are linearly dependent or LA has rank less than r, when L is randomly chosen as described above.

Now consider any fixed matrix $L \in \mathsf{F}^{\ell\times n}$ such that the matrix

$$LA \in \mathsf{F}^{\ell\times m}$$

has rank r. Repeating the above argument, using Corollary 3.5 in place of Lemma 3.4, one can establish that the probability that

$$\widehat{A} = LAR$$

has rank less than r is at most

$$\frac{2q^{-c}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}$$

as well. Thus, if L and R are randomly chosen as above, then the probability that LAR has rank less than r is at most twice the above value, as claimed in part (b). □

3.4.4 Bounding the Number of Nilpotent Blocks

Consider the number of invariant factors of $\widehat{A}$ that are divisible by x^2. This is the same as the number of nontrivial nilpotent blocks in a Jordan form of $\widehat{A}$.

A consideration of the Jordan form establishes that this is also the difference between the rank of $\widehat{A}$ and that of $\widehat{A}^2$. In this section we will establish upper bounds on the expected value of this difference and on the probability that this difference is high.

Lemma 3.7. *Let $L \in \mathsf{F}^{\ell \times n}$, $A \in \mathsf{F}^{n \times m}$, $R \in \mathsf{F}^{m \times \ell}$, and let $\widehat{A} = LAR \in \mathsf{F}^{\ell \times \ell}$. Suppose that*

$$\operatorname{rank}(A) = \operatorname{rank}(\widehat{A}).$$

Then

$$\operatorname{rank}(\widehat{A}^2) = \operatorname{rank}(ARLA).$$

Proof. Since

$$\widehat{A}^2 = (LAR)^2 = L \cdot (ARLA) \cdot R,$$

it is clear that $\operatorname{rank}(\widehat{A}^2) \leq \operatorname{rank}(ARLA)$. It is therefore sufficient to prove that

$$\operatorname{rank}(\widehat{A}^2) \geq \operatorname{rank}(ARLA)$$

as well in order to establish the claim.

Let $s = \operatorname{rank}(ARLA)$. Then the column space of $ARLA$ has dimension s and there exist vectors

$$x_1, x_2, \ldots, x_s \in \mathsf{F}^{m \times 1}$$

such that the vectors

$$y_1, y_2, \ldots, y_s \in \mathsf{F}^{n \times 1}$$

are linearly independent if

$$y_i = ARLAx_i \qquad \text{for } 1 \leq i \leq s.$$

Suppose that

$$\operatorname{rank}(A) = \operatorname{rank}(\widehat{A}).$$

Then, since $\widehat{A} = LAR$,

$$\operatorname{rank}(A) = \operatorname{rank}(LAR) \leq \operatorname{rank}(AR) \leq \operatorname{rank}(A),$$

and $\operatorname{rank}(AR) = \operatorname{rank}(A)$ as well. The matrices AR and A therefore have the same column space.

Let $w_i = Ax_i$ for $1 \leq i \leq s$. Then $w_1, w_2, \ldots, w_s$ belong to the column space of A, so that they belong to the column space of AR. It follows that there exist vectors

$$z_1, z_2, \ldots, z_s \in \mathsf{F}^{\ell \times 1}$$

such that

$$w_i = Ax_i = ARz_i \qquad \text{for } 1 \leq i \leq s.$$

Consequently

$$y_i = ARLAx_i = ARLw_i = ARLARz_i \qquad \text{for } 1 \leq i \leq s$$

as well.

Once again, since $\text{rank}(A) = \text{rank}(\widehat{A}) = \text{rank}(LAR)$,

$$\text{rank}(A) = \text{rank}(LAR) \leq \text{rank}(LA) \leq \text{rank}(A)$$

so that $\text{rank}(LA) = \text{rank}(A)$. It follows that the matrices LA and A have the same right null space: For any vector $v \in \mathsf{F}^{m \times 1}$,

$$LAv = 0 \qquad \Longleftrightarrow \qquad Av = 0.$$

This can be used to establish that if $k \geq 0$ and $v_1, v_2, \ldots v_k \in \mathsf{F}^{m \times 1}$, then

$$LAv_1, LAv_2, \ldots, LAv_k \text{ are linearly independent}$$

if and only if

$$Av_1, Av_2, \ldots, Av_k \text{ are linearly independent.}$$

In particular, this is the case if $k = s$ and $v_i = RLARz_i$ for $1 \leq i \leq s$. That is, since $Av_i = ARLARz_i = y_i$ for $1 \leq i \leq k$,

$$Ly_1, Ly_2, \ldots, Ly_s \text{ are linearly independent,}$$

because

$$y_1, y_2, \ldots, y_s \text{ are linearly independent.}$$

It remains only to note that

$$Ly_i = LARLARz_i = \widehat{A}^2 z_i \qquad \text{for } 1 \leq i \leq s.$$

We have now established that the vectors

$$\widehat{A}^2 z_1, \widehat{A}^2 z_2, \ldots, \widehat{A}^2 z_s$$

are linearly independent, implying that

$$\text{rank}(\widehat{A}) \geq s = \text{rank}(ARLA),$$

as required to complete the proof. □

Lemma 3.8. *Let $A \in \mathsf{F}^{n\times m}$ be a matrix with rank r, and let $L \in \mathsf{F}^{\ell\times n}$ be any matrix such that rows $c+1, c+2, \ldots, r$ of LA are linearly independent and the submatrix of LA including rows*

$$1, 2, \ldots, r \qquad \textit{and} \qquad \min(n,m)+1, \min(n,m)+2, \ldots, \min(n,m)+c$$

has rank r.

Suppose that the matrix $R \in \mathsf{F}^{m\times\ell}$ is randomly selected as described on page 14.

Then, for any integer $i \geq 0$, the probability that the matrix $ARLA$ has rank less than $r-i$ is at most

$$\frac{q^{1-i}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}.$$

Proof. Since A has rank r, there exists a permutation matrix $P \in \mathsf{F}^{m\times m}$ such that the leftmost r columns of the matrix AP are linearly independent. There is therefore a set

$$S \subseteq \{1, 2, \ldots, n\}$$

of size r such that the $r \times r$ submatrix of AP, including the rows in S and the leftmost r columns, is nonsingular. The $r \times m$ submatrix of AP that includes the rows in S therefore has the form

$$\begin{bmatrix} \widetilde{A} & C \end{bmatrix}$$

where $\widetilde{A} \in \mathsf{F}^{r\times r}$ is a nonsingular matrix and where $B \in \mathsf{F}^{r\times(m-r)}$.

Suppose L has the properties given above. Then there exists a permutation matrix $Q \in \mathsf{F}^{\ell\times\ell}$, with $(i,j)^{\text{th}}$ entry $Q_{i,j}$ for $1 \leq i, j \leq \ell$, such that $Q_{i,i} = 1$ for $c+1 \leq i \leq r$ and such that the top r rows of the matrix QLA are linearly independent. There is therefore a set

$$T \subseteq \{1, 2, \ldots, m\}$$

of size r such that the $r \times r$ submatrix of QLA including the top r rows and the columns in T is nonsingular. The $\ell \times r$ submatrix of QLA that includes the columns in T therefore has the form

$$\begin{bmatrix} \widehat{A} \\ D \end{bmatrix}$$

where $\widehat{A} \in \mathsf{F}^{r\times r}$ is a nonsingular matrix and where $D \in \mathsf{F}^{(\ell-r)\times r}$.

Notice that

$$ARLA = AP \cdot (P^{-1}RQ^{-1})(QL)A$$

and — since P and Q are permutation matrices, and $Q_{i,i} = 1$ for $c+1 \leq i \leq r$ — the matrices R and $P^{-1}RQ^{-1}$ are chosen using the same probability distribution. Consequently,

$$P^{-1}RQ^{-1} = \begin{bmatrix} R_0 & R_{1,2} \\ R_{2,1} & R_{2,2} \end{bmatrix}$$

where $R_0 \in \mathsf{F}^{r\times r}$, $R_{1,2} \in \mathsf{F}^{r\times(\ell-r)}$, $R_{2,2} \in \mathsf{F}^{(m-r)\times r}$, $R_{2,2} \in \mathsf{F}^{(m-r)\times(\ell-r)}$, and where the matrix $R_0 \in \mathsf{F}^{r\times r}$ is randomly chosen using the probability distribution described on page 15.

Let us now consider the $r \times r$ submatrix of $ARLA$ that includes the rows in S and the columns in T. This matrix has the form

$$\begin{aligned}
&\begin{bmatrix} \widetilde{A} & C \end{bmatrix} \cdot \begin{bmatrix} R_0 & R_{1,2} \\ R_{2,1} & R_{2,2} \end{bmatrix} \cdot \begin{bmatrix} \widehat{A} \\ D \end{bmatrix} \\
&\qquad = \begin{bmatrix} \widetilde{A} & C \end{bmatrix} \cdot \begin{bmatrix} R_0\widehat{A} + R_{1,2}D \\ R_{2,1}\widehat{A} + R_{2,2}D \end{bmatrix} \\
&\qquad = \widetilde{A}R_0\widehat{A} + \widetilde{A}R_{1,2}D + CR_{2,1}\widehat{A} + CR_{2,2}D \\
&\qquad = \left(\widetilde{A}R_0 + B\right) \cdot \widehat{A},
\end{aligned}$$

where

$$B = \widetilde{A}R_{1,2}D\widehat{A}^{-1} + CR_{2,1} + CR_{2,2}D\widehat{A}^{-1} \in \mathsf{F}^{r\times r}.$$

Fix any choice of values for the entries of the matrices $R_{1,2}$, $R_{2,1}$, and $R_{2,2}$; then the entries of the above matrix $B \in \mathsf{F}^{r\times r}$ are fixed as well. Since R_0 is chosen using the above-mentioned probability distribution, it follows by part (c) of Corollary 3.3 that the matrix $\widetilde{A}R_0 + B$ has rank less than $r - i$ with probability at most

$$\frac{q^{1-i}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}.$$

Since the matrix $\widehat{A}$ is nonsingular, the matrices $\widehat{A}R_{1,1} + B$ and $\left(\widehat{A}R_{1,1} + B\right) \cdot \widehat{A}$ have the same rank. Since the latter matrix is a submatrix of $ARLA$, it follows that $ARLA$ has rank less than $r - i$ with probability at most

$$\frac{q^{1-i}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}$$

as well. □

Theorem 3.9. *Let $A \in \mathsf{F}^{n\times m}$ be a matrix with rank r.*

Suppose the matrices $L \in \mathsf{F}^{\ell\times n}$ and $R \in \mathsf{F}^{m\times\ell}$ are randomly chosen as described in Section 3.4.1, and that $\widehat{A} = LAR$.

(a) The probability that i or more of the invariant factors of $\widehat{A}$ are divisible by x^2 is at most

$$\frac{q^{2-i} + 2q^{-c}}{q-1} + 2N^{1-\frac{\widehat{c}}{\ln q}}.$$

(b) The expected number of invariant factors of $\widehat{A}$ that are divisible by x^2 is at most

$$\frac{q^2}{(q-1)^2} + \frac{2Nq^{-c}}{q-1} + 2N^{2-\frac{\widehat{c}}{\ln q}}.$$

Proof. Let C denote the condition that rows $c+1, c+2, \ldots, r$ of the matrix LA are linearly independent and the submatrix of LA including rows

$$1, 2, \ldots, r \qquad \text{and} \qquad \min(n,m)+1, \min(n,m)+2, \ldots, \min(n,m)+c$$

has rank r. This event depends on L, but not on the choice of R, and it follows by part (a) of Theorem 3.6 that

$$\text{Prob}\left(\neg \mathsf{C}\right) \leq \frac{2q^{-c}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}. \tag{3.6}$$

On the other hand, it is easily established by Lemma 3.8 that

$$\text{Prob}\left(\mathsf{C} \wedge (\text{rank}(ARLA) < r-i)\right) \leq \frac{q^{1-i}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}$$

for any integer $i \geq 0$, because the above quantity bounds the probability that the rank of $ARLA$ is less than $r-i$ for any choice of the matrix L such that condition C holds. Since the rank is an integer value, it is less than $r-i$ if and only if it is less than or equal to $r-(i+1) = r-i-1$. Thus

$$\text{Prob}\left(\mathsf{C} \wedge (\text{rank}(ARLA) \leq r-i)\right) \leq \frac{q^{2-i}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}. \tag{3.7}$$

Notice that if the condition C holds, then the number of invariant factors of $\widehat{A}$ that are divisible by x^2 is

$$\begin{aligned}
\text{rank}(\widehat{A}) - \text{rank}(\widehat{A}^2) &\leq r - \text{rank}(\widehat{A}^2) \\
&= r - \text{rank}(ARLA) \qquad \text{(by Lemma 3.7).}
\end{aligned}$$

Thus the number of invariant factors of $\widehat{A}$ divisible by x^2 can only be greater than or equal to i, when the condition C holds, if the rank of $ARLA$ is less than or equal to $r-i$. It therefore follows, using Equations (3.6) and (3.7), that the probability that $\widehat{A}$ has at least i invariant factors divisible by x^2 is at most

$$\frac{q^{2-i} + 2q^{-c}}{q-1} + 2N^{1-\frac{\widehat{c}}{\ln q}},$$

as required to establish part (a).

Now let X be the number of invariant factors of $\widehat{A}$ that are divisible by x^2, so that X is an integer-valued random variable that can assume values

$$0, 1, \ldots, r.$$

Then the expected value of X is

$$\begin{aligned}
E\left[\mathsf{X}\right] &= \sum_{i=1}^{r} \text{Prob}\left(\mathsf{X} \geq i\right) \\
&= \sum_{i=1}^{r} \Big(\text{Prob}\left(\mathsf{C} \wedge \mathsf{X} \geq i\right) + \text{Prob}\left(\neg\mathsf{C} \wedge \mathsf{X} \geq i\right)\Big) \\
&\leq \sum_{i=1}^{r} \Big(\text{Prob}\left(\mathsf{C} \wedge \mathsf{X} \geq i\right) + \text{Prob}\left(\neg\mathsf{C}\right)\Big) \\
&= \sum_{i=1}^{r} \text{Prob}\left(\mathsf{C} \wedge \mathsf{X} \geq i\right) + r \cdot \text{Prob}\left(\neg\mathsf{C}\right) \\
&\leq \sum_{i=1}^{r} \left(\frac{q^{2-i}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}\right) + \frac{2rq^{-c}}{q-1} + rN^{1-\frac{\widehat{c}}{\ln q}} \qquad \text{(by Equations (3.6) and (3.7))}
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{i\geq 1} \frac{q^{2-i}}{q-1} + rN^{1-\frac{\widehat{c}}{\ln q}} + \frac{2rq^{-c}}{q-1} + rN^{1-\frac{\widehat{c}}{\ln q}} \\
&= \frac{q^2}{(q-1)^2} + \frac{2rq^{-c}}{q-1} + 2rN^{1-\frac{\widehat{c}}{\ln q}} \\
&\leq \frac{q^2}{(q-1)^2} + \frac{2Nq^{-c}}{q-1} + 2N^{2-\frac{\widehat{c}}{\ln q}} && \text{(since } r \leq N\text{)}
\end{aligned}
$$

as required to establish part (b). □

3.4.5 Bounding the Number of Invariant Factors That are Not Powers of x

Next consider the number of invariant factors of $\widehat{A}$ that are not powers of x. Suppose that there are k such factors, so that the first k invariant factors of $\widehat{A}$ are

$$f_1, f_2, \ldots, f_k$$

where f_i is divisible by f_{i+1}, for $1 \leq i \leq k-1$, and where f_k has a nonzero root, λ, in some extension of F. Then λ is a root of f_i as well, for $1 \leq i \leq k$.

It follows that the Jordan form of $\widehat{A}$ (over a suitable extension of F) includes k blocks with eigenvalue λ, and that the matrix

$$\widehat{A} - \lambda I_\ell$$

has nullity k. We will use this observation to bound the number of invariant factors that are not powers of x.

Suppose, for the rest of this section, that E is an algebraic closure of the finite field F.

Lemma 3.10. *Let $\widetilde{A} \in \mathsf{F}^{r\times r}$ be a nonsingular matrix, and that $B \in \mathsf{E}^{r\times r}$.*

Let i be an integer such that $1 \leq i \leq r$.

Suppose that $R_0 \in \mathsf{F}^{r\times r}$ is a matrix whose i^{th} column is chosen as described for the probability distribution on page 15. Let

$$D \in \widetilde{A}R_0 + B \in \mathsf{E}^{r\times r}.$$

Then the probability that column i of D is a linear combination of columns $i+1, i+2, \ldots, r$ is at most

$$q^{-i} + N^{-\frac{\widehat{c}}{\ln q}}.$$

Proof. This can be established using a straightforward modification of the the proof of part (a) of Lemma 3.2 and Corollary 3.3 — which correspond to the case that the above matrix B has elements in the field F.

In the original proof, (after restricting attention to the special case that $\widetilde{A}$ is the identity matrix) one supposes that the submatrix $\widehat{D}_i$ consisting or rows

$$i+1, i+2, \ldots, r$$

of D has rank s_i. Then, after choosing values for all but $r - s_i$ of the entries in column i of R_0 we observe that there is exactly one choice of the remaining values (in F) for the remaining $r - s_i$ entries in column i such that the i^{th} column of D is a linear combination of columns $i+1, i+2, \ldots, r$.

In order to prove the above result one works in almost the same way: One chooses values for all but $r - s_i$ of the entries in column i of R_0. There is now exactly one choice of the remaining values — in the extension E — for the remaining $r - s_i$ entries in column i such that the i^{th} column of D is a linear combination of columns $i+1, i+2, \ldots, r$. If any of the remaining values (to be assigned) lie outside of F, then it is impossible to complete the choice of column i of R_0, in such a way that the i^{th} column of D is a linear combination of columns $i+1, i+2, \ldots, r$. Otherwise (the remaining values to be selected lie in F) the probability that the values are selected in the required way can be bounded using the argument given in Lemma 3.2.

To complete the proof, one removes the assumption that $\widetilde{A}$ is the identity matrix using the same argument as is used in the proof of Lemma 3.2. □

Lemma 3.11. *Let $A \in \mathsf{F}^{n\times m}$ be a matrix with rank r and let $L \in \mathsf{F}^{\ell\times n}$ be a matrix such that rows $c+1, c+2, \ldots, r$ of the matrix LA are linearly independent and the submatrix of LA including rows*

$$1, 2, \ldots, r \qquad \textit{and} \qquad \min(n,m)+1, \min(n,m)+2, \ldots, \min(n,m)+c$$

has rank r.

Suppose the matrix $R \in \mathsf{F}^{m\times\ell}$ is randomly chosen using the probability distribution described in Section 3.4.1.

Let i be an integer such that $1 \leq i \leq r$. Then the probability that the matrix LAR has i or more invariant factors that are not powers of x is at most

$$(r-i+1)\frac{q^{2-i}}{q-1} + (r-i+1)^2 N^{-\frac{\widehat{c}}{\ln q}}.$$

Proof. Suppose that L is as described in the above claim. Then there exists a permutation matrix $P \in \mathsf{F}^{\ell\times\ell}$ with $(i,j)^{\text{th}}$ entry $P_{i,j}$ for $1 \leq i, j \leq \ell$ such that $P_{i,i} = 1$ if $c+1 \leq i \leq \min(n,m)$ and such that the top r rows of the matrix PLA are linearly independent. One can see, by inspection of the probability distribution described in Section 3.4.1, that the matrices R and RP^{-1} are chosen using the same probability distribution.

Since the matrices LAR and $PLARP^{-1}$ are similar, they have the same invariant factors.

It follows that — replacing matrices L and B with the matrices PL and RP^{-1}, respectively — we may assume without loss of generality that the top r rows of the matrix LA are linearly independent.

In this case, there exists a permutation matrix $Q \in \mathsf{F}^{m\times m}$ such that the top left $r \times r$ submatrix of LA is nonsingular. Note that, since Q^{-1} is also a permutation matrix, the i^{th} column of the matrix $Q^{-1}R$ is chosen using the same probability distribution as the i^{th} column of R. Clearly

$$LAR = LAQ \cdot Q^{-1}R.$$

It follows that — replacing matrices L and R with LQ and $Q^{-1}R$, respectively — we may now assume that the top left $r \times r$ submatrix of LA is nonsingular.

We may therefore write LA as

$$LA = \begin{bmatrix} \widetilde{A}_{1,1} & \widetilde{A}_{1,2} \\ \widetilde{A}_{2,1} & \widetilde{A}_{2,2} \end{bmatrix} \tag{3.8}$$

where $\widetilde{A}_{1,1} \in \mathsf{F}^{r\times r}$ is nonsingular, and where $\widetilde{A}_{1,2} \in \mathsf{F}^{r\times(m-r)}$, $\widetilde{A}_{2,1} \in \mathsf{F}^{(\ell-r)\times r}$, and $\widetilde{A}_{2,2} \in \mathsf{F}^{(\ell-r)\times(m-r)}$.

Note that

$$R = \begin{bmatrix} R_0 & R_{1,2} \\ R_{2.1} & R_{2.2} \end{bmatrix} \tag{3.9}$$

where $R_0 \in \mathsf{F}^{r\times r}$ is randomly chosen using the probability distribution given on page 15, and where $R_{1,2} \in \mathsf{F}^{r\times(\ell-r)}$, $R_{2,1} \in \mathsf{F}^{(m-r)\times r}$, and $R_{2,2} \in \mathsf{F}^{(m-r)\times(\ell-r)}$.

Let $B \in \mathsf{F}^{r\times r}$ be the top left $r \times r$ submatrix of LAR. Then if follows by the decompositions given in equations (3.8) and (3.9), above, that

$$B = \begin{bmatrix} \widetilde{A}_{1,1} & \widetilde{A}_{1,2} \end{bmatrix} \cdot \begin{bmatrix} R_0 \\ R_{2,1} \end{bmatrix} = \widetilde{A}_{1,1}R_0 + \widetilde{A}_{1,2}R_{2,1} \in \mathsf{F}^{r\times r}. \tag{3.10}$$

Consequently if λ is an element of an algebraic closure E of F then the top left $r \times r$ submatrix of $LAR - \lambda I_\ell$ is

$$B - \lambda I_r = \widetilde{A}_{1,1}R_0 + (\widetilde{A}_{1,2}R_{2,1} - \lambda I_r).$$

Suppose that $i \geq 2$, and recall that if B has i or more invariant factors that are not powers of x, then there exists a nonzero element λ of E such that

$$\operatorname{rank}(B - \lambda I_r) \leq r - i. \tag{3.11}$$

Let $B_j \in \mathsf{F}^{r\times(r-j)}$ denote the submatrix of B that includes columns $j+1, j+2, \ldots, r$, and let $I_{r,j} \in \mathsf{F}^{r\times(r-j)}$ denote the submatrix of the $r \times r$ identity matrix I_r that includes columns $j+1, j+2, \ldots, r$. Then condition (3.11) clearly implies that

$$\operatorname{rank}(B_{i-2} - \lambda I_{r,i-2}) \leq r - i \tag{3.12}$$

as well.

With this in mind, for $2 \leq i \leq r+1$, let p_i denote the probability that there exists a nonzero element λ of the algebraic closure E such that

$$\operatorname{rank}(B_{i-2} - \lambda I_{r,i-2}) \leq r - i.$$

Clearly

$$p_{r+1} = 0, \tag{3.13}$$

since it is impossible for the matrix $B_{i-2} - \lambda I_{r,i-2}$ to have a negative rank for any choice of λ.

Suppose now that $2 \leq i \leq r$ and that there does not exist any nonzero element λ of E such that

$$\operatorname{rank}(B_{i-1} - \lambda I_{r,i-1}) \leq r - i - 1.$$

Let $C_{i-1} \in \mathsf{F}^{(r-i+1)\times(r-i+1)}$ be the submatrix of B_{i-1} that includes rows $i, i+1, \ldots, r$; then C_{i-1} is also the bottom right $(r-i+1) \times (r-i+1)$ submatrix of B.

Let f_{i-1} be the characteristic polynomial of C_{i-1}. Then f_{i-1} is a nonzero polynomial with degree $r - i + 1$.

Consider any element λ of E. Clearly, either λ is a root of f_{i-1} or it is not; these cases will be considered separately.

First consider the case that λ is not a root of f_{i-1}. In this case, the matrix

$$C_{i-1} - \lambda I_{r-i+1} \in \mathsf{E}^{(r-i+1)\times(r-i+1)}$$

is nonsingular. Since this is the bottom submatrix of the $r \times (r-i+1)$ matrix $B_{i-1} - \lambda I_{r,i-1}$, it follows that

$$\operatorname{rank}(B_{i-1} - \lambda I_{r,i-1}) = r - i + 1.$$

Since $B_{i-1} - \lambda I_{r,i-1}$ is the submatrix of $B_{i-2} - \lambda I_{r,i-2}$ containing the rightmost $r-i+1$ columns, it follows that

$$\operatorname{rank}(B_{i-2} - \lambda I_{r,i-2}) \geq r - i + 1$$

as well. Thus condition (3.12) cannot be satisfied in this case.

Next consider the case that λ is a root of f_{i-1}, and recall the assumption that

$$\operatorname{rank}(B_{i-1} - \lambda I_{r,i-1}) \geq r - i.$$

If $\operatorname{rank}(B_{i-1} - \lambda I_{r,i-1}) \geq r-i+1$ then, as noted in the discussion of the previous case, this implies that

$$\operatorname{rank}(B_{i-2} - \lambda I_{r,i-2}) \geq r - i + 1$$

as well, making condition (3.12) impossible. It is therefore sufficient to consider the case that

$$\operatorname{rank}(B_{i-1} - \lambda I_{r,i-1}) = r - i.$$

It would follow in this case that condition (3.12) is satisfied, for this choice of λ, if and only column $i-1$ of the matrix

$$B - \lambda I_r = \widetilde{A}_{1,1} R_0 + (\widetilde{A}_{1,2} R_{2,1} - \lambda I_r)$$

is a linear combination of columns $i, i+1, \ldots, r$ of this matrix. Applying Lemma 3.10 (for any choice of entries of column $i-1$ of the submatrix $R_{2,1}$), we see that this probability of this is at most

$$q^{1-i} + N^{-\frac{\widehat{c}}{\ln q}}.$$

The polynomial f_{i-1} has at most $r-i+1$ roots. Over-approximating the probability of the union of events by the sum of the probabilities of the events, we may now conclude that

$$p_i \leq p_{i+1} + (r-i+1)\left(q^{1-i} + N^{-\frac{\widehat{c}}{\ln q}}\right) \tag{3.14}$$

for any integer i such that $2 \leq i \leq r$.

Using equations (3.13) and (3.14), it is easily established by induction on $r-i$ that if $2 \leq i \leq r$ then

$$p_i \leq (r-i+1)\frac{q^{2-i}}{q-1} + (r-i+1)^2 N^{-\frac{\widehat{c}}{\ln q}}. \tag{3.15}$$

To conclude, recall the assumption that the top $r \times r$ submatrix $\widetilde{A}_{1,1}$ of LA is nonsingular. The matrix LA must then have rank r, since A does. It follows that if LA is as shown in equation (3.8), and

$$X = -\widetilde{A}_{2,1}\widetilde{A}_{1,1}^{-1} \in \mathsf{F}^{(\ell-r)\times r},$$

then

$$\begin{bmatrix} I_r & 0 \\ X & I_{\ell-r} \end{bmatrix} \cdot LA$$

$$= \begin{bmatrix} I_r & 0 \\ X & I_{\ell-r} \end{bmatrix} \cdot \begin{bmatrix} \widetilde{A}_{1,1} & \widetilde{A}_{1,2} \\ \widetilde{A}_{2,1} & \widetilde{A}_{2,2} \end{bmatrix}$$
$$= \begin{bmatrix} \widetilde{A}_{1,1} & \widetilde{A}_{1,2} \\ 0 & 0 \end{bmatrix},$$

since the choice of X ensures that the bottom left $(\ell - r) \times r$ submatrix of this product is zero, and since the right $(\ell - r)$ columns are linear combinations of the left r columns.

Applying the decomposition of R in equation (3.9) as well, one finds that

$$LAR = \begin{bmatrix} \widetilde{A}_{1,1} & \widetilde{A}_{1,2} \\ \widetilde{A}_{2,1} & \widetilde{A}_{2,2} \end{bmatrix} \cdot \begin{bmatrix} R_0 & R_{1,2} \\ R_{2,1} & R_{2,2} \end{bmatrix},$$

so that

$$\begin{bmatrix} I_r & 0 \\ X & I_{\ell-r} \end{bmatrix} \cdot LAR$$
$$= \begin{bmatrix} I_r & 0 \\ X & I_{\ell-r} \end{bmatrix} \cdot \begin{bmatrix} \widetilde{A}_{1,1} & \widetilde{A}_{1.2} \\ \widetilde{A}_{2,1} & \widetilde{A}_{2,2} \end{bmatrix} \cdot \begin{bmatrix} R_0 & R_{1,2} \\ R_{2,1} & R_{2,2} \end{bmatrix}$$
$$= \begin{bmatrix} \widetilde{A}_{1,1} & \widetilde{A}_{1,2} \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} R_0 & R_{1,2} \\ R_{2,1} & R_{2,2} \end{bmatrix}$$
$$= \begin{bmatrix} B & C \\ 0 & 0 \end{bmatrix}$$

where B is the top left $\ell \times \ell$ submatrix of LAR and where $C \in \mathsf{F}^{r \times (\ell-r)}$ is the top right $r \times (\ell - r)$ submatrix of LAR.

Consequently, if λ is a nonzero element of E, then

$$\begin{bmatrix} I_r & 0 \\ X & I_{\ell-r} \end{bmatrix} \cdot (LAR - \lambda I_\ell) = \begin{bmatrix} B - \lambda I_r & 0 \\ -\lambda X & -\lambda I_{\ell-r} \end{bmatrix}$$

and, clearly,

$$\operatorname{rank}(LAR - \lambda I_\ell) = \operatorname{rank}(B - \lambda I_r) + \ell - r.$$

It follows that $\operatorname{rank}(B - \lambda I_r) \leq r - i$ if and only if $\operatorname{rank}(LAR - \lambda I_\ell) \leq \ell - i$.

Consequently, if L has the properties described in the claim, and R is randomly chosen as described, then the probability that LAR has i or more invariant factors that are not powers of x is the same as the probability that B does. Since this probability is at most p_i, the claim now follows by inequality (3.15), above. □

Theorem 3.12. *Let $A \in \mathsf{F}^{n \times m}$ be a matrix with rank r.*

Suppose the matrices $L \in \mathsf{F}^{\ell \times n}$ and $R \in \mathsf{F}^{m \times \ell}$ are randomly chosen, as described in Section 3.4.1. Let $\widehat{A} = LAR$.

(a) Suppose i is an integer such that $2 \leq i \leq r$. Then the probability that $\widehat{A}$ has at least i invariant factors that are not powers of x is at most

$$\frac{(r - i + 1)q^{2-i} + 2q^{-c}}{q - 1} + (r - i + 2)N^{1 - \frac{\widehat{c}}{\ln q}}.$$

(b) The expected number of invariant factors of $\widehat{A}$ that are not powers of x is at most

$$\log_q r + 4 + \frac{2rq^{-c}}{q-1} + r^2 N^{1-\frac{\widehat{c}}{\ln q}}.$$

Proof. Let $\mathsf{fac}(LAR)$ denote the number of invariant factors of the matrix LAR that are not powers of x.

Consider the condition C that rows $c+1, c+2, \ldots, r$ of the matrix LA are linearly independent that the $(r+c) \times m$ submatrix of LA that includes rows

$$1, 2, \ldots, r \qquad \text{and} \qquad \min(n,m)+1, \min(n,m)+2, \ldots, \min(n,m)+r$$

has rank r. This event depends only on the choice of the matrix L, and it follows by part (a) of Theorem 3.6 that

$$\text{Prob}\left(\neg\mathsf{C}\right) \leq \frac{2q^{-c}}{q-1} + N^{1-\frac{\widehat{c}}{\ln q}}. \tag{3.16}$$

On the other hand, it follows by Lemma 3.11 that

$$\text{Prob}\left(\mathsf{C} \ \wedge \mathsf{fac}(LAR) \geq i\right) \leq (r-i+1)\frac{q^{2-i}}{q-1} + (r-i+1)^2 N^{-\frac{\widehat{c}}{\ln q}}, \tag{3.17}$$

because the above quantity bounds the probability that LAR has at least i invariant factors that are not powers of x, for any choice of the matrix L such that condition C holds.

Since

$$\text{Prob}\left(\neg\mathsf{C} \ \wedge \mathsf{fac}(LAR) \geq i\right) \leq \text{Prob}\left(\neg\mathsf{C}\right),$$

it follows by inequalities (3.16) and (3.17) and the fact that

$$(r-i+1)^2 N^{-\frac{\widehat{c}}{\ln q}} \leq (r-i+1)N^{1-\frac{\widehat{c}}{\ln q}},$$

that if $2 \leq i \leq r$ then

$$\begin{aligned}\text{Prob}\left(\mathsf{fac}(LAR) \geq i\right) &= \text{Prob}\left(\mathsf{C} \ \wedge \mathsf{fac}(LAR) \geq i\right) + \text{Prob}\left(\neg\mathsf{C} \ \wedge \mathsf{fac}(LAR) \geq i\right) \\ &\leq \frac{(r-i+1)q^{2-i} + 2q^{-c}}{q-1} + (r-i+2)N^{1-\frac{\widehat{c}}{\ln q}},\end{aligned}$$

as required to establish part (a).

The number of invariant factors $\mathsf{fac}(LAR)$ is an integer-valued random variable that can assume values between 0 and r. Thus

$$\begin{aligned}E\left[\mathsf{fac}(LAR)\right] &= \sum_{i=1}^{r} \text{Prob}\left(\mathsf{fac}(LAR) \geq i\right) \\ &\leq \sum_{i=1}^{\lfloor \log_q r+3 \rfloor} \text{Prob}\left(\mathsf{fac}(LAR) \geq i\right) + \sum_{i=\lceil \log_q r+3 \rceil}^{r} \text{Prob}\left(\mathsf{fac}(LAR) \geq i\right) \\ &= \sum_{i=1}^{\lfloor \log_q r+3 \rfloor} \text{Prob}\left(\mathsf{fac}(LAR) \geq i\right) + \sum_{i=\lceil \log_q r+3 \rceil}^{r} \text{Prob}\left(\mathsf{C} \ \wedge \mathsf{fac}(LAR) \geq i\right)\end{aligned}$$

$$+ \sum_{i=\lceil \log_q r+3 \rceil}^{r} \text{Prob}\left(\neg \mathsf{C} \ \wedge \mathsf{fac}(LAR) \geq i\right)$$
$$= S_1 + S_2 + S_3,$$

where

$$S_1 = \sum_{i=1}^{\lfloor \log_q r+3 \rfloor} \text{Prob}\left(\mathsf{fac}(LAR) \geq i\right),$$

$$S_2 = \sum_{i=\lceil \log_q r+3 \rceil}^{r} \text{Prob}\left(\mathsf{C} \ \wedge \mathsf{fac}(LAR) \geq i\right),$$

and

$$S_3 = \sum_{i=\lceil \log_q r+3 \rceil}^{r} \text{Prob}\left(\neg \mathsf{C} \ \wedge \mathsf{fac}(LAR) \geq i\right).$$

We will continue by bounding each of S_1, S_2, and S_3 separately.

Clearly

$$\begin{aligned} S_1 &= \sum_{i=1}^{\lfloor \log_q r+3 \rfloor} \text{Prob}\left(\mathsf{fac}(LAR) \geq i\right) \\ &\leq \sum_{i=1}^{\lfloor \log_q r+3 \rfloor} 1 \\ &= \lfloor \log_q r + 3 \rfloor \\ &\leq \log_q r + 3. \end{aligned}$$

Inequality (3.17) can be used to establish that

$$\begin{aligned} S_2 &= \sum_{i=\lceil \log_q r+3 \rceil}^{r} \text{Prob}\left(\mathsf{C} \ \wedge \mathsf{fac}(LAR) \geq i\right) \\ &\leq \sum_{i=\lceil \log_q r+3 \rceil}^{r} \left((r-i+1)\frac{q^{2-i}}{q-1} + (r-i+1)^2 N^{-\frac{\widehat{c}}{\ln q}} \right) \\ &\leq \frac{rq^2}{q-1} \sum_{i \geq \lceil \log_q r+3 \rceil} q^{-i} + \sum_{i=\lceil \log_q r+3 \rceil}^{r} \left((r-2)^2 N^{-\frac{\widehat{c}}{\ln q}} \right) \\ &\leq \frac{rq^2}{q-1} \cdot \frac{q^{1-\lceil \log_q r+3 \rceil}}{q-1} + (r-2)^3 N^{-\frac{\widehat{c}}{\ln q}} \\ &\leq 1 + (r-2)^2 N^{1-\frac{\widehat{c}}{\ln q}}. \end{aligned}$$

Finally, inequality (3.16) can be used to establish that

$$S_3 = \sum_{i=\lceil \log_q r+3 \rceil}^{r} \text{Prob}\left(\neg \mathsf{C} \ \wedge \mathsf{fac}(LAR) \geq i\right)$$

$$
\begin{aligned}
&\leq \sum_{i=\lceil \log_q r+3 \rceil}^{r} \text{Prob}\,(\neg \mathsf{C}) \\
&\leq (r-2)\text{Prob}\,(\neg \mathsf{C}) \\
&\leq \frac{2(r-2)q^{-c}}{q-1} + (r-2)N^{1-\frac{\widehat{c}}{\ln q}} \\
&\leq \frac{2rq^{-c}}{q-1} + (r-2)N^{1-\frac{\widehat{c}}{\ln q}}.
\end{aligned}
$$

The sum of the above bounds for S_1, S_2, and S_3 is less than or equal to

$$\log_q r + 4 + \frac{2rq^{-c}}{q-1} + r^2 N^{1-\frac{\widehat{c}}{\ln q}},$$

as required to establish part (b). □

3.4.6 Conclusion

An invariant factor of the matrix $\widehat{A}$ is nontrivial if and only if it is divisible by x^2, or it is not a power of x. The maximum value of nonnegative integer-valued random variables is always less than or equal to the sum of their values, so that the expected value of the maximum is less than or equal to the sum of the expected values.

The next result is therefore a straightforward consequence of Theorems 3.9 and 3.12.

Theorem 3.13. *Let $A \in \mathsf{F}^{n\times m}$ be a matrix with rank r.*

Suppose the matrices $L \in \mathsf{F}^{\ell\times n}$ and $R \in \mathsf{F}^{m\times \ell}$ are randomly chosen, as described in Section 3.4.1. Let $\widehat{A} = LAR$.

(a) If $2 \leq i \leq r$ then the probability that $\widehat{A}$ has i or more nontrivial invariant factors is at most

$$\frac{(r-i+2)q^{2-i} + 4q^{-c}}{q-1} + (r-i+4)N^{1-\frac{\widehat{c}}{\ln q}}.$$

(b) The expected number of nontrivial invariant factors of $\widehat{A}$ is at most

$$\log_q r + 8 + \frac{4Nq^{-c}}{q-1} + (2N+r^2)N^{1-\frac{\widehat{c}}{\ln q}}.$$

Let

$$\widehat{c} = \begin{cases} 3 & \text{if } q = 2, \\ \lceil 3\ln q \rceil & \text{otherwise,} \end{cases}$$

and let $c = \lceil 2\log_q N \rceil$. It is easily checked that condition (3.4) is satisfied.

We are not interested in small systems of equations; suppose that $N \geq 6$.

In this case, it follows by Theorem 3.6 that

$$\text{Prob}\left(\text{rank}(\widehat{A}) \neq \text{rank}(A)\right) \leq \tfrac{6}{N^2} \leq \tfrac{1}{N}.$$

Theorem 3.9 can be used to establish that the probability that $\widehat{A}$ has at least i invariant factors divisible by x^2 is at most

$$\frac{q^{2-i}}{q-1} + 4N^{-2},$$

so that, in particular, this probability is at most $\frac{1}{2}$ if $i = 4$ and $N \geq 4$. The expected number of invariant factors that are divisible by x^2 is at most $4 + \frac{4}{N} < 5$, since $N \geq 6$.

It follows by Theorem 3.13 that the probability that $\widehat{A}$ has i or more nontrivial invariant factors (for $i \geq 2$) is at most

$$\frac{rq^{2-i}}{q-1} + (r - i + 8)N^{-2},$$

and, since $N \geq 6$, the expected number of nontrivial invariant factors is at most

$$\log_q r + 10.$$

A straightforward modification of an analysis of Wiedemann [15] establishes that, with high probability, L and R are sparse: The expected number of nonzero entries in each is in $O((n+m)(\log N)^2)$.

Suppose that a given matrix A is conditioned, as described above, to produce a matrix $\widehat{A} \in \mathsf{F}^{\ell \times \ell}$, and the Monte Carlo algorithm for matrix rank described in Sections 3.2 and 3.3 is then applied. An analysis of this computation leads to the following.

Theorem 3.14. *Let $A \in \mathsf{F}^{m \times n}$ be a matrix over a finite field F and let $N = \max(n, m)$.*

Then the rank r of A can be computed using a Monte Carlo algorithm such that the expected number of matrix-vector products by A or by A^t is linear in r, the expected number of additional operations over F is in $O(Nr(\log_q N)^2)$, and the expected amount of storage space required is in $O(N(\log_q N)^2)$.

References

[1] 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.

[2] 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.

[3] D. Coppersmith. Solving linear equations over GF(2): Block Lanczos algorithm. *Linear Algebra and Its Applications*, 192:33–60, 1993.

[4] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms.* MIT Press/McGraw-Hill, second edition, 2001.

[5] W. Eberly. Black box Frobenius decompositions over small fields. In *Proceedings, ISSAC '00*, pages 106–113, 2000.

[6] W. Eberly. Asymptotically efficient algorithms for the Frobenius form. Technical Report 2003-723-26, Department of Computer Science, University of Calgary, 2003. Available at `www.cpsc.ucalgary.ca/~eberly/Publications/`.

[7] W. Eberly. Early termination over small fields (extended abstract). In *Proceedings, ISSAC '03*, pages 80–87, 2003. Complete version available at `www.cpsc.ucalgary.ca/~eberly/Publications/`.

[8] W. Eberly and E. Kaltofen. On randomized Lanczos algorithms. In *Proceedings, ISSAC '97*, pages 176–183, 1997. A more complete version is available at `www.cpsc.ucalgary.ca/~eberly/Publications/`.

[9] F. R. Gantmacher. *The Theory of Matrices*, volume one. Chelsea Publishing Company, second edition, 1959.

[10] E. Kaltofen. Analysis of Coppersmith's block Wiedemann algorithm for the parallel solution of sparse linear systems. *Mathematics of Computation*, 64:777–806, 1995.

[11] B. A. LaMacchia and A. M. Odlyzko. Solving large sparse linear systems over finite fields. In *Advances in Cryptology — CRYPTO '90*, volume 537 of *Lecture Notes in Computer Science*, pages 109–133. Springer-Verlag, 1990.

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

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

[14] G. Villard. Further analysis of Coppersmith's block Wiedemann algorithm for the solution of sparse linear systems. In *Proceedings, ISSAC '97*, pages 32–39, 1997.

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