Early Termination over Small Fields II: On the Reliability of Block Krylov-Based Algorithms in a Generic Case

Wayne Eberly*
Department of Computer Science
University of Calgary

June, 2014

Abstract

Block Hankel matrices, generated according to a particular distribution, arise in the analysis of Block Wiedemann and Block Lanczos algorithms. It is shown that if the input matrix A has entries in a small finite field F_q and satisfies a condition that holds generically, then the expected nullities of these matrices are low — as needed to establish the efficiency and reliability of these algorithms. A sparse matrix preconditioner, that ensures that the above-mentioned condition holds with high probability, is also contributed.

Categories and Subject Descriptors

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

General Terms

Algorithms, Performance, Reliability, Theory

Keywords

Black box matrix computations, lookahead in biconditional Lanczos algorithms, early termination in Wiedemann algorithms, randomized algorithms, computations over finite fields

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

1 Introduction

Consider the problem of solving a system of linear equations $Ax = b$ where $A \in \mathsf{F}_q^{n \times n}$ is an $n \times n$ matrix and $b \in \mathsf{F}_q^{n \times 1}$ is a vector with dimension n over the finite field F_q with size q. A related problem which is also of interest is the computation of an element of the null space of such a matrix A. Indeed, instances of these problems are formed and solved in modern algorithms for integer factorization and discrete logarithm computations. In particular, the latter problem arises with $q = 2$ when the number field sieve is applied — as described by Buhler, Lenstra, and Pomerance [2] — while computations over F_q for larger q arise during discrete logarithm computations — as described by LaMacchia and Odlyzko [19].

Several different "Krylov-based" methods have been proposed, implemented and analyzed in recent years. In contrast with elimination-based methods, these do not manipulate the entries of the coefficient matrix A. Instead, these algorithm treat A as an operator on vectors. "Scalar algorithms" work over the vector subspace — generally called a *Krylov space* — generated by $v, Av, A^2v, A^3, v, \ldots$ for a vector $v \in \mathsf{F}_q^{n \times 1}$. Henceforth, this subspace will be called "the Krylov space generated by A and v."

Numerical algorithms of this type have been known and used for years: Both the Lanczos algorithm [20] and the Hestenes-Stiefel conjugate gradient algorithm [15] are examples. A rather different algorithm of this type was proposed more recently for exact computations over arbitrary fields by Wiedemann [27]. The Lanczos method has, subsequently, also been adapted by LaMacchia and Odlyzko [19] and used for matrix computations over large finite fields; a complete analysis of a slightly modified version of this algorithm was subsequently provided by Eberly and Kaltofen [13]. A rather different variant of the Lanczos method was recently proposed for computations over small finite fields by the author of this report [11].

The "scalar" algorithms, mentioned above, are provably reliable when used to solve nonsingular systems. Unfortunately, they are not reliable, in the worst case, when used to solve singular systems or sample from the null space of a given matrix: If the Jordan normal form of the input matrix includes nilpotent blocks with order greater than or equal to two, then a system $Ax = b$ might be consistent, but no solution might exist in the Krylov space generated by A and b — or, indeed, the Krylov space generated by A and v for any vector v that might easily be found. The existence of such nilpotent blocks also biases sampling in searches for elements of the null space of a given matrix. Similarly, algorithms that compute the rank of a matrix by computing the size of a Krylov space will fail if the input matrix has more than one nontrivial invariant factor.

For this, and for various other reasons (discussed below), *block algorithms* have also been used for these computations: These algorithms perform computations over the Krylov space generated by A and a block of vectors $\vec{v} = v_1, v_2, \ldots, v_{k_R}$, that is the subspace of $\mathsf{F}_q^{n \times 1}$ spanned by vectors $A^i v_j$ for $i \geq 0$ (or, for some applications, for $i \geq 1$) and $1 \leq j \leq k_R$. A consultation of the excellent text of Golub and van Loan [14] confirms that, once again,

such algorithms have been used for numerical computations — notably including eigenvalue computations — for years. These algorithms were first adapted for symbolic computations by Coppersmith, who contributed both a block Lanczos algorithm [4] and a block Wiedemann algorithm [5]. Kaltofen [18] subsequently completed the analysis of Coppersmith's block Wiedemann algorithm for computations over large fields, noting, as well, that one could use "rectangular blocks" (different block sizes on the left and the right) in order to significantly improve the algorithm's performance. Villard [25, 26] contributed a more complete analysis of the algorithm for computations over small finite fields as well. More recent work concerning block Lanczos algorithms has included several other, rather different, versions of such an algorithm — including a somewhat simpler variant of Montgomery [23, 24], a "biconditional" block Lanczos algorithm contributed by Hovinen [16, 17], and another "biconditional" algorithm that also included Katofen's idea of rectangular blocking, developed by this author [10].

Unfortunately the reliability of these algorithms is still not completely understood, especially for computations over very small finite fields. In particular, a biconditional block Lanczos algorithm generally selects blocks of vectors

$$\vec{u} = u_1, u_2, \ldots, u_{k_L} \text{ and } \vec{v} = v_1, v_2, \ldots, v_{k_R}$$

uniformly and independently from $\mathsf{F}_q^{n\times 1}$, The algorithm then generates

- vectors $\lambda_1, \lambda_2, \lambda_3, \ldots$, in the Krylov space generated by A and $\vec{v} = v_1, v_2, \ldots, v_{k_R}$, along with
- vectors $\mu_1, \mu_2, \mu_3, \ldots$ in the Krylov space generated by A^T and $\vec{u} = u_1, u_2, u_3, \ldots$

such that, for $i, j \geq 1$,

$$\mu_i^T A \lambda_j = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Additional computations, along the way, allow for systems to be solved or for elements of the null space to be obtained. Once generated, a vector in either of the above Krylov spaces must be stored in memory until it is "matched," that is, a vector in the other Krylov space that is not orthogonal to it has been identified. There is, consequently, a risk that storage requirements will be prohibitive because too many vectors remain unmatched after they are generated.

As noted by Hovinen (and others), the likelihood of this type of failure can be bounded provided that one bounds the expected dimension of the null space of a block Hankel matrix with the form

$$\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R} = \begin{bmatrix} H_{1,1} & H_{1,2} & \ldots & H_{1,k_R} \\ H_{2,1} & H_{2,2} & \ldots & H_{2,k_R} \\ \vdots & \vdots & \ddots & \vdots \\ H_{k_L,1} & H_{k_L,2} & \ldots, & H_{k_L,k_R} \end{bmatrix} \tag{1.1}$$

where $H_{i,j}$ is a Hankel matrix whose entry in row r and column s is $u_i^T A^{r+s-1} v_j$. Matrices $\widehat{\mathcal{H}}_{A,\vec{u},s_L,\vec{v},s_R}$ such that the entry in row r and column s is $u_i^T A^{r+s} v_j$, instead, also arise. These matrices are defined somewhat more precisely in Section 3.

A similar complication arises in some versions of a block Wiedemann algorithm, namely, those that incorporate an "early termination" heuristic. For blocks of vectors $\vec{u}$ and $\vec{v}$ as above, let $\widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},i}$ be the matrix

$$\begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_{k_L}^T \end{bmatrix} \cdot A^i \cdot \begin{bmatrix} v_1 & v_2 & \dots & v_{k_R} \end{bmatrix} \in \mathsf{F}_q^{k_L \times k_R}.$$

In a block Wiedemann algorithm, one generally applies a (block) extension of the Berlekamp-Massey algorithm [1, 22] in order to compute a minimum right generating polynomial[1] of the sequence of matrices

$$\widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},0}, \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},1}, \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},2}, \dots$$

Again, things might change slightly depending on the problem to be solved: It might be necessary to generate a minimum right generating polynomial for the related sequence

$$\widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},1}, \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},2}, \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},3}, \dots$$

instead.

One proceeds by considering successive elements of the whichever of the above sequences is appropriate in order to compute a sequence of iterates

$$g^{(0)}, g^{(1)}, g^{(2)}, \dots$$

for the minimum right generating polynomial $g \in \mathsf{F}_q[x]^{k_R \times k_R}$ of this sequence. The number of terms of the sequence that must be examined, and the number of iterates that should be generated, can be bounded by a function of the order n of the input matrix A and the block sizes k_L and k_R in use — and, indeed, the development of this type of bound has been a significant contribution in the analysis of block Wiedemann algorithms. This bound is not necessarily tight and is, indeed, quite pessimistic if the rank of A is significantly lower than its order.

With that noted, let us say that a *zero discrepancy* has occurred during the i^{th} iteration in the above computation, for $i \geq 1$, if $g^{(i-1)} = g^{(i)} \neq g$. As originally noted by Lobo [21], experimental evidence suggests that a long sequence of zero discrepancies is unlikely to arise when Wiedemann's algorithm is applied, so that one may often conclude that a

[1] In their writings about block Wiedemann algorithms, Coppersmith, Kaltofen and Villard all describe this in some way. I have consulted Villard's report in order to write the above.

minimum right generating polynomial has been computed and terminate the computation after noting that

$$g^{(i)} = g^{(i+1)} = \cdots = g^{(i+k)}$$

for any positive integer i and for reasonably small k: For computations over large fields, it seems to be sufficient to set k to be two when applying this heuristic; depending on the block size used, k might be chosen to be (at most) logarithmic in n for computations over small fields.

Once again, this phenomenon can be related to a condition concerning the rank of matrix, namely, a matrix with either the form

$$\begin{bmatrix} \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},0} & \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},1} & \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},2} & \cdots & \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},r} \\ \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},1} & \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},2} & \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},3} & \cdots & \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},r+1} \\ \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},2} & \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},3} & \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},4} & \cdots & \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},r+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},s} & \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},s+1} & \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},s+2} & \cdots & \widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},r+s} \end{bmatrix}$$

or with a similar block Hankel form, such the top left block is $\widetilde{\mathcal{H}}_{A\vec{u},\vec{v},1}$ instead of $\widetilde{\mathcal{H}}_{A,\vec{u},\vec{v},0}$: Since the rightmost columns of this matrix must be linear combinations of the columns to the left, such a matrix cannot have full rank if a sequence of zero discrepancies has occurred.

Permuting rows a columns of these matrices one obtains either a matrix $\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R}$ or a matrix $\widehat{\mathcal{H}}_{A,\vec{u},s_L,\vec{v},s_R}$ as described above, in the discussion of block Lanczos algorithms. Consequently results that bound the nullity of matrices with this form, for uniformly and randomly generated blocks of vectors $\vec{u}$ and $\vec{v}$, are sufficient both to bound the storage requirements for biconditional block Lanczos algorithms and to establish the reliability of early termination heuristics in block Wiedemann computations.

As part of his Master's research, Hovinen [16, 17] made significant progress in bounding the nullities of such matrices[2]. In an elegant application of module theory, Hovinen established that if vectors

$$\vec{u} = u_1, u_2, \ldots, u_k \text{ and } \vec{v} = v_1, v_2, \ldots, v_k$$

are chosen uniformly and independently from $\mathsf{F}_q^{n\times 1}$, and i is an integer such that ki is less than or equal to the rank of a matrix $A \in \mathsf{F}_q^{n\times n}$, then the probability that the matrix $\widehat{\mathcal{H}}_{A,\vec{u},k(i+\Delta),\vec{v},ki}$ is rank deficient is at most

$$4q^{i-\lceil n/k\rceil} + \tfrac{6}{n} + q^{-k(\Delta-1)}$$

[2]While I was listed as co-author of the conference report of this work, this was not part of my contribution. Instead, the analysis described here was entirely due to Hovinen.

— provided that the number of nontrivial invariant factors of A is less than or equal to $k - \log_q n$, and provided that $q^2/n + q^3/n^2 < 1$. (See, in particular, Corollary 3.21 of [16], which states this in somewhat different form.) Hovinen applied this and other results to complete an analysis of his block Lanczos algorithm — the first such algorithm to be fully analyzed.

It is possible to establish similar bounds by more elementary methods. Indeed, Section 3 presents a method to do this using basic linear algebra and probability theory, along with a well known bound from the theory of finite fields — namely, that the number of distinct monic irreducible polynomials in $\mathsf{F}_q[x]$ with degree d is at most q^d/d. These methods are applied, in Section 4, to generalize Hovinen's results:

- The restriction on field size — that $q^2/n + q^3/n^2 < 1$ — is not serious: If $q^2/n + q^3/n^2 \geq 1$, instead, then $q \geq \frac{1}{2}\sqrt{n}$, and large-field arguments can be used to bound the likelihood that $\widehat{\mathcal{H}}_{A,\vec{u},s_L,\vec{v},s_R}$ is rank deficient, provided that one works over a field extension with degree at most five over F_q.

  However, this restriction is also unnecessary, and it is not used in Section 4. Indeed (as one might suspect) bounds on the likelihood that that the above matrix is rank-deficient improve as q increases.

- The restriction that the number of nontrivial invariant factors of A is less than or equal to $k - \log_q n$ is also significantly relaxed: Bounds for the likelihood that $\widehat{\mathcal{H}}_{A,\vec{u},s_L,\vec{v},s_R}$ are provided in Section 4 whenever the number of nontrivial invariant factors of A is at most $\min(k_L, k_R) - 1$.

- As the above notation may suggest, bounds are given for the likelihood of rank deficiency of a larger class of block Hankel matrices, including those that arise when algorithms using a different block size on the left than on the right are applied. It seems that there are already quite a few variants of "block Wiedemann" and "block Lanczos" algorithms, and still others may emerge. I am hopeful that the bounds given here are general enough to be applicable when analyzing quite a few of these algorithms.

In particular, bounds for the probability that either a matrix $\widehat{\mathcal{H}}_{A,\vec{u},s_L,\vec{v},s_R}$ or a matrix $\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R}$ is rank deficient by at least j, when the number of nontrivial invariant factors is greater than the minimum of the block sizes used, are given at lines (4.6) on page 44 and (4.7) on page 44, respectively. They indicate that at any fixed point in a block Lanczos computation, when the same number of vectors have been produced in the Krylov space generated by A^T and $\vec{u}$ as have been generated in the Krylov space generated by A and $\vec{v}$ (so that $s_L = s_R$), it is unlikely that there are significantly more than $\log \log n$ unmatched vectors produced in either Krylov space. This can be used to argue that it is unlikely that there are ever significantly more than $O(\log n)$ unmatched vectors at any point during the computation at all. It can be also be used to argue that it is unlikely that a sequence

of zero discrepancies with length more than logarithmic in n ever arises during a block Wiedemann computation that employs an early termination heuristic.

As mentioned in various reports — including the original ones of Coppersmith — block sizes are chosen for a variety of reasons when block Lanczos and block Wiedemann algorithms are designed and implemented for multiple reasons. These include more efficient use of storage and improving coarse grained parallelism. Since one block size might not be best for everything I am hopeful that this ongoing work to eliminate or relax conditions to ensure reliability will simplify the design and implementation of block Wiedemann and Lanczos algorithms in the future.

Section 5 introduces a variant of a "sparse Wiedmenn" preconditioner that closely resembles one introduced by Wiedemann [27] and further analyzed by Chen et. al. [3]. Since this preconditioner is slightly different than the one described in these other reports a reasonably complete analysis of its properties is provided. That noted, it is not really either new or surprising that this preconditioner is sparse, that it preserves the rank of the matrix being conditioned, or that the number of nontrivial nilpotent blocks in the Jordan form for the conditioned matrix is probably small — and proofs of these properties are based on similar proofs in the cited reports. It is perhaps more surprising that the number of nontrivial invariant factors of the preconditioned matrix is probably small, as well: This result complements the results of Sections 3–4 by establishing (with high probability) the matrix property used to bound storage requirements for a block Lanczos computation or ensure reliability of an early termination heuristic in a block Wiedemann computation. It is also sufficient to establish the reliability of block algorithms that solve systems, sample from the null space of a matrix, or compute matrix rank.

Section 6 describes additional work in progress as well as a few questions that may be of interest to other researchers.

Much of this work has appeared, in rather different form, in a pair of unpublished technical reports [8, 12]. The properties of the sparse Wiedemann preconditioner, discussed above, were also announced in an extended abstract [9] that included a very brief proof sketch. Apart from that, this is the first conference or journal presentation of this work.

2 Characterizations of the Matrix A

2.1 Main Characterization

The existence of "Frobenius normal form" as well as a (rational) Jordan normal form for a matrix suffice to establish that every matrix $A \in \mathsf{F}_q^{n \times n}$ can be described as follows. The factorization of A, described here, will be used extensively throughout this report.

Suppose now that

$$A = X^{-1} D X \tag{2.1a}$$

for a nonsingular matrix $X \in \mathsf{F}_q^{n\times n}$, and for a block diagonal matrix

$$D = \begin{bmatrix} \Delta_0 & & & 0 \\ & \Delta_1 & & \\ & & \ddots & \\ 0 & & & \Delta_h \end{bmatrix} \in \mathsf{F}_q^{n\times n}. \tag{2.1b}$$

In this case, Δ_i is a matrix whose characteristic polynomial is a power of an irreducible polynomial φ_i with degree $d_i \geq 1$ in $\mathsf{F}_q[x]$, for $0 \leq i \leq h$, so that

$$\varphi_0, \varphi_1, \ldots, \varphi_h$$

are distinct monic irreducible polynomials in $\mathsf{F}_q[x]$.

It will be useful to consider the case that the given irreducible polynomial is x separately from other cases. To facilitate this, let us suppose that $\varphi_0 = x$ (so that $d_0 = 1$), that $h \geq 0$, and that $\varphi_1, \varphi_2, \ldots, \varphi_h$ are the (remaining) monic irreducible divisors of the characteristic polynomial of A.

To continue, suppose that Δ_i is a block diagonal matrix

$$\Delta_i = \begin{bmatrix} Z_{i,1} & & & 0 \\ & Z_{i,2} & & \\ & & \ddots & \\ 0 & & & Z_{i,\ell i} \end{bmatrix} \tag{2.1c}$$

for an integer $\ell_i \geq 0$ for $1 \leq i \leq h$, and that

$$Z_{i,j} = \begin{bmatrix} 0 & & & & -\alpha_{i,j,0} \\ 1 & 0 & & & -\alpha_{i,j,1} \\ & 1 & & & -\alpha_{i,j,2} \\ & & \ddots & & \vdots \\ & & & 1 & -\alpha_{i,j,d_i n_{i,j}-1} \end{bmatrix} \in \mathsf{F}_q^{d_i n_{i,j} \times d_i n_{i,j}} \tag{2.1d}$$

is the companion matrix of the polynomial

$$\varphi_i^{n_{i,j}} = x^{d_i n_{i,j}} + \alpha_{i,j,d_i n_{i,j}-1} x^{d_i n_{i,j}-1} + \cdots + \alpha_{i,j,2} x^2 + \alpha_{i,j,1} x + \alpha_{i,j,0} \tag{2.1e}$$

for $0 \leq i \leq h$ and $1 \leq j \leq \ell_i$, and such that

$$\ell_0 \geq 0 \text{ and } \ell_i \geq 1 \text{ for } 1 \leq i \leq h, \tag{2.1f}$$

$$n_{i,1} \geq n_{i,2} \geq \cdots \geq n_{i,\ell_i} \geq 1 \text{ for } 0 \leq i \leq h, \tag{2.1g}$$

$$d_i(n_{i,1} + n_{i,2} + \cdots + n_{i,\ell_i}) = n_i \text{ for } 0 \leq i \leq h, \tag{2.1h}$$

and

$$n_0 + n_1 + \cdots + n_h = n. \tag{2.1i}$$

If these parameters are defined as above then A is nonsingular if and only if $\ell_0 = 0$, while A is nilpotent if and only if $h = 0$.

Finally, let

$$\ell = \max(\ell_0, \ell_1, \ldots, \ell_h) \tag{2.1j}$$

so that ℓ is the number of invariant factors of A.

2.2 A Decomposition of $\mathsf{F}_q^{n\times 1}$

Lemma 2.1. *If A is as given in equations* (2.1a)–(2.1j) *then there exist vectors $\zeta_{i,j} \in \mathsf{F}_q^{n\times 1}$, for $0 \le i \le h$ and $1 \le j \le \ell_i$, such that*

$$\mathsf{F}_q^{n\times 1} = \bigoplus_{\substack{0\le i\le h\\ 1\le j\le \ell_i}} V_{i,j}$$

where

$$V_{i,j} = \{f(A)\zeta_{i,j} \mid f \in \mathsf{F}_q[x] \text{ and } \deg(f) < d_i n_{i,j}\},$$

is a subspace with dimension $d_i n_{i,j}$ over F_q, and such that

$$\varphi_i^{n_{i,j}}(A)\zeta_{i,j} = 0$$

for all integers i and j such that $0 \le i \le h$ and $1 \le j \le \ell_i$. Furthermore, there is exactly one sequence of polynomials

$$f_{0,1}, f_{0,2}, \ldots, f_{0,\ell_0}, \ldots, f_{h,1}, f_{h,2}, \ldots, f_{h,\ell_h} \in \mathsf{F}_q[x]$$

such that $f_{i,j}$ has degree less than $d_i n_{i,j}$ for $0 \le i \le h$ and $1 \le j \le \ell_i$, and such that

$$y = \sum_{i=0}^{h} \sum_{j=1}^{\ell_i} f_{i,j}(A)\zeta_{i,j}$$

for any given vector $y \in \mathsf{F}_q^{n\times 1}$.

Proof. Let e_i be the i^{th} unit vector (with 1 in position i and 0's elsewhere) for every integer i such that $1 \le i \le n$. Then, since the matrix X^{-1} is nonsingular, the vectors

$$X^{-1}e_1, X^{-1}e_2, \ldots, X^{-1}e_n$$

certainly form a basis for $\mathsf{F}_q^{n\times 1}$. Now let

$$\zeta_{i,j} = X^{-1}e_{n_0+n_1+\cdots+n_{i-1}+d_i(n_{i,1}+n_{i,2}+\cdots+n_{i,j-1})+1}$$

for $0 \leq i \leq h$ and $1 \leq j \leq \ell_i$. It is easily confirmed by an examination of the definitions of the above matrices D, Δ_i and $Z_{i,j}$ (for $0 \leq i \leq h$ and $1 \leq j \leq \ell_i$) that

$$A^{k-1}\zeta_{i,j} = X^{-1}e_{n_0+n_1+\cdots+n_{i-1}+d_i(n_{i,1}+n_{i,2}+\cdots+n_{i,j-1})+k} \tag{2.2}$$

for every integer k such that $1 \leq k \leq d_i n_{i,j}$. Thus, if

$$V_{i,j} = \{f(A)\zeta_{i,j} \mid f \in \mathsf{F}_q[x] \text{ and } \deg(f) < d_i n_{i,j}\}$$

as above, for $0 \leq i \leq h$ and $1 \leq j \leq \ell_i$, then $V_{i,j}$ is spanned by the vectors

$$X^{-1}e_{n_0+n_1+\cdots+n_{i-1}+d_i(n_{i,1}+n_{i,2}+\cdots+n_{i,j-1})+1},$$
$$X^{-1}e_{n_0+n_1+\cdots+n_{i-1}+d_i(n_{i,1}+n_{i,2}+\cdots+n_{i,j-1})+2}\cdots,$$
$$X^{-1}e_{n_0+n_1+\cdots+n_{i-1}+d_i(n_{i,1}+n_{i,2}+\cdots+n_{i,j-1}+n_{i,j})}$$

and, since these vectors are linearly independent, $V_{i,j}$ has dimension $d_i n_{i,j}$ over F_q. It also follows by the equation at line (2.2) that the vectors

$$\zeta_{i,j}, A\zeta_{i,j}, A^2\zeta_{i,j}, \ldots, A^{d_i n_{i,j}-1}\zeta_{i,j}$$

are linearly independent and form a basis for $V_{i,j}$ as well. Now

$$\mathsf{F}_q^{n\times 1} = \bigoplus_{\substack{0\leq i\leq h\\ 1\leq j\leq \ell_i}} V_{i,j}$$

as claimed, and it follows that every vector $y \in \mathsf{F}_q^{n\times 1}$ has a unique representation as

$$y = \sum_{i=0}^{h}\sum_{j=1}^{\ell_i} f_{i,j}(A)\zeta_{i,j}$$

for polynomials $f_{i,j} \in \mathsf{F}_q[x]$ such that $f_{i,j}$ has degree less than $d_i n_{i,j}$ for $0 \leq i \leq h$ and $1 \leq j \leq \ell_i$.

It remains only to argue that $\varphi_i^{n_{i,j}}(A)\zeta_{i,j} = 0$. Set

$$\widehat{e} = e_{n_0+n_1+\cdots+n_{i-1}+d_i(n_{i,1}+n_{i,2}+\cdots+n_{i,j-1})+1}.$$

Then

$$\begin{aligned}\varphi_i^{n_{i,j}}(A)\zeta_{i,j} &= \varphi_i^{n_{i,j}}\left(X^{-1}DX\right)X^{-1}\widehat{e}\\ &= X^{-1}\varphi_i^{n_{i,j}}(D)X\cdot X^{-1}\widehat{e}\end{aligned}$$

$$= X^{-1} \begin{bmatrix} \varphi_i^{n_{i,j}}(\Delta_0) & & & 0 \\ & \varphi_i^{n_{i,j}}(\Delta_1) & & \\ & & \ddots & \\ 0 & & & \varphi_i^{n_{i,j}}(\Delta_h) \end{bmatrix} \widehat{e},$$

and

$$\varphi_i^{n_{i,j}}(\Delta_i) = \begin{bmatrix} \varphi_i^{n_{i,j}}(Z_{i,1}) & & & 0 \\ & \varphi_i^{n_{i,j}}(Z_{i,2}) & & \\ & & \ddots & \\ 0 & & & \varphi_i^{n_{i,j}}(Z_{i,\ell_i}) \end{bmatrix}$$

The result now follows since $Z_{i,j}$ has characteristic polynomial $\varphi_i^{n_{i,j}}$, so that $\varphi_i^{n_{i,j}}(Z_{i,j}) = 0$, and column $n_0 + n_1 + \cdots + n_{i-1} + d_i(n_{i,1} + n_{i,2} + \cdots + n_{i,j-1}) + 1$ of the block diagonal matrix

$$\begin{bmatrix} \varphi_i^{n_{i,j}}(\Delta_0) & & & 0 \\ & \varphi_i^{n_{i,j}}(\Delta_1) & & \\ & & \ddots & \\ 0 & & & \varphi_i^{n_{i,j}}(\Delta_h) \end{bmatrix}$$

is filled with zeroes. □

Corollary 2.2. *If A and the vectors $\zeta_{i,j}$ are as described in Lemma 2.1, and $f_{i,j} \in \mathsf{F}_q[x]$ for $0 \leq i \leq h$ and $1 \leq j \leq \ell_i$, then*

$$\sum_{i=0}^{h} \sum_{j=1}^{\ell_i} f_{i,j}(A)\zeta_{i,j} = 0$$

if and only if $f_{i,j}$ is divisible by $\varphi_i^{n_{i,j}}$ for all i and j such that $0 \leq i \leq h$ and $1 \leq j \leq \ell_i$.

Proof. Suppose first that $f_{i,j}$ is divisible by $\varphi_i^{n_{i,j}}$ for all integers i and j such that $0 \leq i \leq h$ and $1 \leq j \leq \ell_i$. Then $f_{i,j} = g_{i,j} \cdot \varphi_i^{n_{i,j}}$ for a polynomial $g_{i,j} \in \mathsf{F}_q[x]$, so that

$$f_{i,j}(A)\zeta_{i,j} = g_{i,j}(A) \cdot \varphi_i^{n_{i,j}}(A)\zeta_{i,j} = g_{i,j}(A) \cdot 0 = 0,$$

by the results of Lemma 2.1, and since this holds for all i and j,

$$\sum_{i=0}^{h} \sum_{j=1}^{\ell_i} f_{i,j}(A)\zeta_{i,j} = 0$$

as well.

Now suppose, conversely, that there exists at least one pair of integers i and j such that $0 \leq i \leq h$, $1 \leq j \leq \ell_i$, and $f_{i,j}$ is not divisible by $\varphi_i^{n_{i,j}}$. Then

$$f_{i,j} = \varphi_i^m \cdot g$$

for some integer m such that $0 \leq m \leq n_{i,j} - 1$ and for some polynomial $g \in \mathsf{F}_q[x]$ that is relatively prime with the irreducible polynomial φ_i. The polynomial g must then be relatively prime with $\varphi_i^{n_{i,j}}$ as well, so that there exist polynomials $s, t \in \mathsf{F}_q[x]$ such that $s \cdot g + t \cdot \varphi_i^{n_{i,j}} = 1$. Consequently

$$\begin{aligned} s(A) \cdot f_{i,j}(A)\zeta_{i,j} &= s(A) \cdot g(A) \cdot \varphi_i^m(A)\zeta_{i,j} \\ &= (1 - t(A) \cdot \varphi_i^{n_{i,j}}(A)) \cdot \varphi_i^m(A)\zeta_{i,j} \\ &= \varphi_i^m(A)\zeta_{i,j} - t(A) \cdot \varphi_i^m(A) \cdot \varphi_i^{n_{i,j}}(A)\zeta_{i,j} \\ &= \varphi_i^m(A)\zeta_{i,j} - t(A) \cdot \varphi_i^m(A) \cdot 0 \\ &= \varphi_i^m(A)\zeta_{i,j} \in V_{i,j}. \end{aligned}$$

Since φ_i^m is a nonzero polynomial with degree $d_i m < d_i n_{i,j}$, $\varphi_i^m(A)\zeta_{i,j}$ is a nonzero element of $V_{i,j}$, so that $f_{i,j}(A)\zeta_{i,j}$ must certainly be a nonzero element of $V_{i,j}$ as well. Now, since

$$\mathsf{F}_q^{n \times 1} = \bigoplus_{\substack{0 \leq i \leq h \\ 1 \leq j \leq \ell_i}} V_{i,j}$$

and $f_{s,t}(A)\zeta_{s,t} \in V_{s,t}$ for all integers s and t such that $0 \leq s \leq h$ and $1 \leq t \leq \ell_s$, it necessarily follows that

$$\sum_{i=0}^{h} \sum_{j=1}^{\ell_i} f_{i,j}(A)\zeta_{i,j} \neq 0$$

as well, establishing the claim. □

2.3 An Alternative Characterization of A

It will occasionally be useful to consider an alternative characterization of the matrix A. With that noted, suppose that A is as given in the equations at lines (2.1a)–(2.1j), as usual, and let $\widehat{\ell}_0, \widehat{\ell}_1, \ldots, \widehat{\ell}_h$ be integers such that $0 \leq \widehat{\ell}_i \leq \ell_i$ for $0 \leq i \leq h$. Let

$$\widehat{N} = \sum_{i=0}^{h} \sum_{j=1}^{\widehat{\ell}_i} d_i n_{i,j} \qquad \text{and} \qquad \widetilde{N} = \sum_{i=0}^{h} \sum_{j=\widehat{\ell}_i+1}^{\ell_i} d_i n_{i,j} \tag{2.3a}$$

so that $\widehat{N} + \widetilde{N} = n$. Then

$$A = \widehat{X}^{-1} \begin{bmatrix} \widehat{A} & 0 \\ 0 & \widetilde{A} \end{bmatrix} \widehat{X} \tag{2.3b}$$

for a nonsingular matrix $\widehat{X} \in \mathsf{F}_q^{n\times n}$ and for matrices $\widehat{A} \in \mathsf{F}_q^{\widehat{N}\times\widehat{N}}$ and $\widetilde{A} \in \mathsf{F}_q^{\widetilde{N}\times\widetilde{N}}$. In particular,

$$\widehat{A} = \begin{bmatrix} \widehat{\Delta}_0 & & & 0 \\ & \widehat{\Delta}_1 & & \\ & & \ddots & \\ 0 & & & \widehat{\Delta}_h \end{bmatrix} \quad \text{and} \quad \widetilde{A} = \begin{bmatrix} \widetilde{\Delta}_0 & & & 0 \\ & \widetilde{\Delta}_1 & & \\ & & \ddots & \\ 0 & & & \widetilde{\Delta}_h \end{bmatrix} \tag{2.3c}$$

and where

$$\widehat{\Delta}_i = \begin{bmatrix} Z_{i,1} & & & 0 \\ & Z_{i,2} & & \\ & & \ddots & \\ 0 & & & Z_{i,\widehat{\ell}_i} \end{bmatrix} \quad \text{and} \quad \widetilde{\Delta}_i = \begin{bmatrix} Z_{i,\widehat{\ell}_i+1} & & & 0 \\ & Z_{i,\widehat{\ell}_i+2} & & \\ & & \ddots & \\ 0 & & & Z_{i,\ell_i} \end{bmatrix} \tag{2.3d}$$

for $0 \le i \le h$ and for $Z_{i,j}$ as shown at line (2.1d) for $0 \le i \le h$ and $1 \le j \le \ell_i$.

3 Definitions of Matrices and Fundamental Properties

The matrices to be studied will now be defined (see, in particular, Definitions 3.1 and 3.19) and fundamental properties, needed to establish that they are unlikely to have low rank, will be introduced.

3.1 The Matrix $\mathcal{K}_{A,\vec{v},s}$

Definition 3.1. Let k and s be positive integers. Let $s_{k,i} = \lceil s/k \rceil$ for $1 \le i \le (s \bmod k)$ and let $s_{k,i} = \lfloor s/k \rfloor$ for $(s \bmod k) + 1 \le i \le k$, so that

$$s_{k,1} \ge s_{k,2} \ge \cdots \ge s_{k,k}$$

either $s_{k,1} = s_{k,k}$ or $s_{k,1} = s_{k,k} + 1$, and

$$s_{k,1} + s_{k,2} + \cdots + s_{k,k} = s.$$

For vectors $\vec{v} = v_1, v_2, \ldots, v_k \in \mathsf{F}_q^{n\times 1}$, let $\mathcal{K}_{A,\vec{v},s} \in \mathsf{F}_q^{n\times s}$ be the matrix with columns

$$\begin{aligned} v_1,\ Av_1,\ A^2v_1,\ \ldots,\ A^{s_{k,1}-1}v_1, & \\ v_2,\ Av_2,\ A^2v_2,\ \ldots,\ A^{s_{k,2}-1}v_2, & \\ \ldots\ v_k,\ Av_k,\ A^2v_k,\ \ldots,\ A^{s_{k,k}-1}v_k, & \end{aligned}$$

that is, the vectors $A^i v_j$ for integers h and j such that $1 \le j \le k$ and $0 \le i \le s_{k,j} - 1$.

Definition 3.2. Once again, let k and s be positive integers. Let $\mathsf{Null}_{A,k,s}$ be the number of choices of vectors $\vec{v} = v_1, v_2, \ldots, v_k \in \mathsf{F}_q^{n\times 1}$ and scalars

$$d_{1,0}, d_{1,1}, \ldots, d_{1,s_{k,1}-1}, \ldots, d_{k,0}, d_{k,1}, \ldots, d_{k,s_{k,k}-1} \in \mathsf{F}_q$$

such that

$$\mathcal{K}_{A,\vec{v},s}\vec{d} = 0 \qquad \text{for } \vec{d} = \begin{bmatrix} d_{1,0} \\ d_{1,1} \\ \vdots \\ d_{k,s_{k,k}-1} \end{bmatrix}. \tag{3.1}$$

Lemma 3.3. *If A is as given in equations (2.1a)–(2.1j) and k and s are positive integers, then $\mathsf{Null}_{A,k,s}$ is equal to the number of choices of polynomials $f_{i,j,m} \in \mathsf{F}_q[x]$ for $0 \leq i \leq h$, $1 \leq j \leq \ell_i$, and $1 \leq m \leq k$, where $\deg(f_{i,j,m}) < d_i n_{i,j}$ for all i, j and m as above, and of polynomials $g_1, g_2, \ldots, g_k \in \mathsf{F}_q[x]$ such that $\deg(g_j) < s_{k,j}$ for $1 \leq j \leq k$, such that*

$$f_{i,j,1}g_1 + f_{i,j,2}g_2 + \cdots + f_{i,j,k}g_k \equiv 0 \bmod \varphi_i^{n_{i,j}}$$

for all integers i and j such that $0 \leq i \leq h$ and $1 \leq j \leq \ell_i$.

Proof. If follows by Lemma 2.1 that the number of choices of vectors $\vec{v} = v_1, v_2, \ldots, v_k \in \mathsf{F}_q^{n\times 1}$ is the same as the number of choices of polynomials $f_{i,j,m} \in \mathsf{F}_q[x]$ such that $f_{i,j,m}$ has degree less than $d_i n_{i,j}$, for $0 \leq i \leq h$, $1 \leq j \leq \ell_i$, and $1 \leq m \leq k$. Indeed, there is exactly one such choice of polynomials $f_{i,j,m}$ for $0 \leq i \leq h$ and $1 \leq j \leq \ell_i$ such that

$$v_m = \sum_{i=0}^{h} \sum_{j=1}^{\ell_i} f_{i,j,m}(A)\zeta_{i,j} \tag{3.2}$$

for $1 \leq m \leq k$ and for the vectors $\zeta_{i,j}$ described in the lemma.

One can also define a unique set of polynomials $g_1, g_2, \ldots, g_k \in \mathsf{F}_q[x]$, such that g_j has degree less than $s_{k,j}$ for $1 \leq j \leq k$, corresponding to a sequence of scalars

$$d_{1,0}, d_{1,1}, \ldots, d_{1,s_{k,1}-1}, \ldots, d_{k,0}, d_{k,1}, \ldots, d_{k,s_{k,k}-1} \in \mathsf{F}_q,$$

by setting

$$g_j = d_{j,0} + d_{j,1}x + d_{j,2}x^2 + \cdots + d_{j,s_{k,j}-1}x^{s_{k,j}-1} \tag{3.3}$$

for $1 \leq j \leq k$.

It suffices to examine the definition of the matrix $\mathcal{K}_{A,\vec{v},s}$ and the equations at lines (3.2) and (3.3) to confirm that

$$\mathcal{K}_{A,\vec{v},s}\vec{d} = \sum_{i=0}^{h} \sum_{j=1}^{\ell_i} (f_{i,j,1}g_1 + f_{i,j,2}g_2 + \cdots + f_{i,j,k}g_k)(A)\zeta_{i,j}.$$

The claim now follows by an application of Corollary 2.2. □

The following probabilities will also be of interest.

Definition 3.4. Let $\rho_{A,k,s}(j)$ be the probability that the matrix $\mathcal{K}_{A,\vec{v},s} \in \mathsf{F}_q^{n\times s}$ is rank deficient by exactly j if the vectors $\vec{v} = v_1, v_2, \ldots, v_k$ are chosen uniformly and independently from $\mathsf{F}_q^{n\times 1}$ — that is,

$$\rho_{A,k,s}(j) = \mathsf{Pr}[\mathsf{rank}(\mathcal{K}_{A,\vec{v},s}) = \min(n,s) - j] \tag{3.4}$$

Similarly, let $\sigma_{A,k,s}(j)$ be the probability that the matrix $\mathcal{K}_{A,\vec{v},s}$ is rank deficient by *at least* j if the vectors $\vec{v} = v_1, v_2, \ldots, v_k$ are chosen uniformly and independently from $\mathsf{F}_q^{n\times 1}$, that is

$$\sigma_{A,k,s}(j) = \mathsf{Pr}[\mathsf{rank}(\mathcal{K}_{A,\vec{v},s}) \leq \min(n,s) - j]. \tag{3.5}$$

Lemma 3.5. *If the vectors $\vec{v} = v_1, v_2, \ldots, v_k$ are chosen uniformly and independently from $\mathsf{F}_q^{n\times 1}$ then*

$$\mathsf{Null}_{A,k,s} = \sum_{j=0}^{\min(n,s)} q^{nk+s-\min(n,s)+j}\rho_{A,k,s}(j).$$

Proof. Suppose, first, that $s \leq n$, so that $nk + s - \min(n,s) = nk$; then every matrix $\mathcal{K}_{A,\vec{v},s}$ is rank deficient by some integer j such that $0 \leq j \leq s$. Now, if $\mathsf{Null}_{A,k,s}j$ is as defined above then there are $q^{nk}\rho_{A,k,s}(j)$ matrices $\mathcal{K}_{A,\vec{v},s} \in \mathsf{F}_q^{n\times s}$ that are rank deficient by j (corresponding to choices of vectors $\vec{v} = v_1, v_2, \ldots, v_k \in \mathsf{F}_q^{n\times 1}$). The claim for the case that $s \leq n$ now follows by the fact that if $\mathcal{K}_{A,\vec{v},s}$ is rank deficient by j then its right null space has dimension j and includes exactly q^j vectors.

The proof of the claim for the case that $s > n$ (so that $nk+s-\min(n,s) = (k-1)n+s$) is similar: It suffices to observe that every matrix $\mathcal{K}_{A,\vec{v},s}$ is rank deficient by some integer j such that $0 \leq j \leq n$, and that the right null space of this matrix has dimension $s - n + j$ and includes q^{s-n+j} vectors in this case. □

Note that $\sigma_{A,k,s}(0) = 1$, $\rho_{A,k,s}(j) = \sigma_{A,k,s}(j) - \sigma_{A,k,s}(j+1)$ if $0 \leq j < \min(n,s)$, and $\sigma_{A,k,s}(j) = \rho_{A,k,s}(j)$ if $j = \min(n,s)$. The following is therefore a consequence of Lemma 3.5.

Corollary 3.6. *If vectors $\vec{v} = v_1, v_2, \ldots, v_k$ are chosen uniformly and independently from $\mathsf{F}_q^{n\times 1}$ then*

$$\mathsf{Null}_{A,k,s} = q^{nk+s-\min(n,s)}\left(1 + (q-1)\sum_{j=1}^{\min(n,s)} \sigma_{A,k,s}(j)\cdot q^{j-1}\right).$$

Corollary 3.6 implies that

$$\sigma_{A,k,s}(j) \leq \frac{(\mathsf{Null}_{A,k,s} - q^{nk+s-\min(n,s)})}{(q-1)q^{nk+s-\min(n,s)+j-1}} \tag{3.6}$$

if $1 \leq j \leq \min(n, s)$. Bounds for the probability that $\mathcal{K}_{A,\vec{v},s}$ is rank deficient will be obtained by bounding $\mathsf{Null}_{A,k,s}$ and applying this inequality.

Suppose now that $\widehat{\ell}_0, \widehat{\ell}_1, \ldots, \widehat{\ell}_h$ are integers such that $0 \leq \widehat{\ell}_i \leq \ell_i$ for $0 \leq i \leq h$, and consider the decomposition of A described in Subsection 2.3 — considering, in particular, the matrices $\widehat{A} \in \mathsf{F}_q^{\widehat{N} \times \widehat{N}}$ and $\widetilde{A} \in \mathsf{F}_q^{\widetilde{A} \times \widetilde{A}}$ as defined at lines (2.3a)–(2.3d).

Lemma 3.7. *If A, $\widehat{A} \in \mathsf{F}_q^{\widehat{N} \times \widehat{N}}$ and $\widetilde{A} \in \mathsf{F}_q^{\widetilde{N} \times \widetilde{N}}$ are as described above, then*

$$\mathsf{Null}_{A,k,s} \leq q^{\widetilde{N}k} \mathsf{Null}_{\widehat{A},k,s}$$

for all positive integers k and s.

Proof. It follows by Lemma 3.3 that

- $\mathsf{Null}_{A,k,s}$ is the number of choices of polynomials $f_{i,j,m} \in \mathsf{F}_q[x]$ for $0 \leq i \leq h$, $1 \leq j \leq \ell_i$, and $1 \leq m \leq k$ where $\deg(f_{i,j,m}) < d_i n_{i,j}$ for all such i, j and k, and of polynomials $g_1, g_2, \ldots, g_k \in \mathsf{F}_q[x]$ such that $\deg(g_j) < s_{k,j}$ for $1 \leq j \leq k$ and such that
$$f_{i,j,1} g_1 + f_{i,j,2} g_2 + \cdots + f_{i,j,k} g_k \equiv 0 \bmod \varphi_i^{n_{i,j}} \tag{3.7}$$
for all integers i and j such that $0 \leq i \leq h$ and $1 \leq j \leq \ell_i$, while

- $\mathsf{Null}_{\widehat{A},k,s}$ is the number of choices of polynomials $f_{i,j,m} \in \mathsf{F}_q[x]$ for $0 \leq i \leq h$ and $1 \leq j \leq \widehat{\ell}_i$ and of polynomials $g_1, g_2, \ldots, g_k \in \mathsf{F}_q[x]$ such that the above degree constraints are satisfied, and such that the equation at line (3.7) is satisfied for $0 \leq i \leq h$ and $1 \leq j \leq \widehat{\ell}_i$.

It suffices to note that, for any choice of polynomials $f_{i,j,m}$ and $g_1, g_2, \ldots, g_m$ satisfying the second set of constraints, for $0 \leq i \leq h$ and $1 \leq j \leq \widehat{\ell}_i$, there are at most $q^{\widetilde{N}k}$ ways to choose polynomials $f_{i,j,m} \in \mathsf{F}_q[x]$ for $0 \leq i \leq h$ and $\widehat{\ell}_i + 1 \leq j \leq \ell_i$ in order for the first set of constraints to be satisfied — for there are $q^{\widetilde{N}k}$ choices of these polynomials satisfying the given degree constraints. □

3.2 Definitions and Properties Concerning $A \cdot \mathcal{K}_{A,k,s}$

Let r be the rank of A. Then, since $Z_{0,j}$ has nullity one and rank $n_{0,j} - 1$ for all j, the block matrix Δ_0 has rank

$$\sum_{j=0}^{\ell_0} (n_{0,j} - 1) = \left(\sum_{j=0}^{\ell_0} n_{0,j} \right) - \ell_0 = n_0 - \ell_0.$$

On the other hand, if $1 \leq i \leq h$ then the matrix $\Delta_i \in \mathsf{F}_q^{n_i \times n_i}$ is nonsingular, so its rank is n_i. Now, since A is similar to a block diagonal matrix with diagonal blocks $\Delta_0, \Delta_1, \ldots, \Delta_h$, it follows by the above that

$$r = n_0 - \ell_0 + \sum_{h=1}^{h} n_h = n - \ell_0.$$

Definition 3.8. Suppose that s is a positive integer and that $s_{k,1}, s_{k,2}, \ldots, s_{k,k}$ are as in Definition 3.1. Let $A\text{-}\mathsf{Null}_{A,k,s}$ be the number of choices of vectors $v_1, v_2, \ldots, v_k \in \mathsf{F}_q^{n \times 1}$ and scalars

$$d_{1,0}, d_{1,1}, \ldots, d_{1,s_{k,1}-1}, \ldots, d_{k,0}, d_{k,1}, \ldots, d_{k,s_{k,k}-1} \in \mathsf{F}_q$$

such that

$$A \cdot \mathcal{K}_{A,\vec{v},s} \vec{d} = 0 \qquad \text{for } \vec{d} = \begin{bmatrix} d_{1,0} \\ d_{1,1} \\ \vdots \\ d_{k,s_{k,k}-1} \end{bmatrix}.$$

The following lemma replaces Lemma 3.3 in this analysis.

Lemma 3.9. *If A is as given in equations (2.1a)–(2.1j) and k and s are positive integers, then $A\text{-}\mathsf{Null}_{A,k,s}$ is equal to the number of choices of polynomials $f_{i,j,m} \in \mathsf{F}_q[x]$ for $0 \leq i \leq h$, $1 \leq j \leq \ell_i$, and $1 \leq m \leq k$, where $\deg(f_{i,j,m}) < d_i n_{i,j}$ for all i, j and m as above, and of polynomials $g_1, g_2, \ldots, g_k \in \mathsf{F}_q[x]$ such that $\deg(g_j) < s_{k,j}$ for $1 \leq j \leq k$, such that*

$$f_{0,j,1} g_1 + f_{0,j,2} g_2 + \cdots + f_{0,j,k} g_k \equiv 0 \bmod x^{n_{0,j}-1} \tag{3.8}$$

for every integer j such that $1 \leq j \leq \ell_0$, and such that

$$f_{i,j,1} g_1 + f_{i,j,2} g_2 + \cdots + f_{i,j,k} g_k \equiv 0 \bmod \varphi_i^{n_{i,j}} \tag{3.9}$$

for all integers i and j such that $1 \leq i \leq h$ and $1 \leq j \leq \ell_i$.

Proof. The proof of this lemma is quite similar to that of Lemma 3.3; considering the matrix $A \cdot \mathcal{K}_{A,\vec{v},s}$ instead of $\mathcal{K}_{A,\vec{v},s}$, one can apply Lemma 2.1 to confirm that $A\text{-}\mathsf{Null}_{A,k,s}$ is equal to the number of choices of polynomials $f_{i,j,m} \in \mathsf{F}_q[x]$ for $0 \leq i \leq h$, $1 \leq j \leq \ell_i$, and $1 \leq m \leq k$, where $\deg(f_{i,j,m}) < d_i n_{i,j}$ for all i, j and m as above, and of polynomials $g_1, g_2, \ldots, g_k \in \mathsf{F}_q[x]$ such that $\deg(g_j) < s_{k,j}$ for $1 \leq j \leq k$ such that

$$x f_{i,j,1} g_1 + x f_{i,j,2} g_2 + \cdots + x f_{i,j,k} g_k \equiv 0 \bmod \varphi_i^{n_{i,j}}$$

for all integers i and j such that $0 \leq i \leq m$ and $1 \leq j \leq \ell_i$. It now suffices to note that, since $\varphi_0 = x$,

$$x f_{0,j,1} g_1 + x f_{0,j,2} g_2 + \cdots + x f_{0,j,k} g_k \equiv 0 \bmod \varphi_0^{n_{0,j}}$$

if and only if the condition at line (3.8) is satisfied for all j such that $1 \leq j \leq \ell_0$. On the other hand, since the polynomials x and φ_i are relatively prime for $1 \leq i \leq h$,

$$xf_{i,j,1}g_1 + xf_{i,j,2}g_2 + \ldots xf_{i,j,k}g_k \equiv 0 \bmod \varphi_i^{n_{i,j}}$$

if and only if the condition at line (3.9) is satisfied for $1 \leq i \leq h$ and $1 \leq j \leq \ell_i$, as needed to establish the claim. □

Since the rank of $A \cdot \mathcal{K}_{A,\vec{v},s}$ is at most that of A, the following plays the role of Definition 3.4 in this analysis.

Definition 3.10. Let $\widehat{\rho}_{A,k,s}(j)$ be the probability that $A \cdot \mathcal{K}_{A,\vec{v},s}$ is rank deficient by exactly j if the vectors $\vec{v} = v_1, v_2, \ldots, v_k$ are chosen uniformly and independently from $\mathsf{F}_q^{n\times 1}$ — that is,

$$\widehat{\rho}_{A,k,s}(j) = \mathsf{Pr}[\mathsf{rank}(A \cdot \mathcal{K}_{A,\vec{v},s}) = \min(r,s) - j].$$

Let $\widehat{\sigma}_{A,k,s}(j)$ be the probability that $A \cdot \mathcal{K}_{A,\vec{v},s}$ is rank deficient by *at least* j if the vectors $\vec{v} = v_1, v_2, \ldots, v_k$ are chosen uniformly and independently from $\mathsf{F}_q^{n\times 1}$, that is,

$$\widehat{\sigma}_{A,k,s}(j) = \mathsf{Pr}[\mathsf{rank}(A \cdot \mathcal{K}_{A,\vec{v},s}) \leq \min(r,s) - j].$$

The proof of the following claim is virtually identical to that of Lemma 3.5: One should simply replace $\mathcal{K}_{A,\vec{v},s}$ with $A \cdot \mathcal{K}_{A,\vec{v},s}$ and n with r (as the upper bound for maximal rank when $s \geq r$) in the argument — noting, as well, that if the matrix $A \cdot \mathcal{K}_{A,\vec{v},s}$ has rank $r - j$, when $s \geq r$, then the dimension of its right null space is $s - r + j$.

Lemma 3.11. *If the vectors $\vec{v} = v_1, v_2, \ldots, v_k$ are chosen uniformly and independently from $\mathsf{F}_q^{n\times 1}$ then*

$$A\text{-}\mathsf{Null}_{A,k,s} \leq \sum_{j=0}^{\min(r,s)} q^{nk+s-\min(r,s)+j}\widehat{\rho}_{A,k,s}(j).$$

Now $\widehat{\sigma}_{A,k,s}(0) = 1$, $\widehat{\rho}_{A,k,s}(j) = \widehat{\sigma}_{A,k,s}(j) - \widehat{\sigma}_{A,k,s}(j+1)$ if $0 \leq j < \min(r,s)$, and $\widehat{\sigma}_{A,k,s}(j) = \widehat{\rho}_{A,k,s}(j)$ if $j = \min(r,s)$. The following can be concluded from the above lemma, just as Corollary 3.6 can be concluded from Lemma 3.5 using essentially the same relationships between $\rho_{A,k,s}(j)$ and $\sigma_{A,k,s}(j)$.

Corollary 3.12. *If vectors $\vec{v} = v_1, v_2, \ldots, v_k$ are chosen uniformly and independently from $\mathsf{F}_q^{n\times 1}$ then*

$$A\text{-}\mathsf{Null}_{A,k,s} \leq q^{nk+s-\min(r,s)}\left(1 + (q-1)\sum_{j=1}^{\min(r,s)} q^{j-1}\right).$$

Corollary 3.12 implies that

$$\widehat{\sigma}_{A,k,s}(j) \leq \frac{(A\text{-}\mathsf{Null}_{A,k,s} - q^{nk+s-\min(r,s)})}{(q-1)q^{nk+s-\min(r,s)+j-1}}. \tag{3.10}$$

The following lemmas will be of use in obtaining bounds for $A\text{-}\mathsf{Null}_{A,k,s}$ from bounds for $\mathsf{Null}_{A,k,s}$.

Lemma 3.13. *Let $g_1, g_2, \ldots, g_k \in \mathsf{F}_q[x]$ such that $\deg(g_j) < s_{k,j}$ for $1 \leq j \leq k$ and at least one of these polynomials is nonzero. Then the number of polynomials $f_{i,j,m} \in \mathsf{F}_q[x]$ such that $\deg(f_{i,j,m}) < d_i n_{i,j}$ for $0 \leq i \leq h$, $1 \leq j \leq \ell_i$ and $1 \leq m \leq k$, such that*

$$f_{0,j,1}g_1 + f_{0,j,2}g_2 + \cdots + f_{0,j,k}g_k \equiv 0 \bmod x^{n_{0,j}-1}$$

for every integer j such that $1 \leq j \leq \ell_0$, and such that

$$f_{i,j,1}g_1 + f_{i,j,2}g_2 + \cdots + f_{i,j,k}g_k \equiv 0 \bmod \varphi_i^{n_{i,j}}$$

for all integers i and j such that $1 \leq i \leq h$ and $1 \leq j \leq \ell_i$ is less than or equal to the product of q^{ℓ_0} and the number of polynomials $f_{i,j,m}$ satisfying the same degree constraints and such that

$$f_{i,j,1}g_1 + f_{i,j,2}g_2 + \cdots + f_{i,j,k}g_k \equiv 0 \bmod \varphi_i^{n_{i,j}} \tag{3.11}$$

for all integers i and j such that $0 \leq i \leq h$ and $1 \leq j \leq \ell_i$.

Proof. Let S be the set of choices of polynomials $f_{0,j,m}$ such that the given degree constraints are satisfied for $1 \leq j \leq \ell_0$ and $1 \leq m \leq k$ and such that the equation at line (3.11) is satisfied for $1 \leq j \leq \ell_0$. Since the constants on choices of polynomials $f_{i,j,m}$ such that $1 \leq i \leq h$, $1 \leq j \leq \ell_i$ and $1 \leq m \leq k$ are the same for both cases, it is necessary and sufficient to prove that the number of choices of polynomials $f_{0,j,m}$ for $1 \leq j \leq \ell_0$ and $1 \leq m \leq k$, satisfying the given degree constraint and the condition at line (3.8) for $1 \leq j \leq \ell_0$, is at most the product of q^{ℓ_0} and the size of the set S.

With that noted, suppose that $f_{0,j,m}$ are polynomials such that the degree constraints are satisfied for $1 \leq j \leq \ell_i$ and $1 \leq m \leq k$ and the constant at line (3.8) is satisfied for $1 \leq j \leq \ell_0$. Then there exist elements $\alpha_1, \alpha_2, \ldots, \alpha_{\ell_0}$ of F_q such that

$$f_{0,j,1}g_1 + f_{0,j,2}g_2 + \cdots + f_{0,j,k}g_k \equiv \alpha_j x^{n_{0,j}-1} \bmod x^{n_{0,j}} \tag{3.12}$$

for $1 \leq j \leq \ell_0$.

Consider any choice of $\alpha_1, \alpha_2, \ldots, \alpha_{\ell_0}$ as above. Suppose, first, that there exist polynomials $\widehat{f}_{0,j,m}$ such that $\deg(\widehat{f}_{0,j,m}) < n_{0,j}$ for $1 \leq j \leq \ell_0$ and $1 \leq m \leq k$ and such that

$$\widehat{f}_{0,j,1}g_1 + \widehat{f}_{0,j,2}g_2 + \cdots + \widehat{f}_{0,j,k}g_k \equiv \alpha_j x^{n_{0,j}-1} \bmod x^{n_{0,j}}$$

for $1 \leq j \leq \ell_0$. Then there are exactly $|S|$ choices of polynomials satisfying these constraints — for each other choice of polynomials $\overline{f}_{0,j,m}$ (again, for $1 \leq j \leq \ell_0$ and $1 \leq m \leq k$) satisfies these constants if and only if

$$\overline{f}_{0,j,m} = \widehat{f}_{0,j,m} + f_{0,j,m} \qquad \text{for } 1 \leq j \leq \ell_0 \text{ and } 1 \leq m \leq k$$

where $f_{0,j,m}$ (for j and m as above) is one of the choices of polynomials in S.

On the other hand, if there is no choice of polynomials $\widehat{f}_{0,j,m}$ as described above then the number of choices of polynomials satisfying the above condition is at most $|S|$ once again, since $|S| \geq 0$.

The claim now follows because there are q^{ℓ_0} choices of the above sequence of values $\alpha_1, \alpha_2, \ldots, \alpha_{\ell_0} \in \mathsf{F}_q$. □

Lemma 3.14. *Suppose that $g_1 = g_2 = \cdots = g_k = 0 \in \mathsf{F}_q[x]$. Then there are q^{nk} choices of polynomials $f_{i,j,m} \in \mathsf{F}_q[x]$ such that $\deg(f_{i,j,m}) < d_i n_{i,j}$ for $0 \leq i \leq h$, $1 \leq j \leq \ell_i$ and $1 \leq m \leq k$, the constraint at line (3.8) is satisfied for $1 \leq j \leq \ell_0$, and the constraint at line (3.9) is satisfied for $1 \leq i \leq h$ and $1 \leq j \leq \ell_i$.*

Proof. If $g_1 = g_2 = \cdots = g_k = 0$ then the constraint at line (3.8) is satisfied for $1 \leq j \leq \ell_0$ and the constraint at line (3.9) is satisfied for $1 \leq i \leq h$ and $1 \leq j \leq \ell_i$, for all choices of polynomials $f_{i,j,m} \in \mathsf{F}_q[x]$ such that $\deg(f_{i,j,m}) < d_i n_{i,j}$ for $0 \leq i \leq h$, $1 \leq j \leq \ell_i$, and $1 \leq m \leq k$. There are q^{nk} such choices of polynomials. □

Let ψ_1 be the minimal polynomial of A. Since $\gcd(g_1, g_2, \ldots, g_k, \psi_1)$ is a (well-defined) nonzero divisor ξ of ψ_1 with degree less than s/k, for any choice of polynomials $g_1, g_2, \ldots, g_k$ such that $\deg(g_j) < s_{k,j}$ for $1 \leq j \leq k$ such that at least one of these polynomials is nonzero, and there are at most q^{nk} choices of polynomials $f_{i,j,m}$ such that $\deg(f_{i,j,m} < \varphi^{n_{i,j}}$ (for $0 \leq i \leq h$, $1 \leq j \leq \ell_i$, and $1 \leq m \leq k$) to be considered when

$$g_1 = g_2, = \cdots = g_k = 0,$$

such it follows by Lemma 3.3 that

$$\mathsf{Null}_{A,k,s} \leq q^{nk} + \sum_{\substack{\xi \in \mathsf{F}_q[x] \setminus \{0\} \\ \xi \text{ divides } \psi_1 \\ \deg(\xi) < s/k}} \mu(\xi)\nu(\xi) \tag{3.13}$$

and it follows by Lemma 3.9 that

$$A\text{-}\mathsf{Null}_{A,k,s} \leq q^{nk} + \sum_{\substack{\xi \in \mathsf{F}_q[x] \setminus \{0\} \\ \xi \text{ divides } \psi_1 \\ \deg(\xi) < s/k}} \mu(\xi)\widehat{\nu}(\xi) \tag{3.14}$$

where, for any divisor ξ of ψ_1 with degree less than s/k,

- $\mu(\xi)$ is the number of choices of polynomials $g_1, g_2, \ldots, g_k \in \mathsf{F}_q[x]$ such that $\deg(g_j) < s_{k,j}$ for $1 \leq j \leq k$, and
$$\gcd(g_1, g_2, \ldots, g_k, \psi_1) = \xi;$$

- for any choice of polynomials $g_1, g_2, \ldots g_k$ satisfying the above condition, $\nu(\xi)$ is the number of choices of polynomials $f_{i,j,m}$ such that $\deg(f_{i,j,m}) < d_i m_{i,j}$ for $0 \leq i \leq h$, $1 \leq j \leq \ell_i$ and $1 \leq m \leq k$, and such that the condition at line (3.11), above, is satisfied for $0 \leq i \leq h$ and $1 \leq j \leq \ell_i$, and

- for any choice of polynomials $g_1, g_2, \ldots, g_k$ satisfying the above condition, $\widehat{\nu}(\xi)$ is the number of choices of polynomials $f_{i,j,m}$ satisfying the above degree constraint such that the condition at line (3.8) is satisfied for $1 \leq j \leq \ell_0$ and such that the condition at line (3.9) is satisfied for $1 \leq i \leq h$ and $1 \leq j \leq \ell_i$.

It follows by Lemma 3.13 that $\widehat{\nu}(\xi) \leq q^{\ell_0}\nu(\xi)$ for any nonzero divisor ξ of ψ_1., so that

$$\sum_{\substack{\xi \in \mathsf{F}_q[x] \setminus \{0\} \\ \xi \text{ divides } \psi_1 \\ \deg(\xi) < s/k}} \mu(\xi)\widehat{\nu}(\xi) \leq q^{\ell_0} \sum_{\substack{\xi \in \mathsf{F}_q[x] \setminus \{0\} \\ \xi \text{ divides } \psi_1 \\ \deg(\xi) < s/k}} \mu(\xi)\widehat{\nu}(\xi)$$

and (using the equations at lines (3.13) and (3.14), above)

$$A\text{-}\mathsf{Null}_{A,k,s} \leq q^{\ell_0}\mathsf{Null}_{A,k,s} - (q^{\ell_0} - 1)q^{nk}. \tag{3.15}$$

Once again, suppose that $\widehat{\ell}_0, \widehat{\ell}_1, \ldots, \widehat{\ell}_h$ are integers such that $0 \leq \widehat{\ell}_i \leq \ell_i$ for $0 \leq i \leq h$, and consider the decomposition of A described in Subsection 2.3 — considering, in particular, the matrices $\widehat{A} \in \mathsf{F}_q^{\widehat{N} \times \widehat{N}}$ and $\widetilde{A} \in \mathsf{F}_q^{\widetilde{A} \times \widetilde{A}}$ as defined at lines (2.3a)–(2.3d).

The proof of the following lemma is virtually identical to that of Lemma 3.7 — one should simply use Lemma 3.9, instead of Lemma 3.3, in order to relate $A\text{-}\mathsf{Null}_{A,k,s}$ and $A\text{-}\mathsf{Null}_{\widehat{A},k,s}$.

Lemma 3.15. *If A, $\widehat{A} \in \mathsf{F}_q^{\widehat{N} \times \widehat{N}}$ and $\widetilde{A} \in \mathsf{F}_q^{\widetilde{N} \times \widetilde{N}}$ are as described above, then*

$$A\text{-}\mathsf{Null}_{A,k,s} \leq q^{\widetilde{N}k} A\text{-}\mathsf{Null}_{\widehat{A},k,s}$$

for all positive integers k and s.

3.3 Avoidance of a Subspace

It will also be useful to bound the probability that $\mathcal{K}_{A,\vec{v},s}\vec{d}$ belongs to a given subspace of $\mathsf{F}_q^{n \times 1}$.

Lemma 3.16. *If A is as given in equations (2.1a)–(2.1j), k and s are positive integers, and y is a nonzero vector in $\mathsf{F}_q^{n\times 1}$, then there are at most $\mathsf{Null}_{A,k,s} - q^{nk}$ choices of vectors $\vec{v} = v_1, v_2, \ldots, v_k$ and scalars $d_{0,0}, d_{0,1}, \ldots, d_{k,s_{k,k}-1} \in \mathsf{F}_q$ such that*

$$\mathcal{K}_{A,\vec{v},s}\vec{d} = y \qquad \text{for } \vec{d} = \begin{bmatrix} d_{0,0} \\ d_{0,1} \\ \vdots \\ d_{k,s_{k,k}-1} \end{bmatrix}. \tag{3.16}$$

Proof. Once again recall that, by Lemma 3.3, $\mathsf{Null}_{A,k,s}$ is the number of choices of polynomials $g_1, g_2, \ldots, g_k \in \mathsf{F}_q[x]$ such that $\deg(g_j) < s_{k,j}$ and of polynomials $f_{i,j,m} \in \mathsf{F}_q[x]$ such that $\deg(f_{i,j,m}) < d_i n_{i,j}$ for $0 \leq i \leq h$, $1 \leq j \leq \ell_i$ and $1 \leq m \leq k$, and such that

$$f_{i,j,0}g_0 + f_{i,j,1}g_1 + \cdots + f_{i,j,k}g_k \equiv 0 \bmod \varphi_i^{n_{i,j}} \tag{3.17}$$

for $0 \leq i \leq h$ and $1 \leq j \leq \ell_i$. Indeed, as described in Lemma 2.1, the above polynomials $f_{i,j,m}$ for i, j and m as above correspond to different choices of the vectors $v_1, v_2, \ldots, v_k$, while the coefficients of polynomials $g_1, g_2, \ldots, g_k$ are simply the coefficients of a vector $\vec{d}$ as shown above.

Consider a choice of polynomials $g_1, g_2, \ldots, g_k \in \mathsf{F}_q[x]$ such that $\deg(g_j) < s_{k,j}$ for $1 \leq j \leq k$ (and a corresponding vector $\vec{d}$) as above. Suppose first that $g_j \neq 0$ for at least one integer j such that $1 \leq j \leq k$ and let S be the set of choices of polynomials $f_{i,j,m} \in \mathsf{F}_q[x]$ such that $0 \leq i \leq h$, $1 \leq j \leq \ell_i$, $1 \leq m \leq k$, and such that the equation at line (3.17) is satisfied for $0 \leq i \leq h$ and $1 \leq j \leq \ell_i$.

Suppose well that there exists a set of polynomials $\widehat{f}_{i,j,m}$ for $0 \leq i \leq h$, $1 \leq j \leq \ell_i$ and $1 \leq m \leq k$ — corresponding to vectors $\vec{w} = w_1, w_2, \ldots, w_k \in \mathsf{F}_q^{n\times 1}$ — such that $\mathcal{K}_{A,\vec{w},s}\vec{d} = y$. Then the number of polynomials $f_{i,j,m} \in \mathsf{F}_q[x]$ for $0 \leq i \leq h$, $1 \leq j \leq \ell_i$ and $1 \leq m \leq k$ corresponding to vectors $\vec{v} = v_1, v_2, \ldots, v_k$ such that $\mathcal{K}_{A,\vec{v},s}\vec{d} = y$ is equal to the size of the above set S — for vectors $v_1, v_2, \ldots, v_k$ satisfy this condition if and only if $v_m = w_m + \widetilde{v}_m$ for $1 \leq m \leq k$ and the vectors $\widetilde{v}_1, \widetilde{v}_2, \ldots, \widetilde{v}_k \in \mathsf{F}_q^{n\times 1}$ correspond to one of the choices of polynomials $\widetilde{f}_{i,j,m} \in \mathsf{F}_q[x]$ included in S.

On the other hand, if at least one polynomial g_j is nonzero and there does not exist a set of polynomials $\widehat{f}_{i,j,m}$ for $0 \leq i \leq h$, $1 \leq j \leq \ell_i$ and $1 \leq m \leq k$, corresponding to vectors $\widehat{w} = w_1, w_2, \ldots, w_k$ such that $\mathcal{K}_{A,\vec{w},s}\vec{d} = y$, then the number of sets of polynomials $\widehat{f}_{i,j,m}$ satisfying this condition is at most $|S|$ once again, since $|S| \geq 0$.

It remains only to consider the case that $g_j = 0$ for $1 \leq j \leq k$, in which case $\vec{d}$ is the zero vector. There are q^{nk} choices of polynomials $f_{i,j,m} \in \mathsf{F}_q[x]$ such that $\deg(f_{i,j,m}) < d_i n_{i,j}$ for $0 \leq i \leq h$, $1 \leq j \leq \ell_i$ and $1 \leq m \leq k$. Let $\vec{v} = v_1, v_2, \ldots, v_k \in \mathsf{F}_q^{n\times 1}$ be the corresponding vectors; then $\mathcal{K}_{A,\vec{v},s}\vec{d}$ is the zero vector for all such choices of polynomials and since the vector y is nonzero, $\mathcal{K}_{A,\vec{v},s}\vec{d}$ is not equal to y for any of them.

Considering all possible choices of the polynomials $g_1, g_2, \ldots, g_k$ it now follows that the number of choices of vectors $\vec{v} = v_1, v_2, \ldots, v_k \in \mathsf{F}_q^{n\times 1}$ and scalars $d_{0,0}, d_{0,1}, \ldots, d_{k,s_{k,k}-1} \in \mathsf{F}_q$ such that the equation at line (3.16) is satisfied is at most $\mathsf{Null}_{A,k,s} - q^{nk}$, as claimed. □

Definition 3.17. Suppose that $A \in \mathsf{F}_q^{n\times n}$, k and s are positive integers, and that χ is a subspace of $\mathsf{F}_q^{n\times 1}$. Then $\mathsf{Null}_{A,k,s,\chi}$ is the number of choices of vectors $\vec{v} = v_1, v_2, \ldots, v_k \in \mathsf{F}_q^{n\times 1}$ and scalars $d_{0,0}, d_{1,1}, \ldots, d_{k,s_{k,k}-1}$ such that

$$\mathcal{K}_{A,\vec{v},s}\vec{d} \in \chi \qquad \text{for } \vec{d} = \begin{bmatrix} d_{0,0} \\ d_{0,1} \\ \vdots \\ d_{k,s_{k,k}-1} \end{bmatrix}.$$

The following is now immediate from Lemma 3.16.

Corollary 3.18. *If A is as given in equations* (2.1a)–(2.1j)*, k and s are positive integers, and χ is a subspace of $\mathsf{F}_q^{n\times 1}$ with dimension d, then*

$$\mathsf{Null}_{A,k,s,\chi} \leq q^d \cdot \mathsf{Null}_{A,k,s} - \left(q^d - 1\right) \cdot q^{nk}.$$

Proof. If χ has dimension d then χ includes the zero vector along with $(q^d - 1)$ nonzero vectors $y \in \mathsf{F}_q^{n\times 1}$. Now, the number of choices of vectors $\vec{v} = v_1, v_2, \ldots, v_k \in \mathsf{F}_q^{n\times 1}$ and scalars $d_{1,0}, d_{1,1}, \ldots, d_{k,s_{k,k}-1} \in \mathsf{F}_q$ such that $\mathcal{K}_{A,\vec{v},s}\vec{d} = 0$ is, by definition, $\mathsf{Null}_{A,k,s}$. On the other hand, if $y \in \mathsf{F}_q^{n\times 1}$ is a nonzero element of χ then it follows by Lemma 3.16 that the number of choices of vectors $\vec{v} = v_1, v_2, \ldots, v_k \in \mathsf{F}_q^{n\times 1}$ and scalars $d_{1,0}, d_{1,1}, \ldots, d_{k,s_{k,k}-1}$ such that the equation at line (3.16) is satisfied is at most $\mathsf{Null}_{A,k,s} - q^{nk}$. It now follows that

$$\begin{aligned} \mathsf{Null}_{A,k,s,\chi} &\leq \mathsf{Null}_{A,k,s} + \left(q^d - 1\right)\left(\mathsf{Null}_{A,k,s} - q^{nk}\right) \\ &= q^d \cdot \mathsf{Null}_{A,k,s} - \left(q^d - 1\right) \cdot q^{nk} \end{aligned}$$

as claimed. □

3.4 The Matrices $\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R}$ and $\widehat{\mathcal{H}}_{A,\vec{u},s_L,\vec{v},s_R}$

Block Hankel matrices, as defined below, will also be considered.

Definition 3.19. For a matrix $A \in \mathsf{F}_q^{n\times n}$ and for positive integers k_L, s_L, k_R and s_R, and vectors $\vec{u} = u_1, u_2, \ldots, u_{k_L} \in \mathsf{F}_q^{n\times 1}$ and $\vec{v} = v_1, v_2, \ldots, v_{k_R} \in \mathsf{F}_q^{n\times 1}$, let

$$\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R} = \mathcal{K}^T_{A^T,\vec{u},s_L} \cdot \mathcal{K}_{A,\vec{v},s_R} \in \mathsf{F}_q^{s_L\times s_R}.$$

and let

$$\widehat{\mathcal{H}}_{A,\vec{u},s_L,\vec{v},s_R} = \mathcal{K}^T_{A^T,\vec{u},s_L} \cdot A \cdot \mathcal{K}_{A,\vec{v},s_R} \in \mathsf{F}_q^{s_L\times s_R}.$$

Note that, since the rank of $\mathcal{K}_{A^T,\vec{u}s_L} \in \mathsf{F}_q^{n\times s_L}$ cannot exceed $\min(n, s_L)$ and the rank of $\mathcal{K}_{A,\vec{v},s_R} \in \mathsf{F}_q^{n\times s_R}$ cannot exceed $\min(n, s_R)$, the rank of $\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R}$ cannot exceed $\min(n, s_L, s_R)$, and the rank of $\widehat{\mathcal{H}}_{A,\vec{u},s_L,\vec{v},s_R}$ cannot exceed $\min(r, s_L, s_R)$.

Definition 3.20. For a matrix $A \in \mathsf{F}_q^{n\times n}$ and positive integers k_L, s_L, k_R and s_R, let $\mathsf{Null}_{A,k_L,s_L,k_R,s_R}$ be the number of choices of vectors

$$\vec{u} = u_1, u_2, \ldots, u_{k_L} \in \mathsf{F}_q^{n\times 1} \quad \text{and} \quad \vec{v} = v_1, v_2, \ldots, v_{k_R} \in \mathsf{F}_q^{n\times 1}$$

and of scalars

$$d_{1,0}, d_{1,1}, \ldots, d_{1,s_{k_R,1}-1}, \ldots, d_{k_R,0}, d_{k_R,1}, \ldots, d_{k_R,s_{k_R,k_R}-1} \in \mathsf{F}_q$$

such that

$$\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R}\vec{d} = 0 \qquad \text{for } \vec{d} = \begin{bmatrix} d_{1,0} \\ d_{1,1} \\ \vdots \\ d_{k_R,s_{k_R,k_R}-1} \end{bmatrix}.$$

Let A-$\mathsf{Null}_{A,k_L,s_L,k_R,s_R}$ be the number of choices of the above vectors and scalars such that $\widehat{\mathcal{H}}_{A,\vec{u},s_L,\vec{v},s_R}\vec{d} = 0$ for $\vec{d}$ as above.

Definition 3.21. Let $A \in \mathsf{F}_q^{n\times n}$ let k_L, s_L, k_R and s_R be positive integers, and suppose that $k_L + k_R$ vectors

$$\vec{u} = u_1, u_2, \ldots, u_{k_L} \qquad \text{and} \qquad \vec{v} = v_1, v_2, \ldots, v_{k_R}$$

are chosen uniformly and independently from $\mathsf{F}_q^{n\times 1}$. For a nonnegative integer j such that $j \le \min(n, s_L, s_R)$, let $\rho_{A,k_L,s_L,k_R,s_R}(j)$ be the probability that the matrix $\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R}$ is rank deficient by j, that is,

$$\rho_{A,k_L,s_L,k_R,s_R}(j) = \mathsf{Pr}[\mathsf{rank}(\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R}) = \min(n, s_L, s_R) - j].$$

Let $\sigma_{A,k_L,s_L,k_R,s_R}(j)$ be the probability that this matrix is rank deficient by *at least* j, that is,

$$\sigma_{A,k_L,s_L,k_R,s_R}(j) = \mathsf{Pr}[\mathsf{rank}(\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R}) \le \min(n, s_L, s_R) - j].$$

Similarly, let $\rho_{A,k_L,s_L,k_R,s_R}(j)$ be the probability that the matrix $\widehat{\mathcal{H}}_{A,\vec{u},s_L,\vec{v},s_R}$ is rank deficient by j, that is,

$$\widehat{\rho}_{A,k_L,s_L,k_R,s_R}(j) = \mathsf{Pr}[\mathsf{rank}(\widehat{\mathcal{H}}_{A,\vec{u},s_L,\vec{v},s_R}) = \min(r, s_L, s_R) - j].$$

Let $\widehat{\sigma}_{A,k_L,s_L,k_R,s_R}(j)$ be the probability that this matrix is rank deficient by *at least* j, that is,

$$\widehat{\sigma}_{A,k_L,s_L,k_R,s_R}(j) = \mathsf{Pr}[\mathsf{rank}(\widehat{\mathcal{H}}_{A,\vec{u},s_L,\vec{v},s_R}) \le \min(r, s_L, s_R) - j].$$

The next result has essentially the same proof as Lemma 3.5: It suffices to note that there are $q^{n(k_L+k_R)}\rho_{A,k_L,s_L,k_R,s_R}(j)$ matrices $\mathcal{H}_{A,\vec{u},k_L,\vec{v},k_R}$ that are rank deficient by j, and that the right null space of any such matrix has dimension $s_R-\min(n,s_L,s_R)+j$. Similarly, there are $q^{n(k_L+k_R)}\widehat{\rho}_{A,k_L,s_L,k_R,s_R}(j)$ matrices $\widehat{\mathcal{H}}_{A,\vec{u},k_L,\vec{v},k_R}$ that are rank deficient by j, and the right null space of any such matrix has dimension $s_R-\min(r,s_L,s_R)+j$.

Lemma 3.22. *Let $A\in\mathsf{F}_q^{n\times n}$, let k_L, s_L, k_R and s_R be positive integers and suppose that the vectors*

$$\vec{u}=u_1,u_2,\ldots,u_{k_L}\qquad\text{and}\qquad\vec{v}=v_1,v_2,\ldots,v_{k_R}$$

and chosen uniformly and independently from $\mathsf{F}_q^{n\times 1}$. Finally, let j be an integer such that $0\le j\le\min(n,s_L,s_R)$. Then

$$\begin{aligned}\mathsf{Null}_{A,k_L,s_L,k_R,s_R}&=q^{n(k_L+k_R)+s_R-\min(n,s_L,s_R)}\sum_{j=0}^{\min(n,s_L,s_R)}\rho_{A,k_L,s_L,k_R,s_R}(j)\cdot q^j\\&=q^{n(k_L+k_R)+s_R-\min(n,s_L,s_R)}\left(1+(q-1)\sum_{j=1}^{\min(n,s_L,s_R)}\sigma_{A,k_L,s_L,k_R,s_R}(j)\cdot q^{j-1}\right)\end{aligned}$$

and

$$\begin{aligned}A\text{-}\mathsf{Null}_{A,k_L,s_L,k_R,s_R}&=q^{n(k_L+k_R)+s_R-\min(r,s_L,s_R)}\sum_{j=0}^{\min(r,s_L,s_R)}\widehat{\rho}_{A,k_L,s_L,k_R,s_R}(j)\cdot q^j\\&=q^{n(k_L+k_R)+s_R-\min(r,s_L,s_R)}\left(1+(q-1)\sum_{j=1}^{\min(r,s_L,s_R)}\widehat{\sigma}_{A,k_L,s_L,k_R,s_R}(j)\cdot q^{j-1}\right).\end{aligned}$$

Once again, useful inequalities are obtained:

$$\sigma_{A,k_L,s_L,k_R,s_R}(j)\le\frac{(\mathsf{Null}_{A,k_L,s_L,k_R,s_R}-q^{n(k_L+k_R)+s_R-\min(n,s_L,s_R)})}{(q-1)q^{n(k_L+k_R)+s_R-\min(n,s_L,s_R)+j-1}}\tag{3.18}$$

and

$$\widehat{\sigma}_{A,k_L,s_L,k_R,s_R}(j)\le\frac{(A\text{-}\mathsf{Null}_{A,k_L,s_L,k_R,s_R}-q^{n(k_L+k_R)+s_R-\min(r,s_L,s_R)})}{(q-1)q^{n(k_L+k_R)+s_R-\min(r,s_L,s_R)+j-1}}.\tag{3.19}$$

The next results will be of use in bounding $\mathsf{Null}_{A,k_L,s_L}k_Rs_R$ and $A\text{-}\mathsf{Null}_{A,k_L,s_L}k_Rs_R$.

Lemma 3.23. *Let $A\in\mathsf{F}_q^{n\times n}$ and let k_L, s_L, k_R and s_R be positive integers. Then*

$$\mathsf{Null}_{A,k_L,s_L,k_R,s_R}\le q^{n(k_L+k_R)}+q^{n-s_L}\mathsf{Null}_{A,k_L,s_L}(\mathsf{Null}_{A,k_R,s_R}-q^{nk_R})$$

and

$$A\text{-}\mathsf{Null}_{A,k_L,s_L,k_R,s_R}\le q^{n(k_L+k_R)}+q^{n-s_L}A\text{-}\mathsf{Null}_{A,k_L,s_L}(\mathsf{Null}_{A,k_R,s_R}-q^{nk_R}).$$

Proof. Consider vectors $\vec{u} = u_1, u_2, \ldots, u_{k_L}$ and the corresponding matrix $\mathcal{K}_{A^T,\vec{u},s_L}$ whose transpose is included in the definition of $\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R}$, above. This matrix has rank at most $\min(n, s_L)$, and (by definition) if the above vectors are chosen uniformly and independently from $\mathsf{F}_q^{n\times 1}$ then this matrix has rank $\min(n, s_L) - j$ with probability $\rho_{A^T,k_L,s_L}(j)$ for every integer j such that $0 \leq j \leq \min(n, s_L)$. Consequently there are $q^{nk_L}\rho_{A^T,k_L,s_L}(j)$ choices of the vectors $\vec{u}$ such that this is the case.

Fixing any such choice of $\vec{u}$, one should note that the right null space of the matrix $\mathcal{K}^T_{A^T,\vec{u},s_L} \in \mathsf{F}_q^{s_L\times n}$ is a subspace χ of $\mathsf{F}_q^{n\times 1}$ with dimension $n - \min(n, s_L) + j$. Notice, next, that for any choice of vectors

$$\vec{v} = v_1, v_2, \ldots, v_{k_R} \in \mathsf{F}_q^{n\times 1}$$

and scalars $d_{1,0}, d_{1,1}, \ldots, d_{1,s_{k_R,1}-1}, \ldots, d_{k_R,0}, d_{k_R,1}, \ldots, d_{k_R,s_{k_R,k_R}-1} \in \mathsf{F}_q$,

$$\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R}\vec{d} = 0 \qquad \text{for } \vec{d} = \begin{bmatrix} d_{1,0} \\ d_{1,1} \\ \vdots \\ d_{k_R,s_{k_R,k_R}-1} \end{bmatrix}$$

if and only if $\mathcal{K}_{A,\vec{v},s_R}\vec{d} \in \chi$. It follows by Corollary 3.18 that the number of choices of the vectors $\vec{v}$ and scalars $d_{1,0}, d_{1,1}, \ldots, d_{k_R,s_{k_R,k_R}-1}$ such that this is the case is at most

$$q^{n-\min(n,s_L)+j} \cdot \mathsf{Null}_{A,k_R,s_R} - (q^{n-\min(n,s_L)+j} - 1) \cdot q^{nk_R}.$$

Considering all choices of j, it now follows that

$$\begin{aligned}
\mathsf{Null}_{A,k_L,s_L,k_R,s_R} \leq & \sum_{j=0}^{\min(n,s_L)} (q^{nk_L}\rho_{A^T,k_L,s_L}(j) \\
& \qquad \times (q^{n-\min(n,s_L)+j}\mathsf{Null}_{A,k_R,s_R} - (q^{n-\min(n,s_L)+j} - 1)q^{nk_R})) \\
= & \sum_{j=0}^{\min(n,s_L)} (q^{nk_L}\rho_{A^T,k_L,s_L}(j) \\
& \qquad \times (q^{n-\min(n,s_L)+j}(\mathsf{Null}_{A,k_R,s_R} - q^{nk_R}) + q^{nk_R})) \\
= & \, q^{n-s_L}(\mathsf{Null}_{A,k_R,s_R} - q^{nk_R}) \sum_{j=0}^{\min(n,s_L)} q^{nk_L+s_L-\min(n,s_L)+j}\rho_{A^T,k_L,s_L}(j) \\
& \qquad + q^{n(k_L+k_R)} \sum_{j=0}^{\min(n,s_L)} \rho_{A^T,k_L,s_L}(j)
\end{aligned}$$

$$= q^{n-s_L}(\mathsf{Null}_{A,k_R,s_R} - q^{nk_R})\mathsf{Null}_{A^T,k_L,s_L} + q^{n(k_L+k_R)} \sum_{j=0}^{\min(n,s_L)} \rho_{A^T,k_L,s_L}(j)$$

(by Lemma 3.5)

$$= q^{n-s_L}(\mathsf{Null}_{A,k_R,s_R} - q^{nk_R})\mathsf{Null}_{A^T,k_L,s_L} + q^{n(k_L+k_R)}$$

since $\rho_{A^T,k_L,s_L}(j)$ are probabilities of mutually exclusive events, one of which is guaranteed to occur, so that $\sum_{j=0}^{\min(n,s_L)} \rho_{A^T,k_L,s_L}(j) = 1$.

Reorganizing terms, one can see that

$$\mathsf{Null}_{A,k_L,s_L,k_R,s_R} \leq q^{n(k_L+k_R)} + q^{n-s_L}\mathsf{Null}_{A^T,k_L,s_L}(\mathsf{Null}_{A,k_R,s_R} - q^{nk_R}).$$

It remains only to notice that, since the matrices A and A^T have the same invariant factors, it follows by Lemma 3.3 that $\mathsf{Null}_{A^T,k_L,s_L} = \mathsf{Null}_{A,k_L,s_L}$, as needed to establish the first part of the claim.

The proof of the second part of the claim is virtually identical to that of the first. In this case, one should begin by considering matrix $A^T \cdot \mathcal{K}_{A^T,\vec{u},s_L}$, just as the argument given above began by considering $\mathcal{K}_{A^T,\vec{u},s_L}$. In particular, it should be noted that, for $0 \leq j \leq \min(r, s_L)$, the matrix $A^T \cdot \mathcal{K}_{A^T,\vec{u},s_L}$ has rank $\min(r, s_L) - j$ with probability $\widehat{\rho}_{A^T,k_L,s_L}(j)$, so that there are $q^{nk_L}\widehat{\rho}_{A^T,k_L,s_L}(j)$ choices of vectors $\vec{u} = u_1, u_2, \ldots, u_{k_L} \in \mathsf{F}_q^{n\times 1}$ for which this is the case.

Fixing any such choice of $\vec{u}$, one should next note that the right null space of the matrix $\mathcal{K}^T_{A^T,\vec{u},s_L} \cdot A$ is a subspace χ of $\mathsf{F}_q^{n\times 1}$ with dimension $n - \min(r, s_L) + j$. As in the above argument, Corollary 3.18 should be applied to obtain an upper bound for A-$\mathsf{Null}_{A,k_L,s_L,k_R,s_R}$ (instead of $\mathsf{Null}_{A,k_L,s_L,k_R,s_R}$). While $\min(r, s_L)$ replaces $\min(n, s_L)$ in this derivation, and A-$\mathsf{Null}_{A^T,k_L,s_L}$ replaces $\mathsf{Null}_{A^T,k_L,s_L}$, the derivation is otherwise unchanged. □

Lemma 3.23 provides useful bounds for $\mathsf{Null}_{A,k_L,s_L,k_R,s_R}$ and A-$\mathsf{Null}_{A,k_L,s_L,k_R,s_R}$ in a common case, as described next.

Theorem 3.24. *Let $A \in \mathsf{F}_q^{n\times n}$, let k_L, s_L, k_R and s_R, and suppose that*

$$\mathsf{Null}_{A,k_L,s_L} \leq q^{nk_L} + \nu q^{(k_L-1)n+s_L}, \quad \mathsf{Null}_{A,k_R,s_R} \leq q^{nk_R} + \nu q^{(k_R-1)n+s_R}$$

and

$$A\text{-}\mathsf{Null}_{A,k_L,s_L} \leq q^{nk_L} + \widehat{\nu} q^{(k_L-1)n+s_L}$$

where $\nu, \widehat{\nu} \geq 1$. Then

$$\mathsf{Null}_{A,k_L,s_L,k_R,s_R} \leq q^{n(k_L+k_R)}(1 + \nu q^{s_R-s_L} + \nu^2 q^{s_R-n})$$

and

$$A\text{-}\mathsf{Null}_{A,k_L,s_L,k_R,s_R} \leq q^{n(k_L+k_R)}(1 + \nu q^{s_R-s_L} + \nu \cdot \widehat{\nu} q^{s_R-n}).$$

Proof. It follows by the above lemma that

$$\begin{aligned}
\mathsf{Null}_{A,k_L,s_L,k_R,s_R} &\leq q^{n(k_L+k_R)} + q^{n-s_L}\mathsf{Null}_{A,k_L,s_L}(\mathsf{Null}_{A,k_R,s_R} - q^{nk_R}) \\
&\leq q^{n(k_L+k_R)} + q^{n-s_L}\mathsf{Null}_{A,k_L,s_L} \cdot \nu q^{(k_R-1)n+s_R} \\
&= q^{n(k_L+k_R)} + \nu q^{nk_R+s_R-s_L}\mathsf{Null}_{A,k_L,s_L} \\
&\leq q^{n(k_L+k_R)} + \nu q^{nk_R+s_R-s_L}(q^{nk_L} + \nu q^{(k_L-1)n+s_L}) \\
&= q^{n(k_L+k_R)}(1 + \nu q^{s_R-s_L} + \nu^2 q^{s_R-n})
\end{aligned}$$

as claimed. The proof that

$$A\text{-}\mathsf{Null}_{A,k_L,s_L,k_R,s_R} \leq q^{n(k_L+k_R)}(1 + \nu q^{s_R-s_L} + \nu \cdot \widehat{\nu} q^{s_R-n}).$$

is virtually identical: One begins by applying the lemma to conclude that

$$A\text{-}\mathsf{Null}_{A,k_L,s_L,k_R,s_R} \leq q^{n(k_L+k_R)} + q^{n-s_L}A\text{-}\mathsf{Null}_{A,k_L,s_L}(\mathsf{Null}_{A,k_R,s_R} - q^{nk_R}).$$

and applies the given bounds on $A\text{-}\mathsf{Null}_{A,k_L,s_L}$ and $\mathsf{Null}_{A,k_R,s_R}$ in the same way. □

The next objective is to establish an analogue of Lemmas 3.7 and 3.15 — Theorem 3.27, below. Unfortunately the proof of this is somewhat more complicated because of the structure of $\mathcal{H}_{A,\vec{u},s_L,\vec{u},s_R}$ and $\widehat{\mathcal{H}}_{A,\vec{u},s_L,\vec{v},s_R}$.

With that noted suppose, once again, that $\widehat{\ell}_0, \widehat{\ell}_1, \ldots, \widehat{\ell}_h$ are integers such that $0 \leq \widehat{\ell}_i \leq \ell_i$ for $0 \leq i \leq h$ and consider the decomposition of A described in Subsection 2.3. Then

$$A^T = \widehat{X}^T \begin{bmatrix} \widehat{A}^T & 0 \\ 0 & \widetilde{A}^T \end{bmatrix} (\widehat{X}^T)^{-1} \tag{3.20a}$$

as well. Suppose that k_L, s_L, k_R and s_R are positive integers. For any set of vectors

$$\vec{u} = u_1, u_2, \ldots, u_{k_L} \in \mathsf{F}_q^{n\times 1} \tag{3.20b}$$

it is possible to write each vector u_i uniquely as

$$u_i = \widehat{u}_{i,L} + \widetilde{u}_{i,L} \tag{3.20c}$$

where $\widehat{u}_{i,L}$ is a linear combination of the vectors

$$\widehat{X}^T e_1, \widehat{X}^T e_2, \ldots, \widehat{X}^T e_{\widehat{N}}$$

and where $\widetilde{u}_{i,L}$ is a linear combination of

$$\widehat{X}^T e_{\widehat{N}+1}, \widehat{X}^T e_{\widehat{N}+2}, \ldots, \widehat{X}^T e_n$$

(where $e_1, e_2, \ldots, e_n$ are the standard unit vectors in $\mathsf{F}_q^{n\times 1}$). Henceforth let us write

$$\vec{u}_{T,L} = \widehat{u}_{1,L}, \widehat{u}_{2,L}, \ldots, \widehat{u}_{k_L,L} \qquad \text{and} \qquad \vec{u}_{B,L} = \widetilde{u}_{1,L}, \widetilde{u}_{2,L}, \ldots, \widetilde{u}_{k_L,L} \tag{3.20d}$$

for the vectors described above.

Similarly, for any set of vectors

$$\vec{v} = v_1, v_2, \ldots, v_{k_R} \in \mathsf{F}_q^{n\times 1} \tag{3.20e}$$

it is possible to write each vector v_i uniquely as

$$v_i = \widehat{v}_{i,R} + \widetilde{v}_{i,R} \tag{3.20f}$$

where $\widehat{v}_{i,R}$ is a linear combination of the vectors

$$\widehat{X}^{-1}e_1, \widehat{X}^{-1}e_2, \ldots, \widehat{X}^{-1}e_{\widehat{N}}$$

and where $\widetilde{v}_{i,R}$ is a linear combination of the vectors

$$\widehat{X}^{-1}e_{\widehat{N}+1}, \widehat{X}^{-1}e_{\widehat{N}+2}, \ldots, \widehat{X}^{-1}e_n.$$

Henceforth let us write

$$\vec{v}_{T,R} = \widehat{v}_{1,R}, \widehat{v}_{2,R}, \ldots, \widehat{v}_{k_R,R} \qquad \text{and} \qquad \vec{v}_{B,R} = \widetilde{v}_{1,R}, \widetilde{v}_{2,R} \ldots, \widetilde{v}_{k_R,R} \tag{3.20g}$$

for the vectors described above.

It is straightforward to use the above characterization of A and definition of these vectors to establish that

$$\widehat{u}_{i.L}^T A^m \widetilde{v}_{j,R} = \widetilde{u}_{i,L}^T A^m \widehat{v}_{j,R} = 0 \tag{3.20h}$$

for all integers i, j and m such that $1 \le i \le k_L$, $1 \le j \le k_R$, and $m \ge 0$. It follows that if s is a positive integer, $s_{k_R,1}, s_{k_R,2}, \ldots, s_{k_R,k_R}$ are as defined in Section 2, and

$$d_{1,0}, d_{1,1}, \ldots, d_{1,s_{k_R,1}-1}, \ldots, d_{k_R,0}, d_{k_R,1}, \ldots, d_{k_R,k_R} - 1 \in \mathsf{F}_q$$

then

$$\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R}\vec{d} = \mathcal{H}_{A,\vec{u}_{T,L},s_L,\vec{v}_{T,R},s_R}\vec{d} + \mathcal{H}_{A,\vec{u}_{B,L},s_L,\vec{v}_{B,R},s_R}\vec{d}$$

$$\text{for } \vec{d} = \begin{bmatrix} d_{1,0} \\ d_{1,1} \\ \vdots \\ d_{k_R,s_{k_R,k_R}-1} \end{bmatrix}. \tag{3.20i}$$

Lemma 3.25. *Let A be as described in Subsection 2.3, and consider any choice of vectors*

$$\vec{u}_{T,L} = \widehat{u}_{1,L}, \widehat{u}_{2,L}, \ldots, \widehat{u}_{k_L,L}, \quad \vec{u}_{B,L} = \widetilde{u}_{1,L}, \widetilde{u}_{2,L}, \ldots, \widetilde{u}_{k_L,L},$$
$$\text{and} \quad \vec{v}_{B,R} = \widetilde{v}_{1,R}, \widetilde{v}_{2,R} \ldots, \widetilde{v}_{k_R,R}$$

as described above, as well as scalars

$$d_{1,0}, d_{1,1}, \ldots, d_{1,s_{k_R,1}-1}, \ldots, d_{k_R,0}, d_{k_R,1}, \ldots, d_{k_R,s_{k_R,k_R}-1} \in \mathsf{F}_q.$$

The number of choices of vectors

$$\vec{v}_{T,R} = \widehat{v}_{1,R}, \widehat{v}_{2,R}, \ldots, \widehat{v}_{k_R,R}$$

(that are linear combinations of $\widehat{X}^{-1}e_1, \widehat{X}^{-1}e_2, \ldots, \widehat{X}^{-1}e_{\widehat{N}}$ as above) such that

$$\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R}\vec{d} = 0 \qquad \text{for } \vec{d} = \begin{bmatrix} d_{1,0} \\ d_{1,1} \\ \vdots \\ d_{k_R,s_{k_R,k_R}-1} \end{bmatrix} \tag{3.21}$$

where $\vec{u} = u_1, u_2, \ldots, u_{k_L}$, $\vec{v} = v_1, v_2, \ldots, v_{k_R}$, $u_i = \widehat{u}_{i,L} + \widetilde{u}_{i,L}$ for $1 \leq i \leq k_L$, and $v_i = \widehat{v}_{i,R} + \widetilde{v}_{i,R}$ for $1 \leq i \leq k_R$, is less than or equal to the number of choices of vectors

$$\vec{v}_{T,R} = \widehat{v}_{1,R}, \widehat{v}_{2,R}, \ldots, \widehat{v}_{k_R,R}$$

(that are linear combinations of $\widehat{X}^{-1}e_1, \widehat{X}^{-1}e_2, \ldots, \widehat{X}^{-1}e_{\widehat{N}}$ as above) such that

$$\mathcal{H}_{A,\vec{u}_{T,L},s_L,\vec{v}_{T,R},s_R}\vec{d} = 0. \tag{3.22}$$

Similarly, the number of choices of vectors $\vec{v}_{T,R}$ that are linear combinations of the vectors $\widehat{X}^{-1}e_1, \widehat{X}^{-1}e_2, \ldots, \widehat{X}^{-1}e_{\widehat{N}}$, such that $\widehat{\mathcal{H}}_{A,\vec{u},s_L,\vec{v},s_R}\vec{d} = 0$, for $\vec{u}$, $\vec{v}$ and $\vec{d}$ as above, is less than or equal to the number of choices of vectors $\vec{v}_{T,R}$ (that are linear combinations of $\widehat{X}^{-1}e_1, \widehat{X}^{-1}e_2, \ldots, \widehat{X}^{-1}e_{\widehat{N}}$) such that $\widehat{\mathcal{H}}_{A,\vec{u}_{T,L},s_L,\vec{v}_{T,R},s_R}\vec{d} = 0$.

Proof. Fix any choice of the vectors $\vec{u}_{T,L}$, $\vec{u}_{B,L}$, $\vec{v}_{B,R}$ and scalars

$$d_{1,0}, d_{1,1}, \ldots, d_{1,s_{k_R,1}-1}, \ldots, d_{k_R,0}, d_{k_R,1}, \ldots, d_{k_R,s_{k_R,k_R}-1} \in \mathsf{F}_q$$

as above, and let S be the set of choices of vectors $\vec{v}_{T,R} = \widehat{v}_{1,R}, \widehat{v}_{2,R}, \ldots, \widehat{v}_{k_R,R}$ (that are linear combinations of the vectors $\widehat{X}^{-1}e_1, \widehat{X}^{-1}e_2, \ldots, \widehat{X}^{-1}e_{\widehat{N}}$ as above) such that the condition at line (3.22) is satisfied. It is necessary and sufficient to show that the number of choices of vectors $\vec{v}_{T,R} = \widehat{v}_{1,R}, \widehat{v}_{2,R}, \ldots, \widehat{v}_{k_R,R}$ (that are linear combinations of

$\widehat{X}^{-1}e_1, \widehat{X}^{-1}e_2, \ldots, \widehat{X}^{-1}e_{\widehat{N}}$ once again) such that the condition at line (3.21) is satisfied, instead, is less than or equal to the size of the set S.

Now suppose first that there exists a sequence of vectors

$$\vec{w}_{T,R} = \widehat{w}_{1,R}, \widehat{w}_{2,R}, \ldots, \widehat{w}_{k_R,R}$$

that are linear combinations of $\widehat{X}^{-1}e_1, \widehat{X}^{-1}e_2, \ldots, \widehat{X}^{-1}e_{\widehat{N}}$ such that

$$\mathcal{H}_{A,\vec{u}_{T,L},s_L,\vec{w}_{T,R},s_R}\vec{d} = \mathcal{H}_{A,\vec{u}_{B,L},s_L\vec{v}_{B,R},s_R}\vec{d}.$$

In this case it suffices to apply the equation at line (3.20i) to conclude that there are exactly $|S|$ choices of vectors

$$\vec{v}_{T,R} = \widehat{v}_{1,R}, \widehat{v}_{2,R}, \ldots, \widehat{v}_{k_R,R}$$

that are linear combinations of $\widehat{X}^{-1}e_1, \widehat{X}^{-1}e_2, \ldots, \widehat{X}^{-1}e_{\widehat{N}}$ satisfying the condition at line (3.21) — for a sequence of vectors

$$\vec{v}_{T,R} = \widehat{v}_{1,R}, \widehat{v}_{2,R}, \ldots, \widehat{v}_{k_R,R}$$

that are linear combinations of $\widehat{X}^{-1}e_1, \widehat{X}^{-1}e_2, \ldots, \widehat{X}^{-1}e_{\widehat{N}}$ satisfies the condition at line (3.21) if and only if the sequence

$$\widehat{v}_{1,R} - \widehat{w}_{1,R}, \widehat{v}_{2,R} - \widehat{w}_{2,R}, \ldots, \widehat{v}_{k_R,R} - \widehat{w}_{k_R,R}$$

is a sequence of linear combinations of $\widehat{X}^{-1}e_1, \widehat{X}^{-1}e_2, \ldots, \widehat{X}^{-1}e_{\widehat{N}}$ satisfying the condition at line (3.22), instead.

In the remaining case, there is no sequence of vectors

$$\vec{w}_{T,R} = \widehat{w}_{1,R}, \widehat{w}_{2,R}, \ldots, \widehat{w}_{k_R,R}$$

that are linear combinations of $\widehat{X}^{-1}e_1, \widehat{X}^{-1}e_2, \ldots, \widehat{X}^{-1}e_{\widehat{N}}$ such that

$$\mathcal{H}_{A,\vec{u}_{T,L},s_L,\vec{w}_{T,R},s_R}\vec{d} = \mathcal{H}_{A,\vec{u}_{B,L},s_L\vec{v}_{B,R},s_R}\vec{d}.$$

In this case, the equation at one (3.20i) can be applied to establish that there are no vectors

$$\vec{v}_{T,R} = \widehat{v}_{1,R}, \widehat{v}_{2,R}, \ldots, \widehat{v}_{k_R,R}$$

that are linear combinations of the vectors $\widehat{X}^{-1}e_1, \widehat{X}^{-1}e_2, \ldots, \widehat{X}^{-1}e_{\widehat{N}}$ satisfying the condition at line (3.21) at all — and the first part of the claim holds because $|S| \geq 0$.

The proof of the second part of the claim is virtually identical: One can argue as above, with the matrix $\widehat{\mathcal{H}}_{A,\vec{u},s_L,\vec{v},s_R}$ replacing $\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R}$, and with the matrix $\widehat{\mathcal{H}}_{A,\vec{u}_{T,L},s_L,\vec{v}_{T,R},s_R}$ replacing $\mathcal{H}_{A,\vec{u}_{T,L},s_L,\vec{v}_{T,R},s_R}$. □

Lemma 3.26. *If k_L, s_L, k_R and s_R are positive integers and A is as described in Subsection 2.3, then there are $\mathsf{Null}_{\widehat{A},k_L,s_L,k_R,s_R}$ choices of sequences of vectors*

$$\vec{u}_{T,L} = \widehat{u}_{1,L}, \widehat{u}_{2,L}, \ldots, \widehat{u}_{k_L,L} \in \mathsf{F}_q^{n\times 1}$$

that are linear combinations of $\widehat{X}^T e_1, \widehat{X}^T e_2, \ldots, \widehat{X}^T e_{\widehat{N}}$, sequences of vectors

$$\vec{v}_{T,R} = \widehat{v}_{1,R}, \widehat{v}_{2,R}, \ldots, \widehat{v}_{k_R,R} \in \mathsf{F}_q^{n\times 1}$$

that are linear combinations of $\widehat{X}^{-1} e_1, \widehat{X}^{-1} e_2, \ldots, \widehat{X}^{-1} e_{\widehat{N}}$, and sequences of scalars

$$d_{1,0}, d_{1,1}, \ldots, d_{1,s_{k_R,1}-1}, \ldots, d_{k_R,0}, d_{k_R,1}, \ldots, d_{k_R,s_{k_R,k_R}-1} \in \mathsf{F}_q$$

such that

$$\mathcal{H}_{A,\vec{u}_{T,L},s_L,\vec{v}_{T,R},s_R}\vec{d} = 0 \qquad \text{for } \vec{d} = \begin{bmatrix} d_{1,0} \\ d_{11} \\ \vdots \\ d_{k_R,s_{k_R,k_R}-1} \end{bmatrix}, \tag{3.23}$$

and there are A-$\mathsf{Null}_{\widehat{A},k_L,s_L} k_R s_R$ choices of sequences of vectors $\vec{u}_{T,L}$ that are linear combinations of $\widehat{X}^T e_1, \widehat{X}^T e_2, \ldots, \widehat{X}^T e_{\widehat{N}}$, sequences of vectors $\vec{v}_{T,R}$ that are linear combinations of $\widehat{X}^{-1} e_1, \widehat{X}^{-1} e_2, \ldots, \widehat{X}^{-1} e_{\widehat{N}}$, and sequences of scalars $d_{1,0}, d_{1,1}, \ldots, d_{k_R,s_{k_R,k_R}-1} \in \mathsf{F}_q$, such that $\widehat{\mathcal{H}}_{A,\vec{u}_{T,L},s_L,\vec{v}_{T,R},s_R}\vec{d} = 0$ for $\vec{d}$ as above.

Proof. Let

$$0_{\widetilde{N}} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathsf{F}_q^{\widetilde{N}\times 1}.$$

Consider a sequence of vectors $\vec{u} = u_1, u_2, \ldots, u_{k_L} \in \mathsf{F}_q^{\widehat{N}\times 1}$; the sequence of vectors

$$\vec{u}_{\widehat{X},L} = u_{1,\widehat{X},L}, u_{2,\widehat{X},L}, \ldots, u_{k_L,\widehat{X},L} \in \mathsf{F}_q^{n\times 1}$$

such that

$$u_{i,\widehat{X},L} = \widehat{X}^T \begin{bmatrix} u_i \\ 0^{\widetilde{N}} \end{bmatrix}$$

for $1 \le i \le k_L$, is certainly a sequence of k_L vectors in $\mathsf{F}_q^{n\times 1}$ that are linear combinations of $\widehat{X}^T e_1, \widehat{X}^T e_2, \ldots, \widehat{X}^T e_{\widehat{N}}$. Furthermore, every sequence of vectors

$$\vec{u}_{T,L} = \widehat{u}_{1,L}, \widehat{u}_{2,L}, \ldots, \widehat{u}_{k_L,L} \in \mathsf{F}_q^{n\times 1}$$

that are linear combinations of $\widehat{X}^T e_1, \widehat{X}^T e_2, \ldots, \widehat{X}^T e_{\widehat{N}}$ is equal to the sequence $\vec{u}_{\widehat{X},L}$ corresponding to exactly one sequence of vectors $\vec{u} = u_1, u_2, \ldots, u_{k_L} \in \mathsf{F}_q^{\widehat{N}\times 1}$.

Similarly, there is a one-to-one correspondence between sequences of linear combinations of $\widehat{X}^{-1} e_1, \widehat{X}^{-1} e_2, \ldots, \widehat{X}^{-1} e_{\widehat{N}}$ and sequences of vectors in $\mathsf{F}_q^{\widehat{N}\times 1}$: Consider a sequence $\vec{v} = v_1, v_2, \ldots, v_{k_R}$ of vectors in $\mathsf{F}_q^{\widehat{N}\times 1}$. If one sets $\vec{v}_{\widetilde{X},R}$ to be the sequence of vectors

$$\vec{v}_{\widehat{X},R} = v_{1,\widehat{X},R}, v_{2,\widehat{X},R}, \ldots, v_{k_R,\widehat{X},R} \in \mathsf{F}_q^{n\times 1}$$

where

$$v_{i,\widehat{X},R} = \widehat{X}^{-1} \begin{bmatrix} v_i \\ 0_{\widetilde{N}} \end{bmatrix}$$

for $1 \leq i \leq k_R$, then every sequence

$$\vec{v}_{T,R} = \widehat{v}_{1,R}, \widehat{v}_{2,R}, \ldots, \widehat{v}_{k_R,R} \in \mathsf{F}_q^{n\times 1}$$

of vectors that are linear combinations of $\widehat{X}^{-1} e_1, \widehat{X}^{-1} e_2, \ldots, \widehat{X}^{-1} e_{\widehat{N}}$ can be expressed as $\vec{v}_{\widehat{X},R}$ for exactly one sequence of vectors $\vec{v} = v_1, v_2, \ldots, v_{k_R} \in \mathsf{F}_q^{\widehat{N}\times 1}$.

Now consider any sequence of vectors $\vec{u} = u_1, u_2, \ldots, u_{k_R} \in \mathsf{F}_q^{\widehat{N}\times 1}$ and the corresponding sequence of vectors $\vec{u}_{\widehat{X},L} \in \mathsf{F}_q^{n\times 1}$, any sequence of vectors $\vec{v} = v_1, v_2, \ldots, v_{k_R} \in \mathsf{F}_q^{\widehat{N}\times 1}$ and the corresponding sequence of vectors $\vec{v}_{\widehat{X},R} \in \mathsf{F}_q^{n\times 1}$, and any sequence of scalars

$$d_{1,0}, d_{1,1}, \ldots, d_{1,s_{k_R,1}-1}, \ldots, d_{k_R,0}, d_{k_R,1}, \ldots, d_{k_R,s_{k_R,k_R}-1} \in \mathsf{F}_q.$$

One can see by an examination of the equation at line (2.3b) that

$$\mathcal{H}_{A,\vec{u}_{\widehat{X},L},s_L,\vec{v}_{\widehat{X},R},s_R}\vec{d} = \mathcal{H}_{\widehat{A},\vec{u},s_L\vec{v},s_R}\vec{d} \qquad \text{for } \vec{d} = \begin{bmatrix} d_{1,0} \\ d_{1,1} \\ \vdots \\ d_{k_R,s_{k_R,k_R}-1} \end{bmatrix}.$$

The first part of the claim now follows by a consideration a summation over all possible choices of vectors $\vec{u}$ and $\vec{d}$ and all possible coefficients of the above vector $\vec{d}$ and by applying the definition of $\mathsf{Null}_{\widehat{A},k_L,s_L,k_R,s_R}$. The second part of the claim follows by the same argument, with the matrix $\widehat{\mathcal{H}}_{A,\vec{u}_{T,L},s_L,\vec{v}_{T,R},s_R}$ replacing $\mathcal{H}_{A,\vec{u}_{T,L},s_L,\vec{v}_{T,R},s_R}$ and with $\widehat{\mathcal{H}}_{\widehat{A},\vec{u}_{T,L},s_L,\vec{v}_{T,R},s_R}$ replacing $\mathcal{H}_{\widehat{A},\vec{u}_{T,L},s_L,\vec{v}_{T,R},s_R}$ (for various choices of vectors $\vec{u}$ and $\vec{v}$). □

Theorem 3.27. *If k_L, s_L, k_R and s_R are positive integers and A is as described in Subsection 2.3 then*

$$\mathsf{Null}_{A,k_L,s_L,k_R,s_R} = q^{(k_L+k_R)\widetilde{N}} \mathsf{Null}_{\widehat{A},k_L,s_L,k_R,s_R}$$

and

$$A\text{-}\mathsf{Null}_{A,k_L,s_L,k_R,s_R} = q^{(k_L+k_R)\widetilde{N}} A\text{-}\mathsf{Null}_{\widehat{A},k_L,s_L,k_R,s_R}.$$

Proof. By definition, $\mathsf{Null}_{A,k_L,s_L,k_R,s_R}$ is the number of choices of vectors $\vec{u} = u_1, u_2, \ldots, u_{k_L} \in \mathsf{F}_q^{n\times 1}$, vectors $\vec{v} = v_1, v_2, \ldots, v_{k_R} \in \mathsf{F}_q^{n\times 1}$, and scalars

$$d_{1,0}, d_{1,1}, \ldots, d_{1,s_{k_R,1}-1}, \ldots, d_{k_R,0}, d_{k_R,1}, \ldots, d_{k_R,s_{k_R,k_R}-1}$$

such that

$$\mathcal{H}_{A,\vec{u},s_L,\vec{v}s_R}\vec{d} = 0 \qquad \text{for } \vec{d} = \begin{bmatrix} d_{1,0} \\ d_{1,1} \\ \vdots \\ d_{k_R,s_{k_R,k_R}-1} \end{bmatrix}. \tag{3.24}$$

As noted above, each choice of vectors $\vec{u}$ corresponds to exactly one choice of vectors $\vec{u}_{T,L} = \widehat{u}_{1,L}, \widehat{u}_{2,L}, \ldots, \widehat{u}_{k_L,L}$ and vectors $\vec{u}_{B,L} = \widetilde{u}_{1,L}, \widetilde{u}_{2,L}, \ldots, \widetilde{u}_{k_L,L}$ as defined above. Similarly, each choice of vectors $\vec{v}$ corresponds to exactly one choice of vectors $\vec{v}_{T,R} = \widehat{v}_{1,R}, \widehat{v}_{2,R}, \ldots, \widehat{v}_{k_R.R}$ and of vectors $\vec{v}_{B,R} = \widetilde{v}_{1,R}, \widetilde{v}_{2,R}, \ldots, \widetilde{v}_{k_R,R}$ as defined above.

Fix any choice of the vectors $\vec{u}_{B,L}$ and $\vec{v}_{B,R}$. It follows by the first part of Lemma 3.25 (summing over all possible choices of $\vec{u}_{T,L}$ and the coefficients of the above vector $\vec{d}$) that the number of choices of $\vec{u}_{T,L}$, $\vec{v}_{T,R}$ (defining vectors $\vec{u}$ and $\vec{v}$) and the coefficients of $\vec{d}$, such that the equation at line (3.24) is satisfied, is less than or equal to the number of choices of vectors $\vec{u}_{T,L}$, $\vec{v}_{T,R}$, and coefficients of $\vec{d}$ such that

$$\mathcal{H}_{A,\vec{u}_{T,L},s_L,\vec{u}_{T,R},s_R}\vec{d} = 0$$

— and the first part of Lemma 3.26 implies that this number is $\mathsf{Null}_{\widehat{A},k_L,s_L,k_R,s_R}$.

The first part of the claim now follows by the fact that there are $q^{(k_L+k_R)\widetilde{N}}$ possible choices of the vectors $\vec{u}_{B,L}$ and $\vec{v}_{B,R}$.

The second part of the claim follows by the same argument, with applications of the first parts of Lemmas 3.25 and 3.26 replaced by applications of the second parts of these lemmas. □

4 Analysis of the Expected Nullities When There are Few Invariant Factors

Let us next consider the expected nullities of $\mathcal{K}_{A,\vec{v},s}$ and $A \cdot \mathcal{K}_{A,\vec{v},s}$ for $\vec{v} = v_1, v_2, \ldots, v_k$, and of $\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R}$ and $\widehat{\mathcal{H}}_{A,\vec{u},s_L,\vec{v},s_R}$ for $\vec{u} = u_1, u_2, \ldots, u_{k_L}$ and $\vec{v} = v_1, v_2, \ldots, v_{k_R}$, when the number of invariant factors is small — that is, when $k > \ell$ (for consideration of the first and second of these matrices) and when $\min(k_L, k_R) > \ell$ (for consideration of the third and fourth).

4.1 A Bound for $\mathbf{Null}_{A,k,s}$

Lemma 4.1. *Let s be a positive integer, and let ξ be a a nonzero factor of the minimal polynomial $\psi_1 = \varphi_0^{n_{0,1}}\varphi_1^{n_{1,1}}\dots\varphi_h^{n_{h,1}}$ of A whose degree is less than s/k.Then there are at most $q^{s-\deg(\xi)k}$ choices of polynomials $g_1, g_2, \dots, g_k \in \mathsf{F}_q[x]$ such that $\deg(g_j) < s_{k,j}$ for $1 \le j \le k$ and*

$$\gcd(g_1, g_2, \dots, g_k, \psi) = \xi.$$

For every such choice of polynomials there are at most $q^{(k-1)n+\ell\deg(\xi)}$ choices of polynomials $f_{i,j,m}$ such that $\deg(f_{i,j,m}) < d_i n_{i,j}$ for $0 \le i \le h$, $1 \le j \le \ell_i$, and $1 \le m \le k$ and such that

$$f_{i,j,1}g_1 + f_{i,j,2}g_2 + \dots + f_{i,j,k}g_k \equiv 0 \bmod \varphi_i^{n_{i,j}}$$

for all integers i and j such that $0 \le i \le h$ and $1 \le j \le \ell$.

Proof. Consider a polynomial $\xi \in \mathsf{F}_q[x]$ as described in the claim. Since ξ is a divisor of ψ_1,

$$\xi = \varphi_0^{m_0}\varphi_1^{m_1}\dots\varphi_h^{m_h} \tag{4.1}$$

for integers $m_0, m_1, \dots, m_h$ such that $0 \le m_i \le n_{i,1}$ for $1 \le i \le h$.

Let j be an integer such that $1 \le j \le k$. If $s_{k,j} > \deg(\xi)$ then a polynomial $g_j \in \mathsf{F}_q[x]$ with degree less than $s_{k,j}$ is divisible by ξ if and only if $g_j = h_j\xi$ for a polynomial $h_j \in \mathsf{F}_q[x]$ with degree less than $s_{k,j} - \deg(\xi)$. Since the coefficients of h_j can be chosen freely, there are $q^{s_{k,j}-\deg(\xi)}$ choices of $g_j \in \mathsf{F}_q[x]$. On the other hand, if $s_{k,j} = \deg(\xi)$ then a polynomial g_j with degree less than $s_{k,j}$ is divisible by ξ if and only if $g_j = 0$, so that there are $q^{s_{k,j}-\deg(\xi)}$ choices of g_j in this case as well.

Now, if $\deg(\xi) \le s/k$ then $s_{k,j} \ge \deg(\xi)$ for all j such that $1 \le j \le k$ and, since $s_{k,1} + s_{k,2} + \dots + s_{k,k} = s$, it follows that there are there are $q^{s-\deg(\xi)k}$ choices of the sequences of polynomials $g_1, g_2, \dots, g_k$ such that g_j has degree less than $s_{k,j}$ for $1 \le j \le k$ and each is divisible by ξ. It follows that there are also at most $q^{s-\deg(\xi)k}$ choices of polynomials $g_1, g_2, \dots, g_k \in \mathsf{F}_q[x]$ such that g_j has degree less than $s_{k,j}$ for $1 \le j \le k$ and

$$\gcd(g_1, g_2, \dots, g_k, \psi) = \xi, \tag{4.2}$$

as well, since there latter condition implies the former.

Now consider any choice of polynomials $g_1, g_2, \dots, g_k \in \mathsf{F}_q[x]$ such that g_j has degree less than $s_{k,j}$ for $1 \le j \le k$ and such that the equation at line (4.2) is satisfied, and let i be an integer such that $0 \le i \le h$. Since the polynomial ξ has the factorization shown at line (4.1), either

- $m_i < n_{i,1}$, each of $g_1, g_2, \dots, g_k$ is divisible by $\varphi_i^{m_i}$ and at least one of these polynomials is not divisible by $\varphi_i^{m_i+1}$, or
- $m_i = n_{i,1}$ and $g_1, g_2, \dots, g_k$ are all divisible by $\varphi_i^{n_i,1}$.

Consider the first case, so that $m_i < n_{i,1}$ and each of $g_1, g_2, \ldots, g_k$ is divisible by $\varphi_i^{m_i}$ and at least one of these polynomials is not divisible by $\varphi_i^{m_i+1}$. In this case there exist polynomials $\widehat{h}_1, \widehat{h}_2, \ldots, \widehat{h}_k \in \mathsf{F}_q[x]$ such that $g_r = \widehat{h}_r \cdot \varphi_i^{m_i}$ for every integer r such that $1 \leq r \leq k$, and such that $\widehat{h}_r$ is not divisible by φ_i for at least one such integer r. Then, since φ_i is irreducible, $\widehat{h}_r$ is relatively prime with both φ_i and $\varphi^{n_{i,1}-m_i}$ and there exist polynomials $\widehat{s}, \widehat{t} \in \mathsf{F}_q[x]$ such that

$$\widehat{s}\widehat{h}_r + \widehat{t}\varphi_i^{n_{i,1}-m_i} = 1,$$

so that $\widehat{s}\widehat{h}_r \equiv 1 \bmod \varphi_i^{n_{i,1}-m_i}$.

Now consider any integer j such that $1 \leq j \leq \ell_i$ and $n_{i,j} > m_i$ as well; since $n_{i,j} \leq n_{i,1}$, $\widehat{s}\widehat{h}_r \equiv 1 \bmod \varphi_i^{n_{i,j}-m_i}$ as well, and

$$\begin{aligned}
&f_{i,j,1}g_1 + f_{i,j,2}g_2 + \ldots f_{i,j,k}g_k \equiv 0 \bmod \varphi_i^{n_{i,j}} \\
&\quad\iff f_{i,j,1}\widehat{h}_1 + f_{i,j,2}\widehat{h}_2 + \cdots + f_{i,j,k}\widehat{h}_k \equiv 0 \bmod \varphi_i^{n_{i,j}-m_i} \\
&\quad\iff f_{i,j,r}\widehat{h}_r \equiv -(f_{i,j,1}\widehat{h}_1 + f_{i,j,2}\widehat{h}_2 + \cdots + f_{i,j,r-1}\widehat{h}_{r-1} \\
&\qquad\qquad + f_{i,j,r+1}\widehat{h}_{r+1} + f_{i,j,r+2}\widehat{h}_{r+2} + \cdots + f_{i,j,k}\widehat{h}_k) \bmod \varphi_i^{n_{i,j}-m_i} \\
&\quad\iff f_{i,j,r} \equiv -\widehat{s}(f_{i,j,1}\widehat{h}_1 + f_{i,j,2}\widehat{h}_2 + \cdots + f_{i,j,r-1}\widehat{h}_{r-1} \\
&\qquad\qquad + f_{i,j,r+1}\widehat{h}_{r+1} + f_{i,j,r+2}\widehat{h}_{r+2} + \cdots + f_{i,j,k}\widehat{h}_k) \bmod \varphi_i^{n_{i,j}-m_i}.
\end{aligned}$$

Thus $f_{i,j,1}g_1 + f_{i,j,2}g_2 + \cdots + f_{i,j,k}g_k \equiv 0 \bmod \varphi_i^{n_{i,j}}$ if and only if

$$f_{i,j,r} = \widehat{f}_{i,j,r}\varphi_i^{n_{i,j}-m_i} + \upsilon$$

where $\widehat{f}_{i,j,r}$ is a polynomial with degree less than $d_i m_i$ whose coefficients can be chosen freely and where $\upsilon \in \mathsf{F}_q[x]$ is the unique polynomial with degree less than $(n_{i,j} - m_i)d_i$ such that

$$\begin{aligned}
\upsilon \equiv -\widehat{s}(&f_{i,j,1}\widehat{h}_1 + f_{i,j,2}\widehat{h}_2 + \cdots + f_{i,j,r-1}\widehat{h}_{r-1} \\
&+ f_{i,j,r+1}\widehat{h}_{r+1} + f_{i,j,r+2}\widehat{h}_{r+2} + \cdots + f_{i,j,k}\widehat{h}_k) \bmod \varphi_i^{n_{i,j}-m_i}.
\end{aligned}$$

Since each polynomial $f_{i,j,m}$ must have degree less than $d_i n_{i,j}$ and the coefficients of these polynomials can be chosen freely, if $1 \leq m \leq k$ and $m \neq r$, there are $q^{(k-1)d_i n_{i,j}}$ choices of the polynomials

$$f_{i,j,1}, f_{i,j,2}, \ldots, f_{i,j,r-1}, f_{i,j,r+1}, f_{i,j,r+2}, \ldots, f_{i,j,k}$$

and, for each choice of the above polynomials there are $q^{d_i m_i}$ choices of $f_{i,j,r}$ (since the coefficients of the above polynomial $\widehat{f}_{i,j,r}$ can also be chosen freely). The number of choices of $f_{i,j,1}, f_{i,j,2}, \ldots, f_{i,j,k}$ is therefore $q^{(k-1)d_i n_{i,j} + d_i m_i}$ in this case.

Suppose, on the other hand, that j is an integer such that $1 \leq j \leq \ell_i$ and $n_{i,j} \leq m_i$; then

$$\begin{aligned} &f_{i,j,1}g_1 + f_{i,j,2}g_2 + \cdots + f_{i,j,k}g_k \\ &\quad = (f_{i,j,1}\widehat{h}_1 + f_{i,j,2}\widehat{h}_2 + \cdots + f_{i,j,k}\widehat{h}_k)\varphi_i^{m_i} \equiv 0 \bmod \varphi_i^{n_{i,j}} \end{aligned}$$

for every choice of $f_{i,j,1}, f_{i,j,2} + \cdots + f_{i,j,k}$. Since each of these polynomials must have degree less than $d_i n_{i,j}$ but can otherwise be chosen freely, there are $q^{kd_i n_{i,j}}$ choices of these polynomials. Now, since

$$kd_i n_{i,j} = (k-1)d_i n_{i,j} + d_i n_{i,j} \leq (k-1)d_i n_{i,j} + d_i m_i,$$

the number of choices of these polynomials is at most $q^{(k-1)d_i n_{i,j} + d_i m_i}$ in this case as well.

Since $n_i = d_i(n_{i,1} + n_{i,2} + \ldots n_{i,\ell_i})$ it follows that if $n_{i,1} > m_i$ then the number of choices of all of the polynomials $f_{i,j,m}$, such that $1 \leq j \leq \ell_i$ and $1 \leq m \leq k$, is at most

$$\prod_{j=1}^{\ell_i} q^{(k-1)d_i n_{i,j} + d_i m_i} = q^{(k-1)\sum_{j=1}^{\ell_i}(d_i n_{i,j}) + \sum_{j=1}^{\ell_i}(d_i m_i)} = q^{(k-1)n_i + d_i m_i \ell_i}.$$

Consider now the second case, that is, $m_i = n_{i,1}$ and $g_1, g_2, \ldots g_k$ are all divisible by $\varphi_i^{n_{i,1}}$. In this case

$$f_{i,j,1}g_1 + f_{i,j,2}g_2 + \ldots f_{i,j,k}g_k \equiv 0 \bmod \varphi_i^{n_{i,j}}$$

for $1 \leq j \leq \ell_i$ for all choices of polynomials $f_{i,j,m}$ such that $1 \leq j \leq \ell_i$ and $1 \leq m \leq k$, and there are q^{kn_i} choices of these polynomials. Now, since $n_{i,j} \leq n_{i,1} \leq m_i$ for $1 \leq j \leq \ell_i$,

$$n_i = d_i(n_{i,1} + n_{i,2} + \ldots n_{i,\ell_i}) \leq d_i m_i \ell_i,$$

so that $kn_i = (k-1)n_i + n_i \leq (k-1)n_i + d_i m_i \ell_i$, and the number of choices of these polynomials is at most $q^{(k-1)n_i + d_i m_i \ell_i}$ in this case as well.

Since $\ell_i \leq \ell$ for every integer i such that $0 \leq i \leq h$, it follows the number of choices of all polynomials $f_{i,j,m}$ such that $0 \leq i \leq h$, $1 \leq j \leq \ell_i$ and $1 \leq m \leq k$, and the given constraints are satisfied, is at most

$$\begin{aligned} \prod_{i=0}^{h} q^{(k-1)n_i + d_i m_i \ell_i} &= q^{(k-1)\sum_{i=0}^{h} n_i + \sum_{i=0}^{h} d_i m_i \ell_i} \\ &\leq q^{(k-1)\sum_{i=0}^{h} n_i + \sum_{i=0}^{h} d_i m_i \ell} \\ &= q^{(k-1)\sum_{i=0}^{h} n_i + \ell \sum_{i=0}^{h} d_i m_i} \\ &= q^{(k-1)n + \ell \deg(\xi)} \end{aligned}$$

as claimed. □

Lemma 4.2. *Let s be a positive integer, and let ξ be a a factor of the minimal polynomial $\psi_1 = \varphi_0^{n_{0,1}} \varphi_1^{n_{1,1}} \dots \varphi_h^{n_{h,1}}$ of A whose degree is greater than or equal to s/k. If $g_1, g_2, \dots, g_k \in \mathsf{F}_q[x]$ are polynomials such that g_j has degree less than $s_{k,j}$ for $1 \le j \le k$ then $g_1, g_2, \dots, g_k$ are all divisible by ξ if and only if*

$$g_1 = g_2 = \cdots = g_k = 0.$$

If the above equation is satisfied then there are q^{nk} choices of polynomials $f_{i,j,m} \in \mathsf{F}_q[x]$ such that $0 \le i \le h$, $1 \le j \le \ell_i$, $1 \le m \le k$, the degree of $f_{i,j,m}$ is less than $d_i n_{i,j}$ for all such i, j and m, and such that

$$f_{i,j,1} g_1 + f_{i,j,2} g_2 + \dots f_{i,j,k} g_k \equiv 0 \bmod \varphi_i^{n_{i,j}}$$

for all integers i and j such that $0 \le i \le h$ and $1 \le j \le \ell_i$.

Proof. Since the degree of ξ is an integer that is greater than or equal to s/k, the degree of ξ is greater than or equal to $s_{k,m}$ for $1 \le m \le k$. Now, since the degree of the polynomial ξ exceeds the degree of g_m (if $1 \le m \le k$ and the degree of g_m is less than $s_{k,m}$), g_m can only be divisible by ξ if $g_m = 0$. Thus $g_1, g_2, \dots, g_k$ are all divisible by ξ if and only if

$$g_1, g_2, \dots, g_k = 0,$$

as claimed.

If the above equation is satisfied then

$$f_{i,j,1} g_1 + f_{i,j,2} g_2 + \cdots + f_{i,j,k} g_k \equiv 0 \bmod \varphi_i^{n_{i,j}}$$

for all choices of polynomials $f_{i,j,m}$ such that $\deg(f_{i,j,m} < d_i n_{i,j}$ for $0 \le i \le h$, $1 \le j \le \ell_i$, and $1 \le m \le k$. The result now follows because there are q^{nk} choices of polynomials in $\mathsf{F}_q[x]$ satisfying these degree constraints. □

Theorem 4.3. *If A is as given in equations* (2.1a)–(2.1j) *and s and k are positive integers such that $k > \ell$, then*

$$\mathsf{Null}_{A,k,s} \le q^{nk} + \nu q^{(k-1)n+s}$$

where

$$\nu = \begin{cases} 6 \log_q n & \text{if } k = \ell + 1, \\ 4 & \text{if } k = \ell + 2, \\ (1 + 2q^{\ell - k + 1}) & \text{if } k \ge \ell + 3. \end{cases}$$

Proof. It follows once again by Lemma 3.3 that $\mathsf{Null}_{A,k,s}$ is the number of choices of polynomials $f_{i,j,m}$ for $0 \le i \le h$, $1 \le j \le \ell_i$ and $1 \le m \le k$, and of polynomials g_m for $1 \le m \le k$, such that $f_{i,j,m}$ has degree less than $d_i n_{i,j}$ and g_m has degree less than $s_{k,m}$ for all i, j and m as above, such that

$$f_{i,j,1} g_1 + f_{i,j,2} g_2 + \cdots + f_{i,j,k} g_k \equiv \varphi_i^{n_{i,j}}$$

for $0 \le i \le h$ and $1 \le j \le \ell_i$.

Once again, let $\psi_1 = \varphi_0^{n_{0,1}} \varphi_1^{n_{1,1}} \dots \varphi_h^{n_{h,1}}$ be the minimal polynomial of A. Then

$$\gcd(g_1, g_2, \dots, g_k, \psi_1) = \xi \tag{4.3}$$

for some nonzero divisor ξ of ψ_1 in $\mathsf{F}_q[x]$, for any choice of the polynomials $f_{i,j,m}$ and g_m (for i, j and m as above) satisfying these constraints.

Suppose first that the degree of ξ is less than s/k; then it follows by Lemma 4.1 that there are at most $q^{s-\deg(\xi)k}$ choices of $g_1, g_2, \dots, g_k$, such that g_m has degree less than $s_{k,m}$ for $1 \le m \le k$, satisfying the condition at line (4.3). It also follows by this lemma that there are at most $q^{(k-1)n+\ell\deg(\xi)}$ choices of $f_{i,j,m}$ (for all i, j, m) corresponding to any such choice of $g_1, g_2, \dots, g_k$ such that the above constraints are satisfied. Consequently there are at most

$$q^{s-\deg(\xi)k} \cdot q^{(k-1)n+\ell\deg(\xi)} = q^{(k-1)n+s+(\ell-k)\deg(\xi)}$$

choices of $f_{i,j,m}$ and g_m, for all i, j and m as above, satisfying the constraints described in Lemma 3.3 along with the condition at line (4.3).

On the other hand it follows by Lemma 4.2 that the condition at line (4.3) is only satisfied when $\deg(\xi) \ge s/k$ if

$$g_1 = g_2 = \dots = g_k = 0$$

(in which case $\xi = \psi$); there are then q^{nk} corresponding choices of the polynomials $f_{i,j,m}$ satisfying the constraints described in Lemma 3.3.

It follows that

$$\mathsf{Null}_{A,k,s} \le q^{nk} + \sum_{\substack{\xi \in \mathsf{F}_q[x] \setminus \{0\} \\ \xi \text{ divides } \psi \\ \deg(\xi) < s/k}} q^{(k-1)n+s+(\ell-k)\deg(\xi)}. \tag{4.4}$$

Since all terms shown in the above expression are nonnegative, a larger (but correct) bound can be obtained by discarding the condition that $\deg(\xi) < s/k$. Since $\psi = \varphi_0^{n_{0,1}} \varphi_1^{n_{1,1}} \dots \varphi_h^{n_{h,1}}$, ξ is a divisor of ψ in $\mathsf{F}_q[x]$ if and only if

$$\xi = \varphi_0^{m_0} \varphi_1^{m_1} \dots \varphi_h^{m_h}$$

for integers $m_0, m_1, \dots, m_h$ such that $0 \le m_i \le n_{i,1}$ for $0 \le i \le h$ — in which case $\deg(\xi) = \sum_{i=0}^{h} m_i d_i$. It now follows that

$$\begin{aligned} \mathsf{Null}_{A,k,s} &\le q^{nk} + \sum_{\substack{0 \le m_i \le n_{i,1} \\ \text{for } 0 \le i \le h}} q^{(k-1)n+s+(\ell-k)\sum_{i=0}^{h} d_i m_i} \\ &\le q^{nk} + \sum_{\substack{m_i \ge 0 \\ \text{for } 0 \le i \le h}} q^{(k-1)n+s+(\ell-k)\sum_{i=0}^{h} d_i m_i} \end{aligned}$$

$$= q^{nk} + q^{(k-1)n+s} \sum_{\substack{m_i \geq 0 \\ \text{for } 0 \leq i \leq h}} q^{(\ell-k)\sum_{i=0}^{h} d_i m_i}$$

$$= q^{nk} + q^{(k-1)n+s} \sum_{\substack{m_i \geq 0 \\ \text{for } 0 \leq i \leq h}} \prod_{i=0}^{h} q^{(\ell-k) d_i m_i}$$

$$= q^{nk} + q^{(k-1)n+s} \prod_{i=0}^{h} \sum_{m_i \geq 0} q^{(\ell-k) d_i m_i}$$

$$= q^{nk} + q^{(k-1)n+s} \prod_{i=0}^{h} (1 - q^{(\ell-k)d_i})^{-1}.$$

An upper bound for

$$\prod_{i=0}^{h} (1 - q^{(\ell-k)d_i})^{-1}$$

can therefore be used to obtain an upper bound for $\mathsf{Null}_{A,k,s}$. Upper bounds that are sufficient to establish the claim are derived in the rest of this argument.

Suppose first that $k = \ell + 1$. In this case, since $\varphi_0, \varphi_1, \ldots, \varphi_h$ are distinct irreducible polynomials in $\mathsf{F}_q[x]$ that each divide the minimal polynomial ψ of A, and ψ has degree at most n, one can apply Proposition 3 of Wiedemann [27] to conclude that

$$\prod_{i=0}^{h} (1 - q^{(\ell-k)d_i})^{-1} = \prod_{i=0}^{h} (1 - q^{-d_i})^{-1} \leq 6 \log_q n,$$

as needed to establish the bound claimed for this case.

Suppose next that $k = \ell + c$ for some constant $c \geq 2$. In this case, since there are at most q^d/d monic irreducible polynomials in $\mathsf{F}_q[x]$ with degree d for any positive integer d, one can group $\varphi_0, \varphi_1, \ldots, \varphi_h$ by degree to establish that

$$\begin{aligned}
\prod_{i=0}^{h} (1 - q^{(\ell-k)d_i}) &\geq 1 - \sum_{i=0}^{h} q^{(\ell-k)d_i} = 1 - \sum_{i=0}^{h} q^{-cd_i} \\
&\geq 1 - \sum_{d \geq 1} \frac{q^d}{d} q^{-cd} = 1 - \sum_{d \geq 1} \frac{q^{(1-c)d}}{d} \\
&= 1 - q^{1-c} - \sum_{d \geq 2} \frac{q^{(1-c)d}}{d} \\
&\geq 1 - q^{1-c} - \sum_{d \geq 2} \frac{q^{(1-c)d}}{2}
\end{aligned}$$

$$
\begin{aligned}
&= 1 - q^{1-c} - \frac{q^{2(1-c)}}{2} \sum_{d \geq 0} q^{(1-c)d} \\
&= 1 - q^{1-c} - \frac{q^{2(1-c)}}{2} \sum_{d \geq 0} 2^{-d} && \text{(since } q \geq 2 \text{ and } c \geq 2\text{)} \\
&= 1 - q^{1-c} - q^{2(1-c)}.
\end{aligned}
$$

If $c = 2$ then it follows that

$$
\begin{aligned}
\prod_{i=0}^{h} (1 - q^{(\ell-k)d_i}) &\geq 1 - q^{1-c} - q^{2(1-c)} \\
&\geq 1 - \tfrac{1}{2} - \tfrac{1}{4} && \text{(since } q \geq 2\text{).}
\end{aligned}
$$

Thus $\prod_{i=0}^{h} (1 - q^{(\ell-k)d_i})^{-1} \leq 4$ when $k = \ell + 2$, as needed to establish the bound claimed for this case.

Suppose, finally, that $c \geq 3$. Then

$$
\begin{aligned}
(1 + 2q^{1-c})(1 - q^{1-c} - q^{2(1-c)}) &= 1 + q^{1-c} - 3q^{2(1-c)} - 2q^{3(1-c)} \\
&= 1 + q^{1-c}(1 - 3q^{1-c} - 2q^{2(1-c)}) \\
&\geq 1 + q^{1-c}(1 - 3q^{-2} - 2q^{-4}) && \text{(since } c \geq 3\text{)} \\
&\geq 1 + q^{1-c}(1 - \tfrac{3}{4} - \tfrac{1}{8}) && \text{(since } q \geq 2\text{)} \\
&> 1.
\end{aligned}
$$

It follows that $(1 - q^{1-c} - q^{2(1-c)})^{-1} \leq 1 + 2q^{1-c}$, as needed to establish the claimed bound in this final case. □

The following can be obtained using the bounds for $\mathsf{Null}_{A,k,s}$ in the above theorem, together with the inequality shown at line (3.6) on page 15, and the observation that if $s > n$ then $q^{nk} < q^{(k-1)n+s}$.

Corollary 4.4. *Suppose again that A is as given in equations* (2.1a)–(2.1j) *and that s and k are positive integers such that $k > \ell$. Then*

$$\sigma_{A,k,s}(j) \leq \tfrac{\nu}{q-1}\, q^{\min(n,s)-n-j+1}$$

for ν as defined in Theorem 4.3, above, and for $1 \leq j \leq \min(n, s)$.

4.2 A Bound for A-$\mathsf{Null}_{A,k,s}$

Theorem 4.5. *If A is as given in equations* (2.1a)–(2.1j) *and s and k are integers such that $k > \ell$, then*

$$A\text{-}\mathsf{Null}_{A,k,s} \leq q^{nk} + \nu q^{(k-1)n+s+\ell_0}$$

for

$$\nu = \begin{cases} 6\log_q n & \text{if } k = \ell + 1, \\ 4 & \text{if } k = \ell + 2, \\ (1 + 2q^{\ell - k + 1}) & \text{if } k \geq \ell + 3. \end{cases}$$

Proof. This follows from an application of the inequality at line (3.15) on page 21, along with Theorem 4.3, above. □

Since $n = r + \ell_0$, the above bounds can be applied along with the inequality at line (3.10) on page 19 to establish the following.

Corollary 4.6. *If A is as given in equations* (2.1a)–(2.1j) *and s, j and k are integers such that $k > \ell$, then*

$$\widehat{\sigma}_{A,k,s}(j) \leq \tfrac{\nu}{q-1}\, q^{\min(r,s)-r+1-j}$$

for ν as defined in Theorem 4.5, above, and for $1 \leq j \leq \min(r, s)$.

4.3 Bounds for $\mathsf{Null}_{A,k_L,s_L,k_R,s_L}$ and A-$\mathsf{Null}_{A,k_L,s_L,k_R,s_R}$

Theorem 4.7. *If A is as given in equations* (2.1a)–(2.1j) *and s_L, k_L, s_R and k_R are positive integers such that $\min(k_L, k_R) > \ell$, then*

$$\mathsf{Null}_{A,k_L,s_L,k_R,s_R} \leq q^{n(k_L+k_R)}(1 + \nu q^{s_R - s_L} + \nu^2 q^{s_R - n}),$$

where

$$\nu = \begin{cases} 6\log_q n & \text{if } \min(k_L, k_R) = \ell + 1, \\ 4 & \text{if } \min(k_L, k_R) = \ell + 2, \\ 1 + 2q^{\ell - \min(k_L, k_R) + 1} & \text{if } \min(k_L, k_R) \geq \ell + 3. \end{cases}$$

Proof. It follows by Theorem 4.3 that

$$\mathsf{Null}_{A,k_L,s_L} \leq q^{nk_L} + \nu q^{(k_L-1)n+s_L} \quad \text{and} \quad \mathsf{Null}_{A,k_R,s_R} \leq q^{nk_R} + \nu q^{(k_R-1)n+s_R}$$

for all positive integers s_L and s_R, and for ν as above. The claim now follows by an application of Theorem 3.24. □

The bounds in the above theorem can used along with the inequalities at lines (3.18) and (3.19) on page 25 to establish that if $\min(k_L, k_R) > \ell$ then

$$\sigma_{A,k_L,s_L,k_R,s_R}(j) \leq \tfrac{1}{q-1}\,(\nu q^{\min(n,s_L,s_R)-s_L} + \nu^2 q^{\min(n,s_L,s_R)-n})q^{1-j} \tag{4.5}$$

for $0 \leq j \leq \min(n, s_L, s_R)$, and for ν as above.

Now, in order to consider A-$\mathsf{Null}_{A,k_L,s_L,k_R,s_R}$, let $\widehat{\ell}_0 = 0$ if $n_{0,1} = 1$, and let $\widehat{\ell}_0$ be the largest integer such that $1 \leq \widehat{\ell}_0 \leq \ell_0$ and $n_{0,\widehat{\ell}_0} \geq 2$. otherwise — so that

$$n_{0,1} \geq n_{0,2} \geq \cdots \geq n_{0,\widehat{\ell}_0} \geq 2 \qquad \text{and} \qquad n_{0,\widehat{\ell}_0+1} = n_{0,\widehat{\ell}_0+2} = \cdots = n_{0,\ell_0} = 1.$$

Set $\widehat{\ell}_i = \ell_i$ for $1 \le i \le h$ and suppose that A, and the matrices $\widehat{A} \in \mathsf{F}_q^{\widehat{N}\times\widehat{N}}$ and $\widetilde{A} \in \mathsf{F}_q^{\widetilde{N}\times\widetilde{N}}$ are as described in Subsection 2.3. Then $\widehat{N} = n - \ell_0 + \widehat{\ell}_0$, and $\widetilde{A}$ is the zero matrix in $\mathsf{F}_q^{\widetilde{N}\times\widetilde{N}}$ for $\widetilde{N} = \ell_0 - \widehat{\ell}_0$, so that A and $\widehat{A}$ both have rank $r = \widehat{N} - \widehat{\ell}_0$. Furthermore, if

$$\widehat{\ell} = \max(\widehat{\ell}_0, \widehat{\ell}_1, \ldots, \widehat{\ell}_h)$$

then $\widehat{\ell}$ is the number of nontrivial invariant factors (that is, the number of invariant factors different from x) of A.

The following bounds can be established by an application of Theorems 3.24 and 4.3.

Theorem 4.8. *If A is as given in equations* (2.1a)–(2.1j) *and s_L, k_L, s_R and k_R are positive integers such that* $\min(k_L, k_R) > \widehat{\ell}$*, Then*

$$A\text{-}\mathsf{Null}_{A,k_L,s_L,k_R,s_R} \le q^{n(k_L+k_R)}(1 + \nu q^{s_R-s_L} + \nu^2 q^{s_R-r})$$

where

$$\nu = \begin{cases} 6\log_q n & \text{if } \min(k_L,k_R) = \widehat{\ell}+1, \\ 4 & \text{if } \min(k_L,k_R) = \widehat{\ell}+2, \\ 1 + 2q^{\widehat{\ell}-\min(k_L,k_R)+1} & \text{if } \min(k_L,k_R) \ge \widehat{\ell}+3. \end{cases}$$

Proof. Consider the matrix $\widehat{A} \in \mathsf{F}_q^{\widehat{N}\times\widehat{N}}$ described above. It follows by Theorem 4.3 that if k_L, s_L, k_R and s_R are positive integers such that $\min(k_L, k_R) > \widehat{\ell}$, then

$$\mathsf{Null}_{\widehat{A},k_L,s_L} \le q^{\widehat{N}k_L} + \nu q^{(k_L-1)\widehat{N}+s_L} \text{ and } \mathsf{Null}_{\widehat{A},k_R,s_R} \le q^{\widehat{N}k_R} + \nu q^{(k_R-1)\widehat{N}+s_R}$$

for ν as given above. Theorem 4.5 can be applied to establish that

$$A\text{-}\mathsf{Null}_{\widehat{A},k_L,s_L} \le q^{\widehat{N}k_L} + \widehat{\nu} q^{(k_L-1)\widehat{N}+s_L}$$

for $\widehat{\nu} = \nu q^{\widehat{\ell}_0}$.

Theorem 3.24 can now be applied to establish that

$$\begin{aligned} A\text{-}\mathsf{Null}_{\widehat{A},k_L,s_L,k_R,s_R} &\le q^{\widehat{N}(k_L+k_R)}(1 + \nu q^{s_R-s_L} + \nu \cdot \widehat{\nu} q^{s_R-\widehat{N}}) \\ &= q^{\widehat{N}(k_L+k_R)}(1 + \nu q^{s_R-s_L} + \nu^2 q^{s_R-\widehat{N}+\widehat{\ell}_0}) \\ &= q^{\widehat{N}(k_L+k_R)}(1 + \nu q^{s_R-s_L} + \nu^2 q^{s_R-r}). \end{aligned}$$

Finally, since $n = \widehat{N} + \widetilde{N}$, an application of Theorem 3.27 establishes that

$$\begin{aligned} A\text{-}\mathsf{Null}_{A,k_L,s_L,k_R,s_R} &\le q^{(k_L+k_R)\widetilde{N}} A\text{-}\mathsf{Null}_{\widehat{A},k_L,s_L,k_R,s_R} \\ &\le q^{n(k_L+k_R)}(1 + \nu q^{s_R-s_L} + \nu^2 q^{s_R-r}) \end{aligned}$$

as claimed. □

The bounds in the above theorem can used along with the inequalities at lines (3.18) and (3.19) on page 25 to establish that if $\min(k_L, k_R) > \widehat{\ell}$ then

$$\widehat{\sigma}_{A,k_L,s_L,k_R,s_R}(j) \leq \tfrac{1}{q-1}\,(\nu q^{\min(r,s_L,s_R)-s_L} + \nu^2 q^{\min(r,s_L,s_R)-r})q^{1-j} \tag{4.6}$$

for $0 \leq j \leq \min(r, s_L, s_R)$, and for ν as above.

Finally, bounds for $\mathsf{Null}_{A,k_L,s_L,k_R,s_R}$ can also be given when $\widehat{\ell} < \min(k_L, k_R) \leq \ell$, using a different application of Theorems 4.3 and 3.24. In this case, one should set $\widehat{\ell}_0 = \min(k_L, k_R) - 1$, $\widehat{\ell}_i = \ell_i$ for $1 \leq i \leq h$, and consider the corresponding integers $\widehat{N}$ and $\widetilde{N}$ and matrices $\widehat{A} \in \mathsf{F}_q^{\widehat{N}\times\widehat{N}}$ and $\widetilde{A} \in \mathsf{F}_q^{\widetilde{N}\times\widetilde{N}}$ as described in Subsection 2.3. Now

$$\widehat{N} = r + \min(k_L, k_R) - 1 = n - \ell_0 + \min(k_L, k_R) - 1$$

and

$$\widetilde{N} = n - \widehat{N} = \ell_0 - \min(k_L, k_R) + 1;$$

furthermore, the matrix $\widehat{A}$ has exactly $\min(k_L, k_R) - 1$ invariant factors. Arguing as above, with this choice of matrices, one can establish the following.

Theorem 4.9. *If A is as given in equations* (2.1a)–(2.1j) *and s_L, k_L, s_R and k_R are positive integers such that $\widehat{\ell} < \min(k_L, k_R) \leq \ell$, then*

$$\mathsf{Null}_{A,k_L,s_L,k_R,s_R} \leq q^{n(k_L+k_R)}(1 + \nu q^{s_R-s_L} + \nu^2 q^{s_R-r-\min(k_L,k_R)+1}),$$

where $\nu = 6\log_q n$.

Once again, it is appropriate to adjust the definition of "rank deficiency:" A consideration of the Frobenius normal form of A is sufficient to establish that if if $k_L \leq \ell$ then the rank of $\mathcal{K}_{A^T,\vec{u},s_L}$ cannot exceed $\min(r + k_L, s_L)$ for any choice of vectors $\vec{u} = u_1, u_2, \ldots, u_{k_L} \in \mathsf{F}_q^{n\times 1}$. Similarly, the rank of $\mathcal{K}_{A,\vec{v},s_R}$ cannot exceed $\min(r + k_R, s_R)$ for any choice of vectors $\vec{v} = v_1, v_2, \ldots, v_{k_R} \in \mathsf{F}_q^{n\times 1}$. Thus the rank of $\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R}$ cannot exceed $\min(r + \min(k_L, k_R), s_L, s_R)$.

With that noted, let $\sigma'_{A,k_L,s_L,k_R,s_R}(j)$ be the probability that $\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R}$ is "rank deficient by at least j" when $\widehat{\ell} < \min(k_L, k_R) \leq \ell$ and the vectors $\vec{u} = u_1, u_2, \ldots, u_{k_L}$ and $\vec{v} = v_1, v_2, \ldots, v_{k_R}$ are chosen uniformly and independently from $\mathsf{F}_q^{n\times 1}$, that is,

$$\sigma'_{A,k_L,s_L,k_R,s_R}(j) = \mathsf{Pr}[\mathsf{rank}(\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R}) \leq \min(r + \min(k_L, k_R), s_L, s_R) - j].$$

Then

$$\sigma'_{A,k_L,s_L,k_R,s_R}(j) = \sigma_{A,k_L,s_L,k_R,s_R}(j + \min(n, s_L, s_R) - \min(r + \min(k_L, k_R), s_L, s_R)).$$

The bounds in the above theorem can now used along with the inequalities at lines (3.18) and (3.19) on page 25 to establish that if $\widehat{\ell} < k = \min(k_L, k_R) \leq \ell$ then

$$\begin{aligned}&\sigma'_{A,k_L,s_L,k_R,s_R}(j)\\ &\quad\leq \tfrac{1}{q-1}\,(\nu q^{\min(r+\min(k_L,k_R),s_L,s_R)-s_L} + \nu^2 q^{\min(r+\min(k_L,k_R),s_L,s_R)-r-\min(k_L,k_R)})q^{2-j}\end{aligned} \tag{4.7}$$

for $0 \leq j \leq \min(r + \min(k_L, k_R), s_L, s_R)$, and for ν as above.

5 A Sparse Matrix Preconditioner

The goal of this section is to describe a sparse matrix preconditioner that produces a matrix with a small number of nontrivial invariant factors with high probability.

5.1 Definition and Basic Properties

It will be assumed in this section that $A \in \mathsf{F}_q^{n\times m}$ for positive integers n and m. Let r be the (unknown) rank of A, let

$$N = \min(n,m) + \lceil 2\log_q \max(n,m)\rceil \tag{5.1}$$

and let

$$c = \begin{cases} 3 & \text{if } q = 2, \\ \lceil 3\ln q\rceil & \text{otherwise.} \end{cases} \tag{5.2}$$

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

- If $1 \le i \le \min(n,m)$ then each entry in row i of L or column i of R is set to be zero with probability
$$\max\left(1 - \frac{\lceil c\log_q \max(n,m)\rceil}{i}, \frac{1}{q}\right).$$
- If $1 \le i \le \min(n,m)$ 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 $\mathsf{F}_q \setminus \{0\}$.
- Finally, if $\min(n,m) < i \le N$ then each entry of row i of L or column i of R is chosen uniformly and independently from F_q.

Notice that if $1 \le i \le 2\lceil \log_q \max(n,m)\rceil$ then $i \le \lceil c\log_q \max(n,m)\rceil$, as well, so that

$$1 - \frac{\lceil c\log_q \max(n,m)\rceil}{i} \le 0 \le \frac{1}{q}.$$

Thus the entries in the top $\lceil 2\log_q \max(n,m)\rceil$ rows of the matrix L and the left-most $\lceil 2\log_q \max(n,m)\rceil$ columns of the matrix R are chosen uniformly and independently from F_q if L and R are randomly chosen as described above. Indeed, the entries of the top $q\lceil c\log_q \max(n,m)\rceil/(q-1)$ rows of L and columns of R are chosen uniformly and independently from F_q.

As noted in the introduction, this is a minor variant of a sparse matrix preconditioner introduced by Wiedemann [27] — and the following is a minor modification of Wiedemann's argument that such a matrix is sparse.

Lemma 5.1. *If $L \in \mathsf{F}_q^{N\times n}$ and $R \in \mathsf{F}_q^{m\times N}$ then the expected number of nonzero entries in L is in $\Theta(n \ln max(n,m) \ln \min(n,m))$ and the expected number of nonzero entries in R is in $\Theta(m \ln \max(n,m) \ln \min(n,m))$.*

Proof. Let i be an integer such that either

$$1 \le i \le q\lceil c \log_q \max(n,m)\rceil/(q-1) \text{ or } \min(n,m)+1 \le i \le N.$$

Then the entries of row i of L and of column i are chosen uniformly and independently from F_q, so that the expected number of nonzero entries in row i of L is $(q-1)n/q$ and the expected number of nonzero entries in column i of R is $(q-1)m/q$. It follows by linearity of expectation that the expected number of nonzero entries of all such rows of L is

$$\begin{aligned}\left(\frac{q\lceil c\log_q \max(n,m)\rceil}{q-1} + N - \min(n,m)\right) &\cdot \frac{(q-1)n}{q} \\ &\le \frac{2q\lceil c\log_q \max(n,m)\rceil}{q-1} \cdot \frac{(q-1)n}{q} \\ &= 2\lceil c\log_q \max(n,m)\rceil n \\ &\in \Theta(n \ln \max(n,m),\end{aligned}$$

since $c \in \Theta(\ln q)$. Similarly, the expected number of nonzero entries in all such columns of R is in $\Theta(m \ln \max(n,m))$.

Now let i be an integer such that $q\lceil c\log_q \max(n,m)\rceil/(q-1) < i \le \min(n,m)$. Then each entry of row i of L or column i of R is nonzero with probability $\lceil c\log_q \max(n,m)\rceil/i$, instead. It now follows by linearity of expectation that the expected number of nonzero entries in all such rows of L is

$$\begin{aligned}\sum_{i=q\lceil c\log_q \max(n,m)\rceil/(q-1)}^{\min(n,m)} &\frac{\lceil c\log_q \max(n,m)\rceil}{i} n \\ &= \lceil c\log_q \max(n,m)\rceil n \sum_{i=q\lceil c\log_q \max(n,m)\rceil/(q-1)}^{\min(n,m)} (1/i) \\ &= \lceil c\log_q \max(n,m)\rceil n (H_{\min(n,m)} - H_{\lceil c\log_q \max(n,m)\rceil/(q-1)}) \\ &\le \lceil c\log_q \max(n,m)\rceil n \ln \min(n,m) \\ &\in \Theta(n \ln \max(n,m) \ln \min(n,m)),\end{aligned}$$

once again since $c \in \Theta(\ln q)$. In the above derivation H_k is the k^{th} Harmonic number, that is, $H_k = \sum_{i=1}^{k}(1/i)$. Similarly, the expected number of nonzero entries in these columns of R is in $\Theta(m \ln \max(n,m) \ln \min(n,m))$. The result now follows, since all rows of L and all columns of R have been considered, and since $\ln \max(n,m) \in O(\ln \max(n,m) \ln \min(n,m))$. □

The next few lemmas we be used repeatedly in the analysis of these matrices.

Lemma 5.2. *Let i be a positive integer such that $\lceil c \log_q \max(n,m) \rceil \leq i$. Then*

$$\left(1 - \frac{\lceil c \log_q \max(n,m) \rceil}{i}\right)^i \leq \max(n,m)^{-3}.$$

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

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

— see, for example, Section 3.2 of the text of Cormen, Leiserson, Rivest and Stein [6]. Since i is a positive integer such that $\lceil c \log_q \max(n,m) \rceil \leq i$, this implies that

$$\left(1 - \frac{\lceil c \log_q \max(n,m) \rceil}{i}\right) \leq \left(1 - \frac{c \log_q \max(n,m)}{i}\right) \leq e^{-c \log_q \max(n,m)/i}.$$

Therefore

$$\begin{aligned}
\left(1 - \frac{\lceil c \log_q \max(n,m) \rceil}{i}\right)^i &\leq e^{-c \log_q \max(n,m)} \\
&= e^{-c \ln \max(n,m)/\ln q} \\
&\leq e^{-3 \ln \max(n,m)} && \text{(since } c \geq 3 \ln q\text{)} \\
&= \max(n,m)^{-3}. && \square
\end{aligned}$$

Once again, let r be the rank of A. It will be useful to consider two other pairs of matrices that are chosen using a distribution similar to the above. Suppose, in particular, that $L_0, R_0 \in \mathsf{F}_q^{r \times r}$ are randomly chosen as follows.

- If $1 \leq i \leq r$ then each entry of row i or L_0 or column i of R_0 is set to be zero with probability
$$\max\left(1 - \frac{\lceil \log_q \max(n,m) \rceil}{i}, \frac{1}{q}\right).$$
- 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 $\mathsf{F}_q \setminus \{0\}$.

Once again, this implies that the entries of the top $\lceil 2 \log_q \max(n,m) \rceil$ rows of L_0 and the leftmost $\lceil 2 \log_q \max(n,m) \rceil$ columns of R_0 are chosen uniformly and independently from F_q.

Lemma 5.3. *Let $\widetilde{A} \in \mathsf{F}_q^{r \times r}$ be a nonsingular matrix and let $B \in \mathsf{F}_q^{r \times r}$ as well. Suppose that matrix $L_0 \in \mathsf{F}_q^{r \times r}$ is randomly chosen as described above. Let*

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

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

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

$$q^{-i} + \max(n,m)^{-3}.$$

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

$$\frac{q^{1-i}}{q-1} + \max(n,m)^{-2}.$$

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

$$\frac{q^{1-i}}{q-1} + \max(n,m)^{-2}.$$

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

$$\frac{q}{(q-1)^2} + \max(n,m)^{-1}.$$

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 $i+1, i+2, \ldots, r$. Then $s_i \leq r-i$, since there are only $r-i$ rows in this matrix, 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 $i+1, i+2, \ldots, r$.

Suppose that $1 - \lceil c \log_q \max(n,m) \rceil / i \geq 1/q$. Then $\lceil c \log_q \max(n,m) \rceil \leq i$, so that Lemma 5.2 is applicable, and 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

$$\begin{aligned}
\left(1 - \frac{\lceil c \log_q \max(n,m) \rceil}{i}\right)^{r-s_i} &\leq \left(1 - \frac{\lceil c \log_q \max(n,m) \rceil}{i}\right)^{i} && \text{(since } r - s_i \geq i\text{)} \\
&\leq \max(n,m)^{-3} && \text{(by Lemma 5.2).}
\end{aligned}$$

On the other hand, if $1 - \lceil c \log_q \max(n,m) \rceil / i < 1/q$ then the entries of row i of L_0 are chosen uniformly and independently from F_q. 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 by adding together the upper bounds for these cases, one can 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 $q^{-i} + \max(n,m)^{-3}$, as required to establish part (a) of the claim when $\widetilde{A}$ is the identity matrix.

Now let $\widetilde{A}$ be an arbitrary nonsingular matrix in $\mathsf{F}_q^{r \times r}$. Then $C = L_0\widetilde{A} + B = C' \cdot \widetilde{A}$ where $C' = L_0 + B'$ and $B' = B \cdot \widetilde{A}^{-1}$. 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 $q^{-i} + \max(n,m)^{-3}$. 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

$$\begin{aligned}
\sum_{j=i}^{r} \left(q^{-j} + \max(n,m)^{-3}\right) &= \sum_{j=i}^{r} q^{-j} + \sum_{j=i}^{r} \max(n,m)^{-3} \\
&= \sum_{j=i}^{r} q^{-j} + (r-i+1) \cdot \max(n,m)^{-3} \\
&\leq \sum_{j=i}^{r} q^{-j} + \max(n,m)^{-2} \qquad \text{(since } r \leq \max(n,m)\text{)} \\
&\leq \sum_{j \geq i} q^{-j} + \max(n,m)^{-2} \\
&= \frac{q^{1-i}}{q-1} + \max(n,m)^{-2},
\end{aligned}$$

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

$$\begin{aligned}
\sum_{i=1}^{r} \left(\mathsf{Pr}[\mathsf{nullity}(C) \geq i]\right) &\leq \sum_{i=1}^{r} \left(\frac{q^{1-i}}{q-1} + \max(n,m)^{-2}\right) \qquad \text{(by part (c))} \\
&= \sum_{i=1}^{r} \frac{q^{1-i}}{q-1} + \sum_{i=1}^{r} \max(n,m)^{-2}
\end{aligned}$$

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

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 5.4. *Let $\widetilde{A} \in \mathsf{F}_q^{r \times r}$ be a nonsingular matrix and let $B \in \mathsf{F}_q^{r \times r}$ as well. Suppose that the matrix $R_0 \in \mathsf{F}_q^{r \times r}$ is randomly chosen as described on page 47, above. Let*

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

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*

$$q^{-i} + \max(n,m)^{-3}.$$

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

$$\frac{q^{1-i}}{q-1} + \max(n,m)^{-2}.$$

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

$$\frac{q^{1-i}}{q-1} + \max(n,m)^{-2}.$$

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

$$\frac{q}{(q-1)^2} + \max(n,m)^{-1}.$$

Lastly, consider matrices $L_1 \in \mathsf{F}_q^{(r+\lceil 2 \log_q \max(n,m) \rceil) \times r}$ and $R_1 \in \mathsf{F}_q^{r \times (r+\lceil 2 \log_q \max(n,m) \rceil)}$ 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
$$\max\left(1 - \frac{\lceil c \log_q \max(n,m)\rceil}{i}, \frac{1}{q}\right).$$
- 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 $\mathsf{F}_q \setminus \{0\}$.
- Finally, if $r < i \leq r + \lceil 2\log_q \max(n,m)\rceil$ then each entry in row i of L_1 or column i of R_1 is chosen uniformly and independently from F_q.

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 47). Thus the entries in the top $\lceil 2\log_q \max(n,m)\rceil$ rows of L_1 and the leftmost $\lceil 2\log_q \max(n,m)\rceil$ columns of R_1 are chosen uniformly and independently from F_q.

Lemma 5.5. *Let $\widetilde{A} \in \mathsf{F}_q^{r\times r}$ be a nonsingular matrix. Let $B \in \mathsf{F}_q^{(r+\lceil 2\log_q \max(n,m)\rceil)\times r}$, and suppose L_1 is chosen as described above. Let*
$$C = L_1\widetilde{A} + B \in \mathsf{F}_q^{(r+\lceil 2\log_q \max(n,m)\rceil)\times r}.$$
Then the probability that rows $\lceil 2\log_q \max(n,m)\rceil + 1, \lceil 2\log_q \max(n,m)\rceil + 2, \ldots, r$ of C are linearly dependent or C has rank less than r (or both) is at most $3\max(n,m)^{-2}$.

Proof. To begin let us suppose that $\widetilde{A} = I_r$, the $r \times r$ identity matrix. Let $L_0, B_0 \in \mathsf{F}_q^{r\times r}$ 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
$$C_0 = L_0 + B_0 \in \mathsf{F}_q^{r\times r}.$$

Since L_0 is selected using the distribution described on page 47, it follows by part (b) of Lemma 5.3 that rows $\lceil 2\log_1 \max(n,m)\rceil + 1, \lceil 2\log_q \max(n,m)\rceil + 2, \ldots, r$ of C_0 are linearly dependent with probability at most
$$\frac{q^{-\lceil 2\log_q \max(n,m)\rceil}}{q-1} + \max(n,m)^{-2} \leq 2\max(n,m)^{-2}. \tag{5.3}$$

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

Recall that the entries of the matrix L_1 in rows $1, 2, \ldots, \lceil 2\log_q \max(n,m)\rceil$, and in rows $r+1, r+2, \ldots, r+\lceil 2\log_q \max(n,m)\rceil$, are chosen uniformly and independently from F_q. It follows that the entries of C in these rows are chosen uniformly and independently from F_q, as well.

Suppose that rows $\lceil 2\log_q \max(n,m)\rceil + 1, \lceil 2\log_q \max(n,m)\rceil + 2, \ldots, r$ of C are linearly independent. Then there exists a set S of $r - \lceil 2\log_q \max(n,m)\rceil$ columns of C such that

the submatrix of C including rows $\lceil 2\log_q \max(n,m)\rceil + 1, \lceil 2\log_q \max(n,m)\rceil + 2, \ldots, r$ and the columns in set S is nonsingular.

In order the complete the analysis, in the case that $\widetilde{A} = I_r$, consider an assignment of values for the entries of L_1 in rows $\lceil 2\log_q \max(n,m)\rceil + 1, \lceil 2\log_q \max(n,m)\rceil + 2, \ldots, r$ and the columns in S. Since the remaining entries are uniformly and independently selected from F_q, the probability that the columns of the matrix $C \in \mathsf{F}_q^{(r+\lceil 2\log_q \max(n,m)\rceil)\times r}$ are *not* linearly independent is at most

$$\sum_{i=\lceil 2\log_q \max(n,m)\rceil+1}^{2\lceil 2\log_q \max(n,m)\rceil} q^{-i} \leq \sum_{i\geq \lceil 2\log_q \max(n,m)\rceil+1} q^{-i} = \frac{q^{-\lceil 2\log_q \max(n,m)\rceil}}{q-1} \leq \max(n,m)^{-2}. \qquad (5.4)$$

Thus the probability that C has rank less than r, when $\overline{A} = I_r$, is at most the sum of the bounds at lines (5.3) and (5.4), that is, $3\max(n,m)^{-2}$, as claimed.

Suppose, next, that $\widetilde{A}$ is an arbitrary nonsingular matrix in $\mathsf{F}_q^{r\times r}$. 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 $3\max(n,m)^{-2}$. 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 $3\max(n,m)^{-2}$ 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 5.6. *Let $\widetilde{A} \in \mathsf{F}_q^{r\times r}$ be a nonsingular matrix. Let $B \in \mathsf{F}_q^{r\times(r+\lceil 2\log_q \max(n,m)\rceil)}$, and suppose R_1 is chosen as described on page 51. Let*

$$D = \widetilde{A}R_1 + B \in \mathsf{F}_q^{r\times(r+\lceil 2\log_q \max(n,m)\rceil)}.$$

Then the probability that columns $\lceil 2\log_q \max(n,m)\rceil + 1, \lceil 2\log_q \max(n,m)\rceil + 2, \ldots, r$ of D are linearly dependent or D has rank less than r is at most $3\max(n,m)^{-2}$.

5.2 Preservation of Rank

Suppose again that $A \in \mathsf{F}_q^{n\times m}$ for positive integers n and m and that matrices $L \in \mathsf{F}_q^{\mathbb{N}\times n}$ and $R \in \mathsf{F}_q^{m\times N}$ are randomly selected as described in Subsection 5.1 for

$$N = \min(n,m) + \lceil 2\log_q \max(n,m)\rceil.$$

The rest of this section will concern properties of the "conditioned" matrix

$$\widehat{A} = LAR \in \mathsf{F}_q^{N\times N}. \tag{5.5}$$

It should be noted that properties similar to the ones established in this and the next subsection have been proved for a similar matrix preconditioner in Chen et. al. [3]. Indeed, the arguments given here are based on the proofs found in Section 7 of that report.

If $\widehat{A}$ is as above then the rank of $\widehat{A}$ is certainly 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 5.7. *Let $A \in \mathsf{F}_q^{n\times m}$ be a matrix with rank r. Suppose the matrices $L \in \mathsf{F}_q^{N\times n}$ and $R \in \mathsf{F}_q^{m\times N}$ are randomly chosen as described at the beginning of Section 5.1, and that $\widehat{A} = LAR$.*

(a) *The probability that either rows*

$$\lceil 2\log_q \max(n,m)\rceil + 1, \lceil 2\log_q \max(n,m)\rceil + 2, \ldots, r$$

of the matrix LA are linearly dependent, or that the $(r + \lceil 2\log_q \max(n,m)\rceil) \times m$ submatrix of LA that includes rows with indices

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

has rank less than r, is at most $3\max(n,m)^{-2}$.

(b) *The probability that $\widehat{A}$ has rank less than r is at most $6\max(n,m)^{-2}$.*

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 permutation matrices $P \in \mathsf{F}_q^{n\times n}$ and $Q \in \mathsf{F}_q^{m\times m}$ 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

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

Rows $\lceil 2\log_q \max(n,m)\rceil + 1, \lceil 2\log_q \max(n,m)\rceil + 2, \ldots, r$ of the matrix LA are linearly independent if and only if the corresponding rows of $(L \cdot P^{-1}) \cdot PAQ = LAQ$ are. Furthermore, the $(r + \lceil 2\log_q \max(n,m)\rceil) \times m$ submatrix of LA that includes the rows $1, 2, \ldots, r$ and rows $\min(n,m)+1, \min(n,m)+2, \ldots, N$ has rank less than r if and only if the $(r + \lceil 2\log_q \max(n,m)\rceil) \times r$ submatrix of LAQ including rows with these indices has rank less than r, as well. Finally, the ranks of the matrices LA and LAQ are also the same. The matrix A can therefore be replaced with PAQ in the rest of this proof, in order to assume without loss of generality that the leading $r \times r$ submatrix of A is nonsingular.

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 $\mathsf{F}_q^{r \times r}$, and where $A_{1,2} \in \mathsf{F}_q^{r \times (m-r)}$, $A_{2,1} \in \mathsf{F}_q^{(n-r) \times r}$, and $A_{2,2} \in \mathsf{F}_q^{(n-r) \times (m-r)}$.

Consider the $(r + \lceil 2 \log_q \max(n,m) \rceil) \times m$ submatrix of LA that includes rows with indices

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

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 $L_1 \in \mathsf{F}_q^{(r+\lceil 2 \log_q \max(n,m) \rceil) \times r}$, $L_2 \in \mathsf{F}_q^{(r+\lceil 2 \log_q \max(n,m) \rceil) \times (n-r)}$, and where L_1 is chosen using the probability distribution described on page 51. The submatrix consisting of the leftmost r columns of this matrix is

$$\begin{bmatrix} L_1 & L_2 \end{bmatrix} \cdot \begin{bmatrix} A_{1,1} \\ A_{2,1} \end{bmatrix} = L_1 \cdot A_{1,1} + L_2 \cdot A_{2,1} \in \mathsf{F}_q^{(r+\lceil 2 \log_q \max(n,m) \rceil) \times r}.$$

If we choose (and fix) the values in the matrix the above matrix L_2 then, since $A_{1,1} \in \mathsf{F}_q^{r \times r}$ is nonsingular and $L_2 \cdot A_{2,1} \in \mathsf{F}_q^{(r+\lceil 2 \log_q \max(n,m) \rceil) \times r}$, it follows by Lemma 5.5 that (when the entries of L_1 are selected) the likelihood that rows

$$\lceil 2 \log_q \max(n,m) \rceil + 1, \lceil 2 \log_q \max(n,m) \rceil + 2, \ldots, r$$

of this matrix are linearly dependent or that this matrix has rank less than r is at most $3 \max(n,m)^{-2}$. 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 the corresponding rows 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 $L \in \mathsf{F}_q^{N \times n}$ such that the matrix

$$LA \in \mathsf{F}_q^{N \times m}$$

has rank r. Repeating the above argument, using Corollary 5.6 in place of Lemma 5.5, one can establish that the probability that

$$\widehat{A} = LAR$$

has rank less than r is at most $3 \max(n,m)^{-2}$ 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). □

5.3 On 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 $\widehat{A}^2$. This observation can be used to establish upper bounds on the expected value of this difference and on the probability that this difference is high.

Lemma 5.8. *Let $L \in \mathsf{F}_q^{N\times n}$, $A \in \mathsf{F}_q^{n\times m}$, $R \in \mathsf{F}_q^{m\times N}$, and let $\widehat{A} = LAR \in \mathsf{F}_q^{N\times N}$. Suppose that*

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

Then

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

Proof. Since

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

the rank of $\widehat{A}^{\,2}$ is certainly less than or equal to that of $ARLA$. 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 $x_1, x_2, \ldots, x_s \in \mathsf{F}_q^{m\times 1}$ such that the vectors $y_1, y_2, \ldots, y_s \in \mathsf{F}_q^{n\times 1}$ are linearly independent if

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

Suppose, as in the claim, 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 as well. It follows that there exist vectors $z_1, z_2, \ldots, z_s \in \mathsf{F}_q^{N\times 1}$ 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. Now, this can be used to establish that if $k \geq 0$ and $v_1, v_2, \ldots v_k \in \mathsf{F}_q^{m\times 1}$, then $LAv_1, LAv_2, \ldots, LAv_k$ are linearly independent if and only if $Av_1, Av_2, \ldots, Av_k$ 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 $1 \leq i \leq k$, $Ly_1, Ly_2, \ldots, Ly_s$ are linearly independent, because $y_1, y_2, \ldots, y_s$ are linearly independent.

It remains only to note that

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

Thus 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 5.9. *Let $A \in \mathsf{F}_q^{n\times m}$ be a matrix with rank r, and let $L \in \mathsf{F}_q^{N\times n}$ be any matrix such that rows $\lceil 2\log_q \max(n,m)\rceil+1, \lceil 2\log_q \max(n,m)\rceil+2, \ldots, r$ of LA are linearly independent and the submatrix of LA including rows*

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

has rank r.

Suppose that the matrix $R \in \mathsf{F}_q^{m\times N}$ is randomly selected as described at the beginning of Subsection 5.1.

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

$$\frac{q^{1-i}}{q-1} + \max(n,m)^{-2}.$$

Proof. Since A has rank r, there exists a permutation matrix $P \in \mathsf{F}_q^{m\times m}$ 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 $\widetilde{A} \in \mathsf{F}_q^{r\times r}$ is a nonsingular matrix and where $C \in \mathsf{F}_q^{r\times(m-r)}$.

Suppose L has the properties given in the statement of the claim. Then there exists a permutation matrix $Q \in \mathsf{F}_q^{N\times N}$, with $(i,j)^{\text{th}}$ entry $Q_{i,j}$ for $1 \leq i, j \leq N$, such that $Q_{i,i} = 1$

for $\lceil 2\log_q \max(n,m)\rceil + 1 \leq i \leq \min(n,m)$ 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 $N \times r$ submatrix of QLA that includes the columns in T therefore has the form

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

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

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 $c+1 \leq i \leq \min(n,m)$ — 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 $R_0 \in \mathsf{F}_q^{r\times r}$, $R_{1,2} \in \mathsf{F}_q^{r\times(N-r)}$, $R_{2,2} \in \mathsf{F}_q^{(m-r)\times r}$, $R_{2,2} \in \mathsf{F}_q^{(m-r)\times(N-r)}$, and where the matrix $R_0 \in \mathsf{F}_q^{r\times r}$ is randomly chosen using the probability distribution described on page 47.

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

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

where

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

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 $B \in \mathsf{F}_q^{r\times r}$ are fixed as well. Since R_0 is chosen using the above-mentioned probability distribution, it follows by part (c) of Corollary 5.4 that the matrix $\widetilde{A}R_0 + B$ has rank less than $r - i$ with probability at most

$$\frac{q^{1-i}}{q-1} + \max(n,m)^{-2}.$$

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

$$\frac{q^{1-i}}{q-1} + \max(n,m)^{-2}$$

as well. □

Theorem 5.10. *Let $A \in \mathsf{F}_q^{n \times m}$ be a matrix with rank r.*

Suppose the matrices $L \in \mathsf{F}_q^{N \times n}$ and $R \in \mathsf{F}_q^{m \times N}$ are randomly chosen as described in Section 5.1, 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*

$$\frac{q^{2-i}}{q-1} + 4\max(n,m)^{-2}.$$

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

$$\frac{q^2}{(q-1)^2} + 4\max(n,m)^{-1}.$$

Proof. Let us begin by considering the condition that rows

$$\lceil 2\log_q \max(n,m) \rceil + 1, \lceil 2\log_q \max(n,m) \rceil + 2, \ldots, r$$

of the matrix LA are linearly independent and the submatrix of LA including rows

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

has rank r. This event depends on L, but not on the choice of R, and it follows by part (a) of Theorem 5.7 that the probability that this condition is *not* satisfied is at most

$$3\max(n,m)^{-2}. \tag{5.6}$$

On the other hand, it is follows by Lemma 5.9 that the probability that the above condition is satisfied, and $\operatorname{rank}(ARLA) < r - i$, is at most

$$\frac{q^{1-i}}{q-1} + \max(n,m)^{-2}$$

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 the above condition

is satisfied. 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 the above condition is satisfied and $\text{rank}(ARLA) \leq r-(i+1)$ with probability at most

$$\frac{q^{1-i}}{q-1} + \max(n,m)^{-2}$$

as well. Replacing i with $i-1$, one can conclude that this condition is satisfied and $\text{rank}(ARLA) \leq r-i$ with probability at most

$$\frac{q^{2-i}}{q-1} + \max(n,m)^{-2}. \tag{5.7}$$

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

$$\begin{aligned} \text{rank}(\widehat{A}) - \text{rank}(\widehat{A}^2) &\leq r - \text{rank}(\widehat{A}^2) \\ &= r - \text{rank}(ARLA) \qquad \text{(by Lemma 5.8).} \end{aligned}$$

Adding the probability that the above condition is *not* satisfied, it follows that the probability that LAR has i or more invariant factors divisible by x^2 is at most

$$\frac{q^{2-i}}{q-1} + max(n,m)^{-2} + 3\max(n,m)^{-2} = \frac{q^{2-i}}{q-1} + 4\max(n,m)^{-2}$$

as required to establish part (a).

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

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

and such that $\mathsf{NilBlocks} \geq i$ only if $\text{rank}(ARLA) \leq r-i$, as noted above. Then the expected value of NilBlocks is

$$\begin{aligned} \mathsf{E}[\mathsf{NilBlocks}] &= \sum_{i=1}^{r} \mathsf{Pr}[\mathsf{NilBlocks} \geq i] \\ &= \sum_{i=1}^{r} \left(\frac{q^{2-i}}{q-1} + 4\max(n,m)^{-2} \right) \\ &= \frac{q^2}{q-1} \sum_{i=1}^{r} q^{-i} + 4\max(n,m)^{-2} \sum_{i=1}^{r} 1 \\ &= \frac{q^2}{(q-1)^2} + 4r\max(n,m)^{-2} \\ &\leq \frac{q^2}{(q-1)^2} + 4\max(n,m)^{-1} \qquad \text{(since } r \leq \max(n,m)\text{)} \end{aligned}$$

as required to establish part (b). □

5.4 On 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 ℓ such factors, so that the first ℓ invariant factors of $\widehat{A}$ are

$$f_1, f_2, \ldots, f_\ell$$

where f_i is divisible by f_{i+1}, for $1 \leq i \leq \ell - 1$, and where f_ℓ has a nonzero root, λ, in some extension of F_q. Then λ is a root of f_i as well, for $1 \leq i \leq \ell$. It follows that the Jordan normal form of $\widehat{A}$ (over a suitable extension of F_q) includes at least ℓ blocks with eigenvalue λ, and that the matrix

$$\widehat{A} - \lambda I_N$$

has nullity at least ℓ. This observation will be used to bound the expected 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_q.

Lemma 5.11. *Let $\widetilde{A} \in \mathsf{F}_q^{r \times r}$ be a nonsingular matrix, and let $B \in \mathsf{E}^{r \times r}$.*

Suppose that the matrix $R_0 \in \mathsf{F}_q^{r \times r}$ is chosen as described for the probability distribution on page 47. Let

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

Let i be an integer such that $1 \leq i \leq r$. Then the probability that column i of D is a linear combination of columns $i+1, i+2, \ldots, r$ is at most

$$q^{-i} + \max(n, m)^{-3}.$$

Proof. This can be established using a straightforward modification of the the proof of part (a) of Lemma 5.3 and Corollary 5.4 — which correspond to the case that the above matrix B has elements in the field F_q.

As in the original proof, (after restricting attention to the special case that $\widetilde{A}$ is the identity matrix) one should suppose that the submatrix $\widehat{D}_i$ consisting of rows

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

of D has rank s_i. Then, after choosing values from F_q for all but $r - s_i$ of the entries in column i of R_0 one can observe that there is exactly one choice of the remaining values — in this case, in E but possibly not in F_q — 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_q, 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 $i+1, i+2, \ldots, r$. If the remaining values to be selected lie in F_q, instead, then the probability that the values are selected in the required way can be bounded using the argument given in Lemma 5.3.

To complete the proof, one should remove the assumption that $\widetilde{A}$ is the identity matrix using the same argument as is used in the proof of Lemma 5.3. □

Lemma 5.12. *Let $A \in \mathsf{F}_q^{n\times m}$ be a matrix with rank r and let $L \in \mathsf{F}_q^{N\times n}$ be a matrix such that rows $\lceil 2\log_q \max(n,m)\rceil + 1, \lceil 2\log_q \max(n,m)\rceil + 2, \ldots, r$ of the matrix LA are linearly independent and the submatrix of LA including rows*

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

has rank r.

Suppose the matrix $R \in \mathsf{F}_q^{m\times N}$ is randomly chosen using the probability distribution described at the beginning of Subsection 5.1.

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

$$(r-i+1)\frac{q^{2-i}}{q-1} + (r-i+1)^2 \max(n,m)^{-3}.$$

Proof. Suppose that L is as described in the above claim. Then there exists a permutation matrix $P \in \mathsf{F}_q^{N\times N}$ with $(i,j)^{\text{th}}$ entry $P_{i,j}$ for $1 \le i,j \le N$ such that $P_{i,i} = 1$ if $\lceil 2\log_q \max(n,m)\rceil + 1 \le i \le \min(n,m)$ and such that the top r rows of the matrix PLA are linearly independent. Once again, one can see, by inspection of the probability distribution described at the beginning of Section 5.1, 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 R 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 $Q \in \mathsf{F}_q^{m\times m}$ such that the top left $r \times r$ submatrix of LAQ 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. It follows that — replacing matrices A and R with AQ and $Q^{-1}R$, respectively — we may now assume that the top left $r \times r$ submatrix of LA is nonsingular.

The matrix LA can now be written as

$$LA = \begin{bmatrix} \widetilde{A}_{1,1} & \widetilde{A}_{1,2} \\ \widetilde{A}_{2,1} & \widetilde{A}_{2,2} \end{bmatrix} \tag{5.8}$$

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

Note that

$$R = \begin{bmatrix} R_0 & R_{1,2} \\ R_{2.1} & R_{2.2} \end{bmatrix} \tag{5.9}$$

where $R_0 \in \mathsf{F}_q^{r\times r}$ is randomly chosen using the probability distribution given on page 47, and where $R_{1,2} \in \mathsf{F}_q^{r\times(N-r)}$, $R_{2,1} \in \mathsf{F}_q^{(m-r)\times r}$, and $R_{2,2} \in \mathsf{F}_q^{(m-r)\times(N-r)}$.

Let $B \in \mathsf{F}_q^{r\times r}$ be the top left $r \times r$ submatrix of LAR. Then if follows by the decompositions given in equations (5.8) and (5.9), above, that

$$B = \begin{bmatrix} \widetilde{A}_{1,1} & \widetilde{A}_{1,2} \end{bmatrix} \cdot \begin{bmatrix} R_0 \\ R_{2,1} \end{bmatrix} = \widetilde{A}_{1,1}R_0 + \widetilde{A}_{1,2}R_{2,1} \in \mathsf{F}_q^{r\times r}. \tag{5.10}$$

Fix any choice of the rightmost $N - r$ columns of R — fixing the submatrices $R_{1,2}$ and $R_{2,2}$, shown above. Let

$$X = -\widetilde{A}_{2,1}\widetilde{A}_{1,1}^{-1} \in \mathsf{F}_q^{(N-r)\times r} \tag{5.11}$$

and let

$$\widehat{B} = B - (\widetilde{A}_{1,1}R_{1,2} + \widetilde{A}_{1,2}R_{2,2})X \in \mathsf{F}_q^{r\times r}. \tag{5.12}$$

It will next be argued that if the remaining columns of R (including submatrices R_0 and $R_{2,1}$) are chosen using the distribution given in the statement of the claim, then the above matrix $\widehat{B}$ has a large number of invariant factors, that are not powers of x, with small probability.

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

$$\operatorname{rank}(\widehat{B} - \lambda I_r) \leq r - i. \tag{5.13}$$

Let $\widehat{B}_j \in \mathsf{F}_q^{r\times(r-j)}$ denote the submatrix of $\widehat{B}$ that includes columns $j+1, j+2, \ldots, r$, and let $I_{r,j} \in \mathsf{F}_q^{r\times(r-j)}$ denote the submatrix of the $r \times r$ identity matrix I_r that includes columns $j+1, j+2, \ldots, r$. Then the condition at line (5.13) certainly implies that

$$\operatorname{rank}(\widehat{B}_{i-2} - \lambda I_{r,i-2}) \leq r - i \tag{5.14}$$

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

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

Now

$$p_{r+1} = 0, \tag{5.15}$$

since it is impossible for the matrix $\widehat{B}_{r-1} - \lambda I_{r,r-1}$ to have a negative rank for any choice of λ.

Suppose now that $2 \leq i \leq r$. In order to consider the difference between p_{i+1} and p_i, suppose that there does not exist any nonzero element λ of E such that

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

Let $C_{i-1} \in \mathsf{F}_q^{(r-i+1)\times(r-i+1)}$ be the submatrix of $\widehat{B}_{i-1}$ 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 $\widehat{B}$.

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. 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 $\widehat{B}_{i-1} - \lambda I_{r,i-1}$, it follows that

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

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

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

as well. Thus the condition at line (5.14) cannot be satisfied in this case.

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

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

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

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

as well, making the condition at line (5.14) impossible. It is therefore sufficient to consider the case that

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

It would follow in this case that the condition at line (5.14) is satisfied, for this choice of λ, if and only column $i - 1$ of the matrix

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

is a linear combination of columns $i, i+1, \ldots, r$ of this matrix. Applying Lemma 5.11 (for any choice of entries of column $i-1$ of the submatrix $R_{2,1}$, any choice of $R_{1,2}$ and $R_{2,2}$, as noted above, with $\widetilde{A}_{1,1}$ replacing $\widetilde{A}$ and with

$$\widetilde{A}_{1,2} R_{2,1} - (\widetilde{A}_{1,1} R_{1,2} - \widetilde{A}_{1,2} R_{2,2}) X - \lambda I_r$$

replacing the matrix called "B" in the statement of this lemma), one can see that the probability of this is at most

$$q^{1-i} + \max(n,m)^{-3}.$$

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, one can now conclude that

$$p_i \leq p_{i+1} + (r-i+1)\left(q^{1-i} + \max(n,m)^{-3}\right) \tag{5.16}$$

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

Using equations (5.15) and (5.16), it is easily established by induction on $r-i$ that if $2 \leq i \leq r$ then

$$p_i \leq (r-i+1)\frac{q^{2-i}}{q-1} + (r-i+1)^2 \max(n,m)^{-3}. \tag{5.17}$$

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 (5.8), and $X = -\widetilde{A}_{2,1}\widetilde{A}_{1,1}^{-1} \in \mathsf{F}_q^{(N-r)\times r}$, as above, then

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

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

Applying the decomposition of R in equation (5.9) 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

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

$$= \begin{bmatrix} B & C \\ 0 & 0 \end{bmatrix}$$

where B is the top left $r \times r$ submatrix of LAR and where $C = \widetilde{A}_{1,1}R_0 + \widetilde{A}_{1,2}R_{1,2} \in \mathsf{F}_q^{r\times(N-r)}$ s the top right $r \times (N-r)$ submatrix of LAR — and so that $\widehat{B} = B - CX$, for $\widehat{B}$ as defined at line (5.12) above.

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

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

and

$$\begin{aligned} \begin{bmatrix} I_r & 0 \\ X & I_{N-r} \end{bmatrix} \cdot (LAR - \lambda I_N) \cdot \begin{bmatrix} I_r & 0 \\ -X & I_{N-r} \end{bmatrix} & \\ = \begin{bmatrix} B - \lambda I_r & C \\ -\lambda X & -\lambda I_{N-r} \end{bmatrix} \cdot \begin{bmatrix} I_r & 0 \\ -X & I_{N-r} \end{bmatrix} & \\ = \begin{bmatrix} B - CX - \lambda I_r & C \\ 0 & -\lambda I_{N-r} \end{bmatrix} & \\ = \begin{bmatrix} \widehat{B} - \lambda I_r & C \\ 0 & -\lambda I_{N-r} \end{bmatrix}, & \end{aligned}$$

so that (since $\lambda \neq 0$)

$$\operatorname{rank}(LAR - \lambda I_N) = \operatorname{rank}(\widehat{B} - \lambda I_r) + N - r.$$

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

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 $\widehat{B}$ does. Since this probability is at most p_i, the claim now follows by the inequality at line (5.17), above. □

Theorem 5.13. *Let $A \in \mathsf{F}_q^{n\times m}$ be a matrix with rank r.*

Suppose the matrices $L \in \mathsf{F}_q^{N\times n}$ and $R \in \mathsf{F}_q^{m\times N}$ are randomly chosen, as described at the beginning of Section 5.1. 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*

$$\frac{(r-i+1)q^{2-i}}{q-1} + (r-i+4)\max(n,m)^{-2}.$$

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

$$\log_q r + 5.$$

Proof. Once again, consider the condition that rows

$$\lceil 2\log_q \max(n,m)\rceil + 1, \lceil 2\log_q \max(n,m)\rceil + 2, \ldots, r$$

of the matrix LA are linearly independent that the $(r + \lceil 2\log_q \max(n,m)\rceil) \times m$ submatrix of LA that includes rows

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

has rank r. This event depends only on the choice of the matrix L, and it follows by part (a) of Theorem 5.7 that the probability that this condition if *not* satisfied is at most $3\max(n,m)^{-2}$.

On the other hand, it follows by Lemma 5.12 that that the probability that this condition *is* satisfied, and LAR has i or more invariant factors that are not powers of x, is at most

$$(r-i+1)\frac{q^{2-i}}{q-1} + (r-i+1)^2 \max(n,m)^{-3} \leq (r-i+1)\frac{q^{2-i}}{q-i} + (r-i+1)\max(n,m)^{-2},$$

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 the above condition is satisfied.

Now, the probability that LAR has i or more invariant factors is certainly at most the sum of the above probabilities, as needed to establish part (a) of the claim.

Let $\mathsf{nonNilFactors}(LAR)$ denote the number of invariant factors of the matrix LAR that are not powers of x. This is an integer-valued random variable that can assume values between 0 and r. Thus

$$\mathsf{E}[\mathsf{nonNilFactors}(LAR)] = \sum_{i=1}^{r} \mathsf{Pr}[\mathsf{nonNilFactors}(LAR) \geq i] = S_1 + S_2 + S_3,$$

where

$$S_1 = \sum_{i=1}^{\lfloor \log_q r+3 \rfloor} \mathsf{Pr}[\mathsf{nonNilFactors}(LAR) \geq i],$$

$$S_2 = \sum_{i=\lceil \log_q r+3 \rceil}^{r} \mathsf{Pr}[\mathsf{Full} \ \wedge \mathsf{nonNilFactors}(LAR) \geq i],$$

and

$$S_3 = \sum_{i=\lceil \log_q r+3 \rceil}^{r} \mathsf{Pr}[\neg \mathsf{Full} \ \wedge \mathsf{nonNilFactors}(LAR) \geq i].$$

and where Full is the condition mentioned at the beginning of this argument. S_1, S_2 and S_3 will be bounded separately, below.

$$\begin{aligned} S_1 &= \sum_{i=1}^{\lfloor \log_q r+3 \rfloor} \mathsf{Pr}[\mathsf{nonNilFactors}(LAR) \geq i] \\ &\leq \sum_{i=1}^{\lfloor \log_q r+3 \rfloor} 1 \\ &= \lfloor \log_q r + 3 \rfloor \\ &\leq \log_q r + 3. \end{aligned}$$

As noted above, it follows by Lemma 5.12 that

$$\mathsf{Pr}[\mathsf{Full} \wedge \mathsf{nonNilFactors}(LAR) \geq i] \leq (r-i+1)\frac{q^{2-i}}{q-1} + (r-i+1)^2 \max(n,m)^{-3}$$

for any integer i between 1 and r. Thus

$$\begin{aligned} S_2 &= \sum_{i=\lceil \log_q r+3 \rceil}^{r} \mathsf{Pr}[\mathsf{Full} \ \wedge \mathsf{nonNilFactors}(LAR) \geq i] \\ &\leq \sum_{i=\lceil \log_q r+3 \rceil}^{r} \left((r-i+1)\frac{q^{2-i}}{q-1} + (r-i+1)^2 \max(n,m)^{-3} \right) \\ &\leq \frac{rq^2}{q-1} \sum_{i \geq \lceil \log_q r+3 \rceil} q^{-i} \ + \sum_{i=\lceil \log_q r+3 \rceil}^{r} \left((r-2)^2 \max(n,m)^{-3} \right) \\ &\leq \frac{rq^2}{q-1} \cdot \frac{q^{1-\lceil \log_q r+3 \rceil}}{q-1} + (r-2)^3 \max(n,m)^{-3} \\ &\leq 1 + (r-2)^2 \max(n,m)^{-2}. \end{aligned}$$

Once again, as noted above, Theorem 5.7 implies that

$$\mathsf{Pr}[\neg \mathsf{Full} \wedge \mathsf{nonNilFactors}(LAR) \geq i] \leq \mathsf{Pr}[\neg \mathsf{Full}] \leq 3 \max(n,m)^{-2}$$

for any integer i such that $1 \leq i \leq r$. Consequently

$$S_3 = \sum_{i=\lceil \log_q r+3 \rceil}^{r} \mathsf{Pr}[\neg \mathsf{Full} \ \wedge \mathsf{nonNilFactors}(LAR) \geq i]$$

$$\leq \sum_{i=\lceil \log_q r+3 \rceil}^{r} \max(n,m)^{-3}$$
$$\leq (r-2)\max(n,m)^{-2}.$$

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

$$\log_q r + 4 + r^2 \max(n,m)^{-2} \leq \log_q r + 5,$$

as required to establish part (b). □

5.5 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.

Since $q^2/(q-1)^2 \leq 4$, the next result is therefore a straightforward consequence of Theorems 5.10 and 5.13.

Theorem 5.14. *Let $A \in \mathsf{F}_q^{n\times m}$ be a matrix with rank r.*

Suppose the matrices $L \in \mathsf{F}_q^{N\times n}$ and $R \in \mathsf{F}_q^{m\times N}$ are randomly chosen, as described at the beginning of Section 5.1. 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*

$$\frac{(r-i+2)q^{2-i}}{q-1} + (r-i+8)\max(n,m)^{-2}.$$

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

$$\log_q r + 9 + 4\max(n,m)^{-1} \leq \log_q r + 13.$$

6 Questions and Further Work

So far as I know, this work does not contribute to a further analysis of either Coppersmith's or Montgomery's block Lanczos algorithms: Each of these symmetrizes the input matrix (working with A^TA or AA^T instead of A directly) if needed and then applies a somewhat simpler variant of a block Lanczos algorithm than the kind of "biconditional" block Lanczos algorithm whose analysis includes the block Hankel matrices discussed in Sections 3–4.

Unfortunately it is reasonably straightforward to demonstrate that these algorithms *cannot* be used to sample from the null space of A, in the worst case, if A is symmetrized

in either of the above ways, for computations over F_2: Indeed, it suffices to consider block diagonal matrices whose diagonal blocks include

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

(a symmetric matrix whose square is the zero matrix) to confirm that the ranks of both $A^T A$ and AA^T can be significantly lower than the rank of A — making nonzero elements of the null space of A extremely difficult to find by this process.

It is also known that the existence of nilpotent blocks with order at least two, in the Jordan normal form of a matrix, can significantly bias the distribution under which elements of the null space are sampled — such nilpotent blocks must be eliminated if one hopes to sample uniformly. Now consider the matrices

$$U = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} = U^T;$$

it is easily checked that these matrices both have minimal polynomial $x(x+1)$, so that they have rank one and they have no nilpotent blocks with order at least two in their Jordan form. However,

$$UU^T = V^T V = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

so that, once again, symmetrization can decrease the rank of these matrices as well. Furthermore

$$U^T U = VV^T = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

— a matrix that has minimal polynomial x^2, so that it is a nilpotent block with order two. Consequently one can begin with a matrix having no nilpotent blocks with order at least two at all and, through either of the above symmetrizations, produce a matrix for which the number of nilpotent blocks in the Jordan form is quite high.

Now, these algorithms have been used for work in computational number theory for years; there is a well developed code basis and the old adage, "If it ain't broke, don't fix it" probably applies. On the other hand, algorithms that symmetrize in this way should not be included in libraries if users expect them to be efficient and reliable in the worst case and cannot easily check the results that these algorithms provide.

This leads to the following question: Does there exist an efficient process that allows a matrix $A \in \mathsf{F}_2^{n \times n}$ to be symmetrized — in such a way that the resulting matrix has the same rank as A and, furthermore, has no nilpotent blocks with order at least two in its Jordan form?

Regrettably, this would not suffice to allow the results from this report to be applied: These algorithms use the same block of vectors on the left as on the right, so that the block

Hankel matrices arising in their analysis have the form

$$\mathcal{H}_{A,\vec{u},s,\vec{u},s} \qquad \text{or} \qquad \widehat{\mathcal{H}}_{A,\vec{u},s,\vec{u},\vec{s}}$$

for a single block of vectors $\vec{u} = u_1, u_2, \ldots, u_k \in \mathsf{F}_q^{n\times 1}$. On the other hand, the probability bounds given in Section 4 all concern matrices

$$\mathcal{H}_{A,\vec{u},s_L,\vec{v},s_R} \qquad \text{or} \qquad \widehat{\mathcal{H}}_{A,\vec{u},s_L,\vec{v},s_R}$$

where blocks of vectors $\vec{u} = u_1, u_2, \ldots, u_{k_L}$ and $\vec{v} = v_1, v_2, \ldots, v_{k_R}$ are assumed to be chosen uniformly and *independently* from $\mathsf{F}_q^{n\times 1}$. At present, I see no way to overcome this obstacle.

It is curious that a bound on the number of (nontrivial) invariant factors is needed for this analysis: As previously discussed [7, 11], no such restriction is needed for a similar analysis of scalar Wiedemann[3] and Lanczos algorithms. While I do not know how to eliminate this restriction entirely, for block algorithms, it does seem that it can be further relaxed, and eliminated completely in several interesting special cases. Further work along these lines is now in progress.

Finally, it should be noted that the Wiedemann sparse conditioner is probably not as "sparse" as one might like: If it is being used to precondition a matrix $A \in \mathsf{F}_q^{n\times n}$ (so that $n = m$, where Section 5 is concerned) then the expected number of nonzero entries of each preconditioner being used is in $\Theta(n \ln^2 n)$. To my knowledge, this is the only sparse or structured preconditioner that has been proved to be reliable for computations over F_2 or other finite fields F_q when q is extremely small. Do other — cheaper — matrix preconditioners that are reliable for computations over F_2 exist?

References

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

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

[3] L. Chen, W. Eberly, E. Kaltoven, 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, 1993.

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

[3] I should note, here, that Weidemann's original analysis of his scalar algorithm was also general. However, no "early termination" heuristic had been suggested at this time, so it is not considered in his paper.

[5] D. Coppersmith. Solving homogeneous linear equations over GF(2): Block Wiedemann algorithm. *Mathematics of Computation*, 62:333–350, 1994.

[6] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. The MIT Press, third edition, 2009.

[7] W. Eberly. Early termination over small fields. In *Proceedings, 2003 International Symposium for Symbolic and Algebraic Computation (ISSAC '03)*, pages 80–87. ACM Press, 2003.

[8] W. Eberly. Reliable Krylov-based algorithms for matrix null space and rank. Technical Report 2004-749-14, Department of Computer Science, University of Calgary, 2004.

[9] W. Eberly. Reliable Krylov-based algorithms for matrix null space and rank. In *Proceedings, 2004 International Symposium on Symbolic and Algebraic Computation (ISSAC '04)*, pages 127–134. ACM Press, 2004.

[10] W. Eberly. Yet another block Lanczos algorithm: How to simplify the computation and reduce reliance on preconditioners in the small field case. In *Proceedings, 2010 International Symposium on Symbolic and Algebraic Computation (ISSAC '10)*, pages 289–296, 2010.

[11] W. Eberly. Early termination over small fields I: On properties of scalar Krylov-based algorithms. Submitted for publication, 2013.

[12] W. Eberly and B. Hovinen. Bounding the nullities of random block Hankel matrices: An alternate approach. Technical Report 2005-779-10, Department of Computer Science, University of Calgary, 2005.

[13] W. Eberly and E. Kaltofen. On randomized Lanczos algorithms. In *Proceedings, 1997 International Symposium for Symbolic and Algebraic Computation (ISSAC '97)*, pages 176–183. ACM Press, 1997.

[14] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, fourth edition, 2012.

[15] M. R. Hestenes and E Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49:409–436, 1952.

[16] B. Hovinen. Block Lanczos-style algorithms over small finite fields. Master's thesis, Department of Computer Science, University of Waterloo, 2004.

[17] B. Hovinen and W. Eberly. A reliable block Lanczos algorithm over small fields. In *Proceedings, 2005 International Symposium on Symbolic and Algebraic Computation (ISSAC '05)*, pages 177–184. ACM Press, 2005.

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

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

[20] C. Lancos. Solution of systems of linear equations by minimized iterations. *Journal of Research of the National Bureau of Standards*, 49:33–53, 1952.

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

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

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

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

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

[26] G. Villard. Further analysis of Coppersmith's block Wiedemann algorithm using matrix polynomials. In *Proceedings, 1997 International Symposium on Symbolic and Algebraic Computation (ISSAC '97)*, pages 32–39. ACM Press, 1997.

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