RAPPORT DE RECHERCHE

A study of Coppersmith's block Wiedemann algorithm using matrix polynomials

G. Villard

N°975 IM Avril 97

Résumé

Nous analysons un algorithme probabiliste par blocs, proposé par Coppersmith, pour la résolution de grands systèmes linéaires creux $Aw = 0$ donnés sur un corps fini $K = \mathrm{GF}(q)$. Cet algorithme est une modification d'un algorithme de Wiedemann. Coppersmith a justifié sa méthode par plusieurs arguments heuristiques. Une question ouverte était de démontrer que l'algorithme peut effectivement fournir une solution, avec une probabilité strictement plus grande que zéro, sur les corps de petite cardinalité tels que $K = \mathrm{GF}(2)$. Nous répondons à la question à peu près complètement. L'algorithme utilise deux matrices X et Y de dimensions $m \times N$ et $N \times n$. Sur un corps fini quelconque, nous expliquons comment les paramètres m et n peuvent être réglés de manière à ce que, pour un système quelconque en entrée, une solution soit produite avec une bonne probabilité. Inversement, pour certains systèmes particuliers dits *non pathologiques*, nous montrons que les contraintes sur m et n peuvent être relâchées en continuant d'assurer une bonne probabilité de réussite. Accessoirement, nous améliorons une borne donnée par Kaltofen dans les cas des corps finis à grands nombres d'éléments. Enfin, pour rester complets dans notre généralisation des études de Wiedemann au cas matriciel, nous esquissons rapidement une version déterministe de l'algorithme par blocs.

A STUDY OF COPPERSMITH'S BLOCK WIEDEMANN ALGORITHM USING MATRIX POLYNOMIALS

GILLES VILLARD

ABSTRACT. We analyse a randomized block algorithm proposed by Coppersmith for solving large sparse systems of linear equations, $Aw = 0$, over a finite field $K = \mathrm{GF}(q)$. It is a modification of an algorithm of Wiedemann. Coppersmith has given heuristic arguments to understand why the algorithm works. But it was an open question to prove that it may produce a solution, with positive probability, for small finite fields *e.g.* for $K = \mathrm{GF}(2)$. We answer this question nearly completely. The algorithm uses two random matrices X and Y of dimensions $m \times N$ and $N \times n$. Over any finite field, we show how the parameters m and n of the algorithm may be tuned so that, for any input system, a solution is computed with high probability. Conversely, for certain particular input systems, we show that the conditions on the input parameters may be relaxed to ensure the success. We also improve the probability bound of Kaltofen in the case of large cardinality fields. Lastly, for the sake of completeness of the generalization of Wiedemann's work to the matrix case, we will briefly sketch a deterministic block algorithm.

1. INTRODUCTION

The randomized method proposed by Coppersmith [9] solves large sparse systems of homogeneous linear equations $Aw = 0$, $w \neq 0$. Throughout the paper A will be a singular $N \times N$ matrix over the Galois field with q elements $K = \mathrm{GF}(q)$ and w a vector of N unknowns. One fundamental application of this problem is integer and polynomial factorization, where such linear systems arise with N over $200,000$ [23, 25, 19]. This has motivated several authors to develop fast finite-field counterpart to numerical iterative methods. The conjugate gradient method has been used in [23], the Lanczos method in [23, 12] and the block Lanczos method in [8, 29].

But up to now, only the probabilistic analysis of Wiedemann [39] was giving a provably reliable and efficient method to solve $Aw = 0$ over small fields. This method is based on finding relations in Krylov subspaces using the Berlekamp-Massey algorithm [28]. The same analysis could be applied to bound the probability of success of the (bi-orthogonal) Lanczos and conjugate gradient algorithms with look-ahead of Lambert [24]. Anyway, these various approaches are very similar: they can be understood in a unified theory [24].

But since they use *generating polynomials of scalar sequences*, these latter algorithms impose limitations if one wants to perform several operations at a time. To solve this problem, Coppersmith [9] modifies the approach of Wiedemann and uses

LMC-IMAG, B.P. 53 F38041 Grenoble cedex 9, *Gilles.Villard@imag.fr*, April 23, 1997.

1991 *Mathematics Subject Classification.* Primary 15A06, 15A33; Secondary 15-04, 13P99.

Key words and phrases. Sparse linear systems, finite fields, exact arithmetic, probabilistic algorithms, matrix polynomials, minimal polynomials.

matrix sequences. This should be viewed as a block version of the same algorithm. Blocks enable one to take advantage of simultaneous operations: either using the machine word over GF(2) [9] or a parallel machine [18]. Coppersmith's algorithm is very powerful [9, 19, 26] but raises theoretical questions. We are going to answer some of them in this paper.

We refer to §2 for basic definitions and to §3.1 for a detailed presentation of the block algorithm. We only consider the method intuitively in this introduction. In the Wiedemann algorithm [39], one chooses at random a row vector x and a column vector y and one computes the lowest degree polynomial $g(\lambda) = g_0 + g_1\lambda + g_d\lambda^d$ in $K[\lambda]$ that linearly generates the sequence $h_i = xA^iy$, $0 \leq i \leq 2N-1$. We mean that $g(\lambda)$ satisfies for all $0 \leq i \leq 2N-d-1$:

$$g_0h_i + g_1h_{i+1} + \ldots + g_dh_{i+d} = x(g_0A^iy + g_1A^{i+1}y + \ldots + g_dA^{i+d}y) = 0.$$

With high probability, this polynomial is the minimal polynomial $\pi_A^y(\lambda)$ of y with respect to A and is such that $\pi_A^y(0) = 0$ (one does not need the minimal polynomial of A but only a factor of it):

$$\begin{cases} \pi_A^y(\lambda) = g_l\lambda^l + g_{l+1}\lambda^{l+1} + \ldots + g_d\lambda^d, 0 < l \leq d,\ g_l \neq 0, \\ \forall i, 0 \leq i \leq 2N-d-1 : g_lA^{l+i}y + g_{l+1}A^{l+i+1}y + \ldots + g_dA^{d+i}y = 0. \end{cases}$$

Taking $w = g_lA^{l-1}y + \ldots + g_dA^{d-1}y$ above relation shows that w is a solution: $Aw = 0$. Instead of vectors x and y, the modified algorithm of Coppersmith [9] uses a random matrix X with m rows and a random matrix Y with n columns. It first computes the sequence of $m \times n$ matrices $H_i = XA^iY, i = 0, \ldots, L-1$ with $L = N/m + N/n + O(1)$. By analogy with the scalar case we will see in §2 that one may define vector or *matrix generating polynomials* for that sequence. With high probability, such polynomials are also generating polynomials for the sequence $\{A^iY\}_{i\geq 0}$. They will lead to one or even several (when they exist) independent solutions w. Note that computing generating polynomials is a main subproblem of this approach. However, we will not investigate this question in detail in this document, but only make some remarks in relation with our probabilistic analysis (see §3.2 and §9).

The method of Coppersmith is randomized, essentially in the sense that a generating polynomial for $\{XA^iY\}_{0\leq i\leq L-1}$ may not be a generating polynomial for $\{A^iY\}_{i\geq 0}$ and thus may not allow the computation of a solution w. Apart from being influenced by the work of Wiedemann, our study of the block algorithm, will use two main previous results. The first one, of Coppersmith [9], relies on the notion of *pathological* input matrix A. For matrices having "too many" eigenvalues with high multiplicities (compared to m and n) the algorithm might fail. Using heuristic arguments, Coppersmith claims that if the input matrix A is not pathological then the algorithm succeeds. In addition, he observed experimentally that it is sufficient to consider the first $L = N/m + N/n + O(1)$ terms of the sequence $\{H_i\}_i$. The second result has been given by Kaltofen [18]. If $\pi_A(\lambda)$ denotes the minimal polynomial of A, it is possible to precondition A so that $\deg \pi_A(\lambda) = \text{rank } A + 1$. Then, if the field K has enough elements, the algorithm is guaranteed to compute a solution. The problem is to provide a full probabilistic analysis. We are going to establish that for most matrices A and for most choices of the input parameters (or *blocking factors*) m and n, the algorithm succeeds with non-zero probability.

Trying to answer the question "why is there no pathological matrix for Wiedemann's algorithm ?" – in the scalar case – we are first led to generalize his work, knowing that complementary arguments will be necessary – for the matrix case. Wiedemann in §VI of [39], computes the probability that sequences $\{xA^iy\}_{i\geq 0}$ and $\{A^iy\}_{i\geq 0}$ have the same generating polynomials. This will be extended to the matrix case at §8.1. The concept of pathological matrix has no sense at this stage. Then, one must determine the number L of terms H_i actually needed for the computation. In the worst case, it may be necessary to consider $L = N + N/n$, thus $L = O(N)$ terms as in the scalar case. Fortunately, as observed by Coppersmith, for certain matrix A, only $L = N/m + N/n + O(1)$ terms are required in any case – for any value of m and n. Precisely, the matrix A is pathological if this is not true. This concept is thus related only to the latter part of the analysis and to the values of m and n.

From these remarks, we will improve previous results in two directions by emphasizing the role of m and n. On the first hand, we prove that the algorithm may succeed, with a reasonable constant probability, provided that $m \geq n+2$. For that, we theoretically study the additive term $\Delta = O(1)$ in $L = N/m + N/n + O(1)$. This gives an algorithm that works for any field K and for any input system, and thus avoids the notion of pathological matrix. More precisely, theorem 9.1 will show that

$$\mathrm{Prob}_{X,Y} \text{ of success } \geq \tilde{\Phi}(m,n,A) + \tilde{\Theta}(m,n,A)q^{-\Delta}$$

where, for m large enough with respect to n, $\tilde{\Phi}(m,n,A)$ will be close to a constant between $1/4$ and 1 and $\tilde{\Theta}(m,n,A)$ will be close to a constant between 1 and 3. Using Δ additive terms of the sequence the probability of success will be made arbitrarily close to $\tilde{\Phi}(m,n,A)$.

Alternatively, we also show that the condition on m and n can be relaxed. We prove that – as heuristically justified by Coppersmith – the algorithm always works for non pathological matrices. On the other hand, in the case of large fields, we will see that the preconditioning required by Kaltofen is not necessary, the algorithm computes a solution for any input matrix. This will result in a better probability bound.

By analogy with the Wiedemann's deterministic algorithm [39], we will conclude the paper with a modification of the above block version. Using the material of previous sections we will briefly sketch a deterministic block algorithm for computing matrix generating polynomials.

The paper is organized as follows. After basic definitions in §2 and the presentation of the block algorithm in §3, we will characterize the "good blocking matrices" in §4. We will precisely understand which conditions X and Y must satisfy, so that the sequence $\{XA^iY\}_{0\leq i\leq L-1}$ may be used instead of the sequence $\{A^iY\}_{i\geq 0}$. Next, §5 will give two useful technical facts on generating polynomials. We will then characterize the "generic" behaviour of the algorithm by considering matrices X and Y with indeterminate entries in §6. This characterization will immediately apply over large fields to bound the probability of success. It is also useful to explain what are the expected generating polynomials of the input random sequence. The main probabilistic analysis is divided into three sections. After the introductory §7, we will give the first technical results in §8. The reader will then find the final theorems in §9. Paragraph 9.1 is devoted to small fields and §9.2 will focus on large cardinality fields. Finally, before conclusion, §10 will be devoted to the block version of the Wiedemann's deterministic algorithm.

By an abuse of notations, we will use "0" to denote either scalars, vectors or matrices. The dimension will be deduced from context, as it will for I, which denotes the *identity matrix*. The *degree* of a matrix is the maximum degree of its entries, its *determinantal degree* is the degree of its determinant. A *unimodular matrix* is a nonsingular matrix whose determinantal degree is 0. Two matrices are right *equivalent* (resp. left) if they differ by a right (resp. left) unimodular multiplier. Let $\mathcal{M}_{m,n}(K)$ and $\mathcal{M}_N(K)$ respectively denote the set of $m \times n$ matrices and of $N \times N$ matrices over K.

2. About realizations and generating polynomials

This section is intended to give some definitions and facts about realizations and about generating polynomials of matrix sequences. The formalism we introduce was not used by previous authors, but will make easier our presentation.

2.1. **Realizations of rational matrices.** Let $\Sigma = (X, A, Y)$ be a triplet of matrices in $\mathcal{M}_{m,N}(K)$, $\mathcal{M}_N(K)$ and $\mathcal{M}_{N,n}(K)$ respectively. We also consider two polynomial matrices $N(\lambda)$ in $\mathcal{M}_{m,n}(K[\lambda])$ and $D(\lambda)$ nonsingular in $\mathcal{M}_n(K[\lambda])$ such that the *right matrix fraction description* $H(\lambda) = N(\lambda)D^{-1}(\lambda)$ in $\mathcal{M}_{m,n}(K(\lambda))$ is *strictly proper i.e.* the degree of the numerator polynomial of each entry of $H(\lambda)$ is less than the degree of the denominator polynomial.

Definition 2.1. [41]. If $X(\lambda I - A)^{-1}Y = H(\lambda)$ then $\Sigma = (X, A, Y)$ is called an order N realization of $H(\lambda)$. Furthermore, since $H(\lambda)$ is strictly proper, it has a formal expansion at the infinity

$$H(\lambda) = \sum_{i=0}^{\infty} H_i \lambda^{-i-1} \tag{2.1}$$

where the H_i's are matrices in $\mathcal{M}_{m,n}(K)$, we have

$$XA^iY = H_i \text{ for } i = 1, \ldots \tag{2.2}$$

and Σ is also called an order N realization of the above matrix sequence.

The denominator matrix $D(\lambda)$ leads to the notion of generating polynomial. If $D(\lambda) = D_0 + D_1\lambda + \ldots + D_d\lambda^d$ with D_j in $\mathcal{M}_n(K)$, $0 \leq j \leq d$, then by computing $H(\lambda)D(\lambda)$ we get

$$N(\lambda) = (H_0\lambda^{-1} + H_1\lambda^{-2} + \ldots)(D_0 + D_1\lambda + \ldots + D_d\lambda^d).$$

Since $N(\lambda)$ is a matrix polynomial, comparing the coefficient of λ^{-i}, $i \geq 0$, on both sides of the latter equation leads to

$$\forall i \geq 0 : H_iD_0 + H_{i+1}D_1 + \ldots + H_{i+d}D_d = 0. \tag{2.3}$$

Any such nonsingular matrix polynomial $D(\lambda)$ is called a right *generating matrix polynomial* for the matrix sequence $\{H_i\}_{i=0}^{\infty}$. As presented in Coppersmith's paper [9] or in the analysis proposed in [18], we may also consider $D(\lambda)$ column by column. If $D^{(j)}(\lambda) = D_0^{(j)} + D_1^{(j)}\lambda + \ldots + D_d^{(j)}\lambda^d$ is the j-th column of $D(\lambda)$ we get the vector version of (2.3):

$$\forall i \geq 0 : H_iD_0^{(j)} + H_{i+1}D_1^{(j)} + \ldots + H_{i+d}D_d^{(j)} = 0 \tag{2.4}$$

and the vector polynomial is called a right *generating vector polynomial* for the sequence.

To fully characterize and classify the generating polynomials, we use a module theoretic approach as done for instance in [35, 4] for matrix Padé approximants. For the classic facts in the ongoing presentation we refer to [16].

The set of the right generating vector polynomials for the sequence $\{H_i\}_{i=0}^{\infty}$ is a $K[\lambda]$-submodule $\mathcal{W}$ of $K^n[\lambda]$. We know that such a submodule $\mathcal{W}$ has a basis of at most n elements. Since the columns of the diagonal matrix $\mathrm{diag}(\pi_A(\lambda), \ldots, \pi_A(\lambda))$ – where $\pi_A(\lambda)$ is the minimal polynomial of A – are all in $\mathcal{W}$, any basis of $\mathcal{W}$ must have exactly n elements.

All the bases – arranged as columns in a matrix – of $\mathcal{W}$ differ by a right unimodular multiplier. Thus the set of the right generating matrix polynomials of a sequence (2.2) can be uniquely determined by choosing a particular representative. As emphasized in [6, 37] several matrix polynomial normal form can be chosen. In next paragraph we focus on the Popov form which provides a notion of minimal polynomial.

2.2. **Minimum generating polynomials.** From a complexity point of view it is important to handle relations (2.3) or (2.4) of minimal length. We will define minimal bases for $\mathcal{W}$, these will correspond to minimal bases of vector spaces [13, 36]. In addition, to extend the notion of minimal scalar polynomial, we will speak (by abuse of language) of *minimal generating matrix polynomial.*

A basis given by the columns of a matrix $D(\lambda)$ will be minimal (see theorem 2.4 below) when $D(\lambda)$ will be *column reduced.* Uniqueness will be ensured by the *Popov form.* Let us define this latter form. For $D(\lambda)$ in $\mathcal{M}_n(K[\lambda])$ let d_j, $1 \leq j \leq n$, the *j-th column degree*, be the degree of the j-th column of $D(\lambda)$. The coefficient vector of λ^{d_j} is the j-th leading column coefficient vector. We let $[D(\lambda)]_c$ be the matrix of these leading vectors.

Definition 2.2. [33]. A matrix $D(\lambda)$ is said to be *column reduced* if $\mathrm{rank}\,[D(\lambda)]_c = \mathrm{rank}\; D(\lambda)$, thus its determinantal degree is $\deg\det\, D(\lambda) = \Sigma_{j=1}^{n} d_j$. If, in addition, $D(\lambda)$ satisfies the following properties, we shall say that $D(\lambda)$ is in *Popov form*:

I) the column degrees are increasingly ordered;

II) the last entry of degree d_j in each column is monic and is called the pivot of column j with row index r_j;

III) if $d_j = d_k$ and $j < k$ then $r_j < r_k$;

IV) all the entries in a row containing a pivot element have degrees lower than that of the pivot.

The Popov form is normal in the following sense:

Theorem 2.3. [33, 17]. *Any two right equivalent matrices in $\mathcal{M}_n(K[\lambda])$ are right equivalent to a same unique matrix in Popov form.*

Here is an important classical fact that identifies column reduced forms and minimal bases.

Theorem 2.4. [13]. *Let $D(\lambda)$ be a basis of $\mathcal{W}$ with j-th column degree d_j, $1 \leq j \leq n$. For any element $w(\lambda)$ of $\mathcal{W}$ let $w(\lambda) = D(\lambda)v(\lambda)$. Then, $D(\lambda)$ is column reduced if and only if it is a* minimal basis *in the sense that any $w(\lambda)$ satisfies:*

$$\deg\, w(\lambda) = \max_{j \in \underline{v}}\{d_j + \deg\, v_j(\lambda)\} \tag{2.5}$$

where $\underline{v}$ is the set of indices j such that the j-th entry $v_j(\lambda)$ of $v(\lambda)$ is non-zero. Furthermore, any two minimal bases of $\mathcal{W}$ have the same set of column degrees $\{d_j\}_{1\leq j\leq n}$, they are called the Kronecker indices *of $\mathcal{W}$.*

If the indices are arranged in increasing order, identity (2.5) shows that the corresponding elements of $\mathcal{W}$ are the first linearly independent ones with minimal degrees [13, 17]. This motivates the following definition.

Definition 2.5. The unique minimal basis $D_{\mathcal{W}}(\lambda)$ in Popov form is called the *minimal (right) generating matrix polynomial* for the matrix sequence $\{XA^iY\}_{i=0}^{\infty}$.

We may also look at this characterization column by column. For instance, the first column degree d_1 of $D_{\mathcal{W}}(\lambda)$ is the smallest possible length for a vector recurrence of type (2.4):

$$\forall i \geq 0 : H_i D^{(1)}_{\mathcal{W},0} + H_{i+1} D^{(1)}_{\mathcal{W},1} + \ldots + H_{i+d_1} D^{(1)}_{\mathcal{W},d_1} = 0. \tag{2.6}$$

Example 2.6. Consider the matrix sequence $\{H_i\}_{i\geq 0}$ over GF(2):

$$\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \ldots$$

for which

$$D(\lambda) = \begin{bmatrix} \lambda^2 + \lambda + 1 & 1 + \lambda^2 \\ 0 & 1 \end{bmatrix}$$

is a generating polynomial:

$$H_i \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} + H_{i+1} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + H_{i+2} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = 0 \text{ for } 0 \leq i \leq 4.$$

The column Popov form of $D(\lambda)$ is

$$D_{\mathcal{W}}(\lambda) = \begin{bmatrix} \lambda & 1 \\ 1 & 1 + \lambda \end{bmatrix}.$$

For instance, the first column of $D_{\mathcal{W}}(\lambda)$ gives a vector recurrence of smallest length:

$$H_i \begin{bmatrix} 0 \\ 1 \end{bmatrix} + H_{i+1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 0 \text{ for } 0 \leq i \leq 5.$$

Minimal bases appear in various domains and especially in Padé approximation theory. The reader may refer to the uniqueness conditions for the *matrix minimal Padé problem* in [6] and to the *σ-bases* in [3].

3. Coppersmith's block Wiedemann algorithm

Using above terminology, we now give the Coppersmith's version [9] of Wiedemann's algorithm for the solution of $Aw = 0$. The matrix A is singular and square of dimension N over a field K. In §3.1, we follow the notations of Kaltofen [18] and his variant of the method. We will then briefly discuss in §3.2 of the computation of generating polynomials.

3.1. Coppersmith's algorithm for singular systems.

The algorithm picks up a random matrix X in $\mathcal{M}_{m,N}(K)$ – say a *left blocking matrix* – and a random matrix Y in $\mathcal{M}_{N,n}(K)$ – say a *right blocking matrix*. The first step consists in computing the first terms of the sequence $\{XA^iY\}_{i=0}^{\infty}$. Coppersmith has introduced an additive term in $O(1)$ as a safety measure and has recommended to compute $N/m + N/n + O(1)$ terms of the sequence. This additive term will be denoted by Δ and will be called the *shift parameter* of the algorithm. We refer to §8 and to §9 for a detailed study of the theoretical behaviour of the algorithm with respect to that key parameter. We also refer to the experiments reported in [26, 19].

Algorithm. Coppersmith's block Wiedemann.

Input: *A a* $N \times N$ *matrix over* K *and the shift parameter* Δ*, a non-negative integer.*

Step 1. *Pick up random matrices* X*,* Y*. Let* $Z = AY$*.*

Step 2. *Let* $\delta_l = \lceil N/m \rceil$ *and* $\delta_r = \lfloor N/n \rfloor$*. Compute*

$$H_i = XA^iZ, \ i = 0, \ldots, \delta_l + \delta_r + \Delta - 1.$$

Then, solutions w such that $Aw = 0$ are constructed from generating vector polynomials for that sequence.

Step 3. *Compute a generating vector polynomial*

$$g(\lambda) = g_0 + g_1\lambda + \ldots + g_d\lambda^d \in K^n[\lambda]$$

of degree at most δ_r *for the sequence* $\{XA^iZ\}_i$ i.e. *such that:*

$$XA^iZg_0 + XA^{i+1}Zg_1 + \ldots + XA^{i+d}Zg_d = 0. \tag{3.1}$$

for $0 \le i \le \delta_l + \Delta - 1$*.*

With high probability, as we will prove later (see theorem 9.1 and theorem 9.6), the left projection by X does not modify the invariants of the sequence and $g(\lambda)$ is a generating vector polynomial for $\{A^iZ\}_i$:

$$0 \le i \le \delta_l + \Delta - 1 : A^iZg_0 + A^{i+1}Zg_1 + \ldots + A^{i+d}Zg_d = 0.$$

Let g_l be the first non-zero vector coefficient of $g(\lambda)$. Since $Z = AY$, above identities give in particular:

$$A^{l+1}\left(Yg_l + AYg_{l+1} + \ldots + A^{d-l}Yg_d\right) = A^lZg_l + \ldots + A^dZg_d = 0. \tag{3.2}$$

The left-hand side leads to a solution.

Step 4. *Compute* $\hat{w} = Yg_l + AYg_{l+1} + \ldots + A^{d-l}Yg_d$*.*

With high probability, $\hat{w}$ is a non-zero vector (again, see theorem 9.1 and theorem 9.6). From identity (3.2) we know that we can find an integer ι such that $A^\iota\hat{w} = 0$.

Step 5. *Compute the first integer* ι *such that* $A^\iota\hat{w} = 0$*.*

Output: *If* $\iota \ge 1$ *then* $w = A^{\iota-1}\hat{w}$ *else* $w = 0$*.*

The algorithm is randomized concerning two points: identity (3.2) may be false and the algorithm may return the trivial solution. The former point will be the major concern of subsequent sections, the probability of getting a non-trivial solution has been bounded by Coppersmith.

3.2. **One or several generating polynomials.** As noticed in the introduction, a main subproblem of the solution of $Aw = 0$ is the computation of *one* or *several* generating polynomials for the sequence $\{XA^iY\}_i$. Indeed, it seems that for several known methods, one may compute one or several generating polynomials at the same asymptotical cost.

Typically, Coppersmith has proposed a generalization of Berlekamp-Massey algorithm [28] to the matrix case. It can be seen that – as precisely stated in the scalar case in [11] – the algorithm is strongly related to Euclidean division [1, 6]. The reader may refer to [38] for a detailed study of these relations and a proof of the correctness of the algorithm. Its cost is $O((m+n)N^2)$ arithmetic operations in K. If $D_Z(\lambda)$ is the minimal generating polynomial for $\{XA^iZ\}_{i=0}^{\infty}$ and has j

columns of degree less than δ_r, then Coppersmith's method actually computes – with high probability – j independent generating polynomials (independent over $K^n[\lambda]$) [38]. This will motivate us, in §9, to bound the probability of success either to compute one generating polynomial or several ones.

The same remarks remain valid if we use instead a method based on Padé approximation. This approach enables one to use the superfast deterministic algorithm in [3]. Using FFT-based polynomial multiplication, generating polynomials will be computed at deterministic cost $O((m+n)^2 m(N/n)\log^2(mN/n)\log\log(mN/n))$ [38].

Another point of view may be found in [22]. In the current context, the reader will refer to [18] for a survey of the main probabilistic results and other references. Computing a generating polynomial can be done by finding a vector in the kernel of a block-Hankel matrix. Rewritting (3.1) in matrix form gives that a generating polynomial $g(\lambda) = g_0 + g_1\lambda + \ldots + g_d\lambda^d$ in $K^n[\lambda]$ is found by solving:

$$M(\delta_l+\Delta, d+1)\,\vec{g} = \begin{bmatrix} H_0 & H_1 & \ldots & H_d \\ H_1 & H_2 & \ldots & H_{d+1} \\ \vdots & & & \vdots \\ H_{\delta_l+\Delta-1} & \ldots & \ldots & H_{\delta_l+d+\Delta-1} \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \\ \vdots \\ g_d \end{bmatrix} = 0$$

where for any integers d_1 and d_2, $M(d_1, d_2)$ denotes the block-Hankel matrix whose blocks are the $\{H_{i+j-2}\}_{i,j}$'s, $1 \le i \le d_1$ and $1 \le j \le d_2$. Thus, step 3 of the block algorithm can be implemented as the computation of vectors in the kernel of $M(\delta_l+\Delta, \delta_r+1)$ or of the associated block-Toeplitz matrix. Using the Bitmead-Anderson/Morf method [5, 30], this can be done at probabilistic cost $O((m+n)^2 N\log^2 N\log\log N)$ [18].

4. Characterization of good blocking matrices

We have to characterize "good" matrices X and Y. This characterization is divided into two parts. They both study the relationship between the value and properties of the minimal generating polynomials and of the triplet $\Sigma = (X, A, Y)$ that defines the sequence. In §4.1 we see how the minimal polynomial of $\{A^iY\}_i$ and of $\{XA^iY\}_i$ can coincide. In §4.2 we focus on the length of the sequence that must be considered.

4.1. **Blocking matrices and Krylov subspaces.** Let $< Y >= \mathrm{span}(Y, AY, A^2Y, \ldots)$ and denote by $A_{<Y>}$ the restriction to $< Y >$ of the linear operator associated to the matrix A. The non-unity invariant factors (see appendix A) of $D_{\mathcal{W}}(\lambda)$ depend on the spectral structure of A and on the choice of (X, Y) with respect to that structure.

Theorem 4.1. *Let $D_{\mathcal{W}}(\lambda)$ be the minimal generating matrix polynomial of the sequence associated to a strictly proper rational matrix $H(\lambda)$. Any realization $\Sigma = (X, A, Y)$ of $H(\lambda)$ satisfies* $\deg\det D_{\mathcal{W}}(\lambda) \le \mathrm{order}\ \Sigma = \dim A$ *and there exists a realization (X_o, A_o, Y_o) such that the equality holds. The non-unity invariant factors of A_o and of $D_{\mathcal{W}}(\lambda)$ are the same.*

This is a classical result. The proof may be found in several papers. We refer to [7, 40] or to [17] and references therein. Given $D_Y(\lambda)$, next result states precisely how to choose X so that the generating polynomial remains unchanged. It is proven using also classical arguments from linear system theory (see [17] for instance).

Lemma 4.2. *Let $D_Y(\lambda)$ and $D_{\mathcal{W}}(\lambda)$ be the minimal generating matrix polynomials of the sequences $\{A^iY\}_{i=0}^{\infty}$ and $\{XA^iY\}_{i=0}^{\infty}$. Let N_Y denote the dimension of $< Y >$ and A_Y be a matrix associated to $A_{<Y>}$* i.e. *there exists a similarity transformation P such that*

$$P^{-1}AP = [P_Y \; P_2]^{-1}\, A\, [P_Y \; P_2] = \begin{bmatrix} A_Y & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

Then $D_Y(\lambda) = D_{\mathcal{W}}(\lambda)$ if and only if the dimension N_{XY} of the vector subspace of K^N generated by the rows of $\{XP_Y, XP_YA_Y, XP_YA_Y^2, \dots\}$ is equal to N_Y.

Proof. For the *if part* we prove that N_Y is the lowest possible order for a realization of $\{XA^iY\}_{i=0}^{\infty}$, then by theorem 4.1 the assertion will hold. Let us first construct a realization of order N_Y:

$$\begin{aligned} H(\lambda) &= X(\lambda I - A)^{-1}Y = XP(\lambda I - P^{-1}AP)^{-1}P^{-1}Y \\ &= X\,[P_Y \; P_2] \begin{bmatrix} (\lambda - A_Y)^{-1} & T_1 \\ 0 & (\lambda - A_2)^{-1} \end{bmatrix} \begin{bmatrix} \bar{Y} \\ 0 \end{bmatrix} \end{aligned}$$

where T_1 and $\bar{Y}$ are $N_Y \times (N - N_Y)$ and $N_Y \times n$ matrices. It follows that:

$$H(\lambda) = XP_Y(\lambda I - A_Y)^{-1}\bar{Y}. \tag{4.1}$$

Now, let (X', A', Y') be a realization of $H(\lambda)$ of order $N' < N_Y$. Consider the matrices $\mathcal{O}$ and $\mathcal{C}$ constructed from the row vectors of $\{XP_Y, XP_YA_Y, \dots, XP_YA_Y^{N_Y-1}\}$ and the column vectors of $\{\bar{Y}, A_Y\bar{Y}, \dots, A_Y^{N_Y-1}\bar{Y}\}$ (from the Cayley-Hamilton theorem the whole corresponding space is obtained). In the same way we construct $\mathcal{O}'$ and $\mathcal{C}'$ using X', A' and Y'. Since $(XP_Y, A_Y, \bar{Y})$ and (X', A', Y') are two realizations of $H(\lambda)$, we must have (*cf* identity (2.2)) $H_i = XP_YA_Y^i\bar{Y} = X'(A')^iY'$, for $i \geq 1$, and thus $\mathcal{O}C = \mathcal{O}'C'$. Using the assumptions we can compute the ranks of $\mathcal{O}C$ and of $\mathcal{O}'C'$. From the Sylvester's inequality we have

$$\text{rank}(\mathcal{O}) + \text{rank}(\mathcal{C}) - N_Y \leq \text{rank}(\mathcal{O}C) \leq \min\{\text{rank}(\mathcal{O}), \text{rank}(\mathcal{C})\}. \tag{4.2}$$

Using that both $\mathcal{O}$ and $\mathcal{C}$ have rank N_Y, we get that $\mathcal{O}C$ is of rank N_Y. But since A' has dimension N' strictly lower than N_Y: $\text{rank}(\mathcal{O}'C') < N_Y$ which is a contradiction and there is no realization of $H(\lambda)$ of order lower than N_Y. Applying the same reasoning with (I, A, Y), we also see that there is no realization of the sequence $\{A^iY\}_i$ of order lower than N_Y. By theorem 4.1 we get deg det $D_Y(\lambda) =$ deg det $D_{\mathcal{W}}(\lambda) = N_Y$ so the two matrix polynomials differ by a unimodular multiplier. Furthermore, they are both in normal form so $D_Y(\lambda) = D_{\mathcal{W}}(\lambda)$.

For the *only if* part, let us assume that the dimension of the row vector space generated by $\{XP_Y, XP_YA_Y, \dots\}$ is of dimension $N' < N_Y$. We can apply the same reasoning as above. There exists a similarity transformation Q such that

$$Q^{-1}A_YQ = [Q_X \; Q_2]^{-1} \begin{bmatrix} A_{XY} & 0 \\ A_{21} & A_{22} \end{bmatrix} [Q_X \; Q_2]$$

with A_{XY} of dimension N'. Thus we can construct a realization of order N' using $\bar{X}$ given by the first N_{XY} columns of XP_YQ and $\bar{\bar{Y}}$ given by the first N_{XY} rows of $Q^{-1}\bar{Y}$:

$$H(\lambda) = \bar{X}(\lambda I - A_{XY})^{-1}\bar{\bar{Y}}.$$

This is a contradiction and concludes the proof since by theorem 4.1, N_Y, the determinantal degree of $D_{\mathcal{W}}(\lambda)$, is the lowest possible order for a realization of $H(\lambda)$. ∎

4.2. **Length versus degree.** Lemma 4.2 indicates how X must be chosen to ensure that we can consider the sequence $\{XA^iY\}_i$ instead of the sequence $\{A^iY\}_i$. The next problem we address is how many terms of the sequence are required to compute the vector generating polynomials (see step 3 of the block algorithm). Of course, this will heavily depend on the actual degrees of these polynomials.

Lemma 4.3. *Let δ_r be the first index such that* $\mathrm{span}(Y, AY, \ldots, A^{\delta_r-1}Y)$ *is of maximal dimension N_Y. The minimal right generating polynomial $D_Y(\lambda)$ of the sequence $\{A^iY\}_{i=0}^{\infty}$ is of degree exactly δ_r (at least one of the column degree is equal to δ_r).*

Proof. Consider any generating vector polynomial $g(\lambda) = g_0 + g_1\lambda + \ldots + g_{\delta_r+1}\lambda^{\delta_r+1}$ of degree $\delta_r + 1$:

$$\forall i \geq 0 : A^iYg_0 + A^{i+1}Yg_1 + \ldots + A^{i+\delta_r+1}Yg_{\delta_r+1} = 0.$$

By assumption the vector $A^{\delta_r}Yg_{\delta_r+1}$ can be expressed in terms of lower powers A^kY, $k \leq \delta_r - 1$. Thus there exists a generating vector polynomial $h(\lambda) = h_0 + h_1\lambda + \ldots + h_{\delta_r-1}\lambda^{\delta_r-1} + g_{\delta_r+1}\lambda^{\delta_r}$, by definition:

$$\forall i \geq 0 : A^iYh_0 + \ldots + A^{i+\delta_r-1}Yh_{\delta_r-1} + A^{i+\delta_r}Yg_{\delta_r+1} = 0.$$

We now see that

$$g(\lambda) = \lambda h(\lambda) + r(\lambda)$$

where $r(\lambda)$ is also a generating polynomial of degree at most δ_r, in other words, any generating polynomial of degree $\delta_r + 1$ can be expressed using generating polynomials of degree at most δ_r. The same can clearly be done for any higher degree polynomials, this implies that $D_Y(\lambda)$ is at most of degree δ_r since it is minimal (theorem 2.4).

If all the elements of the basis $D_Y(\lambda)$ are of degree δ' strictly lower than δ_r. By definition 2.2 we know that

$$\mathrm{rank}\ [D_Y(\lambda)]_c = n.$$

Up to multiplications by λ we can thus construct n generating polynomials of degree at most $\delta' \leq \delta_r - 1$ whose leading column coefficients form a basis of K^n. These basis and polynomials can be used to express $A^{\delta'}Y$ in terms of lower powers which contradicts the assumption on δ_r. In conclusion, there must be at least one column of $D_Y(\lambda)$ of degree exactly δ_r. □

In this lemma we have considered the overall degree of the minimal generating polynomials, it could be that the degree differ a lot from a column to another. Next lemma consider $D_Y(\lambda)$ column by column.

Lemma 4.4. *Let X be such that the vector subspace generated by the rows of $\{XP_Y, XP_YA_Y, \ldots, XP_YA_Y^{\delta_l-1}\}$ is of dimension N_Y (lemma 4.2) and let δ_l be the first index such that this is true. Any vector polynomial $g(\lambda) = g_0 + g_1\lambda + \ldots + g_d\lambda^d$ such that*

$$XA^iYg_0 + XA^{i+1}Yg_1 + \ldots + XA^{i+d}Yg_d = 0 \tag{4.3}$$

for $0 \leq i \leq \delta_l - 1$, is a generating vector polynomial for the sequence $\{A^iY\}_{i=0}^{\infty}$.

Proof. We keep the same notations as in lemma 4.2. The Sylvester's inequality (4.2) applied to

$$M(\delta_l, d+1) = \begin{bmatrix} XP_Y \\ XP_YA_Y \\ \dots \\ XP_YA_Y^{\delta_l-1} \end{bmatrix} \begin{bmatrix} \bar{Y} & A_Y\bar{Y} & \dots & A_Y^d\bar{Y} \end{bmatrix}$$

gives

$$\begin{aligned} \text{rank } M(\delta_l, d+1) &= \text{rank} \begin{bmatrix} \bar{Y} & A_Y\bar{Y} & \dots & A_Y^d\bar{Y} \end{bmatrix} \\ &= \text{rank} \begin{bmatrix} Y & AY & \dots & A^dY \end{bmatrix}. \end{aligned}$$

Further, since by identity (4.1), $H_i = XA^iY = XP_YA_Y^i\bar{Y}$ for all $i \geq 0$, we have

$$\text{rank} \begin{bmatrix} H_0 & H_1 & \dots & H_d \\ H_1 & H_2 & \dots & H_{d+1} \\ \vdots & & & \vdots \\ H_{\delta_l-1} & \dots & \dots & H_{\delta_l+d-1} \end{bmatrix} = \text{rank} \begin{bmatrix} Y & \dots & A^dY \end{bmatrix}.$$

Any polynomial $g(\lambda)$ that satisfies (4.3) corresponds to a vector $\vec{g} = {}^t[{}^tg_0, \dots, {}^tg_d]$ of $K^{n(d+1)}$ which is in the kernel of the above block-Hankel matrix. From the last rank equality $\vec{g}$ is also in the kernel of the right-hand side Krylov matrix, hence it corresponds to a generating polynomial for $\{A^iY\}_{i=0}^{\infty}$. □

Thus, any generating vector polynomial of degree d for the sequence $\{XA^iY\}_i$ up to the $(\delta_l + d - 1)$-th term, is a generating polynomial for $\{A^iY\}_{i=0}^{\infty}$.

Remark 4.5. Since $D_Y(\lambda)$ has determinantal degree at most N and since it has n columns, for any A and Y we know that there always exists at least one generating vector polynomial of degree less than $\lfloor N/n \rfloor$ for $\{A^iY\}_{i=0}^{\infty}$.

It is well known that identity (4.3) has two main interpretations, in the *reverse sense* in relation with Padé approximation and in the *direct sense* in relation with Euclids's algorithm [1, 11, 6, 38]. As seen in §3.2, these interpretations actually give various methods to compute generating polynomials. For a polynomial $g(\lambda)$ of degree d, denote by $\hat{g}(\lambda)$ its reversal $\hat{g}(\lambda) = \lambda^d g(1/\lambda)$. The same can be defined for vector or matrix polynomials by reversing all the entries with respect to the maximum degree. By abuse, let also $\hat{H}(\lambda)$ denote:

$$\hat{H}(\lambda) = (1/\lambda)N(1/\lambda)D^{-1}(1/\lambda) = \sum_{i=0}^{\infty} H_i\lambda^i.$$

Now, a vector polynomial $g(\lambda)$ of degree d satisfies (4.3) if and only if it satisfies

$$\hat{H}(\lambda)\hat{g}(\lambda) - \hat{h}(\lambda) \equiv 0 \bmod \lambda^{\delta_l+d} \tag{4.4}$$

for a vector polynomial $h(\lambda)$ of degree lower than $d-1$. This relation may be viewed as giving a partial approximation of $\hat{H}(\lambda)$. If we go back to the direct sens, interpreting (4.4) as

$$\hat{H}(\lambda)\hat{g}(\lambda) + \hat{h}'(\lambda)\lambda^{\delta_l+d} = \hat{h}(\lambda),$$

and taking $\bar{H}(\lambda) = H_0\lambda^{\delta_l+d-1} + H_1\lambda^{\delta_l+d-2} + \dots + H_{\delta_l+d-1}$ we obtain

$$\bar{H}(\lambda)g(\lambda) + h(\lambda)\lambda^{\delta_l+d} = h'(\lambda). \tag{4.5}$$

Matrices $\bar{H}(\lambda)$ and $\lambda^{\delta_l+d}I$ may then be viewed as the inputs of a matrix extended Euclid's algorithm.

We conclude this section with an easy consequence of the two previous lemmata. Let δ_r defined as in lemma 4.3 and δ_l and the block-Hankel matrix M defined as in lemma 4.4.

Lemma 4.6. *If the blocking matrices X and Y are such that the two subspaces* $\mathrm{span}(XP_Y, XP_YA_Y, \dots, XP_YA_Y^{\delta_l-1})$ *and* $\mathrm{span}(Y, AY, \dots, A^{\delta_r-1}Y)$ *are of dimension N_Y then the minimal generating polynomial $D_Y(\lambda)$ for $\{A^iY\}_{i=0}^{\infty}$ can be computed from the kernel of $M(\delta_l, \delta_r+1)$.*

Proof. By lemma 4.3 we know that $D_Y(\lambda)$ has degree δ_r, thus by lemma 4.4 we know that each generating polynomial of the basis is found from the kernel of $M(\delta_l, \delta_r+1)$. In addition we may notice that from the Sylvester's inequality (see (4.2) and the assumptions on δ_r and δ_l, the rank of $M(\delta_l, \delta_r)$ is equal to N_Y. □

The assumptions of this lemma imply that the *left* [1] minimal generating polynomial for the sequence $\{XA^iY\}_{i=0}^{\infty}$ is of degree δ_l and that the *right* one is of degree δ_r. The main purpose of the analyses of the block algorithm [9, 18], is precisely to show that these degrees are both small enough.

5. Separation and continuation

We prove two additional facts that will be useful for the computation of minimal generating polynomials. We first show at §5.1, that the correct choices of X and Y can be studied up to a factorization of polynomials. Then at §5.2 we will see that the generating polynomials can be constructed using successive matrices X_k, $k \geq 1$.

To simplify the notations and by analogy with the case of scalar polynomials, we define the action – with respect to A – of a matrix polynomial $D(\lambda)$ on a constant matrix Y. For a matrix polynomial $D(\lambda) = D_0 + D_1\lambda + \dots + D_d\lambda^d$ in $\mathcal{M}_n(K[\lambda])$ and a matrix Y in $\mathcal{M}_{N,n}(K)$ it is natural to consider

$$Y.D(\lambda) = YD_0 + AYD_1 + A^2YD_2 + \dots A^dYD_d \in \mathcal{M}_{N,n}(K).$$

This can be rewritten column by column as

$$Y.D(\lambda) = \left[D_{1,j}(A)Y^{(1)} + D_{2,j}(A)Y^{(2)} + \dots + D_{n,j}(A)Y^{(n)}\right]_{j=1,\dots,n} \in \mathcal{M}_{N,n}(K),$$

where the entries of $D(\lambda)$ are denoted by $D_{i,j}(\lambda)$, $1 \leq i,j \leq n$, and the columns of Y are denoted by $Y^{(l)}$, $1 \leq l \leq n$. This action clearly makes $\mathcal{M}_{N,n}(K)$ a right $\mathcal{M}_n(K[\lambda])$-module. The nonsingular elements in the annihilator of Y are what we have called, the generating matrix polynomials for the sequence $\{A^iY\}_{i=0}^{\infty}$.

5.1. Separation.

In the scalar case, for a matrix A in $\mathcal{M}_N(K)$, if u and u' are two vectors whose minimal polynomials $\mu_{u'}(\lambda)$ and $\mu_{u''}(\lambda)$ with respect to A are relatively prime, then we know (see [14] for instance) that the minimal polynomial of $u'+u''$ is $\mu_{u'}(\lambda)\mu_{u''}(\lambda)$. This result remains valid in the case of matrix polynomials.

A *least common right multiple* $D(\lambda)$ (lcrm) of two matrices $P(\lambda)$ and $Q(\lambda)$ is a common right multiple which is a left divisor of every common right multiple of $P(\lambda)$ and $Q(\lambda)$ [27]. In particular, $P(\lambda)U(\lambda) = Q(\lambda)V(\lambda) = D(\lambda)$ for some matrices $U(\lambda)$ and $V(\lambda)$. Every pair of non-singular matrices $P(\lambda)$ and $Q(\lambda)$ have

[1] all the definitions have been given using column operations (*right* equivalence, *right* fraction description, …). Everything can be done using row operations, we leave the reader to make the appropriate changes.

a lcrm [27]. If $P(\lambda)$ and $Q(\lambda)$ have relatively prime determinants $p(\lambda)$ and $q(\lambda)$ then the determinant of a lcrm is $p(\lambda)q(\lambda)$.

Lemma 5.1. *Let A be in $\mathcal{M}_N(K)$. Let Y' and Y'' in $\mathcal{M}_{N,n}(K)$ such that the minimal generating polynomials $P(\lambda)$ and $Q(\lambda)$ for $\{XA^iY'\}_{i=0}^{\infty}$ and for $\{XA^iY''\}_{i=0}^{\infty}$ have relatively prime determinants. The minimal generating polynomial for the sequence $\{XA^i(Y'+Y'')\}_{i=0}^{\infty}$ is a lcrm of $P(\lambda)$ and $Q(\lambda)$.*

Proof. Let $D(\lambda) = D_0 + \ldots + D_d\lambda^d$ be an arbitrary generating polynomial for $\{XA^iY\}_{i=0}^{\infty}$ where $Y = Y' + Y''$. Applying $D(\lambda)$ to the sequence $\{A_iY'\}_{i=0}^{\infty}$ we let

$$\begin{aligned} Z = Y'.D(\lambda) &= Y'D_0 + AY'D_1 + \ldots + A^dY'D_d \\ &= (Y - Y'')D_0 + \ldots + A^d(Y - Y'')D_d. \end{aligned}$$

By definition of $D(\lambda)$, for all positive i,

$$\begin{aligned} XA^iZ &= XA^i(Y - Y'')D_0 + \ldots + XA^{i+d}(Y - Y'')D_d \\ &= -XA^iY''D_0 - \ldots - XA^{i+d}Y''D_d \\ &= XA^i\bar{Z}, \end{aligned}$$

where $\bar{Z} = -Y''.D(\lambda)$. The columns of $\bar{Z}$ belong to $< Y'' >$ thus by theorem 4.1, the determinant of the minimal generating polynomial $D_Z(\lambda)$ for $\{XA^iZ\}_{i=0}^{\infty} = \{XA^i\bar{Z}\}_{i=0}^{\infty}$ must be a divisor of the determinant of $Q(\lambda)$. Further, using the assumptions $D_Z(\lambda)$ and $P(\lambda)$ must have relatively prime determinants. Now, $Z.D_Z(\lambda) = Y'D(\lambda)D_Z(\lambda)$, thus $D(\lambda)D_Z(\lambda)$ is a generating polynomial for this latter sequence. Therefore, $D(\lambda)D_Z(\lambda)$ is a right multiple of $P(\lambda)$. But $D_Z(\lambda)$ and $P(\lambda)$ are relatively prime determinants thus $D(\lambda)$ is a right multiple of $P(\lambda)$. In the same way, by considering $Y'' = Y - Y'$, it is easily shown that $D(\lambda)$ is a right multiple of $Q(\lambda)$. This establishes that any generating polynomial for $\{XA^iY\}_{i=0}^{\infty}$ is a common right multiple of $P(\lambda)$ and $Q(\lambda)$. A minimal generating polynomial is thus a lcrm of these two matrices. □

As an obvious consequence,

Corollary 5.2. *Assume that A is a block-diagonal matrix $A = \mathrm{diag}\,(A_p, A_q)$ and that the characteristic polynomials $p(\lambda)$ and $q(\lambda)$ of A_p and A_q, are relatively prime. Consider $Y = {}^t[{}^tY_p\ {}^tY_q]$ in $\mathcal{M}_{N,n}(K)$ with corresponding dimensions of blocks. The minimal generating polynomial for the sequence $\{XA^iY\}_{i=0}^{\infty}$ is a lcrm of the minimal generating polynomial $D_p(\lambda)$ for $\{XA_p^iY_p\}_{i=0}^{\infty}$ and of the minimal generating polynomial $D_q(\lambda)$ for $\{XA_q^iY_q\}_{i=0}^{\infty}$.*

Proof. By theorem 4.1 we know that $D_p(\lambda)$ and $D_q(\lambda)$ have relatively prime determinants. They are also the minimal generating polynomials of $\{XA^i\bar{Y}_p\}_{i=0}^{\infty}$ and of $\{XA^i\bar{Y}_q\}_{i=0}^{\infty}$, where $\bar{Y}_p$ and $\bar{Y}_q$ are the block matrices ${}^t[{}^tY_p\ 0]$ and ${}^t[{}^tY_q\ 0]$. Finally, above lemma is applied with A and $Y = \bar{Y}_p + \bar{Y}_q$. □

Example 5.3. Consider the left blocking matrix

$$X = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

that satisfies the assumptions of lemma 4.2 over GF(2) for

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \text{ and } Y = \begin{bmatrix} Y_p \\ Y_q \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

The minimal generating polynomials of $\{XA^i\bar{Y}_p\}_{i=0}^{\infty}$ and of $\{XA^i\bar{Y}_q\}_{i=0}^{\infty}$ are:

$$D_p(\lambda) = \begin{bmatrix} 1+\lambda & 1 \\ 1 & \lambda \end{bmatrix}, \; D_q(\lambda) = \begin{bmatrix} 1+\lambda & 0 \\ 1 & 1+\lambda \end{bmatrix}.$$

One can take

$$D_{\mathcal{W}} = D_p(\lambda) \times \begin{bmatrix} 1 & 1+\lambda \\ 0 & 1+\lambda^2 \end{bmatrix} = D_q(\lambda) \times \begin{bmatrix} 1 & 0 \\ 0 & \lambda^2+\lambda+1 \end{bmatrix} = \begin{bmatrix} 1+\lambda & 0 \\ 1 & 1+\lambda^3 \end{bmatrix}$$

as lcrm of $D_p(\lambda)$ and of $D_q(\lambda)$. Since it is under Popov form, it is the minimal generating polynomial of $\{XA^iY\}_{i=0}^{\infty}$.

The reader should notice that the degree of a lcrm can be much larger than the sum of the degrees of the two matrices (only the determinantal degrees are actually involved). For that reason, corollary 5.2 will be useful to generalize Wiedemann's analysis in §8.1. Contrary to that, it seems difficult to use the same stategy with Coppersmith's analysis, which heavily depends on the actual degrees of the matrices.

5.2. **Continuation.** By analogy with Wiedemann's deterministic algorithm [39], the variant of the block algorithm of Coppersmith we will propose in §10, will be based on

Lemma 5.4. *Let A be in $\mathcal{M}_N(K)$ and Y in $\mathcal{M}_{N,n}(K)$. Let X_1 and X_2 be two left blocking matrices in $\mathcal{M}_{m_1,N}(K)$ and in $\mathcal{M}_{m_2,N}(K)$. If $D_{1,Y}(\lambda)$ is the minimal polynomial for $\{X_1A^iY\}_{i=0}^{\infty}$ and if $D_{2,Y}(\lambda)$ is the minimal polynomial for $\{X_2A^i(Y.D_{1,Y}(\lambda))\}_{i=0}^{\infty}$, then the Popov form of $D_{1,Y}(\lambda)D_{2,Y}(\lambda)$ is the minimal polynomial for $\{{}^t[{}^tX_1, {}^tX_2]A^iY\}_{i=0}^{\infty}$.*

Proof. Clearly, $D_{1,Y}(\lambda)D_{2,Y}(\lambda)$ is a generating polynomial for $\{{}^t[{}^tX_1, {}^tX_2]A^iY\}_{i=0}^{\infty}$. We have to show that it is minimal. Let $D(\lambda)$ be any generating polynomial for this latter sequence. Then, $D(\lambda)$ must be a generating polynomial for $\{X_1A^iY\}_{i=0}^{\infty}$. By definition of $D_{1,Y}(\lambda)$, there exists $U(\lambda)$ such that $D(\lambda) = D_{1,Y}(\lambda)U(\lambda)$. Now, $U(\lambda)$ must be a generating polynomial for $\{X_2A^i(Y.D_{1,Y}(\lambda))\}_{i=0}^{\infty}$, then there exist $V(\lambda)$ such that $U(\lambda) = D_{2,Y}(\lambda)V(\lambda)$. In conclusion, $D(\lambda) = D_{1,Y}(\lambda)D_{2,Y}(\lambda)V(\lambda)$. □

Lemma 5.4 will be used to compute minimal generating polynomials using, if necessary, several left blocking matrices.

Example 5.5. We compute the minimal generating polynomial of $\{A^iY\}_{i=0}^{\infty}$ over GF(2) with

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{bmatrix} \text{ and } Y = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}.$$

Let $m = 2$ and consider three successive left blocking matrices X_1, X_2 and X_3:

$$X_1 = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}, \; X_2 = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}, \; X_3 = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We denote by $D_{1,Y}(\lambda)$, $D_{2,Y}(\lambda)$ and $D_{3,Y}(\lambda)$ the corresponding partial generating polynomials. The minimal generating polynomial for $\{X_1A^iY\}_{i=0}^{\infty}$ is

$$D_{1,Y}(\lambda) = \begin{bmatrix} 1 & \lambda^2+\lambda+1 \\ 1 & 0 \end{bmatrix}$$

which is a left divisor of $D_Y(\lambda)$. To continue the computation we apply $D_{1,Y}(\lambda)$ to Y:

$$Y_1 = \left[I \times Y^{(1)} + I \times Y^{(2)}, (A^2+A+I)Y^{(1)} + 0 \times Y^{(2)}\right]$$

where $Y^{(1)}$ and $Y^{(2)}$ are the first and the second columns of Y. The minimal generating polynomial for $\{X_2 A^i Y_1\}_{i=0}^{\infty}$ is

$$D_{2,Y}(\lambda) = \begin{bmatrix} 1+\lambda & 1 \\ 0 & 1+\lambda \end{bmatrix}.$$

Applying $D_{2,Y}(\lambda)$ to Y_1 gives Y_2 and the minimal polynomial for $\{X_3 A^i Y_2\}_{i=0}^{\infty}$ is found to be

$$D_{3,Y}(\lambda) = \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix}.$$

The Popov form of the product of these three partial generating polynomials is

$$D_Y(\lambda) = \begin{bmatrix} 1+\lambda & \lambda^4+\lambda+1 \\ 1+\lambda & 1 \end{bmatrix}$$

which is the minimal polynomial for $\{A^i Y\}_{i=0}^{\infty}$.

6. Generic degree profiles of minimal polynomials

Given any matrix A, does the block algorithm work for any m and n? In Coppersmith's justification (§6 of [9]) one basic assumption is $m \geq n$. In Kaltofen's analysis (§5 of [18]) there is no restriction on m and n but in return, there are restrictions on A. Indeed, let A^* denote a restriction of A to its range space. If ϕ^* denotes the number of blocks of the Frobenius form of A^* (see theorem 11.1 in appendix A), then one must have $\phi^* = 1$ [18]. In the following we will see that only $m \geq \min\{\phi^*, n\}$ is required (see also proposition 8.2). In §9 we will see that this is not so restrictive.

To justify the probabilistic analysis of two next sections we need to state the result below that give *generic degree profiles* of minimal generating polynomials. This concept catches what are, in general, the column degrees of the generating polynomials for sequences constructed from a given matrix A. This is the same concept than the *generic rank profiles* [10] that catches what are, in general, the ranks of leading principal submatrices of matrices equivalent to a given matrix A. We thus follow the technique from Kaltofen, Pan and Saunders [20, 21, 18]. We call *generic degree profile* of the minimal generating polynomials for sequences $\{XA^iY\}_{i=0}^{\infty}$ (X has m rows and Y has n columns) the column degrees of the minimal generating polynomial for the sequence $\{\mathcal{X}A^i\mathcal{Y}\}_{i=0}^{\infty}$, where the entries of $\mathcal{X}$ and of $\mathcal{Y}$ are indeterminates $\xi_{j,k}$, $1 \leq j \leq m$, and $\upsilon_{k,l}$, $1 \leq l \leq n$, with $1 \leq k \leq N$.

Proposition 6.1. *Let A be a matrix in $\mathcal{M}_N(K)$ whose Frobenius form has ϕ companion blocks. Let $\nu = f_1 + \ldots + f_{\min\{\phi,n\}}$ be the sum of the dimensions of the first n or ϕ blocks of F. The minimal generating polynomial $D_{\mathcal{Y}}(\lambda)$ for the generic sequence $\{A^i\mathcal{Y}\}_{i=0}^{\infty}$ has determinantal degree ν and has degree exactly $\delta_r = \lceil \nu/n \rceil$ (over the rational function field $K(\upsilon_{1,1}, \ldots, \upsilon_{N,n})$) with column degrees*

$$[d_1, \ldots, d_n] = [\delta_r - 1, \ldots, \delta_r - 1, \delta_r, \ldots, \delta_r] \tag{6.1}$$

where $\delta_r - 1$ is repeated $\tau = n\lceil \nu/n \rceil - \nu$ times. Further, the pivot row indices satisfy:

$$r_j = \begin{cases} n - \tau + j, & 1 \leq j \leq \tau, \\ j - \tau, & \tau + 1 \leq j \leq n. \end{cases} \tag{6.2}$$

Proof. By uniqueness of the Frobenius form, we may consider A either as a matrix over K or as a matrix over $K(\upsilon_{1,1},\ldots,\upsilon_{N,n})$. The sum ν of the dimensions of the first n blocks of F is by construction the dimension of the vector subspace $<\mathcal{Y}>$. By theorem 4.1 this dimension is equal to the determinantal degree of $D_{\mathcal{Y}}(\lambda)$ and the first assertion is proven. For the second assertion, by lemma 4.3 we may equivalently show that:

$$\text{rank } \left[\mathcal{Y}, A\mathcal{Y}, \ldots, A^{\delta_r-1}\mathcal{Y}\right] = \nu. \tag{6.3}$$

Indeed, the degree cannot be lower than δ_r because

$$n(\delta_r - 1) < \nu.$$

To prove (6.3) we are going to *specialize* the indeterminates $\upsilon_{1,1},\ldots,\upsilon_{N,n}$. Since the rank cannot be greater than ν, it is sufficient to find a specialization of rank ν for $[\mathcal{Y},\ldots,A^{\delta_r-1}\mathcal{Y}]$. We construct this specialization by intermixing the columns of n Krylov matrices Υ_l over K, $1 \le l \le n$, themselves obtained from a modular cyclic basis with respect to A.

Since the right blocking matrix has n columns, with no lack of generality we may assume that $\phi \le n$ (otherwise, as done during the proof of lemma 4.2, up to a change of basis we can restrict ourselves to the first n blocks). In particular we have $\nu = f_1 + \ldots + f_\phi$. Consider a change of basis P such that $P^{-1}AP = F$, F being the Frobenius normal form of A (see appendix A). We know that P gives a modular cyclic basis generated by ϕ vectors $u_1,\ldots,u_\phi$ *i.e.* P can be written as:

$$P = \left[u_1, Au_1, \ldots A^{f_1-1}u_1, \ldots, u_\phi, \ldots, A^{f_\phi-1}u_\phi\right].$$

If $f_l = \delta_r$ for all l, $1 \le l \le n$, it is sufficient to specialize the columns of $\mathcal{Y}$ to the u_l's. In the general case the blocks of F have different sizes – some are of dimension greater than δ_r and some are of lower dimension – and we need to balance the columns in P.

To simplify the presentation we only detail the case where ν is a multiple of n. We construct n matrices Υ_l with the same number δ_r of columns. Assuming that F has ϕ_1 blocks of dimension greater than δ_r, we define the first ϕ_1 matrices Υ_l by:

$$\Upsilon_l = \left[u_l, Au_l, \ldots, A^{\delta_r-1}u_l\right],\ 1 \le l \le \phi_1. \tag{6.4}$$

For the other $n-\phi_1$ matrices Υ_l, we consider $n-\phi_1$ temporary matrices $\bar{\Upsilon}_l$. These latter matrices are obtained from columns corresponding to the blocks of F of dimension lower than δ_r ("small blocks") and from columns corresponding to the rest of the above blocks of dimension greater than δ_r ("large blocks"). It is easy to see that $\phi-\phi_1$ matrices $\bar{\Upsilon}_l$ with δ_r columns can be chosen under the form

$$\bar{\Upsilon}_l = \left[u_l, \ldots, A^{f_l-1}u_l, A^{s_1}u_{j_1}, \ldots, A^{t_1}u_{j_1}, \ldots, A^{s_k}u_{j_k}, \ldots, A^{t_k}u_{j_k}\right], \tag{6.5}$$

for $\phi_1 + 1 \le l \le \phi$. These matrices are obtained by completing the sets of columns corresponding to each of the $\phi-\phi_1$ small blocks with columns from large blocks (stopping the completion when the number of columns is δ_r). The indices k, $s_1,\ldots,s_k$, $t_1,\ldots,t_k$ and $j_1,\ldots j_k$ depend on l. Anyway, by construction since these indices correspond to large blocks, we have:

$$s_i \ge \delta_r,\ 1 \le i \le k. \tag{6.6}$$

And we may also ensure that

$$\forall\ i,\ 1 \le i \le k-1,\ \exists\ l : t_i = f_l \tag{6.7}$$

if for each completion we use the maximum number of columns corresponding to a given large block. The last $n-\phi$ matrices $\bar{\Upsilon}_l$, also of dimension δ_r, can be chosen as:

$$\bar{\Upsilon}_l = \left[A^{s_1}u_{j_1},\dots,A^{t_1}u_{j_1},\dots,A^{s_k}u_{j_k},\dots,A^{t_k}u_{j_k}\right] \tag{6.8}$$

for $\phi+1\le l\le n$, by taking the rest of the columns corresponding to the large blocks (the indices $k, s_., t_.$ and $j_.$ still depend on l). By construction, the matrix

$$\bar{\Upsilon} = \left[\Upsilon_1,\dots,\Upsilon_{\phi_1},\bar{\Upsilon}_{\phi_1+1},\dots,\bar{\Upsilon}_n\right]$$

is invertible in $\mathcal{M}_N(K)$. We now bring the $\bar{\Upsilon}_l$, $\phi_1+1\le l\le n$, to Krylov form Υ_l by performing elementary invertible transformations on $\bar{\Upsilon}$, so that this latter matrix remains Invertible. From identities (6.5), (6.8) and from properties (6.6), (6.7) we can focus on bringing matrices

$$\bar{T}_1 = \left[u_l,\dots,A^{f_l-1}u_l,0,\dots,0\right],$$

and

$$\bar{T}_2 = \left[0,\dots,0,A^s u_j,A^{s+1}u_j,\dots,A^t u_j,0,\dots,0\right],\ s\ge\delta_r,$$

to Krylov form. Indeed, the $\bar{\Upsilon}_l$'s can then be themselves transformed by linear combination of these two types of matrices. To bring $\bar{T}_1$ to Krylov form, it suffices to add suitable linear combinations of its first f_l columns to its last δ_r-f_l columns to get:

$$T_1 = \left[u_l,\dots,A^{f_l-1}u_l,\dots,A^{\delta_r-1}u_l\right]. \tag{6.9}$$

This is always possible since the dimension of the corresponding invariant subspace is f_l. Surely T_1 is not invertible but we have performed only invertible transformations on $\bar{\Upsilon}$. In the same way, for $\bar{T}_2$, we add to its first zero columns suitable linear combinations of columns of $[\Upsilon_1,\dots,\Upsilon_{\phi_1}]$ to get

$$\left[A^r u_j,A^{r+1}u_j,\dots,A^s u_j,A^{s+1}u_j,\dots,A^t u_j,0,\dots,0\right], \tag{6.10}$$

this is always possible for a non-negative integer r since $s\ge\delta_r$. To finish the construction, we add suitable combinations of columns to the latter zero ones to construct

$$T_2 = \left[A^r u_j,\dots,A^s u_j,\dots,A^t u_j,\dots,A^v u_j\right]. \tag{6.11}$$

These last operations can always be performed since by (6.7) we know that t is the dimension of an invariant subspace (if there actually remains some zero columns). Applying all the corresponding invertible transformations to $\bar{\Upsilon}$ we get:

$$\Upsilon = [\Upsilon_1,\dots,\Upsilon_{\phi_1},\Upsilon_{\phi_1+1},\dots,\Upsilon_n]$$

where Υ_l, $1\le l\le n$, is either a Krylov matrix of the form (6.4) or a linear combination of matrices of the form (6.9) and (6.11). By construction, Υ is invertible:

$$\text{rank }\Upsilon = \nu. \tag{6.12}$$

Let $\Upsilon_l^{(j)}$, $1\le j\le\delta_r$, denotes the j-th column of Υ_l, $1\le l\le n$. By specializing $\mathcal{Y}$ to

$$Y = \left[\Upsilon_1^{(1)},\dots,\Upsilon_n^{(1)}\right]$$

and using (6.12) we have, as announced, a specialization of (6.3) to

$$\text{rank }\left[\Upsilon_1^{(1)},\dots,\Upsilon_n^{(1)},\Upsilon_1^{(2)},\dots,\Upsilon_1^{(\delta_r)},\dots,\Upsilon_n^{(\delta_r)}\right] = \nu.$$

This establishes the second assertion of the lemma when n is an exact divisor of ν. In the general case let $\tau = n\lceil \nu/n \rceil - \nu$. We specialize a submatrix of the left-hand side of (6.3):

$$\text{rank } \left[\mathcal{Y}, A\mathcal{Y}, \ldots, A^{\delta_r-2}\mathcal{Y}, A^{\delta_r-1}\mathcal{Y}^{(1)}, \ldots, A^{\delta_r-1}\mathcal{Y}^{(n-\tau)}\right] = \nu, \tag{6.13}$$

where $\mathcal{Y}^{(l)}$, $1 \leq l \leq n$, denotes the l-th column of $\mathcal{Y}$. Using the same method than above – the proof is omitted – one can construct $n-\tau$ Krylov matrices Υ_l with δ_r columns and τ Krylov matrices Υ_l with $\delta_r - 1$ columns. These matrices provide a suitable specialization of identity (6.13) which is thus proved:

$$\text{rank } \left[\Upsilon_1^{(1)}, \ldots, \Upsilon_n^{(1)}, \ldots, \Upsilon_1^{(\delta_r-1)}, \ldots, \Upsilon_n^{(\delta_r-1)}, \Upsilon_1^{(\delta_r)}, \ldots, \Upsilon_{n-\tau}^{(\delta_r)}\right] = \nu.$$

It remains to see that the column degrees and the corresponding pivot indices of the minimal generating polynomial are as claimed. On the one hand, applying the same reasoning – column by column – than for lemma 4.3, we see that identity (6.13) implies that there exists a generating polynomial for $\{A^i\mathcal{Y}\}_{i=0}^{\infty}$ whose column degrees actually satisfy

$$[d_1, \ldots, d_n] = [\ \underbrace{\delta_r - 1, \ldots, \delta_r - 1}_{\tau},\ \ \delta_r, \ldots, \delta_r].$$

This generating polynomial is given by dependencies between the columns of

$$\mathcal{K} = \left[\Upsilon_1^{(1)}, \ldots, \Upsilon_n^{(1)}, \ldots, \Upsilon_1^{(\delta_r)}, \ldots, \Upsilon_n^{(\delta_r)}, \Upsilon_1^{(\delta_r+1)}, \ldots, \Upsilon_{n-\tau}^{(\delta_r+1)}\right],$$

the entries of the j-th column of the generating polynomial are obtained by writting $\Upsilon_j^{(\delta_r)}$ or $\Upsilon_j^{(\delta_r+1)}$ (depending on j) as a linear combination of the previous columns in $\mathcal{K}$. On the other hand, there is no generating polynomial for $\{A^i\mathcal{Y}\}_{i=0}^{\infty}$ with a column degree strictly lower than $\delta_r - 1$, otherwise the rank of the left-hand side of identity (6.13) should be strictly lower than ν.

Concerning the pivot row indices of $D_{\mathcal{Y}}(\lambda)$ under Popov form, from (6.13) the last τ columns of $\mathcal{Y}$ (indexed by $n - \tau + j$, $1 \leq j \leq \tau$) lead to a column degree $\delta_r - 1$ in $D_{\mathcal{Y}}(\lambda)$ thus

$$r_j = n - \tau + j,\ \ 1 \leq j \leq \tau,$$

and the first $n-\tau$ columns of $\mathcal{Y}$ (indexed by $j-\tau$, $\tau + 1 \leq j \leq n$) lead to a degree δ_r thus

$$r_j = j - \tau,\ \ \tau + 1 \leq j \leq n.$$

This conclude our characterization of the minimal generating polynomial of a generic sequence. □

Example 6.2. Consider a matrix A over GF(2) consisting of $\phi = 2$ companion blocks:

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

If we denote by e_i, $1 \leq i \leq 6$, the i-th canonical vector, we can take $u_1 = e_1$ and $u_2 = e_4$ to generate a modular cyclic basis.

If $n = 3$, the $\phi_1 = 2$ blocks are of dimension greater than $\delta_r = 6/3 = 2$. Following (6.4) we take:

$$\begin{cases} \Upsilon_1 = [e_1, Ae_1] \\ \Upsilon_2 = [e_4, Ae_4] \end{cases}$$

and following (6.8):

$$\bar{\Upsilon}_3 = \left[A^2e_1, A^2e_4\right].$$

This latter matrix is transformed by adding Ae_4 to its first column following (6.10), then by adding $A^3e_1 = Ae_1$ to its second column following (6.11). This leads to

$$\Upsilon_3 = \left[A^2e_1 + Ae_4, A(A^2e_1 + Ae_4)\right].$$

The minimal generating polynomial for $\{A^iY\}_{i=0}^{\infty}$, $Y = [e_1, e_4, A^2e_1 + Ae_4]$ is:

$$n = 3,\ D_Y(\lambda) = \begin{bmatrix} \lambda^2 & \lambda & 0 \\ \lambda & \lambda^2 & 0 \\ 1 & \lambda & 1+\lambda^2 \end{bmatrix}.$$

If $n = 4$ we take

$$\begin{cases} \Upsilon_1 = [e_1, Ae_1], \Upsilon_2 = [e_4, Ae_4] \\ \bar{\Upsilon}_3 = \Upsilon_3 = \left[A^2e_1\right], \bar{\Upsilon}_4 = \Upsilon_4 = \left[A^2e_4\right] \end{cases}$$

The corresponding minimal polynomial is

$$n = 4,\ D_Y(\lambda) = \begin{bmatrix} \lambda & 0 & \lambda^2 & 0 \\ 0 & \lambda & 0 & \lambda^2 \\ \lambda & 0 & 1 & 0 \\ 0 & \lambda & 0 & 1 \end{bmatrix}$$

with $r_1 = 3, r_2 = 4$ and $r_3 = 1, r_4 = 2$.

The properties of $D_{\mathcal{Y}}(\lambda)$ given by proposition 6.1, do not uniquely determine a generating polynomial. Indeed, different specializations may lead to different minimal generating polynomials with the same column degrees and pivot indices. Note also that the entries of $D_{\mathcal{Y}}$ are not, in general, over the ground field K but lie over $K(v_{1,1}, \ldots, v_{N,n})$. This is a main difference with the scalar case – *e.g.* see proposition 2 in [18]. Next example illustrates these remarks.

Example 6.3. We work with the matrix

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

The minimal generating polynomial for the generic sequence $\{A^i\mathcal{Y}\}_{i=0}^{\infty}$ ($n = 2$ over GF(2)) with indeterminate entries $v_{k,l}$, $1 \le l \le 2$ and $1 \le k \le 4$, is

$$D_{\mathcal{Y}}(\lambda) = \begin{bmatrix} \frac{v_{3,2}}{v_{3,1}}\lambda + \frac{v_{3,2}}{v_{3,1}} & \lambda^2 + 1 \\ \lambda + 1 & 0 \end{bmatrix}.$$

Two "good" specializations of $\mathcal{Y}$ may be

$${}^tY_1 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix},\ D_{Y_1}(\lambda) = \begin{bmatrix} 0 & 1+\lambda^2 \\ 1+\lambda & 0 \end{bmatrix}$$

and

$${}^tY_2 = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{bmatrix},\ D_{Y_2}(\lambda) = \begin{bmatrix} 1+\lambda & 1+\lambda^2 \\ 1+\lambda & 0 \end{bmatrix}.$$

The two corresponding minimal polynomials have the same degree profile than the generic one. A lcrm of $D_{\mathcal{Y}}(\lambda)$, $D_{Y_1}(\lambda)$ and $D_{Y_2}(\lambda)$ is:

$$D_{\mathcal{Y}}(\lambda)\begin{bmatrix} 0 & \lambda+1 \\ 1 & \frac{v_{3,2}}{v_{3,1}} \end{bmatrix} = D_{Y_1}(\lambda)\begin{bmatrix} 0 & 1+\lambda \\ 1 & 0 \end{bmatrix} = D_{Y_2}(\lambda)\begin{bmatrix} 0 & 1+\lambda \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1+\lambda^2 & 0 \\ 0 & 1+\lambda^2 \end{bmatrix}$$

where $1+\lambda^2$ is the minimal polynomial of A.

Proposition 6.1 has described the generic situation when lemma 4.3 is applied. In the same way we describe the generic situation corresponding to lemma 4.6:

Corollary 6.4. *Let A be a matrix in $\mathcal{M}_N(K)$ whose Frobenius form has ϕ blocks and let $\nu = f_1 + \ldots + f_{\min\{\phi,n\}}$. If $m \geq \min\{\phi, n\}$ then the minimal generating polynomial $D_{\mathcal{Y}}(\lambda)$ for the generic sequence $\{A^i\mathcal{Y}\}_{i=0}^{\infty}$ can be computed from the kernel of*

$$\mathcal{M}(\delta_l, \delta_r+1) = \begin{bmatrix} \mathcal{H}_0 & \mathcal{H}_1 & \ldots & \mathcal{H}_{\delta_r} \\ \mathcal{H}_1 & \mathcal{H}_2 & \ldots & \mathcal{H}_{\delta_r+1} \\ \vdots & & & \vdots \\ \mathcal{H}_{\delta_l-1} & \ldots & \ldots & \mathcal{H}_{\delta_l+\delta_r-1} \end{bmatrix}$$

where $\mathcal{H}_i = \mathcal{X}A^i\mathcal{Y}$, $i \geq 0$, and where $\delta_r = \lceil \nu/n \rceil$ and $\delta_l = \lceil \nu/m \rceil$. The rank of $\mathcal{M}(\delta_l, \delta_r)$ is equal to ν.

Proof. Let Y be a specialization of $\mathcal{Y}$ provided by the proof of proposition 6.1. We have $< Y > = < \mathcal{Y} >$ over $K(v_{1,1}, \ldots, v_{N,n})$. Let A_Y and $P = [P_Y\ P_2]$ be – as in lemma 4.2 – a corresponding restriction of A and an associated transformation matrix. The matrix $A_Y = A_{\mathcal{Y}}$ has dimension ν and its structure is given by the first n blocks of the Frobenius form of A. By applying proposition 6.1 on the left and using the assumption $m \geq \min\{\phi, n\}$, we get that the left minimal polynomial for $\{\mathcal{X}P_YA_Y^i\}_{i=0}^{\infty}$ has determinantal degree ν and degree $\delta_l = \lceil \nu/m \rceil$. Indeed, we can first specialize a generic matrix $\bar{\mathcal{X}}$ of dimension $m \times \nu$ for the generic sequence $\{\bar{\mathcal{X}}A_Y^i\}_{i=0}^{\infty}$ to a matrix $\bar{X}$. Then $[\bar{X}\ 0]P^{-1}$ is a suitable specialization for $\mathcal{X}$. We conclude as done for lemma 4.6. $\square$

We conclude this section with an easy consequence that will be needed for the block method, which computes generating polynomials for $\{A^iZ\}_{i=0}^{\infty}$ with $Z = AY$ instead of Y (see step 1 in §3.1).

Corollary 6.5. *Let ϕ_0 be the number of singular companion blocks of the Frobenius form F of A, and let $\nu^* = f_1+\ldots+f_{\min\{\phi,n\}} - \min\{\phi_0, n\}$. The minimal generating polynomial $D_{\mathcal{Z}}(\lambda)$ for the generic sequence $\{A^i\mathcal{Z}\}_{i=0}^{\infty}$, $\mathcal{Z} = A\mathcal{Y}$, has determinantal degree ν^* and has degree exactly $\delta_r^* = \lceil \nu^*/n \rceil$. If $m \geq \min\{\phi^*, n\}$ then it can be computed from the kernel of the block-Hankel matrix $\{\mathcal{X}A^{i+j-2}\mathcal{Z}\}_{i,j}$, $1 \leq i \leq \delta_l^*$ and $1 \leq j \leq \delta_r^*+1$, which is of rank ν^*.*

Proof. These results are obtained as done for proposition 6.1 and corollary 6.4. It is sufficient to notice that if the invariant factors of $A_{<\mathcal{Y}>}$ are $s_1(\lambda), \ldots, s_{\min\{\phi,n\}}(\lambda)$, then the invariant factors of $A_{<\mathcal{Z}>}$ are

$$s_1(\lambda)/\lambda, \ldots, s_{\min\{\phi_0,n\}}(\lambda)/\lambda, s_{\min\{\phi_0,n\}+1}(\lambda), \ldots, s_{\min\{\phi,n\}}(\lambda).$$

$\square$

This is a generalization of proposition 3 in [18] to any type of matrix.

7. Following Wiedemann and Coppersmith's analyses

We are going to rely on two main ingredients. The first one is a generalization of Wiedemann's work [39] in §8.1. We bound the probability of picking two matrices X and Y such that $D_Y(\lambda) = D_{\mathcal{W}}(\lambda)$ (following the notations of lemma 4.2). This equality only gives an *incomplete answer with respect to Coppersmith's algorithm.* Indeed, it does not focus on the actual degrees of $D_Y(\lambda)$ and $D_{\mathcal{W}}(\lambda)$ but only on their determinantal degrees. This is a key difference between the scalar and the matrix cases. In the scalar case, the dimension of $\text{span}(y, Ay, A^2y, \ldots)$ is equal to the degree of the minimal polynomial $\pi_y(\lambda)$, of y with respect to A. In the matrix case, there is no such relation between the degrees of the columns of $D_Y(\lambda)$ and the dimension of $\text{span}(Y, AY, A^2Y, \ldots)$. The quantity that should be considered then is the determinantal degree of $D_Y(\lambda)$. To establish his results, Wiedemann has proven the following fact:

Proposition 7.1. [39]. *Let x and y be a row and a column vector over K. Let $\pi_{xy}(\lambda)$ and $\pi_y(\lambda)$ be the minimal generating polynomials for $\{xA^iy\}_i$ and for $\{A^iy\}_i$. Consider the space $\mathcal{R}_1$ of polynomials of degree less than the degree of $\pi_y(\lambda)$ in $K[\lambda]$. There exists a surjective linear map*

$$\zeta_1 : K^N \rightarrow \mathcal{R}_1$$

such that

$$\forall x \in K^N : \pi_{xy}(\lambda) = \pi_y(\lambda) \text{ iff } \gcd(\pi_y(\lambda), \zeta_1(x)) = 1.$$

Then, over $K = \text{GF}(q)$ in analogy with the Euler's Phi function of an integer, the probability that a random polynomial $\zeta(x)$ in $K[\lambda]/(\pi_y(\lambda)K[\lambda])$ is a unit can be computed. Lower bounds for that probability are given in [39] and in [15]. To apply lemma 4.2, in §8.1 below we propose a way of bounding the probability in the matrix case. In particular, this will result in a new proof of Wiedemann's bounds in the scalar case.

The second ingredient of the analysis, will consist in bounding the probability that X and Y lead to generating polynomials with "almost" generic degrees. This will enable us to bound the number of needed terms H_i. In the generic case – as characterized by proposition 6.1 – the minimal generating polynomial has $\tau = n\lceil \nu/n \rceil - \nu$ columns of degree $\delta_r - 1$ and $n - \tau$ columns of degree δ_r. But, as implicitly noticed in [9, 18, 26], to always have exactly the generic degrees seems unlikely. We will prove instead, that with a good probability, X and Y lead to degrees at most Δ plus the generic ones, for Δ a fixed positive integer. This will be done in §8.2 using certain results of Coppersmith [9]. In §3.1, Δ has been called the *shift parameter* of the block algorithm. Fortunately, this relaxed condition on the degrees will be sufficient.

8. Preliminary probability analysis

We keep the notations as in previous sections to investigate the two complementary analyses.

8.1. **Generalization of Wiedemann's analysis.** The probability that $\{A^iY\}_0^\infty$ and $\{XA^iY\}_0^\infty$ have the same minimal generating polynomials is computed in terms of the function $\Phi_n(f, \phi)$. This function is defined over $K = \text{GF}(q)$ for a polynomial

$f(\lambda)$ in $K[\lambda]$ and two positive integers n and ϕ, $n \geq \phi$:

$$\Phi_n(f,\phi) = \prod_{g \text{ irr. } |f} \left((1 - q^{-(\rho+\mu)\deg g}) \prod_{k=2}^{\infty} (1 - q^{-k\mu \deg g}) \right) \tag{8.1}$$

where the product is taken over the irreducible factors of $f(\lambda)$ in $K[\lambda]$ and where $\mu = \lfloor n/\phi \rfloor$ and ρ satisfy ϕ: $n = \mu\phi + \rho$. Lower bounds for $\Phi_n(f,\phi)$ will be given in appendix B.

Lemma 8.1. *Let A be a $N \times N$ matrix over K with minimal polynomial $\pi_A(\lambda)$, and let Y with $n \geq \phi$ columns chosen at random. If $K = GF(q)$ then*

$$\text{Prob}_Y \{\dim \; < Y > \; = \; \nu\} > \Phi_n(\pi_A, \phi).$$

Proof. Up to a change of basis P we may assume that A is under *rational Jordan form* (see theorem 11.4 in appendix A). We mean that K^N is decomposed into a direct sum of invariant subspaces with respect to A whose minimal polynomials are powers of irreducible factors of $\pi_A(\lambda)$ in $K[\lambda]$. This may be denoted by:

$$\bar{A} = P^{-1}AP = \text{diag}(\{A_{g_i}\}_i) = \left(\{A_{g_i}^{(1)}, \ldots, A_{g_i}^{(\phi_i)}\}_i\right) \tag{8.2}$$

where $g_i(\lambda)$ is the i-th irreducible factor of $\pi_A(\lambda)$ and $A_{g_i}^{(j)}$, $1 \leq j \leq \phi_i$, is one the $\phi_i \leq \phi$ square blocks over K associated to $g_i(\lambda)$. By analogy with the dimension ν, we denote by ν_i the dimension of A_{g_i} the block-diagonal matrix formed by all the blocks associated to $g_i(\lambda)$. Since multiplication by P^{-1}: $\mathcal{M}_{N,n}(K) \rightarrow \mathcal{M}_{N,n}(K)$ is a bijection, it follows that Y is chosen uniformly at random over K if and only if $\bar{Y} = P^{-1}Y$ is uniform random over K. Using corollary 5.2, we may thus separate the problem according to (8.2). Let p_Y be the probability that dim$< Y >= \nu$:

$$\begin{aligned} p_Y &= \text{Prob}_Y \left\{\dim \text{span}(\bar{Y}, \bar{A}\bar{Y}, \bar{A}^2\bar{Y}, \ldots) = \nu\right\} \\ &= \textstyle\prod_i \text{Prob}_{Y_i} \left\{\dim \text{span}(Y_i, A_{g_i}Y_i, \ldots) = \nu_i\right\}, \end{aligned} \tag{8.3}$$

where Y_i denotes the $\nu_i \times n$ submatrix of $\bar{Y}$ whose row indices correspond to the row indices of A_{g_i} in $\bar{A}$. We are going to compute a lower bound for each probability in above product. We focus on A_{g_i} and Y_i. Let $g_i(\lambda)$ be of degree d_i and let each block $A_{g_i}^{(j)}$, $1 \leq j \leq \phi_i$, be of minimal polynomial $g_i^{k_j}(\lambda)$ and thus of dimension $k_j d_i$: $(k_1 + \ldots + k_{\phi_i})d_i = \nu_i$. Let us restrict ourselves for the moment to the random choice of the first ϕ_i columns $c_1, \ldots, c_{\phi_i}$ of Y_i ($\phi_i \leq \phi \leq n$). We denote by V_j the vector space generated by $c_1, \ldots, c_j$:

$$V_j = \text{span}(c_1, A_{g_i}c_1, A_{g_i}^2 c_1, \ldots, c_2, \ldots, c_j, A_{g_i}c_j, \ldots),$$

for $1 \leq j \leq \phi_i$ and let $V_0 = \{0\}$. A sufficient condition to ensure that the dimension of span$(Y_i, A_{g_i}Y_i, \ldots)$ is ν_i, is that dim$V_{\phi_i} = \nu_i$. This condition will be satisfied if for any j, the minimal polynomial of c_j modulo V_{j-1} is any power of $g_i(\lambda)$ corresponding to a block of A_{g_i} different from those associated to the previous columns of Y_i. We compute the probability that $c_1, \ldots c_{\phi_i}$ satisfy this property by induction on c_j, $1 \leq j \leq \phi_i$.

For $j = 1$, c_1 must span any of the ϕ_i invariant subspaces, its minimal polynomial has to be any power of $g_i(\lambda)$ corresponding to a block of A_{g_i}. For one block $B_j = A_{g_i}^{(j)}$ the probability of failure is the probability that c_1 satisfies

$$g_i^{k_j - 1}(B_j)c_1 = 0.$$

The rank of $g_i^{k_j-1}(B_j)$ is d_i, for B_j the probability of failure is thus q^{-d_i}. Further, we may look separately at the entries of c_1 corresponding to the different blocks, thus the probability of failure for the choice if c_1 is $q^{-\phi_i d_i}$. We now assume that $c_1, \ldots, c_{j-1}$ satisfy the property. The minimal polynomial of c_j modulo V_{j-1} must be any power of $g_i(\lambda)$ corresponding to one of the remaining $\phi - j + 1$ blocks. Up to a change of basis with respect to V_{j-1} we may follow the same reasoning than for c_1. For one block the probability of failure is q^{-d_i}, thus the probability that c_j does not satisfy the property is $q^{-(\phi_i-j+1)d_i}$. If p_{Y_i} denotes the probability that $\dim \operatorname{span}(Y_i, A_{g_i}Y_i, \ldots) = \nu_i$, this shows that

$$\begin{array}{lcl} p_{Y_i} & \geq & \operatorname{Prob}_{\{c_j\}_j}\{\dim V_{\phi_i} \;=\; \nu_i\} \\ & \geq & (1-q^{-\phi_i d_i})(1-q^{-(\phi_i-1)d_i})\ldots(1-q^{-d_i}). \end{array}$$

Above bound can be easily improved when n is greater than ϕ_i. Indeed, let $n = \mu\phi + \rho$, thus, in particular, we have $\mu\phi_i + \rho \leq n$. Above construction can be done using μ columns of Y_i for each of the first $\phi_i - 1$ blocks of A_{g_i} and $\mu + \rho$ columns for the last one:

$$p_{Y_i} \geq (1-q^{-\phi_i \mu d_i})\ldots(1-q^{-2\mu d_i})(1-q^{-(\mu+\rho)d_i}).$$

Finally, from (8.3) we obtain

$$p_Y \geq \prod_i (1-q^{-\phi_i \mu d_i})\ldots(1-q^{-2\mu d_i})(1-q^{-(\mu+\rho)d_i}) > \Phi_n(\pi_A, \phi).$$

□

The function $\Phi_n(\pi_A, \phi)$ is a rough lower bound, but this will be sufficient to bound the probability of failure. Note that for $n = 1$ our study reduces exactly to the analysis of Wiedemann [39]. Now, applying the lemma on the left:

Proposition 8.2. *Let A be a $N \times N$ matrix over $K = GF(q)$, let X and Y be chosen at random with m rows and n columns. If $m \geq \min\{\phi, n\}$ then $D_Y(\lambda) = D_{\mathcal{W}}(\lambda)$ with probability no less than $\Phi_m(\pi_A, \min\{\phi, n\})$.*

Proof. From lemma 4.2 (with the same notations) we know that the probability $p_{\mathcal{W}}$ that $D_Y(\lambda) = D_{\mathcal{W}}(\lambda)$ is the probability that $\dim \operatorname{span}(XP_Y, XP_YA_Y, \ldots) = N_Y$. We recall that A_Y is a restriction of A to $< Y >$ associated to the similarity transformation P. Since P is invertible, if X is selected uniformly at random in $\mathcal{M}_{m,N}(K)$ then so it is for XP and then, XP_Y is also uniform random in $\mathcal{M}_{m,N_Y}(K)$. If we denote by ϕ_Y the number of non-trivial invariant factors of A_Y and by $\pi_{A_Y}(\lambda)$ its minimal polynomial, we know that $\phi_Y \leq \min\{\phi, n\}$. Further, using the assumptions, $\phi_Y \leq m$. We may thus apply lemma 8.1 on the left with A_Y and XP_Y: $p_{\mathcal{W}} > \Phi_m(\pi_{A_Y}, \phi_Y)$. Now, since $\pi_{A_Y}(\lambda)|\pi_A(\lambda)$, $\Phi_m(\pi_{A_Y}, \phi_Y) \geq \Phi_m(\pi_A, \phi_Y)$ and since $\phi_Y \leq \min\{\phi, n\}$, $\Phi_m(\pi_A, \phi_Y) \geq \Phi_m(\pi_A, \min\{\phi, n\})$. □

To simplify the statement of the proposition and since in general A_Y is unknown, we have given the probability in terms of the structure of A, but as shown by the proof, A_Y plays the key role. Note also that $D_Y(\lambda) = D_{\mathcal{W}}(\lambda)$ implies that $m \geq \phi_Y$ which is thus a necessary condition on m.

8.2. **Using Coppersmith's analysis.** We first focus on the probability that, for a random Y, the column degrees of $D_Y(\lambda)$ are not "too far" from the generic situation. For matrices A and Y, define

$$\mathcal{K}_Y(\delta_r) = \left[Y, AY, \ldots, A^{\delta_r-2}Y, A^{\delta_r-1}Y^{(1)}, \ldots, A^{\delta_r-1}Y^{(n-\tau)}\right]$$

where $Y^{(l)}$, $1 \leq l \leq n$, denotes the l-th column of Y. The column degrees of $D_Y(\lambda)$ are strongly related to the rank of $\mathcal{K}_Y(\delta_r)$. If rank $\mathcal{K}_Y(\delta_r) > \nu - \Delta$ then the determinantal degree of $D_Y(\lambda)$ is also strictly greater than $\nu - \Delta$, and the column degrees of $D_Y(\lambda)$ must be less than $\delta_r + \Delta$. If this is true with Δ small, $D_Y(\lambda)$ should be viewed as *nearly generic*. Next lemma is an interpretation of lemma 6.2 in [9] to fit the current context. We formulate it using the function $\Theta_n(f, \phi)$ defined over K =GF(q) for a polynomial $f(\lambda)$ in $K[\lambda]$, a positive integer n and a matrix A whose Frobenius form has ϕ companion blocks:

$$\Theta_n(f, \phi) = 1 + \sum_{g|f} q^{\nu - n \deg g - \text{rank } g(A)} \tag{8.4}$$

where the sum is taken over all the factors of $f(\lambda)$. This definition is slightly different from the original one in [9]. The reader may refer to the latter article concerning the properties of the function that remain unchanged. We propose lower bounds for $\Theta_n(f, \phi)$ in appendix B.

Lemma 8.3. *Let A be a $N \times N$ matrix over K with minimal polynomial $\pi_A(\lambda)$. The matrix Y is chosen at random with $n \geq \phi$ columns. If K =GF(q) then*

$$\text{Prob}_Y\left\{\text{rank } \mathcal{K}_Y(\delta_r) > \nu - \Delta\right\} \geq 1 - \Theta_n(\pi_A, \phi) q^{-\Delta}.$$

Proof. Following [9] we relate the probability of failure – $D_Y(\lambda)$ has small determinantal degree – to the existence of small degree polynomials $g_1(\lambda), \ldots, g_n(\lambda)$ such that

$$g_1(A)Y^{(1)} + g_2(A)Y^{(2)} + \ldots + g_n(A)Y^{(n)} = 0. \tag{8.5}$$

Intuitively, $D_Y(\lambda)$ is a generating polynomial for $\{A^iY\}_{i=0}^{\infty}$, the columns of $Y.D_Y(\lambda)$ are zero:

$$D_{1,j}(A)Y^{(1)} + D_{2,j}(A)Y^{(2)} + \ldots + D_{n,j}(A)Y^{(n)} = 0,$$

for $1 \leq j \leq n$, where the $D_{i,j}(\lambda)$'s denote the entries of $D_Y(\lambda)$. But since $D_Y(\lambda)$ is the minimal polynomial, its entries must be the lowest degree polynomials such that the above relation is true. More precisely, $D_Y(\lambda)$ satisfies (6.1) and (6.2) *i.e.* rank $\mathcal{K}_Y(\delta_r) = \nu$, if and only if

- **(C):** the trivial collection, $g_i(\lambda) = 0$ for $1 \leq i \leq n$, is the only collection of $n-\tau$ polynomials $g_1(\lambda), \ldots, g_{n-\tau}(\lambda)$ of degrees at most $\delta_r - 1$ and of τ polynomials $g_{n-\tau+1}(\lambda), \ldots, g_n(\lambda)$ of degrees at most $\delta_r - 2$ such that (8.5) holds.

For the "if part", it is easily seen that if $D_Y(\lambda)$ satisfies (6.1) and (6.2) then condition (C) is true. Conversely, for the "only if" part, on the one hand (C) ensures that $D_Y(\lambda)$ has no column of degree less than $\delta_r - 1$. On the other hand, it implies that one can find τ columns of degree $\delta_r - 1$ with pivots as expected (otherwise a collection that violate (C) is exhibited among the columns of degree $\delta_r - 1$) and in the same way, that one can also find $n - \tau$ columns of degree δ_r with pivots as expected.

As done in [9] we may now bound the expected value E_Y of the number $W(n)$ of "wrong collections". We mean the number of choices of collections that satisfy (8.5) but violate (C), plus one for the trivial choice:

$$(8.6)\quad E_Y = \operatorname{Exp}_Y \# \left\{ \{g_i\}_i ; \Sigma_i g_i(A) Y^{(i)} = 0 \right\} = \sum_{g_i} \operatorname{Prob}_Y \left\{ \Sigma_i g_i(A) Y^{(i)} = 0 \right\}.$$

To bound the above sum of probabilities, for any non-trivial collection we consider the polynomial

$$(8.7)\qquad g(\lambda) = \gcd\left(\pi_A(\lambda), g_1(\lambda), \ldots, g_n(\lambda)\right).$$

Given such a $g(\lambda)$, there are $q^{\delta_r - 1 - \deg g}$ possible $g_i(\lambda)$ of degree at most $\delta_r - 2$ such that $g(\lambda)$ divides $g_i(\lambda)$, thus there are at most $q^{n(\delta_r - 1 - \deg g)}$ collections $\{g_i(\lambda)\}_{i=1,\ldots,n}$ of degree at most $\delta_r - 2$ that satisfy (8.7). Adding the collections whose first $n - \tau$ polynomials are of degree $\delta_r - 1$ this gives $q^{n(\delta_r - 1 - \deg g)} + (q^{n-\tau} - 1)q^{n(\delta_r - 1 - \deg g)}$ thus $q^{\nu - n \deg g}$ collections of degrees lower than the generic degrees. For a given collection, the probability that a random Y will satisfy

$$g_1(A)Y^{(1)} + g_2(A)Y^{(2)} + \ldots + g_n(A)Y^{(n)} = 0,$$

is the same that the probability that a random vector y will satisfy $g(A)y = 0$. It is $q^{-\operatorname{rank} g(A)}$. Then using (8.6) we get:

$$\operatorname{Exp}_Y \{W(n)\} \le 1 + \sum_{g|\pi_A} q^{\nu - n \deg g - \operatorname{rank} g(A)} = \Theta_n(\pi_A, \phi).$$

Now, we use the fact that $W(n)$ is related to the rank of $\mathcal{K}_Y(\delta_r)$ so that, $W(n) = q^{\nu - \operatorname{rank} \mathcal{K}_Y(\delta_r)}$. This finally leads to

$$\operatorname{Exp}_Y \left\{ q^{-\operatorname{rank} \mathcal{K}_Y(\delta_r)} \right\} \le \Theta_n(\pi_A, \phi) q^{-\nu}$$

and

$$\operatorname{Prob}_Y \{\operatorname{rank} \mathcal{K}_Y(\delta_r) \le \nu - \Delta\} \le \Theta_n(\pi_A, \phi) q^{-\Delta},$$

which concludes the proof. □

We will also use the same result on the left:

Proposition 8.4. *Let A be a $N \times N$ matrix over K =GF(q), let X and Y be chosen at random with m rows and n columns. Let $\delta_l = \lceil \nu/m \rceil$ and*

$$\mathcal{K}_{XY}(\delta_l) = \left[XP_Y, XP_YA_Y, \ldots, XP_YA_Y^{\delta_l - 1} \right].$$

If $m \ge \min\{\phi, n\}$ then with probability no less than $1 - \Theta_m(\pi_A, \min\{\phi, n\})q^{-\Delta}$ we have rank $\mathcal{K}_{XY}(\delta_l) > N_Y - \Delta$.

Proof. We use the same arguments as in proposition 8.2. Since X is uniform random, so is XP_Y in $\mathcal{M}_{m,N_Y}(K)$. We keep the notations for ϕ_Y and for $\pi_{A_Y}(\lambda)$. Since $m \ge \phi_Y$, by lemma 8.3 we get that

$$\operatorname{Prob}_X \{\operatorname{rank} \mathcal{K}_{XY}(\lceil N_Y/m \rceil) > N_Y - \Delta\} \ge 1 - \Theta_m(\pi_{A_Y}, \phi_Y)q^{-\Delta}.$$

But rank $\mathcal{K}_{XY}(\delta_l) \ge$ rank $\mathcal{K}_{XY}(\lceil N_Y/m \rceil)$ and using that $\pi_{A_Y}(\lambda)$ divides $\pi_A(\lambda)$ and $\phi_Y \le \min\{\phi, n\}$, we conclude with $\Theta_m(\pi_{A_Y}, \phi_Y) \le \Theta_m(\pi_A, \phi)$. □

The remarks after proposition 8.2 remain valid here.

9. Probability of success

The block algorithm uses $Z = AY$ rather than Y. As in §6 we thus define A^* to be a restriction of A to its range space. We let ϕ^* be the number of blocks of the Frobenius form of A^* and ν^* be the dimension of the first ϕ^* or n blocks. Clearly, all previously seen results with A, ϕ and ν, can be applied with A^*, ϕ^* and ν^*.

We do not know whether the block algorithm – as stated in [9] and in [18] – produces a solution with good probability for any A, m and n. In the following, we will work under the assumption that m is at least greater than $\min\{\phi^*, n\}$, fortunately this is not too restrictive in most cases. Besides, there are two ways to bypass this difficuly.

To work with any matrix A: one may simply choose m greater than n. Indeed, even if ϕ^* is large, the blocking factor n on the right always actually limits the number of blocks to $\min\{\phi^*, n\}$. This can be used to apply theorem 9.1 below for computing at least one solution of the system.

To work with any given m and n: the matrix A is either assumed or forced to be non pathological. Coppersmith has assumed $m \geq n > \phi^*$. Using the same notations, Kaltofen has proposed a preconditioning of A to ensure $\phi^* = 1$. This type of assumption is required for computing several solutions as an application of corollary 9.5.

9.1. Over small cardinality fields.

By remark 4.5, to find *at least one solution* to the linear system $Aw = 0$, mainly the choice of X and the shift parameter Δ are relevant:

Theorem 9.1. *Let A be a $N \times N$ singular matrix over $K = GF(q)$. The matrices $X \in \mathcal{M}_{m,N}(K)$ and $Y \in \mathcal{M}_{N,n}(K)$ are chosen at random. Suppose that $m \geq \min\{\phi^*, n\}$ and let $V = |\ker A|$. If w is computed by the block algorithm of §3.1 with shift parameter Δ, then $\mathrm{Prob}_{X,Y}\{w \neq 0, Aw = 0\}$ is greater than*

$$(\Phi_m(\pi_A, \min\{\phi^*, n\}) - \Theta_m(\pi_A, \min\{\phi^*, n\}) q^{-\Delta}) (1 - 1/V). \tag{9.1}$$

Proof. We begin by bounding the probability of having $Aw \neq 0$. Following the notations used in §3.1, we have $\delta_l = \lceil N/m \rceil$ and $\delta_r = \lfloor N/n \rfloor$. We simultaneously apply proposition 8.2 – which controls the dimension ν^* of the corresponding space – and proposition 8.4 – which indicates whether the target dimension is reached. As done for corollary 6.5, these propositions are applied with $Z = AY$ rather than with Y. We denote by N_Z the dimension of $\mathrm{span}(Z, AZ, \ldots)$. If X is such that $\mathrm{rank}\, \mathcal{K}_{XZ}(\delta_l) = N_Z - \Delta$ and if $D_Z(\lambda) = D_{\mathcal{W}}(\lambda)$ then $\mathrm{rank}\, \mathcal{K}_{XZ}(\delta_l + \Delta)$ must be N_Z. Thus by lemma 4.4, any vector in the kernel of the corresponding block-Hankel matrix $M(\delta_l + \Delta, \delta_r + 1)$ must be a vector generating polynomial and must provide, by definition, a w such that $Aw = 0$. Note that – by remark 4.5 – the kernel of $M(\delta_l + \Delta, \delta_r + 1)$ is not trivial. Now, by proposition 8.2 and proposition 8.4 we know that

$$\mathrm{Prob}_X\{Aw \neq 0\} \leq 1 - \Phi_m(\pi_A, \min\{\phi^*, n\}) + \Theta_m(\pi_A, \min\{\phi^*, n\}) q^{-\Delta}.$$

Finally, by an argument of Coppersmith [9] we know that the probability – on the choice of Y – that $w \neq 0$ is more than $(1 - 1/V)$. □

The condition $m \geq n$ and consequently the condition $m \geq \min\{\phi^*, n\}$ of theorem 9.1 are harmless if A^t times a vector can be computed with no loss of efficiency. Indeed, in that case, on can find w^t such that $w^t A^t = 0$ using a left version of the block algorithm. From a theoretical point of view, on may use Tellegen's theorem [32] which states that an algorithm computing A times a vector can be converted to one for A^t times a vector.

However, in the general case we have to make a slight additional restriction especially if the ground field is GF(2). By lemma 11.5, we have a reasonable bound for $\Theta_m(\pi_A, \min\{\phi^*, n\})$ only if $m \geq \min\{\phi^*, n\}+2$. If this is true, by taking $\Delta \geq 8$, the probability of success given by theorem 9.1 is quite low but greater than a fixed ϵ_0. Next corollary is easily obtained from the bounds given in appendix B.

Corollary 9.2. *If $m \geq \min\{\phi^*, n\} + 2$ and $\Delta \geq 8$ then the probability of success is always greater than $\epsilon_0 = 0.03$.*

Fortunately – to explain the good practical behaviour of the algorithm – the probability may be much greater, even over GF(2). Especially when $m >> \phi^*$:

Corollary 9.3. *If $m \geq 4\phi^*$, $\Delta \geq 8$, $V \geq 4$ then the probability of success is greater than 0.6.*

Remark 9.4. . It has been experimentally observed in [19, 26] that blocking may amplify the success probability. This can be easily explained by looking at the values taken by (9.1). Intuitively, the more one uses blocking vectors the more a block-Krylov subspace of dimension ν^* is easy to get (see also Wiedemann about the success of his algorithm 1 in [39]). If, for instance, $\phi^* = \phi_0$ is a constant. When m and n increase ($m - \phi^*$ and $n - \phi^*$ increase), $\Phi_m(\pi_A, \phi_0)$ increases and tends to 1, and $\Theta_m(\pi_A, \phi_0)$ decreases and tends to 2. The probability in theorem 9.1 increases and can be made arbitrarily close to $(1 - 1/V)$.

We may also consider a more general problem than finding at least one solution to the linear system. One may wish for a family of vector generating polynomials and further, for the matrix minimum generating polynomial for the current matrix sequence. This may be used to compute, as actually done in [9], several independant vectors in the kernel of A when they exist, or a multiple of the minimum polynomial of A. The proof of next result – for small fields – reduces to apply the proof of theorem 9.1 on both matrices X and Y. These two matrices now play the same role.

Corollary 9.5. *Let A be a $N \times N$ matrix over K =GF(q), let $\delta_l = \lceil N/m \rceil$ and $\delta_r = \lceil N/n \rceil$. Suppose that the left blocking matrix X with m rows and the right one Y with n columns are chosen at random over K. Suppose that $n \geq \phi$ and $m \geq n$. Then the minimum generating polynomial $D_Y(\lambda)$ for the sequence $\{A^iY\}_{i=0}^{\infty}$ can be computed from the kernel of $M(\delta_l + \Delta, \delta_r + \Delta + 1)$ with probability no less than*

$$\left(\Phi_m(\pi_A, \phi) - \Theta_m(\pi_A, \phi)q^{-\Delta}\right)\left(\Phi_n(\pi_A, \phi) - \Theta_n(\pi_A, \phi)q^{-\Delta}\right).$$

Proof. From lemma 8.1, with probability less than $1 - \Phi_n(\pi_A, \phi)$ the dimension of $< Y >$ is less than ν. From lemma 8.3, with probability less than $\Theta_n(\pi_A, \phi)q^{-\Delta}$ the rank of $\mathcal{K}_Y(\delta_r)$ is lower than $\nu - \Delta$. Thus

$$\text{Prob}_Y\left\{\text{rank } \mathcal{K}_Y(\delta_r + \Delta) = \nu\right\} \geq \Phi_n(\pi_A, \phi) - \Theta_n(\pi_A, \phi)q^{-\Delta}.$$

Using the same arguments on the left for X (see the proof of theorem 9.1) and applying lemma 4.6, the claim is proven. □

9.2. **Large fields – Generalization of Kaltofen's analysis.** For large fields, another technique can be used. We follow the work of [20, 21, 18]. These ideas have been successfully applied to singular matrices A whose minimal polynomial has degree $\deg \pi_A(\lambda) = \mathrm{rank}(A) + 1$. Using the generalization of corollary 6.5, we get for any matrix:

Theorem 9.6. *Let A be a $N \times N$ singular matrix over K and let $m \geq \min\{\phi^*, n\}$ Suppose that X with m rows and Y with n columns are chosen at random over K. If w is a vector computed by the block algorithm of §3.1 with $\Delta = 0$ then*

$$\mathrm{Prob}_{X,Y}\,\{w \neq 0 \text{ and } Aw = 0\} \geq 1 - (2N-1)/|K|.$$

Proof. We take $\delta_l = \lceil N/m \rceil$ and $\delta_r = \lfloor N/n \rfloor$ and $Z = AY$. Let N_Z be the dimension of $\mathrm{span}(Z, AZ, \ldots)$. If the associated Hankel matrix $M(\delta_l, \delta_r)$ is of rank N_Z (we are going to ensure this condition which is stronger condition than necessary) then by lemma 4.6, any vector in the kernel of $M(\delta_l, \delta_r + 1)$ provides a w such that $Aw = 0$. By corollary 6.5 we know that this is true – with $N_Z = \nu^*$ – for matrices $\mathcal{X}$ and $\mathcal{Y}$ which entries are indeterminates $\xi_{j,k}$, $1 \leq j \leq m$, and $\upsilon_{k,l}$, $1 \leq l \leq n$, with $1 \leq k \leq N$. If

$$D^*\left(\xi_{1,1}, \ldots, \xi_{m,N}, \upsilon_{1,1}, \ldots, \upsilon_{N,n}\right)$$

is a minor of rank ν^* of the associated Hankel matrix, then the probability that $Aw = 0$ is greater than the probability of hitting matrices X and Y which entries do not form a root of D^*. By corollary 1 of Schwartz [34] we know that the probability of hitting a zero is less than

$$\deg(D^*)/|K| \leq 2\nu^*/|K| \leq 2N/|K|.$$

Summing this with the probability that $w = 0$ – at most $1/V$ as seen at theorem 9.1 – we conclude the proof. □

Over large fields, as shown in effect during the latter proof, degrees are actually generic both on the right and on the left. A minimum generating matrix polynomial with generic degrees, can thus also be computed from the kernel of $M(\delta_l, \delta_r + 1)$, with the same probability of success than for at least one generating vector polynomial.

10. Alternative variant of the block algorithm

For the sake of completeness of the generalization of Wiedemann's work, this last section is devoted to a matrix version of another algorithm. Indeed, using lemma 5.4 it is not difficult to give an adaptation, in the matrix case, of Wiedemann's deterministic algorithm (see §II of [39]). Again, since this is not the purpose of this paper to study the way minimal generating polynomials for a sequence are computed, the corresponding part of the algorithm is not given.

We define the matrices $U_1, U_2, \ldots, U_{\lfloor N/m \rfloor}$ by dividing the set of the rows of the identity matrix of dimension N, into $\lfloor N/m \rfloor$ subsets of m or $m+1$ rows. Selecting X to be successively $U_1, U_2, \ldots, U_{\lfloor N/m \rfloor}$, we can give a deterministic block algorithm to compute a generating matrix polynomial for a sequence $\{A^iY\}_{0 \leq i \leq L-1}$:

Algorithm. Block Wiedemann's deterministic algorithm.

Input: *A a $N \times N$ matrix over K, Y a $N \times n$ matrix over K and L a non-negative integer.*

Step 1. *Compute A^iY, $0 \leq i \leq L-1$.*

Step 2. *Let $k = 1$ and $D_Y(\lambda) = I$.*
Step 3. *Compute the sequence $\{U_k A^i Y\}_{0 \le i \le L-1}$ from the data produced in step 1.*
Step 4. *Apply $D_Y(\lambda)$ to this sequence.*
Step 5. *Compute a generating matrix polynomial $D(\lambda)$ for this latter sequence.*
Step 6. *Set $D_Y(\lambda) = D_Y(\lambda)D(\lambda)$.*
Step 7. *Set $k = k + 1$ and if $k \le \lfloor N/m \rfloor$ then go to step 3.*
Output: $D_Y(\lambda)$.

Such an algorithm could be also used to solve a linear system $Aw = 0$. In a subsequent work we will try to know whether it can be time-competitive with respect to the algorithm of §3.1.

11. Conclusion

Our approach has been influenced by matrix polynomial theory, where many operations on scalar polynomials are generalized. Especially over GF(2), our contribution is based on a very accurate tuning of parameters (m, n, Δ). The problem is solved because some constraints are relaxed in a first step (see lemma 8.3 concerning dimensions of Krylov subspaces) while others seem to be inevitable (see the comments after theorem 9.1 about m and n). Even if the experiments tend to confirm these facts, a problem is to know whether or why the remaining assumptions are necessary. In passing, one question is to know whether is particular the structure of matrices arizing in factorization algorithms. This could corroborate our analysis.

Future work will concern some extensions of our study. We have focused on a particular block algorithm, the other block ones cited in the introduction may also be considered. We have focused on finite fields, our results may be used for other computable fields.

Appendix A – Some facts from linear algebra

The reader may refer to [14] or to [16] for the following matrix basic facts.

Theorem 11.1. *Any $N \times N$ matrix A with entries in K is similar to a unique block-companion matrix*

$$F = P^{-1}AP = \text{ diag } (C_1, C_2, \dots, C_\phi) \in \mathcal{M}_N(K),$$

where C_i is associated to some $s_i(\lambda)$ in $K[\lambda]$, and $s_i(\lambda)$ is a factor of $s_{i-1}(\lambda)$ for $2 \le i \le \phi$.

The matrix F is called the *Frobenius normal form* of A. The polynomials $s_1(\lambda), \dots, s_\phi(\lambda)$ are called the *invariant factors* of A.

Theorem 11.2. *Let A and Y be a $N \times N$ and a $N \times n$ matrix over K. There exists a transformation matrix P such that*

$$P^{-1}AP = [P_Y \;\; P_2]^{-1} A \, [P_Y \;\; P_2] = \begin{bmatrix} A_Y & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

where A_Y is a restriciton of A to $\text{span}(Y, AY, A^2Y, \dots)$. *The transformation P may be chosen such that A_Y is shift-Hessenberg* [2] *or polycyclic form* [31]. *This form is* *block-triangular:*

$$A_Y = \begin{bmatrix} C_1 & T_{1,2} & T_{1,3} & \dots & T_{1,\phi_Y} \\ 0 & C_2 & T_{2,3} & \dots & \vdots \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & T_{\phi_Y-1,\phi_Y} \\ 0 & \dots & \dots & 0 & C_{\phi_Y} \end{bmatrix}, \; 1 \leq \phi_Y \leq n,$$

where each diagonal block is companion and each upper-triangular block has non-zero entries only in its last column. The structure of $P^{-1}Y$ corresponds to the structure of $P^{-1}AP$. We mean that

$$P^{-1}Y = \begin{bmatrix} \bar{Y} \\ 0 \end{bmatrix},$$

and the first ϕ_Y columns of $\bar{Y}$ form a modular cyclic basis with respect to A_Y. The j-th column, $1 \leq j \leq \phi_Y$, is the k_j-th canonical vector of K^N, k_j being the first row index of block C_j.

Theorem 11.3. *Any $n \times n$ matrix $D(\lambda)$ with entries in $K[\lambda]$ is equivalent to a unique diagonal matrix*

$$S(\lambda) = U(\lambda)D(\lambda)V(\lambda) = \text{ diag } (s_1(\lambda), \dots, s_\phi(\lambda), 1, \dots, 1, 0, \dots, 0) \in \mathcal{M}_n(K[\lambda]),$$

where $U(\lambda)$ and $V(\lambda)$ are unimodular matrices in $\mathcal{M}_n(\lambda)$ and where the $s_i(\lambda)$'s in $K[\lambda]$ are such that $s_i(\lambda)$ is a factor of $s_{i-1}(\lambda)$ for $2 \leq i \leq \phi$.

The polynomials $s_1(\lambda), \dots, s_\phi(\lambda)$ are called the *non-unity invariant factors* of $D(\lambda)$. For any irreducible polynomial $g(\lambda)$ of degree d in $K[\lambda]$, a rational Jordan block associated to $g(\lambda)$ is, for some integer m, a block matrix

$$J_g = \begin{bmatrix} C_g & I & 0 & \dots \\ 0 & C_g & \ddots & 0 \\ \vdots & \ddots & \ddots & I \\ 0 & \dots & 0 & C_g \end{bmatrix} \in \mathcal{M}_{md}(K),$$

where C_g is the companion matrix of $g(\lambda)$.

Theorem 11.4. *Any $N \times N$ matrix A with entries in K is similar to a unique block-diagonal matrix (up to the order of the blocks)*

$$J = Q^{-1}AQ = \text{ diag } \left(J_{g_1}^{(1)}, \dots, J_{g_1}^{(\phi_1)}, \dots, \dots, J_{g_p}^{(1)}, \dots, J_{g_p}^{(\phi_p)} \right) \in \mathcal{M}_N(K),$$

where each $J_{g_j}^{(l)}$ is a rational Jordan block associated to $g_j(\lambda)$, and the $g_j(\lambda)$'s, $1 \leq j \leq p$, are the irreducible factors in $K[\lambda]$ of the minimal polynomial of A.

Appendix B – Bounds on $\Theta_n(f, \phi)$ and $\Phi_n(f, \phi)$

Next two results give lower bounds for the values of the functions $\Theta_n(f, \phi)$ and $\Phi_n(f, \phi)$ defined by identities (8.1) and (8.4).

Lemma 11.5. *For $f(\lambda) | \pi_A(\lambda)$ of degree $N > q$ and for $n \geq \phi + 1$:*

$$\Theta_n(f, \phi) < \begin{cases} 14 \log_q N \text{ if } n = \phi + 1, \\ 1 + \exp\left(1/q^{n-\phi-2} + 1/q^{2(n-\phi)-3}\right), n \geq \phi + 2. \end{cases}$$

Proof. We first look at the values taken by

$$\theta_n(g,\phi) = \nu - n\deg g - \text{rank } g(A)$$

depending on $g(\lambda)$. Let $g_i(\lambda)$ and d_i denote the i-th irreducible factor of $f(\lambda)$ and its degree. Also assume that $g_i(\lambda)$ is a factor of $\phi_i \le \phi$ invariant factors of A *i.e.* is involved in ϕ_i blocks of the Frobenius form of A. For any given $g_i(\lambda)$, we have

$$\theta_n(g_i,\phi) = \nu - nd_i - (\nu - \phi_i d_i) = -(n-\phi_i)d_i \le -(n-\phi)d_i.$$

In particular, the greatest possible value of $\theta_n(g_i,\phi)$ is $-d_i$, it corresponds to the case where $g_i(\lambda)$ is a factor of $\phi = n-1$ invariant factors of A. For two polynomials $g_i(\lambda)$ and $g_j(\lambda)$ such that $g_i^{k_i}(\lambda)$ and $g_j^{k_j}(\lambda)$ are factors of $f(\lambda)$ we obtain in the same way:

$$\begin{aligned}\theta_n(g_i^{k_i}g_j^{k_j},\phi) &\le \nu - n(k_i d_i + k_j d_j) - (\nu - \phi_i k_i d_i - \phi_j k_j d_j)\\ &\le -(n-\phi)(k_i d_i + k_j d_j).\end{aligned} \tag{11.1}$$

The greatest possible value of $\theta_n(g_i^{k_i}g_j^{k_j},\phi)$ is $-k_i d_i - k_j d_j$, it is reached if $g_i(\lambda)$ and $g_j(\lambda)$ are factors of $\phi = n-1$ invariant factors of A. By definition of $\Theta_n(f,\phi)$,

$$\Theta_n(f,\phi) = 1 + \sum_{g|f} q^{\theta_n(g,\phi)}.$$

Using that $\theta_n(1,\phi) = 0$ and using identity (11.1), $\Theta_n(f,\phi)$ is thus bounded by

$$1 + 1 + \sum_{g = g_{i_1}^{k_1}\cdots g_{i_j}^{k_j} | f} q^{-(n-\phi)(k_1 d_{i_1} + \ldots + k_j d_{i_j})}$$

and further, by

$$1 + 1 + \sum_{i_1,\ldots,i_j}\ \sum_{k_1,\ldots,k_j} q^{-(n-\phi)(k_1 d_{i_1} + \ldots + k_j d_{i_j})}$$

where the latter double sum is taken over all the sets of j indices $i_1,\ldots,i_j$ of factors of f and of j corresponding multiplicities $k_1,\ldots,k_j$. Thus:

$$\Theta_n(f,\phi) \le 1 + \prod_i \prod_{k=n-\phi}^{\infty} (1 + q^{-kd_i}). \tag{11.2}$$

We now bound the logarithm of above double product:

$$\log\left(\prod_i \prod_{k=n-\phi}^{\infty}(1+q^{-kd_i})\right) \le \sum_i \sum_{k=n-\phi}^{\infty} q^{-kd_i}. \tag{11.3}$$

We may proceed using the computations done in [39] for the Euler's Phi function for polynomials over finite fields.

We first address the case $n - \phi \ge 2$. The number of irreducible factors of degree i is at most q^i/i thus

$$\begin{aligned}\textstyle\sum_i q^{-kd_i} &< (q^1/1)q^{-k} + (q^2/2)q^{-2k} + \ldots\\ &= \log\left(\frac{1}{1-1/q^{k-1}}\right) = -\log(1 - 1/q^{k-1}).\end{aligned} \tag{11.4}$$

Since if $n-\phi \geq 2$ then $k>1$ and since $-\log(1-x) \leq x+x^2$ when $0 \leq x \leq 1/2$ we bound the right-hand side of (11.3) by

$$\begin{array}{ll} \sum_i \sum_k q^{-kd_i} & < \sum_{k=n-\phi}^{\infty} (1/q^{k-1} + 1/q^{2(k-1)}) \\ & \leq 1/q^{n-\phi-2} + 1/q^{2(n-\phi)-3}. \end{array} \tag{11.5}$$

Using this latter bound in (11.3) then in (11.2), we obtain as announced:

$$n-\phi \geq 2: \ \Theta_n(f,\phi) < 1 + \exp\left(1/q^{n-\phi-2} + 1/q^{2(n-\phi)-3}\right).$$

We terminate the proof by considering the case $n-\phi = 1$. Using the claim for $n-\phi=2$ we get:

$$\sum_i \sum_{k=1}^{\infty} q^{-kd_i} \leq \left(\sum_i q^{-d_i}\right) + (1+1/q)$$

and using proposition 3 in [39]:

$$\sum_i q^{-d_i} < 1 + \log\log_q N,$$

thus,

$$\sum_i \sum_{k=1}^{\infty} q^{-kd_i} \leq 2 + 1/q + \log\log_q N.$$

Using (11.3) and (11.2) we finally obtain

$$n-\phi = 1: \ \Theta_n(f,\phi) \leq 1 + \exp(2+1/q)\log_q N < 14\log_q N$$

which is the claim. □

Lemma 11.6. *For $f(\lambda)$ of degree $N>q$ and $n \geq \phi$:*

$$\Phi_n(f,\phi) > \begin{cases} 1/(45\log_q N) \text{ if } n=\phi, \\ 1/30 \text{ if } \mu = \lfloor n/\phi \rfloor = 1 \text{ and } \rho \geq 1, \\ 1 - 1/q^{\mu-1} - 1/q^{2(\mu-1)} \text{ if } \mu = \lfloor n/\phi \rfloor \geq 2. \end{cases}$$

Proof. For the second term of the product we have:

$$\begin{array}{ll} \prod_i \prod_{k=2}^{\infty} (1-q^{-k\mu d_i}) & \geq (\prod_i (1-q^{-2\mu d_i}))(\prod_i \prod_{k=3}^{\infty} (1-q^{-k\mu d_i})) \\ & > (1-\sum_i q^{-2\mu d_i})(1-\sum_i \sum_{k=3}^{\infty} q^{-k\mu d_i}), \end{array}$$

which is greater, using identities (11.4) and (11.5) of the proof of lemma 11.5, than:

$$\left(1+\log(1-1/q^{2\mu-1})\right)\left(1-1/q^{3\mu-2} - 1/q^{6\mu-3}\right). \tag{11.6}$$

The claim of the lemma is obtained by noticing that this latter expression is always greater than $1/9$ and is greater than $1-1/q^{2\mu-2}$ if $\mu \geq 2$.

For the first term of the main product, if $\rho+\mu=1$ *i.e.* if $n=\phi$ then from theorem 3.15 of [15]:

$$\prod_i (1-q^{-d_i}) > 1/(5\log_q N),$$

which gives with the lower bound for (11.6):

$$n=\phi: \ \Phi_n(f,\phi) > 1/(45\log_q N).$$

If $\rho+\mu>1$ then using (11.4)

$$\prod_i (1-q^{-(\rho+\mu)d_i}) > 1 + \log(1-1/q^{\rho+\mu-1}),$$

we have

$$\mu = 1, \rho \geq 1 : \ \Phi_n(f,\phi) > (1 + \log(1 - 1/2))/9 > 1/30.$$

If $\mu \geq 2$, using (11.5) and the previously given bound for (11.6) then we finally get:

$$\begin{aligned}\Phi_n(f,\phi) \quad &> \left(1 - 1/q^{\rho+\mu-1} - 1/q^{2(\rho+\mu-1)}\right)\left(1 - 1/q^{2\mu-2}\right) \\ &> 1 - 1/q^{\rho+\mu-1} - 1/q^{2(\rho+\mu-1)} > 1 - 1/q^{\mu-1} - 1/q^{2(\mu-1)}.\end{aligned}$$

□

Acknowledgments

Grateful thanks to Erich Kaltofen for his valuable questions.

References

1. A.C. Antoulas, *On recursiveness and related topics in linear systems*, IEEE Trans. Automat. Control **AC-31** (1986), no. 12, 1121–1135.
2. D. Augot and P. Camion, *The minimal polynomials, characteristic subspaces, normal bases and the Frobenius form*, Tech. Report 2006, INRIA France, August 1993.
3. B. Beckermann and G. Labahn, *A uniform approach for the fast computation of matrix-type Padé approximants*, SIAM J. Matrix Anal. Appl. **15** (1994), no. 3, 804–823.
4. ———, *Recursiveness in matrix rational interpolation problems*, Tech. Report Publication ANO 357, Université de Lille France, 1996.
5. R.R. Bitmead and B.D.O. Anderson, *Asymptotically fast solution iof toeplitz and related systems of linear equations*, Linear Algebra and its Appl. **34** (1980), 103–116.
6. A. Bultheel and M. Van Barel, *A matrix Euclidean algorithm and the matrix minimal Padé approximation problem*, Continued Fractions and Padé Approximants (C. Brezinski, ed.), North-Holland, 1990, pp. 11–51.
7. W.A. Coppel, *Matrices of rational functions*, Bull. Austral. Math. Soc. **11** (1974), 89–113.
8. D. Coppersmith, *Solving linear equations over GF(2): block Lanczos algorithm*, Linear Algebra and its Applications **192** (1993), 33–60.
9. ———, *Solving homogeneous linear equations over GF(2) via block Wiedemann algorithm*, Math. Comp. **62** (1994), no. 205, 333–350.
10. P. Delsarte, Y.V. Genin, and Y.G. Kamp, *A generalization of the Levinson algorithm for Hermitian Toeplitz matrices with any rank profile*, IEEE Internat. Conf. Acoust. Speech Signal Process. **33** (1985), 964–971.
11. J.L. Dornstetter, *On the equivalence between Berlekamp's and Euclid's algorithms*, IEEE Trans. Inform. Theory **33** (1987), 428–431.
12. W. Eberly and E. Kaltofen, *On randomized Lanczos algorithms*, International Symposium on Symbolic and Algebraic Computation, Maui, Hawaii, USA, ACM Press, July 1997.
13. G.D. Forney, *Minimal bases of rational vector spaces, with applications to multivariable linear systems*, SIAM J. Control **13** (1975), 493–520.
14. F.R. Gantmacher, *Théorie des matrices*, Dunod, Paris, France, 1966.
15. M. Giesbrecht, *Nearly optimal algorithms for canonical matrix forms* , Ph.D. thesis, Department of Computer Science, University of Toronto, 1993.
16. N. Jacobson, *Basic Algebra I*, W.H. Freeman and Company, 1974.
17. T. Kailath, *Linear systems*, Prentice Hall, 1980.
18. E. Kaltofen, *Analysis of Coppersmith's block Wiedemann algorithm for the parallel solution of sparse linear systems*, Math. Comp. **64** (1995), no. 210, 777–806.
19. E. Kaltofen and A. Lobo, *Distributed matrix-free solution of large sparse linear systems over finite fields*, High Performance Computing 1996, San Diego, CA (A.M. Tentner, ed.), Society for Computer Simulation, Simulation Councils, Inc, 1996, pp. 244–247.
20. E. Kaltofen and V. Pan, *Processor efficient parallel solution of linear systems over an abstract field*, Proc. 3rd Annual ACM Symposium on Parallel Algorithms and Architecture, ACM-Press, 1991.
21. E. Kaltofen and B.D. Saunders, *On Wiedemann's method of solving sparse linear systems*, Proc. AAECC-9, LNCS 539, Springer Verlag, 1991, pp. 29–38.
22. S.Y. Kung, *Multivariable and multidimensional systems: analysis and design*, Ph.D. thesis, Dpt. of Electrical Engineering, Stanford University, June 1977.

23. B.A. LaMacchia and A.M. Odlyzko, *Solving large sparse linear systems over finite fields*, Advances in Cryptology – CRYPTO'90, Springer LNCS 537 (A.J. Menezes and S.A. Vanstone, eds.), 1991, pp. 109–133.
24. R. Lambert, *Computational aspects of discrete logarithms*, Ph.D. thesis, University of Waterloo, Ontario, Canada, 1996.
25. A.K. Lenstra, H.W. Lenstra, M.S. Manasse, and J.M. Pollard, *The factorization of the ninth Fermat number*, Math. Comp. **61** (1993), 319–349.
26. A. Lobo, *Matrix-free linear system solving and applications to symbolic computation*, Ph.D. thesis, Dept. Comp. Sc., Rensselaer Polytech. Instit., Troy, New York, Dec. 1995.
27. C.C. MacDuffee, *The theory of matrices*, Chelsea, New-York, 1956.
28. J.L. Massey, *Shift-register synthesis and BCH decoding*, IEEE Trans. Inform. Theory **15** (1969), 122–127.
29. P.L. Montgomery, *A block Lanczos algorithm for finding dependencies over GF(2)*, EUROCRYPT'95, Heidelberg, Germany. Springer LNCS 921, 1995, pp. 106–120.
30. M. Morf, *Doubling algorithms for Toeplitz and related equations*, IEEE Internat. Conf. Acoust. Speech Signal Process., Piscataway, NJ, 1980, pp. 954–959.
31. P. Ozello, *Calcul exact des formes de Jordan et de Frobenius d'une matrice*, Ph.D. thesis, Université Scientifique et Médicale de Grenoble, France, 1987.
32. P. Penfield Jr., R. Spencer, and S. Duinker, *Tellegen's theorem and electrical networks*, M.I.T. Press, Cambridge, MA, 1970.
33. V.M. Popov, *Invariant description of linear, time-invariant controllable systems*, SIAM J. Control **10** (1972), 252–264.
34. J.T. Schwartz, *Fast probabilistic algorithms for verification of polynomial identities*, J. ACM **27** (1980), 701–717.
35. M. Van Barel and A. Bultheel, *A general module theoretic framework for vector M-Padé and matrix rational interpolation*, Numerical Algorithms **3** (1992), 451–462.
36. G.C. Verghese and Kailath, *Rational matrix structure*, IEEE Trans. Automat. Control **26** (1981), 434–438.
37. G. Villard, *Computing Popov and Hermite forms of polynomial matrices*, International Symposium on Symbolic and Algebraic Computation, Zurich, Suisse, ACM Press, July 1996, pp. 250–258.
38. ———, *Computing minimum generating matrix polynomials*, 1997, Preprint IMAG Grenoble, France.
39. D. Wiedemann, *Solving sparse linear equations over finite fields*, IEEE Transf. Inform. Theory **32** (1986), 54–62.
40. H.K. Wimmer, *A Jordan factorization for polynomial matrices*, Proceedings of the American Math. Soc. **75** (1979), no. 2, 201–206.
41. W.A. Wolovich, *Linear multivariable systems*, Springer-Verlag, New-York, 1974.

List of Research Reports

These reports can be obtained by anonymous ftp at `ftp.imag.fr` in the directory `/pub/CALCUL_FORMEL` or by the WEB at `http://www-lmc.imag.fr/CF` .

RT 130 : Intégration symplectique d'une classe de systèmes hamiltoniens séparables.
D. Rozier, R. Coleman - 1995

RT 134 : Algebraic approach for quasi-linear differential algebraic equations.
Gabriel Thomas - April 1995

RT 138 : A new simultaneous exclusion algorithm.
R. Coleman - 1995

RT 142 : Constructing real canonical forms of nilpotent Hamiltonian matrices.
R. Coleman - 1995

RT 143 : Algorithme de Newton et algorithme de Newton rationnel.
F. Jung - Octobre 1995

RT 153 : Solution générale et solutions singulières des équations différentielles ordinaires scalaires.
Evelyne Hubert - Novembre 1995

RT 154 : DESIR-II
Eckhard Pfluegel - Février 1996.

RT 155 : Constructing real canonical forms of Hamiltonian matrices with two imaginary eigenvalues.
R. Coleman - Février 1996

RT 157 : Some algorithms for matrix polynomials
Gilles Villard - Mars 1996.

RT 158 : Resolvent and canonical forms of matrices.
Jean Della Dora, Françoise Jung - Mars 1996

RT 159 : Estimation de l'erreur globale pour l'intégration numérique d'équations différentielles ordinaires.
R. Aid - Mars 1996

RR 960 M : Séries de Chebyshev formelles
Luc Rebillard - Septembre 1996.

RT 170 : Geometry of Differential Equations and Formal Integrability Theory
Yann Macutan - Janvier 1997