Further Analysis of Coppersmith's Block Wiedemann Algorithm for the Solution of Sparse Linear Systems (Extended abstract)

G. Villard

LMC-IMAG, B.P. 53
F-38041 Grenoble cedex 9
Gilles.Villard@imag.fr
http://www-lmc.imag.fr/~gvillard

Abstract

We analyse the probability of success of the block algorithm proposed by Coppersmith for solving large sparse systems $Aw = 0$ of linear equations over a field K. It is based on a modification of a scheme proposed by Wiedemann. An open question was to prove that the block algorithm may produce a solution for small finite fields *e.g.* for K =GF(2). Our investigations allow us to answer this question nearly completely. We prove that the input parameters of the algorithm may be tuned such that, for any input system, a solution is computed with high probability for any field. Conversely, for particular input systems, we show that the conditions on the input parameters may be relaxed to ensure the success. We also improve the previous probability measurements in the case of large cardinality fields.

Introduction

This paper is an extended abstract of [30]. Results answering the open question over small fields are only partially proven here. Those corresponding to the improvement over large fields are only reproduced without proof[1]. The randomized method proposed by Coppersmith [6] solves large sparse systems of homogeneous linear equations $Aw = 0$, $w \neq 0$. Throughout the paper A will be a $N \times N$ matrix over the Galois field with q elements K =GF(q) and w a vector of N unknowns. One fundamental application of this problem in computer algebra is integer and polynomial factorization, where such systems arise with N over 200,000 [18, 20, 15]. This has motivated several authors to develop fast symbolic counterpart to numerical iterative methods. The conjugate gradient method has been used in [18], the Lanczos method in [18, 8] and the block Lanczos method in [5, 24].

But up to now, only the probabilistic analysis of Wiedemann [32] 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 [23]. 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 [19]. These various algorithms are very similar: they can be understood in a unified theory [19].

Coppersmith [6] has then proposed a block version of Wiedemann's approach to take advantage of simultaneous operations: either using the machine word over GF(2) or a parallel machine [14]. Coppersmith's algorithm is very powerful in practice [15, 21] but raises some theoretical questions. We are going to partially answer them in this paper.

We refer to §1 for the definitions and to §2 for a detailed presentation of the algorithm. We only consider the method intuitively in this introduction. In the Wiedemann algorithm [32], one computes the lowest degree polynomial that linearly generates the sequence $h_i = xA^iy$, $0 \leq i \leq 2N-1$, where x and y are respectively a row and a column random vector. 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):

$$\begin{cases} \pi_A^y(\lambda) = g_l\lambda^l + g_{l+1}\lambda^{l+1} + \ldots + g_d\lambda^d, 0 < l \leq d, g_l \neq 0, \\ g_lA^ly + g_{l+1}A^{l+1}y + \ldots + g_dA^dy = 0 \in K^N. \end{cases}$$

Taking $w = g_lA^{l-1}y + \ldots + g_dA^{d-1}y$ above relation gives that w is a solution: $Aw = 0$. The modified algorithm of Coppersmith [6] uses a random matrix X with m rows and a random matrix Y with n columns and computes first the sequence of $m \times n$ matrices $H_i = XA^iY, i = 0, \ldots, N/m + N/n + O(1)$. By analogy with the scalar case we will see in §1 that one may define vector or matrix generating polynomials for that sequence. With high probability, such a generating polynomial is also a generating polynomial for the sequence $\{A^iY\}_{i\geq 0}$ and leads to a solution w.

To limit the length of our article we will neither address the subproblem of computing a generating polynomial for a given sequence nor study the cost of the method. The reader may refer to [6, 14], or to [31] for an improved complexity obtained using the algorithm of [1] for the computation of Padé approximants. The method of Coppersmith is randomized, essentially in the sense that a generating polynomial for $\{XA^iY\}_{i\geq 0}$ may not be a generating polynomial for $\{A^iY\}_{i\geq 0}$ and thus may not allow the computation of a

[1] The whole proofs has been sent to the referees. They may be found in [30] on http://www-lmc.imag.fr/~gvillard.

solution w. Our computation of the probability of success – apart from being influenced by the one of Wiedemann – will use two main previous results. The first one, by Coppersmith [6], relies on the notion of *pathological* input matrix A. For matrices having too many eigenvalues with high multiplicities (compared to the blocking factors 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. He observed experimentally that it is sufficient to consider the first $N/m + N/n + O(1)$ terms of the sequence $\{H_i\}_i$. The second result has been given by Kaltofen [14]. Up to a preconditioning and if the field K has enough elements, the algorithm is guaranteed to compute a solution.

Trying to answer the question "why is there no pathological matrix for Wiedemann's algorithm ?" we are led to generalize Wiedemann's approach and to use complementary arguments. We 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 $\Delta = O(1)$ experimentally imposed by Coppersmith, for the number of matrices H_i that must be used. This gives an algorithm that works for any field K and any input system, and thus avoids the notion of pathological matrix. More precisely, theorem 4 will show that

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

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 certain 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 with constant probability for any input matrix. This results in a better probability bound in this case.

After basic definitions in §1 and the presentation of the block algorithm in §2, we will characterize the "good blocking matrices" in §3. We will precisely understand which conditions X and Y must satisfy, so that the sequence $\{XA^iY\}_i$ may be used instead of the sequence $\{A^iY\}_i$. We will then characterize the "generic" behaviour of the algorithm by considering matrices X and Y with indeterminate entries in §4. 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 then divided into two sections. We will give the first technical results in §5. The reader will then find the final theorems in §6. Paragraph 6.1 is devoted to small fields and §6.2 will focus on large cardinality fields. 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.

1 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.

1.1 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 $D(\lambda)$ non singular 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 1 [34]. *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*

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

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

$$XA^iY = H_i \quad \text{for} \quad i = 1, \ldots \tag{2}$$

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

$$\forall i \geq 0 : H_iD_0 + H_{i+1}D_1 + \ldots + H_{i+d}D_d = 0. \tag{3}$$

Any such non singular 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 [6] or in the analysis proposed in [14], 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 (3):

$$\forall i \geq 0 : H_iD_0^{(j)} + H_{i+1}D_1^{(j)} + \ldots + H_{i+d}D_d^{(j)} = 0 \tag{4}$$

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 in [27, 2] for matrix Padé approximants. Clearly, 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 (see [12] for instance) that such a submodule $\mathcal{W}$ has a basis of at most n elements. But since the columns of the diagonal matrix $\text{diag}(\chi_A(\lambda), \ldots, \chi_A(\lambda))$ – where $\chi_A(\lambda)$ is the characteristic 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) can be uniquely determined by choosing a particular representative. As emphasized in [3, 29] several matrix

polynomial normal form can be chosen. In next paragraph we focus on the Popov form which provides a notion of minimal polynomial.

1.2 Minimum generating polynomials

From a complexity point of view it is important to handle relations (3) or (4) of minimal length. We will define minimal bases for $\mathcal{W}$, these will correspond to minimal bases of vector spaces [9, 28]. In addition, to extend the notion of minimal scalar polynomial, we will speak (by an 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 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 [26]. *A matrix $D(\lambda)$ is said to be* column reduced *if rank $[D(\lambda)]_c =$ 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 1 [26, 13]. *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 [9]. *Let $\bar{D}(\lambda)$ 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 $w(\lambda) = \bar{D}(\lambda)v(\lambda)$. Then, $\bar{D}(\lambda)$ is column reduced if and only if it is a* minimal basis *in the sense that any $w(\lambda)$ satisfies:*

$$\deg w(\lambda) = \max_{j \in \underline{v}} \{d_j + \deg v_j(\lambda)\} \qquad (5)$$

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 (5) shows that the corresponding elements of $\mathcal{W}$ are the first linearly independent ones with minimal degrees [9, 13]. This motivates the following definition.

Definition 3 . *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 (4):

$$\forall i \geq 0 : H_i D_{\mathcal{W},0}^{(1)} + H_{i+1} D_{\mathcal{W},1}^{(1)} + \ldots + H_{i+d_1} D_{\mathcal{W},d_1}^{(1)} = 0. \quad (6)$$

2 Coppersmith's block Wiedemann algorithm

Using above terminology, we now give here the Coppersmith's version [6] of Wiedemann's algorithm for the solution of $Aw = 0$. The matrix A is square of dimension N over a field K. We follow the notations of Kaltofen [14] and his variant of the method.

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 §5 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 [21, 15].

Input: *A a $N \times N$ matrix over K and the shift parameter: Δ a nonnegative 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*

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

Then, solutions w such that $Aw = 0$ are constructed from generating vector polynomials for the 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:*

$$XA^iZg_0 + XA^{i+1}Zg_1 + \ldots + XA^{i+d}Zg_d = 0. \qquad (7)$$

for $0 \leq i \leq \delta_l + \Delta - 1$.

With high probability, as we will prove later (see theorem 4 and theorem 5), the left projection by X does not modify the invariants of the sequence and $g(\lambda)$ is a generating vector polynomial for $\{A^iZ\}_i$:

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

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

$$\begin{aligned} &A^{l+1}\left(Yg_l + AYg_{l+1} + \ldots + A^{d-l}Yg_d\right) \\ &\quad = A^lZg_l + A^{l+1}Zg_{l+1} + \ldots + A^dZg_d = 0. \end{aligned} \qquad (8)$$

The left-hand side leads to the 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 4 and theorem 5). From identity (8) 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 $\iota \geq 1$ then $w = A^{\iota-1}\hat{w}$ else $w = 0$.*

The algorithm is randomized concerning two points: identity (8) 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 Characterization of good blocking vectors

To answer the problem, 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 §3.1 we see how the minimal polynomial of $\{A^iY\}_i$ and of $\{XA^iY\}_i$ can coincide. In §3.2 we determine the length of the sequence that must be considered. We finally give an additional technical lemma in §3.3.

3.1 Blocking vectors and Krylov subspaces

Let $< Y >= \text{span}(Y, AY, A^2Y, \ldots)$ and denote by $A_{<Y>}$ the restriction to $< Y >$ of the linear operator associated to the matrix A. The nonunity invariant factors 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 3 . *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 $D_{\mathcal{W}}(\lambda) \leq$ order $\Sigma = \dim A$ *and there exists a realization (X_o, A_o, Y_o) such that the equality holds. The nonunity 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 [4, 33] or to [13] 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 in [30] using also classical arguments from linear system theory (see [13] for instance).

Proposition 1 . *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*

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

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

3.2 Length versus degree

Proposition 1 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 1 . *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 (proposition 1) 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*

$$XA^iYg_0 + XA^{i+1}Yg_1 + \ldots + XA^{i+d}Yg_d = 0 \qquad (9)$$

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

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 $\{A^iY\}_{i=0}^{\infty}$.

Note that any polynomial $g(\lambda)$ that satisfies (9) corresponds to a vector $\vec{g} = {}^t[{}^tg_0, {}^tg_1, \ldots, {}^tg_d]$ of K^{nd} which is in the kernel of the block-Hankel matrix

$$M(\delta_l, d+1) = \left\{XA^{i+j-2}Y\right\}_{i,j}, \quad \text{with } 1 \leq i \leq \delta_l,\ 1 \leq j \leq d+1. \qquad (10)$$

Remark 1 . *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 polynomial of degree less than $\lfloor N/n \rfloor$ for $\{A^iY\}_{i=0}^{\infty}$.*

3.3 Separation

We give an additional fact that will be useful for the computation of minimal generating polynomials. The main difficulty is to choose correct blocking matrices. We work by analogy with the analysis of Wiedemann in §VI of [32]. He uses the Euler's Phi function for polynomials and thus their decomposition into primes. We also show below that the correct choices of matrices can be more easily studied after a factorization of the polynomials.

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 [10] 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 $Q(\lambda)$ [22]. 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 a lcrm [22]. If $P(\lambda)$ and $Q(\lambda)$ have relatively prime determinants $p(\lambda)$ and $q(\lambda)$, the determinant of a lcrm is $p(\lambda)q(\lambda)$.

Lemma 2 . *Assume that A is a block-diagonal matrix $A = \text{diag}(A_p, A_q)$ and that the characteristic polynomials $p(\lambda)$ and $q(\lambda)$ of A_p and A_q, are relatively prime. Consider $Y = {}^t[{}^tY_p \quad {}^tY_q]$ 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 $\{XA_q^iY_q\}_{i=0}^{\infty}$.*

This result will be useful to generalize Wiedemann's analysis in §5.1. Contrary to that, it seems difficult to use the same stategy with Coppersmith's analysis, for instance to establish a result similar to lemma 5.

4 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 [6]) one basic assumption is $m \geq n$. In Kaltofen's analysis (§5 of [14]) 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^* (we refer to [10] for the definition), then one must have $\phi^* = 1$ [14]. In the following we will see

that only $m \geq \min\{\phi^*, n\}$ is required and as indicated by proposition 3, this does not seem too strong.

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* [7] 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 [16, 17, 14]. 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 $v_{k,l}$, $1 \leq l \leq n$, with $1 \leq k \leq N$.

Proposition 2 . *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 $K(v_{1,1}, \ldots, v_{N,n})$) with column degrees*

$$[d_1, \ldots, d_n] = [\delta_r - 1, \ldots, \delta_r - 1, \delta_r, \ldots, \delta_r] \qquad (11)$$

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

$$r_j = \begin{cases} n - \tau + j, & 1 \leq j \leq \tau, \\ j - \tau, & \tau + 1 \leq j \leq n. \end{cases} \qquad (12)$$

Note 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 [14]. The following consequence will be needed for the block method which computes generating polynomials for a sequence $\{A^iZ\}_{i=0}^{\infty}$ with $Z = AY$.

Corollary 1 . *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^*.*

This is a generalization of proposition 3 in [14] to any type of matrix.

5 Preliminary probability analysis

We propose two different but complementary analyses. The first one in §5.1 is a direct generalization of Wiedemann's work [32]. We bound the probability of picking two matrices X and Y such that $D_Y(\lambda) = D_{\mathcal{W}}(\lambda)$ (following the notations of proposition 1). If in addition we require that X and Y lead to $D_Y(\lambda)$ with generic degrees – as characterized by proposition 2 – then we may use most of the arguments of Coppersmith [6], to show that this is "almost" true with a good probability. This will allow us in §6.1, especially for small fields, to adapt the initial block algorithm and bound its probability of failure. We define ϕ and ν as in proposition 2.

5.1 Generalization of Wiedemann's analysis

Let us compute the probability that $\{A^iY\}_0^{\infty}$ and $\{XA^iY\}_0^{\infty}$ have the same minimal generating polynomials. Even if this represents a generalization of Wiedemann's study, the 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. The probability is computed in terms of the function $\Phi_n(f, \phi)$ defined over K =GF(q) for a polynomial $f(\lambda)$ in $K[\lambda]$ and two positive integers n and ϕ, $n \geq \phi$. This function is given by:

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

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: $n = \mu\phi + \rho$.

Lemma 3 . *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* [11]. 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:

$$\bar{A} = P^{-1}AP = \text{diag}\left(\{A_{g_i}\}_i\right) = \left(\{A_{g_i}^{(1)}, \ldots, A_{g_i}^{(\phi_i)}\}_i\right) \qquad (13)$$

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}: $\mathcal{M}_{N,n}(K) \to \mathcal{M}_{N,n}(K)$ 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 lemma 2, we may thus separate the problem according to (13). Let p_Y be the probability that $\dim < Y >= \nu$:

$$\begin{aligned} p_Y &= \text{Prob}_Y \left\{\dim \text{span}(\bar{Y}, \bar{A}\bar{Y}, \bar{A}^2\bar{Y}, \ldots) = \nu\right\} \\ &= \textstyle\prod_i \text{Prob}_{Y_i} \left\{\dim \text{span}(Y_i, A_{g_i}Y_i, \ldots) = \nu_i\right\}, \end{aligned} \qquad (14)$$

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 $c_1, \ldots, c_j$:

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

for $1 \leq j \leq \phi_i$ and let $V_0 = \{0\}$. A sufficient condition to ensure that the dimension of $\text{span}(Y_i, A_{g_i}Y_i, \ldots)$ 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 as 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

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

Above bound can be 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 (14) we obtain

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

□

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 [32]. Now, applying the lemma on the left:

Proposition 3 . *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 $D_Y(\lambda) = D_W(\lambda)$ with probability no less than $\Phi_m(\pi_A, \min\{\phi, n\})$.*

Lemma 4 . *For $f(\lambda)$ of degree $N > q$ and $n \geq \phi$:*

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

5.2 Using Coppersmith's analysis

By proposition 2, in the generic case $D_Y(\lambda)$ has determinantal degree ν and has $\tau = n\lceil \nu/n \rceil - \nu$ columns of degree $\delta_r - 1$ and $n - \tau$ columns of degree δ_r. We compute the probability that $D_Y(\lambda)$ is not "too far" from this situation. Indeed, as implicitly noticed in [6, 14, 21], to always have exactly the generic degrees seems unlikely. 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 $\operatorname{rank} \mathcal{K}_Y(\delta_r) > \nu - \rho$ then the determinantal degree of $D_Y(\lambda)$ is also strictly greater than $\nu - \rho$, and the column degrees of $D_Y(\lambda)$ must be less than $\delta_r + \rho$. If this is true with ρ small, $D_Y(\lambda)$ should be viewed as *nearly generic*.

Next lemma is an interpretation of lemma 6.2 in [6] to fit the current context. We formulate it using the function $\Theta_n(f, \phi)$ defined over $K = \mathrm{GF}(q)$ for a polynomial $f(\lambda)$ in $K[\lambda]$, a positive integer n and a matrix A whose Frobenius form has ϕ companion blocks:

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

where the sum is taken over all the factors of $f(\lambda)$. This definition is slightly different from the original one in [6]. The reader may refer to the latter article concerning the properties of the function that remain unchanged.

Lemma 5 . *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*

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

Proof. Following [6] 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

$$g_1(A)Y^{(1)} + g_2(A)Y^{(2)} + \ldots + g_n(A)Y^{(n)} = 0. \qquad (15)$$

Intuitively, $D_Y(\lambda)$ is a generating polynomial for $\{A^i Y\}_{i=0}^{\infty}$, thus considering $D_Y(\lambda)$ column by column we have:

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

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 (11) and (12) *i.e.* $\operatorname{rank} \mathcal{K}_Y(\delta_r) = \nu$, 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 (15) holds.

For the "if part", it is easily seen that if $D_Y(\lambda)$ satisfies (11) and (12) 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 [6] 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 (15) but violate (C), plus one for the trivial choice:

$$\begin{aligned} E_Y &= \operatorname{Exp}_Y \#\left\{\{g_i\}_i ; \textstyle\sum_i g_i(A)Y^{(i)} = 0\right\} \\ &= \textstyle\sum_{g_i} \operatorname{Prob}_Y\left\{\sum_i g_i(A)Y^{(i)} = 0\right\}. \end{aligned} \qquad (16)$$

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

$$g(\lambda) = \gcd\left(\pi_A(\lambda), g_1(\lambda), \ldots, g_n(\lambda)\right). \qquad (17)$$

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 (17). 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 $q^{-\text{ rank } g(A)}$. Then using (16) we get:

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

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

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

and

$$\text{Prob}_Y\{\text{rank } \mathcal{K}_Y(\delta_r) \leq \nu - \rho\} \leq \Theta_n(\pi_A, \phi)q^{-\rho},$$

which concludes the proof. □

We will mainly need the same result on the left:

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

Lemma 6 . *For $f(\lambda)|\pi_A(\lambda)$ of degree $N > q$ and for $n \geq \phi + 1$:*

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

6 Probability of success

The block algorithm uses $Z = AY$ rather than Y. As in §4 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 is correct or not for any A, m and n. We can only work under the assumption that m is at least greater than $\min\{\phi^*, n\}$ and fortunately, this is not too restricting in most cases. Besides, there are two ways to bypass this difficuly.

To work with any given m and n: either the matrix A is assumed or forced to be non pathological. From our point of view, Coppersmith has assumed $m \geq n \geq \phi^*$. Using the same notations, Kaltofen has proposed a preconditioning of A to ensure $m, n \geq \phi^* = 1$.

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 $\min\{\phi, n\}$.

6.1 Over small cardinality fields

To find *at least one solution* to the linear system, mainly the choice of X and the shift parameter Δ are relevant:

Theorem 4 . *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 §2 with shift parameter Δ, then* $\text{Prob}_{X,Y}\{w \neq 0, Aw = 0\}$ *is greater than* $\left(\Phi_m(\pi_A, \min\{\phi^*, n\}) - \Theta_m(\pi_A, \min\{\phi^*, n\})q^{-\Delta}\right)(1 - 1/V)$.

Proof. We begin by bounding the probability of having $Aw \neq 0$. Following the notations used in §2, we have $\delta_l = \lceil N/m \rceil$ and $\delta_r = \lfloor N/n \rfloor$. We simultaneously apply proposition 3 – which controls the dimension ν^* of the corresponding space – and proposition 4 – which indicates whether the target dimension is reached. As done for corollary 1, these propositions are applied with $Z = AY$ rather than with Y. We denote by N_Z the dimension of $\text{span}(Z, AZ, \ldots)$. If X is such that rank $\mathcal{K}_{XZ}(\delta_l) = N_Z - \Delta$ and if $D_Z(\lambda) = D_{\mathcal{W}}(\lambda)$ then rank $\mathcal{K}_{XZ}(\delta_l + \Delta)$ must be N_Z. Thus by lemma 1, any vector in the kernel of the corresponding block-Hankel matrix $M(\delta_l + \Delta, \delta_r + 1)$ (see (10)) must be a vector generating polynomial and must provide, by definition, a w such that $Aw = 0$. Note that – by remark 1 – the kernel of $M(\delta_l + \Delta, \delta_r + 1)$ is not trivial. Now, by proposition 3 and proposition 4 we know that $\text{Prob}_X\{Aw \neq 0\}$ is less than

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

Finally, by an argument of Coppersmith [6] we know that the probability – on the choice of Y – that $w \neq 0$ is more than $(1 - 1/V)$. □

We have to make a slight additional restriction over GF(q). Indeed, from a practical point of view, lemma 6 and lemma 4 will provide realistic bounds only for $m \geq \min\{\phi^*, n\} + 2$. If the ground field is GF(2) and as soon as $\Delta \geq 8$, our theoretical probability is greater than a fixed number $\epsilon_0 = 0.03 > 0$. Fortunately – to explain the good practical behaviour of the algorithm – our bound may be much greater, even over GF(2). Especially when m is a multiple of ϕ^*, for instance:

Corollary 2 . *If $m \geq 4\phi^*$, $\Delta \geq 8$. then the probability of success is greater than* 0.6.

Remark 2 . *Let us explain why – as experimentally noticed in [15, 21] – blocking may amplify the success probability. 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 [32]). 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 4 increases and can be made arbitrarily close to $(1 - 1/V)$.*

Remark 3 . *Referee 3 has noted that the condition $m \geq n$ and consequently the condition $m \geq \min\{\phi^*, n\}$ 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 $w^tA^t = 0$ using a left version of the block algorithm. From a theoretical point of view, on may use Tellegen's theorem [25] which states that an algorithm computing A times a vector can be converted to one for A^t times a vector. However in certain cases one will still need either $m \geq n + 2$ or $n \geq m + 2$.*

6.2 Large fields – Generalization of Kaltofen's analysis

For large fields, another randomization technique can be used. We follow the ideas of Kaltofen [16, 17, 14]. This ideas

have been successfully applied to singular matrices A whose minimal polynomial has degree $\deg \pi_A(\lambda) = \text{rank}(A) + 1$. Using the generalization of corollary 1, we get for any matrix:

Theorem 5 . *Let A be a $N \times N$ singular matrix over K and let $m \geq \min\{\phi^*, n\}$ Suppose that 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 §2 with $\Delta = 0$ then*

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

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 5 concerning dimensions of Krylov subspaces) while others seem to be inevitable (see the comments after theorem 4 about m and n). Even if 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.

Acknowledgments. Grateful thanks to Erich Kaltofen for his valuable questions.

References

[1] Beckermann, B., and Labahn, G. A uniform approach for the fast computation of matrix-type Padé approximants. *SIAM J. Matrix Anal. Appl. 15*, 3 (July 1994), 804–823.

[2] Beckermann, B., and Labahn, G. Recursiveness in matrix rational interpolation problems. Tech. Rep. Publication ANO 357, Université de Lille France, 1996.

[3] Bultheel, A., and Van Barel, M. A matrix Euclidean algorithm and the matrix minimal Padé approximation problem. In *Continued Fractions and Padé Approximants* (1990), C. Brezinski, Ed., North-Holland, pp. 11–51.

[4] Coppel, W. Matrices of rational functions. *Bull. Austral. Math. Soc. 11* (1974), 89–113.

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

[6] Coppersmith, D. Solving homogeneous linear equations over GF(2) via block Wiedemann algorithm. *Math. Comp. 62*, 205 (1994), 333–350.

[7] Delsarte, P., Genin, Y., and Kamp, Y. 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.

[8] Eberly, W., and Kaltofen, E. On randomized Lanczos algorithms. In *International Symposium on Symbolic and Algebraic Computation, Maui, Hawaii, USA* (July 1997), ACM Press.

[9] Forney, G. Minimal bases of rational vector spaces, with applications to multivariable linear systems. *SIAM J. Control 13* (1975), 493–520.

[10] Gantmacher, F. *Théorie des matrices.* Dunod, Paris, France, 1966.

[11] Jacobson, N. *Lectures in Abstract Algebra II, Linear Algebra.* Springer-Verlag, 1953.

[12] Jacobson, N. *Basic Algebra I.* W.H. Freeman and Company, 1974.

[13] Kailath, T. *Linear systems.* Prentice Hall, 1980.

[14] Kaltofen, E. Analysis of Coppersmith's block Wiedemann algorithm for the parallel solution of sparse linear systems. *Math. Comp. 64*, 210 (1995), 777–806.

[15] Kaltofen, E., and Lobo, A. Distributed matrix-free solution of large sparse linear systems over finite fields. In *High Performance Computing 1996, San Diego, CA* (1996), A. Tentner, Ed., Society for Computer Simulation, Simulation Councils, Inc, pp. 244–247.

[16] Kaltofen, E., and Pan, V. Processor efficient parallel solution of linear systems over an abstract field. In *Proc. 3rd Annual ACM Symposium on Parallel Algorithms and Architecture* (1991), ACM-Press.

[17] Kaltofen, E., and Saunders, B. On Wiedemann's method of solving sparse linear systems. In *Proc. AAECC-9* (1991), LNCS 539, Springer Verlag, pp. 29–38.

[18] LaMacchia, B., and Odlyzko, A. Solving large sparse linear systems over finite fields. In *Advances in Cryptology - CRYPTO'90, Springer LNCS 537* (1991), A. Menezes and S. Vanstone, Eds., pp. 109–133.

[19] Lambert, R. *Computational aspects of discrete logarithms.* PhD thesis, University of Waterloo, Ontario, Canada, 1996.

[20] Lenstra, A., Lenstra, H., Manasse, M., and Pollard, J. The factorization of the ninth Fermat number. *Math. Comp. 61* (1993), 319–349.

[21] Lobo, A. *Matrix-free linear system solving and applications to symbolic computation.* PhD thesis, Dept. Comp. Sc., Rensselaer Polytech. Instit., Troy, New York, Dec. 1995.

[22] MacDuffee, C. *The theory of matrices.* Chelsea, New-York, 1956.

[23] Massey, J. Shift-register synthesis and BCH decoding. *IEEE Trans. Inform. Theory 15* (1969), 122–127.

[24] Montgomery, P. A block Lanczos algorithm for finding dependencies over GF(2). In *EUROCRYPT'95, Heidelberg, Germany. Springer LNCS 921* (1995), pp. 106–120.

[25] Penfield Jr., P., Spencer, R., and Duinker, S. *Tellegen's theorem and electrical networks.* M.I.T. Press, Cambridge, MA, 1970.

[26] Popov, V. Invariant description of linear, time-invariant controllable systems. *SIAM J. Control 10* (May 1972), 252–264.

[27] Van Barel, M., and Bultheel, A. A general module theoretic framework for vector M-Padé and matrix rational interpolation. *Numerical Algorithms 3* (1992), 451–462.

[28] Verghese, G., and Kailath. Rational matrix structure. *IEEE Trans. Automat. Control 26* (1981), 434–438.

[29] Villard, G. Computing Popov and Hermite forms of polynomial matrices. In *International Symposium on Symbolic and Algebraic Computation, Zurich, Suisse* (July 1996), ACM Press, pp. 250–258.

[30] Villard, G. A study of Coppersmith's block Wiedemann algorithm using matrix polynomials, Feb. 1997. RR IMAG Grenoble, France.

[31] Villard, G. Computing minimum generating matrix polynomials, 1997. Preprint IMAG Grenoble, France.

[32] Wiedemann, D. Solving sparse linear equations over finite fields. *IEEE Transf. Inform. Theory 32* (1986), 54–62.

[33] Wimmer, H. A Jordan factorization for polynomial matrices. *Proceedings of the American Math. Soc. 75*, 2 (1979), 201–206.

[34] Wolovich, W. *Linear multivariable systems.* Springer-Verlag, New-York, 1974.