Straight-line computation of the polynomial matrix inverse

Claude-Pierre Jeannerod, Gilles Villard

Laboratoire de l'Informatique du Parallélisme

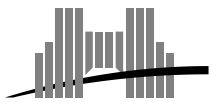
École Normale Supérieure de Lyon
Unité Mixte de Recherche CNRS-INRIA-ENS LYON n° 5668

Straight-line computation of the polynomial matrix inverse

Claude-Pierre Jeannerod
Gilles Villard

December 2002

Research Report N° 2002-47



École Normale Supérieure de Lyon
46 Allée d'Italie, 69364 Lyon Cedex 07, France
Téléphone : +33(0)4.72.72.80.37
Télécopieur : +33(0)4.72.72.80.80
Adresse électronique : `lip@ens-lyon.fr`

Straight-line computation of the polynomial matrix inverse

Claude-Pierre Jeannerod
Gilles Villard

December 2002

Abstract

We present an inversion algorithm for nonsingular $n \times n$ matrices whose entries are degree d polynomials over a field. The algorithm is deterministic and requires $O^{\sim}(n^3 d)$ field operations for a generic input. The "soft Oh" notation $O^{\sim}$ indicates some missing $\log(nd)$ factors. The polynomial matrix inverse can thus be computed by a straight-line program whose size is — asymptotically and up to logarithmic factors — the size of its output.

Keywords: Matrix polynomial, matrix inversion, minimal basis.

Résumé

On présente un algorithme pour inverser les matrices $n \times n$ dont les coefficients sont des polynômes de degré d sur un corps commutatif abstrait K. Cet algorithme est déterministe et nécessite $O^{\sim}(n^3 d)$ opérations sur K dans le cas générique. (La notation $O^{\sim}$ contient les facteurs logarithmiques.) L'inverse d'une matrice polynomiale peut donc être calculé par un programme "straight-line" de longueur égale — asymptotiquement et aux facteurs logarithmiques près — à la taille de sa sortie.

Mots-clés: Matrices polynomiales, inversion matricielle, bases minimales.

1 Introduction

Let K be an abstract commutative field. For two positive integers n and d, consider a nonsingular matrix $A \in \mathsf{K}[x]^{n\times n}$ of degree d. The inverse of A is the $n \times n$ matrix over $\mathsf{K}(x)$ defined by $A^{-1} = A^*/\det A$ where A^* and $\det A$ are the adjoint matrix and the determinant of A. Since the degrees of A^* and $\det A$ are bounded by $(n-1)d$ and nd, representing A^{-1} may require up to $O(n^3d)$ elements of K. In this paper, we present a deterministic algorithm for computing A^{-1} that generically requires $O\tilde{}(n^3d)$ field operations on an algebraic random access machine. By generically, we mean that the algorithm has the above asymptotic complexity for every $n \times n$ matrix polynomial of degree d whose coefficients do not belong to a certain strict subvariety of $\mathsf{K}^{n^2(d+1)}$. Hence we establish a straight-line complexity estimate of $O\tilde{}(n^3d)$. Here and in the following, the $O\tilde{}$ notation indicates some missing $\log(nd)$ factors.

When using either an algebraic random access machine and or the straight-line model, the best previously known complexity estimate was $O\tilde{}(n^{\omega+1}d)$ where ω is the exponent for multiplying two $n\times n$ matrices over K [5, Chap. 1]. We thus improve the straight-line complexity estimate if $\omega > 2$ and the improvement is by a factor n when considering standard matrix multiplication ($\omega = 3$).

We assume that $O(M(d))$ operations in K are sufficient to multiply two polynomials in $\mathsf{K}[x]$ of degree at most d, and use $M(d) = d\log d\log\log d$ [16, 6]. Let us recall how the above classical estimate $O\tilde{}(n^{\omega+1}d)$ for matrix inversion over $\mathsf{K}(x)$ is obtained. The determinant and the entries of the adjoint, whose degrees are bounded by nd, may be recovered for instance using an evaluation / interpolation scheme at nd points [9, §5.5]. A randomized Las Vegas algorithm or a straight-line program — A must be invertible at the nd evaluation points — may thus be based on $O(n^\omega)$ recursive matrix inversion over K [4, 15] and on a fast evaluation / interpolation scheme for univariate polynomials of degree nd in $O\tilde{}(M(nd))$ operations [11], [9, §10]. Many other approaches may be considered such as a direct Gauss /Jordan elimination on truncated power series, Newton iteration [13] or linearization (see for instance [12] and the references therein) but none of them seems to reduce the complexity estimate over K. A deterministic $O\tilde{}(n^{\omega+1}d)$ algorithm is given in [17, §2] for an algebraic random access machine. This algorithm is a fraction-free version over $\mathsf{K}[x]$ (Bareiss' approach [1]) of the recursive inversion algorithms over K cited above. We see that with standard matrix multiplication ($\omega = 3$) the cost of inversion was still about n times higher than the typical size of the inverse.

Let us now give an idea of our approach. The algorithm is iterative and consists in diagonalizing A in $\lceil \log n \rceil$ steps as $UA = B$ with $U \in \mathsf{K}[x]^{n\times n}$ nonsingular and $B \in \mathsf{K}[x]^{n\times n}$ diagonal. The inverse of A is then recovered as $A^{-1} = B^{-1}U$. The first observation is that A can be diagonalized in $\lceil \log n \rceil$ steps of type

$$A = \begin{bmatrix} A_L & A_R \end{bmatrix} \quad \rightarrow \quad UA = \begin{bmatrix} \overline{U} \\ \underline{U} \end{bmatrix} \begin{bmatrix} A_L & A_R \end{bmatrix} = \begin{bmatrix} \overline{U}A_L & \\ & \underline{U}A_R \end{bmatrix} \tag{1}$$

with $A_L, A_R \in \mathsf{K}[x]^{n\times n/2}$ and where $\overline{U}, \underline{U} \in \mathsf{K}[x]^{n/2\times n}$ are bases of, respectively, $\ker A_R$ and $\ker A_L$ considered as $\mathsf{K}[x]$-modules. (Here and hereafter, the blank areas in matrices are assumed to be filled with zeros.) The second and key point is that, generically, there exists among all possible kernel bases $\overline{U}$, $\underline{U}$ some bases whose degree is no more than d; furthermore, if we choose such "low degree bases", this property carries over the next step. Hence, the degree of the working matrix polynomials only doubles at each step, whereas their size is divided by two. These "low degree bases" are so-called minimal bases [7], for which one knows deterministic $O\tilde{}(n^2M(nd))$ algorithms [19, 2]. Combining such algorithms with steps of type (1) eventually allows for A^{-1} to be computed in

$$O(n^2M(nd)\log(nd)) = O\tilde{}(n^3d)$$

field operations by using only standard matrix multiplication.

The paper is organized as follows. In Section 2 we present our inversion algorithm, using block-diagonalization steps as in (1) and assuming that one can compute minimal bases of matrix polynomial kernels. Section 3 surveys minimal bases in the context of the matrix polynomial kernels of interest for our problem: general properties as well as a characterization of the degree of such bases are given in §3.1 and computation as matrix rational approximants is explained in §3.2. We give in Section 4 a sufficient, generically satisfied condition on the coefficients of the input A for the degrees of all the matrices involved at step i of our computation of A^{-1} to be no greater than $2^{i-1}d$. This degree bound allows for the complexity analysis of Section 5.

For the rest of the paper, we assume without loss of generality that $n = 2^p$ and $p \in \mathbb{N}$. Moreover, M_L (resp. M_R) denotes the $2m \times m$ matrix that consists of the first (resp. last) m columns of a given $2m \times 2m$ matrix M; the $m \times 2m$ submatrices $\overline{M}$ and $\underline{M}$ are defined similarly by considering rows instead. For example $\overline{M}_L$ therefore denotes the upper left $m \times m$ submatrix of M.

2 Inversion algorithm

Algorithm `Inverse` is described below. We assume that we have a subroutine `MinimalKernelBasis` for computing a minimal basis of the left kernel of a matrix polynomial.

Algorithm `Inverse`(A)
Input: $A \in \mathsf{K}[x]^{n\times n}$ of degree d
Output: A^{-1}
Condition: $\det A \neq 0$ and $n = 2^p$ with $p \in \mathbb{N}$

(1) $B := \mathrm{copy}(A)$;
$U := I_n$;
(2) **for** i **from** 1 **to** p **do** // $B = \mathrm{diag}(B_i^{(1)}, \ldots, B_i^{(2^{i-1})})$
for j **from** 1 **to** 2^{i-1} **do**
$\underline{U}_i^{(j)} := \texttt{MinimalKernelBasis}(B_{i,L}^{(j)})$; // $\underline{U}_i^{(j)} B_{i,L}^{(j)} = 0$
$\overline{U}_i^{(j)} := \texttt{MinimalKernelBasis}(B_{i,R}^{(j)})$; // $\overline{U}_i^{(j)} B_{i,R}^{(j)} = 0$
od;
$U_i := \mathrm{diag}(U_i^{(1)}, \ldots, U_i^{(2^{i-1})})$; // $U_i^{(j)} = \begin{bmatrix} \overline{U}_i^{(j)} \\ \underline{U}_i^{(j)} \end{bmatrix}$
$B := U_i B$;
$U := U_i U$;
od;
(3) **return** $B^{-1}U$

We now prove that Algorithm `Inverse` is correct. For $i = 1$, it follows from $\det A \neq 0$ that both $\underline{U}_1^{(1)}$ and $\overline{U}_1^{(1)}$ are $n/2 \times n$ matrices over $\mathsf{K}[x]$ of rank $n/2$. Also, $\det U_1^{(1)} \neq 0$ for otherwise $\ker A_L \cap \ker A_R = 0$ which contradicts the fact that $\det A \neq 0$. Therefore, the two $n/2 \times n/2$ blocks of $B = U_1^{(1)} A$ are nonsingular. Repeating the argument for $i = 2, \ldots, p$, we see that at the beginning of step i the matrix B is block-diagonal with 2^{i-1} nonsingular blocks of order 2^{p-i+1}. Hence the 2^{i-1} pairs of kernel bases $\underline{U}_i^{(j)}$, $\overline{U}_i^{(j)}$ with dimensions $2^{p-i} \times 2^{p-i+1}$ are such that $\det U_i^{(j)} \neq 0$. The pth step of stage (2) therefore produces a nonsingular $U \in \mathsf{K}[x]^{n\times n}$ such that $UA = B$ is diagonal and the algorithm is correct.

Notice that correctness of the algorithm does not require the computed kernel bases be minimal, that is, have "small" degrees. However, minimality — whose definition is recalled in Section 3 — is crucial for the complexity of the algorithm. In particular, we prove in Section 4 that minimality

implies that the matrices $B_{i,L}^{(j)}$, $B_{i,R}^{(j)}$, $\underline{U}_i^{(j)}$, $\overline{U}_i^{(j)}$ of Algorithm `Inverse` generically have a degree equal to $2^{i-1}d$ for $1 \leq j \leq 2^{i-1}, 1 \leq i \leq p$. Hence, at step i, one has for these matrices

$$\text{size} \times \text{degree} = n/2^{i-1} \times 2^{i-1}d = nd \quad \text{for } 1 \leq i \leq p.$$

3 Minimal kernel bases

For a positive integer m, let $M \in \mathsf{K}[x]^{2m \times m}$ with rank m and degree d. Let further $U \in \mathsf{K}[x]^{m \times 2m}$ with rows forming a basis of the $\mathsf{K}[x]$-submodule $\ker M$ and denote by $d_1, \ldots, d_m$ its row degrees. When $\sum_{i=1}^m d_i$ is minimal among all the bases of $\ker M$, the matrix U is known as a *minimal basis* of $\ker M$ [7].

First, we characterize in §3.1 the fact that U has degree d exactly when M is generic. In the latter case, we also identify a particular choice for the matrix U and write its entries as rational functions in the entries of M. This will be the main properties used for studying the generic behaviour of the whole inversion algorithm. Then we recall in §3.2 how to compute minimal bases.

3.1 General properties, degree characterization and explicit construction

We refer to Forney [7] and to Kailath [10] for the definition and properties of minimal bases. Recall in particular that such bases are non unique but their row degrees are unique up to ordering [10, §6.5.4]. In this case, we shall refer to $d_1 \leq d_2 \leq \ldots \leq d_m$ as the *minimal row degrees of* $\ker M$. A bound on the sum of the degrees may be seen by linearizing $M = \sum_{i=0}^d M_i x^i$ as

$$L_M = \begin{bmatrix} & & & M_0 \\ -I_{2m} & & & M_1 \\ & \ddots & & \vdots \\ & & -I_{2m} & M_{d-1} \end{bmatrix} + x \begin{bmatrix} I_{2m} & & & \\ & \ddots & & \\ & & I_{2m} & \\ & & & M_d \end{bmatrix} \in \mathsf{K}[x]^{2md \times (2md-m)}.$$

Indeed, one can check first that U is a minimal basis of $\ker M$ if and only if the matrix L_U defined as $L_U = [\, U \,|\, xU \,|\, \cdots \,|\, x^{d-1}U \,]$ is a minimal basis of $\ker L_M$. Second, the row degrees $d_1 + d - 1, \ldots, d_m + d - 1$ of L_U are known to be the left indices of the Kronecker canonical form of the matrix pencil L_M [8, §12.5]. The uniqueness of the d_i's thus corresponds to the uniqueness of the Kronecker indices. By bounding the sum of the Kronecker indices by the column dimension of the pencil L_M it further follows that $\sum_{i=1}^m (d_i + d - 1) \leq 2md - m$. In other words, the minimal row degrees of $\ker M$ satisfy

$$\sum_{i=1}^m d_i \leq md. \tag{2}$$

Degrees of minimal bases may thus range from zero to md. For example, one can take $U = [\, I_m \,|\, O \,]$ for any M whose first m rows are zero; on the other hand, when

$$M = \begin{bmatrix} x & 0 \\ 1 & x \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

a minimal basis of $\ker M$ is given by

$$U = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & -x & x^2 & 0 \end{bmatrix}.$$

Let us now characterize the situation where all the minimal row degrees are equal to d by associating with $M = \sum_{i=0}^{d} M_i x^i$ the block-Toeplitz matrix

$$\mathcal{T}(M) = \begin{bmatrix} M_0 & M_1 & \cdots & M_d & & \\ & \ddots & \ddots & \vdots & \ddots & \\ & & M_0 & M_1 & \cdots & M_d \end{bmatrix} \in \mathsf{K}^{2md \times 2md}. \tag{3}$$

We see that $\ker M$ has a nonzero vector $u = \sum_{i=0}^{d-1} u_i x^i$ of degree less than d if and only if $[u_0^T, \ldots, u_{d-1}^T]^T$ is a nonzero vector of $\ker \mathcal{T}(M)$. Therefore, $d_i \geq d$ for $1 \leq i \leq m$ if and only if $\det \mathcal{T}(M) \neq 0$. The characterization below then follows by using inequality (2).

Lemma 1 *The minimal row degrees of* $\ker M$ *satisfy* $d_i = d$ *for* $1 \leq i \leq m$ *if and only if* $\mathcal{T}(M)$ *is invertible, which is generically the case.*

We use Lemma 1 in the proof of Proposition 2. We shall also use the fact that if $\mathcal{T}(M)$ is invertible then, by uniqueness of the minimal degrees, every basis of $\ker M$ whose degree is at most d must be minimal.

In addition to the bound d on the generic degrees, Proposition 2 relies on the following explicit construction of a kernel basis, assuming further that either $\underline{M}_d$ or $\overline{M}_d$ is nonsingular. In the generic case, we identify the matrix coefficients in both sides of matrix equation $UM = 0$. If $\det \mathcal{T}(M) \neq 0$ and $\det \underline{M}_d \neq 0$ then a minimal basis of $\ker M$ is given by the $m \times 2m$ matrix $N = \sum_{i=0}^{d} N_i x^i$ such that

$$N_d = \left[\; I_m \;\middle|\; -\overline{M}_d \underline{M}_d^{-1} \;\right], \tag{4a}$$

$$\left[\; N_0 \quad \cdots \quad N_{d-1} \;\right] = -N_d \left[\; 0 \;\middle|\; M_0 \quad \cdots \quad M_{d-1} \;\right] \mathcal{T}(M)^{-1}. \tag{4b}$$

If $\det \mathcal{T}(M) \neq 0$ and $\det \overline{M}_d \neq 0$ then one can replace (4a) with

$$N_d = [\, -\underline{M}_d \overline{M}_d^{-1} \mid I_m \,]. \tag{4c}$$

3.2 Computation as matrix rational approximants

We show in Proposition 1 below that computing a minimal basis U of $\ker M$ reduces to computing a suitable Padé approximant basis (called a σ-basis in [2, p. 809]) for the rows of M, a task for which one may use the algorithm of [2, §6]. This follows the idea of [14, Chap. 4] and [19] as applied in [3]. As we detail next, we recover a minimal basis from a minimal approximant $V \in \mathsf{K}[x]^{(2m)\times(2m)}$ such that

$$V(x)M(x) = O(x^\sigma).$$

The approximation order σ we choose is sufficiently large — with respect to a degree bound for the basis — to ensure that m rows of V form a minimal basis U for the kernel.

Let $\sigma \in \mathbb{N}$ and, writing $M^{(i,j)}$ for the (i,j) entry of M, let $f = [\, f^{(i)} \,]_i \in \mathsf{K}[x]^{2m}$ where $f^{(i)}(x) = \sum_{j=1}^{m} x^{j-1} M^{(i,j)}(x^m)$. To define σ-bases with respect to the rows of M, we recall the notion of *order* [2, p. 809] of a polynomial vector $v^T = [\, v^{(i)} \,]_i \in \mathsf{K}[x]^{2m}$:

$$\operatorname{ord} v = \sup\{\tau \in \mathbb{N} \,:\, v(x^m) f(x) = O(x^\tau)\}.$$

A *σ-basis with respect to the rows of M* is a matrix $V \in \mathsf{K}[x]^{2m\times 2m}$ such that:

i) for $1 \leq i \leq 2m$, $\operatorname{ord} V^{(i,*)} \geq \sigma$ where $V^{(i,*)}$ is the ith row of V;

II) every polynomial vector $v \in \mathsf{K}[x]^{2m}$ such that $\operatorname{ord} v \geq \sigma$ admits a unique decomposition $v^T = \sum_{i=1}^{2m} c^{(i)} V^{(i,*)}$ where, for $1 \leq i \leq 2m$, $c^{(i)} \in \mathsf{K}[x]$ and $\deg c^{(i)} + \deg V^{(i,*)} \leq \deg v$ (minimality of the approximant).

The definition and the existence of such a basis V for a given (σ, M) follow from [2]. By the decomposition in II), we see that V must be nonsingular.

The result below shows how to recover a minimal basis of $\ker M$ from a σ-basis when σ is large enough. Although a general version that is not restricted to the full rank $2m \times m$ case can be found in [14, 19], we give a proof for the sake of completeness.

Property 1 *Let V be a σ-basis with respect to the rows of M and let d_i be the ith minimal row degree of $\ker M$. If $\sigma \geq m(\max_i d_i + d + 1)$ then the m rows of V with smallest degrees define a minimal basis of $\ker M$.*

Proof. For $1 \leq i \leq 2m$, one has $V^{(i,*)}(x^m) f(x) = O(x^\sigma)$ where the left hand side is a polynomial of degree at most

$$m(\deg V^{(i,*)} + d + 1) - 1. \tag{5}$$

It thus follows from (5) and from $\sigma \geq m(\max_i d_i + d + 1)$ that a row of V whose degree is no more than $\max_i d_i$ is a vector of $\ker M$. Let us now show that V has m rows of respective degrees $d_1, \ldots, d_m$. By definition, a vector u_1 of $\ker M$ of degree d_1 can be written as $u_1 = \sum_{h=1}^{2m} c_1^{(h)} V^{(h,*)}$ with $\deg c_1^{(h)} + \deg V^{(h,*)} \leq d_1$. Hence there exists h_1 such that $\deg V^{(h_1,*)} \leq d_1$. Now assume that V has $i-1$ rows $V^{(h_1,*)}, \ldots, V^{(h_{i-1},*)}$ of respective degrees $d_1, \ldots, d_{i-1}$ and let u_i be a vector of $\ker M$ that does not belong to the submodule generated by these $i-1$ rows and such that $\deg u_i = d_i$. As for $i = 1$, there exists $h_i \notin \{h_1, \ldots, h_{i-1}\}$ such that $\deg V^{(h_i,*)} \leq d_i$. Therefore V contains m distinct rows (indexed by $h_1, \ldots, h_m$) such that the h_i-th row belongs to $\ker M$ and has degree at most d_i. These m rows are linearly independent in $\ker M$ and, since $\sum_{i=1}^m d_i$ is minimal for any such set of rows, they must form a minimal basis. Notice that the remaining m rows of V cannot belong to $\ker M$ and therefore have degrees greater than $\max_i d_i$. The choice of the m rows with smallest degrees in the statement of the proposition is thus well-defined. □

When multiplying two polynomials of degree d costs $O(M(d))$, one can compute a σ-basis with respect to the rows of M in $O(m^2 M(\sigma) \log \sigma)$ field operations [2, Theorem 6.2]. Lemma 1 and Proposition 1 thus yield the corollary below.

Corollary 1 *Let U be a minimal basis of $\ker M$ of given degree d_U. We have an algorithm for computing U in $O(m^2 M(m(d_U + d)) \log(m(d_U + d)))$ field operations. Thus if $\det \mathcal{T}(M) \neq 0$ this cost is $O(m^2 M(md) \log md)$.*

4 Generic degrees of intermediate minimal bases

In this section we study the generic behaviour of Algorithm `Inverse`. More precisely, we prove the degree bound $2^{i-1}d$ for the intermediate matrices at step i. First, we define in Lemma 2 a rational function Δ in the entries of the algorithm input matrix. The existence of this rational function implicitly yields an inversion straight-line program where minimal kernel bases are computed with scheme (4). Second, we show in Proposition 2 that if Δ is well-defined and nonzero for a given input A then the degree bound $2^{i-1}d$ holds for any choice of minimal basis.

Consider $n^2(d+1)$ indeterminates $\alpha_{i,j,k}$ for $1 \leq i, j \leq n$, $0 \leq k \leq d$, and let

$$A \in \mathsf{K}[\alpha_{1,1,0}, \ldots, \alpha_{i,j,k}, \ldots, \alpha_{n,n,d}][x]^{n \times n}$$

have its (i,j) entry equal to $\sum_{k=0}^{d} \alpha_{i,j,k} x^k$. Recall that $n = 2^p$ and let

$$\nu_i = n/2^{i-1} \quad \text{and} \quad \delta_i = 2^{i-1}d \quad \text{for } 1 \le i \le p.$$

Following Algorithm `Inverse`, we apply (4a-b) and (4b-c) formally and iteratively for $i = 1, \ldots, p$ with degree δ_i at step i. We show in Lemma 2 below that this construction actually leads to a well-defined and nonzero rational funtion Δ in the $\alpha_{i,j,k}$'s. This means that (4a-b) and (4b-c) reflect the quantities computed by the algorithm in the generic case. For $1 \le i \le p$, $1 \le j \le 2^{i-1}$, let the matrices $A_i^{(j)}, N_i^{(j)} \in \mathsf{K}[x]^{\nu_i \times \nu_i}$ be of degree δ_i and such that: $A_1^{(1)} = A$; $\overline{N}_i^{(j)}$ is the minimal basis of $\ker A_{i,R}^{(j)}$ of the form (4a-b); $\underline{N}_i^{(j)}$ is the minimal basis of $\ker A_{i,L}^{(j)}$ of the form (4b-c); for $1 \le i \le p-1$, $1 \le j \le 2^{i-1}$,

$$\begin{bmatrix} A_{i+1}^{(2j-1)} & \\ & A_{i+1}^{(2j)} \end{bmatrix} = \begin{bmatrix} \overline{N}_i^{(j)} \\ \underline{N}_i^{(j)} \end{bmatrix} \begin{bmatrix} A_{i,L}^{(j)} & A_{i,R}^{(j)} \end{bmatrix}. \tag{6}$$

Writing $\mathcal{L}_i^{(j)}$ for the leading matrix of $A_i^{(j)}$ and using the notation of (3) for block-Toeplitz matrices, define further $\Delta_{i,L}^{(j)} = \det \overline{\mathcal{L}}_{i,L}^{(j)} \det \mathcal{T}(A_{i,L}^{(j)})$ and $\Delta_{i,R}^{(j)} = \det \underline{\mathcal{L}}_{i,R}^{(j)} \det \mathcal{T}(A_{i,R}^{(j)})$. Then let

$$\Delta = \prod_{i=1}^{p} \prod_{j=1}^{2^{i-1}} \Delta_{i,L}^{(j)} \Delta_{i,R}^{(j)}.$$

Lemma 2 *For $n \ge 2$, Δ is a nonzero element of $\mathsf{K}(\alpha_{1,1,0}, \ldots, \alpha_{i,j,k}, \ldots, \alpha_{n,n,d})$.*

Proof. We prove the statement by recurrence on the ith stage of the construction. To prove both the existence of Δ — matrix inversions in (4a-b-c) — and the fact that $\Delta \neq 0$, it suffices to show that the successive determinants $\Delta_{i,L}^{(j)}$ and $\Delta_{i,R}^{(j)}$ are nonzero for a particular matrix over $\mathsf{K}[x]$. Define $A_{n,d}, N_{n,d} \in \mathsf{K}[x]^{n \times n}$ of degree d as

$$A_{n,d} = \begin{bmatrix} x^d I_{n/2} & -C_{n/2} \\ -I_{n/2} & x^d I_{n/2} \end{bmatrix} \qquad \text{where} \quad C_{n/2} = \begin{bmatrix} & & & & & & 1 \\ & & & & & 1 & \\ & & & & I_2 & & \\ & & & I_4 & & & \\ & & \iddots & & & & \\ & I_{n/8} & & & & & \\ I_{n/4} & & & & & & \end{bmatrix}.$$

$$N_{n,d} = \begin{bmatrix} x^d I_{n/2} & C_{n/2} \\ I_{n/2} & x^d I_{n/2} \end{bmatrix}$$

We show that the determinants used to define Δ are nonzero by proving that construction (6) yields

$$A_i^{(j)} = A_{\nu_i, \delta_i} \quad \text{and} \quad N_i^{(j)} = N_{\nu_i, \delta_i}. \tag{7}$$

By definition $A_1^{(1)} = A_{n,d}$. Hence $\overline{\mathcal{L}}_{1,L}^{(1)} = \underline{\mathcal{L}}_{1,R}^{(1)} = I_{n/2}$ and one can verify by inspection that $\mathcal{T}(A_{1,L}^{(1)})$ and $\mathcal{T}(A_{1,R}^{(1)})$ are invertible, of determinant ± 1; for example one has

$$\mathcal{T}(A_{1,R}^{(1)}) = \begin{bmatrix} \begin{bmatrix} -C_{n/2} \\ O \end{bmatrix} & & & \begin{bmatrix} O \\ I_{n/2} \end{bmatrix} & & \\ & \ddots & & & \ddots & \\ & & \begin{bmatrix} -C_{n/2} \\ O \end{bmatrix} & & & \begin{bmatrix} O \\ I_{n/2} \end{bmatrix} \end{bmatrix} \in \mathsf{K}^{nd \times nd}.$$

To obtain $N_1^{(1)} = N_n$, notice further that $\mathcal{T}(A_{1,R}^{(1)})^{-1}$ is equal to the transpose of $\mathcal{T}([-C_{n/2}^{-1} \mid x^d I_{n/2}]^T)$:

$$\mathcal{T}(A_{1,R}^{(1)})^{-1} = \begin{bmatrix} \begin{bmatrix} -C_{n/2}^{-1} & O \end{bmatrix} & & \\ & \ddots & \\ & & \begin{bmatrix} -C_{n/2}^{-1} & O \end{bmatrix} \\ \begin{bmatrix} O & I_{n/2} \end{bmatrix} & & \\ & \ddots & \\ & & \begin{bmatrix} O & I_{n/2} \end{bmatrix} \end{bmatrix} \in \mathsf{K}^{nd \times nd}.$$

It then follows from (4a-b) that $\overline{N}_1^{(1)} = \overline{N}_n$; with (4b-c), the fact that $\mathcal{T}(A_{1,L}^{(1)})^{-1}$ is equal to the transpose of $\mathcal{T}([x^d I_{n/2} \mid -I_{n/2}]^T)$ yields $\underline{N}_1^{(1)} = \underline{N}_n$. Hence $N_1^{(1)} = N_n$ and (7) holds for $i = 1$. Now, assuming that (7) holds for $i \in \{1, \ldots, p-1\}$, let us show that this is still true for $i+1$. It follows from the block-diagonalization scheme (6) and from the recurrence formula

$$N_{n,d} A_{n,d} = \begin{bmatrix} A_{n/2,2d} & \\ & A_{n/2,2d} \end{bmatrix}$$

that $A_{i+1}^{(j)} = A_{\nu_{i+1},\delta_{i+1}}$ for $1 \le j \le 2^i$. Hence $\overline{\mathcal{L}}_{i+1,L}^{(j)} = \underline{\mathcal{L}}_{i+1,R}^{(j)} = I_{\nu_{i+2}}$; we further obtain $\det \mathcal{T}(A_{i+1,L}^{(j)}) \neq 0$, $\det \mathcal{T}(A_{i+1,R}^{(j)}) \neq 0$ and $N_{i+1}^{(j)} = N_{\nu_{i+1},\delta_{i+1}}$ by using the same arguments as for $i = 1$. □

Actually, with the construction of Δ we have also shown that Algorithm `Inverse`, with each call to `MinimalKernelBasis` replaced by (4a-b) or (4b-c), leads to a degree bound $2^{i-1}d$ in the generic case. The next proposition shows that the bound remains valid in the generic case for *any* choice of minimal bases. This is clearly a consequence of the uniqueness of the minimal degrees.

Property 2 *Let $A \in \mathsf{K}[x]^{n \times n}$ be nonsingular of degree d. If $\Delta(A) \neq 0$ then, for $1 \le i \le p$, the matrices $B_{i,L}^{(j)}$, $B_{i,R}^{(j)}$, $\underline{U}_i^{(j)}$, $\overline{U}_i^{(j)}$ computed by Algorithm* `Inverse` *have degree δ_i.*

Proof. Let $B_i^{(j)}, U_i^{(j)}$ be the quantities computed by `Inverse`(A) and, since $\Delta(A) \neq 0$, consider $A_i^{(j)}$, $N_i^{(j)}$ as in (6). It then suffices to show that there exists invertible constant matrices $C_i^{(j)}$ such that $B_i^{(j)} = C_i^{(j)} A_i^{(j)}$, $1 \le j \le 2^{i-1}$, $1 \le i \le p$. This indeed implies that $\det \mathcal{T}(B_{i,L}^{(j)}) \neq 0$ if and only if $\det \mathcal{T}(A_{i,L}^{(j)}) \neq 0$ (similarly for $\mathcal{T}(B_{i,R}^{(j)})$) and conclusion follows from $\Delta(A) \neq 0$ and from the "if" part of Lemma 1.

We now prove by recurrence on i that such $C_i^{(j)}$ exist. This is true when $i = 1$, for $B_1^{(1)} = A_1^{(1)} = A$. Now assuming that $B_i^{(j)} = C_i^{(j)} A_i^{(j)}$, let us show that $B_{i+1}^{(2j-1)} = C_{i+1}^{(2j-1)} A_{i+1}^{(2j-1)}$ for some constant invertible matrix $C_{i+1}^{(2j-1)}$ over K. It follows from Algorithm `Inverse` that $B_{i+1}^{(2j-1)} = \overline{U}_i^{(j)} B_{i,L}^{(j)} = \overline{U}_i^{(j)} C_i^{(j)} A_{i,L}^{(j)}$ with $\overline{U}_i^{(j)}$ a minimal basis of $\ker C_i^{(j)} A_{i,R}^{(j)}$. Hence $\overline{U}_i^{(j)} C_i^{(j)}$ is a basis of $\ker A_{i,R}^{(j)}$ and is minimal since $C_i^{(j)}$ having degree zero implies that $\overline{U}_i^{(j)} C_i^{(j)}$ has minimal degree δ_i. Another basis of $\ker A_{i,R}^{(j)}$ being given by $\overline{N}_i^{(j)}$, there exists a unimodular $C_{i+1}^{(2j-1)}$ such that $\overline{U}_i^{(j)} C_i^{(j)} = C_{i+1}^{(2j-1)} \overline{N}_i^{(j)}$. It then follows from (6) that $B_{i+1}^{(2j-1)} = C_{i+1}^{(2j-1)} A_{i+1}^{(2j-1)}$. Additionally, uniqueness of the minimal row degrees and normalization (4a) of the leading matrix of $\overline{N}_i^{(j)}$ imply that the degree of $C_{i+1}^{(2j-1)}$ must be equal to zero. This proves the existence of

$C_{i+1}^{(2j-1)}$ as announced. Using similar arguments, one can verify that $B_{i+1}^{(2j)} = C_{i+1}^{(2j)} A_{i+1}^{(2j)}$ for some invertible constant matrix $C_{i+1}^{(2j)}$. □

The degree bound of Proposition 2 is achieved independently of the way minimal kernel bases are computed. Also note that $\det A \neq 0$ and $\Delta(A) \neq 0$ are two distinct assumptions. As illustrated by

$$A = \begin{bmatrix} x & x \\ 1+x & 1+x \end{bmatrix},$$

$\Delta(A) \neq 0$ does not imply that $\det A \neq 0$. Conversely, every $2n \times 2n$ nonsingular matrix polynomial of the form $A = \mathrm{diag}(\overline{A}_L, \underline{A}_R)$ is such that $\Delta(A) = 0$.

When identifying the matrix set $\{A \in \mathsf{K}[x]^{n\times n} : \deg A \leq d\}$ with $\mathsf{K}^{n^2(d+1)}$, the subset $\{A \in \mathsf{K}[x]^{n\times n} : \deg A \leq d \text{ and } (\Delta(A) \text{ is undefined or } \Delta(A) = 0)\}$ can be identified with the (strict) subvariety of $\mathsf{K}^{n^2(d+1)}$ defined by the zeros of the denominator and the numerator of Δ.

Corollary 2 *Except for a certain subvariety of* $\mathsf{K}^{n^2(d+1)}$*, every nonsingular* $A \in \mathsf{K}[x]^{n\times n}$ *of degree* d *is such that, for* $1 \leq i \leq p$*, the matrices* $B_{i,L}^{(j)}$, $B_{i,R}^{(j)}$, $\underline{U}_i^{(j)}$, $\overline{U}_i^{(j)}$ *computed by Algorithm* `Inverse` *have degree* δ_i*.*

As already mentioned at the beginning of the section, replacing in Algorithm `Inverse` the calls to `MinimalKernelBasis` by the explicit constructions (4a-b) and (4b-c) yields an algebraic straight-line program for inversion.

5 Complexity analysis

We now deduce from Corollaries 1 and 2 the straight-line complexity of Algorithm `Inverse`. When $\Delta(A) \neq 0$, the 2^i minimal bases of step i can be computed by

$$2^i \times O(\nu_i^2 M(\nu_i \delta_i) \log(\nu_i \delta_i)) = O(2^{-i} n^2 M(nd) \log(nd)) \tag{8a}$$

field operations. The update of B consists in multiplying two block-diagonal matrices, each of them having 2^{i-1} diagonal blocks of order ν_i and degree δ_i. This costs $2^{i-1} \times O(\nu_i^\omega M(\delta_i))$ where ω is the exponent for square matrix multiplication. To update the dense matrix U, we update each of its 2^{i-1} block-rows with 2^{i-1} matrix products of order ν_i and degree δ_i. This costs $2^{i-1} \times O(2^{i-1} \nu_i^\omega M(\delta_i))$. The total cost of matrix updates at step i is therefore bounded by

$$2^{i-1} \times O(2^{i-1} \nu_i^\omega M(\delta_i)) = O(2^{(2-\omega)(i-1)} n^\omega M(2^{i-1} d)). \tag{8b}$$

Using $\sum_{i=1}^{p} 2^{-i} \leq 1$, the total cost induced by (8a) is in $O(n^2 M(nd) \log(nd))$. With standard matrix multiplication ($\omega = 3$) and using $M(2^i d) \leq M(2^i) M(d)$, the total cost induced by (8b) is in $O(n^3 M(d))$. It follows that stage (2) of Algorithm `Inverse` has complexity bounded by

$$O(n^2 M(nd) \log(nd))$$

when $\Delta(A) \neq 0$. Stage (3) then consists in computing $B^{-1}U$ with $\deg U = nd - d$ and B diagonal such that $\deg B \leq nd$. This costs

$$O(n^2 M(nd)).$$

Theorem 1 *If* $A \in \mathsf{K}[x]^{n\times n}$ *is nonsingular of degree* d *and* $\Delta(A) \neq 0$ *then Algorithm* `Inverse` *computes* A^{-1} *in* $O(n^2 M(nd) \log(nd))$ *field operations.*

When using the "slow" version of the algorithm in [2] — that is, without Fast Fourier Transform — for computing a minimal basis as in Corollary 1, one would end up with complexity $O(n^3 d^2 \log n)$ instead of $O(n^2 M(nd) \log(nd))$.

The corollary below follows from Theorem 1 by taking $M(d) = d \log d \log\log d$.

Corollary 3 *Except for a subvariety of* $\mathsf{K}^{n^2(d+1)}$*, Algorithm* `Inverse` *computes the inverse of a nonsingular* $A \in \mathsf{K}[x]^{n \times n}$ *of degree* d *in* $O\tilde{}(n^3 d)$ *field operations.*

When ignoring logarithmic factors, we see that (8a) and (8b) respectively read $O\tilde{}(f(i))$ and $O\tilde{}(g_\omega(i))$ where $f(i) = 2^{-i} n^3 d$ and $g_\omega(i) = 2^{(3-\omega)i} n^\omega d$ for $1 \leq i \leq p$. For i ranging from 1 to p, the cost for computing minimal kernel bases therefore decreases from $O\tilde{}(n^3 d)$ to $O\tilde{}(n^2 d)$; simultaneously, the cost of matrix updates increases from $O\tilde{}(n^\omega d)$ to $O\tilde{}(n^3 d)$. Hence basis computations dominate at early stages of the algorithm whereas matrix updates dominate at the end.

We have presented an algorithm for inverting matrix polynomials whose straight-line complexity matches the size of the output up to logarithmic factors. We may note that in a work in progress we show that as a side-effect, our algorithm yields a straight-line program for computing the determinant in $O\tilde{}(n^\omega d)$ operations. To our knowledge, Storjohann's algorithm [18] is the only method that gives this estimate for the determinant problem; this solution is superior to our method since it runs on a random access machine.

A task remaining is to obtain the same complexity estimate for matrix polynomial inversion on a random access machine.

References

[1] E.H. Bareiss, *Computational solution of matrix problems over an integral domain*, J. Inst. Math. Appl., 10 (1972), pp. 68–104.

[2] B. Beckermann and G. Labahn, *A uniform approach for the fast computation of matrix-type Padé approximants*, SIAM Journal on Matrix Analysis and Applications, 15 (1994), pp. 804–823.

[3] B. Beckermann, G. Labahn, and G. Villard, *Shifted normal forms of polynomial matrices*, in Proceedings of the 1999 International Symposium on Symbolic and Algebraic Computation (ISSAC'99), Vancouver, Canada, ACM Press, 1999, pp. 189–196.

[4] J. Bunch and J. Hopcroft, *Triangular factorization and inversion by fast matrix multiplication*, Math. Comp., 28 (1974), pp. 231–236.

[5] P. Bürgisser, M. Clausen and M.A. Shokrollahi, *Algebraic Complexity Theory*, Volume 315, Grundlehren der mathematischen Wissenschaften, Springer-Verlag, 1997.

[6] D.G. Cantor and E. Kaltofen, *On fast multiplication of polynomials over arbitrary algebras*, Acta Informatica, 28(7) (1991), pp. 693–701.

[7] G. D. Forney, Jr., *Minimal bases of rational vector spaces, with applications to multivariable linear systems*, SIAM Journal on Control, 13 (1975), pp. 493–520.

[8] F. R. Gantmacher, *Théorie des matrices*, Editions Jacques Gabay, 1990.

[9] J. von zur Gathen and J. Gerhard, *Modern Computer Algebra*, Cambridge University Press, 1999.

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

[11] J.D. Lipson, *Chinese remainder and interpolation algorithms*, In Proc. 2nd ACM Symposium on Symbolic and Algebraic Manipulation, Los Angeles, CA, ACM Press, 1971, pp. 372–391.

[12] C.-A. Lin, C.-W. Yang and T.-F. Hsieh, *An algorithm for inverting rational matrices*, Systems and Control Letters, 27 (1996), pp. 47–53.

[13] R.T. Moenck and J.H. Carter, *Approximate algorithms to derive exact solutions to systems of linear equations*, In Proc. EUROSAM, LNCS 72, Springer Verlag, 1979, pp. 63–73.

[14] M.-P. Quéré-Stuchlik, *Algorithmique des faisceaux linéaires de matrices - Application à la théorie des systèmes linéaires et à la résolution d'équations algébro-élémentaires*, Thèse de l'Université Paris 6, Paris, France, Juin 1997.

[15] A. Schönhage, *Unitäre Transformationen grosser Matrizen*, Numerishe Math. 20 (1973), pp. 409–417.

[16] A. Schönhage and V. Strassen, *Schnelle Multiplikation Grosser Zahlen*, Computing, 7 (1971), pp. 281–292.

[17] A. Storjohann, *Algorithms for Matrix Canonical Forms*, PhD Thesis, ETH – Swiss Federal Institute of Technology, December 2000.

[18] ——, *High-order lifting*, in Proceedings of the 2002 International Symposium on Symbolic and Algebraic Computation (ISSAC'02), T. Mora, ed., ACM Press, 2002, pp. 246–254.

[19] M.-P. Stuchlik-Quéré, *How to compute minimal bases using Padé approximants*, Tech. Report 1997-035, Laboratoire d'Informatique de Paris 6, December 1997. `http://www.lip6.fr/reports/lip6.1997.035.html`.