J. Austral. Math. Soc. (*Series A*) **27** (1979), 430–436

A LOWER BOUND FOR THE DEGREE OF POLYNOMIALS SATISFIED BY MATRICES

K. R. PEARSON

(Received 2 August 1978; revised 15 January 1979)

Communicated by R. Lidl

Abstract

R. Paré and W. Schelter (1978) have extended the Cayley–Hamilton theorem by showing that for each $n \geqslant 1$ there is an integer k such that all $n \times n$ matrices over any (possibly noncommutative) ring satisfy a monic polynomial of degree k. We give a lower bound for this degree, namely $\pi(n)$, which is defined as the shortest possible length of a sequence with entries from $\{1, 2, \ldots, n\}$ which contains all the permutations of $\{1, 2, \ldots, n\}$.

Subject classification (*Amer. Math. Soc.* (*MOS*) *1970*): 16 A 42, 15 A 24.

Keywords: matrix, noncommutative ring, polynomial, Cayley–Hamilton theorem

1. Introduction and statements of results

Let R denote the free associative $\mathbf{Z}$-algebra generated by the set $\{a_{ij}: i,j \in \mathbf{N}\}$ of indeterminates, and, for all $n \geqslant 1$, let R_n denote the subalgebra generated by the n^2 indeterminates $\{a_{ij}: 1 \leqslant i,j \leqslant n\}$. It has been shown by Paré and Schelter (1978) that for each $n \geqslant 1$ there exists an integer k (depending on n) and a monic (homogeneous) polynomial $f(x) \in R_n\langle x \rangle$ (see below for a precise definition of $R_n\langle x \rangle$) of degree k such that $f(\alpha) = 0$, where α is the $n \times n$ matrix (a_{ij}). (This means, of course, that for all rings S and for all $n \times n$ matrices A over S there is a polynomial $g(x) \in S\langle x \rangle$ of degree k such that $g(A) = 0$.) For each $n \geqslant 1$, we are concerned here with the *least* such k, denoted by $k(n)$. The proof in Paré and Schelter (1978) is by induction on n and yields upper bounds for $k(n)$, namely that $k(1) = 1$ and, for all $n \geqslant 1$, $k(n+1) \leqslant (k(n)+1)^2$. In particular this shows that $k(2) \leqslant 4$, but in fact the existence of Robson's cubic (see Robson (1979)) shows us that $k(2) = 3$. The value of $k(n)$ for $n \geqslant 3$ is unknown.

The purpose of this paper is to give a lower bound for $k(n)$ for all n.

Of all the sequences with entries from $\{1,2,\ldots,n\}$, consider ones of shortest length containing all permutations of $\{1,2,\ldots,n\}$. For example, the sequence 1231231 contains the 6 permutations of $\{1,2,3\}$ (note that a given permutation, for example 213, does not have to have its elements appearing consecutively) while no string of 6 digits contains all the permutations. For each $n \geqslant 1$ we denote the shortest length by $\pi(n)$. (Then $\pi(3) = 7$.) The number $\pi(n)$ has already arisen in combinatorics. While no precise formula for it has been derived, it is known that $\pi(n) \leqslant n^2 - 2n + 4$ (see, for example, Adleman (1974) or Koutas and Hu (1975)) and that, for each $\Delta > 0$, there exists c such that $\pi(n) \geqslant n^2 - cn^{\Delta + 7/4}$ (Kleitman and Kwiatkowski (1976)). For small values of n, $\pi(n)$ is easy to calculate: thus $\pi(1) = 1$, $\pi(2) = 3$, $\pi(3) = 7$, $\pi(4) = 12$, $\pi(5) = 19$.

We prove the following result.

THEOREM. *For all* $n \geqslant 1$, $k(n) \geqslant \pi(n)$.

In particular, this means that $k(3) \geqslant 7$. This lower bound for $k(3)$ was already known as a result of some rather complicated calculations made by the present author (as reported in Robson (1979)). However, these calculations shed no real light on the reason for this lower bound. The methods used in the present paper not only apply for all $n \geqslant 1$, but are also rather intuitive and so give a certain amount of insight into what is going on.

Note that the indeterminate x in the polynomials considered cannot be assumed to commute with the entries from the noncommutative ring R. For this reason we consider polynomials in $R\langle x\rangle$, the free associative $\mathbf{Z}$-algebra generated by x and $\{a_{ij}\colon i,j \in \mathbf{N}\}$. ($R\langle x\rangle$ is also the coproduct of R and $\mathbf{Z}[x]$.)

The proof of the theorem depends on three lemmas which follow. In Section 2 we give the proof of the theorem (assuming these lemmas) while we prove the lemmas in Section 3.

DEFINITION. A monomial in $R\langle x\rangle$ is called *almost Eulerian* (see the proof of Lemma 1 for the reason for this name) if all the a_{ij}'s in it (if any) can be rearranged into an expression of the form

$$(a_{i_1 i_2} a_{i_2 i_3} \ldots a_{i_{s-1} i_s} a_{i_s i_1})(a_{j_1 j_2} a_{j_2 j_3} \ldots a_{j_t j_1}) \ldots (a_{k_1 k_2} a_{k_2 k_3} \ldots a_{k_u k_1}).$$

LEMMA 1. *Let* $n \geqslant 1$ *and suppose that* $f(x) \in R_n\langle x\rangle$ *contains a term* x^m *and let* $g(x)$ *be the sum of all the almost Eulerian monomials in* $f(x)$ *of total degree* m. *(Then* $g(x)$ *also contains the term* x^m.*) If* $f(\alpha) = 0$ *then* $g(\alpha) = 0$.

https://doi.org/10.1017/S1446788700013434 Published online by Cambridge University Press

LEMMA 2. *Consider the generic upper triangular matrix*

$$\alpha_0 = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}.$$

Let T be the subring of R generated by $\{a_{ii}: i \in \mathbf{N}\}$ and let $w(x) \in T\langle x \rangle$. Then $w(\alpha_0) = 0$ if and only if

$$w(x) \in \langle x - a_{11} \rangle \langle x - a_{22} \rangle \dots \langle x - a_{nn} \rangle$$

(where $\langle f \rangle$ denotes the ideal of $T\langle x \rangle$ generated by f).

LEMMA 3. *Let $n \geqslant 1$, let σ be a permutation of $\{1, 2, \dots, n\}$ and let $\alpha\sigma$ denote the $n \times n$ matrix whose (i,j) entry is $a_{i\sigma, j\sigma}$. If $g(x) \in R\langle x \rangle$ is such that $g(\alpha) = 0$ then $g(\alpha\sigma) = 0$ also.*

2. Proof of the theorem

In this section we show how the theorem is a consequence of Lemmas 1, 2 and 3.

PROOF OF THEOREM. Let $n \geqslant 1$ and let $m = k(n)$. Then there is a polynomial $f(x) \in R_n\langle x \rangle$ containing a term x^m such that $f(\alpha) = 0$. By Lemma 1, if $g(x)$ is the sum of all the almost Eulerian monomials in $f(x)$ of total degree m then $g(\alpha) = 0$. Let $h(x)$ denote the sum of all the monomials in $g(x)$ which involve only x, a_{11}, a_{22}, ..., a_{nn}. Then $h(x)$ can be obtained from $g(x)$ by setting all a_{ij}'s with $i \neq j$ equal to 0.

For any $n \times n$ matrix A, let A_0 denote the upper triangular part of A: that is $(A_0)_{ij} = A_{ij}$ if $i \leqslant j$ and $(A_0)_{ij} = 0$ if $i > j$.

Let σ be a permutation of $\{1, 2, \dots, n\}$. We claim that $h((\alpha\sigma)_0) = 0$. For, by Lemma 3, $g(\alpha\sigma) = 0$, and so, if we consider the homomorphism φ from $R_n\langle x \rangle$ to $R_n\langle x \rangle$ which maps a_{ij} to 0 if $i\sigma^{-1} > j\sigma^{-1}$ and leaves x and all other a_{kl} fixed, then φ maps $\alpha\sigma$ to $(\alpha\sigma)_0$ and, because all monomials in g are almost Eulerian, φ maps g to h. (For if we consider a monomial which is in g but not in h, it contains (in some order) a factor

$$a_{i_1 i_2} a_{i_2 i_3} \dots a_{i_{s-1} i_s} a_{i_s i_1},$$

where $i_1 \neq i_2$, $i_2 \neq i_3$, ..., $i_{s-1} \neq i_s$, $i_s \neq i_1$ and clearly for at least one t, $i_t\sigma^{-1} > i_{t+1}\sigma^{-1}$ (where $i_{s+1} = i_1$) so that this term is mapped to zero by φ.) Now φ also acts as an endomorphism of the ring of $n \times n$ matrices over R_n and it is easy to see that, for all $w \in R_n\langle x \rangle$ and for all $n \times n$ matrices β over R_n, $w^{\varphi}(\beta\varphi) = (w(\beta))\varphi$. In

particular,

$$h((\alpha\sigma)_0) = g^{\varphi}((\alpha\sigma)\varphi) = (g(\alpha\sigma))\varphi = 0\varphi = 0,$$

as claimed. We also have an automorphism τ of $R_n\langle x\rangle$ mapping x to x and a_{ij} to $a_{i\sigma^{-1},j\sigma^{-1}}$. Notice that $(\alpha\sigma)_0\tau = \alpha_0$. So, as above,

$$h^{\tau}(\alpha_0) = h^{\tau}((\alpha\sigma)_0\tau) = (h((\alpha\sigma)_0))\tau = 0$$

and we see from Lemma 2 that

$$h^{\tau} \in \langle x-a_{11}\rangle\langle x-a_{22}\rangle \dots \langle x-a_{nn}\rangle$$

which means that

$$h = (h^{\tau})^{\tau^{-1}} \in \langle x-a_{1\sigma,1\sigma}\rangle\langle x-a_{2\sigma,2\sigma}\rangle \dots \langle x-a_{n\sigma,n\sigma}\rangle.$$

The elements

$$y_0 = x, \quad y_1 = x-a_{11}, \quad \dots, \quad y_n = x-a_{nn}, a_{n+1,n+1}, a_{n+2,n+2}, \dots$$

form a free generating set for the free algebra $T\langle x\rangle$ freely generated by $x, a_{11}, a_{22}, \dots$. We have seen above that, for each permutation σ of $\{1,2,\dots,n\}$,

$$h \in \langle y_{1\sigma}\rangle\langle y_{2\sigma}\rangle \dots \langle y_{n\sigma}\rangle$$

so that, when h is expressed in terms of $x, y_1, \dots, y_n$, every term has a $y_{i\sigma}$ to the left of a $y_{(i+1)\sigma}$ for each $1 \leqslant i \leqslant n-1$. Thus if h contains the term $y_{i_1}y_{i_2}\dots y_{i_m}$ $(0 \leqslant i_j \leqslant n)$ with a nonzero coefficient we see that the nonzero integers in $\{i_1, i_2, \dots, i_m\}$ are a sequence from $\{1,2,\dots,n\}$ which contains all permutations of $\{1,2,\dots,n\}$; hence $m \geqslant \pi(n)$.

3. Proofs of Lemmas 1, 2 and 3

I am very grateful to Dr. J. C. Robson for suggesting the following proof of Lemma 1 which is much shorter and neater than the one I originally submitted with this paper.

PROOF OF LEMMA 1. (J. C. Robson.) Suppose that the matrix $g(\alpha)$ contains a nonzero (k,l) entry, and let v be one of the monomials (in the a_{ij}'s) with nonzero coefficient in this entry. Then v is one of the monomials in the (k,l) entry of one of the monomials, say w_1, in $g(x)$; so v has degree m. Let $h(x) = f(x) - g(x)$. Then, since $f(\alpha) = 0$, $h(\alpha) = -g(\alpha)$ and so $h(\alpha)$ has a nonzero (k,l) entry in which v is one of the monomials with nonzero coefficient. Thus v is one of the monomials in the (k,l) entry of one of the monomials, say w_2, in $h(x)$. We will show below that w_2 has total degree m and is almost Eulerian which means that it should have

been included in $g(x)$ (not $h(x)$) and so is a contradiction. Because w_1 is almost Eulerian, we can rearrange the terms in w_1 so that they are of the form $\lambda_1 \lambda_2 \dots \lambda_t x^r$ where each λ_i is a monomial in a diagonal entry of some power of α. Then after rearrangement, v becomes $\lambda_1 \lambda_2 \dots \lambda_t \theta$ where $\theta = a_{kc_1} a_{c_1 c_2} \dots a_{c_{r-1} l}$. Because v is also one of the monomials in the (k,l) entry of w_2, w_2 must be obtained from v by singling out certain of the a_{ij}'s in v, say $a_{kd_1}, a_{d_1 d_2}, \dots, a_{d_{q-1} l}$ and replacing each of them by x. Clearly w_2 has total degree m.

It is Robson's good idea to associate with each monomial u in the a_{ij}'s a directed graph with n vertices $1, 2, \dots, n$ and with one arrow from vertex i to vertex j for each letter a_{ij} in u. If the set of arrows in this directed graph can be decomposed as a disjoint union of cycles, we call the directed graph *almost Eulerian*. Notice that a monomial in $R\langle x \rangle$ is almost Eulerian precisely when the a_{ij}'s in it give rise to an almost Eulerian graph. A simple modification of Euler's Theorem (see, for example, Wilson (1972), Theorem 23A or Harary *et al.* (1965), Theorems 12.5 and 12.6) shows that a directed graph is almost Eulerian if and only if the number of arrows to each vertex equals the number from the vertex (that is, if and only if each connected component is Eulerian). In particular, if any cycle is removed from an almost Eulerian graph, the remaining graph is still almost Eulerian.

Clearly the monomial va_{lk} is almost Eulerian and so, when we delete the cycle $a_{kd_1}, a_{d_1 d_2}, \dots, a_{d_{q-1} l}, a_{lk}$ from its graph we are left with an almost Eulerian graph. As this is the graph of the a_{ij}'s in w_2 we see that w_2 is almost Eulerian, which is the desired contradiction.

PROOF OF LEMMA 2. An element $w(x)$ of $\langle x - a_{11} \rangle \dots \langle x - a_{nn} \rangle$ is a sum of terms of the form

$$q(x) = p_0(x)(x - a_{11}) p_1(x)(x - a_{22}) p_2(x) \dots p_{n-1}(x)(x - a_{nn}) p_n(x).$$

If $r_i(x) = p_{i-1}(x)(x - a_{ii})$ for $1 \leqslant i \leqslant n$, $r_i(\alpha_0)$ has its (i,i) entry zero. Thus it is easy to see by induction on i that if $1 \leqslant i \leqslant n$, $r_1(\alpha_0) r_2(\alpha_0) \dots r_i(\alpha_0)$ has its first i columns zero. Hence $q(\alpha_0) = 0$ and so $w(\alpha_0) = 0$.

The converse is clear if $n = 1$. To prove the converse in general we suppose $n > 1$ and use induction on n. Let $w(x, a_{11}, a_{22}, \dots)$ in $T\langle x \rangle$ be such that $w(\alpha_0, a_{11}, a_{22}, \dots) = 0$. We use induction on the (total) degree of w. (The result is trivial if w has total degree 1.) We use the fact that $T\langle x \rangle$ is freely generated by

$$x, x - a_{11}, \dots, x - a_{nn}, a_{n+1,n+1}, a_{n+2,n+2}, \dots \tag{2}$$

and write w as a polynomial in these elements, say

$$w = g(x, x - a_{11}, \dots, x - a_{nn}, a_{n+1,n+1}, \dots)$$

https://doi.org/10.1017/S1446788700013434 Published online by Cambridge University Press

which we write as

$$xg_0+(x-a_{11})g_1+\ldots+(x-a_{nn})g_n+a_{n+1,n+1}g_{n+1}+\ldots$$

(where the g_i's are polynomials in the elements in (2)) and finally rewrite each g_i as a polynomial in $x, a_{11}, a_{22}, \ldots$, say

$$g_i(x, x-a_{11}, \ldots, x-a_{nn}, a_{n+1,n+1}, \ldots) = w_i(x, a_{11}, a_{22}, \ldots).$$

Thus we have

$$w = xw_0+(x-a_{11})w_1+\ldots+(x-a_{nn})w_n+a_{n+1,n+1}w_{n+1}+\ldots.$$

For each $t\geqslant 0$ we let $c_{ij}^{(t)}$ $(i\leqslant j)$ denote the (i,j) entry of the upper triangular matrix $C^{(t)} = w_t(\alpha_0, a_{11}, a_{22}, \ldots)$. Then we have

$$0 = \alpha_0 C^{(0)}+(\alpha_0-a_{11})C^{(1)}+\ldots+(\alpha_0-a_{nn})C^{(n)}+a_{n+1,n+1}C^{(n+1)}+\ldots. \tag{3}$$

For each $0\leqslant k\leqslant n-1$, we claim that $c_{ij}^{(t)} = 0$ for $1\leqslant i\leqslant j\leqslant \min(n, i+k)$ and $t\geqslant 0$ unless $i = t = 1$ or unless $j = i+k$ and $2\leqslant i = t\leqslant n$. We prove this assertion by induction on k. For the case $k = 0$, notice that, from the (i,i) entries on both sides of (3),

$$0 = a_{ii}c_{ii}^{(0)}+\sum_{t=1, t\neq i}^{n}(a_{ii}-a_{tt})c_{ii}^{(t)}+\sum_{t=n+1}^{\infty}a_{tt}c_{ii}^{(t)}$$

and so $c_{ii}^{(t)} = 0$ if $t = 1, 2, \ldots, i-1, i+1, \ldots$ and hence also if $t = 0$, which is what is claimed in this case. If $0\leqslant k<n-1$ and if the result is assumed true for k then the result for $k+1$ follows by considering the $(i, i+k+1)$ entry of (3) (where $i+k+1\leqslant n$).

In particular, the case $k = n-1$ tells us that $C^{(t)} = 0$ if $t\neq 1$ and that $c_{ij}^{(1)} = 0$ unless $i = 1$. Thus α_0 is a zero of all w_t with $t\neq 1$ and so, by induction on the degree of w, $w_t\in I = \langle x-a_{11}\rangle \ldots \langle x-a_{nn}\rangle$ for $t\neq 1$, which means that

$$xw_0, (x-a_{22})w_2, \ldots, (x-a_{nn})w_n, a_{n+1,n+1}w_{n+1}, \ldots \in I.$$

Also if β_0 is the submatrix formed by the last $(n-1)$ rows and columns of α_0, we see that the last $(n-1)$ rows and columns of $C^{(1)} = w_1(\alpha_0, a_{11}, \ldots)$ are just $w_1(\beta_0, a_{11}, a_{22}, \ldots)$ since α_0 is upper triangular. But we know that all entries in this part of $C^{(1)}$ are zero. Hence $w_1(\beta_0, a_{11}, a_{22}, \ldots) = 0$ and, by induction on n,

$$w_1(x, a_{11}, a_{22}, \ldots)\in\langle x-a_{22}\rangle \ldots \langle x-a_{nn}\rangle.$$

Thus $(x-a_{11})w_1\in I$ and we have shown that $w\in I$.

https://doi.org/10.1017/S1446788700013434 Published online by Cambridge University Press

NOTE. Lemma 2 fails if the hypothesis 'w$(x) \in T\langle x\rangle$' is replaced by '$w(x) \in R\langle x\rangle$' or even '$w(x) \in R_n\langle x\rangle$'. For notice that if $n = 2$ and

$$w(x) = x(x-a_{11})(x-a_{22}) + a_{12}(x-a_{22})(x-a_{11}) - (x-a_{22})a_{12}(x-a_{11}) + (x-a_{22})(x-a_{11})a_{12}$$

then it is easy to see that $w(\alpha_0) = 0$ but $w \notin \langle x-a_{11}\rangle\langle x-a_{22}\rangle$.

PROOF OF LEMMA 3. Let σ be a permutation of $\{1, 2, \ldots, n\}$ and let $P = (p_{ij})$ be the $n \times n$ permutation matrix such that $p_{ij} = 1$ if $j = i\sigma$ and $p_{ij} = 0$ otherwise. It is easy to see that $P\alpha P^{-1} = \alpha\sigma$. Now all entries in P are 0 or 1 so that $a_{ij}P = Pa_{ij}$ for all i and j. Thus if $w(x) \in R\langle x\rangle$,

$$w(\alpha\sigma) = w(P\alpha P^{-1}) = Pw(\alpha)P^{-1}.$$

If $g(\alpha) = 0$ then $g(\alpha\sigma) = Pg(\alpha)P^{-1} = 0$.

References

L. Adleman (1974), 'Short permutation strings', *Discrete Math.* **10**, 197–200.

Frank Harary, Robert Z. Norman and Dorwin Cartwright (1965), *Structural models: an introduction to the theory of directed graphs* (John Wiley, New York).

D. J. Kleitman and D. J. Kwiatkowski (1976), 'A lower bound on the length of a sequence containing all permutations as subsequences', *J. Comb. Theory* **21**, 129–136.

P. J. Koutas and T. C. Hu (1975), 'Shortest string containing all permutations', *Discrete Math.* **11**, 125–132.

R. Paré and W. Schelter (1978), 'Finite extensions are integral', *J. Alg.* **53**, 477–479.

J. C. Robson (1979), 'Polynomials satisfied by matrices', *J. Alg.* **55**, 509–520.

Robin J. Wilson (1972), *Introduction to graph theory* (Oliver and Boyd, Edinburgh).

La Trobe University
Bundoora, Victoria, 3083
Australia