Letters in Mathematical Physics **13** (1987) 299–302.

A Formula for the Determinant of a Sum of Matrices

CHRISTOPHE REUTENAUER★ and MARCEL-PAUL SCHÜTZENBERGER★★
Math-Info, Université du Québec à Montréal, C.P.8888, Succursale "A", Montréal, Canada H3C 3P8

(Received: 20 January 1987)

Abstract. We give a formula, involving circular words and symmetric functions of the eigenvalues, for the determinant of a sum of matrices. Theorem of Hamilton–Cayley is deduced from this formula.

Given a square matrix x over a commutative ring, define functions $\Lambda_i(x)$ by the equality

$$\det(1 - tx) = 1 - t\Lambda_1(x) + t^2\Lambda_2(x) + \cdots + (-1)^n t^n \Lambda_n(x) + \cdots \tag{1}$$

Of course, these functions are the coefficients of the characteristic polynomial of x. In particular, Λ_1 is the trace, and if x is of order n, Λ_n is the determinant and $\Lambda_{n+1} = 0 = \Lambda_{n+2} = \ldots$. Note also that the functions Λ_i are invariant under conjugation, or equivalently

$$\Lambda_i(uv) = \Lambda_i(vu). \tag{2}$$

We shall give a formula expressing $\Lambda_n(x + y + \cdots)$ as a polynomial in the $\Lambda_i(w)$, where $i \leqslant n$ and where w is a product of x and y's.

We start with the example $n = 3$, illustrated by Figure 1. In fact, on this figure, we have drawn all N-sets (sets with multiplicities) of primitive (without period) circular words on x, y of cardinality 3. Each N-set gives rise to a monomial in the $\Lambda_i(w)$, w being determined by the circular words and i being its multiplicity; moreover, the sign is computed as for a permutation (+ for a word of odd length, − for a word of even length).

Then $\Lambda_3(x + y)$ *is the sum of the* 8 *monomials obtained above.*

More generally, let $X = \{x, y, \ldots\}$ be an alphabet. A *circular word* on X is a conjugation class of words on X; recall that two words are *conjugate* if they may be written respectively uv and vu, for some words u and v. A circular word is *primitive* if it has no period. Define the *length* $|c|$ of c to be the length of any word representing it, and its *sign* to be $\text{sgn}(c) = +1$ if its length is odd, -1 if it is even. Hence, $\text{sgn}(c) = (-1)^{|c|+1}$.

Let

$$m = c_1^{i_1} \ldots c_q^{i_q} \tag{3}$$

★ UQAM and LITP, ★★ Université Paris 7 and LITP.

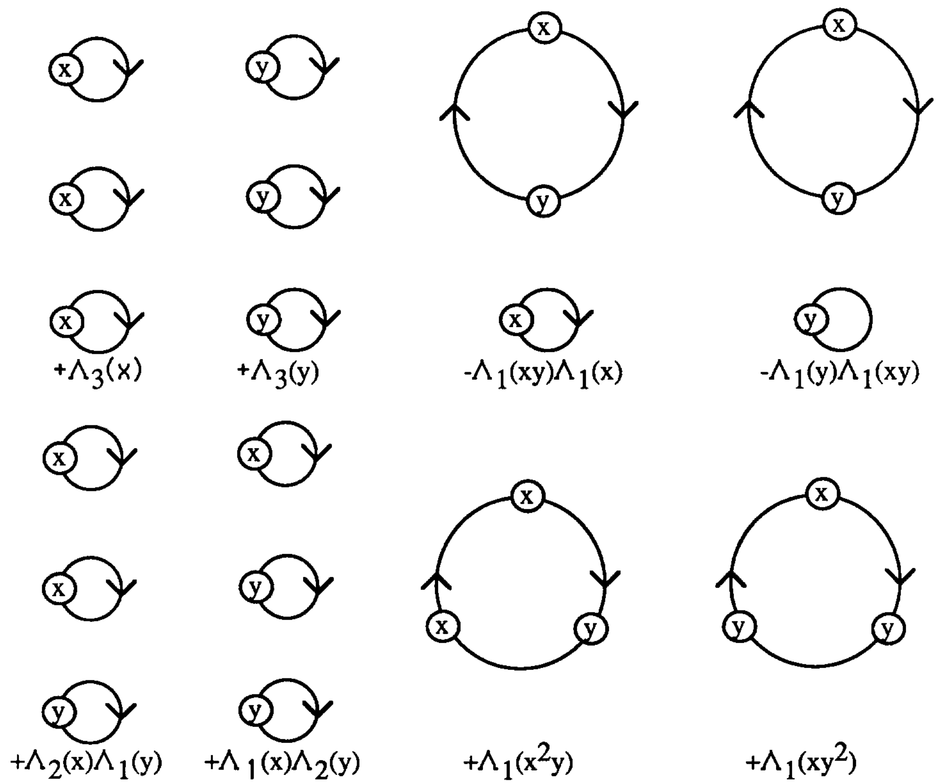

Fig. 1.

be a *monomial of primitive circular words*. Then by definition its length is $\Sigma_j i_j \mid c_j \mid$ and its *sign* is $\Pi_j (\text{sign } c_j)^{i_j}$.

If w is a word and $i \geqslant 1$, define $\Lambda_i(w)$ to be the matrix function obtained in the obvious manner.

Note that if w, w' are conjugate, then $\Lambda_i(w) = \Lambda_i(w')$ in view of Equation (2). Hence, $\Lambda_i(c)$ is a well defined matrix function for any circular word c.

More generally, for m as in Equation (3), let $\Lambda(m)$ be defined by $\Lambda(m) = \Pi_j \Lambda_{i_j}(c_j)$, where the c_j's of Equation (3) are assumed to be distinct.

We can now state and prove our main result.

THEOREM *Let $x_1, \ldots, x_k$ be square matrices of the same order and $n \geqslant 1$. Then*

$$\Lambda_n(x_1 + \cdots + x_k) = \sum_m \text{sgn}(m)\Lambda(m) \tag{4}$$

where the sum is extended to the k^n monomials of length n of primitive circular words on $x_1, \ldots, x_k$.

Proof. Let $A = \{a_1 < \ldots < a_k\}$ be a totally ordered alphabet. A *Lyndon word* is a primitive word which is the smallest element of its conjugation class (for the lexicographic order on the free monoid A^* generated by A). Obviously, Lyndon words are in bijection with primitive circular words. By a theorem of Lyndon (see [2] th. 5.1.5), each word w in A^* may be written uniquely as

$$w = l_1^{i_1} \dots l_q^{i_q}$$

where the l_j's are Lyndon words such that $l_1 > \dots > l_q$ and $i_1, \dots, i_q \geqslant 1$. In the algebra of noncommutative power series on A over $\mathbb{Z}$, this is written

$$(1 - a_1 - \dots - a_k)^{-1} = \prod_l (1 - l)^{-1} \tag{5}$$

where the product is taken over all Lyndon words in decreasing order. Now, let $x_1, \dots, x_k$ be generic matrices (it is enough to prove the theorem in this case). Then invert Equation (5), apply the homomorphism $a_i \to x_i$ and take the determinant. We obtain

$$\det(1 - x_1 - \dots - x_k) = \prod \det(1 - l) \tag{6}$$

where we still write l for the image of l under the above homomorphism. Now, observe that

$$\det(1 - x) = 1 - \Lambda_1(x) + \Lambda_2(x) + \cdots + (-1)^n \Lambda_n(x) + \cdots .$$

Hence, we obtain

$$\begin{aligned} \sum_{i \geqslant 0} (-1)^i \Lambda_i(x_1 + \cdots + x_k) &= \prod_l \left(\sum_{i \geqslant 0} (-1)^i \Lambda_i(l) \right) \\ &= \sum_{\substack{l_1 > \dots > l_q \\ i_1, \dots, i_q \geqslant 0}} (-1)^{i_1 + \dots + i_q} \Lambda_{i_1}(l_1) \dots \Lambda_{i_q}(l_q) . \end{aligned} \tag{7}$$

Taking on both sides the terms of degree n, we obtain almost Equation (4). To conclude, observe that if m is defined by Equation (3) and is of length n, then

$$\operatorname{sgn}(m) = \prod_j \operatorname{sgn}(c_j)^{i_j} = (-1)^{\Sigma i_j |c_j|} (-1)^{\Sigma i_j} = (-1)^{n + \Sigma i_j}$$

which derives completely Equation (4) from Equation (7). □

We now show how the theorem of Hamilton–Cayley may be derived from Equation (4). Take two generic matrices x, y of order n. Then

$$\Lambda_{n+1}(x + y) = 0 .$$

Now, using Equation (4), take in $\Lambda_{n+1}(x + y)$ all the terms of degree n in x, 1 in y. By homogeneity, their sum is equal to 0. But these terms are:

$$\sum_{i=0}^{n} (-1)^i \Lambda_{n-i}(x) \Lambda_1(x^i y) = 0 .$$

Now, Λ_1 is linear, hence

$$\Lambda_1\left(\left(\sum_{i=0}^{n}(-1)^i\Lambda_{n-i}(x)x^i\right)y\right)=0\,.$$

Now, it is well-known that $\operatorname{tr}(ay)=0$ for any matrix y, implies $a=0$.

Thus, we obtain

$$\sum_{i=0}^{n}(-1)^i\Lambda_{n-i}(x)x^i=0$$

which is the Hamilton–Cayley theorem. It is also possible to derive directly from Equation (4) the multilinear version of the HC theorem, well-known to pi-algebraists: take the multilinear part of the equation $\Lambda_{n+1}(x_1,\dots,x_{n+1})=0$, where the x_i's are $n+1$ generic matrices of order n.

REMARKS: (1) In the proof of the theorem, we have used Lyndon words and the fact that they provide a *factorization* of the free monoid (see [2] chapter 5). In fact, in view of corollary 5.4.2 of [2], any complete factorization would also work for the proof. The interest of Lyndon words is that Equation (4) may be efficiently computed using them: generate all the words w of length n on $x_1,\dots,x_k$, then decompose them into Lyndon words using Duval's linear algorithm [1].

(2) Let m_n denote the number of terms in Equation (4) whose sign is $-$ (the total number of terms is k^n). It may be shown that $m_{2n}=(k^{2n}-k^n)/2$ and $m_{2n+1}=(k^{2n+1}-k^{n+1})/2$. Hence, there are asymptotically as many $+$ as $-$ in the formula. Note that m_n is also the number of words of length n having an odd number of Lyndon words of even length in their decomposition (and similarly for any complete factorization of A^*).

Acknowledgements

This work was done during a three weeks stay of the second author, on invitation of the Groupe de Combinatoire de l'UQAM, with support of the Fondation de l'UQAM.

References

1 Duval, J.-P., Factorizing words over an ordered alphabet, *J. Algorithms* **4**, 363–381 (1983).
2 Lothaire, M., *Combinatorics on Words*, Addison-Wesley, 1983.

Added in proof: The authors have learnt that S. A. Amitsur had already proved, by a different method, the formula of the theorem: On the characteristic polynomial of a sum of matrices, *Linear and Multilinear Algebra* **8**, 177–182 (1980).