A GENERALIZATION OF OSTROWSKI'S THEOREM ON MATRIC IDENTITIES[1]

NEAL H. McCOY

The purpose of this note is to generalize a recent theorem due to Ostrowski[2] which is itself a generalization of a theorem proved by Phillips in 1919.[3] We shall first indicate the nature of Ostrowski's result.

Let $A_1=I, A_2, \cdots, A_m$ be square matrices of order n, I being the unit matrix, and let $x_1, \cdots, x_m$ be numerical parameters. Denote by $F(x_1, \cdots, x_m)$ the determinant of the matrix

$$x_1A_1 + x_2A_2 + \cdots + x_mA_m. \tag{1}$$

Let $\Phi(x_1, \cdots, x_m)$ be the greatest common divisor of the n^2 minors of order $n-1$ of the matrix (1), and set $F/\Phi=F^*(x_1, \cdots, x_m)$. We may now state the theorem of Ostrowski in the following form:[4]

THEOREM 1. *If $B_1, \cdots, B_m$ are matrices of order n, commutative with each other and satisfying the equation*

$$A_1B_1 + A_2B_2 + \cdots + A_mB_m = 0, \tag{2}$$

then

$$F^*(B_1, \cdots, B_m) = 0.$$

Further, if $\Psi(x_1, \cdots, x_m)$ is any polynomial with the property that $\Psi(B_1, \cdots, B_m)=0$ for every set of commutative matrices satisfying (2), *then $\Psi(x_1, \cdots, x_m)$ is divisible by $F^*(x_1, \cdots, x_m)$.*

In this theorem it is tacitly assumed that the elements of the matrices as well as the coefficients of the polynomials are real or complex numbers. In Theorem 3 below we find an extension of the first part of Theorem 1, valid if the elements and coefficients are in an arbitrary commutative ring R with unit element 1. To generalize the second part of the theorem, we find it necessary to make an additional restriction on R, namely, that there exists no nonzero polynomial $\phi(\lambda)$,

[1] Presented to the Society, September 8, 1939.

[2] A. Ostrowski, *On a theorem concerning identical relations between matrices*, Quarterly Journal of Mathematics, vol. 9 (1938), pp. 241–245.

[3] H. B. Phillips, *Functions of matrices*, American Journal of Mathematics, vol. 41 (1919), pp. 266–278.

[4] The assumption that $A_1=I$ is not strictly necessary but assures us that $F(x_1, \cdots, x_m)$ does not vanish identically. For the generalization below, we wish to have $A_1=I$ and so we state the theorem at once in this form.

with coefficients in R, such that $\phi(a)=0$ for all elements a of R. The result obtained under this restriction is stated as Theorem 4.

The form of our theorems is suggested by a recent generalization, in another direction, of Frobenius' theorem concerning the minimum equation of a matrix.[5] Since this plays an important part in the proof of Theorem 4, we state it explicitly before proceeding.

THEOREM 2. *Let R be an arbitrary commutative ring with unit element* 1, *and A a matrix of order n with elements in R. Let λ be an indeterminate, denote by $f(\lambda)$ the determinant of the matrix $\lambda I-A$, and let $h_{ij}(\lambda)$ be the minors of $\lambda I-A$ of order $n-1$. Then, an element $g(\lambda)$ of $R[\lambda]$ has the property that $g(A)=0$, if and only if $g(\lambda)h_{ij}(\lambda)\equiv 0\ (f(\lambda))$, $(i, j=1, 2, \cdots, n)$.*

It will be seen that Theorems 3 and 4 bear roughly the same relation to Ostrowski's theorem that Theorem 2 does to Frobenius'.

Henceforth we shall let $A_1=I, A_2, \cdots, A_m$ be matrices of order n with elements in a commutative ring R with unit element 1, and let $x_1, \cdots, x_m$ be indeterminates. Denote by $F(x_1, \cdots, x_m)$ the determinant of the matrix

$$x_1A_1 + x_2A_2 + \cdots + x_mA_m. \tag{3}$$

Let $F_{ij}(x_1, \cdots, x_m)$ be the elements of the adjoint of this matrix, and denote by $\mathfrak{m}$ the ideal of those elements $f(x_1, \cdots, x_m)$ of the ring $R[x_1, \cdots, x_m]$ such that

$$f(x_1, \cdots, x_m)F_{ij}(x_1, \cdots, x_m) \equiv 0\ (F(x_1, \cdots, x_m)),\ i, j = 1, 2, \cdots, n.$$

We may now state the following theorem:

THEOREM 3. *If $f(x_1, \cdots, x_m)\equiv 0\ (\mathfrak{m})$,* and *$B_1, \cdots, B_m$ are commutative matrices of order n, with elements in R, such that*

$$A_1B_1 + A_2B_2 + \cdots + A_mB_m = 0, \tag{4}$$

then $f(B_1, \cdots, B_m)=0$.

The proof is a simple modification of Ostrowski's, and will be only briefly indicated, using his notation so far as possible. Set $A_\mu=(a_{ik}^{(\mu)})$. Now, by hypothesis, we have equations of the form

$$f(x_1, \cdots, x_m)F_{ij}(x_1, \cdots, x_m) = h_{ij}(x_1, \cdots, x_m)F(x_1, \cdots, x_m). \tag{5}$$

But

[5] N. H. McCoy, *Concerning matrices with elements in a commutative ring*, this Bulletin, vol. 45 (1939), pp. 280–284.

$$\sum_{j=1}^{n}\left(\sum_{\mu=1}^{m} x_\mu a_{jk}^{(\mu)}\right)F_{ji}(x_1, \cdots, x_m) = \delta_i^k F(x_1, \cdots, x_m).$$

Multiply this last equation by $f(x_1, \cdots, x_m)$ and use (5), getting

$$F(x_1, \cdots, x_m)\sum_{j=1}^{n}\left(\sum_{\mu=1}^{m} x_\mu a_{jk}^{(\mu)}\right)h_{ji}(x_1, \cdots, x_m) = \delta_i^k F(x_1, \cdots, x_m) f(x_1, \cdots, x_m). \tag{6}$$

Now, if $F(x_1, \cdots, x_m)$ is arranged in terms of decreasing powers of x_1, it is clear that the first term is x_1^n. It follows readily that, in the ring $R[x_1, \cdots, x_m]$, $F(x_1, \cdots, x_m)$ is not a divisor of zero, and can therefore by divided out of equation (6), yielding

$$\sum_{j=1}^{n}\left(\sum_{\mu=1}^{m} x_\mu a_{jk}^{(\mu)}\right)h_{ji}(x_1, \cdots, x_m) = \delta_i^k f(x_1, \cdots, x_m).$$

Let $C_{jk}=\sum_{\mu=1}^{m} B_\mu a_{jk}^{(\mu)}$. Then we have from the preceding equation, and the fact that the B's are commutative,

$$\sum_{j=1}^{n} C_{jk}h_{ji}(B_1, \cdots, B_m) = \delta_i^k f(B_1, \cdots, B_m).$$

Now this equation corresponds to Ostrowski's equation (2.3), and the remainder of the proof will be omitted as from this point the proof coincides with his.

If R is a field, or more generally, a domain of integrity with unique factorization into primes, then it follows readily that $\mathfrak{m}$ is the principal ideal $(F(x_1, \cdots, x_m)/D(x_1, \cdots, x_m))$, where $D(x_1, \cdots, x_m)$ is the greatest common divisor of the $F_{ij}(x_1, \cdots, x_m)$. In this case, our Theorem 3 can be expressed in the same form as the first part of Theorem 1.

Before proceeding, we pause to make a remark which will indicate how, in another way, the ideal $\mathfrak{m}$ has properties generalizing the familiar properties of the minimum function of a single matrix with elements in a field. From the definition of $\mathfrak{m}$, it follows that if $g(x_1, \cdots, x_m) \equiv 0\ (\mathfrak{m})$, then

$$g(x_1, \cdots, x_m)\ \text{adj}\ (x_1A_1 + \cdots + x_mA_m) = KF(x_1, \cdots, x_m),$$

where K is a matrix with elements in $R[x_1, \cdots, x_m]$. By taking determinants and dividing by F^{n-1}, we see that

$$[g(x_1, \cdots, x_m)]^n \equiv 0\ (F(x_1, \cdots, x_m)).$$

In particular, it follows at once that in $R[x_1, \cdots, x_m]$ the ideals $\mathfrak{m}$

and $(F(x_1, \cdots, x_m))$ have precisely the same prime ideal divisors.

We now turn to the problem of generalizing the second part of Theorem 1, and we find it necessary to make the following additional assumption concerning R. *If* $f(\lambda)$ *is an element of* $R[\lambda]$ *such that for every element* a *of* R, $f(a)=0$, *then* $f(\lambda)=0$. It follows readily by induction on the number of indeterminates that if in $R[x_1, \cdots, x_k]$, $f(x_1, \cdots, x_k)$ has the property that $f(a_1, \cdots, a_k)=0$ for all choices of $a_1, \cdots, a_k$ in R, then $f(x_1, \cdots, x_k)=0$, $(k=1, 2, \cdots)$. We now prove the following lemma:

LEMMA. *Let* $x_1, \cdots, x_m$ *be indeterminates, and consider an element of* $R[x_1, \cdots, x_m]$ *of the form*

$$g(x_1, \cdots, x_m) = x_1^n + g_1(x_2, \cdots, x_m)x_1^{n-1} + \cdots + g_n(x_2, \cdots, x_m).$$

If $f(x_1, \cdots, x_m)$ *is an element of* $R[x_1, \cdots, x_m]$ *such that for every choice of* $a_2, \cdots, a_m$ *in* R, $f(x_1, a_2, \cdots, a_m)$ *is divisible by* $g(x_1, a_2, \cdots, a_m)$ *in* $R[x_1]$, *then* $f(x_1, \cdots, x_m)$ *is divisible by* $g(x_1, \cdots, x_m)$ *in the ring* $R[x_1, \cdots, x_m]$.

Let x represent the composite set $x_2, \cdots, x_m$; and a the set of elements $a_2, \cdots, a_m$ of R. Then we may write

$$g(x_1, \cdots, x_m) = x_1^n + g_1(x)x_1^{n-1} + \cdots + g_n(x),$$

and

$$f(x_1, \cdots, x_m) = f_0(x)x_1^p + f_1(x)x_1^{p-1} + \cdots + f_p(x).$$

Now, by hypothesis, we have for arbitrary but fixed a a relation of the form

$$f_0(a)x_1^p + \cdots + f_p(a) = [x_1^n + g_1(a)x_1^{n-1} + \cdots + g_n(a)] \cdot [h_0x_1^{p-n} + h_1x_1^{p-n-1} + \cdots + h_{p-n}], \tag{7}$$

where the h_i are elements of R. But equation (7) is equivalent to a set of $p+1$ equations in R, obtained by equating the coefficients of the different powers of x_1 on each side of (7). These equations take the form

$$\begin{aligned} f_0(a) &= h_0, \\ f_1(a) &= h_1 + h_0g_1(a), \\ &\cdots\cdots\cdots\cdots, \\ f_{p-n}(a) &= h_{p-n} + h_{p-n-1}g_1(a) + \cdots + h_0g_{p-n}(a), \end{aligned} \tag{8}$$

together with the set of equations

$$f_t(a) = \sum_{i+j=t} h_i g_j(a), \qquad t = p - n + 1, \cdots, p. \tag{9}$$

The equations (8) can be solved in turn for the h_i, and these unique solutions take the form

$$h_i = G_i[f_1(a), \cdots, f_{p-n}(a); g_1(a), \cdots, g_{p-n}(a)], \quad i = 0, 1, \cdots, p - n,$$

where the G_i are polynomials with integral coefficients. Thus, in (7), the coefficients h_i are uniquely determined by the choice of a. Now let us set

$$h_i(x) = G_i[f_1(x), \cdots, f_{p-n}(x); g_1(x), \cdots, g_{p-n}(x)], \qquad i = 0, 1, \cdots, p - n, \tag{10}$$

so that our original h_i is $h_i(a)$. Then equations (9) state that for *every* a in R,

$$f_t(a) - \sum_{i+j=t} h_i(a) g_j(a) = 0,$$

and therefore, by hypothesis on R, it follows that

$$f_t(x) = \sum_{i+j=t} h_i(x) g_j(x), \qquad t = p - n + 1, \cdots, p. \tag{11}$$

But equations (10) and (11) are precisely the set of equations which state that

$$f(x_1, \cdots, x_m) = g(x_1, \cdots, x_m)[h_0(x) x_1^{p-n} + \cdots + h_{p-n}(x)],$$

and the lemma is established.

We shall now prove the following theorem under the assumption on R which we have made throughout this section:

THEOREM 4. *If $\phi(x_1, \cdots, x_m)$ is an element of $R[x_1, \cdots, x_m]$ with the property that $\phi(B_1, \cdots, B_m) = 0$ for every $B_1, \cdots, B_m$ which are commutative and satisfy* (4), *then $\phi(x_1, \cdots, x_m) \equiv 0$ (m).*

In the proof of this theorem we shall not distinguish between the ring R and the ring of matrices of the form aI, where a is in R. Accordingly, we identify I with 1, the unit element of R.

Let $a_2, \cdots, a_m$ be arbitrary elements of R, and let us choose

$$B_1 = -a_2 A_2 - \cdots - a_m A_m, \quad B_2 = a_2, \cdots, \quad B_m = a_m. \tag{12}$$

Then clearly condition (4) is satisfied, and by hypothesis we have

$$\phi(B_1, B_2, \cdots, B_m) = \phi(B_1, a_2, \cdots, a_m) = 0.$$

Now if $F(x_1, \cdots, x_m)$ and $F_{ij}(x_1, \cdots, x_m)$ have the same meaning as above, then clearly the determinant of $x_1 - B_1$ is $F(x_1, a_2, \cdots, a_m)$, and the first minors of the matrix $x_1 - B_1$ are precisely the $F_{ij}(x_1, a_2, \cdots, a_m)$, except possibly for sign.

Since $\phi(B_1, a_2, \cdots, a_m) = 0$, it follows by Theorem 2 that, in the ring $R[x_1]$, $\phi(x_1, a_2, \cdots, a_m)F_{ij}(x_1, a_2, \cdots, a_m)$ is divisible by $F(x_1, a_2, \cdots, a_m)$, $(i, j = 1, 2, \cdots, n)$. But $F(x_1, \cdots, x_m)$ is of the form of the $g(x_1, \cdots, x_m)$ in the lemma, and hence the lemma can be applied as the above holds for all choices of $a_2, \cdots, a_m$ in R. Thus

$$\phi(x_1, \cdots, x_m)F_{ij}(x_1, \cdots, x_m) \equiv 0\ (F(x_1, \cdots, x_m)),\ i, j = 1, 2, \cdots, n,$$

that is,

$$\phi(x_1, \cdots, x_m) \equiv 0\ (\mathfrak{m}).$$

This completes the proof of the theorem.

We have established Theorem 4, following Ostrowski, not in fact under the assumption that $\phi(B_1, \cdots, B_m) = 0$ for *every* $B_1, \cdots, B_m$ which are commutative and satisfy (4), but under the weaker assumption that the B's may be restricted to be in the special form (12). Because of the homogeneity of the polynomials considered, it is not difficult to show that we can further restrict our hypothesis by assuming always that $a_m = 1$. If, in the lemma, we consider only homogeneous polynomials, we can also in it assume that $a_m = 1$. Thus, for the case $m = 2$, the lemma and Theorem 4 are true for *any* commutative ring R with unit element. Theorem 4, so interpreted, then yields an actual generalization of Theorem 2.

Note added in proof: The assumption that $A_1 = I$ is used, so far as Theorem 3 is concerned, only to make sure that $F(x_1, \cdots, x_m)$ is not a divisor of zero in $R[x_1, \cdots, x_m]$. The remarks to follow will show that it is certainly sufficient, although by no means necessary, to assume that for some i, $|A_i|$ is not a divisor of zero in R.

If $ab = 0$, $a \neq 0$, a is called an *annihilator* of b. It is quite easy to prove, although this fact does not seem to be in the literature, that *if* $g(x_1, \cdots, x_m)$ *is an element of* $R[x_1, \cdots, x_m]$ *with an annihilator* $h(x_1, \cdots, x_m)$ *in* $R[x_1, \cdots, x_m]$, *then* $g(x_1, \cdots, x_m)$ *has an annihilator* a *in* R. Otherwise expressed, $g(x_1, \cdots, x_m)$ is a divisor of zero in $R[x_1, \cdots, x_m]$ if and only if all coefficients in g are annihilated by the *same* element a of R.

SMITH COLLEGE