TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 347, Number 10, October 1995

AUTOMORPHISM GROUP SCHEMES OF BASIC MATRIX INVARIANTS

WILLIAM C. WATERHOUSE

Abstract. For $3 \leq k < n$, let $E_k(X)$ be the polynomial in n^2 variables defined by $\det(X + \lambda I) = \sum E_k(X)\lambda^{n-k}$. Let R be a ring containing a field of characteristic $p \geq 0$. If p does not divide $n - k + 1$, the invertible linear transformations on matrices preserving $E_k(X)$ up to scalars are (in essence) just the obvious ones arising from scaling, similarities, and transposition. If the power p^s dividing $n - k + 1$ is greater than k, then we have these elements times maps of the form $X \mapsto X + f(X)I$. When smaller powers p^s divide $n - k + 1$, the group scheme is like the first with an infinitesimal part of the second. One corollary is that every division algebra of finite dimension $n^2 > 4$ over its center carries a canonical cubic form that determines it up to anti-isomorphism.

On $n \times n$ matrices, let E_k be the homogeneous polynomial of degree k expressing the sum of all principal $k \times k$ subdeterminants. Thus E_1 gives the trace, E_n gives the determinant, and we have $\det(X + \lambda I) = \sum E_k(X)\lambda^{n-k}$. Our goal in this paper is to determine the automorphisms and generalized automorphisms of each polynomial $E_k(X)$ over commutative rings R, that is, to find

$$\mathrm{Aut}(E_k)(R) = \{\phi \in \mathrm{GL}(M_n(R)) : E_k(\phi X) = E_k(X)\}$$

and

$$\mathrm{GAut}(E_k)(R) = \{\phi \in \mathrm{GL}(M_n(R)) : E_k(\phi X) = cE_k(X) \text{ for some constant } c\}.$$

We shall succeed for all R that contain some field. One interesting corollary of the main result is that the appropriate twisted form of E_k defined on a central simple algebra determines that algebra up to isomorphism or anti-isomorphism. In particular, every central simple algebra of dimension > 4 is so determined by the cubic form that is the coefficient of λ^{n-3} in the reduced norm characteristic polynomial $RN(X - \lambda)$. This is in general a much simpler invariant than the polynomial $RN(X)$ of degree n that was already known to determine the algebra up to anti-isomorphism [3, 10].

Some cases of the problem are special or already familiar. The function E_1 is linear, and E_2 is a quadratic form; these have exceptionally large automorphism groups [6, 4]. If $k = n$, then E_n is the determinant; its automorphisms were determined over fields by Frobenius, and his result was extended to all

Received by the editors April 21, 1994.
1991 *Mathematics Subject Classification*. Primary 14L15; Secondary 15A15.
This work was supported in part by the U. S. NSF/NSA, Grant Number DMS9001223.

commutative R independently by B. R. McDonald [8] and by me [12]. For these reasons, we shall assume throughout this paper that $3 \leq k < n$.

The groups $\mathrm{Aut}(E_k(R))$ have also been known for some time when R is a field of characteristic zero [7,1]. The result could have been (but was not) extended to fields of characteristic $p > n$ by using the work of Hoehnke [2], which implies that all automorphisms of E_k in that case are automorphisms of the determinant. But this implication is definitely false for certain small p depending on n and k. For some p there can be a larger smooth group of automorphisms. More subtly, for some p the automorphism group is not smooth, because it contains an "infinitesimal part" of the larger group; this is invisible over fields but appears for coefficient rings with nilpotent elements. To detect this, we must use affine group schemes. As in [12], they also have technical advantages that allow us to avoid almost all consideration of arbitrary R in the heart of the proof. The different behavior in different characteristics, however, makes it clear that the group scheme is not flat over $\mathbb{Z}$, and consequently the methods of this paper determine the automorphisms only over rings containing a field.

1. Notation and statement of the main theorem

We begin with notation for some subgroup schemes of $\mathrm{GL}(M_n)$. Let C be the scalars, the copy of the multiplicative group inside $\mathrm{GL}(M_n)$ taking X to cX for invertible constants c. Let H be the copy of PGL_n that is the Zariski closure of the group generated by the inner automorphisms $X \mapsto UXU^{-1}$. Let J be the constant group scheme of order 2 given by 1 and tr, where X^{tr} is the transpose of X. Let N be the subgroup consisting of mappings $X \mapsto X + f(X)I$, where f is a linear function and $f(I) + 1$ is invertible. Finally, for any group scheme G in characteristic $p > 0$, let G_r denote the infinitesimal subgroup of G that is the kernel of the rth power of the Frobenius. Thus, for instance, N_r is those elements in N where all coefficients in f have p^rth powers equal to 0.

Next, we list a number of rudimentary facts about these subgroups. Most are familiar, and the proofs can be omitted.

(1) C is central in $\mathrm{GL}(M_n)$.

(2) C, H, and N are smooth algebraic group schemes.

(3) J normalizes C, N, N_r, and H.

(4) H normalizes C, N, and N_r.

(5) HN is the semidirect product of N and H.

(6) CHN is the direct product of C and HN. It is the semidirect product of CN and H. The same results hold with N replaced by N_r.

(7) $CHNJ$ is the semidirect product of CHN and J. The same is true with N replaced by N_r.

(8) $HJ \subseteq \mathrm{Aut}(E_k)$, and $CHJ \subseteq \mathrm{GAut}(E_k)$.

Now we can state the main result. The statement is simplified by the group scheme notation, since the group $J(R)$ in general captures the influence of idempotent elements in R, and $H(R)$ captures the influence of the Picard group of R. See [12] for details. The theorem will be proved in §5 after the intervening sections supply the ingredients for the proof.

Main theorem. *Over the prime field of characteristic $p \geq 0$, consider the polynomial expression E_k on M_n. Assume $3 \leq k < n$. If $p > 0$, let p^s be the highest power of p dividing $n - k + 1$. Then we have the following equalities of group schemes over the prime field.*

(1) *If p does not divide $n-k+1$ (in particular, if $p = 0$), then* $\mathrm{GAut}(E_k) = CHJ$.

(2) *If $p > 0$ and $p^s > k$, then* $\mathrm{GAut}(E_k) = CHNJ$.

(3) *If $p > 0$ and $1 < p^s \leq k$, then* $\mathrm{GAut}(E_k) = CHN_sJ$.

2. Two normalizer computations

Theorem 1. *Let L be an algebraically closed field. Then the normalizer of the group $CH(L)$ inside $\mathrm{GL}(M_n(L))$ is $(CHN^*J)(L)$, where $N^*(L)$ is those mappings in $N(L)$ for which f is a multiple of the trace.*

Proof. We know that $CHJ(L)$ normalizes $CH(L)$. Any element normalizing $CH(L)$ induces an automorphism of the algebraic group $CH/C \cong H \cong \mathrm{PGL}_n$. Those automorphisms are given by inner automorphisms and transpose inverse. It is trivial to check that tr acts on H as transpose inverse. Thus any normalizing element T acts on $(CH/C)(L)$ in the same way as some element in $HJ(L)$, and thus we can modify it to act trivially. Let ϕ_U denote the mapping $X \mapsto UXU^{-1}$. Our condition on T then is that $T\phi_U = c_U\phi_U T$ for some constants c_U. But clearly $\phi_U \mapsto c_U$ is an algebraic homomorphism from $H(L)$ to the multiplicative group. As $H(L)$ is connected and semisimple, this homomorphism must be trivial. Thus we just need to find the T in $\mathrm{GL}(M_n(L))$ for which $T\phi_U = \phi_U T$ for all invertible U in $M_n(L)$.

Write $T(E_{ij}) = \sum t_{ij}^{rs} E_{rs}$, where the E_{ij} are the usual basis of the matrices. Take first $U = \mathrm{diag}(u_1, \dots, u_n)$. This gives

$$u_i u_j^{-1} \sum t_{ij}^{rs} E_{rs} = \sum t_{ij}^{rs} u_r u_s^{-1} E_{rs}.$$

For $i \neq j$, this equation (for all possible U) implies that $t_{ij}^{rs} = 0$ except for $r = i$, $s = j$. Thus each such E_{ij} is an eigenvector of T. As $T(ME_{ij}M^{-1}) = MT(E_{ij})M^{-1}$ for any invertible M, all matrices similar to E_{ij} are eigenvectors with the same eigenvalue. In particular, the various nondiagonal E_{ij} are similar, so $T(E_{ij}) = cE_{ij}$ for some nonzero constant c and all $i \neq j$. Furthermore, for $i \neq j$, we have $E_{ij} + E_{ji}$ similar to $E_{ii} - E_{jj}$ if $\mathrm{char}(L) \neq 2$ and similar to $E_{ii} - E_{jj} + E_{ij}$ if $\mathrm{char}(L) = 2$. Thus in either case we get $T(E_{ii} - E_{jj}) = c(E_{ii} - E_{jj})$. Hence we can conclude that $T(X) = cX$ for all X with trace zero.

Finally, the equation at the start of the last paragraph shows that $T(E_{ii})$ is diagonal. Thus $T(X) - cX$ is diagonal for all X. Again, applying ϕ_M shows that all matrices similar to $T(X) - cX$ must also be diagonal in this same basis. Hence they are all scalar multiples of the identity, and $T(X) - cX = f(X)I$ for some function $f(X)$. Clearly this function is linear. Since it is zero when the trace is zero, it is a multiple of the trace. □

Theorem 2. *Let L be an algebraically closed field. Then the normalizer of the group $CHN(L)$ inside $\mathrm{GL}(M_n(L))$ is $CHNJ(L)$.*

Proof. Suppose T normalizes $CHN(L)$. As CN is solvable and H is semisimple, CN is the radical of the algebraic group CNH. The conjugation by

T is an algebraic automorphism of $CNH(L)$, and hence it maps the radical $CN(L)$ to itself. Hence it induces an automorphism of $CNH/CN \cong H(L)$. As in the previous proof, every such automorphism is induced by an element of $HJ(L)$. Thus we may assume that conjugation by T is trivial on CNH/CN. We need to prove then that T is in $CN(L)$.

Let ϕ_U denote the mapping $X \mapsto UXU^{-1}$. Our condition on T now is that $T\phi_U = \phi_U m_U T$ where $m_U(X) = c_U X + f_U(X)I$ is an element in $CN(L)$. We have

$$\phi_{UV} m_{UV} T = T\phi_{UV} = T\phi_U\phi_V = \phi_U m_U T\phi_V = \phi_U m_U \phi_V m_V T,$$

and hence $m_{UV} = \phi_V^{-1} m_U \phi_V m_V$. It is trivial to compute that

$$\phi_V^{-1} m_U \phi_V m_V(X) = c_U(c_V X + f_V(X)I) + f_U(c_V VXV^{-1} + f_V(X)I)I.$$

Hence in particular $c_{UV} = c_U c_V$ for all invertible U and V, and as before it follows that $c_U = 1$ for all U.

Now set $T(E_{ij}) = \sum t_{ij}^{rs} E_{rs}$ and take $U = \operatorname{diag}(u_1, \dots, u_n)$, as in the previous proof. We have then the equation

$$\begin{aligned} u_i u_j^{-1} \sum t_{ij}^{rs} E_{rs} &= T\phi_U E_{ij} = \phi_U m_U T(E_{ij}) \\ &= \sum u_r u_s^{-1} t_{ij}^{rs} E_{rs} + f_U(T(E_{ij}))I. \end{aligned}$$

For $i \neq j$, this equation implies that $t_{ij}^{rs} = 0$ for $r \neq s$ unless $r = i$, $s = j$. For $r = s$, we see that

$$u_i u_j^{-1} t_{ij}^{rr} = t_{ij}^{rr} + f_U(T(E_{ij})).$$

Choosing $u_i u_j^{-1} \neq 1$, we see that t_{ij}^{rr} is independent of r. Thus $T(E_{ij})$ is a multiple of E_{ij} plus a multiple of I. Taking conjugates and computing modulo scalars, we find as in the previous proof that $T(X) = cX + f(X)I$ for X of trace zero, where f is some linear function. Again the equation tells us that each $T(E_{ii})$ is diagonal, and thus $T(X) - cX$ is diagonal for all X. Again this implies that $T(X) = cX + f(X)I$ for some linear f, and the proof is complete. □

3. The automorphisms of the form $X \mapsto X + f(X)I$

In this section, we almost completely determine the contribution of N to $\operatorname{GAut}(E_k)$. First, we need three simple (and fairly familiar) lemmas.

Lemma 1. *Over any field, E_k is an irreducible polynomial.*

Proof. We prove this by an induction starting back at $k = 1$, where the result is obvious. Inductively, consider E_k as a polynomial in X_{nn}; it is linear, and the coefficient of X_{nn} is the E_{k-1} of the complementary minor. By induction, that coefficient is irreducible in its variables. Thus if E_k factors, this E_{k-1} must divide the sum of all terms not involving X_{nn}. But that is false, as it is easy to write down a matrix where E_k is nonzero but the upper E_{k-1} vanishes. □

Lemma 2. *Let m be a positive integer, p a prime, and p^s the highest power of p dividing m. Then $\binom{m+r-1}{r}$ is divisible by p for $1 \leq r \leq p^s - 1$. It is not divisible by p when $r = p^s$. If $p^s > 1$, it is again divisible by p for $r = p^s + 1$.*

Proof. For $r \leq p^s$, write the binomial coefficient as

$$\frac{m+1}{1} \cdot \frac{m+2}{2} \cdots \frac{m+r-1}{r-1} \cdot \frac{m}{r}$$

and compare powers of p in each numerator and denominator. For $r = p^s + 1$, the value is $(m+p^s)/(p^s+1)$ times the value for $r = p^s$. □

Lemma 3. *Let $f = f(X)$ be a linear polynomial in the X_{ij}. Then*

$$E_k(X + fI) = \sum_{r=0}^{k} \binom{m+r-1}{r} f^r(X) E_{k-r}(X)$$

where $m = n - k + 1$.

Proof. We have

$$\begin{aligned}\det((X + fI) + \lambda I) &= \det(X + (f + \lambda)I) \\ &= \sum_{s=0}^{n} (f+\lambda)^s E_{n-s}(X) = \sum_{s=0}^{n} \sum_{r=0}^{s} \binom{s}{r} f^r \lambda^{s-r} E_{n-s}(X).\end{aligned}$$

We get $E_k(X + fI)$ by taking the coefficient of λ^{n-k}. □

Theorem 3. *Suppose the characteristic is p and the highest power of p dividing $m = n - k + 1$ is p^s. If p^s is greater than k, then $N \subseteq \mathrm{Aut}(E_k)$. In any case $N_s \subseteq \mathrm{Aut}(E_k)$.*

Proof. Take the expansion of $E_k(X + f(X)I)$ in Lemma 3. Lemma 2 implies that all the terms with $1 \leq r < p^s$ have zero coefficients. Under the first hypothesis, we thus have $E_k(X + fI) = E_k(X)$. In any case, N_s involves those f with $f^{p^s} = 0$, and that forces the remaining terms with $r \geq p^s$ to vanish. □

Theorem 4. *If F is a field of characteristic 0 or of characteristic p where the p^s in Theorem 3 is $\leq k$, then $N(F)$ and $\mathrm{GAut}(E_k)(F)$ have trivial intersection.*

Proof. Suppose that for some nonzero linear $f(X)$ and some $\lambda \in F$ we have

$$\lambda E_k = E_k(X + fI) = \sum_{r=0}^{k} \binom{m+r-1}{r} f^r(X) E_{k-r}(X).$$

If $\lambda \neq 1$, we have f dividing a nonzero multiple of E_k; this is impossible, since E_k is irreducible by Lemma 1. Assume now $\lambda = 1$, so the term $r = 0$ in the sum cancels. The hypothesis on k implies that not all of the remaining terms vanish; the first one that does not is at $r = p^s$ (where we set $p^s = 1$ in characteristic 0). Over the field, we can divide by the nonzero polynomial f^{p^s}, and we get a nonzero constant multiple of E_{k-p^s} either equal to zero (if $p^s = k$) or equal to a sum of terms all involving a factor of f. The first case is clearly impossible. The second is impossible unless $k - p^s = 1$, because E_k is irreducible and f is linear. Suppose $k = p^s + 1$, so there are terms only for $r = p^s$ and $r = p^s + 1$. As we are always assuming $k \geq 3$, we have $p^s > 1$. Then by Lemma 2 again the term for $r = p^s + 1$ has zero coefficient, and we have a contradiction. □

Theorem 5. *Let F be a field of characteristic p, and assume $p^s \leq k$. Let $R = F[\tau]/\tau^{1+p^s}$. Then there is no element of* $\mathrm{GAut}(E_k)(R)$ *reducing to a nontrivial map of the form $X \mapsto X + \tau f(X)I$ modulo τ^2.*

Proof. An element in $\mathrm{GAut}(E_k)(R)$ with such a reduction would have the form

$$X \mapsto X + \tau f(X)I + \tau^2 M(X),$$

where $f = \sum b_{uv}X_{uv}$ has coefficients in F and the matrix M has entries that are linear functions of the X_{ij} with coefficients in R. Our hypothesis is that

$$\begin{aligned}\lambda E_k(X) &= E_k(X + \tau f I + \tau^2 M) \\ &= \sum_{r=0}^{k} \binom{m+r-1}{r} \tau^r f^r(X) E_{k-r}(X + \tau^2 M(X))\end{aligned}$$

for some invertible λ in R (necessarily congruent to 1 modulo τ).

Suppose first that $p^s > 2$. The terms for $1 \leq r < p^s$ drop out because of their coefficients, as in Theorem 3, and those for $r > p^s + 1$ drop out because of the power of τ involved. Obviously also $E_{k-p^s}(X + \tau^2 M) \equiv E_{k-p^s}(X) \bmod \tau$. Thus the equation is

$$\lambda E_k(X) = E_k(X + \tau^2 M) + \binom{m+p^s-1}{p^s} \tau^{p^s} f^{p^s} E_{k-p^s}(X).$$

Consider any one variable X_{uv}. As $k - p^s \leq n - 2$, there are nonzero terms in $E_{k-p^s}(X)$ not involving X_{uv}. Thus the second summand on the right contains a nonzero term that is $\tau^{p^s} b_{uv}^{p^s} X_{uv}^{p^s}$ times a factor not involving X_{uv}. There are of course no such terms in $\lambda E_k(X)$. Consider now any monomial term in E_k, and expand the result with each X_{ij} replaced by $X_{ij} + \tau^2 M_{ij}$. The monomial has degree k. In the expansion, we get zero if we take more than $p^s/2$ pieces involving τ^2. The M_{ij} are linear, and X_{uv} occurs at most to the first power in the original monomial, so the nonzero terms in $E_k(X + \tau^2 M)$ involve X_{uv} at most to the exponent $1 + (p^s/2)$. We are supposing that $p^s > 2$, so no term is available to cancel the one we found earlier. Hence we must have $b_{uv} = 0$. Since X_{uv} was arbitrary, we have shown that $f = 0$.

Now suppose that $p^s = 2$. The equation is

$$\begin{aligned}\lambda E_k(X) &= E_k(X + \tau^2 M) + \tau^2 f^2 E_{k-2}(X) \\ &= E_k(X) + \tau^2 \sum M_{ij}(X) \frac{\partial E_k(X)}{\partial X_{ij}} + \tau^2 f^2 E_{k-2}(X).\end{aligned}$$

As $k \geq 3$, there are terms in E_{k-2} involving X_{uu}. Thus the final summand contains monomials involving the factor $\tau^2 b_{uv}^2 X_{uv}^2 X_{uu}$. But each $M_{ij}(X)$ is linear, and no term in $\partial E_k(X)/\partial X_{ij}$ can involve two variables from the same row r. Thus no other term can cancel the beginning one, and hence again $b_{uv} = 0$.

The result for $p^s = 1$ is implied by the Lie algebra computation in the next section, but it is also easy to establish directly. The equation is

$$\lambda E_k(X) = E_k(X) + m\tau f E_{k-1}(X).$$

Looking at the coefficient of τ, we would have to have a nonzero polynomial over L divisible by f and equal to a constant multiple of E_k; we know that is impossible by Lemma 1. □

4. The Lie algebra

Recall that the Lie algebra of $\mathrm{GL}(M_n)$ over a field F can be defined as those mappings over $F[\varepsilon]/(\varepsilon^2)$ that reduce to the identity modulo ε. These have the form $X \mapsto X + \varepsilon T(X)$, and we can identify them with all linear maps T from M_n to itself. Thus $\mathrm{Lie}(C)$ consists of all $T(X) = \lambda X$, while $\mathrm{Lie}(H)$ consists of all $T(X) = UX - XU$, and $\mathrm{Lie}(N)$ consists of all $T(X) = g(X)I$. Our goal in this section is to determine $\mathrm{Lie}(\mathrm{GAut}(E_k))$ for $3 \leq k < n$.

Theorem 6. *The Lie algebra of* $\mathrm{GAut}(E_k)$ *is* $\mathrm{Lie}(C) + \mathrm{Lie}(H)$ *unless the characteristic* p *divides* $n - k + 1$. *In that case, it is* $\mathrm{Lie}(C) + \mathrm{Lie}(H) + \mathrm{Lie}(N)$.

Proof. We know these Lie subalgebras are contained in $\mathrm{Lie}(\mathrm{GAut}(E_k))$; for we always have CH contained in GAut, and under the extra hypothesis we have N_1 in GAut by Theorem 3. Facts (5) and (6) in the introduction imply that the subalgebras are linearly independent. Our procedure is to start with an arbitrary T in Lie(GAut), modifying it as necessary by elements in $\mathrm{Lie}(C)$ and/or $\mathrm{Lie}(N)$, and find enough conditions on the resulting mapping to leave only $n^2 - 1$ parameters free. Since that is the dimension of $\mathrm{Lie}(H)$, the proof will then be finished. I shall number the steps. The basic condition that T be in Lie(GAut) is

$$\sum_{ij} T(X)_{ij} \frac{\partial E_k(X)}{\partial X_{ij}} = cE_k(X).$$

For legibility, I shall write T in terms of coordinates as

$$T(X)_{ij} = \sum \begin{pmatrix} r & s \\ i & j \end{pmatrix} X_{rs}.$$

(1) Consider the monomial $X_{34}X_{21}X_{33}(X_{44}\cdots X_{kk})$. It does not occur in E_k. The combinations $X_{34}X_{33}$ and $X_{34}X_{44}$ cannot occur in the partial derivatives of E_k (nor, if $k = 3$, can $X_{34}X_{21}$). Thus such monomials can occur only when the partial derivative yields $X_{21}X_{33}\cdots X_{kk}$. This happens only in $\partial E_k/\partial X_{12}$. Thus $\begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix} = 0$. Clearly we can apply the same argument for different indices to conclude that $\begin{pmatrix} s & t \\ q & r \end{pmatrix} = 0$ for any four distinct indices q, r, s, t. All our remaining arguments will similarly consider the coefficients of particular monomials in the basic condition. I shall always write out a specific case, following it by the statement of the result for arbitrary indices; and q, r, s, t will always denote distinct indices.

(2) Consider $X_{23}^2X_{34}\cdots X_{k1}$. This is not in E_k and can arise only from the partial derivative $\partial E_k/\partial X_{12}$. Thus $\begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix} = 0$, and in general $\begin{pmatrix} s & t \\ r & s \end{pmatrix} = 0$. Similarly, the terms

$$X_{23}X_{34}\cdots X_{k1}^2, \quad X_{21}^2X_{33}\cdots X_{kk}, \quad X_{21}X_{33}^2X_{44}\cdots X_{kk}$$

can arise only from $\partial E_k/\partial X_{12}$. Thus

$$0 = \begin{pmatrix} k & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 1 & 2 \end{pmatrix},$$

and in general

$$0 = \begin{pmatrix} t & r \\ r & s \end{pmatrix} = \begin{pmatrix} r & s \\ s & r \end{pmatrix} = \begin{pmatrix} r & r \\ s & t \end{pmatrix}.$$

(3) Consider the monomial $X_{32}X_{21}X_{33}\cdots X_{kk}$. It can arise only from $\partial E_k/\partial X_{12}$ and from $\partial E_k/\partial X_{13}$, with opposite signs. Thus $\begin{pmatrix}3&3\\1&3\end{pmatrix}-\begin{pmatrix}3&2\\1&2\end{pmatrix}=0$. Similarly, $X_{13}X_{21}X_{33}\cdots X_{kk}$ can arise only from $\partial E_k/\partial X_{12}$ and $\partial E_k/\partial X_{32}$, and so we must have $\begin{pmatrix}3&3\\3&2\end{pmatrix}-\begin{pmatrix}1&3\\1&2\end{pmatrix}=0$. In general, we have

$$\begin{pmatrix}r&s\\t&s\end{pmatrix}=\begin{pmatrix}r&r\\t&r\end{pmatrix}\quad\text{and}\quad\begin{pmatrix}r&s\\r&t\end{pmatrix}=\begin{pmatrix}s&s\\s&t\end{pmatrix}.$$

(4) Consider $X_{1n}X_{23}X_{34}\cdots X_{k1}$. Here n and 2 are the only indices occurring just once, so the other factors must occur in the partial derivative; we get $\begin{pmatrix}1&n\\1&2\end{pmatrix}+\begin{pmatrix}2&3\\n&3\end{pmatrix}=0$. By step (3), we can rewrite this as $\begin{pmatrix}n&n\\n&2\end{pmatrix}+\begin{pmatrix}2&2\\n&2\end{pmatrix}=0$, or in general

$$\begin{pmatrix}r&r\\s&r\end{pmatrix}=-\begin{pmatrix}s&s\\s&r\end{pmatrix}.$$

(5) Consider $X_{11}X_{12}X_{23}\cdots X_{k-1,1}$. Here X_{11} cannot occur in a partial derivative together with either X_{12} or $X_{k-1,1}$, and thus it is the factor that must come from T. Thus

$$\sum_{r>k-1}\begin{pmatrix}1&1\\r&r\end{pmatrix}=0.$$

A modification of the indices in the monomial gives us also

$$\begin{pmatrix}1&1\\2&2\end{pmatrix}+\sum_{r>k}\begin{pmatrix}1&1\\r&r\end{pmatrix}=0.$$

Thus $\begin{pmatrix}1&1\\2&2\end{pmatrix}=\begin{pmatrix}1&1\\k&k\end{pmatrix}$. In general, we have

$$\begin{pmatrix}r&r\\s&s\end{pmatrix}=\begin{pmatrix}r&r\\t&t\end{pmatrix},$$

and then

$$(n-k+1)\begin{pmatrix}r&r\\s&s\end{pmatrix}=0.$$

(6) Similarly, consider $X_{12}X_{22}X_{33}\cdots X_{kk}$. The X_{12} or X_{22} must come from T, and we get

$$\sum_{r\neq 2,\dots,k}\begin{pmatrix}1&2\\r&r\end{pmatrix}-\begin{pmatrix}2&2\\2&1\end{pmatrix}=0.$$

A modification of indices gives

$$\sum_{r\neq 2,4,5,\dots,k+1}\begin{pmatrix}1&2\\r&r\end{pmatrix}-\begin{pmatrix}2&2\\2&1\end{pmatrix}=0.$$

Thus $\begin{pmatrix}1&2\\3&3\end{pmatrix}=\begin{pmatrix}1&2\\k+1&k+1\end{pmatrix}$, and in general

$$\begin{pmatrix}r&s\\t&t\end{pmatrix}=\begin{pmatrix}r&s\\q&q\end{pmatrix}.$$

(7) The term $X_{12}^2X_{21}X_{44}\cdots X_{kk}$ arises with coefficients $\begin{pmatrix}1&2\\r&r\end{pmatrix}$ for $r\neq 1,2,4,\ldots,k$. Hence the sum of these is zero. Thus by (6) we get

$$(n-k+1)\begin{pmatrix}r&s\\t&t\end{pmatrix}=0.$$

If p does not divide $n-k+1$, then this result together with (5) shows that $\begin{pmatrix}r&r\\s&s\end{pmatrix}$ and $\begin{pmatrix}r&s\\t&t\end{pmatrix}$ are all zero. If p does divide $n-k+1$, we know $\mathrm{Lie}(N)\subseteq \mathrm{Lie}(\mathrm{GAut})$, and thus the values $\begin{pmatrix}c&d\\t&t\end{pmatrix}$ can be arbitrarily prescribed. Subtracting such a map from T, we can henceforth assume in any case that

$$0=\begin{pmatrix}r&r\\s&s\end{pmatrix}=\begin{pmatrix}r&s\\t&t\end{pmatrix}.$$

(8) Returning to the original sum in (6), we see that almost all the terms vanish, and we get just $\begin{pmatrix}2&2\\2&1\end{pmatrix}=\begin{pmatrix}1&2\\1&1\end{pmatrix}$. In general,

$$\begin{pmatrix}r&s\\r&r\end{pmatrix}=\begin{pmatrix}s&s\\s&r\end{pmatrix}.$$

(9) Similarly, consider $X_{12}X_{13}X_{34}\cdots X_{k-1,1}$. It occurs with coefficients $\begin{pmatrix}1&3\\2&3\end{pmatrix}$ and $-\begin{pmatrix}1&2\\r&r\end{pmatrix}$ for $r\neq 1,3,\ldots,k-1$. Dropping the terms that vanish, we find that $\begin{pmatrix}1&2\\2&2\end{pmatrix}$ equals $\begin{pmatrix}1&3\\2&3\end{pmatrix}$, which itself equals $\begin{pmatrix}1&1\\2&1\end{pmatrix}$ by (3). In general,

$$\begin{pmatrix}r&s\\s&s\end{pmatrix}=\begin{pmatrix}r&r\\s&r\end{pmatrix}.$$

(10) Look at the term $X_{11}X_{22}\cdots X_{kk}$. This finally is a term that occurs in cE_k, with coefficient c. Its coefficient on the other side of the equation is

$$\sum_{r\neq 2,\ldots,k}\begin{pmatrix}1&1\\r&r\end{pmatrix}+\sum_{r\neq 1,3,\ldots,k}\begin{pmatrix}2&2\\r&r\end{pmatrix}+\cdots.$$

Almost all these coefficients are zero, by (7), and we get

$$\begin{pmatrix}1&1\\1&1\end{pmatrix}+\cdots+\begin{pmatrix}k&k\\k&k\end{pmatrix}=c.$$

We can do the same computation with one index changed, and so we find that $\begin{pmatrix}r&r\\r&r\end{pmatrix}=\begin{pmatrix}s&s\\s&s\end{pmatrix}$. We recall that all scalar multiplications are in $\mathrm{Lie}(C)$; subtracting such a map from T, we can assume that all

$$\begin{pmatrix}r&r\\r&r\end{pmatrix}=0.$$

This of course implies that $c=0$. Observe that this subtraction does not change any of the other coefficients already known to be zero.

(11) Look at the term $X_{12}X_{21}X_{33}\cdots X_{kk}$. As $c=0$, we get

$$0=\begin{pmatrix}1&2\\1&2\end{pmatrix}+\begin{pmatrix}2&1\\2&1\end{pmatrix}+\sum_{r\neq 1,2,4,\ldots,k}\begin{pmatrix}3&3\\r&r\end{pmatrix}+\cdots=\begin{pmatrix}1&2\\1&2\end{pmatrix}+\begin{pmatrix}2&1\\2&1\end{pmatrix}.$$

Thus

$$\begin{pmatrix} r & s \\ r & s \end{pmatrix} = -\begin{pmatrix} s & r \\ s & r \end{pmatrix}.$$

But consider also the term $X_{12}X_{23}X_{31}X_{44}\cdots X_{kk}$. With $c = 0$, this gives

$$\begin{aligned} 0 &= \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} + \begin{pmatrix} 2 & 3 \\ 2 & 3 \end{pmatrix} + \begin{pmatrix} 3 & 1 \\ 3 & 1 \end{pmatrix} + \sum \begin{pmatrix} 4 & 4 \\ r & r \end{pmatrix} + \cdots \\ &= \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} + \begin{pmatrix} 2 & 3 \\ 2 & 3 \end{pmatrix} + \begin{pmatrix} 3 & 1 \\ 3 & 1 \end{pmatrix}. \end{aligned}$$

Using the previous equality, we get $\begin{pmatrix} 2 & 3 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 1 & 3 \end{pmatrix} - \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$, and in general

$$\begin{pmatrix} r & s \\ r & s \end{pmatrix} = \begin{pmatrix} 1 & s \\ 1 & s \end{pmatrix} - \begin{pmatrix} 1 & r \\ 1 & r \end{pmatrix}$$

for $r, s > 1$. Thus all terms of the form $\begin{pmatrix} r & s \\ r & s \end{pmatrix}$ are expressed in terms of the $n - 1$ terms $\begin{pmatrix} 1 & r \\ 1 & r \end{pmatrix}$ for $r > 1$.

(12) To summarize, we have been able to drop all $\begin{pmatrix} r & s \\ q & t \end{pmatrix}$. All coefficients with three different indices either vanish or are expressed in terms of coefficients with two indices. The $\begin{pmatrix} r & s \\ r & s \end{pmatrix}$ were treated in (11); they involve $n - 1$ parameters. The $\begin{pmatrix} r & r \\ r & r \end{pmatrix}$ have all been reduced to zero. Equations (4), (8), and (9) show that

$$\begin{pmatrix} r & s \\ s & s \end{pmatrix} = \begin{pmatrix} r & r \\ r & s \end{pmatrix} = -\begin{pmatrix} s & s \\ s & r \end{pmatrix} = -\begin{pmatrix} r & s \\ r & r \end{pmatrix},$$

so the number of independent parameters involved in these types is just $n(n-1)$. Thus altogether we have only $n^2 - 1$ parameters remaining for the coefficients. But that is exactly the dimension of $\mathrm{Lie}(H)$, and thus our known maps fill the whole dimension available for $\mathrm{Lie}(\mathrm{GAut}(E_k))$. □

5. Proof of the main theorem

Let $G^0 = \mathrm{GAut}(E_k)^0$ be the connected component of the identity. Suppose first that p does not divide $n - k + 1$. Then we know from Theorem 6 that $\mathrm{Lie}(G^0) = \mathrm{Lie}(\mathrm{GAut}(E_k)) = \mathrm{Lie}(C) + \mathrm{Lie}(H) = \mathrm{Lie}(CH)$. As CH is smooth, $\dim \mathrm{Lie}(CH) = \dim(CH)$. Thus $\dim \mathrm{Lie}(G^0) = \dim(CH) \leq \dim G^0$. Hence G^0 is smooth. As it is connected, it equals the smooth subgroup CH of the same dimension. Now the connected component is normal, and the quotient $\mathrm{GAut}(E_k)/G^0$ is étale. Consider its values in an algebraically closed field L. We know by Theorem 1 that the full normalizer of $G^0(L) = CH(L)$ is $(CHN^*J)(L)$. We know $(CHJ)(L) \subseteq \mathrm{GAut}(E_k)(L)$, and Theorem 4 tells us that $N^*(L)$ meets $\mathrm{GAut}(E_k)(L)$ only in the identity. Thus $\mathrm{GAut}(E_k)/G^0(L) \cong J(L)$, and $\mathrm{GAut}(E_k) = CHJ$. The argument is similar and slightly easier when $p^s > k$, using Theorems 2, 3, and 6.

Now suppose $1 < p^s \leq k$. The main step is to prove that $CHN_s = G^0$; once that is done, we have $G^0(L) = CH(L)$, and the rest of the argument goes as before using Theorems 1 and 4. As CHN_s is certainly a connected subgroup of G^0 by Theorem 3, it suffices [5, pp. 124, 152] to prove that $(CHN_s)_r = G^0_r$

for all $r \geq 1$. We have one included in the other, so all we have to prove is that the ranks of these two finite group schemes are always the same. We use the following lemma, which is surely not new but for which I have no specific reference:

Lemma 4. *Let E be a finite connected group scheme over a perfect field F of characteristic p. Write*

$$E = \operatorname{Spec} F[(Y_i)]/(Y_i^{q(i)}),$$

and let

$$R = F[\tau]/(\tau^{1+p^s}).$$

The mapping sending τ to ε is a homomorphism $R \to F[\varepsilon]/(\varepsilon^2)$ which induces a mapping $E(R) \to \operatorname{Lie}(E)$. Then the image of this mapping is a subspace of dimension equal to the number of $q(i)$ greater than p^s.

Proof. As E is finite and connected, we know [9, p. 112] that it can be written in this form for some p-powers $q(i)$. Clearly also $E(F[\varepsilon]/(\varepsilon^2)) \cong \operatorname{Lie}(E)$. Now an element in $E(R)$ is given by a family of $r(i)$ in R with $r(i)^{q(i)} = 0$. We can choose $r(i)$ to have a nonzero coefficient of τ iff $q(i) > p^s$. □

In particular, of course, the number of Y_i is equal to the dimension of $\operatorname{Lie}(E)$.

Since CH is smooth of dimension n^2, we know that $(CH)_r$ has rank p^{rn^2}, and similarly N_r has rank p^{rn^2}. By Theorem 6, we have

$$\operatorname{Lie}(G_r^0) = \operatorname{Lie}(G^0) = \operatorname{Lie}(CH) \oplus \operatorname{Lie}(N).$$

This shows us first that $\dim \operatorname{Lie}(G^0) = 2n^2$. Hence G_r^0 certainly has rank at most p^{2n^2r}. For $r \leq s$, the rank of $(CHN_s)_r = (CHN)_r$ is equal to p^{2n^2r}, so we have equality. For $r > s$, take G_r^0 as the E in Lemma 4, and consider the mapping

$$G_r^0(R) \to \operatorname{Lie}(G_r^0) = \operatorname{Lie}(CH) \oplus \operatorname{Lie}(N).$$

The image includes $\operatorname{Lie}(CH)$, as CH is a subgroup. But by Theorem 5, the image includes no nonzero element of $\operatorname{Lie}(N)$. Hence the image is equal to $\operatorname{Lie}(CH)$. Of the $2n^2$ variables in the expression of the algebra for G_r^0, the lemma shows that there are n^2 with exponents bigger than p^s. Thus G_r^0 has rank at most $p^{rn^2}p^{sn^2}$. But that is the rank of $(CHN_s)_r$, and we must again have equality. □

6. Corollaries

Corollary 1. *Over the prime field of characteristic $p \geq 0$, consider the polynomial function E_k on M_n. Assume $3 \leq k < n$. If $p > 0$, let p^s be the highest power of p dividing $n - k + 1$. Let μ_k denote the group scheme of kth roots of unity. Then*

(1) *If p does not divide $n - k + 1$ (in particular, if $p = 0$), then* $\operatorname{Aut}(E_k) = \mu_k HJ$.

(2) *If $p > 0$ and $p^s > k$, then* $\operatorname{Aut}(E_k) = \mu_k HNJ$.

(3) *If* $p > 0$ *and* $1 < p^s \leq k$, *then* $\operatorname{Aut}(E_k) = \mu_k H N_s J$.

Proof. We have seen that H and J, and also N and N_s when they occur in $\operatorname{GAut}(E_k)$, actually lie in $\operatorname{Aut}(E_k)$. Thus the scaling of E_k comes entirely from the scalar multiplications in C. They will preserve E_k precisely when the factor is a kth root of unity. □

Corollary 2. *Let* F *be a field of characteristic* $p \geq 0$, *with* $3 \leq k < n$. *If* $p > 0$, *let* p^s *be the highest power of* p *dividing* $n - k + 1$.

(1) *Suppose* $p = 0$ *or* $1 \leq p^s \leq k$. *Then the invertible linear mappings on* $M_n(F)$ *preserving the polynomial expression* E_k *up to scalar are those sending a matrix* X *to either* $\lambda U X U^{-1}$ *or to* $\lambda U X^{\mathrm{tr}} U^{-1}$, *where the scalar* λ *and the matrix* U *are invertible.*

(2) *Suppose* $p^s > k$. *Then the invertible linear mappings on* $M_n(F)$ *preserving* E_k *up to scalar are those sending* X *to* $\lambda U(X + f(X)I)U^{-1}$ *or to* $\lambda U(X^{\mathrm{tr}} + f(X)I)U^{-1}$, *where the scalar* λ *and the matrix* U *are invertible and* f *is a linear function with* $f(I) \neq -1$.

Proof. Over a field, several things become simpler. First, $N_s(F)$ is trivial, as N_s is finite and connected. Similarly, the group $J(F)$ contains only the identity and tr. And finally, we know over fields that $\mathrm{GL}_n(F)$ maps onto $\mathrm{PGL}_n(F) \cong H(F)$. □

Corollary 3. *In the notation of Corollary* 2, *all linear mappings on* $M_n(F)$ *that preserve* E_k *are invertible in case* (1). *In case* (2), *they are the mappings of the forms* $\lambda U(X + f(X)I)U^{-1}$ *and* $\lambda U(X^{\mathrm{tr}} + f(X)I)U^{-1}$ *for* $\lambda^k = 1$ *and invertible* U *and arbitrary linear* f.

Proof. We use some simple results from [11]. If some noninvertible linear mapping preserves E_k, then there is a nontrivial subspace such that $E_k(X + Y) = E_k(X)$ for all X in $M_n(F)$ and all Y in the subspace. The largest such subspace is invariant under automorphisms of E_k. As the similarities $H(F)$ preserve E_k, and E_k is not a function of the trace, it follows that the subspace can only be the span of the identity. Theorems 3 and 4 show that the span of the identity has this property in case (2) but not in case (1). The noninvertible mappings preserving E_k arise by adding to the invertible ones an arbitrary linear mapping with values in the subspace, and thus the previous corollaries tell us what they can be. □

7. Application to mappings on central simple algebras

Let F be a field, and let A be a central simple algebra over F. Recall that there is then a finite Galois extension L of F such that $A \otimes_F L \cong M_n(L)$ for some n. The determinant function on $M_n(L)$ restricts to a polynomial function on A with coefficients in F; it is called the reduced norm, RN^A, and is independent of the choice of L. We can expand

$$RN^A(X + \lambda \cdot 1) = \sum E_k^A(X)\lambda^{n-k}$$

with the $E_k^A(X)$ homogeneous polynomials of degree k over F.

Theorem 7. *Let A and B be two central simple algebras of the same dimension n^2 over a field F of characteristic $p \geq 0$. Let $\psi : A \to B$ be a linear mapping such that*

$$E_k^B(\psi(X)) = cE_k^A(X)$$

for some nonzero constant c and some k with $3 \leq k < n$. If $p > 0$, let p^s be the highest power of p dividing $n - k + 1$.

(1) *Suppose $p = 0$ or $1 \leq p^s \leq k$. Then $\psi(X) = \lambda\phi(X)$ where λ is an invertible scalar in F and $\phi : A \to B$ is either an isomorphism or an anti-isomorphism.*

(2) *Suppose $p^s > k$. Then $\psi(X) = \lambda\phi(X + f(X)1)$ where λ is an invertible scalar in F, ϕ is an isomorphism or anti-isomorphism, and f is an arbitrary linear function on A.*

Proof. If $A \cong M_n(F) \cong B$, the assertion follows from Corollary 2. Furthermore, we observe that the λ, ϕ and (when relevant) f are uniquely determined by ψ. Now in general, we choose a Galois extension L of F with $A \otimes L \cong M_n(L) \cong B \otimes L$. The linear extension $\psi_L : A \otimes L \to B \otimes L$ then preserves E_k on $M_n(L)$. Hence it can be written uniquely as $\lambda_L\phi_L(X)$ or as $\lambda_L\phi_L(X + f_L(X)1)$, depending on the case.

Now consider any g in $\mathrm{Gal}(L/F)$. This acts on $A \otimes L$ through its action on L, and an element in $A \otimes L$ is actually in A iff it is fixed by all g. Similarly, a linear mapping $\psi_L : A \otimes L \to B \otimes L$ is the extension of some $\psi : A \to B$ iff $g\psi_L g^{-1} = \psi_L$ for all g. This is true for our ψ, so in case (1) we have $g(\lambda_L\phi_L(g^{-1}X)) = \lambda_L\phi_L(X)$. The left-hand side equals $g(\lambda_L)(g\phi_L g^{-1})(X)$, so by the uniqueness we get $g(\lambda_L) = \lambda_L$ and $g\phi_L g^{-1} = \phi_L$. Hence we conclude that λ_L is in F and ϕ_L comes from a mapping $\phi : A \to B$. Since ϕ_L is an isomorphism or anti-isomorphism, the same must be true of ϕ. The same kind of argument establishes case (2). □

Corollary 4. *Let A and B be two central simple algebras of the same dimension n^2 over a field. Suppose there is a linear mapping $\psi : A \to B$ such that $E_k^B(\psi(X)) = cE_k^A(X)$ for some nonzero constant c and some k with $3 \leq k < n$. Then A and B are either isomorphic or anti-isomorphic.* □

Jacobson [3] proved that the reduced norm form determines the algebra up to isomorphism or anti-isomorphism; I then gave a different proof of that fact in [10], and the argument here follows the lines of [10]. As mentioned in the introduction, this is our result for the case $k = n$. We can incorporate that to state a final corollary:

Corollary 5. *A central simple algebra of dimension greater than 4 carries a canonical cubic form that determines it up to isomorphism or anti-isomorphism.* □

REFERENCES

1. L. R. Beasley, *Linear transformations on matrices: the invariance of the third elementary symmetric function*, Canad. J. Math. **22** (1970), 746–752.
2. H. J. Hoehnke, *Über komponierbare Formen und konkordante hyperkomplexe Grössen*, Math. Z. **70** (1958), 1–12.
3. N. Jacobson, *Structure groups and Lie algebras of Jordan algebras of symmetric elements of associative algebras with involution*, Adv. Math. **20** (1976), 106–150.

4. D. G. James, *Linear transformations of the second elementary function*, Linear and Multilinear Algebra **10** (1981), 347–349.

5. J. C. Jantzen, *Representations of algebraic groups*, Academic Press, New York, 1987.

6. A. Kovacs, *Trace preserving linear transformations of matrix algebras*, Linear and Multilinear Algebra **4** (1977), 243–250.

7. M. Marcus and R. Purves, *Linear transformations on algebras of matrices: the invariance of the elementary symmetric functions*, Canad J. Math. **11** (1959), 383–396.

8. B. R. McDonald, *R-linear endomorphisms of $(R)_n$ preserving invariants*, Mem. Amer. Math. Soc., vol. 46, No. 287, 1983.

9. W. C. Waterhouse, *Introduction to affine group schemes*, Graduate Texts in Math., no. 66, Springer-Verlag, New York, 1979.

10. _____, *Linear maps preserving reduced norms*, Linear Algebra Appl. **43** (1982), 197–200.

11. _____, *Invertibility of linear maps preserving matrix invariants*, Linear and Multilinear Algebra **13** (1983), 105–113.

12. _____, *Automorphisms of* $\det(X_{ij})$ *: the group scheme approach*, Adv. Math. **65** (1987), 171–203.

Department of Mathematics, The Pennsylvania State University, University Park, Pennsylvania 16802
E-mail address: wcw@math.psu.edu