WILD AUTOMORPHISMS OF FREE P.I ALGEBRAS, AND SOME NEW IDENTITIES

George M. Bergman

Summary. In part I, some large families of automorphisms of of a generic matrix algebra $k < x_1, \ldots, x_n >_d$ $(n, d > 1,$ k a field) are constructed. It is shown that when $n = 2$, these are not all tame. Some criteria for an endomorphism of $k < x_1, \ldots, x_n >_d$ to be an automorphism are discussed.

In part II, identities for $d \times d$ matrices are studied using the trick of diagonalizing one of the generic matrices. Among the results obtained are the nonexistence of nontrivial identities whose total degree in the other indeterminates is $< d$, the existence of an essentially unique identity whose degree in those indeterminates is d, and the existence of elements centralizing the distinguished indeterminate y but not lying in $Z[y]$ (Z the center).

\- \- \-

The two parts of this paper are essentially independent, except that part II uses the notation set up in the first two paragraphs of part I, and provides an example referred to in part I, §2.

Part I. Wild automorphisms.

1. The main example. Let k be a field, and let n and d be positive integers. Let

---

1980 MOS Classifications: Primary, 16A38. Secondary, 05C20, 05C45, 16A06, 16A72.

$$R = k < x_1, \ldots, x_n >_d \subseteq M_d(k[x_{111}, \ldots, x_{ndd}])$$

denote the k-algebra generated by n dxd matrices of commuting indeterminates $x_m = ((x_{mij}))$ ($m \in \{1,\ldots,n\}$, $i,j \in \{1,\ldots,d\}$). Thus, R is the free algebra on n indeterminates in the variety of k-algebras satisfying the identities of dxd matrices over commutative k-algebras.

Z will denote the center of R. For $n > 1$, Z consists of <u>scalar</u> matrices aI ($a \in k[x_{111},\ldots, x_{ndd}]$); thus Z is isomorphic to a subring of $k[x_{111},\ldots,x_{ndd}]$. That this subring is strictly larger than k is a celebrated result of E. Formanek and Y. P. Razmyslov (see [1] §12.6, [2] §1.8, [3] §8.1; cf. [12]).

For $1 \leq m \leq n$, let S_m denote $k < x_1,\ldots, x_{m-1}, x_{m+1},\ldots, x_n >_d \subseteq R$. Also, letting t be an additional commuting indeterminate, let

$$T_m = \{ f(x_1,\ldots,x_n) \in R \mid f(x_1,\ldots,x_m+tI,\ldots, x_n) = f(x_1,\ldots,x_n) \}.$$

Clearly, T_m is a subring of R containing S_m. But if n and d are both > 1, T_m also contains elements not in S_m, such as

$$[x_m, x_p] = x_m x_p - x_p x_m \quad (p \neq m).$$

We claim, in fact, that $Z \cap T_m$ is strictly larger than $Z \cap S_m$. For let f be an element of Z not in k, which is homogeneous in each of $x_1,\ldots, x_n$ and chosen so as to minimize its total degree. Take some m such that f has positive degree h in the variable x_m; in particular $f \notin S_m$. Then for all $j > 0$, the coefficient of t^j in $f(x_1,\ldots, x_m+tI, \ldots, x_n) - f(x_1,\ldots, x_n)$ has degree $h-j$ in x_m, hence must be zero by our minimality assumption on h. Hence $f \in T_m$, so $Z \cap T_m \nsubseteq S_m$.

For $1 \leq m \leq n$, $f \in R$, and $c \in k$, let us define endomorphisms $\eta_{m,f}$ and $\varepsilon_{m,c}$ of R by

$$\eta_{m,f}(x_p) = \begin{cases} x_p & \text{if } p \neq m \\ x_p + f & \text{if } p = m, \end{cases}$$

(1)

$$\varepsilon_{m,c}(x_p) = \begin{cases} x_p & \text{if } p \neq m \\ c\, x_p & \text{if } p = m. \end{cases}$$

It is well known, and not hard to see, that if $f \in S_m$, then $\eta_{m,f}$ is an automorphism of R, with inverse $\eta_{m,-f}$, and that if $c \neq 0$ then $\varepsilon_{m,c}$ is an automorphism, with inverse $\varepsilon_{m,c^{-1}}$. The group generated by these automorphisms, as f, c and m range over all admissible values, is called the group of <u>tame automorphisms</u> of R.

One can clearly define the same types of endomorphism $\eta_{m,f}$ and $\varepsilon_{m,c}$ on the <u>free associative</u> algebra $k < x_1, \ldots, x_n >$, and hence the concept of the group of <u>tame automorphisms</u> of this algebra. In fact, it is mainly for this case, and for the $d = 1$ case, i.e. the commuting polynomial algebra $k < x_1, \ldots, x_n >_1 = k[x_1, \ldots, x_n]$, that the concept has been studied. From the fact that every element $f \in S_m$ is the image of an element of the free associative algebra $k < x_1, \ldots, x_{m-1}, x_{m+1}, \ldots, x_n >$, it is easy to deduce that every tame automorphism of $R = k < x_1, \ldots, x_n >_d$ can be <u>lifted</u> to $k < x_1, \ldots, x_n >$, i.e., is the image of a tame automorphism of this free associative algebra, under the natural homomorphism

$$\mathrm{Aut}(\ k < x_1, \ldots, x_n >\ ) \longrightarrow \mathrm{Aut}(\ R\ ).$$

Jung, Gutwirth, Nagata and van der Kulk ([4]-[7]; cf. attributions in [8]) have proved that all automorphisms of $k[x_1, x_2]$ are tame. This implies that the map $\mathrm{Aut}(k<x_1, x_2>) \to \mathrm{Aut}(k[x_1, x_2])$ is surjective. Czerniakiewicz [8] shows that this map is also injective, and hence that all automorphisms of $k<x_1, x_2>$ are tame. (This tameness result is proved in a slightly different way for char $k = 0$ by L. G. Makar-Limanov [12].)

We shall now construct, for $d \geq 2$, $n \geq 2$, nontrivial automorphisms of $R = k < x_1, \ldots, x_n >_d$ which lie in the kernel of the natural map

$$(2) \qquad \mathrm{Aut}(\, k < x_1, \ldots, x_n >_d \,) \to \mathrm{Aut}(\, k < x_1, \ldots, x_n >_{d-1}).$$

In particular, these go to the identity automorphism of $k[x_1, \ldots, x_n]$; hence for $n = 2$ it follows from Czerniakiewicz's result that they cannot be lifted to $k < x_1, x_2 >$, and so are not tame.

Let us first note that if $g \in Z$ and $f \in T_m$, then

$$(3) \qquad \eta_{m,f}\, \eta_{m,g} = \eta_{m,f+g}.$$

Indeed, both sides of (3) fix x_p for all $p \neq m$. The left-hand-side, applied to x_m, gives $(x_m + f) + g(x_1, \ldots, x_m + f, \ldots, x_n)$, which simplifies to $x_m + f + g$ because $g \in T_m$ is unaffected by adding a scalar matrix to its m^{th} argument. Hence the two sides of (3) also agree on x_m.

It follows from (3) that for $f \in Z \cap T_m$, $\eta_{m,f}$ and $\eta_{m,-f}$ are inverses to one another, and in fact, that for each m, the map $f \mapsto \eta_{m,f}$ is a homomorphism of the additive group of $Z \cap T_m$ into $\mathrm{Aut}(\, R\,)$.

Now any element $f \in Z$ with zero constant term in k lies in the kernel of the natural homomorphism $R \to k < x_1, \ldots, x_n >_{d-1}$ ([3] Prop. VII. 1. 2, (4)). In particular, if we take $f \in Z \cap T_m$ having zero constant term, then $\eta_{m,f}$ will be carried by (2) to $\eta_{m,0} = 1$. For $n = 2$ we can conclude as noted above that $\eta_{m,f}$ is not tame.

Example. On $k < x_1, x_2 >_2$, the automorphism $\eta_{1,[x_1,x_2]^2}$, carrying x_1 to $x_1 + [x_1,x_2]^2$ and fixing x_2, is not tame.

Remark: For $n > 2$, the homomorphism $\mathrm{Aut}(k < x_1, \ldots, x_n >) \to \mathrm{Aut}(k [x_1, \ldots, x_n])$ is not one-to-one, even on tame elements. For instance, $\eta_{1,[x_2,x_3]} \in \mathrm{Aut}(k < x_1, x_2, x_3 >)$ is (by definition) tame, but it is mapped to the identity in $\mathrm{Aut}(k[x_1, x_2, x_3])$. Hence, though the above construction gives automorphisms of R which are of the form $\eta_{m,f}$ for $f \notin S_m$, which go to the identity under (2), and for which one has no obvious lifting to automorphisms of $k < x_1, \ldots, x_n >$ (or even $k < x_1, \ldots, x_n >_{d+1}$), nevertheless I see no way of proving that they are not tame, i.e. that they cannot somehow be written as a product of the automorphisms (1). In particular, a non-tame automorphism of $k < x_1, x_2 >_d$, when extended to an automorphism of $k < x_1, x_2, x_3 >_d$ by letting it fix x_3, might, as far as we can tell, become tame.

For $n > 2$ the questions of whether all automorphisms of $k[x_1, \ldots, x_n]$ are tame, and of whether all automorphisms of $k < x_1, \ldots, x_n >$ are tame, are also open.

2. Variants. We have seen that an endomorphism $\eta_{m,f}$ of R is an automorphism if either $f \in S_m$ or $f \in Z \cap T_m$. Neither of these facts is a special case of the other, but there is a common idea behind both. To see it, let us begin by noting that for any $c_1, \ldots, c_n \in k$, the endomorphism φ of R defined by

$$(4) \qquad \varphi(x_1) = x_1 + c_1 , \ldots, \varphi(x_n) = x_n + c_n$$

clearly has an inverse, given by

$$(5) \qquad \psi(x_1) = x_1 - c_1, \ldots, \psi(x_n) = x_n - c_n.$$

But if for $c_1, \ldots, c_n$ we instead use arbitrary elements of R, then ψ will not in general be an inverse to φ — when we compute $\varphi(\psi(x_m))$, we get $x_i + c_m - \varphi(c_m)$, and the latter two terms may not cancel.

The idea behind the two types of automorphisms noted above is to set up (4) so that φ does not modify the c_m's. In one case (that used in generating conventional "tame automorphisms"), φ changed only one of the generators x_m, and $f = c_m$ was chosen from S_m, so that it was not affected by changes in this generator. In the other case, we chose φ "more" restrictively, to alter x_m only by a central element f, but we were then able to be less restrictive in that f could involve x_m, as long as it was insensitive to central elements added to this generator.

Using the same idea, we can give still more constructions for automorphisms of R. For instance, let $m \neq q \leq n$, let $C_q \subseteq R$ be the centralizer of x_q, and let $U_{m,q} \subseteq R$ denote the subring of elements that involve x_m only via the commutator $[x_q, x_m]$, i.e. $U_{m,q} = S_m < [x_q, x_m] > \subseteq R$. Then we see that for $f \in C_q$, $\eta_{m,f}$ fixes all elements of $U_{m,q}$, so in particular, if $f \in C_q \cap U_{m,q}$ $\eta_{m,f}$ will have the inverse $\eta_{m,-f}$. (More generally, (3) holds for $f \in C_q$, $g \in U_{m,q}$.)

Note that

(6) $$C_q \supseteq Z[x_q], \quad \text{hence} \quad C_q \cap U_{m,q} \supseteq (Z \cap U_{m,q})[x_q].$$

From the latter inclusion we can see that the subring $C_q \cap U_{m,q} \subseteq R$ is not contained in either S_m or $Z \cap T_m$. It also clearly does not contain either of these subrings in general, hence this new class of automorphisms neither lies in nor contains either of the classes discussed above.

In Part II, §8 below, we shall see that for $d > 2$ the inclusions of (6) are both strict, giving still further automorphisms in this new class. (In §10 the same result is noted for $d = 2$, $n > 2$.)

One can also "reverse" the construction of section 1! Suppose $f \in R$ has the property that

(7) $$f(x_1, \ldots, x_m + y, \ldots, x_n) = f(x_1, \ldots, x_n) \text{ for any matrix } y \text{ with } 0 = tr(y) \text{ (trace of } y).$$

Then if also $tr(f) = 0$, $\eta_{m,f}$ will clearly again be an automorphism of R. For an example of an f satisfying (7), (specifically, having the property $f = tr(x_m)\, g$, where the matrix g involves only the entries of $\{x_p \mid p \neq m\}$) see [1] p.467, Exercise 8. This f does not itself has zero trace, but the element $[x_{n+1}f, x_{n+2}]$ in $k< x_1, \ldots, x_{n+2} >_d$ clearly will. (By standard tricks, the number of matrix indeterminates needed can be reduced from $n+2$ (the n used in this example happens to be $3d^2$) to 3.)

Let us note that if the characteristic of k does not divide d, any $d \times d$ matrix over a k-algebra may be written uniquely as the sum of a scalar matrix and a matrix with trace 0. Hence to give an n-tuple of independent $d \times d$ matrices is equivalent to giving $2n$ $d \times d$ matrices, n of which are specified as being scalar and the other n as having zero trace, but otherwise independent. In terms of these "coordinates", the type of automorphisms introduced in section 1, and the above "reversed" example, both look like tame automorphisms: One of the new "coordinates" is being altered by a function of the other "coordinates." Further, note that once a matrix x_1 has been given — assuming it has distinct eigenvalues, and so in particular can be diagonalized over an appropriate extension field, which is true of a matrix of indeterminates — then to give another matrix x_2 is equivalent to giving one matrix which commutes with x_1, i.e. is diagonal with respect to the same basis as x_1, and another which has diagonal entries all zero when expressed in that basis, i.e. is a commutator $[x_1, y]$. In terms of this decomposition, the other class of automorphisms described in this section also looks tame.

The weakness of this viewpoint is that the "components" of our generic matrices with respect to the indicated decompositions do not themselves lie in $R = k < x_1, \ldots, x_n >_d$, so that we cannot really go to a new set of generators for R as we would like. Nevertheless, I find the viewpoint suggestive.

It is interesting to note that if we write $F = k \langle\!\langle x_1, \ldots, x_n \rangle\!\rangle_d$ for the skew field of fractions of R, then the "components" in question will lie in F. Indeed, if we let L denote the center of F, then the trace map is a K-linear map $F \to L$, namely $tr(x) = 1/d$ times the trace of the action of x by left multiplication on the d^2-dimensional L-vector-space F; hence the decomposition $x = (1/d)\, tr(x) + (x - (1/d)\, tr(x))$ makes sense in F. Similarly, if we write $ad(x_1)$ for the operator $y \mapsto [x_1, y]$, then as x_1 has distinct eigenvalues, one can prove, by looking at an extension of scalars to a field containing those eigenvalues, that $F = \ker ad(x_1) \oplus \operatorname{im} ad(x_1)$, which allows one to give the second decomposition.

Observations along these lines allow one to construct large classes of "pseudo-tame" automorphisms of F. Further, using the known fact [2, p.91, Lemma 3] that L is just the field of fractions of Z, one can show that most of these classes have nonidentity members defined on R. We omit the details.

Let us note, however, that the skew field F has still wider classes of easily constructible automorphisms than those mentioned so far! Clearly, for any m, the operation of multiplying x_m by some nonzero member of F depending only on the other indeterminates induces an automorphism of F. To complicat things further, the concept "depending only on the other indeterminates" cannot be interpreted simply as "lying in the sub-skew-field generated by the other indeterminates". For example, we know that $k \langle x_1, x_2 \rangle_d$ contains the scalar $tr(x_1)$ in its center, and one can deduce that it has an automorphism fixing x_1, and carrying x_2 to $x_2\, tr(x_1)$. But $tr(x_1)$ clearly does not lie in the subfield $k(x_1)$. However, this type of peculiar behavior may be limited to the two-variable case. E.g. for $n=3$, $tr(x_1)$ does lie in the sub-skew-field $k \langle x_1, x_3 \rangle_d$ generated by the variables other than x_2.

Let us turn back to the beginning of this section and expand on one point we made. Though we saw that (4) has (5) for an inverse if and only if $c_1, \ldots, c_n$ are fixed by φ, it will have an inverse if and only if the weaker statement

(8) $$c_1, \ldots, c_n \in \varphi(R)$$

holds. Indeed, necessity is clear. Conversely, if we can write $c_m = \varphi(d_m)$ ($d_m \in R$, $m = 1, \ldots, n$), then we claim an inverse to (4) is defined by

$$\psi(x_1) = x_1 - d_1, \ \ldots, \ \psi(x_n) = x_n - d_n.$$

For direct computation shows that $\varphi\psi = 1$. If $\psi\varphi$ were not 1, then φ would have to be a surjective but noninjective endomorphism of R. But arguments involving the transcendence degree of Z show that such endomorphisms cannot exist.

However, I have not been able to use the weaker criterion (8) to find any new examples of automorphisms.

3. <u>A converse</u>. Let $n, d > 1$, and let f be a member of Z, the center of $R = k < x_1, \ldots, x_n >_d$. We have seen that a sufficient condition for $\eta_{m,f}$ to be an automorphism of R is that f also lie in T_m. Let us now show that this is also necessary.

We first note that for any $f(x_1, \ldots, x_n) \in Z$,

(9) $$f(0, \ldots, x_m, \ldots, 0) \in k.$$

For the above element is a polynomial in x_m, and no nonconstant polynomial in one matrix of indeterminates can give a scalar matrix. (This is most easily seen by specializing x_m to a diagonal matrix with distinct indeterminate diagonal entries.)

Now let K denote the algebraic closure of k. The field K is infinite, hence for any $f \in R - T_m$ we can find particular values

(10) $$\xi_1, \ldots, \xi_n \in M_n(K)$$

such that

(11) $$f(\xi_1, \ldots, \xi_m + tI, \ldots, \xi_n) \in M_n(K[t]) \text{ involves } t.$$

In particular, if $f \in Z - T_m$, then the element given in (11) will be the identity matrix times a nonconstant polynomial $F(t) \in K[t]$. We now consider two cases:

<u>Case 1.</u> The matrices (10) can be chosen so that $F(t)$ has degree > 1. Then we can clearly find distinct elements $\tau, \tau' \in K$ such that

(12) $$\tau + F(\tau) = \tau' + F(\tau').$$

Let us define two homomorphisms $\varphi, \varphi' : R \to M_n(K)$, by letting $\varphi(x_p) = \varphi'(x_p) = \xi_p$ for all $p \neq m$, while $\varphi(x_m) = \xi_m + \tau I$, $\varphi'(x_m) = \xi_m + \tau' I$. We claim that

$\varphi\eta_{m,f} = \varphi'\eta_{m,f}$. Indeed, these two homomorphisms $R \to M_n(K)$ clearly agree on x_p $(p \neq m)$. On x_m the first gives $\varphi(x_m + f) = (\xi_m + \tau I) + f(\xi_1,\ldots,\xi_m+\tau I,\ldots,\xi_n)$ $= \xi_m + \tau I + F(\tau)I$, while the second gives $\xi_m + \tau' I + F(\tau')I$; and these agree by (12). Thus φ and φ' agree on the image of $\eta_{m,f}$, hence $\eta_{m,f}$: $R \to R$ cannot be surjective.

Case 2. For all choices of $\xi_1,\ldots,\xi_n$, $F(t)$ has the form $\alpha + \beta t$. Begin by choosing values (10) so that (11) is satisfied. Then introducing another commuting indeterminate s, we see that $f(s\xi_1,\ldots,\xi_m+tI,\ldots, s\xi_n)$ will be a polynomial $\alpha(s) + \beta(s)t$ $(\alpha, \beta \in K[s])$. Putting $s = 0$, we see

$$f(s\xi_1,\ldots,\xi_m+tI,\ldots, s\xi_n) = (\alpha(s) + \beta(s)t)I, \text{ where } \alpha, \beta \in K[s].$$

Putting $s = 0$, we see
from (9) (with ξ_m+tI for x_m) that $\beta(0) = 0$. But by (11), $\beta(1) \neq 0$. So β is a nonconstant polynomial, so we can find $\sigma \in K$ such that $\beta(\sigma) = -1$. So replacing ξ_p by $\sigma\xi_p \in M_n(K)$ $(p \neq m)$, we get $F(t)$ in the form $\alpha - t$. So now (12) holds for all τ, τ', and the proof can be completed as before.

Example. For char $k \neq 2$, the endomorphism $\eta_{1,[x_1^2,\ x_2]^2}$ of $k < x_1, x_2 >_2$ is not an automorphism, in contrast with the example in the first section. In fact, one finds that the homomorphisms φ, φ': $k < x_1, x_2 >_2 \to M_2(k)$ defined by $\varphi(x_1) = e_{12}$, $\varphi'(x_1) = e_{12} - I$, $\varphi(x_2) = \varphi'(x_2) = (1/2)e_{21}$ have the same compositions with $\eta_{1,[x_1^2,\ x_2]^2}$ so this endomorphism cannot have x_1 in its image.

4. An algebraic geometric viewpoint. Again let $R = k < x_1,\ldots,x_n >_d$, and let K be the algebraic closure of k. A k-algebra endomorphism θ of R is determined by the n-tuple of elements

$$\theta(x_1) = a_1(x_1,\ldots,x_n), \quad . \quad . \quad . \quad , \quad \theta(x_n) = a_n(x_1,\ldots,x_n) \in R.$$

This n-tuple in turn determines a set-map θ^*: $M_d(K)^n \to M_d(K)^n$, by

(13) $$\theta^*(\xi_1,\ldots,\xi_n) = (a_1(\xi_1,\ldots,\xi_n),\ldots, a_n(\xi_1,\ldots,\xi_n)).$$

Identifying $M_d(K)$ set-theoretically with K^{d^2}, θ^* may be regarded as a polynomial map $K^{d^2n} \to K^{d^2n}$. This is equivalent to an endomorphism θ' of $k[x_{111},\ldots, x_{ndd}]$. Here θ^* is contravariant in θ, but θ' is again covariant in θ; θ and θ' are related by the formula $\theta(f)_{ij} = \theta'(f_{ij})$ where $f \in R$, and f_{ij} denotes the (i,j) entry of the matrix f.

Clearly a necessary condition for

(14) $\theta \in \mathrm{Aut}(R)$

to hold is

(15) θ^* is bijective, equivalently $\theta' \in \mathrm{Aut}\, k[x_{111},\ldots, x_{ndd}]$.

The two statements of (15) are equivalent by algebraic geometry. A necessary condition for (15) to hold is, in turn

(16) The determinant of the $nd^2 \times nd^2$ Jacobian matrix for θ^*, given by an element of $k[x_{111},\ldots,x_{ndd}]$, is invertible, i.e. lies in $k - \{0\}$.

The implication from (14) to (the first form of) (15) is essentially the criterion we used in the preceding section.

As to whether (15) $\Rightarrow$ (16) is reversible, this is a case of an outstanding question of commutative ring theory, the Jacobian Determinant Problem: If a polynomial map $\varphi : K^r \to K^r$ has constant nonzero Jacobian determinant (say everywhere 1), must it be bijective, i.e. must the corresponding endomorphism of $K[x_1,\ldots,x_r]$ be an automorphism? This is not known even for $r = 2$.

We remark that Czerniakiewicz has conjectured that an endomorphism of the free algebra $k < x_1, x_2 >$ is an automorphism if and only if it preserves the element $[x_1, x_2]$, up to a nonzero scalar factor. ("Only if" follows from her tameness result.) Though there is no apparent generalization of this conjecture to $n > 2$, it seems (as A. Hausknecht has pointed out to me) to be somehow related to the Jacobian determinant conjecture for $k[x_1, x_2]$.

Part II. Identities

The remainder of this paper studies identities in matrix algebras, equivalently equations holding in free p.i. algebras $R = k < x_1, \ldots >_d$. They were inspired by the problem of finding elements centralizing one indeterminate x_q, but not lying in the "obvious" subring $Z[x_q]$. Such elements are indeed obtained, but the methods are probably of as much interest for their ability to show precisely what elements in certain large spaces of polynomials are identities as for these particular examples.

The approach, loosely speaking, is to distinguish one indeterminate y, and consider polynomials of specified degrees, and possibly also orders of occurrence, of the other indeterminates, but arbitrary in the way y can occur. The study of such polynomials is aided by the trick of diagonalizing the matrix y, used before by Procesi.

5. <u>The basic tool: the map</u> α_y. Throughout the remaining sections, k will be a commutative integral domain, and n a positive integer. We shall be using the commutative polynomial ring in $n+1$ indeterminates, $k[t_0,\ldots, t_n]$, which will be abbreviated $k[t_*]$.

If R is any associative k-algebra, and y an element of R, we define a map

$$\alpha_y\colon\ k[t_*] \times R^n \rightarrow R$$

by specifying that for monomial arguments in $k[t_*]$ it be given by

$$(17)\qquad \alpha_y(t_0^{r_0}\, t_1^{r_1}\ldots\, t_n^{r_n},\ x_1, \ldots, x_n) = y^{r_0}\, x_1\, y^{r_1}\ldots\, y^{r_{n-1}}\, x_n\, y^{r_n},$$

and that it be k-linear in each of its $n+1$ arguments. Loosely speaking, then, t_m $(0 \leq m \leq n)$ represents the operation of inserting a factor of y between x_m and x_{m+1} (<u>mutatis mutandis</u> for $m = 0, n$) before multiplying the x's together. Given $x_1,\ldots, x_n$, the image $\alpha_y(k[t_*], x_1,\ldots, x_n) \subseteq R$ can be described as $k[y]\, x_1\, k[y] \ldots k[y]\, x_n\, k[y]$ (i.e. the subgroup spanned by all products of elements from these sets.)

We note that for any $f \in k[t_*]$, and $0 \leq m \leq n$,

$$(18)\qquad \alpha_y(\, t_m\, f,\ x_0,\ldots, x_n\,) = \alpha_y(f,\, x_1,\ldots,\, x_m\, y,\, x_{m+1}, \ldots,\, x_n)$$

$$= \alpha_y(f,\, x_1,\ldots,\, x_m,\, y\, x_{m+1}, \ldots,\, x_n),$$

except that for $m = 0$, the first of the two formulas on the right should be replaced by $y\,\alpha_y(f, x_1,\ldots, x_n)$, and for $m = n$, the second should be replaced by $\alpha_y(f, x_1, \ldots, x_n)\, y$. In particular we see that for $1 \leq m \leq n$,

(19) $\alpha_y(\ (t_{m-1}-t_m)f,\ x_1,\dots,x_n) = \alpha_y(f,\ x_1,\dots,\ [y,\ x_m],\dots,\ x_n)$,

and that

(20) $\alpha_y(\ (t_0-t_n)f,\ x_1,\dots,\ x_n) = [y,\ \alpha_y(f,\ x_1,\ \dots,\ x_n)]$.

6. Identities in $k < x_1,\dots,\ x_n,\ y >_d$. We now fix another integer d, and let R be the free p.i. algebra

$$R = k < x_1,\ \dots,\ x_n,\ y >_d.$$

We shall abbreviate $x_1,\dots,\ x_n$ to x_*, so that we can write R as $k < x_*,\ y >_d$ and $\alpha_y(f,\ x_1,\dots,\ x_n)$ as $\alpha_y(f,\ x_*)$. (Note, however, that t_* represents a string of length $n + 1$, while x_* represents a string of length n.)

It is well-known that R may in fact be represented as a ring of matrices over a commutative polynomial algebra, such that y is a diagonal matrix of indeterminates, while the x_m's have indeterminates for all d^2 entries:

(21) $y = \sum y_{ii}\ e_{ii},\quad x_m = \sum x_{mij}\ e_{ij}\quad (y_{ii},\ x_{mij}$ distinct indeterminates).

(Cf. [9], p.251, paragraph following Lemma 1.6.)

Taking R in this form, let us compute the right hand side of (17) explicitly. We get

(22) $$\sum y^{r_0}_{i_0 i_0} x_{1 i_0 i_1} y^{r_1}_{i_1 i_1} \cdots y^{r_{n-1}}_{i_{n-1} i_{n-1}} x_{n i_{n-1} i_n} y^{r_n}_{i_n i_n} e_{i_0 i_n},$$

where the sum is over all strings of indices

(23) $$i_* = (i_0, \ldots, i_n) \quad (i_0, \ldots, i_n \in \{1, \ldots, d\}).$$

It follows that for any $f \in k[t_*]$, we have

(24) $$\alpha_y(f, x_*) = \sum_{i_*} f(y_{i_0 i_0}, \ldots, y_{i_n i_n})\, x_{1 i_0 i_0} \cdots x_{n i_n i_n}\, e_{i_0 i_n}.$$

In particular, $\alpha_y(f, x_*)$ will be zero if and only if for every i_*, the coefficient of $x_{1 i_0 i_1} \cdots x_{n i_{n-1} i_n} e_{i_0 i_n}$ is (24) is zero, i.e.

(25) $$f(y_{i_0 i_0}, \ldots, y_{i_n i_n}) = 0 \quad \text{for all} \quad i_* \in \{1, \ldots, d\}^{n+1}.$$

For n small, this condition assumes a very simple form.

(26) For $n < d$, $\alpha_y(f, x_*) = 0$ in $R = k \langle x_*, y \rangle_d$ if and only if $f = 0$ in $k[t_*]$.

This follows by looking at the particular string $i_* = (1, 2, \ldots, n+1)$, and recalling that the y_{ii} are independent indeterminates. For $n = d$, we get a more interesting criterion:

(27) For $n = d$, $\alpha_y(f, x_*) = 0$ in $R = k < x_*, y >_d$ if and only if f is divisible by $\prod_{0 \le p < q \le d} (t_p - t_q)$.

To see "if", note that when $n = d$, every string i_* must have at least one repetition: $i_p = i_q$ $(p < q)$. Hence when we evaluate $f(y_{i_0 i_0}, \ldots, y_{i_n i_n})$, the divisibility of f by $t_p - t_q$ will cause the value to be zero. Conversely, if f is not divisible by $\prod (t_p - t_q)$, let us choose a factor $t_p - t_q$ which does not divide it. Then we can form an index-string i_* with $i_p = i_q$, and all other pairs of indices distinct. Then we will clearly have $f(y_{i_0 i_0}, \ldots, y_{i_n i_n}) \neq 0$, so $\alpha_y(f, x_0) \neq 0$.

To formulate the criterion for higher n, we will want some notation. If ρ is an equivalence relation on $\{0, \ldots, n\}$, let $I_\rho \subseteq k[t_*]$ denote the ideal generated by $\{ t_p - t_q \mid (p,q) \in \rho \}$. This is precisely the kernel of the natural homomorphism $k[t_*] \to k[t_{(*/\rho)}]$, where $t_{(*/\rho)}$ denotes a family of indeterminates indexed by the equivalence classes of ρ. (I_ρ can be pictured as the ideal of all polynomials $f \in k[t_*]$ which are zero on the subspace A_ρ of affine space A^{n+1} defined by the equations $t_p = t_q$ for $(p,q) \in \rho$. The dimension of this space is the number of equivalence classes in ρ.) Note that if $\rho \subseteq \rho'$, then $I_\rho \subseteq I_{\rho'}$.

Now for any index-string i_* as in (23), $\{ (p, q) \mid i_p = i_q \}$ is clearly an equivalence relation ρ on $\{0, \ldots, n\}$, and we have $f(y_{i_0 i_0}, \ldots, y_{i_n i_n}) = 0$ if and only if $f \in I_\rho$. Further, if ρ has fewer than d equivalence classes, then we can always refine it to an equivalence relation

ρ' with exactly d equivalence classes, which (assuming $n+1 \geq d$) will in turn be represented by some other index-string i'_*, and will define a smaller ideal. Now let us define $E_{n+1,d}$ to be the set of all equivalence relations on $\{0,\ldots,n\}$ with exactly d equivalence classes. Then we see from the above remarks:

(28) If $n \geq d-1$, then $\alpha_y(f,x_*) = 0$ if and only if $f \in \bigcap_{\rho \in E_{n+1,d}} I_\rho$.

(Remarks: This includes (27) and the $n = d-1$ case of (26). However, for $n > d$ the intersection of ideals in (28) is not generally their product as it is for $n = d$. For instance, if $d = 2$, $n = 3$, then $E_{n+1,d}$ has seven members — three that partition $\{0,1,2,3\}$ into two pairs of elements and four that partition it into a singleton and a three-element set. Hence $\prod I_\rho$ lies in $(t_0, t_1, t_2, t_3)^7$. But $\bigcap I_\rho$ can easily be seen to contain $(t_0-t_1)(t_0-t_2)(t_1-t_2)$, which has degree 3, hence will not lie in the product.)

We now wish to ask which elements $\alpha_y(f, x_*)$ commute with y in R. By (20), this is so if and only if $\alpha_y((t_0-t_n)f, x_*) = 0$. So (26) and (27) immediately give:

(29) If $n < d$, then $\alpha_y(f, x_*)$ centralizes y only if $f = 0$.

(30) If $n = d$, then $\alpha_y(f, x_*)$ centralizes y if and only if f is divisible by all factors $t_p - t_q$ $(0 \leq p < q \leq d)$ except $t_0 - t_d$.

To handle larger n, let $E'_{n+1,d}$ denote $\{\rho \in E_{n+1,d} \mid (0,d) \notin \rho\}$. Then we claim

(31) If $n \geq d-1$, then $\alpha_y(f,x_*)$ centralizes y if and only if $f \in \bigcap_{\rho \in E'_{n+1,d}} I_\rho$.

Indeed, first assume that f lies in the intersection indicated above. Consider any $\rho \in E_{n+1,d}$. If $(0,d) \notin \rho$, then $\rho \in E'_{n+1,d}$, so $f \in I_\rho$. If $(0,d) \in \rho$, then $t_0 - t_d \in I_\rho$. So in either case, $(t_0 - t_d)f \in I_\rho$, hence $(t_0 - t_d)f \in \bigcap_{E_{n+1,d}} I_\rho$, which is the condition for $\alpha_y(f, x_*)$ to centralize y. Conversely, suppose that $\alpha_y(f, x_*)$ centralizes y, i.e. that $(t_0 - t_d)f$ lies in $\bigcap_{E_{n+1,d}} I_\rho$. Then for any $\rho \in E'_{n+1,d}$ we have $(t_0 - t_d)f \in I_\rho$, but $t_0 - t_d \notin I_\rho$. Now I_ρ is prime, because it is the kernel of a homomorphism into an integral domain, $k[t_{(*/\rho)}]$, so we conclude that $f \in I_\rho$.

For the case $n = d$, we can get a precise description of the space of y-centralizing elements of the form $\alpha_y(f, x_*)$.

(32) Let $n = d$, let $u = \prod_{0 \leq p < q \leq d,\ (p,q) \neq (0,d)} (t_p - t_q)$, and let $\alpha_y(u, x_*) = U$. Then the centralizer of y in the k-module $k[y]\ x_1\ k[y] \ldots k[y]\ x_d\ k[y]$ has a basis consisting of the elements $U(y^{r_0} x_1, \ldots, y^{r_{d-1}} x_d)$ $(r_0, \ldots, r_{d-1} \geq 0)$.

For we know that the centralizer consists of all elements $\alpha_y(gu, x_*)$ $(g \in k[x_*])$, and by (27), two elements g, g' yield the same member of the centralizer if and only if they differ by a multiple of $t_0 - t_d$. Now the quotient map $k[t_*] \to k[t_*]/(t_0 - t_d)$ has a section with image $k[t_0, \ldots, t_{d-1}] \subseteq k[t_*]$, and the monomials $t_0^{r_0} \ldots t_{d-1}^{r_{d-1}}$ which form a k-basis for this subring, correspond to the elements described in (32), in view of (18).

(One can similarly obtain bases for the centralizer of y when $n > d$, but the calculations become less trivial.)

The form of our construction considered in this section yields no central polynomials:

(33) For any $d > 1$, $n > 1$, $\alpha_y(f,x_*)$ is central if and only if it is zero.

For if $\alpha_y(f,x_*)$ is central, then so is $\alpha_y(f, yx_1, x_2,\ldots, x_n) = y\,\alpha_y(f, x_*)$. But if A and yA are both scalar matrices, A must be 0.

However, we cannot use (33) to deduce that the elements commuting with y which we have constructed do not lie in $Z[y]$. Indeed, if such an element could be written $\sum y^r c_r$ $(c_r \in Z)$, we could reduce to the case where each c_r was homogeneous of degree 1 in each x_m, but we could not conclude that each c_r lay in $k[y]\, x_1\, k[y] \ldots k[y]\, x_n\, k[y]$. For c_r could be a sum of terms in which the x_m occurred in various orders, and we have not shown that such a sum cannot be central and nonzero.

In the next three sections we shall get out of this predicament by imposing the relations $x_1 = \ldots = x_n$, and prove the existence of y-centralizing elements not in $Z[y]$. In section 10, on the other hand, we shall sketch how one <u>can</u> study expressions in which distinct x_m occur in various orders.

7. <u>Identities in</u> $k\langle x, y\rangle_d$, <u>and directed graphs.</u> We continue to let n, d be fixed positive integers, and $R = k\langle x_*, y\rangle_d$, but now let us also consider the algebra on two generic matrices, $S = k\langle x, y\rangle_d$. We define the algebra homomorphism $\iota: R \to S$, taking all x_m to x, and y to y. Thus, writing $\bar{\bar{x}}$ for the n-tuple $(x,\ldots,x) \in S^n$, we have

(34) $$\iota(\alpha_y(f, x_*)) = \alpha_y(f, \bar{\bar{x}}).$$

In S, as in R, y will be taken diagonal:

(35) $$x = \sum x_{ij}\, e_{ij}, \quad y = \sum y_{ii}\, e_{ii}.$$

Thus we have the analog of (24):

(36) $$\alpha_y(f, \bar{x}) = \textstyle\sum_{i_*} f(y_{i_0 i_0}, \ldots, y_{i_n i_n})\, x_{i_0 i_1} \ldots x_{i_{n-1} i_n}\, e_{i_0 i_n}.$$

Note however that in contrast to the case of (24), the indeterminate factors of the expressions

(37) $$x_{i_0 i_1} \ldots x_{i_{n-1} i_n}\, e_{i_0 i_n}$$

are not taken from n disjoint families $\{x_{1ij}\}, \ldots, \{x_{nij}\}$, but from the single family $\{x_{ij}\}$. Hence the element (37) may not uniquely determine the string i_*, i.e. distinct strings i_*, i_*' can yield the same term (37); hence we can no longer assert that the sum (36) will be zero if and only if each summand is zero.

Nonetheless, this will be true when n is small. We claim

(38) If $n \leq d$, then $\alpha_y(f, \bar{x}) = 0$ in S if and only if $\alpha_y(f, x_*) = 0$ in R. (Equivalently, the restriction of the map $\iota: R \to S$ to $k[y]\, x_1\, k[y] \ldots k[y]\, x_n\, k[y]$ is one-to-one.)

The "if" part of the first sentence is clear, by (34), which also shows the equivalence of the parenthetical version. When n is strictly less than d, "only if" is also easy. Consider the expression (37) corresponding to the index-string $(1, \ldots, n+1)$, namely

(39) $$x_{12} x_{23} \ldots x_{n,n+1}\, e_{1,n+1}.$$

If this element also has the form (37) for some other string i_*, then looking at the factor $e_{1,n+1}$ we see that i_0 must be 1; and if we assume inductively that $i_{m-1} = m$, and then note that (39) has a unique factor with first subscript equal to m, namely $x_{m,m+1}$, we conclude that $i_m = m+1$. Hence the coefficient of (39) in (36) is precisely $f(y_{11},\ldots,y_{n+1,n+1})$, and we conclude that $\alpha_y(f, \bar{x}) = 0$ if and only if $f = 0$, just as for $\alpha_y(f, x_*)$.

The key to showing (38) in the case n=d will be to show that a single repetition among the indices $i_0, \ldots, i_n$ is not enough to prevent us from reconstructing i_* from (37). In proving this it will be convenient to introduce a graph-theoretic viewpoint applicable to the study of the general case $n \geq d$.

As in [13], a graph will mean a directed graph, in which several edges are allowed between vertices, and edges are allowed to connect vertices to themselves. Let us associate to every element (37) a graph G having vertex-set $\{i_0,\ldots,i_n\} \subseteq \{1,\ldots,d\}$. For any two vertices i, j, G will have one edge with initial vertex i and terminal vertex j for each occurrence of a factor equal to x_{ij} in (37), and also one more edge if the matrix-unit factor of (37) is e_{ji}. (Note order of indices!) This last will be called the distinguished edge of G.

To capture some obvious properties of this construction, we define a (d,n)-admissible graph (or when d and n are fixed, an admissible graph) to mean a connected graph G with vertex-set contained in $\{1,\ldots,d\}$, such that G has exactly n+1 edges, one of which is distinguished, and such that the same number of edges lead into each vertex as lead out. (In the language of [13] the last of these conditions says that G is "pseudosymmetric". Note the need, in our construction of a graph G from an element (37), to include the distinguished edge and orient it as we did, if G is to satisfy this condition.)

Two nondistinguished edges e, e' of an admissible graph G will be called equivalent if they have the same initial vertices, and the same terminal vertices.

If G is an admissible graph, a traverse of G will mean a sequence

$i_* = (i_0,\ldots,i_n)$ of vertices of G, such that for each i and j, the number of values $m > 0$ for which $(i_{m-1}, i_m) = (i,j)$ is the number of nondistinguished edges of G with initial vertex i and terminal vertex j, and such that i_n and i_0 are the initial and terminal vertices of the distinguished edge of G. (Note that a traverse does not specify which of a family of equivalent edges is "taken" at a given step. It is only required that the number of times one passes from i to j equal the number of edges leading from i to j. If one takes a traverse and specifies a distinct edge at each step, it becomes equivalent to an Eulerian circuit of G [13]. From [13], p.239 Theorem 6, every admissible graph has a traverse, though we will not use this fact. By [13] p.240 Theorem 8, there are in fact generally a great number of traverses.)

It is now easy to see that if u is a term which can be written in the form (37), G the admissible graph associated with u, and $T(G) \subseteq \{1,\ldots,d\}^{n+1}$ the set of traverses of G, then the total coefficient with which u will occur in $\alpha_y(f, \bar{x})$ is

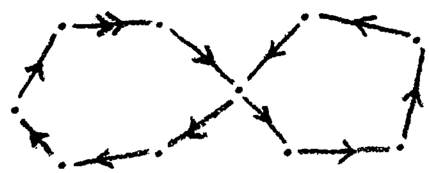
(40) $$\sum_{i_* \in T(G)} f(y_{i_0 i_0}, \ldots, y_{i_n i_n}).$$

Thus

(41) $\alpha_y(f, \bar{x}) = 0$ if and only if f has the property that the sum (40) is zero for all (d,n)-admissible graphs G.

We now specialize to the case $n = d$. (We will come back to general n in section 9.) Consider an index-string i_* with exactly one pair of equal terms, $i_p = i_q$. Then the associated graph G will have essentially the following form, where the double arrow indicates the distinguished edge.

(42)



The graph may be "degenerate" in one or more ways — one of the branches may be reduced to a single loop, or the distinguished edge may have one or both vertices on the "double point". Nevertheless, it is not hard to see that in constructing a traverse for G, when one comes to the "double point" one must switch from one branch to the other, otherwise one will never be able to get onto the other branch. It follows that the given string i_* is the unique traverse of G. Hence if $\alpha_y(f, \bar{x}) = 0$ $(f \in k[t_*])$, we must have $f(y_{i_0 i_0}, \ldots, y_{i_n i_n}) = 0$ for all strings i_* with exactly one pair of equal terms. But this is exactly the set of conditions from which we deduced the necessity of our necessary and sufficient conditions for $\alpha_y(f, x_*) = 0$ in (27). Thus for $n = d$, if $\alpha_y(f, \bar{x}) = 0$, then also $\alpha_y(f, x_*) = 0$, which completes the proof of (38).

8. y-centralizing and central elements of $k < x, y >_d$.

In this section we continue to assume $n = d$. From (20) and (38), it follows also that $\alpha_y(f, \bar{x})$ will centralize y if and only if $\alpha_y(f, x_*)$ does, i.e. if and only if f is divisible by all $t_p - t_q$ $(0 \leq p < q \leq d)$. except $t_0 - t_d$. However, the proof that no nonzero element $\alpha_y(f, x_*)$ is central is not applicable to $\alpha_y(f, \bar{x})$. We shall now investigate conditions for centrality.

Clearly $\alpha_y(f, \bar{x})$ will lie in the center, $Z(S)$, if and only if, first, it is diagonal, i.e. centralizes y, and secondly, has all diagonal entries equal. It is easy to see from our development of the condition for diagonality, (32), that when this holds, the only nonzero terms in the expansion (36) of $\alpha_y(f,\bar{x})$ will be those indexed by strings i_* such that $i_0 = i_d$ is the only repeated term, i.e. strings $(i_0,\ldots, i_{d-1}, i_0)$ where $(i_0,\ldots, i_{d-1})$ is a permutation of $\{1,\ldots,d\}$. For such a sequence, the coefficient of $x_{i_0 i_1}\cdots x_{i_{d-1}i_0}\, e_{i_0 i_0}$ will be $f(y_{i_0 i_0},\ldots, y_{i_{d-1}i_{d-1}}, y_{i_0 i_0})$. For $\alpha_y(f, \bar{x})$ to have equal diagonal entries, this must also be the coefficient of $x_{i_0 i_1}\cdots x_{i_{d-1}i_0}\, e_{i_m i_m}$ for all $m \leq d$. If we rewrite this term as $x_{i_m i_{m+1}}\cdots x_{i_{d-1}i_0} x_{i_0 i_1}\cdots x_{i_{m-1}i_m}\, e_{i_m i_m}$, we see that its coefficient is $f(y_{i_m i_m},\ldots, y_{i_{d-1}i_{d-1}}, y_{i_0 i_0},\ldots, y_{i_m i_m})$.

Since the y_{ii} are independent indeterminates, the condition for these two terms to be equal for all m can be written in a form independent of the particular sequence i_*; it is

(43) The polynomial $f(t_0, t_1,\ldots, t_{d-1}, t_0) \in k[t_0,\ldots,t_{d-1}]$ is invariant under cyclic permutations of the d-tuple of variables $(t_0,\ldots,t_{d-1})$.

Now recall from (30) that for all $m \neq 0, d$, f is divisible by both $t_0 - t_m$ and $t_m - t_d$. Hence $f(t_0,\ldots, t_{d-1}, t_0)$ will be divisible by $(t_0-t_m)^2$. Hence by (43), it must be divisible by $(t_p-t_q)^2$ for all $p \neq q$ in $\{0,\ldots, d-1\}$. Now by (32) and the paragraph following it, f may be chosen to have the form $f = ug$, where u is as in (32) and $g \in k[t_0,\ldots,t_{d-1}]$ is unique for the given central element. Applying the above observations to this f, we easily obtain

(44) Let $n = d$ and let $v = (\prod_{m \neq 0,d} (t_0 - t_m)(t_m - t_d))(\prod_{0<p<q<d} (t_p - t_q)^2)$.

Then any element of the center of $k\langle x, y\rangle_d$ homogeneous in x of degree d (i.e., every member of $k[y] \times k[y] \ldots k[y] \times k[y]$, with d x's) can be written uniquely in the form $\alpha_y(vh, \bar{x})$, where $h \in k[t_0, \ldots, t_{d-1}]$ can be any element invariant under cyclic permutation of the d-tuple of variables $(t_0, \ldots, t_{d-1})$.

In particular, $\deg(v) = d^2 - d$ will be the least degree in y that any nonzero central element of $S = k < x, y >_d$ homogeneous in x of degree d can have. On the other hand, by (32), nonzero y-centralizing elements of degree d in x can have degree in y as small as $\deg(u) = (d^2 + d - 2)/2$. For $d \geq 3$, the latter degree is smaller, hence the y-centralizing element $\alpha_y(u, \bar{x})$ cannot lie in the ring $Z(S)[y]$. From (19) we can see that this element depends on each x_m only via the commutator $[x_m, y]$, showing that the inclusions (6) of Part I are strict, as claimed, at least for $n > 2$.

Note also that if $\alpha_y(f, \bar{x}) \notin Z(S)[y]$, then $\alpha_y(f, x_*) \notin Z(R)[y]$. This shows that the class of y-centralizing elements constructed in section 6 also contains elements not in the ring generated by y over the center.

Examples. $d = n = 2$. In this case, the degrees of the smallest central and y-centralizing elements, computed above, are both 2, and in fact $u = v = (t_0 - t_1)(t_1 - t_2)$. Here $\alpha_y(v, \bar{x})$ is the "ancient" central polynomial $[x,y]^2$. In this case, one can extend the preceding analysis to show that the y-centralizing elements of S homogeneous in x of degree 2 do all lie in $Z(S)[y]$. (However, we shall see in section 10, paragraph following (61), that the y-centralizing element of $k < x_1, x_2, y >_2$, $\alpha_y(v, x_*)$, which in this case is $[x_1,y][x_2,y]$, does not lie in $Z(R)[y]$, completing the proof that the inclusions (6) are strict.)

$d = n = 3$. Then the smallest f such that $\alpha_y(f,\bar{x})$ centralizes y is $u = (t_0-t_1)(t_0-t_2)(t_1-t_2)(t_1-t_3)(t_2-t_3)$, while to get one such that $\alpha_y(f,\bar{x})$ is central, one must put an exponent 2 on the middle factor $t_1 - t_2$, getting the 6th degree polynomial v.

$d = n = 10$. In this case one finds that u has degree 54 and v has degree 90.

9. Some observations on general n. Recall that in studying conditions for a relation $\alpha_y(f,x_*) = 0$ to hold in $k < x_1,\ldots, x_n, y >$, we first obtained one condition for each string of indices $i_* \in \{1,\ldots,d\}^{n+1}$, but then made two simplifying observations: (i) The conditions associated to a string i_* really only depends on which pairs of terms of this string were equal, i.e. on the induced equivalence relation ρ on $\{0,\ldots,n\}$. (ii) If ρ has fewer than d equivalence classes, it can be refined to a relation ρ' with exactly d classes (assuming $n \geq d-1$), and the condition on f corresponding to this refined relation implies the condition obtained from the original relation. Hence we could enumerate our conditions in terms of relations with exactly d equivalence classes.

The condition (41) for a relation $\alpha_y(f,\bar{x}) = 0$ to hold in $k < x, y >_d$ admits corresponding simplifications. The analog of observation (i) above is that the condition on f corresponding to an admissible graph G depends only on the isomorphism class of G as an oriented graph with one distinguished edge. For isomorphic graphs differ only in the numbering of their vertices, and the corresponding sums (40) will then differ only by a relabelling of the indeterminates.

The analog of observation (ii) says that assuming $n \geq d-1$, all the conditions of (41) are implied by those arising from admissible graphs with exactly d vertices. We sketch the proof.

Let G be an admissible graph with fewer than d vertices. Let us call a vertex v of G simple if only one edge enters it, equivalently if only one edge leaves it. When this is so, "the next vertex after v" will mean the terminal vertex of the unique edge originating at v.

Since G has $n+1 \geq d$ edges, but fewer than d vertices, not all its vertices can be simple. Let us find a particular non-simple vertex as follows. We begin our search with the terminal vertex of the distinguished edge of G. If this is simple, we go onto the next vertex after it (as defined above); if that is simple we go on to the next, and so on. From the connectedness of G it is not hard to show that we will eventually hit a non-simple vertex. Let v be the first one that we strike.

Let the equivalence classes of non-distinguished edges that leave v be $E_1, \ldots, E_r$. (Here r may be 1, since several equivalent edges may leave v, or the distinguished edge and a nondistinguished one.) We shall now construct from G new graphs $G_1, \ldots, G_r$, each having one more vertex than G. Namely, we construct G_i by adding to G a new vertex v', diverting the edge along which we arrived at v so that it terminates at v' instead, and similarly diverting one member of the equivalence class E_i by moving its initial vertex to v'. (The first of the two edges mentioned may be the distinguished edge. In any case, the resulting graph G_i is understood to inherit from G a distinguished edge.) Note that there is a natural map from each G_i to G, collapsing v' and v.

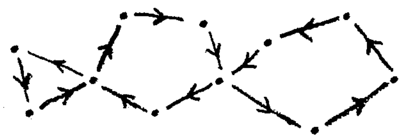
Now we claim that every traverse i_* of G can be lifted to a traverse of exactly one of $G_1, \ldots, G_r$. Indeed, given i_*, let m be the least value such that i_m is not a simple vertex of G. Then one sees from the way we constructed v that $i_m = v$. One can also verify that m must be $< n$. Hence G has a nondistinguished edge from $i_m = v$ to i_{m+1}. If E_i is the equivalence class of this edge, it is easy to see that the given traverse i_* lifts to a unique traverse of G_i (obtained by substituting v' for v in the mth place only), but not to any other $G_{i'}$.

Not all of $G_1, \ldots, G_r$ need be admissible, since some may not be connected. Say $G_1, \ldots, G_s$ are the admissible ones. Since any graph having a traverse is admissible, all of the liftings in question must be to graphs in this subfamily.

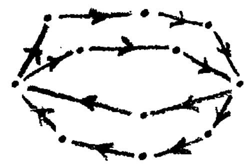
It is not hard to verify that if we write down the expressions (40) associated with these s admissible graphs, sum them, and then substitute y_{vv} for $y_{v'v'}$, the resulting expression is the expression (40) associated with G. Thus the condition on f associated with the graph G is implied by conditions associated with graphs having larger numbers of vertices. Combining with our earlier observation about isomorphic graphs, we get:

(45) For $n \geq d-1$, the family of conditions (41) for $\alpha_y(f, \bar{x})$ to be zero is equivalent to the subfamily defined by a set of representatives of the isomorphism classes of admissible graphs with exactly d vertices.

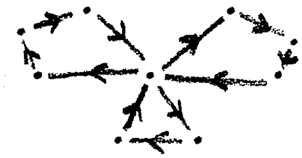
The case $n = d$ treated in preceding sections was so easy to work with because every admissible graph with exactly d vertices had a unique traverse. Let us now take a quick look at the case $n = d+1$. Here an admissible graph with exactly d vertices will have one of the three forms



(46)



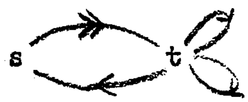
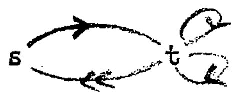
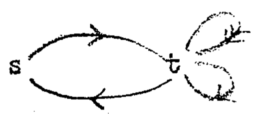
(47)



(48)

(Here we have not shown the location of the distinguished edge, which will not make a difference in most of what we say. As in (42), "degenerate" cases where one or more branches have length 1 can also occur.) One finds that a graph of the form (46) always has a unique traverse. Each graph of the form (47) or (48) has two traverses, unless it has two equivalent nondistinguished edges, i.e. unless two branches having the same initial and terminal vertices both have length 1, but neither is the distinguished edge. In this case there is again only one traverse.

Example: $d = 2$, $n = 3$. Up to isomorphism, there are six (2,3)-admissible graphs with $d=2$ vertices. We list these below followed by the corresponding conditions of (41) on $f(t_0,t_1,t_2,t_3) \in k[t_0,t_1,t_2,t_3]$. The first five graphs have unique traverses, hence the condition which each gives is equivalent to a condition $f \in I_\rho$ as in (28). We describe an equivalence relation ρ below by writing ρ: $S_1,\ldots, S_d$ where $S_1,\ldots, S_d$ are its equivalence classes. Rather than choosing a member of an isomorphism class of graphs on the vertex-set $\{1,2\}$ and labelling the corresponding indeterminates in our condition y_{11}, y_{22}, we write $\{s,t\}$ for both our set of vertices and our set of indeterminates.

graph	condition on f.
	$f(t,s,t,s) = 0$, i.e. $f \in I_\rho$ where ρ: $\{0,2\},\{1,3\}$.
	$f(t,t,s,s) = 0$, i.e. $f \in I_\rho$ where ρ: $\{0,1\},\{2,3\}$.
	$f(t,s,s,t) = 0$, i.e. $f \in I_\rho$ where ρ: $\{0,3\},\{1,2\}$.
	$f(t,t,t,s) = 0$, i.e. $f \in I_\rho$ where ρ: $\{0,1,2\},\{3\}$.
	$f(s,t,t,t) = 0$, i.e. $f \in I_\rho$ where ρ: $\{0\},\{1,2,3\}$.
	$f(t,t,s,t) + f(t,s,t,t) = 0$. (Weaker than the conjunction of the two conditions $f(t,t,s,t) = 0$, $f(t,s,t,t) = 0$, i.e. $f \in I_\rho \cap I_{\rho'}$, where ρ: $\{0,2,3\},\{1\}$ and ρ': $\{0,1,3\},\{2\}$.)

(49)

Note that the equation corresponding to the last graph does not define an <u>ideal</u> of $k[t_*]$. For instance, $f = t_1 - t_2$ satisfies it, but $t_1(t_1-t_2)$ does not. In general, for $n > d$ the set of f such that $\alpha_y(f, \bar{x}) = 0$ is not an ideal of $k[t_*]$. However, it is not hard to show that if we let $A \subseteq k[t_*]$ denote the subring of polynomials symmetric in the n-1 indeterminates $t_1, \ldots, t_{n-1}$, then this set is an A-submodule of $k[t_*]$. The point is that if i_* and i'_* are two traverses of the same admissible graph G, then the terms of $(i'_1, \ldots, i'_{n-1})$ are a permutation of $(i_1, \ldots, i_{n-1})$. So for $a \in A$, $a(t_{i_0}, \ldots, t_{i_n}) = a(t_{i'_0}, \ldots, t_{i'_n})$, so replacing f by af preserves the property that a sum (40) be zero.

10. <u>More general identities.</u> We noted at the end of section 6 that the results of that section dealt with elements of $k[y]\ x_1\ k[y]\ \ldots\ k[y]\ x_n\ k[y] \subseteq k < x_*, y >_d$, but that these were not the most general elements of $k < x_*, y >_d$ homogeneous of degree 1 in each x_m. The general element with that property may be written

$$(50) \qquad \sum_\pi \alpha_y(f_\pi, x_{\pi*})$$

where π ranges over $S(n)$, the permutation group on $\{1,\ldots,n\}$; where further

$$(51) \qquad x_{\pi*} \text{ denotes the string } x_{\pi(1)},\ldots, x_{\pi(n)},$$

and where the f_π $(\pi \in S(n))$ are $n!$ arbitrary elements of $k[t_*] = k[t_0,\ldots,t_n]$.

Still more generally, for any decomposition

$$(52) \qquad n = n_1 + \ldots + n_\nu \qquad (n_\mu \geq 0 \text{ for } \mu = 1,\ldots,\nu)$$

we may consider elements of $k < x_1,\ldots, x_\nu, y >_d$ homogeneous of degree n_μ in each x_μ. These may again be written in the form (50), but with π now ranging over maps $\{1,\ldots,n\} \to \{1,\ldots,\nu\}$ which take on each value $\mu \in \{1,\ldots,\nu\}$ exactly n_μ times.

For the remainder of this section, we shall assume given, in addition to d and n, integers $\nu \geq 1$ and n_μ $(1 \leq \mu \leq \nu)$ satisfying (52). We shall write n_* for $n_1,\ldots, n_\nu$, and $S(n_*)$ for the set of maps $\{1,\ldots,n\} \to \{1,\ldots,\nu\}$ which take on each value μ exactly n_μ times. Note that elements of $S(n_*)$ can be composed on the right with elements of the permutation group $S(n)$. The string $x_1,\ldots, x_\nu$ will be denoted x_*, and R will denote the ring $k < x_*, y >_d$. When we refer to (50), the index π will be understood to range over $S(n_*)$ (rather than $S(n)$ as stated there. These coincide when all $n_\mu = 1$, of course.)

The graph-theoretic methods developed in the last two sections for the study of elements of $k < x, y >_d$ homogeneous in x can be applied to elements of the above more general sort. Note that (50) has the expansion

(53) $$\sum_{\pi \in S(n_*), i_* \in \{1,\dots,d\}^{n+1}} f_\pi(y_{i_0 i_0},\dots,y_{i_n i_n})\, x_{\pi(1) i_0 i_1} \cdots x_{\pi(n) i_{n-1} i_n} e_{i_0 i_n}.$$

Now to each term of the form

(54) $$x_{\pi(1) i_0 i_1} \cdots x_{\pi(n) i_{n-1} i_n} e_{i_0 i_n} \quad (\pi \in S(n_*),\ i_* \in \{1,\dots,d\}^{n+1})$$

let us associate what we shall call an (n_*,d)-admissible graph. This will mean an (n,d)-admissible graph, given with a labeling of the non-distinguished edges by integers from the set $\{1,\dots,\nu\}$ such that the number of edges with label μ is exactly n_μ. Namely, given the term (50), we form a graph having vertex-set $\{i_0,\dots,i_n\}$, with an edge from i to j labeled μ for each factor $x_{\mu i j}$ of (54), and a distinguished edge from i to j as before if the final matrix-unit factor is e_{ji}. Then we see that two expressions (54) are equal if and only if their associated (n_*,d)-admissible graphs are the same.

We shall call two nondistinguished edges of the (n_*,d)-admissible graph G equivalent if they have the same initial vertex i, the same terminal vertex j, and the same label μ; i.e. if they correspond to the same commutative indeterminate $x_{\mu i j}$. If G is an (n_*, d)-admissible graph, we define a traverse of G to mean a string

(55) $$(i_*, \pi_*) = (i_0, \pi(1), i_1, \dots, i_{n-1}, \pi(n), i_n),$$

such that if G has exactly $r_{\mu i j}$ edges from i to j labeled μ, then there are exactly $r_{\mu i j}$ values of m such that $(i_{m-1}, \pi(m), i_m) = (i, \mu, j)$. It is easy to see that if $|G|$ is the (n,d)-admissible graph obtained by dropping edge-labels from the (n_*,d)-admissible graph G, then every traverse i_* of $|G|$

lifts to precisely $\prod_{ij} \left(\frac{(\sum r_{\mu ij})!}{r_{1ij}! \ldots r_{\nu ij}!} \right)$ traverses $(i_*, \pi*)$ of G. Note also that if $(i_*, \pi*)$ is a traverse of G, then $\pi \in S(n_*)$.

It is now easy to see that if G is the (n_*,d)-admissible graph associated with an element (54), and $T(G)$ the set of traverses of G, then the total coefficient of (54) in (53) is

(56) $$\sum_{(i_*, \pi *)\in T(G)} f_\pi(y_{i_0 i_0}, \ldots, y_{i_n i_n}).$$

Let G be an (n_*,d)-admissible graph, and $|G|$ the associated (n,d)-admissible graph. Note that if $|G|$ has no pairs of distinct equivalent edges, then any traverse of $|G|$ lifts to a unique traverse of G. Now, the considerations from which we deduced the necessity of our necessary and sufficient condition for $\alpha_y(f, x_*)$ and $\alpha_y(f, \bar{x})$ to be 0 when $n \leq d$, involved graphs without distinct equivalent edges. (See (42) for the case $n = d$.) Hence the same arguments (together with the arguments by which we previously characterized y-centralizing elements) give

(57) If $n < d$, then the following conditions on a system of elements $f_\pi \in k[t_*]$ $(\pi \in S(n_*))$ are equivalent: (a) $\sum \alpha_y(f_\pi, x_{\pi *})$ centralizes y, (b) $\sum \alpha_y(f_\pi, x_{\pi *}) = 0$, (c) All f_π are 0.

(58) If $n = d$, then $\sum \alpha_y(f_\pi, x_{\pi *}) = 0$ if and only if every f_π is divisible by $t_p - t_q$ for all values $0 \leq p < q \leq n$, while this element centralizes y if and only if every f_π is divisible by all these elements except $t_0 - t_d$.

The arguments by which we characterized central elements of $k < x, y >_d$ of degree d in x generalize to give

(59) If $n = d$, then every <u>central</u> element of R homogeneous of degree n in each x may be written uniquely $\sum \alpha_y(v\ g_{\pi}, x_{\pi *})$, where v is defined in (44), and the g_{π} $(\pi \in S(n_*))$ are elements of $k[t_0,\ldots,t_{d-1}]$ satisfying

$$g_{\pi\sigma}(t_0,\ldots,t_{d-1}) = g_{\pi}(t_1,\ldots,t_{d-1}, t_0)$$

where σ denotes the cyclic permutation $(1,\ldots,n) \in S(n)$.

In particular, we may deduce:

(60) In a generic matrix ring $k < x_1, x_2, \ldots; y >_d$, every central identity of total degree $\leq d$ in the variables other than y is a consequence of the single central identity:

(61) $\sum_{0 \leq m < d} \alpha_y(v; x_{\sigma m *})$ (σ as in (59)).

Note that even for $d = 2$, where <u>central</u> identities and y-<u>centralizing</u> elements first appear in the same degree (see end of section 8), one finds that for $n_0 = n_1 = 1$, the k-module of y-centralizing elements of $k < x_1, x_2, y >_2$ homogeneous of degree 1 in each of x_1, x_2 is 2-dimensional, spanned by $[x_1,y][x_2,y]$ and $[x_2,y][x_1,y]$, while the module of central elements is 1-dimensional, spanned by the sum of these two. Hence one sees that neither of the former two elements lies in $Z[y]$, and one can deduce that the inclusions of (6) are strict for $d = 2$, $n > 2$.

We return to general n and d. Let us call two (n_*,d)-admissible graphs G and G' <u>isomorphic</u> if they differ only by a relabeling of their <u>vertices</u>. Essentially the same arguments used to show (45) give

(62) For $n \geq d-1$, $\sum_{S(n_*)} \alpha_y(f_\pi, x_\pi) = 0$ if and only if the expression (56) is zero for a representative of each equivalence class of (n_*,d)-admissible graphs G with exactly d vertices.

Side remark: If n_* has the property that some $n_\mu = 1$, then from any (n_*,d)-admissible graph G, one can construct another (n_*,d)-admissible graph G' by interchanging the roles of "distinguished edge" and "unique edge labeled μ", and otherwise preserving the graph and its labeling. In fact, this corresponds to a map of the k-submodule of elements of R homogeneous of degree 1 in x_μ into itself, which acts by $A \times B \mapsto B \times A$. That such an operator on $k < x_1, x_2, \ldots >$ was well-defined was discovered by Yu. P. Razmyslov, and its properties exploited in his construction of central polynomials [10]. In [11] I call this operator "Razmyslov's transposition" and further develop its properties. It is also briefly developed in the last section of [1].

The two short remaining sections of this paper are appendices. In section 11 I outline the most convenient way I have found to actually write down the systems of equations given by (45) or (62), in the hope that some other workers may find this useful for practical computations. The validity of the method given rests on a graph-theoretic lemma on how all traverses of an admissible graph may be obtained starting from any one of them, which also seems of interest for its own sake, and which is proved in section 12.

11. Appendix. Some practical considerations. The graph-theoretic viewpoint of the above sections has been useful in making the families of strings of indices we had to work with visualizable, so that we could follow proofs more easily and avoid giving tedious arguments in excessive detail. However, in actually enumerating for particular d and n (or n_*) the equations given by (41), (45) or (62), saying when an element will be zero in our generic matrix rings, we are essentially brought back to strings of indices. For instance, to enumerate the conditions given by (41) on a function $f \in k[t_0, \ldots, t_n]$, we take d variable-symbols, say s, t, ..., z *, write down all ways of substituting these into the n+1 arguments of f, then collect together terms f(...) which agree in the choice of first and last variables, and in the number of transitions between each pair of variables, and set the sum of each of these collections to zero. (E.g. in (49), f(t,t,s,t) and f(t,s,t,t) are the only terms which begin with t, end with t, and have one (t,t), one (t,s) and one (s,t) transition. Hence we set $f(t,t,s,t) + f(t,s,t,t) = 0$.) The simplification (45) says that we get an equivalent system of conditions if we restrict attention to those collections of terms in which all d variable-symbols actually appear, and choose one representative from each equivalence class of such collections under relabeling of variables.

If we followed the above procedure as I have just described it, the most tedious part would be sorting the terms f(...) into families according to number of transitions of each sort. In fact, given one such term, there is a

*In place of the cumbersome $y_{11}, \ldots, y_{dd}$. I kept these above so that we would not forget, in our ring-theoretic considerations, that they were entries of the diagonal matrix y.

shortcut for getting all the others in the same family: We shall show in the next section that these can all be obtained by repeated interchanges of pairs of disjoint substrings A and B of variables, such that the first members of A and of B are the same, and their last members are also the same. For instance from (t,s,t,t) we may obtain (t,t,s,t) by interchanging the initial string (t,s,t) and the final string (t).

So to write down a system of conditions (41), we list in lexicographic order the expressions f(...) which include all d variables and which cannot be transformed, either by relabeling of variables or by interchanges of strings of the sort noted above, to lexicographically lower expressions. We then add to each of these terms all those that can be obtained from it by repeated "interchanges of strings", and set each sum to 0. It is in fact not hard to tell whether a string can be brought to a lexicographically lower one by one of these two sorts of operations, and I have found the above technique useful in calculating with small values of n and d. (Some conditions are still repeated, because it is not so easy to see when a term can be reduced to a lexicographically lower one by a combination of the two types of operation; e.g f(s,t,t,s,u,s) ↦ f(s,u,s,t,t,s) ↦ f(s,t,s,u,u,s). For small n and d these can be weeded out. Whether the above technique can be made into an algorithm which for large n and d is still better than "list and sort", I don't know.)

The statement that all terms in a given family can be obtained from any one by iterated interchanges of appropriate disjoint strings of variables is equivalent to saying that all traverses of an admissible graph G can be gotten from one by a similar kind of cutting and gluing. In the next section this result is stated and proved, using the language of Eulerian circuits, to make it readable independent of of the rest of this paper. (The reader can see from the proof a fact that I implicitly assumed above: that if a term can be carried by a sequence of such interchanges to a lexicographically lower one, then it can in fact be lowered lexicographically by the first of these operations.)

insert where shown: (F. Harary tells me that in his language this should be called an Eulerian diagraph G rooted at an arc d.)

12. Appendix on Eulerian circuits. Let G be a pseudosymmetric graph (cf. [13] or section 7 above) with a distinguished edge d. We shall define an Eulerian circuit of G to mean a sequence $E = (e_0, \ldots, e_n)$ of edges of G such that $e_0 = d$ (not in the standard definition!), each edge appears exactly once, the terminal vertex of e_{i-1} is the initial vertex of e_i $(1 \leq i \leq n)$, and the terminal vertex of e_n is the initial vertex of e_0.

Suppose that E is an Eulerian circuit of G which can be decomposed

$$E = X\,A\,Y\,B\,Z$$

where X, A, Y, B, Z are consecutive strings of edges such that A and B have the same initial vertices, and the same final vertices. Then clearly X B Y A Z will be another Eulerian circuit of G. We may allow some of our substrings to be vacuous. If A is vacuous references to "the initial vertex of A" must be understood as referring to the initial vertex of A Y B Z X, and the corresponding adjustment made for "the final vertex of A"; the analogous statement holds for B. In fact, we shall require

(63) X, Y and B nonempty.

The condition that X be nonempty is needed so that the transformation

(64) $$X\,A\,Y\,B\,Z \mapsto X\,B\,Y\,A\,Z$$

will preserve the property $e_0 = d$. From the conditions that Y and B are nonempty we see that (64) will be nontrivial. We claim

(65) One can get from any Eulerian circuit $E = (e_0, \ldots, e_n)$ of G to any other Eulerian circuit $E' = (e'_0, \ldots, e'_n)$ by a series of transformations (64) satisfying (63).

Indeed, let $E \neq E'$ be given and let $p > 0$ be the first index such that $e_p \neq e'_p$. Then e'_p can be written e_q for some $q > p$. Now let us note that

$$(66) \qquad \{e_p, \ldots, e_{q-1}\} \subseteq \{e_p, \ldots, e_n\} = \{e'_p, \ldots, e'_n\}.$$

The family on the left side of (66) is nonempty, but it does not contain the first-listed member of the right-most family, since that equals e_q. Hence we can find r with $p \leq r < n$ such that e'_r does not lie in the left-hand family of (66), but e'_{r+1} does. Say $e'_r = e_s$ and $e'_{r+1} = e_t$. Thus

$$(67) \qquad 0 < p \leq t < q \leq s \leq n.$$

We now partition E:

$$(\overbrace{e_0, \ldots, e_{p-1}}^{X}, \overbrace{e_p, \ldots, e_{t-1}}^{A}, \overbrace{e_t, \ldots, e_{q-1}}^{Y}, \overbrace{e_q, \ldots, e_s}^{B}, \overbrace{e_{s+1}, \ldots, e_n}^{Z})$$

We see from (67) that X, Y and B are nonempty. The first edges of A (or of AY if A is empty) and B, namely e_p and e_q, are respectively the first noncoinciding edges of the circuits E and E', and so they have the same initial vertex. The last edge of B and the first edge of Y are equal to e'_r and e'_{r+1} respectively; it follows that the final vertex of B and the final vertex of A will coincide. Now if we apply (64) using this decomposition of E, we get a circuit $X\,B\,Y\,A\,Z$ which agrees with E' from its beginning through at least the first edge after X, which is one step longer than E does. Hence by induction, a series of such operations (64) will eventually carry us to E'.

REFERENCES

1. P. M. Cohn, Algebra, volume II, J. Wiley and Sons, 1977.

2. N. Jacobson, PI-Algebras, An Introduction, Springer Lecture Notes in Mathematics, No.441, 1975.

3. C. Procesi, Rings with Polynomial Identities, Marcel Dekker, Inc., 1973.

4. H. W. E. Jung, Über ganze birationale Transformationen der Ebene, J. reine angew. Math. 184 (1942) 161-174.

5. A. Gutwirth, An inequality for certain pencils of plane curves, Proc. A. M. S. 12 (1961) 631-638.

6. M. Nagata, A theorem of Gutwirth, J. Math. Kyoto Univ. 11 (1971) 149-154.

7. W. van der Kulk, On polynomial rings in two variables, Nieuw Arch. Wisk. (3) 1 (1953) 33-41.

8. A. J. Czerniakiewicz, Automorphisms of a free associative algebra of rank 2. II. Trans. A. M. S. 171 (1972) 309-315.

9. C. Procesi, Non-commutative affine rings, Atti Acc. Naz. Lincei, Memorie, ser.VIII, vol.VIII (1967), sez. 1ª, pp.239-255.

10. Yu. P. Razmyslov, On a problem of Kaplansky, Izv.A.N.S.S.S.R., ser. mat., 37 No.3 (1973) 483-501 (Russian).

11. G. M. Bergman, A lemma of Razmyslov, and central polynomials, (unpublished notes.)

12. L. G. Makar-Limanov, On automorphisms of free algebras with two generators, Funkc. anal. i ego priloženiya 4(1970)107-108.

13. C. Berge, Graphs and Hypergraphs, (transl. Edward Minieka), North-Holland, 1973.

Department of Mathematics
University of California
Berkeley, CA 94720
U S A