ISRAEL JOURNAL OF MATHEMATICS 87 (1994), 161–179

ELEMENTS OF REDUCED TRACE 0

BY

S. A. AMITSUR

Department of Mathematics
The Hebrew University of Jerusalem, Jerusalem 91904, Israel

AND

LOUIS H. ROWEN

Department of Mathematics
Bar-Ilan University, Ramat-Gan 52900, Israel

ABSTRACT

Every element r of reduced trace 0 in a simple finite dimensional algebra R is a sum of at most 2 commutators. If R in *not* a division ring then r is a commutator, unless r is a scalar (in which case $\mathrm{char}(R) \neq 0$). The method of proof provides a generic division algebra of transcendence degree $n^2 - 1$.

Introduction

Throughout this paper R is a finite dimensional central simple algebra, with center F. Then $R \otimes_F \overline{F} \approx M_n(\overline{F})$ where $\overline{F}$ is the algebraic closure of F, and the **reduced trace** of an element r in R is defined as the trace of the matrix corresponding to $r \otimes 1$ in $M_n(\overline{F})$. Thus for any commutator $r = [a,b] = ab - ba$ we have $\mathrm{tr}(r) = 0$, and we adddress the converse question

QUESTION 1: *If* $\mathrm{tr}(r) = 0$ *then is* r *a commutator?*

This is obvious for $R = F$ (since then $\mathrm{tr}(r) = r$), and is also true for $R = M_n(F)$, cf. [6],[2]. Although the answer is unknown in general, there are various positive results, including for $n = 2, 3$ (Theorem 0.10). Also it turns out in general that r is a sum of at most two commutators, and we prove a slightly stronger fact. Actually there are two proofs, one of which involves Brauer factor

Received February 15, 1993

sets (§3) and yields considerable extra information about generic matrix algebras, which we include as an appendix at the end.

We cannot yet answer Question 1 for a division algebra D, and its general answer seems to rely on properties of quadratic forms. We do have an affirmative answer for $R = M_n(D)$ whenever $n > 1$, unless the matrix is scalar and n is prime to the characteristic of F. The proof is rather intricate, utilizing the other results of this paper, so we give a weakened result first (Theorem 1.10) and then the full result in section 2. The case where the matrix is scalar but n is prime to $p = \deg D = \mathrm{char}(F)$ is particularly intransigent, and is discussed separately.

The results of section 1 contain some facts concerning normal forms of matrices which might be of independent interest.

0. Some easy special cases

Remark 0.1: If $ac = ca$ then $[a, bc] = [a, b]c$ and $[a, cb] = c[a, b]$.

Remark 0.2: The difference of any element b and any conjugate of b is a commutator. Indeed

$$aba^{-1} - b = [a, b]a^{-1} = [a, ba^{-1}].$$

Conversely, writing $b = ca^{-1}$ we see $[a, b] = aca^{-1} - c$. Thus every commutator is a difference of conjugates, so Question 1 is equivalent to: if $\mathrm{tr}(r) = 0$ then is r a difference of two conjugates? Viewed in this way the question has a rather easy answer in several special cases.

PROPOSITION 0.3: *Suppose K/F is a cyclic field extension with $K \subset R$. Then any element r of K having trace 0 (with respect to the field extension) is a commutator in R.*

Proof: Write $r = \sigma(b) - b$ for suitable b in K, by the additive form of Hilbert's theorem 90. Then there is invertible a in R such that $\sigma(b) = aba^{-1}$, so

$$r = \sigma(b) - b = aba^{-1} - b = [a, ba^{-1}]$$

by Remark 0.2. ∎

Note: If $\alpha = \mathrm{tr}_{K/F}\, r$ then $\mathrm{tr}\, r = \frac{n}{[K:F]}\alpha$; thus if K is a maximal subfield or more generally if $\mathrm{char}(F) \nmid \frac{n}{[K:F]}$ then $\mathrm{tr}_{K/F}\, r = 0$ is equivalent to $\mathrm{tr}\, r = 0$.

On the other hand, we have

Remark 0.4: The map ∂_a: $R \to R$ given by $r \mapsto [a,r]$ is a $C_R(a) - C_R(a)$ bimodule map, where $C_R(a)$ is the centralizer of a. Furthermore, $\ker \partial_a = C_R(a)$, and im $\partial_a = [a, R]$. (The first assertion is by Remark 0.1, and the remainder is immediate.)

In characteristic $p > 0$ we are aided by inseparable elements (if they exist). In the next few results, D is a division ring with center F.

Remark 0.5: Suppose $\mathrm{char}(F) = p$ and $a \in D$ is purely inseparable. Then $a^{p^t} \in F$ for some t. Hence $\partial_a = [a, \ ]$ is a nilpotent derivation, by Leibniz' rule $((\partial_a)^{p^t}(d) = [a^{p^t}, d] = 0$ for all d in D). Thus $D \supset \partial_a D \supset \partial_a^2 D \supset \cdots \supset \partial_a^m D = 0$ for suitable m; by Remark 0.4 this is a chain of vector spaces over the division ring $C_D(a)$. But $\ker(\partial_a) = C_D(a)$ has dimension 1, so each space has $C_D(a)$-codimension 1 in the preceding space (since the chain cannot stabilize before 0), and we conclude $m = [D: C_D(a)] = \deg a$. Hence $\partial_a^u D$ has dimension $m-u$, and is clearly contained in $\ker \partial_a^{m-u}$, which also likewise has dimension $m-u$ (since $\ker \partial_a^u$ has dimension u), so we conclude

$$\partial_a^u D = \ker \partial_a^{m-u} \quad \text{for all } 0 \le u \le m.$$

LEMMA 0.6: *If $a \in D$ is purely inseparable over F and $\partial_a^u(d) = 0$ for $u < \deg a$ then $d \in [a, D]$.*

Proof: Let $m = \deg a$. By Remark 0.5, $d \in \partial_a^{m-u}(D) = [a, \partial_a^{m-u-1}(D)]$. ∎

THEOREM 0.7: *If $d \in D$ commutes with an element a which is <u>not</u> separable over F then $d \in [a, D]$. (Note this hypothesis requires characteristic $p > 0$.)*

Proof: Let F_1 be the maximal separable extension of F inside $F(a)$. Replacing D by $C_D(F_1)$, which contains both a and d and whose center is F_1, we may assume a is purely inseparable over F. Then $\partial_a(d) = [a, d] = 0$, so $d \in [a, D]$ by Lemma 0.6. ∎

COROLLARY 0.8:

(i) *If a is not separable then $a \in [a, D]$.*

(ii) *If D has an inseparable element then 1 is a commutator.*

COROLLARY 0.9: *If* deg $R = 2$ *then every element of R having trace 0 is a commutator.*

Proof: If $R = M_2(F)$ this is a theorem of Shoda and Albert–Muckenkoupt, to be discussed below. Thus assume R is a division ring. Then every subfield of R

is Galois or inseparable over F, so we are done by Proposition 0.3 and Theorem 0.7. ∎

The degree 3 case also is quite straightforward.

THEOREM 0.10: *If* deg $R = 3$ *and* $\operatorname{tr}(r) = 0$ *then* r *is a commutator.*

Proof: Again we may assume R is a division algebra. By [7] there is a in R with $a^3 \in F$ such that the minimal polynomial of r has the form

$$(\lambda - a^2ra^{-2})(\lambda - ara^{-1})(\lambda - r),$$

so $0 = \operatorname{tr}(r) = a^2ra^{-2} + ara^{-1} + r$. Hence

$$0 = ara^{-1} + r + a^{-1}ra = (ara^{-1} - r) + (a^{-1}ra - r) + 3r,$$

implying

$$-3r = [a, ra^{-1} - a^{-1}r] \; (= [a, [r, a^{-1}]]).$$

If $\operatorname{char}(F) \neq 3$ then r is a commutator, as desired.

It remains to consider the case that $\operatorname{char}(F) = 3$. Then $3r = 0$ implying $[r, a^{-1}]$ commutes with a, so is in $F(a)$, since every subfield of R properly containing F is maximal. Thus the inner derivation ∂ on R given by $[a^{-1},\ ]$ satisfies $\partial^2 r = 0$. Note that a is inseparable in this case, so r is a commutator by Lemma 0.6. ∎

Note: In the proof above, one has to be sure that a exists for each r in R. This is implicit in [7], but let us deal with this point in detail. If one puts $r' = uru^{-1}$ then, by [4, Remark 3.2.17],

$$r_2 = (r' - r)r'(r' - r)^{-1} = [u, r]r[u, r]^{-1}$$

(provided $[u, r] \neq 0$) and $a = [r, r_2]$. Thus one needs r_2 does not commute with r for suitable u in F. This is clear unless $r_2 \in F[r]$, but in this case $F[r]$ is Galois over F, and the existence of a is assured by the Skolem–Noether theorem.

Note: Actually this argument shows more generally for any algebraic element r of degree 3 and reduced trace 0, that r is a commutator.

D. Haile has pointed out that for algebras of degree p and characteristic p an affirmative answer to question 1 would imply cyclicity. This follows from Remark 0.2 and

PROPOSITION 0.11 (Haile): *If* $\deg(D) = p = \mathrm{char}(F)$ *then* D *is cyclic iff 1 is a commutator.*

Proof: ($\Rightarrow$) Take any maximal cyclic subfield. It contains 1, so apply Proposition 0.3.

($\Leftarrow$) By Remark 0.2, 1 is a difference of conjugates, i.e.,

$$1 = b - aba^{-1}$$

for suitable a, b in D; thus $aba^{-1} = b-1$, so $F(b)$ has the nontrivial automorphism $\sigma(b) = b - 1$ of order p, implying $F(b)/F$ is cyclic. ∎

An easy result probably due to P.M. Cohn is relevant (and tantalizing):

Remark 0.12 (Cohn): Suppose $d \in D$ is given, and $a, b \in D$ are given and <u>not</u> conjugate. Then there is some $c \in D$ such that $ac - cb = d$. (Indeed, the map $\varphi: D \to D$ given by $\varphi(c) = ac - cb$ is injective, for if $0 \neq c \in \ker\varphi$ then $ac = cb$ so $c^{-1}ac = b$, contrary to hypothesis. Thus $[\varphi(D): F] = [D: F]$ so $\varphi(D) = D$.)

1. Positive results for matrices

In this section we want to show that if question 1 is affirmative for D then it is also true for $M_n(D)$. This generalizes Shoda's theorem [6] (for $D = F$), but will be improved shortly, in §2. We shall use throughout the fact that any conjugate $u[a,b]u^{-1}$ of a commutator is a commutator ($[uau^{-1}, ubu^{-1}]$).

We start with some easy consequences of Remark 0.4.

PROPOSITION 1.1: *If* R *is an algebra over a field* F *then* $\dim C_R(a) + \dim[a, R] = \dim R$. *(Here "*dim*" denotes dimension as vector space over* F.*) In particular if* R *is central simple of degree* n *over* F, *and* $F(a)$ *is a maximal separable commutative subalgebra of* R, *then* $\dim[a, R] = n^2 - n$.

Proof: The first assertion follows from Remark 0.4. The second assertion is true since $\dim R = n^2$ and $\dim C_R(a) = n$ (since $C_R(a) = F(a)$). ∎

LEMMA 1.2: *Suppose* $a = \mathrm{diag}\{\alpha_1, \dots \alpha_n\}$ *is a diagonal matrix in* $R = M_n(F)$, *with* $\alpha_1, \dots, \alpha_n$ *distinct. Then* $[a, R]$ *is the set of matrices whose diagonal entries are all 0.*

Proof: Clearly $C_R(a) =$ {all diagonal matrices}, which has dimension n. By Proposition 1.1 we see $\dim[a, R] = n^2 - n$. But let

$$V = \{\text{matrices whose diagonal entries are all } 0\}.$$

Then $[a,R] \subseteq V$ since for $r = (r_{ij})$ the $i-i$ term of $[a,r]$ is $\alpha_i r_{ii} - \alpha_i r_{ii} = 0$. But also $\dim V = n^2 - n$ so $[a,R] = V$, as desired. ∎

Remark 1.3: Lemma 1.2 in conjunction with Remark 0.4 shows that $R \approx F(a) \oplus V$ as $F(a) - F(a)$ bimodules, where a is any diagonal matrix having distinct entries.

We are now ready to review Shoda's short proof in the characteristic 0 case. First he quickly shows for any matrix of trace 0 (in characteristic 0) there is a suitable similar matrix A for which all the diagonal entries are 0; we shall call this a **0-diagonal matrix**. But then for any diagonal matrix a with distinct diagonal entries we have seen that $[a, M_n(F)]$ consists of all 0-diagonal matrices, so $A \in [a, M_n(F)]$, as desired.

We want to carry out this argument in general.

LEMMA 1.4: *Suppose $R = M_n(D)$ with $\deg(D) = t$, and $a = \operatorname{diag}\{d_1, \dots, d_n\}$, with $d_1, \dots d_n$ nonconjugates each of degree t in D (over $F = Z(D)$). Then*

$$[a,R] = \begin{pmatrix} [d_1,D] & & & \\ & [d_2,D] & & * \\ & * & \ddots & \\ & & & [d_n,D] \end{pmatrix},$$

i.e., the off-diagonal entries are arbitrary, and the i-th diagonal entry is in $[d_i, D]$.

Proof: $\dim[d_i, D] = \dim D - \deg d_i = t^2 - t$. Thus the dimension of the right hand side is $(n^2-n)t^2 + n(t^2-t) = n^2t^2 - nt$. But $F(a)$ is separable in R of dimension nt so is maximal separable, so Proposition 1.1 implies $\dim[a,R] = (nt)^2 - nt$. Thus the left hand side and right hand side have the same dimension, so it remains to show that every element $[a,R]$ has the form of the right hand side. This is clear, for if $r = (d_{ij})$ then the $i-i$ entry of $[a,r]$ is $[d_i, d_{ii}] \in [d_i, D]$. ∎

Let us refine this a bit more.

LEMMA 1.5: *Let $R = M_{m+n}(D)$. Suppose $a \in M_m(D)$, $b \in M_n(D)$, and write $S = [a, M_m(D)]$, $T = [b, M_n(D)]$, and C_a, C_b for the respective centralizers of a, b in $M_m(D)$ and $M_n(D)$. If*

$$C_R \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} = \begin{pmatrix} C_a & 0 \\ 0 & C_b \end{pmatrix},$$

then

$$\left[\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, R \right] = \begin{pmatrix} S & * \\ * & T \end{pmatrix}$$

(where $*$ denotes arbitrary entries).

Proof: Let $t = \deg D$. The left hand side is contained in the right hand side, and the left hand side has dimension $((m+n)t)^2 - \dim C_R\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$. But

$$\begin{aligned}\dim\begin{pmatrix} S & * \\ * & T \end{pmatrix} &= \dim[a, M_m(D)] + \dim[b, M_n(D)] + 2mnt^2 \\ &= (mt)^2 - \dim C_a + (nt)^2 - \dim C_b + 2mnt^2 \\ &= ((m+n)t)^2 - (\dim C_a + \dim C_b) \\ &= ((m+n)t)^2 - \dim C_R\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix},\end{aligned}$$

implying equality holds. ∎

Let us verify the condition of lemma 1.5 for a dense set of matrices. First we need a result about matrices which must be well-known.

LEMMA 1.6: *If $a \in M_m(F)$ and $b \in M_n(F)$ have no common eigenvalue then $av = vb$ cannot hold for any $m \times n$ matrix $v \neq 0$.*

Proof: Passing to the algebraic closure, we may assume F is algebraically closed, so a is triangularizable, i.e. we may assume

$$a = \begin{pmatrix} a_{11} & & * \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix}$$

where a_{ii} are the eigenvalues of a. Writing $v = (v_{ij})$ and $b = (b_{ij})$, we see $a_{mm}v_{mj} = \sum_{i=1}^{n} v_{mi}b_{ij}$ for each j, so $(v_{m1}, \dots, v_{mn})$ is an eigenvector for b^t, with eigenvalue a_{mm}. By our hypothesis $(v_{m1}, \dots, v_{mn}) = 0$; we conclude by induction on m. ∎

Note: In case $a = \operatorname{diag}\{d_1, \dots d_m\}$ is diagonal, one can weaken the hypothesis to "each d_i has an eigenvalue which is not an eigenvalue for b." The proof is similar.

PROPOSITION 1.7: *Suppose $a \in M_m(D)$ and $b \in M_n(D)$ have no common eigenvalues. Then*

$$C_R\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} = \begin{pmatrix} C_a & 0 \\ 0 & C_b \end{pmatrix}.$$

Proof: Any matrix in $C_R\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ can be partitioned as $\begin{pmatrix} u & v \\ w & y \end{pmatrix}$ where $u \in C_a$, $y \in C_b$, and v is $m \times n$, w is $n \times m$, satisfying $av = vb$ and $wa = bw$. By symmetry it suffices to show $v = 0$; passing to $M_n(\overline{F})$ where $\overline{F}$ is the algebraic closure of F, we conclude by Lemma 1.6. ∎

PROPOSITION 1.8: *Any noncentral $r \in M_k(D)$ is similar to a matrix $\begin{pmatrix} A & * \\ * & d \end{pmatrix}$ where A is 0-diagonal of size $(k-1) \times (k-1)$ and $d \in D$.*

Proof: In three steps. (Some of the computations could be deleted by use of the rational canonical form, but we give a self-contained proof for the reader's convenience.)

(i) The case $k = 2$. Let

$$r = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Then

$$\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & -u \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a + uc & -(a+uc)u + (b+ud) \\ c & -cu + d \end{pmatrix}. \tag{1}$$

If $c \neq 0$ we choose $u = -ac^{-1}$, and obtained the desired form. If $b \neq 0$ then an analogous computation works. Finally, if $c = b = 0$ then taking u in (1) such that $-au + ud \neq 0$ reduces us to the case where $b \neq 0$; thus we are done unless $-au + ud = 0$ for all u in D; in particular $-a + d = 0$, i.e. $a = d$, and $a = d \in F$.

(ii) The case $k = 3$. Note r has three principal 2×2 submatrices; since r is not central, one of these submatrices is not central, and thus by (i), exchanging rows and columns if necessary, we may assume

$$r = \begin{pmatrix} 0 & a & b \\ c & u & v \\ d & w & z \end{pmatrix}.$$

In case

$$B = \begin{pmatrix} u & v \\ w & z \end{pmatrix}$$

is noncentral we apply (i) to B, to get r conjugate to a matrix of the form

$$r = \begin{pmatrix} 0 & a & b \\ c & 0 & v \\ d & w & z \end{pmatrix},$$

as desired. Thus we may assume B is central, i.e.

$$r = \begin{pmatrix} 0 & a & b \\ c & \alpha & 0 \\ d & 0 & \alpha \end{pmatrix},$$

with α central. First assume $c \neq 0$. Then

$$\begin{pmatrix} 1 & u & v \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & a & b \\ c & \alpha & 0 \\ d & 0 & \alpha \end{pmatrix} \begin{pmatrix} 1 & u & v \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1}$$
$$= \begin{pmatrix} uc + vd & * & * \\ c & -cu + \alpha & -cv \\ d & -du & -dv + \alpha \end{pmatrix}.$$

If $c \neq 0$ then take $u = -dc^{-1}$ and $v = 1$, and we conclude by (i), applied to the the bottom right-hand 2×2 submatrix. If $d \neq 0$ then argue analagously; similarly if $a \neq 0$ or $b \neq 0$. Thus we may assume $a = b = c = d = 0$, i.e.

$$r = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha \end{pmatrix};$$

conjugating now by $\begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ yields $\begin{pmatrix} 0 & * & * \\ * & 0 & * \\ * & * & 2\alpha \end{pmatrix}$, so we are done.

(iii) The case $k \geq 3$. We replace r by a conjugate $\begin{pmatrix} A & * \\ * & B \end{pmatrix}$, where A is 0-diagonal of maximal possible size $m \times m$ and B has size $n \times n$ (where $n = k - m$). Arguing as in (ii) one has $m \geq 1$. Then in fact $m = k - 1$ and $n = 1$, since otherwise we could choose the principal minor submatrix determined by the last row and column of A and the first two rows and columns of B, and apply (ii) to get a matrix conjugate to r, starting with a 0-diagonal matrix of size $m+1 \times m+1$, contrary to assumption. ∎

Proposition 1.9: *In characteristic $p \neq 0$, if $n \equiv 0 \pmod p$ then any scalar matrix $d \cdot I_n$ is a commutator.*

Proof: $$dI_n = \left[\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 2 & & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & n-1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & \dots & 0 & d \\ d & 0 & \dots & 0 & 0 \\ 0 & d & \dots & 0 & 0 \\ \vdots & & & \vdots & \vdots \\ 0 & 0 & & d & 0 \end{pmatrix} \right].$$ ∎

THEOREM 1.10: *Suppose D is a division ring in which every element of (reduced) trace 0 is a commutator. Then the same holds for $R = M_k(D)$ for any k.*

Proof: We may assume $F = Z(D)$ is infinite (since otherwise $D = F$ so we are reduced to [2] and [6]).

Take any $r \in R = M_k(D)$ of reduced trace 0. We want to show r is a commutator. First assume r is noncentral. By Proposition 1.8 we can put r into the form $\begin{pmatrix} A & * \\ * & d \end{pmatrix}$ where A is a 0-diagonal matrix and $d \in D$. Then tr $d = 0$ so by assumption $d = [v, w]$ for suitable v, w in D. Taking any diagonal $m \times m$ matrix A' whose eigenvalues are distinct from those of v (possible since F is infinite), we see $C_R \begin{pmatrix} A' & 0 \\ 0 & v \end{pmatrix} = \begin{pmatrix} C_{A'} & 0 \\ 0 & C_v \end{pmatrix}$ (in the notation of Proposition 1.7), so

$$\begin{pmatrix} A & * \\ * & d \end{pmatrix} \in \left[\begin{pmatrix} A' & 0 \\ 0 & vI \end{pmatrix}, R \right]$$

by Lemmas 1.4 and 1.5, i.e. r is a commutator.

Thus we are done unless r is central and thus $\mathrm{char}(F)|k$, but then r is a commutator by Proposition 1.9. ∎

Note: Although we reduced to the case F is infinite, we only need $|F| \geq k$ (so that there are "enough" elements of F to miss the eigenvalues of v). Then the proof of this theorem also yields [6] and [2] (for $|F| > k$).

2. Arbitrary central simple algebras

In this section we want to obtain a result for a single commutator for an arbitrary central simple algebra R of degree k.

Note 2.0: As Saltman has observed, a Zariski topology argument (passing to $M_n(\overline{F})$) also shows that the set of commutators is Zariski dense in the set of elements of trace 0.

Unfortunately we have not obtained a single commutator result for an arbitrary element r in a division ring, but can show r is a sum of *two* commutators, with one of the elements in the commutators pre-determined. We need to reduce to the case of matrices; although the Zariski topology would suffice, we shall employ the following method of passing to matrices: Recall the Capelli polynomial (in noncommuting indeterminates)

$$C_{2t-1}\{X_1, \ldots, X_{2t-1}\} = \sum_{\pi \in S_t} (\mathrm{sg}\ \pi) X_{\pi 1} X_{t+1} X_{\pi 2} X_{t+2} \cdots X_{2t-1} X_{\pi t},$$

([4, p.12]), and given a subspace V of R write $C_{2t-1}(V)$ for $\{C_{2t-1}(v_1, \ldots, v_t, x_1, \ldots, x_{t-1})\colon\ v_i \in V,\ x_i \in R\}$. Since C_{2t-1} is t-normal we see $C_{2t-1}(V) = 0$ whenever $\dim V < t$. On the other hand, if $R = M_n(F)$ and $\dim V = t$ then $C_{2t-1}(V) \neq 0$, as seen either as a consequence of [4, proposition 1.4.7] or by modifying its proof. Thus we have

LEMMA 2.1: *Suppose R is central simple of degree n, and*

$$V \subseteq \{\textit{elements of reduced trace 0}\}.$$

Then $C_{2n^2-3}(V) = 0$ unless $V = \{$elements of reduced trace 0$\}$.

Proof: Pass to $R \otimes_F \overline{F} \approx M_n(\overline{F})$ where $\overline{F}$ is a splitting field, and its subspace $V \otimes_F \overline{F}$, which has dimension $\leq n^2 - 1$ over $\overline{F}$. ∎

THEOREM 2.2: *Suppose R is central simple of degree n, and $a \in R$ has n distinct eigenvalues. Then there exists $c \in R$ such that for any $r \in R$ of reduced trace 0 there are b, d in R such that $r = [a, b] + [c, d]$, i.e. $[a, R] + [c, R] = \{$elements of trace 0$\}$.*

Proof: If $F = Z(R)$ is finite then $R \approx M_n(F)$, so we may assume F is infinite. We shall show there is c in R for which $[a, R] + [c, R] = V$, where $V = \{$elements of reduced trace 0$\}$. Indeed, otherwise, $[a, R] + [c, R] < V$, for any c in R, so letting

$$f_Y = C_{2n^2-3}([a, X_{11}] + [Y, X_{12}], \ldots, [a, X_{n^2-1,1}] + [Y, X_{n^2-1,2}], X_{n^2}, \ldots, X_{2n^2-3})$$

for noncommuting indeterminates X_{i1}, X_{i2} $(1 \leq i \leq n^2 - 1)$, X_j $(n^2 \leq j \leq 2n^2 - 3)$, and Y, we see $f_y(V) = 0$ (for any prior substitution y of Y in R). But f_y is linear in the X_{ij}, so $f_y(V \otimes \overline{F}) = 0$ in $R \otimes \overline{F} \approx M_n(\overline{F})$ where $\overline{F}$ is the algebraic closure of F for all y in R.

Since F is infinite, a standard argument shows $f_y(V \otimes \overline{F}) = 0$ for all y in $R \otimes \overline{F}$. (Namely, take a base $b_1, \ldots, b_k$ a base of V; then for $y = \sum \alpha_i b_i$ the fact $f_y(V) = 0$ translates into a set of polynomial conditions in $\alpha_1, \ldots, \alpha_k$, which are satisfied for all α_i in F and thus for all α_i in $\overline{F}$ since F is infinite.) Hence $[a, M_n(\overline{F})] + [c, M_n(\overline{F})]$ has dimension $< n^2 - 1$ for all c in $M_n(\overline{F})$.

On the other hand we could take a base of $\overline{F}^{(n)}$ for which a is diagonal, and with respect to this base, $[a, M_n(\overline{F})]$ is the set of 0-diagonal matrices. Thus, to arrive at a contradiction it suffices to find a matrix c such that $[c, M_n(\overline{F})]$ contains the diagonal matrices of trace 0. But

$$c = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \end{pmatrix}$$

satisfies that property, since $[c, \sum \alpha_i e_{i+1,i}] = \mathrm{diag}\{\alpha_1 - \alpha_2, \alpha_2 - \alpha_3, \dots, \alpha_n - \alpha_1\}$. ∎

We shall now verify question 1 when $R = M_k(D)$ for $k > 1$. Our main result along these lines is

THEOREM 2.4: *Suppose $R = M_k(D)$ is central simple, with $k \geq 2$. Then any noncentral element of R having reduced trace 0 is a commutator.*

Proof: By induction on k. Let $F = Z(D)$.

First we assume $k \geq 3$, and assume that Theorem 2.4 is known to hold for $M_m(D)$ whenever $1 < m < k$. Suppose $r \in M_k(D)$ with $\mathrm{tr}(r) = 0$. By Proposition 1.8, r is conjugate to $\begin{pmatrix} 0 & * \\ * & B \end{pmatrix}$, where B is a noncentral $(k-1) \times (k-1)$ matrix having reduced trace 0. By induction one may write $B = [P, Q]$ in $R' = M_{k-1}(D)$. Taking any α in F distinct from the eigenvalues of P, we see by Proposition 1.7 that

$$C_R \begin{pmatrix} \alpha & 0 \\ 0 & P \end{pmatrix} = \begin{pmatrix} * & 0 \\ 0 & C_{R'}(P) \end{pmatrix}$$

since α commutes with all elements of D. Hence Lemma 1.5 shows

$$\left[\begin{pmatrix} \alpha & 0 \\ 0 & P \end{pmatrix}, R \right] = \begin{pmatrix} 0 & * \\ * & [P, R'] \end{pmatrix}$$

which contains $\begin{pmatrix} 0 & * \\ * & B \end{pmatrix}$ since $B \in [P, R']$. This proves r is conjugate to a commutator, so is itself a commutator, thereby concluding the reduction.

It remains to verify the 2×2 case. We shall show that any matrix of the form $\begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix}$ in $M_2(D)$, having reduced trace 0 (i.e. $\mathrm{tr}(d_{11} + d_{22}) = 0$), but with d_{21} having distinct eigenvalues and $\mathrm{tr}(d_{21}) \neq 0$, can be written as

$$\left[\begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix}, \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \right]$$

for suitable a_{ij}, b_{ij} in D. This will prove the theorem in general since

$$\begin{pmatrix} 1 & 0 \\ 0 & y \end{pmatrix}\begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & y \end{pmatrix}^{-1} = \begin{pmatrix} d_{11} & d_{12} \\ yd_{21} & yd_{22} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & y^{-1} \end{pmatrix} = \begin{pmatrix} d_{11} & d_{12}y^{-1} \\ yd_{21} & yd_{22}y^{-1} \end{pmatrix}$$

so choosing y suitably we can arrange for the $2-1$ position of the conjugate to have distinct eigenvalues, with $\operatorname{tr}(d_{21}) \neq 0$.

The proof is divided into two steps:

Claim 1: *There is some d'_{12} in D and suitable a_{ij}, b_{ij} in D, with a_{11} not conjugate to a_{22} in D, such that*

$$\left[\begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix}, \begin{pmatrix} b_{11} & 0 \\ 1 & b_{22} \end{pmatrix}\right] = \begin{pmatrix} d_{11} & d'_{12} \\ d_{21} & d_{22} \end{pmatrix}. \tag{2}$$

Claim 2: *Given a_{ij} as in Claim 1 and taking d''_{12} aritrarily in D, there is b_{12} in D such that*

$$\left[\begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix}, \begin{pmatrix} 0 & b_{12} \\ 0 & 0 \end{pmatrix}\right] = \begin{pmatrix} 0 & d''_{12} \\ 0 & 0 \end{pmatrix}. \tag{3}$$

Given the two claims, take $d''_{12} = d_{12} - d'_{12}$; then

$$\left[\begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix}, \begin{pmatrix} b_{11} & b_{12} \\ 1 & b_{22} \end{pmatrix}\right] = \begin{pmatrix} d_{11} & d'_{12} \\ d_{21} & d_{22} \end{pmatrix} + \begin{pmatrix} 0 & d_{12} - d'_{12} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix}.$$

Proof of Claim 2:

$$\left[\begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix}, \begin{pmatrix} 0 & b_{12} \\ 0 & 0 \end{pmatrix}\right] = \begin{pmatrix} 0 & a_{11}b_{12} - b_{12}a_{22} \\ 0 & 0 \end{pmatrix}$$

so we need to find b_{12} in D such that $a_{11}b_{12} - b_{12}a_{22} = d''_{12}$. Since a_{11} and a_{22} are presumed nonconjugate, b_{12} can be obtained via Remark 0.12.

Proof of Claim 1: The 1-2 position is irrelevant, so matching the 1-1, 2-1, and 2-2 positions we must solve the following three equations:

$$[a_{11}, b_{11}] + a_{12} = d_{11}, \tag{4}$$

$$a_{22} - a_{11} = d_{21}, \tag{5}$$

$$[a_{22}, b_{22}] - a_{12} = d_{22}. \tag{6}$$

(Then a_{22} and a_{11} will be nonconjugate since d_{21}, their difference, has reduced trace $\neq 0$.)

Solving (4) and (5) give respectively

$$a_{12} = d_{11} - [a_{11}, b_{11}], \tag{4'}$$

$$a_{22} = d_{21} + a_{11}, \tag{5'}$$

so we could use these two equations to define a_{12} and a_{22}; plugging into (6) leaves us with the following equation:

$$[d_{21} + a_{11}, b_{22}] - (d_{11} - [a_{11}, b_{11}]) = d_{22},$$

or

$$[d_{21}, b_{22}] + [a_{11}, b_{22} + b_{11}] = d_{11} + d_{22}. \tag{7}$$

By Theorem 2.2 there are u, v, w in D such that

$$[d_{21}, u] + [v, w] = d_{11} + d_{22}.$$

Take $b_{22} = u$, $a_{11} = v$, and $b_{11} = w - u$. ∎

Digression: Actually we proved the stronger result that any noncentral 2×2 matrix of reduced trace 0 is similar to a commutator $[A, B]$, where

$$A = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix}.$$

Interestingly enough, this might fail for scalar matrices, since it turns out to be equivalent to the cyclicity of the underlying division algebra (which in general is a difficult open question). Indeed, suppose $\deg D = p \neq 2$ is prime and

$$\left[\begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix}, \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}\right] = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{8}$$

Matching the 1-1, 2-1, and 2-2 matrix components yields the equations

$$[a_{11}, b_{11}] + a_{12}b_{21} = 1; \tag{9}$$

$$a_{22}b_{21} - b_{21}a_{11} = 0; \tag{10}$$

$$[a_{22}b_{22}] - b_{21}a_{12} = 1. \tag{11}$$

Then (9) yields $a_{12} = (1 - [a_{11}, b_{11}])b_{21}^{-1}$; plugging into (11) yields

$$[a_{22}, b_{22}] + b_{21}[a_{11}, b_{11}]b_{21}^{-1} = 2.$$

Thus

$$\begin{aligned} 2 &= [b_{21}^{-1}a_{22}b_{21}, b_{21}^{-1}b_{22}b_{21}] + [a_{11}, b_{11}] \\ &= [a_{11}, b_{21}^{-1}b_{22}b_{21}] + [a_{11}, b_{11}] = [a_{11}, b_{21}^{-1}b_{22}b_{21} + b_{11}]. \end{aligned}$$

Thus D is cyclic, by Proposition 0.11, as desired.

To try to solve

$$[A, B] = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

computationally without any prior restrictions, one may assume that A is in rational canonical form, i.e.

$$A = \begin{pmatrix} 0 & 1 \\ a & b \end{pmatrix}.$$

Then we generalize the equation to

$$\left[\begin{pmatrix} 0 & 1 \\ a & b \end{pmatrix}, B\right] = \begin{pmatrix} 1 & 0 \\ u & v \end{pmatrix}.$$

It is easy to see that B must have the form $\begin{pmatrix} d - cb & c \\ 1 + ca & d \end{pmatrix}$, so then

$$\begin{aligned} u &= [a, d] - acb + +b + bca, \\ v &= [a, c] + [b, d] - 1. \end{aligned}$$

The question remains as to whether a, b, c, d can be chosen such that $u = 0$ and $v = 1$. This seems to be a difficult computational question.

There always will be a solution in the following situation: Suppose $\deg(D) = p = \operatorname{char}(F)$, and there is a field K of dimension 2 over F, such that $R_0 = D \otimes_F K$ is cyclic. Then by Proposition 0.11, 1 is a commutator in R_0, and thus in $M_2(D)$. Although this seems like a special case, it is actually quite general, as is seen in the digression after Corollary 3.3′.

3. The use of Brauer factor sets

One can obtain some of the results of §2 more explicitly, using Brauer factor sets. Recall from [3] that if $K = F(a)$ is a maximal separable subfield of a central simple F-algebra R of degree n and E is the normal closure of K, then viewing $G = \mathrm{Gal}(E/F)$ as a transitive subgroup of S_n we can write R as

$$\{(r_{ij}) \in M_n(E)\colon\ r_{\sigma i,\sigma j} = r_{ij} \text{ for all } i,j \text{ and all } \sigma \text{ in } G\};$$

multiplication in R is given in terms of a *Brauer factor set* $\{c_{ijk}\colon\ 1 \le i,j,k \le n\}$ by the formula

$$(r_{ij})(r'_{ij}) = (r''_{ij})$$

where $r''_{ik} = \sum_{j=1}^{n} c_{ijk} r_{ij} r'_{jk}$. The Brauer factor set $\{c_{ijk}\} \subset E - \{0\}$ satisfies the two conditions

$$c_{\sigma i,\sigma j,\sigma k} = c_{ijk} \quad \text{for all } \sigma \text{ in } G \tag{12}$$

$$c_{ijk}c_{ikm} = c_{ijm}c_{jkm}. \tag{13}$$

In particular taking $k = m$ in (13) one sees $c_{ikk} = c_{jkk}$ for all i,j,k; likewise $c_{iij} = c_{iik}$ for all i,j,k. One can in fact normalize and assume $c_{ijj} = c_{iij} = 1$ for all i,j. Furthermore if n is odd then one may assume $c_{iji} = 1$ for all i,j (cf. [5]).

Remark 3.1: $K = \{$elements of R which can be written as diagonal matrices$\}$.

Remark 3.2: If $r = (r_{ij}) \in R$ then $a = \mathrm{diag}\{r_{11},\dots,r_{nn}\} \in R$, so $r - a \in R$. Thus any r in R is written uniquely as the sum of an element of K and an element all of whose diagonal entries are 0.

PROPOSITION 3.3: *Viewed in this notation, $[a,R]$ consists precisely of those elements whose corresponding matrices are 0-diagonal.*

Proof: Let

$$V = \{\text{elements of } R \text{ corresponding to matrices having 0 on the diagonal}\}.$$

Remark 3.2 says $R \approx K \oplus V$ as $K - K$ bimodules in analogy to remark 1.3. But then $\dim V = n^2 - n = \dim[a,R]$. (The centralizer of a is $K = F(a)$.) Furthermore writing $a = \mathrm{diag}\{a_1,\dots,a_n\}$ and $r = (r_{ij})$ we see that the $i - i$ entry of the matrix corresponding to $[a,r]$ is $a_i c_{iii} r_{ii} - r_{ii} c_{iii} a_i = 0$. ∎

COROLLARY 3.3′: *In the above notation* $R = F[a] \oplus [a, R]$.

Digression: In general, the same sort of argument shows that $F[a] \cap [a, R] = 0$, whenever a is a separable element of R such that $F[a]$ is a maximal commutative subalgebra of R (so that a can be written diagonally in terms of $F[a]$). In particular, suppose F has characteristic $p > 0$; note that if $1 = [a, b]$ then a and b must be inseparable. But this means $F[a^p] \subset F[a]$, and thus the centralizer R' of a^p is a cyclic algebra, by a theorem of Albert [1]. For example, if $R = M_2(D)$ with $\deg D = p$ then we see that R contains a cyclic algebra of degree $p = \mathrm{char}(F)$, and the question as to whether the identity 2×2 matrix is a commutator can be translated to the question as to whether R contains such a cyclic subalgebra. This is another famous question about cyclicity, and also is likely to be very difficult.

THEOREM 3.4: *Suppose R is central simple over F, of odd degree n, $r \in R$ has reduced trace 0, and $a \in R$ is arbitrary such that $F(a)$ is a maximal separable subalgebra of R. (For example if $\deg R$ is prime and $\mathrm{char}(F) = 0$ then a could be any element of $R - F$.) Then one can find b, c, d in R for which*

$$r = [a, b] + [c, d].$$

Proof: We represent R as matrices in terms of a normalized Brauer factor set $\{c_{ijk}\}$ and $K = F(a)$, i.e. $c_{iji} = 1$ for all i, j, cf. [5, theorem 4]. The same set of matrices with the trivial Brauer set (i.e. taking 1 for each c_{ijk}) defines a matrix algebra R' isomorphic to $M_n(F)$, cf. [3],[5].) By the theorem of Shoda and Albert–Muckenkoupt, we have matrices u, v such that $r = [u, v]$ inside R'. But writing $u = (u_{ij})$ and $v = (v_{ij})$, we see the $i - i$ diagonal term of $[u, v]$ *taken in* R is $\sum_{j=1}^n u_{ij}c_{iji}v_{ji} - \sum_{j=1}^n v_{ij}c_{iji}u_{ji} = \sum_{j=1}^n (u_{ij}v_{ji} - v_{ij}u_{ji})$, the diagonal term of the matrix of r. Thus the matrix $[u, v] - r$ has diagonal 0, so by proposition 3.3 $[u, v] - r \in [a, R]$. We conclude $r = [a, b] + [u, v]$ for suitable b in R. ∎

Note: Although this result is contained in Theorem 2.7, its proof seems to be more amenable to improvement. Here is another application.

Appendix to 3: Application to generic matrix algebras

Let $UD(n, F)$ be the generic F-division algebra of degree n, which is well-known to be the algebra of central quotients of the algebra $F\{X_1, X_2\}$ of generic matrices. One may assume X_2 is diagonal (since any matrix with distinct eigenvalues is diagonalizable over the algebraic closure of F), so writing $X_2 = \mathrm{diag}(a_1, \ldots, a_n)$ one could take $K = F(X_2)$ and write $UD(n, F)$ in terms of Brauer factor sets. Of course replacing X_2 by $X_2 - \mathrm{tr}\ X_2$ we could take $a_n = (a_1 + \cdots + a_{n-1})$. The point to be made here is that X_1 could then be replaced by an element whose corresponding matrix has zeroes on the diagonal (in view of Remark 3.2), so $UD(n, F)$ contains of the algebra R_0 generated by the matrices

$$X_1 = (\xi_{ij}) \quad \text{where } \xi_{ii} = 0 \quad \text{for } 1 \le i \le n,$$
$$X_2 = \mathrm{diag}(a_1, \ldots a_n) \quad \text{where } a_n = -(a_1 + \cdots + a_{n-1});$$

here the ξ_{ij} $(i \neq j)$ and $a_1, \ldots, a_{n-1}$ are commuting indeterminates over F. Note the transcendence degree is $(n^2 - n) + (n - 1) = n^2 - 1$, which is less than the usual result of $n^2 + 1$ (see [4, corollary 1.10.29]).

4. Generic examples

It remains to see whether Question 1 holds in general, or even for cyclic algebras of degrees > 3. In this section we shall construct the generic "solution" to Question 1, for cyclic algebras. (However, we still cannot answer the question for cyclic algebras.) For convenience we work in characteristic 0.

Let F_0 be a field containing a primitive n-th root ρ of 1, and let F be a purely transcendental field extension over F_0 in commuting indeterminates μ, ν, γ_{ij}, $0 \le i, j \le n - 1$, with $(i, j) \neq (0, 0)$; let R be the symbol $(\mu, \nu)_n$, i.e. R is generated as an algebra by elements x, y such that $x^n = \mu$, $y^n = \nu$, and $xy = \rho yx$. The generic element of trace 0 is then $\sum_{0 \le i,j \le n-1} \gamma_{ij} x^i y^j$ where $\gamma_{00} = 0$. So we want to solve

$$\left[\sum_{0 \le i,j \le n-1} \alpha_{ij} x^i y^j, \sum_{0 \le i,j \le n-1} \beta_{ij} x^i y^j \right] = \sum \gamma_{ij} x^i y^j$$

for suitable α_{ij}, β_{ij} in F; we may assume $\alpha_{00} = \beta_{00} = 0$. Suppose we have a solution for certain β_{ij}. To recover the α_{ij} we could solve a system of $n^2 - 1$ linear equations obtained by matching the coefficients of $x^i y^j$ for $0 \le i, j \le n - 1$ $((i, j) \neq (0, 0))$. Note that

$$[\alpha_{st} x^s y^t, \beta_{uv} x^u y^v] = \alpha_{st} \beta_{uv} (\rho^{tu} - \rho^{sv}) x^{s+u} y^{t+v},$$

so given i, j for any fixed u, v there is one possible s, t such that $s + u \equiv i \pmod{n}$ and $t + v \equiv j \pmod{n}$.

If Question 1 could be answered positively for this algebra then it would be answered positively for all cyclic algebras. However, the calculations for this example are still quite intricate, and seem to become involved with the arithmetic of the field.

References

[1] A. A. Albert, *Structure of Algebras*, AMS Colloq. Pub. **24**, Providence, RI, 1961.

[2] A. A. Albert and B. Muckenkoupt, *On matrices of trace zero*, Mich. Math. J. **4** (1957), 1–3.

[3] N. Jacobson, *Brauer factor sets, Noether factor sets, and crossed products*, in *Emmy Noether, A Tribute to her Life and Work* (J. Brewer and M. Smith, eds.), M. Dekker, New York, 1981.

[4] L. H. Rowen, *Polynomial Identities in Ring Theory*, Pure and Applied Mathematics **83**, Academic Press, Boston, 198.

[5] L. H. Rowen, *Brauer factor sets*, Trans. Amer. Math. Soc. **282** (1984), 765–772.

[6] K. Shoda, *Einige Sätze über Matrizen*, Jap. J. Math. **13** (1936), 361–365.

[7] H.M.S. Wedderburn, *On division algebras*, Trans. Amer. Math. Soc. **22** (1921), 129–135.