THE MERKURYEV-SUSLIN THEOREM

David M. Lewis

This article reports on one of the most important, and to many people, astonishing results in algebra so far this decade. In 1981, a Russian mathematician Merkuryev, virtually unknown in the west, proved a theorem concerning the algebraic K-theory and the Brauer group of a field. This result is now known as Merkuryev's theorem and not long afterwards Merkuryev, together with Suslin, a famous Russian mathematician, generalized the result to what is commonly called the Merkuryev-Suslin theorem. These theorems at once provide answers to some very hard problems in the theory of simple algebras, in the theory of quadratic forms and in algebraic geometry. Thus it seems worthwhile to try and explain, in as elementary a way as possible, what the Merkuryev-Suslin theorem is all about. A good source of background information for this article is [5].

We start with that well-known Dublin product, the real quaternions, discovered in 1843 by Hamilton and usually denoted H. A quaternion is an expression of the form a+bi+cj+dk where a,b,c,d ∈ ℜ, the real numbers, and quaternions can be added in the obvious way and multiplied together using the famous equations i^2=j^2=k^2=ijk=-1. Hamilton's construction may be generalized to give quaternion algebras over any field F. We simply choose non-zero elements a,b in F, (a≠b is allowed), and do exactly as in H except that we require 1^2=a, j^2=-1. For F=Ré, a+b=-1, we have H of course. A quaternion algebra defined as above is usually denoted (a+b-R) as it depends on the choice of a,b and on the base field F. It is always four-dimensional as an F-vector space and it turns out always to be either a skewfield as H is (i.e. a field except that multiplication lacks commutativity) or else is isomorphic to the ring of all 2x2 matrices with entries in F. (In fact it fails to be a skewfield precisely when there exist x,y in F such that ax^2+by^2 = 1.) For F=R, H is the only skewfield
that occurs as a quaternion algebra but for other fields things can be quite different. For example, if $F = \mathbb{Q}$, the rationals, there exist indefinitely many non-isomorphic quaternion algebras which are skewfields.

Any quaternion algebra has a natural involution - on it induced by $\bar{1} = -1, \bar{j} = -j$. (By an involution on an algebra $A$ we mean a map $A \to A, x \to \bar{x}$ such that $x + y = \bar{x} + \bar{y}, x\bar{y} = \bar{y}\bar{x}$ and $\bar{x} = x$, i.e., an anti-automorphism of period two.) On $\mathbb{H}$ this involution is the usual conjugation operation. This kind of involution is called an involution of the first kind because elements of $F$ are fixed by it. We view $F$ as being contained in $(\mathbb{A}, \mathbb{D})$ in the same way as $\mathbb{H}$ lies inside $\mathbb{H}$. An involution of the second kind is one which is non-trivial on $F$.

We must now say a few words about tensor products of algebras. An $F$-algebra is a ring which also is an $F$-vector space, the ring and vector space operations being compatible. Given two $F$-algebras $A_1, A_2$, there exists a unique $F$-algebra $T$ and a map $i : A_1 \times A_2 \to T$ with the following property:

Given any bilinear map $f : A_1 \times A_2 \to W$ into any $F$-vector space $W$ there exists a unique algebra homomorphism $g : W \to T$ such that $gf = i$. $T$ is called the tensor product and is denoted $A_1 \otimes_F A_2$.

For example if $A_1$ and $A_2$ are each quaternion algebras then there are three possibilities for $A_1 \otimes_F A_2$. Firstly $A_1 \otimes_F A_2$ may be a division algebra (i.e., an $F$-algebra which is a skewfield). Secondly $A_1 \otimes_F A_2$ may be the ring of all $2 \times 2$ matrices with entries in a quaternion division algebra and thirdly, $A_1 \otimes_F A_2$ could be the ring of all $4 \times 4$ matrices with entries in $F$. Generally the dimension of $A_1 \otimes_F A_2$ over $F$ is the product of the dimensions of $A_1$ and $A_2$.

Quaternion algebras are special cases of central simple algebras. A central simple algebra $A$ over $F$ is a finite dimensional $F$-algebra whose centre is $F$, i.e., $\{x \in A : xy = yx \text{ for all } y \in A\} = F$, and which has no proper two sided ideals when viewed as a ring. For short, we will write c.s. algebra from now on. The tensor product of two c.s. algebra over $F$ is a c.s. algebra. A celebrated theorem of Wedderburn says that any c.s. algebra over $F$ is isomorphic to $M_n D$, the ring of $n \times n$ matrices with entries in a skewfield $D$. Moreover $n$ is unique and $D$ is unique up to isomorphism. $D$ is a division algebra over $F$. We say that two c.s. algebras are similar if their skewfield parts from Wedderburn's theorem are isomorphic. Similarity is an equivalence relation on the set of c.s. algebras over $F$. In 1929, Brauer discovered that the set of similarity classes of c.s. algebras over $F$ has a group structure, tensor product being the group operation. The class of $F$ itself is the identity element of the group and the inverse of $A$ is the opposite algebra, denoted $A^{op}$, which is identical with $A$ as a set but with multiplication reversed, i.e., $A^{op}A$ as a set with multiplication defined by $ab = ba, ba$ being the usual multiplication in $A$. Then $A \otimes_F A^{op}$ is isomorphic to the ring of all $F$-homomorphisms from $A$ to $A$ and this ring is isomorphic to a full matrix ring $M_n F$, $n = \text{dimension of } A$ over $F$, and thus $A \otimes_F A^{op}$ is similar to $F$. This group is usually called the Brauer group of the field $F$ and is denoted $B(F)$. For $F$ finite $B(F)$ is trivial since finite skew-fields are commutative (by another theorem of Wedderburn). $B(\mathbb{R})$ is cyclic of order 2, $\mathbb{H}$ being the generator. For a local field $B(F) = \mathbb{Q}/\mathbb{Z}$, the rationals modulo one and for an algebraic number field, i.e., a finite algebraic extension of $\mathbb{Q}$, $B(F)$ is extremely large, its calculation being the culmination of work involving Brauer, Hasse, Noether and Albert. See [1] for details.

It should be mentioned that in general, quaternion division algebras are by no means the only kind of division algebras appearing as c.s. algebras. For example a division algebra may be a cyclic algebra defined as follows:

Let $L$ be a cyclic extension of $F$, i.e., a Galois extension field of $F$ such that the group of all automorphisms of $L$ that
fix elements of $F$ is a cyclic group of order $n$. Let $a$ be a
generator of this group. Choose some element $b \in F$. Introduce a symbol $u$ such that $u^n = b$. A typical element of the
cyclic algebra determined by $L$ and $b$ is an $L$-linear combination
$$
\sum_{i=0}^{n-1} x_i u^i,
$$
each $x_i \in L$, with addition defined in the natural way and
multiplication by $u^0 = b$ and $ux = a(x)u$ for all $x \in L$. The
resulting algebra is c.s. and if $b$ is suitably chosen it can be
a division algebra for certain kinds of field $F$. (Note
that for $F = \mathbb{R}$ if we choose $L = C$, $b = -1$ we obtain $\mathbb{H}$, $\mathfrak{c}$ on $C$ then being complex conjugation.) All division algebras
over $\mathbb{Q}$ are cyclic. However there exist fields with central
division algebras that are not cyclic algebras. See [1],
also [2].

For a positive integer $n$ we write $B_n(F) = \{ x \in B(F) : x^n = 1 \}$.
Merkuryev's theorem implies that, for any field $F$ of char $\neq 2$, the
subgroup $B_2(F)$ of $F$ generated by the quaternion algebras
and the Merkuryev-Suslin theorem implies that, provided $F$
contains a primitive $n^{th}$ root of unity, $B_n(F)$ is generated
by cyclic algebras. $B_2(F)$ in fact consists exactly of those
classes of algebras which admit an involution of the first kind.
(An involution of the first kind gives an isomorphism $A = A^{\text{op}}$
and hence $[A]$ has order two in $B(F)$, and conversely $[A]$ has
order two means there exists an isomorphism $A = A^{\text{op}}$ which
yields an involution of the first kind on $A$.) The degree
of a c.s. algebra is defined to be the square root of the
$-\text{dimension}$ of the skewfield part of $A$. A theorem in [1]
shows that the order of $[A]$ in $B(F)$ divides the degree of $A$ and also
that order and degree have the same prime factors apart from
multiplicity. It follows that c.s. algebras admitting an
involution of the first kind must have degree a power of two.

Tensor products of quaternion algebras give elements of
$B_2 F$ and fundamental conjectures studied by some algebraists
were those as to whether an algebra with involution of the
first kind is isomorphic to or else is similar to a tensor
product of quaternion algebras. In 1970, Amitsur, Rowen and
Tignol [3] produced an example of a division algebra over $Q(t)$,
a transcendental extension of $Q$, which has an involution of
the first kind but is not isomorphic to a tensor product of
quaternion algebras. Merkuryev's theorem however gives an
affirmative answer to the above conjecture for similarity.
So any division algebra $D$ with an involution of the first kind
must be such that, for some $n, M_n(D)$ is isomorphic to a tensor
product of quaternion algebras. For the example of [3] $n = 2$
will do, but in general it is not known what the least value
of $n$ be.

So far we have only given part of Merkuryev's theorem.
To describe it fully we must first define the group $K_2 F$
occuring in algebraic K-theory. $K_2 F$ is defined as the additive
abelian group generated by all symbols $(a, b)$, $a, b$ non-zero
elements of $F$, with relations

$$
(ab, c) = (a, c) + (b, c), \quad (a, bc) = (a, b) + (a, c)
$$
and

$$
(a, 1-a) = 0 \text{ for all } a, b, c \in F.
$$

Group theorists may be more familiar with an equivalent def-
inition of $K_2 F$ as the Schur multiplier of the group $E(F)$
generated by the elementary matrices in $F$, [11]. An elementary
matrix is one which coincides with the identity matrix except
for a single off-diagonal entry. Assume char $F \neq 2$. There
is a natural map $K_2 F \to B(F)$ sending $(a, b)$ to the quaternion
algebra $(a/b) F$. This map is easily seen to be trivial on the
subgroup $2 K_2 F = \{ 2x : x \in K_2 F \}$ and Merkuryev's theorem says
that the induced map $a : \frac{K_2 F}{2K_2 F} \to B(F)$ is injective and its image
is precisely $B_2 F$. The surjectivity of $a$ proves the conjecture
stated above.
The injectivity of $\alpha$ also answers a long-standing question in quadratic form theory dating back to work of Pfister [12] in 1966. We describe this briefly. The set of isometry classes of non-singular quadratic forms over a field $F$ can be given a ring structure, the addition (resp. multiplication) being induced by the direct sum (resp. tensor product) of the underlying vector spaces. The quotient, on factoring out by the so-called hyperbolic forms, is known as the Witt ring $\operatorname{W}(F)$ of $F$. See [6] for details. Let $I$ denote the ideal of forms defined on even dimensional spaces. Then powers of this ideal exist, i.e. $I^2$, $I^3$, etc. and clearly $I^{n+1} \subseteq I^n$ for all $n$. The significant connection between algebraic $K$-theory and quadratic forms was shown by Milnor [10] in 1970 when he proved that $I^2/I^3$ is isomorphic to $K_2 F$. There is a natural map $\beta: I^2 \to B(F)$ given by taking the class in $B(F)$ of the Clifford algebra of a quadratic form representing an element of $I^2$. (For anything in $I^2$ the Clifford algebra class can be shown to be trivial.) The map $\beta$ corresponds, under the Milnor isomorphism, to the map mentioned above which Merkurjev showed to be injective. Pfister [12] in 1966 had studied $\beta$ and had shown that in some cases it was injective but since then nobody had come near to a proof in general until Merkurjev's breakthrough. Thus Merkurjev solved a problem which had been regarded by quadratic form theorists as extremely difficult.

We finish by describing the Merkurjev-Suslin theorem. Let $W_n$ be the group of all $n^{th}$ roots of unity. We assume for simplicity that $W_n$ lies in $F$. There exists a unique homomorphism, for each $n$, $K^n F \to B_n(F) W_n$ induced by sending $(a, b)$ to $Aw$ where $w$ is a chosen primitive $n^{th}$ root of unity and $A$ is an algebra, called a norm residue algebra, defined as follows: $A$ is generated by elements $u$ and $v$ with the properties $u^n = a$, $v^n = b$ and $uv = vu$. The tensor product is of abelian groups, defined in a similar fashion to our earlier case, and tensoring on by $W_n$ is necessary in order to obtain a homomorphism which is independent of the choice of $w$. The name 'norm residue algebra' occurs because $A$ will be isomorphic to $M_n F$ precisely when $b$ is a norm from $F(\sqrt[n]{a})$. A will indeed always be similar to a cyclic algebra. Merkurjev and Suslin proved that the above map is in fact an isomorphism. The surjectivity implies that, provided $F$ contains $W_n$, each element of $B_n F$ is represented by a tensor product of cyclic algebras, a result that was somewhat surprising to some algebraists.

Another consequence of the Merkurjev-Suslin theorem is in the realm of algebraic geometry where it leads to new finiteness results about the Chow groups of a rational surface.

We have so far not mentioned the proof of these theorems and to do so would be beyond the scope of this article. The original announcements and proofs are in [7], [8], [9]. The proof requires the Galois cohomological interpretation of $B(F)$. It uses some difficult techniques from Quillen's version of algebraic $K$-theory and from algebraic geometry, in particular an analysis of the Severi-Brauer varieties corresponding to division algebras [13]. There is also now a more elementary proof of the general Merkurjev-Suslin theorem which has been presented in some notes by Merkurjev [7a]. This proof requires much less higher algebraic $K$-theory. In fact, Merkurjev's theorem (i.e. for $n = 2$) can now be more easily done in a couple of ways. Merkurjev himself found an easier proof using Milnor $K$-theory instead of Quillen $K$-theory. (We should explain that algebraic $K$-theory for a field $F$ defines groups $K_n F$ for all non-negative integers $n$, Milnor and Quillen $K$-theory are the same for $K_2$ although it is non-trivial to prove this fact. However, for higher $n$ the two $K$-theories are not always the same.) The Milnor $K_2 F$ is much easier to handle and this simpler proof has been very well written up by Wadsworth [14]. Also the quadratic form version of Merkurjev's theorem has been proved by Aitken [4] avoiding $K$-theory altogether but using some technical results from Galois cohomology.
References


