Journal of the Mathematical Society of Japan Vol. 5, No. 2, July, 1953.

Quadratic forms.

By Irving KAPLANSKY

(Received April 9, 1953)

1. Introduction. The theory of quadratic forms over a p-adic number field is a vital building block of the Minkowski-Hasse theory of quadratic forms over an algebraic number field, and has been expounded from several different points of view. A recent account of the purely number-theoretic attack on the problem appears in [2]. A field with a valuation, subjected to appropriate axioms, is the framework in [1]. A development from the point of view of the theory of algebras is given in [3].

In this last reference it appears that all the major portions of the theory can be deduced just from the fact that a quadratic form in five variables must represent zero. The main point of the present paper is that this in turn can be deduced from the following assumptions on a field F: that F is not formally real, and that its multiplicative group mod squares has exactly order four. It is plausible to make the following more general conjecture: if there are n classes mod squares, then every quadratic form in $n+1$ variables represents zero. We are able to prove this for the next case ($n=8$), and in other cases as well, for instance characteristic p. Beyond these partial results, it seems to be worth while to give a systematic formulation of the problems involved.

Notation: we shall use the symbol $(a_1, \cdots, a_n)$ for the quadratic form $\sum a_i x_i^2$. Equivalence of quadratic forms (or congruence of the corresponding matrices) will be indicated by the notation

$$(a_1, \cdots, a_n) \sim (b_1, \cdots, b_n).$$

2. Three invariants. Throughout the paper F will denote a field of characteristic different from two, and it will be assumed that F is not formally real (that is, -1 is a sum of squares). The formally real case probably has a parallel theory, which may be worth separate study. Of course, quadratic forms of characteristic two are markedly different, and have to be treated specially.

We proceed to define three invariants of a field F, indicating the dependence on F when desirable.

(a) Let F^* denote the multiplicative group of non-zero elements, and write $A(F)$ for the order of the group $F^*/(F^*)^2$. Of course $A(F)$ may be infinite; if it is finite it is evidently a power of 2, for in the group $F^*/(F^*)^2$ the square of every element is the identity.

(b) $B(F)$ is the smallest integer n such that -1 is a sum of n squares. We note that $B(F)\leq 2$ if F has characteristic p, for this is already true in the prime subfield of F.

(c) $C(F)$ is the smallest n such that every quadratic form in $n+1$ variables over F is a null form (that is, the form vanishes for some choice of the variables, not all zero).

We can prove at once the following restriction on the possible values of $B(F)$.

THEOREM 1. *Let F be a field which is not formally real. Then $B(F)=1, 2, 4$ or a multiple of* 8.

PROOF. Write $-1=a_1^2+\cdots+a_n^2$ with n minimal. Suppose first that n is odd, $n>1$. We have

$$-(1+a_1^2)^2=(a_2^2+a_3^2)(1+a_1^2)+(a_4^2+a_5^2)(1+a_1^2)+\cdots. \tag{1}$$

Now each product on the right side of (1) may be condensed to a sum of two squares; indeed we have the identity

$$(a^2+b^2)(c^2+d^2)=(ac-bd)^2+(ad+bc)^2. \tag{2}$$

Also the left side of (1) is non-zero, and we may divide by it. The result is an expression of -1 as a sum of $n-1$ squares, a contradiction.

In just the same way we show that if n exceeds 2 or 4, it is a multiple of 4 or 8 respectively; we use the identities analogous to (2) which compose sums of 4 or 8 squares (obtainable from the quaternions and Cayley numbers). Thus we complete the proof of Theorem 1. We remark that sums of 16 squares cannot be composed, so that this method of proof cannot be pushed any further. Nevertheless it seems unlikely that B for example could be 24. In fact we conjecture that

B is always a power of 2, and we make the same conjecture for C, although here the evidence is more meager.

THEOREM 2. *$C(F)$ cannot be* 3.

PROOF. The assertion $C(F)\geqq 3$ means that there exists a non-null form in 3 variables, which we can assume to be $(1, a, b)$. To prove the theorem we have to exhibit a non-null form in 4 variables, and our choice is $(1, a, b, ab)$. For if

$$x^2+ay^2+bz^2+abt^2=0\,, \tag{3}$$

we repeat the artifice used in Theorem 1: multiply (3) by z^2+at^2 and use the identity

$$(x^2+ay^2)\,(z^2+at^2)=(xz-ayt)^2+a(yz+xt)^2\,.$$

The result is a non-trivial representation of 0 by $(1, a, b)$.

We note finally that fields with valuations provide examples showing that A and C can be any powers of 2. The proof is straightforward and is left to the reader.

THEOREM 3. *Let F be a field complete with respect to a discrete rank one valuation, and suppose that the residue class field K has characteristic different from* 2. *Then $A(F)=2A(K)$ and $C(F)=2C(K)$.*

3. Main results. The principal conjecture of this paper is that $C\leqq A$. In attacking this problem we find that a critical role is played by the value of B. So we begin by noting $B\leqq A$. Indeed write $-1=a_1^2+\cdots+a_n^2$ with n minimal, and set $b_i=a_1^2+\cdots+a_i^2$. Then two different b's must be in different classes mod squares; otherwise the representation of -1 could be shortened. Thus $A\leqq n=B$. Actually we can slightly refine this result.

THEOREM 4. *If $A>2$, then $B<A$.*

PROOF. Suppose on the contrary that $A=B=n$, and use the notation above. Then $b_1, \cdots, b_n$ give us precisely a set of representatives of the non-zero elements mod squares. In particular, for each element $-b_i$ there is a unique index j such that $-b_j/b_i$ is a square c^2. We must have $i+j\geqq n+1$, for otherwise the equation $c^2b_i+b_j=0$ would give us a representation of -1 as a sum of $n-1$ or fewer squares. Since i and j both range precisely over the integers from 1 to n, we must actually have $i+j=n+1$ always. If we write $n=2m$ we get in particular

$$b_m + a^2_{m+1} = b_{m+1} = -c^2 b_m ,$$

$$(1+c^2)^2 b_m + a^2_{m+1} (1+c^2) = 0 .$$

That is, we have a representation of 0 as a sum of $m+2$ squares, and hence a representation of -1 as a sum of $m+1$ squares. This is a contradiction if $m+1 < n$, i. e. $2 < n$.

We turn our attention to inequalities for C, beginning with $C \leq AB$. This is easy to see: let there be given a quadratic form $\sum a_i x_i^2$ in $AB+1$ variables. Since there are A classes mod squares, there must be $B+1$ of the a's which are multiplicatively congruent mod squares. Since -1 is a sum of B squares, it follows that the form represents 0. The argument can be subjected to two successive refinements.

THEOREM 5. (1) *$C \leq AB/2$ if $B \geq 2$.* (2) *$C \leq AB/4$ if $B \geq 4$.*

PROOF. We shall give the proof of (2), and indicate the modifications needed in the (easier) proof of (1).

If $B \geq 4$ we have that -1 is not a sum of 2 squares. Consequently there must exist an element c which is a sum of two squares, but not itself a square. The quadratic form (1, 1) represents c, and it follows from this that (1, 1) is equivalent to (c, c). Let us write G for the multiplicative group of F mod squares; G is a group of order A. The elements ± 1, $\pm c$ map onto a subgroup H of G having order 4. Now let there be given a quadratic form $f = \sum a_i x_i^2$ in $1+AB/4$ variables. If we map the elements a_i into G/H, a group of order $A/4$, it must be the case that at least $1+B$ of the a's map into the same element. After multiplying f by a constant (this does not affect the question as to whether f represents 0), we can assume that $1+B$ of the a's are actually in H. If both 1 and -1, or both c and $-c$ occur, then f trivially represents 0. By another multiplication, if necessary, we can suppose that the $1+B$ elements consist of a certain number of 1's and a remaining group consisting entirely of c's or else entirely of $-c$'s. Now by Theorem 1, B is even and so $1+B$ is odd. Of the two numbers adding up to $1+B$, it must therefore be the case that one is even and the other odd. By still another normalizing mutliplication we arrange that the number of 1's is even. Since $(1, 1) \sim (c, c)$, we can switch all the 1's to c's. This finally gives us $1+B$ elements, all c or $-c$. If there is actually a mixture, f again trivially represents 0. There remains the case of $1+B$ c's, where f again represents 0 since -1 is a sum of B squares.

The proof of (1) is similar but simpler; we use the subgroup H of order 2 consisting of 1 and -1. A form in $1+AB/2$ variables will then have $1+B$ elements in the same coset mod H, and they can be normalized to be all 1's and -1's.

Let us assemble our results concerning the inequality $C \leqq A$. If $B=1$ we deduce $C \leqq A$ from the fact (noted above) that $C \leqq AB$ always holds. If $B=2$ or 4, Theorem 5 shows that $C \leqq A$. So we have achieved success for $B \leqq 4$. Since we always have $B \leqq A$, the case $A \leqq 4$ is accounted for. But we can also look after $A=8$, for then $B<8$ by Theorem 4, and $B \leqq 4$ by Theorem 1. We summarize in the following theorem.

THEOREM 6. *Let F be a field which is not of characteristic two, and is not formally real. Suppose that the multiplicative group of non-zero elements of F, mod squares, is precisely of order A. Then we can assert that every quadratic form in $1+A$ variables over F represents 0 at least in the following two cases:* (1) $A \leqq 8$, (2) -1 *is a sum of four or fewer squares in F. In particular, the conclusion is valid if F has characteristic p, for -1 is then a sum of two squares in F.*

4. Quaternion algebras.

We shall devote this final section to surveying the connection between quaternion algebras and quadratic forms. Let F be a field of characteristic different from 2. By the quaternion algebra $Q(a, b)$ over F we mean the four-dimensional algebra with basis 1, x, y, xy satisfying $x^2=a, y^2=b, xy=-yx$. It is known that $Q(a, b)$ is central simple, and so is either a division algebra or a two by two total matrix algebra. The connection with quadratic forms is summarized in the following three known lemmas (where $\cong$ denotes isomorphism).

LEMMA 1. *$Q(a, b) \cong Q(c, d)$ if and only if $(a, b, -ab) \sim (c, d, -cd)$.*

LEMMA 2. *$(a, b, c) \sim (p, q, r)$ if and only if the determinants of the forms agree up to a square (that is, abc/pqr is a square), and $Q(-ab, -ac) \cong Q(-pq, -pr)$.*

LEMMA 3. *$Q(a, b)$ is a total matrix algebra if and only if $(a, b, -ab)$ represents* 0.

Next we note the following two statements, of which the first is evident, while the second is a consequence of Lemma 3.

(1) F has no quadratic extensions if and only if $C(F)=1$ (that is, every quadratic form in two variables over F represents 0).

(2) F admits no quaternion division algebras if and only if $C(F) \leqq 2$ (that is, every quadratic form in three variables over F represents 0).

It is natural to conjecture that there is a third result to add to this list, stating that $C(F) \leqq 4$ is equivalent to a certain assertion about algebras. We have not discovered such a result, but we do call attention to the following statement:

($*$) F admits exactly one quaternion division algebra (in the sense of isomorphism).

We summarize the facts concerning ($*$).

(a) If F is not formally real and $A(F)=4$, there are just two possibilities. Either $C(F) \leqq 2$, a case of no further interest, or else $C(F)=4$ and ($*$) holds. The p-adic numbers, for instance, fall into the latter category.

(b) $C(F)=4$ does not imply ($*$), as is witnessed by the function fields in one variable over a finite constant field.

(c) However ($*$) does imply $C(F)=4$, as we shall now prove.

THEOREM 7. *Let F be a field which is not of characteristic two, and is not formally real. Suppose that F admits exactly one quaternion division algebra, up to isomorphism. Then every quadratic form in five variables over F represents* 0.

PROOF. We shall break the proof into a number of steps.

I. If $(1, a, b)$ and $(1, c, d)$ both fail to represent 0, then $(a, b, ab) \sim (c, d, cd)$. For if $(1, a, b)$ does not represent 0, neither does $(-a, -b, -ab)$, as we see by multiplying by $-ab$. By Lemma 3, $Q(-a, -b)$ is a division algebra; and the same holds for $Q(-c, -d)$. By hypothesis, they are isomorphic. Then by Lemma 1, $(-a, -b, -ad) \sim (-c, -d, -cd)$. It remains to multiply by -1.

II. If $Q(-1, -1)$ is not a division algebra, -1 is a sum of two squares. This follows at once from Lemma 3.

III. If $Q(-1, -1)$ is a division algebra, then for any element a, either a or $-a$ is a sum of 2 squares. For suppose that $-a$ is not a sum of 2 squares; then $(1, 1, a)$ does not represemt 0. Also $(1, 1, 1)$ does not represent 0, as follows from Lemma 3 and the assumption that $Q(-1, -1)$ is a division algebra. By I, $(1, a, a) \sim (1, 1, 1)$. Hence by Witt's cancellation theorem [3, Satz 4], $(a, a) \sim (1, 1)$. That is, a is a sum of 2 squares.

IV. -1 is a sum of 4 squares. Since F is not formally real, -1 is the sum of a certain number of squares, say $-1 = a_1^2 + \cdots + a_n^2$ with n minimal. If $n > 4$, we can apply III to the element $a_1^2 + a_2^2 + a_3^2$. It is not a sum of 2 squares, and so its negative must be a sum of 2 squares. This gives us a sum of 5 squares equal to 0, whence -1 is a sum of 4 squares.

V. Any element x is a sum of 4 squares. (a) If -1 is a sum of 2 squares, $-1 = a^2 + b^2$, we write

$$4x = (1 + x)^2 + (a^2 + b^2)\,(1 - x)^2,$$

and 3 squares suffice. (b) Otherwise we can apply III. If x is not a sum of 2 squares then $-x$ is, and $(-x, -x) \sim (1, 1)$, whence

$$(1, 1, 1, 1, -x) \sim (-x, -x, -x, -x, -x) = -x\,(1, 1, 1, 1, 1).$$

This last form represents 0 since-1 is a sum of 4 squares. Hence x is a sum of 4 squares.

VI. If $(1, p, q, r, s)$ does not represent 0, then $(p, q) \sim (1, pq)$. To prove this we first show that $(1, -p, -q)$ must represent 0. For if not, we apply I to it and $(1, r, s)$, getting

$$(-p, -q, pq) \sim (r, s, rs).$$

Multiply by -1 and enlarge by r and s:

$$(p, q, r, s, -pq) \sim (r, s, -r, -s, rs). \tag{4}$$

Now $(r, -r) \sim (pq, -pq)$ by [3, Satz 5]. Apply this in (4) and then cancel $-pq$ by Witt's cancellation theorem:

$$(p, q, r, s) \sim (pq, s, -s, rs).$$

But this tells us that (p, q, r, s) represents 0, a contradiction. Hence it must be the case that $(1, -p, -q)$ represents 0, and so does its negative. From [3, Satz 5] we get

$$(-1, p, q) \sim (-1, 1, pq).$$

It remains to cancel -1.

We are ready to complete the proof of Theorem 7. Suppose that $(1, p, q, r, s)$ does not represent 0. By repeated applications of VI we get

$$(1, p, q, r, s) \sim (1, 1, pq, r, s) \sim \cdots \sim (1, 1, 1, 1, pqrs).$$

By V, $-pqrs$ is a sum of 4 squares, and this says that the form represents 0.

University of Chicago

Bibliography

[1] W. H. Durfee, *Quadratic forms over fields with a valuation*, Bull. Amer. Math. Soc., 54 (1948), 338-351.

[2] B. W. Jones, *The Arithmetic Theory of Quadratic Forms*, Carus Monograph No. 10, Math. Assoc. of America, 1950.

[3] E Witt, *Theorie der quadratischen Formen in beliebigen Körpern*, J. Reine Angew. Math., 176 (1937), 31-44.