

On the axiomatic foundations of the theory of Hermitian forms

By C. T. C. WALL

University of Liverpool

(Received 30 October 1968)

In recent work on some topological problems (7), I was forced to adopt a complicated definition of 'Hermitian form' which differed from any in the literature. A recent paper by Tits (5) on quadratic forms over division rings contains a new and simple definition of these. A major objective of this paper is to formulate both these definitions in somewhat more general terms, and to show that they are equivalent.

We also discuss corresponding notions of reflexive sesquilinear forms, which also arose in topological work ((8), section 12); it is no longer true (as it is over division rings) that such forms are equivalent to hermitian or skew-symmetric ones.

It is not claimed that these topics are treated below with the maximum possible generality; however, we do work with arbitrary rings (with unit), so any further generalization is likely to involve additional elements of structure (e.g. a grading or a group of operators) or a higher degree of abstraction (e.g. working over schemes instead of rings). We preface each definition by a discussion, which is intended to show some of the reasons for adopting the definition.

Sesquilinear forms. Let A be a ring (with unit), M a (unital) right A -module. We will discuss bilinear maps $\phi : M \times M \rightarrow A$ satisfying some axioms related to the module structure: these can as well be discussed for a pair M, N of right A -modules. We think of bilinear maps of $M \otimes N$. Now the tensor product inherits any right module structure possessed by N and any left module structure possessed by M . Since A is naturally an $(A-A)$ bimodule, it is natural to require that M be a left A -module, N a right A -module, and that ϕ induce a map $M \otimes N \rightarrow A$ of $(A-A)$ bimodules. But we are only given a right A -module structure on M ; this induces a left module structure over the opposite ring A^{op} . Thus to make M a left A -module, we require an isomorphism $\alpha : A \rightarrow A^{\text{op}}$; which is also to be interpreted as an anti-automorphism of A .

We have thus arrived at a definition. Let A be a ring with anti-automorphism α ; let M, N be right A -modules. Then a map $\phi : M \times N \rightarrow A$ is α -sesquilinear if

$$\begin{aligned}\phi(m, n_1 a_1 + n_2 a_2) &= \phi(m, n_1) a_1 + \phi(m, n_2) a_2, \\ \phi(m_1 a_1 + m_2 a_2, n) &= \alpha(a_1) \phi(m_1, n) + \alpha(a_2) \phi(m_2, n)\end{aligned}$$

for all $m, m_1, m_2 \in M$; $n, n_1, n_2 \in N$ and $a_1, a_2 \in A$. The set of all such maps will be denoted by $S_\alpha(M, N)$: pointwise addition gives it the structure of Abelian group. It does not have a natural A -module structure, though it can be made a module over the centre of A . We will further abbreviate $S_\alpha(M) = S_\alpha(M, M)$. This is of course the usual definition: see e.g. (3), pp. 10-12, (4), p. 10.

Reflexive sesquilinear forms. In the case when A is a field or division ring, an element

ϕ of $S_\alpha(M)$ is termed reflexive when the conditions $\phi(m, n) = 0, \phi(n, m) = 0$ are equivalent. For the present more general situation, a stronger condition is appropriate. Let us require provisionally that $\phi(n, m)$ depends only on $\phi(m, n)$, say

$$\phi(n, m) = F(\phi(m, n)).$$

Then we have $\phi(m, n) = F(\phi(n, m)) = F(F(\phi(m, n)))$

and $F(\phi(m, n)a) = F(\phi(m, na)) = \phi(na, m)$
 $= \alpha(a)\phi(n, m) = \alpha(a)F(\phi(m, n)).$

Thus if ϕ takes the value 1 (and hence all values) we have the identities

$$(1) x = F(F(x)) \quad (2) F(xa) = \alpha(a)F(x).$$

Writing $y = F(x), b = \alpha(a)$, and applying F to (2) we find, using (1), that

$$(3) F(by) = F(y)\alpha^{-1}(b).$$

Now set $u = F(1)$. Then (2) with $x = 1$ and (3) with $y = 1$ give

$$(4) \alpha(a)u = F(a) = u\alpha^{-1}(a).$$

Putting $a = u$, we find $\alpha(u)u = 1 = u\alpha^{-1}(u)$,

so u is a unit, $\alpha(u) = u^{-1}$. Now (4) shows that α^2 is the corresponding inner automorphism of A .

Suppose given a sesquilinear form $\phi \in S_\alpha(M)$. Suppose u a unit of A with $\alpha(u) = u^{-1}$ and $\alpha^2(a) = uau^{-1}$ for all $a \in A$. Then we call ϕ (α, u) -reflexive, and write $\phi \in R_{(\alpha, u)}(M)$, if

$$\phi(n, m) = \alpha(\phi(m, n))u$$

for all $m, n \in M$. We observe that—as usual—the conjunction of this with either of the identities defining $S_\alpha(M)$ implies the other. Particular cases to be noted are $u = 1$, when ϕ is called *Hermitian*, and $u = -1$, when it is termed *skew-Hermitian*; in these cases, of course, α^2 must be the identity.

Our result coincides with the usual description ((1), p. 113, (4), p. 13) of reflexive sesquilinear forms of rank ≥ 2 over a division ring.

If $\phi \in R_{(\alpha, u)}(M)$, and v is any unit in A , then the map $\psi: M \times M \rightarrow A$ defined by

$$\psi(m, n) = v\phi(m, n)$$

is $(\beta, v\alpha(v^{-1}(u)))$ -reflexive, where $\beta(a) = v\alpha(a)v^{-1}$. In the usual case, one can choose v so that ψ is Hermitian or skew-symmetric; we can generalize the argument here as follows (see again (1), pp. 113–114; (4), p. 14). If, for some $m \in M, \phi(m, m) = 1$, then $1 = F(1) = \alpha(1)u$ implies $u = 1$ and ϕ Hermitian. If now $\phi(m, m) = v^{-1}$ is a unit, it follows that the form ψ above is Hermitian. Now for A a division ring, non-units are zero, and $\phi(m, m)$ identically zero implies ϕ skew-symmetric (hence $\alpha = 1$ and A is commutative). In general, we may have $\phi(m, m)$ neither zero nor a unit.

Quadratic forms. We retain the notation of the preceding section. For $\phi \in S_\alpha(M)$, define $T_u(\phi)$ by

$$T_u(\phi)(m, n) = \alpha(\phi(n, m))u.$$

It is easily seen that $T_u(\phi) \in S_\alpha(M)$; thus $T_u: S_\alpha(M) \rightarrow S_\alpha(M)$. Further, T_u is a group

homomorphism, $T_u^2 = 1$, and $T_{-u} = -T_u$. We have called $\phi(\alpha, u)$ -reflexive if $T_u(\phi) = \phi$: this amounts to defining $R_{(\alpha, u)}(M) = \text{Ker}(T_u - 1)$. We now define

$$Q_{(\alpha, u)}(M) = \text{Coker}(T_u - 1),$$

and refer to the elements of $Q_{(\alpha, u)}(M)$ as (α, u) -quadratic forms on M . We owe this terminology to Tits (5): we hope that the reason for it will become clear in the sequel.

Since $T_u^2 - 1 = 0$, multiplication by $T_u + 1$ induces a map b of $Q_{(\alpha, u)}(M) = \text{Coker}(T_u - 1)$ into $\text{Ker}(T_u - 1) = R_{(\alpha, u)}(M)$, called *bilinearization*. In the classical case (A a field, α the identity, $u = 1$, maps represented by matrices), this operation amounts to symmetrizing a matrix. If 2 is invertible in A , b is an isomorphism (an inverse is obtained by dividing the composite map $R_{(\alpha, u)}(M) \subset S_\alpha(M) \rightarrow Q_{(\alpha, u)}(M)$ by 2), so the definition introduces no essentially new concept in this case. In general, however, b is neither injective nor surjective; in particular, over division rings, its image is the set of tracic forms of Dieudonné ((4), p. 19), as we see from the formulae below.

For further discussion of a quadratic form θ we need, as well as the bilinearization b_θ , a certain associated quadratic mapping q_θ . Represent θ by $\phi \in S_\alpha(M)$; then the other representatives are the forms $\phi + T_u(\chi) - \chi$ for $\chi \in S_\alpha(M)$. Now for $m \in M$,

$$\begin{aligned} (\phi + T_u(\chi) - \chi)(m, m) &= \phi(m, m) + T_u(\chi)(m, m) - \chi(m, m) \\ &= \phi(m, m) + \alpha(\chi(m, m))u - \chi(m, m). \end{aligned}$$

Define the value group $V_{(\alpha, u)}$ to be the quotient of A by the additive subgroup of elements $\alpha(a)u - a$, for $a \in A$. Then the class of $\phi(m, m)$ in $V_{(\alpha, u)}$ depends only on the quadratic form θ , not on the choice of representative ϕ ; we denote it by $q_\theta(m)$.

Next, given a form $\theta \in Q_{(\alpha, u)}(M)$, we elucidate the formal properties of q_θ —or rather, of the pair (q_θ, b_θ) . Again let $\phi \in S_\alpha(M)$ represent θ ; and write $\{a\}$ for the class in $V_{(\alpha, u)}$ of $a \in A$. Then

$$\begin{aligned} q_\theta(m_1 + m_2) &= \{\phi(m_1 + m_2, m_1 + m_2)\} \\ &= \{\phi(m_1, m_1)\} + \{\phi(m_1, m_2) + \phi(m_2, m_1)\} + \{\phi(m_2, m_2)\} \\ &= q_\theta(m_1) + \{b_\theta(m_1, m_2) + \phi(m_2, m_1) - \alpha(\phi(m_2, m_1))u\} + q_\theta(m_2) \\ &= q_\theta(m_1) + \{b_\theta(m_1, m_2)\} + q_\theta(m_2). \end{aligned} \tag{1}$$

$$\begin{aligned} q_\theta(ma) &= \{\phi(ma, ma)\} \\ &= \{\alpha(a)\phi(m, m)a\}. \end{aligned}$$

Now $\alpha(a)(\alpha(x)u - x)a = \alpha(\alpha(a)xa)u - \alpha(a)xa$, and so $\{\alpha(a)ya\}$ depends only on $\{y\}$; we may thus write it as $\alpha(a)\{y\}a$. The formula then becomes

$$q_\theta(ma) = \alpha(a)q_\theta(m)a. \tag{2}$$

Finally, we make the analogous observation that for $x \in A$, $x + \alpha(x)u$ depends only on $\{x\}$, so that for $v \in V$ we can write $v + \alpha(v)u$ for the indicated element of A . Then

$$\begin{aligned} b_\theta(m, m) &= \phi(m, m) + T_u(\phi)(m, m) \\ &= \phi(m, m) + \alpha(\phi(m, m))u \\ &= q_\theta(m) + \alpha(q_\theta(m))u. \end{aligned} \tag{3}$$

These are the desired formal properties; we will write $\text{Quad}_{(\alpha, u)}(M)$ for the set of

pairs (b, q) where $b \in R_{(\alpha, w)}(M)$ and $q: M \rightarrow V_{(\alpha, w)}$ satisfy (1), (2) and (3): this, too, can be regarded as an additive group. If we set $f(\theta) = (b_\theta, q_\theta)$, then

$$f: Q_{(\alpha, w)}(M) \rightarrow \text{Quad}_{(\alpha, w)}(M).$$

THEOREM 1. *If M is a projective A -module, f is an isomorphism.*

Proof. In the natural direct sum decomposition

$$S_\alpha(M \oplus N) \cong S_\alpha(M) \oplus S_\alpha(M, N) \oplus S_\alpha(N, M) \oplus S_\alpha(N),$$

the map T_u induces an isomorphism between the middle two summands, so that

$$Q_{(\alpha, w)}(M) \cong Q_{(\alpha, w)}(M) \oplus S_\alpha(M, N) \oplus Q_{(\alpha, w)}(N);$$

and similarly for $R_{(\alpha, w)}$. We claim that there is a corresponding splitting for $\text{Quad}_{(\alpha, w)}$, and that f respects the splittings. Indeed, by taking the restrictions of b to $M \times M$, $M \times N$ and $N \times N$ and of q to M , $\{0\}$ and N respectively we obtain a map

$$\text{Quad}_{(\alpha, w)}(M \oplus N) \rightarrow \text{Quad}_{(\alpha, w)}(M) \oplus S_\alpha(M, N) \oplus \text{Quad}_{(\alpha, w)}(N).$$

Conversely, given a quintuple defining an element of the right-hand side, we define b on $N \times M$ by requiring $b(n, m) = \alpha(b(m, n))u$; then additivity shows how to extend b uniquely to $(M \oplus N) \times (M \oplus N)$, and the result is then necessarily reflexive. We now extend q to $M \oplus N$

$$q((m, n)) = q(m) + \{b(m, n)\} + q(n).$$

This extension is uniquely determined by (1): we now claim that (1), (2) and (3) are satisfied by it. The verification of this presents no difficulty.

Now the splittings were defined so that f does respect them. Hence f induces an isomorphism for $M \oplus N$ if and only if it does so for both M and N . The same argument will also work for infinite direct sums. Hence if we can show the theorem for $M = A$, it will follow first for free modules, and then for arbitrary projective modules.

Finally, take $M = A$. Assigning $\phi(1, 1)$ to ϕ gives an isomorphism of $S_\alpha(A)$ on A which carries T_u to the map $a \rightarrow \alpha(a)u$, and hence induces an isomorphism of $Q_{(\alpha, w)}(A)$ on $V_{(\alpha, w)}$. This isomorphism factors as the composite with f of the map taking (b, q) to $q(1)$. It remains to show this map injective. But $q(a) = \alpha(a)q(1)a$ and

$$b(a, a') = \alpha(a)b(1, 1)a' = \alpha(a)(q(1) + \alpha(q(1))u)a'$$

are both determined by $q(1)$. This completes the proof.

The usual case of quadratic forms is of course when $u = 1$, α is the identity, and A is commutative. In this case, $V_{(\alpha, w)}$ is the quotient of A by the zero subgroup. Thus Axiom 1 for quadratic forms shows that q_θ determines b_θ in this case. The remaining conditions (Axiom 3 is now superfluous) are the usual axioms for quadratic forms (see e.g. (3), p. 54).

The consideration of pairs (b, q) with b reflexive and q satisfying axioms (1)–(3) above was forced on us in (7), (8) as the algebraic expression of certain geometrical facts about intersection and self-intersection numbers.

Apropos of the splitting used in this proof, we may note that a form on $M \oplus N$ whose component in $S_\alpha(M, N)$ is zero in the usual (orthogonal) direct sum of the forms induced on M and on N . Thus a quadratic form splits as an orthogonal direct sum whenever its bilinearization does.

Duality. If M is a right A -module, the dual $\text{Hom}_A(M, A)$ is naturally a left A -module. If α is an anti-automorphism of A we can make the dual a right A -module, M^α , by defining

$$fa(m) = \alpha(a)f(m)$$

for $a \in A$, $m \in M$ and $f: M \rightarrow A$. If $g: M \rightarrow N$ is a map of right A -modules, we will write $g^\alpha: N^\alpha \rightarrow M^\alpha$ for the corresponding dual map: we have, of course, a contravariant functor.

In this notation, the natural map of M to its double dual associates to each $m \in M$ the A -linear map $\omega_{M, \alpha}(m): M^\alpha \rightarrow A$ defined by

$$\omega_{M, \alpha}(m)(f) = \alpha^{-1}(f(m)).$$

Then $\omega_{M, \alpha}$ is an A -module map $M \rightarrow (M^\alpha)^{\alpha^{-1}}$. It is an isomorphism when M is a finitely generated projective module. Whenever it is an isomorphism (and this does not depend on the choice of α) the module M is called *reflexive*.

Now let $\phi \in S_\alpha(M, N)$. We define the *associated homomorphism* $A\phi: M \rightarrow N^\alpha$ by

$$A\phi(m)(n) = \phi(m, n).$$

This is a map of right A -modules. We also define the *transpose* $\phi^t \in S_{\alpha^{-1}}(N, M)$ by

$$\phi^t(n, m) = \alpha^{-1}(\phi(m, n));$$

this has an associated homomorphism $A\phi^t: N \rightarrow M^{\alpha^{-1}}$.

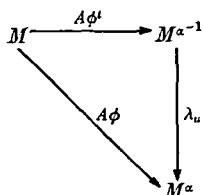
LEMMA 2. *The following diagrams commute*



Proof. By symmetry, it is enough to consider the first. Now

$$\begin{aligned} (A\phi)^{\alpha^{-1}}(\omega_{N, \alpha}(n))(m) &= \omega_{N, \alpha}(n)(A\phi(m)) \\ &= \alpha^{-1}(A\phi(m)(n)) \\ &= \alpha^{-1}(\phi(m, n)) \\ &= \phi^t(n, m) \\ &= A\phi^t(n)(m). \end{aligned}$$

In the case considered above when u is a unit of A with $\alpha(u) = u^{-1}$ and $\alpha^2(a) = uau^{-1}$ for $a \in A$, left multiplication by u gives a natural isomorphism $\lambda_u: M^{\alpha^{-1}} \rightarrow M^\alpha$ for any M . With this notation, (α, u) -reflexivity of $\phi \in S_\alpha(M)$ is expressed by commutativity of the diagram



For this commutativity means that $A\phi(m) = \lambda_u(A\phi^t(m))$ for $m \in M$; applying this to $n \in N$, the formula reduces to

$$\phi(m, n) = u\phi^t(m, n) = u\alpha^{-1}(\phi(n, m)) = \alpha(\phi(n, m))u.$$

Note also that the natural transformation from M to its double dual is now

$$\lambda_u \circ \omega_{M, \alpha} = (\lambda_u^{-1})^\alpha \circ \omega_{M, \alpha^{-1}}: M \rightarrow M^{\alpha\alpha}.$$

Non-singularity. We call $\phi \in S_\alpha(M, N)$ *non-singular* if $A\phi$ and $A\phi^t$ are isomorphisms; *non-degenerate* if both are monomorphisms. The two concepts coincide, of course, over division rings. Bourbaki ((3), p. 13) uses the term ‘non-degenerate’ in the above sense, and (see e.g. p. 23) uses also the concept of non-singularity. We owe the terminology to Hirzebruch, and have used it also in (6). Lemma 2 shows that if ϕ is non-singular, M and N are both reflexive; and that conversely, if $A\phi$ is an isomorphism and N is reflexive, ϕ is non-singular. This argument comes from (6), as does the next remark.

Let $\phi \in R_{(\alpha, u)}(M)$ and let S be a submodule of M such that the restriction of ϕ to S is non-singular. Then if

$$T = \{m \in M : \phi(m, s) = 0 \text{ for all } s \in S\}$$

is the orthogonal complement of S in M , M is the direct sum of S and T . Moreover, ϕ is non-singular if and only if its restriction to T is. For since the composite

$$S \subset M \xrightarrow{A\phi} M^\alpha \xrightarrow{i^\alpha} S^\alpha$$

is an isomorphism, say e , $e^{-1} \circ i^\alpha \circ A\phi: M \rightarrow S$ is a retraction, and clearly has kernel T . Thus M is the direct sum; now by reflexivity, $\phi(S \times T) = 0$, so (M, ϕ) is the orthogonal direct sum of its restrictions to S and T .

Given a non-singular $\phi \in S_\alpha(M, N)$, and $f: N \rightarrow N$, we define the adjoint $J_\phi(f)$ to be the (unique) map which makes the diagram

$$\begin{array}{ccc} M & \xrightarrow{J_\phi(f)} & M \\ \downarrow A\phi & & \downarrow A\phi \\ N^\alpha & \xrightarrow{f^\alpha} & N^\alpha \end{array}$$

commute; equivalently, it is characterized by the identity

$$\phi(m, f(n)) = \phi(J_\phi(f)(m), n).$$

Clearly we have $J_\phi(fg) = J_\phi(g)J_\phi(f)$, and J_ϕ is additive.

If $M = N$, J_ϕ maps the A -endomorphism ring of M to itself; we say that if ϕ is (α, u) -reflexive, J_ϕ is an involutory anti-automorphism of the ring. It remains, in fact only to check that $J_\phi^2 = 1$.

This is equivalent to obtaining the identity

$$\phi(m, J_\phi(f)(n)) = \phi(f(m), n),$$

and this is achieved by the computation

$$\begin{aligned} \phi(m, J_\phi(f)(n)) &= \alpha(\phi(J_\phi(f)(n), m))u \\ &= \alpha(\phi(n, f(m)))u \\ &= \phi(f(m), n). \end{aligned}$$

A quadratic form is called non-singular if its bilinearization is. A submodule on which a quadratic form is non-singular splits as an orthogonal direct summand: this follows by applying an earlier remark to its bilinearization. The concept of non-singular quadratic form is the one we have been leading up to, we suggest that it is the most interesting sort of quadratic (or Hermitian) form to study. We conclude with one result which supports this contention.

For a reflexive A -module M , we define the ‘hyperbolic space’ $H_\alpha(M)$ to be the module $M \oplus M^\alpha$ equipped with the α -sesquilinear form h given by

$$h((p, q), (p', q')) = q(p').$$

The coset of h defines a quadratic form θ_M in $Q_{(\alpha, u)}(M \oplus M^\alpha)$. The bilinearization of θ_M is given by

$$b((p, q), (p', q')) = q(p') + \alpha(q'(p))u,$$

and

$$Ab: M \oplus M^\alpha \rightarrow (M \oplus M^\alpha)^\alpha = M^\alpha \oplus (M^\alpha)^\alpha$$

can be expressed as $Ab(p, q) = (q, \lambda_u \omega_{M, \alpha}(p))$ which, by the remark above, is (for M reflexive) the natural isomorphism. We now give our result (see also (7), 4.5).

THEOREM 3. *Let θ be a non-singular (α, u) -quadratic form on M . Then*

$$(M, \theta) \oplus (M, -\theta) \cong (M \oplus M^\alpha, \theta_M).$$

Proof. Let ϕ represent θ . We define $f: M \oplus M \rightarrow M \oplus M^\alpha$ by

$$f(m, n) = (m - (Ab_\theta)^{-1}(A\phi)(m - n), Ab_\theta(m - n)).$$

Then

$$h(f(m, n), f(m', n')) = \phi(m - n, m') + T_u(\phi)(m - n, n')$$

as is seen after a little calculation, and this differs from $\phi(m, m') - \phi(n, n')$ only by

$$\phi(n, m' - n') - T_u(\phi)(m - n, n') = (1 - T_u)(\psi)((m, n), (m', n'))$$

for

$$\psi((m, n), (m', n')) = \phi(n, m' - n').$$

Thus f takes one quadratic form to the other. It follows, since each quadratic form is non-singular, that f is an isomorphism (alternatively, this can easily be seen directly).

Observe that the corresponding result for $R_{(\alpha, u)}(M)$ is false even in the elementary case of forms over the field with 2 elements (or indeed, also over \mathbf{Z}) with $\alpha = \text{identity}$, $u = 1$. For with $H(M)$, and hence with any submodule or summand, $b_\theta(x, x)$ is always zero (even). But the non-singular form defined on A by $b(x, y) = xy$ does not share this property. In the terminology of Bass (2), the hyperbolic functor is cofinal, but the corresponding functor to reflexive sesquilinear forms is not.

REFERENCES

- (1) ARTIN, E. *Geometric algebra* (Interscience, New York, 1957).
- (2) BASS, H. *Lectures on topics in algebraic K-theory* (Tata Institute Bombay, 1967).
- (3) BOURBAKI, N. Algèbre ch. 9: *Formes sesquilinéaires et formes quadratiques* (Hermann, Paris, 1959).
- (4) DIEUDONNÉ, J. *La géométrie des groupes classiques*, 2nd ed. (Springer, Berlin, 1963).
- (5) TITS, J. Formes quadratiques, groupes orthogonaux, et algèbres de Clifford, *Invent. Math.* 5 (1968), 19–41.
- (6) WALL, C. T. C. Quadratic forms on finite groups and related topics, *Topology*, 2 (1963), 281–298.
- (7) WALL, C. T. C. Surgery of non simply-connected manifolds, *Ann. of Math.* 84 (1966), 217–276.
- (8) WALL, C. T. C. Surgery of compact manifolds, to appear.