ACTA ARITHMETICA
XXIV (1973)

Support polygons and the resolution of modular functional singularities*

by

HARVEY COHN (New York, N. Y.)

Dedicated to Prof. C. L. Siegel on his seventy-fifth birthday

1. Introduction. The geometry of numbers and Hilbert modular functions are two topics not only permanently influenced by the work of C. L. Siegel but permanently interrelated by his work, particularly through the reduction theory of fundamental domains (see [5], [6], [7]).

The purpose of this paper is to consider another phase of this interrelationship as seen through the recent work of F. Hirzebruch on the resolution of singularities (see [3], [4]). A reexamination of these newer methods leads to another elementary idea in the geometry of numbers, namely the "support polygon" for the set of integral lattice points in a sector about the origin.

Many of the results presented here have analogues in results of Hirzebruch, particularly as they apply to sectors whose slopes are rational or conjugate quadratic. Nevertheless a model in the geometry of numbers has the advantage of the natural invariance under $GL_2(\boldsymbol{Z})$, and, with it the inherent facility for computing and generalizing to several dimensions (as we shall do later on).

2. Semigroup of a sector. We begin with an algebraic concept.

DEFINITION 2.1. Let S denote a closed *sector* of the cartesian (x, y)-plane bounded by two rays from the origin of angle $< 180°$.

DEFINITION 2.2. Let S be called *reduced* if it is bounded by a ray of slope λ_2 (>1) in the first quadrant and a ray of slope λ_1 $(0 < \lambda_1 < 1)$ in the third quadrant (thus containing the entire second quadrant at least).

* This work has been sponsored by the U. S. National Science Foundation under grant GP-33910.

DEFINITION 2.3. Let VS (*vectors* of S) denote the integral lattice points contained in S excluding $(0, 0)$ (but naturally including those on any rational ray bounding S).

DEFINITION 2.4. Let SGVS (*semigroup of the sector* S) denote the semigroup of VS under addition.

Remark 2.5. The sector can have an angle $> 180°$, but this matter can be ignored at present.

3. Support polygon of a sector. Now geometric techniques enter.

DEFINITION 3.1. A *support line* of a set VS is a straight line containing no points of VS in one of the open half-planes that this line determines.

DEFINITION 3.2. A *support point* is a point of VS which lies on a support line (which may contain several such points).

DEFINITION 3.3. A pair of support points are *neighbors* if they are end-points of a segment containing no other support points (in between).

DEFINITION 3.4. A *support segment* is the segment joining neighboring support points.

DEFINITION 3.5. A *support polygon* is a sequence of consecutive support segments with common end-points (i.e., with *vertices* as support points).

Remark 3.6. Since support segments are of (euclidean) length at least unity, the support polygon must in the limit extend to infinity. When the boundary ray of S has rational slope the support polygon contains all the integral points of such a ray (except the origin).

DEFINITION 3.7. A *truncated* support polygon is the portion of the support polygon excluding support segments lying in the rational boundaries, if they occur. (See Figure 1 below.)

4. Minimal basis. The following result is immediate:

THEOREM 4.1. *The vertices of a truncated support polygon constitute the unique minimal basis of the semigroup* SGVS.

We single out the most vital step.

LEMMA 4.2. *Any two neighboring support points of a truncated support polygon form a basis of the additive group of integral lattice points (i.e., these support points have a unimodular determinant).*

For proof, let $P_1(x_1, y_1)$, $P_2(x_2, y_2)$ be a support segment. Then with O as the origin, OP_1P_2 bounds a triangular region with no other lattice point in its closure, by definition. Hence the parallelogram OP_1QP_2 (with $Q = P_1+P_2$) is fundamental.

Theorem 4.1 follows from the further observation that every support point (of the truncated polygon) is necessarily a basis element of the semigroup, as it can not be obtained from the others, by construction.

5. Invariant intersection numbers. It is clear that the construction of the support polygon makes it covariant with S under $\mathrm{GL}_2(\boldsymbol{Z})$ or $\mathrm{SL}_2(\boldsymbol{Z})$. We standardize the matrix action as

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad M\begin{pmatrix} y \\ x \end{pmatrix} = \begin{pmatrix} ay+bx \\ cy+dx \end{pmatrix}. \tag{5.1}$$

We also note that the pair of sectors S and $-S$ (180° rotation) are covariant with respect to their support polygons under $\mathrm{PGL}_2(\boldsymbol{Z})$ or $\mathrm{PSL}_2(\boldsymbol{Z})$. The matrices M and $-M$ are identified, or, more practically, both matrices $\pm M$ lead to the same function

$$M(\lambda) = (a\lambda+b)/(c\lambda+d) \qquad (\lambda = y/x). \tag{5.2}$$

DEFINITION 5.3. Let $(x_m, y_m) = v_m$ denote the support points of a sector S lying interior to S (and numbered in sequence of the polygonal vertices). Then the *intersection number* at v_m is defined by scalars b_m for which

$$b_m v_m = v_{m-1} + v_{m+1}. \tag{5.4}$$

DEFINITION 5.5. The *chain* (of intersection numbers) of a sector S is the sequence

$$S \mapsto [\ldots, b_{-2}, b_{-1}, b_0, b_1, b_2, \ldots]. \tag{5.6}$$

The sequence is infinite toward the ends of the sector where irrational slopes occur, while the sequence terminates where rational slopes occur at the last support point interior to the sector. Under this definition, a vacuous chain of intersection numbers describes a sector with boundaries of rational slope determined by a fundamental parallelogram, or by ratios of integers in unimodular relation.

LEMMA 5.7. *The intersection numbers b_m exist ($\in \boldsymbol{Z}$) and satisfy $b_m \geqslant 2$, with no infinite succession of consecutive "twos" possible in any chain.*

Proof. Use Lemma 4.2 to note that $M \in \mathrm{GL}_2(\boldsymbol{Z})$ where

$$M = \begin{pmatrix} -y_m & y_{m-1} \\ -x_m & x_{m-1} \end{pmatrix}, \quad M\begin{pmatrix} Y \\ X \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}. \tag{5.8}$$

Thus let $V_m = (X_m, Y_m)$, the image of v_m under M^{-1}. Then we have $V_{m-1} = (-1, 0)$, $V_m = (0, 1)$. So if we write $V_{m+1} = (X_{m+1}, Y_{m+1})$, we find that $X_{m+1} = 1$ (to preserve unimodularity with V_m), and Y_{m+1} ($\in \boldsymbol{Z}$) is precisely b_m. Thus the obvious relation

$$b_m V_m = V_{m-1} + V_{m+1} \tag{5.9}$$

must transform itself into (5.4). Now $b_m \geqslant 2$ by *convexity* (recall the support segments). Furthermore "$b_m = 2$" is a relation of collinearity of consecutive points, which precludes an infinitude of consecutive "twos".

Remark 5.10. It is still possible to have the relation

$$v_k + v_l = b v_m \quad (2 \leqslant b \in \mathbf{Z}) \tag{5.11}$$

without this equation reducing to the type (5.4) but all we can say then is that m lies between k and l (in the usual enumeration (5.6) of consecutive vertices of the support polygon).

Remark 5.12. The terminology "intersection number" is based on the fact that in (5.4), b_m describes the "intersection" of the coordinate system of (v_{m-1}, v_m) and (v_m, v_{m+1}) as lattice bases. This is in keeping with the idea of (negative) Chern number developed in Hirzebruch's work [3]. From a purely number-theoretic point of view, however, each of these bases gives a b_m-fold covering of the coordinate system formed by the nonbasis (v_{m-1}, v_{m+1}).

THEOREM 5.13. *Every sector S is determined uniquely to within* $\mathrm{SL}_2(\mathbf{Z})$ *(i.e., with orientation), by the (directed) chain of intersection numbers*

$$S \mapsto [\ldots, b_{-2}, b_{-1}, b_0, b_1, b_2, \ldots], \quad b_m \geqslant 2, \tag{5.14}$$

(subject only to the "infinite twos" prohibition of Lemma 5.7).

For proof, assume for simplicity that there are at least two support points interior to S so we can transform neighbors to $(-1, 0)$, $(0, 1)$ under $\mathrm{SL}_2(\mathbf{Z})$. Then by (5.4), the intersection numbers determine the whole polygon uniquely, step by step. (The polygon will determine the sector uniquely, in turn, by the limiting slope.) The further result that any sequence (5.14) subject to the "infinite twos prohibition" is admissible must wait for Lemma 6.1 (below).

COROLLARY 5.15. *The semigroup of the sector S is determined to within isomorphism by the (unoriented) chain of intersection numbers.*

This is a result of the uniqueness of the minimal basis (Theorem 4.1), combined with Theorem 5.13 and Remark 5.10. In effect, if all the basis elements v_m can be linked by equations of type (5.11), then they have to be "numbered properly" and they define the correct intersection numbers.

LEMMA 5.16. *Every sector is equivalent under* $\mathrm{SL}_2(\mathbf{Z})$ *to one in reduced form (with the limiting cases $\lambda_1 = 0$ or $\lambda_2 = \infty$ or both).*

This is a result of the proof of Lemma 5.7.

6. The Hirzebruch algorithm. To find the intersection numbers of a sector we use the following method of Hirzebruch [3]:

LEMMA 6.1. *Every real λ has a unique expansion as continued fraction*

$$\lambda = b_0 - \frac{1}{b_1} - \frac{1}{b_2} - \ldots - \tag{6.2}$$

denoted by

$$\lambda = [b_0, b_1, b_2, \ldots], \tag{6.3}$$

where $b_0 \in \mathbf{Z}$, $b_m \geqslant 2$ for $m > 0$, and no infinite sequence of consecutive twos can occur if the expansion is infinite.

Conversely, every such expansion corresponds uniquely to some real λ. Finally, the expansion is infinite exactly when λ is irrational.

This lemma is quite analogous to the ordinary simple continued fraction. The recursive procedure is set up by writing $\lambda = \lambda_0$ and defining inductively the "partial denominators" b_m and the "remainders" λ_m as follows:

$$b_m = -[-\lambda_m], \quad \lambda_m = b_m - 1/\lambda_{m+1}. \tag{6.4}$$

(Note the inequality $\lambda \leqslant -[-\lambda] < \lambda + 1$.) Also the "infinite twos prohibition" and uniqueness are tied together in the fact that $[b_t, b_{t+1}, \ldots] = \lambda_t \geqslant 1$ $(t > 0)$, with equality only when all $b_t = 2$.

THEOREM 6.5. *Let S denote the sector (for $0 < \lambda \in \mathbf{R}$),*

$$0 \leqslant x/y \leqslant 1/\lambda, \quad y > 0. \tag{6.6}$$

Then the support polygon for S begins with a vertex at $v_0 = (0, 1)$ and ends at $v_M = (x_M, y_M)$ if $\lambda = y_M/x_M$ (reduced positive fraction), or else the support polygon is infinite if $\lambda \notin \mathbf{Q}$. The intermediate vertices $v_m = (x_m, y_m)$ are given by the (reduced positive) fractions

$$y_m/x_m = [b_0, b_1, \ldots, b_{m-1}] \tag{6.7}$$

arising from the expansion of λ. The intersection number at $v_m = (x_m, y_m)$ is b_m, and the chain for the sector S is

$$S \mapsto [b_1, b_2, \ldots]. \tag{6.8}$$

Then chain is vacuous when $0 < \lambda \in \mathbf{Z}$ as the support polygon is then the single segment from $(0, 1)$ to $(1, \lambda)$.

To prove this theorem, we set up an inductive process by transforming the sector from coordinates (x, y) to (x', y') by

$$\begin{pmatrix} y \\ x \end{pmatrix} = \begin{pmatrix} b_0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y' \\ x' \end{pmatrix}. \tag{6.9}$$

Thus the new sector is one with slope λ_1 (where $\lambda = b_0 - 1/\lambda_1$),

$$S'\colon\ y'/x' \leqslant \lambda_1, \ y' > 0. \tag{6.10}$$

The initial support vertex (for $m = 0$) is $(x', y') = (-1, 0)$, and the next one is $(x', y') = (0, 1)$ (corresponding to $m = 1$). The vertex corresponding to $m = 2$ is then $(1, b_1)$ since $b_1 = -[-\lambda_1]$. Thus b_1 is the intersection number for $m = 1$. Similarly, we identify $b_2, b_3, \ldots$

LEMMA 6.11. *If* $y_n/x_n = [b_0, b_1, \ldots, b_{n-1}]$ *with* $x_n > 0$, $y_n > 0$ *and* $(x_n, y_n) = 1$, *then the values* $(X, Y) = (x_{n-1}, y_{n-1})$ *are uniquely determined by the condition*

(6.12) $$Yx_n - Xy_n = -1; \quad 0 \leqslant X,\ 0 < Y \leqslant y_n.$$

The proof is evident. The curious notation is required by the case $n = 1$, where we define formally

(6.13) $$(x_0, y_0) = (0, 1).$$

In Figure 1 an illustration is provided for $\lambda = 5/18$. (Also see the table.) The symbol t_n is explained in Section 9 (below).

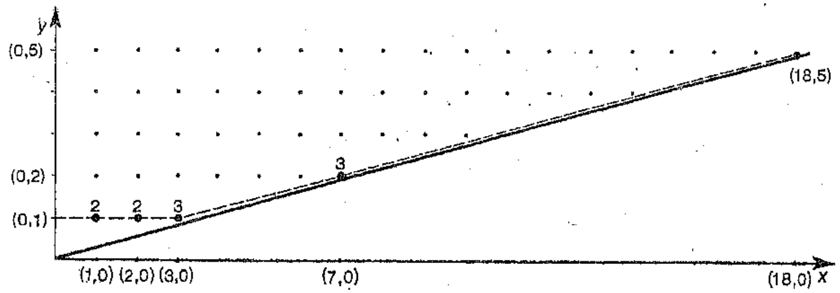

Fig. 1. Support polygon for the sector $0 \leqslant 5x/18 \leqslant y$. The dashed portion denotes the truncated polygon while the full polygon goes to infinity along the boundary rays. Note the intersection numbers at each interior vertex. The polygon is straight exactly when the intersection number is 2. A complete table follows:

n	-1	0	1	2	3	4	5
b_n		1	2	2	3	3	
y_n	0	1	1	1	1	2	5
x_n	-1	0	1	2	3	7	18
t_n		18	13	8	3	1	0

We note the role of intersection numbers in the relations

(6.14) $$b_n\begin{pmatrix} y_n \\ x_n \end{pmatrix} = \begin{pmatrix} y_{n-1} \\ x_{n-1} \end{pmatrix} + \begin{pmatrix} y_{n+1} \\ x_{n+1} \end{pmatrix}.$$

7. The Lagrange algorithm. The ordinary continued fraction (often attributed to Lagrange) has the form

(7.1) $$\lambda = a_0 + \frac{1}{a_1} + \frac{1}{a_2} + \ldots = (a_0, a_1, \ldots)$$

with $\lambda_0 = \lambda\ (\in \boldsymbol{R})$, $a_i = [\lambda_i]$, $\lambda_i = a_i + 1/\lambda_{i+1}$. Here $a_0 \in \boldsymbol{Z}$, $a_j \geqslant 1$ (for $j > 0$) and if λ terminates in a_n ($\lambda \in \boldsymbol{Q}$) $a_n \geqslant 2$ for uniqueness. This algorithm occurs most prominently in number theory because the convergents

(7.2) $$p_n/q_n = (a_0, a_1, \ldots, a_n)$$

provide both the larger and smaller of the best approximations.

By comparison, the Hirzebruch algorithm provides necessarily only the larger of the approximations as well as many which are not the best. Specifically, the convergents y_m/x_m of the Hirzebruch algorithm are the set (for $n = 1, 3, 5, \ldots$) of "neben-fractions"

(7.3) $$\frac{p_n}{q_n}, \quad \frac{p_n + p_{n+1}}{q_n + q_{n+1}}, \quad \ldots, \quad \frac{p_n + a_{n+2}p_{n+1}}{q_n + a_{n+2}q_{n+1}} = \frac{p_{n+2}}{q_{n+2}}.$$

In the standardized configuration (see Figure 1) for $\lambda > 0$, the Hirzebruch algorithm defines an approximating polygon above the ray $y = \lambda x$ ($x > 0$). To find an approximating polygon below this ray we must expand $1/\lambda$ and invert our convergents (or, equivalently expand $-\lambda$ and ignore sign of the convergents; see Remark 8.8 and Corollary 10.17 below).

8. Reduced form of a sector. If we use a reduced sector as defined in Section 2 (above), we can obtain the chain for $\boldsymbol{S}$ directly from Theorem 6.5. The $x \leftrightarrow y$ symmetry yields the following result:

THEOREM 8.1. *Let us expand*

(8.2) $$(1 <)\ \lambda_2 = [b_0, b_1, b_2, \ldots],$$

(8.3) $$(1 <)\ 1/\lambda_1 = [b_{-1}, b_{-2}, b_{-3}, \ldots],$$

then the reduced sector $\boldsymbol{S}$ *containing the second quadrant and lying between* $y = \lambda_1 x$ *in the third and* $y = \lambda_2 x$ *in the first is determined by the chain*

(8.4) $$\boldsymbol{S} \mapsto [\ldots, b_{-2}, b_{-1}, b_0, b_1, b_2, \ldots].$$

THEOREM 8.5. *Let* $\lambda^{(1)}, \lambda^{(2)}$ *be irrationals with the expansions*

$$\lambda^{(i)} = [b_0^{(i)}, b_1^{(i)}, \ldots] \quad (i = 1, 2).$$

The condition for ultimate agreement of these expansions, i.e., for some N_0

(8.6) $$b_i^{(1)} = b_{i+N_0}^{(2)},$$

is exactly that $\lambda^{(1)}$ *and* $\lambda^{(2)}$ *are equivalent in* $\mathrm{PSL}_2(\boldsymbol{Z})$. *In symbols,* $\lambda^{(1)} \approx \lambda^{(2)}$, *or*

(8.7) $$\lambda^{(1)} = (a\lambda^{(2)} + b)/(c\lambda^{(2)} + d), \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\boldsymbol{Z}).$$

It is clear that "ultimate agreement" leads to $\lambda^{(1)} \approx \lambda^{(2)}$. Therefore we have only to show that the equivalence leads to ultimate agreement. Assume $\lambda^{(i)} > 0$ (by adding integers if need be), and consider the sectors $\boldsymbol{S}^{(i)}$ from (say) the ray at 90° to the ray of slope $\lambda^{(i)}$. We know that for some $M \in \mathrm{SL}_2(\boldsymbol{Z})$, $M(\lambda^{(1)}) = \lambda^{(2)}$, so this transformation gives us two sectors terminating on the "right" with the same slope $\lambda^{(2)}$ (but with different slopes on the left). It remains only to show that if the slope is the same on the right, then the support polygons ultimately agree. But one sector now includes the other. Thus any support segment of the larger sector which lies in the smaller is necessarily a support segment of the smaller.

Remark 8.8. Thus if λ has an "upper polygon" as defined in Theorem 6.5, the expansion of $1/\lambda$ (or $-\lambda \approx 1/\lambda$, etc.), produces the "lower support polygon" on expansion and inversion (or negation, etc.).

9. Hirzebruch's resolution of radicals. The singularities

$$w = (z_1 z_2^{R-S})^{1/R}, \quad R > S > 0,\ (R, S) = 1, \tag{9.1}$$

occur at the fixed points of the Hilbert modular function (see [2]). We consider the function-theoretic problem of resolution only in terms of a very superficial aspect: The field $C(z_1, z_2, w)$ has a subring O of elements integral over the polynomial ring $C[z_1, z_2]$. The pure radicals in that ring are the set

$$w_{x,t} = (z_1^x z_2^t)^{1/R}, \quad t \geqslant 0,\ x \geqslant 0,\ t + Sx \equiv 0 \pmod{R}. \tag{9.2}$$

There would be only $R+1$ such radicals if we reduce x and t (mod R), but clearly, even so, all of them are not needed. Some can be constructed multiplicatively from others. What, then, would be the minimal basis of this whole (unreduced) set of pure radicals?

Clearly the (x, t) in (9.2) lead to a sector $\boldsymbol{S}$:

$$Ry = t + Sx, \quad 0/1 \leqslant x/y \leqslant R/S,\ y > 0,\ x \geqslant 0. \tag{9.3}$$

Thus as in Theorem 6.5, we expand

$$S/R = [b_0, b_1, \ldots, b_m], \quad (b_0 = 1,\ b_t \geqslant 2,\ t > 0). \tag{9.4}$$

The minimal basis of SGV$\boldsymbol{S}$ is the $n+1$ vectors (x_m, y_m) with

$$y_m/x_m = [b_0, b_1, \ldots, b_m], \quad 0 \leqslant m \leqslant n. \tag{9.5}$$

Thus for the illustration in Figure 1 (above) the generating radicals for $(z_1 z_2^{13})^{1/18}$ are

$$z_2, \quad (z_1 z_2^{13})^{1/18}, \quad (z_1^2 z_2^8)^{1/18}, \quad (z_1^3 z_2^3)^{1/18}, \quad (z_1^7 z_2)^{1/18}, \quad z_1. \tag{9.6}$$

Remark 9.7. If we renumber the $n+1$ basis radicals as $w_0, w_1, \ldots, w_n$ we can say that the coordinate systems of (w_{m-1}, w_m) and (w_m, w_{m+1}) have intersection number b_m at w_m, or

$$w_{m-1} w_{m+1} = w_m^{b_m}. \tag{9.8}$$

This is interpreted in [3] by the (negative) Chern number. Actually the fraction $R/(R-S) = [b_1, \ldots, b_n]$ gives the successive steps on "blowing up the origin" in [3].

Lemma 9.9. *If $P > Q > 0$, $P > Q' > 0$, $QQ' \equiv 1 \pmod{P}$, then if $[b_0, \ldots, b_s] = P/Q$, it follows that $[b_s, \ldots, b_0] = P/Q'$.*

This "classical lemma" is interesting because its proof can be seen purely algebraically. We note that $(z_1 z_2^Q)^{1/P}$ and $(z_1 z_2^{Q'})^{1/P}$ both lead to the same set of radicals, hence the sectors (9.3) are equivalent and the chains are the same except for order. (See Corollary 5.15.)

10. Quadratic sectors. Next consider the four sectors denoted by $\boldsymbol{S}, \boldsymbol{S}', -\boldsymbol{S}, -\boldsymbol{S}'$ as determined by the real lines

$$y = \lambda x, \quad y = \lambda' x, \quad (\lambda \neq \lambda'), \tag{10.1}$$

where the slopes λ, λ' are taken to be irrational for convenience. The four support polygons will, of course, match in pairs by 180° rotation, but, otherwise, two generally different polygons emerge.

Definition 10.2. Let $\Phi(x, y) = C(y - \lambda x)(y - \lambda' x)$, for $0 \neq C \in \boldsymbol{R}$, then the sectors $\boldsymbol{S}, \boldsymbol{S}', -\boldsymbol{S}, -\boldsymbol{S}'$ are called *sectors of the quadratic form Φ* (and the support polygons are said to belong to that form).

Definition 10.3. A support polygon of sector $\boldsymbol{S}$ (or the sector itself) is called *periodic* if its doubly infinite chain of intersection numbers is periodic, i.e., $b_{t+p} = b_t$ for some $p > 0$. We usually designate the minimal (primitive) period and write

$$\boldsymbol{S} \mapsto ((b_0, b_1, \ldots, b_{p-1})), \quad p > 0. \tag{10.4}$$

Theorem 10.5. *A support polygon belonging to a form Φ is periodic exactly when λ, λ' are real quadratic conjugates, or Φ is proportional to an integral quadratic form*

$$\Phi(x, y) = C(y - \lambda x)(y - \lambda' x) = Ax^2 + Bxy + Cy^2, \tag{10.6}$$

where $(A, B, C) = 1$ and the discriminant $D = B^2 - 4AC$ (> 0) is not a perfect square. Thus when one support polygon is periodic so is the other.

Proof. First assume periodicity, and further assume that the sector is in reduced form (see Lemma 5.16 above). Consider the support points given by the ratios

$$y_{p-1}/x_{p-1} = [b_0, \ldots, b_{p-2}], \quad (y_0/x_0 = 1/0), \tag{10.7a}$$

$$y_p/x_p = [b_0, \ldots, b_{p-1}], \tag{10.7b}$$

as in (6.7). Then the matrix action

$$M_0\begin{pmatrix} y' \\ x' \end{pmatrix} = \begin{pmatrix} y_p & -y_{p-1} \\ x_p & -x_{p-1} \end{pmatrix}\begin{pmatrix} y' \\ x' \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix} \tag{10.8}$$

will shift $(x', y') = (-1, 0)$ to (x_{p-1}, y_{p-1}) and $(x', y') = (0, 1)$ to (x_p, y_p). Thus M_0^{-1} maps S into another *reduced* sector of the same period, or into the *same* oriented sector by Theorem 5.13. Thus, in fractional form

$$M_0(\lambda) = (y_p\lambda - y_{p-1})/(x_p\lambda - x_{p-1}) = \lambda \tag{10.9}$$

for the roots λ, λ'. Thus these roots are real quadratic conjugates.

To show the converse, let

$$A + B\lambda + C\lambda^2 = 0, \quad (A, B, C) = 1, \quad D = B^2 - 4AC. \tag{10.10}$$

We wish to construct a matrix $M_1 \in \mathrm{SL}_2(\boldsymbol{Z})$ such that

$$\lambda = M_1(\lambda) = (a\lambda + b)/(c\lambda + d), \tag{10.11a}$$

or, equivalently

$$c\lambda^2 + (d - a)\lambda - b = 0. \tag{10.11b}$$

Thus we must find some t such that the matrix has the form

$$c = Ct, \quad d = a + Bt, \quad b = -At. \tag{10.12}$$

The unimodularity condition, $ad - bc = 1$, yields

$$(2a + Bt)^2 - Dt^2 = 4. \tag{10.13}$$

We easily see solutions can be constructed from the positively normed units of the ring of discriminant D. Thus at least one matrix M_1 must shift the support polygon into itself because it does not change the boundary slopes (10.11a) or the orientation (10.13). Thus there is a translation on the intersection numbers, which is precisely what a period is.

COROLLARY 10.14. *If the real quadratic conjugates satisfy*

$$\lambda > 1 > \lambda' > 0 \tag{10.15}$$

then the Hirzebruch expansions (6.2) *of λ and $1/\lambda'$ are purely periodic.*

This is a direct consequence of Theorem 8.1.

Remark 10.16. We now have a practical method of "reducing" a quadratic sector. Since any sector is equivalent under $\mathrm{SL}_2(\boldsymbol{Z})$ to a reduced sector, then any quadratic $\lambda \approx \lambda_0$ where λ_0 has an expansion with pure period. Therefore the expansion of λ must ultimately agree with λ_0 by ultimately having a pure period if we follow the elementary numerical operations in (6.4) above. (See Remark 11.14 below for an illustration.)

COROLLARY 10.17. *If λ is equivalent to a certain period, then this period is reversed by $1/\lambda'$ or $-\lambda'$.*

In terms of Remark 8.8, we are saying that the upper polygon of λ is the reverse of the lower polygon of $1/\lambda'$.

The above results on quadratic sectors are contained in essence in the work of Hirzebruch [4] except for the geometric imagery of the support polygons. The next two sections carry this work somewhat further in directions sufficiently broad to extend to several dimensions.

11. Construction of period from units. Actually a surd algorithm can be devised for producing the period analogously with that of the ordinary continued fraction. It is more in keeping with the spirit of ring theory to proceed directly from the units of the ring.

THEOREM 11.1. *Let us consider a sector corresponding to*

$$C\lambda^2 + B\lambda + A = 0, \tag{11.2}$$

where A, B, C are relatively prime integers and the sector is reduced so that $B < 0$, $A > 0$, $C > 0$. The discriminant is a nonsquare given by

$$0 < D = B^2 - 4AC. \tag{11.3}$$

Consider the support polygon of the reduced sector and its various periods (including the nonprimitive ones),

$$((b_0, \ldots, b_{p-1})), \; ((b_0, \ldots, b_{p-1}, b_0, \ldots, b_{p-1})), \; \ldots, \tag{11.4}$$

and at the same time consider the various units of the quadratic ring of discriminant D which have norm 1 and exceed 1 numerically,

$$\omega = (s + tD^{1/2})/2, \quad s^2 - t^2D = 4, \; s > 0, \; t > 0. \tag{11.5}$$

Then there is a biunique correspondance between (nonprimitive) periods by juxtaposition and (nonfundamental) units by multiplication expressed by the ratio

$$[period] = \tfrac{1}{2}(-Bt + s)/Ct. \tag{11.6}$$

Proof. We shall first show that in (11.6) some period must correspond to the given unit (11.5) via the ratio (11.6). Thus we expand the ratio $\frac{1}{2}(-Bt + s)/Ct$ by the Hirzebruch algorithm

$$\tfrac{1}{2}(-Bt + s)/Ct = [b_0, b_1, \ldots, b_{p-1}] = y_p/x_p. \tag{11.7}$$

We, of course, are taking liberties with notation. The value p is just the number of denominators in the expansion of the ratio (11.6), but we do not yet know that the set of b_i constitutes a period (or that the period is primitive). We make the further assertion that the ratio (11.6) is reduced, or

$$y_p = \tfrac{1}{2}(-Bt + s), \quad x_p = Ct. \tag{11.8}$$

This is true because the values

$$(11.9)\qquad y_{p-1} = At, \quad x_{p-1} = \tfrac{1}{2}(-Bt-s)$$

have a unimodular relation by (11.5), namely

$$(11.10)\qquad x_{p-1}y_p - x_p y_{p-1} = 1.$$

We can also apply Lemma 6.11 since

$$y_p/x_p > \tfrac{1}{2}(-Bt+D^{1/2}t)/Ct = \lambda > 1, \quad y_p/y_{p-1} > \tfrac{1}{2}(-Bt+D^{1/2}t)/At = 1/\lambda' > 1$$

by the definition of reduction. Therefore y_{p-1}/x_{p-1} expands into $[b_0, \ldots, b_{p-2}]$ ($=1/0$ if $p=1$).

We now have a matrix of $SL_2(\boldsymbol{Z})$

$$(11.11)\qquad \begin{pmatrix} y_p & -y_{p-1} \\ x_p & -x_{p-1} \end{pmatrix} = \begin{pmatrix} \tfrac{1}{2}(-Bt+s) & -At \\ Ct & -\tfrac{1}{2}(-Bt-s) \end{pmatrix}.$$

If we can show it leaves a sector invariant, it follows from the method of Theorem 10.5 (first part) that $((b_0, \ldots, b_{p-1}))$ is a period. To do this we check that (11.2) is the same as

$$(11.12)\qquad \left(\tfrac{1}{2}(-Bt+s)\lambda - At\right)/\left(Ct\lambda - \tfrac{1}{2}(-Bt-s)\right) = \lambda.$$

Conversely, let us start with a period. The corresponding linear transformation (11.11) stretches an eigenvector at $\lambda = y/x$ by the factor

$$(11.13)\qquad x_p\lambda - x_{p-1} = \omega.$$

Thus any period corresponds to a totally positive unit acting as multiplier on the boundary rays of the sector. The correspondance of units and periods is now complete.

Remark 11.14. We now have an algorithmic procedure for finding the primitive period from the fundamental (totally positive) unit. Let λ, λ' be conjugate surds. Expand them by Lemma 6.1. Take λ ($=\lambda_0$); a reduced surd must equal some remainder λ_M. Thus Theorem 11.1 determined the period "above" λ_M. We repeat the procedure for λ'.

For example, if we take $\lambda = 3^{1/2}$, $\lambda' = -3^{1/2}$, we expand

$$3^{1/2} = [-2, \lambda_1], \quad \lambda_1 = 2+3^{1/2} = ((4)),$$

$$-3^{1/2} = [-1, 2, \lambda_2'], \quad \lambda_2' = (3+3^{1/2})/3 = ((2, 3)).$$

Thus to apply the algorithm to (say) $\lambda_2' = (3+3^{1/2})/3$, we note it is reduced and satisfies $(A, B, C) = (2, -6, 3)$, $D = 12$, $(s, t) = (4, 1)$. Thus (11.7) becomes $5/3 = [2, 3]$, the period of λ_2' (or $-3^{1/2}$). (See Figure 2b, below.)

12. Symmetry types. There are four types of symmetry possible with quadratic sectors. They are illustrated in Figure 2 and the accompanying descriptions are justified by the use of Corollary 10.17.

a. *Rotational symmetry*: $\lambda \approx \lambda'$. This is like 90° rotation of the period, e.g., $\lambda = (21+221^{1/2})/22 \approx \lambda' \approx ((2, 3, 4))$.

b. *Reversing symmetry*: $\lambda \approx -\lambda'$. Each period is reversible, e.g., $\lambda = 3^{1/2} \approx ((4))$, $\lambda' = -3^{1/2} \approx ((2, 3))$.

c. *Reflection symmetry*: $\lambda \approx -\lambda$, $\lambda' \approx -\lambda'$. Here each period can be reflected across the eigenvectors into another period, e.g., $\lambda = (2+82^{1/2})/3 \approx ((4, 2, 2, 2, 2, 3, 2, 7))$, $\lambda' \approx ((7, 2, 3, 2, 2, 2, 2, 4))$.

d. *Total symmetry*: $\pm\lambda \approx \pm\lambda'$ (all four signs). Any two symmetries imply total symmetry; e.g., $\lambda = (3+5^{1/2})/2 = ((3))$.

e. *Total asymmetry*. Here only the trivial 180° rotation occurs as a symmetry; e.g., $\lambda = (2+79^{1/2})/3 \approx ((4, 2, 2, 7, 3))$, $\lambda' \approx ((2, 3, 3, 2, 2, 2, 5))$.

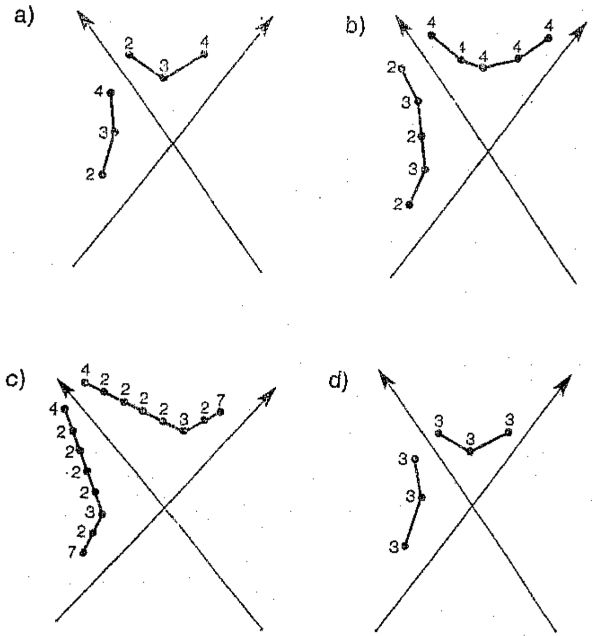

Fig. 2. Symmetries. The drawings are only symbolic and the opposite quadrants have polygons determined by symmetry with respect to the origin. a) Rotational symmetry $\Phi = 5x^2 - 11xy - 5y^2$; b) Reversing symmetry $\Phi = x^2 - 3y^2$; c) Reflection symmetry $\Phi = 3x^2 + 4xy - 26y^2$; d) Total symmetry $\Phi = x^2 - xy - y^2$

13. Symmetry criteria. The following criteria distinguish the symmetries. As before, the symbol $\approx$ denotes equivalence in $SL_2(\boldsymbol{Z})$ as well as $PSL_2(\boldsymbol{Z})$ according to context.

THEOREM 13.1. a. *Rotational symmetry holds exactly when*

$$\Phi \approx Ax^2+Bxy-Ay^2. \tag{13.2}$$

b. *Reversing symmetry holds exactly when Φ is ambiguous, or*

$$\Phi \approx Ax^2+Axy+Cy^2,\ Ax^2+Cy^2,\ Ax^2+Bxy+Ay^2. \tag{13.3}$$

c. *Reflection symmetry holds exactly when*

$$s^2-t^2D = -4 \tag{13.4}$$

is solvable for $s, t \in \boldsymbol{Z}$ (or unit of norm -1 exists in the ring of discriminant D).

d. *Total symmetry follows from any two of the above cases, so that any two symmetries imply the third.*

LEMMA 13.5. *Let $M \in \mathrm{GL}_2(\boldsymbol{Z})$ and*

$$M^2 = -dI, \quad d = \det M = \pm 1. \tag{13.6}$$

Then there exists a $V \in \mathrm{SL}_2(\boldsymbol{Z})$ such that one of these cases occurs:

$$VMV^{-1} = \pm\begin{pmatrix} 0 & 1 \\ -1 & 0\end{pmatrix} \quad (d>0), \tag{13.7}$$

$$VMV^{-1} = \pm\begin{pmatrix} 1 & -b \\ 0 & -1\end{pmatrix} \quad (d<0),\ b = 0, 1, \tag{13.8}$$

$$VMV^{-1} = \pm\begin{pmatrix} 0 & 1 \\ 1 & 0\end{pmatrix} \quad (d<0). \tag{13.9}$$

This lemma is proved by repeated use of $V_0 = \begin{pmatrix} 0 & 1 \\ -1 & 0\end{pmatrix}$ and $V_k = \begin{pmatrix} 1 & -k \\ 0 & 1\end{pmatrix}$ as transforms of M.

Now, to prove Theorem 13.1, we first note that rotational symmetry comes when $\lambda = M(\lambda')$ and $d = 1$, hence from (13.7). The reversing symmetry applies when $-\lambda = M(\lambda')$ and $d = -1$; there the three ambiguous forms in (13.3) follow from (13.8) with $b = 1$, from (13.8) with $b = 0$, and (13.9) respectively. The case of reflection symmetry applies when $-\lambda = M_r(\lambda)$, $M_r \in \mathrm{SL}_2(\boldsymbol{Z})$, which leads to (13.4) by the method of the proof of Theorem 10.5 (second part).

Remark 13.10. The rotational and reversing symmetries have meaning for nonquadratic irrationals (even for rationals), but the reflection symmetry is possible only for λ, λ' real quadratic conjugates.

Remark 13.11. The examples in Section 12 are contrived to have the smallest possible field-discriminants. For instance, the case of reflection symmetry only (Figure 2c) requires that a unit have norm -1, but that there be more than one class per genus (to preclude ambiguity of forms). Hence we arrive at $\boldsymbol{Q}(82^{1/2})$.

LEMMA 13.12. *Let $0 < D_0$ $(\in \boldsymbol{Z})$, $D_0^{1/2} \notin \boldsymbol{Z}$. Then if $D_0 \equiv 1 \pmod 4$, we let $a = 1+[D_0^{1/2}]$ or $2+[D_0^{1/2}]$, so that a is odd. Then*

$$\tfrac{1}{2}(a+D_0^{1/2}) = ((a, b_1, \ldots, b_{p-1})), \quad b_t = b_{p-t}. \tag{13.13}$$

If $D_0 \not\equiv 1 \pmod 4$, let $a = 1+[D_0^{1/2}]$. Then

$$(a+D_0^{1/2}) = ((2a, b_1, \ldots, b_{p-1})), \quad b_t = b_{p-t}. \tag{13.14}$$

This follows from Theorem 13.1b (or (8.2) and (8.3) directly).

14. Formal ring of a sector. We begin with the "canonical" construction of the monoid ring of a sector. We enlarge the semigroup SGVS of the sector S to a monoid by the neutral vector $(0, 0)$. Let $(x, y) \mapsto \boldsymbol{v}$, $(x', y') \mapsto \boldsymbol{v}'$. Then we define

$$\boldsymbol{v}\boldsymbol{v}' = (x+x', y+y') \tag{14.1}$$

while addition $\boldsymbol{v}+\boldsymbol{v}'$ is formal. This creates the monoid ring (over $\boldsymbol{Z}$) $\boldsymbol{R}_{S,\boldsymbol{Z}}$. We focus our attention on the field $\boldsymbol{C}$ and define

$$\boldsymbol{R}_S = \boldsymbol{R}_{S,\boldsymbol{Z}} \otimes \boldsymbol{C}. \tag{14.2}$$

DEFINITION 14.3. Call $\boldsymbol{R}_S$ the *ring of the sector S.*

Remark 14.4. When S is the sector for (9.3), then $\boldsymbol{R}_S$ is the local ring at the origin for $C\big(z_1, z_2, (z_1/z_2^S)^{1/R}\big)$.

DEFINITION 14.5. Call $\hat{\boldsymbol{R}}_S$ the *formal ring of the sector S* the completion of $\boldsymbol{R}_S$ through the use of infinite formal sums of vectors $\boldsymbol{v}$ with coefficients in $\boldsymbol{C}$. Thus $\hat{\boldsymbol{R}}_S = \boldsymbol{C}[[\boldsymbol{v}].]$

15. Quotient ring of a sector. In general a sector S will have automorphisms forming a subgroup $G_S \subseteq \mathrm{GL}_2(\boldsymbol{Z})$. Let G be some subgroup of G_S.

DEFINITION 15.1. Call the *formal quotient ring of S/G*

$$\hat{\boldsymbol{R}}_{S/G} = \boldsymbol{C}[[\boldsymbol{v}^G]] \tag{15.2}$$

where $\boldsymbol{v}^G$ denotes the formal sum of the actions of G on $\boldsymbol{v}$, i.e.,

$$\boldsymbol{v}^G = \sum \boldsymbol{v}^g, \quad g \in G. \tag{15.3}$$

(Here the action $g(x, y) = (x, y)^g \mapsto \boldsymbol{v}^g$ in the obvious way.)

Remark 15.4. It can be seen that $\hat{\boldsymbol{R}}_{S/G}$ is the subset of $\hat{\boldsymbol{R}}_S$ invariant under G. Thus a *formal* ring is needed when G in infinite.

A simple, yet nontrivial case occurs when S is defined by

$$|y| \leqslant \lambda x, \quad x > 0,\ 0 < \lambda\ (\in \boldsymbol{R}). \tag{15.5}$$

If we take $G = \{1, g\}$, $g(x, y) = (x, -y)$ we obtain

$$\boldsymbol{v}^G = \boldsymbol{v}+\boldsymbol{v}^g. \tag{15.6}$$

The relationship of $\hat{R}_{S/G}$ to $\hat{R}_S$ is similar to the quadratic extension of a ring, but this matter will be explored in a later paper.

16. Formal ring of a quadratic sector. Following the work of Siegel [5] and Gundlach [2], we consider the spaces H_+ and H_- defined by the product of half-planes

(16.1) $$\operatorname{Im} z_1 > 0, \quad d \operatorname{Im} z_2 > 0$$

where $d = +1$ for H_+ and $d = -1$ for H_-. Let either of $H_\pm$ be acted upon by a group denoted by $\{L, U\}$ as follows: Here

(16.2) $$L = [\beta_1, \beta_2]$$

a two-dimensional $\boldsymbol{Z}$-lattice of elements of $\boldsymbol{Q}(D_0^{1/2})$, for $D_0 > 0$ square-free; and U is a unit group generated by a totally positive unit ω_0 (> 1) such that

(16.3) $$\omega_0 L = L.$$

According to the classical theory, we can multiply L by a factor so as to make it an ideal in a quadratic ring for which ω_0 is a unit, for convenience. The actions on $(z_1, z_2) \in H_\pm$ are generated by

(16.4) $$T(z_1, z_2) = (z_1 + \beta, z_2 + \beta'),$$

(16.5) $$U(z_1, z_2) = (\omega_0 z_1, \omega_0' z_2)$$

where $\beta \in L$ (and primes denote conjugates). We orient the bases so that

(16.6) $$\beta_2\beta_1' - \beta_1\beta_2' = \Delta > 0.$$

We then consider the ring of functions of z_1, z_2 holomorphic in a neighborhood of infinity defined by

(16.7) $$|(\operatorname{Im} z_1)(\operatorname{Im} z_2)| > \text{const}$$

and this ring can later be completed to a formal ring. For the present, define new variables

(16.8) $$\Delta\zeta_1 = -\beta_2' z_1 + \beta_2 z_2, \quad \Delta\zeta_2 = \beta_1' z_1 - \beta_1 z_2$$

so that these correspondances hold:

(16.9) $$(z_1, z_2) \leftrightarrow (\zeta_1, \zeta_2),$$

(16.10) $$(z_1 + \beta_1, z_2 + \beta_1') \leftrightarrow (\zeta_1 + 1, \zeta_2),$$

(16.11) $$(z_1 + \beta_2, z_2 + \beta_2') \leftrightarrow (\zeta_1, \zeta_2 + 1).$$

Thus we are dealing with a ring of functions of

(16.12) $$\mu(m_1, m_2) = \exp 2\pi i (m_1\zeta_1 + m_2\zeta_2) = \exp 2\pi i (z_1\varrho' - z_2\varrho)/\Delta.$$

Here ϱ is again a general element of L,

(16.13) $$\varrho = -m_1\beta_2 + m_2\beta_1.$$

For boundedness of μ on $H_\pm$ we must have $\varrho' \geqslant 0$, $d\varrho \geqslant 0$ ($d = \pm 1$). This leads to the sector in the (m_1, m_2) coordinates

(16.14) $$S_\pm \begin{cases} -m_1\beta_2' + m_2\beta_1' > 0, \\ \pm(-m_1\beta_2 + m_2\beta_1) > 0. \end{cases}$$

The (monoid of) SGVS$_\pm$ has an isomorphic image in $\mu(m_1, m_2)$ under multiplication and the ring of holomorphic functions on $H_\pm/L$ is canonically imbedded in the formal ring of $S_\pm$.

The unit group U corresponds to periodic translations of the support polygons, indeed isomorphically (by Theorem 11.1) if ω_0 is a fundamental unit for the ring into which L is injected by a proportionality factor.

The ring of holomorphic functions of $H_\pm/\{L, U\}$ is also canonically imbedded in $\hat{R}_{S/\{L,U\}}$, but the fact that the imbedding is injective is not trivial. It involves convergence of the series (15.3), under $\boldsymbol{v} \leftrightarrow \mu(m_1, m_2)$ (see Gundlach [2]).

17. Conjectures of Serre and Hirzebruch. In 1969 Serre conjectured (private communications) that if U is the periodic (translation) group of a quadratic sector S then $\hat{R}_{S/U}$ is an algebraic formal ring of degree of transcendence 2.

This was proved by Hirzebruch in 1971 [4], at least with regard to the subring of holomorphic functions, by the construction of a cyclic covering of the singularity of $H_\pm/\{L, U\}$ at ∞. This covering was based on the period of the intersection numbers.

A direct proof should be possible, at least for the formal rings, based only on number-theoretic manipulations. Unfortunately this has been achieved only in a few cases (see [1], for instance).

A more extended study of the formal rings and quotients is planned for later papers, however, as well as the generalization of the geometric concept of support polygons to several dimensions.

Added in proof: The finiteness of the formal ring $\hat{R}_{S/U}$ is demonstrated by the author in a forthcoming paper. The author was unaware that the equivalence of chains and units (Theorem 11.1) had been proved by Dr. Rohlfs in 1971. (No published reference is available.)

References

[1] H. Cohn, *Some explicit resolutions of modular cusp singularities*, Math. Ann. 198 (1972), pp. 123–130.

[2] K.-B. Gundlach, *Some new results in the theory of Hilbert's modular group*, Contributions to function theory, International Colloquium, Bombay 1960, pp. 165–180.

[3] F. Hirzebruch, *Über vierdimensionale Riemannsche Flächen mehrdeutiger analytischer Funktionen von zwei komplexen Veränderlichen*, Math. Ann. 126 (1953), pp. 1–22.

[4] — *The Hilbert modular group, resolution of the singularities at the cusps and related problems*, Sem. Bourbaki ⋕ 396 (1971/2), pp. 1–14.

[5] C. L. Siegel, *Einführung in die Theorie der Modulfunktionen n-ten Grades*, Math. Ann. 116 (1939), pp. 617–657.

[6] — *Symplectic geometry*, Amer. Journ. Math. 65 (1943), pp. 1–86.

[7] — *Geometry of Numbers*, Lecture notes by B. Friedman, New York University, 1945.

MATHEMATICS DEPARTMENT
CITY COLLEGE OF NEW YORK
New York, N. Y.

Received on 18. 10. 1972 (338)

ACTA ARITHMETICA
XXIV (1973)

Specialization of quadratic and symmetric bilinear forms, and a norm theorem

by

MANFRED KNEBUSCH (Saarbrücken)

Dedicated to Carl Ludwig Siegel on his 75 birthday

Introduction. In the first part of this paper (§ 1–§ 3) we study the specialization of a symmetric bilinear or quadratic form over a field K with respect to a place $\lambda\colon K \to L \cup \infty$, provided the form has "good reduction". We have to distinguish between symmetric bilinear and quadratic forms since we do not exclude fields of characteristic 2. A typical result obtained by this theory is the following: We denote a symmetric bilinear form by the corresponding symmetric matrix of its coefficients. Let $k(t)$ be the field of rational functions in independent variables $t_1, \ldots, t_r$ over a field k. Consider symmetric bilinear forms $(f_{ij}(t))$, $(g_{kl}(t))$ over $k(t)$ whose coefficients $f_{ij}(t)$, $g_{kl}(t)$ are polynomials. Assume that the form $(g_{kl}(t))$ is represented by $(f_{ij}(t))$. Assume further that c is an r-tupel in k^r such that the form $(f_{ij}(c))$ over k is non singular. If $\operatorname{char} k \neq 2$ the following holds true:

(i) If also $(g_{kl}(c))$ is non singular, then this form is represented by $(f_{ij}(c))$ over k (see § 2).

(ii) If $(g_{kl}(t))$ is a diagonal matrix with m rows and columns and if c is a non singular zero of each polynomial $g_{kk}(t)$, then the form $(f_{ij}(c))$ has Witt index $\geqslant m/2$ if m is even and $\geqslant (m+1)/2$ if m is odd (see § 3).

The assertion (i) may be considered as a generalization of the principle of substitution of Cassels and Pfister ([15], p. 365; [10], p. 20). At the end of Section 3 (Proposition 3.6) we shall also generalize the subform theorem of Cassels and Pfister ([15], p. 366; [10], p. 20).

Using the result quoted above and a similar result for $\operatorname{char} k = 2$ we prove in the last section § 4 a theorem about the polynomials in $k[t]$ which can occur as norms of similarity over $k(t)$ for a fixed symmetric bilinear form defined over k. Special cases of this norm theorem have been used in a crucial way by Arason and Pfister in [1] and by Elman and Lam in [5].