8. Tools from the Geometry of Numbers

numbers ω such that there are infinitely many polynomials P with rational integer coefficients of degree $\leqq d$ and with

$$0 < |P(\alpha)| < H(P)^{-\omega}.$$

By Corollary 6D it is clear that $\omega_d \geqq d$ unless α is algebraic of degree $\leqq d$. Furthermore if α is algebraic of degree n, then one can show using the norm of $P(\alpha)$ that $\omega_d \leqq n - 1$ $(d=1, 2, ...)$. Thus Mahler could characterize the algebraic numbers α by the property that $\omega_d(\alpha)$ $(d=1, 2, ...)$ remains bounded.

Koksma (1939) defines $\omega_d^* = \omega_d^*(\alpha)$ as the supremum of the numbers ω^* such that there are infinitely many algebraic numbers β of degree $\leqq d$ with

$$|\alpha - \beta| < H(\beta)^{-1-\omega^*}.$$

It is easy to see that $\omega_d^* \leqq \omega_d$ and Wirsing (1961) showed that $\omega_d^* \geqq \frac{1}{2}(\omega_d + 1)$ if α is transcendental. Hence the algebraic numbers can also be characterized by the property that $\omega_d^*(\alpha)$ $(d=1, 2, ...)$ is bounded. We have $\omega_d^* \leqq \omega_d \leqq n - 1$ if α is algebraic of degree n, and the results of the last section show that $\omega_d^* = d$ if $d \leqq n - 1$. Since ω_d^* and ω_d increase with d, we have for algebraic α of degree n,

$$\omega_d = \omega_d^* = \begin{cases} d & \text{if } d \leqq n-1 \\ n-1 & \text{if } d \geqq n. \end{cases}$$

Thus the exponent in Theorem 7H is best possible precisely if $d < n$.

Another characterization of algebraic numbers by approximation properties was given by Gelfond (1952, §III.4, Lemma VII) and refined by Lang (1965a) and Tijdeman (1971, Lemma 6). This lemma was slightly improved by D. Brownawell (unpublished).

8. Tools from the Geometry of Numbers

8.1. To prove the theorems enunciated in the last section one needs certain results from the Geometry of Numbers. This field was first investigated under this name by Minkowski (1896). Other books on the Geometry of Numbers are Cassels (1959) and Lekkerkerker (1969).

Let K be a symmetric [1]) convex set in Euclidean E^n. For convenience let us assume that K is compact and has a non-empty interior. For $\lambda > 0$ let λK be the set consisting of the points $\lambda \mathbf{x}$ with $\mathbf{x} \in K$. Minkowski defines

[1]) I.e. if $\mathbf{x} \in K$, then also $-\mathbf{x} \in K$.

the *first minimum* λ_1 as the least positive value of λ such that λK contains an integer point $\mathbf{x} \neq \mathbf{0}$. More generally, for $1 \leqq j \leqq n$, the j-th minimum λ_j is the least positive value of λ such that λK contains j linearly independent integer points. It is clear that $0 < \lambda_1 \leqq \lambda_2 \leqq \ldots \leqq \lambda_n < \infty$, and that there are linearly independent integer points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with

$$\mathbf{x}_j \in \lambda_j K \qquad (j = 1, \ldots, n) . \tag{8.1}$$

Minkowski's Theorem 6H is easily seen to be equivalent with the inequality

$$\lambda_1^n V(K) \leqq 2^n .$$

Later Minkowski could refine this to the much stronger

THEOREM 8A (Minkowski's Theorem on Successive Minima).

$$2^n/n\,! \leqq \lambda_1 \ldots \lambda_n V(K) \leqq 2^n . \tag{8.2}$$

Like Theorem 6H this result can be generalized to arbitrary lattices Λ, and then (8.2) is to be replaced by

$$d(\Lambda)\, 2^n/n\,! \leqq \lambda_1 \ldots \lambda_n V(K) \leqq d(\Lambda)\, 2^n . \tag{8.3}$$

Of particular interest to us will be the situation when $L_1(\mathbf{x}), \ldots, L_n(\mathbf{x})$ are linearly independent linear forms and $R_1, \ldots, R_n$ are positive numbers, and when K is the parallelepiped defined by [1])

$$|L_i(\mathbf{x})| \leqq R_i \qquad (i = 1, \ldots, n) . \tag{8.4}$$

In the special case when $R_1 \ldots R_n = 1$ and when $|\det(L_1, \ldots, L_n)| = \Delta$, say, we have $V(K) = 2^n/\Delta$, whence $\Delta/n! \leqq \lambda_1 \ldots \lambda_n \leqq \Delta$. In particular we have

$$1 \ll \lambda_1 \ldots \lambda_n \ll 1 , \tag{8.5}$$

where the notation $A \ll B$ means that $A \leqq cB$ with $c = c(n, \Delta)$. Later on the notation $A \gg \ll B$ will mean that both $A \ll B$ and $B \ll A$.

8.2. We shall need three so-called " transference theorems " which relate the successive minima of certain parallelepipeds to the successive minima of other parallelepipeds.

---

[1]) The case when $R_1 = \ldots = R_n = 1$ is just as general, but the factors $R_1, \ldots, R_n$ will be convenient for later applications.

THEOREM 8B (" Davenport's Lemma " (Davenport, 1937)). *Let* $\lambda_1, \ldots, \lambda_n$ *be the successive minima of the parallelepiped* Π *given by* (8.4). *Let* $\rho_1, \ldots, \rho_n$ *be numbers with*

$$\rho_1 \geqq \rho_2 \geqq \ldots \geqq \rho_n > 0 \quad \text{and} \quad \rho_1\lambda_1 \leqq \ldots \leqq \rho_n\lambda_n .$$

Then there is a permutation $(t_1, t_2, \ldots, t_n)$ *of* $(1, 2, \ldots, n)$ *such that the successive minima* $\lambda_1', \ldots, \lambda_n'$ *of the new parallelepiped* Π' *given by*

$$| L_i(\mathbf{x}) | \leqq R_i\rho_{t_i}^{-1} \qquad (i = 1, \ldots, n)$$

satisfy

$$\lambda_j' \gg \ll \rho_j\lambda_j \qquad (j = 1, \ldots, n) . \tag{8.6}$$

Moreover, let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *be linearly independent integer points with* (8.1), *i.e. with* $R_i^{-1} \, | \, L_i(\mathbf{x}_j) \, | \leqq \lambda_j$ $(i, j = 1, \ldots, n)$. *Let* T_0 *be the subspace consisting of* $\mathbf{0}$, *and for* $1 \leqq j \leqq n$ *let* T_j *be the subspace spanned by* $\mathbf{x}_1, \ldots, \mathbf{x}_j$. *Then every integer point* $\mathbf{x}$ *outside the subspace* T_{j-1} *where* $1 \leqq j \leqq n$ *satisfies*

$$\max \left(R_1^{-1}\rho_{j_1} \, | \, L_1(\mathbf{x}) \, |, \ldots, R_n^{-1}\rho_{j_n} \, | \, L_n(\mathbf{x}) \, | \right) \gg \lambda_j' .$$

Note that the ratios of $\rho_1\lambda_1, \ldots, \rho_n\lambda_n$ are equal to or smaller than the ratios of $\lambda_1, \ldots, \lambda_n$, so that the successive minima have been " pushed closer together ". Usually in transference theorems only inequalities such as (8.6) are given. But the last statement of the theorem will also be needed.

8.3. Every linear form $L(\mathbf{x})$ is of the type $L(\mathbf{x}) = \mathbf{ax}$ where $\mathbf{a}$ is a fixed vector and where $\mathbf{ax}$ denotes the inner product. Now suppose that $L_1(\mathbf{x}), \ldots, L_n(\mathbf{x})$ are linearly independent linear forms. Then if $L_i(\mathbf{x}) = \mathbf{a}_i\mathbf{x}$ $(i = 1, \ldots, n)$, the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are linearly independent. There are unique vectors $\mathbf{a}_1^*, \ldots, \mathbf{a}_n^*$ with

$$\mathbf{a}_i\mathbf{a}_j^* = \delta_{ij} = \begin{cases} 1 \text{ if } i = j , \\ 0 \text{ otherwise} . \end{cases}$$

The linear forms $L_1^*, \ldots, L_n^*$ given by $L_i^*(\mathbf{x}) = \mathbf{a}_i^*\mathbf{x}$ $(i = 1, \ldots, n)$ are called *dual* to $L_1, \ldots, L_n$; they satisfy the identity $L_1(\mathbf{x})\, L_1^*(\mathbf{y}) + \ldots + L_n(\mathbf{x})\, L_n^*(\mathbf{y}) = \mathbf{xy}$. The dual linear forms are again linearly independent, and they have determinant 1 if $L_1, \ldots, L_n$ have determinant 1. The parallelepiped

$$\Pi^* : \; | L_i^*(\mathbf{x}) | \leqq R_i^{-1} \qquad (i = 1, \ldots, n)$$

is called the *dual* of the parallelepiped Π defined by (8.4).

Remark. One can define the *polar* set of any convex symmetric set, and the dual of a parallelepiped is closely related to its polar set. But the polar set of a parallelepiped has the disadvantage that it need not be a parallelepiped.

THEOREM 8C (Mahler 1939). *Let $\lambda_1, \ldots, \lambda_n$ and $\lambda_1^*, \ldots, \lambda_n^*$ be the successive minima of a parallelepiped Π and of its dual Π^*, respectively. Then*

$$\lambda_j^* \gg \ll \lambda_{n+1-j}^{-1} \qquad (j=1, \ldots, n) .$$

Moreover, if $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are linearly independent points with (8.1), *i.e. with $|L_i(\mathbf{x}_j)| \leqq \lambda_j R_i$ $(i, j=1, \ldots, n)$, and if $\mathbf{x}_1^*, \ldots, \mathbf{x}_n^*$ are defined by $\mathbf{x}_i \mathbf{x}_j^* = \delta_{ij}$ $(i, j=1, \ldots, n)$, then*

$$|L_i^*(\mathbf{x}_{n+1-j}^*)| \ll \lambda_j^* R_i^{-1} \qquad (i, j=1, \ldots, n) . \tag{8.7}$$

8.4. Suppose $1 \leqq p \leqq n$ and put $l = \binom{n}{p}$. Vectors in E^n will be denoted as usual by $\mathbf{a}, \mathbf{b}, \ldots$, and vectors in E^l will be denoted by $\mathbf{A}, \mathbf{B}, \ldots$. By

$$\mathbf{a}_1 \wedge \ldots \wedge \mathbf{a}_p$$

we shall denote the exterior product of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_p$, i.e. the vector in E^l whose coordinates are the $(p \times p)$-determinants formed from the matrix with rows $\mathbf{a}_1, \ldots, \mathbf{a}_p$, and arranged in lexicographic order. For example if $n = 4$ and $p = 2$, then $l = 6$, and if $\mathbf{a} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$, $\mathbf{b} = (\beta_1, \beta_2, \beta_3, \beta_4)$, then

$$\mathbf{a} \wedge \mathbf{b} = \left(\begin{vmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \end{vmatrix}, \begin{vmatrix} \alpha_1 & \alpha_3 \\ \beta_1 & \beta_3 \end{vmatrix}, \begin{vmatrix} \alpha_1 & \alpha_4 \\ \beta_1 & \beta_4 \end{vmatrix}, \begin{vmatrix} \alpha_2 & \alpha_3 \\ \beta_2 & \beta_3 \end{vmatrix}, \begin{vmatrix} \alpha_2 & \alpha_4 \\ \beta_2 & \beta_4 \end{vmatrix}, \begin{vmatrix} \alpha_3 & \alpha_4 \\ \beta_3 & \beta_4 \end{vmatrix} \right) .$$

Let $C(n, p)$ be the set of all p-tuples of integers $i_1, \ldots, i_p$ with $1 \leqq i_1 < \ldots < i_p \leqq n$. There are l such p-tuples.

Now suppose that $L_1(\mathbf{x}) = \mathbf{a}_1 \mathbf{x}, \ldots, L_n(\mathbf{x}) = \mathbf{a}_n \mathbf{x}$ are independent linear forms. For $\sigma = \{ i_1, \ldots, i_p \}$ in $C(n, p)$, let $\mathbf{A}_\sigma$ be the vector

$$\mathbf{A}_\sigma = \mathbf{a}_{i_1} \wedge \ldots \wedge \mathbf{a}_{i_p} .$$

Let $L_\sigma^{(p)}$ be the linear form in E^l defined by $L_\sigma^{(p)}(\mathbf{X}) = \mathbf{A}_\sigma \mathbf{X}$. The l linear forms $L_\sigma^{(p)}$ with $\sigma \in C(n, p)$ are again linearly independent, and they have determinant 1 if $L_1, \ldots, L_n$ have determinant 1. Let $R_1, \ldots, R_n$ be positive constants with $R_1 R_2 \ldots R_n = 1$ and define R_σ by $R_\sigma = \prod_{i \in \sigma} R_i$. The inequalities

$$|L_\sigma^{(p)}(\mathbf{X})| \leqq R_\sigma \qquad (\sigma \in C(n, p))$$

define a parallelepiped $\Pi^{(p)}$ in E^l which we shall call the p-th *pseudocompound* of the parallelepiped Π defined by (8.4).

Remarks. Mahler (1955) defined the p-th *compound* of any symmetric convex set, and the pseudocompound of a parallelepiped is closely related to its compound. But the compound of a parallelepiped is not necessarily a parallelepiped. Except for the notation, the $(n-1)$-st pseudocompound is the same as the dual of a parallelepiped, and hence the results of the last subsection may be interpreted as special cases of the results of the present subsection.

THEOREM 8D (Mahler 1955). *Let $\lambda_1, \ldots, \lambda_n$ and $v_1, \ldots, v_l$ be the successive minima of a parallelepiped Π and of its p-th pseudocompound $\Pi^{(p)}$, respectively. For $\sigma \in C(n,p)$ put $\lambda_\sigma = \prod_{i\in\sigma} \lambda_i$ and order the elements of $C(n,p)$ as $\sigma_1, \ldots, \sigma_l$ such that $\lambda_{\sigma_1} \leqq \ldots \leqq \lambda_{\sigma_l}$. Then*

$$v_j \gg \ll \lambda_{\sigma_j} \qquad (j=1, \ldots, l).$$

Moreover, if $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are linearly independent integer points with (8.1), i.e. with $|L_i(\mathbf{x}_j)| \leqq \lambda_j R_i$ $(i,j=1, \ldots, n)$, and if for $\tau = \{j_1, \ldots, j_p\}$ in $C(n,p)$ we put $\mathbf{X}_\tau = \mathbf{x}_{j_1} \wedge \ldots \wedge \mathbf{x}_{j_p}$, then

$$|L_\sigma^{(p)}(\mathbf{X}_\tau)| \ll \lambda_\tau R_\sigma \qquad (\sigma, \tau \in C(n,p)).$$

9. OUTLINE OF THE PROOF OF THE THEOREMS ON SIMULTANEOUS APPROXIMATION TO ALGEBRAIC NUMBERS

9.1. Let us see what happens if we try to generalize Roth's proof to prove, say, Corollary 7B. In Roth's proof we constructed a polynomial $P(x_1, \ldots, x_m)$ in m variables $x_1, \ldots, x_m$ which had a zero of high order at $(\alpha, \ldots, \alpha)$. Hence the natural thing to try would be

(a) to construct a polynomial $P(x_{11}, \ldots, x_{1l}; \ldots; x_{m1}, \ldots, x_{ml})$ in ml variables of total degree $\leqq r_h$ in each block of variables $x_{h1}, \ldots, x_{hl}$ $(h = 1, \ldots, m)$ with a zero of high order at $(\alpha_1, \ldots, \alpha_l; \ldots; \alpha_1, \ldots, \alpha_l)$. Then

(b) one would have to show that if each of m given rational l-tuples $\left(\frac{p_{h1}}{q_h}, \ldots, \frac{p_{hl}}{q_h}\right)$ $(h=1, \ldots, m)$ satisfies (7.2), then P also has a zero of high order at