AN INTRODUCTION TO THE GEOMETRY OF NUMBERS

BY

J. W. S. CASSELS

AN INTRODUCTION TO THE GEOMETRY OF NUMBERS

BY

J. W. S. CASSELS

FELLOW OF TRINITY COLLEGE, CAMBRIDGE

WITH 12 FIGURES

SPRINGER-VERLAG
BERLIN · GOTTINGEN · [redacted]ELBERG
1959

PRINTED IN GERMANY

DRUCK DER UNIVERSITÄTSDRUCKEREI H. STÜRTZ AG., WÜRZBURG

Preface

Of making many bookes there is no end, and much studie is a wearinesse of the flesh.

Ecclesiastes XII, 12.

When I first took an interest in the Geometry of Numbers, I was struck by the absence of any book which gave the essential skeleton of the subject as it was known to the experienced workers in the subject. Since then the subject has developed, as will be clear from the dates of the papers cited in the bibliography, but the need for a book remains. This is an attempt to fill the gap. It aspires to acquaint the reader with the main lines of development, so that he may with ease and pleasure follow up the things which interest him in the periodical literature. I have attempted to make the account as self-contained as possible.

References are usually given to the more recent papers dealing with a particular topic, or to those with a good bibliography. They are given only to enable the reader to amplify the account in the text and are not intended to give a historical picture. To give anything like a reasonable account of the history of the subject would have involved much additional research.

I owe a particular debt of gratitude to Professor L. J. Mordell, who first introduced me to the Geometry of Numbers.

The proof-sheets have been read by Professors K. Mahler, L. J. Mordell and C. A. Rogers. It is a pleasure to acknowledge their valuable help and advice both in detecting errors and obscurities and in suggesting improvements. Dr. V. Ennola has drawn my attention to several slips which survived into the second proofs.

I should also like to take the opportunity to thank Professor F. K. Schmidt and the Springer-Verlag for accepting this book for their celebrated yellow series and the Springer-Verlag for its readiness to meet my typographical whims.

Cambridge, June, 1959 J. W. S. Cassels

Notation

An effort has been made to distinguish different types of mathematical object by the use of different alphabets. It is not necessary to describe the scheme in full since an acquaintance with it is not presupposed. However the following conventions are made throughout the book without explicit mention.

Bold Latin letters (large and small) always denote vectors. The dimensions is n, unless the contrary is explicitly stated: and the letter n is not used otherwise, except in one or two places where there can be no fear of ambiguity. The co-ordinates of a vector are denoted by the corresponding italic letter with a suffix $1, 2, \ldots, n$. If the bold letter denoting the vector already has a suffix, then that is put after the co-ordinate suffix. Thus:

$$\boldsymbol{a} = (a_1, \ldots, a_n)$$

$$\boldsymbol{b}_r = (b_{1r}, \ldots, b_{nr})$$

$$\boldsymbol{X}'_\varepsilon = (X'_{1\varepsilon}, \ldots, X'_{n\varepsilon}).$$

The origin is always denoted by $\boldsymbol{o}$. The length of $\boldsymbol{x}$ is

$$|\boldsymbol{x}| = (x_1^2 + \cdots + x_n^2)^{\frac{1}{2}}.$$

Sanserif Greek capitals, in particular $\mathsf{\Lambda}, \mathsf{M}, \mathsf{N}, \mathsf{\Gamma}$, denote lattices.

The notation $d(\mathsf{\Lambda})$, $\Delta(\mathscr{S})$, $V(\mathscr{S})$ for respectively the determinant of the lattice $\mathsf{\Lambda}$ and for the lattice-constant and volume of a set $\mathscr{S}$ will be standard, once the corresponding concepts have been introduced.

Chapters are divided into sections with titles. These sections are subdivided, for convenience, into subsections, which are indicated by a decimal notation. The numbering of displayed formulae starts afresh in each subsection. The prologue is just subdivided into sections without titles, and it was convenient to number the displayed formulae consecutively throughout.

Contents

Page

Notation VIII

Prologue 1

Chapter I. Lattices 9
1. Introduction 9
2. Bases and sublattices 9
3. Lattices under linear transformation 19
4. Forms and lattices 20
5. The polar lattice 23

Chapter II. Reduction 26
1. Introduction 26
2. The basic process 27
3. Definite quadratic forms 30
4. Indefinite quadratic forms 35
5. Binary cubic forms 51
6. Other forms 60

Chapter III. Theorems of Blichfeldt and Minkowski 64
1. Introduction 64
2. Blichfeldt's and Minkowski's theorems 68
3. Generalisations to non-negative functions 73
4. Characterisation of lattices 78
5. Lattice constants 80
6. A method of Mordell 84
7. Representation of integers by quadratic forms 98

Chapter IV. Distance functions 103
1. Introduction 103
2. General distance-functions 105
3. Convex sets 108
4. Distance functions and lattices 119

Chapter V. Mahler's compactness theorem 121
1. Introduction 121
2. Linear transformations 122
3. Convergence of lattices 126
4. Compactness for lattices 134
5. Critical lattices 141
6. Bounded star-bodies 145
7. Reducibility 152
8. Convex bodies 155
9. Spheres 163
10. Applications to diophantine approximation 165

Chapter VI. The theorem of Minkowski-Hlawka 175
1. Introduction 175
2. Sublattices of prime index 178

Page

3. The Minkowski-Hlawka theorem 181
4. Schmidt's theorems 184
5. A conjecture of Rogers 187
6. Unbounded star-bodies 189

Chapter VII. The quotient space 194
1. Introduction 194
2. General properties 194
3. The sum theorem 198

Chapter VIII. Successive minima 201
1. Introduction 201
2. Spheres 205
3. General distance-functions 207
4. Convex sets 213
5. Polar convex bodies 219

Chapter IX. Packings 223
1. Introduction 223
2. Sets with $V(\mathscr{S}) = 2^n \Delta(\mathscr{S})$ 228
3. Voronoi's results 231
4. Preparatory lemmas 235
5. Fejes Tóth's theorem 240
6. Cylinders 245
7. Packing of spheres 246
8. The product of n linear forms 250

Chapter X. Automorphs 256
1. Introduction 256
2. Special forms 266
3. A method of Mordell 268
4. Existence of automorphs 279
5. Isolation theorems 286
6. Applications of isolation 295
7. An infinity of solutions 298
8. Local methods 301

Chapter XI. Inhomogeneous problems 303
1. Introduction 303
2. Convex sets 309
3. Transference theorems for convex sets 313
4. The product of n linear forms 322

Appendix 332

References 334

Index 343

Prologue

P1. We owe to MINKOWSKI the fertile observation that certain results which can be made almost intuitive by the consideration of figures in n-dimensional euclidean space have far-reaching consequences in diverse branches of number theory. For example, he simplified the theory of units in algebraic number fields and both simplified and extended the theory of the approximation of irrational numbers by rational ones (Diophantine Approximation). This new branch of number theory, which MINKOWSKI christened "The Geometry of Numbers", has developed into an independent branch of number-theory which, indeed, has many applications elsewhere but which is well worth studying for its own sake.

In this prologue we first discuss some of the concepts and results which will play a leading rôle. The arguments we shall use are sometimes rather different from those in the main body of the text: since here we wish to make the geometrical situation intuitive in simple cases without necessarily giving complete proofs, while later we may need to sacrifice picturesqueness for precision. The proofs in the text are independent of this prologue, which may be omitted if desired.

P2. A fundamental and typical problem in the geometry of numbers is as follows:

Let $f(x_1, \ldots, x_n)$ be a real-valued function of the real variables $x_1, \ldots, x_n$. How small can $|f(u_1, \ldots, u_n)|$ be made by suitable choice of the integers $u_1, \ldots, u_n$? It may well be that one has trivially $f(0, \ldots, 0) = 0$, for example when $f(x_1, \ldots, x_n)$ is a homogeneous form; and then one excludes the set of values $u_1 = u_2 = \cdots = u_n = 0$. (The "homogeneous problem".)

In general one requires estimates which are valid not merely for individual functions f but for whole classes of functions. Thus a typical result is that if

$$f(x_1, x_2) = a_{11} x_1^2 + 2a_{12} x_1 x_2 + a_{22} x_2^2 \tag{1}$$

is a positive definite quadratic form, then there are integers u_1, u_2 not both 0 such that

$$f(u_1, u_2) \leqq (4D/3)^{\frac{1}{2}}, \tag{2}$$

where

$$D = a_{11} a_{22} - a_{12}^2$$

is the discriminant of the form. It is trivial that if the result is true then it is the best possible of its kind, since

$$u_1^2 + u_1 u_2 + u_2^2 \geqq 1$$

for all pairs of integers u_1, u_2 not both zero; and here $D = \frac{3}{4}$.

Of course the positive definite binary quadratic forms are a particularly simple case. The result above was known well before the birth of the Geometry of Numbers; and indeed we shall give a proof substantially independent of the Geometry of Numbers in Chapter II, § 3. But positive definite binary quadratic forms display a number of arguments in a particularly simple way so we shall continue to use them as examples.

P3. The result just stated could be represented graphically. An inequality of the type

$$f(x_1, x_2) \leqq k,$$

where $f(x_1, x_2)$ is given by (1) and k is some positive number, represents the region $\mathscr{R}$ bounded by an ellipse in the (x_1, x_2)-plane. Thus our result above states that $\mathscr{R}$ contains a point (u_1, u_2), other than the origin, with integer coordinates provided that $k \geqq (4D/3)^{\frac{1}{2}}$.

A result of this kind but not so precise follows at once from a fundamental theorem of Minkowski. The 2-dimensional case of this states that a region $\mathscr{R}$ always contains a point (u_1, u_2) with integral co-ordinates other than the origin provided that it satisfies the following three conditions.

(i) $\mathscr{R}$ is symmetric about the origin, that is if (x_1, x_2) is in $\mathscr{R}$ then so is $(-x_1, -x_2)$.

(ii) $\mathscr{R}$ is convex, that is if (x_1, x_2) and (y_1, y_2) are two points of $\mathscr{R}$ then the whole line segment

$$\{\lambda x_1 + (1-\lambda) y_1, \ \lambda x_2 + (1-\lambda) y_2\} \qquad (0 \leqq \lambda \leqq 1)$$

joining them is also in $\mathscr{R}$.

(iii) $\mathscr{R}$ has area greater than 4.

Any ellipse $f(x_1, x_2) \leqq k$ satisfies (i) and (ii). Since its area is

$$\frac{k\pi}{(a_{11}a_{22} - a_{12}^2)^{\frac{1}{2}}} = \frac{k\pi}{D^{\frac{1}{2}}},$$

it also satisfies (iii), provided that $k\pi > 4D^{\frac{1}{2}}$. We thus have a result similar to (2), except that the constant $(\frac{4}{3})^{\frac{1}{2}}$ is replaced by any number greater than $4/\pi$.

P4. It is useful to consider briefly the basic ideas behind the proof of Minkowski's theorem, since in the formal proofs in Chapter 3 they may be obscured by the need to obtain powerful theorems which are as widely applicable as possible. Instead of the region $\mathscr{R}$, MINKOWSKI works with the region $\mathscr{S} = \frac{1}{2}\mathscr{R}$ of points $(\frac{1}{2}x_1, \frac{1}{2}x_2)$, where (x_1, x_2) is in $\mathscr{R}$. Thus $\mathscr{S}$ is symmetric about the origin and convex: its area is $\frac{1}{4}$ that of $\mathscr{R}$ and so is greater than 1. More generally, MINKOWSKI considers the set of bodies $\mathscr{S}(u_1, u_2)$ similar and similarly situated to $\mathscr{S}$ but with centres at the points (u_1, u_2) with integer co-ordinates.

We note first that if $\mathscr{S}$ and $\mathscr{S}(u_1, u_2)$ overlap then[1] (u_1, u_2) is in $\mathscr{R}$. For let a point of overlap be (ξ_1, ξ_2). Since (ξ_1, ξ_2) is in $\mathscr{S}(u_1, u_2)$ the point $(\xi_1 - u_1, \xi_2 - u_2)$ must be in $\mathscr{S}$. Hence, by the symmetry of $\mathscr{S}$, the point $(u_1 - \xi_1, u_2 - \xi_2)$ is in $\mathscr{S}$. Finally, the mid-point of $(u_1 - \xi_1, u_2 - \xi_2)$ and (ξ_1, ξ_2) is in $\mathscr{S}$ because of convexity, that is $(\frac{1}{2}u_1, \frac{1}{2}u_2)$ is in $\mathscr{S}$, and (u_1, u_2) is in $\mathscr{R}$, as required. It is clear that $\mathscr{S}(u_1, u_2)$ overlaps $\mathscr{S}(u_1', u_2')$ when and only when $\mathscr{S}$ overlaps $\mathscr{S}(u_1 - u_1', u_2 - u_2')$.

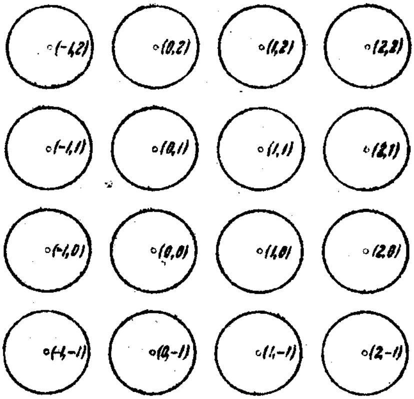

Fig. 1

To prove MINKOWSKI's theorem, it is thus enough to show that when the $\mathscr{S}(u_1, u_2)$ do not overlap then the area of each is at most 1. A little reflection convinces one that this must be so. A formal proof is given in Chapter 3. Another argument, which is perhaps more intuitive is as follows, where we suppose that $\mathscr{S}$ is entirely contained in a square

$$|x_1| \leqq X, \quad |x_2| \leqq X.$$

Let U be a large integer. There are $(2U+1)^2$ regions $\mathscr{S}(u_1, u_2)$ whose centres (u_1, u_2) satisfy

$$|u_1| \leqq U, \quad |u_2| \leqq U.$$

These $\mathscr{S}(u_1, u_2)$ are all entirely contained in the square

$$|x_1| \leqq U + X, \quad |x_2| \leqq U + X$$

of area

$$4(U+X)^2.$$

Since the $\mathscr{S}(u_1, u_2)$ are supposed not to overlap, we have

$$(2U+1)^2 V \leqq 4(U+X)^2,$$

[1] The converse statement is trivially true. If (u_1, u_2) is in $\mathscr{R}$ then $(\frac{1}{2}u_1, \frac{1}{2}u_2)$ is in both $\mathscr{S}$ and $\mathscr{S}(u_1, u_2)$.

where V is the area of $\mathscr{S}$; and so of each $\mathscr{S}(u_1, u_2)$. On letting U tend to infinity we have $V \leqq 1$, as required.

P5. A change in the co-ordinate system in our example of a definite binary quadratic form $f(x_1, x_2)$ leads to another point of view. We may represent $f(x_1, x_2)$ as the sum of the squares of two linear forms:

$$f(x_1, x_2) = X_1^2 + X_2^2, \tag{3}$$

where

$$X_1 = \alpha x_1 + \beta x_2, \qquad X_2 = \gamma x_1 + \delta x_2 \tag{4}$$

and $\alpha, \beta, \gamma, \delta$ are constants, e.g. by putting

$$\alpha = a_{11}^{\frac{1}{2}}, \qquad \beta = a_{11}^{-\frac{1}{2}} a_{12},$$
$$\gamma = 0, \qquad \delta = a_{11}^{-\frac{1}{2}} D^{\frac{1}{2}}.$$

Conversely if $\alpha, \beta, \gamma, \delta$ are any real numbers with $\alpha\delta - \beta\gamma \neq 0$ and X_1, X_2 are given by (4), then

$$X_1^2 + X_2^2 = a_{11} x_1^2 + 2 a_{12} x_1 x_2 + a_{22} x_2^2,$$

with

$$\left.\begin{aligned} a_{11} &= \alpha^2 + \gamma^2, \\ a_{12} &= \alpha\delta + \beta\gamma, \\ a_{22} &= \beta^2 + \delta^2, \end{aligned}\right\} \tag{5}$$

is a positive definite quadratic form with

$$D = a_{11} a_{22} - a_{12}^2 = (\alpha\delta - \beta\gamma)^2. \tag{6}$$

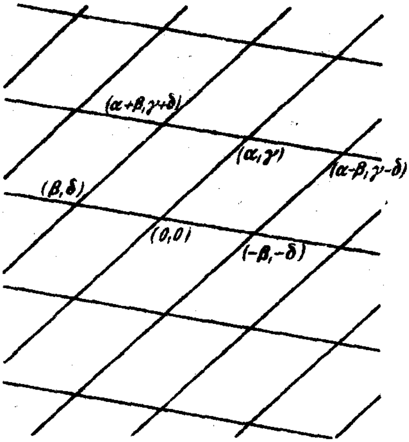

Fig. 2

We now consider X_1, X_2 as a system of rectangular cartesian coordinates. The points X_1, X_2 corresponding to integers x_1, x_2 in (4) are then said to form a (2-dimensional) lattice Λ. In vector notation Λ is the set of points

$$(X_1, X_2) = u_1(\alpha, \gamma) + u_2(\beta, \delta), \tag{7}$$

where u_1, u_2 run through all integer values.

We must now examine the properties of lattices more closely. Since we consider Λ merely as a set of points, it can be expressed in terms of more than one basis. For example

$$(\alpha - \beta, \gamma - \delta), \qquad (-\beta, -\delta)$$

is another basis for Λ. A fixed basis (α, β), (γ, δ) for Λ determines a subdivision of the plane by two families of equidistant parallel lines, the first family consisting of those points (X_1, X_2) which can be expressed in the form (7) with u_2 integral and u_1 only real, while for the lines of the second family the rôles of u_1 and u_2 are interchanged. In this way the plane is subdivided into parallelograms whose vertices are just the points of Λ. Of course the subdivision into parallelograms depends on the choice of basis, but we show that the area of each parallelogram, namely

$$|\alpha\delta-\beta\gamma|,$$

is independent of the particular basis. We can do this by showing that the number $N(X)$ of points of Λ in a large square

$$\mathcal{Q}(X):\quad |X_1|\leqq X,\ |X_2|\leqq X$$

satisfies

$$\frac{N(X)}{4X^2}\to\frac{1}{|\alpha\delta-\beta\gamma|}\qquad (X\to\infty).$$

Indeed a consideration along the lines of the proof of MINKOWSKI'S convex body theorem sketched above shows that the number of points of Λ in $\mathcal{Q}(X)$ is roughly equal to the number of parallelograms contained in $\mathcal{Q}(X)$, which again is roughly equal to the area of $\mathcal{Q}(X)$ divided by the area $|\alpha\delta-\beta\gamma|$ of an individual parallelogram. The strictly positive number

$$d(\Lambda)=|\alpha\delta-\beta\gamma| \tag{8}$$

is called the determinant of Λ. As we have seen, it is independent of the choice of basis.

P6. In terms of the new concepts we see that the statement that there is always an integer solution of $f(x_1, x_2)\leqq(4D/3)^{\frac{1}{2}}$ is equivalent to the statement that every lattice Λ has a point, other than the origin, in

$$X_1^2+X_2^2\leqq(\tfrac{4}{3})^{\frac{1}{2}}d(\Lambda). \tag{9}$$

On grounds of homogeneity this is again equivalent to the statement that the open circular disc

$$\mathcal{D}:\quad X_1^2+X_2^2<1 \tag{10}$$

contains a point of every lattice Λ with $d(\Lambda)<(\frac{3}{4})^{\frac{1}{2}}$, and the fact that there are forms such that equality is necessary in (2) is equivalent to the existence of a lattice Λ_c with determinant $d(\Lambda_c)=(\frac{3}{4})^{\frac{1}{2}}$ having no point in $\mathcal{D}$. So our problem about all definite binary quadratic forms is equivalent to one about the single region $\mathcal{D}$ and all lattices. Similarly consideration of the lattices with points in

$$|X_1X_2|<1$$

gives us information about the minima of indefinite binary quadratic forms:

$$\inf_{\substack{u_1, u_2 \text{ integers} \\ \text{not both } 0}} |f(u_1, u_2)|:$$

and so on.

These considerations prompt the following definitions. A lattice Λ is said to be admissible for a region (point-set) $\mathcal{R}$ in the (X_1, X_2)-plane if it contains no point of $\mathcal{R}$ other than perhaps the origin, if that is a point of $\mathcal{R}$. We may say then that Λ is $\mathcal{R}$-admissible. The lower bound $\Delta(\mathcal{R})$ of $d(\Lambda)$ over all $\mathcal{R}$-admissible lattices is the lattice-constant of $\mathcal{R}$: if there are no $\mathcal{R}$-admissible lattices we put $\Delta(R) = \infty$. Then any lattice Λ with $d(\Lambda) < \Delta(R)$ certainly contains a point of $\mathcal{R}$ other than the origin. An $\mathcal{R}$-admissible lattice Λ with $d(\Lambda) = \Delta(\mathcal{R})$ is called critical (for $\mathcal{R}$): of course critical lattices need not exist in general.

The importance of critical lattices was already recognized by MINKOWSKI. If Λ_c is critical for $\mathcal{R}$ and Λ is obtained from Λ_c by a slight distortion (i.e. by making small changes in a pair of base-points) then either Λ has a point in $\mathcal{R}$ other than the origin or $d(\Lambda) \geqq d(\Lambda_c)$ (or both).

As an example, let us again consider the open circular disc

$$\mathcal{D}: \quad X_1^2 + X_2^2 < 1.$$

Suppose that Λ_c is a critical lattice for $\mathcal{D}$. We outline a proof that a critical lattice, if it exists, must have three pairs of points $\pm(A_1, A_2)$, $\pm(B_1, B_2)$, $\pm(C_1, C_2)$ on the boundary $X_1^2 + X_2^2 = 1$ of $\mathcal{D}$. For if Λ_c had no points on $X_1^2 + X_2^2 = 1$, we could obtain an $\mathcal{D}$-admissible lattice with smaller determinant from Λ_c by shrinking it about the origin, that is by considering the lattice $\Lambda = t\Lambda_c$ of points (tX_1, tX_2), where $(X_1, X_2) \in \Lambda$ and $0 < t < 1$ is fixed. Then $d(\Lambda) = t^2 d(\Lambda_c) < d(\Lambda_c)$, and clearly Λ would be also $\mathcal{D}$-admissible if t is near enough to 1. Hence Λ_c contains a pair of points on $X_1^2 + X_2^2 = 1$, which, after a suitable rotation of the co-ordinate system, we may suppose to be $\pm(1, 0)$. If there were no further points of Λ_c on $X_1^2 + X_2^2 = 1$ then we could obtain a $\mathcal{D}$-admissible lattice Λ of smaller determinant by shrinking Λ_c perpendicular to the X_1-axis, that is by taking Λ to be the lattice of (X_1, tX_2), $(X_1, X_2) \in \Lambda_c$, where t is near enough to 1. Finally, if Λ_c had only two pairs of points $\pm(1, 0)$, $\pm(B_1, B_2)$ on the boundary, then it is not difficult to see that it could be slightly distorted so that $(1, 0)$ remains fixed but (B_1, B_2) moves along $X_1^2 + X_2^2 = 1$ nearer to the X_1-axis, cf. Fig. 3.

This can be verified to decrease the determinant of the lattice [indeed $(1, 0)$ and (B_1, B_2) can be shown to be a basis for Λ_c], and for small distortions the distorted lattice Λ will still be $\mathscr{D}$-admissible. Hence a critical lattice Λ_c (if it exists) must have three pairs of points on $X_1^2+X_2^2=1$: and it is easy to verify that the only lattice with three pairs of points on $X_1^2+X_2^2=1$, one of them being $\pm(1, 0)$, is the lattice Λ' with basis

$$(1,0), \quad \left(\tfrac{1}{2}, \sqrt{\tfrac{3}{4}}\right).$$

This has the vertices of a regular hexagon

$$\pm(1,0),$$
$$\pm\left(\tfrac{1}{2}, \sqrt{\tfrac{3}{4}}\right),$$
$$\pm\left(-\tfrac{1}{2}, \sqrt{\tfrac{3}{4}}\right)$$

on $X_1^2+X_2^2=1$, but no points in $X_1^2+X_2^2<1$. We have thus shown that $\Delta(D)=d(\Lambda')=\left(\frac{3}{4}\right)^{\frac{1}{2}}$ provided that $\mathscr{D}$ has a critical lattice. MINKOWSKI showed that critical lattices exist for a fairly wide set of regions $\mathscr{R}$ by, roughly speaking, showing that any $\mathscr{R}$-admissible lattice Λ can be gradually shrunk and distorted until it becomes critical. In the text we give a more general proof of the existence of critical lattices using concepts due to MAHLER which turn out to have much wider significance.

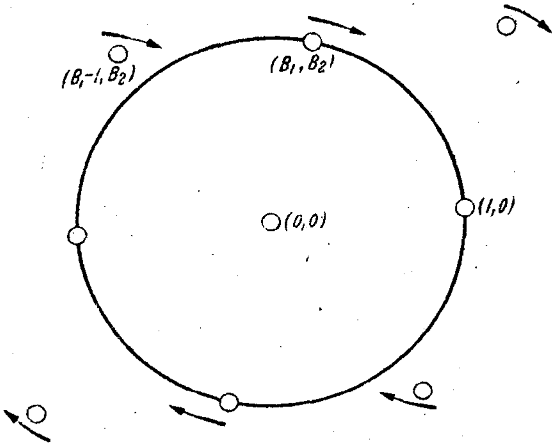

Fig. 3

P7. Another general type of problem is the typical "inhomogeneous problem": Let $f(x_1, \ldots, x_n)$ be some real-valued function of the real variables $x_1, \ldots, x_n$. It is required to find a constant k with the following property: If $\xi_1, \ldots, \xi_n$ are any real numbers there are integers $u_1, \ldots, u_n$ such that

$$|f(\xi_1-u_1, \ldots, \xi_n-u_n)| \leqq k.$$

Questions of this sort turn up naturally, for example in the theory of algebraic numbers. Again there is a simple geometric picture. For simplicity let $n=2$. Let $\mathscr{R}$ be the set of points (x_1, x_2) in the 2-dimensional euclidean plane with

$$|f(x_1, x_2)| \leqq k.$$

Denote by $\mathscr{R}(u_1, u_2)$, where u_1, u_2 are any integers, the region similar to $\mathscr{R}$ but with the displacement u_1, u_2; that is $\mathscr{R}(u_1, u_2)$ is the set of points x_1, x_2 such that

$$|f(x_1-u_1, x_2-u_2)| \leqq k.$$

Then the inhomogeneous problem is clearly to choose k so that the regions $\mathscr{R}(u_1, u_2)$ cover the whole plane. In general one will wish to choose k, and so $\mathscr{R}$, as small as possible so that it still has this covering property. Here we have a contrast with the treatment of the homogeneous problem in § 4, where the objective was to make the regions [there denoted by $\mathscr{S}(u, v)$] as large as possible but so that they did not overlap.

In this book we shall mainly be concerned at first with the homogeneous problem. Only when we have a fairly complete theory of the homogeneous problem will we discuss in Chapter XI the inhomogeneous problem and its relation to the homogeneous one.

Chapter I

Lattices

I.1. Introduction. In this chapter we introduce the most important concept in the geometry of numbers, that of a lattice, and develop some of its basic properties. The contents of this chapter, except § 2.4 and § 5, are fundamental for almost everything that follows.

In this book we shall be concerned only with lattices over the ring of rational integers. A certain amount of work has been done on lattices over complex quadratic fields, see e.g. MULLENDER (1945a) and K. ROGERS (1955a). Many of the concepts should carry over practically unaltered. Again, work on approximation to complex numbers by integers of a complex quadratic field [e.g. MULLENDER (1945a), CASSELS, LEDERMANN and MAHLER (1951a), POITOU (1953a)] and on the minima of hermitian forms when the variables are integers in a quadratic field [e.g. OPPENHEIM (1932a, 1936a, 1953f) and K. ROGERS (1956a)] may be regarded as a generalization of the geometry of numbers to lattices over complex quadratic fields. We shall not have occasion to mention lattices over complex quadratic fields again in this book; we mention them here only for completeness. For lattices over general algebraic number fields see ROGERS and SWINNERTON-DYER (1958a).

I.2. Bases and sublattices. Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be linearly independent real vectors in n-dimensional real euclidean space, so that the only set of numbers $t_1, \ldots, t_n$ for which $t_1\boldsymbol{a}_1 + \cdots + t_n\boldsymbol{a}_n = \boldsymbol{o}$ is $t_1 = t_2 = \cdots = t_n = 0$. The set of all points

$$\boldsymbol{x} = u_1\boldsymbol{a}_1 + \cdots + u_n\boldsymbol{a}_n \tag{1}$$

with integral $u_1, \ldots, u_n$ is called the lattice with basis $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. We note that, since $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ are linearly independent, the expression of any vector $\boldsymbol{x}$ in the shape (1) with real $u_1, \ldots, u_n$ is unique. Hence if $\boldsymbol{x}$ is in Λ and (1) is any expression for $\boldsymbol{x}$ with real $u_1, \ldots, u_n$, then $u_1, \ldots, u_n$ are integers. We shall make use of these remarks frequently, often without explicit reference.

The basis is not uniquely determined by the lattice. For let $\boldsymbol{a}'_i$ be the points

$$\boldsymbol{a}'_i = \sum_j v_{ij}\boldsymbol{a}_j \qquad (1 \leq i,\ j \leq n), \tag{2}$$

where v_{ij} are any integers with

$$\det(v_{ij}) = \pm 1. \tag{3}$$

Then

$$\boldsymbol{a}_i = \sum_j w_{ij} \boldsymbol{a}'_j \tag{4}$$

with integral w_{ij}. It follows easily that the set of points (1) is precisely the set of points

$$u'_1 \boldsymbol{a}'_1 + \cdots + u'_n \boldsymbol{a}'_n$$

where $u'_1, \ldots, u'_n$ run through all integers; that is $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ and $\boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_n$ are bases of the same lattice. We show now that every basis $\boldsymbol{a}'_i$ of a lattice Λ may be obtained from a given basis $\boldsymbol{a}_i$ in this way. For since $\boldsymbol{a}'_i$ belongs to the lattice with basis $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ there are integers v_{ij} such that (2) holds: and since $\boldsymbol{a}_i$ belongs to the lattice with basis $\boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_n$ there are integers w_{ij} such that (4) holds. On substituting (2) in (4) and making use of the linear independence of the $\boldsymbol{a}_i$, we have

$$\sum w_{ij} v_{jl} = \begin{cases} 1 & \text{if} \quad i = l \\ 0 & \text{otherwise}. \end{cases}$$

Hence

$$\det(w_{ij}) \det(v_{jl}) = 1$$

and so each of the integers $\det(w_{ij})$ and $\det(v_{jl})$ must be ± 1; that is (3) holds as required.

We denote lattices by capital sanserif Greek letters, and in particular by Λ, M, N, Γ.

If $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ and $\boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_n$ are bases of the same lattice, so that they are related by (2) and (3), then we have

$$\det(\boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_n) = \det(v_{ij}) \det(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = \pm \det(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n),$$

where, for example, $\det(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ denotes the determinant of the $n \times n$ array whose j-th row is the vector $\boldsymbol{a}_j$. Hence

$$d(\Lambda) = |\det(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)|$$

is independent of the particular choice of basis for Λ. Because of the linear independence of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ we have

$$d(\Lambda) > 0.$$

We call $d(\Lambda)$ the determinant of Λ.

An example of a lattice is the set Λ_0 of all vectors with integral coordinates. A basis for Λ_0 is clearly the set of vectors

$$\boldsymbol{e}_j = (\overbrace{0, \ldots, 0}^{j-1 \text{ zeros}}, 1, \overbrace{0, \ldots, 0}^{n-j \text{ zeros}}) \qquad (1 \leqq j \leqq n);$$

and so

$$d(\Lambda_0) = 1.$$