THE GEOMETRY OF NUMBERS.*

By H. Davenport.

I would like first to thank the Association for the honour they have done me in inviting me to lecture at this meeting. It is, however, an honour which carries with it a certain danger. When a mathematician comes to lecture on a subject in which he has specialised, he is always liable to suffer from the delusion that what has gradually become plain and straightforward to him will at once appear plain and straightforward to others. It may be that I am suffering from this delusion in thinking that I can present the main idea of the geometry of numbers in one lecture.

The theory of numbers is concerned primarily with properties of the integers, or of the rational numbers. It is a remarkable fact that many such properties are most conveniently approached by methods which use irrational numbers, or again by methods which use continuous variables. Also equations in integers can often be best approached from the corresponding inequalities. As a simple example, suppose we wish to prove that the equation

$$x^2 - 2y^2 = \pm 1$$

has an infinity of integral solutions. One way of doing so would be to prove that the inequality

$$| x^2 - 2y^2 | < 2$$

has an infinity of integral solutions; and on factorising this we find that it would suffice to prove that there are an infinity of fractions x/y for which

$$\left| \frac{x}{y} - \sqrt{2} \right| < \frac{c}{y^2},$$

for some constant c less than $1/\sqrt{2}$. Thus a problem of a purely arithmetical character is reduced to a problem of *Diophantine approximation*. True, in this particularly simple example, it is hardly worth while to do so, because it is easier to treat the equation directly; but the principle is one of great generality.

The general problem of Diophantine approximation may be stated as follows. Let $f(x_1, \ldots, x_n)$ be a function of n variables, which is always positive or zero. How small can we make $f(x_1, \ldots, x_n)$ for integral values of the variables? If f is homogeneous, we must exclude the trivial possibility $x_1 = \ldots = x_n = 0$ which makes $f = 0$. In this general form, the question is necessarily somewhat vague, since one naturally asks, how small, in terms of what? When f is an algebraic form, or a combination of such forms, the answer is that one aims at an inequality in terms of the algebraic invariant, or invariants, of f. The problem of Diophantine approximation is of a mixed arithmetic and analytic nature, since the variables are integers and the coefficients of f are supposed to be arbitrary real numbers.

The geometry of numbers is an approach to problems of Diophantine approximation, suggested by interpreting them geometrically. The inequality $f(x_1, \ldots, x_n) < \lambda$ represents a certain region in n dimensional space. Under what conditions does this region contain a point with integral coordinates? The origin O, if it obviously belongs to the region, is to be excluded from consideration.

It was one particular problem of this kind which led Minkowski, in 1891, to his general theorems in the geometry of numbers. Take $f(x_1, \ldots, x_n)$ to be a quadratic form,

$$f(x_1, \ldots, x_n) = a_{11}x_1^2 + \ldots + 2a_{12}x_1x_2 + \ldots,$$

* Lecture delivered on 9th April, 1947, to the General Meeting of the Mathematical Association.

which is positive definite, i.e. strictly positive except when $x_1, \ldots, x_n$ are all zero. In this case, the inequality $f(x_1, \ldots, x_n) < \lambda$ represents the interior of an ellipsoid in n dimensional space, of a given shape and orientation, which expands as λ increases. The problem now is: how far must we expand the ellipsoid in order to be certain that it contains a point with integral co-ordinates, other than O? One possible answer * that suggests itself is that it may suffice to expand until the volume of the ellipsoid is greater than some constant (depending only on n). This is in fact true, but it is not immediately obvious. One may think it possible for an ellipsoid to have a large volume, but to be very long and thin, and to stretch out in some direction which enables it to avoid all the points with integral coordinates.

The theorem I have mentioned, namely that there exists a constant (depending only on n) such that any ellipsoid with centre O whose volume exceeds that constant necessarily contains a point other than O with integral co-ordinates, was first proved by Hermite in 1845. He did not use this geometrical interpretation, but considered the problem in its original formulation, as that of the minimum of a positive definite quadratic form, and proved the result by induction on n. The merit of Minkowski's proof lies in its simplicity and in its far greater generality and power.

Minkowski's argument was as follows. Let E be an ellipsoid, with centre O, which does not contain, in its interior, any point with integral coordinates except O. Let $\frac{1}{2}E$ denote the ellipsoid obtained by contracting E to half the linear dimensions. Place ellipsoids congruent † to the ellipsoid $\frac{1}{2}E$ with their centres at all points with integral coordinates. Then no two of this system of congruent ellipsoids can overlap. For if they did, then by the regularity of their distribution there would be one, with centre P, say, which overlapped the one with centre O. Then the mid-point of OP would lie in $\frac{1}{2}E$, and P would lie in E, contrary to hypothesis.

Thus we have a system of non-intersecting ellipsoids, all congruent to $\frac{1}{2}E$, centred at all points with integral coordinates. Let V denote the volume of E; then the volume of $\frac{1}{2}E$ is $V/2^n$. Intuition suggests that, in view of the non-intersection, the volume of $\frac{1}{2}E$ cannot exceed 1, and this is easily proved by a rigorous argument. Consider those ellipsoids of the system whose centres $(x_1, \ldots, x_n)$ satisfy

$$0 \leqslant x_1 < N, \ldots, \quad 0 \leqslant x_n < N,$$

where N is a large positive integer. Their number is N^n, their total volume is $N^n V/2^n$. They do not overlap, and they all lie in the cube

$$-k \leqslant x_1 < N + k, \ldots, \quad -k \leqslant x_n < N + k,$$

where k depends only on the ellipsoid, and not on N. (k is simply the greatest numerical value of any coordinate of a point on the surface of the ellipsoid $\frac{1}{2}E$.) Thus

$$N^n V/2^n \leqslant (N + 2k)^n,$$

whence

$$V \leqslant 2^n (1 + 2k/N)^n.$$

Since N is arbitrarily large, this implies $V \leqslant 2^n$.

We have therefore proved that *an ellipsoid with centre O, which does not contain any point with integral coordinates except O, has a volume not exceeding*

* This answer would be of the kind envisaged earlier, since the volume of the ellipsoid is expressible in terms of λ and the determinant of the quadratic form, which is the sole algebraic invariant of the form.

† For brevity, I am using the word *congruent* to mean that the one body can be derived from the other by a translation only.

2^n. In other words, *an ellipsoid with centre O and volume greater than* 2^n *must contain a point with integral coordinates, other than O, in its interior.*

As we have seen, the proof is exceedingly simple. At one point only have we used any special property of the ellipsoid ; namely when we asserted that the mid-point of the centres of two congruent intersecting ellipsoids lies in the common part. Now this holds good for many other bodies, e.g. for parallelepipeds. Minkowski showed that it holds for any body * (in n dimensional space) which is

(i) *symmetrical about O*, that is, if a point belongs to the body, so does the image of that point in O ;

(ii) *convex*, that is, if P and Q are two points of the body, so is their mid-point.

The proof is very simple. Suppose R is such a body, and $R(A)$, $R(B)$ denote congruent bodies centred at A, B. Suppose $R(A)$ and $R(B)$ have a point X in common, and complete the parallelogram $AXBY$. Let X', Y' denote the images of X, Y in A. Since X lies in $R(A)$, so does X' by the symmetry, and consequently Y lies in $R(B)$ by the congruence of $R(A)$ and $R(B)$. Similarly Y lies in $R(A)$. By the convexity, the mid-point of XY lies in both $R(A)$ and $R(B)$, and this point is also the mid-point of AB. The proof can be expressed even more briefly by using an appropriate vector notation.

Thus we have, without further proof, the more general theorem :

Any convex body in n dimensional space, which is symmetrical about O and has volume greater than 2^n, *contains a point with integral coordinates, other than O.*

This is Minkowski's fundamental theorem in the geometry of numbers. It is very effective in solving problems of Diophantine approximation, and was applied by Minkowski to give simple proofs of many theorems in that subject and in algebraic number-theory. In this general form, applying to any symmetrical convex body, it is easily seen to be best possible, in that the number 2^n cannot be replaced by a smaller one. For the cube

$$|x_1| < 1, \ldots, |x_n| < 1$$

is a convex symmetrical body, whose volume is exactly 2^n, and it does not contain a point with integral coordinates other than O.

There is a more general geometrical interpretation than the direct one which we have used up to now. Apply to the variables $x_1, \ldots, x_n$ any linear transformation

$$X_1 = \alpha_{11}x_1 + \ldots + \alpha_{1n}x_n,$$
$$\ldots$$
$$X_n = \alpha_{n1}x_1 + \ldots + \alpha_{nn}x_n,$$

with real coefficients α_{ij} whose determinant Δ is not zero. This represents the general affine transformation from the x space to the X space. The system of points with integral coordinates in x space is transformed into a system of points in X space of a more general kind, which we call a *lattice*. If we denote by $A_1 = (\alpha_{11}, \ldots, \alpha_{n1})$ the point in X space which corresponds to the point $(1, 0, \ldots, 0)$ in x space, and so on, the general point P of the lattice is given by

$$\overrightarrow{OP} = x_1\overrightarrow{OA_1} + \ldots + x_n\overrightarrow{OA_n},$$

where $x_1, \ldots, x_n$ take all integral values. The parallelepiped based on the vectors $OA_1, \ldots, OA_n$ may be called the *fundamental parallelepiped* of the lattice.

* To avoid complications which are irrelevant to the argument, I have not defined the term " body ". It will suffice if we understand it as meaning any set of points which is *open* (*i.e.* every point is an interior point), and possesses a *volume*, in any of the various ways in which " volume " can be defined.

A body of volume V in x space is transformed into a body of volume $V' = V \mid \Delta \mid$ in X space. In particular, the volume of the fundamental parallelepiped is $\mid \Delta \mid$, and we call this the *determinant of the lattice*. The properties of convexity and of symmetry in O are preserved by an affine transformation. We can restate Minkowski's fundamental theorem in X space as follows :

Let L be any lattice of determinant Δ in n dimensional space. Then any convex body, symmetrical about O, whose volume exceeds $2^n\Delta$, contains a point of L other than O.

We can use an affine transformation to simplify any particular body with which we are concerned, at the expense of having to consider a general lattice instead of the special lattice of points with integral coordinates. For example, we can transform any ellipsoid into a sphere, and the problem with which we began becomes the problem of finding conditions under which a sphere will necessarily contain a point of any lattice of given determinant. The answer provided by Minkowski's theorem is, in this particular case, clearly not the best possible. For it was obtained by arguing that a set of non-intersecting spheres cannot more than fill up space, and it is plain that in fact an appreciable proportion of space will be left empty. The problem of finding the best possible constant for a sphere, in place of 2^n, is the same as the problem of the *closest packing* of equal spheres in n dimensions, subject to the condition that the centres of the spheres shall form a lattice. We may call this the closest *regular* packing. A considerable advance was made by Blichfeldt in 1914. He used a physical analogy, involving material spheres with a certain law of density, and counted the total quantity of matter instead of the total volume, thus obtaining* an appreciably better constant than 2^n. The true constant is still not known for general n. In two dimensions, the closest of all packings happens to be regular, but this is probably not the case in higher dimensions.

In many problems of Diophantine approximation, the geometrical interpretation does not give us a convex body. In such cases, we can deduce a crude result by applying Minkowski's theorem to a suitable inscribed convex body. A simple example is that of the region

$$\mid XY \mid < 1$$

in the plane. We ask, how small must be the determinant Δ of a lattice in order that we can be certain that a lattice point other than O will lie in the region? If we apply Minkowski's theorem to the inscribed square

$$\mid X \mid + \mid Y \mid < 2,$$

which has area 8, we deduce that the result certainly holds if $\Delta < 2$. In arithmetical form, the problem is that of the minimum of $\mid XY \mid$, where

$$X = \alpha x + \beta y, \quad Y = \gamma x + \delta y$$

are two linear forms of determinant Δ, and x, y take all integral values. A theorem of Markoff tells us that this minimum does not exceed $\Delta/\sqrt{5}$, and that this result is best possible. Consequently, for the geometrical formulation, the true condition is $\Delta < \sqrt{5}$. The corresponding problem for the body

$$\mid XYZ \mid < 1$$

in three-dimensional space was first solved by me in 1937. The answer is that this body contains a point other than O of every lattice whose determinant Δ satisfies $\Delta < 7$, and that 7 is the true constant. It is hardly possible for me in this lecture to say anything useful about this work; or about the subsequent work of Mordell and Mahler, which has given a new lease of life to the subject, and has unfolded problems of an entirely new kind.

* See a later account by Blichfeldt in *Math. Annalen*, 101 (1929), pp. 605–8.

I will conclude by showing how Minkowski's theorem can be used* to prove a classical result in number-theory : Lagrange's theorem that every positive integer is representable as the sum of four integral squares. This is not a very easy theorem to prove : Euler tried many times between 1730 and 1750 and failed. There is one Lemma which is needed in all the direct elementary proofs of the theorem, and which I shall take for granted. This is that for any odd positive integer m there exist integers a, b satisfying

$$a^2 + b^2 + 1 \equiv 0 \pmod{m}. \tag{1}$$

This is proved by simple considerations relating to quadratic residues when m is a prime p, then by induction on v when $m = p^v$, and finally by combination of these results it follows for general m.

We proceed to prove that m is representable as the sum of four integral squares. We define four linear forms X, Y, Z, W in four variables, x, y, z, w as follows :

$$\begin{aligned} X &= mx + az + bw \\ Y &= my + bz - aw \\ Z &= z \\ W &= w \end{aligned}$$

The determinant of these forms is obviously m^2. Thus if x, y, z, w take all integral values, the corresponding points (X, Y, Z, W) form a lattice in four-dimensional space of determinant $\Delta = m^2$.

In virtue of (1), we have

$$X^2 + Y^2 + Z^2 + W^2 \equiv 0 \pmod{m} \tag{2}$$

for all values of x, y, z, w. If we prove that there is a lattice point other than O for which

$$X^2 + Y^2 + Z^2 + W^2 < 2m, \tag{3}$$

it will follow from (2) that m is representable as $X^2 + Y^2 + Z^2 + W^2$. The inequality (3) represents a sphere of radius $\sqrt{(2m)}$. By Minkowski's theorem, it will suffice to prove that the volume of this sphere is greater than $2^4 \Delta = 2^4 m^2$. Now one easily finds, by integration, that the volume of a four-dimensional sphere of radius r is $\frac{1}{2}\pi^2 r^4$. Hence it suffices if

$$\tfrac{1}{2}\pi^2 (2m)^2 > 2^4 m^2,$$

and this is so, since $\pi^2 > 8$.

This proves Lagrange's theorem for any odd positive integer m, and the result extends at once to even integers, since if †

$$m = X^2 + Y^2 + Z^2 + W^2$$

then

$$2m = (X + Y)^2 + (X - Y)^2 + (Z + W)^2 + (Z - W)^2.$$

I do not say that this is the ideal proof of Lagrange's theorem ; far from it. But it is a simple proof, with a minimum of tiresome detail, and it shows how the geometry of numbers can be used to prove purely arithmetical results. I would hazard the conjecture that there are many other applications of this kind awaiting discovery. H. D.

* This proof was given by Hermite in 1853 (*Oeuvres*, I, p. 288). Of course he could not appeal to Minkowski's theorem ; in place of it he used his own result on the minimum of a positive definite quadratic form. Professor Forder has drawn my attention to a somewhat similar proof by Grace, *J. London Math. Soc.*, 2 (1927), pp. 3–8.

† In connection with this step, I cannot refrain from referring to Professor Mordell's note in this *Gazette*, 26 (1942), p. 52, on Lewis Carroll's diary. Lewis Carroll found considerable difficulty in proving that if m is a sum of two squares, so is $2m$.