Chapter 22

Geometry of Numbers

1. The Motivating Problem; Quadratic Forms

The geometry of numbers deals with the use of geometric notions, especially convexity and lattice, to solve problems in number theory, usually via the solutions of inequalities in integers. Its genesis lies in the problem of minimizing the values of a quadratic form for integer values of the variables.

After Gauss, the study of binary quadratic forms was generalized in two directions — first to algebraic number theory (chapters 15–17) and then to general quadratic forms in any number of variables. The theory of general quadratic forms begins by generalizing such notions as representations of integers, equivalence under groups of matrices and reduction theory from the binary case (chapters 12, 13) to the general case. Among the pioneers in these studies were Jacobi, Dirichlet, Eisenstein, Hermite, H. J. S. Smith, H. Minkowski and C. L. Siegel. The theory is active to the present day and is deeply tied to both algebraic theories, e.g., the theory of arithmetic subgroups of Lie groups, and to analytic studies such as the theory of modular functions. Scharlau and Opolka [Sch - Opo] provide a brief and moderately elementary survey from a modern point of view and Borevich and Shafarevich [Bor - Sha] is a good systematic introduction.

Here we will only be concerned with inequalities involving quadratic forms, since trying to solve them led Minkowski to create the geometry of numbers. We will concentrate on **positive definite quadratic forms in n variables**, i.e., the forms

$$Q(x) = \sum_{i,j=1}^{n} a_{i,j} x_i x_j \ , \qquad (1)$$

where $x = (x_1, \ldots, x_n)$, $a_{i,j} \in \mathbf{R}$, $a_{ij} = a_{ji}$, and $Q(x) > 0$ for all $x \neq \mathbf{0} = (0, \ldots, 0)$. $\mathbf{D = det(a_{i,j})}$ is called **the determinant of the form.**

Hermite asked the question raised earlier: given such a form, how small can we make its value for integer values of the variables which are not all zero? A point $g = (g_1, \ldots, g_n) \in \mathbf{R}^n$, with all $g_i \in \mathbf{Z}$, will be called an **integer point**. The set $\mathbf{Z^n}$ of integer points is called the **integer lattice**. Thus we are studying the values of the forms on $\mathbf{Z}^n$ and later we will show, from the geometry, that a positive definite form defined on $\mathbf{Z}^n - \{\mathbf{0}\}$ actually achieves its minimum value at one of these points. Hermite generalized Lagrange's reduction theory for two variables to prove the following:

Theorem: ($\sim$ 1845) *Let Q be a positive definite quadratic form in n variables with determinant D. Then there exists a non-zero integer point g, such that*

$$Q(g_1, \ldots, g_n) \leq \left(\frac{4}{3}\right)^{\frac{n-1}{2}} D^{\frac{1}{n}} .$$

Hermite realized that many important parts of the theory depend on his theorem (see his letter of 1845 to Jacobi — translated in [Sch - Opo]). Thus, for example, he and Eisenstein used it to prove that the number of equivalence classes of positive definite forms in n variables with determinant D, under the action of $Gl_n(\mathbf{Z})$, is finite. These results were proved using arithmetic methods and tools of matrix theory (but without the powerful modern notation of matrices).

Later in the century Hermann Minkowski (1864–1909) and H. J. S. Smith carried the theory much further. There is an interesting story about this latter work as related by C. Reid in her biography of Hilbert [Rei, C].

> "Although Minkowski was still only 17 years old, he was involved in a deep work with which he hoped to win the Grand Priz des Sciences Mathematiques of the Paris Academy.
>
> The Academy had proposed the problem of the representation of a number as the sum of five squares. Minkowski's investigations, however, had led him far beyond the stated problem. By time the deadline of June 1, 1882 arrived, he still had not had his work translated into French as the rules of the competition required. Nevertheless, he decided to submit it. At the last minute, at the suggestion of his older brother Max, he wrote a short prefatory note in which he explained that his neglect had been due to the attractions of his subject and expressed the hope that the Academy would not think "I would have given more if I had given less." ...

> Then, in the spring of 1883, came the announcement that this boy, still only 18 years old, had been awarded jointly with the well-known English mathematician Henry Smith the Grand Priz des Sciences Mathematiques. ...
>
> For a while it seemed, though, that Minkowski might not actually receive his prize. The French newspapers pointed out that the rules of the competition had specifically stated that all entries must be in French. The English mathematicians let it be known that they considered it a reflection upon their distinguished countryman, who had since died, that he should be made to share a mathematical prize with a boy. ("It is curious to contemplate at a distance," an English mathematician remarked some forty years later, " the storm of indignation which convulsed the mathematical circles of England when Smith, bracketed after his death with the then unknown German mathematician, received a greater honor than any that had been paid to him in life.") In spite of the pressures upon them, the members of the prize committee never faltered. From Paris, Camille Jordan wrote to Minkowski: " Work, I pray you, to become a great mathematician." "

Minkowski's close friend Hilbert also wrote about this work

> "The seventeen year old student attacked this topic with all his energy and solved it brilliantly, developing, far beyond the original question, a general theory of quadratic forms, specifically their division in orders and genera, even for arbitrary rank. It is remarkable how well versed Minkowski was in the theory of elementary divisors, and in transcendental tools, such as Dirichlet series and Gauss sums."
>
> (trans. from [Sch -Opo])

Minkowski continued to work on these ideas and on November 6, 1889, he wrote to Hilbert

> " Perhaps you or Hurwitz are interested in the following theorem (which I can prove in half a page): in a positive definite form of determinant D with $n(\geq 2)$, one can always assign such values to the variables that the form is $< nD^{\frac{1}{n}}$ "
>
> (trans. from [Sch - Opo])

This theorem was based on geometric reasoning and it revolutionized the subject. We now explain these ideas, first systematically presented in Minkowski's fundamental book [Min 1].

Unless otherwise stated, all quadratic forms are assumed to be positive definite.

2. Minkowski's Fundamental Theorem

We will not hesitate to assume various geometric properties which are intuitively clear (the reader may first want to think about the cases $n = 2$ or 3). A set S in $\mathbf{R}^n$ is **symmetric with respect to the origin (or has center at 0** or **is centered at the origin)** if $x \in S$ implies that $-x \in S$. The volume of S will be denoted by **V(S)** (we are only dealing with nice sets where all the reasonable definitions of volume coincide).

Minkowski reformulated the problem of minimizing forms in a geometric language. For any positive definite quadratic form and positive real number λ, the set

$$Q_\lambda = \{(x_1, \ldots, x_n) \in \mathbf{R}^n | Q(x_1, \ldots, x_n) < \lambda\}$$

is a set in $\mathbf{R}^n$. Since for $n = 2$ the sets are ellipses and for $n = 3$ they are ellipsoids, we call the general Q_λ **n-dimensional ellipsoids** or just **ellipsoids**. Our diagrams will illustrate the case of ellipses in $\mathbf{R}^2$, which are symmetric with respect to the coordinate axes ($n = 2$ and $a_{12} = a_{21} = 0$ in equation 1.1). As we shall see, nothing is lost by visualizing this very special case.

As λ varies, the Q_λ are all similar and symmetric with respect to the origin (fig. 1).

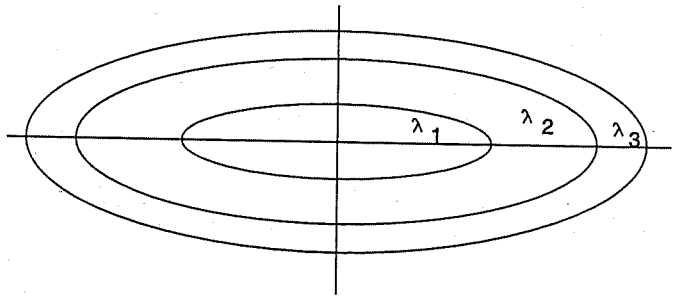

Figure 1

The minimization problem for Q is now equivalent to the question: how big must λ be to guarantee that the ellipsoid Q_λ contains a non-zero integer point?

Minkowski first proved the following theorem.

Theorem: *If the volume of the ellipsoid E is $> 2^n$, then it contains a non-zero integer point.*

This yielded a new proof of Hermite's theorem and, as we shall see, led to far reaching generalizations. We first follow Minkowski's approach to the geometry and delay the applications to quadratic forms.

Proof: We assume that E does not contain any non-zero integer points and show that $V(E) \leq 2^n$.

Let $E' = \frac{1}{2}E = \left\{\frac{1}{2}x | x \in E\right\}$. Then $V(E') = \frac{V(E)}{2^n}$. For each integer point g, consider the ellipsoid $E'_g = E' + g$, which is centered at g (fig. 2).

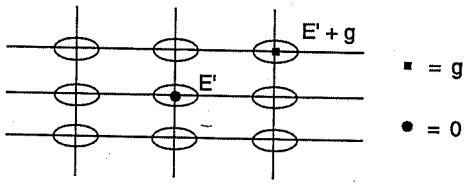

Figure 2

a) The ellipsoids do not overlap.

Suppose $p \in E'_g \cap E'_{g'}$, $g \neq g'$ (fig. 3). Then $p - g \in E'$, $p - g' \in E'$

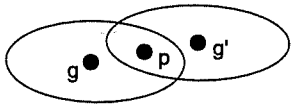

Figure 3

and, since E' is symmetric with respect to 0, $g - p \in E'$. Hence the midpoint

$$\frac{1}{2}[(p - g') + (g - p)] = \frac{1}{2}(g - g') \in E' .$$

Therefore

$$2\left[\frac{1}{2}(g - g')\right] = g - g' \in 2E' = E ,$$

contradicting the assumption that E has no non-zero integer points.

b) The 'density' of the ellipsoids is ≤ 1.

Intuitively, what we are saying is that for a large hypercube C, the ratio of the total volume of the E'_g's, with $g \in C$, to the volume of C is ≤ 1, i.e. if there are m such E'_g's with centers in C, and C is sufficiently large, then we must show that $\frac{mV(E')}{V(C)} \leq 1$. The problem is that pieces of some of the ellipsoids stick out of the sides of C and we must prove that the volume of these pieces is insignificant when C is large.

Now we make these ideas precise. Let N be a positive integer and $C = \{x \in \mathbf{R}^n | 0 \leq x_i \leq N\}$. There are $(N+1)^n$ ellipsoids E'_g, $g = (g_1, \ldots, g_n) \in \mathbf{Z}^n$, with $g \in C$, since there are $N+1$ choices for each integer coordinate with $0 \leq g_i \leq N$. By part a), these ellipsoids do not overlap and thus their total volume is $(N+1)^n V(E')$.

Since E is bounded, so is E' and therefore E' is contained in some hypercube, say $\{x | |x_i| < k\}$ (fig. 4). Then all the E'_g, $g \in C$, are contained

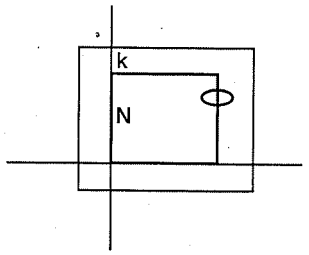

Figure 4

in the hypercube $C' = \{x | -k \leq x_i \leq N + k\}$, where $V(C') = (N + 2k)^n$. Hence

$$(N+1)^n V(E') \leq (N+2k)^n$$

and

$$V(E') \leq \left(\frac{N+2k}{N+1}\right)^n < \left(\frac{N+2k}{N}\right)^n$$
$$= \frac{1+2k}{N} \to 1 , \quad \text{as } N \to \infty .$$

Therefore $V(E') = \frac{V(E)}{2^n} \leq 1$ or $V(E) \leq 2^n$, and the theorem is proved.

Minkowski then observed that the only properties of E used in the proof were the following:

i) E is convex — A set $S \subseteq \mathbf{R}^n$ is **convex** if $x, y \in S$ implies that the line segment $\overline{xy} \in S$, where, analytically, $\overline{xy} = \{\lambda x + \mu y | \lambda, \mu \in \mathbf{R}, \lambda, \mu \geq 0, \lambda + \mu = 1\}$. In fact, we only used the fact that if $x, y \in S$, then $\frac{1}{2}(x + y) \in S$, but this implies convexity (exercise).

ii) E is symmetric with respect to the origin.

iii) E is bounded — E is contained in a box $\{(x_1, \ldots, x_n) | |x_i| < k\}$, for some $k > 0$.

(The techniques for proving that ellipsoids are convex and bounded will be discussed later.)

Thus Minkowski actually proved the following general theorem:

Minkowski's Theorem: *(First form): A bounded convex set C in $\mathbf{R}^n$, with center at $\mathbf{0}$ and volume $V(C) > 2^n$, contains a non-zero integer point.*

But we really have a stronger result. Our ellipsoids $\{Q(x_1, \ldots, x_n) < \lambda\}$ are open sets (the pre-image of an open set under the continuous map from $\mathbf{R}^n$ to $\mathbf{R}$ defined by Q) and our theorem certainly applies to open convex set. However, if the interior of a convex set C is not empty, then it is an open set and also has volume $V(C)$. Hence we have the following.

Minkowski's Theorem: *(Second form): A bounded convex set C in $\mathbf{R}^n$, with center at $\mathbf{0}$ and volume $V(C) > 2^n$, contains a non-zero integer point g in its interior.*

Remarks: i) The hypercube $|x_i| < 1$ of volume 2^n doesn't contain any non-zero integer points, so our constant 2^n is best possible.

ii) A convex set with volume $< 2^n$ *can* contain lattice points. If $n = 2$, consider the rectangle with sides parallel to the axes, centered at the origin, of width k and height $\frac{1}{2k}$. By choosing k sufficiently large, we can make the rectangle contain any preassigned number of lattice points on the x axis. This example obviously generalizes to any dimension.

Minkowski's Theorem: *(Third form): A bounded convex set C in $\mathbf{R}^n$, with center at $\mathbf{0}$ and volume $V(C) \geq 2^n$, contains a non-zero integer point in its interior or on its boundary.*

Proof: Let $C_k = \left(1 + \frac{1}{k}\right) C, k = 1, 2, \ldots$. Then $V(C_k) > 2^n$ and C_k contains an integer point $g^{(k)} \neq \mathbf{0}$. Hence we have a sequence of non-zero integer points, $g^{(1)}, g^{(2)}, \ldots$, which must all lie in $2C$. But $2C$ is bounded and contains only finitely many integer points. Therefore the sequence

takes a constant value g from some point on, and g must be in C or on its boundary (C, together with its boundary, equals the intersection of all the C_k).

This type of continuity argument will sometimes be used in our applications without specific elaboration.

This seemingly innocuous but very pretty theorem would not necessarily impress one as something of great depth. It was Minkowski's genius to recognize that the implications of this theorem and the related geometric ideas went far beyond the application to bounds for quadratic forms; he made it a working tool for the number theorist. In his hands, the theorem blossomed into an important theory with many applications and led to the solutions of outstanding unsolved problems. This work also led to Minkowski's deep contributions to the geometry of convex sets [Min 2]. We shall concentrate on the applications of Minkowski's theorem and only mention other developments.

In 1914, H. F. Blichfeldt gave a new proof of Minkowski's Theorem [Bli] which has led to further research on the geometry of more general sets. The proof proceeds via a more general theorem.

Blichfeldt's Theorem: *Let M be an bounded set in $\mathbf{R}^n$ with $V(M) > 1$. Then M contains two distinct points x, y such that $x - y$ is a non-zero integer point (not necessarily in M).*

Corollary: *Blichfeldt $\Longrightarrow$ Minkowski (First form).*

Proof: Let $K \in \mathbf{R}^n$ be a bounded convex set, centered at the origin, with $V(K) > 2^n$. Let $K' = \frac{1}{2}K$; so $V(K') > 1$. By Blichfeldt's theorem, there exist $x, y \in K'$ such that $g = x - y$ is a non-zero integer point. Then $2x, 2y \in K$ and, by symmetry, $-2y \in K$. Hence, by convexity, the midpoint of $2x$ and $-2y$, namely, $\frac{1}{2}(2x + (-2y)) = x - y$, is a non-zero integer point in K.

Proof of Blichfeldt's theorem: Let A be the hypercube $\{0 < x_i < 1\}$ and $A_g = A + g$, for each integer point g. Since M is bounded, only a finite number of the sets $M_g = A_g \cap M$ are non empty and $M = \cup_g M_g$ (fig.5).

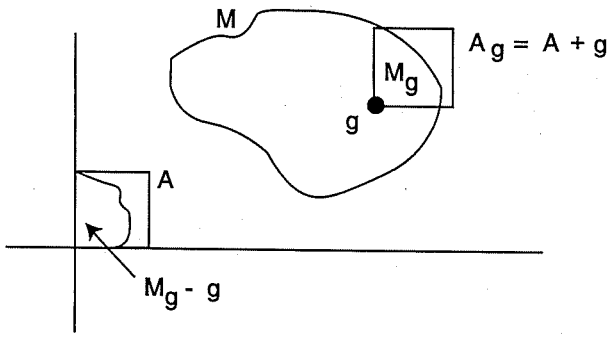

Figure 5

If we translate M_g back inside A (i.e., consider $M_g - g$), then $V(M_g - g) = V(M_g)$ and

$$\sum_g V(M_g - g) = \sum_g V(M_g) \geq V(M) > 1 .$$

Therefore, since $\cup_g (M_g - g) \subseteq A$ and $V(A) = 1$, two of the sets must overlap, and there exist integer points g, g' and a point p such that $p \in (M_g - g) \cap (M_{g'} - g')$. Hence

$$x = p + g \in M_g \quad \text{and} \quad y = p + g' \in M_{g'} ,$$

and

$$x - y = g - g'$$

is an integer point.

Moreover, if M is an open set, then all the other sets in the proof are open; in this case we can choose a p with rational coordinates and then x and y also have *rational* coordinates.

3. Minkowski's Theorem for Lattices

LATTICES

Now we can use Minkowski's theorem to attack our original problem of minimizing quadratic forms (sec. 1).

Example: $n = 2$ — Consider the class of positive definite quadratic forms $Q(x, y) = ax^2 + by^2, a, b \in \mathbf{R}, a, b > 0$, with determinant $D = ab$. The sets $\{Q < k\}$ are ellipses with area $\frac{k\pi}{\sqrt{ab}} = \frac{k\pi}{D^{\frac{1}{2}}}$. By Minkowski's theorem, if

$$\frac{k\pi}{D^{\frac{1}{2}}} \geq 2^2 \quad \text{or} \quad k \geq \frac{4}{\pi} D^{\frac{1}{2}} ,$$

then there is an integer point $(g_1, g_2) \neq (0, 0)$ in the interior or on the boundary of the ellipse; hence $Q(g_1, g_2) \leq k$. Reformulating this, we see that there exist integers g_1, g_2, not both zero, such that

$$Q(g_1, g_2) \leq \frac{4}{\pi} D^{\frac{1}{2}} .$$

We thus have a result similar to Hermite's theorem (sec. 1) with $\left(\frac{4}{3}\right)^{\frac{1}{2}}$ replaced by the weaker bound $\frac{4}{\pi}$.

We could now study bounds for positive definite quadratic forms in n variables by computing volumes of ellipsoids, but instead we will reformulate Minkowski's theorem in terms of lattices in $\mathbf{R}^n$, a generalization of the two dimensional lattices of section 17.2, in order to reduce the problem to the study of n-dimensional spheres. This also leads to an easier way of applying Minkowski's theorem to a variety of number theoretic results.

It is a standard result in linear algebra that a positive definite quadratic form $Q(x)$ can be transformed into a sum of squares by a non-singular linear transformation, $y = Ax$, of its variables, i.e., there exists an $A = (a_{ij}), a_{ij} \in R, \det(A) \neq 0$, such that

$$Q'(y) = Q(A^{-1}y) = \sum y_i^2 .$$

The matrix A maps the set $\mathbf{Z}^n$ to the set $\{y = Ax | x \in \mathbf{Z}^n\}$. Recall that $\mathbf{Z}^n$ is called the **integer lattice**.

Definition: (I) An **n-dimensional lattice** Λ is the image of the integer lattice in $\mathbf{R}^n$ under a non-singular linear transformation $y = Ax$. The points of Λ are called **Λ-lattice points** or, when no confusion can arise, just **lattice points**. Thus, e.g., integer points are $\mathbf{Z}^n$-lattice points by the identity transformation.

Since the ellipsoid $\{Q < \lambda\}$ in x-space is mapped by our matrix A to the sphere $\{Q' = \sum y_i^2 < \lambda\}$ in y-space, our minimization problem for Q

is transformed to the problem of minimizing $\sum y_i^2$ on the lattice Λ. Thus the minimization problem for all positive definite Q of determinant D is changed to the problem of deciding how big λ must be so that the sphere $\{\sum y_i^2 < \lambda\}$ contains a point of every lattice corresponding to the Q's.

The problem of minimizing a class of forms over the integer lattice has now been transformed to the problem of minimizing a single form for a class of lattices, and we shall see that to solve the latter problem we only have to compute the volumes of spheres.

First we study lattices and then we generalize Minkowski's theorem from $\mathbf{Z}^n$ to arbitrary lattices. There are two other equivalent ways of defining a lattice:

(II) An n-dimensional lattice Λ is a set of the form

$$\Lambda = \mathbf{Z}\alpha_1 + \cdots + \mathbf{Z}\alpha_n = \{m_1\alpha_1 + \cdots + m_n\alpha_n | m_i \in \mathbf{Z},\ \alpha_i \in \mathbf{R}^n\}$$

where $\alpha_1, \ldots, \alpha_n$ are linearly independent over $\mathbf{R}$. The α_i can be taken as the column vectors of the matrix A in our original definition.

(III) An n-dimensional lattice Λ is a discrete subgroup of $\mathbf{R}^n$. By **discrete** we mean that the intersection of Λ with any bounded subset of the plane is finite.

Since the proofs of equivalence of definitions I – III are basically the same as those given for the case of two dimensional lattices in section 17.2, we omit them. For a deeper study of lattices, see Lekkerkerker [Lek] or Cassels [Cas 4].

A **sublattice** of a lattice Λ in $\mathbf{R}^n$ is a subset of Λ which is also a lattice in $\mathbf{R}^n$ (so, in particular, it contains n points linearly independent over $\mathbf{R}$).

The set $\{\alpha_1, \ldots, \alpha_n\}$ of (II) is called an **integral** or **lattice basis** of Λ. Λ has infinitely many such bases and we will study them in a moment.

$\Delta = |\det(A)|$ is the **determinant** of the lattice $\Lambda : y = Ax$; we shall prove it is independent of the basis chosen for Λ. Geometrically Δ is the volume of the parallelepiped, $T = \{m_1\alpha_1 + \cdots + m_n\alpha_n | m_i \in \mathbf{Z}, 0 \le m_i \le 1\}$, whose edges are the column vectors $\alpha_1, \ldots, \alpha_n$ of A. This is clear since T is the image of the cube $\{0 \le x_i \le 1\}$, of volume one, under $y = Ax$ and thus $V(T) = |\det(A)| \times V(\text{cube})$. ($T$ is called a **fundamental parallelepiped** for Λ given by the basis $\alpha_1, \ldots, \alpha_n$.)

Δ can also be interpreted as the reciprocal of the "density" of the lattice Λ. We give a very brief sketch of this idea.

Let M be a bounded subset of $\mathbf{R}^n$ which has a volume and let $f(\Lambda) = |\Lambda \cap M|$ be the number of Λ-lattice points in M. Then we claim that

$$d = \lim_{\lambda \to \infty} \frac{f(\lambda M)}{V(\lambda M)}$$

exists and is independent of M; d is the **density** of Λ. First we note that by the map $y = Ax$, λM in y-space corresponds to the set $A^{-1}(\lambda M)$ in x-space, with volume $\frac{V(\lambda M)}{\Delta}$, and lattice points in λM correspond to integer points in $A^{-1}(\lambda M)$. Therefore $f(\lambda M)$ is the number of integer points in $A^{-1}(\lambda M)$ which, as λ gets large, is approximately $\frac{V(\lambda M)}{\Delta}$ (the proof of this is similar to our density argument in the first proof of Minkowski's theorem). Hence $\lim_{\lambda \to \infty} \frac{f(\lambda M)}{V(\lambda M)} = \frac{1}{\Delta}$; so Λ has a density, and $\Delta = \frac{1}{d}$ is independent of the basis.

Change of basis

Now we give a more rigorous proof that Δ is independent of the basis by studying how to change bases.

Let $\{\alpha_1, \ldots, \alpha_n\}$ and $\{\beta_1, \ldots, \beta_n\}$ be bases of Λ. Then Λ is given by $\{y = Ax\} = \{y = Bx\}$, where A (resp. B) is the matrix with column vectors α_i (resp. β_i). Since every β is an integral linear combination of the α's, we have $\beta_k = \sum_i \alpha_i p_{ik}$, for some $p_{ik} \in \mathbf{Z}$, and thus $B = AP$, where $P = (p_{ij})$ is an *integral* matrix ($p_{ij} \in \mathbf{Z}$). Similarly $A = BQ$, for some integral matrix Q. Therefore $A = APQ$ and, since $\det(A) \neq 0$, $PQ = I$. But $\det(P) = \frac{1}{\det(Q)}$ and $\det(P)$, $\det(Q)$ are integers. Hence $\det(P) = \det(Q) = \pm 1$ and we have

Theorem: *If the columns of A and B are bases for the lattice Λ, then $B = AP$, where P is a* **unimodular** *matrix (integral with determinant ± 1).*

Corollary: *Δ is independent of the basis chosen for Λ.*

Proof: $B = AP \implies |\det(B)| = |\det(A)|\,|\det(P)| = |\det(A)|$.

The converse of the last theorem is also true.

Theorem: *If the columns of A are a basis for Λ and P is unimodular, then the columns of $B = AP$ are a basis for Λ.*

Proof: (exercise — this follows from the fact that the inverse of a unimodular matrix is unimodular, which is proved by using the formula for the inverse of a matrix in terms of cofactors.)

MINKOWSKI'S THEOREM REFORMULATED

Let $\Lambda : y = Ax$ be a lattice of determinant Δ. Then A defines a linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^n$ (as we did before, we think of A as a map from x-space with integer points to y-space with Λ-points). Let K be a bounded, convex set in y-space, centered at the origin, with $V(K) > 2^n\Delta$. Then $A^{-1}K$ is a bounded, convex set in x-space, centered at the origin, with $V(A^{-1}K) > 2^n$ (a linear transformation preserves convexity, symmetry and boundedness). By Minkowski's theorem, $A^{-1}K$ contains a non-zero integer point g, and thus the Λ-point Ag is in K. Hence we have

Minkowski's Theorem for Lattices:

1) *A bounded convex set K centered at the origin with $V(K) > 2^n\Delta$ contains a non-zero lattice point of every lattice of determinant Δ.*

2) *Similarly, if $V(K) \geq 2^n\Delta$, then there is a non-zero lattice point in K or on its boundary.*

This form of the theorem suggested new questions to Minkowski. A lattice Λ is said to be **admissible** for a bounded symmetric convex set K if there are no non-zero Λ- points in K. Minkowski's theorem tells us

$$\det \text{ (admissible lattice for K)} \geq \frac{V(K)}{2^n}. \tag{1}$$

$\Delta(\mathbf{K})$, **the critical determinant of K**, is the greatest lower bound of $\det(\Lambda)$, taken over all admissible lattices for K. An admissible lattice Λ for K with $\det(\Lambda) = \Delta(K)$ is a **critical lattice** for K. For a given convex K (and for more general non-convex sets), the exact evaluation of $\Delta(K)$ and finding a critical lattice are central problems of the geometry of numbers, with strong number theoretic consequences. We do not pursue this theme here but refer to [Cas 4] and [Lek].

4. Back to Quadratic Forms

As described at the beginning of the last section, the problem of minimizing all positive definite quadratic forms $Q(x_1, \ldots, x_n)$ of determinant D over the integer lattice is changed to the problem of deciding how big λ must be so that the sphere $\{\sum y_i^2 < \lambda\}$ contains a point of every lattice $\Lambda : y = Ax$, for which A transforms a Q to $\sum y_i^2$.

By Minkowski's theorem for lattices, we need two pieces of information: (i) the volume of $\{\sum_{i=1^n} y_i^2 < \lambda\}$ and (ii) the determinant of Λ.

(i) Let $S_\lambda^n = \{\sum y_i^2 < \lambda\}$. A point y satisfies $\sum y_i^2 < 1$ if and only if $\sqrt{\lambda}\, y$ satisfies $\sum(\sqrt{\lambda}y_i)^2 = \lambda \sum y_i^2 < \lambda$. Hence $S_\lambda^n = \sqrt{\lambda}S_1^n$ and $V(S_\lambda^n) = \lambda^{\frac{n}{2}} V(S_1^n)$. But

$$V(S_1^n) = \frac{2\pi^{\frac{n}{2}}}{n\Gamma\left(\frac{n}{2}\right)}$$

where Γ is the gamma function (see Siegel [Sie 5, pp. 25 - 26]).

(ii) If $y = Ax$, with $\det(A) = \Delta$, transforms Q to $\sum y_i^2$, i.e., $\sum y_i^2 = Q(A^{-1}y)$, then it is known from linear algebra [Bor - Sha, supplement A] that

$$\det\left(\sum y_i^2\right) = \det(Q) \cdot \left(\det(x = A^{-1}y)\right)^2 ;$$

thus

$$1 = D \cdot (\Delta^{-1})^2$$

and

$$\det(\Lambda) = \Delta = D^{\frac{1}{2}} .$$

Hence we have $V(S_\lambda^n) \geq 2^n\Delta = 2^n D^{\frac{1}{2}}$, if $\frac{\lambda^{\frac{n}{2}}\pi^{\frac{n}{2}}}{\Gamma(1+\frac{1}{n})} \geq 2^n D^{\frac{1}{2}}$ or

$$\lambda \geq \frac{4}{\pi}\left(\Gamma\left(1+\frac{1}{n}\right)\right)^{\frac{2}{n}} D^{\frac{1}{n}} .$$

By Minkowski's theorem for lattices, this means that every lattice of determinant $D^{\frac{1}{2}}$ contains a non-zero lattice point y such that

$$\sum y_i^2 \leq \frac{4}{\pi}\left(\Gamma\left(1+\frac{1}{n}\right)\right)^{\frac{2}{n}} D^{\frac{1}{n}} .$$

Transforming back to x-space, we have

Theorem (Minkowski): *A positive definite quadratic form Q in n variables with real coefficients and determinant D assumes a value*

$$Q(x) \leq \frac{4}{\pi}\left(\Gamma\left(1+\frac{1}{n}\right)\right)^{\frac{2}{n}} D^{\frac{1}{n}},$$

for some non-zero integer point x.

The constant in Hermite's theorem (sec. 1) was $\left(\frac{4}{3}\right)^{\frac{n-1}{2}}$. Minkowski's is better for $n \geq 4$.

5. Sums of Two and Four Squares

In section 12.6, we proved

Theorem: *A prime $p \equiv 1 (\mathrm{mod}\, 4)$ is the sum of two integer squares, i.e., $p = \lambda_1^2 + \lambda_2^2$, for some $\lambda_1, \lambda_2 \in \mathbf{Z}$,*

and then generalized this to decide which positive integers are the sum of two squares.

Now we give a geometric proof of this theorem as a prototype for the corresponding theorem about four squares.

Proof: (i) For the moment, we assume that there exists a sublattice Λ of the integer lattice in $\mathbf{R}^2$, with $\det(\Lambda) = p$, such that

$$y_1^2 + y_2^2 \equiv 0 (\mathrm{mod}\, p) ,$$

for all (y_1, y_2) in Λ.

(ii) The circle $y_1^2 + y_2^2 < 2p$ has area

$$4\pi p^2 > 4p = 4 \cdot \det(\Lambda) ,$$

and thus, by Minkowski's theorem for lattices, there is a Λ point $(\lambda_1, \lambda_2) \neq (0, 0)$ in the circle. Therefore we have

$$0 < \lambda_1^2 + \lambda_2^2 < 2p$$

as well as

$$\lambda_1^2 + \lambda_2^2 \equiv 0 (\mathrm{mod}\, p) .$$

Since $\lambda_1^2 + \lambda_2^2$ is a multiple of p which is strictly between 0 and $2p$, it must equal p.

Now we construct Λ and then we are done. For any integer u, let Λ_u be the lattice

$$\begin{aligned} y_1 &= px_1 + ux_2 \\ y_2 &= \qquad x_2 , \end{aligned}$$

where $(x_1, x_2) \in \mathbf{Z}^2$; then $\det(\Lambda_u) = p$. Since $y_1 \equiv uy_2 (\mathrm{mod}\, p)$, we have

$$y_1^2 + y_2^2 \equiv (u^2 + 1) y_2^2 (\mathrm{mod}\, p) ,$$

for all $(y_1, y_2) \in \Lambda_u$. Since $p \equiv 1 (\mathrm{mod}\, 4)$, -1 is a quadratic residue modulo p and there is an integer w such that $w^2 + 1 \equiv 0 (\mathrm{mod}\, p)$. Therefore $\Lambda = \Lambda_w$ is the desired lattice.

Now we use the same idea to prove

Theorem: *(Lagrange): Every positive integer m is the sum of four integer squares,*

$$m = \lambda_1^2 + \lambda_2^2 + \lambda_3^2 + \lambda_4^2 , \quad \lambda_i \in \mathbf{Z} .$$

As in the case of two squares, we can reduce the proof of the theorem to the result for primes by noting that if two integers are each the sum of four squares, then so is their product. This follows from the algebraic identity

$$\begin{aligned} &(x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) \\ &\quad = (x_1 y_1 - x_2 y_2 - x_3 y_3 - x_4 y_4)^2 + (x_1 y_2 + x_2 y_1 + x_3 y_4 - x_4 y_3)^2 \\ &\quad + (x_1 y_3 - x_2 y_4 + x_3 y_1 + x_4 y_2)^2 + (x_1 y_4 + x_2 y_3 - x_3 y_2 + x_4 y_1)^2 . \end{aligned}$$

Just as the analogous identity for sums of two squares expresses the multiplicative property of the norm (absolute value) for complex numbers, this identity expresses the multiplicative property of the norm for the quaternions [Har - Wri].

Now we prove

Theorem (Lagrange): *Every prime p is the sum of four integer squares,*

$$p = \lambda_1^2 + \lambda_2^2 + \lambda_3^2 + \lambda_4^2 , \lambda_i \in \mathbf{Z} ,$$

Proof: (i) For the moment, we assume there exists a sublattice Λ of the integer lattice in $\mathbf{R}^4$, with $\det(\Lambda) = p^2$, such that

$$y_1^2 + y_2^2 + y_3^2 + y_4^2 \equiv 0 (\mathrm{mod}\, p) ,$$

for all (y_1, y_2, y_3, y_4) in Λ.

(ii) From section 4, we know that the solid sphere $S_{2p}^{(4)} = \{y_1^2 + y_2^2 + y_3^2 + y_4^2 < 2p\}$ has volume

$$\frac{1}{2}\pi^2 (2p)^2 = 2p^2\pi^2 > 2^4 p^2 = 2^4 \cdot \det(\Lambda) ,$$

and thus, by Minkowski's theorem for lattices, there is a non-zero Λ-point $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ in this sphere. Therefore we have

$$0 < \lambda_1^2 + \lambda_2^2 + \lambda_3^2 + \lambda_4^2 < 2p$$

as well as

$$\lambda_1^2 + \lambda_2^2 + \lambda_3^2 + \lambda_4^2 \equiv 0 (\mathrm{mod}\ p)\ .$$

Since $\lambda_1^2 + \lambda_2^2 + \lambda_3^2 + \lambda_4^2$ is a multiple of p which is strictly between 0 and $2p$, it must equal p, i.e., $p = \lambda_1^2 + \lambda_2^2 + \lambda_3^2 + \lambda_4^2$.

Now we construct Λ and then we are done. For any integers u, v, let $\Lambda_{u,v}$ be the lattice

$$\begin{aligned} y_1 &= \quad + x_3 \\ y_2 &= \quad\quad + x_4 \\ y_3 &= px_1 \quad + ux_3 + vx_4 \\ y_4 &= \quad px_2 - vx_3 + ux_4\ , \end{aligned}$$

where $(x_1, x_2, x_3, x_4) \in \mathbf{Z}^4$; then $\det(\Lambda_{u,v}) = p^2$. A direct computation, which can be simplified by noting that $y_3 \equiv uy_1 + vy_2 (\mathrm{mod}\ p)$ and $y_4 \equiv -vy_1 + uy_2 (\mathrm{mod}\ p)$, yields

$$y_1^2 + y_2^2 + y_3^2 + y_4^2 \equiv (u^2 + v^2 + 1)(y_1^2 + y_2^2) (\mathrm{mod}\ p)\ ,$$

for all $(y_1, y_2, y_3, y_4) \in \Lambda_{u,v}$. If we can find integers t, w such that $t^2 + w^2 + 1 \equiv 0 (\mathrm{mod}\ p)$, then $\Lambda = \Lambda_{t,w}$ is the required lattice.

By Chevally's theorem (sec. 9.6), $x^2 + y^2 + z^2 \equiv 0 (\mathrm{mod}\ p)$ has a non-zero solution $(a, b, c) (\mathrm{mod}\ p)$; say $c \not\equiv 0 (\mathrm{mod}\ p)$. Then c^{-1} exists modulo p and $(ac^{-1})^2 + (bc^{-1})^2 + 1 \equiv 0 (\mathrm{mod}\ p)$. Setting $t = ac^{-1}$ and $w = bc^{-1}$, we have $t^2 + w^2 + 1 \equiv 0 (\mathrm{mod}\ p)$ and we are done.

There is also a quick direct proof of this last result. If $p = 2$, then $1^2 + 0^2 + 1 \equiv 0 (\mathrm{mod}\ 2)$. If p is odd then $\left\{u^2 \mid 0 \le u \le \frac{p}{2}\right\}$ and $\left\{-1 - v^2 \mid 0 \le v \le \frac{p}{2}\right\}$ are sets of $\frac{p+1}{2}$ integers such that the elements of each set are not congruent modulo p. But there are only p congruence classes modulo p; so there is an integer s congruent to an integer of each set, i.e., there are t, w such that $t^2 \equiv s \equiv -1 - w^2 (\mathrm{mod}\ p)$ or $t^2 + w^2 + 1 \equiv 0 (\mathrm{mod}\ p)$.

There is a common technique behind the proofs of the two and four square theorem, namely, if we take the integer lattice $\mathbf{Z}^n$ and impose m homogeneous linear congruence conditions on the coordinates modulo $k_1, \ldots, k_n$ respectively, then the set of integer points satisfying these conditions is a sublattice of determinant $\le k_1 k_2 \cdots k_n$. Thus for the two square theorem we set $y_1 \equiv uy_2 (\mathrm{mod}\ p)$ to get the lattices Λ_u. For the four square theorem, we set $y_3 \equiv uy_1 + vy_2 (\mathrm{mod}\ p)$ and $y_4 \equiv -vy_1 + uy_2 (\mathrm{mod}\ p)$.

This technique can also be used to prove Legendre's theorem, which yields an algorithm for deciding if a conic contains a point with rational coordinates (sec. 18.3 and [Cas 4, sec. III.7]). However, although we can prove three important theorems using this technique, it has not yet proved capable of yielding more general theorems on the representation of integers by quadratic forms. Thus, for now, it remains a special trick and not a general method.

6. Linear Forms

Minkowski crafted the geometry of numbers into a valuable tool for number theorists. After applying the theory to quadratic forms, he then applied his fundamental theorem to systems of linear forms yielding basic results in algebraic number theory such as Dirichlet's characterization of units in a number field and the finiteness of the class number (see [Hec]). More importantly, he proved a conjecture of Kronecker on 'discriminants' of number fields, which we shall treat in the next section. First we discuss linear forms.

The **box**, $B = \{|y_1| < \lambda_1, \ldots, |y_n| < \lambda_n\}$, is a convex set in $\mathbf{R}^n$ symmetric about the origin, with $V(B) = 2^n \lambda_1 \cdots \lambda_n$. This is intuitive when viewed geometrically and not difficult to prove directly (exercise). We shall provide a general method of proving such results in section 8.

By Minkowski theorem for lattices (sec. 3), a lattice Λ of determinant Δ has a point inside the box if

$$2^n \lambda_1 \cdots \lambda_n > 2^n \Delta\ ,$$

i.e.,

$$\lambda_1 \cdots \lambda_n > \Delta\ .$$

Thus, recalling that a lattice, $\Lambda : y = Ax$, is given by linear forms in x_i, we have

Minkowski's Linear Form Theorem: *If* $y = Ax$, $A = (a_{ij})$, $a_{ij} \in \mathbf{R}$, $\Delta = |\det(A)| \neq 0$, *and if* $\lambda_1, \ldots, \lambda_n > 0$ *satisfy* $\lambda_1 \cdots \lambda_n > \Delta$, *then there*

is a non-zero integer point $(x_1, \ldots, x_n)$ *such that*

$$|y_1| < \lambda_1, \ldots, |y_n| < \lambda_n ,$$

i.e.,

$$|a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n| < \lambda_i , \quad \textit{for all } i.$$

If $\lambda_1 \cdots \lambda_n \geq \Delta$, the result holds with $|y_i| < \lambda_i$ replaced by $|y_i| \leq \lambda_i$, for all i. However, we can do a little more.

Corollary: *If* $\lambda_1 \cdots \lambda_n = \Delta$, *then there is a non-zero integer point such that*

$$|y_1| \leq \lambda_1, |y_2| < \lambda_2, \ldots, |y_n| < \lambda_n ,$$

and similarly for any other λ_i.

Proof: Apply the second form of Minkowski's theorem (sec.2) to the box $\{|y_1| < (1+\epsilon)\lambda_1, |y_2| < \lambda_2, \ldots, |y_n| < \lambda_n\}$ and let $\epsilon \to 0$. This is the same continuity argument used to prove the third form of Minkowski's theorem (sec. 2).

Example: Diophantine Approximation — In sections 4.8 and 21.2, we proved Lagrange's theorem, viz., that for α irrational, there are infinitely many distinct rational solutions $\frac{p}{q}$ of $\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^2}$. Now we give yet another proof using the linear forms theorem.

Proof: Let

$$\begin{aligned} y_1 &= x_1 - \alpha x_2 \\ y_2 &= \qquad x_2 ; \end{aligned}$$

thus $\Delta = 1$. Let $\lambda_1 = \frac{1}{k}$ and $\lambda_2 = k$; so $\lambda_1\lambda_2 = 1$. By the corollary, there is an integer point $(x_1, x_2) \neq (0, 0)$ such that

$$\begin{aligned} |x_1 - \alpha x_2| &< \frac{1}{k} \\ |x^2| &\leq k . \end{aligned}$$

Now if we let $k = 1, 2, \ldots$, there are infinitely many integer solutions to the pair of inequalities; otherwise $|x_1 - \alpha x_2| > k'$ for some k' and $x_1, x_2 \in \mathbf{Z}$, contradicting $|x_1 - \alpha x_2| < \frac{1}{k}$ for k sufficiently large. Furthermore, if x_2 takes any fixed value, say m, there are only finitely many integer solutions,

since only finitely many integers x_1 can satisfy $|x_1 - \alpha m| < 1$ (or $\leq \frac{1}{k}$). Hence there is a sequence $\left(x_1^{(1)}, x_2^{(1)}\right), \left(x_1^{(2)}, x_2^{(2)}\right), \ldots$, with $x_2^{(i)} \to \infty$, such that

$$\left|\frac{x_1^{(i)}}{x_2^{(i)}} - \alpha\right| < \frac{1}{\left|x_2^{(i)}\right|^k} \leq \frac{1}{\left(x_2^{(i)}\right)^2} .$$

Since $x_2^{(i)} \to \infty$, there are infinitely many distinct rationals among the $\frac{x_1^{(i)}}{x_2^{(i)}}$ and we are done.

Minkowski also derived the theory of continued fractions (and the associated theory of indefinite binary quadratic forms) geometrically with the linear form theorem (see [Min 1]). Starting with Klein's interpretation of continued fractions in terms of lattice points in the plane (sec. 4.7), he made a detailed study of the distribution of lattice points on the boundaries of rectangles whose sides are defined by the linear forms $|ax + by| = k_1$ and $|cx + dy| = k_2$. Minkowski expressed considerable pleasure about his proof of periodicity for quadratic irrationalities, which he regarded as the most natural proof of this theorem. This geometric approach was also the basis of Minkowski's generalization of continued fractions [Min 2].

7. Sums and Products of Linear Forms; The Octahedron

The set

$$K = \left\{ \sum_1^n |y_i| < \lambda \right\}$$

is a bounded convex set in $\mathbf{R}^n$, symmetric with respect to the origin (see the next section for a method of proof). For $n = 2$, we have a square (fig. 6), and for $n = 3$, an octahedron. For general n, we also call K the (generalized) **octahedron**. To find its volume, note that the octahedron consists of 2^n congruent pieces, one in each octant. The volume of the piece in the positive octant (all $y_i > 0$) is

$$\lambda^n \int_0^1 \int_0^{1-y_1} \cdots \int_0^{1-y_1-\cdots-y_n} dy_1 dy_2 \cdots dy_n = \frac{\lambda^n}{n!} ;$$

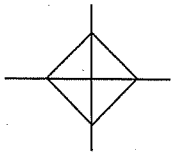

Figure 6

hence $V(K) = 2^n \frac{\lambda^n}{n!}$. If a lattice Λ, with determinant Δ, satisfies

$$2^n \frac{\lambda^n}{n!} \geq 2^n \Delta \quad \text{or} \quad \lambda \geq (n!\Delta)^{\frac{1}{n}} ,$$

then K contains a non-zero point of Λ. Thus, by Minkowski's theorem for lattices, we have

Theorem (Sums of Linear Forms): *If* $y = Ax$, $A = (a_{ij})$, $a_{ij} \in \mathbf{R}$, $\Delta = |\det(A)| \neq 0$, *then there exists a non-zero integer point* $(x_1, \ldots, x_n)$ *such that*

$$|y_1| + \cdots + |y_n| \leq (n!\Delta)^{\frac{1}{n}} .$$

Example: For $n = 2$, the theorem guarantees the existence of a non-zero integer point (x_1, x_2) satisfying

$$|ax_1 + bx_2| + |cx_1 + dx_2| \leq \sqrt{2}\Delta ,$$

where $\Delta = |ad - bc| \neq 0$. As Siegel points out [Sie 5], the case $a = \sqrt{7}$, $b = \sqrt{6}$, $c = \sqrt{15}$, $d = \sqrt{13}$ shows that this is a non trivial result.

Products of linear forms can be reduced to sums by means of the inequality for the arithmetic and geometric means, namely,

$$|y_1 \cdots y_n|^{\frac{1}{n}} \leq \frac{1}{n}(|y_1| + \cdots + |y_n|)$$

or

$$|y_1 \cdots y_n| \leq \frac{1}{n^n}(|y_1| + \cdots + |y_n|)^n .$$

Applying the last theorem, we now have

Theorem (Products of Linear Forms): *Under the same assumptions as in the last theorem, there is a non-zero integer points* $(x_1, \ldots, x_n)$*such that*

$$|y_1 \cdots y_n| \leq \frac{n!\Delta}{n^n} .$$

Example: Discriminants of Algebraic Number Fields — As we mentioned earlier, Minkowski used the geometry of numbers to prove a conjecture of Kronecker, viz.,

the 'discriminant' of a number field ($\neq \mathbf{Q}$) is greater than one.

We defined the notions needed to understand this result in the context of quadratic fields (chapters 16 and 17), where the result is trivial (sec. 17.8). Now we quickly review the basic ideas needed to understand the general conjecture and prove a special case of Minkowski's result.

A complex number θ is an **algebraic number** (of **degree n**) if it is the root of an irreducible polynomial $\sum_0^n a_i x^i$, with integer coefficients. If $a_n = 1$, θ is an **algebraic integer**. An **algebraic number field** K (or just **number field**) is a subfield of $\mathbf{C}$ which is a finite extension of the rationals, i.e., $n = \deg(K/\mathbf{Q}) < \infty$.

Every number field K of degree n over $\mathbf{Q}$ is **generated** by an algebraic number θ of degree n, in the sense that

$$K = \mathbf{Q}(\theta) = \{u_0 + u_1\theta + \cdots + u_{n-1}\theta^{n-1} | u_i \in \mathbf{Q}\} ,$$

i.e., if $\omega \in K$, then $\omega = q(\theta)$, for some $q(x) \in \mathbf{Q}[x]$, deg $q \leq n$. If $p(x)$ is the irreducible equation for θ, with roots $\theta^{(0)} = \theta, \theta^{(1)}, \ldots, \theta^{(n-1)}$ (the **conjugates** of θ), then $\omega^{(0)} = \omega = q(\theta)$, $\omega^{(1)} = q(\theta^{(1)}), \ldots, \omega^{(n-1)} = q(\theta^{(n-1)})$ are the **conjugates of** ω **in K** (they are independent of the choice of which element generates K).

$\mathbf{I_K}$, the set of algebraic integers of K, is a ring with an **integral basis**, i.e., there exist $\omega_0, \ldots, \omega_{n-1} \in I_K$ such that

$$I_K = \left\{ \sum_0^{n-1} v_i \omega_i | v_i \in \mathbf{Z} \right\} .$$

With this basis, let $\omega_i^{(j)}$, $j = 0, \ldots, n-1$, be the n conjugates of ω_i, with $\omega_i^{(0)} = \omega_i$, and consider the linear forms

$$\begin{aligned}
\xi^{(0)} &= x_0\omega_0^{(0)} + \cdots + x_{n-1}\omega_{n-1}^{(0)} \\
\xi^{(1)} &= x_0\omega_0^{(1)} + \cdots + x_{n-1}\omega_{n-1}^{(1)} \\
&\vdots \\
\xi^{(n-1)} &= x_0\omega_0^{(n-1)} + \cdots + x_{n-1}\omega_{n-1}^{(n-1)} .
\end{aligned}$$

The **discriminant of K** is defined by $\mathbf{d_K} = [\det(\omega_i^{(j)})]^2$. The following properties will prove useful:

(i) d_K is a non-zero rational integer, independent of the choice of integral basis,

(ii) if $K = \mathbf{Q}(\theta)$ is **totally real** (θ and all its conjugates are real) then d_K is a positive integer,

(iii) the product $\xi^{(0)} \cdots \xi^{(n-1)}$ is a non-zero integer for any non-zero integer point $(x_0, \ldots, x_{n-1})$.

Now we prove Kronecker's conjecture for totally real fields by finding a lower bound for d_K. The proof for all number fields, given in [Min 1] and [Rib], is not much harder.

By the theorem for products of linear forms applied to the $\xi^{(i)}$, with $\Delta = \sqrt{d_K}$, there exists a non-zero integer point $(x_0, \ldots, x_{n-1})$ such that

$$|\xi^{(0)} \cdots \xi^{(n-1)}| \leq \frac{n!\sqrt{d_K}}{n^n} .$$

By (iii), the left hand side is a positive integer and therefore greater than or equal to one. Hence $1 \leq \frac{n!\sqrt{d_K}}{n^n}$ or $d_K \geq \left(\frac{n^n}{n!}\right)^2$, and $d_K > 1$ for $n > 1 (K \neq \mathbf{Q})$.

Corollary: *The discriminant of a totally real number field $K(\neq \mathbf{Q})$ is divisible by at least one rational prime.*

This has the same significance for general number fields as in the quadratic case (sec. 17.6, 17.12), namely, the existence of 'ramified primes'.

It should be noted that the only known general construction for finding n linear forms whose product is not very small is by taking a form, whose coefficients are in a number field, and constructing the other forms by taking conjugates, as in our example.

8. Gauge Functions; The Equation of a Convex Body

Minkowski introduced the gauge function of a convex set which provides an analytic description of the set. By characterizing which functions are gauge functions, he also produced a general procedure which can often be used to prove that a set is convex. The gauge function also provides a direct way of stating our arithmetic results on minimization of functions (or sets of functions) over integer points, instead of having to translate from geometry to arithmetic. As before, this new point of view leads to yet another formulation of Minkowski's theorem and to new generalizations.

We shall only consider **convex bodies**, i.e., bounded open convex sets containing the origin (*warning*: there is some variation in this terminology among other authors). Since the interior of a convex set is an open convex set, this is not a real restriction. $\partial\mathbf{B}$ will denote the **boundary** of B.

Let B be a convex body in $\mathbf{R}^n$. The **gauge function** of B is a function $f : \mathbf{R}^n \to [0, \infty)$ defined as follows:

(i) $f(\mathbf{0}) = 0$,

(ii) for $x \neq \mathbf{0}$, $f(x) = \frac{\|x\|}{\|x'\|}$, where x' is the intersection of the ray from $\mathbf{0}$ to x with the boundary of B (fig. 7), and $\|y\| = \sqrt{\sum y_i^2}$ denotes the length of $y = (y_1, \ldots, y_n)$.

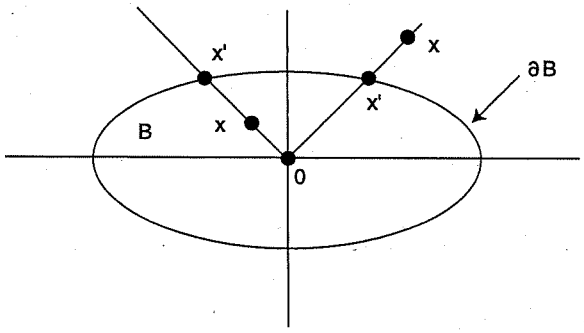

Figure 7

Equivalently, if $x = \mu x'$, $\mu > 0$, $x' \in \partial B$, then $f(x) = \mu$. Note that

(a) $f(x) < 1 \iff x \in B$,
(b) $f(x) = 1 \iff x \in \partial B$,
(c) $f(x) \leq 1 \iff x \in \overline{B}$,

where $\overline{\mathbf{B}} = B \cup \partial B$, the **closure** of B.

Thus (a) can be regarded as an equation (or, more precisely, an inequality) for B and (b) is an equation for the boundary.

The gauge function measures the distance from $\mathbf{0}$ to x with $\|x'\|$ as the unit. If we set $\rho(x, y) = \|x - y\|$, then ρ is a metric on $\mathbf{R}^n$. These metrics are a special class of the so called 'projective metrics', and the 'Minkowski geometry' induced by them is studied in Busemann and Kelly [Bus - Kel].

Theorem: *If B is a convex body with gauge function $f(x)$, then*

(i) $f(x) > 0$ *for* $x \neq \mathbf{0}$, $f(\mathbf{0}) = 0$,
(ii) $f(\lambda x) = \lambda f(x)$, *for* $\lambda \in \mathbf{R}$, $\lambda \geq 0$,
(iii) $f(x + y) \leq f(x) + f(y)$.

Moreover f is continuous.

Proof: (i) follows immediately from the definition of f. To prove (ii), suppose that $x = \mu x'$, $\mu > 0$, $x' \in \partial B$; so f$(x) = \mu$. Then $f(\lambda x) = f(\lambda \mu x') = \lambda \mu = \lambda f(x)$.

The triangle inequality (iii) takes more work. If x or y is $\mathbf{0}$, then it is obvious. Otherwise, we normalize by considering $x' = \frac{x}{f(x)}$, $y' = \frac{y}{f(y)}$; so $f(x') = f(y') = 1$ and $x', y' \in \partial B$. Since $\overline{B}$ is convex (we assume this), we have

$$f(rx' + sy') \leq 1 \,, \tag{1}$$

for any real r and s satisfying $r, s > 0$ and $r + s = 1$. Setting $r = \frac{f(x)}{f(x)+f(y)}$ and $s = \frac{f(y)}{f(x)+f(y)}$ in (1), we have

$$f\left(\frac{x+y}{f(x)+f(y)}\right) \leq 1 \implies \left(\frac{1}{(f(x)+f(y)}\right) f(x+y) \leq 1$$
$$\implies f(x+y) \leq f(x) + f(y) \,.$$

The continuity follows from properties (i) – (iii) and will be proved as part of the next theorem.

Symmetry is incorporated into the gauge function by

Proposition: *B is symmetric $\iff$ f is an even function ($f(x) = f(-x)$).*

Proof: $\implies$) $x = \lambda x', x' \in \partial B \implies -x = \lambda(-x'), -x \in B$ (by symmetry) $\implies f(x) = f(-x) = \lambda$.

$\impliedby$) $x \in B \implies f(x) < 1 \implies f(-x) = f(x) < 1 \implies -x \in B$.

Exercises: (I) Let B be the square in $\mathbf{R}^2$, centered at the origin with sides of length 2 parallel to the axes. Show that the gauge function of B is $f(x) = \max\{|x_1|, |x_2|\}$, where $x = (x_1, x_2)$. Find the gauge function for the general box $\{|x_1| < \lambda_1, |x_2| < \lambda_2\}$.

(II) Let B be the interior of the ellipse $\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$. Show that the gauge function $f(x) = \sqrt{\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2}}$.

The general idea for finding a gauge function of a convex body B is to find an equation for the boundary of B of the form $g(x) = 1$ and then, if necessary, fiddle with g to make it homogeneous (i.e., to satisfy part (ii) of the theorem). Thus in exercise (II), we have $g(x) = \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2}$ and taking the square root makes it homogeneous. $\sqrt{g(x)}$ clearly satisfies part (i) of the theorem, but it is still necessary to verify (iii), the triangle inequality.

In many of our earlier applications we proceeded in the opposite direction, beginning with a function f and assuming that the set $\{f < \lambda\}$ is a convex body. The convexity can often be proved by the converse to the last theorem.

Theorem: *If a function $f : \mathbf{R}^n \to [0, \infty)$ satisfies*

(i) $f(x) > 0$ *for* $x \neq \mathbf{0}$, $f(\mathbf{0}) = 0$,
(ii) $f(\lambda x) = \lambda f(x)$, *for* $\lambda \in \mathbf{R}$, $\lambda \geq 0$,
(iii) $f(x + y) \leq f(x) + f(y)$,

then f is continuous, $B = \{x | f(x) < 1\}$ is a convex body, and f is the gauge function of B.

Proof: 1) f is continuous — First we prove continuity at $\mathbf{0}$. If $e_1 = (1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1)$ denote the unit vectors, then every x

in $\mathbf{R}^n$ can be written in the form

$$x = \sum_1^n \lambda_i e_i = \sum_1^n |\lambda_i|(\pm e_i) .$$

Hence, by (i) – (iii),

$$0 \le f(x) \le \sum f(|\lambda_i|(\pm e_i)) = \sum |\lambda_i| f(\pm e_i) .$$

Therefore, as $x \to \mathbf{0}$, $|\lambda_i| \to 0$, $f(x) \to 0 = f(\mathbf{0})$, and f is continuous at $\mathbf{0}$.

To prove continuity at $x \neq \mathbf{0}$, we write $x = (x+y)+(-y)$, for any $y \neq \mathbf{0}$. Then $f(x) \le f(x+y)+f(-y)$ or $-f(-y) \le f(x+y)-f(x)$. From $f(x+y) \le f(x)+f(y)$, we have $f(x+y)-f(x) \le f(y)$. Combining these inequalities, we have

$$-f(-y) \le f(x+y)-f(x) \le f(y) .$$

By continuity at $\mathbf{0}$, $\lim_{y\to\mathbf{0}} f(y) = \lim_{y\to\mathbf{0}} f(-y) = 0$ and thus f is continuous at x.

2) B is open since it is the pre-image of an open set under the continuous map f.

3) B is convex — Let $x, y \in B$; so $f(x) < 1$, $f(y) < 1$. For $\lambda, \mu > 0$, $\lambda+\mu = 1$, we have $f(\lambda x+\mu y) \le f(\lambda x)+f(\mu y) = \lambda f(x)+\mu f(y) < \lambda+\mu = 1$. Hence $\lambda x+\mu y \in B$ and B is convex.

4) B is bounded — Assume B is unbounded. Then there exists a sequence of points in B, say $x_0, x_1, \ldots$, such that

$$f(x_n) < 1 \text{ and } \|x_n\| \to \infty, \text{ as } n \to \infty .$$

Let $\lambda_n = \frac{1}{\|x_n\|}$. Then as $n \to \infty$,

$$f(\lambda_n x_n) = \frac{f(x_n)}{\|x_n\|} < \frac{1}{\|x_n\|} \to 0 .$$

But $\|\lambda_n x_n\| = 1$ and thus $\lambda_n x_n$ lies on the unit sphere in $\mathbf{R}^n$, which is a compact set. Since f is a continuous function on a compact set, it attains a minimum $f(x')$ at some point x' on the set and, by (i), $f(x') > 0$. This contradicts $f(\lambda_n x_n) \to 0$.

5) f is the gauge function of B (exercise).

Exercise: Show that $f(x) = (\sum |x_i|^r)^{\frac{1}{r}}$, $x \in \mathbf{R}^n$, $r > 1$, is an even gauge function. The triangle inequality is known as the Minkowski inequality (see [Sie 5] or [Har - Lit - Pol]) and this setting is probably the reason Minkowski studied this inequality. If

$r = 1$, we have the generalized octahedron,

$r = 2$, we have the ellipsoid,

$r \to \infty$, we have the square box (all sides equal — why?).

Why does the convexity of the square box imply the convexity of any box?

Now we reformulate Minkowski's fundamental theorem for a symmetric convex body B in $\mathbf{R}^n$ with gauge function f. For any $x \in \partial(\lambda B)$, $f(x) = \lambda$. First we choose λ small enough so that λB doesn't contain any non-zero integer points. Then we increase λ until λB contains an integer point on its boundary and none ($\neq \mathbf{0}$) inside; suppose this occurs at $\lambda = \mu$. Then, since $f(\partial(\lambda B))$ is an increasing function of λ, we have

$$\mu = \text{Min } f(g) ,$$

where the minimum is over all *non-zero* integer points g. Thus we have shown that the minimum is actually achieved at an integer point. We call μ the **first minimum of f** (or B). Since $V(\lambda B) = \lambda^n V(B)$, we have

Minkowski's Theorem:

a) *(geometric form) If B is a bounded symmetric convex body in $\mathbf{R}^n$ centered at the origin, and μ is its first minimum, then*

$$\mu^n V(B) \le 2^n .$$

b) *(arithmetic form) If $f : \mathbf{R}^n \to [0, \infty)$ is a function satisfying the conditions of the last theorem, then both the volume $V(B)$ of $B = \{x | f(x) < 1\}$ and the minimum μ of f over all non-zero integer points exist, and $\mu^n V(B) \le 2^n$.*

The same reformulation can be carried out for convex bodies with respect to any lattice L of determinant Δ. We define the first minimum μ_L of B (with respect to L) in a similar way and μ_L is the minimum of f over all non-zero points of L. Then Minkowski's theorem for lattices becomes $\mu_L^n V(B) \le 2^n \Delta$ and there is a similar arithmetic form.

Minkowski's theorem allows us to state our arithmetic theorems directly, since bounds on the first minimum have always been our arithmetic

objective. For example, in the theorem on sums of linear forms (sec. 7), L is given by $y = Ax$, $f(y) = \sum |y_i|$, the volume of $B = \{y | f(y) < 1)$ is $\frac{2^n}{n!}$, and thus $\mu_L \leq \left(\frac{2^n \Delta}{V(B)}\right)^{\frac{1}{n}} = (n!\Delta)^{\frac{1}{n}}$.

9. Successive Minima

The formulation of Minkowski's theorem as presented in the last section leads to a natural generalization, which was also given by Minkowski.

Let f be an even gauge function defined on $\mathbf{R}^n$ and $B = \{x | f(x) < 1\}$ the corresponding symmetric convex body. We have defined the first minimum of B; for the moment we denote it by α_1. Let $g^{(1)}$ be an integer point in $\partial(\alpha_1 B)$. Suppose there are a maximum of k_1 integer points on $\partial(\alpha_1 B)$, say $g^{(1)}, \ldots, g^{(k_1)}$, which are *linearly independent* over $\mathbf{R}$ (the choice is not necessarily unique); so their f values are all α_1. Now let λ increase from α_1 until $\partial(\lambda B)$ contains an integer point independent of $g^{(1)}, \ldots, g^{(k_1)}$, say for $\lambda = \alpha_2$. Choose a maximum number of integer points on $\partial(\alpha_2 B)$, say $g^{(k_1+1)}, \ldots, g^{(k_2)}$, so that $g^{(1)}, \ldots, g^{(k_2)}$ are linearly independent; the f values of the new points are all α_2. Continue this process by again letting λ increase from α_2 and so on. Since, for λ sufficiently large, λB will contains n linearly independent vectors, our process must stop with the selection of n independent integer points $g^{(1)}, \ldots, g^{(n)}$. Setting $\mu_i = f(g^{(i)})$, we have, by our construction,

$$\mu_1 \leq \mu_2 \leq \cdots \leq \mu_n.$$

The μ_i are called the **successive minima** of B (μ_1 is the first minimum).

Minkowski's Theorem for Successive Minima: *If $\mu_1, \ldots, \mu_n$ are the successive minima of the symmetric convex body B, then*

$$\mu_1 \mu_2 \cdots \mu_n V(B) \leq 2^n .$$

This result is considerably stronger than the basic Minkowski theorem, $\mu_1^n V(B) \leq 2^n$, and has correspondingly stronger arithmetic implications (see [Min 1], [Cas 4] and [Lek]).

We shall not prove the theorem. Minkowski's original proof [Min 1] is quite difficult. Weyl [Wey 4] and Davenport [Dav 4] later gave simpler proofs, but the theorem remains quite deep. Siegel [Sie 5] gives a proof, together with a nice discussion as to why the obvious approach to such a proof doesn't work.

10. Other Directions

We have presented the genesis, proof and many applications of Minkowski's fundamental theorem. Under his brilliant hand these results led to a general theory carried on in the 20th century by such luminaries as Mordell, Weyl, Mahler and Davenport. The theory now includes deeper studies of quadratic forms (including indefinite forms), Mordell's generalizations to non convex sets, packings of $\mathbf{R}^n$ by convex sets and applications to Diophantine approximation, where, e.g., it plays an important role in Schmidt's generalization of the Thue–Siegel–Roth theorem (chap. 21).

Siegel's lectures [Sie 5] provides a good introduction, as well as some of the deeper results on reduction of forms. The books by Cassels' [Cas 4] and Lekkerkerker [Lek] provide comprehensive treatments with exhaustive bibliographies. Minkowski [Min 1] is quite difficult but his more elementary "Diophantische Approximationen" [Min 3] is somewhat more readable and a rich source of the ideas and conjectures which have guided the development of the field. Hancock [Han] gives English translations of substantial parts of Minkowski's books and papers (without explicitly saying so). Unfortunately, his confusing commentaries are often interweaved into Minkowski's text.