

# A THEOREM OF MINKOWSKI; THE FOUR SQUARES THEOREM

PETE L. CLARK

## 1. MINKOWSKI'S CONVEX BODY THEOREM

### 1.1. Introduction.

We have already considered instances of the following type of problem: given a bounded subset  $\Omega$  of Euclidean space  $\mathbb{R}^N$ , to determine  $\#(\Omega \cap \mathbb{Z}^N)$ , the number of integral points in  $\Omega$ . It is clear however that there is no answer to the problem in this level of generality: an arbitrary  $\Omega$  can have any number of lattice points whatsoever, including none at all.

In [Gauss's Circle Problem], we counted lattice points not just on  $\Omega$  itself but on dilates  $r\Omega$  of  $\Omega$  by positive integers  $r$ . We found that for any "reasonable"  $\Omega$ ,

$$(1) \quad L_\Omega(r) := \#(r\Omega \cap \mathbb{Z}^N) \sim r^N \text{Vol}(\Omega).$$

More precisely, we showed that this holds for all bounded sets  $\Omega$  which are **Jordan measurable**, meaning that the characteristic function  $\mathbf{1}_\Omega$  is Riemann integrable.

It is also natural to ask for sufficient conditions on a bounded subset  $\Omega$  for it to have lattice points at all. One of the first results of this kind is a theorem of Minkowski, which is both beautiful in its own right and indispensably useful in the development of modern number theory (in several different ways).

Before stating the theorem, we need a bit of terminology. Recall that a subset  $\Omega \subset \mathbb{R}^N$  is **convex** if for all pairs of points  $P, Q \in \Omega$ , also the entire line segment

$$\overline{PQ} = \{(1-t)P + tQ \mid 0 \leq t \leq 1\}$$

is contained in  $\Omega$ . A subset  $\Omega \subset \mathbb{R}^N$  is **centrally symmetric** if whenever it contains a point  $v \in \mathbb{R}^N$  it also contains  $-v$ , the reflection of  $v$  through the origin.

A **convex body** is a nonempty, bounded, centrally symmetric convex set.

Some simple observations and examples:

- i) A subset of  $\mathbb{R}$  is convex iff it is an interval.
- ii) A regular polygon together with its interior is a convex subset of  $\mathbb{R}^2$ .
- iii) An open or closed disk is a convex subset of  $\mathbb{R}^2$ .
- iv) Similarly, an open or closed ball is a convex subset of  $\mathbb{R}^N$ .

---

Thanks to Laura Nunley and Daniel Smitherman for finding typos.

- v) If  $\Omega$  is a convex body, then  $\exists P \in \Omega$ ; then  $-P \in \Omega$  and  $0 = \frac{1}{2}P + \frac{1}{2}(-P) \in \Omega$ .  
vi) The open and closed balls of radius  $r$  with center  $P$  are convex bodies iff  $P = 0$ .

Warning: The term “convex body” often has a similar but slightly different meaning: e.g., according to Wikipedia, a convex body is a closed, bounded convex subset  $\Omega$  of  $\mathbb{R}^N$  which has nonempty interior (i.e., there exists at least one point  $P$  of  $\Omega$  such that for sufficiently small  $\epsilon > 0$  the entire open ball  $B_\epsilon(P)$  of points of  $\mathbb{R}^N$  of distance less than  $\epsilon$  from  $P$  is contained in  $\Omega$ ). Our definition of convex body is chosen so as to make the statement of Minkowski’s Theorem as clean as possible.

First we record a purely technical result, without proof:

**Lemma 1.** (*Minkowski*) *A bounded convex set  $\Omega \subset \mathbb{R}^N$  is Jordan measurable: that is, the function*

$$\mathbf{1}_\Omega : x \mapsto 1, x \in \Omega; 0, x \notin \Omega$$

*is Riemann integrable. Therefore we can define the **volume** of  $\Omega$  as*

$$\text{Vol}(\Omega) = \int_{\mathbb{R}^N} \mathbf{1}_\Omega.$$

Remark: We are using “volume” as a generic term independent of dimensions. When  $N = 1$  it would be more properly called “length”; when  $N = 2$ , “area”; and, perhaps, “hyper-volume” when  $N > 3$ .

Intuitively speaking, this just says that the boundary of a convex set is not pathologically rugged. In our applications, our bodies will be things like polyhedra and spheres, which are evidently not pathological in this way.

We will also need the following simple result, which ought to be familiar from a course in geometry, multi-variable calculus and/or linear algebra. The reader might try to prove it for herself, but we will not assign it as a formal exercise because we will discuss a more general result in §1.4.

**Lemma 2.** (*Dilation Lemma*) *Recall that for a subset  $\Omega$  of  $\mathbb{R}^N$  and a positive real number  $\alpha$  we define the **dilate** of  $\Omega$*

$$\alpha\Omega := \{\alpha \cdot P = (\alpha x_1, \dots, \alpha x_n) \mid P = (x_1, \dots, x_n) \in \Omega\}.$$

*Then:*

- a)  $\Omega$  is nonempty  $\iff \alpha\Omega$  is nonempty.  
b)  $\Omega$  is bounded  $\iff \alpha\Omega$  is bounded.  
c)  $\Omega$  is Jordan measurable  $\iff \alpha\Omega$  is Jordan measurable, and if so,

$$\text{Vol}(\alpha\Omega) = \alpha^N \text{Vol}(\Omega).$$

- d)  $\Omega$  is convex  $\iff \alpha\Omega$  is convex.  
e)  $\Omega$  is centrally symmetric  $\iff \alpha\Omega$  is centrally symmetric.

An immediate consequence is:

**Corollary 3.** *If  $\Omega \subset \mathbb{R}^N$  is a convex body of volume  $V$ , then for any positive real number  $\alpha$ ,  $\alpha\Omega$  is a convex body of volume  $\alpha^N V$ .*

We saw above that any convex body  $\Omega \subset \mathbb{R}^N$  contains the origin. In particular, such a set contains at least one point in  $\mathbb{Z}^N$ . Must it contain any more?

Of course not. Take in the plane the disk of radius  $r$  centered at the origin. This is a convex body which, if  $r < 1$ , does not intersect any other lattice point besides 0. If  $r = 1$ , it meets the four closest points to 0 if the disk is closed but not if it is open; for  $r > 1$  it necessarily meets other lattice points.

Can we find a convex body in  $\mathbb{R}^2$  which contains no nonzero lattice points but has larger area than the open unit disk, i.e., area larger than  $\pi$ ? Of course we can: the open square

$$(-1, 1)^2 = \{(x, y) \in \mathbb{R}^2 \mid |x|, |y| < 1\}$$

has area 4 but meets no nonzero lattice points. As in the case of circles, this is certainly the limiting case of *its kind*: any centrally symmetric – i.e., with vertices  $(\pm a, \pm b)$  for positive real numbers  $a, b$  – will contain the lattice point  $(1, 0)$  if  $a > 1$  and the lattice point  $(0, 1)$  if  $b > 1$ , so if it does not contain any nonzero lattice points we have  $\max(a, b) \leq 1$  and thus its area is at most 4. But what if we rotated the rectangle? Or took a more elaborate convex body?

One way (not infallible by any means, but a place to start) to gain intuition in a multi-dimensional geometric problem is to examine the problem in a lower dimension. A symmetric convex subset of the real line  $\mathbb{R}^1$  is just an interval, either of the form  $(-a, a)$  or  $[-a, a]$ . Thus by reasoning similar to, but even easier than, the above we see that a centrally symmetric convex subset of  $\mathbb{R}$  must have a nontrivial lattice point if its “one dimensional volume” is greater than 2, and a centrally symmetric convex *body* (i.e., closed) must have a nontrivial lattice point if its one-dimensional volume is at least 2.

Now passing to higher dimensions, we see that the open cube  $(-1, 1)^N$  is a symmetric convex subset of volume  $2^N$  which meets no nontrivial lattice point, whereas for any  $0 < V < 2^N$  the convex body  $[-\frac{V^{1/N}}{2}, \frac{V^{1/N}}{2}]^N$  meets no nontrivial lattice point and has volume  $V$ . After some further experimentation, it is natural to suspect the following result.

**Theorem 4.** (*Minkowski’s Convex Body Theorem*) *Suppose  $\Omega \subset \mathbb{R}^N$  is a convex body with  $\text{Vol}(\Omega) > 2^N$ . Then there exist integers  $x_1, \dots, x_N$ , not all zero, such that  $P = (x_1, \dots, x_N) \in \Omega$ .*

## 1.2. First Proof of Minkowski’s Convex Body Theorem.

Step 0: By Corollary 3,  $\frac{1}{2}\Omega$  is also a convex body of volume

$$\text{Vol}(\frac{1}{2}\Omega) = \frac{1}{2^N} \text{Vol}(\Omega) > 1.$$

Moreover  $\Omega$  contains a nonzero “integral point”  $P \in \mathbb{Z}^N$  iff  $\frac{1}{2}\Omega$  contains a nonzero “half-integral point” – a nonzero  $P$  such that  $2P \in \mathbb{Z}^N$ . So it suffices to show: for any convex body  $\Omega \subset \mathbb{R}^N$  with volume greater than one, there exist integers  $x_1, \dots, x_N$ , not all zero, such that  $P = (\frac{x_1}{2}, \dots, \frac{x_N}{2})$  lies in  $\Omega$ .

Step 1: Observe that if  $\Omega$  contains  $P$  and  $Q$ , by central symmetry it contains

$-Q$  and then by convexity it contains  $\frac{1}{2}P + \frac{1}{2}(-Q) = \frac{1}{2}P - \frac{1}{2}Q$ .

Step 2: For a positive integer  $r$ , let  $L(r)$  be the number of  $\frac{1}{r}$ -lattice points of  $\Omega$ , i.e., points  $P \in \mathbb{R}^N \cap \Omega$  such that  $rP \in \mathbb{Z}^N$ . By Lemma 1,  $\Omega$  is Jordan measurable, and then by [Theorem 3, Gauss's Circle Problem],  $\lim_{r \rightarrow \infty} \frac{L(r)}{r^N} = \text{Vol}(\Omega)$ . Since  $\text{Vol}(\Omega) > 1$ , for sufficiently large  $r$  we must have  $L(r) > r^N$ . Because  $\#(\mathbb{Z}/r\mathbb{Z})^N = r^N$ , by the pigeonhole principle there exist distinct integral points

$$P = (x_1, \dots, x_N) \neq Q = (y_1, \dots, y_N)$$

such that  $\frac{1}{r}P, \frac{1}{r}Q \in \Omega$  and  $x_i \equiv y_i \pmod{r}$  for all  $i$ . By Step 1  $\Omega$  contains

$$R := \frac{1}{2} \left( \frac{1}{r}P \right) - \frac{1}{2} \left( \frac{1}{r}Q \right) = \frac{1}{2} \left( \frac{x_1 - y_1}{r}, \dots, \frac{x_N - y_N}{r} \right).$$

But  $x_i \equiv y_i \pmod{r}$  for all  $i$  and therefore  $\frac{1}{r}(P - Q) = \left( \frac{x_1 - y_1}{r}, \dots, \frac{x_N - y_N}{r} \right) \in \mathbb{Z}^N$  and thus  $R = \frac{1}{2} \left( \frac{1}{r}(P - Q) \right)$  is a half integral point lying in  $\Omega$ : QED!

### 1.3. Second Proof of Minkowski's Convex Body Theorem.

We first introduce some further terminology.

Let  $\Omega \subset \mathbb{R}^N$  be a bounded Jordan measurable set. Consider the following set

$$P(\Omega) := \bigcup_{x \in \mathbb{Z}^N} x + \Omega;$$

that is,  $P(\Omega)$  is the union of the translates of  $\Omega$  by all integer points  $x$ . We say that  $\Omega$  is **packable** if the translates are pairwise disjoint, i.e., if for all  $x \neq y \in \mathbb{Z}^N$ ,  $(x + \Omega) \cap (y + \Omega) = \emptyset$ .

Example: Let  $\Omega = B_0(r)$  be the open disk in  $\mathbb{R}^N$  centered at the origin with radius  $r$ . Then  $\Omega$  is packable iff  $r \leq \frac{1}{2}$ .

Example: For  $r > 0$ , let  $\Omega = [0, r]^N$  be the cube with side length  $r$  and one vertex on the origin. Then  $\Omega$  is packable iff  $r < 1$ , i.e., iff  $\text{Vol}(\Omega) < 1$ . Also the open cube  $(0, 1)^N$  is packable and of volume one.

These examples serve to motivate the following result.

**Theorem 5.** (*Blichfeldt's Theorem*) *If a bounded, Jordan measurable subset  $\Omega \subset \mathbb{R}^N$  is packable, then  $\text{Vol}(\Omega) \leq 1$ .*

*Proof.* Suppose that  $\Omega$  is packable, i.e., that the translates  $\{x + \Omega \mid x \in \mathbb{Z}^N\}$  are pairwise disjoint. Let  $d$  be a positive real number such that every point of  $\Omega$  lies at a distance at most  $d$  from the origin (the boundedness of  $\Omega$  is equivalent to  $d < \infty$ ).

Let  $\bar{B}_r(0)$  be the closed ball of radius  $r$  centered at the origin. It has volume  $c(N)r^N$  where  $c(N)$  depends only on  $N$ .<sup>1</sup> By our work on Gauss's Circle Problem, we know that the number of lattice points inside  $\bar{B}_r(0)$  is asymptotic to  $c(N)r^N$ . Therefore the number of lattice points inside  $\bar{B}_{r-d}(0)$  is asymptotic, as  $r \rightarrow \infty$ ,

<sup>1</sup>The values of  $c(N)$  are known – of course  $c(2) = \pi$  and  $c(3) = \frac{4\pi}{3}$  are familiar from our mathematical childhood, and later on you will be asked to compute  $c(4) = \frac{\pi^2}{2}$ . But as you will shortly see, it would be pointless to substitute in the exact value of  $c(N)$  here.

to  $c(N)(r-d)^N \sim c(N)r^N$ . Therefore for any fixed  $\epsilon > 0$ , there exists  $R$  such that  $r \geq R$  implies that the number of lattice points inside  $\overline{B}_{r-d}(0)$  is at least  $(1-\epsilon)c(N)r^N$ .

Now note that if  $x \in \mathbb{Z}^N$  is such that  $\|x\| \leq r-d$ , then the triangle inequality gives  $x + \Omega \subset \overline{B}_0(r)$ . Then, if  $\Omega$  is packable, then we have at least  $(1-\epsilon)c(N)r^N$  pairwise disjoint translates of  $\Omega$  contained inside  $\overline{B}_0(r)$ . Therefore we have

$$c(N)r^N = \text{Vol}(\overline{B}_r(0)) \geq \text{Vol}(P(\Omega) \cap \overline{B}_r(0)) \geq (1-\epsilon)c(N)r^N \text{Vol}(\Omega),$$

and therefore

$$\text{Vol}(\Omega) \leq \frac{1}{1-\epsilon}.$$

Since this holds for all  $\epsilon > 0$ , we conclude  $\text{Vol}(\Omega) \leq 1$ .  $\square$

Remark: The reader who knows about such things will see that the proof works verbatim if  $\Omega$  is merely assumed to be bounded and Lebesgue measurable.

Now we use Blichfeldt's Theorem to give a shorter proof of Minkowski's Theorem. As in the first proof, after the rescaling  $\Omega \mapsto \frac{1}{2}\Omega$ , our hypothesis is that  $\Omega$  is a convex body with  $\text{Vol}(\Omega) > 1$  and we want to prove that  $\Omega$  contains a nonzero point with half-integral coordinates. Applying Blichfeldt's Lemma to  $\Omega$ , we get  $x, y \in \mathbb{Z}^N$  such that  $(x + \Omega) \cap (y + \Omega)$  is nonempty. In other words, there exist  $P, Q \in \Omega$  such that  $x + P = y + Q$ , or  $P - Q = y - x \in \mathbb{Z}^N$ . But as we saw above, any convex body which contains two points  $P$  and  $Q$  also contains  $-Q$  and therefore  $\frac{1}{2}P - \frac{1}{2}Q = \frac{1}{2}(P - Q)$ , which is a half-integral point.

#### 1.4. Minkowski's Theorem Mark II.

Let  $\Omega \subset \mathbb{R}^N$ . In the last section we considered the effect of a dilation on  $\Omega$ : we got another subset  $\alpha\Omega$ , which was convex iff  $\Omega$  was, centrally symmetric iff  $\Omega$  was, and whose area was related to  $\Omega$  in a predictable way.

Note that dilation by  $\alpha \in \mathbb{R}^{>0}$  can be viewed as a **linear automorphism** of  $\mathbb{R}^N$ : that is, the map  $(x_1, \dots, x_n) \mapsto (\alpha x_1, \dots, \alpha x_n)$  is an invertible linear map. Its action on the standard basis  $e_1, \dots, e_N$  of  $\mathbb{R}^N$  is simply  $e_i \mapsto \alpha e_i$ , so its matrix representation is

$$\alpha : \mathbb{R}^N \rightarrow \mathbb{R}^N, (x_1, \dots, x_n)^t \mapsto \begin{bmatrix} \alpha & 0 & 0 & \dots & 0 \\ 0 & \alpha & 0 & \dots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \dots & \alpha \end{bmatrix} (x_1, \dots, x_n)^t.$$

Now consider a more general linear automorphism  $M : \mathbb{R}^N \rightarrow \mathbb{R}^N$ , which we may identify with its defining matrix  $M \in \text{GL}_N(\mathbb{R})$  (i.e.,  $M = (m_{ij})$  is an  $N \times N$  real matrix with nonzero determinant). We will now state – and prove – the following generalization of the dilation lemma to arbitrary linear automorphisms:

**Lemma 6.** *Let  $\Omega$  be a subset of  $\mathbb{R}^N$  and  $M : \mathbb{R}^N \rightarrow \mathbb{R}^N$  be an invertible linear map. Consider the image*

$$M(\Omega) = \{M(x_1, \dots, x_n)^t \mid (x_1, \dots, x_n) \in \Omega\}.$$

- a)  $\Omega$  is nonempty  $\iff M(\Omega)$  is nonempty.  
 b)  $\Omega$  is bounded  $\iff M(\Omega)$  is bounded.  
 c)  $\Omega$  is convex  $\iff M(\Omega)$  is convex.  
 d)  $\Omega$  is centrally symmetric  $\iff M(\Omega)$  is centrally symmetric.  
 e)  $\Omega$  is Jordan measurable  $\iff M(\Omega)$  is Jordan measurable, and if so,

$$\text{Vol}(M(\Omega)) = |\det(M)| \text{Vol}(\Omega).$$

Proof: Part a) is quite obvious. Part b) holds with  $M$  replaced by any homeomorphism of  $\mathbb{R}^N$ : i.e., a continuous map from  $\mathbb{R}^N$  to itself with continuous inverse, because a subset of  $\mathbb{R}^N$  is bounded iff it is contained in a compact subset, and the image of a compact subset under a continuous function is bounded. Part c) is true because the image of a line segment under a linear map is a line segment. Part d) follows because of the property  $M(-v) = -Mv$  of linear maps. As for part e), the preservation of Jordan measurability follows from the fact that an image of a set of measure zero under a linear map has measure zero. The statement about areas is precisely what one gets by applying the change of variables  $(x_1, \dots, x_N) \mapsto (y_1, \dots, y_N) = M(x_1, \dots, x_N)$  in the integral  $\int_{\mathbb{R}^N} \mathbf{1} dx_1 \cdots dx_N$ .

**Corollary 7.** *If  $\Omega \subset \mathbb{R}^N$  is a convex body and  $M : \mathbb{R}^N \rightarrow \mathbb{R}^N$  is an invertible linear map, then  $M(\Omega)$  is a convex body, and  $\text{Vol}(M(\Omega)) = |\det(M)| \text{Vol}(\Omega)$ .*

Recall that the lattice points inside  $r\Omega$  are precisely the  $\frac{1}{r}$ -lattice points inside  $\Omega$ . This generalizes to arbitrary transformations as follows: for  $M \in \text{GL}_N(\mathbb{R})$ , put

$$\Lambda := M\mathbb{Z}^N = \{M(x_1, \dots, x_N)^t \mid (x_1, \dots, x_N) \in \mathbb{Z}^N\}.$$

The map  $\Lambda : \mathbb{Z}^N \rightarrow M\mathbb{Z}^N$  is an isomorphism of groups, so  $M\mathbb{Z}^N$  is, abstractly, simply another copy of  $\mathbb{Z}^N$ . However, it is embedded inside  $\mathbb{R}^N$  differently. A nice geometric way to look at it is that  $\mathbb{Z}^N$  is the vertex set of a tiling of  $\mathbb{R}^N$  by unit (hyper)cubes, whereas  $\Lambda$  is the vertex set of a tiling of  $\mathbb{R}^N$  by (hyper)parallelopipeds. A single parallelopiped is called a **fundamental domain** for  $\Lambda$ , and the volume of a fundamental domain is given by  $|\det(M)|$ .<sup>2</sup> We sometimes refer to the volume of the fundamental domain as simply the volume of  $\Lambda$  and write

$$\text{Vol}(\Lambda) = |\det(M)|.$$

Now the fundamental fact – a sort of “figure-ground” observation – is the following:

**Proposition 8.** *Let  $\Omega \subset \mathbb{R}^N$  and let  $M : \mathbb{R}^N \rightarrow \mathbb{R}^N$  be an invertible linear map. Then  $M$  induces a bijection between  $M^{-1}(\mathbb{Z}^N) \cap \Omega$  and  $\mathbb{Z}^N \cap M(\Omega)$ .*

If the statement is understood, the proof is immediate!

Applying this (with  $M^{-1}$  in place of  $M$ ) gives the following: if we have a lattice  $\Lambda = M\mathbb{Z}^N$ , and a convex body  $\Omega$ , the number of points of  $\Lambda \cap \Omega$  is the same as the number of points of  $\mathbb{Z}^N \cap M^{-1}(\Omega)$ . Since

$$\text{Vol}(M^{-1}(\Omega)) = |\det(M^{-1})| \text{Vol}(\Omega) = \frac{\text{Vol}(\Omega)}{|\det(M)|} = \frac{\text{Vol}(\Omega)}{\text{Vol}(\Lambda_M)},$$

we immediately deduce a more general version of Minkowski’s theorem.

<sup>2</sup>This is the very important geometric interpretation of determinants, which we would like to assume is familiar from linear algebra. Although we have some concerns as to the validity of this assumption, we will stick with it nonetheless.

**Theorem 9.** (*Minkowski's Theorem Mark II*) Let  $\Omega \subset \mathbb{R}^N$  be a convex body. Let  $M : \mathbb{R}^N \rightarrow \mathbb{R}^N$  be an invertible linear map, and put  $\Lambda_M = M(\mathbb{Z}^N)$ . Suppose that

$$\text{Vol}(\Omega) > 2^N \text{Vol}(\Lambda_M) = 2^N |\det(M)|.$$

Then there exists  $x \in \Omega \cap (\Lambda_M \setminus (0, \dots, 0))$ .

### 1.5. Comments and complements.

Theorem 4 was first proved in an 1896 paper of H. Minkowski, and is treated at further length in Minkowski's 1910 text *Geometrie der Zahlen* [?, pp. 73-76]. Another proof is given in his 1927 *Diophantische Approximationen* [?, pp. 28-30]. Theorem 5 appears in a 1914 paper of H.F. Blichfeldt [?], and the connection to Minkowski's theorem is noted therein. Our first proof of Theorem 4 – which seems to me to be the most direct – is due to Louis Joel Mordell [?].

Blichfeldt's theorem is equivalent to the following result:

**Theorem 10.** Let  $\Omega \subset \mathbb{R}^N$  be a bounded (Jordan or Lebesgue) measurable subset of volume greater than one. Then there exists  $x \in \mathbb{R}^N$  such that the translate  $x + \Omega$  contains at least two integral points.

We leave the proof as an exercise.

There is also a “rotational analogue” of Blichfeldt's theorem:

**Theorem 11.** (*J. Hammer* [?]) Let  $\Omega \subset \mathbb{R}^N$  be a convex body. If the volume of  $\Omega$  is greater than  $c(N)$ , the volume of the unit ball in  $\mathbb{R}^N$ , then there exists an orthogonal matrix  $M \in O(N)$  such that  $M\Omega$  contains a nonzero lattice point.

The proof is not so hard, but it uses some further facts about convex bodies.

Minkowski's theorem is often regarded as the “fundamental theorem” upon which an entire field, the **geometry of numbers**, is based. Because of this, it is not surprising that many mathematicians – including Minkowski himself and C.L. Siegel – have given various refinements over the years. Below we describe one such refinement which can be proved along similar lines.

First, we may allow the nonempty, centrally symmetric convex set  $\Omega \subset \mathbb{R}^N$  to be unbounded. In order to do this, we need to make sense of Jordan measurability and volume for an unbounded subset  $\Omega$ . Since we still want to define  $\text{Vol}(\Omega) = \int_{\mathbb{R}^N} \mathbf{1}_\Omega$ , it comes down to defining what it means for a function defined on an unbounded subset of  $\mathbb{R}^N$  to be Riemann integrable. Evidently what we want is an improper multivariable Riemann integral. Recall that for improper integrals over the real line, if the function  $f$  is allowed to take both positive and negative values then we need to be extremely precise about the sense in which the limits are taken, but if  $f$  is a non-negative function all roads lead to the same answer. Note that characteristic functions are non-negative. So the following definition is simple and reasonable:

Let  $f : \mathbb{R}^N \rightarrow [0, \infty)$  be a function such that the restriction of  $f$  to any rectangle  $[a, b] = \prod_{i=1}^N [a_i, b_i]$  is Riemann integrable. Then we define

$$\int_{\mathbb{R}^N} f = \sup \int_{[a,b]} f,$$

where the supremum ranges all integrals over all rectangles. Note that such an improper integral is always defined although it may be  $\infty$ : for instance it will be if we integrate the constant function 1 over  $\mathbb{R}^N$ .

**Theorem 12.** (*Refined Minkowski Theorem*) *Let  $\Omega \subset \mathbb{R}^N$  be a nonempty centrally symmetric convex subset.*

a) *Then  $\#(\Omega \cap \mathbb{Z}^N) \geq 2(\lceil \frac{\text{Vol}(\Omega)}{2^N} \rceil - 1) + 1$ .*

b) *If  $\Omega$  is closed and bounded, then  $\#(\Omega \cap \mathbb{Z}^N) \geq 2(\lfloor \frac{\text{Vol}(\Omega)}{2^N} \rfloor) + 1$ .*

In other words, part a) says that if for some positive integer  $k$  we have  $\text{Vol}(\Omega)$  is strictly greater than  $k \cdot 2^N$ , then  $\Omega$  contains at least  $2k$  nonzero lattice points (which necessarily come in  $k$  antipodal pairs  $P, -P$ ). Part b) says that the same conclusion holds in the limiting case  $\text{Vol}(\Omega) = k \cdot 2^N$  provided  $\Omega$  is closed and bounded.

There are analogous refinements of Blichfeldt's theorem; moreover, by a linear change of variables we can get a "Refined Mark II Minkowski Theorem" with the standard integral lattice  $\mathbb{Z}^N$  replaced by any lattice  $\Lambda = M\mathbb{Z}^N$ , with a suitable correction factor of  $\text{Vol}(\Lambda)$  thrown in.

We leave the proof of Theorem 12 and the statements and proofs of these other refinements as exercises for the interested reader.

## 2. APPLICATIONS OF MINKOWSKI'S CONVEX BODY THEOREM

### 2.1. The Two Squares Theorem Again.

Suppose  $p = 4k + 1$  is a prime number.

By Fermat's Lemma (Lemma 2 of Handout 4), there exists  $u \in \mathbb{Z}$  such that  $u^2 \equiv -1 \pmod{p}$ : equivalently,  $u$  has order 4 in  $(\mathbb{Z}/p\mathbb{Z})^\times$ . Define

$$M := \begin{bmatrix} 1 & 0 \\ u & p \end{bmatrix}.$$

We have  $\det(M) = p^2$ , so  $\Lambda := M\mathbb{Z}^2$  defines a lattice in  $\mathbb{R}^2$  with

$$\text{Vol}(\Lambda) = \det(M) \text{Vol}(\mathbb{Z}^2) = p.$$

If  $(t_1, t_2) \in \mathbb{Z}^2$  and  $(x_1, x_2)^t = M(t_1, t_2)^t$ , then

$$x_1^2 + x_2^2 = t_1^2 + (ut_1 + pt_2)^2 \equiv (1 + u^2)t_1^2 \equiv 0 \pmod{p}.$$

Now put  $\Omega = \overline{B}_0(\sqrt{\frac{3p}{2}})$ , the closed ball of radius  $\sqrt{\frac{3p}{2}}$  about the origin. We have

$$\text{Vol}(\Omega) = \pi\left(\frac{3p}{2}\right) = \frac{3\pi}{2}p > 2^2 \text{Vol}(\Lambda),$$

so by Minkowski's Theorem Mark II there exists  $(x_1, x_2) \in \Lambda$  with

$$0 < x_1^2 + x_2^2 < \frac{3p}{2}.$$

Since  $p \mid x_1^2 + x_2^2$ , the only possible conclusion is

$$x_1^2 + x_2^2 = p,$$

qed.

## 2.2. The Four Squares Theorem.

The following result is one of the high water marks of classical number theory.

**Theorem 13.** (*Lagrange*) *Every positive integer is a sum of four integral squares.*

Proof: We begin with the following result, whose proof is left as an exercise.

**Lemma 14.** (*Euler*) *For any integers  $a_1, \dots, a_4, b_1, \dots, b_4$ , we have*

$$(a_1^2 + a_2^2 + a_3^2 + a_4^2)(b_1^2 + b_2^2 + b_3^2 + b_4^2) = (a_1b_1 - a_2b_2 - a_3b_3 - a_4b_4)^2 + (a_1b_2 + a_2b_1 + a_3b_4 - a_4b_3)^2 + (a_1b_3 - a_2b_4 + a_3b_1 + a_4b_2)^2 + (a_1b_4 + a_2b_3 - a_3b_2 + a_4b_1)^2.$$

*Thus, if two integers are each sums of four squares, so is their product.*

Also  $1 = 1^2 + 0^2 + 0^2 + 0^2$  is a sum of four squares, so it suffices to show that all prime numbers are sums of four squares.

**Lemma 15.** *For any prime number  $p$  and any  $a \in \mathbb{Z}$ , there exist  $r, s \in \mathbb{Z}$  such that*

$$r^2 + s^2 \equiv a \pmod{p}.$$

Proof: We may and shall assume  $p > 2$ . Because of the cyclicity of  $\mathbb{F}_p^\times$  there are precisely  $\frac{p-1}{2}$  nonzero squares mod  $p$  and hence  $\frac{p-1}{2} + 1 = \frac{p+1}{2}$  squares mod  $p$ . Rewrite the congruence as  $r^2 \equiv a - s^2 \pmod{p}$ . Since the map  $\mathbb{F}_p \rightarrow \mathbb{F}_p$  given by  $t \mapsto a - t$  is an injection, as  $x$  ranges over all elements of  $\mathbb{F}_p$  both the left and right hand sides take  $\frac{p+1}{2}$  distinct values. Since  $\frac{p+1}{2} + \frac{p+1}{2} > p$ , these subsets cannot be disjoint, and any common value gives a solution to the congruence.

By Lemma 15, there are  $r, s \in \mathbb{Z}$  such that  $r^2 + s^2 + 1 \equiv 0 \pmod{p}$ . Define

$$M = \begin{bmatrix} p & 0 & r & s \\ 0 & p & s & -r \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

We have  $\det(M) = p^2$ , so  $\Lambda := M\mathbb{Z}^4$  defines a lattice in  $\mathbb{R}^4$  with

$$\text{Vol}(\Lambda) = \det(M) \text{Vol}(\mathbb{Z}^4) = p^2.$$

If  $(t_1, t_2, t_3, t_4) \in \mathbb{Z}^4$  and  $(x_1, x_2, x_3, x_4) := M(t_1, t_2, t_3, t_4)$  then

$$\begin{aligned} x_1^2 + x_2^2 + x_3^2 + x_4^2 &= (pt_1 + rt_3 + st_4)^2 + (pt_2 + st_3 - rt_4)^2 + t_3^2 + t_4^2 \\ &\equiv t_3^2(r^2 + s^2 + 1) + t_4^2(r^2 + s^2 + 1) \equiv 0 \pmod{p}. \end{aligned}$$

**Lemma 16.** *The (four-dimensional) volume of a ball of radius  $r$  in  $\mathbb{R}^4$  is  $\frac{\pi^2}{2}r^4$ .*

This is just a (single variable!) calculus exercise. We leave the proof to the reader.

Now put  $\Omega = \overline{B_0}(\sqrt{2p})$ , the closed ball of radius  $\sqrt{2p}$  about the origin. We have

$$\text{Vol}(\Omega) = 2\pi^2 p^2 > 2^4 \text{Vol}(\Lambda),$$

so by Minkowski's Theorem Mark II there exists  $(x_1, \dots, x_4) \in \Lambda$  with

$$0 < x_1^2 + x_2^2 + x_3^2 + x_4^2 < 2p.$$

Since  $p \mid x_1^2 + x_2^2 + x_3^2 + x_4^2$ , the only possible conclusion is

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = p,$$

qed!

### 2.3. Vista: Testing for a PID.

The preceding applications were very pretty, but – in that they give new proofs of old theorems – do not really serve to illustrate the power and utility of Minkowski's Convex Body Theorem. A much deeper application is to the computation of the **class number** of a number field  $K$ . Although it will be beyond us to give proofs, we feel the concept is so important that we should at least sketch out the statement.

Let  $K$  be any number field, i.e., a field which contains  $\mathbb{Q}$  as a subfield and is finite-dimensional as a  $\mathbb{Q}$ -vector space, say of dimension  $d$ . To a number field we attach its **ring of integers**  $\mathbb{Z}_K$ . This is the set of all elements  $\alpha$  in  $K$  which satisfy a monic polynomial with integral coefficients: i.e., for which there exist  $a_0, \dots, a_{n-1} \in \mathbb{Z}$  such that

$$\alpha^n + a_{n-1}\alpha^{n-1} + \dots + a_1\alpha + a_0 = 0.$$

It is not hard to show that  $\mathbb{Z}_K$  is indeed a subring of  $K$ : this is shown in Handout A3. But more is true: if  $d$  is the degree of  $K$  over  $\mathbb{Q}$ , then there exist  $\alpha_1, \dots, \alpha_d \in \mathbb{Z}_K$  such that every element  $\alpha \in \mathbb{Z}_K$  can uniquely be expressed as a  $\mathbb{Z}$ -linear combination of the  $\alpha_i$ 's:

$$\alpha = a_1\alpha_1 + \dots + a_d\alpha_d, \quad a_i \in \mathbb{Z}.$$

Such a  $d$ -tuple  $(\alpha_1, \dots, \alpha_d)$  of elements of  $\mathbb{Z}_K$  is called an **integral basis**.

Example: Let  $K = \mathbb{Q}$ . Then  $\mathbb{Z}_K = \mathbb{Z}$ , and  $\alpha_1 = 1$  is an integral basis.

Example: Let  $K/\mathbb{Q}$  be a quadratic extension, so that there exists a squarefree integer  $d \neq 0, 1$  such that  $K = \mathbb{Q}(\sqrt{d})$ . Observe that  $\sqrt{d}$ , satisfying the monic polynomial  $t^2 - d$ , is an element of  $\mathbb{Z}_K$ , as is the entire subring  $\mathbb{Z}[\sqrt{d}] = \{a + b\sqrt{d} \mid a, b \in \mathbb{Z}\}$  that it generates. With  $d = -1$ , this is just the ring of Gaussian integers, which is indeed the full ring of integers of  $\mathbb{Q}(\sqrt{-1})$ .

In general things are more subtle: it turns out that if  $d \equiv 2, 3 \pmod{4}$  then  $\mathbb{Z}_K = \mathbb{Z}[\sqrt{d}]$ ; however if  $d \equiv 1 \pmod{4}$  then the element  $\tau_d := \frac{1+\sqrt{d}}{2}$  may not look like an algebraic integer but it satisfies the monic polynomial  $t^2 + t + \frac{1-d}{4}$  (which has  $\mathbb{Z}$ -coefficients since  $d \equiv 1 \pmod{4}$ ) so in fact it is, and in this case  $\mathbb{Z}_K = \mathbb{Z}[\tau_d] = \{a + b(\frac{1+\sqrt{d}}{2}) \mid a, b \in \mathbb{Z}\}$ .

Example: Let  $K = \mathbb{Q}(\zeta_n)$  obtained by adjoining to  $\mathbb{Q}$  a primitive  $n$ th root of unity. Then it is easy to see that  $\zeta_n$  is an algebraic integer, and in this case it can be shown that  $\mathbb{Z}_K = \mathbb{Z}[\zeta_n]$  is the full ring of integers.

It is rare to be able to write down an integral basis by pure thought; however, there exists a straightforward algorithm which, given any single number field  $K$ , computes an integral basis for  $K$  (the key word here is “discriminant”, but no more about that!).

**Question 1.** *For which number fields  $K$  is  $\mathbb{Z}_K$  a principal ideal domain?*

This is absolutely one of the deepest and most fundamental number-theoretic questions because, as we have seen, in trying to solve a Diophantine equation we are often naturally led to consider arithmetic in a ring of integers  $\mathbb{Z}_K$  – e.g., in studying

the equation  $x^2 - dy^2 = n$  we take  $K = \mathbb{Q}(\sqrt{d})$  and in studying  $x^n + y^n = z^n$  we take  $K = \mathbb{Q}(\zeta_n)$ . If  $\mathbb{Z}_K$  turns out to be a PID, we can use Euclid's Lemma, a formidable weapon. Indeed, it turns out that a common explanation of each of the classical success stories regarding these two families of equations (i.e., theorems of Fermat, Euler and others) is that the ring  $\mathbb{Z}_K$  is a PID.

Gauss conjectured that there are infinitely many squarefree  $d > 0$  such that the ring of integers of the real quadratic field  $K = \mathbb{Q}(\sqrt{d})$  is a PID. This is still unknown; in fact, for all we can prove there are only finitely many number fields  $K$  (of any and all degrees!) such that  $\mathbb{Z}_K$  is a PID. In this regard two important goals are:

- (i) To give an algorithm that will decide, for any given  $K$ , whether  $\mathbb{Z}_K$  is a PID;
- (ii) When it isn't, to "quantify" the failure of uniqueness of factorization in  $\mathbb{Z}_K$ .

For this we define the concept of **class number**. If  $R$  is any integral domain, we define an equivalence relation on the set  $\mathcal{I}(R)$  of nonzero ideals of  $R$ . Namely we put  $I \sim J$  iff there exist nonzero elements  $a, b \in R$  such that  $(a)I = (b)J$ . This partitions all the nonzero ideals into equivalence classes, simply called **ideal classes**.<sup>3</sup> The **class number** of  $R$  is indeed the number of classes of ideals. For an arbitrary domain  $R$ , the class number may well be infinite.

The point here is that there is one distinguished class of ideals: an ideal  $I$  is equivalent to the unit ideal  $R = (1)$  iff it is principal. It follows that  $R$  is a PID iff its class number is equal to one. Therefore both (i) and (ii) above would be addressed if we can compute the class number of an arbitrary ring of integers  $\mathbb{Z}_K$ .

This is exactly what Minkowski did:

**Theorem 17.** (*Minkowski*) *Let  $K$  be any number field.*

- a) The ideals of the ring  $\mathbb{Z}_K$  of integers of  $K$  fall into finitely many equivalence classes; therefore  $K$  has a well-defined class number  $h(K) < \infty$ .*
- b) There is an explicit upper bound on  $h(K)$  in terms of invariants of  $K$  which can be easily computed if an integral basis is known.*
- c) There is an algorithm to compute  $h(K)$ .*

The proof is not easy; apart from the expected ingredients of more basic algebraic number theory, it also uses, crucially, Theorem 4!

As an example of the usefulness of the class number in "quantifying" failure of factorization even when  $\mathbb{Z}_K$  is not a UFD, we note that Lamé erroneously believed he could prove FLT for all odd primes  $p$  because he assumed (implicitly, since the concept was not yet clearly understood) that  $\mathbb{Z}[\zeta_p]$  was always a PID. Lamé's proof is essentially correct when the class number of  $\mathbb{Q}(\zeta_p)$  is equal to one, which is some progress from the previous work on FLT, but unfortunately this happens iff  $p \leq 19$ . Kummer on the other hand found a sufficient condition for FLT(p) to hold which turns out to be equivalent to: the class number of  $\mathbb{Q}(\zeta_p)$  is not divisible by  $p$ .

---

<sup>3</sup>In fact, the use of the term "class" in mathematics in the context of equivalence relations can be traced back to this very construction in the case of  $R = \mathbb{Z}_K$  the ring of integers of an imaginary quadratic field  $K$ , which was considered by Gauss in his *Disquisitiones Arithmeticae*.

This condition, in turn, is satisfied for all  $p < 200$  *except* for 37, 59, 67, 101, 103, 131, 149, and 157; and *conjecturally* for a subset of the primes of relative density  $e^{-\frac{1}{2}} \approx 0.61$ . Note finally that this remains conjectural to this day while FLT has been proven: the study of class numbers really is among the deepest and most difficult of arithmetic questions.

#### 2.4. Comments and complements.

As is the case for many of the results we have presented, one of the attractions of Theorem 13 is its simple statement. There is no question that anyone who is inquisitive enough to wonder which integers can be written as a sum of four squares will eventually conjecture the result, but the proof of course is another matter. Apparently the first recorded statement – without proof – is to be found in the *Arithmetica* of Diophantus of Alexandria, some time in the third century AD. Diophantus’ text entered into the mathematical consciousness of Renaissance Europe through Gaspard Bachet’s 1620 Latin translation of the *Arithmetica*.

Famously, Fermat was an ardent reader of Bachet’s book, and he saw and claimed a proof of the Four Squares Theorem. As we have already mentioned, with one exception (FLT for  $n = 4$ ) Fermat *never* published proofs, making the question of exactly which of his “theorems” he had actually proved a subject of perhaps eternal debate. In this case the consensus among mathematical historians seems to be skepticism that Fermat actually had a proof. In any case, the proof was still much sought after Fermat’s death in 1665. Euler, one of the greatest mathematicians of the 18th century, labored for 40 years without finding a proof. Finally the theorem was proved and published in 1770 by the younger and equally great<sup>4</sup> Italian-French mathematician Joseph-Louis Lagrange.

There are many quite different looking proofs of the Four Squares Theorem. In particular it is possible to give a proof which is “completely elementary” in that it neither requires nor introduces any extraneous concepts like lattices. The most pedestrian proof begins as ours did with Euler’s identity and Lemma 15. From this we know that it suffices to represent any prime as a sum of four squares and also that for any prime  $p$ , some positive integer multiple  $mp$  is of the form  $r^2 + s^2 + 1^2$  and in particular a sum of four squares, and the general strategy is to let  $m_0$  be the smallest integral multiple of  $p$  which is a sum of four squares and to show, through a “descent” argument, that  $m_0 = 1$ . Presumably Lagrange’s original proof followed this broad strategy, and many elementary number theory texts contain such a proof, including Hardy and Wright.

Another proof, which has the virtue of explaining the mysterious identity of Lemma 14 proceeds in analogy to the first proof we gave of the two squares theorem: it works in a certain ring of **integral quaternions**. Believe it or not, quaternions have a simply vital role to play in modern number theory, but although it is not too hard to introduce enough quaternionic theory to prove the Four Squares Theorem (see again Hardy and Wright for this), one has to dig deeper to begin to appreciate what is really going on, too deeply for the scope of this course.

---

<sup>4</sup>In quality at least. No one has ever equalled Euler for quantity, not even the famously prolific and relatively long-lived twentieth century mathematician Paul Erdős, although there are one or two living mathematicians that might eventually challenge Euler.

Yet another proof uses the arithmetic properties of theta series; this leads to an exact formula for  $r_4(n)$ , the number of representations of a positive integer as a sum of four squares. In this case, to understand what is really going on involves discussion of the arithmetic theory of modular forms, which is again too rich for our blood (but we will mention that modular forms and quaternions are themselves quite closely linked!); and again Hardy and Wright manage to give a proof using only purely formal power series manipulations, which succeeds in deriving the formula for  $r_4(n)$ .

Regarding generalizations of Theorem 13, we will only briefly mention one: a few months *before* Lagrange's proof, Edward Waring asserted that "every number is a sum of four squares, nine cubes, nineteen biquadrates [i.e., fourth powers] and so on." In other words, Waring believed that for every positive integer  $k$  there exists a number  $n$  depending only on  $k$  such that every positive integer is a sum of  $n$  non-negative  $k$ th powers. If so, we can define  $g(k)$  to be the least such integer  $k$ . Evidently the Four Squares Theorem together with the observation that 7 is not a sum of three squares, give us  $g(2) = 4$ . That  $g(k)$  actually exists for all  $k$  is by no means obvious, and indeed was first proven by David Hilbert in 1909. We now know the exact value of  $g(k)$  for all  $k$ ; that  $g(3) = 9$  was established relatively early on (Wieferich, 1912), but  $g(4)$  was the last value to be established, by Balasubramanian in 1986: indeed  $g(4) = 19$ , so all of Waring's conjectures turned out to be correct.

Of much more enduring interest is the similar quantity  $G(k)$ , defined to be the least positive integer  $n$  such that every sufficiently large positive integer can be written as a sum of  $n$  non-negative  $k$ th powers: in other words, we allow finitely many exceptions. Since for all  $k$ ,  $8k+7$  is not even a sum of three squares modulo 8, none of these infinitely many integers are sums of three squares, so  $G(2) = G(2) = 4$ . On the other hand it is known that  $G(3) \leq 7 < 9 = g(3)$ , and it moreover *suspected* that  $G(3) = 4$ , but this is far from being proven. Indeed, the only computation of  $G(k)$  for  $k > 2$  thus far is the following:

**Theorem 18.** (*Davenport, 1939*)  $G(4) = 16$ .

Getting better bounds on  $G(k)$  is very much an active topic in twenty-first century number theory.