Math 153: The Four Square Theorem

Rich Schwartz

April 12, 2014

The purpose of these notes is to explain Lagrange's famous 4 square theorem. Every positive integer n can be written as the sum of 4 integer squares. $n = a^2+b^2+c^2+d^2$. These notes follow Herstein's proof in Chapter 7 to some extent, but they simplify the argument and explore some of the beautiful underlying geometry.

1 Reduction to Odd Primes

Given a quaternion $q = a + bi + cj + dk$ we define

$$N(q) = q\overline{q} = a^2 + b^2 + c^2 + d^2, \qquad \overline{q} = a - bi - bj - bk. \tag{1}$$

Lagrange's Theorem is equivalent to the statement that every positive integer arises as $N(q)$ for some quaternion in the ring $\boldsymbol{Z}[i,j,k]$ of integer quaternions.

We have

$$\overline{qr} = (\overline{r})(\overline{q}). \tag{2}$$

This identity is easily checked in the 16 cases of $q, r \in \{1, i, j, k\}$ and then the general case follows from the distributive law. The identity above implies that

$$N(qr) = (qr)(\overline{qr}) = q(r\overline{r})\overline{q} = qN(r)\overline{q} = q\overline{q}N(r) = N(q)N(r).$$

In short

$$N(qr) = N(q)N(r). \tag{3}$$

So, if we can prove that every prime arises as $N(q)$ for some quaternion q, then Equation 3 and the existence of factoring over $\boldsymbol{Z}$ finishes the proof for us. Note that $2 = 1^2 + 1^2 + 0^2 + 0^2$, so we just have to worry about odd primes.

2 Counting Lemmas

Lemma 2.1 *In $\mathbf{Z}/p$ the equation $x^2 = b$ has at most 2 solutions.*

Proof: Let $p(x) = x^2 - b$. If $p(a) = 0$ then $p(-a) = 0$ as well. But then we can write $p(x) = (x-a)(x+a)$ in $\mathbf{Z}/p[x]$. But then $p(c) = (c-a)(c+a) \neq 0$ as long as c is distinct from a and $-a$ in $\mathbf{Z}/p$. ♠

Here is the key lemma.

Lemma 2.2 *For any odd prime p, then there exist integers a and b such that $a^2 + b^2 \equiv -1$ mod p.*

Proof: Consider the set of all elements of the form a^2 in $\mathbf{Z}/p$. By the previous result, there are at least $(p+1)/2$ distinct elements. Likewise consider the set of all elements of the form $-1 - b^2$ in $\mathbf{Z}/p$. Again, there are at least $(p+1)/2$ distinct elements. By the pidgeonhole principle, some element a^2 coincides with some element $-1-b^2$ in $\mathbf{Z}/p$. But then $a^2+b^2 \equiv -1 \bmod p$. ♠

3 The Half-Integral Quaternions

Let

$$\zeta = \frac{1+i+j+k}{2}. \tag{4}$$

Here ζ is a certain half-integral quaternion. Geometrically, ζ is the center of the cube $[0,1]^4$. Let R be the ring of element of the form

$$a\zeta + bi + cj + dk, \qquad a,b,c,d \in \mathbf{Z}. \tag{5}$$

Geomtrically, one can think of $\mathbf{Z}[i,j,k]$ as the set of vertices of the tiling of $\mathbf{R}^4$ by unit cubes. One obtains R by adjoining to $\mathbf{Z}[i,j,k]$ the centers of all these cubes. We will explore the beautiful geometry of R below. But for now, we prove some basic facts about R.

Lemma 3.1 *R is a ring, and in particular a left PID. That is, every left ideal has the form Ra for some $a \in R$.*

Proof: Considering the several cases, we see that R is closed under addition, multiplication, and taking additive inverses. Hence R is a ring. The equation $N(qr) = N(q)N(r)$ implies that R has no zero divisors.

First we show that R is a left Euclidean Domain: Given any nonzero elements $a, b \in R$ there are $q, r \in R$ such that $a = qb + r$ where $N(r) < N(b)$. We think of the elements of R as points in $\boldsymbol{R}^4$. The distance between $q, r \in R$ is $\sqrt{N(q-r)}$. This is always a square-root of an integer and hence at least 1. We just have to show that every element in $\boldsymbol{R}^4$ is closer than 1 to a point corresponding to an element of R. The only point in the 4D cube which is as far from the vertices as the side length is the center of the cube. Thus, $\boldsymbol{Z}[i, j, k]$ barely fails to be a left ED. The only failure points are the centers of the cubes. But we get R by adding in these centers.

The same argument which shows that an ED is a PID shows that a left ED is a left PID. ♠

Lemma 3.2 *The ideal Rp is not maximal.*

Proof: If Rp is maximal, then R/Rp is a division ring, and in particular has no zero divisors. But let $\alpha = a + bi + j$, where a and b solve $a^2 + b^2 \equiv -1 \mod p$. We have $\alpha\overline{\alpha} \in Rp$. But then $[\alpha]$ is a zero divisor in R/Rp. This is a contradiction. ♠

Since Rp is not a maximal ideal, there is some new ideal M such that $Rp \subset M \subset R$ and M is not equal to either Rp or R. Since M is also a left ideal, $M = R\alpha$ for some $\alpha \in R$. Hence $p = c\alpha$ for some $c \in R$. We have

$$p^2 = N(p) = N(c)N(\alpha).$$

If $N(c) = 1$ then c is a unit and $M = R\alpha$. This is false. If $N(\alpha) = 1$ then α is a unit and $M = R$. Hence $N(\alpha) = p$. But then, we can write $\alpha = a + bi + cj + dk$ and $p = N(\alpha) = a^2 + b^2 + c^2 + d^2$.

The numbers a, b, c, d might be half integers rather than integers. In this case, let

$$a' = (a+b)/2, \qquad b' = (a-b)/2, \qquad c' = (c+d)/2, \qquad d' = (c-d)/2$$

These are all integers, and $(a')^2 + (b')^2 + (c')^2 + (d')^2 = p$.

4 Heavenly Beauty

When I was an undergraduate, I didn't appreciate the heavenly beauty of the ring R of half-integral quaternions. So, in this section I'm going to describe it.

Let's talk about polyhedra first. Given a convex polyhedron P, we can form a new polygon P^* whose vertices are the centers of each face. In order to make sense of this, the notion of a center of a face must make sense. In general, we could take the center of mass of the face, but in the cases of interest – platonic solids – it is clear what "center" means just from symmetry.

Consider what this does to the platonic solids:

- When P is a tetrahedron, P^* is a tetrahedron.
- When P is a octahedron, P^* is a cube
- When P is a cube, P^* is a octahedron
- When P is a icosahedron, P^* is a dodecahedron.
- When P is a dodecahedron, P^* is a icosahedron.

From the point of view of this construction, known as duality, the tetrahedron is best: It is self-dual.

Now consider another possible property of polyhedra. Does the polyhedron tile all of space? The cube certainly tiles all of space, but the other 4 platonic solids do not. So, from the point of view of the tiling property, the cube is best.

Even though the five platonic solids are perfectly symmetric, in the sense their group of symmetries takes any vertex to any vertex and any edge to any edge and any face to any face, none of them is both self-dual and a space tiler. They are all "flawed" in some sense.

In 4 dimensions, it turns out that there is a convex polyhedron (or polytope, as it is usually called) which is perfectly symmetric, self-dual, and a space tiler. This polytope is called the 24-cell and it is precisely *the group of units in our ring* R! There are 24 such units. Call this group Γ. Up to sign and permutation, elements of Γ are represented by points of the form $(1,0,0,0)$ or $(1/2,1/2,1/2,1/2)$. Given $a, b \in \Gamma$, we have the symmetry $T_{a,b} : \Gamma \to \Gamma$ given by $T_{a,b}(q) = aqb$. We also the symmetry $T_{c,c}$ where $c = (1+i)/\sqrt{2}$. Finally, ther have the symmetry $q \to \overline{q}$. Each of the 1152 possible symmetries of Γ is a composition of the ones just described.

Lemma 4.1 $\boldsymbol{R}^4$ *is tiled by smaller copies of the* 24*-cell.*

Proof: For each point $q \in \boldsymbol{R}^4$ corresponding to an element of R, there is the *Voronoi cell* V_q consisting of points which are closer to q than to any other point representing an element of R. All of $\boldsymbol{R}^4$ is tiled by the union of these Voronoi cells. So, to finish the proof, we have to show that each Voronoi cell is a 24-cell. Since R is a ring (or even just an additive group) all the Voronoi cells are translates of each other. So, we just have to check that V_0 is a 24-cell.

To see this, suppose that $x = (a, b, c, d)$ is some point in $\boldsymbol{R}^4$. Subtracting off elements of R and using reflection symmetries, we can arrange that $a, b, c, d \in [0, \pi/2]$ and also $a + b + c + d \leq 1$. Conversely, any such point is closer to 0 than to any other point representing an element of R. From this it is not hard to deduce that V_0 is the polytope with vertices that have the form $(1/2, 1/2, 0, 0)$ up to sign and permutation. There are 24 such points and they are the vertices of a smaller (rotated) copy of the 24-cell. For later reference, we call this smaller 24 cell Q. ♠

Lemma 4.2 *The* 24 *cell is self-dual.*

Proof: Let P be the 24-cell whose vertices represent the elements of Γ. Let Q be the 24-cell from Lemma 4.1. Consider the linear functional

$$f(x, y, z, w) = (1, 1, 0, 0) \cdot (x, y, z, w) = x + y.$$

The level sets of f are hyperplanes in $\boldsymbol{R}^4$. There are 6 points of P which maximize f on P, namely

$$(1, 0, 0, 0), \quad (0, 1, 0, 0), \quad (1/2, 1/2, \pm 1/2, \pm 1/2).$$

By computing distances you can check that these 6 points are the vertices of a regular octahedron. The midpoint of this face is just the average of these 6 vertices, namely the point $(1/2, 1/2, 0, 0)$. Permuting the entries in the vector defining f and choosing all possible signs, we see that P has 24 faces and that the centers of all these faces make up the vertices of Q. In short $P^* = Q$, which means that P is self-dual. ♠