A Quaternionic Gem

Anthony G. O'Farrell
Mathematics Department
NUI, Maynooth
Co. Kildare
Ireland
e-mail: anthonyg.ofarrell@gmail.com

February 16, 2006

1 Introduction

It is generally accepted nowadays that Hamilton's greatest achievment is his general theory of dynamics. By comparison, quaternions have had less impact. At the same time, people continue to use and develop the theory and techniques of quaternionic algebra and analysis, and they continue to find new applications, so one cannot say what the verdict may be on the relative importance of the two inventions, in the long run.

It is in the nature of mathematics that its abstract concepts find use far from their origins. Quaternions were invented because Hamilton wanted an algebra that would facilitate geometric work in three dimensions. Number theory (the theory of whole numbers) is quite a different area of mathematics, so I particularly like the fact that quaternions may be used to make an important step in the proof of a theorem in number theory. Here is the theorem:

Theorem 1.1 (Lagrange) *Each positive integer is the sum of at most four square positive integers.*

[0] Supported by Science Foundation Ireland, under grant 05/RFP/MAT0003

(Fermat claimed to have a proof of this theorem, but we usually give the credit to the first person who publishes a proof, and, as in other cases, Fermat did not reveal his proof.)

It is slightly more convenient to discuss this in the equivalent form: *Each nonnegative integer is the sum of four square integers.* The point is that, by allowing zero into consideration, we can always use exactly four numbers to represent a given number. For instance, 6 is not the sum of four positive squares, but it is $2^2 + 1^2 + 1^2 + 0^2$.

2 Sums of Squares

The squares $1, 4, 9, \ldots$ are relatively rare among positive integers. The sums of at most two squares are

$$0 = 0^2 + 0^2,\ 1 = 0^2 + 1^2,\ 2 = 1^2 + 1^2,\ 4 = 0^2 + 2^2,\ 5 = 1^2 + 2^2,\ 8 = 2^2 + 2^2,$$

and so on. It is a useful exercise for a programming class to write code to generate the numbers of this form that are less than or equal to 100 (or 1000, 10000,...). For instance, using Maple:

```
> S2S:= proc( N )
          # returns a list of the sums of 2 squares <= N

          local A,B,C,n,m;

          A:= seq( seq( m^2+n^2,
                        m=0..floor(sqrt(N-n^2))
                       ),
                   n=0..sqrt(N)
                 );  #  sequence of all the sums <= N

         B:= {A}; # cast to set; removes duplicates

         C:= sort( [B[]] );  # cast to list and sort.

       end proc;

> S2S(100):  # output supressed.
```

The sorting step is probably unnecessary, in practice, since Maple implementations will probably store the elements of a set of numbers in the natural order.

Inspection of this list, and longer ones, reveals that there are substantially more sums of two squares than of one square, but there are gaps: the numbers 3,6,7,11,12,14,15, etc. are missing.

What is the pattern?

Mathematicians always look for structure, and the key structure here is the *semigroup.* By definition, a set S of nonnegative numbers is a *multiplicative semigroup* if and only if it has the product of each pair of elements as another element, i.e.

$$\left.\begin{array}{rcl} x & \in & S \\ y & \in & S \end{array}\right\} \Rightarrow xy \in S.$$

For example:

1. The set $\mathbb{N}$ of all positive integers is a multiplicative semigroup.
2. The set $\mathbb{Z}^+$ of all nonnegative integers is a multiplicative semigroup.
3. The set $\mathbf{10}^\bullet$ of all nonnegative powers of 10 is a multiplicative semigroup.
4. The set E of all even positive numbers is a multiplicative semigroup.
5. The set O of all odd positive numbers is a multiplicative semigroup.
6. The set $\langle 6 \rangle$ of all positive multiples of 6 is a multiplicative semigroup.
7. The only finite multiplicative semigroups are the empty set $\emptyset$, and $\{0\}$, $\{1\}$, and $\{0,1\}$.

If S is a multiplicative semigroup, then so are

$$S \cup \{0\},\ S \cup \{1\},\ S \sim \{0\},\ S \sim \{1\},$$

i.e. a semigroup remains a semigroup if 0 or 1 (or both) are added or removed.

The following fact is basic:

Lemma 2.1 *If S is a multiplicative semigroup of nonnegative integers, and all the primes belong to S, then all integers greater than* 1 *belong to S.*

Proof. This is just a restatement of part of the Fundamental Theorem of Arithmetic: every integer greater than 1 is a product of prime numbers. ∎

The connection between semigroups and sums of two squares is:

Theorem 2.2 *The set*

$$S_2 = \{m^2 + n^2 : n \in \mathbb{Z}^+, m \in \mathbb{Z}^+\}$$

of sums of two nonnegative squares is a multiplicative semigroup.

Proof. We use complex numbers $c = a + ib$, and recall that the modulus, or absolute value is given by

$$|c| = |a + ib| = \sqrt{a^2 + b^2},$$

where $\sqrt{\ }$ denotes the nonnegative square root. Also, the complex conjugate is

$$\bar{c} = \overline{a + ib} = a - ib,$$

and is related to the absolute value by

$$|a + ib|^2 = (a + ib)(a - ib),$$

i.e. $|c|^2 = c\bar{c}$.

Recall that a is called the real part of c, denoted $\Re c$, and b the imaginary part, $\Im c$.

Finally, note that

$$\overline{c_1 c_2} = \overline{c_1} \cdot \overline{c_2},$$

i.e. the complex conjugate of a product is the product of the complex conjugates.

Now suppose m_1, n_1, m_2, n_2 are nonnegative integers. We want to show that

$$(m_1^2 + n_1^2)(m_2^2 + n_2^2)$$

is the sum of two squared integers, i.e. has the form

$$m_3^2 + n_3^2$$

for some integers m_3, n_3.

Take

$$\begin{aligned} c_1 &= m_1 + in_1, \\ c_2 &= m_2 + in_2, \\ m_3 &= \Re(c_1c_2), \\ n_3 &= \Im(c_1c_2). \end{aligned}$$

Then

$$m_3 = m_1m_2 - n_1n_2, \text{ and } n_3 = m_1n_2 + m_2n_1 \tag{1}$$

are obviously integers, and

$$\begin{aligned} (m_1^2 + n_1^2) \cdot (m_2^2 + n_2^2) &= |c_1^2| \cdot |c_2^2| \\ &= c_1 \cdot \overline{c_1} \cdot c_2 \cdot \overline{c_2} \\ &= (c_1 \cdot c_2) \cdot (\overline{c_1} \cdot \overline{c_2}) \\ &= (c_1 \cdot c_2) \cdot \overline{(c_1 \cdot c_2)} \\ &= |c_1 \cdot c_2|^2 \\ &= m_3^2 + n_3^2. \end{aligned}$$

Thus we are done. ■

For instance, $5 = 1^2 + 2^2$, $13 = 2^2 + 3^2$, so 65 is the sum of two squares, and the formula (1) gives us two numbers that do the trick:

$$m_3 = 1 \times 2 - 2 \times 3 = -4, \; n_3 = 1 \times 3 + 2 \times 2 = 7,$$

$$65 = 4^2 + 7^2.$$

A person could perhaps impress their friends by pushing the boat a bit further out:

$$85 = 9^2 + 2^2, \; 97 = 9^2 + 4^2,$$

so

$$8245 = (9 \times 9 - 2 \times 4)^2 + (9 \times 4 + 2 \times 9)^2 = 73^2 + 54^2.$$

Starting from this point, standard texts on number theory go on to characterise the numbers that are sums of two squares. The end result is:

Theorem 2.3 *(***Euler***) A number $n \in \mathbb{Z}^+$ belongs to S_2 if and only if*

$$n = r^2 p_1 \cdots p_k,$$

where r is an integer and the p_i are primes congruent to 1 *or* 2 *modulo* 4.

■

Note that 2 is the only prime conguent to 2 modulo 4. The primes excluded are those congruent to 3 modulo 4, such as 3, 7, 11, 19, etc.

We omit the proof of this result. If you wish to pursue it, you could look at one of the references at the end of the paper.

3 Quaternions

We would like to use a similar approach for

$$S_4 = \{m_1^2 + m_2^2 + m_3^2 + m_4^2 : m_i \in \mathbb{Z}\},$$

the set of all numbers expressible as the sum of four nonnegative square integers, and in fact we can do this by replacing the complex numbers by the quaternions.

A quaternion q takes the form

$$q = a + bi + cj + dk,$$

where a, b, c and d are real numbers, and the (distinct!) quaternion units i, j, k satisfy

$$\begin{array}{rclcrclcrcl} i^2 &=& j^2 &=& k^2 &=& -1 & , & & & \\ ij &=& k & , & jk &=& i & , & ki &=& j, \\ ji &=& -k & , & ki &=& -i & , & ik &=& -j. \end{array}$$

Apart from the famous oddity that the multiplication of i, j, and k is non-commutative, the usual rules of arithmetic apply:

$$\begin{aligned} q(q_1 + q_2) &= qq_1 + qq_2, \\ (q_1 + q_2)q &= q_1 q + q_2 q, \end{aligned}$$

and, provided a is real,

$$(aq)q_1 = a(qq_1) = q(aq_1) = (qq_1)a.$$

Thus if $q = a + bi + cj + dk$ and $q_1 = a_1 + b_1 i + c_1 j + d_1 k$, then we calculate

$$
\begin{aligned}
qq_1 &= (a + bi + cj + dk)(a_1 + b_1 i + c_1 j + d_1 k) \\
&= aa_1 + ab_1 i + ac_1 j + ad_1 k \\
&\quad + ba_1 i + bb_1 i^2 + bc_1 ij + bd_1 ik \\
&\quad + ca_1 j + cb_1 ji + cc_1 j^2 + cd_1 jk \\
&\quad + da_1 k + db_1 ki + dc_1 kj + dd_1 k^2 \qquad (2) \\
&= (aa_1 - bb_1 - cc_1 - dd_1) \\
&\quad + (ab_1 + ba_1 + cd_1 - dc_1) i \\
&\quad + (ac_1 - bd_1 + ca_1 + db_1) j \\
&\quad + (ad_1 + bc_1 - cb_1 + da_1) k.
\end{aligned}
$$

The conjugate of q is defined as

$$\bar{q} = a - bi - cj - dk,$$

and a calculation gives

$$q\bar{q} = a^2 + b^2 + c^2 + d^2.$$

The *norm* of q is defined as

$$|q| = \sqrt{a^2 + b^2 + c^2 + d^2},$$

so $|q|^2 = q\bar{q}$. Note also that $|\bar{q}| = |q|$.

The conjugate of a sum is the sum of the conjugates, and $\overline{aq} = a\overline{q}$ whenever a is real, but it is not usually true that the conjugate of a product is the product of the conjugates; the correct formula is

$$\overline{q_1 \cdot q_2} = \overline{q_2} \cdot \overline{q_1}.$$

This can be verified by another horrendous direct calculation, but the elegant way to see it is to observe that both sides are linear (over the reals) in both factors q_1, q_2, so it suffices to check the equation for the 16 cases in which q_1, q_2 are drawn from the basis $\{1, i, j, k\}$. The cases involving equal factors or the factor 1 are trivial, so given the symmetries of the system, this boils down to one equation:

$$\overline{i \cdot j} = \bar{k} = -k = j \cdot i = \bar{j} \cdot \bar{i}.$$

This allows us to imitate the calculations for complex numbers (taking care about the order of factors!), and get

$$\begin{aligned}|q_1 \cdot q_2|^2 = (q_1 \cdot q_2) \cdot \overline{(q_1 \cdot q_2)} &= q_1 \cdot q_2 \cdot \overline{q_2} \cdot \overline{q_1} \\ &= q_1 \cdot |q_2|^2 \cdot \overline{q_1} \\ &= q_1 \cdot \overline{q_1} \cdot |q_2|^2 \\ &= |q_1|^2 \cdot |q_2|^2 \\ &= (|q_1| \cdot |q_2|)^2,\end{aligned}$$

so taking square roots,

$$|q_1 \cdot q_2| = |q_1| \cdot |q_2|.$$

4 Sums of Four Squares

We are not going to prove Lagrange's Theorem, but we can now prove one of the main ingredients in the proof:

Lemma 4.1 (Euler) *The set S_4 of sums of four square integers is a multiplicative semigroup.*

Proof. Suppose we are given integers m_1, m_2, m_3, m_4, n_1, n_2, n_3, and n_4. We have to show that

$$(m_1^2 + m_2^2 + m_3^2 + m_4^2) \cdot (n_1^2 + n_2^2 + n_3^2 + n_4^2)$$

is the sum of four squares.

Take $q_1 = m_1 + m_2 i + m_3 j + m_4 k$ and $q_2 = n_1 + n_2 i + n_3 j + n_4 k$, and let

$$q_1 \cdot q_2 = p_1 + p_2 i + p_3 j + p_4 k.$$

Then from the formula (2) it is clear that the p_i are all integers, and

$$(m_1^2+m_2^2+m_3^2+m_4^2)\cdot(n_1^2+n_2^2+n_3^2+n_4^2) = |q_1|^2\cdot|q_2|^2 = |q_1\cdot q_2|^2 = p_1^2+p_2^2+p_3^2+p_4^2,$$

so we are done. ∎

This lemma reduces the problem of proving Lagrange's Theorem to the problem of proving that each prime is the sum of four squares. This is not a trivial remainder, and in fact it defeated Euler.

If you wish to get to the end of the story, you can read an account of the rest in Burton [B] or Herstein [H], for instance.

It is remarkable that Euler found his lemma long before the invention of quaternions. The above proof makes it really transparent, but he actually came up with what is effectively the same formula for the components p_i by sheer ingenuity. Armed with a knowledge of quaternions, you can easily reconstruct his formulas by substituting q_1 and q_2 into the formulas (2), getting:

$$\begin{aligned}
& (m_1^2+m_2^2+m_3^2+m_4^2)\cdot(n_1^2+n_2^2+n_3^2+n_4^2) \\
\\
= \quad & (m_1n_1-m_2n_2-m_3n_3-m_4n_4)^2 \\
& +(m_1n_2+m_2n_1+m_3n_4-m_4n_3)^2 \\
& +(m_1n_3+m_3n_1-m_2n_4+m_4n_2)^2 \\
& +(m_1n_4+m_4n_1+m_2n_3-m_3n_2)^2.
\end{aligned}$$

Now you can really amaze your friends by writing huge numbers as the sum of four squares. For instance, take 6469693230, which is the product of the first ten primes. We have

$$\begin{aligned}
2 &= 0^2+0^2+1^2+1^2 \\
3 &= 0^2+1^2+1^2+1^2 \\
5 &= 0^2+0^2+1^2+2^2 \\
7 &= 1^2+1^2+1^2+2^2 \\
11 &= 0^2+1^2+1^2+3^2 \\
13 &= 0^2+0^2+2^2+3^2 \\
17 &= 0^2+0^2+1^2+4^2 \\
19 &= 1^2+1^2+1^2+4^2 \\
23 &= 1^2+2^2+3^2+3^2 \\
29 &= 0^2+0^2+2^2+5^2.
\end{aligned}$$

Following the quaternion method we used to prove Euler's Lemma, we form the quaternions

$$\begin{aligned} a_2 &= 0 + 0i + 1j + 1k \\ a_3 &= 0 + 1i + 1j + 1k \\ a_5 &= 0 + 0i + 1j + 2k \\ a_7 &= 1 + 1i + 1j + 2k \\ a_{11} &= 0 + 1i + 1j + 3k \\ a_{13} &= 0 + 0i + 2j + 3k \\ a_{17} &= 0 + 0i + 1j + 4k \\ a_{19} &= 1 + 1i + 1j + 4k \\ a_{23} &= 1 + 2i + 3j + 3k \\ a_{29} &= 0 + 0i + 2j + 5k. \end{aligned}$$

Then we multiply them all together, getting

$$-70176 - 37594i + 6997j + 9097k.$$

A quick calculation verifies that

$$70176^2 + 37594^2 + 6997^2 + 9097^2 = 6469693230,$$

as expected.

Here are a few exercises that can be tackled using the methods developed above:

1. Express the following numbers as the sum of four squares:

   (a) 2640330.

   (b) 200560490130.

   (c) 277945762500.

2. (a) 1462500.

   (b) 2371330.

   (c) 44240625.

(Hint: Start by factorising the number, in each case.)

We close by remarking that the sums of *three* squares cannot be tackled in this kind of way, and the theorem that characterises these (due to Gauss) is quite deep. You can find it in Serre [S].

Acknowledgment

The author is grateful to Ian Short and the editor for their comments on a draft of this paper.

References

[B] D.M. Burton. Elementary Number Theory. Revised printing. Allyn and Bacon. 1980.
[H] I.N. Herstein. Topics in Algebra. Blaisdell. 1964.
[S] J.-P. Serre. A Course in Arithmetic. Springer. 1973.