Primes of the form $x^2 + ny^2$

Daniel Fretwell

School of Mathematics and Statistics, University of Sheffield

Semester 2, 2010/2011

Outline of talk

1 History of the problem : Fermat and Euler

2 History of the problem: Lagrange and Gauss

3 Picking up the pieces

Fermat's beautiful result

To get to the roots of this problem we need to time travel back to Christmas Day, 1640. This is the day when Fermat made a nice discovery.

Fermat had probably just unwrapped his 41st present (yet another tie) when he realised something nice, that $41 = 4^2 + 5^2$ and also that $41 \equiv 1 \bmod 4$.

Why is this of any importance? Well he was interested in which positive integers could be written as a sum of two integer squares. So far he had found that only the numbers $1, 2, 4, 5, 8, 9, 10, 13, 16, 17, 18, 20, 25, 26, 29, 32, 34, 36, 37, 40$ and 41 had this property.

Fermat's beautiful result

To get to the roots of this problem we need to time travel back to Christmas Day, 1640. This is the day when Fermat made a nice discovery.

Fermat had probably just unwrapped his 41st present (yet another tie) when he realised something nice, that $41 = 4^2 + 5^2$ and also that $41 \equiv 1 \bmod 4$.

Why is this of any importance? Well he was interested in which positive integers could be written as a sum of two integer squares. So far he had found that only the numbers $1, 2, 4, 5, 8, 9, 10, 13, 16, 17, 18, 20, 25, 26, 29, 32, 34, 36, 37, 40$ and 41 had this property.

Fermat's beautiful result

To get to the roots of this problem we need to time travel back to Christmas Day, 1640. This is the day when Fermat made a nice discovery.

Fermat had probably just unwrapped his 41st present (yet another tie) when he realised something nice, that $41 = 4^2 + 5^2$ and also that $41 \equiv 1 \bmod 4$.

Why is this of any importance? Well he was interested in which positive integers could be written as a sum of two integer squares. So far he had found that only the numbers $1, 2, 4, 5, 8, 9, 10, 13, 16, 17, 18, 20, 25, 26, 29, 32, 34, 36, 37, 40$ and 41 had this property.

If we look at the prime numbers in this list we find that $2, 5, 13, 17, 29, 37, 41$ are all expressible as a sum of two square numbers. A closer look shows that all of these (except 2) are congruent to 1 mod 4. What is even more startling is that every prime that is congruent to 1 mod 4 is in this list so far.

Fermat was excited by this. He checked that the pattern held further and conjectured the following in a letter to Mersenne:

Fermat's two square theorem

An odd prime number p can be written as $x^2 + y^2$ for integers x, y if and only if $p \equiv 1 \bmod 4$.

If we look at the prime numbers in this list we find that $2, 5, 13, 17, 29, 37, 41$ are all expressible as a sum of two square numbers. A closer look shows that all of these (except 2) are congruent to 1 mod 4. What is even more startling is that every prime that is congruent to 1 mod 4 is in this list so far.

Fermat was excited by this. He checked that the pattern held further and conjectured the following in a letter to Mersenne:

Fermat's two square theorem

An odd prime number p can be written as $x^2 + y^2$ for integers x, y if and only if $p \equiv 1 \bmod 4$.

If we look at the prime numbers in this list we find that $2, 5, 13, 17, 29, 37, 41$ are all expressible as a sum of two square numbers. A closer look shows that all of these (except 2) are congruent to 1 mod 4. What is even more startling is that every prime that is congruent to 1 mod 4 is in this list so far.

Fermat was excited by this. He checked that the pattern held further and conjectured the following in a letter to Mersenne:

Fermat's two square theorem

An odd prime number p can be written as $x^2 + y^2$ for integers x, y if and only if $p \equiv 1 \bmod 4$.

Actually this result lets us decide when any given integer can be written as a sum of squares. This is done by a clever use of the identity:

$$(a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (ad - bc)^2$$

telling us that the product of two numbers that are sums of squares is again a sum of squares.

The full result is as follows:

Theorem

A non-negative integer m can be written in the form $x^2 + y^2$ if and only if for every prime $p = 3 \bmod 4$ that divides m, we have that p divides m to an even power.

Actually this result lets us decide when any given integer can be written as a sum of squares. This is done by a clever use of the identity:

$$(a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (ad - bc)^2$$

telling us that the product of two numbers that are sums of squares is again a sum of squares.

The full result is as follows:

Theorem

A non-negative integer m can be written in the form $x^2 + y^2$ if and only if for every prime $p \equiv 3 \bmod 4$ that divides m, we have that p divides m to an even power.

To demonstrate this result take the positive integer 612. This has prime factorisation $2^2 \cdot 17 \cdot 3^2$.

Now we see that the only odd primes dividing 612 are 17 and 3. We find that $17 \equiv 1 \bmod 4$ and $3 \equiv 3 \bmod 4$. The prime 3 occurs to an even power in the prime factorisation above so the result tells us that 612 is a sum of squares. In fact $612 = 6^2 + 24^2$ (we could have found this using the identity mentioned earlier but trial and error works better here).

Take another positive integer, say $14 = 2 \cdot 7$. Here we have that $7 \equiv 3 \bmod 4$ occurs to an odd power in the factorisation and so 14 is not a sum of two squares. This is easily checked by hand.

To demonstrate this result take the positive integer 612. This has prime factorisation $2^2 \cdot 17 \cdot 3^2$.

Now we see that the only odd primes dividing 612 are 17 and 3. We find that $17 \equiv 1 \bmod 4$ and $3 \equiv 3 \bmod 4$. The prime 3 occurs to an even power in the prime factorisation above so the result tells us that 612 is a sum of squares. In fact $612 = 6^2 + 24^2$ (we could have found this using the identity mentioned earlier but trial and error works better here).

Take another positive integer, say $14 = 2 \cdot 7$. Here we have that $7 \equiv 3 \bmod 4$ occurs to an odd power in the factorisation and so 14 is not a sum of two squares. This is easily checked by hand.

Similar problems

Fermat studied a few related problems. He found the following nice results for rational primes p:

$$p = x^2 + 2y^2 \iff p = 2 \text{ or } p \equiv 1, 3 \bmod 8$$

$$p = x^2 + 3y^2 \iff p = 3 \text{ or } p \equiv 1 \bmod 3$$

These were proved by Euler using descent methods.

This is remarkable, each of these results seems to classify the primes that can be written in the form $x^2 + 2y^2$ or $x^2 + 3y^2$ by one simple congruence condition on p.

Similar problems

Fermat studied a few related problems. He found the following nice results for rational primes p:

$$p = x^2 + 2y^2 \iff p = 2 \text{ or } p \equiv 1, 3 \bmod 8$$

$$p = x^2 + 3y^2 \iff p = 3 \text{ or } p \equiv 1 \bmod 3$$

These were proved by Euler using descent methods.

This is remarkable, each of these results seems to classify the primes that can be written in the form $x^2 + 2y^2$ or $x^2 + 3y^2$ by one simple congruence condition on p.

One question arises. Given a positive integer n, can we find a corresponding congruence on the primes p that can be written in the form $x^2 + ny^2$?

Before discussing this, it should first be mentioned that, as in the case where $n = 1$, we can use the primes that can be written in the form $x^2 + ny^2$ to determine which positive integers can be written in this form. This is done by use of a more general identity than what we used earlier:

$$(a^2 + nb^2)(c^2 + nd^2) = (ac + nbd)^2 + n(ad - bc)^2$$

One question arises. Given a positive integer n, can we find a corresponding congruence on the primes p that can be written in the form $x^2 + ny^2$?

Before discussing this, it should first be mentioned that, as in the case where $n = 1$, we can use the primes that can be written in the form $x^2 + ny^2$ to determine which positive integers can be written in this form. This is done by use of a more general identity than what we used earlier:

$$(a^2 + nb^2)(c^2 + nd^2) = (ac + nbd)^2 + n(ad - bc)^2$$

Returning to the problem, it seems that in all of the cases we have discussed, the primes p that could be written in the form $x^2 + ny^2$ were ones such that either $p|n$ or $\left(\frac{-n}{p}\right) = 1$.

This is actually only true one way:

Theorem

We have that:

$$p = x^2 + ny^2 \Longrightarrow p|n \text{ or } \left(\frac{-n}{p}\right) = 1$$

Proof

Suppose that $p = x^2 + ny^2$, then $x^2 + ny^2 \equiv 0 \bmod p$. If $x \not\equiv 0 \bmod p$ then $y \not\equiv 0 \bmod p$ and so rearranging gives $(xy^{-1})^2 \equiv -n \bmod p$. Hence $\left(\frac{-n}{p}\right) = 1$.

If $x \equiv 0 \bmod p$ then clearly $p|n$.

Returning to the problem, it seems that in all of the cases we have discussed, the primes p that could be written in the form $x^2 + ny^2$ were ones such that either $p|n$ or $\left(\frac{-n}{p}\right) = 1$.

This is actually only true one way:

Theorem

We have that:

$$p = x^2 + ny^2 \Longrightarrow p|n \text{ or } \left(\frac{-n}{p}\right) = 1$$

Proof

Suppose that $p = x^2 + ny^2$, then $x^2 + ny^2 \equiv 0 \bmod p$. If $x \not\equiv 0 \bmod p$ then $y \not\equiv 0 \bmod p$ and so rearranging gives $(xy^{-1})^2 \equiv -n \bmod p$. Hence $\left(\frac{-n}{p}\right) = 1$.

If $x \equiv 0 \bmod p$ then clearly $p|n$.

Returning to the problem, it seems that in all of the cases we have discussed, the primes p that could be written in the form $x^2 + ny^2$ were ones such that either $p|n$ or $\left(\frac{-n}{p}\right) = 1$.

This is actually only true one way:

Theorem

We have that:

$$p = x^2 + ny^2 \Longrightarrow p|n \text{ or } \left(\frac{-n}{p}\right) = 1$$

Proof

Suppose that $p = x^2 + ny^2$, then $x^2 + ny^2 \equiv 0 \bmod p$. If $x \not\equiv 0 \bmod p$ then $y \not\equiv 0 \bmod p$ and so rearranging gives $(xy^{-1})^2 \equiv -n \bmod p$. Hence $\left(\frac{-n}{p}\right) = 1$.

If $x \equiv 0 \bmod p$ then clearly $p|n$.

To see why the converse of this theorem does not work consider the case where $n = 5$. We have (by quadratic reciprocity) that $\left(\frac{-5}{p}\right) = 1$ exactly when $p \equiv 1, 3, 7, 9 \bmod 20$, yet the primes 3 and 7 cannot be written in the form $x^2 + 5y^2$.

Euler conjectured the actual result here:

$$p = x^2 + 5y^2 \iff p = 5 \text{ or } p \equiv 1, 9 \bmod 20$$

What is the significance of the two congruence classes we had to throw away and in general how do we decide which ones to keep?

This question is very difficult to answer.

To see why the converse of this theorem does not work consider the case where $n = 5$. We have (by quadratic reciprocity) that $\left(\frac{-5}{p}\right) = 1$ exactly when $p \equiv 1, 3, 7, 9 \bmod 20$, yet the primes 3 and 7 cannot be written in the form $x^2 + 5y^2$.

Euler conjectured the actual result here:

$$p = x^2 + 5y^2 \iff p = 5 \text{ or } p \equiv 1, 9 \bmod 20$$

What is the significance of the two congruence classes we had to throw away and in general how do we decide which ones to keep?

This question is very difficult to answer.

To see why the converse of this theorem does not work consider the case where $n = 5$. We have (by quadratic reciprocity) that $\left(\frac{-5}{p}\right) = 1$ exactly when $p \equiv 1, 3, 7, 9 \bmod 20$, yet the primes 3 and 7 cannot be written in the form $x^2 + 5y^2$.

Euler conjectured the actual result here:

$$p = x^2 + 5y^2 \iff p = 5 \text{ or } p \equiv 1, 9 \bmod 20$$

What is the significance of the two congruence classes we had to throw away and in general how do we decide which ones to keep?

This question is very difficult to answer.

Outline of talk

1 History of the problem : Fermat and Euler

2 History of the problem: Lagrange and Gauss

3 Picking up the pieces

The problem took a new direction after Fermat and Euler. Gauss and Lagrange independently decided to study the behaviour of integer binary quadratic forms, i.e. functions $f(x, y) = ax^2 + bxy + cy^2$ of an integer variable (and with integer coefficients). If a, b, c all share no common factor bigger than 1 then we call the form primitive.

New methods of study were arising. An equivalence relation can be placed on the set of integer binary quadratic forms by declaring that two such forms $f(x, y)$ and $g(x, y)$ are equivalent if and only if there exists integers a, b, c, d with $ad - bc = \pm 1$ such that $g(x, y) = f(ax + by, cx + dy)$. When we can achieve a positive sign in the above it is called a proper equivalence.

The problem took a new direction after Fermat and Euler. Gauss and Lagrange independently decided to study the behaviour of integer binary quadratic forms, i.e. functions $f(x, y) = ax^2 + bxy + cy^2$ of an integer variable (and with integer coefficients). If a, b, c all share no common factor bigger than 1 then we call the form primitive.

New methods of study were arising. An equivalence relation can be placed on the set of integer binary quadratic forms by declaring that two such forms $f(x, y)$ and $g(x, y)$ are equivalent if and only if there exists integers a, b, c, d with $ad - bc = \pm 1$ such that $g(x, y) = f(ax + by, cx + dy)$. When we can achieve a positive sign in the above it is called a proper equivalence.

It turns out that every primitive positive definite form is properly equivalent to a specific kind of form, called a reduced form. When n is a positive integer, the form $x^2 + ny^2$ is always one of these reduced forms.

It also turns out that the primes with $\left(\frac{-n}{p}\right) = 1$ are partitioned by considering the reduced forms of discriminant $-4n$ that represent them. This is why when deciding which primes can be written in the form $x^2 + ny^2$ we had to throw away certain classes of integers mod $4n$, often there is more than one reduced form of discriminant $-4n$.

So it is now clear that when there is only one reduced form it must represent all such primes. This happens when $n = 1, 2, 3, 4$ and 7, explaining why our first few examples worked perfectly and why the $n = 5$ one didn't (we find that $x^2 + 5y^2$ and $2x^2 + 2xy + 3y^2$ are the reduced forms of discriminant -20).

It turns out that every primitive positive definite form is properly equivalent to a specific kind of form, called a reduced form. When n is a positive integer, the form $x^2 + ny^2$ is always one of these reduced forms.

It also turns out that the primes with $\left(\frac{-n}{p}\right) = 1$ are partitioned by considering the reduced forms of discriminant $-4n$ that represent them. This is why when deciding which primes can be written in the form $x^2 + ny^2$ we had to throw away certain classes of integers mod $4n$, often there is more than one reduced form of discriminant $-4n$.

So it is now clear that when there is only one reduced form it must represent all such primes. This happens when $n = 1, 2, 3, 4$ and 7, explaining why our first few examples worked perfectly and why the $n = 5$ one didn't (we find that $x^2 + 5y^2$ and $2x^2 + 2xy + 3y^2$ are the reduced forms of discriminant -20).

It turns out that every primitive positive definite form is properly equivalent to a specific kind of form, called a reduced form. When n is a positive integer, the form $x^2 + ny^2$ is always one of these reduced forms.

It also turns out that the primes with $\left(\frac{-n}{p}\right) = 1$ are partitioned by considering the reduced forms of discriminant $-4n$ that represent them. This is why when deciding which primes can be written in the form $x^2 + ny^2$ we had to throw away certain classes of integers mod $4n$, often there is more than one reduced form of discriminant $-4n$.

So it is now clear that when there is only one reduced form it must represent all such primes. This happens when $n = 1, 2, 3, 4$ and 7, explaining why our first few examples worked perfectly and why the $n = 5$ one didn't (we find that $x^2 + 5y^2$ and $2x^2 + 2xy + 3y^2$ are the reduced forms of discriminant -20).

Gauss took this further and formed a finite group out of the equivalence classes of forms, the so called form class group. The operation here was a special operation called composition of forms. We saw an example of this earlier when we saw the identity:

$$(a^2 + nb^2)(c^2 + nd^2) = (ac + nbd)^2 + n(ad - bc)^2 = c^2 + nd^2$$

This is telling us that the equivalence class of the form $x^2 + ny^2$ is the identity in the form class group since composing $x^2 + ny^2$ with itself gives the same form.

Studies of the form class group (sometimes) told Gauss exactly which classes mod $4n$ to throw away to tell us when $p = x^2 + ny^2$.

You might think that this solves the problem fully but actually it doesn't. It turns out that in some cases two reduced forms are indistinguishable by simply looking at the classes of numbers that they represent. This causes trouble however, when this behaviour doesn't occur we have a full solution!

We need a better way.

Studies of the form class group (sometimes) told Gauss exactly which classes mod $4n$ to throw away to tell us when $p = x^2 + ny^2$.

You might think that this solves the problem fully but actually it doesn't. It turns out that in some cases two reduced forms are indistinguishable by simply looking at the classes of numbers that they represent. This causes trouble however, when this behaviour doesn't occur we have a full solution!

We need a better way.

Studies of the form class group (sometimes) told Gauss exactly which classes mod $4n$ to throw away to tell us when $p = x^2 + ny^2$.

You might think that this solves the problem fully but actually it doesn't. It turns out that in some cases two reduced forms are indistinguishable by simply looking at the classes of numbers that they represent. This causes trouble however, when this behaviour doesn't occur we have a full solution!

We need a better way.

Outline of talk

1 History of the problem : Fermat and Euler

2 History of the problem: Lagrange and Gauss

3 Picking up the pieces

Moving into more abstract realms

The full solution of this problem comes from a quite sophisticated branch of algebraic number theory called class field theory, the topic of my project.

Basically, the idea is to work in the number field $K = \mathbb{Q}(\sqrt{-n})$. This field contains the ring $\mathbb{Z}[\sqrt{-n}]$, which is perfect for our problem since being able to write $p = x^2 + ny^2$ is the same as being able to factorise p in this ring as:

$$p = (x + y\sqrt{-n})(x - y\sqrt{-n})$$

Moving into more abstract realms

The full solution of this problem comes from a quite sophisticated branch of algebraic number theory called class field theory, the topic of my project.

Basically, the idea is to work in the number field $K = \mathbb{Q}(\sqrt{-n})$. This field contains the ring $\mathbb{Z}[\sqrt{-n}]$, which is perfect for our problem since being able to write $p = x^2 + ny^2$ is the same as being able to factorise p in this ring as:

$$p = (x + y\sqrt{-n})(x - y\sqrt{-n})$$

Next comes the introduction of a new field K_R, the ring class field of K with respect to $\mathbb{Z}[\sqrt{-n}]$ (the Hilbert class field is a special case of this).

This is a finite field extension of K and the degree of this extension is the same as the size of the form class group mentioned earlier (for forms of discriminant $-4n$).

This property, along with the existence and uniqueness of K_R, is provided by the theorems of class field theory.

Next comes the introduction of a new field K_R, the ring class field of K with respect to $\mathbb{Z}[\sqrt{-n}]$ (the Hilbert class field is a special case of this).

This is a finite field extension of K and the degree of this extension is the same as the size of the form class group mentioned earlier (for forms of discriminant $-4n$).

This property, along with the existence and uniqueness of K_R, is provided by the theorems of class field theory.

It turns out that being able to write $p = x^2 + ny^2$ is exactly the same as the ideal $\langle p \rangle$ of the ring of integers of K_R factorising as much as possible (into $[K_R : K]$ distinct prime ideals). This also comes from class field theory.

Then, other theorems from algebraic number theory tell us exactly when this can happen. The full result is as follows:

Theorem

Let $n > 0$ be an integer and p be a prime number. Then there is some monic irreducible polynomial $f_n(x)$ over $\mathbb{Z}$ such that if $p \nmid n$ and p does not divide the discriminant of $f_n(x)$ then:

$$p = x^2 + ny^2 \iff \begin{cases} \left(\frac{-n}{p}\right) = 1 \text{ and} \\ f_n(a) \equiv 0 \bmod p \text{ for some integer } a \end{cases}$$

It turns out that being able to write $p = x^2 + ny^2$ is exactly the same as the ideal $\langle p \rangle$ of the ring of integers of K_R factorising as much as possible (into $[K_R : K]$ distinct prime ideals). This also comes from class field theory.

Then, other theorems from algebraic number theory tell us exactly when this can happen. The full result is as follows:

Theorem

Let $n > 0$ be an integer and p be a prime number. Then there is some monic irreducible polynomial $f_n(x)$ over $\mathbb{Z}$ such that if $p \nmid n$ and p does not divide the discriminant of $f_n(x)$ then:

$$p = x^2 + ny^2 \iff \begin{cases} \left(\frac{-n}{p}\right) = 1 \text{ and} \\ f_n(a) \equiv 0 \bmod p \text{ for some integer } a \end{cases}$$

The polynomial $f_n(x)$ can be taken to be the minimal polynomial over $\mathbb{Z}$ of any real algebraic integer generator of K_R over K.

Example

When $n = 27$ it turns out that the ring class field of $K = \mathbb{Q}(\sqrt{-14})$ with respect to $\mathbb{Z}[\sqrt{-14}]$ is $K_R = K(\sqrt[3]{2})$. Thus we can take the polynomial to be $f_{27}(x) = x^3 - 2$ (this has discriminant $-2^2 \cdot 3^3$). Also note that $\left(\frac{-27}{p}\right) = \left(\frac{-3}{p}\right)$, which is 1 exactly when $p \equiv 1 \bmod 3$ (by quadratic reciprocity).

So we get that if $p \neq 2, 3$ then:

$$p = x^2 + 27y^2 \iff p \equiv 1 \bmod 3 \text{ and } a^3 \equiv 2 \bmod p$$

for some integer a.

The polynomial $f_n(x)$ can be taken to be the minimal polynomial over $\mathbb{Z}$ of any real algebraic integer generator of K_R over K.

Example

When $n = 27$ it turns out that the ring class field of $K = \mathbb{Q}(\sqrt{-14})$ with respect to $\mathbb{Z}[\sqrt{-14}]$ is $K_R = K(\sqrt[3]{2})$. Thus we can take the polynomial to be $f_{27}(x) = x^3 - 2$ (this has discriminant $-2^2 \cdot 3^3$). Also note that $\left(\frac{-27}{p}\right) = \left(\frac{-3}{p}\right)$, which is 1 exactly when $p \equiv 1 \bmod 3$ (by quadratic reciprocity).

So we get that if $p \neq 2, 3$ then:

$$p = x^2 + 27y^2 \iff p \equiv 1 \bmod 3 \text{ and } a^3 \equiv 2 \bmod p$$

for some integer a.

To illustrate the previous example, take the prime $p = 31$. Now 31 can be written in the form $x^2 + 27y^2$ if we let $x = 2$ and $y = 1$. Checking the other side we find that $31 \equiv 1 \bmod 3$ and $a = 4$ is such that $a^3 \equiv 2 \bmod 31$.

Conversely we find that if we take $p = 43 \equiv 1 \bmod 3$ then $a = 9$ is such that $a^3 \equiv 2 \bmod 43$ and so 43 should be expressible in the form $x^2 + 27y^2$. A quick check shows that it is, taking $x = 4$ and $y = 1$.

To illustrate the previous example, take the prime $p = 31$. Now 31 can be written in the form $x^2 + 27y^2$ if we let $x = 2$ and $y = 1$. Checking the other side we find that $31 \equiv 1 \bmod 3$ and $a = 4$ is such that $a^3 \equiv 2 \bmod 31$.

Conversely we find that if we take $p = 43 \equiv 1 \bmod 3$ then $a = 9$ is such that $a^3 \equiv 2 \bmod 43$ and so 43 should be expressible in the form $x^2 + 27y^2$. A quick check shows that it is, taking $x = 4$ and $y = 1$.

So is the problem solved entirely?

Theoretically yes, but to get a practical solution we would have to be able to find these ring class fields.

Fortunately, this problem has been solved too and we do have algorithms for finding it. These rely on elliptic curves having complex multiplication by $\mathbb{Z}[\sqrt{-n}]$ and their corresponding j-invariants.

So is the problem solved entirely?

Theoretically yes, but to get a practical solution we would have to be able to find these ring class fields.

Fortunately, this problem has been solved too and we do have algorithms for finding it. These rely on elliptic curves having complex multiplication by $\mathbb{Z}[\sqrt{-n}]$ and their corresponding j-invariants.

So is the problem solved entirely?

Theoretically yes, but to get a practical solution we would have to be able to find these ring class fields.

Fortunately, this problem has been solved too and we do have algorithms for finding it. These rely on elliptic curves having complex multiplication by $\mathbb{Z}[\sqrt{-n}]$ and their corresponding j-invariants.

That's all folks

The end.