PROOFS BY DESCENT

KEITH CONRAD

> *As ordinary methods, such as are found in the books, are inadequate to proving such difficult propositions, I discovered at last a most singular method . . . that I called the infinite descent.*   Fermat, 1659.

1. Introduction

The method of descent is a technique developed by Fermat for proving certain equations have no (or few) integral solutions. The idea is to show that if there is an integral solution to an equation then there is another integral solution that is smaller in some way. Repeating this process and comparing the sizes of the successive solutions leads to an infinitely decreasing sequence

$$a_1 > a_2 > a_3 > \cdots$$

of positive integers, and that is impossible. Let's take a look at two examples.

Example 1.1 (Euler)**.** We will show the equation $x^3 + 2y^3 + 4z^3 = 0$ has no solution in integers other than the obvious solution $(0, 0, 0)$. Assume there is a solution $(x, y, z) \neq (0, 0, 0)$, so at least one of x, y, and z is not 0. The equation tells us x^3 is even, so x is even. Write $x = 2x'$. Then $8x'^3 + 2y^3 + 4z^3 = 0$. Dividing by 2 and rearranging terms, we get $y^3 + 2z^3 + 4x'^3 = 0$. This is just like our original equation, with (x, y, z) replaced by (y, z, x'). Since y is now playing the role previously played by x, the argument used before on x shows y is even. Writing $y = 2y'$, substituting this in, and removing a common factor of 2, we get $z^3 + 2x'^3 + 4y'^3 = 0$. Therefore z is even, so $z = 2z'$. Substituting this in and simplifying, $x'^3 + 2y'^3 + 4z'^3 = 0$. Thus (x', y', z') fits the original equation and at least one of x', y' or z' is nonzero (corresponding to whichever of x, y, and z is nonzero). Since $0 < \max(|x'|, |y'|, |z'|) = (1/2) \max(|x|, |y|, |z|)$, we have produced a smaller integral solution measured by the maximum absolute value, which is a positive integer. This process can be repeated infinitely often, leading to a contradiction.

The same proof shows for each prime p that the equation $x^3 + py^3 + p^2z^3 = 0$ has no integral solution other than $(0, 0, 0)$. Indeed, if (x, y, z) fits the equation then $p \mid x^3$, so $p \mid x$ and we can proceed exactly as in the special case $p = 2$.

In Section 2 we will give proofs by descent that certain numbers are irrational. In Section 3 we will show the equation $a^4 + b^4 = c^4$ (a special case of Fermat's Last Theorem) has no solution in positive integers using descent. In Section 4 we will use descent to show certain equations have no solution in nonconstant rational functions. In a positive direction, descent will be used in Section 5 to show each prime p such that $-1 \equiv \square \bmod p$ is a sum of two squares and in Section 6 to show each positive integer is a sum of four squares. In Section 7 we will argue by descent that for all integers $k > 0$ other than 1 or 3, the equation $x^2 + y^2 + z^2 = kxyz$ has no integral solutions (x, y, z) besides $(0, 0, 0)$.

While descent may appear to be something like "reverse induction," it is not as widely applicable in the whole of mathematics as induction. Descent is nevertheless quite central to some important developments in number theory.

2. Irrationality by descent

Here is the usual proof that $\sqrt{2}$ is irrational, expressed using the idea of descent.

Example 2.1. We assume $\sqrt{2}$ is rational, so $\sqrt{2} = a/b$ with positive integers a and b. Squaring both sides and clearing the denominator, $2b^2 = a^2$. (This is an equation we want to show is not solvable in positive integers.) Since $2 \mid a^2$, $2 \mid a$. Write $a = 2a'$ for some positive integer a', so $2b^2 = 4a'^2$, which is the same as $b^2 = 2a'^2$. Thus $2 \mid b^2$, so $2 \mid b$. Write $b = 2b'$, so $4b'^2 = 2a'^2$, which is the same as $2b'^2 = a'^2$. Since a' and b' are positive, we have $\sqrt{2} = a'/b'$, so

$$\sqrt{2} = \frac{a}{b} = \frac{a'}{b'}.$$

Since $b = 2b'$ and both b and b' are positive, $0 < b' < b$, so we started with one rational expression for $\sqrt{2}$ and found another rational expression with a smaller (positive) denominator. Now we can repeat this process and obtain a sequence of rational expressions for $\sqrt{2}$ with decreasing positive denominators. This can't go on forever, so we have a contradiction.

Where would this proof break down if we tried to adapt it to prove $\sqrt{4}$ is irrational by contradiction? Starting from $\sqrt{4} = a/b$ for a and b in $\mathbf{Z}^+$, we'd get $4 \mid a^2$, so $2 \mid a$ (*not* $4 \mid a$), and writing $a = 2a'$, $4a'^2 = a^2 = 4b^2$, so $a'^2 = b^2$ and we can't show b is even too.

The way this proof usually is written starts with $\sqrt{2} = a/b$ where the fraction is in lowest terms. Then the fact that $a = 2a'$ and $b = 2b'$, as shown in the theorem, is a contradiction since it means the fraction wasn't in lowest terms. The method of descent bypassed having to put the fraction in lowest terms, obtaining a contradiction in a different way.

Let's take a look at another proof by descent that $\sqrt{2}$ is irrational. We assume $\sqrt{2}$ is rational. Since $1 < \sqrt{2} < 2$, we can write

$$(2.1) \qquad\qquad \sqrt{2} = 1 + \frac{m}{n},$$

where m and n are positive integers with $0 < m/n < 1$, so $0 < m < n$. Squaring both sides of (2.1) and clearing the denominator,

$$2n^2 = n^2 + 2mn + m^2,$$

so $m^2 = n^2 - 2mn = n(n - 2m)$. Since m^2 and n are positive, so is $n - 2m$, and

$$\frac{m}{n} = \frac{n - 2m}{m}.$$

This in between 0 and 1, by the definition of m/n, so $0 < n - 2m < m$. We have reached the descent step: the fractional part m/n of $\sqrt{2}$ has been written as a fraction $(n - 2m)/m$ with a smaller denominator than before: $0 < m < n$. We can repeat this process again and again, eventually reaching a contradiction.

This proof by descent that $\sqrt{2}$ is irrational is not the same as the proof by descent in Example 2.1, since it does not use anything about even and odd numbers. It also generalizes nicely to other square roots.

Theorem 2.2. *If $d \in \mathbf{Z}^+$ and d is not a perfect square then $\sqrt{d}$ is irrational.*

Proof. (Dedekind, 1858) Suppose $\sqrt{d}$ is rational. Since d is *not* a perfect square, its positive square root lies strictly between two consecutive integers. Let ℓ be the integer such that $\ell < \sqrt{d} < \ell + 1$. (Note ℓ is uniquely determined by $\sqrt{d}$.) Write

$$\sqrt{d} = \ell + \frac{m}{n},$$

where m and n are positive integers with $0 < m/n < 1$, so $0 < m < n$. Squaring both sides and clearing the denominator,

$$dn^2 = n^2\ell^2 + 2mn\ell + m^2,$$

so $m^2 = nq$, where $q = n(d - \ell^2) - 2m\ell$. Since m^2 and n are positive, q is positive. Then $m/n = q/m$, so

$$\sqrt{d} = \ell + \frac{m}{n} = \ell + \frac{q}{m}.$$

Since $q/m = m/n$, $0 < q/m < 1$, so $0 < q < m$. The fraction q/m has a smaller (positive) denominator than m/n, so from one representation $\sqrt{d} - \ell = m/n$ we get another representation $\sqrt{d} - \ell = q/m$ with a smaller (positive) denominator. This leads to a contradiction by repeating this process enough times. □

Here is another proof of Theorem 2.2, using descent in $\mathbf{Z}^2$ rather than in $\mathbf{Z}$. The argument is taken from [8].

Proof. Set $A = \left(\begin{smallmatrix} 0 & d \\ 1 & 0 \end{smallmatrix}\right)$. Its characteristic polynomial is $\det(\lambda I_2 - A) = \lambda^2 - d$, with an eigenvalue $\sqrt{d}$ and associated eigenvector $\left(\begin{smallmatrix} \sqrt{d} \\ 1 \end{smallmatrix}\right)$. Assuming $\sqrt{d}$ is rational, write $\sqrt{d} = a/b$ with nonzero integers a and b. A scalar multiple of an eigenvector is an eigenvector, and $\left(\begin{smallmatrix} \sqrt{d} \\ 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} a/b \\ 1 \end{smallmatrix}\right)$ can be scaled to $\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$. This is also an eigenvector of A: $A\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) = \left(\begin{smallmatrix} db \\ a \end{smallmatrix}\right) = \sqrt{d}\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$. Let ℓ be the integer such that $\ell < \sqrt{d} < \ell + 1$. Then

$$(A - \ell I_2)\begin{pmatrix} a \\ b \end{pmatrix} = \sqrt{d}\begin{pmatrix} a \\ b \end{pmatrix} - \ell\begin{pmatrix} a \\ b \end{pmatrix} = (\sqrt{d} - \ell)\begin{pmatrix} a \\ b \end{pmatrix},$$

where $\sqrt{d} - \ell$ lies between 0 and 1. The integral vector $\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$ is an eigenvector of the integral matrix $A - \ell I_2$ with eigenvalue between 0 and 1.

Since $\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$ is an eigenvector of $A - \ell I_2$, it is also an eigenvector of $(A - \ell I_2)^r$ for each $r \geq 1$, with eigenvalue $(\sqrt{d} - \ell)^r$:

$$(A - \ell I_2)^r\begin{pmatrix} a \\ b \end{pmatrix} = (\sqrt{d} - \ell)^r\begin{pmatrix} a \\ b \end{pmatrix}.$$

On the left side, for all $r \geq 1$ we have a vector in $\mathbf{Z}^2$ since A has integer entries and a, b, and ℓ are integers. On the right side we have a *nonzero* vector (since a, b, and $\sqrt{d} - \ell$ are nonzero) and it is getting arbitrarily small as r grows since $|\sqrt{d} - \ell| < 1$. So we have a sequence of nonzero vectors in $\mathbf{Z}^2$ with length shrinking to 0 (the descent idea). This is impossible, so we have a contradiction. □

We can extend the same proof to cube roots, using descent in $\mathbf{Z}^3$.

Theorem 2.3. *If $d \in \mathbf{Z}$ and d is not a perfect cube then $\sqrt[3]{d}$ is irrational.*

Proof. Suppose $\sqrt[3]{d} = a/b$ with nonzero integers a and b. Let

$$A = \begin{pmatrix} 0 & 0 & d \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad v = \begin{pmatrix} a^2 \\ ab \\ b^2 \end{pmatrix},$$

so $\det(\lambda I_3 - A) = \lambda^3 - d$ and $Av = (a/b)v = \sqrt[3]{d}v$.

Let $\ell \in \mathbf{Z}$ satisfy $\ell < \sqrt[3]{d} < \ell + 1$, so $(A - \ell I_3)v = (\sqrt[3]{d} - \ell)v$. Then

$$(2.2) \qquad\qquad (A - \ell I_3)^r v = (\sqrt[3]{d} - \ell)^r v$$

for all $r \geq 1$. Since $v \in \mathbf{Z}^3$ and A has integer entries, the left side of (2.2) is a vector in $\mathbf{Z}^3$. Since $v \neq \mathbf{0}$ and $0 < \sqrt[3]{d} - \ell < 1$, the right side of (2.2) is nonzero and its length is tending to 0 as r grows. Thus, as $r \to \infty$, the nonzero vectors $(A - \ell I_3)^r v$ are a sequence in $\mathbf{Z}^3$ with length shrinking to 0. This is impossible, so $\sqrt[3]{d}$ must be irrational. $\qquad\square$

Remark 2.4. In a similar way one can deal with higher roots: if $d \in \mathbf{Z}$ and $k \geq 2$ (with $d > 0$ if k is even) and d is not a kth power in $\mathbf{Z}$ then $\sqrt[k]{d}$ is irrational. Just assume $\sqrt[k]{d} = a/b$ is rational and use the $k \times k$ matrix and vector

$$A = \begin{pmatrix} 0 & d \\ I_{k-1} & 0 \end{pmatrix}, \quad v = \begin{pmatrix} a^{k-1} \\ a^{k-2}b \\ \vdots \\ b^{k-1} \end{pmatrix}.$$

3. Fermat's Last Theorem for $n = 4$

We will use descent to prove the exponent 4 case of Fermat's Last Theorem: the equation $a^4 + b^4 = c^4$ has no solution in positive integers. Fermat proved something more general, allowing a square and not just a fourth power on the right side.

Theorem 3.1 (Fermat). *There is no solution to the equation $x^4 + y^4 = z^2$ in positive integers. In particular, the equation $a^4 + b^4 = c^4$ has no solution in positive integers.*

Proof. We will use the parametrization of primitive Pythagorean triples, so let's recall that: a primitive solution to $a^2 + b^2 = c^2$ where a, b, and c are positive integers with b even is

$$a = k^2 - \ell^2, \quad b = 2k\ell, \quad c = k^2 + \ell^2,$$

where $k > \ell$, $(k, \ell) = 1$, and $k \not\equiv \ell \bmod 2$.[1]

Assume there is a solution to $x^4 + y^4 = z^2$ where x, y, and z are positive integers. If p is a common prime factor of x and y then $p^4 \mid z^2$, so $p^2 \mid z$. Then we can cancel the common factor of p^4 throughout and get a similar equation with smaller positive values of x, y, and z. Doing this enough times, we may suppose that $(x, y) = 1$. Then $(x, z) = 1$ and $(y, z) = 1$ too.

We will find a second positive integer solution (x', y', z') with $(x', y') = 1$ that is smaller in a suitable sense.

Since $x^4 + y^4 = z^2$ and $(x, y) = 1$, at least one of x and y is odd. They can't both be odd, since otherwise $z^2 \equiv 2 \bmod 4$, which has no solution. Without loss of generality, say x is odd and y is even. Then z is odd. Since $(x^2)^2 + (y^2)^2 = z^2$, (x^2, y^2, z) is a primitive

---

[1] For a proof of this, see https://kconrad.math.uconn.edu/blurbs/ugradnumthy/pythagtriple.pdf.

Pythagorean triple with y^2 the even term, so by the formula for primitive triples we can write

(3.1) $$x^2 = k^2 - \ell^2, \quad y^2 = 2k\ell, \quad z = k^2 + \ell^2,$$

where $k > \ell > 0$ and $(k, \ell) = 1$ (also $k \not\equiv \ell \bmod 2$, but we don't need this). The first equation in (3.1) says $x^2 + \ell^2 = k^2$. Since $(k, \ell) = 1$, (x, ℓ, k) is another primitive Pythagorean triple. Since x is odd, using the formula for primitive Pythagorean triples once again tells us

(3.2) $$x = a^2 - b^2, \quad \ell = 2ab, \quad k = a^2 + b^2,$$

where $a > b > 0$ and $(a, b) = 1$. The second equation in (3.1) now says

$$y^2 = 4(a^2 + b^2)ab.$$

Since y is even,

$$\left(\frac{y}{2}\right)^2 = (a^2 + b^2)ab.$$

Since $(a, b) = 1$, the three factors on the right are pairwise relatively prime. They are all positive, so their product being a square means each one is a square:

(3.3) $$a = x'^2, \quad b = y'^2, \quad a^2 + b^2 = z'^2,$$

where x', y', and z' can all be taken as positive. From $(a, b) = 1$, $(x', y') = 1$. The last equation in (3.3) can be rewrittten as $x'^4 + y'^4 = z'^2$, so we have another solution to our original equation with $(x', y') = 1$. Now we compare z' to z. Since

$$0 < z' \leq z'^2 = a^2 + b^2 = k \leq k^2 < z,$$

measuring the size of positive integer solutions (x, y, z) by the size of z leads to a contradiction by descent. $\square$

Remark 3.2. At the end of the proof a simple estimate showed $z > z'^2$. We can also get a formula for z in terms of x', y', and z' that explains this inequality. By (3.1), (3.2), and (3.3),

$$z = k^2 + \ell^2 = (a^2 + b^2)^2 + (2ab)^2 = z'^4 + 4x'^4 y'^4,$$

so in fact $z > z'^4$, not just $z > z'^2$ as we found before.

Let's write x and y in terms of x', y', and z' too. From (3.2) and (3.3),

$$x = a^2 - b^2 = x'^4 - y'^4$$

and $y^2 = 2k\ell = 2(a^2 + b^2)(2ab) = 4z'^2 (x'y')^2$, so

$$y = 2x'y'z'.$$

This formula for y shows x', y', and z' are all less than y, so $0 < \max(x', y', z') < y \leq \max(x, y, z)$. Using $\max(x, y, z)$ rather than z to measure the size of a solution (x, y, z) is another way to get a contradiction for Theorem 3.1 by descent.

Our proof of Theorem 3.1 used the parametric formula for primitive Pythagorean triples twice. For a proof that does not explicitly use this parametrization, see [2, pp. 55–56].

If we apply the descent technique for $x^4 + y^4 = z^2$ to $a^4 + b^4 = c^4$, with a *fourth* power on the right side, then the proof breaks down. The reason is that the descent step will not return another solution of $a^4 + b^4 = c^4$; the smaller c that comes out will only show up as a square, not a 4th power. So the extra generality of dealing with $x^4 + y^4 = z^2$ is essential for the descent to work as above.

Elementary number theory books that discuss Fermat's Last Theorem for exponent 4 introduce the equation $x^4 + y^4 = z^2$ out of the blue, like we did, as if it were the most natural thing in the world to look at this equation instead of $a^4 + b^4 = c^4$. Of course it isn't! The reason Fermat was thinking about $x^4 + y^4 = z^2$ was *not* in order to solve $a^4 + b^4 = c^4$ in integers but for an entirely different reason, and it was natural to consider the right side as z^2 for that other problem. See Appendix A for more details.

Now we present a number of corollaries to Theorem 3.1, concerning solvability of certain equations in integers or rationals. None of the proofs (which are mostly short) will involve descent. They are presented here simply to show Theorem 3.1 has uses other than Fermat's Last Theorem for exponent 4.

Corollary 3.3. *Each rational solution to $x^4 + y^4 = z^2$ has x or y equal to 0.*

Proof. Assume x and y are both nonzero. Then $z^2 > 0$, so $z \neq 0$ too. Write $x = a/d$, $y = b/d$, and $z = c/d$ with $a, b, c, d \in \mathbf{Z}$ and $d > 0$. Then a, b, and c are nonzero. Clearing the denominator in $x^4 + y^4 = z^2$, we have $a^4 + b^4 = (cd)^2$. Changing signs if necessary, a, b, and cd are positive. Then we have a contradiction with Theorem 3.1. $\square$

Corollary 3.4. *The only rational solutions to $y^2 = x^4 + 1$ are $(0, \pm 1)$.*

Proof. Use Corollary 3.3 to see $x = 0$. $\square$

Corollary 3.5. *The only rational solutions to $2y^2 = x^4 - 1$ are $(\pm 1, 0)$.*

Proof. Squaring both sides, $4y^4 = x^8 - 2x^4 + 1$. Add $4x^4$ to both sides and divide by 4 to get $y^4 + x^4 = ((x^4 + 1)/2)^2$. Since $x \neq 0$ in the original equation, we can divide by x^4 to get $(y/x)^4 + 1 = ((x^4 + 1)/2x^2)^2$. By Corollary 3.4, $y/x = 0$, so $y = 0$ and therefore $x = \pm 1$. $\square$

Corollary 3.6. *The integral solutions of $x^4 - y^4 = 2z^2$ are $(x, \pm x, 0)$ for $x \in \mathbf{Z}$.*

Proof. If $y = 0$ then $x = z = 0$ since $\sqrt{2}$ is irrational. If $y \neq 0$, then divide by y^4 to get $(x/y)^4 - 1 = 2(z/y^2)^2$. By Corollary 3.5, $z/y^2 = 0$, so $z = 0$ and therefore $y = \pm x$. $\square$

Corollary 3.7. *The only rational solutions to $y^2 = x^3 - 4x$ are $(0, 0)$, $(\pm 2, 0)$.*

Proof. There is a bijection between solutions of $y^2 = x^3 - 4x$ with $x \neq 0$ and solutions to $v^2 = u^4 + 1$ by

$$(x, y) \mapsto \left(\frac{y}{2x}, \frac{y^2 + 8x}{4x^2} \right), \quad (u, v) \mapsto \left(\frac{2}{v - u^2}, \frac{4u}{v - u^2} \right).$$

Since each rational solution to $v^2 = u^4 + 1$ has $u = 0$, each rational solution to $y^2 = x^3 - 4x$ has $y = 0$, so $x = 0$ or $x = \pm 2$. $\square$

Corollary 3.8. *The only rational solution to $y^2 = x^3 + x$ is $(0, 0)$.*

Proof. Writing the equation as $y^2 = x(x^2 + 1)$, we see $x = 0$ if and only if $y = 0$. Assume there is a rational solution other than $(0, 0)$ so $x \neq 0$ and $y \neq 0$. From the equation, x must be positive.

Write x and y in reduced form as $x = a/b$ and $y = c/d$ where b and d are positive. Clearing denominators in $(c/d)^2 = (a/b)^3 + a/b$, we get

$$b^3 c^2 = d^2(a^3 + ab^2).$$

Therefore $d^2 \mid b^3 c^2$. Since $(c, d) = 1$, $d^2 \mid b^3$. Also $b^3 \mid d^2(a^3 + ab^2)$. Since $(a, b) = 1$, b^3 is relatively prime to $a^3 + ab^2$, so $b^3 \mid d^2$. Thus $b^3 = d^2$, so by unique factorization $b = t^2$ and $d = t^3$ for some positive integer t. Then $(a, t) = 1$ and $(c, t) = 1$.

In the equation $y^2 = x^3 + x$ with $x = a/t^2$ and $y = c/t^3$, we get $c^2 = a^3 + t^4 a = a(a^2 + t^4)$ after clearing the denominator. Since $(a, t) = 1$, a and $a^2 + t^4$ are relatively prime and positive. Their product is a square, so each factor is a square:

$$a = u^2, \quad a^2 + t^4 = v^2.$$

Thus $u^4 + t^4 = v^2$. By Theorem 3.1, u or t is 0. Since $t \neq 0$, $u = 0$ so $x = 0$ and then $y = 0$. $\qquad\square$

Remark 3.9. Conversely, Corollary 3.8 implies Theorem 3.1. If $x^4 + y^4 = z^2$ in positive integers then multiplying through by x^2/y^6 gives us $(x/y)^6 + (x/y)^2 = (xz/y^3)^2$, so $Y^2 = X^3 + X$ for $X = (x/y)^2$ and $Y = xz/y^3$. Since X is a nonzero rational number, we have a contradiction with Corollary 3.8.

Here is another theorem about fourth powers and squares proved by Fermat using descent.

Theorem 3.10 (Fermat). *There is no solution to $x^4 - y^4 = z^2$ in positive integers.*

Proof. We will argue by descent in a similar style to the proof of Theorem 3.1. In particular, we will use the formula for primitive Pythagorean triples twice. Since now we have $z^2 + y^4 = x^4$ while in Theorem 3.1 we had $x^4 + y^4 = z^2$, the roles of x^2 and z basically get interchanged. For example, we will use descent on x^2 (or equivalently, on x) rather than on z as we did in Theorem 3.1.

Assume $x^4 - y^4 = z^2$ with x, y, and z in $\mathbf{Z}^+$. There must be a solution with x, y, and z pairwise relatively prime (see the start of the proof of Theorem 3.1; the same argument there applies here), so we suppose this is the case. Since $x^4 - y^4 > 0$, $x > y$.

There are two cases to consider: z odd and z even.

Case 1: z is odd. Since $z^2 + y^4 = x^4$ and z is odd, y must be *even*. (Otherwise $z^2 + y^4 \equiv 1 + 1 \equiv 2 \bmod 4$, but 2 is not a 4th power modulo 4.) Since $(x, y) = 1$, (z, y^2, x^2) is a primitive Pythagorean triple with y^2 the even term, so the formula for primitive Pythagorean triples says

$$(3.4) \qquad z = k^2 - \ell^2, \quad y^2 = 2k\ell, \quad x^2 = k^2 + \ell^2,$$

where $k > \ell > 0$, $(k, \ell) = 1$, and $k \not\equiv \ell \bmod 2$. The third equation in (3.4) says (k, ℓ, x) is a Pythagorean triple. Since $(k, \ell) = 1$, this triple is primitive. One of k or ℓ is odd and the other is even. If k is odd, the formula for primitive Pythagorean triples says

$$(3.5) \qquad k = a^2 - b^2, \quad \ell = 2ab, \quad x = a^2 + b^2,$$

where $a > b > 0$, $(a, b) = 1$, and $a \not\equiv b \bmod 2$. If ℓ is odd the formula says

$$(3.6) \qquad \ell = a^2 - b^2, \quad k = 2ab, \quad x = a^2 + b^2,$$

where $a > b > 0$, $(a, b) = 1$, and $a \not\equiv b \bmod 2$. Using whichever of (3.5) or (3.6) is correct (depending on the parity of k and ℓ), the second equation in (3.4) becomes

$$(3.7) \qquad y^2 = 4(a^2 - b^2)ab.$$

Since y is even, we can divide by 4 (in $\mathbf{Z}$):

$$\left(\frac{y}{2}\right)^2 = (a^2 - b^2)ab.$$

Since $(a, b) = 1$, the three factors on the right are pairwise relatively prime. They are all positive, so their product being a square means each one is a square:

$$(3.8) \qquad a = x'^2, \quad b = y'^2, \quad a^2 - b^2 = z'^2,$$

where x', y', and z' can all be taken as positive. From $(a, b) = 1$, $(x', y') = 1$. The last equation in (3.8) can be rewrittten as $x'^4 - y'^4 = z'^2$, so we have another solution to our original equation. Moreover, $z'^2 = a^2 - b^2$ is odd, so our new solution again has an odd square on the right and we are still in Case 1. Now we compare x' to x:

$$0 < x' \leq x'^2 = a < a^2 + b^2 = x.$$

Since $x' < x$, by descent we have a contradiction.

Case 2: z is even. (This has no analogue in the proof of Theorem 3.1.)

Since $y^4 + z^2 = x^4$, we have a primitive Pythagorean triple (y^2, z, x^2) with even z. Thus

$$y^2 = m^2 - n^2, \quad z = 2mn, \quad x^2 = m^2 + n^2,$$

where m and n are positive and $(m, n) = 1$. Multiplying the first and third equations,

$$(xy)^2 = m^4 - n^4,$$

with xy odd. This expresses a square as the difference of two fourth powers, with the square being odd, so by Case 1 we have a contradiction. $\square$

Remark 3.11. In Case 1 we can solve for x, y, and z in terms of x', y', and z'. From (3.5) or (3.6), $x = a^2 + b^2$. This becomes, by (3.8),

$$x = x'^4 + y'^4.$$

From (3.7) and (3.8), $y^2 = 4(a^2 - b^2)ab = 4z'^2(x'^2 y'^2) = (2x'y'z')^2$, so

$$y = 2x'y'z'.$$

Lastly, by (3.4), (3.5) or (3.6), and (3.8),

$$z = k^2 - \ell^2 = \pm((a^2 - b^2)^2 - (2ab)^2) = \pm(z'^4 - 4x'^4 y'^4),$$

so $z = |z'^4 - 4x'^4 y'^4|$. From the formula $y = 2x'y'z'$ we get $0 < \max(x', y', z') < y \leq \max(x, y, z)$, so using $\max(x, y, z)$ rather than x as a measure of the size of a positive integer solution is another way of reaching a contradiction in Case 1 by descent. This parallels Remark 3.2.

Theorems 3.1 and 3.10 together lead to the following two results.

Corollary 3.12. *There is no Pythagorean triple in which two of the terms are squares.*

Proof. Such a triple would give a solution in positive integers to either $x^4 + y^4 = z^2$ (the two legs are squares) or $x^4 = y^4 + z^2$ (a leg and hypotenuse are squares), but such solutions do not exist by Theorems 3.1 and 3.10. $\square$

Many primitive Pythagorean triples have just one term equal to a square. See Table 1.

Corollary 3.13. *The only $m \in \mathbf{Z}^+$ such that $1 + 2 + 3 + \cdots + m$ is a fourth power is 1.*

Proof. Since $1 + 2 + 3 + \cdots + m = m(m + 1)/2$, we are trying to solve $m(m + 1)/2 = n^4$. Clearing the denominator, $m(m + 1) = 2n^4$. Since m and $m + 1$ are relatively prime,

$$\{m, m + 1\} = \{x^4, 2y^4\}$$

a	b	c
3	**4**	5
7	24	**25**
9	40	41
16	63	65
17	**144**	145
225	272	353
161	240	**289**

TABLE 1. Pythagorean triples with a square term

for some positive integers x and y, which must be relatively prime. Therefore $x^4 - 2y^4 = \pm 1$. Rewrite as $y^4 = (x^4 \mp 1)/2$ and square both sides followed by a little algebra to get

$$(3.9) \qquad y^8 \pm x^4 = \left(\frac{x^4 \pm 1}{2}\right)^2.$$

The right side is an integer since x is odd from $x^4 = 2y^4 \pm 1$. Equation (3.9) expresses a sum or difference of positive fourth powers as a perfect square. Using Theorems 3.1 and 3.10, we see this is impossible for a sum, and for a difference the square must be 0, so $x = 1$ (since $x > 0$). Therefore $m = 1$. $\qquad\square$

Now we present consequences of Theorem 3.10 alone, paralleling the consequences of Theorem 3.1.

Corollary 3.14. *Each rational solution to $x^4 - y^4 = z^2$ has y or z equal to 0.*

Proof. Similar to Corollary 3.3. $\qquad\square$

Corollary 3.15. *The only rational solutions to $y^2 = x^4 - 1$ are $(\pm 1, 0)$.*

Proof. Similar to Corollary 3.4. $\qquad\square$

Corollary 3.16. *The only rational solutions to $2y^2 = x^4 + 1$ are $(\pm 1, \pm 1)$.*

Proof. Squaring both sides, $4y^4 = x^8 + 2x^4 + 1$. Subtract $4x^4$ from both sides and divide by 4 to get $y^4 - x^4 = ((x^4 - 1)/2)^2$. Since $x \neq 0$ in the original equation, we can divide by x^4 to get $(y/x)^4 - 1 = ((x^4 - 1)/2x^2)^2$. By Corollary 3.15, $(x^4 - 1)/2x^2 = 0$, so $x = \pm 1$ and therefore $y = \pm 1$. $\qquad\square$

Corollary 3.17. *The integral solutions of $x^4 + y^4 = 2z^2$ are $(x, \pm x, \pm x^2)$.*

Proof. If $y = 0$ then $x = z = 0$ since $\sqrt{2}$ is irrational. If $y \neq 0$, then divide by y^4 to get $(x/y)^4 + 1 = 2(z/y^2)^2$. By Corollary 3.16, $x/y = \pm 1$ and $z/y^2 = \pm 1$, so $y = \pm x$ and $z = \pm y^2 = \pm x^2$. $\qquad\square$

Corollary 3.18. *The only rational solutions to $y^2 = x^3 + 4x$ are $(0,0)$ and $(2, \pm 4)$.*

Proof. A bijection between solutions of $y^2 = x^3 + 4x$ with $x \neq 0$ and solutions of $v^2 = u^4 - 1$ is given by

$$(x, y) \mapsto \left(\frac{y}{2x}, \frac{y^2 - 8x}{4x^2}\right), \quad (u, v) \mapsto \left(\frac{2}{u^2 - v}, \frac{4u}{u^2 - v}\right).$$

Each rational solution to $v^2 = u^4 - 1$ has $v = 0$, so each rational solution to $y^2 = x^3 + 4x$ has $x = 0$ or $y^2 = 8x$. The second case implies $x^3 = 4x$, so $x = \pm 2$. Only $x = 2$ leads to a solution, $(x, y) = (2, \pm 4)$. $\qquad\square$

Corollary 3.19. *The only rational solutions to $y^2 = x^3 - x$ are $(0,0)$ and $(\pm 1, 0)$.*

Proof. This stands in relation to $x^4 - y^4 = z^2$ in the same way that $y^2 = x^3 + x$ does to $x^4 + y^4 = z^2$ in Corollary 3.8. Details are left to the reader as an exercise. $\square$

Remark 3.20. In the other direction, Corollary 3.19 implies Theorem 3.10. The argument is like that in Remark 3.9.

4. Equation with no rational function solution

There are many rational solutions to $x^2 + y^2 = 1$, like $(x, y) = (3/5, 4/5)$ or $(5/13, 12/13)$. These solutions are nearly all described by a single parametric formula: $x = 2t/(1+t^2)$ and $y = (1 - t^2)/(1 + t^2)$ for some $t \in \mathbf{Q}$. Try $t = 1/3$, $1/4$, and $1/5$. Algebraically, we have an identity of rational functions

$$\left(\frac{2t}{1+t^2}\right)^2 + \left(\frac{1-t^2}{1+t^2}\right)^2 = 1$$

and specializing t to rational numbers gives rational solutions to $x^2 + y^2 = 1$.[2]

However, it is not true that rational solutions to other equations can always be fit into a parametric formula using rational functions. For example, the equation $y^2 = x^3 - 2$ has infinitely many rational solutions (two of them are $(x, y) = (3, 5)$ and $(129/100, 383/1000)$), but there are no rational functions $f(t)$ and $g(t)$ in $\mathbf{Q}(t)$ such that $g(t)^2 = f(t)^3 - 2$ other than constant functions, which "parametrize" only one solution (so really do not provide a parametric formula at all). To prove negative results like this we will use descent on the degree in a hypothetical nonconstant solution.

Several times we will need the following lemma.

Lemma 4.1. *If $f(t)$ and $g(t)$ in $\mathbf{C}[t]$ are relatively prime and $fg = \square$ in $\mathbf{C}[t]$ then $f = \square$ and $g = \square$ in $\mathbf{C}[t]$.*

Proof. By unique factorization, the multiplicity of each irreducible factor of fg is even. Since f and g are relatively prime, it follows that the multiplicity of each irreducible factor of f is even and likewise for g. Therefore f and g are squares up to a nonzero scaling factor. Every nonzero complex number is the square of a complex number, so f and g are squares. $\square$

This lemma easily extends, by induction on the number of terms, to a product of any finite number of polynomials in $\mathbf{C}[t]$ that are pairwise relatively prime.

Theorem 4.2. *For distinct complex numbers r, r', r'', each solution to the equation $y^2 = (x - r)(x - r')(x - r'')$ in rational functions $x = f(t)$ and $y = g(t)$ in $\mathbf{C}(t)$ is a constant solution: $f(t)$ and $g(t)$ are in $\mathbf{C}$.*

The example $y^2 = x^3 - 2$ that we discussed above is the special case where r, r', and r'' are the three cube roots of 2.

Proof. Our argument is adapted from [3, pp. 75–76].

Assume there is a solution in rational functions: $x = p_1(t)/p_2(t)$ and $y = q_1(t)/q_2(t)$ where $p_1(t)$, $p_2(t)$, $q_1(t)$, and $q_2(t)$ are polynomials in $\mathbf{C}[t]$. Without loss of generality we can assume $(p_1, p_2) = 1$ and $(q_1, q_2) = 1$ in $\mathbf{C}[t]$.

---

[2]This method produces all the rational solutions to $x^2 + y^2 = 1$ except for $(0, -1)$, but we will not show that here.

Substituting the formulas for x and y into the equation $y^2 = (x - r)(x - r')(x - r'')$, we have

$$\frac{q_1^2}{q_2^2} = \left(\frac{p_1}{p_2} - r\right)\left(\frac{p_1}{p_2} - r'\right)\left(\frac{p_1}{p_2} - r''\right) = \frac{(p_1 - rp_2)(p_1 - r'p_2)(p_1 - r''p_2)}{p_2^3},$$

so after clearing denominators

(4.1) $$p_2^3 q_1^2 = (p_1 - rp_2)(p_1 - r'p_2)(p_1 - r''p_2)q_2^2.$$

Since $(p_1, p_2) = 1$, the factors $p_1 - rp_2$, $p_1 - r'p_2$, and $p_1 - r''p_2$ are relatively prime to p_2 (why?). Thus by (4.1), $p_2^3 \mid q_2^2$. Since $(q_1, q_2) = 1$, $q_2^2 \mid p_2^3$, so $p_2^3 = cq_2^2$ for some nonzero $c \in \mathbf{C}$. Since c is a square in $\mathbf{C}$, p_2^3 is a square in $\mathbf{C}[t]$. That implies p_2 is a square by unique factorization in $\mathbf{C}[t]$. Write $p_2 = f_2^2$.

Substituting cq_2^2 for p_2^3 in (4.1) and cancelling q_2^2 from both sides,

$$cq_1^2 = (p_1 - rp_2)(p_1 - r'p_2)(p_1 - r''p_2).$$

The factors on the right side are pairwise relatively prime since $(p_1, p_2) = 1$ (why?) and the numbers r, r', and r'' are distinct, so by an extension of Lemma 4.1 to a product of three terms, $p_1 - rp_2$, $p_1 - r'p_2$, and $p_1 - r''p_2$ are all squares in $\mathbf{C}[t]$. Since $p_2 = f_2^2$,

$$p_1 - rf_2^2 = \square, \quad p_1 - r'f_2^2 = \square, \quad p_1 - r''f_2^2 = \square.$$

Writing the first equation as $p_1 - rf_2^2 = f_1^2$, the second and third equations become

(4.2) $$f_1^2 - (r' - r)f_2^2 = \square, \quad f_1^2 - (r'' - r)f_2^2 = \square,$$

where $r' - r$ and $r'' - r$ are nonzero and distinct. We want to show f_1 and f_2 are constant.

Now we set up our descent statement, based on (4.2): we will show for all distinct a and b in $\mathbf{C}$ that relatively prime polynomials g_1 and g_2 in $\mathbf{C}[t]$ that satisfy

(4.3) $$g_1^2 - ag_2^2 = \square, \quad g_1^2 - bg_2^2 = \square$$

must both be constant. Note (4.2) is a special case of this.

Assume for some a and b that there is a solution (g_1, g_2) to (4.3) where g_1 or g_2 is not constant. In (4.3), write $a = c^2$ and $b = d^2$ with c and d in $\mathbf{C}$. Since $a \neq b$, $c \neq \pm d$. We can rewrite (4.3) as

(4.4) $$(g_1 + cg_2)(g_1 - cg_2) = \square, \quad (g_1 + dg_2)(g_1 - dg_2) = \square.$$

A common factor of $g_1 + cg_2$ and $g_1 - cg_2$ is a factor of both g_1 and g_2 (why?), so it is constant since $(g_1, g_2) = 1$. Therefore the factors on the left side of the first equation in (4.4) are relatively prime, and the product of the factors is a square, so by Lemma 4.1

(4.5) $$g_1 + cg_2 = h_1^2, \quad g_1 - cg_2 = h_2^2,$$

where h_1 and h_2 are relatively prime. Adding and subtracting the equations in (4.5), $g_1 = (h_1^2 + h_2^2)/2$ and $g_2 = (h_1^2 - h_2^2)/(2c)$. Since g_1 or g_2 is not constant, h_1 or h_2 is not constant.

Arguing in a similar way with the second equation in (4.4),

(4.6) $$g_1 + dg_2 = \square, \quad g_1 - dg_2 = \square.$$

Substituting the formulas for g_1 and g_2 in terms of h_1 and h_2 into (4.6),

$$\frac{1}{2}\left(1 + \frac{d}{c}\right)h_1^2 + \frac{1}{2}\left(1 - \frac{d}{c}\right)h_2^2 = \square, \quad \frac{1}{2}\left(1 - \frac{d}{c}\right)h_1^2 + \frac{1}{2}\left(1 + \frac{d}{c}\right)h_2^2 = \square.$$

The numbers $1 \pm d/c$ are nonzero since $c \neq \pm d$, so we can divide by the coefficient of h_1^2:

$$h_1^2 + \frac{1 - d/c}{1 + d/c}h_2^2 = \square, \quad h_1^2 + \frac{1 + d/c}{1 - d/c}h_2^2 = \square.$$

Set $a' = -(1 - d/c)/(1 + d/c)$ and $b' = -(1 + d/c)/(1 - d/c) = 1/a'$. Both a' and b' are nonzero, $a' \neq b'$, and

$$(4.7) \qquad\qquad h_1^2 - a'h_2^2 = \square, \quad h_1^2 - b'h_2^2 = \square.$$

From (4.5), $2 \deg h_1 \leq \max(\deg g_1, \deg g_2)$ and $2 \deg h_2 \leq \max(\deg g_1, \deg g_2)$. Therefore

$$(4.8) \qquad\qquad 0 < \max(\deg h_1, \deg h_2) \leq \frac{1}{2}\max(\deg g_1, \deg g_2).$$

We can now repeat the argument leading from (4.3) to (4.7) with h_1, h_2, a', b' in place of g_1, g_2, a, b. Each repetition leads to a new version of (4.8) where the maximum degree in the new solution is positive and smaller than the maximum degree in the previous solution. By descent this leads to a contradiction (positive degrees can't strictly drop forever), so each solution to (4.3) in relatively prime polynomials g_1 and g_2 must be a constant solution. What this tells us in (4.2) is that f_1 and f_2 are both constant, so $p_1 = f_1^2 + rf_2^2$ and $p_2 = f_2^2$ are constant. That makes $x = p_1/p_2$ constant, so the equation $y^2 = (x-r)(x-r')(x-r'')$ implies that $y^2 = (q_1/q_2)^2$ is constant. Therefore the rational function $y = q_1/q_2$ is constant. $\square$

Corollary 4.3. *Let $F(x)$ be a polynomial with coefficients in $\mathbf{C}$ of degree 3 or 4 that has distinct roots. If x and y in $\mathbf{C}(t)$ satisfy $y^2 = F(x)$ then $x \in \mathbf{C}$ and $y \in \mathbf{C}$.*

Proof. First suppose $F(x)$ has degree 3. In factored form $F(x) = c(x - r)(x - r')(x - r'')$ where the roots r, r', and r'' are distinct and $c \neq 0$.[3] In $\mathbf{C}$, c is a square, so the equation

$$(4.9) \qquad\qquad y^2 = c(x - r)(x - r')(x - r'')$$

can be scaled to $Y^2 = (x - r)(x - r')(x - r'')$ where $Y = y/\sqrt{c}$. If there is a solution to (4.9) where $x, y \in \mathbf{C}(t)$ then $Y \in \mathbf{C}(t)$. By Theorem 4.2, $x \in \mathbf{C}$ and $Y \in \mathbf{C}$, so $y \in \mathbf{C}$.

Next suppose $F(x)$ has degree 4 with distinct roots r, r', r'', and r'''. In factored form

$$F(x) = c(x - r)(x - r')(x - r'')(x - r'''),$$

where $c \neq 0$. Suppose we can solve

$$(4.10) \qquad\qquad y^2 = c(x - r)(x - r')(x - r'')(x - r''')$$

where $x, y \in \mathbf{C}(t)$. The roots of F are distinct, so without loss of generality $x \neq r'''$. Dividing through (4.10) by $(x - r''')^4$,

$$\begin{aligned}
\frac{y^2}{(x - r''')^4} &= c\left(\frac{x - r}{x - r'''}\right)\left(\frac{x - r'}{x - r'''}\right)\left(\frac{x - r''}{x - r'''}\right) \\
&= c\left(1 + \frac{r''' - r}{x - r'''}\right)\left(1 + \frac{r'' - r}{x - r'''}\right)\left(1 + \frac{r' - r}{x - r'''}\right).
\end{aligned}$$

Set $X = 1/(x - r''')$ and $Y = y/(x - r''')^2$, so X and Y are in $\mathbf{C}(t)$. Then

$$Y^2 = c(1 + (r''' - r)X)(1 + (r''' - r')X)(1 + (r''' - r'')X),$$

where the right side is cubic in X with distinct roots. By the cubic case we already discussed, $X \in \mathbf{C}$ and $Y \in \mathbf{C}$, so $x \in \mathbf{C}$ and $y \in \mathbf{C}$. $\square$

---

[3]The difference between this case and Theorem 4.2 is that we now allow a leading coefficient besides 1.

5. Sum of two squares

We have used descent to prove "negative" theorems, which say certain equations have no (or very few) solutions of a certain type. Fermat's initial applications of descent had this flavor.[4] He later found descent could be used to prove "positive" theorems like the following.

Theorem 5.1. *For a prime p, if $-1 \equiv \square \bmod p$ then we can write p as a sum of two squares: $p = x^2 + y^2$ for some x and y in $\mathbf{Z}$.*

Proof. Write $-1 \equiv s^2 \bmod p$. Since $s^2 \equiv (-s)^2 \bmod p$ we can take $1 \le s \le p/2$, so $s^2 + 1 = pd$ with $d \in \mathbf{Z}$. Since $s^2 + 1 \le p^2/4 + 1 < p^2$, we have $0 < d < p$. So a multiple of p is a sum of two squares and the multiplying factor d is less than p. If $d = 1$ we are done.

If $d > 1$, we will show that pd' is a sum of two squares for some integer d' such that $0 < d' < d$. Then repeating the argument, by descent eventually we must reach $d = 1$, so $p \cdot 1 = p$ is a sum of two squares.

To make our descent step recursive, generalize the specific equation $pd = s^2 + 1$ to

$$pk = x^2 + y^2$$

where $0 < k < p$. (For $k = d$ this holds with $x = s$ and $y = 1$, so such an equation does exist for some choice of k.) If $k > 1$, we want to produce another equation of this form where k is smaller. Reduce x and y modulo k: $x \equiv r \bmod k$ and $y \equiv r' \bmod k$, where $|r|, |r'| \le k/2$. Squaring and adding,

$$r^2 + r'^2 \equiv x^2 + y^2 \equiv 0 \bmod k,$$

so

$$r^2 + r'^2 = k\ell$$

for integer $\ell \ge 0$. Let's show r or r' is not 0, so $\ell > 0$. If $r = 0$ and $r' = 0$ then $k \mid x$ and $k \mid y$, so k^2 divides $x^2 + y^2 = kp$, so $k \mid p$, but this is not true since $0 < k < p$ and p is prime. Now let's bound ℓ from above. From the bounds on $|r|$ and $|r'|$,

$$k\ell = r^2 + r'^2 \le \frac{k^2}{4} + \frac{k^2}{4} = \frac{k^2}{2} < k^2,$$

so $0 < \ell < k$. We will show $p\ell$ is a sum of two squares.

Since

$$(pk)(k\ell) = (x^2 + y^2)(r^2 + r'^2) = (x^2 + y^2)(r'^2 + r^2) = (xr' - yr)^2 + (xr + yr')^2.$$

Modulo k, $xr' - yr \equiv xy - yx \equiv 0$ and $xr + yr' \equiv r^2 + r'^2 \equiv 0$. Therefore $xr' - yr$ and $xr + yr'$ are multiples of k, so we can write

$$pk^2\ell = (ka)^2 + (kb)^2$$

for some positive integers a and b. Dividing by k^2, $p\ell = a^2 + b^2$. This completes our descent step, since $0 < \ell < k$. $\qquad \square$

Remark 5.2. Fermat's own proof by descent of Theorem 5.1 was based on counterexamples: assuming an odd prime p with $-1 \equiv \square \bmod p$ is not a sum of two squares, Fermat wrote to Huygens (without giving details) that he could show there is a smaller prime with the same property, so by descent 5 is not a sum of two squares, but it is: a contradiction [12, p. 67].

---

[4]Almost none of Fermat's proofs are known in detail, but he did include in letters the statements of some propositions that he said he had established with descent.

6. The four-square theorem

A famous theorem of Lagrange [10], called his four-square theorem, says every positive integer is a sum of four squares in $\mathbf{Z}$. (By dropping any term equal to 0, we get a sum of *at most* four squares in $\mathbf{Z}^+$).

Example 6.1. Some numbers, like 2, 7, 15, and 23, can be written as a sum of four squares in only one way up to the order of the terms, while other numbers like 65 have multiple different-looking four-square representations:

$$65 = 1^2 + 8^2 + 0^2 + 0^2 = 4^2 + 7^2 + 0^2 + 0^2 = 2^2 + 5^2 + 6^2 + 0^2 = 2^2 + 3^2 + 4^2 + 6^2.$$

Euler tried to prove the four-square theorem for 40 years, ever since he first read about it in 1730 in the work of Fermat [12, p. 173].[5] Euler's efforts led to results such as the next two lemmas, which get used in proofs of the four-square theorem.

Lemma 6.2 (Euler, 1748)**.** *Sums of four squares in* $\mathbf{Z}$ *are closed under multiplication:*

$$(x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) = z_1^2 + z_2^2 + z_3^2 + z_4^2,$$

where

$$z_1 = x_1 y_1 + x_2 y_2 + x_3 y_3 + x_4 y_4,$$
$$z_2 = x_1 y_2 - x_2 y_1 - x_3 y_4 + x_4 y_3,$$
$$z_3 = x_1 y_3 + x_2 y_4 - x_3 y_1 - x_4 y_2,$$
$$z_4 = x_1 y_4 - x_2 y_3 + x_3 y_2 - x_4 y_1.$$

Proof. It is left to the reader to expand both sides to check they are equal.   □

Remark 6.3. Other choices of signs in the z_i's are possible. Above we used the first one Euler wrote down, in a letter to Goldbach [5, p. 452]. In a later article (see the proof of [6, Theorem 19]), Euler said multiple choices of signs can be used and gave the one above as well as another with the same z_1 and

$$z_2 = x_1 y_2 - x_2 y_1 + x_3 y_4 - x_4 y_3,$$
$$z_3 = x_1 y_3 - x_2 y_4 - x_3 y_1 + x_4 y_2,$$
$$z_4 = x_1 y_4 + x_2 y_3 - x_3 y_2 - x_4 y_1.$$

A third four-square product identity was discovered by Hamilton in 1848 in his work on quaternions, and it uses

$$z_1 = x_1 y_1 - x_2 y_2 - x_3 y_3 - x_4 y_4,$$
$$z_2 = x_1 y_2 + x_2 y_1 + x_3 y_4 - x_4 y_3,$$
$$z_3 = x_1 y_3 - x_2 y_4 + x_3 y_1 + x_4 y_2,$$
$$z_4 = x_1 y_4 + x_2 y_3 - x_3 y_2 + x_4 y_1.$$

---

[5]Fermat made a more general assertion, called the polygonal number theorem: each positive integer is a sum of at most 3 triangular numbers, at most 4 squares, and more generally at most k k-gonal numbers. Pictures of triangular, square, and pentagonal numbers are in Section 3 of https://kconrad.math.uconn.edu/blurbs/ugradnumthy/pelleqn1.pdf. The case of triangular numbers was settled by Gauss in 1796 and the general polygonal number theorem was first proved by Cauchy in 1813.

Lemma 6.4 (Euler, 1760). *Let p be prime. The congruence*

$$x^2 + y^2 + 1 \equiv 0 \bmod p$$

has at least one solution.

Proof. This is obvious when $p = 2$, so let $p > 2$. Rewrite the congruence as

$$x^2 \equiv -1 - y^2 \bmod p.$$

Let $A = \{x^2 \bmod p : x \in \mathbf{Z}/(p)\}$, so $|A|$ counts the number of possible values of the left side as x varies and let $B = \{-1 - y^2 \bmod p : y \in \mathbf{Z}/(p)\}$, so $|B|$ counts the number of possible values of the right side as y varies. We want to show $A \cap B \neq \emptyset$.

There are $(p-1)/2$ nonzero squares mod p (since $p > 2$), so counting 0 makes the number of squares mod p equal to $1 + (p-1)/2 = (p+1)/2$. Thus $|A| = (p+1)/2$. Likewise, $|B| = |\{-1 - y^2 \bmod p : y \in \mathbf{Z}/(p)\}| = |\{y^2 \bmod p : y \in \mathbf{Z}/(p)\}| = (p+1)/2$. Since $|A| + |B| = (p+1)/2 + (p+1)/2 = p + 1 > |\mathbf{Z}/(p)|$, A and B can't be disjoint, so A and B contain a common value: there are some x_0 and y_0 in $\mathbf{Z}/(p)$ such that $x_0^2 \equiv -1 - y_0^2 \bmod p$, so $x_0^2 + y_0^2 + 1 \equiv 0 \bmod p$. $\qquad\square$

Remark 6.5. The way Euler stated Lemma 6.4 [6, Theorem 18] was that the congruence $x^2 + y^2 + z^2 \equiv 0 \bmod p$ has a solution where some term is nonzero mod p. Dividing all the terms by that nonzero square converts the congruence into the one in Lemma 6.4, where some term is 1.

Using these two lemmas, Euler showed each positive integer is a sum of four squares in $\mathbf{Q}$ (see Appendix C), but doing this in $\mathbf{Z}$ was frustratingly out of reach until he saw Lagrange's descent proof of the four-square theorem. Euler then found a descent [7, Theorem 4] that is simpler than Lagrange's, and it in essence is what we use next.

Theorem 6.6 (Lagrange, 1770). *For each $n \in \mathbf{Z}^+$, we can write $n = a^2 + b^2 + c^2 + d^2$ for some $a, b, c, d \in \mathbf{Z}$.*

Proof. The result is obvious for 1. Since each integer bigger than 1 is a product of primes, and sums of four squares are closed under multiplication by Lemma 6.2, it suffices to show each prime number p is a sum of four squares.

Step 1: There is $k \in \mathbf{Z}^+$ such that $k < p$ and $pk = x^2 + y^2 + 1$ for some x and y in $\mathbf{Z}$.

To show this, we start with Lemma 6.4: $x^2 + y^2 + 1 \equiv 0 \bmod p$ for some integers x and y. Adjusting x and y modulo p, we can assume they are in $\{0, 1, \ldots, p-1\}$. Since $x^2 \equiv (-x)^2 \bmod p$, we can also assume $0 \leq x, y \leq p/2$. The congruence mod p tells us $x^2 + y^2 + 1 = pk$ where $k \geq 1$. By the bounds on x and y,

$$x^2 + y^2 + 1 \leq \left(\frac{p}{2}\right)^2 + \left(\frac{p}{2}\right)^2 + 1 = \frac{p^2}{2} + 1 < p^2,$$

so $pk < p^2$. Thus $k < p$.

Step 2: Let m be a positive integer less than p such that pm is a sum of four squares. If $m > 1$ then there is a positive integer $n < m$ such that pn is a sum of four squares.

By hypothesis, we can write $pm = a^2 + b^2 + c^2 + d^2$ with $1 < m < p$. We are going to look at $a, b, c, d \bmod m$. They are not all 0 mod m, as otherwise $a^2 + b^2 + c^2 + d^2$ would be divisible by m^2, so $m^2 \mid pm$, making $m \mid p$, but $1 < m < p$ and p is prime.

Using remainders under division by m that are smallest in absolute value (this allows negative remainders), $\mathbf{Z}/(m)$ is represented by integers in the range $\{x : |x| \leq m/2\}$,

so we can set $a \equiv a' \bmod m$, $b \equiv b' \bmod m$, $c \equiv c' \bmod m$, and $d \equiv d' \bmod m$, where $|a'|, |b'|, |c'|, |d'| \leq m/2$ and at least one of a', b', c', and d' is not 0. Then

$$a'^2 + b'^2 + c'^2 + d'^2 \equiv a^2 + b^2 + c^2 + d^2 \equiv 0 \bmod m$$

and the left side is positive, so $a'^2 + b'^2 + c'^2 + d'^2 = mn$ where $n \geq 1$. Using upper bounds,

$$(6.1) \qquad mn = a'^2 + b'^2 + c'^2 + d'^2 \leq \left(\frac{m}{2}\right)^2 + \left(\frac{m}{2}\right)^2 + \left(\frac{m}{2}\right)^2 + \left(\frac{m}{2}\right)^2 = m^2,$$

so $n \leq m$.

Let's show $n = m$ is impossible, so $n < m$. If $n = m$ in (6.1) then the inequality there is an equality, so $|a'| = |b'| = |c'| = |d'| = m/2$. (In particular, this requires m to be even.) Then, since $-m/2 \equiv m/2 \bmod m$, we have $a, b, c, d \equiv m/2 \bmod m$, so[6]

$$pm = a^2 + b^2 + c^2 + d^2 \equiv 4\left(\frac{m}{2}\right)^2 = m^2 \equiv 0 \bmod m^2,$$

so $p \equiv 0 \bmod m$, which is impossible since p is prime and $1 < m < p$.

We have $pm = a^2 + b^2 + c^2 + d^2$ and $mn = a'^2 + b'^2 + c'^2 + d'^2$ where $1 < m < p$ and $1 \leq n < m$. Multiply these equations and use Lemma 6.2:

$$pm^2 n = (a^2 + b^2 + c^2 + d^2)(a'^2 + b'^2 + c'^2 + d'^2) = A^2 + B^2 + C^2 + D^2,$$

where

$$A = aa' + bb' + cc' + dd',$$
$$B = ab' - ba' - cd' + dc',$$
$$C = ac' + bd' - ca' - db',$$
$$D = ad' - bc' + cb' - da'.$$

The numbers A, B, C, and D are each divisible by m:

$$A = aa' + bb' + cc' + dd' \equiv a^2 + b^2 + c^2 + d^2 = pm \equiv 0 \bmod m,$$
$$B = ab' - ba' - cd' + dc' \equiv ab - ba - cd + dc \equiv 0 \bmod m,$$
$$C = ac' + bd' - ca' - db' \equiv ac + bd - ca - bd \equiv 0 \bmod m,$$
$$D = ad' - bc' + cb' - da' \equiv ad - bc + cb - da \equiv 0 \bmod m.$$

Set $A = mA'$, $B = mB'$, $C = mC$, and $D = mD'$, so $pm^2 n = m^2(A'^2 + B'^2 + C'^2 + D'^2)$. Divide by m^2 and we have $pn = A'^2 + B'^2 + C'^2 + D'^2$ where $1 \leq n < m$.

Step 3: The prime p is a sum of four squares.

This is the descent step. By Step 1, there is a positive integer $m < p$ such that pm is a sum of three, and thus also four, squares. If $m = 1$ then we're done. If $m > 1$ then apply Step 2 repeatedly, replacing m with n each time until $n = 1$. □

Remark 6.7. Our use of Lemma 6.2 in the proof of the four-square theorem would work (to show $A, B, C, D \equiv 0 \bmod m$) using the first alternate choices of signs for Lemma 6.2 in Remark 6.3, but it would *not* work with the second alternate choice of signs related to quaternions. There is a separate proof of the four-square theorem using quaternions.

---

[6]If $a \equiv m/2 \bmod m$, say $a = m/2 + mr$, then $a^2 = (m/2)^2 + m^2(r + r^2) \equiv (m/2)^2 \bmod m^2$.

7. Markov's equation

The Markov equation, introduced by Markov in 1880 [11, Sect. 10], is

$$x^2 + y^2 + z^2 = 3xyz.$$

A solution (x, y, z) in positive integers is called a *Markov triple*, and every positive integer arising as part of a Markov triple is called a *Markov number*. For example, 1 and 2 are Markov numbers because of the solutions $(1, 1, 1)$ and $(2, 1, 1)$.

Markov's insight about this equation is that each Markov triple (x, y, z) other than $(1, 1, 1)$ can be created from a "smaller" Markov triple, as follows. Bringing $3xyz$ to the left side lets us interpret x as a root of the quadratic polynomial

$$T^2 - (3yz)T + (y^2 + z^2) = 0.$$

There is a second root of this equation (besides x), which we can find without the quadratic formula by thinking about relations between roots and coefficients. Letting the other root be r, our polynomial is $(T - x)(T - r) = T^2 - (x + r)T + xr$. Therefore $x + r = 3yz$, so $r = 3yz - x$. From one solution (x, y, z) of Markov's equation we get a second solution: $M_1(x, y, z) := (3yz - x, y, z) = ((y^2 + z^2)/x, y, z)$. Interchanging the roles of x, y, and z, we similarly get the additional solutions $M_2(x, y, z) := (x, 3xz - y, z)$ and $M_3(x, y, z) := (x, y, 3xy - z)$. From $(1, 1, 1)$ we can successively generate, for instance, $(2, 1, 1)$, $(2, 5, 1)$, and $(2, 5, 29)$. Applying the mappings M_i to get new Markov triples from old ones leads to a tree of solutions, the start of which is below. (Don't apply an M_i twice in a row since it is its own inverse.)
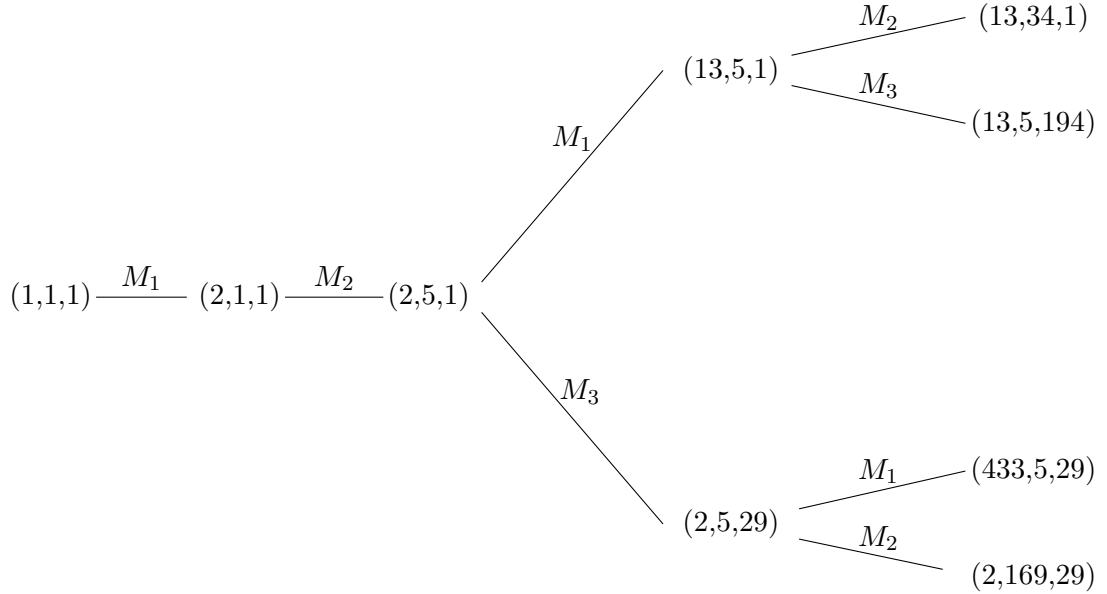
Markov proved every Markov triple can be produced by iteration from $(1, 1, 1)$. An account of his proof, which uses descent, is in [9]. Besides $(1, 1, 1)$, $(2, 1, 1)$, and the rearrangements $(1, 2, 1)$ and $(1, 1, 2)$, all other Markov triples (x, y, z) have distinct terms. Each Markov number arises as the maximum number in some Markov triple, and it is an open problem going back to 1913 [1], called the uniqueness conjecture, to prove each Markov

number other than 1 or 2 is the maximum term of exactly one Markov triple (regarding a triple with reordered terms as the same triple).

We want to use descent for a different purpose, also taken from [9]. We will prove a theorem of Frobenius and Hurwitz that shows the special role of 3 as a coefficient on the right side of Markov's equation.

Theorem 7.1. *For a positive integer k other than 1 or 3, the equation $x^2 + y^2 + z^2 = kxyz$ has no integral solution except $(0, 0, 0)$.*

Proof. First we will treat the case $k > 3$, returning later to $k = 2$.

Suppose a, b, and c satisfy $a^2 + b^2 + c^2 = kabc$. If a, b, or c is 0 then the equation says the sum of the squares of the other two terms is 0, so a, b, and c are all 0. Thus, assuming $(a, b, c) \neq (0, 0, 0)$ means a, b, and c are all nonzero. At least one of them is positive (otherwise the right side of the equation is negative). The other two are both positive or both negative, and in the negative case we can change their signs to get a solution where all are positive. So without loss of generality a, b, and c are all positive.

The numbers a, b, and c are distinct. To show this, we argue by contradiction. Suppose (without loss of generality) that $a = b$. Then $2a^2 + c^2 = ka^2c$, so $a^2(kc - 2) = c^2$. Therefore $kc - 2$ is a rational square, hence an integral square. Write $kc - 2 = d^2$ with $d \geq 1$, so $kc = 2 + d^2$. Therefore $2a^2 + c^2 = (2 + d^2)a^2$, so $c^2 = d^2a^2$, so $c = da$. Now $d^2 = kc - 2 = k(da) - 2$, so $2 = d(ka - d)$, which means $d \mid 2$, so d is 1 or 2. In either case we get $ka = 3$, which contradicts $k > 3$.

Without loss of generality, say $a > b > c \geq 1$. The triple $(kbc - a, b, c)$ is also a solution to $x^2 + y^2 + z^2 = kxyz$, and $kbc - a$ is positive since $a(kbc - a) = b^2 + c^2$ and $a > 0$. Which coordinate in $(kbc - a, b, c)$ is maximal? We know $b > c$ by design. Is $kbc - a > b$ or is $b > kbc - a$? We answer this by looking at the polynomial $f(x) = x^2 - (kbc)x + b^2 + c^2$. The roots of $f(x)$ are a and $kbc - a$, and

$$f(b) = 2b^2 + c^2 - kb^2c \leq 2b^2 + c^2 - kb^2 < 3b^2 - kb^2 = (3 - k)b^2 < 0.$$

The region where f is negative is between its two roots. Thus b lies between a and $kbc - a$. Since $b < a$ we must have $kbc - a < b < a$, so

$$0 < \max(kbc - a, b, c) = b < a = \max(a, b, c),$$

Repeating this construction, by descent we get a contradiction, so the equation $a^2 + b^2 + c^2 = kabc$ has only $(0, 0, 0)$ as an integer solution when $k > 3$.

Now we look at $k = 2$. Suppose $a^2 + b^2 + c^2 = 2abc$ with integers a, b, and c. Since $a^2 + b^2 + c^2$ is even, a, b, and c are not all odd. If exactly 1 of them is even then reducing both sides of the equation modulo 4 gives $2 \equiv 0 \bmod 4$, a contradiction. If exactly 2 are even then reducing modulo 2 gives $1 \equiv 0 \bmod 2$, another contradiction. Therefore a, b, and c are all even. Write $a = 2a'$, $b = 2b'$, and $c = 2c'$, so $a'^2 + b'^2 + c'^2 = 4a'b'c'$. This is the case $k = 4$, which we have already shown has no integral solution except $(0, 0, 0)$, so $(a, b, c) = (2a', 2b', 2c') = (0, 0, 0)$. $\square$

Why do we avoid $k = 1$ in Theorem 7.1? Looking at $x^2 + y^2 + z^2 = xyz$ modulo 3 shows x, y, and z are all multiples of 3 (check!). Writing $x = 3x'$, $y = 3y'$, and $z = 3z'$ yields $x'^2 + y'^2 + z'^2 = 3x'y'z'$, so a solution to $x^2 + y^2 + z^2 = xyz$ in $\mathbf{Z}$ is 3 times an integral solution of Markov's equation.

Here is another "positive" use of descent based on passing from one root of a quadratic polynomial to the other root.

Theorem 7.2. *For $m \in \mathbf{Z}^+$, if there are a and b in $\mathbf{Z}^+$ such that $m = (a^2 + b^2)/(ab + 1)$ then m is a perfect square.*

This was the final question on the International Math Olympiad in 1988, and there are Numberphile videos about it, such as https://www.youtube.com/watch?v=Y30VF3cSIYQ and https://www.youtube.com/watch?v=NcaYEaVTA4g.

Proof. We will present an argument by descent as described by Zagier[7].

Rewrite the equation as $a^2 + b^2 = m(ab + 1)$, or $a^2 + b^2 - mab - m = 0$. We will consider not just solutions (a, b) in positive integers, but solutions (a, b) in nonnegative integers. There's no solution where a and b are both 0 (why?), but a solution where one of a or b is 0 might occur and in fact finding such a solution is the whole point: if $a = 0$ then $m = b^2$ and if $b = 0$ then $m = a^2$, so either way m is a perfect square. We will show that from a solution in positive integers a and b, by descent there is a solution where a or b is 0.

Since $a^2 + b^2 - mab - m$ is symmetric in a and b, we can assume $a \geq b \geq 1$. We will find a solution (a', b) in *nonnegative* integers where $a' < a$.

From $a^2 + b^2 - mab - m = 0$, a is a root of $x^2 - mbx + (b^2 - m)$. This polynomial has two roots, a and a', where $a + a' = mb$ and $aa' = b^2 - m$ (the formulas linking roots and coefficients). So $a'^2 + b^2 - ma'b - m = 0$. Obviously $a' \in \mathbf{Z}$, since $a' = mb - a$, but why is $a' \geq 0$ and $a' < a$?

Claim: $a' \geq 0$. If $a' < 0$ then $-a' > 0$ and $mb < a$, so $m = b^2 - aa' > b^2 + a > a > bm$, so $1 > b$, which is a contradiction.

Claim: $a' < a$. Since $b \leq a$ (an initial assumption), $aa' = b^2 - m < b^2 \leq a^2$, so $a' < a$.

From the solution (a, b) in positive integers (where $a \geq b$ by symmetry), we got a solution (a', b') in nonnegative integers where $a' < a$ and $b' = b$, so $0 < a' + b' < a + b$: the *sum of terms* has gotten smaller. If $a' > 0$ then we can repeat the process and get a solution (a'', b'') where $0 < a'' + b'' < a' + b$. Eventually we must reach a solution $(a^{(n)}, b^{(n)})$ where $a^{(n)}$ or $b^{(n)}$ is 0, and that implies m is a perfect square. $\qquad\square$

Appendix A. Areas of Right Triangles

In Section 3 we saw Fermat's Last Theorem for exponent 4 follows from $x^4 + y^4 = z^2$ having no solution in $\mathbf{Z}^+$. Here we will explain what led Fermat to this equation. It was not Fermat's Last Theorem, but the following problems about areas of right triangles:

(1) Can a right triangle and square with side lengths in $\mathbf{Z}$ have the same area?
(2) Can a right triangle with side lengths in $\mathbf{Z}$ have twice the area of a square with side lengths in $\mathbf{Z}$?

Algebraically, if (a, b, c) is a Pythagorean triple we are asking if $ab/2$ can be a perfect square or twice a perfect square.

The first question is connected with $x^4 - y^4 = z^2$ and the second question is connected with $x^4 + y^4 = z^2$. This is explained in Table 2. The first column shows how to turn a Pythagorean triple (a, b, c) such that $ab/2$ is a perfect square into a positive integer solution of $x^4 - y^4 = z^2$. In the second column we turn such a solution (x, y, z) into a Pythagorean triple (a, b, c) such that $ab/2$ is a perfect square. In the next two columns we turn Pythagorean triples (a, b, c) with $ab/2$ being twice a perfect square into positive integer solutions of $x^4 + y^4 = z^2$ and *vice versa*. Note d in the fourth column is an integer since x or y must be even (otherwise $z^2 \equiv 2 \bmod 4$, which is impossible)

---

[7]See http://www-groups.mcs.st-andrews.ac.uk/%7Ejohn/Zagier/Solution1.3.html.

$a^2 + b^2 = c^2$, $ab/2 = d^2$	$x^4 - y^4 = z^2$	$a^2 + b^2 = c^2$, $ab/2 = 2d^2$	$x^4 + y^4 = z^2$
$x = c$	$a = z^2$	$x = b$	$a = x^2$
$y = 2d$	$b = 2x^2y^2$	$y = 2d$	$b = y^2$
$z = |a^2 - b^2|$	$c = x^4 + y^4$	$z = bc$	$c = z$
	$d = xyz$		$d = xy/2$

TABLE 2.

The transformations in the table between (a, b, c) and (x, y, z) are not inverses, but they show there's an integral right triangle with a certain kind of area exactly when a certain equation has solutions in $\mathbf{Z}^+$.

We showed by descent in Theorems 3.1 and 3.10 that $x^4 \pm y^4 = z^2$ has no solution in $\mathbf{Z}^+$, so no integral right triangle has an area that is a perfect square or twice a perfect square.

APPENDIX B. ANOTHER DESCENT WITH SUMS OF TWO SQUARES

By descent we will show solvability of a certain equation in $\mathbf{Q}$ implies solvability in $\mathbf{Z}$.

Theorem B.1. *Let $n \in \mathbf{Z}^+$. If n is a sum of two squares in $\mathbf{Q}$ then it is a sum of two squares in $\mathbf{Z}$.*

This says that if $n = x^2 + y^2$ has a solution for some $x, y \in \mathbf{Q}$ then it has a solution where $x, y \in \mathbf{Z}$. Before proving the theorem, we'll illustrate it with an example.

Example B.2. Check $193 = (933/101)^2 + (1048/101)^2$. The point $P = (933/101, 1048/101)$ lies on the circle $C : x^2 + y^2 = 193$ (with radius $\sqrt{193} \approx 13.8$). Since $P \approx (9.23, 10.37)$, the nearest integral point to P is $Q = (9, 10)$, which does *not* lie on C. See the figure below.[8] The line L through P and Q has equation $y = (19/12)x - 17/4$ and it meets C in a second point besides P: $P' = (-27/5, -64/5)$. Note P' is on C and has rational coordinates with common denominator 5, which is less than the previous common denominator 101.

Now repeat the process: since $P' = (-5.4, -12.8)$, the nearest integral point to P' is $Q' = (-5, -13)$; note we round -12.8 to -13, the nearest integer, rather than to -12. The point Q' is not on C, and the line L' through P' and Q' has equation $y = -(1/2)x - 31/2$ and it meets C in a second point besides P': the integral point $P'' = (-7, -12)$. We have found an integral solution $(-7, -12)$ to $x^2 + y^2 = 193$.
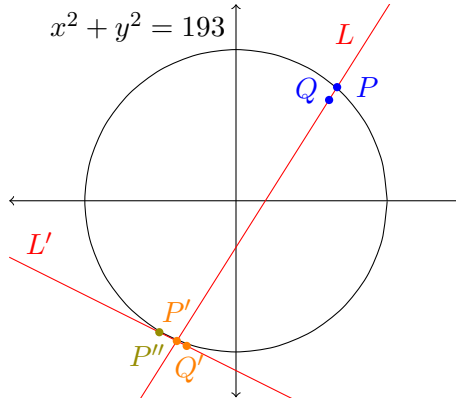
_____
[8]The figure is not drawn to scale.

As a check on your understanding, starting with $193 = (83/109)^2 + (1512/109)^2$, and $P = (83/109, 1512/109) \approx (.76, 13.87)$, apply the method above to find rational points on C with decreasing denominators until you reach an integral point: $(249/65, 868/65)$, $(64/5, -27/5)$, and finally $(12, -7)$.

Now let's prove Theorem B.1.

Proof. Our argument is from [12, App. II, Chap. III]. The main idea is due to Aubry (1912).

Suppose $n = r_1^2 + r_2^2$ with fractions r_1 and r_2. If r_1 and r_2 are in $\mathbf{Z}$, we're done, so assume at least one is not in $\mathbf{Z}$. Write the r_i's with a common denominator: $r_1 = a_1/b$ and $r_2 = a_2/b$, where a_1, a_2, and b are in $\mathbf{Z}$ and $b > 1$. We will show $n = r_1'^2 + r_2'^2$ where r_1' and r_2' are fractions with a common denominator *less* than b. Repeat this enough times to get a common denominator of 1, so n is a sum of integral squares.

The point (r_1, r_2) lies on the circle $x^2 + y^2 = n$. Pick a nearby $\mathbf{Z}$-point in the plane: choose k_1 and k_2 in $\mathbf{Z}$ such that $|r_i - k_i| \leq 1/2$. Since r_1 and r_2 are not both in $\mathbf{Z}$, $(r_1, r_2) \neq (k_1, k_2)$, so there is a line through these two points. This line meets the circle $x^2 + y^2 = n$ in the point (r_1, r_2). We will show this line meets the circle in a second point with rational coordinates having a smaller common denominator than b (the common denominator of the coordinates of (r_1, r_2)).

First let's check the line through (r_1, r_2) and (k_1, k_2) meets $x^2 + y^2 = n$ in a second point. We argue by contradiction. If the line only meets the circle at (r_1, r_2) then the line is tangent to the circle at (r_1, r_2), so the three points (r_1, r_2), (k_1, k_2), and $(0, 0)$ are the vertices of a right triangle (a tangent line to a point on a circle is always perpendicular to the line connecting the origin to the point of tangency). By the Pythagorean theorem,

$$k_1^2 + k_2^2 = (r_1^2 + r_2^2) + ((k_1 - r_1)^2 + (k_2 - r_2)^2) = n + ((k_1 - r_1)^2 + (k_2 - r_2)^2).$$

Both $k_1^2 + k_2^2$ and n are integers, so $(k_1 - r_1)^2 + (k_2 - r_2)^2$ is an integer. However, $|k_1 - r_1|$ and $|k_2 - r_2|$ are both less than $1/2$, so $0 \leq (k_1 - r_1)^2 + (k_2 - r_2)^2 \leq (1/2)^2 + (1/2)^2 = 1/2$. Then the only way $(k_1 - r_1)^2 + (k_2 - r_2)^2$ could be an integer is if it is 0, which forces $r_1 = k_1$ and $r_2 = k_2$. However, we are supposing r_1 and r_2 are not both integers, so we have a contradiction.

Now let's look more closely at the algebraic formula for the squared distance between (r_1, r_2) and (k_1, k_2). This number, which is positive, equals

$$
\begin{aligned}
(r_1 - k_1)^2 + (r_2 - k_2)^2 &= \left(\frac{a_1}{b} - k_1\right)^2 + \left(\frac{a_2}{b} - k_2\right)^2 \\
&= \left(\frac{a_1}{b}\right)^2 + \left(\frac{a_2}{b}\right)^2 - \frac{2(a_1 k_1 + a_2 k_2)}{b} + k_1^2 + k_2^2 \\
&= n - \frac{2(a_1 k_1 + a_2 k_2)}{b} + k_1^2 + k_2^2 \\
&= n + k_1^2 + k_2^2 - \frac{2(a_1 k_1 + a_2 k_2)}{b}.
\end{aligned}
$$

We can write this fraction in the form b'/b, where

$$b' := b(n + k_1^2 + k_2^2) - 2(a_1 k_1 + a_2 k_2) \in \mathbf{Z}$$

and

$$(r_1 - k_1)^2 + (r_2 - k_2)^2 = \frac{b'}{b} \implies (a_1 - bk_1)^2 + (a_2 - bk_2)^2 = bb'.$$

We will show the line through (r_1, r_2) and (k_1, k_2) meets $x^2 + y^2 = n$ in a rational point whose coordinates have b' as a common denominator. Since $(r_1 - k_1)^2 + (r_2 - k_2)^2 \leq (1/2)^2 + (1/2)^2 = 1/2$, we have $0 < b'/b \leq 1/2$, so

$$0 < b' \leq \frac{b}{2} < b.$$

The line through (r_1, r_2) and (k_1, k_2) can be described parametrically by

(B.1) $$L(t) = (k_1 + (r_1 - k_1)t, k_2 + (r_2 - k_2)t).$$

This meets $x^2 + y^2 = n$ at $t = 1$ ($L(1) = (r_1, r_2)$). Where else does it meet the circle? To find out, we solve for t in

$$
\begin{aligned}
n &= (k_1 + (r_1 - k_1)t)^2 + (k_2 + (r_2 - k_2)t)^2 \\
&= k_1^2 + 2k_1(r_1 - k_1)t + (r_1 - k_1)^2 t^2 + k_2^2 + 2k_2(r_2 - k_2)t + (r_2 - k_2)^2 t^2 \\
&= k_1^2 + k_2^2 + 2(k_1 r_1 - k_1^2 + k_2 r_2 - k_2^2)t + ((r_1 - k_1)^2 + (r_2 - k_2)^2)t^2 \\
&= k_1^2 + k_2^2 + 2(k_1 r_1 + k_2 r_2 - (k_1^2 + k_2^2))t + ((r_1 - k_1)^2 + (r_2 - k_2)^2)t^2 \\
&= k_1^2 + k_2^2 + \frac{2(a_1 k_1 + a_2 k_2) - 2b(k_1^2 + k_2^2)}{b}t + \frac{(a_1 - bk_1)^2 + (a_2 - bk_2)^2}{b^2}t^2.
\end{aligned}
$$

Using the definition of b' to rewrite the coefficients of t and t^2, we have

$$
\begin{aligned}
n &= k_1^2 + k_2^2 + \frac{b(n + k_1^2 + k_2^2) - b' - 2b(k_1^2 + k_2^2)}{b}t + \frac{b'}{b}t^2 \\
&= k_1^2 + k_2^2 + \left(n - (k_1^2 + k_2^2) - \frac{b'}{b}\right)t + \frac{b'}{b}t^2.
\end{aligned}
$$

Bringing all terms to the right side,

$$0 = \frac{b'}{b}t^2 + \left(n - (k_1^2 + k_2^2) - \frac{b'}{b}\right)t + k_1^2 + k_2^2 - n.$$

This has a root at $t = 1$ (because $L(1)$ is on the circle, but it can also be seen algebraically), so we know $t - 1$ is a factor on the right, leading to

$$0 = (t - 1)\left(\frac{b'}{b}t + n - (k_1^2 + k_2^2)\right).$$

Thus the second point of intersection of the line $L(t)$ with the circle $x^2 + y^2 = n$ is at

$$t = \frac{b(k_1^2 + k_2^2 - n)}{b'}.$$

Feeding this value of t into (B.1), and writing r_i as a_i/b, we get the point

$$L\left(\frac{b(k_1^2 + k_2^2 - n)}{b'}\right) = \left(k_1 + \frac{(a_1 - bk_1)(k_1^2 + k_2^2 - n)}{b'}, k_2 + \frac{(a_2 - bk_2)(k_1^2 + k_2^2 - n)}{b'}\right),$$

which shows this second point of intersection of the line and circle is a rational point and b' is a common denominator for its coordinates. We noted earlier that $0 < b' \leq b/2 < b$, so the common denominator for this new rational point on $x^2 + y^2 = n$ is smaller than that for (r_1, r_2), and we are done. $\square$

The proof of Theorem B.1 works for a sum of three squares, using lines and spheres in three dimensions. That is, an integer that is a sum of three rational squares is also a sum of three integral squares. The only change to be made in the proof is the following: now we have $|a_i/b - k_i| \leq 1/2$ for $i = 1, 2, 3$, and $(a_1/b - k_1)^2 + (a_2/b - k_2)^2 + (a_3/b - k_3)^2 \leq 3/4$ instead of $\leq 1/2$. So the new rational point on the sphere $x^2 + y^2 + z^2 = n$ will have a common denominator $b' \leq (3/4)b$, which is still less than b, which means everything still works in the proof when it is done for sums of three squares.

Theorem B.1 can fail for some algebraic expressions other than $x^2 + y^2$. For example, $x^2 + 82y^2 = 2$ has no solution in $\mathbf{Z}$ but infinitely many solutions in $\mathbf{Q}$, such as $(4/7, 1/7)$, $(16/13, 1/13)$, $(40/29, 1/29)$, and $(20/29, 7/47)$. And $x^3 + y^3 = 13$ has no solution in $\mathbf{Z}$ but infinitely many solutions in $\mathbf{Q}$, such as $(7/3, 2/3)$, $(2513/1005, -1388/1005)$, and $(26441619018689/18636783082845, 40343602894936/18636783082845)$. If we try to apply the geometric method of Example B.2 to these curves, then we run into problems.

- The nearest integral point to $(4/7, 1/7)$ is $(1, 0)$ and the line through those two points meets $x^2 + 82y^2 = 2$ in $(16/13, -1/13)$, so the denominator has gone up, not down. The line through $(16/13, -1/13)$ and its nearest integral point, which is again $(1, 0)$, meets $x^2 + 82y^2 = 2$ in the original point $(4/7, 1/7)$, so we return to where we started.
- The nearest integral point to $(7/3, 2/3)$ is $(2, 1)$ and the line through those two points meets $x^3 + y^3 = 13$ in $(2/3, 7/3)$. Next, the line through $(2/3, 7/3)$ and its nearest integral point $(1, 2)$ meets $x^3 + y^3 = 13$ in $(7/3, 2/3)$, which is the original point.

Appendix C. Some results related to the four-square theorem

Before Lagrange proved the four-square theorem, Euler made partial progress [6, Theorem 20]: the theorem below.

Theorem C.1 (Euler)**.** *If $n \in \mathbf{Z}^+$, then $n = a^2 + b^2 + c^2 + d^2$ where a, b, c, d are in $\mathbf{Q}$.*

Proof. We will be using Lemma 6.2 for $\mathbf{Q}$ (same proof as in $\mathbf{Z}$: it's a polynomial identity).

The theorem is obvious at $n = 1$, so by Lemma 6.2 (for $\mathbf{Q}$) it suffices to show each prime number is a sum of four rational squares.

Assume there are primes that are not a sum of four squares in $\mathbf{Q}$ and let p be the least such prime. As in Step 1 in the proof of Theorem 6.6, there is a positive integer $k < p$ such that $pk = x^2 + y^2 + 1$ where $x, y \in \mathbf{Z}$ (this uses Lemma 6.4).

By the minimality of p, each prime less than p is a sum of four squares in $\mathbf{Q}$, so by Lemma 6.2 (for $\mathbf{Q}$) each positive integer less than p is a sum of four squares in $\mathbf{Q}$. Thus k is a sum of four squares in $\mathbf{Q}$. Writing

$$p = \frac{x^2 + y^2 + 1}{k} = \frac{(x^2 + y^2 + 1)k}{k^2},$$

we have $(x^2 + y^2 + 1)k = A^2 + B^2 + C^2 + D^2$ for some $A, B, C, D \in \mathbf{Q}$ by Lemma 6.2 (for $\mathbf{Q}$). Then

$$p = \frac{A^2 + B^2 + C^2 + D^2}{k^2} = \left(\frac{A}{k}\right)^2 + \left(\frac{B}{k}\right)^2 + \left(\frac{C}{k}\right)^2 + \left(\frac{D}{k}\right)^2 = a^2 + b^2 + c^2 + d^2,$$

where $a = A/k$, $b = B/k$, $c = C/k$, and $d = D/k$. This contradicts the definition of p, so every prime is a sum of four squares in $\mathbf{Q}$. $\square$

A summary of Lagrange's proof of the four-square theorem is in [4, pp. 279–281]. Here is his generalization of Lemma 6.4, which was used in the proof.

Theorem C.2. *Let p be an odd prime and $a, b \not\equiv 0 \bmod p$. For each $c \in \mathbf{Z}$, the congruence*

$$ax^2 + by^2 \equiv c \bmod p$$

has at least one solution.

Proof. We'll see that the proof of Lemma 6.4 about $x^2 + y^2 + 1 \equiv 0 \bmod p$ (where $a = b = 1$ and $c = -1$) remains applicable to the more general situation above.

Rewrite $ax^2 + by^2 \equiv c \bmod p$ as $ax^2 \equiv c - by^2 \bmod p$ and let $A = \{ax^2 \bmod p : x \in \mathbf{Z}/(p)\}$ and $B = \{c - by^2 \bmod p : y \in \mathbf{Z}/(p)\}$.

Since there are $(p + 1)/2$ squares mod p (including 0^2) and $a \not\equiv 0 \bmod p$, we have $|A| = (p+1)/2$. Similarly, $|B| = (p+1)/2$. Since $|A| + |B| = p + 1 > p$, A and B can't be disjoint in $\mathbf{Z}/(p)$, so some x_0 and y_0 in $\mathbf{Z}/(p)$ satisfy $ax_0^2 \equiv c - by_0^2 \bmod p$. Thus $ax_0^2 + by_0^2 \equiv c \bmod p$. $\square$

References

[1] M. Aigner, *Markov's Theorem and 100 Years of the Uniqueness Conjecture*, Springer-Verlag, New York, 2013.

[2] J. W. S. Cassels, *Lectures on Elliptic Curves*, Cambridge Univ. Press, Cambridge, 1991.

[3] H. Clemens, *A Scrapbook of Complex Curve Theory*, 2nd ed., Amer. Math. Soc., 2003.

[4] L. E. Dickson, *History of the Theory of Numbers*, Vol. II, Chelsea, New York, 1971.

[5] L. Euler, Letter to Goldbach on May 4, 1748. URL http://eulerarchive.maa.org//correspondence/letters/OOO841.pdf.

[6] L. Euler, "Demonstratio theorematis Fermatiani omnem numerum sive integrum sive fractum esse summam quatuor pauciorumve quadratorum," Novi Commentarii academiae scientiarum Petropolitanae **5** (1760), 13–58. URL https://scholarlycommons.pacific.edu/euler-works/242/.[9]

[7] L. Euler, "Novae demonstrationes circa resolutionem numerorum in quadrata," Nova Acta Eruditorum **17** (1773), 193–211. URL https://scholarlycommons.pacific.edu/euler-works/445/.[10]

[8] D. Kalman, R. Mena, and S. Shahriari, "Variations on an Irrational Theme – Geometry, Dynamics, Algebra," Math. Mag. **70** (1997), 93–104.

[9] M. G. Krein, "Markov's Diophantine Equation," pp. 121–126 in *Kvant Selecta: Algebra and Analysis, I* (S. Tabachnikov, ed.), Amer. Math. Soc., Providence, 1991.

[10] J.-L. Lagrange "Démonstration d'un théorème d'arithmétique," Nouveaux mémoires de l'Acadmie royale des sciences et belles-lettres de Berlin, Année 1770 (1772), 123–133. Oeuvres t. 3, Gauthier-Villars, Paris (1869), 189–201. URL https://gallica.bnf.fr/ark:/12148/bpt6k229222d/f190.

[11] A. Markoff, "Sur les formes quadratiques binaires indefinies (Second mémoire)," Math. Annalen **17** (1880), 379–399. URL https://eudml.org/doc/156934.

[12] A. Weil, *Number Theory: An Approach Through History from Hammurapi to Legendre*, Birkhäuser, Boston, 1984.

---

[9]English translation is in a link at this page in the upper right.

[10]English translation is in a link at this page in the upper right.