The ABC's of Number Theory

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation	Elkies, Noam D. 2007. The ABC's of number theory. The Harvard College Mathematics Review 1(1): 57-76.
Citable link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:2793857
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

FACULTY FEATURE ARTICLE

6

The ABC's of Number Theory

Prof. Noam D. Elkies[†]
Harvard University
Cambridge, MA 02138
elkies@math.harvard.edu

Abstract

The ABC conjecture is a central open problem in modern number theory, connecting results, techniques and questions ranging from elementary number theory and algebra to the arithmetic of elliptic curves to algebraic geometry and even to entire functions of a complex variable. The conjecture asserts that, in a precise sense that we specify later, if A, B, C are relatively prime integers such that $A + B = C$ then A, B, C cannot all have many repeated prime factors. This expository article outlines some of the connections between this assertion and more familiar Diophantine questions, following (with the occasional scenic detour) the historical route from Pythagorean triples via Fermat's Last Theorem to the formulation of the ABC conjecture by Masser and Oesterlé. We then state the conjecture and give a sample of its many consequences and the few very partial results available. Next we recite Mason's proof of an analogous assertion for polynomials $\mathsf{A}(t), \mathsf{B}(t), \mathsf{C}(t)$ that implies, among other things, that one cannot hope to *dis*prove the ABC conjecture using a polynomial identity such as the one that solves the Diophantine equation $x^2 + y^2 = z^2$. We conclude by solving a Putnam problem that predates Mason's theorem but is solved using the same method, and outlining some further open questions and fragmentary results beyond the ABC conjecture.[‡]

6.1 Pythagorean triples: $x^2 + y^2 = z^2$

An ordered triple (x, y, z) of integers is called a **Pythagorean triple** if and only if it solves the Diophantine equation $x^2 + y^2 = z^2$; that is, if and only if $|x|$ and $|y|$ are the lengths of the sides, and $|z|$ the length of the hypotenuse, of a right triangle. (We allow degenerate triangles with a "side" of length zero.) It is well-known that every such triple is proportional to

$$(x, y, z) = (m^2 - n^2, 2mn, m^2 + n^2) \tag{6.1}$$

for some integers m, n. Equivalently (dividing by n^2 to obtain polynomials in the single rational variable $t = m/n$), the solution (x, y, z) is proportional to $(t^2 - 1, 2t, t^2 + 1)$ for some $t \in \mathbb{Q}$, or to $(1, 0, 1)$ which arises for "$t = \infty$" (corresponding to $(m, n) = (1, 0)$). That is, all Pythagorean triples are accounted for by the single polynomial identity

$$(t^2 - 1)^2 + (2t)^2 = (t^2 + 1). \tag{6.2}$$

[†]Noam D. Elkies earned his doctorate in mathematics in 1987 at Harvard, where his advisors where Professors Barry Mazur and Benedict H. Gross. After three years in Harvard's Society of Fellows he joined the Mathematics faculty and has remained at Harvard since. Most of his research is in number theory, usually Diophantine geometry (the combination of algebraic geometry and Diophantine equations) and/or computational number theory. Other interests include some combinatorial mathematics (lattices and codes, incidence geometry, and combinatorial games) and, outside of mathematics, classical music (mostly composition and piano) and chess (usually chess problems and endgames).

[‡]Supported in part by NSF grant DMS-0501029.

This classical fact can be profitably approached from many points of view.[1] In one familiar approach, illustrating an important method in algebraic geometry, we first divide by z^2 to obtain the equivalent $(x/z)^2 + (y/z)^2 = 1$, so we now seek rational solutions of $X^2 + Y^2 = 1$, or geometrically a **rational point** (a point with both coordinates rational) on the unit circle. Note that two nonzero solutions $(x : y : z)$ in integers yield the same solution (X, Y) in rationals if and only if they are proportional, so that by going from $x^2 + y^2 = z^2$ to $X^2 + Y^2 = 1$ we have automatically identified proportional Pythagorean triples (corresponding to similar right triangles). The unit vector $(1, 0)$ is an obvious rational point on the circle. This point yields only a degenerate Pythagorean triple, but we can use it to find any other rational point (X, Y) using the straight line through (X, Y) and $(1, 0)$. The general such line is $Y = -t(X - 1)$, where the slope $-t$ must be rational if X and Y are. (We choose $-t$ rather than t for consistency with equation (6.2).) Substituting $-t(X - 1)$ for Y in $X^2 + Y^2 = 1$ we get the quadratic equation $X^2 + t^2(X - 1)^2 = 1$, one of whose solutions must be $X = 1$. The other solution is then the root of

$$\frac{X^2 + t^2(X-1)^2 - 1}{X - 1} = (t^2 + 1)X - (t^2 - 1),$$

that is, $X = (t^2 - 1)/(t^2 + 1)$. Then $Y = -t(X - 1) = 2t/(t^2 + 1)$, so we have recovered the rational point corresponding to the solution $(t^2 - 1, 2t, t^2 + 1)$ of $x^2 + y^2 = z^2$. See Figure 1, which shows this construction for $t = 2$.

This procedure readily generalizes: instead of $X^2 + Y^2 - 1$ we can use any irreducible polynomial $P(X, Y)$ of degree 2, and instead of the initial point $(1, 0)$ we can use any rational solution (X_0, Y_0) of $P(X, Y) = 0$; the lines through (X_0, Y_0) not tangent to the curve $P(X, Y) = 0$ at that point then parametrize all other rational points on the curve. [Try $X^2 + Y^2 = 2$ and $X_0 = Y_0 = 1$. What goes wrong if we attempt this for $P(X, Y) = X^2 + Y^2$ and $X_0 = Y_0 = 0$? Note that $X^2 + Y^2$ is irreducible over the rationals, but not over $\mathbb{C}$ where it factors as $(X + iY)(X - iY)$.] The technique even works in some settings beyond plane curves of degree 2, including notably degree-3 plane curves with a double point; see Figure 2 for the example of the double point $(0, 0)$ on the curve $(X + Y)^3 = XY$. In our special case of $X^2 + Y^2 = 1$ and $(X_0, Y_0) = (1, 0)$ we can make yet another connection: if $(X, Y) = (\cos\theta, \sin\theta)$ then our line $Y = -t(X - 1)$ makes an angle of $\theta/2$ with the vertical. This can be seen by elementary plane geometry for $0 < \theta < \pi$, starting from the fact that $(0, 0)$, $(1, 0)$ and (X, Y) are vertices of an isosceles triangle (this too is shown in Figure 1); in general one must remember that θ is defined only up to integer multiples of 2π. In any case, this gives $t = \cot(\theta/2)$, so our parametrization is equivalent to the trigonometric half-angle formulas that give $\cot(\theta/2)$ as a rational function of $(\sin\theta, \cos\theta)$ and vice versa:

$$\cot\frac{\theta}{2} = \frac{\sin\theta}{1 - \cos\theta}; \qquad \cos\theta = \frac{\cot^2(\theta/2) - 1}{\cot^2(\theta/2) + 1}, \ \ \sin\theta = \frac{2\cot(\theta/2)}{\cot^2(\theta/2) + 1}. \tag{6.3}$$

These formulas reappear in integral calculus in the guise of the universal substitution that converts $\int f(\sin\theta, \cos\theta)\, d\theta$ (where f is any rational function) into $\int F(t)\, dt$ for some rational function $F \in \mathbb{R}(t)$, which can then be expanded in partial fractions to obtain an elementary antiderivative. Equivalently this lets us integrate any rational function of X and $\sqrt{1 - X^2}$ with respect to X, and the generalization to quadratic $P(X, Y) = 0$ lets us replace $\sqrt{1 - X^2}$ by the square root of any quadratic polynomial.

[1] Besides the algebro-geometric method we follow, at least four others come to mind, which suggest various perspectives on and generalizations of the result. The most elementary may be to begin with the trigonometric identities (6.3), or with an equivalent geometric calculation with isosceles and right triangles. An elementary derivation from unique factorization in $\mathbb{Z}$ is obtained by removing common factors from (x, y, z), switching x, y if necessary to make x odd, and using the factorization $x^2 = z^2 - y^2 = (z - y)(z + y)$ and the fact that $\gcd(z - y, z + y) = 1$ to write $z \pm y = (m \pm n)^2$ for some coprime integers m, n. See for instance [IR, p.23, Exercise 12]. Alternatively, factor $z^2 = (x + iy)(x - iy)$ in the ring $\mathbb{Z}[i]$ of Gaussian integers, and use unique factorization in $\mathbb{Z}[i]$; this explains why x and y are the real and imaginary parts of $(m + in)^2$. Finally, for $X, Y \in \mathbb{Q}$ we have $X^2 + Y^2 = 1$ if and only if the element $X + iY$ of $\mathbb{Q}(i)$ has norm 1, which by Hilbert's Theorem 90 is equivalent to $X + iY = w/\overline{w}$ for some nonzero $w \in \mathbb{Q}(i)$. Taking $t = \mathrm{Re}(w)/\mathrm{Im}(w)$ we recover $X + iY = (t^2 - 1 + 2it)/(t^2 + 1)$. See [Ta].

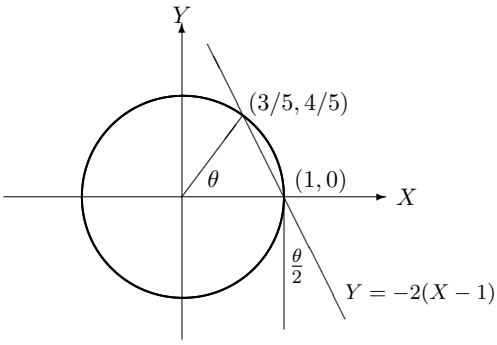

Figure 1: $X^2 + Y^2 = 1$

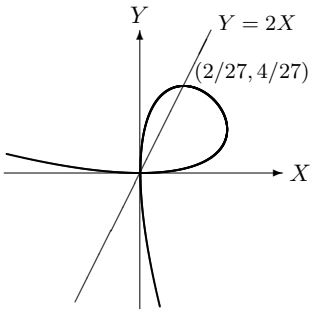

Figure 2: $(X + Y)^3 = XY$

But we have digressed from our main plot, to which we now return by looking at $x^2 + y^2 = z^2$ and the parametrization (6.1) or (6.2) from another point of view. We ask: *How many solutions does the Diophantine equation $x^2 + y^2 = z^2$ have in integer triples (x, y, z)?* Our parametrizations provide infinitely many (x, y, z) even when we identify proportional solutions, but we can still ask how common these solutions are. To make this vague question more precise, for all $N > 0$ define $\mathcal{C}(N)$ to be the number of solutions of $x^2 + y^2 = z^2$ in integers such that x^2, y^2, z^2 are relatively prime and of absolute value at most N. (We give the condition on x, y, z in this form because of the way we intend to generalize it to other Diophantine equations, though of course for $x^2 + y^2 = z^2$ the absolute value condition is equivalent to the single inequality $z^2 \leq N$.) Then the existence of infinitely many non-proportional Pythagorean triples is equivalent to the fact that $\mathcal{C}(N) \to \infty$ as $N \to \infty$, and we ask: *How quickly does $\mathcal{C}(N)$ grow?*

Using either of the forms (6.1) and (6.2) of our parametrization of Pythagorean triples we see that $\mathcal{C}(N)$ should grow as some multiple of $N^{1/2}$. For instance, (6.1) gives points (m, n) in the circle $m^2 + n^2 \leq N^{1/2}$, whose number is asymptotic to the area $\pi N^{1/2}$ of the circle. This is not quite right because we must count only relatively prime (m, n), and if both m and n are odd then we must remove a common factor of 2; but each of these corrections changes the asymptotic formula only by a constant factor. As it happens this factor is $2/(3\zeta(2)) = 4/\pi^2$, making $\mathcal{C}(N) \sim (4/\pi)N^{1/2}$. But it is the exponent $1/2$ that concerns us here, and we could have guessed this exponent much more easily as follows. Let $A = x^2$, $B = y^2$, and $C = z^2$. Then

$$A + B = C,$$

and the number of solutions of $A + B = C$ in relatively prime integers in $[-N, N]$ is asymptotically proportional to N^2. Of the $2N + 1$ integers in $[-N, N]$, approximately $N^{1/2}$ are squares (and all but one are squares in two different ways, but this will not affect the exponent of N, only the coefficient of that power). So, if we pick A, B, C independently and uniformly at random from the integers in $[-N, N]$, the probability that all three will be squares is asymptotically proportional to $N^{-3/2}$. While we actually choose A, B, C not at random but subject to $A + B = C$, it seems a reasonable guess that the fraction of such (A, B, C) all of which are squares is still roughly $N^{-3/2}$, giving a total of roughly $N^{2-\frac{3}{2}} = N^{1/2}$ such triples in that range.

If you think this seems suspiciously easy, you are right: we are only guessing the correct answer (up to a constant factor), not proving it. This kind of heuristic is quite naïve, and can easily fail. For instance, for the equations $x^2 + y^2 + z^2 = 0$ or $x^2 + y^2 = 3z^2$ we might similarly expect the number of solutions with all three terms in $[-N, N]$ to grow at the same $N^{1/2}$ rate. But neither of these equations has any solution other than the trivial $(0, 0, 0)$: the first obviously so, because the terms x^2, y^2, z^2 are all nonnegative; and the second because after removing common factors from (x, y, z) we get a contradiction mod 3.[2] In the other direction, the heuristic might grossly underestimate the

[2]In fact these two obstructions are more similar than they might seem: $x^2 + y^2 + z^2 = 0$ has no nontrivial solution in the

number of solutions. Consider for example solutions in relatively prime integers of $(x+y)^3 = xyz$ (the homogeneous form of the curve $(X+Y)^3 = XY$ shown in Figure 2). We might expect very few solutions, on the grounds that there are about $8H^3$ triples (x,y,z) of integers in $[-H,H]$, and in that range $(x+y)^3 - xyz$ can be as large as a multiple of H^3, so should vanish with probability only c/H^3 for some $c > 0$, leaving a constant expected number of solutions no matter how large H is. Somewhat more reasonably, we could start with the number of solutions in $\max(|x|,|y|,|z|) \in (2^{h-1}, 2^h]$ and then sum over $h \le \log_2 H$; but even then we would guess that the number of solutions with $\max(|x|,|y|,|z|) \le H$ grows only logarithmically. But in fact the rational parametrization by lines through the origin shows that the correct order of growth is $H^{2/3}$. Here the failure of the naïve heuristic can be attributed to the singularity of our curve at the origin. In higher dimensions, examples are known where our heuristic fails for other, subtler reasons.

Still, such failures are not surprising. What is remarkable is how often such a naïve heuristic gives the correct answer when this answer can be established, and an answer consistent with or close to the predictions of more refined conjectures and heuristics when the correct answer is not known but the problem fits into a suitable mathematical framework. In the next few sections we illustrate this by successively generalizing the problem of solving $x^2 + y^2 = z^2$ until we reach the ABC conjecture.

6.2 Fermat's "Last Theorem" (FLT): $x^n + y^n = z^n$

Of the many fruitful generalizations of $x^2 + y^2 = z^2$, one of the most natural and by far the best known is the Fermat equation $x^n + y^n = z^n$ for $n \ge 2$. Again we seek solutions in nonzero integers, or equivalently solutions of $X^n + Y^n = 1$ in rational numbers $X = x/z$, $Y = y/z$. The locus of $X^n + Y^n = 1$ is known as the **n-th Fermat curve**; Figures 3 and 4 show part of the real locus for $n = 3$ and the entire real locus for $n = 4$, and are typical of the visual appearance (albeit not necessarily of the arithmetic or algebraic geometry) of Fermat curves with $n \ge 3$ odd or even respectively.

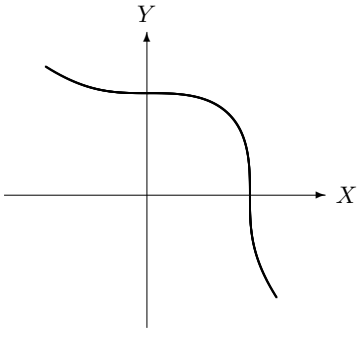

Figure 3: $X^3 + Y^3 = 1$

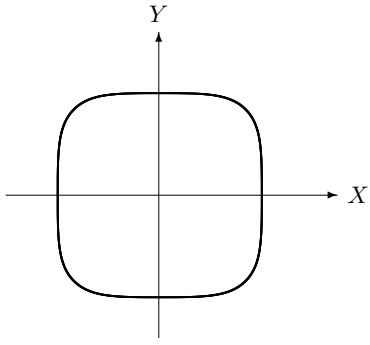

Figure 4: $X^4 + Y^4 = 1$

Fermat's "Last Theorem" (FLT) is the assertion, recorded by Fermat in 1637 and proved by him at least for $n = 4$, that for $n \ge 3$ there are no solutions of $x^n + y^n = z^n$ in nonzero integers; equivalently,

real field $\mathbb{R}$, and $x^2 + y^2 = 3z^2$ has no nontrivial solution in the field $\mathbb{Q}_3$ of 3-adic numbers. Since we live in the real world rather than the 3-adic world, the former obstruction is more intuitive to us, but both $\mathbb{R}$ and $\mathbb{Q}_3$ (and more generally $\mathbb{Q}_p$ for any prime p) are completions of $\mathbb{Q}$ with respect to the corresponding valuations on $\mathbb{Q}$, and decades of experience have shown the advantage of regarding the real and p-adic valuations of $\mathbb{Q}$ on as equal a footing as possible.

At this point we cannot resist another digression. Both $x^2 + y^2 + z^2 = 0$ and $x^2 + y^2 = 3z^2$ are obstructed not just over $\mathbb{R}$ and $\mathbb{Q}_3$ respectively, but also over $\mathbb{Q}_2$. It turns out that for *any* irreducible homogeneous quadratic $P(x,y,z)$ there are at most finitely many completions of $\mathbb{Q}$ in which there are no nonzero solutions of $P(x,y,z) = 0$, and that the number — call it ν — of such completions (either real or p-adic) is always even; this is equivalent to Quadratic Reciprocity. Conversely, any finite subset of $\{\mathbb{R}, \mathbb{Q}_2, \mathbb{Q}_3, \mathbb{Q}_5, \mathbb{Q}_7, \ldots\}$ of even size can arise this way, a fact that ultimately amounts to the determination of the 2-torsion of the Brauer group of $\mathbb{Q}$. Finally, if $\nu = 0$ then $P(x,y,z) = 0$ does in fact have nontrivial rational solutions; that is, the Hasse principle holds for homogeneous quadratics in three variables over $\mathbb{Q}$.

that the n-th Fermat curve has no rational points other than $(\pm 1, 0)$ and $(0, \pm 1)$ (with minus signs allowed only when n is even). Why should $n \geq 3$ behave so differently from $n = 2$? Let us consult our heuristic for estimating the expected number of solutions of $x^n + y^n = z^n$ with $\max(|x^n|, |y^n|, |z^n|) \in (N/2, N]$. (Every solution (x, y, z) will satisfy this condition with $N = 2^h$ for a unique nonnegative integer h.) As before we write $(A, B, C) = (x^n, y^n, z^n)$, and observe that $A + B = C$, and that the number of triples (A, B, C) of integers with $A + B = C$ and $\max(|A|, |B|, |C|) \in (N/2, N]$ is asymptotically proportional to N^2. But now we want each of them to be not a square but an n-th power for some $n \geq 3$, and n-th powers get rarer as n increases. Indeed the number of n-th powers in $[-N, N]$ grows only as $N^{1/n}$, so the probability that three integers A, B, C chosen independently and uniformly at random in that range are all n-th powers is asymptotically proportional to $N^{3((1/n)-1)}$. We thus expect roughly $N^{2+3((1/n)-1)} = N^{(3-n)/n}$ such triples with $A + B = C$. The exponent $(3 - n)/n$ is positive, zero, or negative according as $n < 3$, $n = 3$, or $n > 3$. Taking $N = 2^h$ and summing over h, we thus expect the solutions to be plentiful for $n < 3$ (the number of solutions up to N growing as a positive power of N), sparse for $n = 3$, and finite in number for $n > 3$. The same should be true of primitive[3] integral solutions of $A_0 x^n + B_0 y^n = C_0 z^n$ for any fixed choice of A_0, B_0, C_0, corresponding to rational points on the curve $A_0 X^n + B_0 Y^n = C_0$.

It turns out that each of these predictions is essentially correct. For $n = 1$ the result is almost trivial. For $n = 2$ we saw that, once the curve $A_0 X^2 + B_0 Y^2 = C_0$ has a rational point P, the lines through P yield the expected plenty of rational points on the curve. For $n \geq 3$ we must appeal to more advanced and recent results on Diophantine equations. When $n = 3$, the curve $E : A_0 X^3 + B_0 Y^3 = C_0$ is a nonsingular cubic plane curve, and thus an **elliptic curve** assuming it has a rational point P.[4] Here it is not so easy to get new rational points, because a typical line through P meets E at two more points, which in general are not rational. To obtain a new rational point we must use the line joining two rational points on E, or tangent to one rational point. This is shown in Figure 5 for the curve with $(A_0, B_0, C_0) = (1, 1, 91)$: the line through the rational points[5] $(3, 4)$ and $(6, -5)$ meets E again at $(9/2, -1/2)$, and the tangent at $(6, -5)$ meets E again at $(-204/341, 1535/341)$.

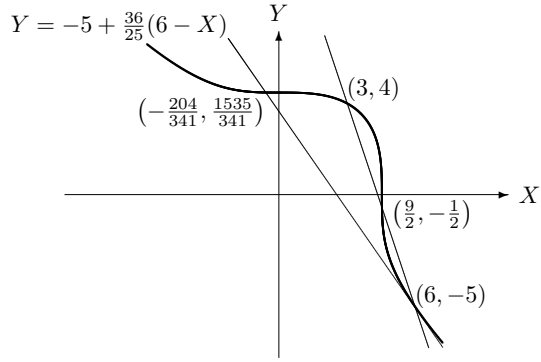

Figure 5: some rational points on $X^3 + Y^3 = 91$

[3] An integer solution (x, y, z) of a homogeneous polynomial equation $p(x, y, z) = 0$ is said to be **primitive** if $\gcd(x, y, z) = 1$. Every integer solution other than $(0, 0, 0)$ can be written uniquely as (kx, ky, kz) for some primitive solution (x, y, z) and some positive integer k.

[4] It is known that in characteristic zero such a curve is always isomorphic to one of the more familiar form $Y^2 = P_3(X)$ for some polynomial P_3 with distinct roots. See [Si1, Chapter III, §3] for such isomorphisms, and [Si1, Chapter III, §1] for standard formulas for elliptic curves.

[5] The value $C_0 = 91$ was chosen so that our curve has two simple rational points $(3, 4)$ and $(6, -5)$. This required a simple but nontrivial solution of $X^3 + Y^3 = X'^3 + Y'^3$. It would have been nice to use the famous "Ramanujan taxicab" example $C_0 = 1729 = 1^3 + 12^3 = 9^3 + 10^3$; but this would make it hard to draw a clear and accurate Figure 5, because $(1, 12)$ is too close to an inflection point of E and $(10, 9)$ too close to the middle of the curve. Our example with $(3, 4)$ and $(6, -5)$ relies instead on another famous identity $3^3 + 4^3 + 5^3 = 6^3$, which is tantalizingly reminiscent of $3^2 + 4^2 = 5^2$ but alas does not generalize further: $\sum_{m=3}^{n+2} m^n \neq (n + 3)^n$ once $n > 3$.

By drawing more lines and tangents we can generate infinitely many rational points on $X^3+Y^3=91$, and it can be shown that every rational point can be obtained this way. As one might guess from the case of $(-204/341, 1535/341)$, the resulting primitive solutions of $x^3+y^3=91z^3$ grow rapidly, and it turns out that the number of primitive solutions with all variables in $[-N,N]$ is asymptotic only to $R\log N$ for some $R>0$. There are similar results for any elliptic curve E. By a famous theorem of Mordell [Mo] there is a finite list of rational points on E from which all other points can be recovered by repeatedly drawing chords and tangents through points already known or constructed. More precisely, Mordell uses the chords-and-tangents construction to give the set $E(\mathbb{Q})$ of rational points on E the structure of an abelian group,[6] and proves that this group is finitely generated. It then follows from the Néron-Tate theory of canonical heights that the number of rational points $(x/z, y/z)$ with each of x,y,z in $[-N,N]$ is asymptotic to $R(\log N)^{\rho/2}$, where ρ is the rank of the abelian group $E(\mathbb{Q})$ and R is a positive constant depending on E. The curve has finitely many rational points if and only if $\rho=0$. It is known that this happens for the cubic Fermat curve $X^3+Y^3=1$, whose only rational points are the obvious $(1,0)$, $(0,1)$, and the point at infinity $(X:Y:1)=(1:-1:0)$.

Finally, for $n>3$ the curve $A_0X^n+B_0Y^n=C_0$ is a smooth plane curve of degree at least 4. Mordell conjectured that (as our heuristics suggest) every such curve has only finitely many rational points.

At any rate there is no longer a general method for constructing new points out of known ones; even the line through two known points, or tangent to one known point, meets the curve in $n-2$ more points (allowing points with complex coordinates), and those points need not be rational once $n-2>1$. For example, the line $X+Y=1$ through the rational points $(X,Y)=(0,1)$ and $(1,0)$ on the Fermat quartic $X^4+Y^4=1$ meets the curve again in a pair of Galois-conjugate points, each defined only over $\mathbb{Q}(\sqrt{-7})$, namely $\left(\frac{1}{2}(1\pm\sqrt{-7}), \frac{1}{2}(1\mp\sqrt{-7})\right)$. More generally, Mordell conjectured that any algebraic curve of genus at least 2 has only finitely many rational points. (The genus of a curve is a measure of its complexity[7]; an irreducible plane curve of degree d has genus $(d-1)(d-2)/2$ at most, with equality if and only if the curve is smooth; an elliptic curve has genus 1, and rationally parametrized curves have genus 0.) Mordell's conjecture was finally proved by Faltings, who gave two entirely different proofs [F1, F2]. Like Mordell's proof of the finite generation of $E(\mathbb{Q})$ for an elliptic curve E, both of Faltings' proofs are "ineffective": Mordell's proof yields an upper bound on the rank, and either of Faltings' proofs yields an upper bound on the number of rational points, but in general there may be no way to find a list of points and prove that it accounts for all the rational points on the curve. While much more is known now than at the time of Mordell's or even Faltings' proof, the general problems of making those theorems effective remain open.

A final note on Mordell's and Faltings' theorems: while they share the mystery of ineffectivity, the proofs are of quite a different flavor. Mordell's proof for elliptic curves can be traced back to Fermat's proof of the case $n=4$ of FLT (showing in effect that the elliptic curves $Y^2=X^4\pm1$ associated to the Diophantine equations $x^4\pm y^4=z^2$ have rank zero), and can be regarded as the culmination of Fermat's work in this direction. On the other hand, Faltings' proofs, together with the proof of FLT by Wiles and Taylor [Wil, TW], depend heavily on some of the most abstract and difficult results and techniques of late twentieth-century number theory; it would take an expository paper at least as long as this one to even give a sense of these methods to a reader not already acquainted with them.

6.3 The Darmon-Granville theorem: $x^p+y^q=z^r$

Another natural way to generalize the Fermat equation is to allow different exponents, changing $x^n+y^n=z^n$ to $x^p+y^q=z^r$. Here p,q,r are fixed positive integers that are not necessarily equal, and x,y,z are integer unknowns. Solving this equation is equivalent to solving $A+B=C$ under the

[6] While the chord-and-tangent method has been known at least since the time of Fermat, the construction of an abelian group law from it is not obvious. See [Si1, Chapter III, §2] for the details.

[7] At this point it is almost obligatory for an expository paper to cite the fact that an algebraic curve of genus g is one whose graph over $\mathbb{C}$ is an orientable surface with g holes; if nothing else, that is one indication that g measures the curve's complexity.

condition that A be a p-th power, B be a q-th power, and C be a r-th power. The Fermat equation with exponent n is the special case $p = q = r = n$. Applying our heuristic to general (x, y, z), we find that if A, B, C are random integers with $\max(|A|, |B|, |C|) \in (N/2, N]$ then they are respectively p-th, q-th, and r-th powers with probability asymptotically proportional to $N^{((1/p)-1)+((1/q)-1)+((1/r)-1)}$, and thus that of the roughly N^2 solutions of $A + B = C$ in that range we might expect about

$$N^{((1/p)-1)+((1/q)-1)+((1/r)-1)}N^2 = N^{((1/p)+(1/q)+(1/r)-1)}$$

to yield solutions of $x^p + y^q = z^r$. As before, the same analysis applies (to the extent we believe it) to the equation

$$A_0x^p + B_0y^q = C_0z^r \tag{6.4}$$

for fixed nonzero A_0, B_0, C_0. This leads us to introduce

$$\delta = \delta(p, q, r) := 1 - \frac{1}{p} - \frac{1}{q} - \frac{1}{r}. \tag{6.5}$$

Our expected number of solutions with $\max(|A|, |B|, |C|) \in (N/2, N]$ is now roughly $N^{-\delta}$, and as before we vary N and expect the solutions to be plentiful, sparse, or bounded according as $\delta < 0$, $\delta = 0$, or $\delta > 0$. The corresponding values of (p, q, r) are as follows.

Exercise 6.3.1. We have $\delta(p, q, r) < 0$ if and only if one of the following conditions holds: the smallest of p, q, r equals 1; the two smallest of p, q, r both equal 2; or (p, q, r) is a permutation of $(2, 3, 3)$, $(2, 3, 4)$, or $(2, 3, 5)$. In this case, if $\min(p, q, r) = 2$ then $1/\delta$ is a negative integer. We have $\delta(p, q, r) = 0$ if and only if (p, q, r) is a permutation of $(3, 3, 3)$, $(2, 4, 4)$, or $(2, 3, 6)$. Otherwise $\delta(p, q, r) \geq 1/42$, with equality if and only if (p, q, r) is a permutation of $(2, 3, 7)$.

The new borderline cases $(2, 4, 4)$ and $(2, 3, 6)$ again yield elliptic curves, with equations $Y^2 = X^4 \pm 1$ and $Y^2 = X^3 \pm 1$ in the simplest case $A_0 = B_0 = C_0 = 1$. It so happens that again each of these elliptic curves has rank zero, and thus only finitely many rational points. For $Y^2 = X^4 \pm 1$ the only rational points not at infinity are obvious ones with $XY = 0$; this is equivalent to Fermat's result that there are no solutions of $x^4 \pm y^4 = z^2$ in nonzero integers. For $Y^2 = X^3 \pm 1$ there is one additional solution[8] $3^2 = 2^3 + 1$, giving rise to a single set of equivalent solutions of $x^2 + y^3 = z^6$ in nonzero integers, namely $(x, y) = (3z^3, -2z^2)$ for nonzero $z \in \mathbb{Z}$. For general A_0, B_0, C_0 there may be infinitely many such equivalence classes, but again their minimal representatives will be quite sparse, with the number of representatives in the range $\max(|A|, |B|, |C|) \leq N$ growing only as a multiple of $(\log(N)^{\rho/2})$ (where as before ρ is the rank of the corresponding elliptic curve).

But for general p, q, r our prediction can be very wide of the mark: there are cases where $\delta > 0$ but solutions are plentiful. For example, the equation $x^3 + y^4 = z^5$ has the solution

$$(x, y, z) = (209952, 11664, 1944) = (2^5 3^8, 2^4 3^6, 2^3 3^5), \tag{6.6}$$

with (A, B, C) proportional to $(1, 2, 3)$ — and indeed every integer solution of $A + B = C$ is proportional to (x^3, y^4, z^5) for some (and thus for infinitely many) integer triples (x, y, z). More generally we have:

Exercise 6.3.2. Suppose the natural numbers p, q, r are pairwise relatively prime, and A_0, B_0, C_0 are any nonzero integers. Then every integer solution of $A+B = C$ is proportional to (A_0x^p, B_0y^q, C_0z^r) for some (and thus for infinitely many) integer triples (x, y, z), and given the initial A, B, C (not all zero) the number of such (x, y, z) with $\max(|A_0x^p|, |B_0y^q|, |C_0z^r|) \leq N$ is asymptotically proportional to $N^{1/(pqr)}$ as $N \to \infty$. Moreover there are triples (p, q, r) of relatively prime numbers for which δ is arbitrarily close to 1.

[8]The elliptic curve $Y^2 = X^3 + 1$ still has rank zero, but with six rational points: one at infinity, one with $X = -1$, and two each with $X = 0$ and $X = 2$. The reader can check that no further points are obtained by intersecting the curve with the tangent line at any of these points, or the line through any two of them. For instance, $(X, Y) = (2, 3)$ is the third point of intersection of $Y^2 = X^3 + 1$ with the line $Y = X + 1$ through the obvious points $(-1, 0)$ and $(0, 1)$.

The exponent $1/(pqr)$, though usually small, is positive for all p, q, r; hence if p, q, r are pairwise relatively prime our equation $A_0x^p + B_0y^q = C_0z^r$ has "plentiful solutions" by our standards, even when the value of δ is almost as positive as it can be. This seems to utterly demolish our heuristic, which suggests that when $\delta > 0$ there should be only finitely many solutions, and moreover that this tendency should be more pronounced the larger δ gets. But even in favorable cases like the "twisted Fermat curves" $A_0x^n + B_0y^n = C_0z^n$ our heuristic holds only for primitive solutions, those with x, y, z pairwise relatively prime. Indeed we should not expect the heuristic to hold when x and y have a large common factor, say d, because then $A = A_0x^n$ and $B = B_0y^n$ are both multiples of d^n, which makes $A + B$ much likelier to be of the form C_0z^n than a random number of the same size. Our construction of plentiful solutions such as (6.6) likewise exploits large common factors. We thus restrict attention to solutions with $(A, B, C) = (A_0x^p, B_0y^q, C_0z^r)$ *relatively prime*.[9,10] In this case our heuristic agrees precisely with the remarkable theorem of Darmon and Granville (1995):

Theorem 1. *[DG]: Let p, q, r be natural numbers such that $\delta(p, q, r) > 0$, and let A_0, B_0, C_0 be any nonzero integers. Then there are finitely many triples (x, y, z) of integers with $\gcd(x, y, z) = 1$ satisfying the equation (6.4).*

As with FLT and Faltings' theorem, the proof is alas much too advanced for us to be able to even outline the main ingredients here — though we do note that one key step is an application of Faltings' theorem itself!

Exercise 6.3.3. The Darmon-Granville theorem may seem a bit stronger than what we suggested, because (A, B, C) might still have a common factor coming from the coefficients A_0, B_0, C_0. But given those coefficients there are only finitely many possible values of $d := \gcd(A, B, C)$. Use this to show that there are also only finitely many equations $A_1x_1^p + B_1y_1^q = C_1z_1^r$ whose integer solutions satisfying $\gcd(A_1x_1^p, B_1y_1^q, C_1z_1^r) = 1$ account for all solutions of (6.4) with $\gcd(x, y, z) = 1$. Therefore if we knew Darmon-Granville only under the more restrictive hypothesis that A_0x^p, B_0y^q, C_0z^r be relatively prime, we could deduce the result in the form quoted above.

Seeing that the Darmon-Granville theorem for equation (6.4) generalizes Faltings' finiteness result for the case $p = q = r$ of twisted Fermat curves, can we also generalize FLT to the special case $A_0 = B_0 = C_0 = 1$ of (6.4), finding all solutions of $x^p + y^q = z^r$ with $\delta(p, q, r) > 0$ and $\gcd(x, y, z) = 1$? Our heuristic analysis suggests that there should be only finitely many such triples (x^p, y^q, z^r), but we have no reason to expect that there should be none at all — and we would not be surprised if some of them are quite large, especially for those choices of (p, q, r) that make δ positive but small. Note that the Darmon-Granville theorem gives finiteness for any particular choice of (p, q, r) but (like Faltings' theorem vis-a-vis FLT) leaves open the possibility of infinitely many solutions with (p, q, r) varying as well.

The full answer is still beyond reach, so we report on the known partial results and conjectures. The simplest example is the identity $1 + 8 = 9$ already noted in connection with $(p, q, r) = (2, 3, 6)$; it yields a solution $1^r + 2^3 = 3^2$ for all r, satisfying $\delta(2, 3, r) > 0$ for all $r > 6$. Computer searches reveal 9 more solutions: $13^2 + 7^3 = 2^9$ with $\delta(2, 3, 9) = 1/18$; two solutions

$$2^5 + 7^2 = 3^4, \quad 3^5 + 11^4 = 122^2 \tag{6.7}$$

with $\{p, q, r\} = \{2, 4, 5\}$ and $\delta = 1/20$; two solutions

$$33^8 + 1549034^2 = 15613^3, \quad 43^8 + 96222^3 = 30042907^2 \tag{6.8}$$

[9] We need not specify *pairwise* relatively prime, because the relation $A + B = C$ forces any factor of two of A, B, C to divide the third.

[10] The failure of our naïve heuristic when A, B, C can have large common factors is related to the failure we noted earlier for a singular cubic curve. Here the *surface* $A_0x^p + B_0y^q = C_0z^r$ is highly singular at the origin, and a solution with A, B, C all divisible by a high power of p yields a point (x, y, z) on that surface that is close to that singularity in the p-adic metric.

with $\{p,q,r\} = \{2,3,8\}$ and $\delta = 1/24$; and four solutions

$$\begin{gathered} 2^7 + 17^3 = 71^2, \quad 17^7 + 76271^3 = 21063928^2, \\ 1414^3 + 2213459^2 = 65^7, \quad 9262^3 + 15312283^2 = 113^7 \end{gathered} \tag{6.9}$$

with $\{p,q,r\} = \{2,3,7\}$ and the minimal δ value of $1/42$. These computations are reported in [DG], with the discovery of the five large solutions credited to Beukers and Zagier. This list is conjectured to be complete, based both on further computer searches that revealed no other solutions and on various partial results that prove special cases of the conjecture. In particular it would follow from this conjecture (plus FLT for $n = 3$) that $x^p + y^q = z^r$ has no solution in integers $p,q,r \geq 3$ and relatively prime integers x,y,z; this is the **Tijdeman-Zagier conjecture**, for whose solution Andrew Beal offers a $50,000 prize [Mau].

The most recent of the partial results in the direction of the conjecture that there are no further solutions with $\delta(p,q,r) > 0$ is [PSS], a *tour de force* proving that there are no further solutions for $\{p,q,r\} = \{2,3,7\}$. This paper also gives an extensive list (Table 1 at the end of the Introduction) of triples (p,q,r) for which the corresponding result had been proved earlier, including the triples with $\{p,q,r\} = \{2,4,5\}$ and $\{2,3,8\}$ seen in the other known solutions (6.7, 6.8). Another special case is Catalan's conjecture that 8 and 9 are the only consecutive powers of integers, recently proved by Mihăilescu [Mi]; this shows that there are no other solutions with $x = 1$. The proofs of these partial results call on a vast range of number-theoretical techniques, including classical methods of elementary, algebraic, and analytic number theory, Galois representations and modularity as in the proof of FLT, and algebraic geometry of curves. This huge theoretical arsenal is complemented by sophisticated computational and algorithmic tools that are often essential for carrying out algebraic manipulations or for completing a proof that has been reduced to a finite but nontrivial calculation.

What about $\delta(p,q,r) < 0$, when we expect that the number of relatively prime solutions of (6.4) with $\max(|A|,|B|,|C|) \leq N$ can grow as a multiple of $N^{-\delta}$ as $N \to \infty$? We easily dispose of the case where at least one of p,q,r is 1, when we can simply solve (6.4) for the corresponding variable x, y, or z in terms of the other two. So we assume that each of p,q,r is at least 2. In Exercise 6.3.1, we saw that then $-\delta = 1/d$ for some integer $d > 0$. There are choices of the coefficients A_0, B_0, C_0 for which (6.4) has no solutions at all — we already saw the examples $x^2 + y^2 + z^2 = 0$ and $x^2 + y^2 = 3z^2$ with $p = q = r = 2$. But if we assume that there is at least one solution of (6.4) in relatively prime integers then Beukers showed [Beu] that the $N^{1/d}$ guess is correct. Moreover, for each A_0, B_0, C_0 there are finitely many polynomial identities in degree $2d$ that together account for all the relatively prime solutions, in the same way that the single identity (6.2) accounts for all Pythagorean triples. (In fact the Pythagorean parametrization illustrates the special case $A_0 = B_0 = C_0 = 1$, $p = q = r = 2$ of Beukers' result; note that here $\delta = -1/2$ and both sides of the identity are polynomials of degree 4.)

Unlike the Faltings and Darmon-Granville finiteness results, Beukers' is effective: at least in principle one can find all the parametrizations by carrying out a computation whose length is bounded by an explicit function of p,q,r,A_0,B_0,C_0. Doing this in practice still requires some work. For the three exceptional cases where only one of p,q,r equals 2, this work was recently completed by Edwards [Ed]. In particular he gave for the first time the complete solution for $\{p,q,r\} = \{2,3,5\}$ in the case $A_0 = B_0 = C_0 = 1$. There are 27 inequivalent identities, of which the simplest (which Beukers had already obtained) is $X(t)^2 + Y(t)^3 = Z(t)^5$ where

$$\begin{aligned} X(t) &= (t^{10} + 12^4)\,(t^{20} - 12^2 522\,t^{15} - 12^4 10006\,t^{10} + 12^6 522\,t^5 + 12^8), \\ Y(t) &= -t^{20} - 12^2 228\,t^{15} - 12^4 494\,t^{10} + 12^6 228\,t^5 - 12^8, \\ Z(t) &= 12(-t^{11} + 12^2 11\,t^6 + 12^4\,t). \end{aligned} \tag{6.10}$$

For any $m,n \in \mathbb{Z}$ we recover an integer solution of $x^2 + y^3 = z^5$ by taking $x = n^{30}X(m/n)$, $y = n^{20}Y(m/n)$, and $z = n^{12}Z(m/n)$, and these x,y,z are relatively prime if and only if

$$\gcd(m,6n) = 1 \quad \text{and} \quad m \not\equiv 2n \bmod 5. \tag{6.11}$$

For example, $m = n = 1$ yields $36934790165857^2 + 240546239^3 = 267828^5$. To make it such that $\max(|x^2|, |y^3|, |z^5|)$ less than N it is enough to make both $|m|$ and $|n|$ less than some multiple of $N^{1/60}$; the number of such (m, n) satisfying (6.11) is asymptotically proportional to $N^{1/30} = N^{-\delta(2,3,5)}$ as expected.

We conclude this section with another scenic detour: a view of two surprisingly pertinent alternative descriptions of the triples (p, q, r) of integers greater than 1 for which $\delta(p, q, r) < 0$. First, p, q, r satisfy this condition if and only if there exists a spherical triangle Δ with angles π/p, π/q, π/r on the unit sphere Σ, in which case the triangle has area $\pi \cdot (-\delta)$. Second, we have $\delta(p, q, r) < 0$ if and only if the group $\Gamma = \Gamma_{p,q,r}$ with the presentation

$$\Gamma_{p,q,r} := \langle \alpha, \beta, \gamma \mid \alpha^p = \beta^q = \gamma^r = \alpha\beta\gamma = 1 \rangle \tag{6.12}$$

is finite, in which case it has $2d$ elements, where $d = -1/\delta$ as before. The first equivalence follows from the well-known fact that the sum of the angles of Δ exceeds π by an amount equal to the area of Δ. In this case we can take the generators α, β, γ of Γ to be rotations about the vertices of Δ through angles $2\pi/p$, $2\pi/q$, $2\pi/r$, or equivalently the products of pairs of reflections in the edges of Δ. If we identify Σ with the Riemann sphere $\mathbb{CP}^1$ and let t be a complex coordinate on Σ then Γ becomes a finite group of automorphisms of $\mathbb{CP}^1$, which is to say a finite group of fractional linear transformations $t \mapsto (at + b)/(a't + b')$. Then for each of our identities $X(t)^p + Y(t)^q = Z(t)^r$ in degree $2d$ the ratios X^p/Z^r, Y^q/Z^r, etc. are invariant under Γ for a suitable choice of spherical triangle Δ! Moreover, by Galois theory any such ratio T actually *generates* the field of Γ-invariant rational functions of t; that is, $\mathbb{C}(T) = (\mathbb{C}(t))^\Gamma$. For example, when $p = q = r = 2$ our Pythagorean parametrization (6.2) yields functions such as $(t^2 - 1)^2/(2t)^2$ and $(t^2 + 1)^2/(2t)^2$ invariant under the 4-element group isomorphic with $\Gamma_{2,2,2}$ and generated by $t \leftrightarrow -t$ and $t \leftrightarrow 1/t$. For $(p, q, r) = (2, 3, 5)$, we have $\Gamma \cong A_5$, the group of rotational symmetries of a regular icosahedron inscribed in Σ, and the roots of the polynomials[11] X, Y, Z of (6.10) are precisely the 30 edge centers, 20 face centers, and 12 vertices of that icosahedron!

When $\delta(p, q, r) = 0$ or $\delta(p, q, r) > 0$ the triangle Δ is respectively planar or hyperbolic rather than spherical, and the group $\Gamma = \Gamma_{p,q,r}$ generated by pairs of reflections in its edges is no longer finite. But Γ is still intimately connected with $x^p + y^q = z^r$ via automorphisms of Riemann surfaces. When $\delta(p, q, r) = 0$, we can regard Γ as a group of affine linear transformations $t \mapsto at + b$ of $\mathbb{C}$; its finite-index subgroup of translations (with $a = 1$) is then a lattice, and the quotient of $\mathbb{C}$ by this lattice is the elliptic curve we obtained from $x^p + y^q = z^r$. When $\delta = \delta(p, q, r)$ is positive, Δ is a hyperbolic triangle of area $\pi\delta$ and Γ is a discrete group of isometries of the hyperbolic plane $\mathcal{H}$; the quotient $\mathcal{H}/\Gamma$ can be identified with $\mathbb{CP}^1$ via a Γ-invariant meromorphic function on $\mathcal{H}$ analogous to the functions T of the previous paragraph, and quotients of $\mathcal{H}$ by subgroups of finite index in Γ yield finite extensions of $\mathbb{C}(T)$ that are used in the proof of the Darmon-Granville theorem and in the solution of some special cases such as $x^2 + y^3 = z^7$.

6.4 The ABC conjecture: $A + B = C$

Masser and Oesterlé noted that a solution of the Fermat equation, or of a natural generalization such as the equation (6.4) addressed by Darmon and Granville, yields relatively prime numbers A, B, C (such as x^n, y^n, z^n for a primitive Fermat solution) such that $A + B = C$ and each of A, B, C has many repeated prime factors. This inspired them to guess a vastly more general constraint on repeated prime factors in $A, B, A + B$ for coprime integers A, B, and to formulate a precise conjecture on the nature of this constraint, now known as the **ABC conjecture**. This conjecture is stated in terms of an arithmetic function called (for reasons whose explanation would take us too far afield here) the "conductor", defined as follows:

[11] Note that X, Y, Z are regarded as homogeneous polynomials of degrees 30, 20, and 12 respectively, so we count $t = \infty$ among the roots of Z. The other roots of Z are 0 and the ten values of $\varrho\phi$ where ϕ is $(1 \pm \sqrt{5})/2$ (the golden ratio or its algebraic conjugate) and ϱ is one of the five fifth roots of 12^2 in $\mathbb{C}$.

Definition 2. The **conductor** $N(D)$ of a nonzero integer D is the product of the (positive) primes dividing D, counted *without* multiplicity. Equivalently, $N(D)$ is the largest squarefree factor of N.

Example 6.4.1. $N(D_1D_2) \leq N(D_1)N(D_2)$ for all nonzero integers D_1, D_2, with equality if and only if they are relatively prime; $N(D^n) = N(D)$ for all nonzero integers D and $n \geq 1$. The following brief table gives $N(D)$ for $24 \leq D \leq 32$:

D	24	25	26	27	28	29	30	31	32
$N(D)$	6	5	26	3	14	29	30	31	2

The size of the integer $|D|/N(D)$ should be regarded a measure of how far D is from being squarefree, that is, of how rich D is in repeated prime factors.

Conjecture 3. *(Masser-Oesterlé [Oe]): For every real $\epsilon > 0$ there exists $c_\epsilon > 0$ such that*

$$N(ABC) > c_\epsilon C^{1-\epsilon} \tag{6.13}$$

holds for all relatively prime natural numbers A, B, C such that $A + B = C$; equivalently, for every real $\epsilon > 0$ there exists $c_\epsilon > 0$ such that

$$N(ABC) > c_\epsilon \max(|A|, |B|, |C|)^{1-\epsilon} \tag{6.14}$$

holds for all relatively prime integers A, B, C such that $\pm A \pm B \pm C = 0$.

The equivalence is elementary, and the more symmetrical form $\pm A \pm B \pm C = 0$ will let us avoid repeating some arguments twice or thrice according to the signs of A, B, C.

In the following exercises, we detail how the ABC conjecture implies "asymptotic FLT" (that is, FLT for sufficiently large n) as well as its generalizations by Darmon-Granville and Tijdeman-Zagier, and then give an equivalent formulation in terms of the "ABC exponent", and explain why the ϵ in (6.13,6.14) cannot be removed.

Exercise 6.4.1. The ABC conjecture applied to $(A, B, C) = (A_0x^p, B_0y^q, C_0z^r)$ implies the Darmon-Granville theorem; moreover, for any p, q, r such that $\delta = \delta(p, q, r) > 0$ and any positive $\epsilon < \delta$, the inequality (6.13) with an explicit value of c_ϵ yields an explicit upper bound on relatively prime integers x, y, z such that $A_0x^p + B_0y^q = C_0z^r$.

Exercise 6.4.2. The ABC conjecture implies the Tijdeman-Zagier conjecture with at most finitely many exceptions; moreover, for any positive $\epsilon < 1/12$ the inequality (6.13) with an explicit value of c_ϵ yields an explicit upper bound on x^p, y^q, z^r in any counterexample to the conjecture.[12]

Exercise 6.4.3. The ABC conjecture for *any* $\epsilon < 1$ implies that Fermat's Last Theorem holds for all but finitely many exponents n. Again, an explicit value of c_ϵ yields an explicit n_0 such that FLT holds for all $n \geq n_0$.

Exercise 6.4.4. The ABC conjecture for *any* $\epsilon < 1$ implies that any finitely generated multiplicative subgroup G of $\mathbb{Q}^*$ contains only finitely many solutions (s, s') of $s + s' = 1$. [Choose generators for G, and let S be the set of primes that divide the numerator or denominator of at least one generator; then $s + s' = 1$ yields $A + B = C$ with $N(ABC) \mid \prod_{p \in S} p$.]

Remark. For this problem, as with the first exercise in this list, the finitude of solutions is already a theorem, without assuming ABC or any other unproved conjecture. Better yet, explicit upper bounds have been given on C as a function of $N(ABC)$ — whereas no such bound is known for the Darmon-Granville theorem without an ABC hypothesis. Still, the proved bounds are much worse than what would follow from (6.13); see below.

[12] The bound $1/12$ can be raised to $1/6$ because Bruin showed [Br] that there are no solutions of $x^3+y^3 = z^4$ or $x^3+y^3 = z^5$ in relatively prime integers.

Exercise 6.4.5. For relatively prime natural numbers A, B, C such that $A + B = C$, define the **ABC exponent** $\theta(A, B, C)$ by

$$\theta(A, B, C) := (\log C)/(\log N(ABC))$$

(so that $C = N(ABC)^{\theta(A,B,C)}$); for example $\theta(1, 8, 9) = \log 9/\log 6 = 1.226+$. Set

$$\Theta := \limsup_{(A,B,C)} \theta(A, B, C),$$

the limsup running over all triples $(A, B, A + B)$ of natural numbers. Then the ABC conjecture is equivalent to $\Theta \leq 1$. In fact the ABC conjecture is equivalent to $\Theta = 1$, because it is elementary that $\Theta \geq 1$ (for instance we may take $(A, B) = (1, 2^r - 1)$ with $r \to \infty$).

Remark. If it is true that $\limsup \theta(A, B, C) = 1$ then the convergence must be very slow: it is known that there are infinitely many examples of $\theta(A, B, C) > 1 + c/\sqrt{\log C}$ for some universal constant $c > 0$; and it is expected, based on probabilistic heuristics such as applied earlier to $A_0x^p + B_0y^q = C_0z^r$, that in fact $\theta(A, B, C) - 1 > (\log C)^{-\vartheta}$ holds infinitely often for all $\vartheta > 1/3$, but only finitely often for each $\vartheta < 1/3$. In particular, the ABC conjecture is consistent with those heuristics. The largest numerical value known for $\theta(A, B, C)$ is $1.6299+$, for $2 + 3^{10}109 = 23^5$ (found by Eric Reyssat in 1987). See [Ni] for other large $\theta(A, B, C)$.

Exercise 6.4.6. The inequality (6.13) cannot hold for $\epsilon = 0$ and any positive value of c_0. (One way to prove this is to find for each $\alpha > 0$ a natural number r such that $3^\alpha | 2^r - 1$.)

The ABC conjecture, like FLT, is formulated over $\mathbb{Z}$ but has an equivalent statement over $\mathbb{Q}$ obtained by considering ratios of the variables. If $A + B = C$, consider $F = A/C$, so $1 - F = B/C$. Both fractions are in lowest terms because A, B, C are assumed relatively prime. The conductor $N(A)$ is the product of the primes p such that $F \equiv 0 \bmod p$, and likewise $N(B)$ is the product of the primes p such that $F \equiv 1 \bmod p$. As for $N(C)$, that is the product of primes p for which $F \bmod p$ cannot be found in $\mathbb{Z}/p\mathbb{Z}$ because the denominator C vanishes mod p. Since in this case $p \nmid A$, we say that these are the primes such that "$F \equiv \infty \bmod p$". Hence $N(ABC)$, the LHS of the ABC conjecture (6.13), is the product of primes p such that $F \bmod p$ is one of $0, 1, \infty$. The RHS is $c_\epsilon C^{1-\epsilon}$, in which C is simply the denominator of F. This assumes that A, B, C are positive, that is, that $0 < F < 1$; in the general case we replaced C by $\max(|A|, |B|, |C|)$ (see (6.14)), so now we replace the denominator of F by the **height** $h(F)$. By definition, the height of a rational number m/n with $\gcd(m, n) = 1$ is $\max(|m|, |n|)$. This need not exactly equal $\max(|A|, |B|, |C|)$, but is within a factor of 2, which can be accommodated by changing the constant c_ϵ of (6.14). Thus the ABC conjecture is equivalent to the assertion that for every $\epsilon > 0$ there exists $c_\epsilon > 0$ such that, for all $F \in \mathbb{Q}$, the product of the primes at which F reduces to 0, 1, or ∞ is at least $c_\epsilon h(F)^{1-\epsilon}$.

Geometrically, the reduction of FLT to ABC in Exercise 6.4.3 amounts to applying the ABC conjecture to the value of the rational function $F = (x/z)^n = X^n$ on the n-th Fermat curve. This succeeds because F and $1 - F$ have multiple poles and zeros (some defined only over an algebraic closure $\bar{\mathbb{Q}}$) — that is, the preimages of $0, 1, \infty$ under F have large multiplicities, which makes the total number of preimages counted *without* multiplicity small compared to the degree of F as a rational function on the curve. It turns out that here the degree is n^2, and the number of preimages is $3n$, which is less than n^2 once $n > 3$, and indeed less than δn^2 once $n > 3/\delta$. When we try to generalize this argument to rational points on a general algebraic curve $\mathcal{X}$, we find that it is rare for there to be a rational function F on $\mathcal{X}$ whose degree exceeds the size of $F^{-1}(\{0, 1, \infty\})$ by a large factor, so we cannot usually expect to deduce Mordell's conjecture (finiteness of rational points) for $\mathcal{X}$ from an ABC inequality with ϵ near 1. But Belyi [Bel] shows how to construct a function F satisfying $\deg(F) > \#\left(F^{-1}(\{0, 1, \infty\})\right)$ whenever $\mathcal{X}$ is a curve of genus at least 2 defined by an equation with coefficients in $\bar{\mathbb{Q}}$, and then Mordell's conjecture follows from ABC with ϵ sufficiently small [El1]. Recall that Faltings already proved this conjecture twice without any unproved hypothesis, but the proofs are ineffective; the argument in [El1]

shows that the ABC conjecture with effective constants c_ϵ would yield a completely effective finiteness result for rational points on $\mathcal{X}$.

Many other consequences of the ABC conjecture are known, ranging from elementary special cases (there are only finitely many integers N such as $N = 4, 5, 7$ for which $N! + 1$ is a perfect square) to applications that give unexpected connections with other problems in number theory. A striking example is Silverman's application to Wieferich primes, that is, primes p for which $2^{p-1} \equiv 1 \bmod p^2$, such as 1093 and 3511. (Note that the congruence always holds mod p by Fermat's little theorem. In 1909 Wieferich proved [Wie] that a FLT counterexample $x^p + y^p = z^p$ with $p \nmid xyz$ for some prime p would imply $2^{p-1} \equiv 1 \bmod p^2$.) Such primes are expected to be very rare; indeed none is known other than 1093 and 3511, and any further such prime must exceed $1.25 \cdot 10^{15}$ according to computations reported by Richard McIntosh (`http://www.loria.fr/~zimmerma/records/Wieferich.status`). But it is not even known that the set of *non*-Wieferich primes is infinite! Silverman [Si2] proves the infinitude of non-Wieferich primes under the hypothesis of the ABC conjecture, and shows further that this conjecture implies that for every integer $\alpha \neq 0, \pm 1$ there exist constants c_α, x_α such that for all $x > x_\alpha$ there are at least $c_\alpha \log x_\alpha$ primes $p < x$ satisfying $\alpha^{p-1} \not\equiv 1 \bmod p^2$.

Unfortunately a proof of the ABC conjecture still seems a very distant prospect; it is even much too hard to prove the existence of any $\epsilon < 1$ for which the inequality (6.13) holds for some $c_\epsilon > 0$. To show just how far we are, consider the situation suggested by Exercise 6.4.4: we know $N = N(ABC)$, and want all possible (A, B, C). Let S be the set of primes dividing N. Then the inequality (6.13) for any $\epsilon < 1$ gives an upper bound on solutions of $A + B = C$ in relatively prime integers all of whose prime factors are contained in S. (This is often called the "S-unit equation", because it is equivalent to solving $a + b = 1$ in rational numbers $(a, b) = (A/C, B/C)$ that are units in the ring $\mathbb{Z}[1/N]$ obtained from $\mathbb{Z}$ by inverting all the primes in S.) In particular, there should be only a finite number of solutions. This result is known [La1], but already far from trivial. It was not much harder to give an explicit bound on the *number* of solutions [LM], and by now there are bounds depending only on the size of S, as in [Ev]. But that still gives no control over the *size* of the largest solution, which is what the ABC conjecture addresses. Stewart and Tijdeman gave such a bound in [ST], and the bound was recently improved by Stewart and Yu [SY]. But even the best bounds remain exponential: the *logarithm* of C is only known to be bounded by a multiple of $N^{1/3}(\log(N))^3$. Even these results can be useful; for instance the Stewart-Tijdeman bound $\log C = O(N^{15})$ is already enough to compute in practice the full solution of the S-unit equation when S is not too large (see for instance [dW]). But the known results are still very weak compared with the inequalities that the ABC conjecture predicts and that we need for applications such as the Tijdeman-Zagier conjecture and explicit upper bounds in the Darmon-Granville theorem.

6.5 Mason's theorem: $\mathsf{A}(t) + \mathsf{B}(t) = \mathsf{C}(t)$

A curious feature of the ABC conjecture is that not only does it seem very hard to prove but it is not at all obvious how one might try to *dis*prove the conjecture. If FLT were false, a single counterexample would expose the falsity; likewise for the Catalan and Tijdeman-Zagier conjectures, or the Riemann hypothesis and its variants. But there can be no single counterexample for the ABC conjecture, even for a specific value of ϵ, because the inequality (6.13) can accommodate any given triple (A, B, C) by simply decreasing c_ϵ. Likewise for the formulation of the conjecture in terms of ABC exponents $\theta(A, B, C)$: a single example may break the record for the maximal θ, but has no bearing on Θ which is defined as a lim sup of $\theta(A, B, C)$. Proving that the conjecture is false would require the existence of an *infinite family* of (A, B, C)'s whose ABC exponents approach a limit greater than 1 (or approach ∞), just as we had to construct an infinite family such as $\{(1, 2^r - 1, 2^r)\}_{r=1}^{\infty}$ to prove $\Theta \geq 1$.

The most natural families to try arise from identities $\mathsf{A}(t) + \mathsf{B}(t) = \mathsf{C}(t)$ relating polynomials $\mathsf{A}, \mathsf{B}, \mathsf{C} \in \mathbb{Z}[t]$, not all constant. Recall that we already used such polynomials to construct infinitely many Pythagorean triples, or relatively prime solutions of $x^2 + y^3 = z^5$; in effect we solved these

Diophantine equations in $\mathbb{Z}[t]$, then specialized to $t \in \mathbb{Q}$ and multiplied by powers of the denominator of t to recover integer solutions. Similarly, from polynomials $\mathsf{A}(t), \mathsf{B}(t), \mathsf{C}(t)$ satisfying $\mathsf{A}+\mathsf{B}=\mathsf{C}$ for which $D := \max(\deg(\mathsf{A}), \deg(\mathsf{B}), \deg(\mathsf{C}))$ is positive we get a family of integer solutions A, B, C as follows: for any pair (m, n) of relatively prime integers we take

$$(A, B, C) = n^D(\mathsf{A}(m/n), \mathsf{B}(m/n), \mathsf{C}(m/n)). \tag{6.15}$$

Thus A, B, C are homogeneous polynomials of degree D in (m, n). If $\mathsf{A}, \mathsf{B}, \mathsf{C}$ have repeated factors then so do A, B, C, and with enough repeated factors we can hope to get a sequence with

$$\limsup \theta(A, B, C) > 1.$$

We must assume that $\mathsf{A}(t), \mathsf{B}(t), \mathsf{C}(t)$ are relatively prime as polynomials, else A, B, C will have a common factor for most choices of (m, n). This also means that D is the degree of the quotient $\mathsf{F} = \mathsf{A}/\mathsf{C} \in \mathbb{Q}(t)$ as a rational function of t. Conversely, if the polynomials have no common factors then $\gcd(A, B)$ is bounded above,[13] so dividing each of our triples (A, B, C) of (6.15) by its greatest common divisor yields relatively prime solutions of $A + B = C$ with asymptotically the same ABC exponent as the ratio

$$\frac{\log\max(|A|, |B|, |C|)}{\log N(ABC)} = \frac{\log\max(|A|, |B|, |C|)}{\log(N(A)N(B)N(C))} \tag{6.16}$$

that we would compute if A, B, C were relatively prime.

The numerator of this ratio is easy to estimate: it is $D \log h(m, n) + e$, where

$$D = \max(\deg(\mathsf{A}), \deg(\mathsf{B}), \deg(\mathsf{C}))$$

as above, $h(m, n)$ is the height $|\max(m, n)|$ of (m, n) (or of the rational number m/n as before), and e is an error of bounded absolute value. What of the denominator? Let us try some examples using polynomial identities that we have already encountered. If

$$(\mathsf{A}, \mathsf{B}, \mathsf{C}) = \big((t^2-1)^2, (2t)^2, (t^2+1)^2\big)$$

as in (6.2), then $D = 4$ and we get $(A, B, C) = ((m^2-n^2)^2, (2mn)^2, (m^2+n^2)^2)$ (the squares of the entries of the Pythagorean triple (6.2)), and then $N(ABC)$ is a factor of $(m^2-n^2)2mn(m^2+n^2)$. Hence $N(ABC)$ is bounded above by a multiple of $h(m, n)^6$. We can save two factors of $h(m, n)$ in various ways, for instance by making $(m, n) = (1, 2^r)$ as in Exercise 6.4.5; but that still leaves both numerator and denominator of (6.16) asymptotic to $4 \log h(m, n)$, giving the same lower bound of 1 on Θ that we obtained in that Exercise. Might we do better with the more complicated example $(A, B, C) = (X(t)^2, Y(t)^3, Z(t)^5)$, where X, Y, Z are the polynomials of (6.10)? Now $D = 60$ and A, B, C are respectively a square, a cube, and a fifth power, so $N(ABC)$ is bounded by a multiple of $h(m, n)^{30+20+12} = h(m, n)^{62}$. Again we can save a factor $h(m, n)^2$ thanks to the factor mn of C, but that still brings our bound on $N(ABC)$ only down to a multiple of $h(m, n)^{60} = h(m, n)^D$, and again we fail to improve on $\Theta \geq 1$.

In general, suppose A factors as $\mathsf{A}_0 \prod_i \mathsf{x}_i^{e_i}$ where A_0 is a scalar and the x_i are distinct irreducible polynomials. Let $x_i = n^{\deg \mathsf{x}_i}\mathsf{x}_i(m/n)$. Then $A = n^D\mathsf{A}(m/n) = \mathsf{A}_0 n^{e_\infty} \prod_i x_i^{e_i}$, where $e_\infty := D - \sum_i e_i \deg(\mathsf{x}_i)$ is the multiplicity of n as a factor of the homogeneous polynomial $n^D\mathsf{A}(m/n)$ (which may also be regarded as the "order of vanishing at $t = \infty$" when A is regarded as a polynomial of degree D). Hence $N(A)$ is bounded by a constant multiple of $\prod_i x_i$ or $n\prod_i |x_i|$ according as $e_\infty = 0$ or $e_\infty > 0$. Each $|x_i|$ is in turn bounded by a constant multiple of $(h(m, n))^{\deg \mathsf{x}_i}$, and of course $|n| \leq h(m, n)$. It follows that $N(A) \leq h(m, n)^{\nu_D(A)}$ where $\nu_D(A) = \nu_{D,\infty}(A) + \sum_i \deg \mathsf{x}_i$

[13] By the Euclidean algorithm for polynomials there exist $\mathsf{X}, \mathsf{Y} \in \mathbb{Z}[t]$ such that $\mathsf{AX} - \mathsf{BY} = d$ for some nonzero $d \in \mathbb{Z}$, and then $\gcd(A, B) \mid n^D d$ for all $m, n \in \mathbb{Z}$. Repeating this argument with A, B replaced by the relatively prime polynomials $t^D\mathsf{A}(1/t), t^D\mathsf{B}(1/t)$ yields a nonzero integer d' such that $\gcd(A, B) \mid m^D d'$. Thus if $\gcd(m, n) = 1$ then $\gcd(A, B) \mid dd'$.

and $\nu_{D,\infty}(A) = 0$ or 1 according as $e_\infty = 0$ or $e_\infty > 0$. More succinctly, $\nu_D(A)$ is the number of solutions of $\mathsf{F}(t) = 0$ in $\mathbb{CP}^1$, counted *without multiplicity* (note in particular that $e_\infty > 0$ if and only if $\mathsf{F}(\infty) = 0$). We define $\nu_D(B)$ and $\nu_D(C)$ likewise, and observe that they are the numbers of solutions in $\mathbb{CP}^1$ of $\mathsf{F}(t) = 1$ and $\mathsf{F}(t) = \infty$. Putting these together we find that $N(A, B, C)$ is bounded by a constant multiple of $h(m, n)^\nu$ where $\nu = \nu_D(A) + \nu_D(B) + \nu_D(C)$ is the size of $\mathsf{F}^{-1}(\{0, 1, \infty\})$. Moreover, if at least two points in $\mathsf{F}^{-1}(\{0, 1, \infty\})$ are rational then we can save an extra factor of $h(m, n)^2$ as we did before; in fact we expect to save this factor in any case, because there are about H^2 choices of (m, n) with $h(m, n) \in (H/2, H]$, and it is not too hard to show that in fact this $h(m, n)^2$ saving is available for all nonconstant rational functions F. In other words, we can make the denominator of (6.16) no larger than $(\nu - 2) \log h(m, n) + e'$, where e' is another bounded error.

Combining our estimates and letting $h(m, n) \to \infty$, we find that the polynomial identity $\mathsf{A}+\mathsf{B} = \mathsf{C}$ will yield a disproof the ABC conjecture if $\nu < D + 2$. We have already given several examples of $\nu = D + 2$, and there are many others, some of which are very easy to construct (try $(\mathsf{A}, \mathsf{B}, \mathsf{C}) = (1, t^D - 1, t^D)$ for instance). Might we attain $\nu < D + 2$ if we are just a little more clever, or look harder? This is where Mason's theorem enters:

Theorem 4. *[Mas]: If $\mathsf{F} \in \mathbb{C}(t)$ is a rational function of degree $D > 0$ on $\mathbb{CP}^1$ then $\mathsf{F}^{-1}(\{0, 1, \infty\})$ has cardinality at least $D + 2$.*

This ruins our hope for an easy refutation of the ABC conjecture. Viewed more positively, it is evidence for the truth of the conjecture, and indeed can be viewed as an "ABC theorem" for polynomials or rational functions. To make the comparison explicit, we again take logarithms in the conjectured inequality (6.13) to write it as $\log N(ABC) > (1 - \epsilon) \log C - \log(1/c_\epsilon)$. We saw that for polynomials ν and D play the roles of $\log N(ABC)$ and $\log C$ respectively. Thus Mason's theorem is an even stronger statement, because the troublesome terms $-\epsilon \log C$ and $-\log(1/c_\epsilon)$ in the lower bound for $\log N(ABC)$ have been replaced by the helpful $+2$ in the lower bound on ν.

Moreover, while the ABC conjecture seems intractable at present, Mason's theorem can be proved easily. There are several related routes, all exploiting the idea of detecting multiple roots of a polynomial or rational function using its *derivative* — a tool not available for integers or rational numbers. The route we choose uses the logarithmic derivative, for which it will be convenient to assume that ∞ is not a preimage of 0, 1, or ∞. We ensure this by applying to t a fractional linear transformation that moves all the preimages of $\{0, 1, \infty\}$ away from infinity.

Proof. Fix a number t_0 not in $\mathsf{F}^{-1}(\{0, 1, \infty\})$, and let $\mathsf{F}_1(t) = \mathsf{F}(t_0 + (1/t))$, a rational function also of degree D and with the same number of preimages of $\{0, 1, \infty\}$ as F, none of which are at infinity. Let ν_0, ν_1, ν_∞ be the number of preimages of $0, 1, \infty$ respectively. Let Λ be the logarithmic derivative $\mathsf{F}_1'/\mathsf{F}_1$. Then Λ is not identically zero because F_1 is nonconstant, and Λ has a simple pole (that is, has a denominator with a simple root) at each preimage of 0 or ∞, regardless of its multiplicity. Hence the denominator of Λ has degree $\nu_0 + \nu_\infty$. Any root of $\mathsf{F}_1' - 1$ of multiplicity e is a root of Λ of multiplicity $e - 1$. Summing over the roots, we find the the numerator of Λ has at least $D - \nu_1$ roots counted *with* multiplicity, and therefore has degree at least $D - \nu_1$. But the difference between the denominator's and numerator's degrees is the order of vanishing of Λ at infinity, which is at least 2 (to see this, expand F_1 at infinity as $\sum_{i=0}^{\infty} a_i t^{-i} = a_0 + a_1 t^{-1} + a_2 t^{-2} + a_3 t^{-3} + \cdots$ with $a_0 \neq 0$, and calculate $\mathsf{F}_1' = -a_1 t^{-2} - 2a_2 t^{-3} - 3a_3 t^{-4} - \cdots$). Hence $D - \nu_1 \leq \nu_0 + \nu_\infty - 2$, which is equivalent to the desired inequality $\nu_0 + \nu_1 + \nu_\infty \geq D + 2$. □

Since the numerator of the derivative or the logarithmic derivative of A/C is (up to sign) the Wronskian

$$W_2(\mathsf{A}, \mathsf{C}) = \det \begin{vmatrix} \mathsf{A} & \mathsf{C} \\ \mathsf{A}' & \mathsf{C}' \end{vmatrix} = \mathsf{A}\mathsf{C}' - \mathsf{A}'\mathsf{C},$$

the proof can also be formulated in terms of Wronskians. The key fact that F and $\mathsf{F} - 1$ have the same derivative then corresponds to the identity $W_2(\mathsf{A}, \mathsf{C}) = W_2(\mathsf{A} - \mathsf{C}, \mathsf{C})$, which holds because $W_2(\cdot, \cdot)$ is

bilinear and alternating, and forces $W_2(\mathsf{A}, \mathsf{C})$ to vanish at multiple zeros of B. Also equivalent, though not as transparently so, is the proof obtained by applying the Riemann-Hurwitz formula to F. This approach explains the "+2" in Mason's inequality as the Euler characteristic of $\mathbb{CP}^1$, and generalizes to rational functions F of degree $D > 0$ on other compact Riemann surfaces, for which Mason finds the inequality $\#\left(\mathsf{F}^{-1}(\{0,1,\infty\})\right) \geq D + \chi = D + 2 - 2g$, where g is the genus and χ the Euler characteristic of the surface. This is why the rational functions F constructed by Belyi cannot satisfy $\deg(F) > \#\left(F^{-1}(\{0,1,\infty\})\right)$ unless $g \geq 2$. For an elliptic curve we have $g = 1$, so $\deg(F) = \#F^{-1}(\{0,1,\infty\})$ is possible, and if the elliptic curve has positive rank then its rational points yield another kind of infinite family of (A, B, C) triples with $\limsup \theta(A, B, C) \geq 1$ (such as $(x^3, y^3, 91z^3)$ for primitive solutions of $x^3 + y^3 = 91z^3$); but the points are too sparse for us to prove that the limsup strictly exceeds 1, and again we come just short of a disproof of the ABC conjecture.

6.6 A Putnam problem: minding our P's and Q's

The last problem of the 1956 William Lowell Putnam Mathematical Competition asks [GGK, p.47]:

> The polynomials $P(z)$ and $Q(z)$ with complex coefficients have the same set of numbers for their zeros but possibly different multiplicities. The same is true of the polynomials $P(z) + 1$ and $Q(z) + 1$. Prove that $P(z) \equiv Q(z)$.

As noted in [GGK, p.431], it must be assumed that at least one of P and Q is not constant, else the claim is false. We thus assume $\max(\deg(P), \deg(Q)) > 0$, and by symmetry may take $m = \deg(P) \geq \deg(Q) = n$. The claim is clearly true if P has distinct roots, because then $Q = cP$ for some $c \in \mathbb{C}$, and if λ is any root of $P + 1$ then $0 = Q(\lambda) + 1 = cP(\lambda) + 1 = -c + 1$ implies $c = 1$. Likewise if $P + 1$ has distinct roots. We must then contend with the case that P and $P + 1$ both have multiple roots — and we know already that the derivative $P' = (P + 1)'$ detects multiple roots of either P or $P + 1$. We proceed as in [GGK, p.431–432]. Let $\lambda_1, \ldots, \lambda_r$ be the distinct roots of P (and thus also of Q), and $\mu_1, \ldots, \mu_s$ the distinct roots of $P + 1$ (and thus also of $Q + 1$). By an argument we can now recognize as the special case of Mason's theorem in which F is a polynomial — and which would fail if $m = 0$ were allowed — we have $m - 1 = \deg(P') \geq 2m - r - s$, whence $r + s \geq m + 1$. But each root of P or $P + 1$ is also a root of $P - Q$, a polynomial of degree at most m. Therefore $P - Q$ is the zero polynomial, and we are done.

The corresponding statement for integers instead of polynomials would be that a positive integer n is determined uniquely by the sets (without the multiplicities) of prime factors of n and of $n + 1$, that is, by the conductors $N(n)$ and $N(n+1)$. We might expect that this should be false, because the proof in the polynomial case hinges on an inequality stronger than can be true for integers. Indeed there are infinitely many counterexamples, the smallest with natural numbers being $n = 2$ and $n' = 8$ (this is yet another appearance of $1 + 8 = 9$), which begins the infinite family $\{n, n'\} = \{2^m - 2, 2^m(2^m - 2)\}$ $(m = 2, 3, 4, \ldots)$. Still, such examples seem quite rare; an exhaustive search finds that the only case with $0 < n, n' < 10^8$ not of the form $\{2^m - 2, 2^m(2^m - 2)\}$ is $\{75, 1215\}$ (with $N(75) = N(1215) = 15$ and $N(76) = N(1216) = 38$). When we allow also negative integers, the identity $N(-n) = N(n)$ gives an involution $\{n, n'\} \leftrightarrow \{-1 - n, -1 - n'\}$ on the set of solutions. Modulo this involution, we find one more infinite family $\{2^m + 1, -(2^m + 1)^2\}$ $(m = 1, 2, 3, \ldots)$, and one more sporadic pair in $(-10^8, 10^8)$, namely $\{35, -4375\}$. The infinite families intersect at $\{2, -4, 8\}$ and $\{-3, 3, -9\}$, which may be the only three-element subsets of $\mathbb{Z}$ mapped to a single point under $n \mapsto (N(n), N(n + 1))$.

Might we generalize the Putnam problem to rational functions F? Since a polynomial is just a rational function with $F^{-1}(\{\infty\}) = \{\infty\}$, we might guess that more generally if F and G are non-constant rational functions with complex coefficients that, when considered as maps from the Riemann sphere $\mathbb{CP}^1$ to itself, satisfy $F^{-1}(\{w\}) = G^{-1}(\{w\})$ for each of $w = 0, 1, \infty$, then $F = G$. (In the Putnam problem, F and G would be the polynomials $P + 1$ and $Q + 1$.) Alas this natural guess is false.

An explicit counterexample is[14]

$$F(z) = \frac{(z-1)^3(z+3)}{16z}, \quad G(z) = h(-3/z) = \frac{(z-1)(z+3)^3}{16z^3},$$

with $F(z) - 1 = (z-3)(z+1)^3/16z$ and $G(z) - 1 = (z-3)^3(z+1)/16z^3$. Here F and G are rational functions of degree 4. Is this the smallest possible? It is probably much harder to completely describe all counterexamples, or even to decide whether there any with $\deg(F) \neq \deg(G)$.

6.7 Further problems and results

In number theory most things that can be done in $\mathbb{Q}$ or $\mathbb{Z}$ generalize, with some additional effort, to number fields K (finite-degree field extensions of $\mathbb{Q}$) and their rings O_K of algebraic integers. This is true of the ABC conjecture, which can be naturally formulated over any K or O_K, and has much the same consequences there as we saw over $\mathbb{Q}$ or $\mathbb{Z}$. Much of the extra effort in making this generalization arises because O_K need not have unique factorization, so some solutions in K of $A + B = C$ may not be proportional to any solution in relatively prime elements of O_K. Thus it is more natural to formulate the conjecture in terms of the ratio $F = A/C$, which is invariant under scaling (A, B, C). Briefly, we replace $N(ABC)$ in the LHS of (6.13) or (6.14) by the product of the *norms* of all prime ideals of O_K at which F is congruent to one of $0, 1, \infty$, and in the RHS we take the $(1-\epsilon)$-th power of the height of F, appropriately defined, rather than of C or of $\max(|A|, |B|, |C|)$. See [Vo, p.84] for the details. Mason's theorem still defeats attempts at easy disproofs — recall that the coefficients of the rational function F were allowed to be arbitrary complex numbers.

More subtle is the question of how the constant c_ϵ in the ABC conjecture should depend on K. In the context of Mason's theorem, if we replace $\mathbb{C}(t)$ by a finite-degree extension we get the function field of a compact Riemann surface of some genus g, and then the lower bound $D + 2$ on the size of $\mathsf{F}^{-1}(\{0, 1, \infty\})$ is lowered by $2g$. Granville and Stark [GS] propose an anologous "uniform ABC conjecture", in which the LHS of (6.13) or (6.14) is multiplied by $|\mathrm{disc}(K/\mathbb{Q})|^{1/[K:\mathbb{Q}]}$ and then the constant c_ϵ in the RHS is independent of K. Remarkably, they then show that this uniform ABC conjecture implies the long-standing conjecture that the class number of an imaginary quadratic field $\mathbb{Q}(\sqrt{-d})$ (with $d > 0$ a squarefree integer) is bounded below by a constant multiple of $d^{1/2}/\log d$, and thus that the Dirichlet L-function attached to an odd character has no "Siegel-Landau zero" (a zero s with $1 - s \ll 1/\log(d)$; the nonexistence of such zeros is an important special case of the Riemann Hypothesis for such L-functions). The proof uses special values of modular functions arising from elliptic curves with complex multiplication by the ring of algebraic integers in $\mathbb{Q}(\sqrt{-d})$.

Finally we consider the generalization to more than three variables, to integers satisfying $\pm A \pm B \pm C \pm D = 0$ and beyond. In each case we ask: Given $\max(|A|, |B|, |C|, \ldots)$, how small can the product $N(A)N(B)N(C)\cdots$ get? As before we must assume that the integers have no common factor. With more than three variables, it no longer follows that they are relatively prime in pairs, but we must at least assume that no proper sub-sum of $\pm A \pm B \pm C \pm \cdots$ vanishes, to avoid such trivialities as $2^r + 1 - 2^r - 1 = 0$. It is then known that an upper bound on $N(A)N(B)N(C)\cdots$ implies an upper bound on $\max(|A|, |B|, |C|, \ldots)$, but again this known bound is much too large for our purpose. Even in the special case $A = A_0 w^n$, $B = B_0 x^n$, etc. we have a difficult question: How

[14]The reader who got this far may well wonder where this counterexample comes from. It arises naturally in the theory of elliptic modular functions. For τ in the upper half-plane $\mathcal{H}$, let $\eta(\tau)$ be the Dedekind eta function $e^{\pi i/12}\prod_{n=1}^{\infty}(1-e^{2\pi i n\tau})^{24}$, and define $\lambda(\tau) = 16(\eta_2^2\,\eta_{1/2}/\eta_1^3)^8 = 16q\prod_{n=1}^{\infty}(1+q^{2n})/(1+q^{2n-1})$ where $\eta_k = \eta(k\tau)$ and $q = e^{\pi i\tau}$. Then λ generates the field of modular functions invariant under the ideal hyperbolic triangle group $\Gamma(2)$, and takes the values $0, 1, \infty$ at the cusps of that group. The function F expresses λ in terms of the generator $-3(\eta_3/\eta_1)^{10}(\eta_{1/2}\eta_2/\eta_{3/2}\eta_6)^4$ of the modular functions for $\Gamma(2)\cap\Gamma_0(3)$, and thus gives explicitly the map from the corresponding modular curve to the modular curve $X(2)$ corresponding to $\Gamma_0(2)$. The coordinate λ of $X(2)$ parametrizes elliptic curves $E : Y^2 = X(X-1)(X-\lambda)$ with all their 2-torsion points rational; z parametrizes 3-isogenies $E \to E'$ between pairs of such curves; and the involution $z \leftrightarrow -3/z$ takes the isogeny $E \to E'$ to the dual isogeny $E' \to E$. See [El2].

are the nontrivial primitive solutions of $A_0w^n + B_0x^n + C_0y^n = D_0z^n$ distributed? Our heuristics suggest that solutions should be plentiful for $n < 4$ (if there is a nonzero solution to begin with), sparse for $n = 4$, and bounded for $n > 4$. Likewise for N variables, with critical exponent $n = N$.

Unfortunately this guess is at best close to the truth. Euler already found a polynomial solution for $w^4+x^4 = y^4+z^4$, giving plentiful solutions for that equation, starting with $133^4+134^4 = 59^4+158^4$. There is even a polynomial family of solutions of $w^5 + x^5 = y^5 + z^5$, though sadly not over $\mathbb{Q}$:

$$w, x = 2t \pm (t^2 - 2), \quad y, z = 2t \pm i(t^2 + 2). \tag{6.17}$$

For $n = 6$ one can still obtain infinitely many primitive solutions for some choices of (A_0, B_0, C_0, D_0), using the polynomial identity

$$(t^2 + t - 1)^3 + (t^2 - t - 1)^3 = 2t^6 - 2.$$

Indeed let $(A_0, B_0, C_0, D_0) = (\alpha^3, \beta^3, 2, 2)$. Then if there are infinitely many rational solutions (t, u, v) of

$$t^2 + t - 1 = \alpha u^2, \quad t^2 - t - 1 = \beta v^2 \tag{6.18}$$

then each yields a rational solution $(u, v, 1, t)$ of $A_0w^6 + B_0x^6 + C_0y^6 = D_0z^6$, and thus a primitive integer solution by clearing common factors. Now it can be shown that (6.18) is an elliptic curve, which has positive rank if it has a single rational point with $t \notin \{0, \pm1, \infty\}$. The simplest such (α, β) is $(5, 1)$ with $t = 2$, giving $125 + 1 + 2 = 2 \cdot 2^6$. The next few t values for $(\alpha, \beta) = (5, 1)$ are $-82/19$, $-148402/91339$, and $-10458011042/1213480199$, giving the solutions[15]

$$(31, 19, 89, 82), \quad (5009, 91339, 165031, 148402),$$
$$(4363642319, 1213480199, 10981259039, 10458011042).$$

Note that, unlike the ABC conjecture, our naïve guess for $A_0w^n + B_0x^n + C_0y^n = D_0z^n$ was disproved by polynomial identities. Thus even Mason's theorem has no good analogue here. One can use a 3×3 Wronskian to get an "ABCD theorem", and likewise for more variables, but these inequalities are no longer sharp. For example, if (w, x, y, z) is a nontrivial solution in $\mathbb{C}[t]$ of $w^n + x^n = y^n + z^n$ then one can show that $n < 8$ by counting roots of $W_3(w^n, x^n, y^n)$, but it is not known whether $n = 6$ or $n = 7$ can occur, nor whether all nontrivial solutions for $n = 5$ are equivalent with (6.17).

Can we salvage from our predicament a conjecture that is both plausible and sharp? Lang [La2] suggested that such conjectures should still be true "on a nonempty Zariski-open set", that is, when we exclude variables that satisfy some algebraic condition. This may well be true, though the possibility of an unpredictable exceptional set makes Lang's conjectures even harder to test. As an indication of the power of these conjectures, we conclude by citing one striking application. Recall that Mordell conjectured, and Faltings proved, that an algebraic curve of genus $g > 1$ over $\mathbb{Q}$ has only finitely many rational points. The conjecture and proofs are silent on how the number of points can vary with the curve. But Caporaso, Harris, and Mazur showed [CHM] that Lang's conjectures imply a uniform upper bound $B(g)$, depending only on g, on the number of rational points of any genus-g curve over $\mathbb{Q}$!

References

[Bel] G[ennadii] V[ladimirovich] Belyi: On the Galois extensions of the maximal cyclotomic field (in Russian), *Izv. Akad. Nauk. SSSR* **43** (1979), 267–276.

[Beu] Frits Beukers: The Diophantine Equation $Ax^p + By^q = Cz^r$, *Duke Math. J.* **91** (1998), 61–88.

[Br] Nils Bruin: On powers as sums of two cubes, pages 169–184 in *Algorithmic Number Theory (Leiden, 2000)*, Berlin: Springer, 2000 (Wieb Bosma, ed.; *Lecture Notes in Computer Science* **1838**).

[15] I do not know where this construction originated. I must have noticed it by 1988, because my computer files include a listing of these solutions dated May 1988.

[CHM] Lucia Caporaso, Joe Harris, and Barry Mazur: Uniformity of rational points, *J. Amer. Math. Soc.* **10** (1997) #1, 1–35.

[DG] Henri Darmon and Andrew Granville: On the equations $x^p + y^q = z^r$ and $z^m = f(x, y)$, *Bull. London Math. Soc.* #129 (**27** part 6, Nov.1995), 513–544.

[Ed] Johnny Edwards: A Complete Solution to $X^2 + Y^3 + Z^5 = 0$, *J. f. d. reine u. angew. Math.* **571** (2004), 213–236 (also online at `http://www.math.uu.nl/people/edwards/icosahedron.pdf`).

[El1] Noam D. Elkies: ABC implies Mordell, *International Math. Research Notices* **1991** #7, 99–109 [bound with *Duke Math. J.* **64** (1991)].

[El2] Noam D. Elkies: Wiles minus epsilon implies Fermat, pages 38–40 in *Elliptic Curves, Modular forms, and Fermat's Last Theorem* (J. Coates and S.-T. Yau, eds.; Boston: International Press, 1995; proceedings of the 12/93 conference on elliptic curves and modular forms at the Chinese University of Hong Kong).

[Ev] Jan-Hendrik Evertse: On equations in S-units and the Thue-Mahler equation, *Invent. Math.* **75** (1984), 561–584 (1994).

[F1] Gerd Faltings: Endlichkeitssätze für abelsche Varietäten über Zahlkörpern, *Invent. Math.* **73** (1983), 349–366.

[F2] Gerd Faltings: Diophantine Approximation on Abelian Varieties, *Ann. Math.* (2) **133** (1991), 549–576.

[GGK] Andrew M. Gleason, R.E. Greenwood, and Leon M. Kelly: *The William Lowell Putnam Mathematical Competition — Problems and Solutions: 1938–1964*. Washington, D.C.: Math. Assoc. of America, 1980.

[GS] Andrew Granville and Harold M. Stark: ABC implies no 'Siegel zero' for L-functions of characters with negative discriminant, *Invent. Math.* **139** #3 (2000), 509–523.

[IR] Kenneth Ireland and Michael Rosen, *A Classical Introduction to Modern Number Theory*, 2nd ed. New York: Springer, 1990 (Graduate Texts in Math. **84**).

[La1] Serge Lang: Integral points on curves, *Publ. Math. IHES* **6** (1960), 27–43.

[La2] Serge Lang: Hyperbolic and diophantine analysis, *Bull. Amer. Math. Soc.* **14** #2 (1986), 159–205.

[LM] D[onald] J. Lewis and Kurt Mahler: Representation of integers by binary forms, *Acta Arith.* **6** (1960/61), 333–363.

[Mas] R[ichard] C. Mason: *Diophantine Equations over Function Fields*, London Mat. Soc. Lect. Notes Ser. **96**, Cambridge Univ. Press 1984. See also pp.149–157 in Springer LNM **1068** (1984) [=proceedings of Journées Arithmétiques 1983, Noordwijkerhout].

[Mau] R. Daniel Mauldin: A Generalization of Fermat's Last Theorem: The Beal Conjecture and Prize Problem, *Notices of the Amer. Math. Soc.* **44** #11 (1997), 1436–1437. `http://www.ams.org/notices/199711/beal.pdf`

[Mi] Preda Mihăilescu: Primary Cyclotomic Units and a Proof of Catalan's Conjecture, *J. reine angew. Math.* **572** (2004), 167–195.

[Mo] Louis J. Mordell: On the rational solutions of the indeterminate equations of the third and fourth degrees, *Proc. Cambridge Phil. Soc.* **21** (1922), 179–192.

[Ni] Abderrahmane Nitaj: The ABC Conjecture Home Page. `http://www.math.unicaen.fr/kern1mm/~nitaj/abc.html`

[Oe] Joseph Oesterlé: Nouvelles approches du "théorème" de Fermat, *Sém. Bourbaki* 2/1988, exposé #694.

[PSS] Bjorn Poonen, Edward F. Schaefer, and Michael Stoll: Twists of $X(7)$ and primitive solutions to $x^2+y^3 = z^7$. Preprint, 2005 (online at `http://arxiv.org/math.NT/0508174`).

[Si1] Joseph H. Silverman: *The Arithmetic of Elliptic Curves*. New York: Springer 1986 (GTM **106**).

[Si2] Joseph H. Silverman: Wieferich's criterion and the *abc*-conjecture, *J. Number Theory* **30** #2 (1988), 226–237.

[ST] Cameron L. Stewart and Robert Tijdeman: On the Oesterlé-Masser conjecture, *Monatsh. Math.* **102** (1986), 251–257.

[SY] Cameron L. Stewart and Kunrui Yu: On the abc conjecture, II, *Duke Math. J.* **108** (2001), 169–181.

[Ta] Olga Taussky: Sums of squares, *Amer. Math. Monthly* **77** #8 (Oct.1970), 805–830.

[TW] Richard Taylor and Andrew Wiles: Ring-theoretic properties of certain Hecke algebras, *Ann. Math.* **141** (1995), 553–572.

[Vo] Paul Vojta: *Diophantine Approximations and Value Distribution Theory*. Berlin: Springer 1987 (Lect. Notes Math. **1239**).

[dW] Benne de Weger: *Algorithms for Diophantine equations.* Amsterdam: Centrum voor Wiskunde en Informatica, 1989 (CWI tract **65**).

[Wie] Arthur Wieferich: Zum letzten Fermat'schen Theorem, *J. f. d. reine u. angew. Math.* **136** (1909), 293–302.

[Wil] Andrew Wiles: Modular elliptic curves and Fermat's Last Theorem, *Ann. Math.* **141** (1995), 443–551.