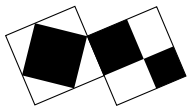
NOTES ON FERMAT'S LAST THEOREM VII



ceNTRe
Macquarie University
NSW 2109 Australia

A. J. van der Poorten

Centre for Number Theory Research, Macquarie University

ellipsis *the omission from a sentence of a word or words which would complete or clarify the construction.*

Today's story is a child's introduction to elliptic functions. Since I'll be covering a few years' coursework in a few pages — my remarks will be elliptical — we had best fasten our seatbelts.

When there is talk of periodic functions one thinks of $\sin 2\pi z$ and $\cos 2\pi z$; I've put in the 2π as a normalisation so that these functions have *primitive* period 1, rather than some random ω, or whatever. The adjective 'primitive' is there to acknowledge that they actually have periods 0, ± 1, ± 2, $\ldots$, but fundamentally the period is 1 as said. Of course we say that f has period ω if $f(z+\omega) = f(z)$ for all z. It is not a very sophisticated remark to observe that the *circular functions* are periodic because $e^{2i\pi z}$ is periodic; after all, they are just its imaginary and real part respectively. But it is not quite trivial to add that in fact *every* periodic function is periodic because it is itself a function of $e^{2i\pi z}$. This manifests itself in practice by reasonable periodic functions having a Fourier expansion

$$\sum_{-\infty}^{\infty} c_n e^{2ni\pi z} = c_0 + \sum_{n=1}^{\infty} \big((c_n + c_{-n}) \cos 2n\pi z + i(c_n - c_{-n}) \sin 2n\pi z \big).$$

Mind you, it is worthwhile to pause to ask just how we might have known in the first place that $e^{2i\pi z}$ is periodic. Surely this is not obvious from its power series definition. Let me suggest two different 'explanations'. In the first we treat sin as original by computing the length of an arc of the circle $x^2 + y^2 = 1$ between the ordinates $x = 0$ and $x = \sin z$, and define sin by

$$z = \int_0^{\sin z} \frac{dx}{\sqrt{1 - x^2}}\,.$$

Then the circularity of the circle entails the periodicity of sin.

1991 *Mathematics Subject Classification.* 11D41, 11-01.
Work supported in part by grants from the Australian Research Council and a research agreement with Digital Equipment Corporation.

Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX

I prefer the following illustration. Here we admit that $\sin \pi z$ has simple zeros exactly at 0, ± 1, ± 2, ..., and — rather wildly thinking of it as just a polynomial of infinite degree — we factorise it and write

$$\sin \pi z = \pi z \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{n^2}\right) .$$

Of course that multiplier π (which, after all, might have been any decent function that never vanishes) needs rather calmer justification*.

With this evil deed done, we acknowledge that we are frightened of products, so we take the logarithm; and being bothered by logarithms, we differentiate. That yields

$$\pi \cot \pi z = \frac{1}{z} - \sum_{n=1}^{\infty} \left(\frac{1}{n-z} - \frac{1}{n+z}\right) .$$

Unfortunately, as we catch our breath, we see that this is a mildly nasty partial fraction expansion† in that it only converges conditionally — that is, on condition that we don't muck about with those parentheses. So we differentiate again and contemplate

$$\pi^2 \operatorname{cosec}^2 \pi z = \sum_{-\infty}^{\infty} \frac{1}{(n-z)^2} ,$$

and see that it shouts its periodicity. If we now backtrack, carefully, we are done.

Of course I've told this story to motivate its generalisation. I'd better also announce a principle. Loosely speaking, a function is 'good' if it is a convergent sum of good functions. That's why the sums above are good if we stay away from their *poles* — the points at which a term blows up, the integers $\mathbb{Z}$ in our examples. All this is given that 'good' — *analytic* — sort of means 'differentiable'; or better said, that the function may be expanded as a convergent power series. At the poles our functions — examples of *meromorphic* functions — are quite bad, but not very bad. They just take the value ∞; in other words, their reciprocal is 0. That's not too

---

* The trick is to notice that De Moivre's theorem, and then replacing $\cos^2 \theta$ by $1 - \sin^2 \theta$, allows us to write

$$\sin(2m+1)\theta = (-1)^m \sin\theta P_{2m}(\sin\theta) ,$$

with P_{2m} a monic polynomial of degree $2m$ and constant term $(-1)^m(2m+1)$. Of course its $2m$ zeros are $\pm \sin \pi k/(2m+1)$ for $k = 1, 2, \ldots, m$ so we have

$$\sin(2m+1)\theta = (2m+1)\sin\theta \prod_{k=1}^{m} \left(1 - \frac{\sin^2\theta}{\sin^2 \pi k/(2m+1)}\right) .$$

Now there is little more to do than to set $\pi z = (2m+1)\theta$ and to let m go to ∞.

† But now I can tell about Euler's evaluation of $\zeta(2k)$. We obtain

$$i\pi z \cot i\pi z = \pi z + \frac{2\pi z}{e^{2\pi z} - 1} = \pi z + \sum_{m=0}^{\infty} \frac{B_m}{m!}(2\pi z)^m = 1 + 2\sum_{k=0}^{\infty}(-1)^k \sum_{n=1}^{\infty} \frac{1}{n^{2(k+1)}} z^{2(k+1)} ,$$

and comparing coefficients of z^{2k} we have the claim ending the Notes III.

bad. Incidentally, after taking a logarithm, infinite products are just infinite sums. I'll try only to deal with functions that are good everywhere, except possibly for the odd pole; such pretty good functions are called *meromorphic* functions. These functions may well be very bad at $z = \infty$, that's permitted. A function without any singularity in the finite complex plane is called an entire function. I will use the fact that the constants comprise the only bounded entire functions.

If a good function f, not just a constant, has essentially different periods 1 and τ, then τ must be a dinkum* complex number. The point is that τ cannot be rational, either because then 1 is not a primitive period or, anyhow, then the periods are not essentially different; and if τ were real irrational then, because there are now $\mathbb{Z}$-linear combinations of 1 and τ that are arbitrarily small, f would have to be just a miserable constant. Thus τ must be $\mathbb{R}$-linearly independent of 1 and we may choose it in the upper half plane $\mathcal{H}$, that is, with positive imaginary part.

Just as $\mathbb{Z}$ is all the periods of $e^{2i\pi z}$, so all the periods of a doubly-periodic function with primitive periods 1 and τ is the *lattice* $\Omega = \{\omega = n\tau + m : n, m \in \mathbb{Z}\}$. Then

$$\wp'(z) = 2\sum_{\omega\in\Omega} \frac{1}{(\omega - z)^3}$$

defines a meromorphic doubly-periodic function with period lattice Ω. One confirms that, just as $\sum n^{-k}$ converges provided that $k > 1$, so $\sum' \omega^{-k}$ converges absolutely if $k > 2$. The $'$ tells one not to be silly. Integrating, we obtain

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega\in\Omega}{}' \left(\frac{1}{(\omega - z)^2} - \frac{1}{\omega^2}\right).$$

The periodicity of the Weierstraß $\wp$-function follows by observing that certainly $\wp(z + \omega) - \wp(z)$ is constant. Now take ω a primitive period and $z = -\frac{1}{2}\omega$, to see that the constant is zero because $\wp(z)$ is an even function of z.

Just as we produced the Riemann ζ-function in expanding $\operatorname{cosec}^2 \pi z$ as a power series, so expanding $\wp'(z)$ yields the Eisenstein series G_{2k},

$$\wp'(z) = -2z^{-3} + 6G_4 z + 20G_6 z^3 + \cdots \quad \text{where} \quad G_{2k} = G_{2k}(\tau) = \sum_{\omega\in\Omega}{}' \frac{1}{\omega^{2k}}.$$

There is now nothing for it other than brute computation to discover that

$$\big(\wp'(z)\big)^2 = 4\big(\wp(z)\big)^3 - 60G_4\wp(z) - 140G_6\,,$$

because the difference of the sinister and dexter sides of this equation is a doubly-periodic function without poles and vanishing at $z = 0$. An amusing corollary is that the G_{2k} must all be polynomials in just G_4 and G_6. Thus, for example, we must have

$$\left(\sum_{\omega\in\Omega}{}' \frac{1}{\omega^4}\right)\left(\sum_{\omega\in\Omega}{}' \frac{1}{\omega^6}\right) = c\sum_{\omega\in\Omega}{}' \frac{1}{\omega^{10}},$$

* *Macquarie Dictionary*: **dinkum** true, honest, genuine: as in *dinkum Aussie*.

a child's dream; c is a constant, independent of Ω, which is easy to determine, but which I have been too lazy to compute.

Put $y = \wp'(z)$ and $x = \wp(z)$. Then

$$y^2 = 4x^3 - g_2 x - g_3, \quad \text{with} \quad g_2 = 60G_4 \text{ and } g_3 = 140G_6,$$

is the equation of an *elliptic curve*. In truth, this equation has nothing much to do with ellipses. The background is that the integral that gives the length of an arc of an ellipse gives rise to a doubly-periodic function; thence the terminology whereby the Weierstraß $\wp$-function (and indeed all doubly-periodic meromorphic functions) are called *elliptic functions*. All this goes back to Fagnano's *rectification* of the *lemniscate*, the locus of a point which moves such that the product of its distances from two given points remains a positive constant, which leads to the integral

$$\int_w^\infty \frac{dr}{\sqrt{1-r^4}}\,.$$

Our curve is an *elliptic* curve because it may be *parametrised* by elliptic functions. I will eventually try to explain how it became believed that these curves can also be parametrised by modular forms.

My earlier remarks readily provide the factorisation

$$4x^3 - g_2 x - g_3 = 4\big(x - \wp(\tfrac{\tau}{2})\big)\big(x - \wp(\tfrac{1}{2})\big)\big(x - \wp(\tfrac{\tau+1}{2})\big) = (x-e_1)(x-e_2)(x-e_3)\,,$$

showing that the cubic has distinct zeros. Thus its discriminant $g_2^3 - 27g_3^2$, the square of the difference product $(e_1-e_2)(e_2-e_3)(e_3-e_1)$, is non-zero.

Now, a word about the lattice Ω. Let a, b, c and d be integers so that $ad-bc=1$. Then the ordered pair $(a\tau+b, c\tau+d)$ generates the same lattice as did $(\tau,1)$, and $(a\tau+b)/(c\tau+d)$ is still in $\mathcal{H}$, the upper half-plane. Set

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{and define} \quad M\tau = (a\tau+b)/(c\tau+d)\,.$$

Then

$$G_{2k}(M\tau) = (c\tau+d)^{2k} G_{2k}(\tau)\,.$$

The Eisenstein series are the basic examples of *modular forms*, of functions exhibiting an invariance under transformations by a discrete group, in this case the full modular group $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ of integer matrices of determinant 1.

In the mid-eighties Gerhard Frey suggested that if $a^p + b^p + c^p = 0$ then the elliptic curve $y^2 = x(x-a^p)(x+b^p)$ — note that its discriminant is essentially $(abc)^{2p}$, so the definition is symmetric — would have trouble existing. Indeed, in 1987, Ken Ribet showed on the presumption of the conjecture of Taniyama-Weil-Shimura — to the effect that every elliptic curve is *modular*, namely that it is parametrised by certain modular forms — that the *Frey curve* cannot exist. Andrew Wiles has demonstrated the Taniyama-Weil-Shimura conjecture for *semi-stable* elliptic curves, a class that includes the Frey curve.

André Weil, *Elliptic functions according to Eisenstein and Kronecker*, Ergebnisse der Math. 88, Springer-Verlag, 1976 makes instructive reading.