The Zeta Function

The Real Zeta Function

One way of defining the zeta function is as follows

$$\zeta(s) = \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \ldots \qquad (1)$$

When $s = 1$, this boils down to the harmonic series which is (famously) divergent. However, for all values of s greater than 1, the series converges.

In 1736 Euler proved that when $s = 2$ the sum of the series is equal to $\pi^2/6$. Interestingly, while the sum is known if s is an even integer, there is no known general formula for the sum of this series.

This is in contrast to the, superficially similar (reciprocal) geometric series:

$$G(s) = 1 + \frac{1}{s^1} + \frac{1}{s^2} + \frac{1}{s^3} + \frac{1}{s^4} + \ldots \qquad (2)$$

Multiplying this series by s we obtain

$$s\,G(s) = s + 1 + \frac{1}{s^1} + \frac{1}{s^2} + \frac{1}{s^3} + \ldots$$

$$s\,G(s) = s + G(s)$$

from which we deduce that

$$G(s) = \frac{s}{s-1} \qquad (3)$$

Putting $s = 2$, for example we have

$$G(2) = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots = \frac{2}{2-1} = 2$$

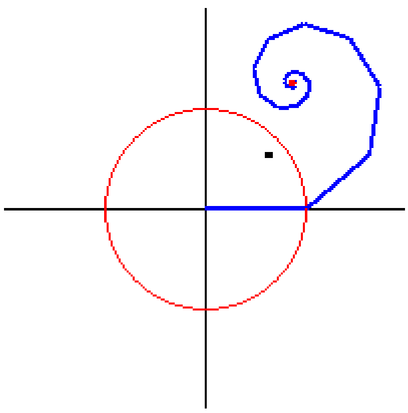
Obviously this formula is only valid for s > 1. When $s < 1$, the reciprocal series becomes a simple power series which diverges and the value of $G(s)$ becomes negative.

The Complex Geometric Series

If s is complex ($= a + ib = r\,e^{i\theta}$), a different situation arises. Each term of the geometric series can be written as

$$g_n = \frac{1}{r^n} e^{-in\theta}$$

and when we sum all of these terms to infinity, what we are doing is adding a whole series of rotating vectors which get smaller and smaller (assuming that $r > 1$) and which turn by an angle θ at each step – like this:



The black dot represents $\frac{1}{z}$ and the red dot shows the sum of $1 + \frac{1}{z} + \frac{1}{z^2} + \frac{1}{z^3} + \frac{1}{z^4} + \dots$

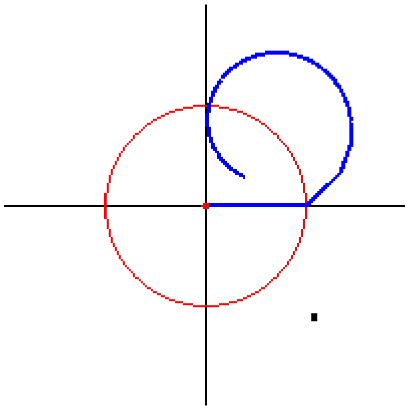
It is not difficult to see that the series will diverge whenever the modulus of $z < 1$. On the other hand, the formula for the sum of this geometric series $G(z) = \frac{z}{z - 1}$ has no such limitations (except when z actually equals 1). So although it is not true to say that this formula *is* the sum of the series, it does *extend* the function into regions where the brute force approach breaks down. The formula is called the *analytic continuation* of the series and it has been shown that any formula which correctly determines the sum of the series in the regions where the series converges will always give the same result in the regions where it does not. In other words the analytic continuation of a series formula is *unique*.

The complex Zeta function

It is of great interest to study the complex version of the Zeta function and provided that the *real* part of $z > 1$, a brute force may be used to calculate the approximate values.

$$\zeta(z) = \frac{1}{1^z} + \frac{1}{2^z} + \frac{1}{3^z} + \frac{1}{4^z} + \frac{1}{5^z} + \dots$$

This is what the successive terms look like when z is approximately equal to $(1.1, -i)$.



It turns out that whenever the real part of $z > 1$, the series converges, spiralling in to a fixed point; but when $\mathbf{R}(z) < 1$ it spirals out to infinity. Also, for values of $\mathbf{R}(z) < 2$ the series converges very slowly. Nevertheless, although the sum of the series must be infinite at $z = 1$, it appears to behave perfectly normally whenever the imaginary part is non zero. Why should the series suddenly blow up as the $\mathbf{R}(z) = 1$ line is approached?

What we need is a way to continue the function into the 'forbidden' region in the same way that we found an analytic continuation of the geometric series when $|z| < 1$.

The analytic continuation of the Zeta function

One way of doing this is to try the same trick with the Zeta series as we did with the geometric one – that is to say, try to find a way of manipulating the series algebraically so that we generate the same series in a different form. Watch.

Here is the Zeta series: $$\zeta(z) = \frac{1}{1^z} + \frac{1}{2^z} + \frac{1}{3^z} + \frac{1}{4^z} + \frac{1}{5^z} + \dots \quad (4)$$

Negate all the even terms. Call this series $\eta(z)$ (eta).

$$\eta(z) = \frac{1}{1^z} - \frac{1}{2^z} + \frac{1}{3^z} - \frac{1}{4^z} + \frac{1}{5^z} + \frac{1}{6^z} + \dots$$

Multiply the zeta series by $\frac{2}{2^z}$

$$\frac{2\zeta(z)}{2^z} = \frac{2}{2^z} + \frac{2}{4^z} + \frac{2}{6^z} + \frac{2}{8^z} + \ldots$$

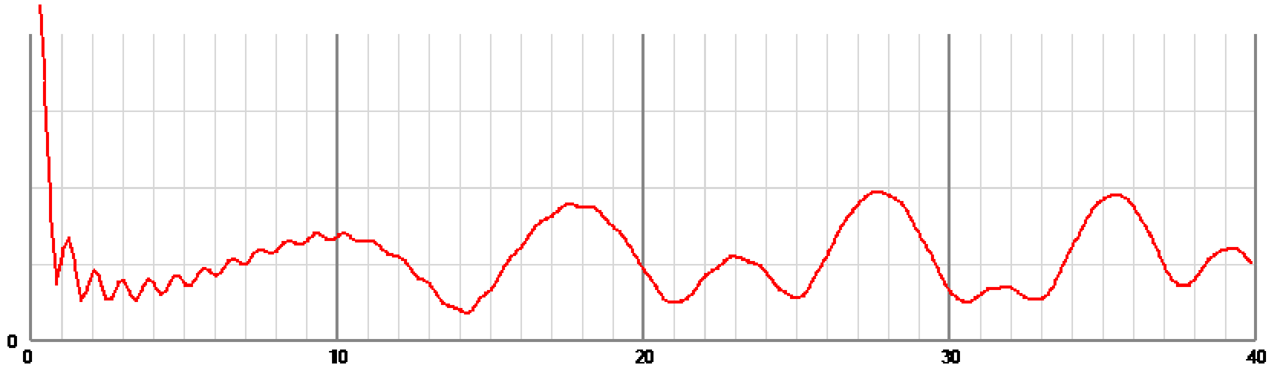
Now if we add this series to the eta series, all the negative terms get turned back into positive ones.

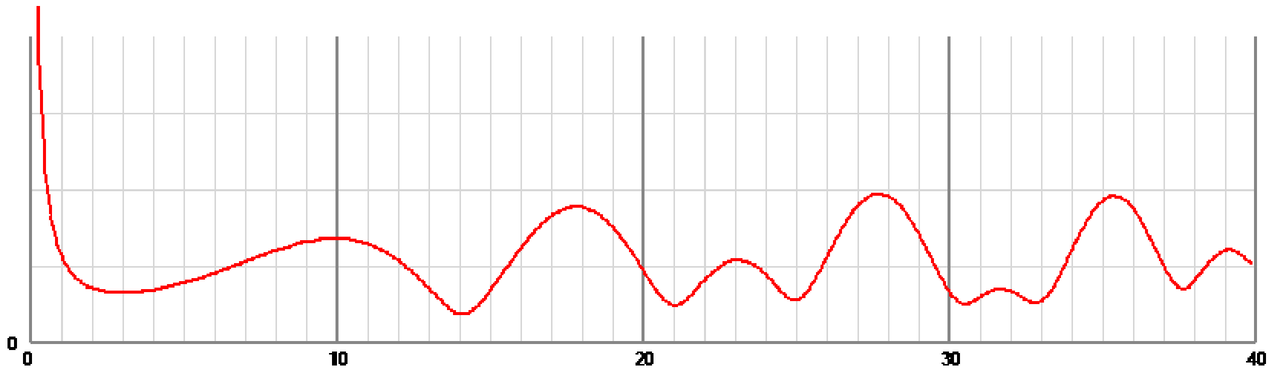
What we are saying is that $$\frac{2\zeta(z)}{2^z} + \eta(z) = \zeta(z)$$

or, to put it another way $$\zeta(z) = \eta(z) \times \frac{2^z}{2^z - 2} \qquad (5)$$

Now you might think this has got us nowhere because in order to evaluate $\zeta(z)$ we need to know $\eta(z)$ but it turns out that, because of the alternating terms, not only does $\eta(z)$ converge a lot quicker that $\zeta(z)$, it also converges for **R**$(z) > 0$, not just > 1.

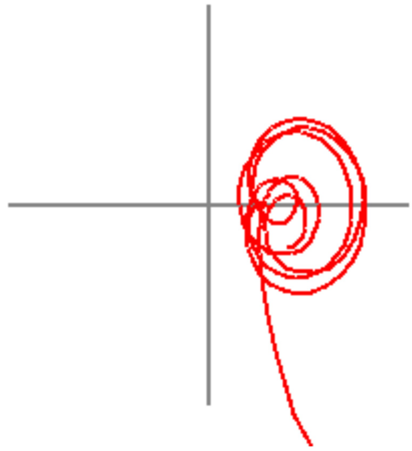
Compare the following two graphs which plot the modulus of the zeta function along the line **Re**$(z) = 1.1$ for imaginary values from 0 to 40.





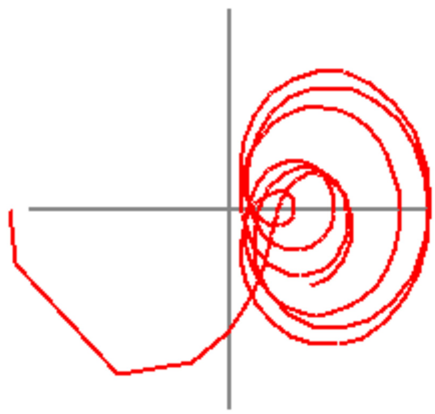
The first is calculated using the standard algorithm. The second uses the analytic continuation. The improvement is obvious.

The shape of this is a bit surprising and shows little obvious regularity. The reason for the dips and bumps is best explained using a plot showing how $\zeta(z)$ moves in the complex plane as **Im**(z) varies from 0 to 40. The point starts at ∞ and then starts to spiral around the point (1, 0) in rather irregular circles. These circles are always in the real-is-positive region so the modulus of $\zeta(z)$ never reaches zero. The circles get smaller and smaller as **Im**(z) is increased and eventually spirals into the point (1, 0-)

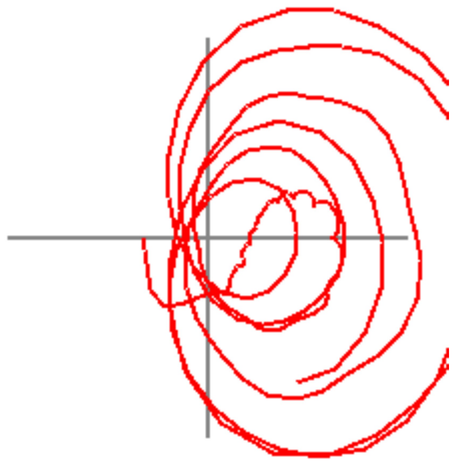


***Re**(z) = 1.1*

Now if we look at what happens when we pass from **Re**(*z*) > 1 to **Re**(*z*) < 1 nothing very surprising happens (as long as we are using the analytic continuation, of course). The circles shift slightly to the left but they remain in the real-is-positive region. On the other hand, Something very interesting (and as it turns out, very significant) happens when move from **Re**(*z*) = 0.7 to **Re**(*z*) = 0.3.

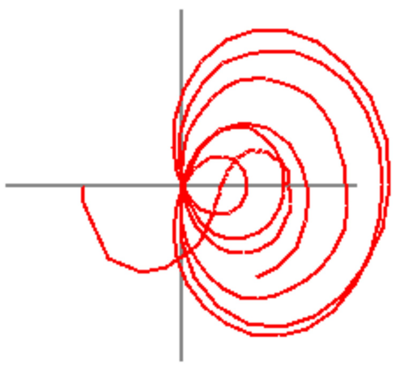


***Re**(z) = 0.7*



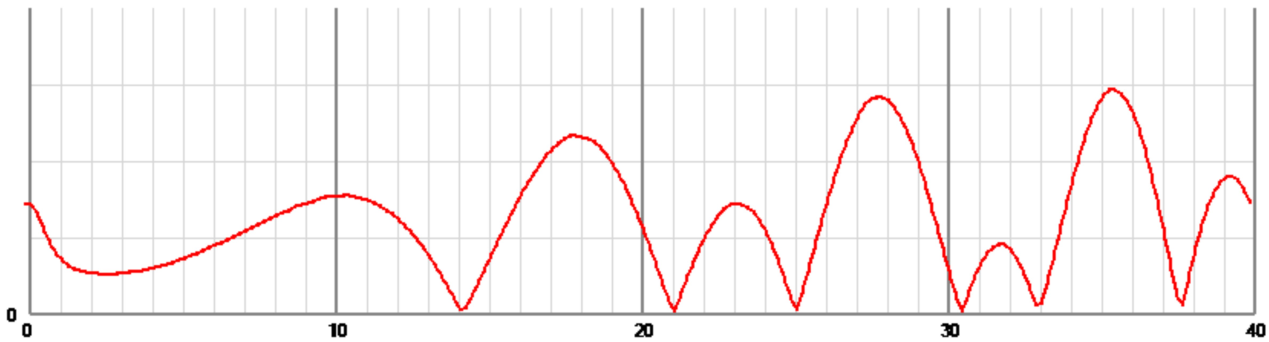
***Re**(z) = 0.3*

When **Re**(*z*) = 0.7 the circles are still wholly in the real-is-positive region (at least up to **Im**(*z*) = 40.). But when **Re**(*z*) = 0.3, the circles are orbiting round the origin. Clearly there must be some value (or values) where the circles cross through the origin exactly and this is exactly what happens when **Re**(*z*) = 0.5.



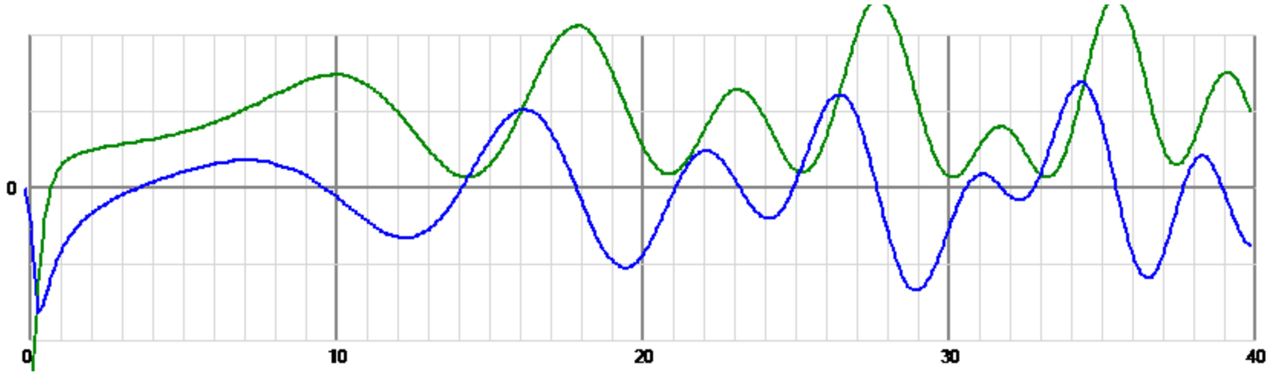
***Re**(z) = 0.5*

The really surprising thing about this is that, apparently, *all* of the circles pass through the origin when **Re**(z) = 0.5. What this means is that, at this special value, the modulus of ζ(z) dips down to exactly zero. This happens when the imaginary part of z is equal to approximately 14.13, 21.02, 25.01, 30.42, 32.94 and 37.59 et. etc. These are the famous Riemann zeros.

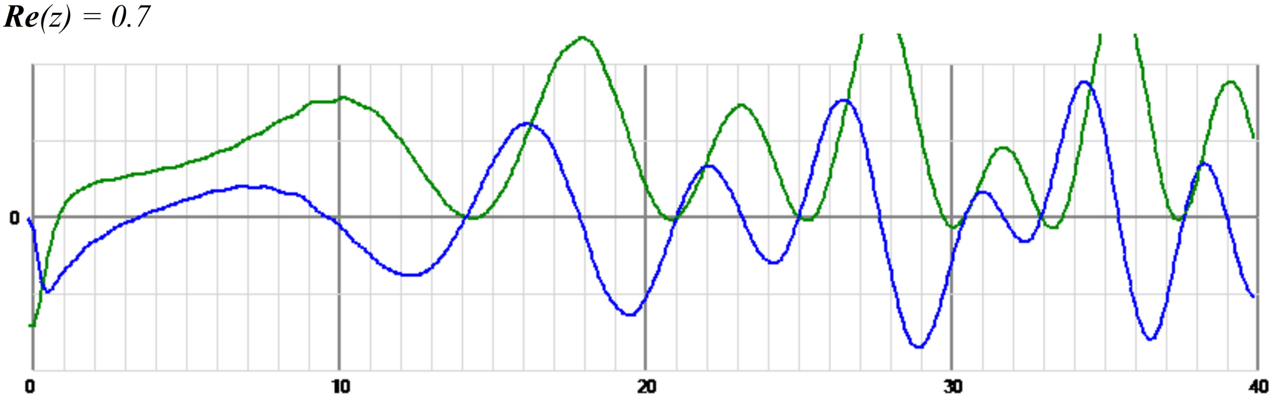


Re*(z) = 0.7*

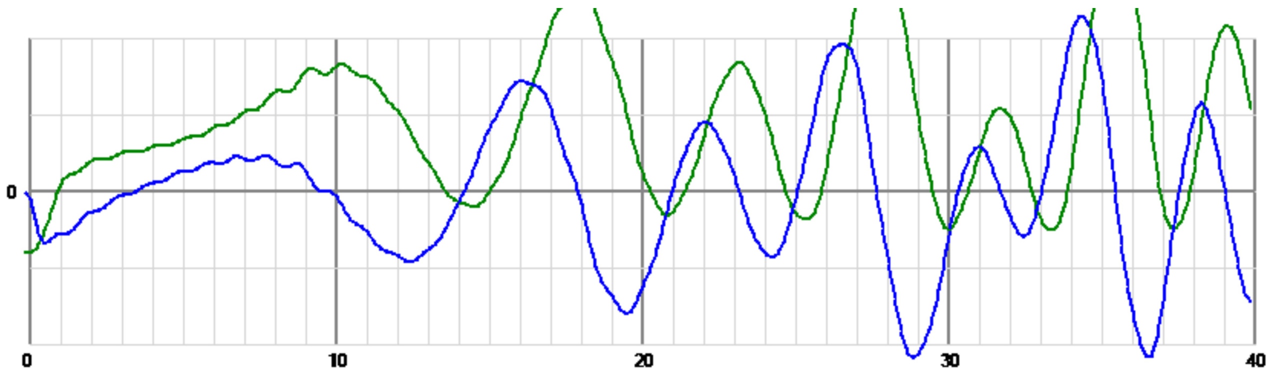
To see how surprising this is, it is worth comparing the graphs of the real (green) and imaginary (blue) parts of ζ(z) at the three values **Re**(z) = 0.7, 0.5 and 0.3. Here they are:



Re*(z) = 0.7*



Re*(z) = 0.5*



Re*(z) = 0.3*

On the face of it the real and imaginary parts oscillate up and down with the same frequency but apparently random phase and amplitude. If they were totally independent oscillations, the chances that the imaginary oscillation would be zero at the *precise instant* that the real oscillation crosses the axis would be zero.

But of course, they are not independent and as we have seen, there *must* be places where this happens. But there is no obvious reason why the *only* places this happens are when **Re**(z) = 0.5. The statement that all the Riemann zeros lie on the line **Re**(z) = 0.5 is known as the Riemann hypothesis and is probably the most famous unproved mathematical theorem of all time.

The connection between the zeta function and prime numbers

In 1737 Euler discovered the first link. He showed that $\zeta(s)$ (where s is a real number) was related to a certain product involving all the prime numbers. His proof is relatively straightforward and will help us to understand what it is that he actually proved.

The (real) zeta function looks like this:

$$\zeta(s) = \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \ldots \tag{6}$$

We can select all the even terms by multiplying all the term by $1/2^s$.

$$\zeta(s) \times \frac{1}{2^s} = \frac{1}{2^s} + \frac{1}{4^s} + \frac{1}{6^s} + \frac{1}{8^s} + \frac{1}{10^s} + \ldots \tag{7}$$

Now if we subtract all these terms from equation (6) we remove all the even terms:

$$\zeta(s) \times \left(1 - \frac{1}{2^s}\right) = \frac{1}{1^s} + \frac{1}{3^s} + \frac{1}{5^s} + \frac{1}{7^s} + \frac{1}{9^s} + \ldots \tag{8}$$

The idea now is to remove all the terms which are multiples of 3 by doing the same sort of thing. First multiply by $1/3^s$:

$$\zeta(s) \times \left(1 - \frac{1}{2^s}\right) \times \frac{1}{3^s} = \frac{1}{3^s} + \frac{1}{9^s} + \frac{1}{15^s} + \frac{1}{21^s} + \ldots \tag{9}$$

and subtract this from equation (8)

$$\zeta(s) \times \left(1 - \frac{1}{2^s}\right) \times \left(1 - \frac{1}{3^s}\right) = \frac{1}{1^s} + \frac{1}{5^s} + \frac{1}{7^s} + \frac{1}{11^s} + \frac{1}{13^s} + \ldots \tag{10}$$

What we are left with now is a series in which all the terms which are either multiples of 2 or multiples of 3 have been removed. You can see that what we are doing is a form of sieve, weeding out all the composite numbers, prime by prime.

Obviously if we were to go on doing this for all the primes, the only term left would be the first, namely $1/1^s = 1$. We conclude therefore that

$$\zeta(s) \times \prod_p \left(1 - \frac{1}{p^s}\right) = 1$$

or, if you prefer

$$\zeta(s) = \prod_p \left(\frac{1}{1 - p^s}\right)$$

(Incidentally, if we put $s = 1$, we find that the product of $\frac{2}{1} \times \frac{3}{2} \times \frac{5}{4} \times \frac{7}{6} \times \frac{11}{10} \ldots$ must be infinite. - which proves the infinitude of primes.)

In 1859 the 33 year old German mathematician Bernhard Riemann wrote a 8-page paper which revolutionised the study of prime number theory. The Holy Grail would be a formula for the n^{th} prime number. He did not achieve this but he used Euler's product formula to devise a formula $\boldsymbol{R}(x)$ for the number of prime numbers less than or equal to any real x. (In principle therefore you could use his formula to test whether a given integer was or was not prime. Simply (!) calculate $\boldsymbol{R}(x)$ and $\mathbf{R}(x - 1)$. If they differ by 1, x must be prime.) The problem is, his formula involves using the non-trivial zeros of the zeta function and therefore relies of the truth of the Riemann Hypothesis.

Riemann's prime number counting formula is not the only theorem which depends on the truth of the hypothesis.

Some thoughts on the Riemann Hypothesis

Riemann's hypothesis was included in Hilbert's famous list of unsolved problems in 1900. It remains unproved to this day. Mathematicians can be divided into three camps. There are those who firmly believe that the hypothesis is provable and that one day they will have the proof. Certainly there is a lot of evidence that the hypothesis is true. Using modern computers the positions of the first 2 million zeros have been calculated and all have been found to be on the line $\mathbf{Re}(z) = 0.5$. In addition, many zeros way beyond this have been calculated, so far in fact that on a scale of 1mm to the unit, the furthest zero would be over a light year away – and all have so far been found to lie on the line. But specific examples do not constitute a proof. They do, however, give many mathematicians encouragement that there is some deep reason *why* all the zeros lie on the line and that it should not be beyond the wit of man to discover it.

On the other hand, the wit of man has been unequal to the task for 150 years. This has caused another group of mathematicians to speculate that, perhaps, the Riemann hypothesis is one of those famous Gǒdel statement that are true but unprovable. If so the situation is similar to the situation facing geometers in the early 19th century. Euclid's famous 'parallel postulate' was proving to be unprovable on the basis of the other axioms of the system – but 'obviously' true. Then came Lobochevsky who showed that it was perfectly possible to construct an alternative geometry in which the parallel postulate was false. What everyone had assumed up to that point was that geometry had to take place on a 'flat' surface. Could it be that there is some axiom of logic at the basis of number theory which we have so far unwittingly assumed which makes the Riemann hypothesis true but unprovable? And could it be that a future Lobochevsky will identify this axiom and thereby generate a new mathematics which contains all the familiar theorems of arithmetic but in which the Riemann hypothesis is false? Somehow I find this idea very unpalatable. Surely either the hypothesis is true – in which case there must be a *reason* why it is true; or it is false, in which case it is worth continuing the search for a counter-example.

There is a third option. The Riemann hypothesis is true and a proof of the hypothesis exists but the proof is so complicated that it will forever remain beyond the wit of man to devise it. If this is, in fact, the case, we shall have to rely on artificial intelligence to generate the proof for us. Even then, we will not necessarily be able to understand or check the proof which the computer has generated for us.

Carr Bank, August 2018 (updated December 2019)