Space, but not as we know it

Wiskundigen kennen allerlei soorten getallen – en veel daarvan worden ook in de natuurkunde gebruikt. Heel fascinerend zijn de zogeheten p-adische getallen. In deel 2 van deze serie legt Lizzy Rieth uit hoe de p-adische afstandsmaat leidt tot heel bijzondere 'ruimtes', waarin bijvoorbeeld driehoeken er heel anders uit zien dan wij gewend zijn!

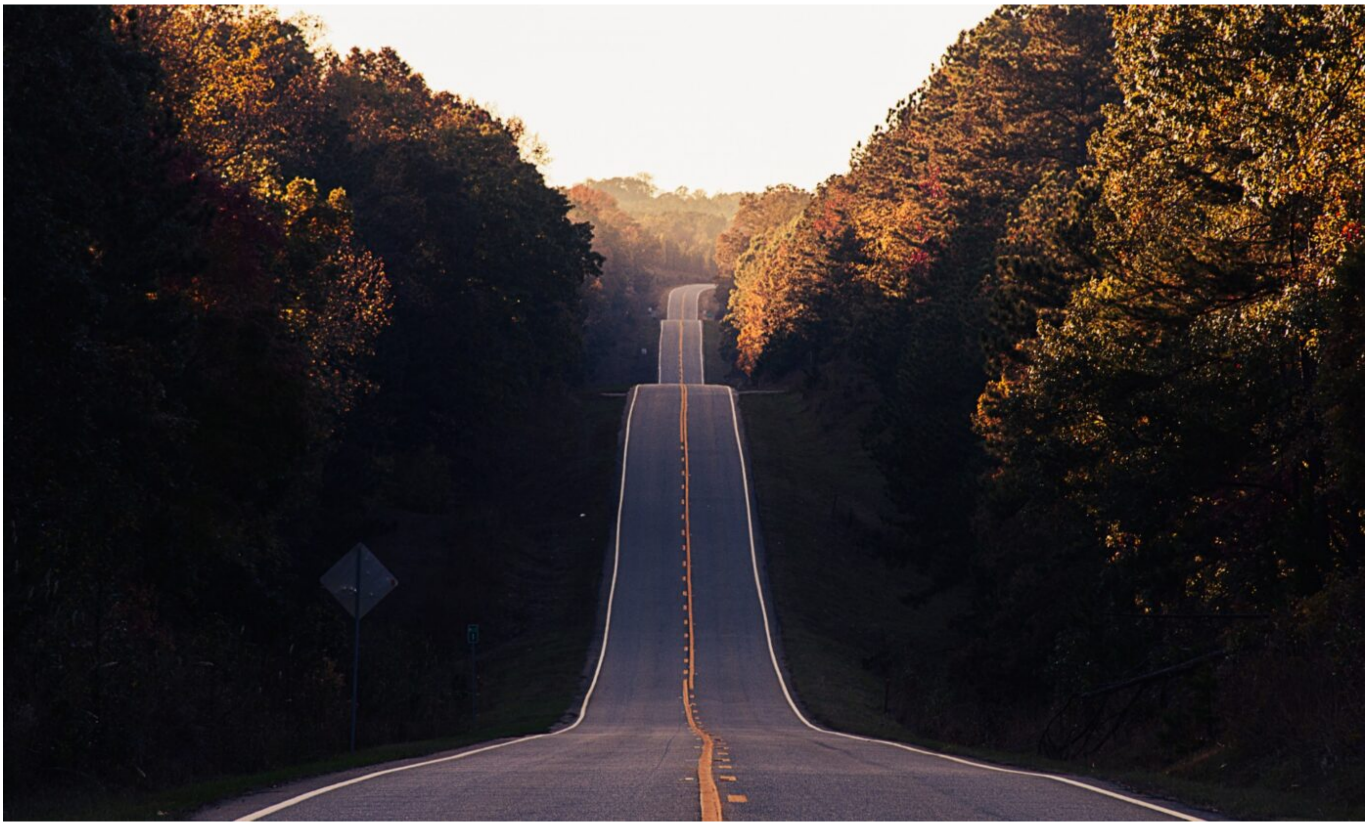

Figure 1. Distance matters. We often think of numbers as situated on a real line – think of measuring distances on a road. However, if we change the shape of the road – and therefore: change the distance concept – the notion of numbers also changes. The p-adic numbers are what you get if you do a very rigorous change of this type, much less smooth than the relatively minor deformation of the road in this picture. [Image source: Public domain]

In a previous article, we explored p-adic numbers and some of their surprising properties. In particular, I introduced the idea of p-adic distance and briefly mentioned that p-adic spaces obey a stronger version of the familiar triangle inequality, known as the **ultrametric**

property. In this article, we shall take a closer look at what this means and see examples, from biology and physics, of how these ideas are used in practice.

i) What is a Metric Space?

Mathematicians use the term **metric space** to denote a collection of objects, often thought of as points, equipped with a rule that tells us how to measure the distance between any two of them. In this way, metric spaces provide a mathematical framework for studying geometry and also many physical systems.

Before exploring ultrametric spaces, for which distance is defined in the p-adic way, let us take a step back and consider the kinds of spaces that are more familiar to us. These are spaces in which distance is defined like the absolute value on the real number line. Familiar examples include smooth geometries like the Euclidean plane, the surface of a sphere, and hyperbolic space. Even the spacetime geometries used in Einstein’s theory of relativity, such as Minkowski or (Anti-)de Sitter spacetimes, fit naturally into this broader picture.

One of the most important properties of metric spaces is that distances within them satisfy the **triangle inequality**. As the name suggests, it tells us something about the side lengths of triangles “living” in a given space and is one of the fundamental requirements that any sensible definition of distance should satisfy. In words, the triangle inequality states: for any triangle, the sum of the lengths of any two sides must always be greater than or equal to the length of the remaining side. The triangle inequality, for a triangle with vertices (corners) A, B and C, can be expressed mathematically by the equation below, while Figure 1 provides a simple visual illustration of the same geometric idea.

$$\large d(A,C) \leq d(A,B) + d(B,C).$$

Note that for a given triangle there are really *three* triangle inequalities: we could have written analogous statements for d(A,B) and for d(B,C).

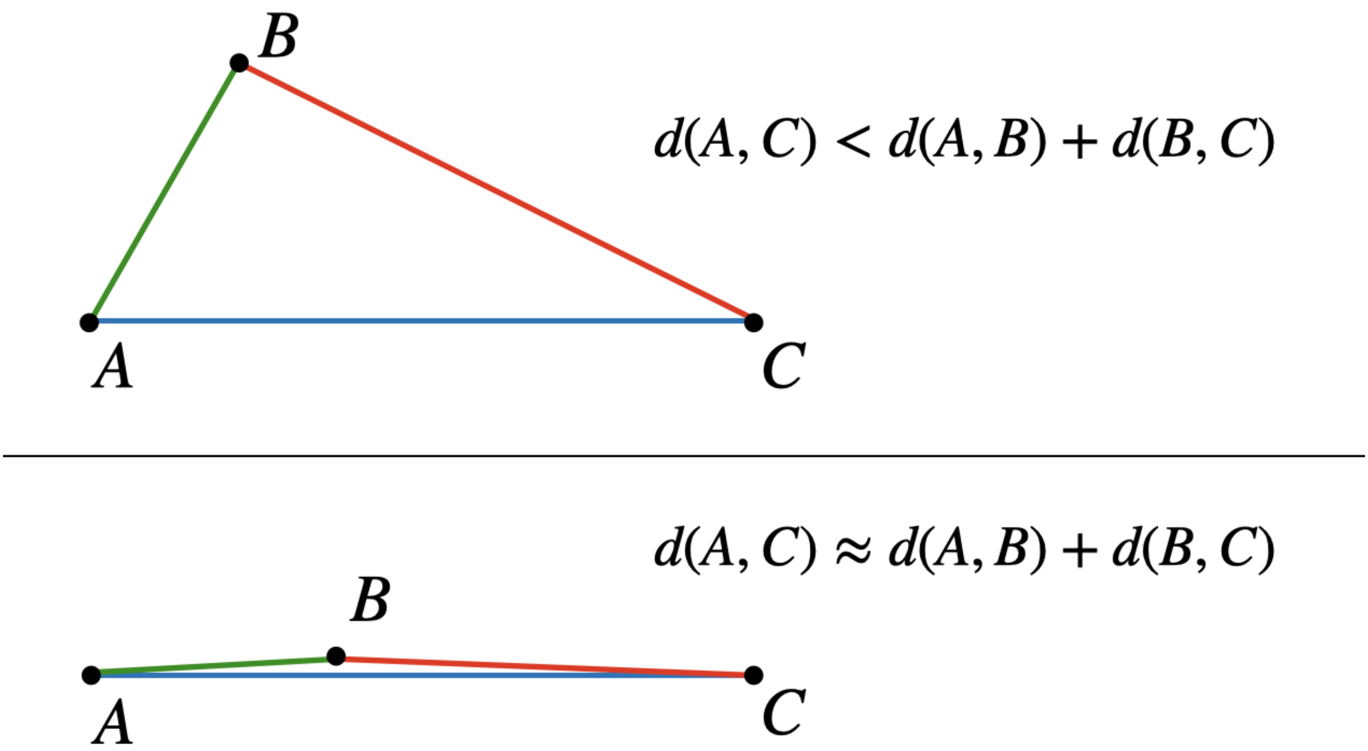

Figure 2. Two examples illustrating the triangle inequality in metric spaces. In the top example, the distance d(A,C) between points A and C is smaller than the sum of the other two sides, d(A,B)+d(B,C). In the bottom example, both sides of the inequality are almost equal. If d(A,C) were exactly equal to d(A,B)+d(B,C), the three points would lie on a straight line and the area of the triangle would shrink to zero.

The importance of the triangle inequality lies in the fact that it allows us to consistently define the concept of "closeness" in a given space. The intuition it provides us goes as follows: If two points are both close to a third point, they must also be close to each other. For example, if a point A is close to B, and B is close to C, then A cannot possibly be very far away from C. The triangle inequality captures this idea by saying that in fact, A can at most be the sum of the other two distances away from C. This consistency is also essential when mathematically describing convergence. If several points in a sequence become increasingly close to the same limiting value, the triangle inequality guarantees that they must also become increasingly close to one another.

Perhaps surprisingly, the triangle inequality is not just restricted to flat planes and spaces, which are also known as Euclidean geometries, but also holds in curved spaces. The main difference is that, in such spaces, the shortest path between two points is not necessarily a straight line, as it is on the Euclidean plane, but instead follows the geometry of the space itself. For example, on the surface of a sphere, the shortest path between two points is an arc along a great circle. Although triangles drawn on a sphere can have angle sums greater than 180 degrees (while in negatively curved spaces they can have angle sums less than 180 degrees), the triangle inequality itself remains valid. In other words, even in curved geometries, a path that goes via a third point can never be shorter than the shortest direct path between the two endpoints.

In the smooth and continuous metric spaces familiar to us, the distance between two points is represented by a non-negative real number. In these spaces, the distances relating any set of three points always satisfy the triangle inequality and therefore provide a consistent way of deciding when two points are close to or far from one another. However, things change if we employ the p-adic norm to characterise distances between points in a space.

ii) Moving on to ultrametric spaces

In the previous article, we saw that defining what we mean by “distance” plays a crucial role in constructing both the real numbers and the p-adic numbers as completions of the rational numbers. For the real numbers, distance is measured in the familiar way: two numbers are close if they lie close together on the number line, and the absolute value of the difference of the two numbers is the way to measure this distance. In terms of the p-adic numbers, however, two numbers are considered close whenever their difference is not necessarily “small” in the traditional sense, but rather divisible by a large power of the prime number p. Equivalently, they are close when they share many leading digits in their p-adic expansion (that I discussed in the previous article), even if they are very far apart in the usual, intuitive sense.[1]

This unusual p-adic definition of distance has remarkable consequences: Suppose that two p-adic numbers are both close to a third number. Since their p-adic expansions each share many leading digits with that third number, they must also share many leading digits with one another. For example, if one number shares ten leading digits with a third number and another shares six, then the first two numbers must share at least those six digits with each other. In terms of distance, this means that the separation between the first two numbers can never be greater than the larger of their two distances from the third. This is a much stronger statement than the ordinary triangle inequality, under which the third side can take any intermediate real value permitted by the inequality, allowing triangles with a much wider range of side lengths and shapes. Mathematically, p-adic distances thus satisfy the more stringent **strong triangle inequality**, also known as the **ultrametric property**:

$$\large d(A,C) \leq \mbox{max} \{ d(A,B), d(B,C) \}.$$

In words, the ultrametric inequality ensures that the length of a given side of a triangle can never be longer than the larger of the other two sides. As a consequence, among the three sides of a triangle, the longest side can never be unique: it must always occur at least twice.

Let’s clarify this statement using a concrete example: consider the 5-adic distances between integers 5, 25, and 50. Remember that the 5-adic distance between two rational numbers x and y is generally defined as follows:

$$\large d_5(x,y)=|x-y|_5 = 5^{- v_5(|x-y|)},$$

where $v_5(|x-y|)$ is the exponent of 5 in $|x-y|$. Two numbers x, y whose difference is divisible by a large power of 5 are considered 5-adically close. In our concrete example, the differences 25-5=20 and 50-5=45 are each divisible by 5, but not by 25. So we can determine the 5-adic distances $d_5(5,25)$ and $d_5(5,50)$ as follows;

$$\large d_5(5,25)=d_5(5,50)=\frac{1}{5}.$$

By contrast, 50-25=25 is also divisible by 5^2=25, giving us a different (closer) 5-adic distance:

$$\large d_5(25,50)=\frac{1}{25}.$$

Thus, in terms of 5-adic distance, the integer number 5 is equally close to 25 and to 50. However, perhaps against our intuition about the real numbers, the integer 25 is 5-adically closer to 50 than it is to 5. The three considered 5-adic distances are therefore $d_5(5,25)=1/5, d_5(5,50)=1/5$, and $d_5(25,50)=1/25$, illustrating that the largest distance occurs twice, as stated by the strong triangle inequality.

Enforcing this more stringent triangle inequality places much stronger restrictions on the shapes that triangles can take in ultrametric spaces: they can only be either equilateral; i.e., all three sides have the same length; or isosceles, with the two longer sides having exactly the same length. In particular, scalene triangles, those with three different side lengths, are impossible to construct if one respects the strong triangle inequality.

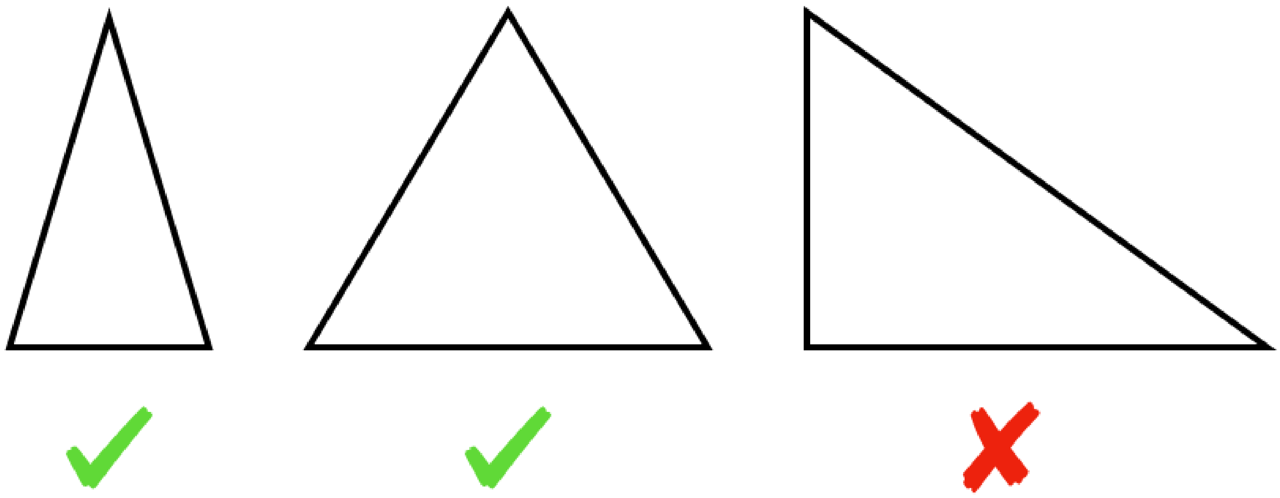

Figure 3. Allowed triangles. The only triangles permitted in an ultrametric space are equilateral (left) and isosceles with the two longer sides equal (centre), both of which satisfy the strong triangle inequality. A scalene triangle (right), which is perfectly admissible in an ordinary metric space, cannot satisfy the ultrametric inequality.

At first sight, the difference between the ordinary and strong triangle inequalities may seem like a minor detail. However, imposing the strong triangle inequality results in spaces whose structure is very different from familiar metric spaces in everyday geometry. Since every triangle must be either equilateral or isosceles, many of the geometric intuitions we have developed from everyday experience no longer apply. An example is the more stringent way in which closeness is preserved. As mentioned above, in an ordinary metric space, if A is close to B and B is close to C, then A and C cannot be arbitrarily far apart. However, in an ultrametric space, their distance cannot even exceed the larger of the two original distances. For example, suppose that d(A,B)=2 and d(B,C)=5. The strong triangle inequality tells us that d(A,C) ≤ max{2,5} = 5, unlike the ordinary triangle inequality, under which we would need to satisfy the bound d(A,C) ≤ 7. This leaves no possibility for A and C to be further apart than their largest distance to B. In fact, in this example the ultrametric inequality forces d(A,C)=5 (it is a fun exercise to prove this to yourself by writing out the other two triangle inequalities between the three points), producing an isosceles triangle.

Thus, instead of spreading out smoothly and continuously, as points do in ordinary Euclidean geometry, ultrametric spaces organise themselves into a hierarchy of nested clusters of points that all have the exact same distance to one another. The resulting geometry is therefore much more naturally represented by a branching tree than by a flat plane. This brings us back to the p-adic numbers. As we saw in the previous article, two p-adic numbers are considered close when they share many leading digits in their p-adic expansion. It is precisely this notion of distance that satisfies the strong triangle inequality, making the set of p-adic numbers an ultrametric space.

iii) What's next?

We have seen that besides 'ordinary' (metric) spaces, the mathematical world also contains stranger, ultrametric spaces. How can we visualise these ultrametric spaces, and do they also play a role in the real, *physics* world? Those are questions I will address in the third episode of this series on p-adic numbers, which will appear on this website on 14 August.