The Importance of Diversity in Mathematical Research

Alan F. Beardon,
University of Cambridge and AIMS

Kigali, Tuesday 9th July 2019

The speakers were asked to speak

- on the Future of Science
- to a non-specialist audience
- from a broad variety of backgrounds.

In my view, mathematics inhabits two worlds:
(i) the world of ideas (what we think about in our head), and
(ii) the world of reason (what we write down on paper and, perhaps, publish).

(i) is a very personal thing, and no one else has access to this unless, and until, we participate in an informal discussion.

The purpose of (ii) is to *communicate* ideas to others, in a way that is *beyond dispute*.

the world of IDEAS	the world of PRECISION
imaginative ideas in the mind	carefully written account
thinking with diagrams is effective	no dependence on diagrams
free of grammar	obey grammatical rules
free of logic	obey logical rules
no calculations needed	precise calculations required
ambiguities require clarification	no ambiguities allowed
contradictions stimulate thoughts	no contradictions allowed
intuition,guesses,creative thoughts	not part of the argument

I believe that we should discuss this distinction with *all* students of mathematics.

So how should we teach mathematics?

We should actively seek, **and teach**, *as many connections between different parts of mathematics as we can*, and not avoid them (as many texts do).

In this lecture I will focus on the ‘world of ideas’, and some connections between different parts of mathematics.

A Möbius map is a map of the form

$$z \mapsto \frac{az+b}{cz+d}, \quad z \in \mathbb{C}.$$

With students we focus on the computational aspects, and relegate the issues at $-d/c$ and ∞ to little more than footnotes. There are different ways round this problem (e.g. stereographic projection, and projective spaces), and *these ideas are important in their own right*.

Stereographic projection

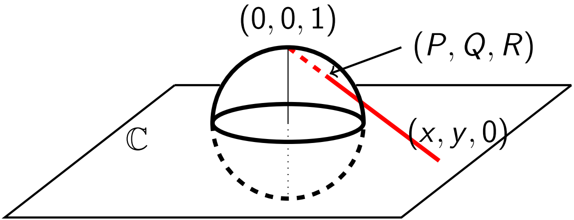

Figure: Stereographic projection

$$z = x + iy \mapsto (P, Q, R), \qquad \infty \mapsto (0, 0, 1)$$

The hyperbolic plane

The upper half $\mathbb{H}^2$ of $\mathbb{C}$ is $\{x + iy : y > 0\}$, and the hyperbolic metric on $\mathbb{H}^2$ is $ds = |dz|/y$. With this, $\mathbb{H}^2$ is a model of the hyperbolic plane.

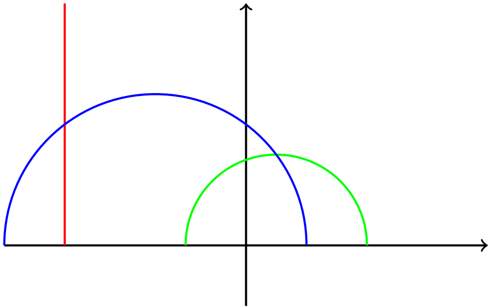

Figure: The hyperbolic plane $\{x + iy \colon y > 0\}$

Hyperbolic 3-space

The upper half $\mathbb{H}^3$ of $\mathbb{R}^3$ is $\{(x, y, t) : t > 0\}$, and the hyperbolic metric on $\mathbb{H}^3$ is $ds = |dx|/t$. With this, $\mathbb{H}^3$ is a model of hyperbolic 3-space.

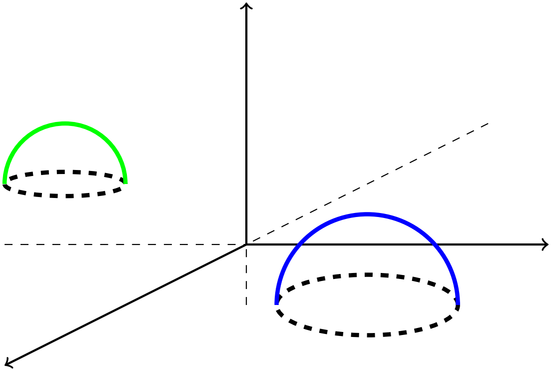

Figure: Hyperbolic 3-space

It is well known that Möbius maps form the entire group of isometries of hyperbolic 3-space (the upper half of $\mathbb{R}^3$), so why do we insist on studying them in the complex plane (where they are not isometries) instead of studying them in $\mathbb{H}^3$ (where they are isometries)?

For example, a hyperbolic plane is a hemisphere whose bounding circle is in $\mathbb{C}$ and (like other geometries) the hyperbolic plane is the set of points equidistant from two given points. Thus circles map to circles. A similar comment holds for inverse points.

These ideas have profound implication for number theory. The (much studied) modular group is the group of Möbius maps

$$z \mapsto \frac{az+b}{cz+d}, \quad a, b, c, d \in \mathbb{Z}, ad - bc = 1,$$

and this is a discrete group of isometries acting on the hyperbolic plane $\mathbb{H}^2$.

The modular group

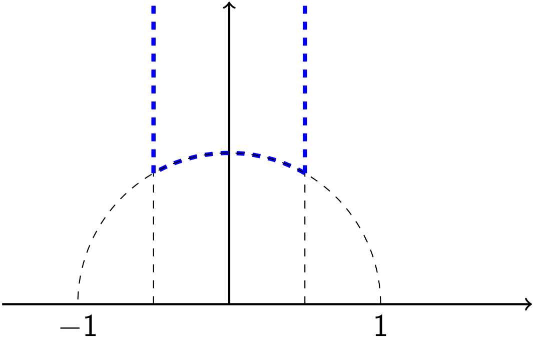

Figure: A fundamental region for the modular group

The modular group

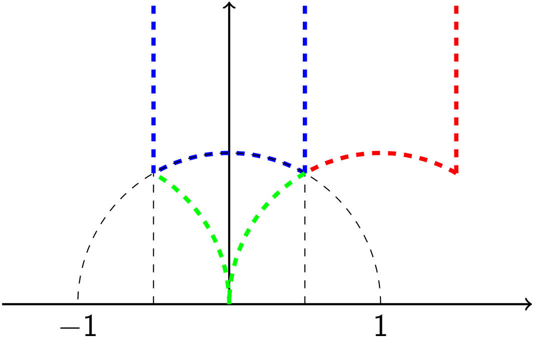

Figure: A fundamental region for the modular group

A lot of number theory *can be interpreted, and understood, as the boundary behaviour of the action of the modular group on* $\mathbb{H}^2$.

Indeed, *every* discrete group of isometries of the hyperbolic plane (and not just the modular group) has its own "number theory" on the real line!

Here is an application to number theory: **Pell's equation** is the famous Diophantine equation $X^2 - NY^2 = 1$.

As far back as AD 628 the Indian mathematician Brahmagupta shows how to combine two solutions to get a third solution:

$$(XU + NYV)^2 - N(XV + YU)^2 = (X^2 - NY^2)(U^2 - NV^2)$$

The Pell hyperbola

The set of solutions of Pell's equation form an infinite cyclic group.

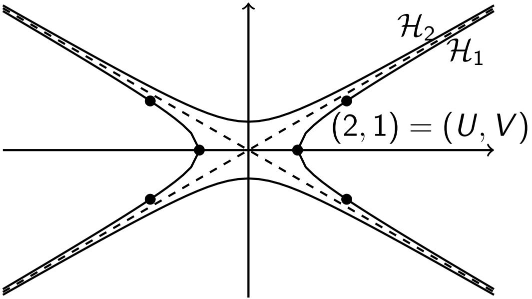

Figure: $X^2 - 3Y^2 = \pm 1$

From a more modern viewpoint, we have

$$\det \begin{pmatrix} XU + NVY & N(XV + YU) \\ XV + YU & XU + NYV \end{pmatrix} = \det \begin{pmatrix} X & NY \\ Y & X \end{pmatrix} \det \begin{pmatrix} U & NV \\ V & U \end{pmatrix}$$

Now if $X^2 - NY^2 = 1$, then $\begin{pmatrix} X & NY \\ Y & X \end{pmatrix}$ is the matrix for the Möbius map $z \mapsto \frac{Xz+NY}{Yz+X}$ in the modular group, and the entire theory of Pell's equation can be studied through the action of the modular group on the hyperbolic plane $\mathbb{H}^2$ (that is, by using hyperbolic geometry!)

Solutions of Pell's equation

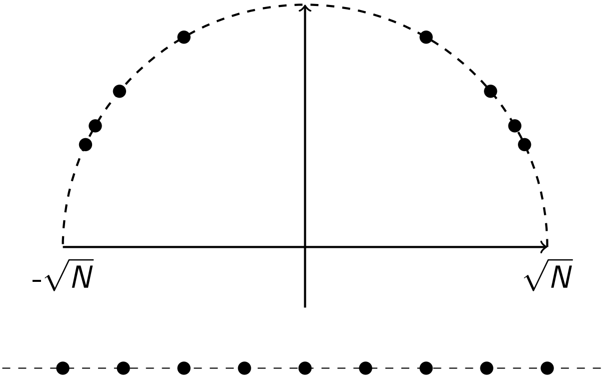

Figure: Solutions of Pell's equation

The advantages of this approach are even greater when we come to study the 'negative' Pell equation $X^2 - NY^2 = -1$. From a number-theoretic point of view there are difficulties with this equation: it has integer solutions for some values of N, but not for others, and there is no known simple condition which tells us when this equation does have integer solutions.

In this case we should study the Möbius map derived from the matrix $A = \begin{pmatrix} X & NY \\ Y & X \end{pmatrix}$, where $\det(A) = -1$. This Möbius map is

$$f(z) = \frac{iXz + iNY}{iYZ + iX}, \det \begin{pmatrix} iX & iNY \\ iY & ix \end{pmatrix} = 1,$$

and this takes us out of the modular group and into the situation of Möbius maps acting on $\mathbb{H}^3$. So again we need three-dimensional hyperbolic geometry!

Euclidean frieze groups

L L L	$\langle z+1 \rangle$
N N N	$\langle z+1, -z \rangle$
V V V	$\langle z+1, -\overline{z} \rangle$
▷ ▷ ▷	$\langle z+1, \overline{z} \rangle$
L Γ L Γ L Γ	$\langle \overline{z} + \frac{1}{2} \rangle$
O O O	$\langle z+1, -z, \overline{z} \rangle$
V Λ V Λ V Λ	$\langle \overline{z} + \frac{1}{2}, -\overline{z} \rangle$

Here is another application to number theory. Let us look at continued fractions, say

$$[b_0, b_1, \ldots, b_n] = b_0 + \cfrac{1}{b_1 + \cfrac{1}{b_2 + \cfrac{1}{\ddots}}}. \tag{1}$$

Typically, b_0 is an integer, and $b_1, b_2, \ldots$ are positive integers.

But why should the b_j be (i) positive, and (ii) integers?

In fact, there is **no reason** why they should be. There is a rich theory of complex continued fractions yet these are hardly ever mentioned (even in texts on continued fractions).

If we return to complex continued fractions (where the b_j are complex numbers) we see that continued fractions arise naturally from Möbius maps:

$$f_0(z) = b_0 + \frac{1}{z}, \quad f_1(z) = b_1 + \frac{1}{z},$$

so that

$$f_0 f_1(z) = b_0 + \cfrac{1}{b_1 + \cfrac{1}{z}}$$

and so on.

Therefore (I claim) continued fractions should be seen in the context of three-dimensional hyperbolic geometry – this is what the mathematics is telling us! It is also telling us that the b_j can be zero (recall that $1/0$ is acceptable in the theory of Möbius maps).

In fact, there are problems raised in the conventional continued fraction theory whose solution lies (apparently necessarily) in considering a continued fraction as a sequence of Möbius maps acting on hyperbolic 3-space $\mathbb{H}^3$, and then passing to the boundary $\mathbb{C}$ of $\mathbb{H}^3$ to obtain results in $\mathbb{C}$.

Hamilton studied continued fractions over the quaternions – albeit with not much success. As multiplication of quaternions is not commutative, this **is a reason** for not taking this further.

We give an illustration which links mathematical economics, set theory and topology. The classical view of consumer theory in economics was that if a consumer wanted a certain amount (x, y) of a pair of goods
for example, x amount of cheese, and y amount of wine,
then his/her preferences gave rise to a relation $\prec$, where $(x, y) \prec (x', y')$ means that

(x', y') *gives greater satisfaction than* (x, y).

Until 1954 it was assumed that this 'level of satisfaction' could be quantified by a real value function that increases in value as the level of satisfaction increase. A simple example is when the quantity (x, y) of goods has the level of satisfaction measured by the function xy. If we decrease one good, we can achieve the same level of satisfaction by increasing the amount of the other good.

Consumer's preferences

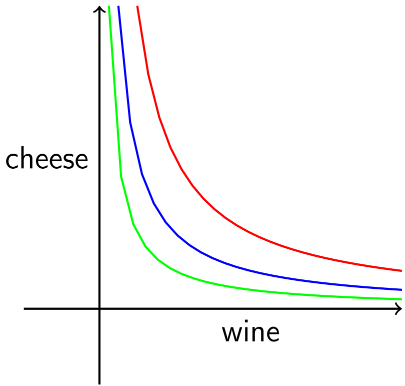

Figure: Typical indifference curves

In 1954, the French economist G. Debreu produced a simple example to show that we cannot always represent the preference relation by a real-valued function in this sense. As Debreu's example was based on a simple topological idea, this lead to a complete re-assessment of preference theory which is now governed by topology and the theory of binary relations and logic. In this case, mathematical economists did indeed take note of what the subject was telling them!

Debreu's example is very simple; it is the lexicographic ordering, where $(x, y) \prec (x', y')$ if $x' > x$, or $x' = x$ and $y' > y$. Informally, (x', y') is preferred to (x, y) if it lies to the right of (x, y), or is in the same vertical line as (x, y) but higher up.

The lexicographic ordering

NB: each point has its own level of satisfaction!

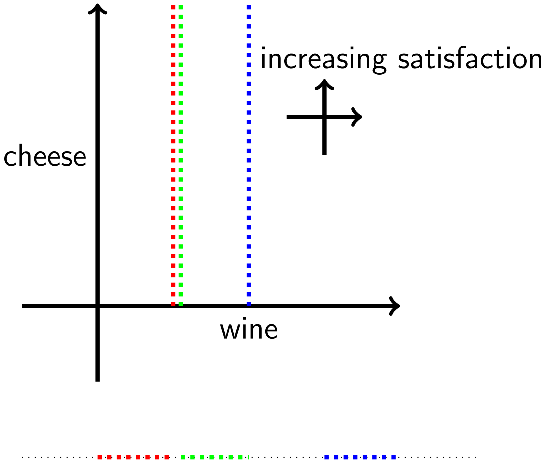

Figure: Debreu's example

It turns out that the problems associated with this area are related to the following problem in topology. Suppose that a particle moves along a path from A to B. Can it also move from A to B in a way that it is never stationary, and never returning to where it has been before? The answer is 'yes' in reasonable topological spaces, but the solution is not quite as elementary as one might imagine.

Intuitively, if the particle is stationary for a certain time interval, we could insist that it simply 'moves on' without waiting.

However,....

The Cantor function

The particle is ‘moving’ ONLY for a set of times of zero length.

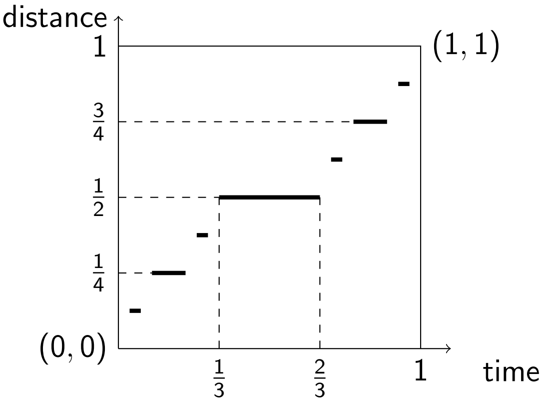

Figure: The Cantor function $\Phi\colon [0,1] \to [0,1]$

Thus one cannot simply collapse each time interval where it remains stationary to a point for, if one does, then the time remaining is zero!

This example is also relevant in probability theory where it gives rise to a probability distribution that has no discrete point probabilities, and no probability density function either.

Let me end on a more light-hearted note, addressed to the younger members of the audience.

I submitted my first paper for publication in 1962 so, as you can imagine, over the intervening years I have had a variety of responses from referees. Here are three of these responses.

This submitted manuscript was with the referee for **two** years:

"I apologize for working hard on this paper, which I found interesting, for such a long time, but it is not of sufficient interest to warrant publication."

(2) A referee's comment on a manuscript which answered a problem (in mathematical economics) which he himself had posed in a previous publication:

"If this is the answer then I am not interested in it.
I recommend rejection."

(3) On another submission the referee said
"This paper is not interesting."

This paper was subsequently awarded the prize for the best paper of the year in another journal.

My advice to young scientists is not to take the word of a referee, or indeed, and editor, at face value.
Referees, and Editors, are human like the rest of us!.

Thank you