Conjecture generation using Machine Intelligence

The University of Nottingham

Challenger Mishra, Computer Laboratory, Cambridge: July 2023

Hilbert's 23 problems

The thirteenth problem:

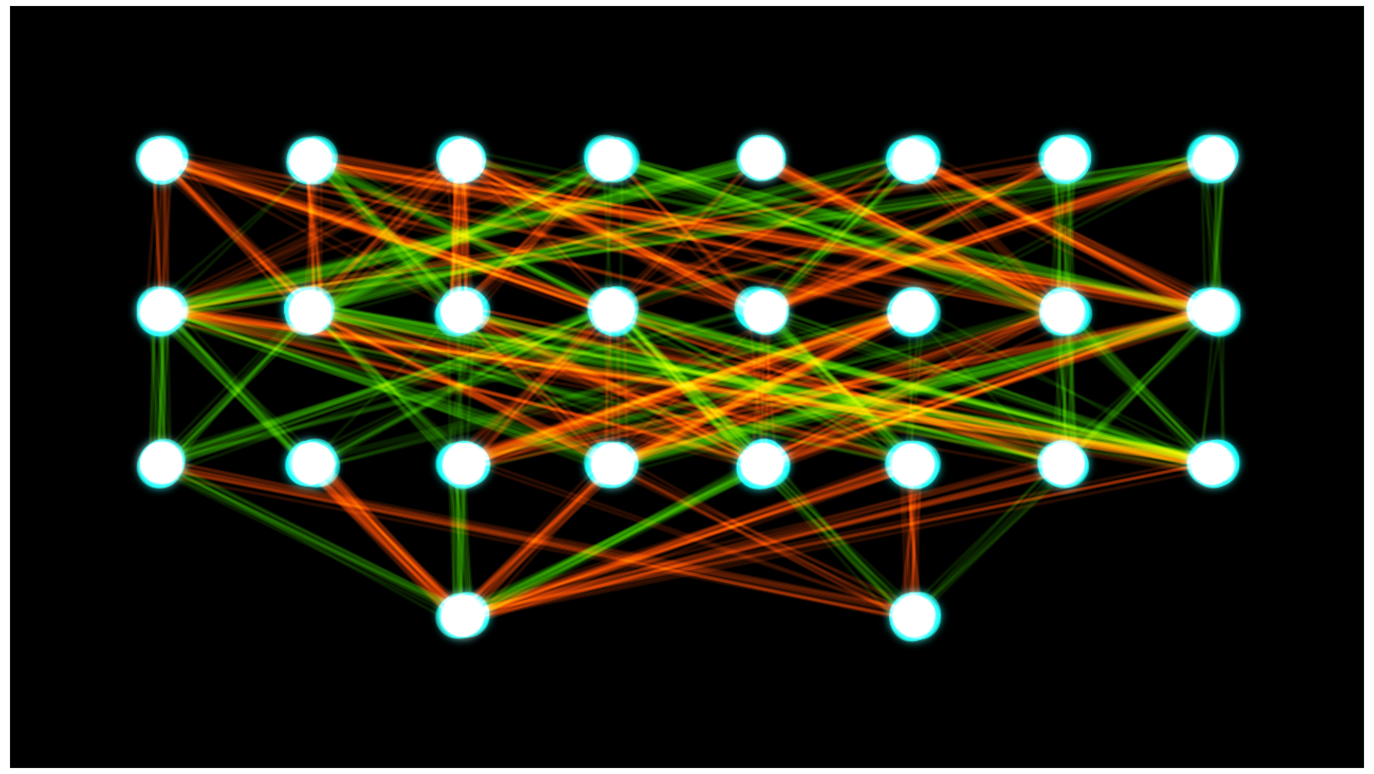
$x^7 + a x^3 + b x^2 + c x + 1 = 0.$

Can one write down the solution (x) as a composition of a finite number of 2 variable algebraic/continuous functions? – Hilbert (1900)

It is possible to construct any continuous function involving multiple variables using a finite set of three-variable functions [Kolmogorov, 1956]

Only two variable functions are required! [Arnold, 1957 (aged 19)]

A mathematical supernova



Neurocomputing

Kolmogorov Arnold representation

$$f(\mathbf{x}) = f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right).$$

$$f(\mathbf{x}) = \sum_{q=0}^{2n} \Phi \left(\sum_{p=1}^{n} \lambda_p \phi(x_p + \eta q) + q \right).$$

"The Kolmogorov theorem was discovered during a friendly mathematical duel between Kolmogorov and fellow Soviet Mathematician V.I. Arnol'd ... Kolmogorov won." - Robert Hecht-Nielsen

Hilbert's conjectures beg the question what constitutes a good conjecture?

Geometric Sampling Techniques –
Theo Long, MPhil, Cambridge

The subtle art of making Conjectures

Conway: Oberwolfach 1987 (with Hirzebruch)

Robbert Dijkgraaf: https://www.quantamagazine.org/the-subtle-art-of-the-mathematical-conjecture-20190507/

Good conjectures are milestones in mathematics. They are

1. nontrivial;
2. with potentially substantial evidence in favour of it (e.g., Goldbach's conjecture);
3. terse (e.g., Collatz conjecture);
4. can potentially unlock new theorems (e.g., RH);
5. "outrageous" — John Conway.

If a conjecture is proved within a few months, then perhaps its creator should have pondered it a bit longer before announcing it to the world. — Robbert Dijkgraaf (we might ignore this sagely advice atm).

Let's consider a class of conjectures ...

Conjectures on Inequalities

INEQUALITIES

By

G. H. HARDY

J. E. LITTLEWOOD

G. PÓLYA

It is often really difficult to trace the origin of a familiar inequality. It is quite likely to occur first as an auxiliary proposition, often without explicit statement, in a memoir on geometry or astronomy; it may have been rediscovered, many years later, by half a dozen different authors; and no accessible statement of it may be quite complete. We have almost always found, even with the most famous inequalities, that we have a little new to add.

CAMBRIDGE

AT THE UNIVERSITY PRESS

1934

$$f < g$$

Wide ranging applications result from an ability to bound functions in mathematics.

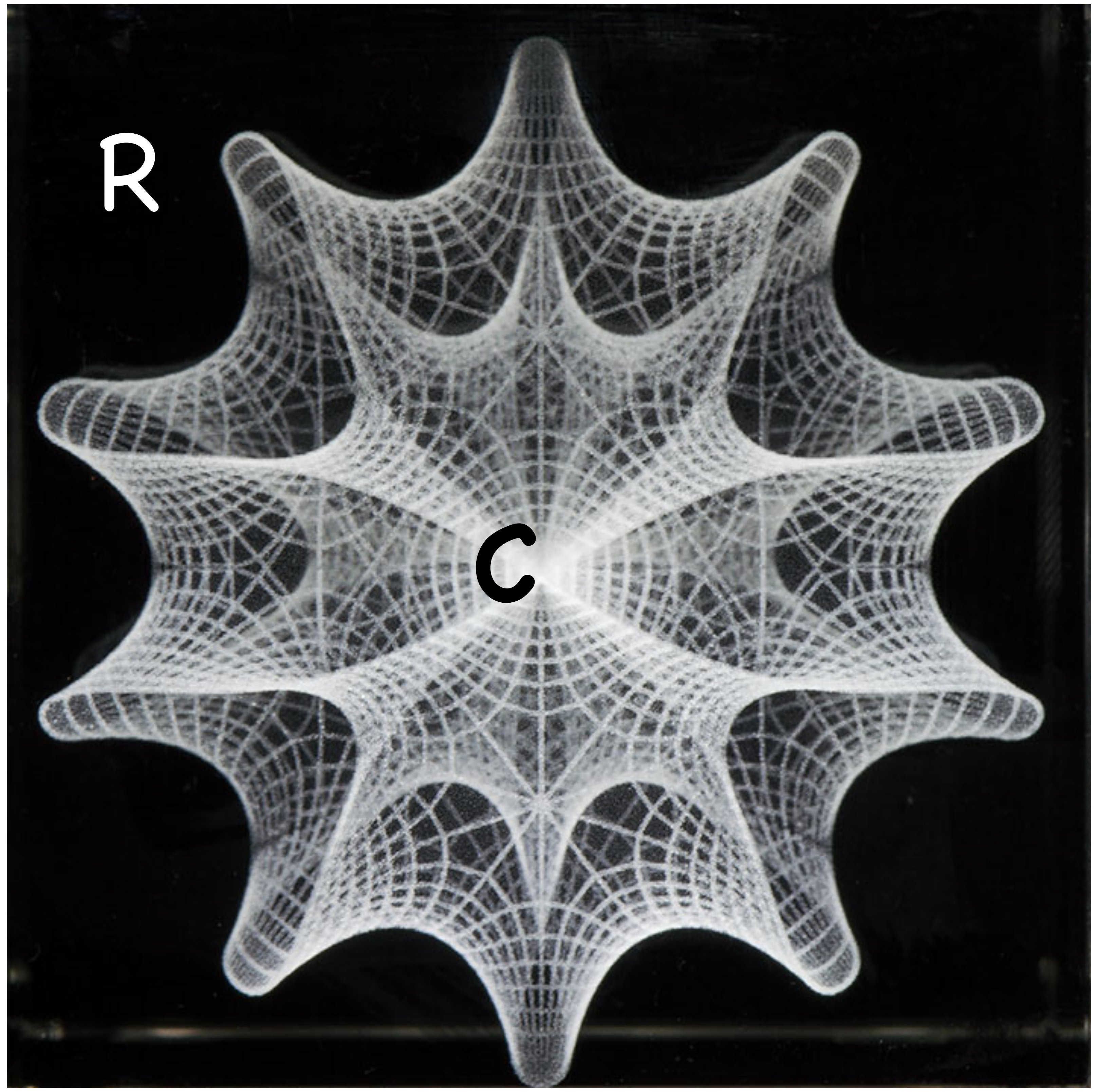
We aim to build an oracle that interacts with mathematicians to generate novel conjectures about inequalities which agree on a large amount of data. First we address the question: is there any structure in this space of relations ?

Symmetry as an organising principle for conjectures

Symmetries are a guiding principle for understanding nature through modern theoretical physics.

We now seek to understand whether tools from classical invariant theory help us give structure to a space of inequality relations.

The space of Relations and Conjectures



Let f, g be continuous real valued functions over a compact set D.

Space of relations (R): tuples (f, g).

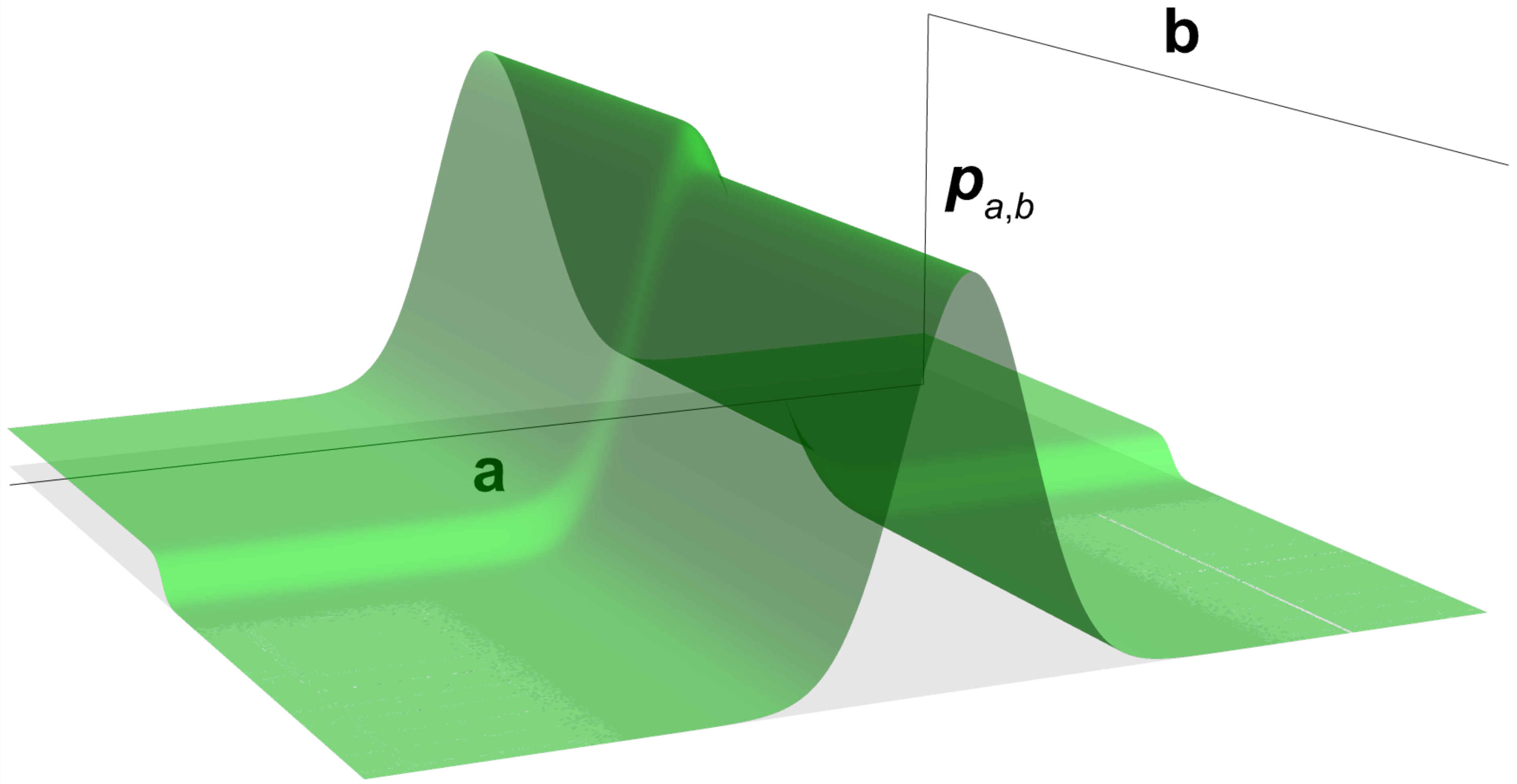
Space of conjectures (C): tuples (f, g) such that $f < g$ over D.

Note $(0,0) \in R$ but $(0,0) \notin C$.

We pose the following questions:

1. What is the largest group acting linearly on C?
2. Are there any free group actions on this space?

The space of conjectures (C)



A visualisation of a smooth approximation of $p(a,b) := 1 + \delta(a) - \text{sgn}(b)$. When the smooth approximation approaches $p(a,b)$, the conditions $a \neq 0$ and $b > 0$ are met strictly.

Consider a set of linear transformations acting on C.

$\mathscr{G} := \{A \in \mathbf{GL}(2,\mathbb{R}) : A(f,g) \in \text{C}, \forall (f,g) \in \text{C}\}$, where

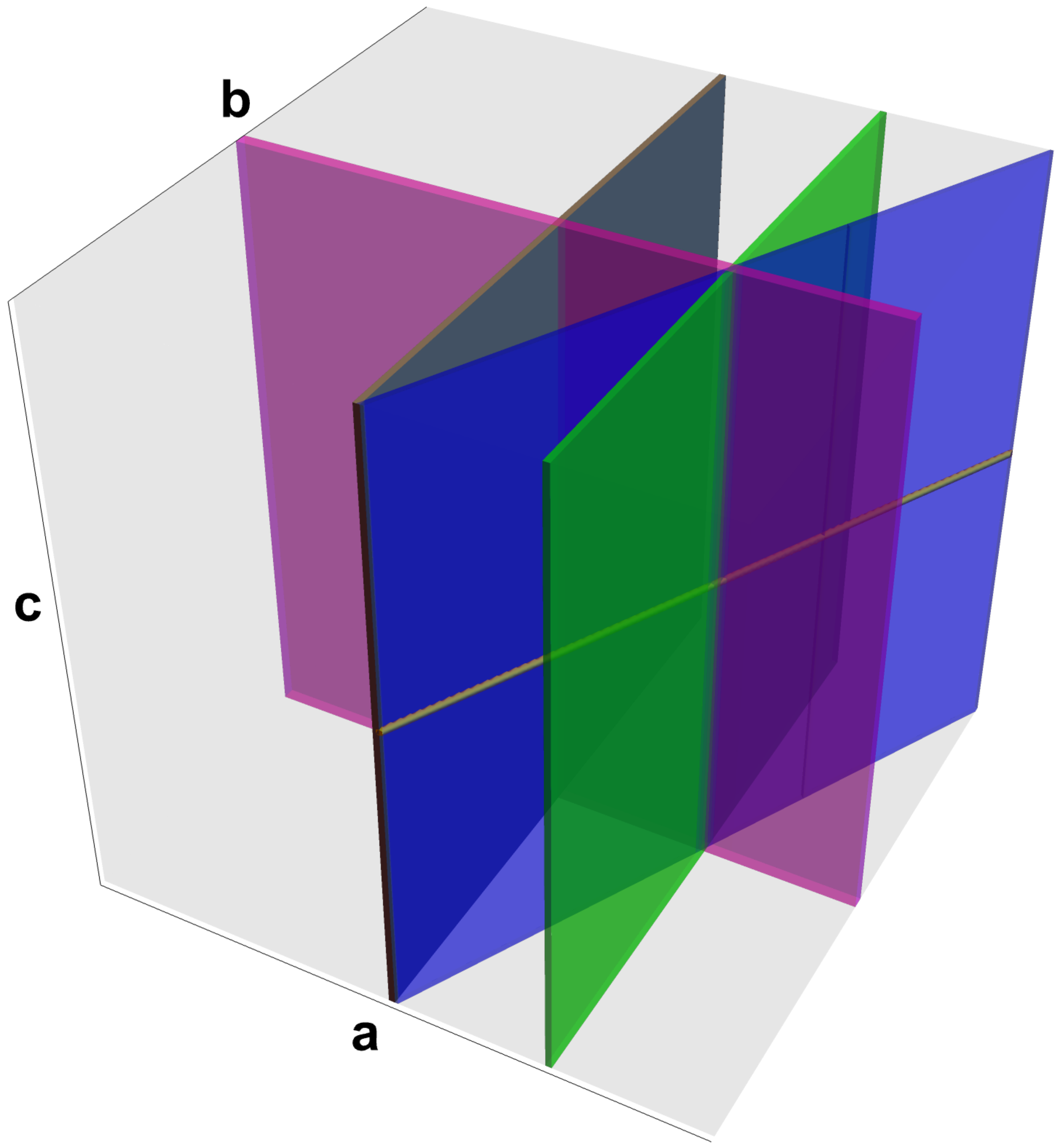
$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a & c \\ 0 & b \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \; ; \quad a, b, c \in \mathbb{R},\ a \neq 0,\ b > 0.$$

This forms a group and admits a semi direct product structure: $\mathscr{G} = \mathscr{T} \rtimes \mathscr{H}$, where $\mathscr{T}$ is a group of positive dilations parameterised by b, and $\mathscr{H}$ is the subgroup corresponding to b=1/2.

Group elements can be thought of as zeroes of the function $p(a,b) := 1 + \delta(a) - \text{sgn}(b)$.

Does this group or any subgroup act freely on C? If so, perhaps we could study quotients of C.

The space of conjectures (C)



Various subgroups of $\mathscr{G}$.

$\mathscr{H} := \mathscr{G}|_{b=1/2}$

$\mathscr{J}_1 := \mathscr{G}|_{a=b}$

$\mathscr{J}_2 := \mathscr{G}|_{a=\frac{1}{2}}$

$\mathscr{T} := \mathscr{G}|_{a=b,\ c=0}$

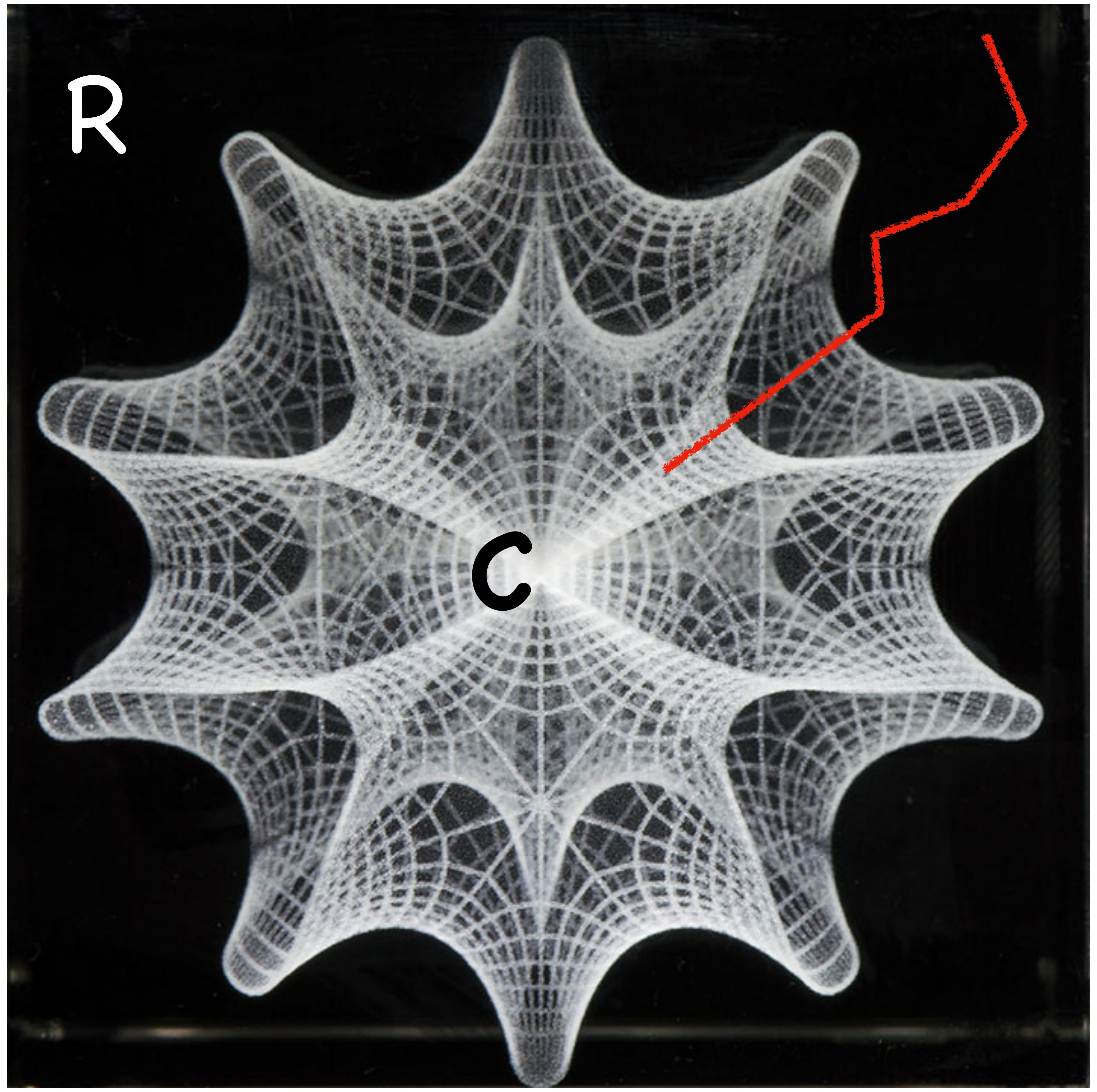
The subgroups $\mathscr{J}_1$, $\mathscr{J}_2$, and $\mathscr{T}$ act freely on C.

$\mathscr{J}_1$, and $\mathscr{J}_2$ are maximal.

These groups are matrix Lie groups.

These open up a number of questions, but the insights feed into the algorithm that powers the "oracle".

The Oracle



The conjecture space (C) can be seen as a subset of the space of relations (R).

This is a sampling problem! One approach is a naive search ala geometric gradient descent (our space admits a metric). Every point in C and its orbit are conjectures.

Algorithm 1 Oracle

Inputs: *Mathematical features* $\{x_i\}_{i=1}^{N}$, *function class* $\mathcal{F}_{d_1,d_2}$, and *hyperparameters*: tolerance (tol), batch-size (b), maximum epochs (emax), learning rate (η).
$\theta \leftarrow$ random real vector.
Parameterise: $c_\theta := (f_\theta, g_\theta) := \mathcal{P}(\theta)$; $f_\theta, g_\theta \in \mathcal{F}(K^{\mathcal{T}}[x])$.
for $a \leftarrow 1$ to emax **do**
    for $b \leftarrow 1$ to b **do**
        if $\mathcal{L}(c_\theta) \leq$ tol **then**
            return θ
        $\theta \leftarrow \theta - \eta\, \mathcal{J}_\theta^{-1}\, \nabla\mathcal{L}(\theta)$
        $\omega(c_\theta) \leftarrow \frac{2}{b}\sum_{i\in\text{rand}} \text{sgn}(f_\theta(x_i) - g_\theta(x_i))$
        $\mathcal{L}(c_\theta) \leftarrow (1 - \omega(c_\theta)^2)^2$
Output: $c_\theta = \mathcal{P}(\theta)$; $f_\theta < g_\theta$.

Some number theoretic conjectures!

#	Conjecture (c_θ)
1	$\pi(ab) \geq \pi(a) + \pi(b)$
2	$\pi(ab) + \pi(a+b) + \pi(a) + \pi(b) \leq 2ab$
3	$4(\pi(a) + \pi(b)) + \pi(a+b) \leq 4\pi(ab)$
4	$\pi(ab) + 2\pi(a+b) \geq \pi(a) + \pi(b)$
5	$\pi(a+b) \leq \pi(a) + \pi(b)$
6	$\pi(a\ b\ c) \geq \pi(a)\ \pi(b)\ \pi(c)$
$7^{\dagger}$	$\pi(a\ b\ c \ldots) \geq \pi(a)\ \pi(b)\ \pi(c) \ldots$
8	$\pi(a+b+c) \leq \pi(a) + \pi(b) + \pi(c)$
$9^{\dagger}$	$\pi(a+b+c+\ldots) \leq \pi(a) + \pi(b) + \pi(c) + \ldots$
10	$\pi(a)\pi(b)\pi(c) \leq \sqrt{(\pi(a)+\pi(b)+\pi(c))^2 + (\pi(abc))^2}$
$11^{\dagger\star}$	$(\ (\Sigma_i \pi(a_i))^2 + (\pi(\Pi_i a_i))^2)^{1/3} \leq \Pi_i \pi(a_i) \leq (\ (\Sigma_i \pi(a_i))^2 + (\pi(\Pi_i a_i))^2\ )^{1/2}$
12	$\pi(ab+bc+ca)^3 \geq 2\pi(abc)^2 + \pi(abc) + \pi(a+b+c)^2$
13	$\pi(a+b+c)^2 + \pi(abc) \geq \pi(ab+bc+ca)^3 + 2\pi(abc)^2$
14	$\pi(ab+bc+ca)^7 \geq \pi(a+b+c)^7$
15	$\pi\left(\alpha_1\alpha_2\alpha_3\right)^2 + \pi\left(\alpha_1\alpha_2 + \alpha_2\alpha_3 + \alpha_3\alpha_1\right) \geq \pi\left(\alpha_1\alpha_2 + \alpha_2\alpha_3 + \alpha_3\alpha_1\right)^3 + \pi\left(\alpha_1 + \alpha_2 + \alpha_3\right)$
16	$\pi\left(\alpha_1\alpha_2\alpha_3\right)^3 + 4\pi\left(\alpha_1\alpha_2\alpha_3\right)^2 + 4\pi\left(\alpha_1 + \alpha_2 + \alpha_3\right)^3 + 3 + ldots \geq \pi\left(\alpha_1\alpha_2 + \ldots\right)$
17°	$\pi\left(\chi_2\right)^3 - \pi\left(\chi_3\right)^2 + \pi\left(\chi_3\right) \geq \pi\left(\chi_1\right)$
18	$\pi\left(\chi_1\right)^3 \geq \pi\left(\chi_2\right)^3$
19	$5\pi\left(\chi_2\right)^3 \geq \pi\left(\chi_1\right)$
20	$\pi\left(\chi_1\right) + \pi\left(\chi_3\right) < \pi\left(\chi_2\right)$
21	$5\pi\left(\pi\left(\chi_1\right)\right) + 2\pi\left(\pi\left(\chi_3\right)\right) \geq \pi\left(\pi\left(\chi_2\right)\right)$
22	$\pi\left(\pi\left(\chi_2\right)\right) \geq 2\pi\left(\pi\left(\chi_3\right)\right) + 3\pi\left(\pi\left(\chi_1\right)\right)$
22	$11\left(\pi(ab) + \frac{ab}{\log(ab)}\right) > 9\pi(a+b) + \frac{9(a+b)}{\log(a+b)}$
23	$\pi\left(x + \sqrt{x}\right) \leq 3\pi\left(x\right) + 1$
$24^{\dagger\dagger}$	$\pi\left(x\right)^2 > x^3 + 2x + 2$
25	$\pi\left(x + \sqrt{x}\right) < 3\pi(x)$
26	$\pi\left(x + \sqrt{x}\right) < \frac{12}{5}\pi(x) + 1$

A theorem for the prime counting function

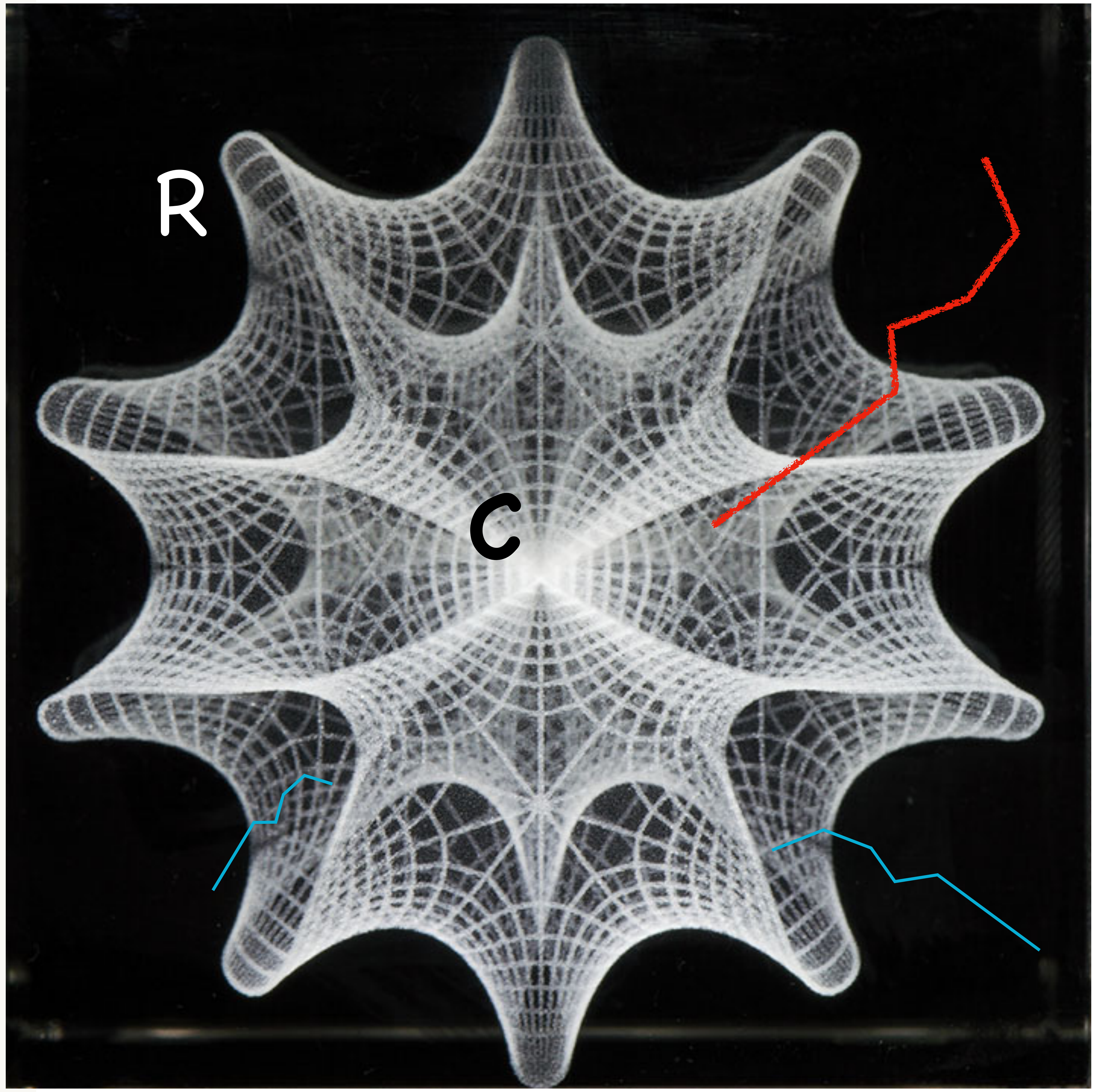
#	Conjecture (c_θ)
1	$\pi(ab) \geq \pi(a) + \pi(b)$
2	$\pi(ab) + \pi(a+b) + \pi(a) + \pi(b) \leq 2ab$
3	$4(\pi(a) + \pi(b)) + \pi(a+b) \leq 4\pi(ab)$
4	$\pi(ab) + 2\pi(a+b) \geq \pi(a) + \pi(b)$
5	$\pi(a+b) \leq \pi(a) + \pi(b)$
6	$\pi(a\ b\ c) \geq \pi(a)\ \pi(b)\ \pi(c)$
7†	$\pi(a\ b\ c \ldots) \geq \pi(a)\ \pi(b)\ \pi(c) \ldots$
8	$\pi(a+b+c) \leq \pi(a) + \pi(b) + \pi(c)$
9†	$\pi(a+b+c+\ldots) \leq \pi(a) + \pi(b) + \pi(c) + \ldots$

Theorem: $\pi(ab) \geq \pi(a) + \pi(b), \ \forall\, a, b \geq 17$

CM, RS, SRM (2023)

Second Hardy-Littlewood conjecture (1923)

Outlook



1. Representations: symbolic, HOL, NLP
2. Coupling with proof assistants
3. Quantum processor
4. Brining in machine architectures
5. Applications: Physical systems, Machine Learning, Number theory, Group theory, Random Matrix models, ...
6. Conjectures beget conjectures?
7. Rethinking mathematics education.

Limitations:

1. Curse of dimensionality!
2. Needs easily computable functions.
3. Human readability.

Mathematical conjecture generation using machine intelligence

Challenger Mishra, Subhayan Roy Moulik, Rahul Sarkar

Mathematical conjecture generation using machine intelligence

Challenger Mishra[1], Subhayan Roy Moulik[2], and Rahul Sarkar[3]

[1]The Computer Laboratory, University of Cambridge,Cambridge, CB3 0FD
[2]DAMTP, Centre for Mathematical Sciences, University of Cambridge, Cambridge, CB3 0WA
[3]Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305

cm2099@cam.ac.uk, sr2068@cam.ac.uk, rsarkar@stanford.edu

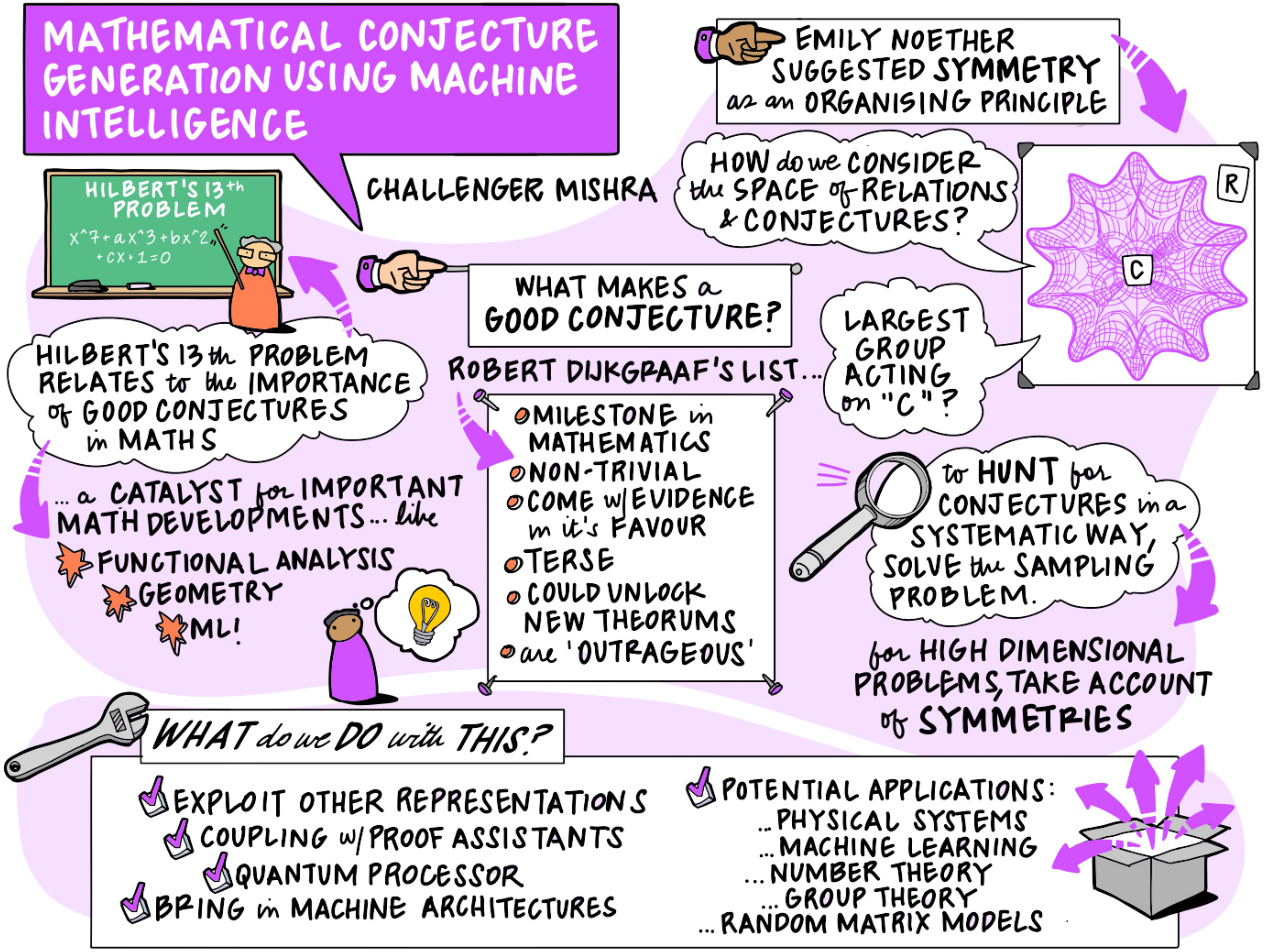
Abstract

Conjectures have historically played an important role in the development of pure mathematics. We propose a systematic approach to finding abstract patterns in mathematical data, in order to generate conjectures about mathematical inequalities, using machine intelligence. We focus on strict inequalities of type $f < g$ and associate them with a vector space. By geometerising this space, which we refer to as a *conjecture space*, we prove that this space is isomorphic to a Banach manifold. We develop a structural understanding of this *conjecture space* by studying linear automorphisms of this manifold and show that this space admits several free group actions. Based on these insights, we propose an algorithmic pipeline to generate novel conjectures using geometric gradient descent, where the metric is informed by the invariances of the *conjecture space*. As proof of concept, we give a toy algorithm to generate novel conjectures about the prime counting function and diameters of Cayley graphs of non-abelian simple groups. We also report private communications with colleagues in which some conjectures were proved, and highlight that some conjectures generated using this procedure are still unproven. Finally, we propose a pipeline of mathematical discovery in this space and highlight the importance of domain expertise in this pipeline.

Subhayan Roy Moulik, Dept of Applied Mathematics and Theoretical Physics, Cambridge

Rahul Sarkar, Institute for Computational and Mathematical Engineering, Stanford

Being humbled in Chess @Berkeley, 2023



MATHEMATICAL CONJECTURE GENERATION USING MACHINE INTELLIGENCE
CHALLENGER MISHRA
HILBERT'S 13th PROBLEM
X^7+aX^3+bX^2 +CX+1=0
HILBERT'S 13th PROBLEM RELATES to the IMPORTANCE of GOOD CONJECTURES in MATHS
... a CATALYST for IMPORTANT MATH DEVELOPMENTS... like
FUNCTIONAL ANALYSIS
GEOMETRY
ML!
WHAT MAKES a GOOD CONJECTURE?
ROBERT DIJKGRAAF'S LIST...
MILESTONE in MATHEMATICS
NON-TRIVIAL
COME w/ EVIDENCE in it's FAVOUR
TERSE
COULD UNLOCK NEW THEORUMS
are 'OUTRAGEOUS'
EMILY NOETHER SUGGESTED SYMMETRY as an ORGANISING PRINCIPLE
HOW do we CONSIDER the SPACE of RELATIONS & CONJECTURES?
R
C
LARGEST GROUP ACTING on "C"?
to HUNT for CONJECTURES in a SYSTEMATIC WAY, SOLVE the SAMPLING PROBLEM.
for HIGH DIMENSIONAL PROBLEMS, TAKE ACCOUNT of SYMMETRIES
WHAT do we DO with THIS?
EXPLOIT OTHER REPRESENTATIONS
COUPLING w/ PROOF ASSISTANTS
QUANTUM PROCESSOR
BRING in MACHINE ARCHITECTURES
POTENTIAL APPLICATIONS:
...PHYSICAL SYSTEMS
...MACHINE LEARNING
...NUMBER THEORY
...GROUP THEORY
...RANDOM MATRIX MODELS

Let's make some conjectures?