The Collatz conjecture

Anna

Tea talk - 12th September 2019

Collatz map and Collatz conjecture

Let $n \in \mathbb{N}$. Define the **Collatz map**:

if n is even $n \Longrightarrow Col(n) = \frac{n}{2}$

if n is odd $n \Longrightarrow Col(n) = 3n + 1$

What happens then? does it grows to infinity? does it get small? Let's see what happens with 7.

Collatz map and Collatz conjecture

Let $n \in \mathbb{N}$. Define the **Collatz map**:

if n is even $n \Longrightarrow Col(n) = \frac{n}{2}$

if n is odd $n \Longrightarrow Col(n) = 3n + 1$

What happens then? does it grows to infinity? does it get small? Let's see what happens with 7.

People have tried lots of numbers beginning this way, and so far all have gone to 1 (up to 10^{20})

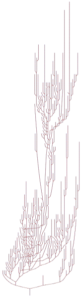
This leads to the **Collatz conjecture** (1937):
every number will "fall" on 1:
$Col_{\min}(n) = \inf_{j \in \mathbb{N}} Col^j(n) = \inf\{Col(n), Col^2(n)...\} = 1$

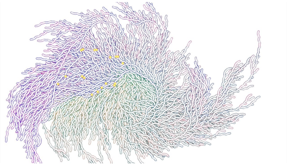
Collatz trajectories

Some numbers have particularly nice trajectories: e.g. $n = 27$ goes up to 9232, but after 111 steps ("flight duration") it falls onto 1.

"Hailstone numbers": patterns the rise and fall of hailstones in the clouds.

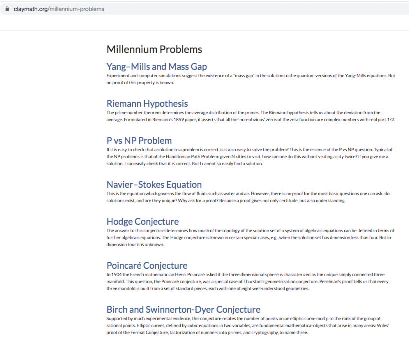
`http://l.pellegrino.free.fr/syracuse/index.php`





on the left: Orbits of the first 1000 numbers. *on the right:* Another visualization for n's smaller than 10000 (does it help understanding why no one's got a clue?)

Interest around Collatz conjecture



claymath.org/millennium-problems

Millennium Problems

Yang–Mills and Mass Gap

Experiment and computer simulations suggest the existence of a "mass gap" in the solution to the quantum versions of the Yang-Mills equations. But no proof of this property is known.

Riemann Hypothesis

The prime number theorem determines the average distribution of the primes. The Riemann hypothesis tells us about the deviation from the average. Formulated in Riemann's 1859 paper, it asserts that all the 'non-obvious' zeros of the zeta function are complex numbers with real part 1/2.

P vs NP Problem

If it is easy to check that a solution to a problem is correct, is it also easy to solve the problem? This is the essence of the P vs NP question. Typical of the NP problems is that of the Hamiltonian Path Problem: given N cities to visit, how can one do this without visiting a city twice? If you give me a solution, I can easily check that it is correct. But I cannot so easily find a solution.

Navier–Stokes Equation

This is the equation which governs the flow of fluids such as water and air. However, there is no proof for the most basic questions one can ask: do solutions exist, and are they unique? Why ask for a proof? Because a proof gives not only certitude, but also understanding.

Hodge Conjecture

The answer to this conjecture determines how much of the topology of the solution set of a system of algebraic equations can be defined in terms of further algebraic equations. The Hodge conjecture is known in certain special cases, e.g., when the solution set has dimension less than four. But in dimension four it is unknown.

Poincaré Conjecture

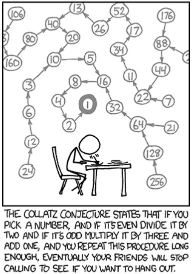
In 1904 the French mathematician Henri Poincaré asked if the three dimensional sphere is characterized as the unique simply connected three manifold. This question, the Poincaré conjecture, was a special case of Thurston's geometrization conjecture. Perelman's proof tells us that every three manifold is built from a set of standard pieces, each with one of eight well-understood geometries.

Birch and Swinnerton-Dyer Conjecture

Supported by much experimental evidence, this conjecture relates the number of points on an elliptic curve mod p to the rank of the group of rational points. Elliptic curves, defined by cubic equations in two variables, are fundamental mathematical objects that arise in many areas: Wiles' proof of the Fermat Conjecture, factorization of numbers into primes, and cryptography, to name three.

It is **the** simplest open problem in mathematics, yet still not solved.

Paul Erdös : *"Mathematics is not yet ready for such problem"*.



Why would we care about this conjecture?

- It is expected to answer to the question: *In what ways does the prime factorization of n affects the prime factorization of $n+1$?*
  (adding 1 "shuffles" the prime factors). Virtually all of modern security relies upon the current limitations of the understanding of prime numbers.

Why would we care about this conjecture?

- It is expected to answer to the question: *In what ways does the prime factorization of n affects the prime factorization of $n+1$?*
  (adding 1 "shuffles" the prime factors). Virtually all of modern security relies upon the current limitations of the understanding of prime numbers.
- Even if the problem does not seem important, it might bring new important branches in mathematics:
  Example: Fermat's last theorem $\nexists a, b, c \in \mathbb{N}^*$ s.t. $a^n + b^n = c^n$ for $n \geq 2$. Wiles's proof in 1994 involved new techniques that revolutionized number theory and implied new results on elliptic curves

Why would we care about this conjecture?

- It is expected to answer to the question: *In what ways does the prime factorization of n affects the prime factorization of $n+1$?*
  (adding 1 "shuffles" the prime factors). Virtually all of modern security relies upon the current limitations of the understanding of prime numbers.
- Even if the problem does not seem important, it might bring new important branches in mathematics:
  Example: Fermat's last theorem $\nexists a, b, c \in \mathbb{N}^*$ s.t. $a^n + b^n = c^n$ for $n \geq 2$. Wiles's proof in 1994 involved new techniques that revolutionized number theory and implied new results on elliptic curves
- Some mathematicians think it might be an example of a statement unprovable with the current axioms of mathematics

Two subproblems

Collatz's conjecture is that every trajectory ends on 1. It is natural to ask ourselves what is the long-term behavior of a trajectory. There are thus two subproblems (none of them solved yet)

1. is there any divergent trajectory? (going to ∞)
2. If a trajectory is upperbounded, it contains repetitions and get stuck in a cycle. Is $1 \implies 4 \implies 2$ the only cycle?

Some known results

- It has been proved that the Collatz map admits no cycle of length between 4 and around 17 000 000 000 on $\mathbb{N}^+$.

  There are three (maybe the only ones) known cycles on the negative integers:

  - trajectory of -1 is a cycle of length 2: $-1 \Longrightarrow 3 \times (-1) = -2 \Longrightarrow -1$
  - trajectory of -5 is a cycle of length 5: $-5 \Longrightarrow -14 \Longrightarrow -7 \Longrightarrow -20 \Longrightarrow -10 \Longrightarrow -5$
  - -17 gives a cycle of length 18

Some known results

- It has been proved that the Collatz map admits no cycle of length between 4 and around 17 000 000 000 on $\mathbb{N}^+$.

  There are three (maybe the only ones) known cycles on the negative integers:

  - trajectory of -1 is a cycle of length 2: $-1 \Longrightarrow 3 \times (-1) = -2 \Longrightarrow -1$
  - trajectory of -5 is a cycle of length 5: $-5 \Longrightarrow -14 \Longrightarrow -7 \Longrightarrow -20 \Longrightarrow -10 \Longrightarrow -5$
  - -17 gives a cycle of length 18

- what about other related maps? e.g. Conway's generalized maps: $g(n) = a_i n + b_i, n \equiv i \pmod{P}$ where $a_0, b_0, \ldots, a_{P-1}, b_{P-1} \in \mathbb{Q}$.

  It was shown that this family is undecidable (=some elements in the family are): no algorithm can take as input a Collatz-like function and decide yes/no to whether every integer iterates to 1 under the inputted Collatz function.

  But it does not say anything on the decidability of the particular Collatz problem...

Some news this week

Terence Tao (2006 Field's medal) came with new results this week.

ALMOST ALL ORBITS OF THE COLLATZ MAP ATTAIN ALMOST BOUNDED VALUES

TERENCE TAO

ABSTRACT. Define the *Collatz map* Col: $\mathbb{N}+1 \to \mathbb{N}+1$ on the positive integers $\mathbb{N}+1 = \{1, 2, 3, \dots\}$ by setting $\mathrm{Col}(N)$ equal to $3N+1$ when N is odd and $N/2$ when N is even, and let $\mathrm{Col}_{\min}(N) := \inf_{n \in \mathbb{N}} \mathrm{Col}^n(N)$ denote the minimal element of the Collatz orbit $N, \mathrm{Col}(N), \mathrm{Col}^2(N), \dots$. The infamous *Collatz conjecture* asserts that $\mathrm{Col}_{\min}(N) = 1$ for all $N \in \mathbb{N}+1$. Previously, it was shown by Korec that for any $\theta > \frac{\log 3}{\log 4} \approx 0.7924$, one has $\mathrm{Col}_{\min}(N) \leq N^\theta$ for almost all $N \in \mathbb{N}+1$ (in the sense of natural density). In this paper we show that for *any* function $f\colon \mathbb{N}+1 \to \mathbb{R}$ with $\lim_{N \to \infty} f(N) = +\infty$, one has $\mathrm{Col}_{\min}(N) \leq f(N)$ for almost all $N \in \mathbb{N}+1$ (in the sense of logarithmic density). Our proof proceeds by establishing an approximate transport property for a certain first passage random variable associated with the Collatz iteration (or more precisely, the closely related Syracuse iteration), which in turn follows from estimation of the characteristic function of a certain skew random walk on a 3-adic cyclic group at high frequencies. This estimation is achieved by studying how a certain two-dimensional renewal process interacts with a union of triangles associated to a given frequency.

on the left: Terence Tao with Paul Erdös in 1985. *on the right:* Terence's paper (8th September 2019)

Theorem 2 Let $f : \mathbf{N} + 1 \to \mathbf{R}$ be any function with $\lim_{N\to\infty} f(N) = +\infty$. Then we have $\mathrm{Col}_{\min}(N) < f(N)$ for almost all N (in the sense of logarithmic density).

Definition

An almost bounded grows more slowly than any function that tends to infinity.

For example, $N^{0.00000001}$, or $\log\log\log\log N$ grow "slowly" but they tend towards infinity as N gets larger. An almost bounded function grows more slowly than both of them by definition (it can also be bounded).

Definition

Let $A \subset \mathbb{N}$ set of integers and x a cutoff. The natural density and log density of A are respectively:

$$\Delta(A) = \lim_{x\to\infty} \frac{|A \cap [1, x]|}{x} \quad \delta(A) = \lim_{x\to\infty} \frac{\sum_{n\in A, n\leq x} \frac{1}{n}}{\log x}$$

So we still don't have a "for all N" result...