1932

The $3x + 1$ Problem

Introduction

The *Collatz function* $T : \mathbb{N} \to \mathbb{N}$ is defined by

$$T(x) = \begin{cases} \dfrac{x}{2} & \text{if } x \text{ is even,} \\ 3x + 1 & \text{if } x \text{ is odd.} \end{cases}$$

Now pick a *seed*, a natural number n, and consider the corresponding *Collatz sequence* $n, T(n), T^2(n), \ldots$, in which $T^k(n)$ denotes the k-fold iterate $T(T(\cdots(T(n))))$. This is also called the *orbit* of n under T. For example, $n = 21$ yields the Collatz sequence

$$21, \quad 64, \quad 32, \quad 16, \quad 8, \quad 4, \quad 2, \quad 1, \quad 4, \quad 2, \quad 1, \quad 4, \quad 2, \quad 1, \quad \ldots$$

and $n = 24$ provides

$$24, \quad 12, \quad 6, \quad 3, \quad 10, \quad 5, \quad 16, \quad 8, \quad 4, \quad 2, \quad 1, \quad 4, \quad 2, \quad 1, \quad \ldots.$$

Both sequences eventually settle down to the repeating pattern $4, 2, 1, 4, 2, 1, \ldots$, a periodic orbit of period three.

It appears that for every initial seed n, the Collatz sequence eventually reaches the number 1; that is, every Collatz sequence ends with $4, 2, 1, 4, 2, 1, \ldots$. Proving this is the famed $3x+1$ *problem* (or the $3x+1$ *conjecture*), often credited to Lothar Collatz (1910–1990). It goes by an astounding number of other names as well: Ulam's conjecture, Kakutani's problem, the Thwaites conjecture, and the Syracuse problem. We are not going to debate the origins of this problem and are content to call it the $3x + 1$ problem.

One way to visualize the $3x + 1$ problem is with a directed graph; see Figure 1. Each natural number is a vertex in the *Collatz graph* and there is an arrow from j to k whenever $T(j) = k$. The $3x + 1$ conjecture asserts that no matter which vertex you start on, following the arrows in the Collatz graph always leads to $4, 2, 1, 4, 2, 1, \ldots$.

Some seeds take a long time to reach 1. For example, $n = 27$ requires 111 iterations. Its Collatz sequence climbs all the way up to 9,232 before coming back down (see the notes for 1929 for an example of another sequence that demonstrates this sort of behavior). This highlights one of our main obstacles: there is no simple way to predict how high the Collatz sequence for a given seed reaches. As of 2015, the $3x + 1$ conjecture has been verified for all seeds less than 2^{60}. Although this is an overwhelming amount of numerical evidence, it is not a proof (see the notes for 1930 for examples of misleading computations).

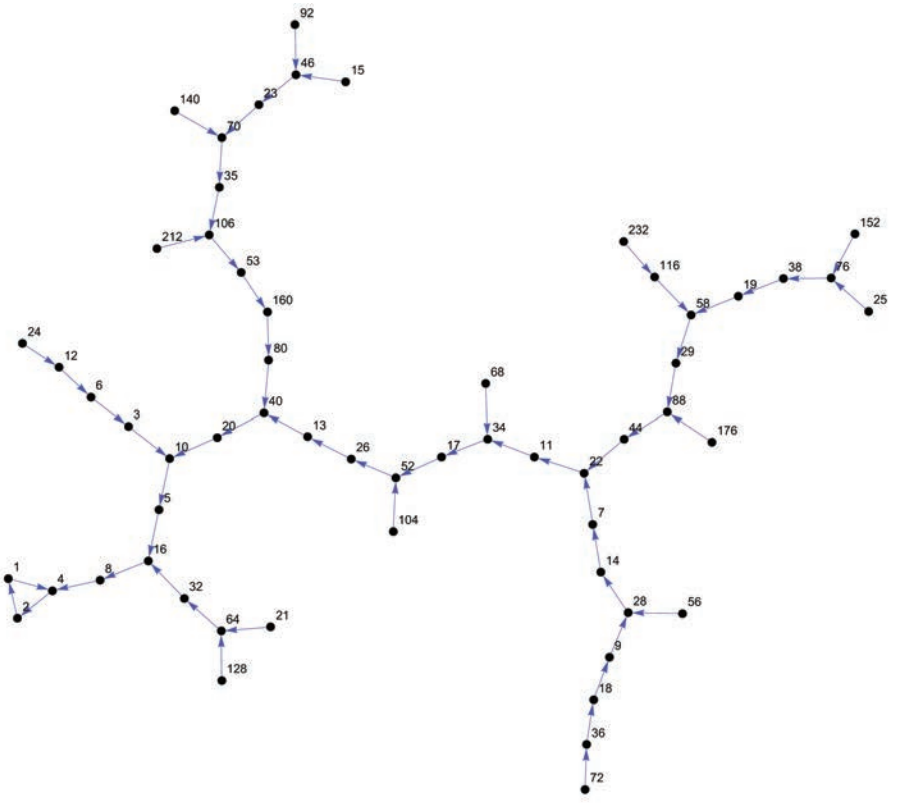


FIGURE 1. The map $T : \mathbb{N} \to \mathbb{N}$ induces a directed graph on the natural numbers. A portion of this graph is represented here. The $3x + 1$ problem suggests that any starting point eventually leads to the cycle $1, 4, 2, 1, 4, 2, \ldots$.

One can prove that the distribution of digits converges, in an appropriate sense, to Benford's law on digit bias (see the 1938 entry and [**4**, **7**]). That is, if you take a large starting seed and look at all the iterates until it hits the cycle $4, 2, 1$, then with probability tending to 1 the digit distribution converges to Benford's law. One can make this precise: given any tolerance, the number of starting seeds in an interval $[1, X]$ that are more than this tolerance away from Benford tends to zero as $X \to \infty$.

How can one disprove the $3x + 1$ conjecture? There are two ways in which the conjecture could be false. There might be a seed whose Collatz sequence is unbounded. Or there might be a periodic orbit other than $4, 2, 1$ (it is known that there are no other periodic orbits of length 100,000,000 or less [**3**]).

Some have described the $3x + 1$ problem as a Soviet conspiracy to slow down American mathematics since so many people tried working on it, tempted by its apparent simplicity. Paul Erdős said that mathematics is not yet ready to address questions such as the the $3x + 1$ problem.

Centennial Problem 1932

Proposed by Jeffrey Lagarias, University of Michigan.

Here we consider the original function $G : \mathbb{Z} \to \mathbb{Z}$, defined by

$$G(3n) = 2n, \qquad G(3n+1) = 4n+1, \quad \text{and} \quad G(3n+2) = 4n+3,$$

that Collatz wrote down on July 1, 1932. It is a permutation of $\mathbb{Z}$ and its inverse is given by

$$G^{-1}(2n) = 3n, \qquad G^{-1}(4n+1) = 3n+1, \quad \text{and} \quad G^{-1}(4n+3) = 3n+2.$$

One can show that G maps $\mathbb{N}$ onto $\mathbb{N}$, so it induces a permutation of $\mathbb{N}$ too. One finds that $G(1) = 1$ is a fixed point, that $G(2) = 3$ and $G(3) = 2$ form a periodic orbit of period 2, and that

$$G(4) = 5, \qquad G(5) = 7, \qquad G(7) = 9, \qquad G(9) = 6, \quad \text{and} \quad G(6) = 4$$

form a periodic orbit of period 5.

(a) What happens for $n = 8$? Computation indicates that the *forward orbit* $\{G^k(n) : k \geq 0\}$ of $n = 8$ includes numbers larger than 10^{400}. But is the orbit infinite? This question is the original Collatz problem and it has been proposed independently several times, starting with Murray S. Klamkin (1921–2004) in 1963 [**2**]. It too is unsolved and could be as hopeless as the $3x + 1$ problem.

(b) For $N = 1, 2, \ldots$, let

$$S_N \;=\; \{n \in [1, N] : G^k(8) = n \text{ for some } k \in \mathbb{Z}\}.$$

It is conjectured that

$$\lim_{N\to\infty} \frac{|S_N|}{N} \;=\; 0. \tag{1932.1}$$

Probabilistic models suggest that $|S_N| = O(\log N)$ as $N \to \infty$ and computer experiments support this. So there seems to be "room to spare" in trying to establish (1932.1). Nevertheless, this problem seems difficult. The reader is warned.

(c) Consider the full forward and backward orbit of $n = 8$:

$$S_\infty \;=\; \{n \in \mathbb{N} : G^k(8) = n \text{ for some } k \in \mathbb{Z}\}.$$

Disprove that there are only finitely many natural numbers that are not in S_∞. This assertion sounds simple to resolve and it is much weaker than (1932.1). Nevertheless, it is an open problem and may be as intractable as the $3x + 1$ problem.

1932: Comments

A heuristic approach. When stuck on a difficult conjecture, one can try to give heuristic arguments for or against its validity. To simplify our model, we omit the troublesome +1 in the definition of the Collatz function. Since half of the even numbers are divisible by 2 and not by 4, and a fourth are divisible by 4 and not by 8, and so on, we consider the functions $H_2(x) = 3x/2$, $H_4(x) = 3x/4$, $H_8(x) = 3x/8$, and so forth. Our heuristic approximation to the Collatz function is denoted H; it is obtained by applying H_{2^k} with probability $1/2^k$ for $k = 1, 2, \ldots$. The hope is that this related problem is easier to analyze and that its behavior will shed light on the original problem.

It is more appropriate to consider the expected value of $\log H(x)$ since there are products involved. According to our model,[1]

$$\begin{aligned}\mathbb{E}[\log H(x)] &= \sum_{k=1}^{\infty} \frac{1}{2^k} \log H_{2^k}(x) = \sum_{k=1}^{\infty} \frac{\log(3x/2^k)}{2^k} \\ &= \log x + \left(\log 3 - \sum_{k=1}^{\infty} \frac{k \log 2}{2^k} \right) = \log x + \log(3/4) \\ &< \log x.\end{aligned}$$

Consequently, iterating H once decreases the size of the expected outcome. Repeated iterations should continue to decrease. Not only does such an argument lead to heuristic support for the $3x+1$ conjecture, it also suggests roughly how many steps one needs to iterate until we reach 1. Since each iteration tends to replace x with $\frac{3}{4}x$, the expected number of iterations should satisfy $(3/4)^m x = 1$; that is,

$$m \approx \frac{\log x}{\log 4/3}.$$

Numerical data strongly supports this rate; see [**5**, **6**] for more on these ideas.

The idea of replacing a deterministic problem with a random one is applicable in many other settings. One can do this with prime numbers to build intuition about a host of problems. However, one must be careful. Just as the $3x+1$ problem has some structure that is lost in the conversion to a random model, the actual sequence of primes has additional structure not present in random analogues. While random models are useful, they sometimes give the wrong answer in certain regimes.

Lychrel numbers. We end with an example that leads to another simply stated open problem. Consider the function $L : \mathbb{N} \to \mathbb{N}$ defined by $L(n) = n + R(n)$, in which $R(n)$ is the number formed by reversing the decimal representation of n. For instance, $L(89) = 89 + 98 = 187$,

$$L^2(89) = L(187) = 187 + 781 = 968,$$

and so forth. This leads to the following sequence:

89, 187, 968, 1837, 9218, 17347, 91718, 173437, 907808, 1716517,
8872688, 17735476, 85189247, 159487405, 664272356, 1317544822,
3602001953, 7193004016, 13297007933, 47267087164, 93445163438,
176881317877, 955594506548, 1801200002107, 8813200023188....

The number $L^{24}(89)$ is the *palindrome* 8,813,200,023,188; it is the same read forward or backward. Most natural numbers eventually appear to reach a palindrome after repeated applications of L.

A *Lychrel number* is a natural number for which this process never yields a palindrome. Brute force computations show that no $n \leq 195$ is a Lychrel number, but no one is sure about 196 (this leads to an alternative name for this iteration: the *196-algorithm*). Nobody knows whether Lychrel numbers exist, but 196 sure

[1] To sum $\sum_{k=1}^{\infty} \frac{k}{2^k}$, differentiate the identity $\sum_{n=0}^{\infty} z^n = (1-z)^{-1}$, valid for $|z| < 1$, multiply the result by z, and obtain $\sum_{n=1}^{\infty} n z^n = z/(1-z)^2$. Then substitute $z = 1/2$.

looks like a strong candidate:

196, 887, 1675, 7436, 13783, 52514, 94039, 187088, 1067869, 10755470, 18211171, 35322452, 60744805, 111589511, 227574622, 454050344, 897100798, 1794102596, 8746117567, 16403234045, 70446464506, 130992928913, 450822227944, 900544455998....

Over a billion iterates have been computed without reaching a palindrome. Extensive computation suggests that the following integers are Lychrel numbers [**9**]:

196, 295, 394, 493, 592, 689, 691, 788, 790, 879, 887, 978, 986, 1495, 1497, 1585, 1587, 1675, 1677, 1765, 1767, 1855, 1857, 1945, 1947, 1997, 2494, 2496, 2584, 2586, 2674, 2676, 2764, 2766, 2854, 2856, 2944, 2946, 2996, 3493, 3495, 3583, 3585, 3673, 3675.

Curiously, Lychrel numbers are known to exist in other bases. For example, in binary the number 10110 (which is 22 in decimal) is a Lychrel number. Can you prove it?

Bibliography

[1] R. K. Guy, *Don't try to solve these problems!*, Amer. Math. Monthly **90** (1983), 35–41. `http://www.jstor.org/discover/10.2307/2975688?uid=3739256&uid=2&uid=4&sid=21102550539183`.

[2] M. S. Klamkin, *Problem 63-13**, SIAM Review **5** (1963), 275–276.

[3] L. Halbeisen and N. Hungerbühler, *Optimal bounds for the length of rational Collatz cycles*, Acta Arith. **78** (1997), no. 3, 227–239, DOI 10.4064/aa-78-3-227-239. MR1432018

[4] A. V. Kontorovich and S. J. Miller, *Benford's law, values of L-functions and the* $3x + 1$ *problem*, Acta Arith. **120** (2005), no. 3, 269–297, DOI 10.4064/aa120-3-4. `http://arxiv.org/pdf/math/0412003v2`. MR2188844

[5] J. C. Lagarias, *The* $3x + 1$ *problem and its generalizations*, Amer. Math. Monthly **92** (1985), no. 1, 3–23, DOI 10.2307/2322189. MR777565

[6] J. C. Lagarias (ed.), *The ultimate challenge: the* $3x + 1$ *problem*, American Mathematical Society, Providence, RI, 2010. MR2663745

[7] J. C. Lagarias and K. Soundararajan, *Benford's law for the* $3x + 1$ *function*, J. London Math. Soc. (2) **74** (2006), no. 2, 289–303, DOI 10.1112/S0024610706023131. `http://arxiv.org/pdf/math/0509175.pdf`. MR2269630

[8] H. L. Montgomery and K. Soundararajan, *Primes in short intervals*, Comm. Math. Phys. **252** (2004), no. 1-3, 589–617, DOI 10.1007/s00220-004-1222-4. MR2104891

[9] The On-Line Encyclopedia of Integer Sequences, *A023108 (Positive integers which apparently never result in a palindrome under repeated applications of the function* $f(x) = x + (x$ *with digits reversed*), `http://oeis.org/A023108`.

1933

Skewes's Number

Introduction

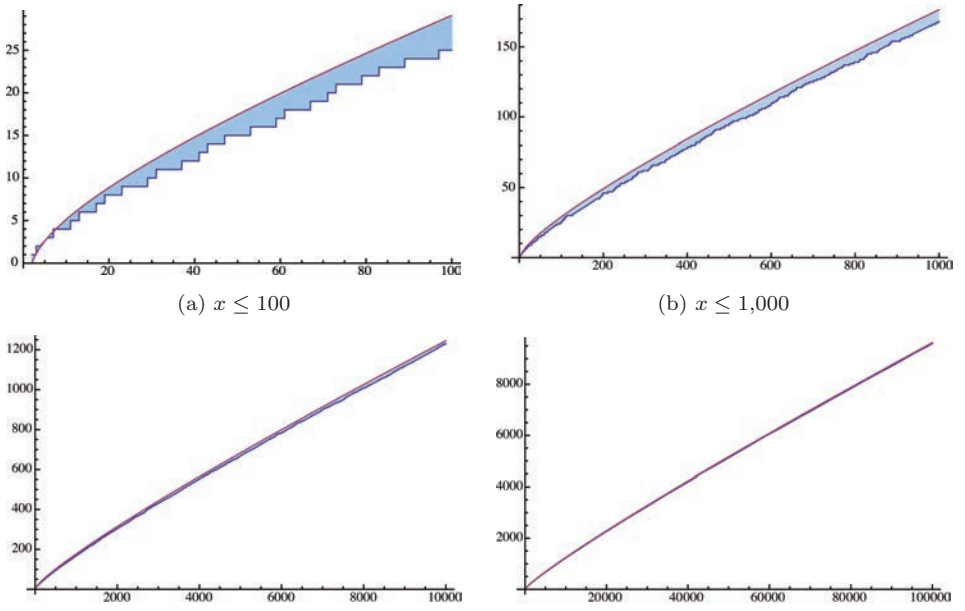
For a few decades, Skewes's number held the record as the largest finite number to meaningfully appear in a mathematical research paper. Let $\pi(x)$ denote the number of primes at most x and let

$$\mathrm{Li}(x) \;=\; \int_2^x \frac{dt}{\log t} \tag{1933.1}$$

denote the offset logarithmic integral function. One version of the prime number theorem (see the 1913 and 1919 entries) says that

$$\lim_{x\to\infty} \frac{\pi(x)}{\mathrm{Li}(x)} \;=\; 1.$$

This is illustrated in Figure 1. The logarithmic integral gives a better approximation to $\pi(x)$ than $x/\log x$, which is used in other formulations of the prime number theorem; see Table 1.



(a) $x \le 100$ (b) $x \le 1{,}000$

(c) $x \le 10{,}000$ (d) $x \le 100{,}000$

FIGURE 1. Graphs of $\mathrm{Li}(x)$ versus $\pi(x)$ on various scales.

TABLE 1. The logarithmic integral $\mathrm{Li}(x)$ is a better approximation to the prime-counting function $\pi(x)$ than is $x/\log x$. The entries in the table have been rounded to the nearest integer.

x	$\pi(x)$	$\mathrm{Li}(x)$	$x/\log x$
1000	168	177	145
10,000	1,229,	1,245	1,086
100,000	9,592	9,629	8,686
1,000,000	78,498	78,627	72,382
10,000,000	664,579	664,917	620,421
100,000,000	5,761,455	5,762,208	5,428,681

For all practically computable values of x, the function $\mathrm{li}(x) = \mathrm{Li(x)} + \log 2$ satisfies $\mathrm{li}(x) > \pi(x)$. Based upon overwhelming numerical evidence, it was conjectured that this held for all x. In 1914, John Edensor Littlewood (1885–1977) showed that $\mathrm{li}(x) - \pi(x)$ changes sign infinitely many times. Littlewood asked one of his students, a South African named Stanley Skewes (1899–1988), to compute how high one must go to find the first integer s_0 for which $\pi(s_0) > \mathrm{li}(s_0)$. Assuming the truth of the Riemann hypothesis,[1] Skewes proved in 1933 that

$$s_0 \ < \ e^{e^{e^{79}}}.$$

In 1955, he showed that if the Riemann hypothesis is false, then

$$s_0 \ < \ e^{e^{e^{e^{7.705}}}}.$$

Both of these extraordinary numbers are sometimes referred to as *Skewes's number*. While much progress has been made, the best upper bounds on s_0 are still on the order of e^{728} (or about 10^{316}). It seems hopeless to expect the first sign change to be found by computer.

Since Skewes's second bound is larger than the first, we can conclude that $\mathrm{li}(x) - \pi(x)$ changes sign somewhere before $\exp(\exp(\exp(\exp(7.705))))$. Why? There are two cases. Either the Riemann hypothesis is true or it is false, and Skewes covered both cases! Voilà! For another striking example of this sort of "magical" reasoning, see the 1935 entry.

Are we overlooking a third possibility? Could the Riemann hypothesis (see the 1942 and 1945 entries) be undecidable, say in ZFC (Zermelo–Fraenkel set theory with the axiom of choice)? If it is false, then it must be provably false in ZFC. Why? Because it is known to be equivalent, under ZFC, to various elementary statements about natural numbers. Let

$$H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$$

denote the nth *harmonic number*. In 2002, Lagarias showed that the statement

$$\text{"for each } n \geq 1, \quad \sum_{d|n} d \ \leq \ H_n + e^{H_n} \log H_n\text{"}$$

[1] The Riemann hypothesis, one of the seven Clay Millennium Problems (see the comments for the 2000 entry), is one of the most important open problems in mathematics. Its veracity would have numerous applications throughout number theory and cryptography. It's going to take a while to build up to! See below and the entries for 1942, 1945, 1948, 1967, and 1987.

is equivalent to the Riemann hypothesis [**3**]. Thus, if the Riemann hypothesis (RH) is false, there is a natural number n for which the preceding inequality is violated and hence there is a finite computation that disproves the Riemann hypothesis. On the other hand, if RH is undecidable in ZFC, then it is true (but just not provable in ZFC; see the 1929 entry on Gödel's work). Why? If the RH were undecidable in ZFC, then no natural number n violating Lagarias's condition exists (the existence of such an n would lead to a quick proof of the falsehood of the Riemann hypothesis). Thus, if the RH is undecidable in ZFC, then Lagarias's condition holds, so the RH is true (just not provable). See the 1924, 1929, and 1963 entries for more information on axiom systems, and the 1987 entry for connections between the Riemann hypothesis and counting primes.

Centennial Problem 1933

Proposed by Steven J. Miller, Williams College.

Let

$$e_{\uparrow n}(x) \;=\; \exp(\exp(\cdots\exp(\exp(x)))),$$

in which there are n iterated exponentials. Thus, Skewes's 1955 result is the bound $s_0 \le e_{\uparrow 4}(7.705)$. If we were to write this as 10^y, what would y equal? More generally, if

$$e_{\uparrow n}(x) \;=\; 10^{f(x;n)},$$

how fast does f grow with n? With x? The functions $e_{\uparrow n}(x)$ are also known as *iterated towers*. For more rapidly growing quantities, see the 1926 and 1992 entries.

1933: Comments

A proof technique. Skewes's arguments use a powerful proof technique: break the problem into an exhaustive set of cases, where in each case you have additional facts at your disposal. For another example of this approach, see the 1935 entry.

Term-by-term multiplication and Mertens's theorem. Here are some facts about infinite series that we will need shortly. Suppose that $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$ are two convergent series of complex numbers. Naively multiplying the two series term-by-term suggests that

$$\begin{aligned}\Big(\sum_{i=0}^{\infty} a_i\Big)\Big(\sum_{j=0}^{\infty} b_j\Big) &= (a_0+a_1+a_2+\cdots)(b_0+b_1+b_2+\cdots)\\ &= a_0b_0+(a_0b_1+a_1b_0)+(a_0b_2+a_1b_1+a_2b_0)+\cdots\\ &= \sum_{n=0}^{\infty} c_n,\end{aligned}$$

in which $c_n = \sum_{k=0}^{n} a_k b_{n-k}$. The series $\sum_{n=0}^{\infty} c_n$ is the *Cauchy product* of $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$. This term-by-term multiplication of series is permissible if both of

the series involved are *absolutely convergent*.[2] This is used implicitly in calculus, complex variables, and differential equations whenever power series methods are involved.

If both series are *conditionally convergent* (convergent but not absolutely convergent), then their Cauchy product series can diverge. An example is furnished by $a_n = b_n = \frac{(-1)^n}{\sqrt{n+1}}$. The alternating series test confirms that $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$ converge. However,

$$\begin{aligned} |c_n| &= \Big|\sum_{k=0}^{n} a_k b_{n-k}\Big| = \sum_{k=0}^{n} \frac{1}{\sqrt{(k+1)(n-k+1)}} \\ &\geq \sum_{k=0}^{n} \frac{1}{\sqrt{\left(\frac{n}{2}+1\right)^2}} = \sum_{k=0}^{n} \frac{1}{\frac{n}{2}+1} = \sum_{k=0}^{n} \frac{2}{n+2} \\ &= (n+1)\frac{2}{n+2} = \frac{2n+2}{n+2} \end{aligned}$$

does not tend to zero, so $\sum_{n=1}^{\infty} c_n$ diverges.

Mertens's theorem, due to Franz Mertens (1840–1927), ensures that if at least one of the two series involved is absolutely convergent, then term-by-term multiplication is permissible. To be more specific, if $\sum_{n=0}^{\infty} a_n = A$ and $\sum_{n=0}^{\infty} b_n = B$ are convergent series of complex numbers, at least one of which is absolutely convergent, then their Cauchy product series $\sum_{n=0}^{\infty} c_n$ converges to AB. Proving Mertens's theorem is a good exercise in analysis. Here is a sketch. Let A_n, B_n, and C_n be the nth partial sums of the three series involved and consider the identity $C_n = A_n B + \sum_{i=0}^{n}(B_i - B)a_{n-i}$. Since $A_n \to A$, the key is to show that $\sum_{i=0}^{n}(B_i - B)a_{n-i} \to 0$ as $n \to \infty$.

The Riemann zeta function and the Euler product formula. In homage to Riemann, who wrote $s = \sigma + it$ to denote his complex variable, we follow him and use the letter s below to refer to a complex number. The Riemann hypothesis concerns the location of the complex zeros of the *Riemann zeta function*

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \tag{1933.2}$$

which is defined initially for $\operatorname{Re} s > 1$. It might at first appear strange to call (1933.2) by such a fancy name. Indeed, (1933.2) is the familiar *p-series* from calculus. However, the Riemann zeta function is the critical function that links analysis and number theory. In particular, the deepest properties of the prime numbers are encoded in the Riemann zeta function.

The connection between the innocuous looking Riemann zeta function and the prime numbers is furnished by the *Euler product formula*. If $\operatorname{Re} s > 1$, then

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right)^{-1}. \tag{1933.3}$$

[2] A series $\sum_{n=0}^{\infty} a_n$ is absolutely convergent if $\sum_{n=0}^{\infty} |a_n|$ converges. Absolute convergence implies convergence, but the converse is not true. The alternating harmonic series $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$ converges to $\log 2$, but the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges.

Since quite a few of our entries (1928, 1942, 1945, 1967, and 1987) involve the Riemann zeta function, we can take the liberty to develop the topic slowly and deliberately.

If p is a fixed prime number and $s > 1$, then the series

$$\sum_{n=0}^{\infty} \frac{1}{(p^n)^s} \;=\; \sum_{n=0}^{\infty} \frac{1}{p^{ns}} \;=\; \sum_{n=0}^{\infty} \left(\frac{1}{p^s}\right)^n \;=\; \left(1 - \frac{1}{p^s}\right)^{-1}$$

converges absolutely since $|1/p^s| < 1$. By Mertens's theorem,

$$\begin{aligned}\left(1 - \frac{1}{2^s}\right)^{-1} \left(1 - \frac{1}{3^s}\right)^{-1} &= \left(1 + \frac{1}{2^s} + \frac{1}{4^s} + \frac{1}{8^s} + \cdots\right)\left(1 + \frac{1}{3^s} + \frac{1}{9^s} + \frac{1}{27^s} + \cdots\right) \\ &= 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{6^s} + \frac{1}{8^s} + \frac{1}{9^s} + \frac{1}{12^s} + \cdots,\end{aligned}$$

in which the last sum includes terms corresponding exactly to those numbers whose prime factorizations involve only 2 or 3. Since $\operatorname{Re} s > 1$, the preceding series is absolutely convergent. Similarly,

$$\begin{aligned}&\left(1 - \frac{1}{2^s}\right)^{-1} \left(1 - \frac{1}{3^s}\right)^{-1} \left(1 - \frac{1}{5^s}\right)^{-1} \\ &\quad= 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \frac{1}{6^s} + \frac{1}{8^s} + \frac{1}{9^s} + \frac{1}{10^s} + \frac{1}{12^s} + \frac{1}{15^s} + \cdots,\end{aligned}$$

in which the sum involves those numbers whose only prime factors are 2, 3, or 5, and so forth. Since the tail end of a convergent series tends to zero,

$$\Big|\sum_{n=1}^{\infty} \frac{1}{n^s} - \prod_{\substack{p \text{ prime} \\ p \le N}} \left(1 - \frac{1}{p^s}\right)^{-1}\Big| \;\le\; \sum_{n=N}^{\infty} \frac{1}{n^s} \;\to\; 0$$

as $N \to \infty$. This establishes the Euler product formula (1933.3).

We get Euclid's theorem on the infinitude of the primes as a corollary. If there were only finitely many primes, then the right-hand side of (1933.3) would converge to a finite limit as $s \to 1^+$. However, the left-hand side of (1933.3) diverges as $s \to 1^+$ since its terms tend to those of the harmonic series.

Bibliography

[1] T. M. Apostol, *Introduction to analytic number theory*, Undergraduate Texts in Mathematics, Springer-Verlag, New York-Heidelberg, 1976. MR0434929

[2] H. Davenport, *Multiplicative number theory*, 3rd ed., revised and with a preface by Hugh L. Montgomery, Graduate Texts in Mathematics, vol. 74, Springer-Verlag, New York, 2000. MR1790423

[3] J. C. Lagarias, *An elementary problem equivalent to the Riemann hypothesis*, Amer. Math. Monthly **109** (2002), no. 6, 534–543, DOI 10.2307/2695443. MR1908008

[4] S. J. Miller and R. Takloo-Bighash, *An invitation to modern number theory*, with a foreword by Peter Sarnak, Princeton University Press, Princeton, NJ, 2006. MR2208019

[5] S. Skewes, *On the Difference pi(x) − li(x) (I)*, J. London Math. Soc. **8** (1933), no. 4, 277–283, DOI 10.1112/jlms/s1-8.4.277. MR1573970

[6] S. Skewes, *On the difference* $\pi(x) - \operatorname{li} x$. *II*, Proc. London Math. Soc. (3) **5** (1955), 48–70, DOI 10.1112/plms/s3-5.1.48. MR0067145

1942

Zeros of $\zeta(s)$

Introduction

The Riemann zeta function is perhaps the most important function in number theory; see the 1928, 1933, 1939, 1942, 1945, 1967, and 1987 entries. It is initially defined for $\operatorname{Re} s > 1$ by the series

$$\zeta(s) \;=\; \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

The Euler product formula (1933.3) is the product representation

$$\zeta(s) \;=\; \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right)^{-1},$$

also valid for $\operatorname{Re} s > 1$; see the 1933 entry for a proof. Although Euler and others studied the zeta function first, it is named after Georg Friedrich Bernhard Riemann (1826–1866) because of his 1859 masterpiece that relates the distribution of the zeros of $\zeta(s)$ to the fine properties of the prime-counting function $\pi(x)$ [**8**].

The Euler product formula confirms that the zeta function has no zeros in the half plane $\operatorname{Re} s > 1$. However, neither the series nor the product representation given above converges if $\operatorname{Re} s \le 1$. So what do we mean by the zeros of $\zeta(s)$? To resolve this issue and to understand Riemann's contribution, we must discuss analytic continuation.

An *analytic function* is a differentiable function $f : U \to \mathbb{C}$ defined on a nonempty, connected open set $U \subseteq \mathbb{C}$. By "differentiable," we mean that

$$f'(z_0) \;=\; \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

exists for every $z_0 \in U$. This is the complex version of the single-variable calculus definition. For instance, the zeta function is analytic on $\operatorname{Re} s > 1$ with derivative

$$\zeta'(s) = -\sum_{n=1}^{\infty} \frac{\log n}{n^s}.$$

An *analytic continuation* of an analytic function $f : U \to \mathbb{C}$ is an analytic function $g : V \to \mathbb{C}$, defined on an open set V that contains U, so that f and g agree on U. That is, g is an analytic "extension" of f to the larger set V.

An example of analytic continuation involves the geometric series. The summation formula

$$\sum_{n=0}^{\infty} z^n \;=\; \frac{1}{1 - z} \tag{1942.1}$$

is valid for $|z| < 1$.[1] There is an important asymmetry in (1942.1): the series converges only for $|z| < 1$, whereas the function $(1-z)^{-1}$ is defined for all $z \neq 1$. Thus, $(1-z)^{-1}$ provides an analytic continuation of $\sum_{n=0}^{\infty} z^n$ from the open disk $|z| < 1$ to the much larger region $\mathbb{C}\backslash\{1\}$.

Obtaining an analytic continuation of the zeta function is more difficult. We first construct an analytic continuation to $\operatorname{Re} s > 0$. Observe that

$$\begin{aligned}\zeta(s) - \frac{1}{s-1} &= \sum_{n=1}^{\infty} n^{-s} - \int_1^{\infty} x^{-s}\,dx = \sum_{n=1}^{\infty}\left(n^{-s} - \int_n^{n+1} x^{-s}\,dx\right)\\ &= \sum_{n=1}^{\infty}\int_n^{n+1}\left(n^{-s} - x^{-s}\right)dx\\ &= \sum_{n=1}^{\infty}\int_n^{n+1}\left(s\int_n^x y^{-1-s}\,dy\right)dx. \qquad (1942.2)\end{aligned}$$

Since

$$\left|\int_n^{n+1}\left(s\int_n^x y^{-1-s}\,dy\right)dx\right| \leq |s| n^{-1-\operatorname{Re} s},$$

it follows that the series (1942.2) converges absolutely and uniformly on each half-plane $\operatorname{Re} s \geq \delta > 0$. Each summand is an analytic function of s, so (1942.2) provides an analytic continuation of $\zeta(s) - (s-1)^{-1}$ to the half-plane $\operatorname{Re} z > 0$; the presence of the term $(s-1)^{-1}$ on the left-hand side ensures that $\zeta(s)$ has a simple pole at $s = 1$ with residue 1. That is, near the point $s = 1$, the zeta function behaves like the function $(s-1)^{-1}$.

The next, and most complicated, step is to show that the zeta function satisfies the *functional equation*

$$\zeta(s) = 2^s \pi^{s-1} \sin\left(\frac{\pi s}{2}\right)\Gamma(1-s)\zeta(1-s), \qquad (1942.3)$$

in which

$$\Gamma(s) \;=\; \frac{e^{-\gamma s}}{s}\prod_{n=1}^{\infty}\left(1+\frac{s}{n}\right)^{-1} e^{\frac{s}{n}} \qquad (1942.4)$$

is the gamma function and

$$\gamma \;=\; \lim_{N\to\infty}\left(\sum_{n=1}^{N}\frac{1}{n} - \log N\right) \approx 0.5772156\ldots \qquad (1942.5)$$

is the Euler–Mascheroni constant. For the sake of brevity, we omit this step. Since the product (1942.4) is an analytic function on $\mathbb{C}\backslash\{0,-1,-2,-3,\ldots\}$, the functional equation (1942.3) permits us to define $\zeta(s)$ for $\operatorname{Re} s \leq 0$ since the function on the right-hand side of (1942.3) is now defined for $s \neq 1$ with $\operatorname{Re} s \geq 1$. Thus, we have obtained an analytic continuation of the zeta function to $\mathbb{C}\backslash\{0\}$.

The product representation (1942.4) of the gamma function and (1942.3) ensure that ζ has zeros at $-2, -4, -6, \ldots$ These are the *trivial zeros of the zeta function*. Any remaining zeros must be in the *critical strip*

$$\{s \in \mathbb{C} : 0 < \operatorname{Re} s < 1\}.$$

[1] The radius of convergence of the series $\sum_{n=0}^{\infty} z^n$ is 1. What students of calculus do not often realize is that the "radius" referred to is the radius of the disk $|z| < 1$ in the complex plane.

These are the *nontrivial zeros of the zeta function*. It turns out that the nontrivial zeros govern the main terms in our error estimates of the $\pi(x)$. Neglecting some logarithmic factors, if

$$\theta = \sup\{\operatorname{Re} s : 0 < \operatorname{Re} s < 1,\ \zeta(s) = 0\},$$

then the maximum deviation[2] $|\pi(x) - \operatorname{Li}(x)|$ from the prediction of the prime number theorem is essentially of size at most x^θ. Thus, the nontrivial zeros of the zeta function have an enormous influence in number theory: they control the large-scale distribution of the prime numbers.

To a few decimal places, these are the first twenty nontrivial zeros that lie in the upper half-plane:

$$0.5 + 14.1347i,\ 0.5 + 21.0220i,\ 0.5 + 25.0109i,\ 0.5 + 30.4249i,\ 0.5 + 32.9351i,$$
$$0.5 + 37.5862i,\ 0.5 + 40.9187i,\ 0.5 + 43.3271i,\ 0.5 + 48.0052i,\ 0.5 + 49.7738i,$$
$$0.5 + 52.9703i,\ 0.5 + 56.4462i,\ 0.5 + 59.3470i,\ 0.5 + 60.8318i,\ 0.5 + 65.1125i,$$
$$0.5 + 67.0798i,\ 0.5 + 69.5464i,\ 0.5 + 72.0672i,\ 0.5 + 75.7047i,\ 0.5 + 77.1448i.$$

Notice a pattern? Numerical calculations have confirmed that the first 10^{13} nontrivial zeros lie on the *critical line* $\operatorname{Re} s = \frac{1}{2}$; see Figure 1. The *Riemann hypothesis*, one of the seven Clay Millennium Problems, asserts that the nontrivial zeros all lie on the critical line. Riemann wrote in [**8**]:

> . . . and it is very probable that all roots are real.[3] Certainly one would wish for a stricter proof here; I have meanwhile temporarily put aside the search for this after some fleeting futile attempts, as it appears unnecessary for the next objective of my investigation.

The Riemann hypothesis, which was one of Hilbert's problems [**10**] (see the 1935, 1963, 1970, 1980, and 1983 entries), is considered by many mathematicians to be the most important open problem in mathematics.

In 1914, Godfrey Harold Hardy (see the 1920, 1923, and 1940 entries) proved there are infinitely many nontrivial zeros on the critical line. However, he was unable to ascertain whether a positive proportion of them are on the critical line. The situation changed in 1942, when Atle Selberg (1917–2007) showed that a small, but positive, proportion of the zeros of $\zeta(s)$ are on the critical line; see the 1948 entry. A major advance came in 1974 with the work of Norman Levinson (1912–1975), who proved more than a third of these zeros are on the line. The best results today are around 40%; there is still a long way to go. Even if we can prove that 100% of the zeros are on the critical line, that still would be insufficient to prove the Riemann hypothesis. There could still be infinitely many zeros in the critical strip that do not lie on the critical line. This is meant in the same sense that "100% of natural numbers are not perfect squares." The proportion of natural numbers at most x that are not perfect squares is approximately $(x - \sqrt{x})/x = 1 - 1/\sqrt{x}$, which tends to zero as $x \to \infty$.

It is still unknown whether or not there is a $c < 1$ such that all nontrivial zeros of the zeta function have real part at most c; the Riemann hypothesis is equivalent to being able to take $c = \frac{1}{2}$ (the nontrivial zeros are symmetric about the line

[2]Here $\operatorname{Li}(x)$ denotes the offset logarithmic integral function (1933.1).

[3]Riemann was considering a variant of the zeta function, for which the corresponding conjecture is that the zeros are real.

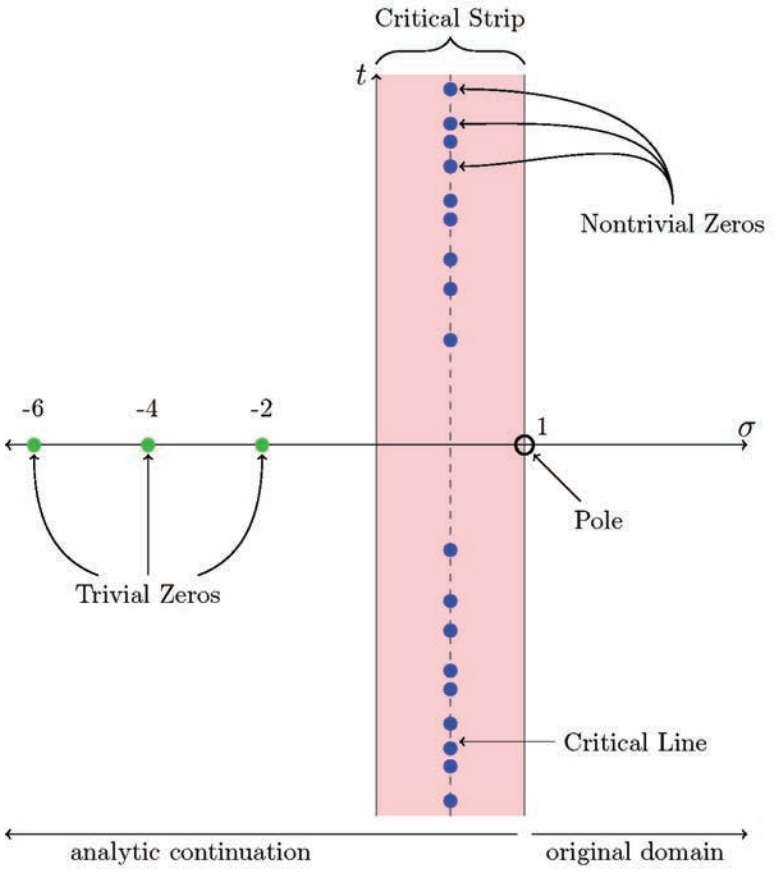


FIGURE 1. The nontrivial zeros of the Riemann zeta function lie in the critical strip $0 < \operatorname{Re} s < 1$. The Riemann hypothesis asserts that they all lie on the critical line $\operatorname{Re} s = \frac{1}{2}$.

$\operatorname{Re} s = \frac{1}{2}$). The best results are zero-free regions where how far to the left of the line $\operatorname{Re} s = 1$ we can go tends to zero rapidly with the height t, giving regions where

$$\zeta(\sigma + it) \neq 0 \quad \text{if } \sigma > 1 - A(\log|t|)^{-r_1}(\log\log|t|)^{-r_2}$$

for some positive constants A, r_1, r_2.

Centennial Problem 1942

Proposed by Steven J. Miller, Williams College.

What is wrong with the following "proof" of the Riemann hypothesis?[4]

(a) For each prime p let $h_p(s) = (1 - p^{-2s})^{-1}/(1 - p^{-s})^{-1}$. Note that $h_p(s)$ is never zero or infinity for $\operatorname{Re} s > 0$.

[4]This is not a valid proof, nor can it be salvaged.

(b) Let $\zeta_2(s) = h_2(s)\zeta(s)$. The analytic continuation of $\zeta_2(s)$ is simply $h_2(s)$ times the analytic continuation of $\zeta(s)$. Furthermore, $\zeta_2(s)$ and $\zeta(s)$ have the same zeros for $\operatorname{Re} s > 0$. Observe that

$$\zeta_2(s) \;=\; \left(1 - 2^{-2s}\right)^{-1} \prod_{\substack{p \text{ prime} \\ p \geq 3}} \left(1 - p^{-s}\right)^{-1}.$$

(c) Similarly set $\zeta_3(s) = h_3(s)\zeta_2(s)$, and observe that $\zeta_3(s)$ and $\zeta_2(s)$ (and hence also $\zeta(s)$) have the same zeros in the region $\operatorname{Re} s > 0$. Note that

$$\zeta_3(s) \;=\; \left(1 - 2^{-2s}\right)^{-1} \left(1 - 3^{-2s}\right)^{-1} \prod_{\substack{p \text{ prime} \\ p \geq 5}} \left(1 - p^{-s}\right)^{-1}.$$

(d) We continue this process, working initially in the region $\operatorname{Re} s > 2$ so that all the products involved converge uniformly. We let $\zeta_\infty(s)$ be the limit of $\zeta_p(s)$ as $p \to \infty$. This limit exists and equals $\zeta(2s)$ for $\operatorname{Re} s > 2$.

(e) Since $\zeta(2s)$ has an analytic continuation that does not vanish for $\operatorname{Re} s > 1/2$ (because $\zeta(s)$ does not vanish if $\operatorname{Re} s > 1$), each $\zeta_p(s)$ does not vanish for $\operatorname{Re} s > 1/2$. Since all these functions have the same zeros in this region, none of them vanish for $\operatorname{Re} s > 1/2$. Thus, $\zeta(s)$ does not vanish in this region and the Riemann hypothesis is true.

1942: Comments

Solution to the problem. The approach sketched above is fundamentally flawed. The error is that the analytic continuation of the limit is not necessarily the limit of the analytic continuation. Moreover, there is no hope of salvaging the argument above. If instead of replacing each prime with its square we used its cube, we would then deduce that $\zeta(s)$ has no zeros for $\operatorname{Re} s > 1/3$. However, this is impossible since the zeta function has infinitely many zeros on the critical line.

Bibliography

[1] E. Bombieri, *Problems of the millennium: the Riemann hypothesis*, Clay Mathematics Institute, `http://www.claymath.org/sites/default/files/official_problem_description.pdf`.

[2] Clay Mathematics Institute, *Millennium problems*, `http://www.claymath.org/millennium-problems`.

[3] H. Davenport, *Multiplicative number theory*, 2nd ed., revised by Hugh L. Montgomery, Graduate Texts in Mathematics, vol. 74, Springer-Verlag, New York-Berlin, 1980. MR606931

[4] H. M. Edwards, *Riemann's zeta function*, Pure and Applied Mathematics, Vol. 58, Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1974. MR0466039

[5] G. H. Hardy, *Sur les zéros de la fonction* $\zeta(s)$, Comp. Rend. Acad. Sci. **158** (1914), 1012–1014.

[6] H. Iwaniec and E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, vol. 53, American Mathematical Society, Providence, RI, 2004. MR2061214

[7] N. Levinson, *More than one third of zeros of Riemann's zeta-function are on* $\sigma = 1/2$, Advances in Math. **13** (1974), 383–436, DOI 10.1016/0001-8708(74)90074-7. MR0564081

[8] G. F. B. Riemann, *Über die Anzahl der Primzahlen unter einer gegebenen Grösse*, Monatsber. Königl. Preuss. Akad. Wiss. Berlin, Nov. 1859, 671–680. `http://www.maths.tcd.ie/pub/HistMath/People/Riemann/Zeta/EZeta.pdf`.

[9] A. Selberg, *Contributions to the theory of the Riemann zeta-function*, Arch. Math. Naturvid. **48** (1946), no. 5, 89–155. MR0020594

[10] Wikipedia, *Hilbert's problems*, `http://en.wikipedia.org/wiki/Hilbert's_problems`.

1960

The Unreasonable Effectiveness of Mathematics

Introduction

This year honors a groundbreaking, influential article by Eugene Wigner [**12**], the Nobel laureate in physics whose work in random matrix theory eventually led to astonishing connections between the seemingly diverse fields of number theory and nuclear physics; see the 1928 entry. In his article, Wigner discusses the use of mathematics in physics:[1]

> A possible explanation of the physicist's use of mathematics to formulate his laws of nature is that he is a somewhat irresponsible person. As a result, when he finds a connection between two quantities which resembles a connection well-known from mathematics, he will jump at the conclusion that the connection is that discussed in mathematics simply because he does not know of any other similar connection. It is not the intention of the present discussion to refute the charge that the physicist is a somewhat irresponsible person. Perhaps he is. However, it is important to point out that the mathematical formulation of the physicist's often crude experience leads in an uncanny number of cases to an amazingly accurate description of a large class of phenomena. This shows that the mathematical language has more to commend it than being the only language which we can speak; it shows that it is, in a very real sense, the correct language.

Mathematics is so ubiquitous in physics that the American Journal of Physics asked, "Does any piece of mathematics exist for which there is *no application whatsoever in physics?*" To this, physicist Dwight E. Neuenschwander (1952–) responded:

> While constructing such a "useless" piece of mathematics would be the delight of a mathematical purist, it seems we physicists have always managed to foil this lofty goal. It seems that even the most esoteric mathematical inventions of the human mind are eventually used to model physical systems. Why *that* should be true is of course a deep and fascinating question. [**9**]

The catchphrase "unreasonable effectiveness" has spawned innumerable imitators and it is difficult to catalogue them all. Some of the most influential were discussed by economist K. Vela Velupillai (1947–) [**11**]:

> Eugene Wigner's Richard Courant Lecture in the Mathematical Sciences, delivered at New York University on 11 May 1959, was titled,

[1] The repeated use of "his" and "he" to refer to a generic physicist is regrettable.

> picturesquely and, perhaps, with intentional impishness *The Unreasonable Effectiveness of Mathematics in the Natural Sciences* [**12**]. Twenty years later, another distinguished scientist, Richard W. Hamming, gave an invited lecture to the Northern California Section of the Mathematical Association of America with the slightly truncated title *The Unreasonable Effectiveness of Mathematics* [**5**]. A decade or so later, Stefan Burr tried a different variant of Wigner's title by organising a short course on *The Unreasonable Effectiveness of Number Theory* [**2**]. Another decade elapsed before Arthur Lesk, a distinguished molecular biologist at Cambridge, gave a lecture at the Isaac Newton Institute for Mathematical Sciences at Cambridge University where yet another twist to the Wigner theme was added: *The Unreasonable Effectiveness of Mathematics in Molecular Biology* [**8**].[2]

The words "unreasonable" and "effectiveness" are often slightly modified to fit the author's point. For example, there is *The Reasonable Ineffectiveness of Research in Mathematics Education* [**7**]. In *The Reasonable Effectiveness of Mathematics in Economics* [**3**], Frank J. Fabozzi (1948–) and Sergio M. Focardi tell us:

> In a nutshell, we believe that the reason that mathematics is only reasonably effective in economics is because we apply mathematics to study large engineered artefacts (i.e., economies or financial markets), that have been designed to allow a lot of freedom so as to encourage change and innovation. The level of unpredictability and control is clearly different when considering systems governed by immutable natural laws as opposed to artefacts constructed by humans. Some systems, such as economies or financial markets, are prone to crises. Mathematics does a reasonably good job in describing these systems. But the mathematics involved is not that of physics: It is the mathematics of learning and complexity.

Mathematics is often called the language of the universe. However, some dispute how far this universe extends beyond physics and astronomy and how much is actually needed to describe the world and make significant contributions; see the article [**13**] by biologist Edward Osborne Wilson (1929–). Wigner's article influenced even those who profoundly disagree with him. For example, Israel Gelfand, who worked both in pure mathematics (see the 1941 entry) and mathematical biology, said:

> Eugene Wigner wrote a famous essay on the unreasonable effectiveness of mathematics in natural sciences. He meant physics, of course. There is only one thing which is more unreasonable than the unreasonable effectiveness of mathematics in physics, and this is the unreasonable ineffectiveness of mathematics in biology.

The engineer Derek Abbott (1960–) wrote the influential rebuttal *The Reasonable Ineffectiveness of Mathematics* [**1**], in which he writes:

> Science is a modern form of alchemy that produces wealth by producing the understanding for enabling valuable products from base ingredients. Science is merely functional alchemy that has had a few incorrect assumptions fixed, but has in its arrogance replaced them

[2]The punctuation and citation style has been slightly modified.

> with more insidious ones. The real world of nature has the uncanny habit of surprising us; it has always proven to be a lot stranger than we give it credit for. Mathematics is a product of the imagination that sometimes works on simplified models of reality. Platonism is a viral form of philosophical reductionism that breaks apart holistic concepts into imaginary dualisms.... Mathematics is a human invention for describing patterns and regularities. It follows that mathematics is then a useful tool in describing regularities we see in the universe. The reality of the regularities and invariances, which we exploit, may be a little rubbery, but as long as they are sufficiently rigid on the scales of interest to humans, then it bestows a sense of order.

Certainly many mathematicians would disagree with Abbott's account!

Centennial Problem 1960

Proposed by Stanislav Molchanov and Harold Reiter, UNC Charlotte.

The following four problems illustrate Wigner's principle that a single mathematical idea often appears in several different areas.

Problem 1. The Catalan numbers are defined for integers $n \geq 0$ by

$$C_n = \frac{1}{n+1}\binom{2n}{n}. \tag{1960.1}$$

The first several Catalan numbers are

$$1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786, 208012, \ldots.$$

Prove that C_n is always an integer.

Problem 2. The probability density

$$p(x) \;=\; \begin{cases} \dfrac{1}{2\pi}\sqrt{4-x^2} & \text{if } |x| \leq 2, \\ 0 & \text{otherwise,} \end{cases}$$

arises in Wigner's semicircle law, which he proposed for the description of the spectra of heavy atomic nuclei. Show that its moments are

$$\frac{1}{2\pi}\int_{-2}^{2} x^n \sqrt{4-x^2}\,dx \;=\; \begin{cases} C_{n/2} & \text{if } n \text{ is even,} \\ 0 & \text{if } n \text{ is odd.} \end{cases}$$

Problem 3. A *tree* is a graph in which any two vertices are connected by exactly one path. An *ordered tree* is a rooted tree in which the children of each vertex are given a fixed left-to-right order. Show that C_n is the number of nonisomorphic orderded trees with n vertices; see Figure 1.

Problem 4. Suppose that we must multiply $n \geq 2$ symbols $a_1, a_2, \ldots, a_n$ using a binary but not necessarily associative operation $b(x, y)$. Consequently, we must keep track of order. We are interested in the number of structurally different ways we can combine the symbols, and not the number of different ways we can then input the n objects into the possibilities. If we let S_{n-1} denote the number of different structures we can use to multiply n symbols using our binary operation $n-1$ times, then $S_1 = 1$ since the only way to combine two symbols is $b(a_1, a_2)$; we do not count $b(a_2, a_1)$ since it is structurally the same as $b(a_1, a_2)$.

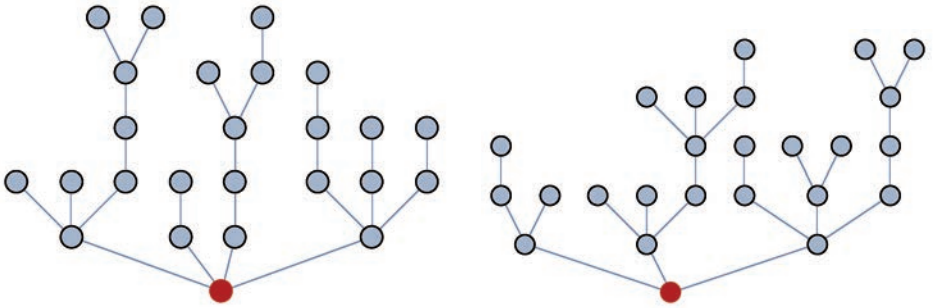


FIGURE 1. Two rooted trees on 23 vertices. The root vertices are highlighted in red. If we had to choose names for the trees, they would be Telperion the Silver and Laurelin the Golden.

Similarly, $S_2 = 2$ since we have only two structurally different approaches:

$$b\,(a_1, b(a_2, a_3)) \qquad \text{and} \qquad b\,(b(a_1, a_2), a_3)\,.$$

A little more work shows that $S_3 = 5$:

$$b\big(b(a_1, a_2), b(a_3, a_4)\big), \qquad \big(b\big(b(a_1, a_2), a_3\big), a_4\big), \qquad b\big(a_1, b\big(b(a_2, a_3), a_4\big)\big),$$

$$b\big(b\big(a_1, b(a_2, a_3)\big), a_4\big), \quad \text{and} \quad b\big(a_1, b\big(a_2, b(a_3, a_4)\big)\big).$$

Show that $S_n = C_n$.

1960: Comments

Catalan numbers. There is a wealth of interesting facts known about the Catalan numbers. First of all, they are named after the French-Belgian mathematician Eugène Charles Catalan (1814–1894), who does not appear to be Catalonian. Nevertheless, the term "Catalonian" has been used by a few authors to refer to subjects related to the Catalan numbers [**4**, p. 254] (at least the authors think it a good idea and are not above flagrant self-reference). The Catalan numbers appear in many different places in mathematics; over fifty such occurrences are discussed in [**10**].

It turns out that C_n is the number of ways to write n left parentheses and n right parentheses so that, as we move from left to right, we never see more right parentheses than left parentheses. We see that $C_1 = 1$ since the only possible arrangement is (). Similarly, $C_2 = 2$ since there are only two permissible configurations: ()() and (()). For $n = 3$, we have exactly five options:

$$((())), \quad (()()), \quad (())(), \quad ()(()), \quad \text{and} \quad ()(()).$$

Thus, $C_3 = 5$. See the comments for the 2008 entry for the asymptotic rate of growth of the Catalan numbers.

Another interesting interpretation of C_n is that it is the number of "staircase walks" from $(0, 0)$ to (n, n) that never rise above the main diagonal; that is, $j \le k$ whenever (j, k) is on our path. Such a path is called a *Dyck path*, in honor of Walther von Dyck (1856–1934); see Figure 2.

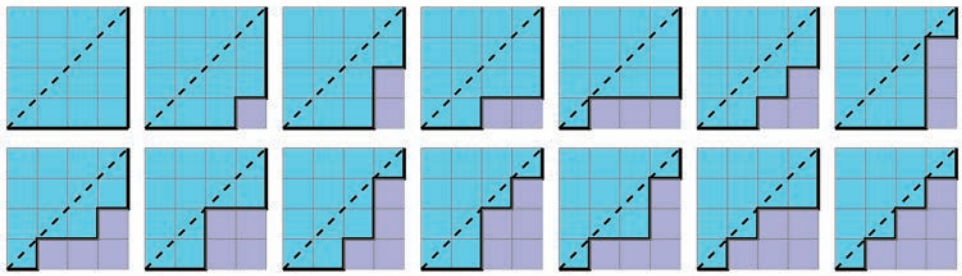


FIGURE 2. There are $C_4 = 14$ Dyck paths of order 4.

Bibliography

[1] D. Abbott, *The reasonable ineffectiveness of mathematics*, Proceedings of the IEEE, Vol. 101, no. 10, October 2013.

[2] S. A. Burr (ed.), *The unreasonable effectiveness of number theory*, papers from the American Mathematical Society Short Course held in Orono, Maine, August 6–7, 1991, Proceedings of Symposia in Applied Mathematics, vol. 46, American Mathematical Society, Providence, RI, 1992. MR1195838

[3] S. M. Focardi and F. J. Fabozzi, *The reasonable effectiveness of mathematics in economics*, American Economist **1** (2010), no. 55, 19–30.

[4] S. R. Garcia and S. J. Miller, *100 Years of Math Milestones: The Pi Mu Epsilon Centennial Collection*, American Mathematical Society, 2019.

[5] R. W. Hamming, *The unreasonable effectiveness of mathematics*, Amer. Math. Monthly **87** (1980), no. 2, 81–90, DOI 10.2307/2321982. MR559142

[6] A. Harvey, *The Reasonable Effectiveness of Mathematics in the Physical Sciences*, Relativity and Gravitation, **43** (2011), 3057–3064.

[7] J. Kilpatrick, *The reasonable ineffectiveness of research in mathematics education*, For the Learning of Mathematics **2** (1981), no. 2, 22–29.

[8] A. M. Lesk, *The unreasonable effectiveness of mathematics in molecular biology*, Math. Intelligencer **22** (2000), no. 2, 28–37, DOI 10.1007/BF03025372. MR1764266

[9] D. E. Neuenschwander, *Does any piece of mathematics exist for which is no application whatsoever in physics?*, Amer. J. Phys. **63** (1996), 63.

[10] R. P. Stanley, *Enumerative combinatorics. Vol. 2*, with a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin, Cambridge Studies in Advanced Mathematics, vol. 62, Cambridge University Press, Cambridge, 1999. MR1676282

[11] K. V. Velupillai, *The unreasonable* in*effectiveness of mathematics in economics*, Cambridge Journal of Economics **29** (2005), 849–872.

[12] E. P. Wigner, *The unreasonable effectiveness of mathematics in the natural sciences [Comm. Pure Appl. Math.* **13** *(1960), 1–14; Zbl 102, 7]*, Mathematical analysis of physical systems, Van Nostrand Reinhold, New York, 1985, pp. 1–14. `https://www.dartmouth.edu/~matc/MathDrama/reading/Wigner.html`. MR824292

[13] E. O. Wilson, *Great Scientist ≠ Good at Math: E. O. Wilson shares a secret: Discoveries emerge from ideas, not number-crunching*, Wall Street Journal (online). `http://www.wsj.com/articles/SB10001424127887323611604578398943650327184`.

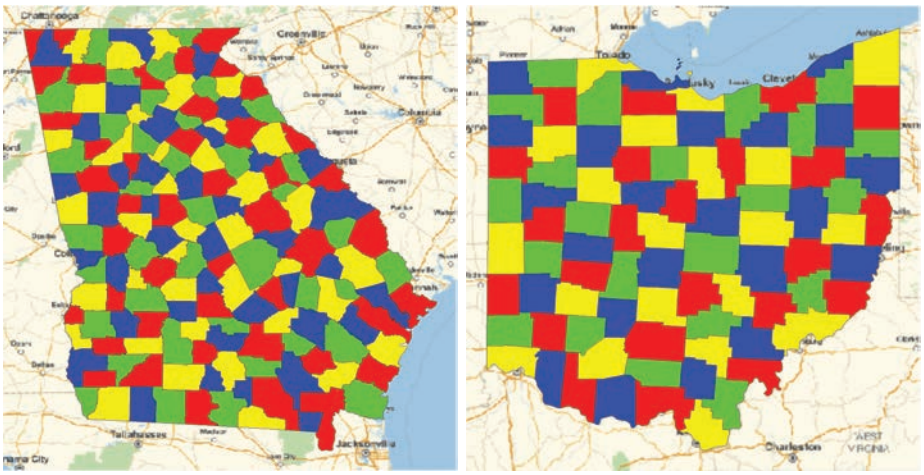
1976

Four Color Theorem

Introduction

The *four color theorem* states that every planar map can be colored with four colors in such a way that no two adjacent countries share the same color; see Figure 1. However, we should be precise about what this means. First of all, each country must be connected. For example, the United States does not count because Alaska and Hawaii are not connected to the lower forty-eight states. Second, we do not consider countries that touch "at corners" to be adjacent. Thus, Arizona and Colorado do not share a border as far as we are concerned; neither do Utah and New Mexico. Finally, we prohibit countries with infinitely long boundaries since otherwise one can construct bizarre maps that require more than four colors [**8**].

The year 1976 marked the end of the long search for a (correct) proof of the four color theorem, which was initially conjectured in 1852 by Francis Guthrie (1831–1899). The conjecture was prompted by his attempt to color a map of English counties. Today most people know the theorem in the form "no more than four colors are needed to color a map." Despite this common understanding of the theorem, cartographers claim that it does not matter since there is no reason to limit the number of colors used. Moreover, only three colors are needed for most



(a) Georgia (b) Ohio

FIGURE 1. Four colorings of the counties in two US states.

maps that arise in practice. Despite its pragmatic insignificance, the four color theorem has great historical importance.

To make the problem more precise, one converts statements about maps into statements about graphs. Assign each country a vertex. Place an edge between two vertices if and only if the two corresponding countries share a common border. This permits us to phrase the four color theorem in terms of graph theory: the vertices of any graph that can be drawn in the plane without edge crossings can be colored with at most four colors so that no two adjacent vertices share the same color.

The four color theorem has the dubious honor of having been "proved" twice before 1976. Proofs by Alfred Kempe (1849–1922) in 1879 and by Peter Guthrie Tait (1831–1901) in 1880 each stood unchallenged for 11 years before fatal flaws were found. It is much easier to prove that five colors suffice [**7**]; see [**9**, Chapter 19] for details.

It was not until 1976 that mathematicians again claimed to have a proof of the elusive theorem. Kenneth Appel (1932–2013) and Wolfgang Haken (1928–) at the University of Illinois proved the four color theorem with computer assistance, through which they reduced the problem to 1,936 special cases, each of which was checked by computer. This was greeted with controversy by the mathematical community (see also the 1998 entry on the Kepler conjecture). Is a proof valid if it is so long and computationally intensive that no human can understand it in totality? Although the theorem has since been verified by the Coq interactive theorem prover [**6**], there are some who still find the prospect of computer-aided proofs unsettling. Perhaps a more elegant, humanly understandable proof of the four color theorem exists. Try to find it!

Centennial Problem 1976

Proposed by Alexandra Jensen, Steven J. Miller, and Pamela Mishkin, Williams College.

We know that four colors suffice to color a planar map so that no two countries with a common border share the same color. What if we add the constraint that no color is used too often? For what $p \in [25, 100]$ does a four coloring exist that uses each color for at most $p\%$ of the countries? The four color theorem says we may take $p = 100$ and the pigeonhole principle tells us we cannot have $p < 25$. What if we only require at most $p\%$ of each color when there are at most N regions?

1976: Comments

Heawood conjecture. The four color theorem tells us that we can color any planar map using at most four colors. What about map colorings on the torus, the Klein bottle (see the 1958 entry), or other surfaces? Percy J. Heawood (1861–1955), who spent most of his career attempting to prove the four color theorem and found the fatal flaw in Kempe's 1879 proof, conjectured in 1890 that the minimum

TABLE 1. Computation of the Euler characteristics of the five Platonic solids. Here v denotes the number of vertices, e the number of edges, and f the number of faces of the solid. Since all five solids are homeomorphic, their Euler characteristics are equal.

S	v	e	f	χ
tetrahedron	4	6	4	2
cube	8	12	6	2
octahedron	6	12	8	2
dodecahedron	20	30	12	2
isosahedron	12	30	20	2

number of colors required to color any map on a two-dimensional surface S is

$$\left\lfloor \frac{7+\sqrt{49-24\chi}}{2} \right\rfloor, \tag{1976.1}$$

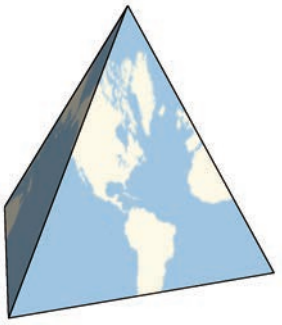
in which χ denotes the *Euler characteristic* of S [**7**]. To compute χ, triangulate S and use the formula

$$\chi(S) \;=\; v - e + f,$$

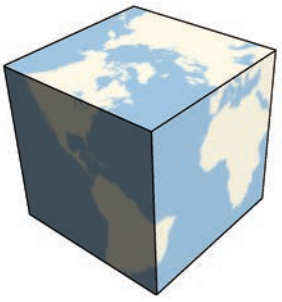
in which v denotes the number of vertices, e the number of edges, and f the number of faces in the triangulation. It turns out that any triangulation of S produces the same value; that is, the Euler characteristic is a topological invariant of S.[1] For example, the five Platonic solids are all homeomorphic (see p. 22) to a sphere and all have $\chi = 2$; see Figure 2 and Table 1. Substituting this into (1976.1) suggests that any map on a sphere can be colored with at most four colors.

What is the status of the Heawood conjecture? Technically, it was disproved in 1934 when Philip Franklin (1898–1965) proved that any map on the Klein bottle (for which $\chi = 0$) can be colored with only six colors, as opposed to the seven predicted by the conjecture [**5**]. This bound is tight since the Franklin graph (Figure 3) can be embedded on the surface of the Klein bottle and the resulting map cannot be colored with fewer than six colors. Morally speaking, however, the conjecture is true 100% of the time since Gerhard Ringel (1919–2008) and John W. T. Youngs (1910–1970) proved that it holds for all surfaces other than the Klein bottle [**10**]. For example, any map on the torus (which has $\chi = 0$) can be colored with only seven colors, and this is minimal; see Figure 4.

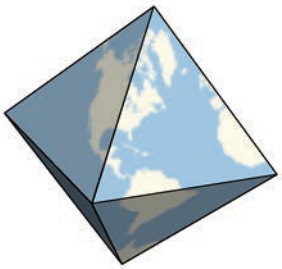
[1]It is important to note that nonhomeomorphic surfaces may have the same Euler characteristic. For example, the torus and the Klein bottle both have Euler characteristic zero. They are not homeomorphic since, for example, they have different fundamental groups ($\mathbb{Z}^2$ for the torus and $\langle a, b : ab = b^{-1}a\rangle$ for the Klein bottle). We refrain from further discussion since that would take us too far afield.



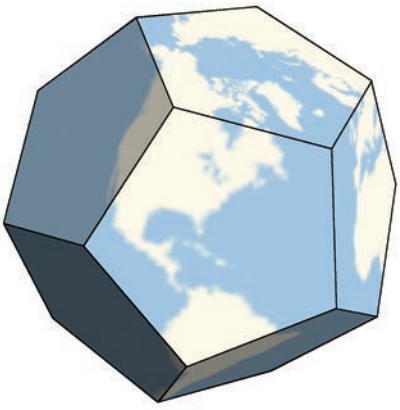
(a) Tetrahedron $(4, 6, 4)$.



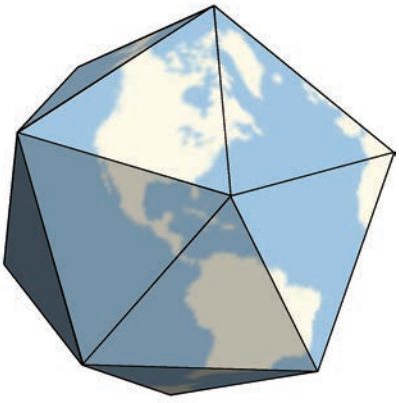
(b) Cube $(8, 12, 6)$



(c) Octahedron $(6, 12, 8)$



(d) Dodecahedron $(20, 30, 12)$



(e) Icosahedron $(12, 30, 20)$

FIGURE 2. The five Platonic solids along with (v, e, f), in which v denotes the number of vertices, e the number of edges, and f the number of faces. The surface of each Platonic solid is homeomorphic to a two-dimensional sphere. Since the Euler characteristic of a surface is a topological invariant, $v - e + f = 2$ for all five surfaces. Readers who prefer the terminology d4, d6, d8, d10, d12, and d20, respectively, for these objects gain 100 experience points.

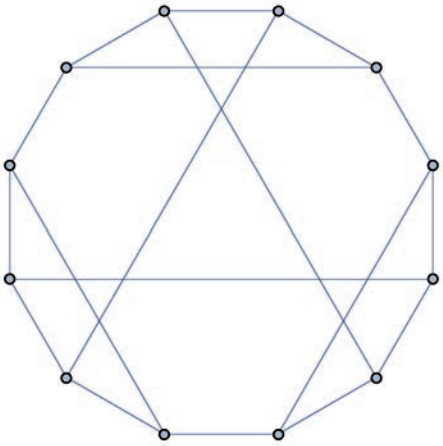


FIGURE 3. The Franklin graph can be embedded on the surface of the Klein bottle. The resulting map cannot be colored with fewer than six colors. Since Franklin proved that every map on a Klein bottle can be colored with at most six colors, this example shows that his bound is sharp.

FIGURE 4. The map at left can be wrapped onto the surface of a torus. This example shows that not every map on the torus can be colored with fewer than seven colors.

Bibliography

[1] K. Appel and W. Haken, *Every planar map is four colorable. I. Discharging*, Illinois J. Math. **21** (1977), no. 3, 429–490. http://www.projecteuclid.org/euclid.ijm/1256049011. MR0543792

[2] K. Appel, W. Haken, and J. Koch, *Every planar map is four colorable. II. Reducibility*, Illinois J. Math. **21** (1977), no. 3, 491–567. http://projecteuclid.org/euclid.ijm/1256049012. MR0543793

[3] K. Appel and W. Haken, *The solution of the four-color-map problem*, Sci. Amer. **237** (1977), no. 4, 108–121, 152, DOI 10.1038/scientificamerican1077-108. MR0543796

[4] K. Appel and W. Haken, *Every planar map is four colorable*, with the collaboration of J. Koch, Contemporary Mathematics, vol. 98, American Mathematical Society, Providence, RI, 1989. MR1025335

[5] P. Franklin, *A six color problem*, J. Math. Phys. **13** (1934), 363–379.

[6] G. Gonthier, *Formal proof—the four-color theorem*, Notices Amer. Math. Soc. **55** (2008), no. 11, 1382–1393. http://www.ams.org/notices/200811/tx081101382p.pdf. MR2463991
[7] P. J. Heawood, *Map-colour theorems*, Quarterly Journal of Mathematics, Oxford **24** (1890), 332–338.
[8] H. Hudson, *Four colors do not suffice*, Amer. Math. Monthly **110** (2003), no. 5, 417–423.
[9] S. J. Miller, *Mathematics of optimization: how to do things faster*, Pure and Applied Undergraduate Texts, vol. 30, American Mathematical Society, Providence, RI, 2017. MR3729274
[10] G. Ringel and J. W. T. Youngs, *Solution of the Heawood map-coloring problem*, Proc. Nat. Acad. Sci. U.S.A. **60** (1968), 438–445, DOI 10.1073/pnas.60.2.438. MR0228378
[11] R. Thomas, *An update on the four-color theorem*, Notices Amer. Math. Soc. **45** (1998), no. 7, 848–859. http://www.ams.org/notices/199807/thomas.pdf. MR1633714
[12] Wikipedia, *Four color theorem*, http://en.wikipedia.org/wiki/Four_color_theorem.

1977

RSA Encryption

Introduction

Alice and Bob wish to communicate without letting an eavesdropper, Eve, understand their conversation. Any information that they wish to exchange can be encoded with numbers (see the comments for the 1936 entry). Instead of sending one large number that represents an entire message, information is typically broken up into smaller blocks of fixed size. Thus, Alice and Bob want to securely send and receive nonnegative integers less than or equal to a fixed threshold while Eve is eavesdropping. Moreover, they need to do this without first exchanging a secret key for their code: otherwise Eve will know the key!

The RSA cryptosystem, invented by Ronald Rivest (1947–), Adi Shamir (1952–), and Leonard Adleman (1945–) in 1977 and, independently, by Clifford Cocks (1950–) of the UK intelligence agency GCHQ (Government Communications Headquarters) in 1973, addresses this issue (Cocks's work remained classified until 1997). Eve can listen to the entire RSA-encrypted communication and she will be unable to decipher it! Without algorithms such as RSA, modern e-commerce would be impossible: we can buy things online without meeting the seller in person to agree on a secret key for the transaction. To perform this amazing feat, Alice and Bob require some number theory.

To describe the RSA cryptosystem, we need Euler's generalization of Fermat's little theorem. Fermat's little theorem tells us that

$$a^{p-1} \equiv 1 \pmod{p}$$

if p is prime and $\gcd(a, p) = 1$; see the 2002 entry. Let $\phi(n)$ denote the number of $j \in \{1, 2, \ldots, n\}$ that are relatively prime to n. For example, $\phi(15) = 8$ since there are eight numbers, namely $1, 2, 4, 7, 8, 11, 13, 14$, in the specified range that are relatively prime to 15. The function ϕ is called the *Euler totient* function. It is multiplicative, in the sense that $\phi(mn) = \phi(m)\phi(n)$ if m and n are relatively prime. For example, $\phi(15) = \phi(3)\phi(5) = 2 \cdot 4 = 8$. Moreover, $\phi(p) = p - 1$ whenever p is prime, since $1, 2, \ldots, p - 1$ are relatively prime to p. Euler's theorem states that if $\gcd(a, n) = 1$, then

$$a^{\phi(n)} \equiv 1 \pmod{n}.$$

We are now ready to state the RSA algorithm.

RSA algorithm.

- Alice secretly selects distinct large primes p and q. Their product $n = pq$ is her *enciphering modulus*.
- Alice picks a *public key* (also called an *encryption key*) e. This is a positive integer such that $\gcd(e, \phi(n)) = 1$. She knows $n = pq$, so she can compute $\phi(n) = \phi(p)\phi(q) = (p-1)(q-1)$ and check if $\gcd(e, \phi(n)) = 1$ rapidly via the Euclidean algorithm.
- Alice's *private key* (also called a *decryption key*) d is the inverse of $e \pmod{\phi(n)}$. Thus, $de = j\phi(n) + 1$ for some integer j.
- Alice makes n and e known to the public. She does not disclose p, q, or d.
- To send the message $M \in \{1, 2, \ldots, n\}$ to Alice, Bob computes[1] $E \equiv M^e \pmod{n}$. He sends E to Alice.
- Alice recovers M from E as follows:[2]

$$E^d \equiv (M^e)^d \equiv M^{de} \equiv M^{j\phi(n)+1} \equiv (M^{\phi(n)})^j M \equiv M \pmod{n}.$$

Since n and e are publicly available, anyone can send messages to Alice. Only she can decrypt these messages because only she knows the private key d. Here is an example. Alice selects secret primes $p = 7{,}919$ and $q = 9{,}733$. Then $n = pq = 77{,}075{,}627$ and $\phi(n) = (p-1)(q-1) = 77{,}057{,}976$. Alice chooses $e = 47$ and checks that $\gcd(47, \phi(n)) = 1$. The multiplicative inverse of $47 \pmod{\phi(n)}$ is $d = 68{,}860{,}319$. Bob wants to send the message $M = 12{,}345$ to Alice. He computes

$$E \equiv M^e = (12{,}345)^{47} \equiv 18{,}269{,}972 \pmod{n}$$

and sends this to Alice, who receives it and computes

$$E^d = (18{,}269{,}972)^{68{,}860{,}319} \equiv 12{,}345 \pmod{n}.$$

Suppose that Eve wants to find M, knowing only E and Alice's public information, n and e. She needs Alice's private key d, so Eve must solve $de \equiv 1 \pmod{\phi(n)}$. To do this, Eve needs to know $\phi(n) = (p-1)(q-1)$. Since

$$(p-1)(q-1) = pq - p - q + 1 = n - (p+q) + 1,$$

knowing $\phi(n)$ is equivalent to knowing $p+q$. However, knowing $p+q$ is equivalent to knowing p and q since the roots of

$$(x-p)(x-q) = x^2 - (p+q)x + pq = x^2 - (p+q)x + n,$$

namely p and q, can be found by the quadratic formula. Thus, finding $\phi(n) = (p-1)(q-1)$ is as hard as factoring $n = pq$.

The security of RSA is based upon the assumption that it is hard to factor large numbers (even though it is easy to multiply them). If a method for fast factorization were to be found, then RSA would cease to be secure. Peter Shor (1959–) found such an algorithm for fast factorization, but it requires a quantum computer. Although quantum computers have so far only been able to factor relatively small

[1] Although exponentiating M modulo n appears to be a daunting task, it can be done rapidly by repeated squaring and modular reduction; see the 2002 entry.

[2] One can prove that $E^d \equiv M \pmod{n}$ even if $\gcd(M, n) \neq 1$.

numbers, the potential exists for them to one day factor RSA moduli. Other cryptographic systems, such as lattice-based methods, are believed to be more secure against quantum-computer attacks.

Centennial Problem 1977

Proposed by Steven J. Miller, Williams College.

Rivest, Shamir, and Adleman formed RSA Laboratories to market and further develop applications of the RSA cryptosystem, which was granted U.S. Patent 4,405,829. In 1991, the company announced fifty-four factoring challenges to encourage cryptographic research and to monitor the state of contemporary factoring algorithms and technology.

Each challenge number is the product of two large primes. These RSA challenge numbers were generated by an isolated computer, with no access to the internet, whose hard drive was immediately destroyed. Thus, we can be certain that if someone presents a factorization of an RSA challenge number, there was no cheating involved. Cash prizes were offered, ranging from $1,000 to $200,000. The challenge was officially closed in 2007, although many people continue to try to factor the RSA numbers.

As of 2017, the smallest unfactored RSA challenge number is RSA-230

$$\begin{aligned}&1796949159794106673291612844957324615636756180801260007088\\&9188355317264603414909334933722478686507552308558641999292\\&2181443668472287405206525793749569434838926317115252252565 44\\&1098081917061174250970244071801036483163828851885 2689,\end{aligned}$$

which has 230 digits. The largest of the challenge numbers is RSA-2048, which has 617 decimal digits (2048 bits). Without a major advance in quantum computation, RSA-2048 will probably never be factored.

The smallest RSA challenge number is RSA-100

$$\begin{aligned}&1522605027922533360535618378132637429718068114961380688 6579\\&0849458012296325895289765400035069200613 9.\end{aligned}$$

This 100-digit number was factored less than a month after the challenge began. Find the factors yourself!

1977: Comments

Poor choices and Pollard's $p-1$ algorithm. The security of RSA rests on the assumption that factoring $n = pq$ is computationally infeasible. However, there are some choices of p and q that render n susceptible to certain factorization algorithms. Suppose that $p-1$ has only small prime factors. For instance, the prime $p = 614{,}657$ is "large" but $p - 1 = 614{,}656 = 2^8 \cdot 7^4$ has only "small" prime factors. In this situation, *Pollard's $p-1$ algorithm* might be able to factor n in a reasonable amount of time. In what follows, we do not require that n is a product of two distinct primes.

The starting point of Pollard's algorithm is the observation that if $p-1$ does not have any large prime factors, then $(p-1)|k!$ for some small k. For example, if $p = 181$, then

$$p - 1 \;=\; 180 \;=\; 2^2 \cdot 3^2 \cdot 5$$

contains only small prime factors and $p-1$ divides $6! = 720 = 180 \cdot 4$. On the other hand, if $p = 179$, then $p-1 = 178 = 2 \cdot 89$ has a relatively large prime factor. Because of this, $p-1$ does not divide $k!$ for $k = 1, 2, \ldots, 88$, although it divides $89!$.

Suppose that p is a prime factor of n and $(p-1)|k!$. Then $k! = (p-1)r$ for some $r \in \mathbb{N}$ and Fermat's little theorem yields

$$2^{k!} = 2^{(p-1)r} \equiv (2^{p-1})^r \equiv 1^r \equiv 1 \pmod{p},$$

so $p|(2^{k!}-1)$. Although other bases may be used, the base 2 is preferred in practice since exponentiation with base 2 is particularly amenable to computation.

Let $m_k \equiv 2^{k!} - 1 \pmod{n}$ with $1 \le m_k \le n$. Since m_k and $2^{k!} - 1$ differ by a multiple of n, we have

$$\gcd(m_k, n) \;=\; \gcd(2^{k!} - 1, n) \;\ge\; p.$$

If n does not divide $2^{k!} - 1$, then $\gcd(m_k, n)$ is a proper divisor of n. In the preceding, we insisted that m_k is the least positive residue of $2^{k!} - 1$ modulo n since $m_k = 0$ implies that $\gcd(m_k, n) = n$ and hence we do not obtain a proper factor of n.

To implement Pollard's algorithm, fix a threshold K and compute $\gcd(m_k, n)$ for $k = 2, 3, \ldots, K$ and hope that a proper divisor of n is found. Observe that

$$m_k \;\equiv\; 2^{k!} - 1 \;\equiv\; (2^{(k-1)!})^k - 1 \;\equiv\; (m_{k-1} + 1)^k - 1 \pmod{n},$$

so the m_k can be computed iteratively without computing $k!$. This shortcut is important, since the rapid growth of $k!$ prevents the direct evaluation of m_k.

Here is an example. If $n = 26{,}016{,}619$, then

$$\begin{aligned}
&2^{2!} \equiv 4 \pmod{n}, && m_2 = 3, && \gcd(m_2, n) = 1,\\
&2^{3!} \equiv 4^3 \equiv 64 \pmod{n}, && m_3 = 63, && \gcd(m_3, n) = 1,\\
&2^{4!} \equiv 64^4 \equiv 16{,}777{,}216 \pmod{n}, && m_4 = 16{,}777{,}215, && \gcd(m_4, n) = 1,\\
&2^{5!} \equiv 16{,}777{,}216^5 \equiv 6{,}730{,}144 \pmod{n}, && m_5 = 6{,}730{,}143, && \gcd(m_5, n) = 1,\\
&2^{6!} \equiv 6{,}730{,}144^6 \equiv 14{,}067{,}788 \pmod{n}, && m_6 = 14{,}067{,}787, && \gcd(m_6, n) = 1,\\
&2^{7!} \equiv 14{,}067{,}788^7 \equiv 20{,}137{,}005 \pmod{n}, && m_7 = 20{,}137{,}004, && \gcd(m_7, n) = 5{,}419,
\end{aligned}$$

so $5{,}419|n$. In fact, $n = pq$, in which $p = 5{,}419$ and $q = 4{,}801$ are prime. Neither

$$p - 1 = 5{,}418 = 2 \cdot 3^2 \cdot 7 \cdot 43 \qquad \text{nor} \qquad q - 1 = 4{,}800 = 2^6 \cdot 3 \cdot 5^2$$

divides $7! = 5{,}040$. That is, the Pollard $p-1$ method was successful before our initial analysis predicted that it should be. This is because $2^{k!} - 1$ might be divisible by p by chance, as opposed to being divisible by p because $k!$ is a multiple of $p-1$. This is the case here, since $2^{7!} - 1$ happens to be divisible by p.

If Alice is careful in selecting her primes p and q, she can prevent Eve from factoring her RSA modulus $n = pq$ using Pollard's $p-1$ algorithm. Let p_0, q_0 be large primes. Then let p and q be even larger primes of the form

$$p = ip_0 + 1 \qquad \text{and} \qquad q = jq_0 + 1.$$

Dirichlet's theorem on primes in arithmetic progressions ensures that there are infinitely many such primes; see the comments for the 1913 entry. By construction,

$p-1=ip_0$ and $q-1=jq_0$ have the large prime factors p_0 and q_0, respectively. This prevents Eve from applying the Pollard $p-1$ algorithm effectively.

Answer to the problem. The factorization of RSA-100 is

$$3797522793694367392280887275544562785456553663819 9$$
$$\times 4009469095092088103068373529276146838921489972 4061$$

This was found in 1991 by Mark Manasse (1958–) and Arjen K. Lenstra (1956–) [**3**].

Bibliography

[1] R. Rivest, A. Shamir, and L. Adleman, *US Patent 4,405,829* (1977). `http://www.google.com/patents/US4405829`.

[2] R. L. Rivest, A. Shamir, and L. Adleman, *A method for obtaining digital signatures and public-key cryptosystems*, Comm. ACM **21** (1978), no. 2, 120–126, DOI 10.1145/359340.359342. `https://people.csail.mit.edu/rivest/Rsapaper.pdf`. MR700103

[3] RSA Laboratories, *RSA Honor Roll*, `http://www.ontko.com/pub/rayo/primes/hr_rsa.txt`

[4] Wikipedia, *Pollard's $p-1$ algorithm*, `https://en.wikipedia.org/wiki/Pollard's_p_-_1_algorithm`.

[5] Wikipedia, *RSA Factoring Challenge*, `http://en.wikipedia.org/wiki/RSA_Factoring_Challenge`.

[6] Wikipedia, *Shor's Algorithm*, `http://en.wikipedia.org/wiki/Shor's_algorithm`.

1978

Mandelbrot Set

Introduction

The Mandelbrot set is an example of a *fractal*, a mathematical object that possesses a great deal of self-similarity. It is constructed as follows. For each complex number c, form the sequence $z_{n;c}$, in which

$$z_{0;c} \;=\; c \qquad \text{and} \qquad z_{n+1;c} \;=\; z_{n;c}^2 + c.$$

The simplest pictures of the Mandelbrot set are obtained by coloring a point c black if the sequence defined above is bounded and white otherwise; see Figure 1. For finer detail, we can color points c whose sequences $z_{n;c}$ appear unbounded based upon how many iterations are needed to exceed a fixed, large threshhold; see Figure 2. One can zoom in on the Mandelbrot set and obtain a variety of beautiful and bewildering images; see Figure 3 and the links at [**9**].

One of the most important things to address with any iterative problem is the existence and classification of fixed points. If w is a fixed point of the map $p(z) = z^2 + c$, in which c is a constant, then $p(w) = w$; that is,

$$w^2 - w + c \;=\; 0.$$

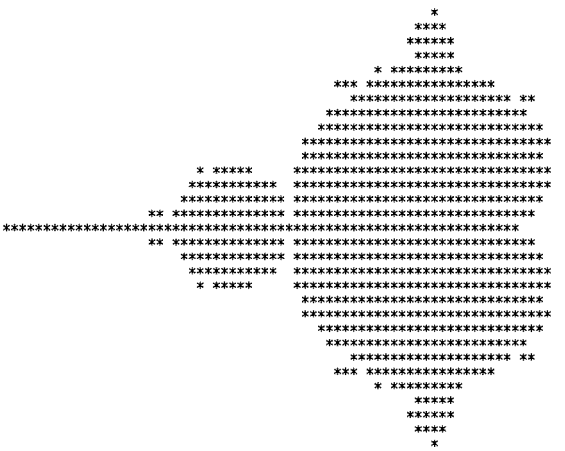


FIGURE 1. The first visualization of the Mandelbrot set was produced in 1978 by Robert W. Brooks (1952–2002) and J. Peter Matelski [**1**]. Image public domain.

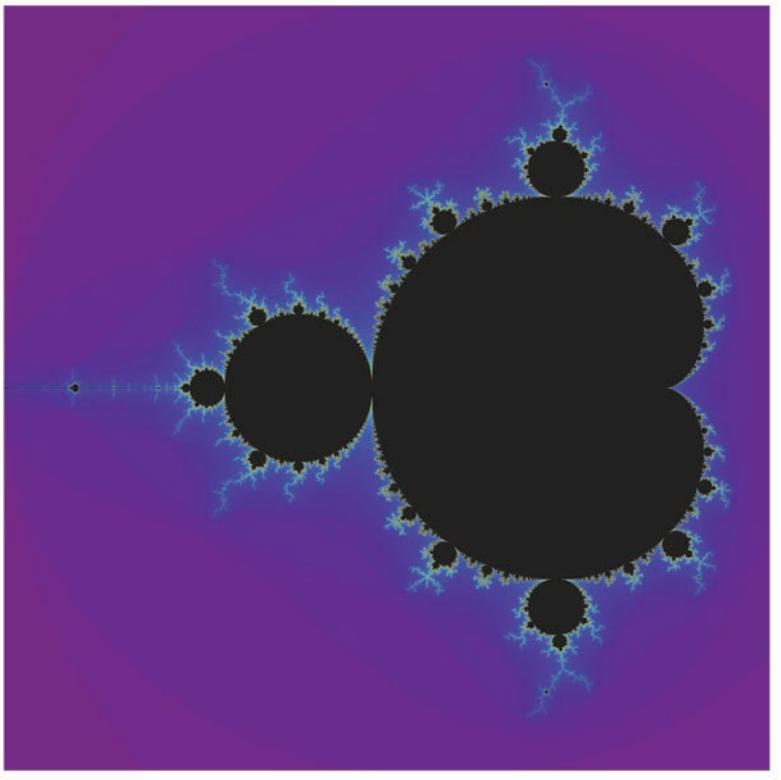


FIGURE 2. The Mandelbrot set.

This yields two fixed points (which coincide if $c = \frac{1}{4}$):

$$w \;=\; \frac{1 \pm \sqrt{1-4c}}{2}.$$

The magnitude of

$$p'(w) \;=\; 1 \pm \sqrt{1-4c}$$

determines the nature of the fixed point w. If $|p'(w)| < 1$, then w is an *attracting fixed point* and values that start out close to w will iterate toward w. If $|p'(w)| > 1$, then w is a *repelling fixed point* and values that start out close to w will iterate away from w. If $|p'(w)| = 1$, then the situation is more complicated and the argument of the complex number $p'(w)$ comes into play.

What about polynomials of higher degree? If p is a polynomial of degree n, then $p(w) = w$ means that w is a zero of the polynomial $h(z) = p(z) - z$, which has degree at most n. The fundamental theorem of algebra asserts that a polynomial of degree n has exactly n zeros, counted according to multiplicity, in the complex plane. Thus, p has at most n fixed points. What if the polynomial p is replaced with a slightly more exotic function?

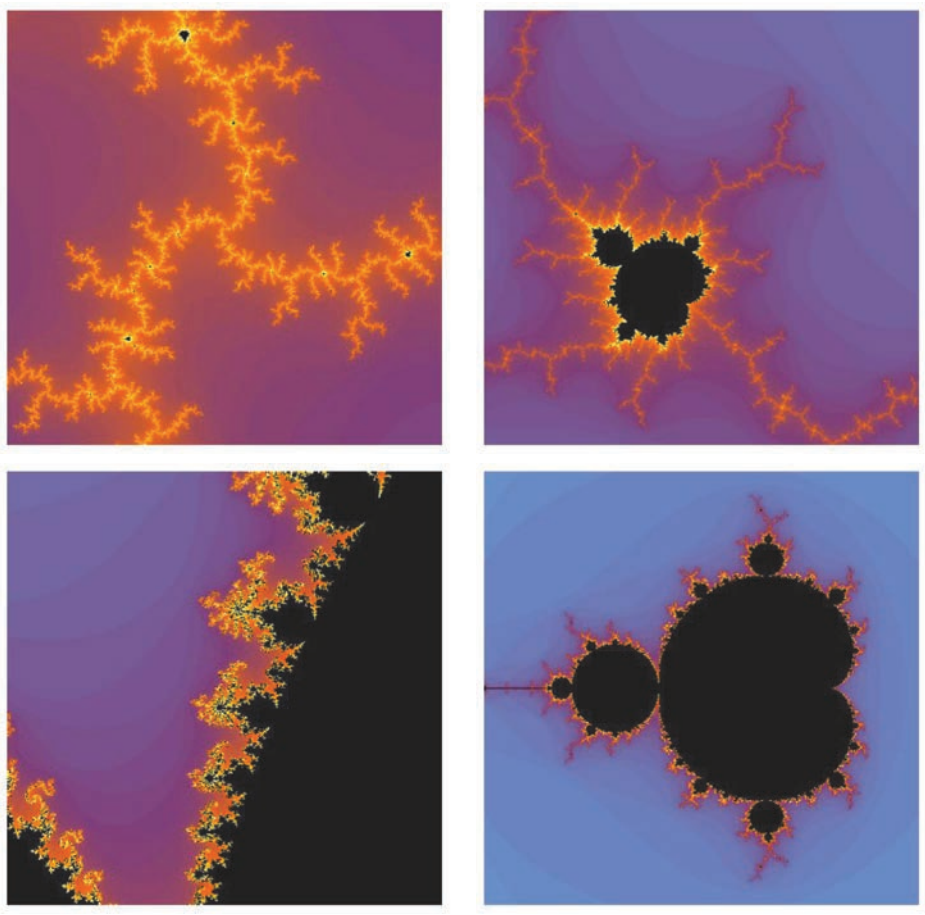


FIGURE 3. Several close-up images of the Mandelbrot set.

Centennial Problem 1978

Proposed by Stephan Ramon Garcia, Pomona College.

Let $p(z)$ be a complex polynomial of degree n. How many fixed points can $\overline{p(z)}$ have? That is, how many roots can the equation $p(z) = \overline{z}$ have? At most n? Infinitely many? Or something in between?

1978: Comments

Space Invaders. A strong contender for this year's topic was the video game *Space Invaders*[1] [**10**]. Created by Tomohiro Nishikado (1944–) and released in 1978, this mega-blockbuster game revolutionized the industry. Interestingly, one of the defining features of the game was due to hardware limitations. In the game, alien ships are attacking the Earth. As more and more of them are destroyed, the

[1]A common misconception is that the line "And the space he invades he gets by on you" from the 1981 Rush song *Tom Sawyer* is "And the space invaders get by on you." Certainly, the second is the more amusing interpretation.

remaining ships move faster and faster until the last few ships move at incredible speeds. This feature was due to a computational bottleneck. The fewer the number of ships that need to be drawn, the faster the computer could display them! Nishikado decided that he liked this and incorporated it into the game.

A continuous, nowhere-differentiable function. Self-similarity is a key ingredient in the construction of the *blancmange function*, a continuous, nowhere-differentiable function $f : \mathbb{R} \to \mathbb{R}$; see Figure 4. Since the original construction is due to Teiji Takagi (1875–1960) [**8**], this function is also called the *Takagi function*.

The first step is to prove that if $f : \mathbb{R} \to \mathbb{R}$ is differentiable at x, then

$$\lim_{n\to\infty} \frac{f(v_n) - f(u_n)}{v_n - u_n} \;=\; f'(x)$$

whenever u_n, v_n are sequences such that

(a) $u_n \le x < v_n$ for all $n \in \mathbb{N}$,

(b) $u_n < v_n$ for all $n \in \mathbb{N}$, and

(c) $\lim_{n\to\infty}(v_n - u_n) = 0$.

To do this, use the definition of the derivative as a limit of difference quotients and argue that it suffices to consider the case $f'(x) = 0$.

Given $x \in \mathbb{R}$, let $g(x)$ denote the distance from x to the nearest integer. The graph of $g(x)$ looks like a "sawtooth wave" with each "tooth" of height $1/2$ and width 1; see Figure 5. Use the Weierstrass M-test to prove that the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) \;=\; \sum_{n=0}^{\infty} \frac{g(2^n x)}{2^n} \tag{1978.1}$$

is continuous and bounded. Since $g(x)$ is periodic with period 1, it follows that $g(2^n x)$ is periodic with period $\frac{1}{2^n}$. If x is a dyadic rational number (that is, its denominator is a power of 2), then $2^k x$ is an integer whenever $k \ge n$ and hence

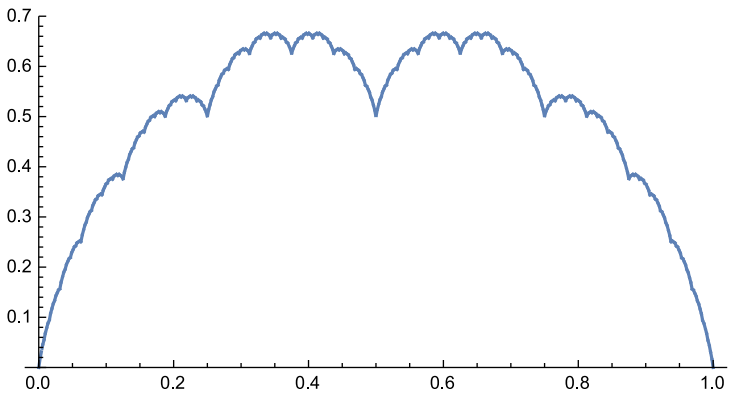


FIGURE 4. Graph of the blancmange function on $[0, 1]$. This function is continuous, but nowhere differentiable.

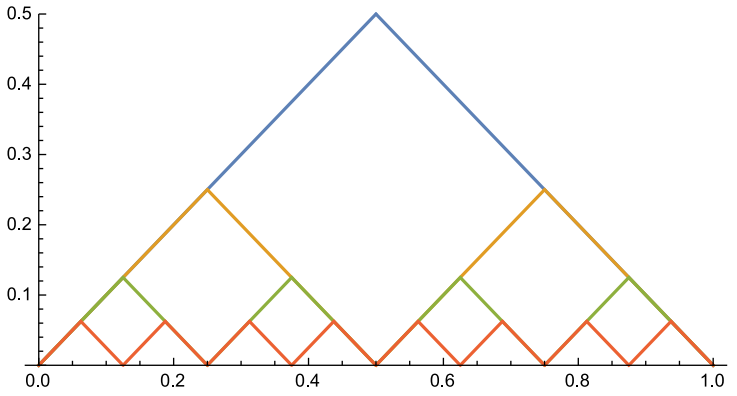


FIGURE 5. Graphs of the summands $g(x)$, $\frac{1}{2}g(2x)$, $\frac{1}{4}g(4x)$, $\frac{1}{8}g(8x)$ for $n = 1, 2, 3, 4$.

$g(2^k x) = 0$ for all $k \geq n$. Fix $x \in \mathbb{R}$. For each $n \in \mathbb{N}$, let

$$u_n \;=\; \frac{m_n}{2^n} \qquad \text{and} \qquad v_n \;=\; \frac{m_n + 1}{2^n}$$

be dyadic rational numbers that satisfy

$$u_n \;\leq\; x \;<\; v_n \qquad \text{and} \qquad v_n - u_n \;=\; \frac{1}{2^n}.$$

By the preceding remarks, the series for f reduces to

$$\frac{f(v_n) - f(u_n)}{v_n - u_n} \;=\; \sum_{k=0}^{n-1} \frac{g(2^k v_n) - g(2^k u_n)}{2^k v_n - 2^k u_n}. \tag{1978.2}$$

However,

$$\begin{aligned} 2^k u_n \;&=\; 2^{k-n} 2^n u_n \;=\; 2^{k-n} m_n, \quad \text{and} \\ 2^k v_n \;&=\; 2^{k-n}(m_n + 1), \end{aligned}$$

for some $m_n \in \mathbb{Z}$. Since $2^{k-n} \leq \frac{1}{2}$ for $k < n$, this means that g is linear on the interval $[2^k u_n, 2^k v_n]$. Thus, each of the difference quotients on the right side of (1978.2) is ± 1 (depending on whether m_n is even or odd). In other words,

$$\frac{f(v_n) - f(u_n)}{v_n - u_n} \;=\; \sum_{k=0}^{n-1} \pm 1 \tag{1978.3}$$

for some sequence of signs $\pm$. Since the terms of a convergent series must tend to zero, it follows that (1978.3) does not tend to a finite limit as $n \to \infty$. In light of (1978.1), we conclude that $f'(x)$ does not exist.

Answer to the problem. Let $p(z)$ and $q(z)$ be polynomials with $\deg p = n$, $\deg q = m$, and $m < n$. What is the maximum number of zeros of

$$h(z) = p(z) - \overline{q(z)}?$$

Terence Sheil-Small conjectured in 1992 that the sharp upper bound was n^2. This is indeed the case if $m = n$ or $m = n - 1$, as his student A. S. Wilmshurst proved [**11**]. What if $m < n - 1$? Wilmshurst conjectured that if $m = 1$, that is,

$$h(z) = p(z) - \overline{z},$$

then the number of zeros of h is at most $3n - 2$. This was proved in 2002 by Dmitry Khavinson (1956–) and Grzegorz Świątek using techniques from complex dynamics [**4**]; see [**3**] for an elegant exposition of this result and an application to gravitational lensing (also see the 1915 entry). The sharpness of the upper bound $3n - 2$ was proved in 2008 by Lukas Geyer [**2**].

Bibliography

[1] R. Brooks and J. P. Matelski, *The dynamics of* 2*-generator subgroups of* PSL(2, **C**), Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference (State Univ. New York, Stony Brook, N.Y., 1978), Ann. of Math. Stud., vol. 97, Princeton Univ. Press, Princeton, N.J., 1981, pp. 65–71. MR624805

[2] L. Geyer, *Sharp bounds for the valence of certain harmonic polynomials*, Proc. Amer. Math. Soc. **136** (2008), no. 2, 549–555, DOI 10.1090/S0002-9939-07-08946-0. MR2358495

[3] D. Khavinson and G. Neumann, *From the fundamental theorem of algebra to astrophysics: a "harmonious" path*, Notices Amer. Math. Soc. **55** (2008), no. 6, 666–675. MR2431564

[4] D. Khavinson and G. Świątek, *On the number of zeros of certain harmonic polynomials*, Proc. Amer. Math. Soc. **131** (2003), no. 2, 409–414, DOI 10.1090/S0002-9939-02-06476-6. MR1933331

[5] B. Mandelbrot, *Fractal aspects of the iteration of* $z \mapsto \lambda z(1 - z)$ *for complex* λ, z, Annals of the New York Academy of Sciences **357**, 249–259.

[6] B. B. Mandelbrot, *The fractal geometry of nature*, Schriftenreihe für den Referenten [Series for the Referee], W. H. Freeman and Co., San Francisco, Calif., 1982. MR665254

[7] Team Fresh, *Last Lights On—Mandelbrot fractal zoom to 6.066 e228 (*2^{760}*).* `http://vimeo.com/12185093`.

[8] T. Takagi, *A simple example of the continuous function without derivative*, Proc. Phys. Math. Japan, **1** (1901), 176–177.

[9] Wikipedia, *Mandelbrot set*, `http://en.wikipedia.org/wiki/Mandelbrot_set`.

[10] Wikipedia, *Space invaders*, `http://en.wikipedia.org/wiki/Space_Invaders`.

[11] A. S. Wilmshurst, *The valence of harmonic polynomials*, Proc. Amer. Math. Soc. **126** (1998), no. 7, 2077–2081, DOI 10.1090/S0002-9939-98-04315-9. MR1443416

1998

The Kepler Conjecture

Introduction

What is the densest way to pack spheres into n-dimensional space? In one dimension, each sphere is a line segment of length two and hence the densest packing consists of infinitely many line segments placed end to end. Thus, the *packing density* in one dimension is 1. In two dimensions the problem is somewhat harder. Here the "spheres" are disks of radius one. Joseph-Louis Lagrange (1736–1813) proved in 1773 that the hexagonal lattice packing (see Figure 1) is the densest possible lattice-based sphere packing in the plane. Its density is

$$\frac{\pi\sqrt{3}}{6} \approx 0.9069,$$

so about 90.7% of the plane is covered. Although Axel Thue had provided a flawed proof back in 1890, a complete proof that the hexagonal lattice packing is the densest of all possible packings, including irregular, non-lattice-based packings, came only in 1940, when it was established by László Fejes Tóth (1915–2005) [**14**].

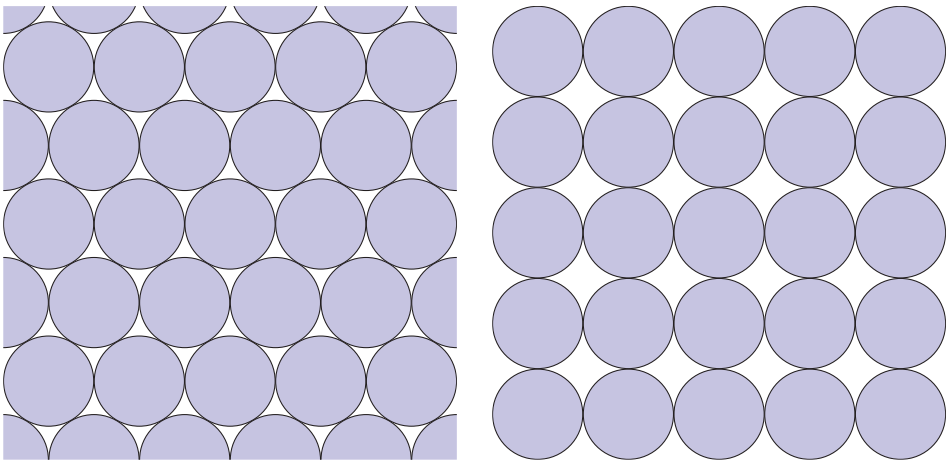


FIGURE 1. (LEFT) The densest sphere packing in two dimensions is the hexagonal lattice (honeycomb) packing. It covers approximately 90.7% of the plane. (RIGHT) The square lattice packing has density $4 - \pi \approx 0.8584$, so only around 85.8% of the plane is covered.

In 1611, Johannes Kepler (1571–1630) conjectured that the densest packing of identical spheres in three-dimensional space has density

$$\frac{\pi}{3\sqrt{2}} \approx 0.74048; \tag{1998.1}$$

that is, the spheres occupy about 74.05% of the available space. This is the famed Kepler conjecture. What made Kepler think of the number (1998.1)?

There are two familiar sphere packings in three dimensions: the hexagonal close and cubic close packings; see Figure 2. Both of these packings have density equal to (1998.1) and it seems impossible to do better.[1] Kepler was aware of the cubic close packing and conjectured that its density cannot be beaten [**9**, **10**]. The hexagonal close packing was only identified as a different packing by William Barlow (1845–1934) in 1883 [**2**].

The problem was brought to Kepler's attention by Thomas Harriot (ca. 1560–1621), who had been asked by Walter Raleigh (1554–1618) about the best way to stack cannonballs; see Figure 3. The problem was posed earlier (1611) than Fermat's last theorem (1637) and was solved shortly afterwards, making it an open, active problem for a longer period of time.

A proof of the conjecture was announced by Thomas C. Hales and his student Samuel P. Ferguson in 1998 (see [**13**, **15**] for summaries of the key ideas). Although it required a large number of computer-assisted computations, the proof did not spark nearly the level of philosophical debate that the proof of the four color theorem did over two decades earlier (see the 1976 entry).

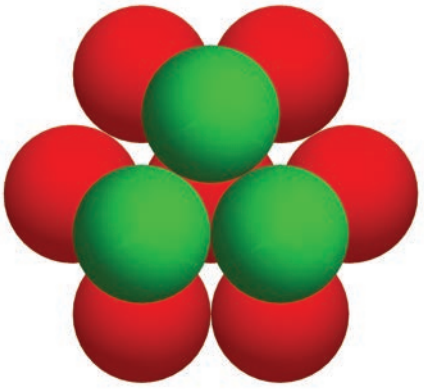
> [T]he proof was a 300-page monster that took 12 reviewers four years to check for errors. Even when it was published in the journal *Annals of Mathematics* in 2005, the reviewers could say only that they were "99 per cent certain" the proof was correct. [**1**]

Although the final paper was eventually published in a top peer reviewed journal [**4**], the entire process prompted an important question. How does one referee an argument where a significant amount of the argument is the result of running tens of thousands of lines of code? To address this, Hales began a collaborative project in 2003 to create a formal proof verifiable through automated proof checking software. Called Project Flyspeck (the "F," "P," and "K" stand for a "Formal Proof of Kepler"), it was successfully completed in 2014:

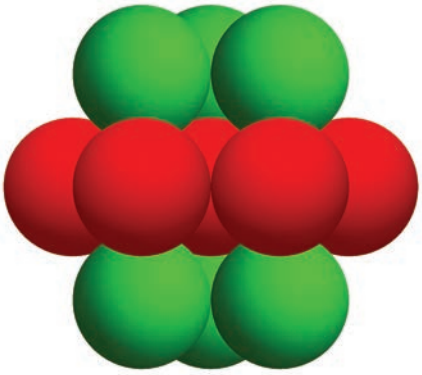
> So in 2003, Hales started the Flyspeck project, an effort to vindicate his[2] proof through formal verification. His team used two formal proof software assistants called Isabelle and HOL Light, both of which are built on a small kernel of logic that has been intensely scrutinised for any errors—this provides a foundation which ensures the computer can check any series of logical statements to confirm they are true . . . the Flyspeck team announced they had finally translated the dense mathematics of Hale's proof into computerised form, and verified that it is indeed correct.

[1] There are uncountably many packings that do just as well: study the key difference between the two packings in Figure 2 and see if you can use it to build more packings of the same density.

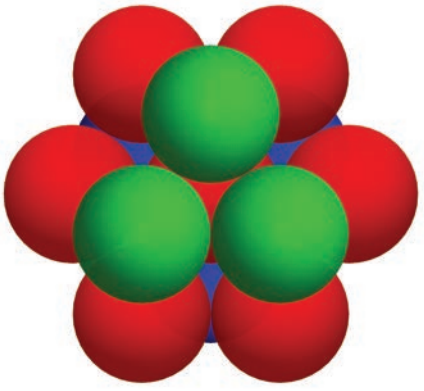
[2] Actually, the proof in the Flyspeck project involves a different local inequality based on later work of Christian Marchal [**12**]. In converting the proof ideas to formal form, Hales took advantage of this to get a local inequality that was cleaner and easier to prove by computer [**11**].



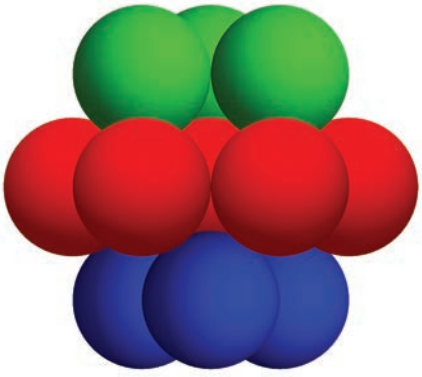
Hexagonal close packing (above)



Hexagonal close packing (front)



Cubic close packing (above)



Cubic close packing (front)

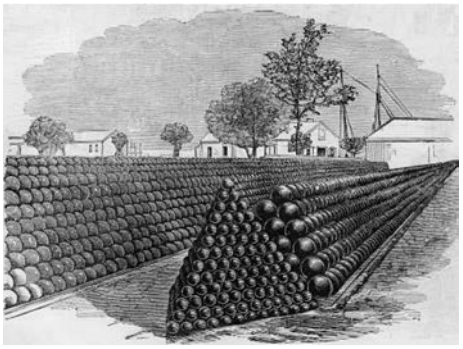
FIGURE 2. The cubic close and hexagonal close packings both have density $\pi/(3\sqrt{2}) \approx 0.74048$. The difference between the two packings is in the relative orientation of every other layer. The spheres in the hexagonal packing lie directly above the spheres two layers below. The spheres in the cubic close packing do not: consider the relative orientation of the green and blue triangles suggested by the top and bottom layers.

"This technology cuts the mathematical referees out of the verification process," says Hales. "Their opinion about the correctness of the proof no longer matters." [**1**]

Centennial Problem 1998

Proposed by Jeffrey Lagarias, University of Michigan.

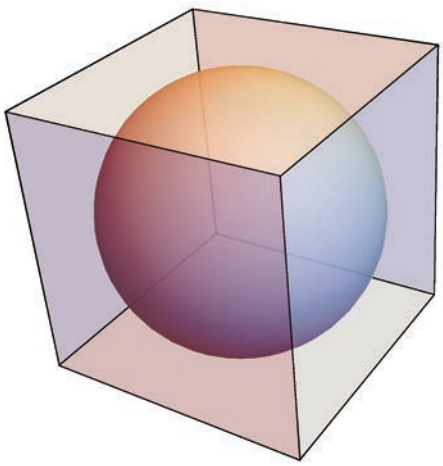
For $1 \leq n \leq 20$, determine the minimal side length $R(n)$ of a cube in which one can completely pack n unit-radius spheres. If you cannot get exact answers, determine upper and lower bounds.



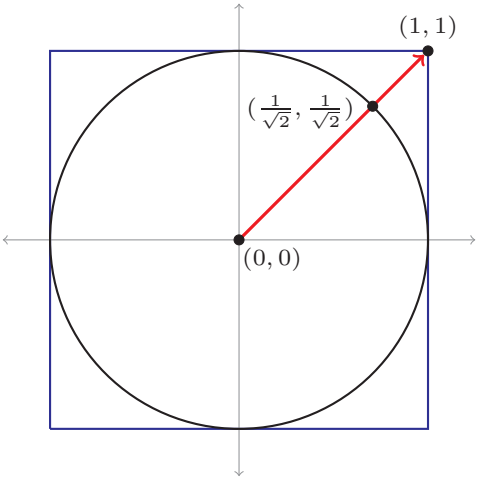
(a) A cubic close packing of cannonballs at Fort Monroe in Hampton, Virginia, in 1861 (image public domain).

(b) Snowballs packed in hexagonal close (front) and cubic close packings (rear) (image public domain).

FIGURE 3. Packings of cannonballs and snowballs.



(a) In two dimensions, the sphere occupies approximately 52.36% of the box that contains it.



(b) How does the distance between the corner of the cube to the nearest point of the sphere change as the dimension increases?

FIGURE 4. What proportion of an n-dimensional cube with side length 2 is taken up by the n-dimensional unit sphere?

1998: Comments

Cubes and spheres. What fraction of the n-dimensional cube (with sides of length 2) is taken up by the n-dimensional unit sphere? In two dimensions the area of the circle is π, giving a ratio of $\pi/4 \approx 0.785398$, while in three dimensions the volume of the sphere is $4\pi/3$, giving a ratio of $\pi/6 \approx 0.523599$; see Figure 4(a). One can show that in n dimensions the sphere has volume

$$V_n = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)},$$

TABLE 1. The ratio of the volume of the n-dimensional sphere to the n-dimensional cube tends to zero rapidly as n tends to infinity.

n	$r(n)$	$r(n)$ approx	n	$r(n)$	$r(n)$ approx	n	$r(n)$	$r(n)$ approx
1	1	1.	6	$\frac{\pi^3}{384}$	0.0807455	11	$\frac{\pi^5}{332640}$	0.000919973
2	$\frac{\pi}{4}$	0.785398	7	$\frac{\pi^3}{840}$	0.0369122	12	$\frac{\pi^6}{2949120}$	0.000325992
3	$\frac{\pi}{6}$	0.523599	8	$\frac{\pi^4}{6144}$	0.0158543	13	$\frac{\pi^6}{8648640}$	0.000111161
4	$\frac{\pi^2}{32}$	0.308425	9	$\frac{\pi^4}{15120}$	0.0064424	14	$\frac{\pi^7}{82575360}$	0.0000365762
5	$\frac{\pi^2}{60}$	0.164493	10	$\frac{\pi^5}{122880}$	0.00249039	15	$\frac{\pi^7}{259459200}$	0.0000116407

in which

$$\Gamma(s) \;=\; \int_0^\infty e^{-x}x^{s-1}dx, \qquad \operatorname{Re} s > 0,$$

is the gamma function. For positive integers n, we have

$$\Gamma(x) \;=\; \begin{cases} (n-1)! & \text{if } x = n, \\ \sqrt{\pi}\dfrac{(n-2)!!}{2^{\frac{n-1}{2}}} & \text{if } x = n + \frac{1}{2}, \end{cases}$$

in which $n!!$ denotes the product of every other term of the corresponding factorial. For example, $6!! = 6 \cdot 4 \cdot 2$ and $7!! = 7 \cdot 5 \cdot 3 \cdot 1$.

Using Stirling's formula (see the comments for the 1934 entry)

$$n! \;\approx\; n^n e^{-n}\sqrt{2\pi n},$$

it follows that the ratio

$$r(n) \;=\; \frac{\pi^{n/2}/\Gamma(\frac{n}{2}+1)}{2^n}$$

of the volumes of the n-dimensional sphere and cube tends to zero rapidly; see Table 1. Thus, in higher dimensions the sphere occupies very little of the cube. How can this be? Our low-dimensional intuition misleads us in higher dimensions. For example, the point

$$\frac{1}{\sqrt{n}}(1, 1, \ldots, 1) \in \mathbb{R}^n$$

lies on the n-dimensional sphere. Its distance to the corner $(1, 1, \ldots, 1)$ of the n-dimensional cube is

$$\sqrt{\sum_{i=1}^{n}\left(1 - \frac{1}{\sqrt{n}}\right)^2} \;=\; \sqrt{n}\left(1 - \frac{1}{\sqrt{n}}\right) \;=\; \sqrt{n} - 1,$$

which tends to infinity! This unexpected behavior is not evident in Figure 4(b).

Remark on the problem. One can show that $R(1) = 2$, and we think that $R(2) = 1 + \frac{2}{\sqrt{3}}$. Then things rapidly get tricky. There are some $n \le 20$ for which the exact answer is unknown. Some records for $1 \le n \le 32$ are in [**3**, **8**].

Bibliography

[1] J. Aron, *Proof confirmed of 400-year-old fruit-stacking problem*, New Scientist (August 12, 2014), https://www.newscientist.com/article/dn26041-proof-confirmed-of-400-year-old-fruit-stacking-problem.

[2] W. Barlow, *Probable nature of the internal symmetry of crystals*, Nature **29** (1883), 186–188.

[3] Th. Gensane, *Dense packings of equal spheres in a cube*, Electron. J. Combin. **11** (2004), no. 1, Research Paper 33, 17. http://www.combinatorics.org/ojs/index.php/eljc/article/view/v11i1r33/pdf. MR2056085

[4] T. C. Hales, *A proof of the Kepler conjecture*, Ann. of Math. (2) **162** (2005), no. 3, 1065–1185, DOI 10.4007/annals.2005.162.1065. http://annals.math.princeton.edu/2005/162-3/p01. MR2179728

[5] T. C. Hales, *Historical overview of the Kepler conjecture*, Discrete Comput. Geom. **36** (2006), no. 1, 5–20, DOI 10.1007/s00454-005-1210-2. http://link.springer.com/article/10.1007%2Fs00454-005-1210-2. MR2229657

[6] T. C. Hales and S. P. Ferguson, *A formulation of the Kepler conjecture*, Discrete Comput. Geom. **36** (2006), no. 1, 21–69, DOI 10.1007/s00454-005-1211-1. http://link.springer.com/article/10.1007%2Fs00454-005-1211-1. MR2229658

[7] T. C. Hales, J. Harrison, S. McLaughlin, T. Nipkow, S. Obua, and R. Zumkeller, *A revision of the proof of the Kepler conjecture*, Discrete Comput. Geom. **44** (2010), no. 1, 1–34, DOI 10.1007/s00454-009-9148-4. http://link.springer.com/article/10.1007%2Fs00454-009-9148-4. MR2639816

[8] A. Joós, *On the packing of fourteen congruent spheres in a cube*, Geom. Dedicata **140** (2009), 49–80, DOI 10.1007/s10711-008-9308-3. http://link.springer.com/article/10.1007%2Fs10711-008-9308-3. MR2504734

[9] C. Hardie, translation of J. Kepler's *Strena, seu de nive sexangula*, Oxford University Press, 2014.

[10] J. Kepler, *Strena, seu de nive sexangula*, Francofurti ad Moenum apud Godfefridum Tampach, 1611.

[11] J. C. Lagarias, *Dense sphere packings: a blueprint for formal proofs [book review of MR3012355]*, Bull. Amer. Math. Soc. (N.S.) **53** (2016), no. 1, 159–166, DOI 10.1090/bull/1502. MR3443950

[12] C. Marchal, *Study of the Kepler's conjecture: the problem of the closest packing*, Math. Z. **267** (2011), no. 3-4, 737–765, DOI 10.1007/s00209-009-0644-2. MR2776056

[13] J. Lagarias (ed.), *The Kepler Conjecture: The Hales-Ferguson Proof*, Springer-Verlag, 2011.

[14] L. F. Tóth, *Über die dichteste Kugellagerung*, Math. Z. **48** (1940), 676–684.

[15] S. J. Miller, *Mathematics of optimization: how to do things faster*, Pure and Applied Undergraduate Texts, vol. 30, American Mathematical Society, Providence, RI, 2017. MR3729274

2002

PRIMES in P

Introduction

Given a large integer n, how quickly can one determine whether it is prime or composite? The naive method is to divide n by each prime $2, 3, 5, 7, \ldots$. If one reaches $\sqrt{n}$ without finding a factor, then n is prime. However, if n has a few hundred digits, this approach can take longer than the age of the universe!

A more efficient approach is based upon Fermat's little theorem, which says that if p is prime and p does not divide a, then[1]

$$a^{p-1} \equiv 1 \pmod{p}. \tag{2002.1}$$

First, select an integer a. The Euclidean algorithm rapidly computes $\gcd(a, n)$ without the need to factor either number.[2] If $\gcd(a, n) \neq 1$, then n is composite. If $\gcd(a, n) = 1$ and $a^{n-1} \not\equiv 1 \pmod{n}$, then Fermat's little theorem ensures that n is composite (although it does not provide a specific factor of n). If $a^{n-1} \equiv 1 \pmod{n}$, then the test is inconclusive. In this case, repeat the test with another base a.

This can be implemented rapidly on a computer since a^{n-1} need not be computed directly. An example illustrates the approach. Suppose that we wish to determine whether $n = 1763$ is prime or composite. We first write 1763 in binary. Divide $n = 1763$ by the largest power of 2 that is at most n and repeat:

$$\begin{aligned} 1762 &= 1024 + 738, \\ 738 &= 512 + 226, \\ 226 &= 128 + 98, \\ 98 &= 64 + 34, \\ 34 &= 32 + 2. \end{aligned}$$

Thus,

$$\begin{aligned} 1763 &= 1024 + 512 + 128 + 64 + 32 + 2 \\ &= 2^{10} + 2^9 + 2^7 + 2^6 + 2^5 + 2^1 \\ &= (11011100010)_2. \end{aligned}$$

[1] To prove Fermat's little theorem, first show that $a, 2a, 3a, \ldots, (p-1)a$ are distinct and nonzero modulo p. Then $a, 2a, 3a, \ldots, (p-1)a$ are congruent modulo p to $1, 2, 3, \ldots, (p-1)$, in some order. Thus, $a \cdot 2a \cdot 3a \cdots (p-1)a \equiv 1 \cdot 2 \cdot 3 \cdots (p-1) \pmod{p}$, and hence $a^{p-1}(p-1)! \equiv (p-1)! \pmod{p}$. Since p does not divide $(p-1)!$, we obtain $a^{p-1} \equiv 1 \pmod{p}$.

[2] A theorem of Gabriel Lamé says that the number of steps in the Euclidean algorithm is at most five times the number of base-10 digits of $\min\{a, n\}$.

Repeated squaring and reduction modulo 1763 provide

$$\begin{aligned}
2^{2^0} &= 2^1 \equiv 2 \pmod{1763},\\
2^{2^1} &= 2^2 \equiv 4 \pmod{1763},\\
2^{2^2} &= 2^4 \equiv 16 \pmod{1763},\\
2^{2^3} &= 2^8 \equiv 256 \pmod{1763},\\
2^{2^4} &= 2^{16} \equiv 305 \pmod{1763},\\
2^{2^5} &= 2^{32} \equiv 1349 \pmod{1763},\\
2^{2^6} &= 2^{64} \equiv 385 \pmod{1763},\\
2^{2^7} &= 2^{128} \equiv 133 \pmod{1763},\\
2^{2^8} &= 2^{256} \equiv 59 \pmod{1763},\\
2^{2^9} &= 2^{512} \equiv 1718 \pmod{1763},\\
2^{2^{10}} &= 2^{1024} \equiv 262 \pmod{1763}.
\end{aligned}$$

Reducing modulo 1763 at each step, we obtain

$$\begin{aligned}
2^{1762} &\equiv 2^{2^{10}+2^9+2^7+2^6+2^5+2^1} \equiv 2^{2^{10}}2^{2^9}2^{2^7}2^{2^6}2^{2^5}2^{2^1} \pmod{1763}\\
&\equiv 262\cdot 1718\cdot 133\cdot 385\cdot 1349\cdot 4 \pmod{1763}\\
&\equiv 262\cdot 1718\cdot 133\cdot 385\cdot 107 \pmod{1763}\\
&\equiv 262\cdot 1718\cdot 133\cdot 646 \pmod{1763}\\
&\equiv 262\cdot 1718\cdot 1294 \pmod{1763}\\
&\equiv 262\cdot 1712 \pmod{1763}\\
&\equiv 742 \pmod{1763}.
\end{aligned}$$

Since $2^{1762} \not\equiv 1 \pmod{1763}$, Fermat's little theorem implies that 1763 is composite.

There are several important points here.

- We have proved that n is composite without providing a factor of n (for those dying of curiosity: $1763 = 41\cdot 43$).
- Judicious reduction modulo n means that our computations do not involve numbers that are significantly larger than n.
- The number of steps is proportional to $\log_2 n$ and not $\sqrt{n}$, as in the naive method.

Although the Fermat-based algorithm is fast, it is not always conclusive. For example,

$$n = 341 = 11\cdot 41 \qquad \text{and} \qquad 2^{340} \equiv 1 \pmod{341}.$$

We say that 341 is a *pseudoprime* for the base 2. There are infinitely many such numbers; see the comments below. The first few are

341, 561, 645, 1105, 1387, 1729, 1905, 2047, 2465, 2701, 2821,
3277, 4033, 4369, 4371, 4681, 5461, 6601, 7957, 8321, 8481, 8911,
10261, 10585, 11305, 12801, 13741, 13747, 13981, 14491, 15709,
15841, 16705, 18705, 18721, 19951.

Fortunately, $3^{340} \equiv 56 \pmod{341}$ and hence 3 is a *witness* to the fact that 341 is composite; that is, 341 is not a pseudoprime for the base 3. By testing an integer n with several different bases, we can weed out more pseudoprimes. Unfortunately, there are composite numbers n that are pseudoprime for all bases $2 \le k \le n-1$ with $\gcd(k,n) = 1$.[3] These *Carmichael numbers* always fool our Fermat-based primality test; see the 2010 entry.

Is there a polynomial-time algorithm that distinguishes primes and composites? By *polynomial time* we mean that there are constants $A, B > 0$ such that the number of elementary steps performed by the algorithm on the input n is at most $A(\log n)^B$. The focus on $\log n$ is because the length of the decimal (or binary) representation of n is proportional to $\log n$.

There are algorithms that depend upon randomly selected parameters that can do the job. One example is the *Miller–Rabin test*, named after Gary Lee Miller and Michael Oser Rabin (1931–). Let $n > 2$ and write $n - 1 = 2^r m$, in which $m \ge 1$ is odd and $r \ge 0$. If

$$b^m \equiv 1 \pmod{n} \quad \text{or} \quad b^{2^j m} \equiv -1 \pmod{n} \quad \text{for some } j \in \{0, 1, 2, \ldots, r-1\},$$

then n *passes Miller's test for the base* b. If n fails the test for some base b, then it is composite. It can be shown that if n is an odd composite number, then n passes Miller's test for at most $(n-1)/4$ bases b with $1 \le b \le n-1$.[4] This yields the Miller–Rabin probabilistic primality test: if n passes Miller's test for k different bases, then the probability that n is composite is at most $1/4^k$. For example, if n passes the test for $k = 50$ bases, then this probability is $1/4^{50} \approx 7.89 \times 10^{-31}$. Although we are not 100% certain that n is a prime, our level of confidence is sufficient for most industrial applications. Sometimes speed is more important than absolute certainty.

It is conceivable, although highly unlikely, that n is composite but that we continually pick from among the $(n-1)/4$ "bad" bases. Thus, we cannot guarantee that the Miller–Rabin test will work in polynomial time. On the other hand, the Adleman–Huang test is a random procedure that is guaranteed to find a proof of primality for a prime input in polynomial time [**1**].

What we really want is a deterministic, polynomial-time algorithm that distinguishes primes and composites. Over the years there were some close calls, but it was not until an electrifying announcement from India in 2002 that we had an answer. Manindra Agrawal (1966–) and his two undergraduate honors students Neeraj Kayal (1979–) and Nitin Saxena (1981–) gave a fairly simple deterministic, polynomial-time algorithm that distinguishes primes from composites. It involves a generalization of Fermat's little theorem to the ring of polynomials over a finite field of prime order modulo an irreducible polynomial.

We follow the description of the AKS primality test (named for Agrawal, Kayal, and Saxena) in [**3**], which also contains a number of worked examples. We first require some preliminaries. Recall that

$$(\mathbb{Z}/n\mathbb{Z})^\times = \{k \in \{1, 2, \ldots, n-1\} : \gcd(k,n) = 1\}$$

[3] If $\gcd(k,n) \ne 1$, then we already know that n is composite.

[4] If the generalized Riemann hypothesis is true, then for every composite integer n, there is a $b < 2(\log_2 n)^2$ for which n fails Miller's test for the base b.

is a group under multiplication modulo n. The *order* of $x \in (\mathbb{Z}/n\mathbb{Z})^\times$ is the smallest natural number k such that $x^k \equiv 1 \pmod{n}$. For example, $(\mathbb{Z}/12\mathbb{Z})^\times = \{1, 5, 7, 11\}$. Each element has order 2 since

$$1^2, 5^2, 7^2, 11^2 \equiv 1 \pmod{12}.$$

For polynomials $f(x)$, $g(x)$, and $m(x)$ with integer coefficients and $\deg m(x) \geq 1$, we say that

$$f(x) \equiv g(x) \pmod{m(x)} \qquad \Longleftrightarrow \qquad m|(f-g),$$

that is, if and only if there is a polynomial $h(x)$ with integer coefficients such that

$$h(x)m(x) = f(x) - g(x).$$

For example,

$$3x^2 + 7x + 4 \equiv x^2 + 2x + 1 \pmod{(x+1)}$$

since

$$(3x^2 + 7x + 4) - (x^2 + 2x + 1) = (2x+3)(x+1).$$

The great insight of Agrawal–Kayal–Saxena was to combine regular and polynomial congruences. We say that

$$f(x) \equiv g(x) \pmod{n, m(x)}$$

if there is an $h(x)$ with

$$f(x) - g(x) - h(x)m(x) \equiv 0 \pmod{n}.$$

Although we can describe the AKS primality test, showing that it runs in polynomial time would take us too far afield. See the original paper [**2**] or the exposition in [**3**].

AKS primality test. Let $N > 1$ be a positive integer.

1. Test if N is a perfect kth power for some $k \geq 2$. If it is, then N is composite and stop. Else proceed to step 2.
2. Find the smallest prime r such that the order of N modulo r is greater than $(\log_2 N)^2$.
3. If any of the numbers in $\{2, 3, \ldots, r\}$ share a common divisor other than 1 with N, then N is composite and stop. Else proceed to step 4.
4. If $N \leq r$, then N is prime and stop. Else proceed to step 5.
5. For each positive integer a at most $\sqrt{\phi(r)}\log_2 N$, check if

   $$(x+a)^N \equiv x^N + a \pmod{x^r - 1, N}.$$

   If there is an a for which the congruence fails, then N is composite; if the congruence holds for all such a at most $\sqrt{\phi(r)}\log_2 N$, then N is prime.

If the AKS primality test terminates in either step 1 or 3, then it produces a factor of N. This is done by applying the Euclidean algorithm in step 3 to r and N to find their greatest common divisor. If the program ends in step 5, then N is composite but we do not obtain a factor.

Agrawal, Kayal, and Saxena were successful in de-randomizing the prime recognition problem. Here is another problem for which there is a random polynomial-time algorithm, yet for which we do not know if there is a deterministic polynomial-time algorithm.

Centennial Problem 2002

Proposed by Carl Pomerance, Dartmouth College.

An integer a is a *quadratic nonresidue* modulo p if $x^2 \equiv a \pmod{p}$ has no solutions. Exactly half of the nonzero residues modulo p fit the bill. A candidate can be checked (in polynomial time) via Euler's criterion or the reciprocity law for Jacobi symbols. Thus, randomly selecting nonzero residues a until you get a quadratic nonresidue should succeed in around two tries!

A possible deterministic algorithm sequentially tries small a until a quadratic nonresidue is found. This works well for a large proportion of the primes. For example, one of $-1, 2, 3, 5$ is a quadratic nonresidue for an odd prime p unless $p \equiv 1$ or $49 \pmod{120}$. It is believed that this procedure works in polynomial time, but this is only known under the extended Riemann hypothesis.

Another possible strategy is to start with -1 and sequentially take square roots modulo p until a nonsquare is found. Unfortunately, we know no method to take modular square roots in deterministic polynomial time, unless one has an oracle that provides quadratic nonresidues!

Is there a deterministic, polynomial-time algorithm to produce a quadratic nonresidue modulo an odd prime p?

2002: Comments

Infinitude of base-2 psuedoprimes. To demonstrate that there are infinitely many pseudoprimes for the base 2, it suffices to show that for each odd, base-2 pseudoprime, there is a larger odd one. We start our construction with $n = 341$. Let n be an odd pseudoprime for the base 2 and let

$$M_n \;=\; 2^n - 1$$

denote the nth Mersenne number, which is known to be composite (see the 1996 entry). Because $2^{n-1} \equiv 1 \pmod{n}$, we have

$$M_n - 1 \;=\; 2^n - 2 \;=\; 2(2^{n-1} - 1) \;=\; 2dn$$

for some d. Thus,

$$\begin{aligned} 2^{(M_n-1)/2} - 1 \;&=\; 2^{dn} - 1 \\ &=\; (2^n - 1)(2^{n(d-1)} + 2^{n(d-2)} + \cdots + 2^n + 1) \\ &=\; M_n(2^{n(d-1)} + 2^{n(d-2)} + \cdots + 2^n + 1) \\ &\equiv\; 0 \pmod{M_n}. \end{aligned}$$

Since $M_n > n$ is composite and

$$2^{M_n-1} \;\equiv\; (2^{(M_n-1)/2})^2 \;\equiv\; 1^2 \;\equiv\; 1 \pmod{M_n},$$

we conclude that M_n is a pseudoprime for the base 2.

Although there are infinitely many pseudoprimes for the base 2, our method does not provide an efficient method for producing them. Indeed, $M_{341} = 2^{341} - 1$ is far larger than 561, the smallest pseudoprime for the base 2 after 341. The number 561 is also the first Carmichael number; see the 2010 entry.

Carl Pomerance alerted us to a simpler proof of the infinitude of base-2 pseudoprimes. We claim that if $p \geq 5$ is prime, then $n = (4^p - 1)/3$ is a base-2 pseudoprime. First observe that $4^p \equiv 1 \pmod 3$, so n is indeed an integer. Moreover, $(2^p + 1)/3$ is an integer and hence $n = (2^p - 1)((2^p + 1)/3)$ is composite. Fermat's theorem ensures that

$$n \;\equiv\; (2^p - 1)(2^p + 1)3^{-1} \;\equiv\; (2 - 1)(2 + 1)3^{-1} \;\equiv 1 \pmod p.$$

Since $n - 1$ is even, we have $(n-1)/2 \equiv 0 \pmod p$ and hence $2^{n-1} - 1 = 4^{(n-1)/2} - 1$ is divisible by $4^p - 1$. Thus, n is a base-2 pseudoprime.

Bibliography

[1] L. M. Adleman and M.-D. A. Huang, *Primality testing and abelian varieties over finite fields*, Lecture Notes in Mathematics, vol. 1512, Springer-Verlag, Berlin, 1992. MR1176511

[2] M. Agrawal, N. Kayal, and N. Saxena, *PRIMES is in P*, Ann. of Math. (2) **160** (2004), no. 2, 781–793, DOI 10.4007/annals.2004.160.781. `http://www.cse.iitk.ac.in/users/manindra/algebra/primality_v6.pdf`. MR2123939

[3] M. Cozzens and S. J. Miller, *The mathematics of encryption: An elementary introduction*, Mathematical World, vol. 29, American Mathematical Society, Providence, RI, 2013. MR3098499

[4] R. Crandall and C. Pomerance, *Prime numbers: A computational perspective*, 2nd ed., Springer, New York, 2005. MR2156291

[5] A. Granville, *It is easy to determine whether a given integer is prime*, Bull. Amer. Math. Soc. (N.S.) **42** (2005), no. 1, 3–38, DOI 10.1090/S0273-0979-04-01037-7. `http://www.dms.umontreal.ca/~andrew/PDF/Bulletin04.pdf`. MR2115065

[6] C. Pomerance, *Primality testing: variations on a theme of Lucas*, Congr. Numer. **201** (2010), 301–312. `https://math.dartmouth.edu/~carlp/PDF/lucastalk.pdf`. MR2598366

2003

Poincaré Conjecture

Introduction

In 2003, Grigori Perelman, building upon seminal work of Richard S. Hamilton (1943–), proved the Poincaré conjecture, one of the million-dollar Clay Millennium Problems (see the comments for the 2000 entry). The conjecture asserts that every smooth, compact, simply connected, closed 3-manifold is homeomorphic to the 3-sphere

$$\big\{(x, y, z, w) \in \mathbb{R}^4 : x^2 + y^2 + z^2 + w^2 = 1\big\}.$$

Two manifolds are *homeomorphic* if there is a continuous bijection between them that has a continuous inverse. For example, a circle and the trefoil knot (see the 1985 entry) are homeomorphic 1-manifolds, even though they cannot be continuously deformed into each other when embedded in $\mathbb{R}^3$. On the other hand, the 2-sphere and the Euclidean plane are not homeomorphic: one is compact and the other is not.[1]

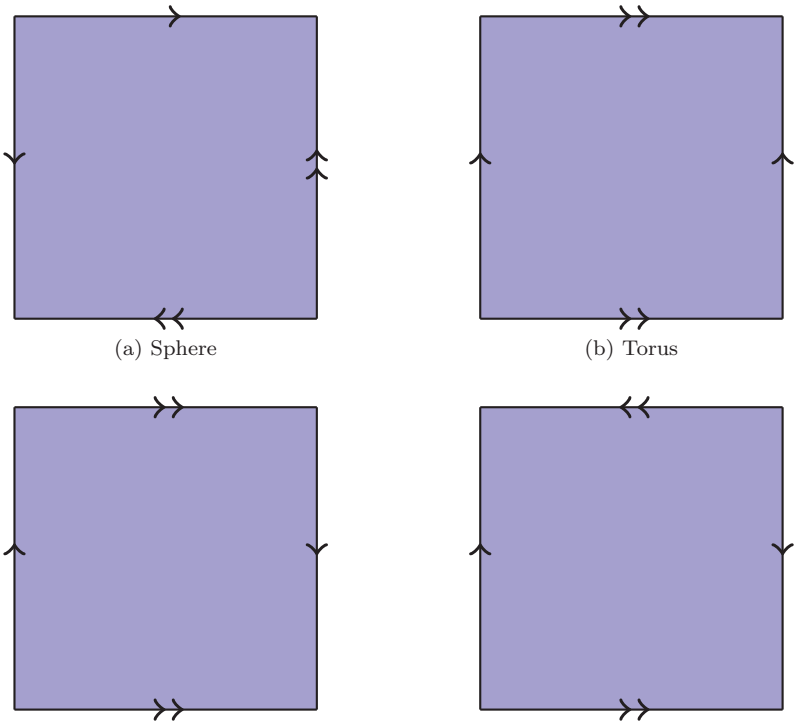
A particularly down-to-earth explanation of the main difficulty behind the Poincaré conjecture was recalled by Gerry Myerson in 2012 [**3**]:

> I once heard an expert "explain" the difficulty of the $n = 3$ case to a general audience by saying something like this: when $n \leq 2$, there isn't enough room for anything to go wrong, while for $n \geq 4$, there's enough room to fix anything that goes wrong; for $n = 3$, there's enough room for something to go wrong, and... it's not clear whether there's enough room to fix things when they go wrong.

The cases $n = 1$ and $n = 2$ are classical and date back to the foundations of algebraic topology. Stephen Smale proved the conjecture for $n \geq 4$ in 1961 and Michael Freedman (1951–) proved it for $n = 4$ in 1982. Since both of them received Fields Medals for their work, one can claim that the Poincaré conjecture resulted in either two or three medals, depending upon how one accounts for the enigmatic Perelman (see the comments below).

Although the resolution of the Poincaré conjecture is hopelessly beyond the level of this book and the expertise of its authors, we can discuss its analogue for 2-manifolds. By a *surface*, we mean a smooth, connected, two-dimensional manifold. Think of this as a nice topological space that locally resembles $\mathbb{R}^2$ and does not consist of multiple disjoint pieces. For example, a microscopic observer on a torus or Klein bottle will believe their local environment is flat and two dimensional, much as we perceive the ground around us as flat. A surface is *closed* if it is compact

[1] However, the 2-sphere with a point removed is homeomorphic to the plane via stereographic projection.



(a) Sphere (b) Torus

(c) Klein bottle (d) Projective plane

FIGURE 1. Fundamental polygons for several 2-manifolds. More sides may be necessary for more complicated manifolds, such as a sphere with several handles attached.

and has no boundary. For example, the sphere is closed, but the Möbius strip is not since its boundary is a circle (see the 1958 entry). A closed surface can be diagrammed using a *fundamental polygon*, an even-sided polygon with some of its edges identified; see Figure 1 and the comments for the 1958 entry.

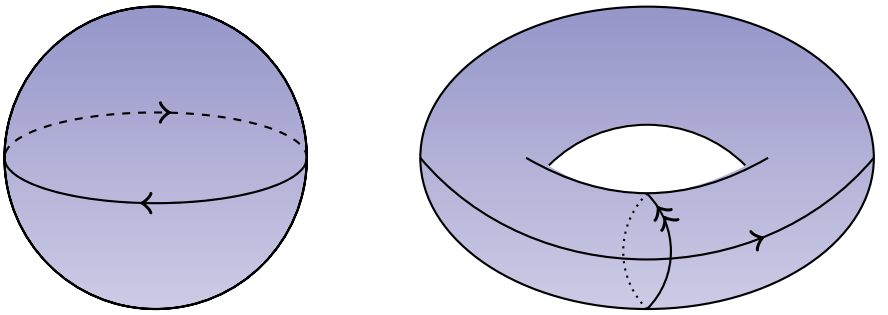
A surface is *simply connected* if every loop on the surface can be contracted to a point without leaving the surface. For example, the sphere is simply connected and the torus is not; see Figure 2. The analogue of the Poincaré conjecture for 2-manifolds asserts that every simply connected, closed surface is homeomorphic to the sphere. This is a consequence of the classification of surfaces from algebraic topology, which says that every closed surface is homeomorphic to one of the following:

(a) the sphere,

(b) the connected sum of tori, or

(c) the connected sum of real projective planes;

see the comments below for information about the connected sum of manifolds. Every surface in the first two classes is orientable; every surface in the third class is nonorientable. Of these, the only simply connected surface is the sphere; this implies the Poincaré conjecture for 2-manifolds.



(a) Every path on the sphere is contractible to a point. Thus, the sphere is simply connected.

(b) Neither of these two paths on the torus is contractible to a point.

FIGURE 2. The sphere is simply connected and the torus is not.

The problem for this year was originally posed by Frank Morgan of Williams College and it concerned 4-manifolds. However, he felt that the statement was too imprecise to be included here. Instead, we present a simple combinatorial problem with a visual twist that builds upon the comments to the 1980 entry. See below for the solution.

Centennial Problem 2003

Proposed by Stephan Ramon Garcia, Pomona College.

We saw in the 1980 entry that it is impossible to tile, with nonoverlapping 2×1 black-and-white dominoes, a chessboard that has two corners removed (while respecting the underlying black-and-white pattern). Is such a tiling possible if two squares of different colors are removed instead (see Figure 3)?

2003: Comments

Perelman's Fields Medal. Contrary to popular belief, Perelman did not receive the prestigious Fields Medal for his resolution of the Poincaré conjecture. He declined the award and did not even attend the award ceremony:

> In May 2006, a committee of nine mathematicians voted to award Perelman a Fields Medal for his work on the Poincaré conjecture. However, Perelman declined to accept the prize. Sir John Ball, president of the International Mathematical Union, approached Perelman in Saint Petersburg in June 2006 to persuade him to accept the prize. After 10 hours of attempted persuasion over two days, Ball gave up. Two weeks later, Perelman summed up the conversation as follows: "He proposed to me three alternatives: accept and come; accept and don't come, and we will send you the medal later; third, I don't accept the prize. From the very beginning, I told him I have chosen the third one... [the prize] was completely irrelevant for me. Everybody understood that if the proof is correct, then no other recognition is needed." [**9**]

In 2010, Perelman also declined the million-dollar prize offered by the Clay foundation (see the comments for the 2000 entry).

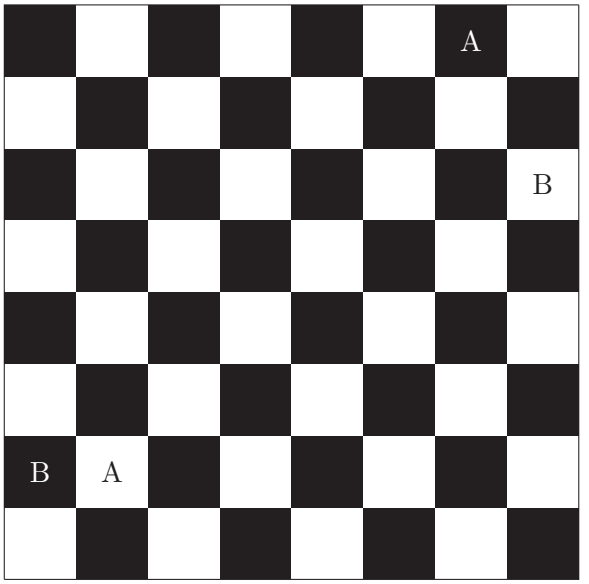


FIGURE 3. Is it possible to tile, with 2×1 black-and-white dominoes, a chessboard that has two squares of different colors removed? What if both squares marked "A" are removed? What if both squares marked "B" are removed?

A monoid of manifolds. A *monoid* is an algebraic structure similar to a group, except that inverses need not exist. To be more specific, a monoid is a set that is closed under an associative binary operation for which an identity element exists. What is the relationship between monoids and surfaces?

Given two surfaces M and N, their *connected sum* $M\#N$ is the surface obtained by removing a disk from each of M and N and then gluing the resulting boundary circles together [**10**]. One can show that the homeomorphism class of the resulting surface is independent of the location of the excised disks.

Let S denote the (two-dimensional) sphere, K the Klein bottle, T the torus, and P the (real) projective plane. The sphere is the identity element for the connected sum operation, in the sense that $S\#M = M$ for all surfaces M. This is because if we remove a disk from S, then the resulting surface can be deformed into a disk that takes the place of the disk removed from M.

What about the connected sum of a surface with a torus? In visual terms, $M\#T$ is "M with a handle attached." What does attaching a Klein bottle to a surface mean? If M is orientable, then $M\#K$ can be regarded as "M with a handle whose ends are attached to opposite sides of M." The projective plane P is not orientable, so the notion of "side" is meaningless. This is reflected in the algebraic relation $P\#T = P\#K$. One can also show that $P\#P = K$ and $P\#K = P\#T$. These computations can be summarized succinctly as follows. The monoid of homeomorphism classes of surfaces is the commutative monoid with

identity S that is generated by P and T, modulo the single relation

$$P\#P\#P \;=\; P\#T.$$

This identity is called Dyck's theorem, after Walther von Dyck.

The connected sum operation is compatible with the Euler characteristic (see the comments for the 1976 entry) in the following sense:

$$\chi(M\#N) \;=\; \chi(M) + \chi(N) - 2.$$

Since $\chi(S) = 2$, $\chi(P) = 1$, and $\chi(T) = 0$, it follows that

$$\chi(\underbrace{T\#T\#\cdots\#T}_{k \text{ times}}) \;=\; 2 - 2k \qquad \text{and} \qquad \chi(\underbrace{P\#P\#\cdots\#P}_{k \text{ times}}) \;=\; 2 - k.$$

Putting this all together, we see that a closed surface is completely determined, up to homeomorphism, by its Euler characteristic and orientability. If a surface is nonorientable, then it is homeomorphic to a connected sum of projective planes. On the other hand, an orientable surface is homeomorphic either to a sphere or a connected sum of tori. The number of summands, in both cases, can be discerned by computing the Euler characteristic of the given surface [**10**].

Solution to the problem. The elegant solution to our problem is due to Ralph E. Gomory (1929–) [**1**]. Consider the path llustrated in Figure 4. Suppose that two squares of different colors are removed from the board. Then they are separated, in either direction along the path, by an even number of squares, half of which are black and half of which are white. Thus, the desired tiling exists and, moreover, Figure 4 suggests an algorithm to efficiently produce it.

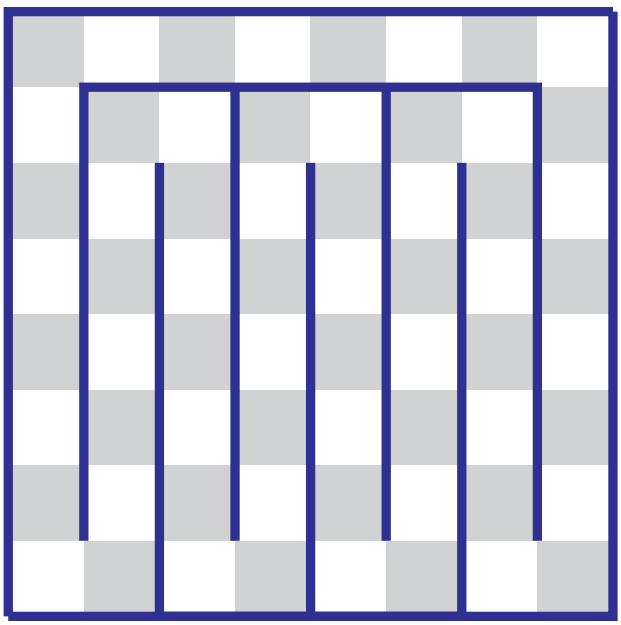


FIGURE 4. The chessboard can be regarded as a cycle graph of length 64.

What happens if we replace the standard 8×8 board with an 8×9 board? A 9×9 board? More generally, when can we tile an $m \times n$ board that has two squares of the same color removed?

Bibliography

[1] R. Honsberger, *Mathematical Gems I*, Mathematical Association of America, 1974.

[2] J. Milnor, *Poincare Conjecture*, http://www.claymath.org/millennium-problems/poincare-conjecture.

[3] G. Myerson, *Poincare conjecture for $n = 2$* (answer), https://math.stackexchange.com/questions/103182/poincare-conjecture-for-n-2.

[4] S. Nasar and D. Gruber, *Manifold Destiny: A legendary problem and the battle over who solved it*, The New Yorker, https://www.newyorker.com/magazine/2006/08/28/manifold-destiny.

[5] G. Perelman, *The entropy formula for the Ricci flow and its geometric applications*, https://arxiv.org/abs/math/0211159.

[6] G. Perelman, *Ricci flow with surgery on three-manifolds*, https://arxiv.org/abs/math/0303109.

[7] G. Perelman, *Finite extinction time for the solutions to the Ricci flow on certain three-manifolds*, https://arxiv.org/abs/math/0307245.

[8] T. Tao, *Perelman's proof of the Poincaré conjecture: a nonlinear PDE perspective*, https://arxiv.org/abs/math/0610903.

[9] Wikipedia, *Grigori Perelman*, https://en.wikipedia.org/wiki/Grigori_Perelman.

[10] Wikipedia, *Surface (topology)*, https://en.wikipedia.org/wiki/Surface_(topology).

2004

Primes in Arithmetic Progression

Introduction

2004 is another year that witnessed the announcement of two major results, each of which is worthy of a whole entry in its own right. One was the culmination of decades of work by dozens of mathematicians: the classification of finite simple groups. The other is the celebrated Green–Tao theorem, which guarantees the existence of arbitrarily long arithmetic progressions in the primes [**8**, **17**]. Alas, we can choose only one to focus on. However, we do have a few words to say about finite simple groups; see the comments below.

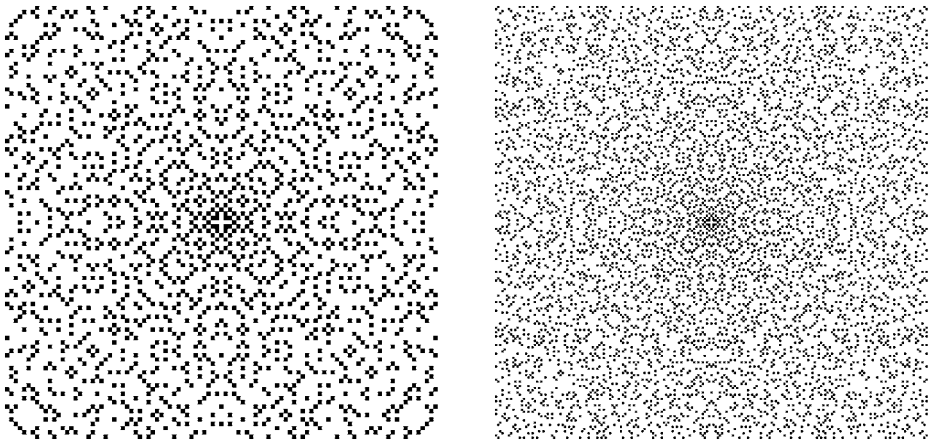
What does the Green–Tao theorem say? It asserts that for any length ℓ, there is an initial prime p and a common difference k so that the length-ℓ arithmetic progression $p, p+k, p+2k, \ldots, p+(\ell-1)k$ consists entirely of primes. Ben Green and Terence Tao proved this amazing result using a "relative" version of Szemerédi's theorem (see the 1975 entry). Szemerédi's theorem tells us that a subset of the natural numbers with positive upper density contains arbitrarily long arithmetic progressions. Unfortunately, the prime numbers have density zero and hence Szemerédi's theorem does not immediately apply. Green and Tao proved a version of Szemerédi's theorem that applies to sets of natural numbers that are pseudorandom in a certain technical sense. The final step of their proof is the construction of a pseudorandom subset of the natural numbers that contains the primes as a relatively dense subset. A recent overview of the theorem and its proof is [**3**].

Can the Green–Tao theorem be used to find arithmetic progressions in the primes? Yes and no. The proof provides numerical bounds that guarantee the existence of such an arithmetic progression in a certain range. However, the numbers produced are so astronomically large that they are well beyond the limit of modern computation. As of mid-2018, the longest known arithmetic progression in the primes has length twenty-six. The first such example,

$$43142746595714191 + 5283234035979900k, \qquad k = 0, 1, 2, \ldots, 25,$$

was discovered in 2010 by Benoît Perichon on a PlayStation 3 equipped with special software produced for the purpose [**11**, **18**].

There are now many generalizations and extensions of the Green–Tao theorem [**7**, **9**, **13**–**15**]. We focus here on one of them that has a particularly nice visual appeal to it [**13**]. A *Gaussian integer* is a number of the form $a + bi$, in which $a, b \in \mathbb{Z}$ and $i^2 = -1$. The set of Gaussian integers forms a ring, denoted $\mathbb{Z}[i]$, under the usual operations inherited from the complex number system. A *Gaussian prime* is



(a) $\rho = 50$ (b) $\rho = 100$

FIGURE 1. Gaussian primes $a + bi$ in the range $|a|, |b| \leq \rho$.

a prime in the ring $\mathbb{Z}[i]$. Thus, $z \in \mathbb{Z}[i]$ is a Gaussian prime if

$$z = xy \text{ with } x, y \in \mathbb{Z}[i] \quad \Longrightarrow \quad x \in \{1, -1, i, -i\} \text{ or } y \in \{1, -1, i, -i\}.$$

For example, 2 is not a Gaussian prime since $2 = (1 + i)(1 - i)$. One can show that a Gaussian integer is prime if and only if it is of the form $\pm p$ or $\pm pi$, in which $p \equiv 3 \pmod 4$ is prime in the usual sense, or if it is of the form $a + bi$, in which $a^2 + b^2$ is prime in the usual sense; see Figure 1. In 2005, Terence Tao showed that given any distinct $v_0, v_1, \ldots, v_{k-1} \in \mathbb{Z}[i]$, then there are infinitely many sets $\{a + rv_0, a + rv_1, \ldots, a + rv_{k-1}\}$, in which $a \in \mathbb{Z}[i]$ and $r \in \mathbb{Z}\backslash\{0\}$, all of whose elements are Gaussian primes.

The Green–Tao theorem, along with many other famous theorems and difficult conjectures, follows from the Bateman–Horn conjecture. See the comments for the 2005 entry for more information about this tantalizing conjecture.

Centennial Problem 2004

Proposed by Steven J. Miller, Williams College.

The Green–Tao theorem implies that for each natural number N, there is an even number $2m$, in which m depends on N, such that there are at least N pairs of primes whose common difference is $2m$. Prove this without appealing to the Green–Tao theorem.

2004: Comments

Four squares in arithmetic progression. The Green–Tao theorem addresses primes in arithmetic progressions. What about perfect squares? The comments to the 1913 entry show how to construct three squares in arithmetic progression. Mathematical folklore credits Fermat with the proof that there does

not exist an arithmetic progression of four perfect squares [**6**], although Leonhard Euler is attributed the observation in 1780 [**4**]. A proof using Fermat's method of descent can be found in [**16**]. The more modern approach to the problem involves elliptic curves. The crux of the matter is that the rational quadruples (a, b, c, d) so that a^2, b^2, c^2, d^2 form an arithmetic progression can be parametrized by the rational points on the elliptic curve

$$y^2 \;=\; (x-1)(x-2)(x+2).$$

One can show that the curve has only eight rational points, all of which give rise to trivial solutions to the original problem. Consequently, there are no rational perfect squares in arithmetic progression. The details can be found in [**4**].

Euclid's theorem revisited. There is a lot that we do not understand about prime numbers. Even Euclid's theorem (see p. 4) still holds some mystery. Let $a_1 = 2$, the first prime. Then $a_1 + 1 = 3$, which is also prime; set $a_2 = 3$. Next, observe $a_1a_2 + 1 = 7$, which is another prime; set $a_3 = 7$. In the next stage, we see that $a_1a_2a_3 + 1 = 43$, another prime, which we denote by a_4. Now things get interesting. Observe that

$$a_1a_2a_3a_4 + 1 \;=\; 1{,}807 \;=\; 13 \cdot 139;$$

set $a_5 = 13$. In general, let a_n be the smallest prime in the factorization of $a_1a_2 \cdots a_{n-1} + 1$ that is not among $a_1, a_2, \ldots, a_{n-1}$. This yields the *Euclid–Mullin sequence* [**5**, **10**]:

> 2, 3, 7, 43, 13, 53, 5, 6221671, 38709183810571, 139, 2801, 11, 17,
> 5471, 52662739, 23003, 30693651606209, 37, 1741, 1313797957,
> 887, 71, 7127, 109, 23, 97, 159227, 643679794963466223081509857,
> 103, 1079990819, 9539, 3143065813, 29, 3847, 89, 19, 577, 223,
> 139703, 457, 9649, 61, 4357,. . . .

Does this sequence contain every prime? Without a major breakthrough in our understanding of prime numbers, this question will likely remain unanswered for many years to some.

Classification of finite simple groups. The year 2004 witnessed the completion of the classification of finite simple groups, a decades-long quest. A group is *simple* if it contains no normal subgroups other than itself and the trivial subgroup (see the 1992 entry for more background). Consequently, a simple group cannot be decomposed further using the quotient group construction. The finite simple groups are the "atoms" from which more complicated finite groups, "molecules" if you will, can be constructed. In contrast to atoms, which come in only a hundred or so varieties, there are infinitely many finite simple groups.

In 1972, Daniel Gorenstein (1923–1992) proposed a sixteen-step program to complete the classification, an odyssey first (unintentionally) begun by Évariste Galois (1811–1832) with his discovery of groups and of two families of finite simple groups. In 2004, Michael Aschbacher (1944–) and Stephen D. Smith published a massive two-volume book, over a thousand pages in total, that handled the classification of "quasithin groups" [**1**, **2**]. This was the only missing piece in the Gorenstein program and it finally completed the classification of finite simple groups.

The classification theorem states that every finite simple group is isomorphic to one of the following:

(a) a cyclic group of prime order,

(b) an alternating group A_n with $n \geq 5$,

(c) a group of Lie type,

(d) one of the 26 sporadic groups.

There is a lot to unpack here and we can only sketch the details. The alternating group A_n is the subgroup of S_n, the group of permutations on n symbols, that consists of all even permutations. The groups A_n are simple if $n \geq 5$ (the simplicity of these groups is closely related to the fact that there is no analogue of the quadratic formula for polynomial equations of degree five and higher; see the 1973 entry).

The technical definition of a "group of Lie type" would take us too far afield, so we content ourselves with some broad strokes. Here "Lie" refers to Sophus Lie (1842–1899), whose name is pronounced "Lee." There are sixteen families of finite simple groups of Lie type, most of which were discovered long ago. Many can be realized as matrix groups over finite fields and several are closely related to exotic Lie algebras. For the sake of illustration, here is one such example. Start with the special linear group $SL_n(\mathbb{F}_q)$ of all $n \times n$ matrices with determinant 1 and entries in the finite field $\mathbb{F}_q$ of q elements. The quotient of $SL_n(\mathbb{F}_q)$ by the subgroup of nonzero multiples of the identity is the projective special linear group $PSL_n(\mathbb{F}_q)$. If $n \geq 2$ and $q \neq 2, 3$, then one can show that $PSL_n(\mathbb{F}_q)$ is a finite simple group of Lie type.

Most interesting are the 26 sporadic groups.[1] These are outliers that do not fit neatly into any classification scheme. The sporadic groups are divided into two broad classes: the pariahs and the happy family. The pariahs are not subquotients of the monster group M (see the 1992 entry); that is, a pariah cannot be obtained as a quotient group of some subgroup of M. These are the six vertices that do not have upward paths toward the monster group M in Figure 2. In contrast, all twenty members of the happy family are subquotients of M. They are divided into three "generations," with the monster group being of the third generation.

No single human in 2004 could comprehend the proof of the classification theorem in its entirety (see the 1976 and 1998 entries for other instances of this phenomenon). It was spread over hundreds of journal articles, written by many dozens of authors, over the course of several decades. Moreover, the final piece of the puzzle was the two-volume book of Aschbacher and Smith, which weighs in at well

[1]There is another group, named after Jacques Tits (1930–), that is occasionally regarded as the "27th sporadic group." However, it is usually considered an unusual group of Lie type.

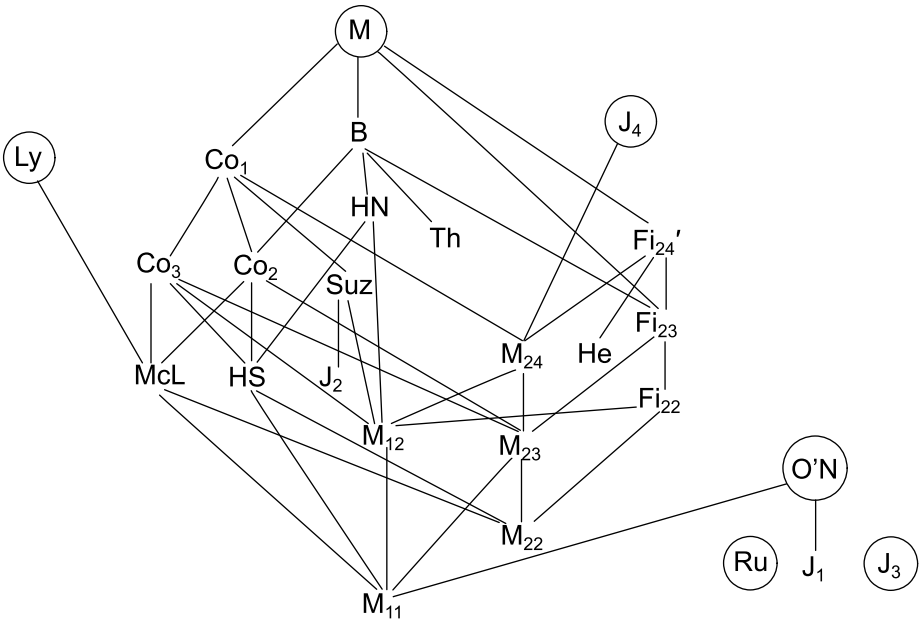


FIGURE 2. Table of sporadic groups and their subquotient relationships (groups that are maximal with respect to this relation are circled). The monster group M contains 20 of the sporadic groups as subquotients. Image by user Drschawrz `https://en.wikipedia.org/wiki/File:SporadicGroups.svg` and used under Creative Commons Attribution-Share Alike 3.0 Unported license.

over 1,000 pages. A massive effort to compile a complete and largely self-contained proof of the classification theorem is well underway:

> In 1981 the monumental project to classify all of the finite simple groups appeared to be nearing its conclusion. Danny Gorenstein had dubbed the project the "Thirty Years' War" dating its inception from an address by Richard Brauer at the International Congress of Mathematicians in 1954. He and Richard Lyons agreed that it would be desirable to write a series of volumes that would contain the complete proof of this Classification Theorem, modulo a short and clearly specified list of background results. As the existing proof was scattered over hundreds of journal articles, some of which cited other articles that were never published, there was a consensus that this was indeed a worthwhile project. [**12**]

The project is expected to be completed in 2023. Perhaps one day soon the entire proof will be verified by computer.

Solution to the problem. Although we could use the prime number theorem to solve the problem, a weaker result due to Chebyshev suffices. He proved that there are constants $A \approx 0.9212$ and $B \approx 1.1055$ so that

$$\frac{Ax}{\log x} \;\le\; \pi(x) \;\le\; \frac{Bx}{\log x}$$

for sufficiently large x, in which $\pi(x)$ denotes the prime-counting function. Suppose that x is even and large enough for Chebyshev's estimate to hold. Then the number of distinct pairs of primes (p, q) with $2 < p < q \leq x$ is

$$\binom{\pi(x)-1}{2} = \frac{(\pi(x)-1)(\pi(x)-2)}{2} > \frac{\pi(x)^2}{3} > \frac{A^2x^2}{3\log^2 x}.$$

Since the number of possible even differences between primes at most x is bounded above by $x/2$, the average number of occurrences of each difference is

$$\frac{\binom{\pi(x)-1}{2}}{x/2} \geq \frac{A^2x^2/(3\log^2 x)}{x/2} = \frac{2A^2x}{3\log^2 x}, \tag{2004.1}$$

which tends to infinity. At least one of these differences occurs at least the average number of times. Given N, let x be an even number that is large enough to ensure that Chebyshev's estimates are valid and that the right-hand side of (2004.1) is larger than N. Then there is a common difference $2m$ for which at least N pairs of primes (p, q) with $p - q = 2m$ exist.

Bibliography

[1] M. Aschbacher and S. D. Smith, *The classification of quasithin groups. I, Structure of strongly quasithin K-groups*, Mathematical Surveys and Monographs, vol. 111, American Mathematical Society, Providence, RI, 2004. MR2097623

[2] M. Aschbacher and S. D. Smith, *The classification of quasithin groups. II, Main theorems: the classification of simple QTKE-groups*, Mathematical Surveys and Monographs, vol. 112, American Mathematical Society, Providence, RI, 2004. MR2097624

[3] D. Conlon, J. Fox, and Y. Zhao, *The Green-Tao theorem: an exposition*, EMS Surv. Math. Sci. **1** (2014), no. 2, 249–282, DOI 10.4171/EMSS/6. https://arxiv.org/abs/1403.2957. MR3285854

[4] K. Conrad, *Arithmetic progressions of four squares*, http://www.math.uconn.edu/~kconrad/blurbs/ugradnumthy/4squarearithprog.pdf.

[5] R. Crandall and C. Pomerance, *Prime numbers:A computational perspective*, Springer-Verlag, New York, 2001. MR1821158

[6] L. E. Dickson, *History of the theory of numbers. Vol. II, Diophantine analysis*, reprinted by AMS, 1992.

[7] J. Fox and Y. Zhao, *A short proof of the multidimensional Szemerédi theorem in the primes*, Amer. J. Math. **137** (2015), no. 4, 1139–1145, DOI 10.1353/ajm.2015.0028. MR3372317

[8] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) **167** (2008), no. 2, 481–547, DOI 10.4007/annals.2008.167.481. http://arxiv.org/abs/math.NT/0404188. MR2415379

[9] B. Green and T. Tao, *Linear equations in primes*, Ann. of Math. (2) **171** (2010), no. 3, 1753–1850, DOI 10.4007/annals.2010.171.1753. MR2680398

[10] On-Line Encyclopedia of Integer Sequences, *A000945 (Euclid-Mullin sequence:* $a(1) = 2$, $a(n+1)$ *is smallest prime factor of* $1 + \prod_{k=1}^{n} a(k)$, https://oeis.org/A000945.

[11] On-Line Encyclopedia of Integer Sequences, *A204189 (Benoît Perichon's 26 primes in arithmetic progression)*, https://oeis.org/A204189.

[12] R. Solomon, *The classification of finite simple groups: a progress report*, Notices Amer. Math. Soc. **65** (2018), no. 6, 646–651. https://www.ams.org/journals/notices/201806/rnoti-p646.pdf. MR3792856

[13] T. Tao, *The Gaussian primes contain arbitrarily shaped constellations*, J. Anal. Math. **99** (2006), 109–176, DOI 10.1007/BF02789444. https://arxiv.org/abs/math/0501314. MR2279549

[14] T. Tao and T. Ziegler, *The primes contain arbitrarily long polynomial progressions*, Acta Math. **201** (2008), no. 2, 213–305, DOI 10.1007/s11511-008-0032-5. MR2461509

[15] T. Tao and T. Ziegler, *A multi-dimensional Szemerédi theorem for the primes via a correspondence principle*, Israel J. Math. **207** (2015), no. 1, 203–228, DOI 10.1007/s11856-015-1157-9. MR3358045
[16] A. van der Poorten, *Fermat's Four Squares Theorem*, `https://arxiv.org/abs/0712.3850v1`.
[17] Wikipedia, *Green–Tao theorem*, `https://en.wikipedia.org/wiki/Green-Tao_theorem`.
[18] Wikipedia, *Primes in arithmetic progression*, `https://en.wikipedia.org/wiki/Primes_in_arithmetic_progression`.