Bonus read: A review of prime problems

5.11. Prime problems

In this bonus section we will discuss various natural sequences that are expected to contain infinitely many primes, highlighting recent progress.

> *Mathematicians have tried in vain to discover some order in the sequence of the prime numbers and we have every reason to believe that there are some mysteries that the human mind shall never penetrate.*
>
> — Leonhard Euler (1740)

Prime values of polynomials in one variable

In section 5.6 we mentioned the twin prime conjecture, that there are infinitely many pairs of primes that differ by 2. What about other pairs? Obviously there can be no more than one pair of primes that differ by an odd integer k (as one of the two integers must be divisible by 2), but when the difference is an even integer k there is no such obstruction. Calculations then suggest that:

For all even integers $2m > 0$ there are infinitely many pairs of primes that differ by $2m$. That is, there are infinitely many prime pairs $p, p + 2m$.

Here we asked for simultaneous prime values of two *monic* linear polynomials x and $x+2m$. What if we select polynomials with different leading coefficients, like x and $2x+1$? Such *prime pairs* come up naturally in Sophie Germain's Theorem 7.11 (of section 7.27 in appendix 7F) and calculations support the guess that there are many (like 3 and 7; 5 and 11; 11 and 23; 23 and 47; ...). We therefore conjecture:

There are infinitely many pairs of primes $p, 2p + 1$.

One can generalize this to other pairs of linear polynomials but we might again have the problem that at least one is even, as with $p, 3p + 1$.

Exercise 5.11.1. Give conditions on integers a, b, c, d with $a, c > 0$, assuming that $(a, b) = (c, d) = 1$, which guarantee that there are infinitely many integers n for which $an + b$ and $cn + d$ are different and both positive and odd. We conjecture, under these conditions that:

There are infinitely many pairs of primes $am + b, cm + d$.

For triples of linear forms and even k-tuplets of linear forms, there are more exceptional cases. For example, the three polynomials $n, n + 2, n + 4$ can all simultaneously take odd values but, for each integer n, one of them is divisible by 3. We call 3 a *fixed prime divisor*, which plays the same role as 2 in the example $n, n + k$ with k odd. In general we need that a given set of linear forms $a_1x + b_1, a_2x + b_2, \dots, a_kx + b_k$ with integer coefficients is *admissible*; that is, there is no fixed prime divisor p. Specifically, for each prime p, there exists an integer n_p for which none of the $a_jn_p + b_j$ is divisible by p, which implies that p does not divide $a_jn + b_j$ for $1 \le j \le k$ for every integer $n \equiv n_p \pmod p$. This leads us to

The prime k-tuplets conjecture. *Let $a_1x + b_1, \dots, a_kx + b_k$ be an admissible set of k linear polynomials with integer coefficients, such that each a_j is positive. Then there are infinitely many positive integers m for which*

$$a_1m + b_1, \dots, a_km + b_k \text{ are all prime.}$$

Exercise 5.11.2.† Assuming the prime k-tuplets conjecture deduce that there are infinitely many pairs of *consecutive* primes $p, p + 100$.

Exercise 5.11.3.† Assuming the prime k-tuplets conjecture deduce that there are infinitely many triples of *consecutive* primes in an arithmetic progression.

Exercise 5.11.4.† Assuming the prime k-tuplets conjecture deduce that there are infinitely many quadruples of *consecutive* primes formed of two pairs of prime twins.

Exercise 5.11.5.† Let $a_{n+1} = 2a_n + 1$ for all $n \ge 0$. Fix an arbitrarily large integer N. Use the prime k-tuplets conjecture to show that we can choose a_0 so that $a_0, a_1, \dots, a_N$ are all primes.

Exercise 5.11.6. Show that the set of linear polynomials $a_1m + 1, a_2m + 1, \dots, a_km + 1$, with each a_j positive, is admissible.

There is more on prime k-tuplets of linear polynomials in appendix 5E.

What about other polynomials? For example, the polynomial $n^2 + 1$ takes prime values $2, 5, 17, 37, 101, \dots$ seemingly on forever, so we conjecture that:

There are infinitely many primes of the form $n^2 + 1$.

The polynomial $x^2 + 2x$ cannot be prime for many integer values since it is reducible (recall Theorem 5.4 and exercise 5.8.14(c)). This is a different reason (from the fixed prime factors above) for a polynomial not to take more than finitely many prime values. These are the only reasons known for a polynomial not to take infinitely many prime values and, if neither of them holds, then we believe that the polynomial does take on infinitely many prime values. More precisely:

Polynomial prime values conjecture. *Let $f_1(x), \dots, f_k(x) \in \mathbb{Z}[x]$, each irreducible, with positive leading coefficients. If $f_1 \cdots f_k$ has no fixed prime divisor, then:*

There are infinitely many integers m for which $f_1(m), \dots, f_k(m)$ are all prime.

To be precise, if $f_1, \dots, f_k$ have "no fixed prime divisor" then we mean that for every prime p there exists an integer n_p such that $f_1(n_p) \cdots f_k(n_p)$ is not divisible

by p. The polynomial prime values conjecture specialized to linear polynomials is the prime k-tuplets conjecture.[20]

Exercise 5.11.7. Prove that the only prime pair p, p^2+2 is $3, 11$.

Exercise 5.11.8. (a) Prove that if $f_1 \cdots f_k$ has no fixed prime divisor, then, for each prime p, there are infinitely many integers n such that $f_1(n) \cdots f_k(n)$ is not divisible by p.
(b)† Show that if $p > \deg(f_1(x) \cdots f_k(x))$ and p does not divide $f_1(x) \cdots f_k(x)$, then n_p exists.
(c) Prove that if $f_j(x) = x + h_j$ for given integers $h_1, \dots, h_k$, then n_p exists for a given prime p if and only if $\#\{\text{distinct } h_j \pmod p\} < p$.

The only case of the polynomial prime values conjecture that has been proved is when $k = 1$ with $f_1(.)$ is linear. The hypothesis ensures that $f(x) = qx + a$ with $q \geq 1$ and $(a, q) = 1$. This is Dirichlet's Theorem (that there are infinitely many primes $\equiv a \pmod q$ whenever $(a, q) = 1$, which we discuss in sections 8.17 of appendix 8D and 13.7).

Distinguishing primes and P_k's from other integers. The Möbius function was introduced in section 4.5, and in Corollary 4.5.1 we saw that the sum

$$\sum_{d|n} \mu(d)$$

is non-zero only if $n = 1$ and so allows us to distinguish the integer 1 from all other positive integers. In section 4.11 of appendix 4B we saw that if the sum

$$\sum_{d|n} \mu(d) \log(n/d)$$

is non-zero, then n has exactly one prime factor and so allows us to distinguish primes and prime powers from all other positive integers. A positive integer is called a "P_k" if it has no more than k distinct prime factors. In the next exercise we will see how an analogous sum allows us to distinguish P_k's.

Exercise 5.11.9.† (a)‡ Let $x_0, \dots, x_m$ be variables. Prove that if $m > k \geq 0$, then

$$\sum_{S \subset \{1,2,\dots,m\}} (-1)^{|S|} \Big(x_0 + \sum_{j \in S} x_j\Big)^k = 0.$$

(b) Deduce that if n has more than k different prime factors, then

$$\sum_{d|n} \mu(d)(\log(n/d))^k = 0.$$

(c)‡ What value does this take when n has exactly k different prime factors?

Exercise 5.11.10. Show that if each prime factor of n is $> n^{1/3}$, then n is either prime or the product of two primes.

[20] This conjecture was first formulated by Andrzej Schinzel in 1958. He called it "*Hypothesis H*" in that paper, and the name has stuck.

Prime values of polynomials in several variables

One can ask for prime values of polynomials in two or more variables, for example, primes of the form m^2+n^2 or the form a^2+b^2+1 or more complicated polynomials of mixed degree like $4a^3+27b^2$. What is known?

The proof of the prime number theorem can be adapted to many situations, for example to primes of the form m^2+n^2 or the form $2u^2+2uv+3v^2$ or indeed the prime values of any irreducible binary quadratic form (which are discussed in chapters 9 and 12) without a fixed prime divisor. The proof for m^2+n^2 uses the fact that $m^2+n^2=(m+in)(m-in)$, the *norm* of $m+in$. One can develop this to prove that any such *norm form* (the appropriate generalization[21] of m^2+n^2 to higher degree) takes on infinitely many prime values as long as it has no fixed prime factor. A norm form is always a degree d polynomial in d variables.

One can then ask for prime values of norm forms in which we fix some of the variables (perhaps to 0). For example, if $m=1$ in m^2+n^2, we are back to the open question about prime values of n^2+1. *However* in 2002 Heath-Brown was able to prove that a^3+2b^3 takes on infinitely prime values and then extended this, with Moroz, to any irreducible cubic form in two variables. In 2018, Maynard proved such a result for a family of norm forms[22] in $3m$ variables of degree $4m$ (or less).

These results on norm forms were all inspired by Friedlander and Iwaniec's 1998 breakthrough in which they took n to be a square in m^2+n^2 (and therefore found prime values of u^2+v^4), following Fouvry and Iwaniec's 1997 paper in which they took n to be prime (and therefore obtained infinitely many prime pairs p, m^2+p^2). This was the first example in which the polynomial in question is *sparse* in that the number of integer values it takes up to x is roughly x^c for some $c<1$. The current record sparsity is $c=\frac{2}{3}$ from the work of Heath-Brown and Moroz. In 2017, Heath-Brown and Xiannan Li went beyond the Fouvry-Iwaniec and Friedlander-Iwaniec results by showing that there are infinitely many prime pairs p, m^2+p^4.

In every case we expect that the proportion of values of the polynomial up to x which are prime is about $c/\log x$, where c is a constant which depends on how often each prime divides values of the polynomial.

Back in 1974, Iwaniec had shown how versatile sieve methods could be by showing that any quadratic polynomial in two variables (which is irreducible and has no fixed prime divisor) takes on infinitely many prime values, for example, m^2+n^2+1. We will see this result put to good use in appendix 12G when tiling a circle with smaller circles.

What about the prime values of more than one polynomial in several variables? We can generalize our conjectures as follows:

Multivariable polynomial prime values conjecture. *Let $f_1(x_1,\dots,x_n),\dots, f_k(x_1,\dots,x_n)\in\mathbb{Z}[x_1,\dots,x_n]$, each of which is irreducible. Suppose that there are infinitely many n-tuplets of integers $m_1,\dots,m_n$ for which each $f_j(m_1,\dots,m_n)$ is positive. If $f_1\cdots f_k$ has no fixed prime divisor, then there are*

Infinitely many n-tuplets of integers $m_1,\dots,m_n$ for which

[21] More precisely the norm of $\sum_i x_i\omega_i$ where the ω_i are a basis for the ring of integers of some number field of degree d and the x_i are the variables.

[22] The norm of $\sum_{i=1}^{3m} x_i\omega^i$ where the field, of degree $4m$, is generated by ω over $\mathbb{Q}$.

$$f_1(m_1, \dots, m_n), \dots, f_k(m_1, \dots, m_n) \text{ are all prime.}$$

In 1939, van der Corput showed that there are infinitely many three-term arithmetic progressions of primes, which can be written as

$$a, a+d, a+2d,$$

three degree-one polynomials in two variables. For a long time, methods seemed inadequate to extend this to length four arithmetic progressions, but this was resolved in 2008 by Green and Tao, who proved that for any fixed integer $k \geq 3$ there are infinitely many prime k-tuplets of the form

$$a, a+d, a+2d, \dots, a+(k-1)d.$$

The methods used were quite new to the search for prime numbers and this has led to widespread interest. In 2012, along with Ziegler, they were able to prove a very general result for linear polynomials, which is as good as one can hope for, given that there has been no progress directly on the prime k-tuplets conjecture:

Until we prove the twin prime conjecture we will be unable to prove the multivariable polynomial prime values conjecture, in full generality, even for linear polynomials, since two of the polynomials might differ by two, for example if $x+3y$ and $x+3y+2$ are in our set. More generally, without progress on the prime k-tuplets conjecture, we must avoid any linear relation between two of our polynomials.

Theorem 5.8 (The Green-Tao-Ziegler Theorem)**.** *Suppose that $f_1(\mathbf{x}), \dots, f_k(\mathbf{x})$ are linear polynomials which satisfy the hypothesis of the multivariable polynomial prime values conjecture. Moreover assume that if $1 \leq i < j \leq k$, there do not exist integers a, b, c, not all zero, for which $af_i + bf_j = c$. Then there are infinitely many $\mathbf{m} \in \mathbb{Z}^n$ for which $f_1(\mathbf{m}), \dots, f_k(\mathbf{m})$ are all prime.*

We will discuss applications of the Green-Tao-Ziegler Theorem in appendix 5E.

It is not difficult to show that there are infinitely many primes of the form $b^2 - 4ac$, the discriminant of an arbitrary quadratic polynomial. However we do not know how to prove that there are infinitely many primes of the form $4a^3 + 27b^2$, the discriminant of the cubic polynomial $x^3 + ax + b$. Proving this would have a significant impact on our understanding of various questions about degree 3 Diophantine equations.

Exercise 5.11.11. Let $g(x) = 1 + \prod_{j=1}^{k}(x-j)$. Prove that there exist integers a and b such that the reducible polynomial $f(x) = (ax+b)g(x)$ is prime when $x = n$ for $1 \leq n \leq k$. Compare this to the result in exercise 5.8.14(c) (with $d = k+1$).

Goldbach's conjecture and variants

Goldbach's 1742 conjecture is the statement that every even integer ≥ 4 can be written as the sum of two primes. It is still an open question though it has now been verified for all even numbers $\leq 4 \times 10^{18}$.

Great problems motivate mathematicians to think of new techniques, which can have great influence on the subject, even if they fail to resolve the original question. For example, although there have been few plausible ideas for proving Goldbach's conjecture, it has motivated some of the development of sieve theory,

and there are some beautiful results on modifications of the original problem. The most famous are:

In 1975 Montgomery and Vaughan showed that if there are any exceptions to Goldbach's conjecture (that is, even integers n that are not the sum of two primes), then there are very few of them.

In 1973 Jingrun Chen showed that every sufficiently large even integer is the sum of a prime and an integer that is the product of at most two primes. Here "sufficiently large" means enormous.

In 1934 I. M. Vinogradov proved that every sufficiently large odd integer is the sum of three primes. The "sufficiently large" has recently been removed: Harald Helfgott, with computational assistance from David Platt, proved that every odd integer > 1 is the sum of at most three primes.

Exercise 5.11.12. Show that the Goldbach conjecture is equivalent to the statement that every integer > 1 is the sum of at most three primes.[23]

Other questions

Before this chapter we asked if there are infinitely many primes of the form $2^p - 1$ (Mersenne primes) or of the form $2^{2^n} + 1$ (Fermat primes). We can ask other questions in this vein, for example prime values of second-order linear recurrences which start $0, 1$ (like the Fibonacci numbers) or their companion sequences (see exercise 3.9.3) or prime values of high-order linear recurrence sequences.

Mersenne primes written in binary look like $111\ldots111$, and so are palindromic. Some people have been interested in primes of the form $\frac{1}{9}(10^n - 1)$ which equal $111\ldots111$ in base 10 and so are palindromic. We are unable to prove there are infinitely many Mersenne primes, so how about the easier question, are there infinitely many palindromic primes when written in binary or in decimal or indeeed in any other base? Also open.

We saw earlier that it is not difficult to show that there are infinitely many primes with the first few digits given. But how about missing digits? Can one find infinitely many primes which have no 7 in their decimal expansion or no 9 or no consecutive digits 123? These questions are all answered in a remarkable recent paper of Maynard [4].

Let M be a given n-by-n matrix. The (i, j)th entry of $M, M^2, \ldots$ can all be described by an nth-order linear recurrence sequence. To see this think of the powers of $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$. We have already asked whether the trace can take infinitely many prime values. A recent question of interest is to take two (or more) such matrices M and N say, and then look at the entries of all "words" created by M and N, for example $M^a N^b M^c \cdots N^z$, and ask whether the entries are infinitely

[23] This was in fact the form in which Goldbach made his conjecture. Goldbach was a friend of Euler, arguably the greatest mathematician of the 18th century, and would often send Euler mathematical questions. In one letter Goldbach asked whether every integer > 1 is the sum of at most three primes, and Euler observed that this is equivalent to showing that every even number ≥ 4 is the sum of two primes. Why then does Goldbach get credit for this conjecture that he did not make? Perhaps because "*Euler is rich, and Goldbach is poor.*"

often prime (see section 9.15 of appendix 9D and appendix 12G for a beautiful example).

Guides to conjectures and the Green-Tao Theorem

[1] David Conlon, Jacob Fox, and Yufei Zhao, *The Green-Tao theorem: An exposition*, EMS Surv. Math. Sci. **1** (2014), 249–282.

[2] G. H. Hardy and J. E. Littlewood, *Some problems of 'Partitio Numerorum'; III: On the expression of a number as a sum of primes*, Acta Math. **44** (1923), 1–70.

[3] Bryna Kra, *The Green-Tao theorem on arithmetic progressions in the primes: An ergodic point of view*, Bull. Amer. Math. Soc. **43** (2006), 3–23.

[4] James Maynard, *Small gaps between primes*, Annals Math. **181** (2015), 383–413.

[5] A. Schinzel and W. Sierpiński, *Sur certaines hypothèses concernant les nombres premiers*, Acta Arith. **4** (1958), 185–208; erratum **5** (1958), 259.

Appendices. The extended version of chapter 5 has the following additional appendices:

Appendix 5B. *An important proof of infinitely many primes.* We give Euler's proof that there are infinitely many primes (which yields that the sum of the reciprocals of the primes diverges) and use this to show that the primes make up a vanishing proportion of the integers. We use this to introduce the Riemann zeta-function, as well as Riemann's program for proving the prime number theorem.

Appendix 5C. *What should be true about primes?* Here we explain Cramér's model for the distribution of primes based on Gauss's thoughts and determine what it predicts about the expected longest gaps between primes.

Appendix 5D. *Working with Riemann's zeta-function.* We further develop Riemann's program for proving the prime number theorem, detailing how the zeros of the Riemann zeta-function relate to the count of primes. We are therefore able to state the Riemann Hypothesis and discuss some attractive reformulations.

Appendix 5E. *Prime patterns: Consequences of the Green-Tao Theorem.* We look for all sorts of prime patterns and at fun questions about primes, for example magic squares of primes like

17	89	71
113	59	5
47	29	101

41	71	103	61
97	79	47	53
37	67	83	89
101	59	43	73

Examples of magic squares of primes.

Appendix 5F. *A panoply of prime proofs* presents several further proofs that there are finitely many primes, one by point-set topology, another using irrationality, and yet another via a counting argument.

Appendix 5G. *Searching for primes and prime formulas.* We look for formulas for primes, including Matijasevic's amazing polynomial in 26 variables, discuss their value, explore Conway's prime-producing machine and patterns in Ulam's spiral.

Appendix 5H. *Dynamical systems and infinitely many primes.* Developing a perspective on Euclid's original proof, we show that there are many different polynomials for which there are infinitely many prime divisors of the iterated values of the polynomial, starting from a non-periodic point.