Small gaps between prime numbers

Yitang Zhang

University of California @ Santa Barbara
Mathematics Department
University of California, Santa Barbara

April 20, 2016

Infinitude of prime numbers

- A prime number is an integer greater than 1 which is only divisible by 1 and itself. For example, 2, 17, 1009 are prime numbers.
- **The fundamental theorem of arithmetic**: Every integer greater than 1 is either a prime number or can be uniquely written as the product of prime numbers.
- **Euclid** (ca. 2300 years ago): There are infinitely many prime numbers.
- **Euclid's proof of the infinitude of prime numbers**

Assume that there are only finitely many prime numbers, say they are p_1, p_2,...,p_n. Then the integer $N := p_1 p_2 ... p_n + 1$ is not divisible by any of p_1, p_2,...,p_n, contradicting that N is either a prime number or a product of prime numbers.

The distribution of prime numbers

The distribution of prime numbers is a major topic in modern number theory. There are many interesting and unsolved problems in this field which look simple and can be explained to beginning students and amateurs.

For example, a pair of prime numbers $\{p, q\}$ is said to be a twin if $q = p + 2$. The twin prime pairs less than 100 are

$$\{3, 5\}, \{5, 7\}, \{11, 13\}, \{17, 19\}, \{29, 31\}, \{41, 43\}, \{59, 61\}, \{71, 73\}.$$

Numerical calculation shows that there are many twin prime pairs. The **Twin Prime Conjecture** asserts that there are infinitely many twin prime pairs.

While many mathematicians have made contributions towards this conjecture and a number of partial results have been obtained, we still do not have a proof.

In particular, until recently, we had not been able to prove that there are infinitely many pairs of prime numbers whose gap is bounded by a fixed number.

Now let me give an indication of how in 2013 I was able to prove:

There are infinitely many pairs of prime numbers $\{p, q\}$ *such that*

$$0 < q - p < 70,000,000, \tag{1}$$

which is a weaker form of the twin prime conjecture.

For each positive integer n, let

$$\varrho(n) = \begin{cases} 1, & \text{if} \quad n \quad \text{is a prime number} \\ 0, & \text{if} \quad n \quad \text{is not a prime number.} \end{cases}$$

For example,

$$\varrho(2) = \varrho(3) = \varrho(101) = \varrho(1009) = \varrho(65537) = 1,$$

and

$$\varrho(1) = \varrho(4) = \varrho(1001) = \varrho(2047) = \varrho(10^{100}) = 0.$$

The reason for introducing the function $\varrho(n)$ is due to the simple fact: For any n, $\varrho(n)\varrho(n+2)$ is 1 exactly when n and $n+2$ form a twin prime pair, and is 0 otherwise.
Thus the Twin Prime Conjecture is to show that $\varrho(n)\varrho(n+2) = 1$ for infinitely many n. In other words, one needs to show that the sum

$$\sum_{n=1}^{N} \varrho(n)\varrho(n+2)$$

tends to infinity as $N \to \infty$.
In practice, products like $\varrho(n)\varrho(n+2)$ are difficult, while sums like $\varrho(n) + \varrho(n+2)$ are easier to handle.

The Twin Prime Conjecture is equivalent to

$$\varrho(n) + \varrho(n+2) = 2 \quad \text{infinitely often,}$$

or even to

$$\varrho(n) + \varrho(n+2) - 1 > 0 \quad \text{infinitely often.}$$

Thus, for any large N, if one can find *non-negative* real numbers c_n and show that

$$\sum_{n=N}^{2N} c_n[\varrho(n) + \varrho(n+2) - 1] > 0, \tag{2}$$

then there exists at least one term in the sum satisfying

$$c_n[\varrho(n) + \varrho(n+2) - 1] > 0.$$

Since $c_n \geq 0$, this implies that

$$\varrho(n) + \varrho(n+2) - 1 > 0,$$

so that n and $n+2$ form a twin prime pair. Since $n \geq N$ and N can be arbitrarily large. This proves the Twin Prime Conjecture.

In modern number theory, solutions of many problems are reduced to finding some numbers c_n such that a certain relation, like (2), holds. Meanwhile, the c_n should satisfy certain constraints , such as the non-negative condition. The construction of the c_n is often technical, and, in many cases, the optimal choice of c_n is unknown. For example, in order that $c_n \geq 0$, one may start with real numbers d_n and take $c_n = d_n^2$.

Now let us come back to the weaker form (1). Following Goldston, Pintz and Yildirim, we choose a set $\{h_1, h_2, ..., h_k\}$ of integers satisfying certain "admissible" conditions. In our paper, we use $k = 3,500,000$ and

$$0 = h_1 < h_2 < h_3 < \cdots < h_k < 70,000,000.$$

To prove (1), it suffices to show that for infinitely many n, the k-tuple

$$n + h_1,\ n + h_2,\ n + h_3, ..., n + h_k$$

contains two or more prime numbers. This occurs precisely when

$$\sum^{k} \varrho(n + h_j) - 1 > 0. \quad (3)$$

Further, it suffices to show that for all large N, there is a $n \geq N$ such that (3)holds. To prove this, we choose certain real numbers c_N, c_{N+1},....,c_{2N} with

$$c_n \geq 0,$$

and show that for all large N,

$$S := \sum_{n=N}^{2N} c_n \left(\sum_{j=1}^{k} \varrho(n + h_j) - 1 \right) > 0. \tag{4}$$

This would then show that, between N and $2N + 70,000,000$, there are two distinct prime numbers whose gap is less than $70,000,000$.

In the prior work of Goldston, Pintz and Yildirim, they found appropriate choices of the coefficients c_n which are of the form

$$c_n = \Big(\sum_{\substack{d<D \\ d|P(n)}} \mu(d)(\log D/d)^g \Big)^2,$$

where g is a certain positive integer (constant),

$$D = x^b, \qquad 0 < b < \frac{1}{2}, \tag{5}$$

$$P(n) = (n+h_1)(n+h_2)....(n+h_k),$$

and $\mu(d)$ is the Möbius function. That is, $\mu(1) = 1$, $\mu(d) = (-1)^r$ if d is the product of r distinct primes, and $\mu(d) = 0$ otherwise.

The evaluation of the sum S in (4) yields a relation of the form

$$S = M + E,$$

where M represents the "main" term and E represents the "error" term. There is an explicit expression for the main term M. On the other hand, however, for the error term E, one can only obtain upper bounds for its absolute value.

With the above notation, the proof of (4) is therefore reduced to showing that

$$M > |E|. \tag{6}$$

The result of Goldston, Pintz and Yildirim was close to (6). In particular, they were able to show that $M > 0$. However, due to an essential obstacle on the estimation of E, they were unable to complete the proof.

This obstacle is related to the distribution of prime numbers in arithmetic progressions. Recall the exponent b appearing in (5). Using the classical estimates, such as the Bombieri-Vinogradov theorem, one can efficiently bound $|E|$ with $b \leq 1/4$ only. However, the proof of (6) requires an efficient upper bound for $|E|$ with some $b > 1/4$. This was achieved in my paper using various tools from modern number theory, and a number of innovations were introduced. In particular, I also used deep results from algebraic geometry.

Since my paper, many mathematicians have improved the result in different ways. In particular, today the constant $70,000,000$ has been reduced to 246.

Thank you!