CONSECUTIVE PRIMES IN ARITHMETIC PROGRESSIONS

ARANYA KARIGHATTAM

1. Introduction

An important result concerning primes in any arithmetic progression is Dirichlet's theorem which was proven by Dirichlet in 1837.

Theorem 1.1 (Dirichlet)**.** *Let a and m be relatively prime integers. Then there are infinitely many primes p such that $p \equiv a \pmod{m}$.*

Chowla conjectured the following stronger statement in 1920.

Conjecture 1.2 (Chowla)**.** *There exists infinitely many integers n such that the consecutive primes p_n and p_{n+1} are congruent to $a \pmod{m}$ where a and m are relatively prime.*

In this paper, we will review the proof of a theorem, proven by Shiu in 1997 [1], which is a generalization of Dirichlet's theorem.

Theorem 1.3 (Shiu)**.** *Let a and m be relatively prime integers. For every positive integer k, there exists a string of k consecutive primes $p_n, \dots, p_{n+k}$ for some positive integer n such that*

$$p_n \equiv p_{n+1} \equiv \dots \equiv p_{n+k} \equiv a \pmod{m}.$$

Example 1.4. Consider the arithmetic progression of integers congruent to $2 \pmod{3}$. This contains

$$2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, \dots$$

We can immediately see that the prime 2 forms a string of length 1. The first string of length 2 is formed by the consecutive primes 23 and 29 which are both congruent to $2 \pmod{3}$. The consecutive primes 47, 53, and 59 are all congruent to $2 \pmod{3}$ and form a string of length 3. The first string of length 4 appears much later, with consecutive primes 251, 257, 263, and 269 that are all congruent to $2 \pmod{3}$. Note that the primes 5, 11, 17, 23, and 29 do not form a string of length 5 since, although they are all congruent to 2 $\pmod{3}$, they are not consecutive.

The proof of Theorem 1.3 is based on Maier's matrix method, described in [2]. Shiu also proved another theorem which is an extension of the first and states that there are infinitely many strings of length k of consecutive primes in an arithmetic progression. We refer the reader to [1] for the proof of this theorem.

Theorem 1.5. *Let a and m be relatively prime integers and k be a positive integer. There exists infinitely many positive integers n such that the primes $p_n, \dots, p_{n+k}$ satisfy*

$$p_n \equiv p_{n+1} \equiv \dots \equiv p_{n+k} \equiv a \pmod{m}.$$

To prove Theorem 1.3, we will prove that we can create a lower bound on the length k of the longest string of consecutive primes less than some large x. It turns out that there is a stronger lower bound whenever a is in either of the following sets:

$$A_+ := \{a : a \equiv 1 \pmod{p} \text{ for all } p|m\}$$

and

$$A_- := \{a : a \equiv -1 \pmod{p} \text{ for all } p|m\},$$

where a and m are relatively prime integers. The notation $A_\pm$ refers to the union of the two sets. We can now restate Theorem 1.3.

Theorem 1.6. *For any relatively prime positive integers a and m, and some large x, there is a string of k primes $p_n, \ldots, p_{n+k}$, for some positive integer n, satisfying*

$$p_n \equiv p_{n+1} \equiv \ldots \equiv p_{n+k} \equiv a \pmod m,$$

where $p_{n+k} < x$ and

$$k \gg \left(\frac{\log\log x \log\log\log\log x}{(\log\log\log x)^2}\right)^{1/\phi(m)}.$$

Further, for each $a \in A_{\pm}$, there exists a string of k primes $p_n, \ldots, p_{n+k}$, for some positive integer n, satisfying

$$p_n \equiv p_{n+1} \equiv \ldots \equiv p_{n+k} \equiv a \pmod m,$$

where $p_{n+k} < x$ and

$$k \gg \left(\frac{\log\log x}{\log\log\log x}\right)^{1/\phi(m)}.$$

2. Background

We will state some important lemmas that we will be using in the proof of Theorem 1.6. First, we need the following definition.

Definition 2.1. For any positive integer y and prime q, define $P(y, q)$ to be

$$P(y, q) := m \prod_{p \leq y, p \neq q} p.$$

Lemma 2.2. *There is a constant C such that for every positive integer m and large x, there is a positive integer y and prime q with $q \gg \log y$ such that*

$$x < P(y, q) \ll x(\log x)^2,$$

and no L-function modulo $P(y, q)$ has a zero in the region

$$1 \geq \mathfrak{Re}(s) > 1 - \frac{C}{\log P(y, q)(\mathfrak{Im}(s) + 1)}. \tag{1}$$

We refer the reader to [1] for the proof of this lemma and the next two, and [3] for more on L-functions.

Lemma 2.3. *Choose a constant C and positive integer m' so that no L-functions modulo m' have a zero in the region*

$$1 \geq \mathfrak{Re}(s) > 1 - \frac{C}{\log m'(\mathfrak{Im}(s) + 1)}. \tag{2}$$

There exists a constant D and constants C_1 and C_2 such that the inequality

$$C_1\left(\frac{x}{\phi(m')\log x}\right) \leq \pi(x; m', a') \leq C_2\left(\frac{x}{\phi(m')\log x}\right)$$

is true for all $x \geq m'^D$ and a' relatively prime to m'.

The proof of this lemma uses the Brun-Titchmarsh inequality (see [4]) and Gallagher's Theorem (stated and proved in [2]).

Lemma 2.4. *Let m be a positive integer. Define $K(x)$ to be the set of positive integers $n \leq x$ such that all prime factors of n are congruent to $1 \pmod m$. As x approaches ∞,*

$$|K(x)| = \left(c_0 + O\left(\frac{1}{\log x}\right)\right)\frac{x}{\log x}(\log x)^{1/\phi(m)},$$

where

$$c_0 := \frac{1}{\Gamma(1/\phi(m))}\lim_{s\to 1}(s-1)^{1/\phi(m)} \prod_{p \equiv 1 \pmod m}\left(1 - \frac{1}{p^s}\right)^{-1}.$$

3. Consecutive primes congruent to $a \pmod m$

Proof of Theorem 1.6. By Lemma 2.2, for positive integers a, m, x and large D, we can choose y and q so that

$$x^{1/D} \leq P(y,q) \ll x^{1/D}(\log x)^2,$$

and no L-function modulo $P(y,q)$ has an exceptional zero. We now define two sets consisting of a subset of the primes less than y satisfying specific conditions. First, if $z < y$ and $t \leq \sqrt{yz}$, for all values of a, let

$$\begin{aligned} R := &\{p \leq y : p \neq q, p \not\equiv 1, a \pmod m\} \\ &\cup \{t \leq p \leq y : p \neq q, p \equiv 1 \pmod m\} \\ &\cup \{p \leq yz/t : p \neq q, p \equiv a \pmod m\}. \end{aligned}$$

Next, for $a \in A_{\pm}$, let

$$R_0 := \{p \leq y : p \neq q, p \not\equiv 1 \pmod m\}.$$

Define the following functions:

$$Q(y) := m \prod_{p \in R} p,$$

and

$$Q_0(y) := m \prod_{p \in R_0} p.$$

The primes in R and R_0 are a subset of the primes less than y, so $Q(y)$ and $Q_0(y)$ divide $P(y,q)$. We also note that $\log P(y,q) < 3 \log Q(y), 3 \log Q_0(y)$. This implies that the regions

$$1 \geq \mathfrak{Re}(s) > 1 - \frac{C}{3 \log Q(y)(\mathfrak{Im}(s) + 1)} \tag{3}$$

and

$$1 \geq \mathfrak{Re}(s) > 1 - \frac{C}{3 \log Q_0(y)(\mathfrak{Im}(s) + 1)} \tag{4}$$

are contained in the region given by (1). If there is an L-function modulo $Q(y)$ or $Q_0(y)$ with a zero in the region given by (3) or (4) respectively, then the corresponding L-function modulo $P(y,q)$ would contain a zero at the same point in the region given by (1). Since there is no L-function modulo $P(y,q)$ with a zero in this region, there must not be any L-functions modulo $Q(y)$ or $Q_0(y)$ with a zero in the regions given by (3) and (4). Note that $Q(y)$ and $Q_0(y)$ satisfy

$$x^{1/2D} < Q(y), Q_0(y) < x^{1/D}.$$

We now use Maier's matrix method to find strings of consecutive primes. At this point, we will split the proof into two parts, the first (Part 1) for all values of a and the second (Part 2) for when $a \in A_{\pm}$.

(Part 1): Let a and m be any relatively prime positive integers. Let I be an interval with $I = (0, yz]$. Define M to be the set of elements in the array (or matrix)

$$\begin{array}{cccc} 1 + Q(y) & 2 + Q(y) & \ldots & yz + Q(y) \\ 1 + 2Q(y) & 2 + 2Q(y) & \ldots & yz + 2Q(y) \\ \vdots & \vdots & \ddots & \vdots \\ 1 + Q(y)^D & 2 + Q(y)^D & \ldots & yz + Q(y)^D \end{array}$$

We will refer to the rows of this array as intervals. We are looking for strings of consecutive primes congruent to $a \pmod m$ which are in the set

$$P_1 := \{p \in M : p \equiv a \pmod m \text{ where } p \text{ is prime}\}.$$

Note that all other primes are in the set

$$P_2 := \{p \in M : p \not\equiv a \pmod m \text{ where } p \text{ is prime}\}.$$

We will find a lower bound for the size of P_1 and an upper bound for the size of P_2, since the larger $|P_1|$ is and the smaller $|P_2|$ is, the more likely we are to find longer strings. To do this, consider the sets S and T, where

$$S := \{i \in I : \gcd(i, Q(y)) = 1, i \equiv a \pmod m\},$$

$$T := \{i \in I : \gcd(i, Q(y)) = 1, i \not\equiv a \pmod m\}.$$

By Lemma 2.4, the lower bound for $|S|$ is

$$|S| \gg \frac{yz(\log t)^{1/\phi(m)}}{\log y},$$

and the upper bound for $|T|$ is

$$|T| \ll \frac{yz(\log z)^{1/\phi(m)}}{\log y}.$$

By Lemma 2.3, we can bound $|P_1|$ below by

$$|P_1| \gg |S| \frac{x}{\phi(Q(y)) \log x},$$

and $|P_2|$ above by

$$|P_2| \ll |T| \frac{x}{\phi(Q(y)) \log x},$$

where $x \geq Q(y)^D$. Define M' to be the union of the intervals in the array that contain primes in P_1. There are two possible cases:

- Case I: Interval I_0 exists in M' in which the number of primes in P_1 is at least $|P_1|/2|P_2|$ times the number of primes in P_2. So

$$|I_0 \cap P_1| \geq \frac{1}{2} \frac{|P_1|}{|P_2|} |I_0 \cap P_2|.$$

- Case II: At least $1/2$ the primes in P_1 are outside M'. So

$$|(M \setminus M') \cap P_1| \geq \frac{1}{2}|P_1|.$$

We can visualize this using Figure 1. The outer gray square consists of the elements of M and the blue rectangles inside it consist of the elements of M' (the elements in all the intervals that contain elements of P_2). The elements of interval I_0 are in the yellow rectangle.

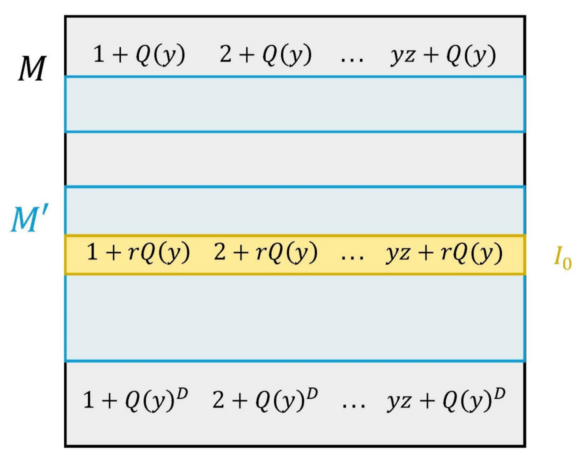

Figure 1. Diagram of the elements of the array and the interval I_0.

According to Case I, in the yellow rectangle, the number of primes in P_1 is much more than the number of primes in P_2, by a factor of $|P_1|/2|P_2|$. From Case II, at least $1/2$ the primes in P_1 are in the gray sections of the outer square. If both cases are not true, we get the following contradiction:

$$\begin{aligned}|P_1| &= |P_1 \cap (M \setminus M')| + |P_1 \cap M'| \\ &= |P_1 \cap (M \setminus M')| + \sum_{I_0 \subseteq M'} (|P_1 \cap I_0|) \\ &< \frac{|P_1|}{2} + \left(\frac{|P_1|}{2|P_2|}\right) \sum_{I_0 \subseteq M'} (|P_2 \cap I_0|) \\ &= \left(\frac{|P_1||P_2|}{2|P_2|}\right) + \frac{|P_1|}{2} \\ &= |P_1|.\end{aligned}$$

Therefore either Case I or II must be true. For Case I, by the Pigeonhole Principle, since the number of primes in I_0 that are in P_1 to the number of primes in I_0 that are in P_2 is $|P_1|/|P_2|$, the interval I_0 must contain a string of length k of primes in P_1 so that $k \gg |P_1|/|P_2|$. Similarly, for Case II, there are up to $x/Q(y)$ intervals outside M' and one of these intervals must contain a string of length k so that $k \gg Q(y)|P_1|/x$. Since

$$\frac{Q(y)}{\phi(Q(y))} = \frac{m}{\phi(m)} \prod_{p \in R} \frac{p}{p-1} = \frac{m}{\phi(m)} \prod_{p \in R} \left(1 - \frac{1}{p}\right)^{-1},$$

we find that $Q(y)/\phi(Q(y)) \gg (\log t)^{1/\phi(m)}/\log y$ using a generalization of Merten's Theorem. Since $\log x \ll Q(y) \ll y$, we get

$$\frac{|P_1|Q(y)}{x} \gg \frac{yz}{\log x} \gg z.$$

We can choose a lower bound for the length k of a string of consecutive primes congruent to $a \pmod m$ in the set M that satisfies both cases, so

$$k \gg \min\left(\frac{|P_1|}{|P_2|}, z\right). \tag{5}$$

Using the lower bounds for $|S|$ and $|P_1|$, and upper bounds for $|T|$ and $|P_2|$, we see that

$$|P_1| \gg |S| \frac{x}{\phi(Q(y)) \log x} \gg \left(\frac{yz(\log t)^{1/\phi(m)}}{\log y}\right)\left(\frac{x}{\phi(Q(y)) \log x}\right),$$

and

$$|P_2| \ll |T| \frac{x}{\phi(Q(y)) \log x} \ll \left(\frac{yz(\log z)^{1/\phi(m)}}{\log y}\right)\left(\frac{x}{\phi(Q(y)) \log x}\right).$$

Therefore (5) is equivalent to

$$k \gg \min\left(\frac{yz(\log t)^{1/\phi(m)}}{yz(\log z)^{1/\phi(m)}}, z\right) = \min\left(\left(\frac{\log t}{\log z}\right)^{1\phi(m)}, z\right).$$

Since $z < y$, $t \leq \sqrt{yz}$, and $\log x < y$, let $z = \log\log x$. Then we get

$$k \gg \min\left(\left(\frac{\log\log x \log\log\log\log x}{(\log\log\log x)^2}\right)^{1/\phi(m)}, \log\log x\right) = \left(\frac{\log\log x \log\log\log\log x}{(\log\log\log x)^2}\right)^{1/\phi(m)}$$

as desired.

(Part 2): Let $a \in A_{\pm}$. Let J be defined as

$$J := \begin{cases} (r_1, r_1 + yz] & a \in A_+, \\ [r_2 - yz, r_2) & a \in A_-, \end{cases}$$

where $r_1 \equiv 0 \pmod p$ and $r_1 \equiv a-1 \pmod m$ for all $a \in A_+$, and $r_2 \equiv 0 \pmod p$ and $r_2 \equiv a+1 \pmod m$ for all $a \in A_-$. Define M_0 to be the set of elements in the array

$$\begin{array}{cccc} (c+1)+Q_0(y) & (c+2)+Q_0(y) & \dots & (c+yz)+Q_0(y) \\ (c+1)+2Q_0(y) & (c+2)+2Q_0(y) & \dots & (c+yz)+2Q_0(y) \\ \vdots & \vdots & \ddots & \vdots \\ (c+1)+Q_0(y)^D & (c+2)+Q_0(y)^D & \dots & (c+yz)+Q_0(y)^D \end{array}$$

where $c = r_1$ for $a \in A_+$ and $c = r_2 - yz - 1$ for $a \in A_-$.

Define

$$P_1 := \{p \in M_0 : p \equiv a \pmod m \text{ where } p \text{ is prime}\},$$

and

$$P_2 := \{p \in M_0 : p \not\equiv a \pmod m \text{ where } p \text{ is prime}\}.$$

As in Part 1, we will find a lower bound for the size of P_1 and an upper bound for the size of P_2. Define the sets S and T to be

$$S := \{j \in J : \gcd(j, Q(y)) = 1, j \equiv a \pmod m\},$$

$$T := \{j \in J : \gcd(j, Q(y)) = 1, j \not\equiv a \pmod m\}.$$

By Lemma 2.4, the lower bound for $|S|$ is

$$|S| \gg \frac{yz(\log y)^{1/\phi(m)}}{\log y},$$

and

$$|T| \ll \frac{yz(\log z)^{1/\phi(m)}}{\log y}.$$

By Lemma 2.3, we can bound $|P_1|$ below by

$$|P_1| \gg |S| \frac{x}{\phi(Q_0(y)) \log x},$$

and $|P_2|$ above by

$$|P_2| \ll |T| \frac{x}{\phi(Q_0(y)) \log x},$$

where $x \geq Q_0(y)^D$. Define M_0' to be the union of the intervals in the array that contain primes in P_1. We have the same two possible cases from Part 1:

- Case I: Interval I_0 exists in M_0' in which the number of primes in P_1 is at least $|P_1|/2|P_2|$ times the number of primes in P_2. So

$$|I_0 \cap P_1| \geq \frac{1}{2} \frac{|P_1|}{|P_2|} |I_0 \cap P_2|.$$

- Case II: At least 1/2 the primes in P_1 are outside M_0'. So

$$|(M_0 \setminus M_0') \cap P_1| \geq \frac{1}{2} |P_1|.$$

We know that either Case I or II must be true. For Case I, the interval I_0 must contain a string of length k of primes in P_1 so that $k \gg |P_1|/|P_2|$. Similarly, for Case II, there are up to $x/Q_0(y)$ intervals outside M_0' and one of these intervals must contain a string of length k so that $k \gg Q_0(y)|P_1|/x$. Since

$$\frac{Q_0(y)}{\phi(Q_0(y))} = \frac{m}{\phi(m)} \prod_{p \in R_0} \frac{p}{p-1} = \frac{m}{\phi(m)} \prod_{p \in R_0} \left(1 - \frac{1}{p}\right)^{-1},$$

we find that $Q_0(y)/\phi(Q_0(y)) \gg (\log y)^{1/\phi(m)}/\log y$ using a generalization of Merten's Theorem. Since $\log x \ll Q_0(y) \ll y$, we get

$$\frac{|P_1| Q_0(y)}{x} \gg \frac{yz}{\log x} \gg z.$$

As we did in Part 1, we can choose a lower bound for the length k of a string of consecutive primes congruent to $a \pmod m$ in the set M_0 that satisfies both cases, so

$$k \gg \min\left(\frac{|P_1|}{|P_2|}, z\right). \tag{6}$$

Using the lower bounds for $|S|$ and $|P_1|$, and upper bounds for $|T|$ and $|P_2|$, we see that

$$|P_1| \gg |S|\frac{x}{\phi(Q_0(y))\log x} \gg \left(\frac{yz(\log y)^{1/\phi(m)}}{\log y}\right)\left(\frac{x}{\phi(Q_0(y))\log x}\right),$$

and

$$|P_2| \ll |T|\frac{x}{\phi(Q_0(y))\log x} \ll \left(\frac{yz(\log z)^{1/\phi(m)}}{\log y}\right)\left(\frac{x}{\phi(Q_0(y))\log x}\right).$$

Therefore (6) is equivalent to

$$k \gg \min\left(\frac{yz(\log y)^{1/\phi(m)}}{yz(\log z)^{1/\phi(m)}}, z\right) = \min\left(\left(\frac{\log y}{\log z}\right)^{1\phi(m)}, z\right).$$

Since $z < y$ and $\log x < y$, let $z = \log\log x$. Then we get

$$k \gg \min\left(\left(\frac{\log\log x}{\log\log\log x}\right)^{1/\phi(m)}, \log\log x\right) = \left(\frac{\log\log x}{\log\log\log x}\right)^{1/\phi(m)}$$

as desired.

□

References

[1] Shiu, Daniel KL. "Strings of congruent primes." Journal of the London Mathematical Society 61, no. 2 (2000): 359-373.
[2] Maier, Helmut. "Chains of large gaps between consecutive primes." Advances in Mathematics 39, no. 3 (1981): 257-269.
[3] Iwaniec, Henryk, and Emmanuel Kowalski. Analytic Number Theory. Vol. 53. American Mathematical Soc., 2021.
[4] Halberstam, Heini, and Hans-Egon Richert. Sieve methods. Courier Corporation, 2011.