ON STRINGS OF CONGRUENT PRIMES

ETHAN YANG

1. Introduction

This expository account serves as an overview of the work surrounding Shiu's result on strings of congruent primes. [KLS00] In 1920, Chowla conjectured that there were infinitely many pairs of consecutive primes such that they were congruent to a modulo q, for any relatively prime a and q. In particular, let $p_1 = 2, p_2 = 3, p_3 = 5 \ldots$ be the sequence of primes. Then Chowla's conjecture says that for any relatively prime q, a, there are infinitely many pairs of consecutive primes p_n and p_{n+1} such that

$$p_n \equiv p_{n+1} \equiv a \pmod{q}.$$

Shiu proved in 2000 a stronger version of Chowla's conjecture, that for any $k \in \mathbb{N}$ there exists a string of k congruent primes such that

$$p_{n+1} \equiv p_{n+2} \equiv \cdots \equiv p_{n+k} \equiv a \pmod{q}.$$

He gave a bound on the size of the string versus the size of the largest prime. In particular, he split his argument into two cases of choices for a. We define

$$A_+ = \{a : \forall p \mid q, a \equiv 1 \mod p\}$$

and

$$A_- = \{a : \forall p \mid q, a \equiv -1 \mod p\}.$$

and $A_\pm = A_+ \cup A_-$. Shiu found that longer strings in terms of the largest prime could be found when considering residues from $A_\pm$.

The main theorem we will prove is the following:

Theorem 1.1. (1) *For each q and $a \in A_\pm$ and large x, there exists a string of primes*

$$p_{n+1} \equiv p_{n+2} \equiv \cdots \equiv p_{n+k} \equiv a \pmod{q},$$

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

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

(2) *For each q and a with $(q, a) = 1$ and large x, there exists a string of primes*

$$p_{n+1} \equiv p_{n+2} \equiv \cdots \equiv p_{n+k} \equiv a \pmod{q},$$

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

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

Further work on similar results to this theorem by Freiberg in 2010, where he found that that consecutive primes can be made close together as well [Fre11].

Theorem 1.2. *Let $q \geq 3$ and a be integers with $(q, a) = 1$, and fix any $\varepsilon > 0$. There exist infinitely many pairs of consecutive primes p_r, p_{r+1} such that $p_r \equiv p_{r+1} \equiv a \bmod q$ and $p_{r+1} - p_r < \varepsilon \log p_r$.*

From state of the art work on the theory of small gaps between primes developed by Maynard, Tao, Zhang, and others in [May15], Freiberg was further able to show in 2015 a stronger version of Shiu's result [Fre13].

Theorem 1.3. *Let $p_1 = 2 < p_2 = 3 < \cdots$ be the sequence of all primes. Let $q \geq 3$ and a be a coprime pair of integers, and let $m \geq 2$ be an integer. There exists a constant $B = B(q, a, m)$, depending only on q, a and m, such that the following holds. There exist infinitely many n such that*

$$p_{n+1} \equiv p_{n+2} \equiv \cdots \equiv p_{n+m} \equiv a \bmod q \quad \text{and} \quad p_{n+m} - p_{n+1} \leq B.$$

Although we won't prove these two theorems, the reader can read the proofs of them.

2. Outline

The rest of the paper focuses on proving Theorem 1.1. Section 3 defines useful notation and states necessary lemmas without proof. The proof of the theorem defines $Q(y)$ as a product of q and a subset of primes less than y, where y is carefully chosen such that the L-functions modulo $Q(y)$ have no Siegel zeros. This is done so that the distribution of primes in residue classes modulo $Q(y)$ are all the same and what we expect it to be.

After creating intervals such that they are dense in primes equivalent to $a \mod q$ and sparse in other primes, a matrix M is created similar to in Maier's proof of the existance of chains of large gaps between consecutive primes [Mai85]. The matrix has rows of consecutive integers, and the columns are arithmetic progressions with common difference $Q(y)$.

We prove that most of the primes in M are congruent to $a \mod q$, exceeding the other primes by a constant factor. After considering two cases, this implies that there is a string of primes that are congruent to $a \mod q$ in one of the columns of the matrix. The length of the string is estimated as well.

3. Background

Throughout the proof, we will use the following function

$$P(y, p_0) = q \prod_{\substack{p \leq y \\ p \neq p_0}} p,$$

where q is as in the statement of Theorem 1.1.

We provide lemmas used in the proof that will not be proven here, but can be found in the original paper by Shiu.

Lemma 3.1. *There exists a fixed constant C such that for all $q \in \mathbb{N}$ and large X there exists y and prime $p_0 \gg \log y$ such that none of the L-functions modulo $P(y, p_0)$ has a zero in the region*

$$1 \geq \Re(s) \geq 1 - \frac{c}{\log(P(y, p_0)(|\Im(s)| + 1))}$$

and

$$X < P(y, p_0) \ll X(\log X)^2.$$

Lemma 3.2. *Let C be a constant and let q' be a natural number such that the L-functions induced by characters mod q' have no zeros in the region*

$$1 \geq \Re(s) \geq 1 - \frac{c}{\log(q'(|\Im(s)|+1))}.$$

Then there exists a constant D depending on at most C such that the estimates

$$\frac{x}{\phi(q')\log x} \ll \pi(x; q', a') \ll \frac{x}{\phi(q')\log x}$$

hold uniformly for $(q', a') = 1$ and $x \geq q'^D$, where $\pi(x; q', a')$ counts the number of primes less than or equal to x congruent to $a' \mod q'$.

Lemma 3.3. *Let q be a natural number and $\mathscr{S}(x)$ denote the set of positive integers $n \leq x$ which only have primes congruent to $1 \mod q$ in its prime factorization. Then as $x \to \infty$, we have*

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

where c_0 is a constant depending at at most q.

Before stating the next lemma, we must define what it means to be a smooth number.

Definition 3.4. For a positive real number y and positive integer n, n is said to be *y-smooth* if every prime factor of n is $\leq y$. Let $\Psi(x, y)$ be the number of y-smooth numbers $n \leq x$.

The next lemma attempts to estimate $\Psi(x, y)$ as both x and y go to infinity. More precisely, we have the following inequality due to de Bruijin:

Lemma 3.5. *For $y \leq x$ and y approaching infinity with x, we have*

$$\Psi(x, y) \leq x(\log y)^2 \exp(-u \log u - u \log\log u + O(u)),$$

where $u = \log x / \log y$.

Armed with these lemmas, we are ready to prove Shiu's theorem.

4. Main Result

We now proceed to prove Theorem 1.1. For a given q, a, x from the statement of the theorem and sufficiently large D, Lemma 3.1 gives us a y and prime p_0 such that

$$x^{1/D} < P(y, p_0) \ll x^{1/D} (\log x)^2.$$

and such that there is no L-function that has a zero in the region described in the lemma. We define the product $Q(y) = q \prod_{p \in \mathscr{P}_a} p$, where the product is over primes in the set $\mathscr{P}_a$, which is defined next. We further define $z \leq y$ and $t \leq (yz)^{1/2}$ and define a set of primes less than y, depending on whether a is in $A_\pm$.

$$\mathscr{P}_a = \begin{cases} \{p \leq y : p \neq p_0, p \not\equiv 1 \mod q\} & \text{for } a \in A_\pm, \\ \{p \leq y : p \neq p_0, p \not\equiv 1, a \mod q\} & \\ \cup \{t \leq p \leq y : p \neq p_0, p \equiv 1 \mod q\} & \\ \cup \{p \leq yz/t : p \neq p_0, p \equiv a \mod q\} & \text{otherwise.} \end{cases}$$

For either of these cases, since $\mathscr{P}_a$ only excludes primes that are less than y, we have that $Q(y) \mid P(y, p_0)$ and that $\log P \leq 3 \log Q$. As a result, there are no L-functions modulo $Q(y)$ such that

$$1 \geq \Re(s) \geq 1 - \frac{c}{3 \log(Q(y)(|\Im(s)| + 1))},$$

because an L-function modulo $Q(y)$ that has a zero in that region would imply a zero in the same region for an L-function modulo $P(y, p_0)$ and thus contradict the assumption made on choice of y and p_0 from Lemma 3.1.

An interval I of length yz is defined in various cases of the residue class of a

$$I = \begin{cases} (m, m + yz] & \text{for } a \in A_+ \\ [n - yz, n) & \text{for } a \in A_- \\ (0, yz] & \text{otherwise,} \end{cases}$$

where m and n satisfy

$$m \equiv n \equiv \begin{cases} 0 & \mod p \text{ for } pq \mid Q \\ a - 1 & \mod q\ . \end{cases}$$

We now construct a matrix M that has dimensions of $Q(y)^{D-1}$ rows and yz columns of integers

$$M = \bigcup_{k=1}^{Q(y)^{D-1}} \bigcup_{i \in I} (i + kQ(y)).$$

The columns of the matrix M are an arithmetic progression with common difference $Q(y)$, so we are trying to find a column that contains our string of primes. We also define the sets

$$S = \{i \in I : (i, Q) = 1, i \equiv a \mod q\},$$
$$T = \{i \in I : (i, Q) = 1, i \not\equiv a \mod q\},$$
$$P_1 = \{p \in M : p \text{ prime}, p \equiv a \mod q\},$$
$$P_2 = \{p \in M : p \text{ prime}, p \not\equiv a \mod q\}.$$

Our goal is to estimate $|P_1|$ and $|P_2|$, showing that there are more primes in P_1. We do this by first estimating $|S|$ and $|T|$. We first estimate the two sets for the case $a \in A_\pm$.

In particular, when $a \in A_\pm$, we have that

$$S = |\{j \in (0, yz] : (j, Q) = 1, j \equiv 1 \mod q\}|,$$
$$T = |\{j \in (0, yz] : (j, Q) = 1, j \not\equiv 1 \mod q\}|.$$

This can be seen with a bijection from S to the interval $(0, yz]$. If $a \in A_+$, then $i \in I$ and $i \equiv a \mod q$ if and only if $i - m \equiv 1 \mod q$. Furthermore, $(i, Q) = 1$ if and only if $(i - m, Q) = 1$ since $m \mid Q$ which completes the bijection for the case $a \in A_+$. In the other case where $a \in A_-$, we can take $n - i$ instead of $i - m$ and the same arguments follow.

We know that if $n \in (0, yz]$ and only has primes $p \equiv 1 \mod q$ in its prime factorization, then $n \equiv 1 \mod q$ and $(n, Q) = 1$ by our construction of $\mathscr{P}$. This implies that $n \in S$ when $n \in \mathscr{S}(yz)$ and thus we have the following bound from Lemma 3.3.

$$|S| \geq |\mathscr{S}(yz)| \gg \frac{yz(\log y)^{1/\phi(q)}}{\log y}.$$

For every $j \not\equiv 1 \mod q$, there exists a prime $p \mid j$ such that $p \not\equiv 1 \mod q$. Elements of T are estimated by estimating the number of elements of the form pn where $p \not\equiv 1 \mod q$ and $n \in \mathscr{S}(z)$ and multiples of p_0 in $(0, yz]$. There are $O(yz/\log y)$ such multiples which is less than the elements of the first type. We estimate those such elements by splitting the interval $(y, yz]$ into $O(\log z)$ intervals of length $2^l y$.

$$\begin{aligned}|T| &\ll \sum_{l \ll \log z} \sum_{2^{l-1}y < p \leq 2^l y} |\mathscr{S}(z/2^l)| \\ &\ll \sum_{l \ll \log z} \frac{2^{l-1}y}{\log y} \frac{z(\log z)^{1/\phi(q)}}{2^l \log z} \\ &\ll \frac{yz(\log z)^{1/\phi(q)}}{\log y}.\end{aligned}$$

When $a \notin A_{\pm}$, we can employ a similar strategy as the above splitting argument to estimate $|S|$. From our construction of $\mathscr{P}$, the elements of S are of the form pn such that $p > yz/t, p \equiv a \mod q, n \in \mathscr{S}(t)$ in the interval $(0, yz]$. Using the splitting argument, replacing z with t, we have the following bound.

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

Elements of T are now of three possible types. They are composed of multiples of p_0, special multiples of a prime larger than y that we estimated using our splitting argument, or only have prime factors less than t and congruent to $1 \mod q$. We already estimated the first two types, and the third type is estimated using de Bruijin's bound for smooth numbers in Lemma 3.5. We take

$$t = \exp\left(\theta \frac{\log y \log\log\log y}{\log\log y}\right).$$

From Lemma 3.5, we have the inequality

$$\Psi(yz, t) \leq yz(\log t)^2 \exp(-\theta^{-1} \log\log y + o(\log\log y)) \ll \frac{yz}{\log y},$$

when θ is sufficiently small ($\theta = 1/4$ works). The dominating term is thus still the one from the splitting argument, so in both cases of Theorem 1.1, we have the estimate

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

Every element of S and T is the first term in an arithmetic progression $\mod Q(y)$. We now estimate $|P_1|$ and $|P_2|$ by using Lemma 3.2. We can choose D such that

$$|P_1| \gg |S| \frac{x}{\phi(Q)\log x}, |P_2| \ll |T| \frac{x}{\phi(Q)\log x}$$

for sufficiently large x such that $x \geq Q^d$. There are two cases that we argue one of which happens. Let M' be the subset of columns of M which contain a prime in P_2. The first is that there exists an interval in M' where the primes belonging to P_1 exceed those belonging to P_2 by a factor of $|P_1|/2|P_2$ or the number of primes in $M \setminus M'$ is at least $|P_1|/2$. Formally, there either exists

$$I_0 \in M' : |I_0 \cap P_1| \geq \frac{1}{2} \frac{|P_1|}{|P_2|} |I_0 \cap P_2|$$

or

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

One of these two cases must be true, as if we assume by contradiction that both of them are false, then

$$\begin{aligned}
|P_1| &= |P_1 \cap M'| + |P_1 \cap (M \setminus M')| \\
&= \sum_{I \in M'} |P_1 \cap I| + |P_1 \cap (M \setminus M')| \\
&< \frac{1}{2}\frac{|P_1|}{|P_2|} \sum_{I \in M'} |I \cap P_2| + \frac{1}{2}|P_1| \\
&= \frac{1}{2}\frac{|P_1|}{|P_2|}|P_2| + \frac{1}{2}|P_1| = |P_1|.
\end{aligned}$$

Since this leaves us with $|P_1| < |P_1|$, we have reached a contradiction and thus at least one of our two cases must arise.

If the first case is true, then our interval I_0 must contain a string of primes of length k where $k \gg |P_1|/|P_2|$. If the second case is true, we first note that there can be at most x/Q intervals in $M \setminus M'$ and thus one of them must contain a string of primes of length k where $k \gg Q|P_1|/x$. From our above bound for $|P_1|$,

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

Substituting our definition of Q,

$$\frac{Q}{\phi(Q)} = \frac{q}{\phi(q)} \prod_{p \in \mathscr{P}} \frac{p}{p-1} = \frac{q}{\phi(q)} \prod_{p \in \mathscr{P}} \left(1 - \frac{1}{p}\right)^{-1}.$$

By a generalization of Merten's theorem, we have $Q/\phi(Q) \gg (\log y)^{1/\phi(q)}/\log y$ if $a \in A_{\pm}$ and $Q/\phi(Q) \gg (\log t)^{1/\phi(q)}/\log y$ otherwise. Then

$$\frac{Q|P_1|}{x} \gg \frac{yz}{\log x} \gg z$$

because $\log x \ll Q \ll y$.

Combining the bounds on k for the two cases, there exists a column in m with a string of length k where

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

Substituting our bounds for $|P_1|, |P_2|, |S|$, and $|T|$, we end up with the following when $a \in A_{\pm}$:

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

Otherwise when $a \notin A_{\pm}$,

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

Setting $z = \log\log x$ proves Theorem 1.1.

Shiu also provided an estimate for the number of such strings in his paper through Theorem 2, which we will not prove here. Similar to Theorem 1.1, Shiu split his theorem into two cases on the residue class of a. In particular, when $a \in A_{\pm}$, there are more strings.

Theorem 4.1. *Define*

$$\varepsilon_1(x) = C(q)k \left(\frac{\log\log\log x}{\log\log x} \right)^{1/\phi(q)}$$

and

$$\varepsilon_2(x) = C'(q)k \left(\frac{(\log\log\log x)^2}{\log\log x \log\log\log\log x} \right)^{1/\phi(q)}$$

(1) *Given $q, k, a \in A_{\pm}$, the number B of strings of primes of the form*

$$p_{n+1} \equiv p_{n+2} \equiv \cdots \equiv p_{n+k} \equiv a \pmod{q},$$

where $p_{n+k} < x$, has the asymptotic bound

$$B \gg x^{1-\varepsilon_1(x)}.$$

(2) *Given q, k, a, where q and a are relatively prime, the number B of strings of primes of the form*

$$p_{n+1} \equiv p_{n+2} \equiv \cdots \equiv p_{n+k} \equiv a \pmod{q},$$

where $p_{n+k} < x$, has the asymptotic bound

$$B \gg x^{1-\varepsilon_2(x)}.$$

References

[Fre11] Tristan Freiberg. Strings of congruent primes in short intervals. *Journal of the London Mathematical Society*, 84(2):344–364, 07 2011.

[Fre13] Tristan Freiberg. A note on the theorem of maynard and tao. 77, 11 2013.

[KLS00] D K. L. SHIU . Strings of congruent primes. *Journal of the London Mathematical Society*, 61:359 – 373, 04 2000.

[Mai85] Helmut Maier. Primes in short intervals. *Michigan Math. J.*, 32(2):221–225, 1985.

[May15] James Maynard. Small gaps between primes. *Annals of Mathematics*, 181(1):383–413, 2015.