

BOUNDING THE PRIME GAPS

VISHAL ARUL

ABSTRACT. The aim of this expository article is to explain recent ideas in sieve theory which have been applied to prove results about prime gaps. We start by discussing the sieve of Eratosthenes, Brun's pure sieve, and Selberg's sieve. Next we move onto the GPY sieve and then explain Maynard's refinement [1] which allowed him to prove that there exist prime m -tuples in bounded intervals; that is, $\liminf_n (p_{n+m} - p_n) < \infty$ for any integer m . Afterwards, we explain a recent improvement on a result about large prime gaps, and present two proofs of this theorem: Maynard's proof [2] and a proof by Ford, Green, Konyagin, and Tao [3].

1. NOTATION

Unless otherwise mentioned, (n, m) refers to the greatest common divisor of n and m , and $[n, m]$ refers to their least common multiple. Sometimes these may denote the open or closed interval; in those cases, we will make this difference clear.

An expression of the form $n \equiv a \pmod{q}$ is equivalent to stating that q divides $n - a$. The "residue class" $a \pmod{q}$ is the subset of all integers containing all n such that $n \equiv a \pmod{q}$.

For a positive real x , $\pi(x)$ denotes the number of prime numbers in the interval $[1, x]$ and $\pi(x; q, a)$ denotes the number of primes in this interval which are $a \pmod{q}$. For a natural number n , p_n denotes the n th prime.

Expressions of the form $A = X + O(E)$ or $A - X \ll E$ are equivalent to saying that there exists some constant c such that $|A - X| \leq cE$. If the constant c itself depended on some other variables d_1, \dots, d_n but is otherwise uniform, then we express this by writing $A = X + O_{d_1, \dots, d_n}(E)$ or $A - X \ll_{d_1, \dots, d_n} E$.

An expression of the form $F(n) \sim G(n)$ is equivalent to $\lim_{n \rightarrow \infty} F(n)/G(n) = 1$. If the expressions F and G depend on many variables other than n , it will be clear from context which variable is tending to infinity.

2. PRELIMINARIES

2.1. Multiplicative functions. A multiplicative function $f : \mathbb{N} \rightarrow \mathbb{C}$ is defined as one for which $f(mn) = f(m)f(n)$ for any pair of coprime integers m, n . A totally multiplicative function is one for which this equation holds for all m and n . Consequently the values of any multiplicative function are completely determined by its values on prime powers; a totally multiplicative function is determined by its values on the primes. A basic property is that for two (totally) multiplicative functions f and g , the product fg and the Dirichlet convolution $(f \star g)(n) = \sum_{ab=n} f(a)g(b)$ are both (totally) multiplicative functions. In fact the (totally) multiplicative functions form a group under convolution, with identity given by δ which is given by $\delta(1) = 1$, $\delta(n) = 0$ for $n > 1$.

We define $\mu(n)$ to be the multiplicative function given by $\mu(1) = 1$, and for any prime p , $\mu(p) = -1$, $\mu(p^n) = 0$ for $n > 1$. This is the inverse of the function which is always 1. That

is,

$$(2.1) \quad \sum_{d|n} \mu(d) = \delta(n) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n > 1. \end{cases}$$

We define $\omega(n)$ to be the function which gives the number of distinct primes dividing n . For squarefree n , we have

$$(2.2) \quad \mu(n) = (-1)^{\omega(n)}.$$

We define $\varphi(n)$ to be the number of integers in $[1, n]$ which are coprime to n . This is a multiplicative function given by $\varphi(p^k) = p^{k-1}(p-1)$.

Finally, we will use the following result. It is a simplified form of Lemma 6.1 in [1].

Lemma 2.1. *Let $\kappa, A_1, A_2, L > 0$. Let γ be a multiplicative function satisfying*

$$0 \leq \frac{\gamma(p)}{p} \leq 1 - A_1,$$

and

$$-L \leq \sum_{w \leq p \leq z} \frac{\gamma(p) \log p}{p} - \kappa \log z/w \leq A_2$$

for any $2 \leq w \leq z$. Let g be the totally multiplicative function defined on primes by $g(p) = \gamma(p)/(p - \gamma(p))$. Then

$$(2.3) \quad \sum_{d < z} \mu(d)^2 g(d) = \mathfrak{O} \left(\frac{(\log z)^\kappa}{\Gamma(\kappa + 1)} + O_{A_1, A_2, \kappa}(\mathfrak{O} L (\log z)^{\kappa-1}) \right),$$

where

$$\mathfrak{O} = \prod_p \left(1 - \frac{\gamma(p)}{p} \right)^{-1} \left(1 - \frac{1}{p} \right)^\kappa.$$

2.2. The prime k -tuples conjecture. Given a set of positive integers $\mathcal{H} = \{h_1, \dots, h_k\}$ and a positive real x , a fundamental problem in prime number theory is to understand how many integers n in $[1, x]$ there are such that $n + h_1, n + h_2, \dots, n + h_k$ are all prime. Of course, if \mathcal{H} occupies all the residue classes $(\text{mod } p)$ for some prime p , then there will be at most finitely many such n . All other n would give at least one multiple of p which is at least $2p$. In the case that this scenario does not occur for any primes p , we call such a set admissible.

For a one-element (admissible) set \mathcal{H} , the answer is given by the celebrated prime number theorem, which asserts that

$$\pi(x) \sim \frac{x}{\log x}.$$

The error here is $\frac{x}{(\log x)^2}$; that is, $\pi(x) = \frac{x}{\log x} + O\left(\frac{x}{(\log x)^2}\right)$. A more precise version of the prime number theorem (proven in chapter 6 of [4]) asserts that

$$(2.4) \quad \pi(x) = \int_2^x \frac{dt}{\log t} + O(x \exp(-C \sqrt{\log x}))$$

for a constant C . The main term in this expression is known as the logarithmic integral and is also denoted $\text{li}(x)$. The reason this version of the prime number theorem is more precise is because the error term is much smaller. In fact, for any $A > 0$, $x \exp(-C \sqrt{\log x}) \ll_A \frac{x}{(\log x)^A}$. Therefore $\text{li}(x)$ is a much better approximation to $\pi(x)$. The Riemann Hypothesis is equivalent to showing that the error in (2.4) is actually the much smaller $O(x^{1/2} \log x)$.

For admissible sets \mathcal{H} which are larger, the answer is not yet known. In fact, the still unresolved twin-primes conjecture asserts that there are infinitely many such n for the two-element set $\mathcal{H} = \{0, 2\}$. However, number theorists are able to conjecture a lot more than this. To see the motivation behind the following conjecture, see [5].

Conjecture (Hardy-Littlewood). *The number of integers n in $[1, x]$ such that $n+h_1, \dots, n+h_k$ are all prime is $\sim \mathfrak{S}(\mathcal{H}) \frac{x}{(\log x)^k}$, where $\mathfrak{S}(\mathcal{H})$ is a constant that depends only on \mathcal{H} . Letting $\nu_p(\mathcal{H})$ be the number of distinct residue classes occupied by \mathcal{H} , the constant $\mathfrak{S}(\mathcal{H})$ is given by*

$$\mathfrak{S}(\mathcal{H}) = \prod_p \left(1 - \frac{\nu_p(\mathcal{H})}{p}\right) \left(1 - \frac{1}{p}\right)^{-k}.$$

We will be interested in the much weaker assertion that for an admissible set \mathcal{H} , there are infinitely many n such that all the $n+h_i$ are simultaneously prime. Maynard's work shows that for any m , there is a k large enough such that for any admissible set \mathcal{H} of size k , there are infinitely many n such that at least m of the $n+h_i$ are simultaneously prime. In particular, that implies $\liminf_n (p_{n+m} - p_n) < \infty$ for any m . The proof is surprisingly elementary, and the only real result used is the Bombieri-Vinogradov theorem.

2.3. Primes in arithmetic progressions. To understand the Bombieri-Vinogradov theorem, consider the problem of determining how many primes in the interval $[1, x]$ there are for a reduced residue class $a \pmod{q}$. We only consider reduced residue classes because otherwise, the integer $d = (a, q) > 1$ would divide every element of the residue class, and can only be prime for at most one entry. Intuition suggests that the primes should not conspire in any way to put more primes in one class more than another. In fact, this turns out to be the case. The prime number theorem for arithmetic progressions states that for a fixed reduced residue class modulo $a \pmod{q}$,

$$(2.5) \quad \pi(x; q, a) \sim \frac{\pi(x)}{\varphi(q)}.$$

However, in applications we usually need the modulus q to grow as x does, so (2.5) is insufficient for these purposes. The best kind of result that bounds this error directly is the Siegel-Walfisz theorem, which states that for $q \leq (\log x)^N$, there is a constant $C(N)$ depending on N such that

$$\pi(x; q, a) = \frac{\text{li}(x)}{\varphi(q)} + O\left(x \exp\left(-C(N)(\log x)^{1/2}\right)\right).$$

Unfortunately, the restriction that $q \leq (\log x)^N$ is quite severe and we would like to do better. In fact, the Generalized Riemann Hypothesis would give an error term of $O(x^{1/2} \log(qx))$ for $q \leq x$, which is a much better range. However, it seems difficult to be able to tackle this error term directly for q in a larger range.

Thankfully, we do have the Bombieri-Vinogradov theorem (proven in chapter 28 of [7]) which is often a sufficient substitute for the Generalized Riemann Hypothesis in many applications. The idea is that even though we may have difficulties bounding the error term in (2.5) uniformly for q in some range, we do know a result which deals with the average error for q in a good range.

Theorem 2.2 (Bombieri-Vinogradov). *For $A > 0$ there is some constant $B = B(A)$ depending on A such that*

$$(2.6) \quad \sum_{q \leq \frac{x^{1/2}}{(\log x)^B}} \max_{y \leq x} \max_{(a,q)=1} \left| \pi(y; q, a) - \frac{\text{li}(y)}{\varphi(q)} \right| \ll_A \frac{x}{(\log x)^A}.$$

The left hand side of (2.6) is summing, for each q in the indicated range, the worst possible error over all values $y \leq x$ and reduced residue classes $a \pmod{q}$. This is equivalent to stating that the “mean worst possible error” is no more than $O(x^{1/2}(\log x)^{B-A})$. In this way, the Bombieri-Vinogradov theorem is often utilized as an averaged version of the Generalized Riemann Hypothesis.

For $\theta > 0$, we say that the primes have “level of distribution θ ” if for any $A > 0$,

$$\sum_{q \leq x^\theta} \max_{(a,q)=1} \left| \pi(x; q, a) - \frac{\pi(x)}{\varphi(q)} \right| \ll_A \frac{x}{(\log x)^A}.$$

The definition of “level of distribution” extends in a straightforward manner to any sequence of integers as well, not just the primes. The idea is that a sequence of integers is “distributed well” if it divides evenly into all the reduced residue classes modulo q for every q . Therefore, knowing that the primes have level of distribution for larger θ would mean that the error term in the prime number theorem for arithmetic progressions is smaller.

The Bombieri-Vinogradov theorem implies that the primes have level of distribution θ for any $\theta < \frac{1}{2}$. In fact, number theorists actually conjecture quite a bit more. The following conjecture was originally formulated in 1968 by Elliott and Halberstam in [6].

Conjecture (Elliott-Halberstam). *The primes have level of distribution θ for any $\theta < 1$.*

Therefore if this conjecture were true, we would know that the “mean worst possible error” in the prime number theorem for arithmetic progressions would actually be $O(x^{1-\theta}(\log x)^{-A})$ for any $\theta < 1$, even though the Generalized Riemann Hypothesis states that the errors are only $O(x^{1/2} \log(qx))$.

2.4. Prime gaps. We will see in this expository article how the level of distribution of the primes influences the gaps between them. The prime number theorem implies that $p_n \sim n \log n$, so the average prime gaps are about $\log n$ in size. The work of Goldston, Pintz, and Yıldırım [8] shows that if we assume the primes have level of distribution θ for any $\theta < \frac{1}{2}$, then there do exist infinitely many prime gaps of size $\epsilon \log p_n$ for infinitely many n . That is,

$$(2.7) \quad \liminf_n \frac{p_{n+1} - p_n}{\log p_n} = 0.$$

That is, we can find gaps which are ϵ times the average prime gap for any $\epsilon > 0$. Furthermore, the work of Goldston, Pintz, and Yıldırım shows that if we assume that the primes have level of distribution $\frac{1}{2}$, then there are indeed infinitely many bounded gaps between primes. Maynard has been able to circumvent a difficulty in their work which allows him to prove bounded gaps between primes as long as the primes have level of distribution θ for some $\theta > 0$. In particular, Bombieri-Vinogradov suffices for his application.

The goal of this paper is to explain the ideas involved in Maynard’s proof without worrying too much about the details in the calculations. For those details, we refer the reader to Maynard’s original work [1].

3. IDEAS FROM SIEVE THEORY

We first consider the general problem of sieve theory. Consider any finite set \mathcal{A} of integers and \mathcal{P} to be a finite set of primes. For each $p \in \mathcal{P}$ we fix a subset $\mathcal{A}_p \subseteq \mathcal{A}$. We would like to get asymptotic bounds on the size of the set

$$\mathcal{S}(\mathcal{A}, \mathcal{P}) := \mathcal{A} \setminus \cup_{p \in \mathcal{P}} \mathcal{A}_p.$$

The exact size of the set can be given by the inclusion-exclusion principle. Setting $\mathcal{A}_d = \cap_{p|d} \mathcal{A}_p$ for any integer d and P to be the product of the primes in \mathcal{P} , we have

$$\#\mathcal{S}(\mathcal{A}, \mathcal{P}) = \sum_{d|P} (-1)^{\omega(d)} \#\mathcal{A}_d = \sum_{d|P} \mu(d) \#\mathcal{A}_d,$$

where the second equality follows from (2.2).

3.1. The sieve of Eratosthenes. Let \mathcal{A} be the set of integers in $(x, x + y]$, \mathcal{P} be the set of all primes $p \leq z$ for some z which may depend on x , and $\mathcal{A}_p = \mathcal{A} \cap p\mathbb{Z}$. Then $\mathcal{S}(\mathcal{A}, \mathcal{P})$ would consist of the subset of \mathcal{A} which consists of those integers which are free of any prime factors less than z . We would think of this subset as the subset of \mathcal{A} which are almost primes – it certainly contains all the primes in this interval. Then \mathcal{A}_d is the subset of \mathcal{A} which consists of the multiples of d , so

$$(3.1) \quad \#\mathcal{A}_d = \left\lfloor \frac{x+y}{d} \right\rfloor - \left\lfloor \frac{x}{d} \right\rfloor = \frac{y}{d} + O(1).$$

Therefore, from (3) and (3.1) we have

$$\begin{aligned} \pi(x+y) - \pi(x) &\leq \#\mathcal{S}(\mathcal{A}, \mathcal{P}) \\ &= \sum_{d|P} \mu(d) \#\mathcal{A}_d \\ &= \sum_{d|P} \mu(d) \left(\frac{y}{d} + O(1) \right) \\ &= y \sum_{d|P} \frac{\mu(d)}{d} + O\left(\sum_{d|P} 1 \right) \\ &= y \sum_{d|P} \frac{\mu(d)}{d} + O(\omega(P)). \end{aligned}$$

Here, $\omega(n)$ is the number of prime divisors dividing n (defined back in section 2.1). Since $P = \prod_{p \leq z} p$, we know $\omega(P) = 2^{\pi(z)} \leq 2^z$. Furthermore, note $\mu(d)/d$ is a multiplicative function, so the main term is a multiplicative function as well whose behavior on primes is $1 - \frac{1}{p}$ and zero on all non-squarefree numbers. Hence

$$(3.2) \quad \pi(x+y) - \pi(x) \leq y \prod_{p \leq z} \left(1 - \frac{1}{p} \right) + O(2^z).$$

Intuitively, the main term in (3.2) is the one obtained by assuming that all the primes dividing P act “independently” in the interval $(x, x + y]$. That is, there is a probability of $1 - \frac{1}{p}$ that any integer is not divisible by p , so we could guess that the number of integers in $(x, x + y]$ not divisible by any $p \leq z$ is given by $y \prod_{p \leq z} \left(1 - \frac{1}{p} \right)$. However, the error term associated with this naive guess is quite costly as it is exponential in z . The main term itself turns out to be asymptotic to $\frac{Cy}{\log z}$ for a constant C , so in order for the error term to be

smaller than this, we must take z to be roughly of size $\log y$. Taking $z = \log y - \log \log y$ shows that

$$(3.3) \quad \pi(x+y) - \pi(x) \ll \frac{y}{\log \log y}.$$

The prime number theorem implies that the left hand side of (3.3) is actually asymptotic to $\frac{y}{\log y}$, so the upper bound we have is not optimal. The difficulty stems from the fact that the error is exponential in z , so we seek to find some way to reduce the error.

3.2. The pure Brun sieve. An idea of Viggo Brun [10], fundamentally combinatorial in nature, gives such a way to reduce this error. The principle of inclusion-exclusion can be modified to give upper and lower bounds for $\#\mathcal{S}(\mathcal{A}, \mathcal{P})$ that involve fewer terms in the overall sum than the exact expansion (3). For any nonnegative integer r ,

$$(3.4) \quad \sum_{\substack{d|P \\ \omega(d) \leq 2r+1}} \mu(d) \#\mathcal{A}_d \leq \#\mathcal{S}(\mathcal{A}, \mathcal{P}) \leq \sum_{\substack{d|P \\ \omega(d) \leq 2r}} \mu(d) \#\mathcal{A}_d.$$

To see why this holds true, for any $a \in \mathcal{A}$ choose the maximal n for which $a \in \mathcal{A}_n$. Therefore if $a \in \mathcal{A}_d$ for some d , then $d|n$. Consequently this element $a \in \mathcal{A}$ gets counted $\sum_{d|n, \omega(d) \leq 2r} \mu(d)$ times on the right hand side. From combinatorics and (2.2), this simplifies to

$$\sum_{\substack{d|n \\ \omega(d) \leq 2r}} \mu(d) = \sum_{k=0}^{2r} \sum_{\substack{d|n \\ \omega(d)=k}} \mu(d) = \sum_{k=0}^{2r} (-1)^k \sum_{\substack{d|n \\ \omega(d)=k}} 1 = \sum_{k=0}^{2r} (-1)^k \binom{\omega(n)}{k} = (-1)^{2r} \binom{\omega(n)-1}{2r}.$$

If $n = 1$, then this equals 1 and if $n > 1$, it is nonnegative. Hence the right hand side of (3.4) does provide an upper bound for $\#\mathcal{S}(\mathcal{A}, \mathcal{P})$. Similar considerations show the left hand side gives a lower bound.

The benefit this new upper bound gives is that now the error can be made significantly smaller if r is chosen properly. We redo the calculation from the previous section with this new idea in mind.

$$\begin{aligned} \pi(x+y) - \pi(x) &\leq \#\mathcal{S}(\mathcal{A}, \mathcal{P}) \\ &\leq \sum_{\substack{d|P \\ \omega(d) \leq 2r}} \mu(d) \#\mathcal{A}_d \\ &= \sum_{\substack{d|P \\ \omega(d) \leq 2r}} \mu(d) \left(\frac{y}{d} + O(1) \right) \\ &= y \sum_{\substack{d|P \\ \omega(d) \leq 2r}} \frac{\mu(d)}{d} + O \left(\sum_{\substack{d|P \\ \omega(d) \leq 2r}} 1 \right). \end{aligned}$$

Since P has exactly $\pi(z)$ prime factors, the error will be exactly $\sum_{k=0}^{2r} \binom{\pi(z)}{k} \leq \sum_{k=0}^{2r} \frac{\pi(z)^k}{k!} \leq e^{\pi(z)}$, and an upper bound for this quantity is $O(z^{2r})$. It can now be shown via detailed calculation that for some $\alpha < 1$, if

$$\log z = \frac{\alpha \log x}{\log \log x} \quad \text{and} \quad r = \frac{\log x}{2 \log z}$$

then the main term and error term are both $\ll \frac{y \log \log y}{\log y}$. This is closer to the asymptotic $\frac{y}{\log y}$ given by the prime number theorem, and this is due to the fact that we had an error of $O(z^{2r})$ in the pure Brun sieve compared to the original Eratosthenes' sieve's $O(z^2)$ error.

3.3. Selberg's sieve. Let us now frame the discussions of the previous two sieves in a slightly different way. Again consider the range $(x, x + y]$ but now consider the number of these integers which are coprime to some fixed integer N . In the previous two examples, $N = \prod_{p < z} p$ but now we let it be any integer. Note that any integer a is coprime to N if and only if $(a, N) = 1$. We sum over all a in the interval $(x, x + y]$, adding 1 to our count if $(a, N) = 1$ and adding nothing to our count otherwise. Therefore, the quantity of interest is

$$\sum_{x < a \leq x+y} \delta((a, N)).$$

From the inversion formula (2.1), this equals

$$\sum_{x < a \leq x+y} \delta((a, N)) = \sum_{x < a \leq x+y} \sum_{d|(a, N)} \mu(d).$$

Since $d|(a, N)$ if and only if $d|a$ and $d|N$, we can switch the order of the sums here to get

$$\sum_{x < a \leq x+y} \delta((a, N)) = \sum_{d|N} \mu(d) \sum_{\substack{x < a \leq x+y \\ d|a}} 1,$$

and note that the inner sum is exactly what we defined \mathcal{A}_d to be previously. Hence this calculation turns out to give exactly the same result as in the sieve of Eratosthenes. Furthermore, the fundamental inequality in the pure Brun sieve is that for any r ,

$$\sum_{\substack{d|(a, N) \\ \omega(d) \leq 2r+1}} \mu(d) \leq \delta((a, N)) \leq \sum_{\substack{d|(a, N) \\ \omega(d) \leq 2r}} \mu(d).$$

Motivated by this realization, we return to our general situation where \mathcal{A} is any general set of integers and \mathcal{P} is any finite set of primes. To place upper and lower bounds on $\#\mathcal{S}(\mathcal{A}, \mathcal{P})$, we need to find functions which approximate $\delta(n)$ from above and below. Selberg's idea was to consider an upper bound by choosing some function $\lambda_n : \mathbb{N} \rightarrow \mathbb{R}$ (not necessarily multiplicative) with $\lambda_1 = 1$, and then noting that

$$\delta(n) \leq \left(\sum_{d|n} \lambda_d \right)^2.$$

This equation holds since both sides are 1 when $n = 1$, and when $n > 1$ the left side is zero while the right is nonnegative as it is the square of a real number. Now we can calculate

the result of this approximation.

$$\begin{aligned}
\#\mathcal{S}(\mathcal{A}, \mathcal{P}) &= \sum_{a \in \mathcal{A}} \delta((a, P)) \\
&\leq \sum_{a \in \mathcal{A}} \left(\sum_{d|(a, P)} \lambda_d \right)^2 \\
&= \sum_{a \in \mathcal{A}} \sum_{d_1, d_2 | (a, P)} \lambda_{d_1} \lambda_{d_2} \\
&= \sum_{d_1, d_2 | P} \lambda_{d_1} \lambda_{d_2} \left(\sum_{\substack{d_1, d_2 | a \\ a \in \mathcal{A}}} 1 \right).
\end{aligned}$$

Note that d_1, d_2 both divide a if and only if their least common multiple $[d_1, d_2]$ does, so the inner quantity is precisely $\mathcal{A}_{[d_1, d_2]}$. Hence,

$$(3.5) \quad \#\mathcal{S}(\mathcal{A}, \mathcal{P}) \leq \sum_{d_1, d_2 | P} \lambda_{d_1} \lambda_{d_2} \mathcal{A}_{[d_1, d_2]}.$$

In order to get the best possible upper bound, we would like to choose the function λ , subject to the constraint that $\lambda_1 = 1$ in such a way that the right hand side is minimized. The right hand side is a quadratic form, so we can either diagonalize it via an appropriate change of coordinates or apply Lagrange multipliers in order to find the right minimum.

In the usual situation $\mathcal{A} = \mathbb{N} \cap (x, x+y]$ and $\mathcal{P} = \{p : p \leq z\}$, we recall that $\mathcal{A}_d = \frac{y}{d} + O(1)$ and substitute this relation into (3.5) to get

$$\#\mathcal{S}(\mathcal{A}, \mathcal{P}) \leq y \sum_{d_1, d_2 | P} \frac{\lambda_{d_1} \lambda_{d_2}}{[d_1, d_2]} + O\left(\sum_{d_1, d_2 | P} |\lambda_{d_1}| \cdot |\lambda_{d_2}| \right).$$

To deal with the error term effectively, let us suppose that $\lambda_d = 0$ for all $d > z$. If $\lambda_{max} = \max_d |\lambda_d|$, then the error is $O(z^2 \lambda_{max})$. We will see later that $\lambda_{max} = 1$, so that the error is $O(z^2)$, which is already better than the $O(z^2)$ given by the sieve of Eratosthenes and the $O(z^{2r})$ given by the pure Brun sieve. It turns out that if we make the change of coordinates (details can be found in [9])

$$(3.6) \quad u_\delta = \sum_{\substack{d \leq z \\ \delta | d}} \frac{\lambda_d}{d} \iff \lambda_\delta = \delta \sum_{\delta | d} \mu\left(\frac{d}{\delta}\right) u_d,$$

then the constraints transform to $u_\delta = 0$ for $\delta > z$ and $\sum_{\delta \leq z} \mu(\delta) u_\delta = 1$. The quadratic form in the main term becomes diagonalized to give

$$\#\mathcal{S}(\mathcal{A}, \mathcal{P}) \leq y \sum_{d_1, d_2 \leq z} \varphi(\delta) u_\delta^2 + O\left(z^2 \lambda_{max}^2\right).$$

From calculus or Cauchy's inequality, the sum in the main term has minimum $1/V(z)$ where

$$V(z) := \sum_{d \leq z} \frac{\mu(d)^2}{\varphi(d)}.$$

The optimal choice of u_δ will be

$$(3.7) \quad u_\delta = \frac{\mu(\delta)}{\varphi(\delta) V(z)}.$$

Under this choice one can check that $|\lambda_d| \leq 1$ for all d , so that $\lambda_{max} = 1$. Hence,

$$\#\mathcal{S}(\mathcal{A}, \mathcal{P}) \leq \frac{y}{V(z)} + O(z^2).$$

We can show $V(z)$ is roughly $\log z$ via (2.3) with $\gamma(p) = \kappa = 1$ for all p . Therefore, in order to get the best possible upper bound here we would like to choose z such that $y \sim z^2 \log z$. Choosing $z = \left(\frac{y}{\log y}\right)^{1/2}$ gives

$$\#\mathcal{S}(\mathcal{A}, \mathcal{P}) \ll \frac{y}{\log y}.$$

We know we cannot do better than this from the prime number theorem.

3.4. Summary. The three sieves we discussed in this section were increasingly more complex, but the main terms became increasingly more precise and the error terms became smaller and smaller.

Sieve	Main term	Error term	Estimate for $\pi(x+y) - \pi(x)$
Eratosthenes	$y \sum_{d P} \frac{\mu(d)}{d} = y \prod_{p \leq z} \left(1 - \frac{1}{p}\right) = y \frac{\varphi(P)}{P}$	$O(2^z)$	$\ll \frac{y}{\log \log y}$
Brun	$y \sum_{d P, \omega(d) \leq 2r} \frac{\mu(d)}{d}$	$O(z^{2r})$	$\ll \frac{y \log \log y}{\log y}$
Selberg	$y \left(\sum_{d \leq z} \frac{\mu(d)^2}{\varphi(d)}\right)^{-1}$	$O(z^2)$	$\ll \frac{y}{\log y}$

3.5. Using Selberg's sieve to pick out prime tuples. Let us now try to better understand the values of λ_δ which gave the minimum in Selberg's sieve. Substituting (3.7) into (3.6),

$$(3.8) \quad \lambda_\delta = \frac{\delta}{V(z)} \sum_{\substack{\delta|d \\ d \leq z}} \frac{\mu(d/\delta)\mu(d)}{\varphi(d)}.$$

We set $d = k\delta$ in this sum. Since $\mu(k\delta) = 0$ whenever k and δ are not coprime, we sum over all $k \geq 1$ with $(k, \delta) = 1$ and use multiplicativity of μ and φ here.

$$\lambda_\delta = \frac{\delta}{V(z)} \sum_{\substack{k \leq z/\delta \\ (k, \delta) = 1}} \frac{\mu(k)\mu(k\delta)}{\varphi(k\delta)} = \frac{\delta\mu(\delta)}{V(z)\varphi(\delta)} \sum_{\substack{k \leq z/\delta \\ (k, \delta) = 1}} \frac{\mu(k)^2}{\varphi(k)}.$$

Applying (2.3) with $\gamma(p) = \delta((p, \delta))$ and $\kappa = 1$ will show that the sum in the above equation is approximately $\varphi(\delta) \log(z/\delta)/\delta$, so

$$(3.9) \quad \lambda_\delta \sim \frac{\mu(\delta) \log(z/\delta)}{\log z}.$$

Let $\lambda'_\delta = \frac{\mu(\delta) \log(z/\delta)}{\log z}$. Then for $n > 1$,

$$\begin{aligned} \sum_{\delta|n} \lambda'_\delta &= \sum_{\delta|n} \frac{\mu(\delta)(\log(z/n) + \log(n/\delta))}{\log z} \\ &= \frac{\log(z/n)}{\log z} \delta(n) + \frac{1}{\log z} (\mu \star \log)(n) \\ &= \frac{1}{\log z} (\mu \star \log)(n). \end{aligned}$$

It can be shown that $\mu \star \log$ is a function whose behavior on prime powers p^k is $\log p$ and zero otherwise. Therefore, the optimal Selberg weights are essentially picking out the primes and nothing else, since the contribution from prime powers is minimal. This is precisely what we wanted to do, since we wanted to count the primes in some range

$(x, x + y]$. In this way, the optimal weights form a kind of approximation to the identity $(\mu \star \log)(n) = 0$ when $\omega(n) > 1$. We prove a more general version of this identity.

Lemma 3.1. *When $\omega(n) > k$, $(\mu \star (\log)^k)(n) = 0$.*

Proof. Suppose $n = p_1^{e_1} \cdots p_\ell^{e_\ell}$ for $\ell > k$. Then

$$\begin{aligned} \sum_{d|n} \mu(d) \log^k \left(\frac{n}{d} \right) &= \sum_{I \subseteq \{1, \dots, \ell\}} (-1)^{\#I} \log^k \left(\prod_{i \in I} p_i^{e_i - 1} \right) \\ &= \sum_{I \subseteq \{1, \dots, \ell\}} (-1)^{\#I} \left(\sum_{i \in I} \log p_i^{e_i - 1} \right)^k \\ &= \sum_{I \subseteq \{1, \dots, \ell\}} (-1)^{\#I} \sum_{i_1, \dots, i_k \in I} \prod_{j=1}^k \log p_{i_j}^{e_{i_j} - 1} \\ &= \sum_{i_1, \dots, i_k \in I} \left(\prod_{j=1}^k \log p_{i_j}^{e_{i_j} - 1} \right) \left(\sum_{\{i_1, \dots, i_k\} \subseteq I \subseteq \{1, \dots, \ell\}} (-1)^{\#I} \right). \end{aligned}$$

If there are m distinct i_j , then to calculate the inner sum we can sum over $s := \#I$ to get

$$\sum_{s=m}^{\ell} (-1)^s \sum_{\substack{\{i_1, \dots, i_k\} \subseteq I \subseteq \{1, \dots, \ell\} \\ \#I=s}} 1 = \sum_{s=m}^{\ell} (-1)^s \binom{\ell - m}{s - m} = (1 - 1)^{\ell - m} = 0.$$

Hence the entire sum is zero since each inner sum is. Note the importance of $\ell > k$; this allowed us to assume $\ell > m$ so that we could calculate the inner sum in the above fashion. \square

The form of this lemma suggests that perhaps we could modify Selberg's sieve in a way that produces weights $\lambda_\delta \sim \mu(\delta) \left(\frac{\log(z/\delta)}{\log z} \right)^k$ which would essentially pick out prime k -tuples. Indeed, fix an admissible set $\mathcal{H} = \{h_1, \dots, h_k\}$ with $0 \leq h_1 \leq \dots \leq h_k$, and suppose we want to place an upper bound on the number of $n \in [x, 2x]$ such that the $n + h_i$ are simultaneously prime. We have just solved this problem in the case that \mathcal{H} is a one element set. If λ is again a real-valued function on the naturals with $\lambda_1 = 1$ supported in the range $[1, z]$ for some $z < x$, note that

$$\delta(X(n + h_1) \cdots X(n + h_k)) \leq \left(\sum_{d|(n+h_1) \cdots (n+h_k)} \lambda_d \right)^2.$$

To see why this holds true, note the right hand side is always nonnegative and the left hand side is either 0 or 1. When the left hand side is 1, all the $n + h_i$ are simultaneously prime, so that $d|(n + h_1) \cdots (n + h_k)$ implies either $d = 1$ or $d \geq n + h_1 \geq x + 0 > z$. The latter case contributes nothing due to the restricted support of λ , so when the left hand side is 1, so is the right hand side. Therefore,

$$\begin{aligned} \#\{x \leq n \leq 2x : n + h_i \text{ all prime}\} &= \sum_{x \leq n \leq 2x} \delta(X(n + h_1) \cdots X(n + h_k)) \\ &\leq \sum_{x \leq n \leq 2x} \left(\sum_{d|(n+h_1) \cdots (n+h_k)} \lambda_d \right)^2. \end{aligned}$$

Analyzing this sum in the exact same way we did it for the one-element case gives

$$\#\{x \leq n \leq 2x : n + h_i \text{ all prime}\} \leq \sum_{d_1, d_2 \leq z} \lambda_{d_1} \lambda_{d_2} \left(\sum_{\substack{x \leq n \leq 2x \\ [d_1, d_2] | (n+h_1) \cdots (n+h_k)}} 1 \right).$$

The condition $[d_1, d_2] | (n + h_1) \cdots (n + h_k)$ is equivalent to stating that n needs to lie in a certain number of residue classes modulo $[d_1, d_2]$. This number only depends on \mathcal{H} and $[d_1, d_2]$, so let it be $f([d_1, d_2])$. It is not hard to show via the Chinese remainder theorem that for squarefree d , $f(d) = \prod_{p|d} \nu_{\mathcal{H}}(p)$. Returning to the estimate, we can proceed as we did in the one-element case to get

$$(3.10) \quad \#\{x \leq n \leq 2x : n + h_i \text{ all prime}\} \leq x \sum_{d_1, d_2 \leq z} \frac{f([d_1, d_2])}{[d_1, d_2]} \lambda_{d_1} \lambda_{d_2} + O(\lambda_{\max}^2 z^2).$$

We skip the calculation to find the minimum and cite the result instead. We get an expression for λ_δ and as we expected, its approximate value is a suitable generalization of (3.9):

$$\lambda_d \sim \mu(d) \left(\frac{\log(z/d)}{\log z} \right)^k.$$

So $\lambda_{\max} = 1$ again, and if we choose $z \leq \sqrt{x}/(\log x)^{2k}$, the main term will dominate the error term in (3.10). In that case, it is possible to show that (see [5])

$$x \sum_{d_1, d_2 \leq z} \frac{f([d_1, d_2])}{[d_1, d_2]} \lambda_{d_1} \lambda_{d_2} \sim 2^k k! \mathfrak{S}(\mathcal{H}) \frac{x}{(\log x)^k}.$$

Substituting this into (3.10) gives an upper bound that is $2^k \cdot k!$ times the value predicted by the Hardy-Littlewood conjecture.

4. FINDING SMALL PRIME GAPS

We now return to the question of finding small prime gaps. Let \mathcal{H} be an admissible set of size k , and order them such that $h_1 \leq \cdots \leq h_k$. If we could show that for infinitely many n , at least two of the $n + h_i$ are prime, then we would know there are infinitely many prime gaps of size no larger than $h_k - h_1$. Suppose we try to show that for large enough x , there is always such an n in the range $[x, 2x]$. Then there would certainly be infinitely many such n by choosing large powers of 2 for x . Letting $X(n)$ be the prime indicator (1 if n is prime, 0 otherwise), we are interested in the quantity

$$\sum_{x \leq n \leq 2x} \left(\sum_{i=1}^k X(n + h_i) - 1 \right).$$

In particular, if we can show that this quantity is strictly positive, then there must be some value of n for which $\sum_{i=1}^k X(n + h_i) - 1 > 0$, so that will be the desired value of n .

More generally, consider for some $\rho > 0$ and nonnegative weights $w_n : \mathbb{N} \rightarrow \mathbb{R}^{\geq 0}$ the quantity

$$(4.1) \quad S(x, \rho) := \sum_{x \leq n \leq 2x} \left(\sum_{i=1}^k X(n + h_i) - \rho \right) w_n.$$

If we can show that $S(x, \rho) > 0$ for all x , then by a similar argument as before, we can find infinitely many n for which at least $\lfloor \rho + 1 \rfloor$ of the $n + h_i$ are simultaneously prime.

4.1. **The GPY sieve.** We consider $\rho = 1$ and in order to choose w_n such that (4.1) is strictly positive, we instead will tackle the more difficult task of finding the w_n such that for each i ,

$$\sum_{\substack{x \leq n \leq 2x \\ X(n+h_i)=1}} w_n > \frac{1}{k} \sum_{x \leq n \leq 2x} w_n.$$

Summing this up over all the i would then give (4.1). Consequently, we would like to maximize the ratio

$$(4.2) \quad \left(\sum_{\substack{x \leq n \leq 2x \\ X(n+h_i)=1}} w_n \right) / \left(\sum_{x \leq n \leq 2x} w_n \right).$$

Motivated by our work with Selberg's sieve in Section 3.5, we would like to find weights of the form $w_n = (\sum_{d|n} \lambda_d)^2$ given by

$$\lambda_d = \mu(d) P\left(\frac{\log(z/d)}{\log z}\right).$$

where $P(y)$ is a polynomial with $P(1) = 1$ which vanishes to order at least k at $y = 0$. The analysis of the denominator proceeds in essentially the same way, and making the assumption that z is a little below \sqrt{x} as before, one can show that

$$(4.3) \quad \sum_{x \leq n \leq 2x} w_n \sim \frac{x}{(\log z)^k} \mathfrak{O}(\mathcal{H}) \int_0^1 \frac{y^{k-1}}{(k-1)!} P^{(k)}(1-y)^2 dy.$$

The numerator then becomes

$$(4.4) \quad \sum_{d_1, d_2 \leq z} \lambda_{d_1} \lambda_{d_2} \left(\sum_{\substack{x \leq n \leq 2x \\ [d_1, d_2] | (n+h_1) \cdots (n+h_k) \\ n+h_i \text{ prime}}} 1 \right).$$

We have this extra constraint on the inner sum that $n + h_i$ must be prime. As before, n needs to belong to one of $f([d_1, d_2])$ residue classes modulo $[d_1, d_2]$. Considering what this means for $n + h_i$, we only care about the residue classes for n for which $n + h_i$ lies in a reduced residue class. Suppose there are then $g([d_1, d_2])$ such reduced residue classes, so that by the prime number theorem for arithmetic progressions,

$$(4.5) \quad \sum_{\substack{x \leq n \leq 2x \\ [d_1, d_2] | (n+h_1) \cdots (n+h_k) \\ n+h_i \text{ prime}}} 1 \sim \frac{\pi(x)}{\varphi([d_1, d_2])} g([d_1, d_2]) \sim \frac{x}{\log x} \cdot \frac{g([d_1, d_2])}{\varphi([d_1, d_2])}.$$

Note that for a prime p , $g(p) = \nu_H(p) - 1$, and the Chinese remainder theorem gives that when d is squarefree, $g(d) = \prod_{p|d} (\nu_H(p) - 1)$. To deal with the errors involved with substituting (4.5) back into (4.4), we can apply the Bombieri-Vinogradov theorem since we are adding up a lot of these errors for moduli in a range. However Bombieri-Vinogradov only lets us go up to $[d_1, d_2] \leq \sqrt{x}$, so suppose $z \leq x^{1/4}$ instead (since $[d_1, d_2] \leq d_1 d_2 \leq z^2$). This allows us to take care of the error terms and so the estimate for the numerator becomes

$$(4.6) \quad \sum_{\substack{x \leq n \leq 2x \\ X(n+h_i)=1}} w_n \sim \frac{x}{(\log x)(\log z)^{k-1}} \mathfrak{O}(\mathcal{H}) \int_0^1 \frac{y^{k-2}}{(k-2)!} P^{(k-1)}(1-y)^2.$$

With the estimates (4.3) and (4.6), one can show that the ratio (4.2) is optimized by choosing $P(y) = y^{k+r}$ for some nonnegative integer $r \sim \sqrt{k}/2$, in which case the ratio (4.2) equals

$$\left(\frac{\log z}{\log x}\right) \left(\frac{4(\sqrt{k}+1)}{(\sqrt{k}+2)(k+\sqrt{k}+1)}\right).$$

Since $z \leq x^{1/4}$ in order for the estimate of the numerator to work properly, we see that the above fraction is just a little bit less than $1/k$. The reason we could not make z any larger was because Bombieri-Vinogradov only tells us that the primes have level of distribution θ for any $\theta < 1/2$. If we were to assume Elliott-Halberstam, then this method would already give bounded gaps between primes. Indeed, all we really needed was to be able to extend z to a slightly bigger range, so knowing the primes have level of distribution θ for some $\theta \geq 1/2$ would do the job.

However, a slight modification of this method shows that if the size of our admissible set \mathcal{H} were $\epsilon \log x$ for any $\epsilon > 0$, we can guarantee the existence of at least two primes among the $n + h_i$. The extra widening of \mathcal{H} 's range allows us to find the extra prime, but at the cost of losing boundedness of the potential prime gaps. This calculation can be found in Soundararajan's exposition of the GPY sieve method in [5].

In summary, the GPY sieve method shows that if we assume the primes have level of distribution θ for $\theta < 1/2$ (as guaranteed by Bombieri-Vinogradov), then

$$\liminf_n \frac{p_{n+1} - p_n}{\log p_n} = 0.$$

If we were to assume Elliott-Halberstam, then

$$\liminf_n p_{n+1} - p_n < \infty.$$

4.2. Finding bounded gaps: Maynard's work. Maynard's extension of the GPY sieve is to consider weights of the form

$$w_n = \left(\sum_{d_i | n + h_i} \lambda_{d_1, \dots, d_k} \right)^2.$$

The extra flexibility for the λ -parameters to depend on the divisors of each $n + h_i$ is what gives the extra push needed to overcome the deficiency in the GPY sieve.

Furthermore, a technical issue requires that $w_n = 0$ unless n lies in a fixed residue class $v_0 \pmod{W}$ where $W = \prod_{p \leq D_0} p$. It is possible to choose $D_0 = \log \log \log x$ so that by the prime number theorem, $W \ll (\log \log x)^2$. The reason we consider n to be in a fixed residue class modulo W is so that we can not worry about the effect of small primes $p \leq D_0$. In any case, most of the analysis is unaffected since W is so small in comparison to x .

Using the Chinese remainder theorem and the fact that \mathcal{H} is admissible, it is possible to choose v_0 such that $(v_0 + h_i, W) = 1$ for all i . Therefore we would like to maximize the ratio S_2/S_1 , where

$$(4.7) \quad S_1 = \sum_{\substack{x \leq n \leq 2x \\ n \equiv v_0 \pmod{W}}} \left(\sum_{d_i | n + h_i} \lambda_{d_1, \dots, d_k} \right)^2,$$

$$(4.8) \quad S_2 = \sum_{\substack{x \leq n \leq 2x \\ n \equiv v_0 \pmod{W}}} \left(\sum_{i=1}^k X(n + h_i) \right) \left(\sum_{d_i | n + h_i} \lambda_{d_1, \dots, d_k} \right)^2.$$

If we can show $S_2/S_1 > \rho$, then $S(x, \rho) > 0$ as desired.

To maximize the ratio S_2/S_1 , it is more convenient to work in slightly better coordinates. Recall that during the derivation of Selberg's sieve, we switched to the “ u ” coordinates which allowed us to diagonalize a particular quadratic form which ensured that our bounds would be optimal. In the case of Maynard's sieve, a similar multi-dimensional change of coordinates makes the calculations simpler. That is, we first define the new coordinates

$$y_{r_1, \dots, r_k} = \left(\prod_{i=1}^k \mu(r_i) \varphi(r_i) \right) \sum_{\substack{d_1, \dots, d_k \\ r_i | d_i \forall i}} \frac{\lambda_{d_1, \dots, d_k}}{\prod_{i=1}^k d_i},$$

$$y_{r_1, \dots, r_k}^{(m)} = \left(\prod_{i=1}^k \mu(r_i) g(r_i) \right) \sum_{\substack{d_1, \dots, d_k \\ r_i | d_i \forall i \\ d_m = 1}} \frac{\lambda_{d_1, \dots, d_k}}{\prod_{i=1}^k \varphi(d_i)},$$

where g is the totally multiplicative function defined on primes p by $g(p) = p - 2$. These coordinates are related as follows

$$(4.9) \quad y_{r_1, \dots, r_{m-1}, 1, r_{m+1}, \dots, r_k}^{(m)} \sim \sum_{a_m} \frac{y_{r_1, \dots, r_{m-1}, a_m, r_{m+1}, \dots, r_k}}{\varphi(a_m)}.$$

Then Maynard derives asymptotics for the S_i in these new coordinates. Note that the S_i are now diagonalized in these new coordinates.

$$S_1 \sim \frac{N}{W} \sum_{r_1, \dots, r_k} \frac{y_{r_1, \dots, r_k}^2}{\prod_{i=1}^k \varphi(r_i)}$$

$$S_2 \sim \frac{N}{\varphi(W) \log N} \sum_{m=1}^k \sum_{r_1, \dots, r_k} \frac{(y_{r_1, \dots, r_k}^{(m)})^2}{\prod_{i=1}^k g(r_i)}.$$

Since we have now diagonalized S_1 and S_2 , we can apply Lagrange multipliers (and ignoring any error terms) to obtain a choice of y which will be near-optimal. In order to maximize the ratio S_2/S_1 , we can treat S_1 to be the constant and S_2 the quantity to maximize. Doing so, we arrive at the condition

$$\lambda y_{r_1, \dots, r_k} = \left(\prod_{i=1}^k \frac{\varphi(r_i)}{g(r_i)} \right) \sum_{m=1}^k \frac{g(r_m)}{\varphi(r_m)} y_{r_1, \dots, r_{m-1}, 1, r_{m+1}, \dots, r_k}^{(m)}$$

for some fixed constant λ . Recall that the y 's are supported only when the product of the r_j 's is squarefree and does not contain any prime factors smaller than D_0 . Since $g(r) \approx \varphi(r) \approx r$ for most integers r which are coprime to the small primes, the above condition is essentially equivalent to

$$\lambda y_{r_1, \dots, r_k} \approx \sum_{m=1}^k y_{r_1, \dots, r_{m-1}, 1, r_{m+1}, \dots, r_k}^{(m)}.$$

From (4.9), this is equivalent to

$$(4.10) \quad \lambda y_{r_1, \dots, r_k} \approx \sum_{m=1}^k \sum_{a_m} \frac{y_{r_1, \dots, r_{m-1}, a_m, r_{m+1}, \dots, r_k}}{\varphi(a_m)}.$$

We see that (4.10) is a smooth condition in the r_i . That is, this choice uses nothing about the primes dividing the r_i . Consequently, (4.10) will be satisfied if we set

$$(4.11) \quad y_{r_1, \dots, r_k} = F\left(\frac{\log r_1}{\log R}, \dots, \frac{\log r_k}{\log R}\right),$$

where the support of y_{r_1, \dots, r_k} is restricted to when $\prod r_i$ is squarefree and coprime to W .

Using this choice of y (and hence λ), Maynard's main result is the following.

Theorem 4.1. *Let the primes have exponent of distribution $\theta > 0$, and let $z = x^{\theta/2-\delta}$ for some small fixed $\delta > 0$. Let $\lambda_{d_1, \dots, d_k}$ be defined in terms of a fixed piecewise differentiable function F by*

$$\lambda_{d_1, \dots, d_k} = \left(\prod_{i=1}^k \mu(d_i) d_i\right) \sum_{\substack{r_1, \dots, r_k \\ d_i | r_i \forall i \\ (r_i, W) = 1 \forall i}} \frac{\mu(\prod_{i=1}^k r_i)^2}{\prod_{i=1}^k \varphi(r_i)} F\left(\frac{\log r_1}{\log z}, \dots, \frac{\log r_k}{\log z}\right),$$

whenever $(\prod_{i=1}^k d_i, W) = 1$, and let $\lambda_{d_1, \dots, d_k} = 0$ otherwise. Moreover, let F be supported on $\mathcal{R}_k = \{(x_1, \dots, x_k) \in [0, 1]^k : \sum_{i=1}^k x_i \leq 1\}$. Then we have

$$S_1 \sim \frac{\varphi(W)^k x (\log z)^k}{W^{k+1}} I_k(F),$$

$$S_2 \sim \frac{\varphi(W)^k x (\log z)^{k+1}}{W^{k+1} \log x} \sum_{j=1}^k J_k^{(m)}(F),$$

provided $I_k(F) \neq 0$ and $J_k^{(m)}(F) \neq 0$ for each m , where

$$I_k(F) = \int_0^1 \dots \int_0^1 F(t_1, \dots, t_k)^2 dt_1 \dots dt_k,$$

$$J_k^{(m)}(F) = \int_0^1 \dots \int_0^1 \left(\int_0^1 F(t_1, \dots, t_k) dt_m\right)^2 dt_1 \dots dt_{m-1} dt_{m+1} \dots dt_k.$$

Using the above theorem, we see that our desired ratio is

$$(4.12) \quad \frac{S_2}{S_1} = \left(\frac{\log z}{\log x}\right) \frac{\sum_{j=1}^k J_k^{(m)}(F)}{I_k(F)} = \left(\frac{\theta}{2} - \delta\right) \frac{\sum_{j=1}^k J_k^{(m)}(F)}{I_k(F)}.$$

Therefore, if we can show that

$$\frac{\theta}{2} \sup_F \frac{\sum_{j=1}^k J_k^{(m)}(F)}{I_k(F)} > \rho,$$

then since we can make the right hand side of the above equation as close to $\frac{\theta}{2} \sup_F \frac{\sum_{j=1}^k J_k^{(m)}(F)}{I_k(F)}$ as we like, we would be done. Since θ and ρ are fixed, it suffices to show that

$$M_k := \sup_F \frac{\sum_{j=1}^k J_k^{(m)}(F)}{I_k(F)}.$$

can be made to be as large as possible when k grows. So we would like to find smooth functions F which can maximize this ratio now; the problem is essentially reduced to calculus now.

We now consider the optimal choice of F . For small k , Maynard has used F of the form

$$F(t_1, \dots, t_k) = \begin{cases} P(t_1, \dots, t_k), & \text{if } (t_1, \dots, t_k) \in \mathcal{R}_k \\ 0, & \text{otherwise,} \end{cases},$$

where P is a symmetric polynomial in the t_i -variables. To work with symmetric polynomials, one uses the general fact that such a P must necessarily be a polynomial expression in expressions of the form $P_n = \sum_{i=1}^k t_i^n$. For simplicity, consider only P of the form

$$P = \sum_{i=1}^d a_i (1 - P_1)^{b_i} P_2^{c_i},$$

for reals a_i and nonnegative integers b_i and c_i . For P of this form, it is possible to compute I_k and $J_k^{(m)}$ quite explicitly. Note that by symmetry, $J_k^{(m)}(P)$ will only depend on k . Then one can write down $d \times d$ matrices $N_1(\mathbf{b}, \mathbf{c})$ and $N_2(\mathbf{b}, \mathbf{c})$ depending only on the b_i and c_i such that $\sum_{m=1}^k J_k^{(m)}(F) = \mathbf{a}^\top N_2 \mathbf{a}$ and $I_k(F) = \mathbf{a}^\top N_1 \mathbf{a}$. It is then a fact of linear algebra that expressions of the form

$$\frac{\mathbf{a}^\top N_2 \mathbf{a}}{\mathbf{a}^\top N_1 \mathbf{a}}$$

are maximized when \mathbf{a} is an eigenvector of $N_1^{-1} N_2$, with the ratio then being the largest eigenvalue of $N_1^{-1} N_2$. Using a computer program, one can now compute that $M_5 > 2$ and $M_{105} > 4$.

However for general k , computations using this method are difficult to perform. Instead, a choice of F that would work is the following.

$$(4.13) \quad F(t_1, \dots, t_k) = \begin{cases} \prod_{i=1}^k g(kt_i), & \text{if } \sum_{i=1}^k t_i \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

$$(4.14) \quad g(t) = \begin{cases} 1/(1 + At), & t \in [0, T], \\ 0, & \text{otherwise,} \end{cases}$$

and where $A = \log k - 2 \log_2 k$ and $T = (e^A - 1)/A$. This choice of F then shows that

$$M_k > \log k - 2 \log \log k - 2.$$

This completes his proof of bounded gaps between primes.

Unraveling what the specific constants will be, we choose $\rho = 1$ and can set $\theta = 1/2$ by Bombieri-Vinogradov. Hence if we have an admissible set of size 105 whose entries are $0 \leq h_1 \leq \dots \leq h_{105}$, then at least two of the $n + h_i$ are prime for infinitely many n . Calculations by Thomas Engelsma have found an explicit such admissible set with $h_{105} - h_1 = 600$. On the other hand, if we assume Elliott-Halberstam then we merely need $M_k > 2$, so that the admissible set $\{0, 2, 6, 8, 12\}$ of size five is sufficient. Therefore we know unconditionally that there are infinitely many prime gaps of size 600 or less, and if we assume Elliott-Halberstam, this can be improved to gaps of size 12 or less.

5. APPLICATION TO LARGE PRIME GAPS

5.1. Preliminaries. We now consider the related question of finding large gaps between the primes. We know that the quantity $G(x) = \sup_{p_n \leq x} (p_{n+1} - p_n)$ grows in x : for example, the numbers $n! + 2, \dots, n! + n$ are all composite since k divides $n! + k$ for every integer $1 \leq k \leq n$, so this produces a prime gap of size at least n for every integer n . In fact, this construction was a bit inefficient since it produces a prime gap of size $n \approx \log \log n! = \log \log x$, and we already know from the prime number theorem that average prime gaps should be of size $\log x$. However, we can do better than this. Let \log_ν denote the ν -fold logarithm, so that

$$\log_n(x) = \underbrace{\log \log \dots \log x}_{n \text{ times}}$$

Theorem 5.1 (Erdős-Rankin). *There is a constant $c > 0$ such that*

$$G(x) \geq (c + o(1))(\log x)(\log_2 x)(\log_4 x)(\log_3 x)^{-2},$$

so that

$$(5.1) \quad \limsup \frac{P_{n+1} - P_n}{(\log p_n)(\log_2 p_n)(\log_4 p_n)(\log_3 p_n)^{-2}} > 0.$$

Proof. (Outline) Set $P_x = \prod_{p \leq x} p$. The idea is to find the largest possible U such that the interval $[1, U]$ can be covered by arithmetic progressions of the form $a_p \pmod{p}$ for primes $p \leq x$. If we can do this, then by the Chinese Remainder Theorem there exists some $a \in [U - 1, U + P_x - 2]$ such that $a \equiv -a_p \pmod{p}$ for all primes p . For any integer n in the interval $[a + 1, a + U]$, we have $n - a \in [1, U]$, so there is some prime $p \leq x$ with $n - a \equiv a_p \pmod{p}$. Hence $n \equiv a + a_p \equiv 0 \pmod{p}$, so n cannot be prime (assuming U is significantly larger than x so that $n \neq p$). This produces a prime gap of size at least U among numbers which are less than $a + U < 2U + P_x \leq 3P_x$. Note $P_x \sim x$ by the prime number theorem, so that it suffices to find U which is at least

$$U \gg \frac{(\log P_x)(\log_2 P_x)(\log_4 P_x)}{(\log_3 P_x)^2} \sim \frac{x(\log x)(\log_3 x)}{(\log_2 x)^2}.$$

Indeed, fix $\epsilon \in (0, 1)$ and now define

$$y = \exp\left((1 - \epsilon) \frac{\log x \log_3 x}{\log_2 x}\right), \quad z = \frac{x}{\log_2 x}, \quad U = \frac{x \log y}{\log_2 x}.$$

Note $0 \leq y \leq z \leq x$. The idea is to use different strategies for finding a_p depending on whether $p \leq y$, $p < y \leq z$, or $p > z$. In the ‘‘medium prime’’ case $p < y \leq z$, choose $a_p = 0$. Each sieve $a_p \pmod{p}$ for p in this range will knock out all elements divisible by p . For the small primes $p \leq y$, choose the a_p which will knock out the most elements from the remaining set; this gives us a reduction of a factor of about

$$\prod_{p \leq y} \left(1 - \frac{1}{p}\right) \sim \frac{C}{\log y}.$$

Finally, it will turn out that the size of the remaining set is no larger than the number of ‘‘large primes’’ $p \in (z, x]$ we have left, so we can trivially just pick the a_p that will take care of at least one element of the remaining set to finish. \square

The last step in the Erdős-Rankin construction is a bit wasteful since each large prime only gets rid of one element from the remaining set. Indeed, this is the step that Maynard’s sieve allows us to improve.

5.2. Maynard’s improvement of the large prime sieve. If each large prime on average could get rid of a constant C_U number of remaining elements, then it stands to reason that we could take

$$(5.2) \quad y = \exp\left((1 - \epsilon) \frac{\log x \log_3 x}{\log_2 x}\right), \quad z = \frac{x}{\log_2 x}, \quad U = C_U \frac{x \log y}{\log_2 x},$$

run the argument through until the final step, and use our large primes more efficiently this time. The hope would be that such an improvement would allow us to show that

$$(5.3) \quad \limsup \frac{P_{n+1} - P_n}{(\log p_n)(\log_2 p_n)(\log_4 p_n)(\log_3 p_n)^{-2}} = \infty.$$

Indeed, this is precisely Maynard’s improvement of the Erdős-Rankin result which he shows in [2].

In fact we can even relax the choice of the a_p for the small primes to

$$(5.4) \quad a_p = 0, \quad \text{for every prime } p \in (y, z],$$

$$(5.5) \quad a_p = 1, \quad \text{for every prime } p \leq y.$$

To describe what kinds of numbers are left after these two sieving processes, we introduce a definition.

Definition 5.2. *An integer n is called m -smooth if n is coprime to all primes $p \leq m$.*

Call the y -smooth subset of remaining elements \mathcal{R}' , and suppose the rest are \mathcal{R} . Then for any $m \in \mathcal{R}'$, $m-1$ is also y -smooth since we knocked out the residue classes $1 \pmod{p}$ for $p \leq y$. For $n \in \mathcal{R}'$, n must contain a large prime factor since it is not y -smooth and also does not contain any primes $y < p \leq z$ (due to the first sieving process). Clearly $z^2 > x$, so n contains no more than one large prime factor, and the rest has to be free of primes less than y . Hence the remaining elements are $\mathcal{R} \cup \mathcal{R}'$, where

$$(5.6) \quad \mathcal{R} = \{mp \leq U : p > z, m \text{ is } y\text{-smooth}, (mp-1, P_y) = 1\},$$

$$(5.7) \quad \mathcal{R}' = \{m \leq U : m \text{ is } y\text{-smooth}, (m-1, P_y) = 1\}.$$

Also note that the condition $(mp-1, P_y) = 1$ in the definition of \mathcal{R} means m has to be even, since otherwise $2|mp-1$ would imply $(mp-1, P_y) \geq 2$. Therefore, for even m set

$$(5.8) \quad \mathcal{R}_m = \{z < p \leq U/m : (mp-1, P_y) = 1\}, \quad \mathcal{R} = \cup_{m \text{ even}} \mathcal{R}_m.$$

We would like to be able to bound the sizes of \mathcal{R}_m , \mathcal{R}' , and ensure that we can cover $\mathcal{R} \cup \mathcal{R}'$ with arithmetic progressions whose common difference is a large prime. Maynard obtains the following estimates on the sizes of these sets.

Lemma 5.3. *We have*

$$(5.9) \quad |\mathcal{R}'| \ll \frac{x}{(\log x)^{1+\epsilon}}.$$

We have uniformly for $z + z/\log x \leq V \leq x(\log x)^2$ and $m \leq x$

$$(5.10) \quad \#\{z < p \leq V : (mp-1, P_y) = 1\} = \frac{V-z}{\log x} \left(\prod_{\substack{p \leq y \\ p \nmid m}} \frac{p-2}{p-1} \right) (1 + O(\exp(-(\log_2 x)^{1/2}))).$$

In particular, uniformly for even $m \leq U(1 - 1/\log x)/z$ we have

$$(5.11) \quad |\mathcal{R}_m| = \frac{2e^{-\gamma}U(1+o(1))}{m(\log x)(\log y)} \left(\prod_{p>2} \frac{p(p-2)}{(p-1)^2} \right) \left(\prod_{p|m, p>2} \frac{p-1}{p-2} \right).$$

Furthermore,

$$(5.12) \quad \sum_{U/(z(\log_2 x)^2) \leq m < U/z} |\mathcal{R}_m| = o\left(\frac{x}{\log x}\right),$$

$$(5.13) \quad \sum_{1 \leq m < U/(z(\log_2 x)^2)} |\mathcal{R}_m| = O\left(\frac{C_U x}{\log x}\right).$$

Armed with these estimates, Maynard reduces the problem of choosing residue classes for large primes to the following result.

Proposition 5.4. *Fix $\epsilon, \delta > 0$. Let $m < Uz^{-1}(\log_2 x)^{-2}$ be even and let $I_m \subseteq [x/2, x]$ be an interval of length at least $\delta|\mathcal{R}_m| \log x$. Then for $x > x_0(\epsilon, \delta, C_U)$, there exists a choice of residue classes $a_q \pmod{q}$ for each prime $q \in I_m$ such that all but $\epsilon|\mathcal{R}_m|$ elements p of \mathcal{R}_m satisfy $p \equiv a_q \pmod{q}$ for some prime $q \in I_m$.*

Proof of (5.3) using Proposition 5.4. From 5.13 we know

$$\sum_{m < Uz^{-1}(\log_2 x)^{-2}} \delta' |\mathcal{R}_m| \log x \ll \delta' C_U x.$$

Therefore choosing $\delta' = (4C_U)^{-1}$, we can find disjoint intervals $\mathcal{I}_m \subseteq [x/2, x]$ of length $\delta' |\mathcal{R}_m| \log x$ for each even $m < Uz^{-1}(\log_2 x)^{-2}$. Let J_m be an interval contained in the middle one-third of \mathcal{I}_m , and apply Proposition 5.4 to J_m with $\epsilon = \delta = \delta'/3$. Therefore for large enough x , there is a choice of $a_q \pmod{q}$ for each prime in J_m which covers all but $\epsilon |\mathcal{R}_m|$ elements of \mathcal{R}_m . With the remaining $2\epsilon |\mathcal{R}_m| \log x$ elements in \mathcal{I}_m , we can cover those $\epsilon |\mathcal{R}_m|$ elements with a residue class. Therefore the primes in $\cup_{m < Uz^{-1}(\log_2 x)^{-2}} \mathcal{I}_m \subseteq [x/2, x]$ are enough to cover $\cup_{m < Uz^{-1}(\log_2 x)^{-2}} \mathcal{R}_m$. From Lemma 5.3 there can be at most $o(x/\log x)$ elements in \mathcal{R} left that need to be covered. Since there are at least that many primes in $[z, x/2]$ via the prime number theorem, we can use one residue class for each prime in that interval to cover the rest.

5.3. Applying the Maynard sieve to large prime gaps. For each \mathcal{I}_m for an even $m < Uz^{-1}(\log_2 x)^{-2}$ we need to somehow choose a residue class $a_q \pmod{q}$ for each $q \in \mathcal{I}_m$. The idea is to show the existence of such a choice probabilistically. In other words, the goal is to put a probability measure $\mu_{m,q}$ on the residue classes $a \pmod{q}$ such that the probability of any given $p \in \mathcal{R}_m$ to not lying in $\cup_{m,q \in \mathcal{I}_m} (a_q + q\mathbb{Z})$ is small. This probability equals

$$\prod_{q \in \mathcal{I}_m \text{ prime}} (1 - \mu_{m,q}(p)) = \exp\left(\sum_{q \in \mathcal{I}_m \text{ prime}} \log(1 - \mu_{m,q}(p)) \right) \leq \exp\left(- \sum_{q \in \mathcal{I}_m \text{ prime}} \mu_{m,q}(p) \right).$$

If we can choose the probability measure $\mu_{m,q}$ in such a way that for almost every $p \in \mathcal{R}_m$,

$$t \leq \mathbb{E}_\mu \left[\# \text{ of } q \text{ for which } p = a_q \pmod{q} \right] = \sum_{q \in \mathcal{I}_m \text{ prime}} \mu_{m,q}(p),$$

then for almost every p we would know that the probability of p not being covered by some $a_q \pmod{q}$ is no more than e^{-t} . Taking t sufficiently large, the expected number of $p \in \mathcal{R}_m$ not being covered would be less than any desired small fraction of $|\mathcal{R}_m|$. Hence there would exist at least one arrangement of the $a_q \pmod{q}$ which would ensure that Proposition 5.4 holds.

The key is to ensure that $\mu_{m,q}(a)$ is large when $a \pmod{q}$ contains many primes in \mathcal{R}_m ; this will ensure that such residue classes contribute more to the desired expectation $\mathbb{E}_\mu \left[\# \text{ of } q \text{ for which } p = a_q \pmod{q} \right]$, so that we can find some configuration of $a_q \pmod{q}$ which works. Maynard's sieve provides the right weights for such a probability measure.

Let w be a quantity that grows very slowly with respect to x ; say $w = \log_4 x$ so that $P_w = o(\log_2 x)$. For $1 \leq j \leq k$ let $h_j = p_{\pi(k)+j} P_w$, so that h_j is the product of some very small primes and then $p_{\pi(k)+j}$, which is the j th prime larger than k . We will treat k essentially like a constant, so that w is larger than k . So each h_j is not squarefree as it contains $p_{\pi(k)+j}^2$. The set $\mathcal{H} = \{h_1, \dots, h_k\}$ is seen to be admissible. Now choose

$$(5.14) \quad \mu_{m,q}(a) = \alpha_{m,q} \sum_{\substack{n \leq U/m \\ n \equiv a \pmod{q} \\ (n(mn-1), P_w) = 1}} \left(\sum_{\substack{d_1, \dots, d_k \\ d_i | n + h_i q}} \lambda_{d_1, \dots, d_k}^{(1)} \sum_{\substack{e_1, \dots, e_k \\ e_i | m(n+h_i q) - 1}} \prod_{i=1}^k \lambda_{e_i}^{(2)} \right)^2,$$

where the $\lambda_{d_1, \dots, d_k}^{(1)}$ and $\lambda_{e_i}^{(2)}$ are weights we choose later, and $\alpha_{m,q}$ is the appropriate normalizing constant chosen so that the $\mu_{m,q}$ sum to 1.

The structure of the weights 5.14 reveals why we hope they will have the desired properties. The $\lambda_{d_1, \dots, d_k}^{(1)}$ are chosen to look like a GPY sieve, and we will chose them to be large only if all the $n + h_i q$ are close to being prime; that is, they have no small prime factors. This will mean that $\mu_{m,q}(a)$ is large only if there is some $n \equiv a \pmod{p}$ such that many of the $n + h_i q$ are actually prime. In such a situation, $a \pmod{q}$ contains many primes in \mathcal{R}_m so we would want $\mu_{m,q}(a)$ to be large.

The $\lambda_{e_j}^{(2)}$ are chosen to imitate Selberg weights to ensure that $m(n + h_i q) - 1$ does not have any prime factors smaller than y^ϵ . We want this property in order to ensure that $mp - 1$ is indeed free of prime factors less than y (recall this was a condition for membership in \mathcal{R}_m). Toward this end, Maynard chooses

$$(5.15) \quad \lambda_{d_1, \dots, d_k}^1 = \left(\prod_{i=1}^k \mu(d_i) \right) \sum_{j=1}^J c_j \left(\prod_{\ell=1}^k F_{\ell,j} \left(\frac{\log d_\ell}{\log x} \right) \right),$$

$$(5.16) \quad \lambda_{e_i}^2 = \mu(e_i) G \left(\frac{\log e_i}{\log y} \right),$$

for some smooth non-negative functions $F_{i,j}, G : [0, \infty) \rightarrow \mathbb{R}$, and constants $c_j > 0$. Furthermore the support of the $F_{i,j}$ and G will be restricted so that

$$\sup \left\{ \sum_{i=1}^k u_i : F_{i,j}(u_i) \neq 0 \right\} \leq 1/10,$$

$$\text{supp } G \subseteq [0, 1].$$

With these choices, Maynard is able to prove the following lower bound for our desired probability.

Lemma 5.5. *Let $m < Uz^{-1}(\log x)^{-2}$ be even and let $p_0 \in \mathcal{R}_m$ with $h_k x < p_0 < U/m - h_k x$. Then*

$$(5.17) \quad \sum_{q \in \mathcal{I}_m} \mu_{m,q}(p_0) \gg (1 + o_k(1)) \frac{k |\mathcal{I}_m| J_k^{(1)}(F) J_k^{(2)}(G)}{(\log x) |\mathcal{R}_m| I_k^{(1)}(F) I_k^{(2)}(G)},$$

where

$$I_k^{(1)}(F) = \int \cdots \int_{t_1, \dots, t_k \geq 0} F(t_1, \dots, t_k)^2 dt_1 \dots dt_k, \quad I_k^{(2)}(G) = \left(\int_0^\infty G'(t)^2 dt \right)^k,$$

$$J_k^{(1)}(F) = \int \cdots \int_{t_1, \dots, t_{k-1} \geq 0} \left(\int_{t_k \geq 0} F(t_1, \dots, t_k) dt_k \right)^2 dt_1 \dots dt_{k-1}, \quad J_k^{(2)}(G) = G(0)^2 \left(\int_0^\infty G'(t)^2 dt \right)^{k-1}.$$

Making the same choices for the smooth functions F and G as in Maynard's small prime gap paper, he uses

$$(5.18) \quad F_k(t_1, \dots, t_k) = \begin{cases} \prod_{i=1}^k g(kt_i), & \text{if } \sum_{i=1}^k t_i \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

$$(5.19) \quad g(t) = \begin{cases} 1/(1 + At), & t \in [0, T], \\ 0, & \text{otherwise,} \end{cases}$$

and where $A = \log k - 2 \log_2 k$ and $T = (e^A - 1)/A$. Plugging this choice of smooth functions into (5.17) results in

$$\sum_{q \in \mathcal{I}_m} \mu_{m,q}(p_0) \gg \log k.$$

From (5.10) one can show that the number of $p \in \mathcal{R}_m$ which do not satisfy $h_k x < p < U/m - h_k x$ is no larger than $o_k(|\mathcal{R}_m|)$ for even $m < Uz^{-1}(\log_2 x)^{-2}$. Hence we have showed Proposition 5.4, so the proof is complete.

5.4. Another Approach toward Large Prime Gaps. Ford, Green, Konyagin, and Tao [3] have found another way to get Maynard's estimate on large prime gaps. Again the approach is to show that we can find arithmetic progressions $a_p \pmod{p}$ for primes $p \leq x$ that cover the first U integers, where

$$U \sim \frac{C_U x (\log x) (\log_3 x)}{(\log_2 x)^2}.$$

More specifically, they show that for any integer $r \geq 13$, one can choose x large enough (depending on r) such that U can be taken to be

$$U := \frac{r}{6 \log r} \frac{x (\log x) (\log_3 x)}{(\log_2 x)^2}.$$

Set

$$w = x^{\log_3 x / (3 \log_2 x)}$$

and separate the primes $p \leq x$ into the four classes

$$\mathcal{S}_1 := \{p \text{ prime} : p \leq \log x \text{ or } w < p \leq x/4\}$$

$$\mathcal{S}_2 := \{p \text{ prime} : \log x < p \leq w\}$$

$$\mathcal{S}_3 := \mathcal{P} := \{p \text{ prime} : x/2 < p \leq x\}$$

$$\mathcal{S}_4 := \{p \text{ prime} : x/4 < p \leq x/2\}.$$

From the prime number theorem, there are $\left(\frac{1}{4} + o(1)\right) \frac{x}{\log x}$ primes in \mathcal{S}_4 . We can trivially pick $a_p \pmod{p}$ to knock out at least one element in any subset of U , so we are left to show that we can pick $a_p \pmod{p}$ for $p \in \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$ such that

$$\{1, \dots, U\} \setminus \bigcup_{i=1}^3 \bigcup_{p \in \mathcal{S}_i} (a_p + p\mathbb{Z}) \leq \left(\frac{1}{5} + o(1)\right) \frac{x}{\log x}.$$

For $p \in \mathcal{S}_1$, we choose $a_p = 0$. Then the remaining numbers have no prime factors smaller than $\log x$, and are either w -smooth or contain a prime factor larger than $x/4$. Those in the latter category must necessarily be prime, or else any other prime factor is at least $\log x$, so they are at least $(x/4) \log x > U$, a contradiction. Letting \mathcal{Q} be those primes in this latter category; that is,

$$\mathcal{Q} = \{p \text{ prime} : x/4 < p < U\},$$

we see that the remaining numbers belong to \mathcal{Q} or are z -smooth. If we set t to be such that $U = z^t$, de Bruijn's estimate gives that the number of z -smooth numbers less than U is no more than

$$U e^{-t \log t + O(t \log \log(t+2))} = o\left(\frac{x}{\log x}\right).$$

Therefore we can effectively discard this contribution, so the primes from \mathcal{S}_1 have helped us reduce to showing

$$\mathcal{Q} \setminus \bigcup_{i=2}^3 \bigcup_{p \in \mathcal{S}_i} (a_p + p\mathbb{Z}) \leq \left(\frac{1}{5} + o(1)\right) \frac{x}{\log x}.$$

The main improvement from previous methods now comes from how the primes in \mathcal{S}_2 are chosen. For each prime $p \in \mathcal{S}_2$, we would like to find $a_p \pmod{p}$ which contains at least

r primes in \mathcal{Q} . Hardy-Littlewood implies that there should be about a constant multiple of $\frac{x}{(\log x)^r}$ such tuples. The hard theorem shown by these authors and Ziegler is to establish a bound of this kind “on average.” Precisely, the result that will be relevant for large prime gaps is the following.

Theorem 5.6. *For $p \in \mathcal{P} = \mathcal{S}_3$ and $q \in \mathcal{Q}$, write “ $p \dashv\!\! \dashv q$ ” if $\{p, p+r!q, \dots, p+(r-1)r!q\} \subseteq \mathcal{Q}$, but $p+rr!q \notin \mathcal{Q}$. There exists a constant α_r and subsets $\mathcal{P}_0 \subseteq \mathcal{P}$, $\mathcal{Q}_0 \subseteq \mathcal{Q}$ such that*

$$(5.20) \quad \#\mathcal{P}_0 \sim \frac{x}{2 \log x}; \quad \#\mathcal{Q}_0 \sim \frac{U}{\log x},$$

such that

$$(5.21) \quad \#\{q \in \mathcal{Q} : p \dashv\!\! \dashv q - ir!p\} \sim \alpha_r \frac{U}{\log^r x}$$

for all $p \in \mathcal{P}_0$ and $0 \leq i \leq r-1$, and similarly that

$$(5.22) \quad \#\{p \in \mathcal{P}_0 : p \dashv\!\! \dashv q - ir!p\} \sim \alpha_r \frac{x}{2 \log^r x}$$

for all $q \in \mathcal{Q}_0$ and $0 \leq i \leq r-1$.

The “average” aspect of the hard theorem manifests itself in the restriction that we need to consider \mathcal{P}_0 and \mathcal{Q}_0 instead of only using \mathcal{P} and \mathcal{Q} .

Given Theorem 5.6, the next step is to consider a random choice $\vec{\mathbf{a}} = (\mathbf{a}_p)_{p \in \mathcal{S}_2}$ of residue classes for each modulo $p \in \mathcal{S}_2$. The probability that an integer survives this random sieving process will be

$$\gamma := \prod_{p \in \mathcal{S}_2} \left(1 - \frac{1}{p}\right) = \prod_{\log x < p \leq w} \left(1 - \frac{1}{p}\right)$$

which by Mertens’ theorem, is asymptotic to

$$\gamma \sim \frac{\log_2 x}{\log w} \sim \frac{3(\log_2 x)^2}{\log x \log_3 x} \sim \frac{r}{2 \log r} \left(\frac{x}{U}\right)$$

We are interested in $\mathcal{Q}(\vec{\mathbf{a}})$, defined to be the primes remaining after removing the arithmetic progressions $\mathbf{a}_p \pmod{p}$. However, Theorem 5.6 allows us to control prime arithmetic progressions in the slightly smaller set \mathcal{Q}_0 , so define $\mathcal{Q}_0(\vec{\mathbf{a}}) = \mathcal{Q}(\vec{\mathbf{a}}) \cap \mathcal{Q}_0$. Therefore from linearity of expectation (with respect to the random choice in $\vec{\mathbf{a}}$), we get

$$(5.23) \quad \mathbb{E}\#\mathcal{Q}(\vec{\mathbf{a}}) = \gamma\#\mathcal{Q} \sim \frac{r}{2 \log r} \frac{x}{\log x}$$

$$(5.24) \quad \mathbb{E}\#(\mathcal{Q}(\vec{\mathbf{a}}) \setminus \mathcal{Q}_0(\vec{\mathbf{a}})) = \gamma(\#\mathcal{Q} \setminus \mathcal{Q}_0) = o\left(\gamma \frac{U}{\log x}\right) = o\left(\frac{x}{\log x}\right).$$

Therefore, once we have found a satisfactory random choice of $\vec{\mathbf{a}}$, it suffices to choose residue classes for primes $p \in \mathcal{S}_3 = \mathcal{P}$ such that

$$(5.25) \quad \mathcal{Q}(\vec{\mathbf{a}}) \setminus \bigcup_{p \in \mathcal{S}_3} (a_p + p\mathbb{Z}) \leq \left(\frac{1}{5} + o(1)\right) \frac{x}{\log x}.$$

From equation (5.24), we expect that we should be able to replace $\mathcal{Q}(\vec{\mathbf{a}})$ with $\mathcal{Q}_0(\vec{\mathbf{a}})$ with probability 1 in the choice of $\vec{\mathbf{a}}$, since the $o(x/\log x)$ will just go to the error term. Let $1 - \delta(x)$ be the the probability that

$$(5.26) \quad \#(\mathcal{Q}(\vec{\mathbf{a}}) \setminus \mathcal{Q}_0(\vec{\mathbf{a}})) = o\left(\frac{x}{\log x}\right).$$

Then $\delta(x)$ should be small, and in fact, one can show from Markov's inequality applied to equation (5.24) that

$$\delta(x) = o(x).$$

The claim is that it suffices to show the following result, which guarantees the existence of $q_p \pmod{p}$ for each prime $p \in \mathcal{P}_0 \in \mathcal{S}_3$ which, when combined with the random choice $(\mathbf{a}_p)_{p \in \mathcal{S}_2}$, give us exactly the desired (5.25).

Theorem 5.7. *Suppose $\varepsilon(x) > 0$ is going to 0 arbitrarily slowly as x grows, so that $\varepsilon(x) = o(1)$. Then with probability at least $\varepsilon(x)$ in the random choice of $\vec{\mathbf{a}}$, there exists a length r arithmetic progression $\{q_p + ir!p : 0 \leq i \leq r-1\}$ for each $p \in \mathcal{P}_0 \subseteq \mathcal{P} = \mathcal{S}_3$ such that*

$$(5.27) \quad \mathcal{Q}_0(\vec{\mathbf{a}}) \setminus \bigcup_{p \in \mathcal{S}_3} (q_p + p r! \mathbb{Z}) \leq \left(\frac{1}{5} + o(1) \right) \frac{x}{\log x},$$

where this $o(1)$ depends on $\varepsilon(x)$.

So we can take $\varepsilon(x) = 2\delta(x)$, so that $1 - \delta(x) + \varepsilon(x) > 1$ implies that granted Theorem 5.7, we can find $\vec{\mathbf{a}}$ satisfying both equations (5.26) and (5.27). This will show equation (5.25), and we will be done.

We would like a variant of the relation \dashv between \mathcal{P} and \mathcal{Q} to a random relation between \mathcal{P}_0 and $\mathcal{Q}(\vec{\mathbf{a}})$. So for $p \in \mathcal{P}_0$ and $q \in \mathcal{Q}(\vec{\mathbf{a}})$, we define $p \overset{\vec{\mathbf{a}}}{\dashv} q$ if $p \dashv q$ and the entirety of the r -term arithmetic progression $\{q, q + r!p, \dots, q + (r-1)r!p\}$ is contained in $\mathcal{Q}(\vec{\mathbf{a}})$; that is, if this entire arithmetic progression survives the sieving operation by the primes in $\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_4$. This is precisely the situation when we can really use our result about r -term prime arithmetic progressions to the fullest.

Our intuition about independence would be that if $p \dashv q$, then $p \overset{\vec{\mathbf{a}}}{\dashv} q$ should have probability γ^r . In fact, this is true ‘‘on average,’’ which we formalize by taking random subsets again as follows.

Lemma 5.8. *Suppose $\varepsilon(x)$ is a quantity going to zero arbitrarily slowly as x grows. With probability at least $\varepsilon(x)$, there exist random subsets $\mathcal{P}_1(\vec{\mathbf{a}}) \subseteq \mathcal{P}_0$ and $\mathcal{Q}_1(\vec{\mathbf{a}}) \subseteq \mathcal{Q}_0(\vec{\mathbf{a}})$ with*

$$(5.28) \quad \#\mathcal{P}_1(\vec{\mathbf{a}}) \sim \frac{x}{2 \log x}; \quad \#\mathcal{Q}_1(\vec{\mathbf{a}}) \sim \#\mathcal{Q}_0(\vec{\mathbf{a}}) \sim \frac{r}{2 \log r} \frac{x}{\log x}.$$

Furthermore, define for $p \in \mathcal{P}_1(\vec{\mathbf{a}})$ and $0 \leq i \leq r-1$

$$\mathcal{Q}_1(\vec{\mathbf{a}}, p; i) = \{q \in \mathcal{Q}(\vec{\mathbf{a}}) : p \overset{\vec{\mathbf{a}}}{\dashv} q - ir!p\}$$

and define for $q \in \mathcal{Q}_1(\vec{\mathbf{a}})$ and $0 \leq i \leq r-1$

$$\mathcal{P}_1(\vec{\mathbf{a}}, q; i) := \{p \in \mathcal{P}_1(\vec{\mathbf{a}}) : p \overset{\vec{\mathbf{a}}}{\dashv} q - ir!p\}.$$

Then

$$(5.29) \quad \#\mathcal{Q}_1(\vec{\mathbf{a}}, p; i) \sim \gamma^r \alpha_r \frac{U}{\log^r x}$$

and

$$(5.30) \quad \#\mathcal{P}_1(\vec{\mathbf{a}}, q; i) \sim \gamma^{r-1} \alpha_r \frac{x}{2 \log^r x}$$

The (implicit) $o(1)$ errors depend on $\varepsilon(x)$.

Assuming Lemma 5.8, we can show Theorem 5.7. We choose $\vec{\mathbf{a}}$ and corresponding $\mathcal{P}_1(\vec{\mathbf{a}})$, $\mathcal{Q}_1(\vec{\mathbf{a}})$ satisfying the properties of Lemma 5.8, and will find q_p for $p \in \mathcal{P}_1(\vec{\mathbf{a}}) \subseteq \mathcal{S}_3$ (this is the last sieving process) satisfying the conclusion of Theorem 5.7.

This $\vec{\mathbf{q}} = (\mathbf{q}_p)_{p \in \mathcal{S}_3}$ will be chosen randomly (independently for each p) once $\vec{\mathbf{a}}$ has been fixed such that

$$\mathbf{q}_p \in \mathcal{Q}(\vec{\mathbf{a}}, p; 0) = \{q \in \mathcal{Q}(\vec{\mathbf{a}}) : p \stackrel{\vec{\mathbf{a}}}{\parallel} q\}.$$

This means we really want to knock out r primes in a progression from the remaining set $\mathcal{Q}(\vec{\mathbf{a}})$. Write $\mathbb{P}_{\vec{\mathbf{q}}}$ and $\mathbb{E}_{\vec{\mathbf{q}}}$ for the probability measure and expectation induced by the random choice of $\vec{\mathbf{q}}$. Finally, let

$$\mathcal{Q}_1(\vec{\mathbf{a}}, \vec{\mathbf{q}}) = \mathcal{Q}_1(\vec{\mathbf{a}}) \setminus \bigcup_{p \in \mathcal{P}_1(\vec{\mathbf{a}})} \{\mathbf{q}_p, \mathbf{q}_p + r!p, \dots, \mathbf{q}_p + i(r-1)r!p\}$$

be the unsieved elements after sieving out by $\mathbf{a}_p \pmod{p}$ for $p \in \mathcal{S}_2$ and $\mathbf{q}_p \pmod{p}$ for $p \in \mathcal{S}_3$.

Fixing $q \in \mathcal{Q}_1(\vec{\mathbf{a}})$, it is straightforward to see that $\mathcal{P}_1(\vec{\mathbf{a}}, q; i)$ for $0 \leq i \leq r-1$ are disjoint from the definition of the relation $\stackrel{\vec{\mathbf{a}}}{\parallel}$. For $p \in \mathcal{P}_1(\vec{\mathbf{a}}, q; i)$, we know by definition that $q - ir!p \in \mathcal{Q}(\vec{\mathbf{a}}, p; 0)$. Therefore from the random choice of \mathbf{q}_p and Lemma 5.8, we have

$$\mathbb{P}_{\vec{\mathbf{q}}}(\mathbf{q}_p = q - ir!p) = \frac{1}{\mathcal{Q}(\vec{\mathbf{a}}, p; 0)} = \frac{1 + o(1)}{\gamma^r \alpha_r \frac{U}{\log^r x}}$$

Using independence, we compute $\mathbb{P}_{\vec{\mathbf{q}}}(q \in \mathcal{Q}_1(\vec{\mathbf{a}}, \vec{\mathbf{q}}))$ as follows.

$$\begin{aligned} \mathbb{P}_{\vec{\mathbf{q}}}(q \in \mathcal{Q}_1(\vec{\mathbf{a}}, \vec{\mathbf{q}})) &= \mathbb{P}_{\vec{\mathbf{q}}}(\mathbf{q}_p \neq q - ir!p \text{ for all } i, p \in \mathcal{P}_1(\vec{\mathbf{a}})) \\ &\leq \mathbb{P}_{\vec{\mathbf{q}}}(\mathbf{q}_p \neq q - ir!p \text{ for all } i, p \in \mathcal{P}_1(\vec{\mathbf{a}}, q; i)) \\ &= \prod_{i=0}^{r-1} \prod_{p \in \mathcal{P}_1(\vec{\mathbf{a}}, q; i)} \mathbb{P}_{\vec{\mathbf{q}}}(\mathbf{q}_p \neq q - ir!p) \\ &= \prod_{i=0}^{r-1} \prod_{p \in \mathcal{P}_1(\vec{\mathbf{a}}, q; i)} \left(1 - \frac{1 + o(1)}{\gamma^r \alpha_r \frac{U}{\log^r x}}\right) \\ &= \exp\left(-\frac{1 + o(1)}{\gamma^r \alpha_r \frac{U}{\log^r x}} r(1 + o(1)) \cdot \#\mathcal{P}_1(\vec{\mathbf{a}}, q; i)\right) \\ &= \exp\left(-\frac{1 + o(1)}{\gamma^r \alpha_r \frac{U}{\log^r x}} r(1 + o(1)) \gamma^{r-1} \alpha_r \frac{x}{2 \log^r x}\right) \\ &= \exp\left(-(1 + o(1)) \frac{rx}{2U\gamma}\right) \\ &= \exp(-(1 + o(1)) \log r) \\ &\leq \frac{1 + o(1)}{r}. \end{aligned}$$

Hence, from linearity of expectation and Lemma 5.8,

$$\mathbb{E}_{\vec{\mathbf{q}}}(\#\mathcal{Q}_1(\vec{\mathbf{a}}, \vec{\mathbf{q}})) = \mathbb{P}_{\vec{\mathbf{q}}}(q \in \mathcal{Q}_1(\vec{\mathbf{a}}, \vec{\mathbf{q}})) \cdot \#\mathcal{Q}_1(\vec{\mathbf{a}}) \leq \left(\frac{1 + o(1)}{r}\right) \left(\frac{r}{\log r} \frac{x}{\log x}\right) = \frac{1}{2 \log r} \left(\frac{x}{\log x}\right).$$

From this estimate on expectation, we can find a deterministic $\vec{\mathfrak{q}}$ (choice of sieve for primes in \mathcal{S}_3) for every $\vec{\mathfrak{a}}$ (choice of sieve for primes in \mathcal{S}_2) such that

$$\#Q_1(\vec{\mathfrak{a}}, \vec{\mathfrak{q}}) \leq \frac{1 + o(1)}{2 \log r} \frac{x}{\log x}.$$

Taking $r \geq 13$, this gives the desired

$$\#Q_1(\vec{\mathfrak{a}}, \vec{\mathfrak{q}}) \leq \left(\frac{1}{5} + o(1)\right) \frac{x}{\log x}.$$

6. ACKNOWLEDGEMENTS

I am very grateful to my thesis advisor, Professor Kannan Soundararajan, for his help, guidance, and valuable comments toward the writing of this article.

REFERENCES

- [1] James Maynard, *Small gaps between primes*, Preprint.
- [2] ———, *Large gaps between primes*, Preprint.
- [3] K. Ford, B. Green, S. Konyagin, and T. Tao, *Large gaps between consecutive prime numbers*, Preprint.
- [4] Hugh L. Montgomery and Robert C. Vaughan, *Multiplicative Number Theory I: Classical Theory*, 2006.
- [5] K. Soundararajan, *Small gaps between prime numbers: the work of Goldston-Pintz-Yildirim*, Bull. Amer. Math. Soc. **44** (Jan. 2007), no. 1, 1–18.
- [6] P. Elliott and H. Halberstam, *A conjecture in prime number theory*, Symp. Math. **4**, 59–72.
- [7] Harold Davenport, *Multiplicative Number Theory*, 1980.
- [8] Daniel A. Goldston, János Pintz, and Cem Y. Yıldırım, *Primes in tuples. I*, Ann. of Math. (2) **170** (2009), no. 2, 819–862. MR2552109 (2011c:11146)
- [9] M. Ram Murty and Alina Cojocaru, *An Introduction to Sieve Methods and Their Applications*, 2006.
- [10] Viggo Brun, *Über das Goldbachsche Gesetz und die Anzahl der Primzahlpaare*, Archiv for Mathematik og Naturvidenskab **B34**.

E-mail address: varul@stanford.edu