ON GAPS BETWEEN PRIMES

EDGAR ASSING

1. Introduction

Many of the most interesting results in (analytic) number theory are in one way and another concerned with the distribution of primes. While our knowledge has vastly improved over the years, there are still many interesting problems and conjectures that remain untouched.

Presumably everybody knows Euclid's proof that there are infinitely many prime numbers. Let us give an alternative proof of this:[1]

Lemma 1.1. *There are infinitely many primes. Even more, the sum* $\sum_p \frac{1}{p}$ *diverges.*

Proof. Suppose that $\sum_p \frac{1}{p}$ is convergent. Then there is $X > 0$ such that

$$\sum_{p>X} \frac{1}{p} < \frac{1}{2}.$$

Multiplying this by an integer N yields

$$\sum_{p>X} \frac{N}{p} < \frac{N}{2}.$$

We define the two numbers

$$N_X = \sharp\{0 < n \leq N : p \mid n \text{ for some } p > X\} \text{ and}$$
$$N^X = \sharp\{0 < n \leq N : (p,n) = 1 \text{ for all } p > X\}.$$

Of course we have

$$N_X + N^X = N \tag{1}$$

The key observation is that

$$N_X \leq \sum_{p>X} \left\lfloor \frac{N}{p} \right\rfloor \leq \frac{N}{2}.$$

In order to estimate N^X we need to make some observations. First suppose $\sharp\{p \leq X\} = k$. Then every integer n contributing to the count N^X can be written as $n = a \cdot b$ for a square-free number a. Obviously there are at most 2^k possibilities

[1]This proof is due to Erdös and can be found in [AZ]. Note that it was Euler who first proved that $\sum_p \frac{1}{p}$ diverges.

to choose square-free numbers with prime divisors $\leq X$. To estimate the number of possible square-parts we simply observe that $b \leq \sqrt{n} \leq \sqrt{N}$. We thus have

$$N^X \leq 2^k \cdot \sqrt{N}.$$

Combining our two estimates yields

$$N_X + N^X < \frac{N}{2} + 2^k\sqrt{N}.$$

For N sufficiently large (for example $N = 2^{2k+2}$) we have $N_X + N^X < N$. This is a contradiction to (1). □

Nowadays we can not only say that there are infinitely many primes, but we also have a good asymptotic understanding of their number:

Theorem 1.2 (Prime Number Theorem)**.** *There is a constant $c > 0$ such that*

$$\pi(X) = \sharp\{p \leq X\} = \mathrm{Li}(X) + O(Xe^{-c\sqrt{\log(X)}}).$$

Remark 1.3. Here we use

$$\mathrm{Li}(X) = \int_0^X \frac{dy}{\log(y)} = \frac{X}{\log(X)} \cdot \left(\sum_{0\leq l<m} l! \log(X)^{-l} + O(\log(X)^{-m}) \right).$$

In particular, we have the maybe more familiar statement

$$\pi(X) = \frac{X}{\log(X)} + O\left(\frac{X}{\log(X)^2}\right).$$

In order to truly understand the distribution of prime numbers the next natural question is how they are distributed in arithmetic progressions. More precisely we can ask for an asymptotic understanding of the counting function

$$\pi(X; q, a) = \sharp\{p \leq X : p \equiv a \bmod q\}.$$

Of course, if $(a, q) \neq 1$, then this is rather un-interesting. On the other hand, it was shown by Siegel, that every arithmetic progression with $(a, q) = 1$ contains infinitely many primes. This can be made quantitative as follows:

Theorem 1.4 (Siegel-Walfisz)**.** *Let $A > 0$ be fixed. Then there is a constant $C = C(A) > 0$ such that*

$$\pi(X; a, q) = \frac{1}{\varphi(q)}\mathrm{Li}(X) + O(Xe^{-C\sqrt{\log(X)}})$$

for all $q \leq \log(X)^A$ and all a with $(a, q) = 1$.

The results mentioned so far are just the starting point and (in some sense) much more can be said. But let us get to some open problems. At the 1912 ICM in Cambridge listed the following four problems as *unattackable* with current technology:[2]

[2]All of these are still open so nothing much has changed...

(1) (Goldbach Conjecture) Every (positive) even integer is the sum of two primes.
(2) (Twin Prime Conjecture) There are infinitely many primes p such that $p+2$ is also prime.
(3) (Legendre's Conjecture) For every $n \in \mathbb{N}$ there is a prime between n^2 and $(n+1)^2$.
(4) There are infinitely many primes of the form n^2+1.

Another very important (and hard) conjecture in the field is due to Hardy and Littlewood. To state it we need the following definition:

Definition 1.1. A set $\mathcal{H} = \{h_1, \ldots, h_k\}$ of distinct non-negative integers is called admissible if, for every prime p, there is an integer a_p such that $a_p \not\equiv h_i \bmod p$ for all $i = 1, \ldots, k$.

Conjecture 1.1 (Prime k-tuple conjecture). *Let $\mathcal{H} = \{h_1, \ldots, h_k\}$ be admissible, Then there are infinitely many integers n such that $n+h_1, \ldots, n+h_k$ are prime.*

Using Selberg's sieve it can be seen relatively easily that

$$\begin{aligned}\pi(X, \mathcal{H}) &= \sharp\{n \le X \colon n + h_i \text{ is prime } \forall i\} \\ &\le 2^k k! \cdot H \cdot \frac{X}{\log(X)^k} + O_{k,\mathcal{H}}\left(\frac{X \log\log(X)}{\log(X)^{k+1}}\right),\end{aligned}$$

where $\mathcal{H}$ is an admissible k-tuple and

$$H = \prod_p \left(1 - \frac{\omega(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-1}.$$

The conjectural asymptotic would be

$$\pi(x, \mathcal{H}) \sim H \cdot \frac{x}{\log(x)^k}.$$

Note that 2-tuple conjecture implies the Twin Prime Conjecture, by taking the tuple $\{0, 2\}$. For $k > 1$ there is no case of the Prime k-tuple conjecture known. However, the slightly weaker problem of showing that small prime gaps exist was rather successfully studied over the years. In order to record some milestone results let is number the primes as

$$p_1 = 2 < p_2 = 3 < p_3 = 5 < p_4 < \ldots < p_n < p_{n-1} < \ldots.$$

We have

- Goldston, Pintz and Yildrim (GPY) shew that
$$\liminf_n \frac{p_{n+1} - p_n}{\log(p_n)} = 0.$$
- Zhang extended this work to
$$\liminf_n (p_{n+1} - p_n) \le 70000000.$$

- Polymath 8(a) pushed Zhang's argument to obtain
$$\liminf_n (p_{n+1} - p_n) \leq 4680.$$
- Maynard introduced exciting new ideas[3] and found that
$$\liminf_n (p_{n+1} - p_n) \leq 600.$$
- Polymath 8b refined these ideas to establish
$$\liminf_n (p_{n+1} - p_n) \leq 246.$$

Note that the GPY-method fails to prove that bounded intervals can contain two or more primes. Indeed, the Goldston, Pintz and Yildrim could only show
$$\liminf_n \frac{p_{n+2} - p_n}{\log(p_n)} = 0.$$
conditional on the Elliott-Halberstam conjecture (see Conjecture 4.1 below). The new ideas by Maynard alluded to above allow us to produce much stronger and unconditional results:

Theorem 1.5 (Maynard 2015). *For $m \in \mathbb{N}$ we have*
$$\liminf_n (p_{n+m} - p_n) \ll m^3 e^{4m}.$$

The goal of this lecture is to reproduce Maynard's argument. Before doing so we will have to cover some preliminaries. While we will take the Prime Number Theorem (PNT) and the Siegel-Walfisz Theorem for granted we will give a proof of the Bombieri-Vinogradov Theorem. As we will see the latter is a crucial ingredient for Maynard's argument.

1.1. **Notation.** Let us summarize some of the most common notations used in the lectures:

- Throughout the letter p will be reserved fro prime numbers. We write $\mathcal{P}$ for the set of all prime numbers and
$$\mathcal{P}^\infty = \{p^\alpha \colon \alpha \in \mathbb{N},\, p \in \mathcal{P}\}.$$
- We write (n, m) for the greatest common divisor (i.e. the gcd) of n and m. The least common multiple (i.e. the lcm) is written as $[n, m] = \frac{nm}{(n,m)}$.
- We write $e(x) = e^{2\pi i x}$. This defines an additive character on $\mathbb{R}/\mathbb{Z}$ and we will often use character orthogonality in the form
$$\frac{1}{d}\sum_{k=1}^{d} e\left(n\frac{k}{d}\right) = \begin{cases} 1 & \text{if } d \mid n, \\ 0 & \text{else.} \end{cases}$$
for $n, d \in \mathbb{N}$.

[3] Similar ideas were independently developed by Tao.

- We use the asymptotic o, O-notation. Furthermore we write $f \ll g$ if $f = O(g)$. We indicate dependencies in the implicit constants by adding subscripts.

2. Arithmetic Functions

A function $f\colon \mathbb{N} \to \mathbb{C}$ is called an **arithmetic function**. Given two arithmetic functions f, g we can define the convolution

$$[f * g](n) = \sum_{d|n} f(d) g\left(\frac{n}{d}\right).$$

We call an arithmetic function f **multiplicative** if

$$f(nm) = f(n) \cdot f(m) \text{ for } n, m \in \mathbb{N} \text{ with } (n, m) = 1. \tag{2}$$

Furthermore, we say that f is **completely multiplicative** if $f(nm) = f(n)f(m)$ for all $n, m \in \mathbb{N}$.

Remark 2.1. Note that a multiplicative $f \neq 0$ must satisfy $f(1) = 1$. Furthermore, it is completely determined by its values on prime powers $f(p^\alpha)$ for $p^\alpha \in \mathcal{P}^\infty$. Similarly, completely multiplicative f are characterized by their values $f(p)$ for $p \in \mathcal{P}$.

Lemma 2.2. *Let f and g be multiplicative, then $f * g$ is multiplicative.*

Proof. Let $n, m \in \mathbb{N}$ with $(n, m) = 1$. Then the divisors $d \mid nm$ are in one to one correspondence with tuples (d_n, d_m) such that $d_n \mid n$ and $d_m \mid m$. We have

$$\begin{aligned}[f * g](nm) &= \sum_{d|nm} f(d) g\left(\frac{nm}{d}\right) = \sum_{d_n|n} \sum_{d_m|m} f(d_n d_m) g\left(\frac{nm}{d_n d_m}\right) \\ &= \sum_{d_n|n} f(d_n) g(n/d_n) \sum_{d_m|m} f(d_m) g(m/d_m) = [f * g](n) \cdot [f * g](m).\end{aligned}$$

In the last step we have exploited multiplicativity of f and g. ☐

Let us give some examples of (more ore less) interesting arithmetic functions:

- The simplest (completely multiplicative) arithmetic function is the constant function $\mathbf{1}(n) = 1$ and the identity $\mathrm{id}(n) = n$.
- We define $\delta_1(n) = 1$ if $n = 1$ and 0 otherwise. This is the identity for the convolution:
$$f * \delta_1 = f.$$
- We also have the divisor function
$$\tau(n) = \sharp\{d \mid n\} = [\mathbf{1} * \mathbf{1}](n).$$
Similarly we have higher divisor functions
$$\tau_k = [\underbrace{\mathbf{1} * \ldots * \mathbf{1}}_{k \text{ times}}](n) = \sharp\{(d_1, \ldots, d_k) \in \mathbb{N}^k \colon d_1 \cdots d_k = n\}.$$

Of course $\tau_2 = \tau$.

- We define $\varphi(n) = \sharp(\mathbb{Z}/n\mathbb{Z})^\times$. By the Chinese Remainder Theorem this is a multiplicative (not completely multiplicative) arithmetic function given by
$$\varphi(p^\alpha) = p^{\alpha-1}(p-1).$$
- A central role in the theory of prime numbers is played by the von Mangoldt function
$$\Lambda(n) = \begin{cases} \log(p) & \text{if } n = p^\alpha \in \mathcal{P}^\infty, \\ 0 & \text{else.} \end{cases}$$
- Finally we have the Möbius function given by
$$\mu(n) = \begin{cases} (-1)^r & \text{if } n = p_1 \cdots p_r \text{ for } r \text{ distinct primes } p_1, \ldots, p_r \in \mathcal{P}, \\ 0 & \text{else.} \end{cases}$$

There are many interesting convolution identities between these arithmetic functions. For example
$$\log(n) = \sum_{p^\alpha || n} \alpha \cdot \log(p) = \sum_{\substack{p \in \mathcal{P} \\ \alpha \in \mathbb{N} \\ p^\alpha | n}} \log(p) = \sum_{d|n} \Lambda(d) = [\Lambda * \mathbf{1}](n). \tag{3}$$
We also claim that
$$\varphi(n) = [\mu * \mathrm{id}](n). \tag{4}$$
Since we know that both sides are multiplicative it suffices to check this on prime powers:
$$\varphi(p^\alpha) = p^{\alpha-1}(p-1) = p^\alpha - p^{\alpha-1} = \sum_{0 \le \beta \le \alpha} \mu(p^\beta) p^{\alpha-\beta} = [\mu * \mathrm{id}](p^\alpha).$$
More importantly we have
$$[\mu * \mathbf{1}](n) = \sum_{d|n} \mu(d) = \delta_1(n) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{else.} \end{cases}.$$
This is also easily checked on prime powers. We deduce the following important technique:

Lemma 2.3 (Möbius Inversion)**.** *For two arithmetic functions f and g the following two relations are equivalent:*
$$g(n) = \sum_{d|n} f(d) \text{ and } f(n) = \sum_{d|n} \mu(d) g(n/d).$$

As an easy application we find that
$$\Lambda(n) = \sum_{d|n} \mu(d) \log(n/d).$$

This indeed follows from (3) and Möbius inversion.

Let us get used to some estimates. First we recall **Mertens' formula**:

$$\prod_{p\le R}\left(1-\frac{1}{p}\right)=\frac{e^{-\gamma}}{\log(R)}\left(1+O\left(\frac{1}{\log(R)}\right)\right). \tag{5}$$

This can be proven in an elementary way and we omit the details. We will however record the following estimate:

Lemma 2.4. *Let $R>0$. For a squarefree number W with prime divisors less than R we have*

$$\sum_{\substack{u\le R\\(u,W)=1}}\frac{\mu(u)^2}{\varphi(u)}\ll\frac{\varphi(W)}{W}\log(R).$$

Proof. We first observe that $\mu(u)^2$ is precisely the characteristic function on squarefree (positive) integers. Thus we can estimate

$$\sum_{\substack{u\le R\\(u,W)=1}}\frac{\mu(u)^2}{\varphi(u)}\le\prod_{\substack{p\le R\\p\nmid W}}\left(1+\frac{1}{\varphi(p)}\right)=\prod_{\substack{p\le R\\p\nmid R}}\frac{p}{p-1}=\frac{\varphi(W)}{W}\left[\prod_{p\le R}(1-\frac{1}{p})\right]^{-1}$$

We conclude by applying Mertens' formula to the remaining product. □

In another direction we will encounter the estimate

Lemma 2.5. *We have*

$$\sum_{n\le R}\tau_k(n)\ll_k R\cdot\log(R)^k.$$

Proof. We estimate

$$\sum_{n\le R}\tau_k(n)\le R\cdot\sum_{n\le R}\frac{\tau_k(n)}{n}\le R\cdot\prod_{p\le R}\left(\sum_{\alpha\ge 0}\tau_k(p^\alpha)p^{-\alpha}\right). \tag{6}$$

At this point we observe that

$$\sum_{\alpha\ge 0}\tau_k(p^\alpha)p^{-\alpha}=\left(\sum_{\alpha\ge 0}p^{-\alpha}\right)^k=\left(1-\frac{1}{p}\right)^{-k}$$

In particular, we have

$$\sum_{n\le R}\tau_k(n)\le R\cdot\left[\prod_{p\le R}\left(1-\frac{1}{p}\right)\right]^{-k}.$$

Thus we are done by Mertens' formula. □

Remark 2.6. With a little trick one can actually improve the estimate above to

$$\sum_{n \leq R} \tau_k(n) \ll_k R \cdot \log(R)^{k-1}.$$

Indeed, we write $\tau_k(n) = [\mathbf{1} * \tau_{k-1}](n)$. Opening the convolution gives

$$\sum_{n \leq R} \tau_k(n) = \sum_{n \leq R} \sum_{d|n} \tau_{k-1}(d) = \sum_{d \leq R} \tau_{k-1}(d) \cdot \left\lfloor \frac{R}{d} \right\rfloor \leq R \sum_{d \leq R} \tau_{k-1}(d) d^{-1}.$$

From here we can continue the argument as above, ultimately saving one logarithm.[4] This is essentially sharp. Indeed one has the asymptotic

$$\sum_{n \leq R} \tau_k(n) = R \cdot P_k(\log(R)) + O(R^{1-\frac{1}{k}}),$$

where P_k is a polynomial of degree $k-1$.

We turn to another very important class of completely functions. These are obtained from characters of $(\mathbb{Z}/q\mathbb{Z})^\times$. More precisely we take a group homomorphism

$$\chi\colon (\mathbb{Z}/q\mathbb{Z})^\times \to S^1$$

and associate the completely multiplicative function arithmetic (also denoted by χ):

$$\chi(n) = \begin{cases} \chi(n \bmod q) & \text{if } (n,q) = 1, \\ 0 & \text{else.} \end{cases}$$

This makes χ a function on $\mathbb{Z}$ (and in particular on $\mathbb{N}$), which is periodic modulo q. We call these functions Dirichlet characters modulo q. We write

$$\chi_0(m) = \delta_1((m,q))$$

for the principal Dirichlet character. (This corresponds to the trivial character of $(\mathbb{Z}/q\mathbb{Z})^\times$.)

Lemma 2.7 (Character Orthogonality)**.** *We have*

$$\frac{1}{\varphi(q)} \sum_{a \bmod q} \chi(a) = \begin{cases} 1 & \text{if } \chi = \chi_0, \\ 0 & \text{else.} \end{cases}$$

and

$$\frac{1}{\varphi(q)} \sum_{\chi \bmod q} \chi(a) = \begin{cases} 1 & \text{if } a \equiv 1 \bmod q, \\ 0 & \text{else.} \end{cases}$$

[4]The same trick works if we want to estimate a multiplicative function f by writing $f = \mathbf{1} * h$ for $h = \mu * f$. One only needs that h is non-negative.

Proof. The two statements are dual to each other. We will prove the first one. The second one is similar. First note that

$$\sum_{a \bmod q} \chi_0(a) = \sum_{\substack{a \bmod q \\ (a,q)=1}} 1 = \varphi(q).$$

On the other hand, if $\chi \neq \chi_0$, then there is b with $(b,q) = 1$ and $\chi(b) \neq 1$. We now write

$$\chi(b) \sum_{a \bmod q} \chi(a) = \sum_{a \bmod q} \chi(ab) = \sum_{a \bmod q} \chi(a).$$

However, this identity can only be true if the a-sum vanishes. □

Definition 2.1. Let χ be a Dirichlet character modulo q. The conductor q^* of χ is defined to be the smallest divisor of q such that we can write $\chi = \chi^* \cdot \chi_0$. where χ^* is a character modulo q^*. If $q^* = q$, then we say that χ is primitive.

Finally, given a Dirichlet character χ modulo q, we define the Gaußsum by

$$\tau(\chi) = \sum_{a \bmod q} \chi(a) e(\frac{a}{q}).$$

Lemma 2.8. *Let χ be a primitive Dirichlet character modulo q. Then we have*

$$\tau(\chi)\tau(\overline{\chi}) = \chi(-1)q.$$

In particular, $|\tau(\chi)| = \sqrt{q}$.

Proof. We write

$$\tau(\chi)\tau(\overline{\chi}) = \sum_{a \bmod q} \sum_{b \bmod q} \chi(a)\overline{\chi(b)} e\left(\frac{a+b}{q}\right) = \sum_{\substack{a \bmod q \\ (a,q)=1}} \sum_{b \bmod q} \chi(a)\overline{\chi(b)} e\left(\frac{a+b}{q}\right).$$

A change of variables in the b-sum yields

$$\tau(\chi)\tau(\overline{\chi}) = \sum_{b \bmod q} \overline{\chi(b)} \underbrace{\sum_{\substack{a \bmod q \\ (a,q)=1}} e\left(\frac{a(b+1)}{q}\right)}_{=c_q(b+1)}.$$

The inner sum is a Ramanujan sum. In general we can observe that

$$\sum_{d|q} c_{q/d}(x) = \sum_{a \bmod q} e(\frac{ax}{q}) = q \cdot \delta_{q|x}.$$

By Möbius inversion we get

$$c_q(x) = \sum_{d|(q,x)} \mu(q/d)d.$$

Inserting this above yields

$$\tau(\chi)\tau(\overline{\chi}) = \sum_{d|q} d\mu(q/d) \sum_{\substack{b \bmod q \\ b \equiv -1 \bmod d}} \overline{\chi(b)}.$$

If χ is primitive, then the remaining b-sum vanishes unless $d = q$. This proves the first statement. Finally,

$$\tau(\overline{\chi}) = \sum_{a \bmod q} \overline{\chi(a)} e(\frac{a}{q}) = \overline{\chi(-1)} \sum_{a \bmod q} \overline{\chi(a)} e(-\frac{a}{q}) = \overline{\chi(-1)\tau(\chi)}.$$

With the statement shown above this implies $|\tau(\chi)|^2 = q$ and the proof is complete. □

3. The large sieve

The main result in this section is the multiplicative large sieve inequality stated in Lemma 3.5. We will approach this in an ad-hoc manner, but we will later see how useful such a technical estimate turns out to be.

We start by establishing the following useful inequality.

Lemma 3.1. *Let $f\colon [\alpha - \frac{1}{2}\delta, \alpha + \frac{1}{2}\delta] \to \mathbb{R}$ be continuously differentiable. Then we have*

$$|f(\alpha)| \leq \delta^{-1} \int_{\alpha - \frac{1}{2}\delta}^{\alpha + \frac{1}{2}\delta} |f(\beta)| d\beta + \frac{1}{2} \int_{\alpha - \frac{1}{2}\delta}^{\alpha + \frac{1}{2}\delta} |f'(\beta)| d\beta.$$

Proof. After a change of variables it is sufficient to show that

$$\left| f\left(\frac{1}{2}\right) \right| \leq \int_0^1 \left(|f(t)| + \frac{1}{2} |f'(t)| \right) dt.$$

To see this we write

$$f\left(\frac{1}{2}\right) = \int_0^1 f(t) dt + \int_0^{\frac{1}{2}} t f'(t) dt + \int_{\frac{1}{2}}^1 (t-1) f'(t) dt.$$

This can be rewritten as

$$f\left(\frac{1}{2}\right) = \int_0^1 f(t) dt + \int_0^1 \rho(t) f'(t) dt.$$

We are done by simply observing that $|\rho(t)| \leq \frac{1}{2}$ for $t \in [0, 1]$. □

With this at hand we can prove the following:

Lemma 3.2. *Suppose $\alpha_1, \ldots, \alpha_R$ are distinct real numbers that are distinct modulo 1. Let $\delta = \min_{i \neq j} \|\alpha_i - \alpha_j\|$.*[5] *Then, for an arithmetic function a supported in*

[5]Here $\|\beta\|$ denotes the distance of β to the nearest integer. This defines a metric on $\mathbb{R}/\mathbb{Z}$.

$M < n \leq M + N$ we have

$$\sum_{r=1}^{R} \left| \sum_n a(n) \cdot e(\alpha_r n) \right|^2 \leq (\delta^{-1} + \pi N) \|a\|_2^2.$$

Proof. We define $S(\alpha) = \sum_n a(n) \cdot e(\alpha n)$. Note that this is essentially a trigonometric polynomial. We apply Lemma 3.1 with $f(\alpha) = S(\alpha)^2$. This gives us

$$|S(\alpha_r)|^2 \leq \delta^{-1} \int_{\alpha - \frac{1}{2}\delta}^{\alpha + \frac{1}{2}\delta} |S(\beta)|^2 d\beta + \int_{\alpha - \frac{1}{2}\delta}^{\alpha + \frac{1}{2}\delta} |S(\beta) \cdot S'(\beta)| d\beta.$$

Summing this over R and recalling the definition of δ allows us to estimate

$$\sum_{r=1}^{R} |S(\alpha_r)|^2 \leq \delta^{-1} \int_0^1 |S(\beta)|^2 d\beta + \int_0^1 |S(\beta) \cdot S'(\beta)| d\beta. \tag{7}$$

We only have to estimate the right hand side. First consider

$$\int_0^1 |S(\beta)|^2 d\beta = \sum_{n_1, n_2} a(n_1)\overline{a(n_1)} \int_0^1 e((n_1 - n_2)\beta) d\beta = \sum_n |a(n)|^2 = \|a\|_2^2.$$

Next we note that

$$S'(\beta) = 2\pi \sum_n n a(n) \cdot e(n\beta).$$

Thus by Cauchy-Schwarz and (7) we have

$$\sum_{r=1}^{R} |S(\alpha_r)|^2 \leq \delta^{-1} \|a\|_2^2 + 2\pi \|a\|_2 \cdot \left(\sum_n n^2 |a(n)|^2 \right)^{\frac{1}{2}}.$$

For $M = -\lfloor \frac{1}{2} N \rfloor$ we have $n^2 \leq \frac{1}{4} N^2$ so that we are done in this case. By a simple shifting argument it is easy to deduce the case of general M. □

Remark 3.3. In the set up of Lemma 3.2 we can take $R = 1$ and $a_n = e(-n\alpha_1)$. Then $\|a\|_2^2 = N$ and $S(\alpha_1) = N$. With this choice we have

$$|S(\alpha_1)|^2 = N \cdot \|a\|_2^2.$$

On the other hand we can compute

$$\int_0^1 \sum_{r=1}^{R} |S(\alpha + \alpha_r)|^2 d\alpha = R \cdot \|a\|_2^2.$$

In particular, there is $\alpha_0 \in [0, 1]$ with

$$\sum_{r=1}^{R} |S(\alpha + \alpha_r)|^2 \geq R \cdot \|a\|_2^2.$$

We can take $\alpha_r = \frac{r}{R}$ and $\alpha'_r = \alpha_0 + \alpha_r$. These points have distance $\delta = R^{-1}$. For this choice of points we have the lower bound

$$\sum_{r=1}^{R} |S(\alpha'_r)|^2 \geq \delta^{-1} \cdot \|a\|_2^2.$$

We conclude that, if we are looking for an estimate of the form

$$\sum_{r=1}^{R} |S(\alpha_r)|^2 \leq \Delta(N, \delta) \cdot \|a\|_2^2$$

which is uniform in all parameters, then we must have

$$\Delta(N, \delta) \geq \max(N, \delta^{-1}).$$

Thus, our large sieve estimate is close to optimal in many situations. Note that it can be slightly improved, but this is irrelevant for our purposes.

More important for our needs is the following special case.

Corollary 3.4. *For any arithmetic function a supported in $M \leq n < M + N$ we have*

$$\sum_{q \leq Q} \sum_{\substack{a \bmod q \\ (a,q)=1}} \left| \sum_{M < n \leq M+N} a(n) e\left(\frac{an}{q}\right) \right|^2 \leq (Q^2 + \pi N) \|a\|_2^2.$$

Proof. This follows from Lemma 3.2 applied to the numbers

$$\{\alpha_1, \ldots, \alpha_R\} = \{\frac{a}{q} \colon q \leq Q,\ 1 \leq a < q \text{ with } (a,q) = 1\}$$

It is easy to see that in this case we can use $\delta = Q^2$. Indeed

$$\left\| \frac{a_1}{q_1} - \frac{a_2}{q_2} \right\| = \left\| \frac{a_1 q_2 - a_2 q_1}{q_1 q_2} \right\| \geq (q_1 q_2)^{-1}.$$

This completes the proof. ☐

Lemma 3.5. *For any arithmetic function a supported in $M < n \leq M + N$ we have*

$$\sum_{q \leq Q} \frac{q}{\varphi(q)} \sum_{\substack{\chi \bmod q \\ \text{primitive}}} \left| \sum_{M < n \leq M+N} a(n) \chi(n) \right|^2 \leq (Q^2 + \pi N) \cdot \|a\|_2^2.$$

Proof. Let χ be a primitive Dirichlet character modulo q. Then, for any n, we have

$$\tau(\chi) \overline{\chi}(n) = \sum_{a \bmod q} \chi(a) e\left(\frac{an}{q}\right).$$

When $(n,q)=1$, then this follows from the definition of the Gauß sum and a simple change of variables. When $(n,q)\neq 1$, we leave it as an exercise to check that both sides are actually zero.

Using this we can write

$$\sum_{M\leq n<M+N} a(n)\overline{\chi}(n)=\frac{1}{\tau(\chi)}\sum_{b \bmod q}\chi(b)\underbrace{\sum_{M\leq n<M+N} a(n)\cdot e\left(\frac{bn}{q}\right)}_{=S(b/q)}.$$

Recall that, since χ is primitive, we have $|\tau(\chi)|=\sqrt{q}$ by Lemma 2.8. We get

$$\left|\sum_{M\leq n<M+N} a(n)\overline{\chi}(n)\right|^2=\frac{1}{q}\left|\sum_{b \bmod q}\chi(b)S\left(\frac{b}{q}\right)\right|^2.$$

Summing both sides over primitive characters χ modulo q and over $q\leq Q$ (with weight $q/\varphi(q)$ yields

$$\begin{aligned}\sum_{q\leq Q}\frac{q}{\varphi(q)}\sum_{\substack{\chi \bmod q\\ \text{primitive}}}\left|\sum_{M\leq n<M+N} a(n)\overline{\chi}(n)\right|^2&=\sum_{q\leq Q}\frac{1}{\varphi(q)}\sum_{\substack{\chi \bmod q\\ \text{primitive}}}\left|\sum_{b \bmod q}\chi(b)S\left(\frac{b}{q}\right)\right|^2\\ &\leq\sum_{q\leq Q}\frac{1}{\varphi(q)}\sum_{\chi \bmod q}\left|\sum_{b \bmod q}\chi(b)S\left(\frac{b}{q}\right)\right|^2\end{aligned}$$

At this point we open the square on the right hand side and execute the χ-sum:

$$\begin{aligned}\sum_{\chi \bmod q}\left|\sum_{b \bmod q}\chi(b)S\left(\frac{b}{q}\right)\right|^2&=\sum_{\chi \bmod q}\sum_{b_1,b_2 \bmod q}\chi(b_1)\overline{\chi}(b_2)S\left(\frac{b_1}{q}\right)\overline{S\left(\frac{b_2}{q}\right)}\\ &=\sum_{\substack{b_1,b_2 \bmod q\\ (b_1b_2,q)=1}}S\left(\frac{b_1}{q}\right)\overline{S\left(\frac{b_2}{q}\right)}\sum_{\chi \bmod q}\chi(b_1)\overline{\chi}(b_2)\\ &=\varphi(q)\sum_{\substack{b \bmod q\\ (b,q)=1}}\left|S\left(\frac{b}{q}\right)\right|^2.\end{aligned}$$

Thus we have seen that

$$\sum_{q\leq Q}\frac{q}{\varphi(q)}\sum_{\substack{\chi \bmod q\\ \text{primitive}}}\left|\sum_{M\leq n<M+N} a(n)\overline{\chi}(n)\right|^2\leq\sum_{q\leq Q}\sum_{\substack{b \bmod q\\ (b,q)=1}}\left|\sum_{M\leq n<M+N} a(n)\cdot e\left(\frac{bn}{q}\right)\right|^2.$$

We are done after applying Corollary 3.4. □

4. The Bombieri-Vinogradov Theorem

In the introduction we already mentioned the Siegel-Walfisz Theorem about primes in arithmetic progressions. One problem with this result is that it only works for relatively short arithmetic progressions. However, in practice one often faces the problem of having to deal with longer arithmetic progressions. Instead of looking to extend the range of the modulus in the Siegel-Walfisz Theorem, which is very hard, one can often work with certain average statements. This motivates us to make the following definition:

Definition 4.1 (Level of Distribution)**.** We say that the primes have level of distribution $\theta > 0$, if for every $A > 0$, we have

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

Conjecture 4.1 (Elliott-Halberstam)**.** *The primes have level of distribution θ for every $\theta < 1$.*

While this conjecture is out of reach of current technology one can go essentially half way. This the consequence of the classical Bombieri-Vinogradov Theorem, which we want to prove in this section:[6]

Theorem 4.1 (Bombieri-Vinogradov)**.** *The primes have level of distribution θ for all $\theta < \frac{1}{2}$. More precisely we will show that for any $A \geq 0$ there is $B = B(A)$ such that*

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

as long as $Q \leq x^{\frac{1}{2}} \log(x)^{-B}$.

Remark 4.2. Recent developments in (analytic) number theory allow one to establish level of distribution beyond $\frac{1}{2}$. This comes at the cost that one has to restrict the range of q in other ways. Results in this direction are very deep and have many interesting applications.

Let us write

$$D_f(x; q, a) = \sum_{\substack{n \leq x \\ n \equiv a \bmod q}} f(n) - \frac{1}{\varphi(q)} \sum_{\substack{n \leq x \\ (n,q)=1}} f(n).$$

We will start by showing that in order to prove the Bombieri-Vinogradov Theorem it is sufficient to study $D_{\Lambda}(x, q, a)$.

[6] We follow the argument given in [IK, Sectio 17.2 and 17.3].

Lemma 4.3. *The Bombieri-Vinogradov Theorem follows from the estimate*

$$\sum_{q\leq Q}\max_{1\leq t\leq x}\max_{(a,q)=1}|D_\Lambda(t;q,a)| \ll_A \frac{x}{\log(x)^A} \tag{8}$$

for $Q \leq x^{\frac{1}{2}}\log(x)^{-B}$.

Passing between $\pi(x)$ resp. $\pi(x;q,a)$ and

$$\psi(x)=\sum_{n\leq x}\Lambda(n) \text{ resp. } \psi(x;q,a)=\sum_{\substack{n\leq x\\ n\equiv a \bmod q}}\Lambda(n)$$

is a rather standard procedure. For completeness we provide the details.

Proof. We define

$$f_a(n;q)=\mathbb{1}_{(a+q\mathbb{Z})\cap\mathcal{P}}(n)-\frac{1}{\varphi(q)}\mathbb{1}_{\mathcal{P}}(n).$$

In particular, we have

$$\pi(x;q,a)-\frac{\pi(x)}{\varphi(q)}=\sum_{n\leq x}f_a(n;q).$$

At this point we want to include a logarithm. To do so we need **partial summation**:[7]

$$\sum_{n\leq x}f(n)g(n)=g(x)\sum_{n\leq x}f(n)-\int_1^x\left(\sum_{n\leq t}f(n)\right)g'(t)dt, \tag{9}$$

for continuously differentiable g. We apply this with $g(n)=\frac{1}{\log(n)}$ and $f(n)=f_a(n;q)\log(n)$ and obtain

$$D_{\mathbb{1}_{\mathcal{P}}}(x,q,a)=\log(x)^{-1}\sum_{n\leq x}f_a(n;q)\log(n)-\int_1^x\left(\sum_{n\leq t}f_a(n;q)\log(n)\right)\frac{dt}{t\log(t)^2}.$$

We only need to account for prime powers and divisors of q:

$$\begin{aligned}\sum_{n\leq t}f_a(n;q)\log(n) &= D_\Lambda(x,q,a)-\sum_{k\geq 2}\left(\sum_{\substack{p^k\leq x\\ p^k\equiv a \bmod q}}\log(p)-\frac{1}{\varphi(q)}\sum_{p^k\leq x}\log(p)\right)\\ &\quad+O\left(\frac{\tau(q)}{\varphi(q)}\log(q)\right)\\ &= D_\Lambda(x,q,a)+O\left(x^{\frac{1}{2}}\log(x)^2+\frac{\tau(q)}{\varphi(q)}\log(q)\right).\end{aligned}$$

[7] This is an easy consequence of partial integration and an indispensable tool for every analytic number theorist.

This is a very wasteful estimate, but good enough for our purposes. So far we have seen that

$$\pi(x;q,a) - \frac{\pi(x)}{\varphi(q)} = \log(x)^{-1} D_\Lambda(x,q,a) - \int_1^x D_\Lambda(t,q,a)\frac{dt}{t\log(t)^2} + O\left(x^{\frac{1}{2}}\log(x) + \frac{\tau(q)\log(q)}{\varphi(q)}\log(x)\right).$$

We immediately get the bound

$$\max_{(a,q)=1}\left|\pi(x;q,a) - \frac{\pi(x)}{\varphi(q)}\right| \ll \max_{t\le x}\max_{(a,q)=1}|D_\Lambda(t,q,a)| + x^{\frac{1}{2}}\log(x).$$

Summing over q, applying (8) and adjusting $B = B(A)$ produces the claim of the Bombieri-Vinogradov Theorem. □

We return to the general setting studying $D_f(x,q,a)$ for a wide class of arithmetic functions f. We start with a little preparatory lemma:

Lemma 4.4. *Let β be arithmetic function supported on $1 \le n \le N$ such that*

$$|D_\beta(N;q,a)| \le N^{\frac{1}{2}}\Delta^9 \cdot \|\beta\|_2,$$

for some $0 < \Delta \le 1$ and all $(a,q) = 1$. Then, for a non-trivial character χ modulo r and $s \in \mathbb{N}$ we have

$$\left|\sum_{(n,s)=1}\beta(n)\chi(n)\right| \le N^{\frac{1}{2}}\Delta^3 r \cdot \tau_3(s)\cdot\|\beta\|_2 \qquad (10)$$

Proof. We first remove the condition $(n,s) = 1$ using Möbius inversion:

$$\sum_{(n,s)=1}\beta(n)\chi(n) = \sum_{k|s}\mu(k)\sum_{n\equiv 0 \bmod k}\beta(n)\chi(n).$$

Now we will fix a parameter K and split the k-sum accordingly:

$$\sum_{(n,s)=1}\beta(n)\chi(n) = \sum_{\substack{k|s\\k\le K}}\mu(k)\sum_{l|k}\mu(l)\sum_{(n,l)=1}\beta(n)\chi(n) + \sum_{\substack{k|s\\k>K}}\mu(k)\sum_{n\equiv 0 \bmod k}\beta(n)\chi(n).$$

By [HLP, Theorem 7] we estimate the second sum by

$$\sum_{\substack{k|s\\k>K}}\mu(k)\sum_{n\equiv 0 \bmod k}\beta(n)\chi(n) \ll \|\beta\|_2\cdot N^{\frac{1}{2}}\sum_{\substack{k|s\\k>K}}k^{-\frac{1}{2}} \le N^{\frac{1}{2}}K^{-\frac{1}{2}}\tau(s)\cdot\|\beta\|_2.$$

To estimate the first sum we will use our assumption on β. However, we first must reinsert a co-primality condition. This is done by inclusion-exclusion:

$$\sum_{\substack{k|s\\k\le K}}\mu(k)\sum_{l|k}\mu(l)\sum_{(n,l)=1}\beta(n)\chi(n) = \sum_{\substack{k|s\\k\le K}}\mu(k)\sum_{l|k}\mu(l)\sum_{\substack{a \bmod rl\\(a,rl)=1}}\chi(a)D_\beta(N;rl,a)$$

Note that we have artificially inserted the leading term in the definition of $D_\beta(N; rl, a)$. This is possible because χ is non-trivial. Estimating this in view of our assumption yields

$$\sum_{\substack{k|s \\ k\le K}} \mu(k) \sum_{l|k} \mu(l) \sum_{(n,l)=1} \beta(n)\chi(n) \ll \|\beta\|_2 \cdot N^{\frac{1}{2}}\Delta^9 \sum_{\substack{k|s \\ k\le K}} |\mu(k)| \sum_{l|k} |\mu(l)|\varphi(rl)$$
$$\ll N^{\frac{1}{2}}\Delta^9 K r \tau_3(s) \cdot \|\beta\|_2.$$

Combining the estimates for both ranges of K gives

$$\sum_{(n,s)=1} \beta(n)\chi(n) \ll N^{\frac{1}{2}} K^{-\frac{1}{2}} \tau(s) \cdot \|\beta\|_2 + N^{\frac{1}{2}}\Delta^9 K r \tau_3(s) \cdot \|\beta\|_2.$$

The result follows after choosing $K = \Delta^{-6}$. ☐

Proposition 4.5. *Let β be an arithmetic function supported on $1 \le n \le N$ such that*

$$|D_\beta(N; q, a)| \le N^{\frac{1}{2}}\Delta^9 \cdot \|\beta\|_2,$$

for some $0 < \Delta \le 1$ and all $(a, q) = 1$. Furthermore, let α be an arithmetic function supported in $1 \le n \le M$. Then we have

$$\sum_{q\le Q} \max_{(a,q)=1} |D_{\alpha*\beta}(MN; q, a)| \ll (\Delta\sqrt{MN} + \sqrt{M} + \sqrt{N} + Q)\log(Q)^4 \cdot \|\alpha\|_2\|\beta\|_2.$$

Proof. We start by opening the definition of $D_{\alpha*\beta}(MN; q, a)$ and applying character orthogonality:

$$\begin{aligned} D_{\alpha*\beta}(MN; q, a) &= \sum_{\substack{n\le MN \\ n\equiv a \bmod q}} [\alpha*\beta](n) - \frac{1}{\varphi(q)} \sum_{\substack{n\le MN \\ (n,q)=1}} [\alpha*\beta](n) \\ &= \frac{1}{\varphi(q)} \sum_{\substack{\chi \bmod q \\ \text{non-trivial}}} \overline{\chi(a)} \sum_{n\le MN} [\alpha*\beta](n)\chi(n) \\ &= \frac{1}{\varphi(q)} \sum_{\substack{\chi \bmod q \\ \text{non-trivial}}} \overline{\chi(a)} \left(\sum_m \alpha(m)\chi(m)\right)\left(\sum_n \beta(n)\chi(n)\right). \end{aligned}$$

In the last step we have simply used the definition of the convolution and the complete multiplicativity of χ. Inserting this in the quantity we are aiming to estimate leaves us to treat

$$\sum_{q\le Q} \frac{1}{\varphi(q)} \sum_{\substack{r|q \\ r\ne 1}} \sum_{\substack{\chi \bmod r \\ \text{primitive}}} \left|\sum_{(m,q)=1} \alpha(m)\chi(m)\right| \cdot \left|\sum_{(n,q)=1} \beta(n)\chi(n)\right|.$$

We fix a parameter R and first treat the part with $r \leq R$ essentially trivially. Indeed in view of our assumption on β we can estimate

$$\sum_{q \leq Q} \frac{1}{\varphi(q)} \sum_{\substack{r \mid q \\ 1 < r \leq R}} \sum_{\substack{\chi \bmod q \\ \text{primitive}}} \underbrace{\left| \sum_{(m,s)=1} \alpha(m)\chi(m) \right|}_{\leq M^{\frac{1}{2}} \|\alpha\|_2} \cdot \underbrace{\left| \sum_{(n,s)=1} \beta(n)\chi(n) \right|}_{\ll N^{\frac{1}{2}} \Delta^3 r \tau_3(s) \cdot \|\beta\|_2}$$
$$\ll \|\alpha\|_2 \|\beta\|_2 \cdot \sqrt{NM} \Delta^3 \sum_{q \leq Q} \frac{\tau_3(q)}{\varphi(q)} \sum_{\substack{r \mid q \\ 1 < r \leq R}} r \ll \|\alpha\|_2 \|\beta\|_2 \cdot \sqrt{NM} \Delta^3 R^2 \log(Q)^4.$$

Here we have estimated $\sum_{\substack{r \mid q \\ r \leq R}} r \ll R^2$ trivially. Estimating the remaining q-sum is a good exercise:

$$\begin{aligned} \sum_{q \leq Q} \frac{\tau_3(q)}{\varphi(q)} &\leq \prod_{p \leq Q} \left(1 + \sum_{k=1}^{\infty} \frac{\tau_3(k)}{p^k (1 - 1/p)} \right) \\ &= \prod_{p \leq Q} \left(1 - (1 - 1/p)^{-1} + (1 - 1/p)^{-1} \sum_{k=0}^{\infty} \frac{\tau_3(k)}{p^k} \right) \\ &= \prod_{p \leq Q} \left(1 - (1 - 1/p)^{-1} + (1 - 1/p)^{-4} \right) \leq \prod_{p \leq Q} (1 - 1/p)^{-4}. \end{aligned}$$

An application of Mertens' formula (5) gives

$$\sum_{q \leq Q} \frac{\tau_3(q)}{\varphi(q)} \ll \log(Q)^4$$

as desired.

For the remaining part (i.e. $r > R$) we want to use the large sieve from Lemma 3.5. We need some preparations. First note that we have $\varphi(q)\varphi(r) \leq$

$\varphi(qr)$.[8] We make the following reformulations:

$$\sum_{q\le Q}\frac{1}{\varphi(q)}\sum_{\substack{r|q\\ r\neq 1}}\sum_{\substack{\chi \bmod r\\ \text{primitive}}}\left|\sum_{(m,q)=1}\alpha(m)\chi(m)\right|\cdot\left|\sum_{(n,q)=1}\beta(n)\chi(n)\right|$$

$$=\sum_{R<r\le Q}\sum_{\substack{q\le Q\\ r|q}}\frac{1}{\varphi(q)}\sum_{\substack{\chi \bmod r\\ \text{primitive}}}\left|\sum_{(m,q)=1}\alpha(m)\chi(m)\right|\cdot\left|\sum_{(n,q)=1}\beta(n)\chi(n)\right|$$

$$\le\sum_{R<r\le Q}\sum_{q\le Q/r}\frac{1}{\varphi(q)\varphi(r)}\sum_{\substack{\chi \bmod r\\ \text{primitive}}}\left|\sum_{(m,q)=1}\alpha(m)\chi(m)\right|\cdot\left|\sum_{(n,q)=1}\beta(n)\chi(n)\right|$$

$$\le\sum_{q\le Q}\frac{1}{\varphi(q)}\sum_{R<r\le Q}\frac{1}{\varphi(r)}\sum_{\substack{\chi \bmod r\\ \text{primitive}}}\left|\sum_{(m,q)=1}\alpha(m)\chi(m)\right|\cdot\left|\sum_{(n,q)=1}\beta(n)\chi(n)\right|$$

At this point we need another trick. We split the r-sum into dyadic pieces. This is we write

$$\sum_{R<r\le Q}=\sum_{R<r\le 2R}+\sum_{2R<r\le 4R}+\ldots=\sum_{R<P\le Q}^{\text{dyadic}}\sum_{P<r\le 2P}.$$

Note that the dyadic P-sum has $\ll\log(Q)$ terms. We can write

$$\sum_{q\le Q}\frac{1}{\varphi(q)}\sum_{\substack{r|q\\ r\neq 1}}\sum_{\substack{\chi \bmod r\\ \text{primitive}}}\left|\sum_{(m,q)=1}\alpha(m)\chi(m)\right|\cdot\left|\sum_{(n,q)=1}\beta(n)\chi(n)\right|$$

$$\le\sum_{q\le Q}\frac{1}{\varphi(q)}\sum_{R<P\le Q}^{\text{dyadic}}\frac{1}{2P}\sum_{P<r\le 2P}\frac{r}{\varphi(r)}\sum_{\substack{\chi \bmod r\\ \text{primitive}}}\left|\sum_{(m,q)=1}\alpha(m)\chi(m)\right|\cdot\left|\sum_{(n,q)=1}\beta(n)\chi(n)\right|$$

$$\le\sum_{q\le Q}\frac{1}{\varphi(q)}\sum_{R<P\le Q}^{\text{dyadic}}\frac{1}{2P}\left(\sum_{P<r\le 2P}\frac{r}{\varphi(r)}\sum_{\substack{\chi \bmod r\\ \text{primitive}}}\left|\sum_{(m,q)=1}\alpha(m)\chi(m)\right|^2\right)^{\frac{1}{2}}$$

$$\cdot\left(\sum_{P<r\le 2P}\frac{r}{\varphi(r)}\sum_{\substack{\chi \bmod r\\ \text{primitive}}}\left|\sum_{(n,q)=1}\beta(n)\chi(n)\right|^2\right)^{\frac{1}{2}}.$$

[8] By multiplicativity it is sufficient to check this on prime powers.

In the last step we have applied Cauchy-Schwarz in order to separate the n and m-sum. Applying the large sieve in both pieces yields

$$\sum_{q\le Q}\frac{1}{\varphi(q)}\sum_{\substack{r|q\\ r\neq 1}}\sum_{\substack{\chi \bmod r\\ \text{primitive}}}\left|\sum_{(m,q)=1}\alpha(m)\chi(m)\right|\cdot\left|\sum_{(n,q)=1}\beta(n)\chi(n)\right|$$
$$\ll \sum_{q\le Q}\frac{1}{\varphi(q)}\sum_{R<P\le Q}^{\text{dyadic}}\frac{1}{P}(P^2+M)^{\frac12}(P^2+N)^{\frac12}\|\alpha\|_2\|\beta\|_2$$
$$\ll (P+M^{\frac12}+N^{\frac12}+(MN)^{\frac12}R^{-1})\log(Q)^2.$$

The result follows with $R=\Delta^{-1}$. □

We are closing in on being able to prove the Bombieri-Vinogradov Theorem. The key is Proposition 4.5. In order to apply it we have to write Λ (at least approximately) as a convolution of two arithmetic functions with the required properties. This can be done using the following curious identity:

Proposition 4.6 (Vaughan's Identity)**.** *Let $z\ge 1$. Then for any $n>z$ we have*

$$\Lambda(n)=\sum_{\substack{b|n\\ b\le z}}\mu(b)\log(\frac{n}{b})-\sum_{\substack{b|n\\ b\le z}}\sum_{\substack{c|\frac{n}{b}\\ c\le z}}\mu(b)\Lambda(c)+\sum_{\substack{b|n\\ b>z}}\sum_{\substack{c|\frac{n}{b}\\ c>z}}\mu(b)\Lambda(c).$$

Proof. Our starting point are the well known identities

$$\Lambda(n)=\sum_{b|n}\mu(b)\log(\frac{n}{b})\text{ and }\log(n)=\sum_{d|n}\Lambda(d).$$

We keep the terms with $b\le z$ in the first identity and transform the rest as follows:

$$\sum_{\substack{b|n\\ b>z}}\mu(b)\log(\frac{n}{b})=\sum_{\substack{b|n\\ b>z}}\sum_{c|\frac{n}{b}}\mu(b)\Lambda(c).$$

Now we keep the part $c>z$ in the c-sum. The rest can be rewritten as

$$\sum_{\substack{b|n\\ b>z}}\sum_{\substack{c|\frac{n}{b}\\ c\le z}}\mu(b)\Lambda(c)=\sum_{b|n}\sum_{\substack{c|\frac{n}{b}\\ c\le z}}\mu(b)\Lambda(c)-\sum_{\substack{b|n\\ b\le z}}\sum_{\substack{c|\frac{n}{b}\\ c\le z}}\mu(b)\Lambda(c)$$
$$=\sum_{\substack{c|n\\ c\le z}}\Lambda(c)\underbrace{\sum_{b|\frac{n}{c}}\mu(b)}_{=0}-\sum_{\substack{b|n\\ b\le z}}\sum_{\substack{c|\frac{n}{b}\\ c\le z}}\mu(b)\Lambda(c)$$
$$=-\sum_{\substack{b|n\\ b\le z}}\sum_{\substack{c|\frac{n}{b}\\ c\le z}}\mu(b)\Lambda(c).$$

Gathering all the pieces concludes the proof. □

Exercise 4.1. *Suppose* $f(t) = \log(t)$*, the we have*

$$|D_f(y;q,a)| \leq 2|f(y)|.$$

Solution: By partial summation we can write

$$D_f(y;q,a) = \left(\sum_{\substack{n\leq y,\\ n\equiv a \bmod q}} 1 - \frac{1}{\varphi(q)} \sum_{\substack{n\leq y\\ (n,q)=1}} 1 \right) \log(y) - \int_1^y \left(\sum_{\substack{n\leq t,\\ n\equiv a \bmod q}} 1 - \frac{1}{\varphi(q)} \sum_{\substack{n\leq t\\ (n,q)=1}} 1 \right) \frac{dt}{t}.$$

By cutting the sums into intervals $[qk+1, q(k+1)]$ we see that only the tail $[q\lfloor \frac{t}{y} \rfloor + 1, t]$ survives. Estimating this contribution trivially completes the argument. □

We are now ready for the proof of the Bombieri-Vinogradov Theorem:

Proof of Theorem 4.1. We first fix $Q \leq x^{\frac{1}{2}} \log(x)^{-B}$, $t \leq x$ and $z = x^{\frac{1}{5}}$.[9] Note that B will be chosen later in terms of A. Finally, recall that we need to bound

$$\sum_{q\leq Q} \max_{(a,q)=1} |D_\Lambda(t;q,a)|.$$

The result will then follow from Lemma 4.3.

First, for $n \leq t$ write

$$\Lambda(n) = \Lambda(n) \cdot \mathbb{1}_{(0,z]}(n) + \Lambda(n) \cdot \mathbb{1}_{(z,t]}(n).$$

We can trivially estimate

$$\sum_{q\leq Q} \max_{(a,q)=1} |D_{\Lambda\cdot \mathbb{1}_{(0,z]}}(t;q,a)| \ll zQ\log(x) \ll x^{\frac{7}{10}} \log(x).$$

In particular, we have seen that

$$\sum_{q\leq Q} \max_{(a,q)=1} |D_\Lambda(t;q,a)| \ll \sum_{q\leq Q} \max_{(a,q)=1} |D_{\Lambda\cdot \mathbb{1}_{(z,t]}}(t;q,a)| + x^{\frac{7}{10}} \log(x)$$

We can now apply Vaughan's identity (i.e. Proposition 4.6) with z as above to write

$$\Lambda \cdot \mathbb{1}_{(z,t]} = \Lambda_1(n) - \Lambda_2(n) + \Lambda_3(n),$$

[9] A lot of other choices for z would work. We fix this one for convenience.

for

$$\Lambda_1(n) = \sum_{\substack{b|n \\ b\leq z}} \mu(b)\log(n/b) = [(\mu\cdot \mathbb{1}_{[1,z]}) * \log](n),$$

$$\Lambda_2(n) = \sum_{\substack{b|n \\ b\leq z}} \sum_{\substack{c|\frac{n}{b} \\ c\leq z}} \mu(b)\Lambda(c) = [(\Lambda\cdot \mathbb{1}_{[1,z]}) * (\mu\cdot \mathbb{1}_{[1,z]}) * 1](n) \text{ and}$$

$$\Lambda_3(n) = \sum_{\substack{b|n \\ b> z}} \sum_{\substack{c|\frac{n}{b} \\ c> z}} \mu(b)\Lambda(c) = [(\Lambda\cdot \mathbb{1}_{(z,\infty)}) * (\mu\cdot \mathbb{1}_{(z,\infty)}) * 1](n).$$

Even though we have not indicated this explicitly the functions Λ_1, Λ_2 and Λ_3 appearing in the decomposition are also restricted to $n \in (z,t]$.

We continue by estimating the contributions of Λ_1 and Λ_2. We first bound

$$\begin{aligned} D_{\Lambda_1}(t;q,a) &= \sum_{\substack{n\leq t \\ n\equiv a \bmod q}} \sum_{\substack{b|n \\ b\leq z}} \mu(b)\log(n/b) - \frac{1}{\varphi(q)} \sum_{\substack{n\leq t \\ (n,q)=1}} \sum_{\substack{b|n \\ b\leq z}} \mu(b)\log(n/b) \\ &= \sum_{\substack{b\leq z \\ (b,q)=1}} \mu(b) \sum_{\substack{m\leq \frac{t}{b} \\ (m,q)=1}} \log(m)\cdot \left(\mathbb{1}_{a+q\mathbb{Z}}(bm) - \frac{1}{\varphi(q)}\right) \\ &= \sum_{\substack{b\leq z \\ (b,q)=1}} \mu(b) D_{\log}(t/b;q,a\overline{b}). \end{aligned}$$

Here $\overline{b}$ is the inverse of b modulo q. Estimating the result using Exercise 4.1 and summing over Q yields

$$\sum_{q\leq Q} \max_{(a,q)=1} |D_{\Lambda_1}(t;q,a)| \ll zQ\log(x) \ll x^{\frac{7}{10}}\log(x).$$

We proceed similar to handle the Λ_2-part. Indeed, we can write

$$D_{\Lambda_2}(t;q,a) = \sum_{\substack{b,c\leq z \\ (bc,q)=1}} \mu(b)\Lambda(c) D_1\left(\frac{t}{bc};q,a\overline{bc}\right).$$

An easy observation, which was also used in the solution to Exercise 4.1, yields $D_1(t/(bc);q,a) \ll 1$. We obtain

$$D_{\Lambda_2}(t;q,a) \ll z^2\log(z).$$

Summing everything over $q \leq Q$ gives

$$\sum_{q\leq Q} \max_{(a,q)=1} |D_{\Lambda_2}(t;q,a)| \ll z^2 Q\log(x) \ll x^{\frac{9}{10}}\log(x).$$

So far we have seen that

$$\sum_{q\le Q}\max_{(a,q)=1}|D_\Lambda(t;q,a)| \ll \sum_{q\le Q}\max_{(a,q)=1}|D_{\Lambda_3}(t;q,a)| + x^{\frac{9}{10}}\log(x)$$

The remaining task is to handle the Λ_3-part. This requires some further preparation.

We fix $\delta = \log(x)^{-C}$. Here C is a suitable parameter that can be chosen in terms of A later on. Put $\lambda = 1+\delta$. We define

$$\alpha_L = [(\mu\cdot \mathbb{1}_{(z,\infty)}) * 1]\cdot \mathbb{1}_{(L,\lambda L]}$$

and similarly

$$\beta_M = \Lambda\cdot \mathbb{1}_{(M,\lambda M]}.$$

Our goal is to decompose

$$\Lambda_3 \approx \sum_{L,M}\alpha_L * \alpha_M.$$

We first note that since $z<n\le t$ we can assume that $LM\le t$ but we also have the individual restrictions $z<L, M<t/z$. We have written $\approx$ instead of $=$ above, because we can do this up to some overlap close to t. Indeed, similar to the dyadic decompositions used earlier we can take L and M of the shape λ^j for $j\in\mathbb{N}$. We get

$$D_{\Lambda_3}(t;q,a) = \sum_{\substack{z<L,M<t/z\\ LM\le t}}^{\lambda-\text{adic}} D_{\alpha_L*\beta_M}(t;q,a) + O(q^{-1}\delta x\log(x))$$

The error comes from estimating the overlaps at the ends of the ranges trivially. We have used the non-standard notation λ-adic to indicate that L and M are of the form λ^j. Note that the L-sum and the M-sum have a combined length $\ll \delta^2$. We obtain

$$\sum_{q\le Q}\max_{(a,q)=1}|D_\Lambda(t;q,a)| \ll \sum_{\substack{z<L,M<t/z\\ LM\le t}}^{\lambda-\text{adic}}\sum_{q\le Q}\max_{(a,q)=1}|D_{\alpha_L*\beta_M}(t;q,a)| + \delta x\log(x)^2 + x^{\frac{9}{10}}\log(x).$$

We are ready to apply Proposition 4.5 to each of the remaining pieces. Note that by the Siegel-Walfisz Theorem β_M satisfies (4.5) with $\Delta = \log(x)^{-D}$ for arbitrary D.[10] By the prime number theorem we can estimate

$$\|\beta_M\|_2^2 \ll \delta M\log(x)^2$$

[10] Here we are omitting implicit constants that can depend on D.

On the other hand we have

$$\|\alpha_L\|_2^2 = \sum_{L<n\leq\lambda L}\left|\sum_{\substack{l|n\\ l>z}}\mu(l)\right|^2 \leq \sum_{L<n\leq\lambda L}\tau(n)^2 \ll L\log(x)^3.$$

By Proposition 4.5 we obtain

$$\sum_{q\leq Q}\max_{(a,q)=1}|D_{\alpha_L*\beta_M}(t;q,a)| \ll (\Delta\sqrt{LM}+\sqrt{L}+\sqrt{M}+Q)\delta^{\frac{1}{2}}\sqrt{LM}\log(x)^7$$
$$\ll (\Delta + z^{-\frac{1}{2}}+\log(x)^{-B})x\log(x)^7.$$

Where we have used our restrictions on L and M. Summing this up and recalling that $\delta = \log(x)^{-C}$ gives

$$\sum_{q\leq Q}\max_{(a,q)=1}|D_{\Lambda}(t;q,a)| \ll (\log(x)^{-D}+x^{-\frac{1}{10}}+\log(x)^{-B})x\log(x)^{2C+7}+x\log(x)^{2-C}+x^{\frac{9}{10}}\log(x).$$

We easily conclude by taking B, C and D appropriately. □

5. Bounded gaps in primes

We finally turn to the proof of Maynard's theorem. Among other things our goal is to proof Theorem 1.5 stated in the introduction. To do so we will closely follow Maynard's original argument from [Ma]. Note that a similar result was independently proven by Tao. Furthermore, Maynard's results have been refined. The state of the art and essentially the limit of the method was achieved in an impressive Polymath Project [Po].

5.1. **The Set-Up.** Let $\mathcal{H} = \{h_1,\dots,h_k\}$ be admissible. As in the GPY-method we consider the sum

$$S(N,\rho) = \sum_{N\leq n<2N}\left(\sum_{i=1}^{k}\mathbb{1}_{\mathcal{P}}(n+h_i)-\rho\right)\omega_n$$

for $\rho>0$ and non-negative weights ω_n. If we can show that $S(N,\rho)>0$, then there is $N\leq n_0<2N$ such that

$$\sum_{i=1}^{k}\mathbb{1}_{\mathcal{P}}(n+h_i)>\rho.$$

We conclude that at least $\lfloor\rho+1\rfloor$ of the numbers

$$n_0+h_1,\dots,n_0+h_k$$

are prime. The new idea that appears in [Ma] is the choice of the sieve weights:

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

The precise choice of $\lambda_{d_1,\ldots,d_k}$ will be discussed below.

We will now set up some notation, which we will use for the rest of this section. Define

$$D_0 = \log\log\log(N) \text{ and } W = \prod_{p \leq D_0} p.$$

Note that $W \ll (\log\log(N))^2$. Let θ be the level of distribution of the primes and set

$$R = N^{\frac{\theta}{2}-\delta}, \text{ for } \delta > 0.$$

We choose v_0 so that $v_0 + h_i$ is co-prime to W for all $i = 1, \ldots, k$. To see that this is possible one uses that $\mathcal{H}$ is admissible and the Chinese Remainder Theorem. We will now slightly modify $S(N, \rho)$. Indeed we will actually estimate

$$S_{v_0}(N,\rho) = \sum_{\substack{N \leq n < 2N, \\ n \equiv v_0 \bmod W}} \left(\sum_{i=1}^{k} \mathbb{1}_{\mathcal{P}}(n+h_i) - \rho \right) \omega_n = S_2 - \rho \cdot S_1 \tag{11}$$

for

$$S_1 = \sum_{\substack{N \leq n < 2N, \\ n \equiv v_0 \bmod W}} \omega_n \text{ and}$$

$$S_2 = \sum_{\substack{N \leq n < 2N, \\ n \equiv v_0 \bmod W}} \left(\sum_{i=1}^{k} \mathbb{1}_{\mathcal{P}}(n+h_i) \right) \omega_n.$$

Furthermore, we define ω_n as in (5.1) with

$$\lambda_{d_1,\ldots,d_k} = \left(\prod_{i=1}^{k} \mu(d_i) d_i \right) \cdot \sum_{\substack{r_1,\ldots,r_k, \\ d_i | r_i, \\ (r_i, W)=1}} \frac{\mu(\prod_{i=1}^{k} r_i)^2}{\prod_{i=1}^{r} \varphi(r_i)} F\left(\frac{\log(r_1)}{\log(R)}, \ldots, \frac{\log(r_k)}{\log(R)} \right). \tag{12}$$

Here F is a fixed smooth function. Note that if $(\prod_{i=1}^{k} d_i, W) \neq 1$, then the $r_1, \ldots, r_k$-sum is empty and we define the $w_n = 0$ in this case. We will usually write $d = \prod_{i=1}^{k} d_i$. Let us record that the support of the weights $\lambda_{d_1,\ldots,d_k}$ is restricted to $d_1, \ldots, d_k$ with the properties:

- $d \leq R$;
- $(d, W) = 1$; and
- $\mu(d)^2 = 1$ (in other words $(d_i, d_j) = 1$ for $i \neq j$ and d_i- square free).

We will encounter the integrals

$$I_k(F) = \int_0^1 \cdots \int_0^1 F(t_1, \ldots, t_k)^2 dt_1 \cdots dt_k \text{ and} \tag{13}$$

$$J_k^{(m)}(F) = \int_0^1 \cdots \int_0^1 \left(\int_0^1 F(t_1, \ldots, t_k) dt_m \right)^2 dt_1 \cdots dt_{m-1} dt_{m+1} \cdots dt_k. \tag{14}$$

5.2. **Sieve Manipulations.** In this subsection we will prove asymptotic estimates for the sums S_1 and S_2. The arguments are generalizations of the original GPY-arguments.

We define

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

and

$$y_{\max} = \sup_{r_1,\ldots,r_k} |y_{r_1,\ldots,r_k}|.$$

Lemma 5.1. *We have*

$$S_1 = \frac{N}{W} \sum_{r_1,\ldots,r_k} \frac{y_{r_1,\ldots,r_k}^2}{\prod_{i=1}^k \varphi(r_i)} + O\left(\frac{y_{max}^2 \varphi(W)^k N \log(R)^k}{W^{k+1} D_0} \right).$$

Proof. We start from the definition of S_1, open the square and exchange summation:

$$\begin{aligned} S_1 &= \sum_{\substack{N \le n < 2N, \\ n \equiv v_0 \bmod W}} \left(\sum_{d_i | n+h_i} \lambda_{d_1,\ldots,d_k} \right)^2 \\ &= \sum_{\substack{d_1,\ldots,d_k \\ e_1,\ldots,e_k}} \lambda_{d_1,\ldots,d_k} \lambda_{e_1,\ldots,e_k} \sum_{\substack{N \le n < 2N, \\ n \equiv v_0 \bmod W, \\ [d_i,e_i] | n+h_i}} 1. \end{aligned}$$

We first suppose that the numbers $W, [d_1, e_1], \ldots, [d_k, e_k]$ are pairwise co-prime. In this case we define

$$q = W \cdot \prod_{i=1}^k [d_i, e_i].$$

According to the Chinese Remainder Theorem there is γ such that

$$\sum_{\substack{N \le n < 2N, \\ n \equiv v_0 \bmod W, \\ [d_i,e_i] | n+h_i}} 1 = \sum_{\substack{N \le n < 2N, \\ n \equiv \gamma \bmod q}} 1 = \frac{N}{q} + O(1).$$

If the integers are not pairwise co-prime, then the inner sum is empty. Recall that $\lambda_{d_1,\ldots,d_k}$ is supported on tuple $(d_1, \ldots, d_k)$ with $(d_i, d_j) = 1$ for $i \neq j$ (i.e. pairwise co-prime) and $(d_i, W) = 1$. Thus the condition that $W, [d_1, e_1], \dots, [d_k, e_k]$ are pairwise co-prime boils down to $(d_i, e_j) = 1$ for $i \neq j$. We arrive at

$$S_1 = \frac{N}{W} \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k \\ (d_i,e_j)=1 \text{ for } i\neq j}} \frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{[d_1,e_1]\cdots[d_k,e_k]} + O\left(\sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k \\ (d_i,e_j)=1 \text{ for } i\neq j}} |\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}| \right).$$

We first look at the error. Put

$$\lambda_{\max} = \sup_{d_1,\dots,d_k} |\lambda_{d_1,\dots,d_k}|$$

and observe that $\lambda_{d_1,\dots,d_k}$ are non-zero only when $d_1 \cdots d_k < R$. We can therefore estimate

$$\sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k \\ (d_i,e_j)=1 \text{ for } i\neq j}} |\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}| \ll \lambda_{\max}^2 \left(\sum_{d<R} \tau_k(d) \right)^2 \ll \lambda_{\max}^2 R^2 \log(R)^{2k}.$$

This turns out to be an acceptable bound.

We return to the main term. The next step is to decouple d_i and e_i appearing together in the smallest common multiple $[d_i, e_i]$. To do so we use the basic identity

$$\frac{1}{[d_i,e_i]} = \frac{(d_i,e_i)}{d_i \cdot e_i} = \frac{1}{d_i \cdot e_i} \sum_{u_i|(d_i,e_i)} \varphi(u_i).$$

With this at hand we can write

$$\frac{N}{W} \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k \\ (d_i,e_j)=1 \text{ for } i\neq j}} \frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{[d_1,e_1]\cdots[d_k,e_k]}$$
$$= \frac{N}{W} \sum_{u_1,\dots,u_k} \varphi(u_1)\cdots\varphi(u_k) \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k, \\ u_i|(d_i,e_i) \\ (d_i,e_j)=1 \text{ for } i\neq j}} \frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{d_1\cdots d_k \cdot e_1 \cdots e_k}.$$

The remaining conditions $(d_i, e_j) = 1$ can be detected by

$$\delta_{(d_i,e_j)=1} = \sum_{s_{i,j}|(d_i,e_j)} \mu(s_{i,j}).$$

We arrive at

$$\frac{N}{W}\sum_{\substack{d_1,\dots,d_k\\ e_1,\dots,e_k\\ (d_i,e_j)=1 \text{ for } i\neq j}}\frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{[d_1,e_1]\cdots[d_k,e_k]}$$

$$=\frac{N}{W}\sum_{u_1,\dots,u_k}\varphi(u_1)\cdots\varphi(u_k)\sum_{s_{i,j},\, i\neq j}\left(\prod_{\substack{1\le i,j\le k,\\ i\neq j}}\mu(s_{i,j})\right)\cdot\sum_{\substack{d_1,\dots,d_k\\ e_1,\dots,e_k,\\ u_i|(d_i,e_i)\\ s_{i,j}|(d_i,e_j) \text{ for } i\neq j}}\frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{d_1\cdots d_k\cdot e_1\cdots e_k}.$$

Due to the support of the weights $\lambda_{d_1,\dots,d_k}$ we can make the following simplifying assumptions. First, we can assume that $s_{i,j}$ is co-prime to u_i and u_j. Similarly we can restrict our attention to $s_{i,j}$ that are co-prime to $s_{i,a}$ and $s_{b,j}$ for $a\neq j$ and $b\neq i$. We will always make these assumptions in the $s_{*,*}$-sums below.

We now make the following change of variables:

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

It turns out that we can invert this. Indeed, for $d_1\cdots d_k$ square free we have

$$\sum_{\substack{r_1,\dots,r_k\\ d_i|r_i}}\frac{y_{r_1,\dots,r_k}}{\varphi(r_1)\cdots\varphi(r_k)}=\sum_{\substack{r_1,\dots,r_k\\ d_i|r_i}}\left(\prod_{i=1}^k\mu(r_i)\right)\cdot\sum_{\substack{e_1,\dots,e_k\\ r_i|e_i}}\frac{\lambda_{e_1,\dots,e_k}}{e_1\cdots e_k}$$
$$=\sum_{e_1,\dots,e_k}\frac{\lambda_{e_1,\dots,e_k}}{e_1\cdots e_k}\sum_{\substack{r_1,\dots,r_k\\ d_i|r_i\\ r_i|e_i}}\prod_{i=1}^k\mu(r_i)=\frac{\lambda_{d_1,\dots,d_k}}{\mu(d_1)\cdots\mu(d_k)d_1\cdots d_k} \tag{15}$$

by Möbius inversion. We put

$$y_{\max}=\sup_{r_1,\dots,r_k}|y_{r_1,\dots,r_k}|.$$

Recall that for square free d we have[11]

$$\frac{d}{\varphi(d)} = \sum_{e|d} \frac{1}{\varphi(e)}.$$

Set $r' = \prod_{i=1}^{k} \frac{r_i}{d_i}$. We compute the relation between $\lambda_{\max}$ and $y_{\max}$:

$$\begin{aligned}
\lambda_{\max} &\leq \sup_{\substack{d_1,\dots,d_k \\ \prod_{i=1}^k d_i\, \square\text{-free}}} y_{\max} \left(\prod_{i=1}^{k} d_i \right) \cdot \sum_{\substack{r_1,\dots,r_k \\ d_i | r_i \\ r_1 \cdots r_k < R \\ r_1 \cdots r_k\, \square\text{-free}}} \left(\prod_{i=1}^{k} \frac{\mu(r_i)^2}{\varphi(r_i)} \right) \\
&\leq y_{\max} \sup_{\substack{d_1,\dots,d_k \\ \prod_{i=1}^k d_i\, \square\text{-free}}} \left(\prod_{i=1}^{k} \frac{d_i}{\varphi(d_i)} \right) \sum_{\substack{r' \leq R/\prod_{i=1}^k d_i \\ (r', d_1 \cdots d_k) = 1}} \frac{\mu(r')\tau_k(r')}{\varphi(r')} \\
&\leq y_{\max} \sup_{d_1,\dots,d_k} \sum_{d | \prod_{i=1}^k d_i} \frac{\mu(d)^2}{\varphi(d)} \sum_{\substack{r' \leq R/\prod_{i=1}^k d_i \\ (r', d_1 \cdots d_k) = 1}} \frac{\mu(r')\tau_k(r')}{\varphi(r')} \\
&\leq y_{\max} \cdot \sum_{u < R} \frac{\mu(u)^2 \tau_k(u)}{\varphi(u)} \ll y_{\max} \cdot \log(R)^k.
\end{aligned}$$

In particular, we can replace the error term $O(\lambda_{\max}^2 R^2 \log(R)^{2k})$ by $O(y_{\max}^2 R^2 \log(R)^{4k})$.

Inserting everything back in our formula for S_1 leads to

$$\begin{aligned}
S_1 = \frac{N}{W} &\sum_{u_1,\dots,u_k} \left(\prod_{i=1}^{k} \varphi(u_i) \right) \\
&\cdot \sum_{s_{i,j}\ i \neq j} \left(\prod_{\substack{1 \leq i,j \leq k \\ i \neq j}} \mu(s_{i,j}) \right) \cdot \left(\prod_{i=1}^{k} \frac{\mu(a_i)\mu(b_i)}{\varphi(a_i)\varphi(b_i)} \right) y_{a_1,\dots,a_k} y_{b_1,\dots,b_k} \\
&\qquad + O(y_{\max}^2 R^2 \log(R)^{4k})
\end{aligned}$$

Here we use

$$a_j = u_j \prod_{i \neq j} s_{j,i} \text{ and } b_j = u_j \prod_{i \neq j} s_{i,j}.$$

[11] This is easily verified for d prime. Since both sides are multiplicative we obtain the general formula immediately.

Note that in order to make the formulation above correct we use the restrictions in the $s_{i,j}$-sum, which we discussed above. We can further rewrite this as

$$S_1 = \frac{N}{W} \sum_{u_1,\dots,u_k} \left(\prod_{i=1}^{k} \frac{\mu(u_i)^2}{\varphi(u_i)} \right) \cdot \sum_{s_{i,j}\ i\neq j} \left(\prod_{\substack{1\le i,j\le k \\ i\neq j}} \frac{\mu(s_{i,j})}{\varphi(s_{i,j})^2} \right) \cdot y_{a_1,\dots,a_k} y_{b_1,\dots,b_k} + O(y_{\max}^2 R^2 \log(R)^{4k})$$

Note that the support of $y_{a_1,\dots,a_k}$ allows us to assume that $(s_{i,j}, W) = 1$. Thus, we can either have $s_{i,j} = 1$ or $s_{i,j} > D_0$.

If there is one tuple (i,j) (with $i \neq j$) with one $s_{i,j} > D_0$, then we can estimate the corresponding contribution trivially by

$$\ll y_{\max}^2 \frac{N}{W} \left(\sum_{\substack{u<R \\ (u,W)=1}} \frac{\mu(u)^2}{\varphi(u)} \right)^k \left(\sum_{s_{i,j}>D_0} \frac{\mu(s_{i,j})^2}{\varphi(s_{i,j})^2} \right) \cdot \left(\sum_{s\ge 1} \frac{\mu(s)^2}{\varphi(s)^2} \right)^{k^2-k-1} \ll y_{\max}^2 \frac{\varphi(W)^k N \log(R)^k}{W^{k+1} D_0}. \quad (16)$$

Here we used Lemma 2.4 to estimate the u-sum.

Thus, the main contribution comes from $s_{i,j} = 1$ for all $i \neq j$. We simply have

$$S_1 = \frac{N}{W} \sum_{u_1,\dots,u_k} \frac{y_{u_1,\dots,u_k}^2}{\prod_{i=1}^k \varphi(u_i)} + O\left(y_{\max}^2 \frac{\varphi(W)^k N \log(R)^k}{W^{k+1} D_0} + y_{\max}^2 R^2 \log(R)^{4k} \right).$$

Recall that $R^2 = N^{\theta-2\delta} \le N^{1-2\delta}$ and $W \ll N^\delta$, so that the second error term is absorbed into the first. This completes the proof. $\square$

We turn towards S_2. First, decompose

$$S_2 = \sum_{m=1}^{k} S_2^{(m)} \quad (17)$$

for

$$S_2^{(m)} = \sum_{\substack{N\le n<2N, \\ n\equiv v_0 \bmod W}} \mathbb{1}_{\mathcal{P}}(n+h_m) \left(\sum_{\substack{d_1,\dots,d_k \\ d_i|(n+h_i)}} \lambda_{d_1,\dots,d_k} \right)^2.$$

We define a completely multiplicative function g by setting $g(p) = p-2$ and define

$$y^{(m)}_{r_1,\dots,r_k} = \left(\prod_{i=1}^{k} \mu(r_i)g(r_i)\right) \cdot \sum_{\substack{d_1,\dots,d_k \\ r_i|d_i \\ d_m=1}} \frac{\lambda_{d_1,\dots,d_k}}{\varphi(d_1)\cdots\varphi(d_k)}. \tag{18}$$

Finally set $y^{(m)}_{\max} = \sup_{r_1,\dots,r_k} |y^{(m)}_{r_1,\dots,y_k}|$. Towards the evaluation of $S^{(m)}_2$ we obtain

Lemma 5.2. *For any $A > 0$ we have*

$$S^{(m)}_2 = \frac{N}{\varphi(W)\log(N)} \cdot \sum_{r_1,\dots,r_k} \frac{(y^{(m)}_{r_1,\dots,r_k})^2}{g(r_1\cdots r_k)} + O\left(\frac{(y^{(m)}_{\max})^2\varphi(W)^{k-2}N\log(N)^{k-2}}{W^{k-1}D_0}\right) + O\left(\frac{y^2_{\max}N}{\log(N)^A}\right).$$

Proof. As earlier we start by opening the square and exchanging order of summation:

$$S^{(m)}_2 = \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,d_k}} \lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k} \sum_{\substack{N\le n<2N \\ n\equiv v_0 \bmod W \\ [d_i,e_i]|(n+h_i)}} \mathbb{1}_{\mathcal{P}}(n+h_m).$$

As before put $q = W\prod_{i=1}^{k}[d_i,e_i]$. If $W, [d_1,e_1],\dots,[d_k,e_k]$ are pairwise co-prime and $d_m = e_m = 1$, then we can write

$$\sum_{\substack{N\le n<2N \\ n\equiv v_0 \bmod W \\ [d_i,e_i]|(n+h_i)}} \mathbb{1}_{\mathcal{P}}(n+h_m) = \sum_{\substack{N\le n<2N \\ n\equiv\gamma \bmod q}} \mathbb{1}_{\mathcal{P}}(n+h_m) = \frac{1}{\varphi(q)} \underbrace{\sum_{N\le n<2N} \mathbb{1}_{\mathcal{P}}(n)}_{=X_N} + O(E(N,q)).$$

for

$$E(N,q) = 1 + \sup_{(\gamma,q)=1} \left| \sum_{\substack{N\le n<2N \\ n\equiv\gamma \bmod q}} \mathbb{1}_{\mathcal{P}}(n) - \frac{X_N}{\varphi(q)} \right|.$$

If $W, [d_1,e_1],\dots,[d_k,e_k]$ are not pairwise co-prime, then the inner sum simply vanishes. Recall that, due to the support of $\lambda_{d_1,\dots,d_k}$ we can assume that d_i and e_i are $\square$-free. In this case the co-primality condition reduces to $(d_i,e_j) = 1$ for $i \ne j$. We obtain

$$S^{(m)}_2 = \frac{X_N}{\varphi(W)} \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k \\ (d_i,e_j)=1\, i\ne j \\ e_m=d_m=1}} \frac{\lambda_{d_1\dots,d_k}\lambda_{e_1,\dots,e_k}}{\varphi([d_1,e_1]\cdots[d_k,e_k])} + O\left(\sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k}} |\lambda_{d_1\dots,d_k}\lambda_{e_1,\dots,e_k}| \cdot E(N,q)\right).$$

We focus on estimating the error. First note that for a square-free integer r there are at most $\tau_{3k}(r)$ possibilities for $d_1,\dots,d_k,e_1,\dots,d_k$ with $W[d_1,e_1]\cdots[d_k,e_k] =$

r. Furthermore, by exploiting the support of $\lambda_{d_1,\dots,d_k}$ we find that $q \le R^2W$. After recalling that $\lambda_{\max} \ll y_{\max} \cdot \log(R)^k$ we can estimate the error by

$$\sum_{\substack{d_1,\dots,d_k\\e_1,\dots,e_k}} |\lambda_{d_1\dots,d_k}\lambda_{e_1,\dots,e_k}| \cdot E(N,q) \ll y_{\max}^2 \log(R)^{2k} \sum_{r<R^2W} \mu(r)^2\tau_{3k}(r)E(N,r)$$
$$\ll \left(\sum_{r<R^2W} \mu(r)^2\tau_{3k}(r)^2E(N,r)\right)^{\frac{1}{2}} \left(\sum_{r<R^2W} E(N,r)\right)^{\frac{1}{2}}.$$

Here we have used Cauchy-Schwarz. The first sum can be treated by estimating $E(N,r) \ll \frac{N}{\varphi(r)}$ trivially. Indeed this gives a contribution of

$$\sum_{r<R^2W} \mu(r)^2\tau_{3k}(r)^2E(N,r) \ll N \sum_{r<R^2W} \frac{\mu(r)^2}{\varphi(r)}\tau_{3k}(r)^2$$
$$\ll \left(\sum_{d\le R^2W} \frac{\mu(r)^2}{\varphi(r)}\right)^{9k^2} \ll \log(R^2W)^{9k^2} \ll \log(N)^{9k^2}.$$

To treat the second one we use that the primes have level of distribution θ. Indeed, by choice of R we have

$$R^2W \ll N^{\theta-2\delta} \log\log(N).$$

Thus, for any $A' > 0$ we get

$$\sum_{r<R^2W} E(N,r) \ll \frac{N}{\log(N)^{A'}}.$$

All together we have obtained

$$\sum_{\substack{d_1,\dots,d_k\\e_1,\dots,e_k}} |\lambda_{d_1\dots,d_k}\lambda_{e_1,\dots,e_k}| \cdot E(N,q) \ll y_{\max}^2 \frac{N}{\log(N)^{A'/2-9k^2/2}}.$$

We are done after choosing A' appropriately in terms of A and k.

We turn towards the main term. Here we will start as in the proof of Lemma 5.1 and use similar notation. For square-free d_i, e_i we have

$$\frac{1}{\varphi([d_i,e_i])} = \frac{1}{\varphi(d_i)\varphi(e_i)} \sum_{u_i|(d_i,e_i)} g(u_i).$$

This is again easily verified on primes.[12] So far we can write

$$S_2^{(m)} = \frac{X_N}{\varphi(W)} \sum_{u_1,\dots,u_k} \left(\prod_{i=1}^k g(u_i) \right) \sum_{s_{i,j}\, i\neq j} \left(\prod_{\substack{1\leq i,j\leq k\\ i\neq j}} \mu(s_{i,j}) \right)$$
$$\cdot \sum_{\substack{d_1,\dots,d_k\\ e_1,\dots,e_k\\ u_i|(d_i,e_i)\\ s_{i,j}|(d_i,e_j)\, i\neq j\\ d_m=e_m=1}} \frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k \varphi(d_i)\varphi(e_i)} + O\left(y_{\max}^2 \frac{N}{\log(N)^A} \right).$$

At this point we can substitute $y_{r_1,\dots,r_k}$ and obtain

$$S_2^{(m)} = \frac{X_N}{\varphi(W)} \sum_{u_1,\dots,u_k} \left(\prod_{i=1}^k \frac{\mu(u_i)^2}{g(u_i)} \right) \sum_{s_{i,j}\, i\neq j} \left(\prod_{\substack{1\leq i,j\leq k\\ i\neq j}} \frac{\mu(s_{i,j})}{g(s_{i,j})^2} \right) \cdot y_{a_1,\dots,a_k} y_{b_1,\dots,b_k}$$
$$+ O\left(y_{\max}^2 \frac{N}{\log(N)^A} \right),$$

for $a_j = u_j \prod_{i\neq j} s_{j,i}$ and $b_j = u_j \prod s_{i,j}$.[13]

The contributions with $s_{i,j} \neq 1$ for some some tuple (i,j) with $i \neq j$, can be estimated trivially as in (16) above. The corresponding error is

$$O\left(\frac{(y_{\max}^{(m)})^2 \varphi(W)^{k-2} N \log(R)^{k-1}}{W^{k-1} D_0 \log(N)} \right).$$

Note that we get slightly different powers because we have the additional constraint $u_m = 1$. Furthermore, we have to replace the estimate from Lemma 2.4 with

$$\sum_{\substack{u\ll R\\ (u,W)=1}} \frac{\mu(u)^2}{g(u)} \ll \frac{\varphi(W)}{W} \log(R),$$

[12] If $(d_i, e_i) = 1$, then the formula is true by multiplicativity of ϖ. For $e_i = d_i = p$ we have

$$\frac{1}{\varphi([p,p])} = \frac{1}{\varphi(p)} = \frac{1}{p-1}$$

and on the other hand

$$\frac{1}{\varphi(p)\varphi(p)} \sum_{d|(p,p)} g(d) = \frac{1+p-2}{(p-1)^2} = \frac{1}{p-1}.$$

The general case can easily be deduced similarly.

[13] Note that as in the proof of Lemma 5.1 we are using that the numbers contributing to the $s_{i,j}$-sums satisfy certain co-primality conditions, such as $(s_{i,j}, W) = 1$ for example. In particular $2 \nmid s_{i,j}$ so that $g(2) = 0$ is no issue for the formula presented above.

which is easy to prove.

Thus we have

$$S_2^{(m)} = \frac{X_N}{\varphi(W)} \sum_{u_1,\dots,u_k} \frac{(y^{(m)}_{u_1,\dots,u_k})^2}{g(u_1\cdots u_k)} + O\left(\frac{(y^{(m)}_{\max})^2\varphi(W)^{k-2}N\log(R)^{k-2}}{W^{k-1}D_0}\right) + O\left(y_{\max}^2\frac{N}{\log(N)^A}\right).$$

We conclude the proof by the prime number theorem:

$$X_N = \frac{N}{\log(N)} + O\left(\frac{N}{\log(N)^2}\right).$$

The error term produced this way is easily seen to be negligible. □

The next result establishes the link between $y_{r_1,\dots,r_k}$ that appear in our evaluation of S_1 and the weights $y^{(m)}_{r_1,\dots,r_k}$ encountered here.

Lemma 5.3. *For $r_m = 1$ we have*

$$y^{(m)}_{r_1,\dots,r_k} = \sum_{a_m} \frac{y_{r_1,\dots r_{m-1},a_m,r_{m+1},\dots,r_k}}{\varphi(a_m)} + O\left(\frac{y_{\max}\varphi(W)\log(R)}{WD_0}\right).$$

Proof. We combine the definition of $y^{(m)}_{r_1,\dots,r_m}$ (see (18)) with the change of variables (15). This yields

$$y^{(m)}_{r_1,\dots,r_k} = \left(\prod_{i=1}^k \mu(r_i)g(r_i)\right)\cdot \sum_{\substack{d_1,\dots,d_k\\ r_i|d_i\\ d_m=1}} \left(\frac{\mu(d_i)d_i}{\varphi(d_i)}\right)\cdot \sum_{\substack{a_1,\dots,a_k\\ d_i|a_i}} \frac{y_{a_1,\dots,a_k}}{\varphi(a_1)\cdots\varphi(a_k)}.$$

The only available move is to interchange the sums. We get

$$y^{(m)}_{r_1,\dots,r_k} = \left(\prod_{i=1}^k \mu(r_i)g(r_i)\right)\cdot \sum_{\substack{a_1,\dots,a_k\\ r_i|a_i}} \frac{y_{a_1,\dots,a_k}}{\varphi(a_1)\cdots\varphi(a_k)}\cdot \sum_{\substack{d_1,\dots,d_k\\ d_i|a_i,\, r_i|d_i\\ d_m=1}} \prod_{i=1}^k \frac{\mu(d_i)d_i}{\varphi(d_i)}.$$

The d_i-sums can be evaluated exactly and we get

$$y^{(m)}_{r_1,\dots,r_k} = \left(\prod_{i=1}^k \mu(r_i)g(r_i)\right)\cdot \sum_{\substack{a_1,\dots,a_k\\ r_i|a_i}} \frac{y_{a_1,\dots,a_k}}{\varphi(a_1)\cdots\varphi(a_k)} \prod_{i\neq m}\frac{\mu(a_i)r_i}{\varphi(a_i)}.$$

(Note that the d_m-sum is trivial!) At this point we observe that the support of $y_{a_1,\dots,a_k}$ allows us to assume that $(a_j, W) = 1$. But this implies that either $a_j = r_j$

or $a_j > D_0 r_j$. If $a_j \neq r_j$ for some $j \neq m$, then we can estimate everything trivially by

$$\ll y_{\max} \left(\prod_{i=1}^{k} g(r_i) r_i \right) \left(\sum_{\substack{a_j > D_0 r_j \\ r_j | a_j}} \frac{\mu(a_j)^2}{\varphi(a_j)^2} \right) \left(\sum_{\substack{a_m < R \\ (a_m, W)=1}} \frac{\mu(a_m)^2}{\varphi(a_m)} \right) \prod_{\substack{1 \leq i \leq k \\ i \neq j, m}} \left(\frac{\mu(a_i)^2}{\varphi(a_i)^2} \right)$$

$$\ll y_{\max} \left(\prod_{i=1}^{k} \frac{g(r_i) r_i}{\varphi(r_i)^2} \right) \frac{\varphi(W) \log(R)}{W D_0} \ll y_{\max} \frac{\varphi(W) \log(R)}{W D_0}.$$

We conclude that the only relevant contribution comes from $a_j = r_j$ for all $j \neq m$. More precisely

$$y^{(m)}_{r_1,\ldots,r_k} = \left(\prod_{i=1}^{k} \frac{g(r_i) r_i}{\varphi(r_i)^2} \right) \cdot \sum_{a_m} \frac{y_{a_1,\ldots,a_k}}{\varphi(a_m)} + O\left(y_{\max} \frac{\varphi(W) \log(R)}{W D_0} \right).$$

Except for the first product this is exactly what we want. To finish the proof we observe that

$$\frac{g(p)p}{\varphi(p)^2} = \frac{p(p-2)}{(p-1)^2} = 1 + O(p^{-2}).$$

Thus, taking $(r_i, W) = 1$ into account yields

$$\prod_{i=1}^{k} \frac{g(r_i) r_i}{\varphi(r_i)^2} = 1 + O(D_0^{-1}).$$

Inserting this and observing that error is negligible completes the proof. □

The next step is to make a meaningful Ansatz for the choice of the weights y. This choice should maximize S_2/S_1 (or at least the main terms thereof). On can arrive at a reasonable guess, which is most likely close to optimal, by using Lagrangian multipliers. We will not go into this and only make the following ad-hoc definition.

Let $F\colon \mathbb{R}^k \to \mathbb{R}$ be a smooth function with support in

$$\mathcal{R}_k = \{(x_1, \ldots, x_k) \in [0,1]^k \colon \sum_{i=1}^{k} x_i \leq 1\}.$$

Then we set

$$y_{r_1,\ldots,r_k} = \begin{cases} F(\frac{\log(r_1)}{\log(R)}, \ldots, \frac{\log(r_k)}{\log(R)}) & \text{if } (r, W) = 1 \text{ and } \mu(r) \neq 0, \\ 0 & \text{else,} \end{cases} \tag{19}$$

where we write $r = r_1 \cdots r_k$. This choice gives the weights $\lambda_{d_1,\ldots,d_k}$ indicated in (12). It will be useful to introduce the Sobolev norms

$$\|G\|_{W^{k,\infty}(\Omega)} = \max_{|\alpha| \leq k} \sup_{x \in \Omega} |\frac{\partial^{|\alpha|}}{\partial x^{\alpha}} G(x)|,$$

where $G\colon \Omega \to \mathbb{C}$ is smooth and $\Omega \subseteq \mathbb{R}^n$.

We will require some further lemmata before we can come to the main result of this subsection.

Lemma 5.4. *Let $A_1, A_2, L > 0$. Let γ be a multiplicative function such that*

$$0 \leq \frac{\gamma(p)}{p} \leq 1 - A_1 \tag{20}$$

and

$$-L \leq \sum_{w \leq p \leq z} \frac{\gamma(p)\log(p)}{p} - \log(\frac{z}{w}) \leq A_2 \tag{21}$$

for any $2 \leq w \leq z$. Define the completely multiplicative function g by setting $g(p) = \frac{\gamma(p)}{p-\gamma(p)}$. Then, for a smooth function $G\colon [0,1] \to \mathbb{R}$, we have

$$\sum_{d<z} \mu(d)^2 g(d) G\left(\frac{\log(d)}{\log(z)}\right) = \mathfrak{S}\log(z)\int_0^1 G(x)dx + O_{A_1,A_2}(\mathfrak{S}L \cdot \|G\|_{W^{1,\infty}([0,1])}),$$

where

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

This is a standard result in sieve theory and a detailed proof can be found in [HR, Chapter 5]. Note that in loc. cit. the condition stated in (20) is referred to as Ω_1 while the condition from (21) is called $\Omega(1, L)$.

Proof. We first aim to convince ourselves that the product $\mathfrak{S}$ is absolutely convergent. To do so we will prove that

$$\prod_{z \leq p < w}\left(1 + \frac{g(p)}{p^s}\right)\left(1 - \frac{1}{p^{s+1}}\right) = 1 + O\left(\frac{1}{\log(z)}\right) \text{ uniformly in } s \geq 0. \tag{22}$$

Indeed, noting

$$1 + g(p) = \left(1 - \frac{\gamma(p)}{p}\right)^{-1},$$

taking $s = 0$ and letting w go to ∞ gives the tail bound

$$\prod_{p \geq z}\left(1 - \frac{\gamma(p)}{p}\right)^{-1}\left(1 - \frac{1}{p}\right) = 1 + O\left(\frac{L}{\log(z)}\right) \tag{23}$$

This shows convergence of $\mathfrak{S}$. Establish (22) relies on the two inequalities

$$\sum_{z \leq p < w} g(p)^2 = O\left(\frac{1}{\log(z)}\right) \text{ and} \tag{24}$$

$$\sum_{z \leq p < w} \frac{g(p)}{p^s} - \sum_{z \leq p < w} \frac{1}{p^{s+1}} = O\left(\frac{L}{\log(z)}\right) \tag{25}$$

These are both straight forward consequences of (20) and (21) and we leave the proof as an exercise.[14] Finally we recall the Taylor expansion of the logarithm:

$$\log(1+x) = x + O(x^2). \tag{26}$$

This gives

$$\begin{aligned}\log&\left[\prod_{z\le p<w}\left(1+\frac{g(p)}{p^s}\right)\left(1-\frac{1}{p^{s+1}}\right)\right]\\ &= \sum_{z\le p<w}\frac{g((p)}{p^s} - \sum_{z\le p<w}\frac{1}{p^{s+1}} + O\left(\sum_{z\le p<w}(g(p)^2+p^{-2})\right)\\ &= O\left(\frac{L}{\log(z)}\right),\end{aligned}$$

where we used (24) and (25) in the last step. We arrive at (22) by exponentiation.

Next we define

$$S(z) = \sum_{d<z}\mu(d)^2 g(d).$$

Our next goal is to show that[15]

$$S(z) = \mathfrak{G}\log(z) + O(\mathfrak{G}L). \tag{27}$$

We start proving (27) with some elementary observations. First, put

$$S_p(z) = \sum_{\substack{d<z\\(d,p)=1}}\mu(d)^2 g(d).$$

For $p<z$ we compute

$$S(z) = \sum_{\substack{d<z\\(d,p)=1}}\mu(d)^2 g(d) + \sum_{\substack{d<z\\p\mid d}}\mu(d)^2 g(d) = S_p(z) + g(p)S_p(z/p).$$

Using this expression and the definition of g yields

$$\left(1-\frac{\gamma(p)}{p}\right)S(z) = S_p(z) - \frac{\gamma(p)}{p}\left(S_p(z)-S_p(z/p)\right).$$

This can be rewritten as

$$S_p(z/p) = \left(1-\frac{\gamma(p)}{p}\right)S(z/p) + \frac{\gamma(p)}{p}\left(S_p(z/p)-S_p(z/p^2)\right). \tag{28}$$

We can now study the sum

$$\sum_{d<z}\mu(d)^2 g(d)\log(d) = \sum_{d<z}\mu(d)^2 g(d)\sum_{p\mid d}\log(p) = \sum_{p<z} g(p)\log(p)S_p(z/p).$$

---

[14] See Remark 5.5 below for a sketch of the solution.

[15] This is essentially the content of [HR, Lemma 5.4].

Inserting (28) gives

$$\begin{aligned}\sum_{d<z}\mu(d)^2g(d)\log(d) &= \sum_{p<z}\frac{\gamma(p)\log(p)}{p}\sum_{d<z/p}\mu(d)^2g(d)\\ &\quad+\sum_{p<z}\frac{g(p)\gamma(p)}{p}\log(p)\sum_{\substack{z/p^2\le d<z/p\\(d,p)=1}}\mu(d)^2g(d)\\ &=\sum_{d<z}\mu(d)^2g(d)\sum_{p<z/d}\frac{\gamma(p)\log(p)}{p}\\ &\quad+\sum_{d<z}\mu(d)^2g(d)\sum_{\substack{\sqrt{z/d}\le p<z/d\\ p\nmid d}}\frac{g(p)\gamma(p)}{p}\log(p).\end{aligned}$$

We see that our assumption (21) can be written as

$$\sum_{p<y}\frac{\gamma(p)\log(p)}{p}=\log(y)+O(L).$$

On the other one can deduce from (20) and (21) that[16]

$$\sum_{\substack{\sqrt{z/d}\le p<z/d\\ p\nmid d}}\frac{g(p)\gamma(p)}{p}\log(p)\ll\sum_{\sqrt{z/d}\le p<z/d}g(p)\ll 1. \tag{29}$$

In particular we get

$$\sum_{d<z}\mu(d)^2g(d)\log(d)=\sum_{d<z}\mu(d)^2g(d)\log(z/d)+O\left(LS(z)\right). \tag{30}$$

On the other hand we can use partial summation (i.e. (9)) to see that

$$\sum_{d<z}\mu(d)^2g(d)\log(d)=S(z)\log(z)-\int_1^z S(t)\frac{dt}{t}.$$

We define

$$T(z)=\int_1^z S(t)\frac{dt}{t}$$

and observe that by partial summation (i.e. (9)) we have

$$T(z)=\sum_{s<z}\mu(d)^2g(d)\log\left(\frac{z}{d}\right).$$

[16] This is a byproduct of the solution to the exercise posed earlier in this proof. See Remark 5.5 for details.

We can thus write (30) as

$$S(z)\log(z) = 2T(z) + S(z)\log(z)\cdot r(z),$$

where $r(z) \ll \frac{L}{\log(z)}$ is some error, which we think of as small.[17] We write this as

$$S(z) = \frac{1}{1-r(z)}\frac{2}{\log(z)}T(z) \text{ and } E(y) = \log\left(\frac{2}{\log(y)^2}T(y)\right).$$

By differentiating E we get

$$\begin{aligned}E'(y) = -\frac{2}{\log(y)y} + \frac{S(y)}{yT(y)} &= -\frac{2}{\log(y)y} + \frac{1}{1-r(y)}\frac{2}{y\log(y)} \\ &= \frac{2}{y\log(y)}\frac{r(y)}{1-r(y)} \ll \frac{L}{y\log(y)^2}.\end{aligned}$$

We conclude that the integral

$$\int_z^\infty E'(y)dy$$

is convergent. We conclude that

$$\frac{2}{\log(z)^2}T(z) = \exp(E(z)) = C\cdot\exp\left(-\int_z^\infty E'(y)dy\right) = C\cdot\left(1+O\left(\frac{L}{\log(z)}\right)\right).$$

We rewrite this as

$$T(z) = \frac{C}{2}\log(z)^2\left(1+O\left(\frac{L}{\log(z)}\right)\right).$$

Since

$$\frac{1}{1-r(z)} = 1 + \frac{r(z)}{1-r(z)} = 1 + O\left(\frac{L}{\log(z)}\right)$$

we deduce

$$S(z) = C\cdot\log(z)\left(1+O\left(\frac{L}{\log(z)}\right)\right).$$

It remains to determine the constant C. To do so we recall

$$\int_1^\infty \frac{\log(y)^{\lambda-1}}{y^{s+1}}dy = \frac{1}{s^\lambda}\int_0^\infty e^{-t}t^{\lambda-1}dt = \frac{\Gamma(\lambda)}{s^\lambda} \text{ for } \lambda, s > 0.$$

Furthermore,

$$\prod_p\left(1+\frac{g(p)}{p^s}\right) = \sum_{d=1}^\infty \frac{\mu(d)^2 g(d)}{d^s} = s\int_1^\infty \frac{S(y)}{y^{s+1}}dy \text{ for } s > 0.$$

Combining this with our previous result gives

$$\prod_p\left(1+\frac{g(p)}{p^s}\right) = C\cdot s\int_1^\infty \frac{\log(y)+O(L)}{y^{s-1}}dy = \frac{C}{s} + O(L).$$

[17]The opposite case is handled by a separate argument. See Remark 5.6 for a sketch.

We conclude that

$$C = \lim_{s\to 0} s \prod_p \left(1 + \frac{g(p)}{p^s}\right).$$

Recall that for $s > 0$ we can define the Riemann zeta function by its Euler product

$$\zeta(1+s) = \prod_p \left(\frac{1}{1-p^{-1-s}}\right).$$

On the other hand it is well known that

$$1 = \lim_{s\to 0} s\zeta(1+s).$$

Inserting this yields

$$C = \lim_{s\to 0} \prod_p \left(1 + \frac{g(p)}{p^s}\right)\left(1 - \frac{1}{p^{s+1}}\right) = \prod_p \left(1 - \frac{\gamma(p)}{p}\right)^{-1}\left(1 - \frac{1}{p}\right) = \mathfrak{G}.$$

This concludes the proof of (27).

Finally we upgrade (27) to the desired statement by partial summation. This is an easy application of (9). Indeed we get

$$\sum_{d<z} \mu(d)^2 g(d) G\left(\frac{\log(d)}{\log(z)}\right) = S(z)G(1) - \int_1^z S(t) G'\left(\frac{\log(t)}{\log(z)}\right)\frac{dt}{t\log(z)}.$$

Inserting (27) and estimating the error trivially yields

$$\sum_{d<z} \mu(d)^2 g(d) G\left(\frac{\log(d)}{\log(z)}\right) = \mathfrak{G}\log(z)G(1) - \mathfrak{G}\int_1^z \frac{\log(t)}{\log(z)} G'\left(\frac{\log(t)}{\log(z)}\right)\frac{dt}{t} + O\left(\mathfrak{G}L\|G\|_{W^{1,\infty}([0,1])}\right).$$

We now make a change of variables $\log(t) = s\log(z)$, which yields

$$\sum_{d<z} \mu(d)^2 g(d) G\left(\frac{\log(d)}{\log(z)}\right) = \mathfrak{G}\log(z)G(1) - \mathfrak{G}\log(z)\int_0^1 sG'(s)ds + O\left(\mathfrak{G}L\|G\|_{W^{1,\infty}([0,1])}\right).$$

We are done by partial integration in the s-integral. □

Remark 5.5. Let us sketch the parts of the argument from the proof of Lemma 5.4 that were left as an exercise. First, by applying partial summation to (21) one obtains

$$\sum_{z\le p<w} \frac{\gamma(p)}{p} = \log\left(\frac{\log(w)}{\log(z)}\right) + O\left(\frac{L}{\log(z)}\right) \quad \text{and}$$

$$\sum_{z\le p<w} \frac{\gamma(p)}{p\log(p)} \ll \frac{L}{\log(z)}.$$

In particular we have $\frac{\gamma(p)}{p} \ll \log(p)^{-1}$. On the other hand (20) implies $g(p) \ll \gamma(p)/p$. This allows us to conclude that

$$\sum_{z\leq p<w} \frac{\gamma(p)g(p)}{p} \ll \sum_{z\leq p<w} \frac{\gamma(p)}{p\log(p)} \ll \frac{L}{\log(z)}. \tag{31}$$

After observing that

$$g(p) = \frac{\gamma(p)}{p} + \frac{\gamma(p)}{p}g(p)$$

we conclude that

$$\sum_{z\leq p<w} g(p) = \log\left(\frac{\log(w)}{\log(z)}\right) + \left(\frac{L}{\log(z)}\right). \tag{32}$$

Note that this already implies (29).

Recall that a version of Mertens' formula reads

$$\sum_{z\leq p<w} \frac{1}{p} = \log\left(\frac{\log(w)}{\log(z)}\right) + O\left(\frac{1}{\log(z)}\right).$$

(One can compare this to (5).) This allows us to write (32) as

$$\sum_{z\leq p<w} g(p) - \sum_{z\leq p<w} \frac{1}{p} \ll \frac{L}{\log(z)}.$$

One obtains (25) uniformly in s by another application of partial summation.Finally, since $g(p)^2 \ll \gamma(p)g(p)p^{-1}$m we see that (24) follows directly from (31).

Remark 5.6. Here we will discuss the situation $\log(z) \ll L$. In this case the error $r(z)$ appearing in the proof above is potentially large and another argument is required. We define

$$W(z) = \prod_{p<z}\left(1 - \frac{\gamma(p)}{p}\right) \text{ and } V(z) = \prod_{p<z}\left(1 - \frac{1}{p}\right).$$

From (23) and Mertens' formula (see (5)) we obtain

$$W(z)^{-1} = \mathfrak{G}e^{\gamma}\log(z)\left(1 + O\left(\frac{L}{\log(z)}\right)\right) \ll \mathfrak{G}L.$$

We note that in (27) the error term dominates. In particular it is sufficient to show that

$$S(z) \ll W(z)^{-1}.$$

But this is easy to see. Indeed, if we define $P(z) = \prod_{p<z} p$, then we can observe that

$$S(z) = \frac{1}{W(z)} - \sum_{\substack{d\geq z \\ p|P(z)}} \mu(d)^2 g(d).$$

Dropping the negative terms directly yields $S(z) \leq W(z)^{-1}$ as desired.

Lemma 5.7. *Let $F\colon \mathcal{R}_k \to \mathbb{R}$ be a smooth function and define the weights $y_{r_1,\dots,r_k}$ by (19). Then we have*

$$S_1 = \frac{\varphi(W)^k N \log(R)^k}{W^{k+1}} I_k(F) + O\left(\frac{\varphi(W)^k N \log(R)^k)}{W^{k+1} D_0} \|F\|^2_{W^{1,\infty}(\mathcal{R}_k)}\right),$$

where $I_k(F)$ is as in (13).

Proof. We start by inserting the formula (19) into the asymptotic expansion from Lemma 5.1. Doing so we obtain

$$S_1 = \frac{N}{W} \sum_{\substack{u_1,\dots,u_k \\ (u_i,u_j)=1 \\ (u_i,W)=1}} \left(\prod_{i=1}^k \frac{\mu(u_i)^2}{\varphi(u_i)}\right) F\left(\frac{\log(u_1)}{\log(R)}, \dots, \frac{\log(u_k)}{\log(R)}\right)^2 + O\left(\frac{\varphi(W)^k N \log(R)^k)}{W^{k+1} D_0} \|F\|^2_{W^{1,\infty}(\mathcal{R}_k)}\right).$$

If we drop the condition $(u_i, u_j) = 1$, then we introduce an error

$$\begin{aligned}
&\ll \|F\|^2_{W^{1,\infty}(\mathcal{R}_k)} \frac{N}{W} \sum_{p>D_0} \sum_{\substack{u_1,\dots,u_k<R \\ p|(u_i,u_j) \\ (u_i,W)=1}} \prod_{i=1}^k \frac{\mu(u_i)^2}{\varphi(u_i)} \\
&\ll \|F\|^2_{W^{1,\infty}(\mathcal{R}_k)} \frac{N}{W} \sum_{p>D_0} \frac{1}{(p-1)^2} \left(\sum_{\substack{u<R \\ (u,W)=1}} \frac{\mu(u)^2}{\varphi(u)}\right)^k \\
&\ll \frac{\varphi(W)^k N \log(R)^k}{W^{k+1} D_0} \|F\|^2_{W^{1,\infty}(\mathcal{R}_k)}.
\end{aligned}$$

So far we have seen that

$$S_1 = \frac{N}{W} \sum_{\substack{u_1,\dots,u_k \\ (u_i,W)=1}} \left(\prod_{i=1}^k \frac{\mu(u_i)^2}{\varphi(u_i)}\right) F\left(\frac{\log(u_1)}{\log(R)}, \dots, \frac{\log(u_k)}{\log(R)}\right)^2 + O\left(\frac{\varphi(W)^k N \log(R)^k)}{W^{k+1} D_0} \|F\|^2_{W^{1,\infty}(\mathcal{R}_k)}\right).$$

This remaining sum is evaluated using Lemma 5.4 for each variable u_i separately. In each application we take $\gamma(p) = \delta_{p\nmid W}$, $z = R$ and

$$L \ll 1 + \sum_{p|W} \frac{\log(p)}{p} \ll \log(D_0).$$

Note that with this choice

$$\mathfrak{S} = \prod_{p|W}(1 - \frac{1}{p}) = \frac{\varphi(W)}{W}.$$

As a result we get

$$\sum_{\substack{u_1,\dots,u_k\\(u_i,W)=1}} \left(\prod_{i=1}^k \frac{\mu(u_i)^2}{\varphi(u_i)}\right) F\left(\frac{\log(u_1)}{\log(R)},\dots,\frac{\log(u_k)}{\log(R)}\right)^2$$

$$= \frac{\varphi(W)^k}{W^k}\log(R)^k I_k(F) + O\left(\frac{\varphi(W)^k \log(R)^{k-1}\log(D_0)}{W^k}\|F\|^2_{W^{1,\infty}(\mathcal{R}_k)}\right).$$

Inserting this above gives the desired estimate. □

Lemma 5.8. *Let $F\colon \mathcal{R}_k \to \mathbb{R}$ be a smooth function and define the weights $y_{r_1,\dots,r_k}$ by* (19)*. Then we have*

$$S_2^{(m)} = \frac{\varphi(W)^k N \log(R)^{k+1}}{W^{k+1}\log(N)} J_k^{(m)}(F) + O\left(\frac{\varphi(W)^k N \log(R)^k)}{W^{k+1}D_0}\|F\|^2_{W^{1,\infty}(\mathcal{R}_k)}\right),$$

for $J_k^{(m)}$ as in (14)*.*

Proof. The first step is to estimate $y^{(m)}_{r_1,\dots,r_k}$. In view of the support properties of $y^{(m)}_{r_1,\dots,r_k}$ we can assume that $r_m = 1$, $(r, W) = 1$ and $\mu(r) \neq 0$ (for $r = r_1 \cdots r_k$). Recall the result from Lemma 5.3 and substitute our choice of $y_{r_1,\dots,r_k}$. We get

$$y^{(m)}_{r_1,\dots,r_k} = \sum_{(u,Wr)=1} \frac{\mu(u)^2}{\varphi(u)} F\left(\frac{\log(r_1)}{\log(R)},\dots,\frac{\log(r_{m-1})}{\log(R)},\frac{\log(u)}{\log(R)},\frac{\log(r_{m+1})}{\log(R)},\dots,\frac{\log(r_k)}{\log(R)}\right)$$

$$+ O\left(\frac{\varphi(W)\log(R)}{WD_0}\|F\|_{W^{1,\infty}(\mathcal{R}_k)}\right).$$

We deduce that

$$y^{(m)}_{\max} \ll \frac{\varphi(W)\log(R)}{W}\|F\|_{W^{1,\infty}(\mathcal{R}_k)}.$$

Next we define $\gamma(p) = \delta_{(p,Wr)=1}$ so that

$$\mathfrak{S} = \prod_p \left(1 - \frac{\gamma(p)}{p}\right)^{-1}(1 - \frac{1}{p}) = \frac{\varphi(Wr)}{Wr}.$$

Further, choose suitable constants A_1, A_2 and

$$L \ll 1 + \sum_{p|Wr}\frac{\log(p)}{p} \ll \sum_{p<\log(R)}\frac{\log(p)}{p} + \sum_{\substack{p|Wr,\\p>\log(R)}}\frac{\log\log(R)}{\log(R)} \ll \log\log(N).$$

Applying Lemma 5.4 yields

$$y^{(m)}_{r_1,\dots,r_k} = \log(R)\frac{\varphi(W)}{W}\left(\prod_{i=1}^{k}\frac{\varphi(r_i)}{r_i}\right)F^{(m)}_{r_1,\dots,r_k} + O\left(\frac{\varphi(W)\log(R)}{WD_0}\|F\|_{W^{1,\infty}(\mathcal{R}_k)}\right)$$

with

$$F^{(m)}_{r_1,\dots,r_k} = \int_0^1 F\left(\frac{\log(r_1)}{\log(R)},\dots,\frac{\log(r_{m-1})}{\log(R)},t_m,\frac{\log(r_{m+1})}{\log(R)},\dots,\frac{\log(r_k)}{\log(R)}\right)dt_m.$$

This expression for $y^{(m)}_{r_1,\dots,r_k}$ holds whenever it is non-zero. Inserting it in Lemma 5.2 reveals

$$S_2^{(m)} = \frac{N\varphi(W)\log(R)^2}{W^2\log(N)}\sum_{\substack{r_1,\dots,r_k\\(r_i,W)=1\\(r_i,r_j)=1\, i\neq j\\ r_m=1}}\left(\prod_{i=1}^{k}\frac{\mu(r_i)^2\varphi(r_i)^2}{g(r_i)r_i^2}\right)(F^{(m)}_{r_1,\dots,r_k})^2$$
$$+ O\left(\frac{\varphi(W)^kN\log(R)^k}{W^{k+1}D_0}\|F\|^2_{W^{1,\infty}(\mathcal{R}_k)}\right).$$

As in the proof of Lemma 5.7 we remove the condition $(r_i,r_j)=1$ for $i\neq j$. This introduces an acceptable error. The remaining sum will be evaluated once again using Lemma 5.4. This time we use

$$\gamma(p) = \delta_{(p,W)=1}\cdot\left(1-\frac{p^2-3p+1}{p^3-p^2-2p+1}\right)$$

and

$$L \ll 1 + \sum_{p|W}\frac{\log(p)}{p} \ll \log(D_0).$$

We obtain

$$S_2^{(m)} = \frac{\varphi(W)^kN\log(R)^{k+1}}{W^{k+1}\log(N)}J_k^{(m)} + O\left(\frac{\varphi(W)^kN\log(N)^k}{W^{k+1}D_0}\|F\|^2_{W^{1,\infty}(\mathcal{R}_k)}\right)$$

as desired. □

Let us summarize the results obtained in this section:

Proposition 5.9 (Proposition 4.1, [Ma]). *Suppose the primes have level of distribution $\theta>0$ and let $R=N^{\theta/2-\delta}$ for some small fixed $\delta>0$. Let $F\colon\mathcal{R}_k\to\mathbb{R}$ be a smooth function and define the weights $\lambda_{d_1,\dots,d_k}$ as in (12). Then we have*

$$S_1 = (1+o(1))\cdot\frac{\varphi(W)^kN\log(R)^k}{W^{k+1}}\cdot I_k(F)\ \text{and}$$
$$S_2 = (1+o(1))\cdot\frac{\varphi(W)^kN\log(R)^{k+1}}{W^{k+1}\log(N)}\cdot\sum_{m=1}^{k}J_k^{(m)}(F),$$

where $I_k(F)$ *(resp.* $J_k^{(m)}(F)$*) is defined as in* (13) *(resp.* (14)*).*

Proof. This is a direct consequence of Lemma 5.7 and Lemma 5.8 above. □

5.3. **Choosing the smooth weight.** At this point we have arrived at a purely analytic problem of choosing the smooth fixed function F optimally. This is an interesting and not to difficult optimization problem.

Recall the integrals $I_k(F)$ and $J_k^{(m)}(F)$ defined in (13) and (14) above:

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

We let $\mathcal{S}_k$ denote the set of Riemann-integrable functions $F\colon [0,1]^k \to \mathbb{R}$ with support in $\mathcal{R}_k$ such that $I_k(F) \neq 0$ and $J_k^{(m)}(F) \neq 0$. Define

$$M_k = \sup_{F \in \mathcal{S}_k} \frac{\sum_{m=1}^k J_k^{(m)}(F)}{I_k(F)}. \tag{33}$$

The goal of this section is to establish the following:

Proposition 5.10 (Proposition 4.3, [Ma]). *For* $k \in \mathbb{N}$ *we have*

1. $M_5 > 2$*;*
2. $M_{105} > 4$*; and*
3. $M_k > \log(k) - 2\log\log(k) - 2$ *for* k *sufficiently large.*

Proving this will take up the rest of the section. Before diving into the argument let us make the following general remarks.

1. Let $G_i\colon \mathcal{R}_k \to (0, \infty)$ be such that $\int_0^1 G_i(t_1, \dots, t_k) dt_i \leq 1$, then we have
$$M_k \leq \sup_{(t_1,\dots,t_k)\in\mathcal{R}_k} \sum_{i=1}^k G_i(t_1, \dots, t_k)^{-1}. \tag{34}$$
This is seen by an easy application of Cauchy-Schwarz.
2. Using (34) with $G_i(t_1, \dots, t_k) = \frac{k-1}{\log(k)} \cdot \frac{1}{1-t_1-\dots-t_k+kt_i}$ yields
$$M_k \leq \frac{k}{k-1}\log(k).$$
For $k = 2$ one can actually solve the optimization problem and obtain
$$M_2 = \frac{1}{1-W(1/e)} = 1,38593\dots.$$

(3) We define the operator $\mathcal{L}\colon \mathcal{S}_k \to \mathcal{S}_k$ by

$$[\mathcal{L}F](t_1,\dots,t_k) = \sum_{i=1}^{k}\int_0^1 F(t_1,\dots,t_i',\dots,t_k)dt_i'.$$

If F is an eigenfunction of $\mathcal{L}$ with positive eigenvalue λ (i.e. $\mathcal{L}F = \lambda F$), then $\lambda = M_k$. Using (34) with

$$G_i(t_1,\dots,t_k) = F(t_1,\dots,t_k)\left(\int_0^\infty F(t_1,\dots,t_i',\dots,t_k)dt_i'\right)^{-1}$$

yields the inequality $M_k \le \lambda$. On the other hand we can compute

$$\lambda I_k(F) = \lambda\langle F,F\rangle = \langle \mathcal{L}F, F\rangle = \sum_{m=1}^{k} J_k^{(m)}(F),$$

so that $M_k \ge \lambda$.

We first consider the case when k is large. Here we make the following Ansatz:

$$F(t_1,\dots,t_k) = \begin{cases} \prod_{i=1}^{k} g(kt_1) & \text{if } (t_1,\dots,t_k)\in\mathcal{R}_k,\\ 0 & \text{else.}\end{cases}$$

for a function $g\colon \mathbb{R}_{\ge 0}\to\mathbb{R}$ supported in $[0,T]$. Note that this choice makes F symmetric, so that

$$\sum_{m=1}^{k} J_k^{(m)}(F) = k\cdot J_k^{(1)}(F).$$

We also define

$$\gamma = \int_{\mathbb{R}_{\ge 0}} g(u)^2du \text{ and } \mu = \frac{\int_{\mathbb{R}_{\ge 0}} ug(u)^2du}{\int_{\mathbb{R}_{\ge 0}} g(u)^2}.$$

Finally we make the following assumption concerning the center of mass of g:

$$\mu < 1 - \frac{T}{k}. \tag{35}$$

We first record the trivial estimate

$$I_k(F) = \int\cdots\int_{\mathcal{R}_k} F(t_1,\dots,t_k)^2dt_1\cdots dt_k \le \left(\int_{\mathbb{R}_{\ge 0}} g(kt)^2dt\right)^k = k^{-k}\gamma^k. \tag{36}$$

Next we turn towards giving lower bounds for $J_k^{(1)}(F)$. By positivity we can estimate

$$J_k^{(1)}(F) \ge \int\cdots\int_{\substack{t_2,\dots,t_k\ge 0\\ t_2+\dots+t_k\le 1-T/k}} \left(\int_0^{T/k}\prod_{i=1}^{k} g(kt_i)dt_1\right)^2 dt_2\cdots dt_k$$

Without the restriction on the $t_2, \dots, t_k$ variables the integral is easily computed:

$$\begin{aligned}\widetilde{J}_k^{(1)}(F) &= \int_{\mathbb{R}_{\geq 0}} \cdots \int_{\mathbb{R}_{\geq 0}} \left(\int_0^{T/k} \prod_{i=1}^k g(kt_i) dt_1 \right)^2 dt_2 \cdots dt_k \\ &= \left(\int_{\mathbb{R}_{\geq 0}} g(kt_1) dt_1 \right)^2 \left(\int_{\mathbb{R}_{\geq 0}} g(kt)^2 dt \right)^{k-1} = k^{-k-1} \gamma^{k-1} \left(\int_{\mathbb{R}_{\geq 0}} g(t) dt \right)^2 .\end{aligned}$$

We now express the right hand side of (5.3) as $\widetilde{J}_k^{(1)}(F) - E_k$. More precisely, E_k is given by

$$\begin{aligned}E_k &= \int \cdots \int_{\substack{t_2, \dots, t_k \geq 0 \\ t_2 + \dots + t_k > 1 - T/k}} \left(\int_0^{T/k} \prod_{i=1}^k g(kt_i) dt_1 \right)^2 dt_2 \cdots dt_k \\ &= k^{-k-1} \left(\int_0^\infty g(u) du \right)^2 \int \cdots \int_{\substack{t_2, \dots, t_k \geq 0 \\ t_2 + \dots + t_k > k - T}} \prod_{i=2}^k g(u_i)^2 du_2 \cdots du_k .\end{aligned}$$

We will now use assumption (35) to show that E_k is small. We start with some observations. Define

$$\eta = \frac{k-T}{k-1} - \mu > 0$$

In particular, we can artificially write

$$k - T = (k-1)(\mu + \eta).$$

We write the restriction on the $t_2 \dots t_k$ integrals as

$$u_2 + \dots + u_k > (k-1)(\mu + \eta)$$

and obtain the inequality

$$1 < \eta^2 \left(\frac{1}{k-1} \sum_{i=2}^k u_i - \mu \right)^2 .$$

Using this inequality we can estimate E_k as

$$E_k \leq \eta^{-2} k^{-k-1} \left(\int_0^\infty g(u) du \right)^2 \int_0^\infty \cdots \int_0^\infty \left(\frac{u_2 + \dots + u_k}{k-1} - \mu \right)^2 \prod_{i=2}^k g(u_i)^2 du_2 \cdots du_k$$

We can compute

$$\int_0^\infty \cdots \int_0^\infty \left(\frac{2}{(k-1)^2} \sum_{2 \leq i < j \leq k} u_i u_j - \frac{2\mu}{k-1} \sum_{i=2}^k u_i + \mu^2 \right) \prod_{i=2}^k g(u_i)^2 du_2 \cdots du_k = -\frac{\mu^2 \gamma^{k-1}}{k-1} .$$

This accounts for all the terms in our upper bound for E_k that do not contain u_i^2 for some $2 \le i \le k$. To handle the remaining terms we observe that $u_j^2 g(u_j)^2 \le T u_j g(u_j)^2$. This leads to

$$\int_0^\infty \cdots \int_0^\infty u_j^2 \cdot \prod_{i=2}^k g(u_i)^2 du_2 \cdots du_k \le T\mu\gamma^{k-1}.$$

Combining everything gives

$$E_k \le \eta^{-2} k^{-k-1} \left(\int_0^\infty g(u)du \right)^2 \cdot \frac{T\mu\gamma^{k-1} - \mu^2\gamma^{k-1}}{k-1} \le \frac{\mu T \gamma^{k-1}}{\eta^2 (k-1) k^{k+1}} \left(\int_0^\infty g(u)du \right)^2.$$

Using

$$(k-1)\eta^2 \ge k(1 - T/k - \mu)^2 \text{ and } \mu \le 1$$

and combining all our observations so far gives is the rather clean result

$$\frac{kJ_k^{(1)}(F)}{I_k(F)} \ge \frac{\left(\int_0^\infty g(u)du\right)^2}{\int_0^\infty g(u)^2 du} \cdot \left(1 - \frac{T}{k(1-T/k-\mu)^2}\right).$$

We have arrived at the problem of maximizing $\int_0^T g(u)du$ under the constraints $\int_0^T g(u)^2 du = \gamma$ and $\int_0^T u g(u)^2 du = \mu\gamma$. We guess[18] that

$$g(t) = \frac{1}{1+At}.$$

With this choice we compute

$$\int_0^T g(u)du = \frac{\log(1+AT)}{A},$$
$$\int_0^T g(u)^2 du = \frac{1}{A}\left(1 - \frac{1}{1+AT}\right) \text{ and}$$
$$\int_0^T u g(u)^2 du = \frac{1}{A^2}\left(\log(1+AT) - 1 + \frac{1}{1+AT}\right).$$

We choose T such that $1 + AT = e^A$.[19] This leads to

$$\mu = (1 - e^{-A})^{-1} - A^{-1},\ T \le e^A/A \text{ and}$$
$$1 - T/k - \mu \ge A^{-1}(1 - A/(e^A - 1) - E^A/k).$$

[18] One can arrive at this by rewriting the problem as maximizing

$$\int_0^T g(u)du - \alpha\left(\int_0^T g(u)^2 du - \gamma\right) - \beta\left(\int_0^T u g(u)^2 du - \mu\gamma\right)$$

with respect to $\alpha, \beta > 0$ and g and then considering the Euler-Lagrange equation.

[19] This turns out to be close to optimal.

Inserting everything above yields

$$\frac{kJ_k^{(1)}(F)}{I_k(F)} \geq \frac{A}{1-e^{-A}}\left(1-\frac{T}{k(1-T/k-\mu)^2}\right) \geq A\left(1-\frac{Ae^A}{k(1-A/(e^A-1)-e^A/k)^2}\right),$$

where we need to make sure that the right hand side is positive. Our final choice for A is

$$A = \log(k) - 2\log\log(k) > 0.$$

For sufficiently large k we obtain

$$1 - T/k - \mu \geq A^{-1}\left(1 - \frac{\log(k)^3}{k} - \frac{1}{\log(k)^2}\right) > 0.$$

In particular $\mu < 1 - T/k$ as required in (35). We end up with the estimate

$$M_k \geq \frac{kJ_k}{I_k} \geq (\log(k) - 2\log\log(k))\left(1 - \frac{\log(k)}{\log(k)^2 + O(1)}\right) \geq \log(k) - 2\log\log(k) - 2$$

for k sufficiently large. This completes the case of large k.

If k is small we let P be a symmetric polynomial and suppose that

$$F(t_1, \ldots, t_k) = \begin{cases} P(t_1, \ldots, t_k) & \text{if } (t_1, \ldots, t_k) \in \mathcal{R}_k, \\ 0 & \text{else.} \end{cases} \tag{37}$$

Recall that such a symmetric polynomial P can be written as a polynomial in

$$P_j(t_1, \ldots, t_k) = \sum_{i=1}^{k} t_i^j.$$

Lemma 5.11 (Lemma 8.1, [Ma]). *We have*

$$\int \cdots \int_{\mathcal{R}_k} (1 - P_1)^a P_j^b \, dt_1 \cdots dt_k = \frac{a!}{(k + jb + a)!} G_{b,j}(k),$$

where

$$G_{b,j}(x) = b! \sum_{r=1}^{b} \binom{x}{r} \sum_{\substack{b_1, \ldots, b_r \geq 1 \\ b_1 + \ldots + b_r = b}} \prod_{i=1}^{r} \frac{(jb_i)!}{b_i!}$$

is a polynomial of degree b which depends only on b and j.

Proof. The first step is to show

$$\int \cdots \int_{\mathcal{R}_k} (1 - t_1 - \ldots - t_k)^a \prod_{i=1}^{k} t_i^{a_i} \, dt_1 \cdots dt_k = \frac{a! \prod_{i=1}^{k} a_i!}{(k + a + a_1 + \ldots + a_k)!}.$$

This is done by induction over k. The case $k = 1$ follows directly from the beta function identity

$$\int_0^1 t^a (1-t)^b \, dt = \frac{a! b!}{(a + b + 1)!}.$$

For the induction step we want to substitute $v = t_1/(1 - t_2 - \ldots - t_k)$ in the first integral. We compute

$$\int_0^{1-t_2-\ldots-t_k} \left(1 - \sum_{i=1}^k t_i\right)^a \left(\prod_{i=1}^k t_i^{a_i}\right) dt_1$$
$$= \left(\prod_{i=2}^k t_i^{a_i}\right) \left(1 - \sum_{i=2}^k t_i\right)^{a+a_1+1} \int_0^1 (1-v)^a v^{a_1} dv$$
$$= \frac{a!a_1!}{(a+a_1+1)!} \left(\prod_{i=2}^k t_i^{a_i}\right) \left(1 - \sum_{i=2}^k t_i\right)^{a+a_1+1}.$$

Inserting this we see that

$$\underset{\mathcal{R}_k}{\int \cdots \int} (1 - t_1 - \ldots - t_k)^a \prod_{i=1}^k t_i^{a_i} dt_1 \cdots dt_k$$
$$= \frac{a!a_1!}{(a+a_1+1)!} \underset{\mathcal{R}_{k-1}}{\int \cdots \int} (1 - t_2 - \ldots - t_k)^{a+a_1+1} \prod_{i=2}^k t_i^{a_i} dt_2 \cdots dt_k.$$

It is easy to conclude using the induction hypothesis.

Now the proof is easily completed by expanding

$$P_j^b = \sum_{\substack{b_1,\ldots,b_k \\ a_1+\ldots+b_k=b}} \frac{b!}{\prod_{i=1}^k b_i!} \prod_{i=1}^k t_i^{jb_i}.$$

Inserting this in our original integral allows us to compute that

$$\underset{\mathcal{R}_k}{\int \cdots \int} (1 - P_1)^a P_j^b dt_1 \cdots dt_k = \frac{b!a!}{(k+a+jb)!} \sum_{\substack{b_1,\ldots,b_k \\ a_1+\ldots+b_k=b}} \prod_{i=1}^k \frac{(jb_i)!}{b_i!}.$$

This only needs to be reformulated slightly by splitting the sum into pieces depending on how many of the b_i are non-zero. Given r there are $\binom{k}{r}$ ways to choose r of $b_1, \ldots, b_k$ non-zero. We arrive at

$$\sum_{\substack{b_1,\ldots,b_k \\ a_1+\ldots+b_k=b}} \prod_{i=1}^k \frac{(jb_i)!}{b_i!} = \sum_{r=1}^b \binom{k}{r} \sum_{\substack{b_1,\ldots,b_r \geq 1 \\ b_1+\ldots+b_r=b}} \prod_{i=1}^r \frac{(jb_i)!}{b_i!}$$

as desired. □

This can be used to get reasonable expressions for our integrals:

Lemma 5.12 (Lemma 8.2, [Ma]). *Let*

$$P = \sum_{i=1}^{d} a_i (1-P_1)^{b_i} P_2^{c_i} \text{ with } a_i \in \mathbb{R} \text{ and } b_i, c_i \in \mathbb{Z}_{\geq 0}$$

and suppose that F *is defined as in* (37)*. Then, for each* $1 \leq m \leq k$ *we have*

$$I_k(F) = \sum_{1 \leq i,j \leq d} a_i a_j \frac{(b_i+b_j)! G_{c_i+c_j,2}(k)}{(k+b_i+b_j+2c_i+2c_j)!} \text{ and}$$

$$J_k^{(m)}(F) = \sum_{1 \leq i,j \leq d} a_i a_j \sum_{c_1'=0}^{c_i} \sum_{c_2'=0}^{c_j} \binom{c_i}{c_i'} \binom{c_j}{c_2'} \cdot \frac{\gamma_{b_i,b_j,c_i,c_j,c_1',c_2'} G_{c_1'+c_2',2}(k-1)}{(k+b_i+b_j+2c_i+2c_j+1)!},$$

where

$$\gamma_{b_i,b_j,c_i,c_j,c_1',c_2'} = \frac{b_i! b_j! (2c_i - 2c_1')! (b_i+b_j+2c_i+2c_j-2c_1'-2c_2'+1)!}{(b_i+2c_i-2c_1'+1)!(b_j+2c_j-2c_2'+1)!}.$$

Proof. We will freely apply the previous Lemma. Let us start by evaluating $I_k(F)$:

$$\begin{aligned} I_k(F) &= \int \cdot \int_{\mathcal{R}_k} P^2 dt_1 \cdots dt_k \\ &= \sum_{1 \leq i,j \leq d} a_i a_j \int \cdot \int_{\mathcal{R}_k} (1-P_1)^{b_i+b_j} P_2^{c_i+c_j} dt_1 \cdots dt_k \\ &= \sum_{1 \leq i,j \leq d} a_i a_j \frac{(b_i+b_j)! G_{c_i+c_j,2}(k)}{(k+b_i+b_j+2c_i+2c_j)!}. \end{aligned}$$

Turning to $J_k^{(m)}(F)$ we first note that by symmetry this is independent of m. Thus we do the computation for $m=1$. Let us again first compute the t_1-integral

$$\begin{aligned} &\int_0^{1-t_2-\ldots-t_k} (1-P_1)^b P_2^c dt_1 \\ &= \sum_{c'=0}^{c} \binom{c}{c'} \left(\sum_{i=2}^{k} t_i^2 \right)^{c'} \int_0^{1-t_2-\ldots-t_k} \left(1 - \sum_{i=1}^{k} t_i \right)^b t_1^{2c-2c'} dt_1 \\ &= \sum_{c'=0}^{c} \binom{c}{c'} (P_2')^{c'} (1-P_1')^{b+2c-2c'+1} \int_0^1 (1-u)^b u^{2c-2c'} du \\ &= \sum_{c'=0}^{c} \binom{c}{c'} (P_2')^{c'} (1-P_1')^{b+2c-2c'+1} \frac{b!(2c-2c')!}{(b+2c-2c'+1)!}. \end{aligned}$$

Here we have abbreviated $P_1' = t_2 + \ldots + t_k$ and $P_2' = t_2^2 + \ldots + t_k^2$. This allows us to express

$$\left(\int_0^1 F dt_1\right)^2 = \sum_{1\le i,j\le d} a_i a_j \sum_{c_1'=0}^{c_i}\sum_{c_2'=0} c_j \binom{c_i}{c_1'}\binom{c_j}{c_2'}(P_2')^{c_1'+c_2'}(1-P_1')^{b_i+b_j+2c_i+2c_j-2c_1'-2c_2'+2} \cdot \frac{b_i!b_j!(2c_i-2c_1')!(2c_j-2c_2')!}{(b_i+2c_i-2c_1'+1)!(b_j+2c_j-2c_2'+1)!}.$$

The remaining $t_2, \ldots, t_k$-integrals over $\mathcal{R}_{k-1}$ can be computed using the previous result. After a bit of bookkeeping one arrives at the required expression. □

We put $\mathbf{a} = (a_1, \ldots, a_d)$. The upshot is that we can express $I_k(F)$ and $\sum_{m=1}^k J_k^{(m)}(F)$ as real positive definite quadratic forms in $\mathbf{a}$. Thus we can write

$$\frac{\sum_{m=1}^k J_k^{(m)}(F)}{I_k(F)} = \frac{\mathbf{a}^\top A_2 \mathbf{a}}{\mathbf{a}^\top A_1 \mathbf{a}}.$$

The upshot is that the matrices A_1 and A_2 can be computed explicitly in terms of k and the exponents b_i, c_i. Maximizing expressions as in (5.3) is well studied. We recall the following result:

Lemma 5.13. *Let A_1 and A_2 be real symmetric positive definite matrices. Then*

$$\frac{\mathbf{a}^\top A_2 \mathbf{a}}{\mathbf{a}^\top A_1 \mathbf{a}}$$

is maximal when $\mathbf{a}$ is an eigenvector of $A_1^{-1}A_2$ corresponding to the largest eigenvalue of $A_1^{-1}A_2$. The value of the ration at its maximum is this largest eigenvalue.

Proof. The function we are trying to maximize is scaling invariant. Thus we can assume things are normalized such that $\mathbf{a}^\top A_1 \mathbf{a} = 1$. Applying the law of Lagrangian multipliers we see that for $\mathbf{a}^\top A_2 \mathbf{a}$ to have a local extreme point subject to $\mathbf{a}^\top A_1 \mathbf{a} - 1 = 0$ it is necessary that

$$L(\mathbf{a}, \lambda) = \mathbf{a}^\top A_2 \mathbf{a} - \lambda(\mathbf{a}^\top A_1 \mathbf{a} - 1)$$

has a critical point. This happens exactly when

$$0 = \frac{\partial L}{\partial a_i} = ((2A_2 - 2\lambda A_1)\mathbf{a})_i$$

for all $1 \le i \le d$. This happens precisely when

$$A_1^{-1} A_2 \mathbf{a} = \lambda \mathbf{a}.$$

In this case it is clear that $\mathbf{a}^\top A_1 \mathbf{a} = \lambda^{-1} \mathbf{a}^\top A_2 \mathbf{a}$ and the proof is complete. □

It should be clear from here how to complete the proof of Proposition 5.10 by treating the cases $k = 5$ and $k = 105$. However the computations get already slightly involved and are best done using the help of computers.[20]

[20] We rely on Maynard's computations, which we have not checked ourselves.

First, if $k = 105$, then one computes all possible polynomials P with degree at most 11 that are of the shape considered in Lemma 5.12. There are 42 of these. For each polynomial one computes the corresponding matrices A_1 and A_2 as well as the maximal eigenvalue of $A_1^{-1}A_2$. It turns out that there is a polynomial for which we have

$$\lambda \approx 4.0020697\ldots > 4.$$

This gives the first result.

If $k = 5$, then we take

$$P = (1 - P_1)P_2 + \frac{7}{10}(1 - P_1)^2 + \frac{1}{14}P_2 - \frac{3}{14}(1 - P_1).$$

With this choice one computes that

$$M_5 \geq \frac{\sum_{m=1}^{k} J_k^{(m)}(F)}{I_k(F)} = \frac{1417255}{708216} > 2.$$

This completes the proof of Proposition 5.10 as well as this section.

5.4. **The Endgame.** We are now ready to put everything together and establish our main results.

Theorem 5.14 (Theorem 1.3, [Ma])**.** *We have*

$$\liminf_{n}(p_{n+1} - p_n) \leq 600.$$

Proof. We pick $k = 105$ and recall that by Proposition 5.10 we have $M_{105} > 4$. Furthermore, by the Bombieri-Vinogradov Theorem we can take $\theta = \frac{1}{2} - \epsilon$ for every $\epsilon > 0$. Making $\epsilon > 0$ sufficiently small we can achieve that

$$\frac{\theta}{2}M_{105} > 1.$$

Recall that $R = N^{\theta/2-\delta}$.

Having set up everything we can recall our basic strategy. We consider the sum $S_{v_0}(N, \rho) = s_2 - \rho S_1$ defined in (11). If we can show

$$S_{v_0}(N, \rho) > 0,$$

then there are at least $\lfloor \rho + 1 \rfloor$ primes among the set $\{n + h_1, \ldots, n + h_k\}$ for some $N \leq n < 2N$. Thus we have have to show $S_2 > \rho S_1$ with our set-up above. But this follows from Proposition 5.9 (for N sufficiently large). In our case we were able to achieve $\lfloor \rho + 1 \rfloor \geq 2$.

Note that this is true for any admissible 105-tuple $\mathcal{H} = \{h_1, \ldots, h_{105}\}$. Thus we have

$$\liminf_{n}(p_{n+1} - p_n) \leq \max_{1 \leq i,j \leq 105}(h_i - h_j).$$

By choosing the tuple explicitly one establishes the theorem.[21] □

[21]Optimizing the diameter of an admissible 105-tuple can be done numerically. However, we have not checked the computations ourselves!

The next result is established similarly:

Theorem 5.15 (Theorem 1.4, [Ma]). *Assuming the Elliott-Halberstam Conjecture (i.e. Conjecture 4.1) we get*

$$\liminf_n (p_{n+2} - p_n) \le 600 \quad \textit{and} \qquad \liminf_n (p_{n+1} - p_n) \le 12.$$

Proof. For the first statement we again take $k = 105$. This time, according to Conjecture 4.1, we are allowed to pick $\theta = 1 - \epsilon$. Thus we can make $\epsilon > 0$ sufficiently small so that

$$\frac{\theta}{2} M_{105} > 2.$$

Therefore we can run the above argument arriving at $\rho > 2$ (i.p. $\lfloor \rho + 1 \rfloor \ge 3$), so that

$$\liminf_n (p_{n+2} - p_n) \le \max_{1 \le i,j \le 105} (h_i - h_j).$$

We get the desired bound by choosing an appropriate admissible 105-tuple.

Alternatively we can run the argument with $k = 5$ and $\theta = 1 - \epsilon$. In this situation Proposition 5.10 allows us to obtain

$$\frac{\theta}{2} M_5 > 1.$$

Thus, for any admissible 5-tuple $\mathcal{H}$ we get

$$\liminf_n (p_{n+1} - p_n) \le \max_{1 \le i,j \le 5} (h_i - h_j).$$

Here we can write down $\mathcal{H} = \{0, 2, 6, 8, 12\}$ obtaining the desired result.[22] □

Theorem 5.16 (Theorem 1.1, [Ma]). *For $m \in \mathbb{N}$ we have*

$$\liminf_n (p_{n+m} - p_n) \ll m^3 e^{4m}.$$

Proof. The idea of proof is as above. We take $\theta = \frac{1}{2} - \epsilon$ as above, but we now take k sufficiently large. Then, by Proposition 5.10 we get

$$\frac{\theta}{2} M_k \ge \left(\frac{1}{4} - \frac{\epsilon}{2}\right)(\log(k) - 2\log\log(k) - 2).$$

Here we can put $\epsilon = 1/k$ and observe that $\frac{\theta}{2} M_k > m$ if $k \ge C m^2 e^{4m}$. The same argument we have seen essentially three times already now shows that at least $m+1$ of the number $n + h_1, \ldots, n + h_k$ must be prime. We can now choose the admissible k-tuple by putting

$$\mathcal{H} = \{p_{\pi(k)+1}, \ldots, p_{\pi(k)+k}\}.$$

We claim that this is admissible. Indeed, none of the elements is a multiple of primes less than k. Since there are only k elements we can also not cover all residue classes for primes bigger than k.

[22] It is an easy exercise to check that this is admissible.

We conclude that

$$\liminf_n (p_{n+m} - p_n) \ll (p_{\pi(k)+1} - p_{\pi(k)+k}) \ll k \log(k) \ll m^3 e^{4m},$$

with $k = \lceil Cm^2 e^{4m} \rceil$. This finishes the proof. □

Theorem 5.17 (Theorem 1.2, [Ma])**.** *Let $m \in \mathbb{N}$ and let $r \in \mathbb{N}$ be sufficiently large (in terms of m). For a set $\mathcal{A} = \{a_1, \dots, a_r\}$ of r distinct integers we have*

$$\frac{\sharp\{\mathcal{H} \subseteq \mathcal{A} \colon n + h_1, \dots, n + h_m \text{ are simultaneously prime infinitely often}\}}{\sharp\{\mathcal{H} \subseteq \mathcal{A}\}} \gg_m 1$$

Proof. As above we take $k = \lceil Cm^2 e^{4m} \rceil$. Our argument from the proof of Theorem 5.16 shows that for each admissible k-tuple $\mathcal{H} = \{h_1, \dots, h_k\}$ there exists a subset $\{h'_1, \dots, h'_m\} \subseteq \mathcal{H}$ such that there are infinitely many integers n for which $n + h'_1, \dots, n + h'_m$ are all prime.

Starting from $\mathcal{A}$ we will now construct a set $\mathcal{A}_2$ such that any subset of $\mathcal{A}_2$ with k elements must be admissible. This is done by simply removing the residue class modulo p containing the fewest elements for all primes $p \leq k$. A quick count shows

$$s = \sharp\mathcal{A}_2 \geq r \prod_{p \leq k} (1 - 1/p) \gg_m r.$$

We can now finish that argument by first observing that there are $\binom{s}{k}$ subsets $\mathcal{H} \subseteq \mathcal{A}_2$ all of which are admissible by construction. Each of them contains at least one subset $\{h'_1, \dots, h'_m\}$ with the desired property. Observe that any admissible set $\mathcal{B} \subseteq \mathcal{A}_2$ of size m is contained in $\binom{s-m}{k-m}$ sets $\mathcal{H} \subseteq \mathcal{A}_2$ of size k. Thus there are at least

$$\binom{s}{k} \binom{s-m}{k-m}^{-1} \gg_m s^m \gg_m r^m$$

admissible sets $\mathcal{B} \subseteq \mathcal{A}_2$ of size m which satisfy the prime m-tuple conjecture. We are done since in total there are $\binom{r}{m} \leq r^m$ subsets $\{h_1, \dots, h_m\} \subseteq \mathcal{A}$. □

References

[AZ] M. Aigner, G. M. Ziegler, *Proofs from The Book*, Sixth edition, Springer, Berlin, 2018. viii+326 pp.

[FI] J. Friedlander, H. Iwaniec, *Opera de cribro*, American Mathematical Society Colloquium Publications, **57**. American Mathematical Society, Providence, RI, 2010.

[HLP] G. H. Hardy, J. E. Littlewood, G. Pólya, *Inequalities*, Reprint of the 1952 edition. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1988. xii+324 pp.

[HR] H. Halberstam, H.-E. Richert, *Sieve methods*, London Mathematical Society Monographs, No. 4. Academic Press [Harcourt Brace Jovanovich, Publishers], London-New York, 1974.

[IK] H. Iwaniec, E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, 53. American Mathematical Society, Providence, RI, 2004. xii+615 pp.

[Ma] J. Maynard, *Small gaps between primes*, Ann. of Math. (2) **181** (2015), no. 1, 383–413.

[Mo] Y. Motohashi, *Lectures on sieve methods and prime number theory*, Tata Institute of Fundamental Research Lectures on Mathematics and Physics, **72**. Published for the Tata Institute of Fundamental Research, Bombay; by Springer-Verlag, Berlin, 1983.

[Po] D. H. J. Polymath, *Variants of the Selberg sieve, and bounded intervals containing many primes*, `arXiv:1407.4897`.

[Te] G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, Translated from the second French edition (1995) by C. B. Thomas. Cambridge Studies in Advanced Mathematics, 46. Cambridge University Press, Cambridge, 1995.