EXPLORING RECENT DEVELOPMENTS IN GAPS BETWEEN PRIME NUMBERS

Seth Hardy

Supervised by Dr Ian Petrow

University College London

March 2022

Abstract

One of the most celebrated results regarding prime numbers in recent years states that there are infinitely many consecutive primes no more than 246 apart. This essay will give a detailed proof of the weaker statement that there are infinitely many consecutive primes no more than 270 apart, first studying sieve theory to motivate the ideas behind the proof. A recent result conjectured by Erdős regarding large gaps between primes will also be proved, with a particular focus on the surprising connection between small and large gaps between primes. One vital theorem used in these proofs is the Bombieri-Vinogradov theorem, which will also be motivated, and a proof will be given.

Contents

1 Introduction

The central object of study in this essay will be gaps between prime numbers. We let p_n denote the n-th prime, so that $p_1 = 2,\ p_2 = 3,\ p_3 = 5,\ p_4 = 7,\ p_5 = 11...$. The "n-th prime gap" is then defined by $g_n = p_{n+1} - p_n$, forming an infinite sequence

$$1, 2, 2, 4, 2, 4, 2, 4, 6, 2, 6... \ .$$

Elements of this sequence are referred to throughout as "prime gaps" or "gaps between consecutive primes".

1.1 Conjectures on Small Gaps

In 1849, French mathematician Alphonse De Polignac conjectured that for any even number n, there are infinitely many prime gaps of size n. Indeed, with the exception of 2, primes are odd, and so the gap between any two consecutive primes above 2 is even. De Polignac's conjecture eluded the efforts of mathematicians for over a century, when, in 2013, multiple breakthroughs were made. The best current result proves that De Polignac's conjecture is true for some $n \leq 246$, courtesy of the Polymath8b project [1].

The fact that primes get increasingly sparse as we move up the number line means that the average spacing between the primes gets larger and larger. Therefore, proving that some fixed gap appears infinitely often is no trivial matter. The hope is to reduce 246 down to 2, proving what is knows as the twin prime conjecture: the special case of De Polignac's conjecture for $n = 2$. A pair of primes that are 2 apart is called a twin prime pair, and a twin prime is a number that is part of some twin prime pair.

Despite the simplicity of the statement, the twin prime conjecture is one of the most notorious open problems in number theory. In the early 1900's, in an attempt to prove the twin prime conjecture, Viggo Brun tried to estimate the sum of the reciprocals of all twin prime pairs. That is

$$\sum_{p:\, p+2 \text{ prime}} \left(\frac{1}{p} + \frac{1}{p+2}\right) = \left(\frac{1}{3} + \frac{1}{5}\right) + \left(\frac{1}{5} + \frac{1}{7}\right) + \left(\frac{1}{11} + \frac{1}{13}\right) + \dots \ .$$

If Brun could show that this sum was infinite, then there must be infinitely many twin primes! Unfortunately, Brun found that this sequence actually converges[1]. We know that $\sum_p 1/p = \infty$, and so twin primes are certainly very rare in the set of primes. Despite this, the twin prime conjecture is still generally believed to be true, and recent developments have encouraged this sentiment.

If we consider the set $\mathcal{H} = \{0, 2\}$, then the twin prime conjecture is equivalent to the existence of infinitely many $n \in \mathbb{N}$ where the shifts of $\mathcal{H}$ by n (i.e. $n + 0$ and $n + 2$) are both prime.

[1] The constant to which the sum converges is known as Brun's constant, $B_2 \approx 1.902160583104$. This is an approximation using computational estimates, but it has been proven unconditionally that $B_2 < 2.4$.

More generally, it's conjectures that for certain sets $\mathcal{H} = \{h_1, ..., h_k\}$, there exist infinitely many n where every element of $\{n + h_1, ..., n + h_k\}$ is prime. Indeed, our choice of $\mathcal{H}$ is important. If we take $\mathcal{H} = \{0, 1\}$, then one of our shifts n or $n+1$ is even, so both cannot simultaneously be prime for any $n > 2$. Similarly, if we take $\mathcal{H} = \{0, 2, 4\}$, then one of our shifts $n, n+2$ or $n+4$ is always a multiple of 3. The underlying factor here is that $\{0, 1\}$ covers all residue classes modulo 2, and $\{0, 2, 4\}$ covers all residue classes modulo 3.

In fact, if $\mathcal{H} = \{h_1, ..., h_k\}$ covers $\mathbb{Z}/p\mathbb{Z}$ for any prime p (in the sense that every element of $\mathbb{Z}/p\mathbb{Z}$ is congruent to some $h_1, ..., h_k$ modulo p), then for any $n \in \mathbb{N}$, we find that $p | n + h_i$ for some $i \in \{1, ..., k\}$. Therefore we know at least one of $n + h_1, ..., n + h_k$ is always a multiple of p, regardless of n. To avoid this case, we only consider sets $\mathcal{H} = \{h_1, ..., h_k\}$ that do not cover $\mathbb{Z}/p\mathbb{Z}$ for any prime p.

Definition 1.1. A set $\mathcal{H} = \{h_1, ..., h_k\}$ is called *admissible* if it does not cover $\mathbb{Z}/p\mathbb{Z}$ for any prime p. Without loss of generality, we will assume that $h_1, ..., h_k$ are in increasing order.

This aforementioned conjecture is known as the prime k-tuples conjecture. It states that for any admissible set $\mathcal{H} = \{h_1, ..., h_k\}$, there are infinitely many $n \in \mathbb{N}$ such that all of $n + h_1, ..., n + h_k$ are prime. Building on this, the Hardy-Littlewood conjecture gives a precise asymptotic estimate for the size of the set $\{n \leq x : n+h_1, ..., n+h_k \text{ are all prime}\}$. These more general conjectures are far beyond the reach of our current techniques.

1.2 Conjectures on Large Gaps

Gaps between consecutive primes can get arbitrarily large. Indeed, for any $n \in \mathbb{N}$ we can manufacture a gap of size $n-1$ by observing that the string of numbers $\{n! + i\}_{i=2}^{n}$ are all composite, as the number being added divides both itself and $n!$. Large gap conjectures subsequently focus on estimating the size of the largest prime gap amongst primes $\leq X$ as $X \in \mathbb{R}$ grows. In 1936, Cramér conjectured results about large gaps [2] using his random model of primes.

Recall that the prime number theorem tells us that there are $\pi(n) \sim n/\log n$ prime numbers less than or equal to n. Therefore, we expect the average spacing of primes $\leq n$ to be about $\log n$. If we have a prime for every $\approx \log n$ integers, we may expect that the probability of some randomly chosen n being prime is about $1/\log n$.

Subsequently, Cramér's random model consists of independent random variables $(X_n)_{n \geq 3}$, where $X_n = 1$ models the event that n is prime, and $X_n = 0$ the event that n is composite. Motivated by this, the model takes $\mathbb{P}(X_n = 1) = 1/\log n$ and $\mathbb{P}(X_n = 0) = 1 - 1/\log n$. We set $X_2 = 1$ a.s. for convenience.

We let $P_n = \min\{m : \sum_{i \leq m} X_i = n\}$ be a random variable modelling the position of the n-th prime. Also, for some $c \in \mathbb{R}_{>0}$ we let E_m denote the event that all of $X_{m+1}, X_{m+2}, ..., X_{\lfloor m+c(\log m)^2 \rfloor}$ are zero, modelling the occurrence of prime gap of size at least $c(\log m)^2$ following the integer m. It is therefore intuitive that the following events have the same probability:

(i) $P_{n+1} - P_n > c(\log P_n)^2$ for infinitely many n.

(ii) E_n are realised an infinite number of times (or (E_n i.o.)).

Concentrating on (ii), we note that

$$\mathbb{P}(E_m) = \prod_{i=1}^{\lfloor c(\log m)^2 \rfloor} \Big(1 - \frac{1}{\log(m+i)}\Big).$$

Now using the facts that $e^{-x} \geq 1 - x$ and $(1 - 1/n)^n \to e^{-1}$, we can find $A, B > 0$ so that for large m we have

$$\frac{A}{m^c} < \mathbb{P}(E_m) < \frac{B}{m^c}.$$

Therefore, if $c > 1$, we have $\sum \mathbb{P}(E_m) < \infty$, so Borel-Cantelli Lemma 1 implies that $\mathbb{P}(E_m \text{ i.o.}) = 0$.

Considering the case $c < 1$, we observe that events $E_{m_1}, E_{m_2}, ..$ where $m_1 = 2$ and $m_{i+1} = m_i + c(\log m_i)^2 + 1$ are disjoint, and subsequently independent (as X_i are independent). By induction we can show that there exists $K > 0$ (dependent on c) such that $m_r < Kr(\log r)^2$. Then using the above and recalling $c < 1$, we have

$$\sum_{r \geq 1} \mathbb{P}(E_{m_r}) > \sum_{r \geq 1} \frac{A}{K^c r^c (\log r)^{2c}} = \infty.$$

As (E_{m_i}) are mutually independent, Borel-Cantelli Lemma 2 implies that $\mathbb{P}(E_{m_i} \text{ i.o.}) = 1$. Therefore $\mathbb{P}(E_m \text{ i.o.}) = 1$.

Putting this all together using the equal probability of the aforementioned events, we have

$$\mathbb{P}(P_{n+1} - P_n > c(\log P_n)^2 \text{ i.o.}) = \begin{cases} 0, & \text{if } c > 1 \\ 1, & \text{if } c < 1 \end{cases}.$$

Therefore,

$$\mathbb{P}\Big(\limsup_{n\to\infty} \frac{P_{n+1} - P_n}{(\log P_n)^2} = 1\Big) = 1.$$

This heuristic lead Cramér to conjecture that the primes satisfy

$$\limsup_{n\to\infty} \frac{p_{n+1} - p_n}{(\log p_n)^2} = 1,$$

which, using the fact that $p_n \sim n \log n$ from the prime number theorem, suggests the estimate

$$\max_{p_n \leq x} (p_{n+1} - p_n) \geq (1 - o(1))(\log x)^2.$$

We note that Cramér's model certainly has its flaws. For example, it gives a nonzero probability that both n and $n+1$ are prime, despite the fact that one of these numbers is even. Granville [3] has since improved Cramér's model to acknowledge the effect of relatively small primes, leading to the conjecture that

$$\max_{p_n \leq x} (p_{n+1} - p_n) \geq (1 - o(1)) 2e^{-\gamma} (\log x)^2,$$

where γ is the Euler-Mascheroni constant. Both of these conjectures are far beyond our best current results.

1.3 Developments in Small Prime Gaps

Following De-Polignac's conjecture in 1849, very little progress was made towards any results for over a century. Then, in 2005, a paper published by Goldston, Pintz and Yildirim (referred to throughout as GPY) [4] revolutionised how we approach the problem of bounded gaps between primes. The main result of the paper showed that

$$\liminf_{n\to\infty} \frac{p_{n+1}-p_n}{\log n} = 0.$$

Recalling that $\log n$ is the average gap between p_{n+1} and p_n, the result shows that for any fixed multiple of the average prime gap, say $\epsilon \log n$, there are infinitely many n-th prime gaps that are within $\epsilon \log n$ of one another. This brings us very close to bounded gaps between primes.

In 2013, building on the work of GPY, Zhang [5] published the first proof showing bounded gaps between prime numbers, with the result

$$\liminf_{n\to\infty} (p_{n+1}-p_n) < 7 \times 10^7,$$

proving that there are infinitely many examples of consecutive prime numbers that are less than 7×10^7 apart. Only a few months later, Maynard [6] released his paper showing bounded gaps between primes, with a smart idea to generalise the approach in the GPY paper. This approach led to the result

$$\liminf_{n\to\infty} (p_{n+1}-p_n) \leq 600, \tag{1.01}$$

utilising far more elementary techniques than those in Zhang's paper. The arguments of Maynard were later optimised by the Polymath8b project [1] to arrive at the aforementioned

$$\liminf_{n\to\infty} (p_{n+1}-p_n) \leq 246.$$

In Section 4 we shall follow Maynard's proof of (1.01) to prove the result

$$\liminf_{n\to\infty} (p_{n+1}-p_n) \leq 270,$$

with help in the final stages from the Polymath project.

1.4 Developments in Large Prime Gaps

In 1938, only 2 years after Cramér published his conjecture on the size of prime gaps, it was shown by Rankin [7] that for some small constant $c > 0$, we have

$$\limsup_{n\to\infty} \frac{p_{n+1}-p_n}{L(n)} > c, \; L(n) = \frac{\log(n)\log_2(n)\log_4(n)}{(\log_3(n))^2},$$

where $\log_2(n) := \log(\log(n))$, with $\log_3, \log_4$ defined analogously. It was conjectured by Erdős that this constant can be taken arbitrarily large. Whilst small results were made improving the constant c, the question remained open for the remainder of the 1900's.

In 2014, on consecutive days, Ford, Green, Konyagin and Tao [8], followed by Maynard [9], released their papers, proving, by different methods, the conjectured result

$$\limsup_{n\to\infty} \frac{p_{n+1}-p_n}{L(n)} = \infty,\ L(n) = \frac{\log(n)\log_2(n)\log_4(n)}{(\log_3(n))^2}. \tag{1.02}$$

In essence, this states that there are infinitely many examples of n-th primes p_n such that $p_{n+1} - p_n$ is larger than any fixed multiple of $L(n)$. The result was later improved by Maynard, Ford, Green, Konyagin and Tao [10], to arrive at

$$\limsup_{n\to\infty} \frac{p_{n+1}-p_n}{L(n)} > C,\ L(n) = \frac{\log(n)\log_2(n)\log_4(n)}{\log_3(n)},$$

for some constant $C > 0$.

In Section 5 we will present the main details of Maynard's proof for (1.02), with a particular focus on how the proof hinges on previous developments regarding small gaps between primes.

To gain a broader understanding, we will first look at the foundations of these proofs: in particular sieve theory (Section 2) and the Bombieri-Vinogradov Theorem (Section 3). We will utilise many estimates throughout this essay, which may be found in the Appendix (Section 6).

1.5 Notation

The following notation will be used throughout, unless stated otherwise.

p, q	Denote prime numbers, e.g. $\sum_{q\in N} 1$ will count the number of primes in N.
p_n	The n-th prime number.
$\mathcal{H}$	An admissible set, often denoted $\{h_1, ..., h_k\}$.
$\#\mathcal{T}, \lvert\mathcal{T}\rvert$	The size of a set $\mathcal{T}$
τ_k	The $k-$divisor function, equal to the k-fold convolution $1 * 1 * ... * 1$.
$\mathbb{N}$	The set of natural numbers $\{1, 2, 3, ...\}$
P_W	The primorial of W, $\prod_{p\le W} p$.
$\sum_{a_1,...,a_k}$	The k-fold summation over natural numbers $\sum_{a_1\ge 1} \cdots \sum_{a_k\ge 1}$.
$O_k, \ll_k$	Standard asymptotic upper bounds where the implied constant depends on k.
o_k	The standard loose asymptotic upper bound where the implied constant depends on k.
$P^+(n)$	The largest prime factor of n.
$\log_k(x)$	The k-times iterated logarithm $\log(\log ...(\log x))$.
id	The (arithmetic) identity function, $id(n) = n$.
$\mathbf{1}$	The indicator function. For example, $\mathbf{1}_p(n)$ will denote the indicator for the primes, whereas $\mathbf{1}(n\mid b)$ denotes an indicator for divisors of b.

1.6 Acknowledgements

This essay is mainly based on the two papers: *"Small gaps between primes"* [6] and *"Large gaps between primes"* [9] by James Maynard.

The author would like to express his thanks to Dr I.Petrow for many great conversations, and for the initial suggestion of the project. He is especially grateful to Dr Petrow for introducing him to the subject, inspiring the author's own pursuit in the field.

2 Sieve Theory and Sieve Weights

2.1 Sieve Theory

Sieve theory is a branch of mathematics that aims to estimate the number of elements in a set that are either prime or have no small prime factors. We begin with a finite set of integers $\mathcal{A}$, a set of primes $\mathcal{P}$, and some number $z \geq 2$. If we denote $\mathcal{P}_z = \{p \in \mathcal{P} : p \leq z\}$, then the standard problem in sieve theory is trying to estimate the quantity

$$\mathcal{S}(\mathcal{A}; \mathcal{P}, z) = \sum_{a \in \mathcal{A}} \delta((a, \mathcal{P}_z)), \tag{2.01}$$

where $(a, \mathcal{P}_z); = (a, \prod_{p \in \mathcal{P}_z} p)$. This quantity describes the number of elements in $\mathcal{A}$ that have no prime factors that are both in $\mathcal{P}$ and less than or equal to z. In the language of sieve theory, we sift out from $\mathcal{A}$ all multiples of primes ($\leq z$) that are in our sieve $\mathcal{P}$.

To motivate this, let us try and formulate $\mathcal{S}(\mathcal{A}; \mathcal{P}, z)$ so that it tells us the number of twin prime pairs in a certain interval. We will take $\mathcal{A} = \{n(n+2) : 1 < n \leq x\}$, $\mathcal{P}$ the set of all primes and $z = \sqrt{x+2}$. Note that for $1 < n \leq x$, n and $n+2$ are either both prime or one of them has a prime factor $p \leq \sqrt{n+2} \leq \sqrt{x+2}$. Therefore if we sift out all elements of $\mathcal{A}$ with prime factors $p \leq \sqrt{x+2}$ then we are left with cases where $n(n+2)$ is the product of two primes that are each bigger than $\sqrt{x+2}$, hence n and $n+2$ are both prime. Therefore our sifted set is counting the number of twin prime pairs $(n, n+2)$ for $\sqrt{x+2} < n \leq x$. A good lower bound for this quantity could allow one to prove the twin prime conjecture. Methods for estimating $\mathcal{S}(\mathcal{A}; \mathcal{P}, z)$ are known as sieve methods, which can take many forms.

In the 1960's, Chen utilised sieve methods to prove that every sufficiently large even number is equal to a prime plus a product of two primes [11]. This statement is remarkably close to proving Goldbach's conjecture, which demonstrates the strength that sieve methods possess. In this essay, we will mainly study a form of the Selberg sieve, a sieve method developed by Atle Selberg in the 1940's. This method and variants of it are essential in the arguments of [4], [5], [6] and [9].

2.2 The Selberg Sieve

The Selberg sieve is a sieve method used to find an upper bound for $\mathcal{S}(\mathcal{A}; \mathcal{P}, z)$. Suppose we have some arithmetic function $\lambda : \mathbb{N} \to \mathbb{R}$ where $\lambda_1 = 1$. Then we have the inequality

$$\delta(k) = \sum_{d|k} \mu(d) \leq (\sum_{d|k} \lambda_d)^2.$$

We proceed by inserting this inequality into (2.01) to find an upper bound. The advantages of doing this will be made clear in the following example.

Example 2.1. Consider $\mathcal{A} = \{n \in \mathbb{N} : n \leq x\}$, $\mathcal{P}$ is the set of all prime numbers and $z \in \mathbb{R}$. Now $\mathcal{S}(\mathcal{A}; \mathcal{P}, z)$ is counting the number of integers $\leq x$ with no prime factors $\leq z$.

We use the notation $d|\mathcal{P}_z$ to describe d that are multiples of elements of $\mathcal{P}_z$. Now by the above inequality, we have

$$\begin{aligned}\mathcal{S}(\mathcal{A};\mathcal{P},z) &\leq \sum_{n\leq x}\Big(\sum_{d|(n,\mathcal{P}_z)}\lambda_d\Big)^2 \leq \sum_{n\leq x}\Big(\sum_{d_1,d_2|(n,\mathcal{P}_z)}\lambda_{d_1}\lambda_{d_2}\Big)\\ &\leq \sum_{\substack{d_1|\mathcal{P}_z\\ d_2|\mathcal{P}_z}}\lambda_{d_1}\lambda_{d_2}\sum_{\substack{n\leq x\\ [d_1,d_2]|n}}1 \leq \sum_{\substack{d_1|\mathcal{P}_z\\ d_2|\mathcal{P}_z}}\lambda_{d_1}\lambda_{d_2}\Big(\frac{x}{[d_1,d_2]}+O(1)\Big)\\ &\leq x\sum_{\substack{d_1|\mathcal{P}_z\\ d_2|\mathcal{P}_z}}\frac{\lambda_{d_1}\lambda_{d_2}}{[d_1,d_2]}+O\Big(\sum_{d_1,d_2|\mathcal{P}_z}|\lambda_{d_1}||\lambda_{d_2}|\Big).\end{aligned}$$

To make calculations feasible, we assume that $\lambda_d = 0$ for $d > z$, then we have the estimate

$$\mathcal{S}(\mathcal{A};\mathcal{P},z) \leq x\sum_{d_1,d_2\leq z}\frac{\lambda_{d_1}\lambda_{d_2}}{[d_1,d_2]}+O\Big(\sum_{d_1,d_2\leq z}|\lambda_{d_1}||\lambda_{d_2}|\Big). \tag{2.02}$$

The brilliance of the Selberg sieve is that the sum in our main term can now be realised as a quadratic form in $(\lambda_d)_{d\leq z}$, which we seek to minimise to obtain a good upper bound. Unfortunately, the term $[d_1, d_2]$ links the variables in the sum. It is tempting to write $\frac{1}{[d_1,d_2]} = \frac{(d_1,d_2)}{d_1 d_2}$ and use möbius inversion to untangle (d_1, d_2). Instead, we shall utilise the fact[2] that $\sum_{d|n}\varphi(d) = n$. This leads to the estimate

$$\begin{aligned}\sum_{d_1,d_2\leq z}\frac{\lambda_{d_1}\lambda_{d_2}}{[d_1,d_2]} &= \sum_{d_1,d_2\leq z}\frac{\lambda_{d_1}\lambda_{d_2}(d_1,d_2)}{d_1d_2}\\ &= \sum_{d_1,d_2\leq z}\frac{\lambda_{d_1}\lambda_{d_2}}{d_1d_2}\sum_{r|(d_1,d_2)}\varphi(r)\\ &= \sum_{r\leq z}\varphi(r)\sum_{\substack{d_1,d_2\leq z\\ r|(d_1,d_2)}}\frac{\lambda_{d_1}\lambda_{d_2}}{d_1d_2}\\ &= \sum_{r\leq z}\varphi(r)\Big(\sum_{\substack{d\leq z\\ r|d}}\frac{\lambda_d}{d}\Big)^2.\end{aligned}$$

We now perform a substitution

$$u_r = \sum_{\substack{d\leq z\\ r|d}}\frac{\lambda_d}{d}.$$

It is clear that u_r inherits the property from λ_d that $u_r = 0$ for $r > z$. We have now diagonalised the quadratic form to obtain a main term

$$x\sum_{r\leq z}\varphi(r)u_r^2, \tag{2.03}$$

[2]Note that for $d|n$ there are $\varphi(d)$ elements of order d in the cyclic group of order n.

which we shall minimise subject to the condition that $\lambda_1 = 1$ It is important that this change of variables is invertible, so that we can also estimate our error term. Using möbius inversion, we have

$$
\begin{aligned}
\lambda_d &= \sum_{a\geq 1} \frac{\lambda_{ad}}{a}\delta(a) = \sum_{a\geq 1} \frac{\lambda_{ad}}{a} \sum_{r|a} \mu(r) \\
&= \sum_{r\geq 1} \mu(r) \sum_{\substack{a\geq 1 \\ r|a}} \frac{\lambda_{ad}}{a} = \sum_{r\geq 1} \frac{\mu(r)}{r} \Big(\sum_{b\geq 1} \frac{\lambda_{bdr}}{b} \Big) \\
&= d \sum_{r\geq 1} \mu(r) u_{dr}.
\end{aligned}
$$

So we now wish to minimise $\sum_{r\leq z} \varphi(r) u_r^2$ subject to $\lambda_1 = 1$, which is gives the relation

$$
1 = \sum_{r\leq z} \mu(r) u_r. \tag{2.04}
$$

Applying Cauchy-Schwarz to this sum gives

$$
1 \leq \Big(k \sum_{r\leq z} \frac{\mu(r)^2}{\varphi(r)} \Big)^{1/2} \Big(\sum_{r\leq z} \frac{\varphi(r)}{k} u_r^2 \Big)^{1/2},
$$

for constant k. Therefore

$$
\sum_{r\leq z} \varphi(r) u_r^2 \geq \Big(\sum_{r\leq z} \frac{\mu(r)^2}{\varphi(r)} \Big)^{-1}.
$$

Here the left hand side controls the size of our main term (2.03), which we wish to minimise. Cauchy-Schwarz would have given equality if we had

$$
u_r = \frac{k\mu(r)}{\varphi(r)},
$$

therefore this is our choice of u_r to minimise $\sum_{r\leq z} \varphi(r) u_r^2$. Furthermore, the constraint (2.04) implies that

$$
k = \Big(\sum_{r\leq z} \frac{\mu(r)^2}{\varphi(r)} \Big)^{-1}.
$$

We substitute this choice of u_r into our main term (2.03), and (2.02) becomes

$$
\mathcal{S}(\mathcal{A}; \mathcal{P}, z) \leq x \Big(\sum_{r\leq z} \frac{\mu(r)^2}{\varphi(r)} \Big)^{-1} + O\Big(\sum_{d_1, d_2 \leq z} |\lambda_{d_1}||\lambda_{d_2}| \Big).
$$

It can be shown using our relation $\lambda_d = d\sum_r \mu(r) u_{dr}$ that $\sum_{d\leq z} |\lambda_d| \ll z$. Also (6.04) gives

$$
\sum_{r\leq z} \frac{\mu(r)^2}{\varphi(r)} = \log z + O(1),
$$

so we have

$$\mathcal{S}(\mathcal{A};\mathcal{P},z) \le \frac{x}{\log z + O(1)} + O(z^2).$$

Now taking $z = x^{1/2}/\log x$, we find that for large x,

$$\pi(x) - \pi(\sqrt{x}) \le \frac{2x}{\log x} + O\Big(\frac{x}{(\log x)^2}\Big).$$

Which is only a constant out from what we could wish for. We achieved this estimate using very little machinery from analytic number theory.

2.3 Probability Measures and Small Gaps between Primes

Instead of simply using the Selberg sieve to solve classical sieve problems by estimating $\mathcal{S}(\mathcal{A},\mathcal{P},z)$, we can use it to create a probability measure. Suppose we have an admissible set $\mathcal{H} = \{h_1, ..., h_k\}$ and we want to prove that the shifted set $n+\mathcal{H} = \{n+h_1, ..., n+h_k\}$ contains at least ρ primes for infinitely many n. Consider some weights $w_n \ge 0$ on the set $[N, 2N)$. If we can show

$$\sum_{N\le n<2N} \Big(\sum_{i=1}^{k} \mathbf{1}_p(n+h_i) - \rho\Big) w_n > 0, \tag{2.05}$$

then we know that there is some $n \in [N, 2N)$ such that more than ρ of $n+h_1, ..., n+h_k$ are prime. Now, if we can show that this is true for infinitely many N, then we can find infinitely many $n \in \mathbb{N}$ such that more than ρ of $n+h_1, ..., n+h_k$ are prime. If $\rho \ge 1$ this implies that there are infinitely many pairs of primes that are within $h_k - h_1$ of one another, and so we obtain bounded gaps between primes!

It is helpful to view this method from a probabilistic perspective, where we are trying to obtain a probability measure on $[N, 2N)$ that is concentrated on n where the set $\{n+h_1, ..., n+h_k\}$ contains many primes. In our case, the probability measure would be

$$\mathbb{P}_{[N,2N)}(n) = \frac{w_n}{\sum_{N\le n<2N} w_n}.$$

And so if we consider $\sum_{i=1}^{k} \mathbf{1}_p(n+h_i)$ as a random variable on the space $[N, 2N)$, we find that showing (2.05) is equivalent to proving

$$\mathbb{E}_{[N,2N)}\Big(\sum_{i=1}^{k} \mathbf{1}_p(n+h_i)\Big) = \sum_{N\le n<2N} \Big(\sum_{i=1}^{k} \mathbf{1}_p(n+h_i)\Big)\mathbb{P}_{[N,2N)}(n) > \rho.$$

Therefore the problem of proving there are infinitely many bounded gaps between primes can be rephrased as trying to maximise the expected value of $\sum_{i=1}^{k} \mathbf{1}_p(n+h_i)$ on $[N, 2N)$ by creating a suitable probability measure. However, this task is not trivial. We need our weights w_n to be nice enough so that we can calculate them unconditionally. This rules out choices such at $w_n = \prod_i \mathbf{1}_p(n+h_i)$ (which would certainly be concentrated on the desired set), as knowing that such a choice had a nonzero sum over $[N, 2N)$ would be equivalent to proving that all of $n+h_1, ..., n+h_k$ are primes for some $n \in [N, 2N)$, which is harder than our original problem! On the other hand, if we take $w_n = 1$ to induce a uniform probability distribution, the decreasing density of the primes means that (2.05) will never be achieved.

This is where the Selberg Sieve proves incredibly useful, allowing us to produce weights which we can calculate unconditionally whilst also allowing the flexibility to optimise the weights so that they are concentrated on n where many of $n+h_1, ..., n+h_k$ are prime.

2.4 The GPY Sieve for Small Gaps

The first people to obtain a significant result by working with the likes of (2.05) were GPY in [12]. Their method was to optimise the Selberg sieve weights $w_n = (\sum_{d|P(n)} \lambda_d)^2$, where $P(n) = \prod_{i=1}^k (n+h_i)$, to create weights that are concentrated on n where this polynomial has few prime factors. They found that optimal weights had

$$\lambda_d = \mu(d)\Big(\log \frac{R}{d}\Big)^{k+l},$$

where $d < R$ and $l = k + \lfloor \sqrt{k}/2 \rfloor$, giving weights of the form

$$w_n = \Big(\sum_{\substack{d|P(n)\\ d<R}} \mu(d)\Big(\log \frac{R}{d}\Big)^{k+l}\Big)^2. \tag{2.06}$$

It shouldn't be too surprising that the weights look like this. With the knowledge that the Von-Mangoldt function $\Lambda(n) = \sum_{d|n} \mu(d) \log(n/d)$ is supported on prime powers, it is natural to think that closely related functions may be good candidates for the weight w_n (after being squared, to ensure positivity).

We need to make adjustments to the classic Von-Mangoldt function to obtain such weights. The function

$$\Lambda(P(n)) = \sum_{d|P(n)} \mu(d) \log \frac{n}{d}$$

will be zero unless $P(n)$ is a prime power, meaning shifts n where more than one of $n+h_i$ are prime will not survive. Therefore we wish to enlarge our support, to potentially allow all of $n+h_1, ..., n+h_k$ to be prime.

Remark 2.2. The function $\Lambda_k(n) = \sum_{d|n} \mu(d)\Big(\log \frac{n}{d}\Big)^k$ is supported on n with at most k distinct prime factors. If we consider

$$\frac{d}{ds}\Big(\frac{\zeta^{(k)}(s)}{\zeta(s)}\Big) = \frac{\zeta^{(k+1)}(s)}{\zeta(s)} - \frac{\zeta^{(k)}(s)}{\zeta(s)}\frac{\zeta'(s)}{\zeta(s)},$$

we find that the left hand side is the Dirichlet series of $\log n \cdot \Lambda_k(n)$ and the right hand side is the Dirichlet series of $\Lambda_{k+1} - \Lambda_k * \Lambda$. We therefore have the relationship

$$\Lambda_{k+1}(n) = \Lambda_k(n) \log n + \Lambda_k * \Lambda(n).$$

It is fairly straightforward to use induction on the number of prime factors to show that Λ_k is supported on integers with $\leq k$ prime factors.

If we instead consider the k-th generalised Von-Mangoldt function, $\Lambda_k(n)$, we have support on n with at most k distinct prime factors. Therefore, weights of the form $w_n = (\Lambda_k(P(n)))^2$ will be positive on n where all of $n+h_i$ are prime, and zero if $P(n)$ has more than k prime factors. This seems like a good candidate. However, this gives zero weight on n where many of $n+h_i$ are prime and, for example, some badly behaved $n+h_j$ is composite with many prime factors. Subsequently, it happens that if we allow for a slightly increased support we obtain better results. Therefore the GPY weights look instead like (2.06), which is analogous to $(\Lambda_{k+l}(P(n)))^2$ but with the divisors of $P(n)$ being truncated at R and the logarithm being a function of R/d.

Indeed, we wish to take R as large as possible, but similarly to Example 2.1 (in which we took z small enough to control the error term), we need R to be small enough to allow calculation of the sum in (2.05) without introducing a large error term. In-fact, the limiting factor here is the requirement that R is small enough to allow for use of the Bombieri-Vinogradov theorem, which is a key result that allows us to estimate the size of the weighted sum.

3 The Bombieri-Vinogradov Theorem

The following section is based on lectures by Soundararajan [13], which have been made available online. In the interest of brevity, this section, unlike others, will rely on a number of results which are given without proof. The aim is therefore to motivate why certain results are important, and the connections they allow us to make.

Throughout this section, we shall use the notation $\log_k$ to denote the logarithm with base k, as oppose to the iterated logarithm.

Definition. We define the quantity

$$\pi(x; q, a) = \sum_{\substack{n \leq x \\ n \equiv a \pmod q}} \mathbf{1}_p(n),$$

and $\psi(x; q, a)$ in an analogous way using the Von-Mangoldt function.

Recent proofs regarding bounded gaps hinge on a strong result: The Bombieri-Vinogradov Theorem. The theorem provides an average for errors arising from the estimate $\pi(x; a, q) \sim \pi(x)/\varphi(q)$ over many residue classes q. In Maynard's proof for small gaps between primes, our weights w_n naturally give rise to an arithmetic progression in (2.05). This allows the use of the Bombieri-Vinogradov theorem, which gives a sufficiently small error to prove bounded gaps.

The proof of the Bombieri-Vinogradov theorem requires an explicit error term in the prime number theorem. We will use the following without proof.

Theorem 3.1. ***The Prime Number Theorem (with error term)*** *For some positive constant $c > 0$, we have*

$$\psi(x) = x + O(xe^{-c\sqrt{\log x}}).$$

And the statement of the theorem follows.

Theorem 3.2. ***The Bombieri-Vinogradov Theorem*** *For any $\theta < 1/2$, we have*

$$\sum_{q \leq x^\theta} \max_{(a,q)=1} \left| \pi(x; q, a) - \frac{\pi(x)}{\varphi(q)} \right| \ll_A \frac{x}{(\log x)^A}. \tag{3.01}$$

Definition 3.3. We define the *level of distribution* of the primes to be the supremal value of θ for which (3.01) holds. Therefore the Bombieri-Vonogradov theorem tells us that the primes have level of distribution $\geq 1/2$.

Conjecture 3.4. ***The Elliot-Halberstam Conjecture*** *The primes have level of distribution 1. That is, for any $\theta < 1$,* (3.01) *holds.*

3.1 The Large Sieve

One key ingredient in the proof of the Bombieri-Vinogradov theorem is a relation known as the multiplicative large sieve inequality. This result gives us a very strong upper bound on the average size of *twisted* series: that is, the average of series of the form $\sum \chi(n)a_n$. This strong bound is very useful, as averages over more general series in arithmetic progression can be written as twisted series using orthogonality of characters. This is the route we shall take to prove the Bombieri-Vinogradov theorem. We denote

$$e(x) = e^{2\pi i x},$$

take $(a_n)_{n=1}^N$ a vector in $\mathbb{C}^N$ and let $(\alpha_m)_{m=1}^M$ be a vector containing elements from $\mathbb{R}/\mathbb{Z}$. The space $\mathbb{R}/\mathbb{Z}$ is essentially the real numbers modulo 1, with the norm being the distance to the nearest integer. For example $\|0.6\| = 0.4$. We assume that our $\alpha_1, ..., \alpha_M$ are δ-spaced, meaning that $\|\alpha_j - \alpha_k\| \geq \delta \ \ \forall j \neq k$. Therefore if we picture $(\alpha_m)_{m=1}^M$ as points on a circle of circumference 1, each α_i will be separated from its neighbours by an arc of length at least δ. Finally, we define

$$S_1(x) = \sum_{n \leq N} a_n e(nx).$$

By relating exponentials to characters, the hope is that averages of this quantity can be utilised to prove results for averages of twisted series.

Theorem 3.5. ***The Additive Large Sieve Inequality*** *With $(a_n)_{n=1}^N$ complex numbers and $(\alpha_m)_{m=1}^M$ elements of $\mathbb{R}/\mathbb{Z}$ that are δ-spaced so that $\|\alpha_j - \alpha_k\| \geq \delta \ \ \forall j \neq k$, we have*

$$\sum_{m=1}^M |S_1(\alpha_m)|^2 \leq \Big(N - 1 + \frac{1}{\delta}\Big) \sum_{n=1}^N |a_n|^2.$$

This bound is tight: it's the best we could hope for. We shall not prove this theorem, but we shall instead prove the following lemma which is slightly weaker, but sufficient for our purpose.

Lemma 3.6. *Under the same assumptions as the above theorem, we have*

$$\sum_{m=1}^M |S_1(\alpha_m)|^2 \leq \Big(N + \frac{C \log M}{\delta}\Big) \sum_{n=1}^N |a_n|^2.$$

Proof. The sum that we wish to estimate is

$$\sum_{m=1}^M \Big| \sum_{n=1}^N a_n e(n\alpha_m) \Big|^2.$$

If we consider the Hilbert space $\mathbb{C}^M$ with the standard inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i \leq M} x_i \, \overline{y_i}$, this sum can be written as $\|\mathbf{A}\mathbf{a}\|^2$, where $\mathbf{A} \in \mathbb{C}^{M \times N}$ with $(\mathbf{A})_{m,n} = e(n\alpha_m)$ and $\mathbf{a} = (a_n)_{n=1}^N$.

We wish to obtain an estimate of the form $\|\mathbf{A}\mathbf{a}\|^2 \le \Delta^2\|\mathbf{a}\|^2$. From functional analysis, we know that the adjoint of $\mathbf{A}$ is $\mathbf{A}^* = \mathbf{A}^H$, the Hermitian matrix of $\mathbf{A}$. Furthermore, $\|\mathbf{A}\| = \|\mathbf{A}^*\|$, making following equivalent:

(i) $\|\mathbf{A}\mathbf{a}\|^2 \le \Delta^2\|\mathbf{a}\|^2 \;\; \forall \mathbf{a} \in \mathbb{C}^N$

(ii) $\|\mathbf{A}^*\mathbf{b}\|^2 \le \Delta^2\|\mathbf{b}\|^2 \;\; \forall \mathbf{b} \in \mathbb{C}^M$

Now instead we just need to get an estimate of the form (ii). This is in fact easier, as we have an outer sum up to N that gives us the N in the inequality. We have

$$(3.02) \qquad \|\mathbf{A}^*\mathbf{b}\| = \sum_{n=1}^{N}\Big|\sum_{m=1}^{M} b_m\overline{e(n\alpha_m)}\Big|^2 = \sum_{m_1=1}^{M}\sum_{m_2=1}^{M} b_{m_1}\overline{b_{m_2}}\sum_{n=1}^{N} e(n(\alpha_{m_2}-\alpha_{m_1})).$$

Observe that $\sum_{n=1}^{N} e(n(\alpha_{m_2}-\alpha_{m_1}))$ is a geometric series when $m_1 \neq m_2$, and $\sum_{n=1}^{N} e(nx) = (e(x)-e((N+1)x))/(1-e(x))$ in this case. Noting that $|1-e(x)| = |2\sin(\pi x)| \ge 2||x||_{\mathbb{R}/\mathbb{Z}}$ for $x \in [0,1]$, we have

$$\Big|\sum_{n=1}^{N} e(n(\alpha_{m_2}-\alpha_{m_1}))\Big| \le \begin{cases} N & m_1 = m_2 \\ \frac{1}{\|\alpha_{m_2}-\alpha_{m_1}\|} & m_1 \neq m_2 \end{cases}.$$

So from (3.02) we obtain

$$\|\mathbf{A}^*\mathbf{b}\| \le N\sum_{m=1}^{M}|b_m|^2 + \sum_{m_1=1}^{M}\sum_{\substack{m_2=1\\ m_2\neq m_1}}^{M}|b_{m_1}||b_{m_2}|\frac{1}{\|\alpha_{m_2}-\alpha_{m_1}\|}.$$

We use the inequality $2|b_{m_1}||b_{m_2}| \le |b_{m_1}|^2 + |b_{m_2}|^2$ and the symmetry of the sum to write

$$\begin{aligned}\|\mathbf{A}^*\mathbf{b}\| &\le N\sum_{m=1}^{M}|b_m|^2 + \sum_{m_1=1}^{M}|b_{m_1}|^2\sum_{\substack{m_2=1\\ m_2\neq m_1}}^{M}\frac{1}{\|\alpha_{m_2}-\alpha_{m_1}\|} \\ &\le N\sum_{m=1}^{M}|b_m|^2 + \sum_{m_1=1}^{M}|b_{m_1}|^2\sum_{\substack{m_2=1\\ m_2\neq m_1}}^{M}\frac{2}{m_2\delta}.\end{aligned}$$

To obtain the last line, we have used the δ-spaced property to deduce that for a given m_1, there are at-most 2 such m_2 with $m_2\delta \le \|\alpha_{m_2}-\alpha_{m_1}\| < (m_2+1)\delta$ for $m_2 = 1, ..., M$. We then have

$$\|\mathbf{A}^*\mathbf{b}\| \le \Big(N + \frac{C\log M}{\delta}\Big)\sum_{m=1}^{M}|b_m|^2,$$

and the required result follows by the equivalence above. □

This result will now allow us to gain a similar result for twisted series. We define

$$S_2(\chi) = \sum_{n=1}^{N} a_n\chi(n).$$

When relating exponentials to characters, it is convenient for χ to be a *primitive* Dirichlet character. These are a special class of characters that form the foundation of all Dirichlet characters, in a similar way to the primes.

For $r \in \mathbb{N}_{\geq 2}$, note that if χ is a Dirichlet Character mod q, we find that

$$\psi(n) = \mathbf{1}_{(n,rq)=1}\chi(n),$$

is a Dirichlet character mod rq. When such a relation holds, we say that χ *induces* ψ.

Definition 3.7. If a Dirichlet character χ mod q is not induced by any character χ' mod d for any $d|q$, we call χ a *primitive* Dirichlet character.

Theorem 3.8. ***The Multiplicative Large Sieve Inequality*** *We have*

$$\sum_{q \leq Q} \sideset{}{^*}\sum_{\chi \pmod q} |S_2(\chi)|^2 \leq \Big(N + 2Q^2 \log Q\Big) \sum_{n=1}^{N} |a_n|^2,$$

where $\sum^*_{\chi \pmod q}$ *is the sum over primitive Dirichlet characters* χ *modulo* q*.*

Proof. To connect these inequalities we need a relation between Dirichlet characters and the exponential function. This comes in the form of the Gauss sum

$$\mathcal{G}(\chi) = \sideset{}{^*}\sum_{b \pmod q} \chi(bn)e(bn/q),$$

where $\sum^*_{b \pmod q}$ is the sum over elements of $(\mathbb{Z}/q\mathbb{Z})^\times$, and n satisfies[3] $(n, q) = 1$. From this we deduce that for $n \in \mathbb{N}$ coprime to q, we have

$$\chi(n) = \frac{1}{\mathcal{G}(\overline{\chi})} \sideset{}{^*}\sum_{b \pmod q} \overline{\chi}(b)e(bn/q).$$

For χ a primitive character, we may analyse the case where $(n, q) > 1$ to deduce that the above actually holds for all $n \in \mathbb{N}$ ([14] Theorem 10.3). Noting that the sum in Theorem 3.8 is over primitive characters, we substitute the above into the definition of $S_2(\chi)$ to obtain

$$\begin{aligned} S_2(\chi) &= \frac{1}{\mathcal{G}(\overline{\chi})} \sideset{}{^*}\sum_{b \pmod q} \overline{\chi}(b) \sum_{n=1}^{N} a_n e(bn/q) \\ &= \frac{1}{\mathcal{G}(\overline{\chi})} \sideset{}{^*}\sum_{b \pmod q} \overline{\chi}(b) S_1(b/q). \end{aligned}$$

It is a standard result that for χ a primitive character we have $|\mathcal{G}(\chi)| = \sqrt{q}$ (for a proof see [14], Theorem 10.4).

[3] The Gauss sum is often defined by taking $n = 1$, but the choice of n has no effect on the value of $\mathcal{G}(\chi)$ for $(n, q) = 1$.

Therefore we have the relation

$$\begin{aligned}
\sideset{}{^*}\sum_{\chi \pmod q} |S_2(\chi)|^2 &= \frac{1}{q} \sideset{}{^*}\sum_{\chi \pmod q} \Big| \sideset{}{^*}\sum_{b \pmod q} \overline{\chi}(b) S_1(b/q) \Big|^2 \\
&\le \frac{1}{q} \sum_{\chi \pmod q} \Big| \sideset{}{^*}\sum_{b \pmod q} \overline{\chi}(b) S_1(b/q) \Big|^2, \\
&\le \frac{1}{q} \sideset{}{^*}\sum_{b \pmod q} S_1(b/q) \sideset{}{^*}\sum_{a \pmod q} \overline{S_1(a/q)} \sum_{\chi \pmod q} \chi(a)\overline{\chi}(b) \\
&\le \frac{\varphi(q)}{q} \sideset{}{^*}\sum_{b \pmod q} \sideset{}{^*}\sum_{a \pmod q} S_1(b/q)\overline{S_1(a/q)}\delta_{a \equiv b \pmod q} \\
&\le \frac{\varphi(q)}{q} \sideset{}{^*}\sum_{b \pmod q} |S_1(b/q)|^2.
\end{aligned}$$

We have extended the sum to all Dirichlet characters modulo q, allowing us to utilise orthogonality of characters. Now summing over $q \le Q$, we obtain

$$\begin{aligned}
\sum_{q \le Q} \sideset{}{^*}\sum_{\chi \pmod q} |S_2(\chi)|^2 &\le \sum_{q \le Q} \frac{\varphi(q)}{q} \sideset{}{^*}\sum_{b \pmod q} |S_1(b/q)|^2 \\
&\le \sum_{q \le Q} \sideset{}{^*}\sum_{b \pmod q} |S_1(b/q)|^2 \\
&\le \Big(N + 2CQ^2 \log Q\Big) \sum_{n=1}^{N} |a_n|^2,
\end{aligned}$$

where we have used the additive large sieve inequality on the last line. We note that our α_m take the form of reduced fractions with denominator q for each $q \le Q$. Such fractions are $1/Q^2$-spaced, and there are at most Q^2 of these fractions (so $M \le Q^2$ in the statement of Lemma 3.6). □

3.2 Proof of the Bombieri-Vinogradov Theorem

The multiplicative large sieve inequality is quintessential in the proof of the Bombieri-Vinogradov theorem. We shall use the following two results without proof, which may be found in Koukoulopoulos [14], Chapters 12 and 10 respectively.

Theorem 3.9. ***Siegel-Walfisz Theorem*** *For any $A > 0$ there exists $C > 0$ such that if $1 \le q \le (\log x)^A$ then for $a \in (\mathbb{Z}/q\mathbb{Z})^\times$ we have*

$$\psi(x; q, a) = \frac{x}{\varphi(q)} + O_A\Big(xe^{-C\sqrt{\log x}}\Big).$$

Theorem 3.10. ***Pólya-Vinogradov inequality*** *For a Dirichlet character χ modulo q, $\chi \ne \chi_0$, and any $N \in \mathbb{N}$, we have*

$$\sum_{n=1}^{N} \chi(n) \ll \sqrt{q} \log q.$$

And we shall prove that for $Q = x^\theta$ with $0 \le \theta < 1/2$, we have

$$\sum_{q\le Q} \max_{(a,q)=1} \left|\psi(x;q,a) - \frac{x}{\varphi(q)}\right| \ll_A \frac{x}{(\log x)^A}. \tag{3.03}$$

Remark. Equation (3.03) can be shown to imply Theorem 3.2 by a very similar argument to the proof that $\psi(x) \sim x \implies \pi(x) \sim x/\log x$ in the prime number theorem, by translating through Chebyshev's θ function.

Proof. Let $Q = x^\theta$ with $0 \le \theta < 1/2$. First note that it's enough to prove that

$$\sum_{(\log x)^E < q \le Q} \max_{(a,q)=1} \left|\psi(x;q,a) - \frac{x}{\varphi(q)}\right| \ll_A \frac{x}{(\log x)^A}, \tag{3.04}$$

for some fixed power $E > 0$. This is because the sum over small modulus $q \le (\log x)^E$ can be estimated using the Siegel-Walfisz theorem. We concentrate on the inner term in (3.04). First of all we define

$$\psi(x;\chi) = \sum_{n\le x} \Lambda(n)\chi(n).$$

Orthogonality of characters gives

$$\psi(x;q,a) = \frac{1}{\varphi(q)} \sum_{\chi \pmod q} \overline{\chi}(a)\psi(x;\chi).$$

By the triangle inequality, we have

$$\begin{aligned}
\max_{(a,q)=1} \left|\psi(x;q,a) - \frac{x}{\varphi(q)}\right| &= \max_{(a,q)=1} \left|\frac{1}{\varphi(q)} \sum_{\chi \pmod q} \overline{\chi}(a)\psi(x;\chi) - \frac{x}{\varphi(q)}\right| \\
&\le \max_{(a,q)=1} \Big(\Big|\frac{1}{\varphi(q)}\overline{\chi_0}(a)\psi(x;\chi_0) - \frac{x}{\varphi(q)}\Big| + \Big|\frac{1}{\varphi(q)} \sum_{\substack{\chi \pmod q \\ \chi\ne\chi_0}} \overline{\chi}(a)\psi(x;\chi)\Big|\Big) \\
&\le \left|\frac{\psi(x)-x}{\varphi(q)}\right| + \Big|\frac{1}{\varphi(q)}\sum_{p\mid q}\log p \sum_{\substack{\alpha\ge 0\\ p^\alpha \le x}} 1\Big| + \max_{(a,q)=1}\Big|\frac{1}{\varphi(q)} \sum_{\substack{\chi \pmod q \\ \chi\ne\chi_0}} \overline{\chi}(a)\psi(x;\chi)\Big|.
\end{aligned}$$

By the prime number theorem (Theorem 3.1) and the fact that $\sum_{q\le Q} 1/\varphi(q) = O(\log x)^2$ (from estimate (6.14)) we have $\sum_{q\le Q} |\psi(x) - x/\varphi(q)| \ll x/(\log x)^P$. Furthermore, using simple estimates analogous to those in Mertens' Theorem for primes in arithmetic progression, followed by the estimate (6.14), we find that the sum over $q \le Q$ of the middle term is $\ll Q(\log x)^3$. Therefore, summing over $(\log x)^E < q \le Q$ on both sides, it remains to prove that

$$\sum_{(\log x)^E \le q \le Q} \max_{(a,q)=1} \Big|\frac{1}{\varphi(q)} \sum_{\substack{\chi \pmod q \\ \chi\ne\chi_0}} \overline{\chi}(a)\psi(x;\chi)\Big| \ll \frac{x}{(\log x)^D}.$$

By the triangle inequality, it is sufficient to show

$$\sum_{(\log x)^E<q\le Q}\frac{1}{\varphi(q)}\sum_{\substack{\chi \pmod q\\ \chi\neq\chi_0}}|\psi(x;\chi)|\ll\frac{x}{(\log x)^D},\tag{3.05}$$

for some $D>0$. With the desire to utilise the multiplicative large sieve inequality, we make two observations about the above. First, we note that our $\psi(x;\chi)=\sum_{n\le x}\Lambda(n)\chi(n)$ is of the same form as $S_2(\chi)$, taking $a_n=\Lambda(n)$. However, we need the sum to be over primitive characters to have the desired form. Our second observation is that we also need to obtain $|\psi(x;\chi)|^2$ to apply the large sieve. A natural way to obtain this would be to apply Cauchy-Schwarz.

First we shall make the sum over only primitive characters. Note that any primitive character χ modulo q will induce Q/q characters (including itself). If a primitive character χ modulo q induces a character χ' modulo qr, then we have the following relation

$$\begin{aligned}|\psi(x;\chi')|&=|\psi(x;\chi)-\sum_{n\le x}\Lambda(n)\chi(n)\mathbf{1}_{(n,r)>1}|\le|\psi(x;\chi)|+O\Big(\sum_{\substack{p|r\\ p\nmid q}}\sum_{\alpha\le\log_p x}\chi(p^\alpha)\log p\Big)\\ &\le|\psi(x;\chi)|+O\Big(\log x\sum_{p|r}\log p\Big)=|\psi(x;\chi)|+O(\log x\log r)\\ &\le|\psi(x;\chi)|+O\Big((\log x)^2\Big),\end{aligned}$$

where we have used the fact that $r\le Q/q\le x$. Any character modulo qr will carry a weight of $\frac{1}{\varphi(qr)}$ in (3.05). Therefore, working with the left hand side of (3.05), we have the relation

$$\sum_{(\log x)^E<q\le Q}\frac{1}{\varphi(q)}\sum_{\substack{\chi \pmod q\\ \chi\neq\chi_0}}|\psi(x;\chi)|\le\sum_{\substack{(\log x)^E\\ <q\le Q}}\ \sideset{}{^*}\sum_{\substack{\chi \pmod q\\ \chi\neq\chi_0}}\Big(|\psi(x;\chi)|+O((\log x)^2)\Big)\sum_{r\le Q/q}\frac{1}{\varphi(rq)}.$$

And by the estimate (6.13) we have

$$\sum_{r\le Q/q}\frac{1}{\varphi(rq)}\ll\sum_{r\le Q/q}\frac{\log rq}{rq}\ll\frac{1}{q}\Big(\log Q\Big)^2\ll\frac{1}{q}(\log x)^2.$$

Inserting this into the above, we have obtained a sum over primitive characters as desired.

$$\sum_{(\log x)^E<q\le Q}\frac{1}{\varphi(q)}\sum_{\substack{\chi \pmod q\\ \chi\neq\chi_0}}|\psi(x;\chi)|\ll(\log x)^2\sum_{(\log x)^E<q\le Q}\frac{1}{q}\ \sideset{}{^*}\sum_{\substack{\chi \pmod q\\ \chi\neq\chi_0}}|\psi(x;\chi)|+x^{1/2}(\log x)^4.\tag{3.06}$$

We note that the second term on the right hand side is smaller our desired upper bound $x/(\log x)^D$, and so we focus on showing that the first term is no larger than this bound. We split the interval $\sum_{(\log x)^E<q\le Q}$ into dyadic blocks, writing $\sum_{(\log x)^E<q\le Q}=\sum_{(\log x)^E<q\le 2(\log x)^E}$

$+\sum_{2(\log x)^E < q \le 4(\log x)^E} + \ldots$. The number of blocks is $\ll \log_2 Q \ll \log x$. Therefore, we obtain

(3.07)

$$\sum_{(\log x)^E < q \le Q} \frac{1}{\varphi(q)} \sum_{\substack{\chi \pmod q \\ \chi \neq \chi_0}} |\psi(x;\chi)| \ll (\log x)^3 \max_{(\log x)^E \le R \le Q} \Big(\frac{1}{R} \sum_{R < q \le 2R} \sideset{}{^*}\sum_{\substack{\chi \pmod q \\ \chi \neq \chi_0}} |\psi(x;\chi)| \Big).$$

We shall now utilise Vaughan's identity to split the sum $\psi(x;\chi) = |\sum_{n \le x} \Lambda(n)\chi(n)|$ into multiple sums that we can bound: either in a straightforward way, or using the multiplicative large sieve. Similar to Remark 2.2 we can use L-functions to give the decomposition

$$L(s,\Lambda) = -\frac{\zeta'(s)}{\zeta(s)}. \tag{3.08}$$

We define $M(s)$ as the truncated Möbius Dirichlet series $M(s) = \sum_{n \le M} \mu(n)/n^s$, with $\mu_{\le M}$ as the corresponding Möbius function supported on $\mathbb{N}_{\le M}$ so that $M(s) = L(s, \mu_{\le M})$. Therefore $\zeta(s)M(s) = L(s, 1 * \mu_{\le M})$ where $1 * \mu_{\le M}(n)$ agrees with the δ-function for $n \le M$. Explicitly, we have

$$1 * \mu_{\le M}(n) = \sum_{\substack{d|n \\ d \le M}} \mu(n) = \begin{cases} 1 & n = 1 \\ 0 & 1 < n \le M \\ f(n) & n > M \end{cases}.$$

and f is some arithmetic function bounded above by the divisor function τ_2 by using the triangle inequality. We then have $1 - \zeta(s)M(s) = \sum_{n > M} -f(n)/n^s = L(s, -f_{>M})$ where $f_{>M}$ denotes the function f with the explicit reminder that the function is taken as zero on $n \le M$. Therefore

$$\begin{aligned} L(s,\Lambda) &= -\frac{\zeta'(s)}{\zeta(s)} \Big(1 - \zeta(s)M(s) + \zeta(s)M(s)\Big) \\ &= -\frac{\zeta'(s)}{\zeta(s)} \Big(1 - \zeta(s)M(s)\Big) - \zeta'(s)M(s) \\ &= \Big(\sum_{n \le P} \frac{\Lambda(n)}{n^s} + \sum_{n > P} \frac{\Lambda(n)}{n^s} \Big) \Big(1 - \zeta(s)M(s)\Big) - \zeta'(s)M(s), \end{aligned}$$

where we have used the relation (3.08) on the last line and split the series at some number P which will depend on x. This gives the relation

$$L(s,\Lambda) = \sum_{n \le P} \frac{\Lambda(n)}{n^s} - \zeta(s)M(s) \sum_{n \le P} \frac{\Lambda(n)}{n^s} + \Big(\sum_{n > P} \frac{\Lambda(n)}{n^s} \Big) \Big(\sum_{n > M} \frac{f(n)}{n^s} \Big) - \zeta'(s)M(s).$$

We define the functions $\Lambda_{\le P}$ and $\Lambda_{>P}$ in an analogous way to $\mu_{\le M}$ and $f_{>M}$. We then have the decomposition

$$\Lambda = \Lambda_{\le P} - 1 * \mu_{\le M} * \Lambda_{\le P} + \Lambda_{>P} * f_{>M} + \log * \mu_{\le M}.$$

Inserting this the right hand side of (3.07) and applying the triangle inequality, we have

$$\sum_{(\log x)^E<q\le Q} \frac{1}{\varphi(q)} \sum_{\substack{\chi \pmod q\\ \chi\neq\chi_0}} |\psi(x;\chi)| \ll (\log x)^3 \max_{\substack{(\log x)^E\\ \le R\le Q}} \Big(\frac{1}{R} \sum_{R<q\le 2R} \sideset{}{^*}\sum_{\substack{\chi \pmod q\\ \chi\neq\chi_0}} \Big| \sum_{n\le x} \Lambda(n)\chi(n) \Big| \Big)$$

$$\ll (\log x)^3 \max_{\substack{(\log x)^E\\ \le R\le Q}} \Bigg(\frac{1}{R} \sum_{R<q\le 2R} \sideset{}{^*}\sum_{\substack{\chi \pmod q\\ \chi\neq\chi_0}} \Bigg(\Big| \sum_{n\le P} \Lambda(n)\chi(n) \Big| + \Big| \sum_{n\le x} 1 * \mu_{\le M} * \Lambda_{\le P}(n)\chi(n) \Big|$$

(3.09)

$$+ \Big| \sum_{n\le x} \log *\mu_{\le M}(n)\chi(n) \Big| + \Big| \sum_{n\le x} \Lambda_{>P} * f_{>M}(n)\chi(n) \Big| \Bigg)\Bigg).$$

We concentrate on the first 3 sums: known as type I sums. Such a name is derived from the fact we can often arrange them as sums over single variables. For the first sum, we note that $\sum_{n\le P}\Lambda(n) \ll P$ by the prime number theorem (Theorem 3.1), so the triangle inequality gives us

$$\frac{1}{R} \sum_{R<q\le 2R} \sideset{}{^*}\sum_{\substack{\chi \pmod q\\ \chi\neq\chi_0}} \Big| \sum_{n\le P} \Lambda(n)\chi(n) \Big| \ll RP.$$

For the second sum, noting that M and P will be small compared to x, we have

$$\begin{aligned}
&\frac{1}{R} \sum_{R<q\le 2R} \sideset{}{^*}\sum_{\substack{\chi \pmod q\\ \chi\neq\chi_0}} \Big| \sum_{n\le x} 1 * \mu_{\le M} * \Lambda_{\le P}(n)\chi(n) \Big| \\
&\qquad = \frac{1}{R} \sum_{R<q\le 2R} \sideset{}{^*}\sum_{\substack{\chi \pmod q\\ \chi\neq\chi_0}} \Big| \sum_{n\le x} \sum_{abc=n} \mu_{\le M}(a)\Lambda_{\le P}(b)\chi(abc) \Big| \\
&\qquad = \frac{1}{R} \sum_{R<q\le 2R} \sideset{}{^*}\sum_{\substack{\chi \pmod q\\ \chi\neq\chi_0}} \Big| \sum_{a\le M} \mu(a)\chi(a) \sum_{b\le P} \Lambda(b)\chi(b) \sum_{c\le x/ab} \chi(c) \Big| \\
&\qquad \ll MPR^{3/2}\log R.
\end{aligned}$$

To obtain the last line, we have used the triangle inequality, followed by the Pólya-Vinogradov inequality (Theorem 3.10), the prime number theorem, and trivial bounds on the Möbius function. Similarly, for the third sum we have

$$\begin{aligned}
\frac{1}{R} \sum_{R<q\le 2R} \sideset{}{^*}\sum_{\substack{\chi \pmod q\\ \chi\neq\chi_0}} \Big| \sum_{n\le x} \log *\mu_{\le M}(n)\chi(n) \Big| &= \frac{1}{R} \sum_{R<q\le 2R} \sideset{}{^*}\sum_{\substack{\chi \pmod q\\ \chi\neq\chi_0}} \Big| \sum_{ab\le x} (\log a)\mu_{\le M}(b)\chi(ab) \Big| \\
&\le \frac{1}{R} \sum_{R<q\le 2R} \sideset{}{^*}\sum_{\substack{\chi \pmod q\\ \chi\neq\chi_0}} M \Big| \sum_{a\le x} (\log a)\chi(a) \Big| \\
&\ll MR^{3/2}(\log x)^2.
\end{aligned}$$

The last line here follows from Abel summation and the Pólya-Vinogradov inequality. Plugging these estimates into (3.09), and taking $M = P = x^{0.1}$ we have

$$\sum_{(\log x)^E < q \le Q} \frac{1}{\varphi(q)} \sum_{\substack{\chi \pmod q \\ \chi \ne \chi_0}} |\psi(x;\chi)| \ll x^{0.99} + (\log x)^3 \max_{(\log x)^E \le R \le Q} \frac{1}{R} \sum_{R < q \le 2R} \sideset{}{^*}\sum_{\substack{\chi \pmod q \\ \chi \ne \chi_0}} \Big| \sum_{n \le x} \Lambda_{>P} * f_{>M}(n) \chi(n) \Big|.$$

Therefore to prove (3.05) and complete the proof, it is sufficient to show

$$\max_{(\log x)^E \le R \le Q} \frac{1}{R} \sum_{R < q \le 2R} \sideset{}{^*}\sum_{\substack{\chi \pmod q \\ \chi \ne \chi_0}} \Big| \sum_{n \le x} \Lambda_{>P} * f_{>M}(n) \chi(n) \Big| \ll_C \frac{x}{(\log x)^C}. \tag{3.10}$$

This is known as a type II sum: we must estimate it as a bilinear sum (using methods such as Cauchy-Schwarz and Theorem 3.8). By dyadic decomposition, we have

$$\Big| \sum_{n \le x} \Lambda_{>P} * f_{>M}(n) \chi(n) \Big| = \Big| \sum_{P \le a \le x/M} \Lambda(a) \chi(a) \sum_{\substack{M < b \le x/P \\ ab \le x}} f(b) \chi(b) \Big|$$
$$\le \log(x)^2 \max_{\substack{x^{0.1} \le A \le x^{0.9} \\ x^{0.1} \le B \le x^{0.9} \\ AB \le x}} \Big| \sum_{A \le a \le 2A} \Lambda(a) \chi(a) \sum_{B < b \le 2B} f(b) \chi(b) \Big|.$$

Now for some $x^{0.1} \le A, B \le x^{0.9}$ with $AB \le x$, we apply Cauchy-Schwarz and the large sieve inequality (Theorem 3.8) to find that the left hand side of (3.10) is bounded above by

$$(\log x)^2 \max_{\substack{(\log x)^E \\ \le R \le Q}} \frac{1}{R} \Big(\sum_{R < q \le 2R} \sideset{}{^*}\sum_{\substack{\chi \pmod q \\ \chi \ne \chi_0}} \Big| \sum_{A \le a \le 2A} \Lambda(a) \chi(a) \Big|^2 \Big)^{1/2} \Big(\sum_{R < q \le 2R} \sideset{}{^*}\sum_{\substack{\chi \pmod q \\ \chi \ne \chi_0}} \Big| \sum_{B \le b \le 2B} f(b) \chi(b) \Big|^2 \Big)^{1/2}$$
$$\ll (\log x)^2 \max_{\substack{(\log x)^E \\ \le R \le Q}} \frac{1}{R} \left(\Big(A + R^2 \log R \Big) \sum_{A \le a \le 2A} \Big| \Lambda(a) \Big|^2 \right)^{1/2} \left(\Big(B + R^2 \log R \Big) \sum_{B \le b \le 2B} \Big| f(b) \Big|^2 \right)^{1/2}$$
$$\ll (\log x)^4 \max_{\substack{(\log x)^E \\ \le R \le Q}} \frac{1}{R} \Big(A^2 + R^2 A \log R \Big)^{1/2} \Big(B^2 + R^2 B \log R \Big)^{1/2},$$

where we have used the trivial bound $\sum |a(n)|^2 \le (\sum |a(n)|)^2$ and the fact that $|f(n)| \le$

$\tau_2(n)$ with (6.15). Now using the fact that $\sqrt{A^2+B^2} \leq A+B$, we obtain the upper bound

$$\begin{aligned}
&\ll (\log x)^4 \max_{(\log x)^E \leq R \leq Q} \frac{1}{R}(A + R\sqrt{A})(B + R\sqrt{B}) \\
&\ll (\log x)^4 \max_{(\log x)^E \leq R \leq Q} \Big(R\sqrt{AB} + \frac{AB}{\sqrt{B}} + \frac{AB}{\sqrt{A}} + \frac{AB}{R} \Big) \\
&\ll (\log x)^4 \max_{(\log x)^E \leq R \leq Q} \Big(R\sqrt{x} + \frac{x}{\sqrt{B}} + \frac{x}{\sqrt{A}} + \frac{x}{R} \Big) \\
&\ll \frac{x}{(\log x)^C},
\end{aligned}$$

and so (3.05) holds when E is chosen to be sufficiently large, completing the proof. In the last step, we have used the fact that $R \geq (\log x)^E$ in order to bound the x/R term. This is why the Siegel-Walfisz theorem was an important ingredient in our proof: without it we would not be able to control the case of small moduli. □

With the Bombieri-Vinogradov theorem proved, we may now utilise the ideas of Section 2.3 to proceed with proving bounded gaps between primes.

4 Small Gaps Between Primes

In this section we shall mainly follow Maynard proof in [6] to prove the following theorems.

Theorem 4.1. *We have the following results unconditionally*

$$\liminf_{n\to\infty}(p_{n+1}-p_n) \le 270,$$
$$\liminf_{n\to\infty}(p_{n+m}-p_n) \le e^{4m}m^5.$$

Theorem 4.2. *Assuming the primes have a level of distribution $\theta = 1$ (see definition 3.3), we have*

$$\liminf_{n\to\infty}(p_{n+1}-p_n) \le 12.$$

The idea in Maynard's paper was also to choose good weights that allow for the largest possible ρ in the inequality

$$\sum_{N\le n<2N}\Big(\sum_{i=1}^{k}\mathbf{1}_p(n+h_i)-\rho\Big)w_n > 0.$$

The key idea that allowed for a better result than GPY was to consider multidimensional Selberg sieve weights

$$w_n = \Big(\sum_{d_i|n+h_i\ \forall i}\lambda_{d_1,\dots,d_k}\Big)^2. \tag{4.01}$$

With this choice of weights, the proof splits into 3 main parts:

1. Estimation of the weighted sum using standard Selberg sieve techniques and the Bombieri-Vinogradov theorem.
2. Transformation from sums to integrals.
3. Optimisation to maximise the value of ρ:
   (a) For small k,
   (b) For large k.

4.1 Estimation of weighted sum

First we consider the sum on the left hand side of (2.05) with $N \in \mathbb{N}$ large. We wish to remove the effect of small prime factors, and so we take w_n to be zero if any of $n + h_1, \dots, n + h_k$ have a small prime factor. To do this, we consider some relatively small number $D_0 = \log\log\log N$, and require that $n + h_i \not\equiv 0 \pmod p$ for any $p \le D_0$. The fact that $\mathcal{H}$ is admissible allows for such a construction, as for any prime $p \le D_0$ we can find $u_p \in \mathbb{N}$ such that $u_p + h_i \not\equiv 0 \pmod p$ for all $i = 1, \dots, k$. With $W = \prod_{p\le D_0} p$, the

Chinese Remainder theorem allows us to find $v_0 \pmod{W}$ satisfying $v_0 \equiv u_p \pmod{p}$ for all $p \leq D_0$. Therefore $n \equiv v_0 \pmod{W}$ ensures all of $n + h_i$ are all coprime to W.

To find the size of W, we take logarithms and apply Abel summation and the prime number theorem to find that for sufficiently large N we have

$$\log W = \sum_{p \leq D_0} \log p \leq 2D_0.$$

Hence $W \ll (\log\log N)^2$.

Subsequently, taking $w_n = 0$ unless $n \equiv v_0 \pmod{W}$, the inequality (2.05) becomes

$$\sum_{\substack{N \leq n \leq 2N \\ n \equiv v_0 \pmod{W}}} \Big(\sum_{i=1}^{k} \mathbf{1}_p(n + h_i) - \rho \Big) \Big(\sum_{d_i | n+h_i \ \forall i} \lambda_{d_1,...,d_k} \Big)^2 > 0. \tag{4.02}$$

We take $R = N^{\theta/2 - \delta}$ for θ the level of distribution of the primes from Definition 3.3. Any occurrence of δ and ϵ will assume that they are sufficiently small fixed constants. Furthermore, we restrict the support of $\lambda_{d_1,...,d_k}$ to tuples $(d_1, ..., d_k)$ with $d = \prod_{i=1}^{k} d_i$ square free, $d < R$ and $(d, W) = 1$.

Remark. Note that d being square free implies $(d_i, d_j) = 1$ for all $i \neq j$. Such a condition allows us to invert the substitutions that we make when we diagonalise the quadratic forms. Furthermore, the condition $d < R$ allows for use of the Bombieri Vinogradov theorem (Theorem 3.2) later in the proof. To implement the Bombieri-Vinogradov theorem we need to fix an arithmetic progression for our prime indicator function. We therefore begin by swapping the order of summation and considering independently each prime candidate $n + h_i$.

Definition 4.3. To simplify (4.02), we define

$$S_1 = \sum_{\substack{N \leq n \leq 2N \\ n \equiv v_0 \pmod{W}}} \Big(\sum_{d_i | n+h_i \ \forall i} \lambda_{d_1,...,d_k} \Big)^2,$$

$$S_2^{(m)} = \sum_{\substack{N \leq n \leq 2N \\ n \equiv v_0 \pmod{W}}} \mathbf{1}_p(n + h_m) \Big(\sum_{d_i | n+h_i \ \forall i} \lambda_{d_1,...,d_k} \Big)^2.$$

With these definitions, (4.02) becomes $\sum_{i=1}^{k} S_2^{(i)} - \rho S_1 > 0$. It is now clear to see that maximising ρ is equivalent to maximising the ratio

$$\frac{\sum_{i=1}^{k} S_2^{(i)}}{S_1}.$$

So we wish to estimate S_1 and $S_2^{(m)}$, proceeding in a similar way to Example 2.1.

Lemma 4.4. *With*

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

and $u_{\max} := \sup_{r_1,\dots,r_k} |u_{r_1,\dots,r_k}|$, *for some constant* $C_1 > 0$ *we have the estimate*

$$S_1 = \frac{N}{W} \sum_{r_1,\dots,r_k} \frac{u_{r_1,\dots,r_k}^2}{\prod_{i=1}^{k} \varphi(r_i)} + O\Big(\frac{u_{\max}^2 \varphi(W)^k N (\log R)^k (\log D_0)^{C_1}}{W^{k+1} D_0}\Big).$$

Lemma 4.5. *For* g *a totally multiplicative function defined on primes by* $g(p) = p-2$, *and with*

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

and $u^{(m)}_{\max} := \sup_{r_1,\dots,r_k} |u^{(m)}_{r_1,\dots,r_k}|$, *then for any fixed* $A > 0$ *and some constant* $C_2 > 0$, *we have the estimate*

$$S_2^{(m)} = \frac{N}{\varphi(W)\log N} \sum_{r_1,\dots,r_k} \frac{(u^{(m)}_{r_1,\dots,r_k})^2}{\prod_{i=1}^{k} g(r_i)} + O\Big(\frac{(u^{(m)}_{\max})^2 \varphi(W)^{k-2} N (\log N)^{k-2} (\log D_0)^{C_2}}{W^{k-1} D_0}\Big) + O_A\Big(\frac{u_{\max}^2 N}{(\log N)^A}\Big).$$

Remark. Such substitutions are very similar to those seen in Example 2.1. The motivation is identical: we want to express S_1 and $S_2^{(m)}$ as quadratic forms.

Here we shall give a full proof for Lemma 4.5. Analogous steps can be followed to prove Lemma 4.4.

Proof. (Lemma 4.5) Expanding out the square and swapping the order of summation in Definition 4.3 gives

(4.03)
$$S_2^{(m)} = \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k}} \lambda_{d_1,\dots,d_k} \lambda_{e_1,\dots,e_k} \sum_{\substack{N \le n < 2N \\ n \equiv v_0 \pmod{W} \\ [d_i,e_i] | n+h_i \forall i}} \mathbf{1}_p(n+h_m).$$

Here we make two observations about the inner sum:

First of all, any cases that contribute to the overall sum must have $d_m = e_m = 1$. To see this note that to simultaneously have $d_m, e_m | n+h_m$ and $n+h_m$ prime we require that d_m, e_m equal either 1 or $n+h_m$. However, if $d_m = n+h_m$ then $\prod_{i=1}^{k} d_i \ge n+h_m \ge N > R$, assuming a level of distribution $\theta \le 1$. Therefore we have $\lambda_{d_1,\dots,d_k} = 0$ as we are outside the

the support, so the only contribution comes from $d_m = 1$. An identical line or reasoning allows us to only consider when $e_m = 1$.

Secondly, we only need to consider the case when $W, [d_1, e_1], ..., [d_k, e_k]$ are all pairwise coprime. We split this into two cases. If $p|W, [d_j, e_j]$ then at-least one of d_j and e_j have a prime factor that divides W. However, the support condition for $\lambda_{d_1,...,d_k}$ that $(\prod_{i=1}^k d_i, W) = 1$ (and similarly for $\lambda_{e_1,...,e_k}$) mean that such a case will not contribute to our sum. Alternatively, if $p|[d_i, e_i], [d_j, e_j]$ for some $1 \leq i, j \leq k$, $i \neq j$, then conditions on the inner sum give $p|n + h_i, n + h_j$. Therefore we must have $p||h_j - h_i|$. However, for N sufficiently large we have $|h_j - h_i| < D_0$, hence $p|W$. Here we have used the fact that $\mathcal{H}$ is a fixed set of finite cardinality. Similarly to the first case, such values lie outside the support of either $\lambda_{d_1,...,d_k}$ or $\lambda_{e_1,...,e_k}$.

Remark. This second observations would not be possible without the introduction of D_0 (and therefore W) to remove the effect of small prime factors.

Noting that $[d_i, e_i]|n + h_i$ is equivalent to writing $n \equiv -h_i \pmod{[d_i, e_i]}$, we can now have

$$S_2^{(m)} = \sideset{}{'}\sum_{\substack{d_1,...,d_k \\ e_1,...,e_k \\ d_m=e_m=1}} \lambda_{d_1,...,d_k} \lambda_{e_1,...,e_k} \sum_{\substack{N \leq n < 2N \\ n \equiv a \pmod{q}}} \mathbf{1}_p(n + h_m), \tag{4.04}$$

where $\sum'$ denotes the fact that we are only summing over cases where $W, [d_1, e_1], ..., [d_k, e_k]$ are coprime. We have denoted $q = W \prod_{i=1}^k [d_i, e_i]$, and a is such that $(a, q) = 1$ by the Chinese Remainder theorem.

The inner sum is now over primes in arithmetic progression, so we hope to use the Bombieri-Vinogradov theorem. With this motivation in mind, we have the following estimates for the inner sum

$$\begin{aligned}
&\Big| \sum_{\substack{N \leq n < 2N \\ n \equiv a \pmod{q}}} \mathbf{1}_p(n + h_m) - \frac{\sum_{N \leq n < 2N} \mathbf{1}_p(n)}{\varphi(q)} \Big| \\
&= \Big| \sum_{\substack{N \leq n < 2N \\ n \equiv h_m + a \pmod{q}}} \mathbf{1}_p(n) - \frac{\sum_{N \leq n < 2N} \mathbf{1}_p(n)}{\varphi(q)} - \sum_{\substack{N \leq n < N + h_m \\ n \equiv h_m + a \pmod{q}}} \mathbf{1}_p(n) + \sum_{\substack{2N \leq n < 2N + h_m \\ n \equiv h_m + a \pmod{q}}} \mathbf{1}_p(n) \Big| \\
&\leq 2h_m + \sup_{(a,q)=1} \Big| \sum_{\substack{N \leq n < 2N \\ n \equiv a \pmod{q}}} \mathbf{1}_p(n) - \frac{\sum_{N \leq n < 2N} \mathbf{1}_p(n)}{\varphi(q)} \Big|.
\end{aligned}$$

Therefore, writing

$$E(N, q) = 1 + \sup_{(a,q)=1} \Big| \sum_{\substack{N \leq n < 2N \\ n \equiv a \pmod{q}}} \mathbf{1}_p(n) - \frac{\sum_{N \leq n < 2N} \mathbf{1}_p(n)}{\varphi(q)} \Big|, \tag{4.05}$$

and

$$X_N = \sum_{N \le n < 2N} \mathbf{1}_p(n),$$

we have

$$\sum_{\substack{N \le n < 2N \\ n \equiv a \pmod q}} \mathbf{1}_p(n + h_m) = \frac{X_N}{\varphi(q)} + O(E(N, q)),$$

where we have used the fact that $\mathcal{H}$ is a fixed set of finite cardinality to write $O(2h_m) = O(1)$. Recalling that $q = W \prod_{i=1}^{k} [d_i, e_i]$, (4.04) gives

$$S_2^{(m)} = \frac{X_N}{\varphi(W)} \sideset{}{'}\sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k \\ d_m = e_m = 1}} \frac{\lambda_{d_1,\dots,d_k} \lambda_{e_1,\dots,e_k}}{\varphi(\prod_{i=1}^{k} [d_i, e_i])} + O\Big(\sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k}} |\lambda_{d_1,\dots,d_k} \lambda_{e_1,\dots,e_k}| E(N, q) \Big). \tag{4.06}$$

As desired for the Bombieri-Vinogradov theorem, the error term contains primes in arithmetic progression in the form of $E(N, q)$. Focusing now on the main term of (4.06), we realise that the restriction $\sum'$ to $W, [d_1, e_1], \dots, [d_k, e_k]$ coprime allows us to use the multiplicative property of φ, writing $\varphi(\prod [d_i, e_i]) = \prod \varphi([d_i, e_i])$. We wish to untangle the variables $\varphi([d_i, e_i])$ to allow for a substitution, so we use the identity

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

This follows fairly easily by performing a decomposition of $\varphi(d_i)\varphi(e_i)$ into prime powers and noting that $\varphi(p^\alpha)\varphi(p^\beta) = \varphi(p^{\min(\alpha,\beta)})\varphi(p^{\max(\alpha,\beta)})$.

We have d_i and e_i square free on the support of $\lambda_{d_1,\dots,d_k}$ and $\lambda_{e_1,\dots,e_k}$, therefore the effect of φ on any term above is simply to reduce all primes in their factorisation by 1. Subsequently,

$$\varphi((d_i, e_i)) = \prod_{p | (d_i, e_i)} (p - 1) = \prod_{p | d_i, e_i} (1 * g)(p) = \prod_{p | d_i, e_i} \sum_{n_i | p} g(n_i) = \sum_{t_i | d_i, e_i} g(t_i),$$

where g is a completely multiplicative function defined on primes by $g(p) = p - 2$. Using the previous two results, we have

$$\frac{1}{\prod_{i=1}^{k} \varphi([d_i, e_i])} = \frac{1}{\prod_{i=1}^{k} \varphi(d_i) \varphi(e_i)} \sum_{t_1 | d_1, e_1} \dots \sum_{t_k | d_k, e_k} g(t_1) \dots g(t_k).$$

Hence our main term from (4.06) becomes

$$\frac{X_N}{\varphi(W)} \sideset{}{'}\sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k \\ d_m = e_m = 1}} \sum_{\substack{t_1,\dots,t_k \\ t_i | d_i, e_i \forall i}} \Big(\prod_{i=1}^{k} g(t_i) \Big) \frac{\lambda_{d_1,\dots,d_k} \lambda_{e_1,\dots,e_k}}{\prod_{i=1}^{k} \varphi(d_i) \varphi(e_i)},$$

which we write as

$$\frac{X_N}{\varphi(W)} \sum_{t_1,\dots,t_k} \Big(\prod_{i=1}^{k} g(t_i) \Big) \sideset{}{'}\sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k \\ d_m = e_m = 1 \\ t_i | d_i, e_i \forall i}} \frac{\lambda_{d_1,\dots,d_k} \lambda_{e_1,\dots,e_k}}{\prod_{i=1}^{k} \varphi(d_i) \varphi(e_i)}. \tag{4.07}$$

We now wish to undo the restriction of the sum $\sum'$ to make a substitution easier. Recall that $\sum'$ restricts the summation to cases where $W, [d_1, e_1], ..., [d_k, e_k]$ are coprime. It is simple to remove W from this condition, as indeed if d_i or e_i share any factors with W then $\lambda_{d_1,...,d_k} = 0$ or $\lambda_{e_1,...,e_k} = 0$ respectively.

Therefore we may extend the sum to $[d_1, e_1], ..., [d_k, e_k]$ coprime with no effect. Using a breakdown into prime factors, it can be shown that $[d_i, e_i]$ is coprime to $[d_j, e_j]$ for $i \neq j$ if and only if $(d_i, d_j) = 1$, $(d_i, e_j) = 1$, $(e_i, d_j) = 1$ and $(e_i, e_j) = 1$ for $i \neq j$. Notice that if $p|d_i, d_j$ then $\prod_{i=1}^k d_i$ is not square-free, and so $\lambda_{d_1,...,d_k} = 0$. Similarly if $p|e_i, e_j$ then $\lambda_{e_1,...,e_k} = 0$. Therefore any nonzero contributions to the sum have $(d_i, d_j) = 1$ and $(e_i, e_j) = 1$ for all $i \neq j$. Hence the only condition that we need to consider from $\sum'$ is $(d_i, e_j) = 1$ for all $i \neq j$. We can enforce such a condition by multiplying the inner sum by

$$\prod_{\substack{1\le i,j\le k \\ i\neq j}} \delta((d_i, e_j)) = \prod_{\substack{1\le i,j\le k \\ i\neq j}} \Big(\sum_{s_{i,j}|d_i,e_j} \mu(s_{i,j}) \Big) = \sum_{s_{1,2}|d_1,e_2} \cdots \sum_{s_{k,k-1}|d_k,e_{k-1}} \mu(s_{1,2}) ... \mu(s_{k,k-1}).$$

This makes the conditions invoked from $\sum'$ explicit and subsequently removes the notation. This transforms the main term of (4.07) into

$$\frac{X_N}{\varphi(W)} \sum_{t_1,...,t_k} \Big(\prod_{i=1}^k g(t_i) \Big) \sum_{\substack{d_1,...,d_k \\ e_1,...,e_k \\ d_m=e_m=1 \\ t_i|d_i,e_i \forall i}} \frac{\lambda_{d_1,...,d_k} \lambda_{e_1,...,e_k}}{\prod_{i=1}^k \varphi(d_i)\varphi(e_i)} \sum_{\substack{s_{1,2},...,s_{k,k-1} \\ s_{i,j}|d_i,e_j \forall i\neq j}} \Big(\prod_{\substack{1\le i,j\le k \\ i\neq j}} \mu(s_{i,j}) \Big).$$

This equation can be written as

$$\text{(4.08)} \quad \frac{X_N}{\varphi(W)} \sum_{t_1,...,t_k} \Big(\prod_{i=1}^k g(t_i) \Big) \sideset{}{^*}\sum_{s_{1,2},...,s_{k,k-1}} \Big(\prod_{\substack{1\le i,j\le k \\ i\neq j}} \mu(s_{i,j}) \Big) \sum_{\substack{d_1,...,d_k \\ e_1,...,e_k \\ d_m=e_m=1 \\ t_i|d_i,e_i \forall i \\ s_{i,j}|d_i,e_j \forall i\neq j}} \frac{\lambda_{d_1,...,d_k} \lambda_{e_1,...,e_k}}{\prod_{i=1}^k \varphi(d_i)\varphi(e_i)},$$

where we have written $\sum^*_{s_{1,2},...,s_{k,k-1}}$ to denote the restriction of the $k^2 - k$ sums $\sum_{s_{1,2},...,s_{k,k-1}}$ to cases where each $s_{i,j}$ is coprime to t_i and t_j, and $s_{i,j}$ is coprime to $s_{i,a}$ and $s_{b,j}$ for $a \neq j$ and $b \neq i$.

We can make such restrictions because the sum is only nonzero when $s_{i,j}$ is coprime to t_i and t_j. Suppose $p|s_{i,j}$ and $p|t_i$, then conditions on the innermost sum of (4.08) imply $p|e_j$ and $p|e_i$. But $\lambda_{e_1,...,e_k}$ is only supported on square-free $\prod_{i=1}^k e_i$, so the contribution to the sum is 0. An identical argument is true if $s_{i,j}$ is not coprime to t_j.

Furthermore, the sum is only nonzero when $s_{i,j}$ is coprime to $s_{i,a}$ and $s_{b,j}$ for $a \neq j$ and $b \neq i$. For example, nonzero terms will have $s_{1,2}$ coprime to $s_{1,3}, s_{1,4}, ...$ and $s_{3,2}, s_{4,2}, ...$. Similarly to above, if we have $p|s_{i,j}$ and $p|s_{i,a}$ for $a \neq j$ then $p|e_j, e_a$, so we are outside the support of $\lambda_{e_1,...,e_k}$ due to the square-free condition.

Remark. We realise that in (4.08) the conditions on the innermost sum $t_i|d_i, e_i\ \forall i$ and $s_{i,j}|d_i, e_j\ \forall i \neq j$ may be written as a single condition using this restriction to coprime integers, by introducing two new variables.

Let $a_j = t_j \prod_{i\neq j} s_{j,i}$ and $b_j = t_j \prod_{i\neq j} s_{i,j}$. Under the condition that $s_{i,j}$ are coprime to t_i and t_j for $i \neq j$, and $s_{i,j}$ is coprime to $s_{i,a}$ and $s_{b,j}$ for $a \neq j$ and $b \neq i$, we have the following:

$$t_j|d_j, e_j\ \forall j \text{ and } s_{i,j}|d_i, e_j\ \forall i \neq j \iff a_j|d_j \text{ and } b_j|e_j\ \forall j.$$

The reverse implication is trivial. For the forward implication, we note that each term on the left hand side contributes a different prime factor to the products on the right hand side. Now $a_j = t_j \prod_{i\neq j} s_{j,i}$ and $b_j = t_j \prod_{i\neq j} s_{i,j}$ transform (4.08) into

$$\text{(4.09)} \qquad \frac{X_N}{\varphi(W)} \sum_{t_1,\dots,t_k} \Big(\prod_{i=1}^k g(t_i)\Big) \sideset{}{^*}\sum_{s_{1,2},\dots,s_{k,k-1}} \Big(\prod_{\substack{1\le i,j\le k\\ i\neq j}} \mu(s_{i,j})\Big) \sum_{\substack{d_1,\dots,d_k\\ e_1,\dots,e_k\\ d_m=e_m=1\\ a_j|d_j\ \forall j\\ b_j|e_j\ \forall j}} \frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k \varphi(d_i)\varphi(e_i)},$$

Motivated by the form of the inner sum, we consider the substitution

$$\text{(4.10)} \qquad u^{(m)}_{r_1,\dots,r_k} = \Big(\prod_{i=1}^k \mu(r_i) g(r_i)\Big) \sum_{\substack{d_1,\dots,d_k\\ r_i|d_i\ \forall i\\ d_m=1}} \frac{\lambda_{d_1,\dots,d_k}}{\prod_{i=1}^k \varphi(d_i)},$$

noting that $u^{(m)}_{r_1,\dots,r_k} = 0$ unless $r_m = 1$. Let us show that this substitution is invertible. For $d_1, \dots, d_k$ with $d_m = 1$ and $\prod_{i=1}^k d_i$ square-free, we have

$$\begin{aligned}
\lambda_{d_1,\dots,d_k} &= \sum_{\substack{e_1,\dots,e_k\\ d_i|e_i\ \forall i\\ e_m=1}} \Big(\prod_{i=1}^k \varphi(d_i)\Big) \frac{\lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k \varphi(e_i)} \Big(\prod_{i=1}^k \delta(e_i/d_i)\Big)\\
&= \prod_{i=1}^k \Big(\varphi(d_i)\mu(d_i)^2\Big) \sum_{\substack{e_1,\dots,e_k\\ d_i|e_i\ \forall i\\ e_m=1}} \frac{\lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k \varphi(e_i)} \sum_{\substack{t_1,\dots,t_k\\ t_i|e_i/d_i\ \forall i}} \Big(\prod_{i=1}^k \mu(t_i)\Big)\\
&= \prod_{i=1}^k \Big(\varphi(d_i)\mu(d_i)\Big) \sum_{\substack{e_1,\dots,e_k\\ d_i|e_i\ \forall i\\ e_m=1}} \frac{\lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k \varphi(e_i)} \sum_{\substack{t_1,\dots,t_k\\ d_i t_i|e_i\ \forall i}} \Big(\prod_{i=1}^k \mu(d_i t_i)\Big)\\
&= \prod_{i=1}^k \Big(\varphi(d_i)\mu(d_i)\Big) \sum_{\substack{e_1,\dots,e_k\\ d_i|e_i\ \forall i\\ e_m=1}} \frac{\lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k \varphi(e_i)} \sum_{\substack{r_1,\dots,r_k\\ d_i|r_i\ \forall i\\ r_i|e_i\ \forall i}} \Big(\prod_{i=1}^k \mu(r_i)\Big),
\end{aligned}$$

where we have used the fact that with d_i square free we have $1 = \prod_{i=1}^k \mu(d_i)^2$ to introduce the latter term on the second line. Furthermore, in the third line we used the support conditions of $\lambda_{e_1,\ldots,e_k}$ to merge an occurrence of $\mu(d_i)$ with a trailing term $\mu(t_i)$. If these arguments are not coprime then e_i is not square free. We also note that when $r_i|e_i$, the condition $d_i|r_i$ is stronger than $d_i|e_i$. Therefore we write

$$\begin{aligned}\lambda_{d_1,\ldots,d_k} &= \prod_{i=1}^k \Big(\varphi(d_i)\mu(d_i)\Big) \sum_{\substack{r_1,\ldots,r_k \\ d_i|r_i\ \forall i}} \Big(\prod_{i=1}^k \mu(r_i)\Big) \sum_{\substack{e_1,\ldots,e_k \\ r_i|e_i \\ e_m=1}} \frac{\lambda_{e_1,\ldots,e_k}}{\prod_{i=1}^k \varphi(e_i)} \\ &= \prod_{i=1}^k \Big(\varphi(d_i)\mu(d_i)\Big) \sum_{\substack{r_1,\ldots,r_k \\ d_i|r_i\ \forall i}} \frac{u^{(m)}_{r_1,\ldots,r_k}}{\prod_{i=1}^k g(r_i)}. \qquad (4.11)\end{aligned}$$

We wish to be able to find optimal $u^{(m)}_{r_1,\ldots,r_k}$ with corresponding $\lambda_{d_1,\ldots,d_k}$ supported on the desired set. Therefore, by the above relations, any choice of $u^{(m)}_{r_1,\ldots,r_k}$ supported on $r_1, \ldots, r_k$ with $r = \prod_{i=1}^k r_i$ square-free, $r_m = 1$, $r < R$ and $(r, W) = 1$ will give a suitable choice of $\lambda_{d_1,\ldots,d_k}$.

Remark. We find a similar result relating $u_{r_1,\ldots,r_k}$ and $\lambda_{d_1,\ldots,d_k}$ when proving Lemma 4.4. This result is given in the appendix, Lemma 6.1. Such a result will be useful later on in the proof.

Now applying the substitution (4.10) to the main term (4.09), we have

$$\frac{X_N}{\varphi(W)} \sum_{t_1,\ldots,t_k} \Big(\prod_{i=1}^k g(t_i)\Big) \sideset{}{^*}\sum_{s_{1,2},\ldots,s_{k,k-1}} \Big(\prod_{\substack{1\le i,j\le k \\ i\neq j}} \mu(s_{i,j})\Big) \prod_{i=1}^k \Big(\frac{\mu(a_i)\mu(b_i)}{g(a_i)g(b_i)}\Big) u^{(m)}_{a_1,\ldots,a_k} u^{(m)}_{b_1,\ldots,b_k}, \qquad (4.12)$$

where we have used the fact that a_i and b_i are square-free on the support of $u^{(m)}_{a_1,\ldots,a_k}$ and $u^{(m)}_{b_1,\ldots,b_k}$ so we may write $\mu(a_i) = 1/\mu(a_i)$. Using the restriction $\sum^*$, we can use multiplicative property of μ to separate out $\mu(a_i)$ and $\mu(b_i)$. We also separate out $g(a_i)$ and $g(b_i)$ to obtain

$$\frac{X_N}{\varphi(W)} \sum_{t_1,\ldots,t_k} \Big(\prod_{i=1}^k \frac{\mu(t_i)^2}{g(t_i)}\Big) \sideset{}{^*}\sum_{s_{1,2},\ldots,s_{k,k-1}} \Big(\prod_{\substack{1\le i,j\le k \\ i\neq j}} \frac{\mu(s_{i,j})}{g(s_{i,j})^2}\Big) u^{(m)}_{a_1,\ldots,a_k} u^{(m)}_{b_1,\ldots,b_k}. \qquad (4.13)$$

On nonzero contributions, we necessarily have $t_m = 1$ to ensure that $u_{a_1,\ldots,a_k}$ and $u_{b_1,\ldots,b_k}$ are nonzero. Note that the condition that $u^{(m)}_{a_1,\ldots,a_k}$ is supported on $a_1, \ldots, a_k$ with $(\prod_{i=1}^k a_i, W) = 1$ implies that $(t_i, W) = 1$ on nonzero contributions. Furthermore, it implies either $s_{i,j} = 1$ for all $i \neq j$ or we have some $s_{i,j} > D_0$ (with $(s_{i,j}, W) = 1$). In the latter case, noting that on the support of $u^{(m)}_{a_1,\ldots,a_k}$ and $u^{(m)}_{b_1,\ldots,b_k}$ we have $t_i \le R$, the contribution to the sum is of size

$$
\begin{aligned}
&\ll \Big|\frac{X_N}{\varphi(W)}\Big| \sum_{\substack{t_1,\dots,t_k\le R\\ (t_i,W)=1\\ t_m=1}} \Big(\prod_{i=1}^k \frac{\mu(t_i)^2}{g(t_i)}\Big)\Big(\sum_{\substack{s_{1,2}\\ (s_{1,2},W)=1}} \frac{\mu(s_{1,2})}{g(s_{1,2})^2}\Big)\dots\Big(\sum_{\substack{s_{k,k-1}\\ (s_{k,k-1},W)=1}} \frac{\mu(s_{k,k-1})}{g(s_{k,k-1})^2}\Big)\\
&\qquad \times \sum_{\substack{s_{i,j}>D_0\\ (s_{i,j},W)=1}} \frac{\mu(s_{i,j})}{g(s_{i,j})^2}(u^{(m)}_{\max})^2\\
&\ll \frac{N(u^{(m)}_{\max})^2}{\varphi(W)\log N}\Big(\sum_{\substack{t\le R\\ (t,W)=1}} \frac{\mu(t)^2}{g(t)}\Big)^{k-1}\Big(\sum_{\substack{s\ge 1\\ (s,W)=1}} \frac{\mu(s)^2}{g(s)^2}\Big)^{k(k-1)-1} \sum_{\substack{s_{i,j}>D_0\\ (s_{i,j},W)=1}} \frac{\mu(s_{i,j})^2}{g(s_{i,j})^2}\\
&\ll \frac{N(u^{(m)}_{\max})^2\varphi(W)^{k-2}(\log R)^{k-1}(\log D_0)^{C_2}}{W^{k-1}D_0\log N},
\end{aligned}
$$

where we have used the notation $u^{(m)}_{\max} = \sup_{r_1,\dots,r_k}|u^{(m)}_{r_1,\dots,r_k}|$, and the estimates (6.03), (6.08) and (6.07) to obtain the last line, where $C_2 > 0$ is a sufficiently large constant.

Remark. Here we require that D_0, and hence W, are large enough give a sufficient error when restricting ourselves to cases where all $s_{i,j} = 1$. This restriction is another motivation for introducing the parameter W in the initial support conditions.

Now with all $s_{i,j} = 1$, we have that $a_i = b_i = t_i$. Furthermore, we may assume $\mu(t_i)^2 = 1$, which is required to be in the support of $u^{(m)}_{t_1,\dots,t_k}$. Putting this all together, we may write (4.06) as

$$
\begin{aligned}
S_2^{(m)} =& \frac{X_N}{\varphi(W)} \sum_{t_1,\dots,t_k} \frac{(u^{(m)}_{t_1,\dots,t_k})^2}{\prod_{i=1}^k g(t_i)} + O\Big(\frac{N(u^{(m)}_{\max})^2\varphi(W)^{k-2}(\log R)^{k-1}(\log D_0)^{C_2}}{W^{k-1}D_0\log N}\Big)\\
&+ O\Big(\sum_{\substack{d_1,\dots,d_k\\ e_1,\dots,e_k}} |\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}|E(N,q)\Big). \qquad (4.14)
\end{aligned}
$$

We now shift focus to the error term $O(\sum|\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}|E(N,q))$ with $E(N,q)$ defined in (4.05), and $q = W\prod_{i=1}^k [d_i,e_i]$. We notice that $q \le W\prod_{i=1}^k d_ie_i < WR^2$ on the support of $\lambda_{d_1,\dots,d_k}$ and $\lambda_{e_1,\dots,e_k}$, and we may similarly restrict q to be square free. Note that for $r = W\prod_{i=1}^k [d_i,e_i]$ square-free, there are at-most $\tau_k(r) > \tau_k(r/W)$ choices for $[d_1,e_1],\dots,[d_k,e_k]$ giving this product. Furthermore, given some n, there are at-most $\tau_3(n)$ choices of d_i, e_i such that $[d_i,e_i] = n$. These 3 divisors of n comprise of a choice for (d_i,e_i), and choices for $d_i/(d_i,e_i)$ and $e_i/(d_i,e_i)$, all of which must multiply to give n. Any two d_i, e_i with $[d_i,e_i] = n$ must correspond to one of these combinations. This gives at-most

$\tau_{3k}(r)$ choices of $d_1, ..., d_k, e_1, ..., e_k$ in total , where $r = W\prod_{i=1}^k [d_i, e_i]$. Now

$$\sum_{\substack{d_1,...,d_k\\e_1,...,e_k}} |\lambda_{d_1,...,d_k}\lambda_{e_1,...,e_k}|E(N,q) \ll \lambda_{\max}^2 \sum_{r<R^2W} \mu(r)^2\tau_{3k}(r)E(N,r)$$

$$\ll u_{\max}^2(\log R)^{2k} \sum_{r<R^2W} \mu(r)^2\tau_{3k}(r)E(N,r), \tag{4.15}$$

where we have used the estimate $\lambda_{\max} \ll u_{\max}(\log R)^k$ from Lemma 6.1, where $u_{\max}$ is as defined in Lemma 4.4. Using the triangle inequality in (4.05), we get the bound $E(N,r) \ll N/\varphi(r)$. By Cauchy-Schwarz, we then have

$$\sum_{r<R^2W} \mu(r)^2\tau_{3k}(r)E(N,r) \leq \Big(\sum_{r<R^2W} \mu(r)^2\tau_{3k}^2(r)\frac{N}{\varphi(r)}\Big)^{1/2}\Big(\sum_{r<R^2W} \mu(r)^2E(N,r)\Big)^{1/2}$$
$$\ll_A \frac{N}{(\log N)^A},$$

for some $A > 0$ large. Here we have used the assumption that primes have level of distribution θ and the estimate (6.17) to bound the second term. Inserting this into (4.15) gives a bound on the error term in (4.14), therefore we have

$$S_2^{(m)} = \frac{X_N}{\varphi(W)} \sum_{t_1,...,t_k} \frac{(u_{t_1,...,t_k}^{(m)})^2}{\prod_{i=1}^k g(t_i)} + O\Big(\frac{N(u_{\max}^{(m)})^2\varphi(W)^{k-2}(\log R)^{k-1}(\log D_0)^{C_2}}{W^{k-1}D_0\log N}\Big)$$
$$+ O_A\Big(\frac{u_{\max}^2N}{(\log N)^A}\Big). \tag{4.16}$$

Finally we use the prime number theorem to deduce $X_N = N/\log N + O(N/(\log N)^2)$, where X_N is the number of primes in $[N, 2N)$. The error term here is smaller than the first error term of (4.16), and so we have

$$S_2^{(m)} = \frac{N}{\varphi(W)\log N} \sum_{r_1,...,r_k} \frac{(u_{r_1,...,r_k}^{(m)})^2}{\prod_{i=1}^k g(r_i)} + O\Big(\frac{(u_{\max}^{(m)})^2\varphi(W)^{k-2}N(\log N)^{k-2}(\log D_0)^{C_2}}{W^{k-1}D_0}\Big)$$
$$+ O_A\Big(\frac{u_{\max}^2N}{(\log N)^A}\Big),$$

which completes the proof of Lemma 4.5.

Remark. Application of the Bombieri-Vinogradov theorem requires that W is sufficiently small, so that R^2W is smaller than $N^{\theta/2}$.

□

Remark. The proof of Lemma 4.4 is almost identical, but with easier estimates because no prime indicator function is present in S_1. Therefore a subtle difference is that the estimate for S_1 contains $\varphi(n)$'s in place of $g(n)'s$ and n's in place of $\varphi(n)$'s compared to the proof for $S_2^{(m)}$. The underlying factor here is that the integers may lie in any of q residue class modulo q, whereas prime are restricted to $\varphi(q)$ residue class modulo q. We note that in the proof of Lemma 4.4, the estimates (6.05), (6.09) and (6.10) are used in an identical way to their corresponding results containing $g(n)$.

To allow us to compare our estimates for S_1 and $S_2^{(m)}$ from Lemma 4.4 and Lemma 4.5 we need a relation between $u_{r_1,\ldots,r_k}$ and $u^{(m)}_{r_1,\ldots,r_k}$.

Lemma 4.6. *For some constant $C_3 > 0$, we have the relationship*

$$u^{(m)}_{r_1,\ldots,r_k} = \sum_{e\geq 1} \frac{u_{r_1,\ldots,r_{m-1},e,r_{m+1},\ldots,r_k}}{\varphi(e)} + O\Big(\frac{u_{\max}\varphi(W)\log R(\log D_0)^{C_3}}{WD_0}\Big).$$

Proof. We begin by assuming that $r_m = 1$, as otherwise $u^{(m)}_{r_1,\ldots,r_k} = 0$. In an analogous way to showing the result (4.11), we can show that

$$\lambda_{d_1,\ldots,d_k} = \prod_{i=1}^{k} \mu(d_i)d_i \sum_{\substack{e_1,\ldots,e_k\\ d_i|e_i\ \forall i}} \frac{u_{e_1,\ldots,e_k}}{\prod_{i=1}^k \varphi(e_i)}, \tag{4.17}$$

where $u_{e_1,\ldots,e_k}$ is defined in Lemma 4.4. Similarly to in the case of $u^{(m)}_{r_1,\ldots,r_k}$, we have $u_{e_1,\ldots,e_k}$ supported on tuples with $(\prod_{i=1}^k e_i, W) = 1$, $\prod_{i=1}^k e_i$ square-free and $\prod_{i=1}^k e_i < R$, to ensure that $\lambda_{d_1,\ldots,d_k}$ has the desired support. Inserting (4.17) into the definition of $u^{(m)}_{r_1,\ldots,r_k}$ from Lemma 4.5 or (4.10) gives

$$\begin{aligned} u^{(m)}_{r_1,\ldots,r_k} &= \prod_{i=1}^k \mu(r_i)g(r_i) \sum_{\substack{e_1,\ldots,e_k\\ r_i|e_i\ \forall i}} \frac{u_{e_1,\ldots,e_k}}{\prod_{i=1}^k \varphi(e_i)} \sum_{\substack{d_1,\ldots,d_k\\ r_i|d_i,\ d_i|e_i\ \forall i\\ d_m=1}} \prod_{i=1}^k \frac{\mu(d_i)d_i}{\varphi(d_i)} \\ &= \prod_{i=1}^k \mu(r_i)g(r_i) \sum_{\substack{e_1,\ldots,e_k\\ r_i|e_i\ \forall i}} \frac{u_{e_1,\ldots,e_k}}{\prod_{i=1}^k \varphi(e_i)} \prod_{\substack{1\leq i\leq k\\ i\neq m}} \sum_{\substack{d_i\geq 1\\ r_i|d_i,\ d_i|e_i}} \frac{\mu(d_i)d_i}{\varphi(d_i)}. \end{aligned} \tag{4.18}$$

We focus on the innermost sum. Due to the support of $u_{e_1,\ldots,e_k}$ we may assume that the d_i in the innermost sum are square-free. Therefore, may write this sum as

$$\begin{aligned} \sum_{\substack{d_i\geq 1\\ r_i|d_i,\ d_i|e_i}} \frac{\mu(d_i)d_i}{\varphi(d_i)} = \sum_{\substack{t_i\geq 1\\ t_i|(e_i/r_i)}} \frac{\mu(r_it_i)r_it_i}{\varphi(r_it_i)} &= \frac{r_i\mu(r_i)}{\varphi(r_i)} \sum_{t_i|(e_i/r_i)} \frac{\mu(t_i)t_i}{\varphi(t_i)} \\ &= \frac{r_i\mu(r_i)}{\varphi(r_i)}\Big(1 * \frac{\mu\cdot id}{\varphi}\Big)\Big(\frac{e_i}{r_i}\Big). \end{aligned} \tag{4.19}$$

As e_i/r_i is square-free on the support of $u_{e_1,\ldots,r_k}$ and $1 * \mu\cdot id/\varphi$ is multiplicative, we only need to consider its value on primes. We have

$$\Big(1 * \frac{\mu\cdot id}{\varphi}\Big)(p) = 1 - \frac{p}{p-1} = \frac{\mu(p)}{\varphi(p)},$$

therefore

$$\Big(1 * \frac{\mu \cdot id}{\varphi}\Big)\Big(\frac{e_i}{r_i}\Big) = \frac{\mu(e_i/r_i)}{\varphi(e_i/r_i)}.$$

Substituting into (4.19) and using the fact that e_i are square-free, we have

$$\sum_{\substack{d_i \geq 1 \\ r_i | d_i,\, d_i | e_i}} \frac{\mu(d_i) d_i}{\varphi(d_i)} = \frac{r_i \mu(e_i)}{\varphi(e_i)},$$

and so (4.18) becomes

$$u^{(m)}_{r_1,\dots,r_k} = \prod_{i=1}^{k} \mu(r_i) g(r_i) r_i \sum_{\substack{e_1,\dots,e_k \\ r_i | e_i\ \forall i}} \frac{u_{e_1,\dots,e_k}}{\prod_{i=1}^{k} \varphi(e_i)} \prod_{\substack{1 \leq i \leq k \\ i \neq m}} \frac{\mu(e_i)}{\varphi(e_i)}. \tag{4.20}$$

On the inner sum we consider e_i which are multiples of r_i with no prime factors below D_0 on the support of $u_{e_1,\dots,e_k}$. Therefore we either have all $e_i = r_i$, where r_i could potentially equal 1, or some $e_j > 1$ and subsequently $e_j > D_0$ for $j \neq m$. As $r_m = 1$, the latter cases contribute

$$\ll u_{\max} \Big(\prod_{i=1}^{k} g(r_i) r_i\Big) \prod_{\substack{1 \leq i \leq k \\ i \neq m}} \Big(\sum_{r_i | e_i} \frac{\mu(e_i)^2}{\varphi(e_i)^2}\Big)\Big(\sum_{\substack{e_m < R \\ (e_m, W) = 1}} \frac{\mu(e_m)^2}{\varphi(e_m)}\Big)\Big(\sum_{\substack{e_j > D_0 \\ r_j | e_j}} \frac{\mu(e_j)^2}{\varphi(e_j)^2}\Big)$$

$$\ll \Big(\prod_{i=1}^{k} \frac{g(r_i) r_i}{\varphi(r_i)^2}\Big) \frac{u_{\max} \varphi(W) \log R (\log D_0)^{C_3}}{W D_0} \ll \frac{u_{\max} \varphi(W) \log R (\log D_0)^{C_3}}{W D_0}, \tag{4.21}$$

for some $C_3 > 0$, where we have used estimates (6.05), (6.11) and (6.12) to bound these terms. Inserting this into (4.20) with the main term from the case where $e_i = r_i$ for $j \neq m$, we have

$$u^{(m)}_{r_1,\dots,r_k} = \Big(\prod_{i=1}^{k} \frac{r_i g(r_i)}{\varphi(r_i)^2}\Big) \sum_{e \geq 1} \frac{u_{r_1,\dots,r_{m-1},e,r_{m+1},\dots,r_k}}{\varphi(e)} + O\Big(\frac{u_{\max} \varphi(W) \log R (\log D_0)^{C_3}}{W D_0}\Big). \tag{4.22}$$

On the support of $u^{(m)}_{r_1,\dots,r_k}$ we have all of r_i square-free, $\prod_{i=1}^{k} r_i < R$ and $(\prod_{i=1}^{k} r_i, W) = 1$. Therefore, as $(id \cdot g)/\varphi^2$ is multiplicative, we have

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

therefore

$$\prod_{i=1}^{k}\frac{r_i g(r_i)}{\varphi(r_i)^2} = \prod_{\substack{p\mid \prod_{i=1}^k r_i \\ p>D_0}}\Big(1+O(p^{-2})\Big) \leq \exp\Big(\log \prod_{\substack{p\mid \prod_{i=1}^k r_i \\ p>D_0}}\Big(1+\frac{C}{p^2}\Big)\Big)$$

$$\leq \exp\Big(\sum_{\substack{p\mid \prod_{i=1}^k r_i \\ p>D_0}} \log\Big(1+\frac{C}{p^2}\Big)\Big) \leq \exp\Big(\sum_{\substack{p\mid \prod_{i=1}^k r_i \\ p>D_0}} \frac{C}{p^2}\Big)$$

$$\leq \exp\Big(\sum_{n>D_0}\frac{C}{n^2}\Big) \leq \exp(O(D_0^{-1})) = 1+O(D_0^{-1}). \tag{4.23}$$

Inserting this into (4.22) gives the required result. □

4.2 Transformation into integrals

We recall that our goal is to maximise the ratio $\sum_{i=1}^k S_2^{(i)}/S_1$, thus maximising the value of ρ for which (4.02) can hold. So far, we have managed to obtain estimates for S_1 and $S_2^{(m)}$, and now we wish to be able to compare these estimates. We find it convenient to transform our k-dimensional arithmetic function $u_{r_1,...,r_k}$ into some k-dimensional smooth function F which inherits the properties of u. Specifically, we let

$$u_{r_1,...,r_k} = F\Big(\frac{\log r_1}{\log R}, ..., \frac{\log r_k}{\log R}\Big), \tag{4.24}$$

whenever $r = \prod_{i=1}^k r_i$ satisfies $(r, W) = 1$ and $\mu(r)^2 = 1$, and zero otherwise. Here F is a smooth function supported on $\mathcal{R}_k = \{(x_1, ..., x_k) \in [0, 1]^k \ : \ \sum_{i=1}^k x_i \leq 1\}$ (which is equivalent to the condition that $u_{r_1,...,r_k}$ is supported on $r < R$).

This transformation allows us to write S_1 and $S_2^{(m)}$ in terms of sums over products of arithmetic and smooth functions, which we know how to evaluate (Abel summation is an example of such an evaluation). We may then try different choices of F within our restrictions imposed by u that maximise the desired ratio. This approach is far easier than optimising the ratio in its current form.

To proceed, we prove a less general version of Lemma 4 from [12]. We utilise Lemma 6.2 here. A proof of this more general version can be found in Halberstam and Richert, [15], Lemmas 5.3 and 5.4.

Lemma 4.7. *Let f a totally multiplicative function defined by $f(p) = \frac{\gamma(p)}{p-\gamma(p)}$, following the assumptions of Lemma 6.2. If, in addition to these assumptions, we have $G : [0, 1] \to \mathbb{R}$ a smooth function with $G_{max} := \sup_{x\in[0,1]}(|G(x)| + |G'(x)|)$, then*

$$\sum_{n\leq x} f(n)\mu(n)^2 G\left(\frac{\log n}{\log x}\right) = c_\gamma \log(x)\int_0^1 G(t)dt + O_\gamma(G_{max}),$$

where

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

Proof. Abel summation gives

$$\sum_{n\le x} f(n)\mu(n)^2 G\left(\frac{\log n}{\log x}\right) = G(1)M_{\mu^2 f}(x) - \int_1^x \frac{1}{t\log(x)} G'\left(\frac{\log t}{\log x}\right) M_{\mu^2 f}(t)dt.$$

Substituting $u = \frac{\log t}{\log x}$ and using Lemma 6.2 , the right hand side is

$$G(1)M_{\mu^2 f}(x) - \int_0^1 G'(u)M_{\mu^2 f}(x^u)du = G(1)M_{\mu^2 f}(x) - c_\gamma \log(x)\int_0^1 uG'(u)du + O_\gamma(G_{max}),$$

and integration by parts gives

$$G(1)M_{\mu^2 f}(x) - G(1)c_\gamma \log(x) + c_\gamma \log(x)\int_0^1 G(t)dt + O_\gamma(G_{max}).$$

Application of Lemma 6.2 then gives the desired result. □

Corollary 4.8. *We have*

$$\begin{aligned}(4.25)\qquad &\sum_{\substack{n\ge 1\\ (n, W\prod_{i=1}^k r_i)=1}} \frac{\mu(n)^2}{\varphi(n)} F\Big(\frac{\log r_1}{\log R},\dots,\frac{\log r_{m-1}}{\log R},\frac{\log n}{\log R},\dots,\frac{\log r_k}{\log R}\Big)\\ &= \frac{\varphi(W)}{W}\Big(\prod_{i=1}^k \frac{\varphi(r_i)}{r_i}\Big)\log R\int_0^1 F\Big(\frac{\log r_1}{\log R},\dots,\frac{\log r_{m-1}}{\log R},t,\dots,\frac{\log r_k}{\log R}\Big)dt + O(F_{\max}),\end{aligned}$$

by taking $\gamma(p) = \mathbf{1}_{p\nmid W\prod_{i=1}^k r_i}(p)$. Also, on square-free $\prod_{i=1}^k r_i$, we have

$$\begin{aligned}(4.26)\qquad &\sum_{\substack{n\ge 1\\ (n,W)=1}} \frac{\mu(n)^2}{\varphi(n)} F\Big(\frac{\log r_1}{\log R},\dots,\frac{\log r_{m-1}}{\log R},\frac{\log n}{\log R},\dots,\frac{\log r_k}{\log R}\Big)\\ &= \frac{\varphi(W)}{W}\log R\int_0^1 F\Big(\frac{\log r_1}{\log R},\dots,\frac{\log r_{m-1}}{\log R},t,\dots,\frac{\log r_k}{\log R}\Big)dt + O(F_{\max}),\end{aligned}$$

by taking $\gamma(p) = \mathbf{1}_{p\nmid W}(p)$. Finally we have,

$$(4.27)\qquad \sum_{\substack{r_i\ge 1\\ (r_i,W)=1}} \frac{\mu(r_i)^2\varphi(r_i)^2}{r_i^2 g(r_i)} G\Big(\frac{\log r_i}{\log R}\Big) = \frac{\varphi(W)\log R}{W}\int_0^1 G(t)dt + O\Big(\frac{G_{\max}\log R}{D_0}\Big),$$

by taking $\gamma(p) = (p^3-2p^2+p)/(p^3-p^2-2p+1)\mathbf{1}_{p\nmid W}(p)$. This error term comes from the fact that with this γ we have $c_\gamma = \varphi(W)/W + O(D_0^{-1})$.

This corollary allows us to prove the following lemmas.

Lemma 4.9. *Take $u_{r_1,...,r_k}$ as described by (4.24). Let*

$$F_{\max} := \sup_{(t_1,...,t_k)\in[0,1]^k} \Big(|F(t_1,...,t_k)| + \sum_{i=1}^{k} \Big|\frac{\partial F}{\partial t_i}(t_1,...,t_k)\Big|\Big),$$

then we have

$$S_1 = \frac{\varphi(W)^k N(\log R)^{k+1}}{W^{k+1}} I_k(F) + O\Big(\frac{F_{\max}^2 \varphi(W)^k N(\log R)^k (\log D_0)^C}{W^{k+1} D_0}\Big), \tag{4.28}$$

where $C > 0$ is a fixed constant and

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

Lemma 4.10. *Furthermore,*

$$S_2^{(m)} = \frac{\varphi(W)^k N(\log R)^{k+1}}{W^{k+1}\log N} J_k^{(m)}(F) + O\Big(\frac{F_{\max}^2 \varphi(W)^k N(\log R)^k (\log D_0)^{\tilde{C}}}{W^{k+1} D_0}\Big), \tag{4.29}$$

where $\tilde{C} > 0$ is a fixed constant and

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

In this section we shall once again focus only on the estimate for $S_2^{(m)}$, as the corresponding result for S_1 is easier.

Proof. We use the relation from Lemma 4.6 to relate $u^{(m)}$ to F via (4.24). When $r_m = 1$, this yields

$$u^{(m)}_{r_1,...,r_k} = \sum_{\substack{n\geq 1\\ (n,W\prod_{i=1}^k r_i)=1}} \frac{\mu(n)^2}{\varphi(n)} F\Big(\frac{\log r_1}{\log R}, ..., \frac{\log r_{m-1}}{\log R}, \frac{\log n}{\log R}, ..., \frac{\log r_k}{\log R}\Big) + O\Big(\frac{F_{\max}\varphi(W)\log R(\log D_0)^{C_3}}{WD_0}\Big), \tag{4.30}$$

where the conditions of our sum have been adapted to take into account the support conditions for $u_{r_1,...,r_k}$ mentioned following (4.24). Now, using (4.25) from Corollary 4.8, we have

$$u^{(m)}_{r_1,...,r_k} = \frac{\varphi(W)\log R}{W}\Big(\prod_{i=1}^{k}\frac{\varphi(r_i)}{r_i}\Big)\int_0^1 F\Big(\frac{\log r_1}{\log R}, ..., \frac{\log r_{m-1}}{\log R}, t_m, ..., \frac{\log r_k}{\log R}\Big) dt_m + O\Big(\frac{F_{\max}\varphi(W)\log R(\log D_0)^{C_3}}{WD_0}\Big), \tag{4.31}$$

which holds when $r_m = 1$, $(\prod_{i=1}^k r_i, W) = 1$, and $(r_i, r_j) = 1$ for $i \neq j$, otherwise $u^m_{r_1,...,r_k} = 0$. From this we can recover the estimate $u^{(m)}_{\max} \ll F_{\max}\varphi(W)\log R/W$. We wish to substitute this into Lemma 4.5. Noting that $\varphi(n)/n \leq 1$, we calculate

(4.32)

$$(u^{(m)}_{r_1,...,r_k})^2 = \frac{\varphi(W)^2(\log R)^2}{W^2}\Big(\prod_{i=1}^k \frac{\varphi(r_i)^2}{r_i^2}\Big)F^{(m)}_{r_1,...,r_k} + O\Big(\frac{F^2_{\max}\varphi(W)^2(\log R)^2(\log D_0)^{C_3})}{W^2 D_0}\Big),$$

where we have denoted

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

We recall that Lemma 4.5 states

$$S_2^{(m)} = \frac{N}{\varphi(W)\log N} \sum_{\substack{r_1,...,r_k \\ r_i \text{ square free} \\ r_m=1 \\ (r_i,W)=1 \\ \prod_{i=1}^k r_i \leq R}} \frac{(u^{(m)}_{r_1,...,r_k})^2}{\prod_{i=1}^k g(r_i)} + O\Big(\frac{(u^{(m)}_{\max})^2\varphi(W)^{k-2}N(\log N)^{k-2}(\log D_0)^{C_2}}{W^{k-1}D_0}\Big)$$
$$+ O_A\Big(\frac{u^2_{\max}N}{(\log N)^A}\Big),$$

where we have explicitly written the support conditions for $u^{(m)}_{r_1,...,r_k}$. Inserting (4.32) into this equation , we find that the error induced by the O term in (4.32) is

$$\ll \frac{N}{\varphi(W)\log N} \sum_{\substack{r_1,...,r_k \\ r_i \text{ square-free} \\ r_m=1 \\ (r_i,W)=1 \\ \prod_{i=1}^k r_i \leq R}} \frac{F^2_{\max}\varphi(W)^2(\log R)^2(\log D_0)^{C_3}}{W^2 D_0} \prod_{i=1}^k \frac{1}{g(r_i)}$$
$$\ll \frac{F^2_{\max}\varphi(W)N(\log R)^2(\log D_0)^{C_3}}{W^2 D_0 \log N}\Big(\sum_{\substack{r \leq R \\ (r,W)=1}} \frac{\mu(r)^2}{g(r)}\Big)^{k-1}$$
$$\ll \frac{F^2_{\max}\varphi(W)N(\log R)^{k+1}(\log D_0)^{C_3}}{W^2 D_0 \log N} \ll \frac{F^2_{\max}\varphi(W)^k N(\log R)^k(\log D_0)^{C_3}}{W^{k+1}D_0},$$

where we have used the estimate (6.03) to go from the second to the third line. We now take A to be a sufficiently large constant. Now, using the aforementioned fact from (4.31) that $u^{(m)}_{\max} \ll F_{\max}\varphi(W)\log R/W$, when we inset (4.32) into our $S_2^{(m)}$ estimate, we obtain

(4.33)
$$S_2^{(m)} = \frac{\varphi(W)N(\log R)^2}{W^2\log N} \sum_{\substack{r_1,...,r_{m-1},r_{m+1},...,r_k \\ (r_i,W)=1}} \Big(\prod_{i=1}^k \frac{\mu(r_i)^2\varphi(r_i)^2}{r_i^2 g(r_i)}\Big)(F^{(m)}_{r_1,...,r_k})^2$$
$$+ O\Big(\frac{F^2_{\max}\varphi(W)^k N(\log R)^k(\log D_0)^{\tilde{C}}}{W^{k+1}D_0}\Big),$$

where $\tilde{C} = \max\{C_2, C_3\}$. We note that the condition $\prod_{i=1}^k r_i \leq R$ is enforced naturally by the assumed support of F. We focus on the term

$$\sum_{\substack{r_1,\ldots,r_{m-1},r_{m+1},\ldots,r_k \\ (r_i,W)=1}} \Big(\prod_{i=1}^{k} \frac{\mu(r_i)^2\varphi(r_i)^2}{r_i^2 g(r_i)}\Big)(F^{(m)}_{r_1,\ldots,r_k})^2, \tag{4.34}$$

noting that $r_m = 1$ is fixed, we write this as the $k-1$ sums

$$\sum_{\substack{r_1\geq 1 \\ (r_1,W)=1)}} \frac{\mu(r_1)^2\varphi(r_1)^2}{r_1^2 g(r_1)} \ldots \sum_{\substack{r_k\geq 1 \\ (r_k,W)=1)}} \frac{\mu(r_k)^2\varphi(r_k)^2}{r_k^2 g(r_k)} (F^{(m)}_{r_1,\ldots,r_k})^2. \tag{4.35}$$

Now performing $k-1$ applications of (4.27) from Corollary 4.8, we obtain

$$\frac{\varphi(W)^{k-1}(\log R)^{k-1}}{W^{k-1}} J_k^{(m)}(F) + O\Big(\frac{F_{\max}^2 \varphi(W)^{k-2}(\log R)^{k-1}}{W^{k-2}D_0}\Big), \tag{4.36}$$

where $J_k^{(m)}(F)$ is as defined in Lemma 4.10. Inserting this back into (4.33) gives

$$S_2^{(m)} = \frac{\varphi(W)^k N(\log R)^{k+1}}{W^{k+1}\log N} J_k^{(m)}(F) + O\Big(\frac{F_{\max}^2 \varphi(W)^k N(\log R)^k(\log D_0)^{\tilde{C}}}{W^{k+1}D_0}\Big),$$

as the error term from (4.33) dominates. This is the required result. □

Remark. When we perform corresponding estimates for S_1 to prove Lemma 4.9, we can substitute F directly into Lemma 4.4, which introduces an F^2 term in the integrand. This is in contrast to our $S_2^{(m)}$ estimate, where we first had to relate $u^{(m)}$ to F, which introduced $\int F dt_m$ before substitution into Lemma 4.5. This is the reason behind the different integral operators $I_k(F)$ and $J_k^{(m)}(F)$. The only integral estimate needed to prove Lemma 4.9 is (4.26).

4.3 Optimisation

Combining the results from Lemma 4.9 and Lemma 4.10, we obtain that the expected number of primes in $n+h_1, \ldots, n+h_k$ for $N \leq n < 2N$ using the chosen weights is

$$\frac{\sum_{i=1}^k S_2^{(i)}}{S_1} = \Big(\frac{(1+o(1))\log R}{(1+o(1))\log N}\Big)\Big(\frac{\sum_{i=1}^k J_k^{(i)}(F)}{I_k(F)}\Big), \tag{4.37}$$

for some smooth function F supported on $\mathcal{R}_k = \{(x_1, \ldots, x_k) \in [0,1]^k \ : \ \sum_{i=1}^k x_i \leq 1\}$. Recall that $R = N^{\theta/2-\delta}$ where θ is the level of distribution of the primes. Therefore this ratio is equal to

$$(1+(o(1))\Big(\frac{\theta}{2}-\delta\Big)\frac{\sum_{i=1}^k J_k^{(i)}(F)}{I_k(F)}. \tag{4.38}$$

We therefore have positivity in the sum (4.02) for any ρ strictly less than this value. To get the best results regarding bounded gaps between primes, we wish to maximise this value, and so we wish to show that there are some choices of F that make this value large. We define

$$M_k = \sup_{F \in \mathcal{S}_k} \frac{\sum_{i=1}^k J_k^{(i)}(F)}{I_k(F)} \tag{4.39}$$

where $\mathcal{S}_k$ is the set of square-integrable functions supported on $\mathcal{R}_k = \{(x_1, ..., x_k) \in [0,1]^k \ : \ \sum_{i=1}^k x_i \leq 1\}$.

Lemma 4.11. *Let $r_k = \lceil \theta M_k/2 \rceil$. Then there are infinitely many integers n such that at least r_k of our shifts $n+h_1, ..., n+h_k$ are prime. Therefore we have* $\liminf_n (p_{n+1} - p_n) \leq |h_k - h_1|$.

Proof. Let $\rho = \theta M_k/2 - \epsilon$ for any $\epsilon > 0$. Using approximation by smooth functions, we can choose F_1 smooth, supported on $\mathcal{R}_k$, so that $M_k - 2\delta < \sum_{i=1}^k J_k^{(i)}(F_1)/I_k(F_1)$. With this choice of function F_1 in our estimates for S_1 and $S_2^{(m)}$, we find that

$$\begin{aligned}
\frac{\sum_{i=1}^k S_2^{(i)}}{S_1} - \rho &= (1 + (o(1))\Big(\frac{\theta}{2} - \delta\Big)\frac{\sum_{i=1}^k J_k^{(i)}(F_1)}{I_k(F_1)} - \frac{\theta M_k}{2} + \epsilon \\
&> (1 + (o(1))\Big(\frac{\theta}{2} - \delta\Big)\Big(M_k - 2\delta\Big) - \frac{\theta M_k}{2} + \epsilon \\
&> \epsilon + \delta(2\delta - \theta - M_k).
\end{aligned}$$

For any fixed k and any $\epsilon > 0$, we can choose δ sufficiently small so that the right hand side is ≥ 0. Therefore we find that the inequality (4.02) can be achieved for any $\rho < \theta M_k/2$. Therefore we have infinitely many bounded gaps of length $|h_k - h_1|$ containing $\lfloor \rho + 1 \rfloor \geq r_k$ prime numbers, as required. □

All that is left to do is to find good lower bounds for the value M_k for different values of k, allowing us to take r_k large in Lemma 4.11. We will consider 2 cases: the case where k is small (in which we may find explicit lower bounds) and the case where k is large (where we will find asymptotic lower bounds, allowing the likes of the second result in Theorem 4.1).

First we shall focus on the case when k is small. Specifically, using the unconditional result of Bombieri-Vinogradov (Theorem 3.2), we wish to obtain $r_k = 2$ in Lemma 4.11 and prove bounded gaps between primes. Subsequently we want to find the smallest possible k for which $M_k > 4$. This is equivalent to showing that there is some smooth $F \in \mathcal{S}_k$ such that

$$\sum_{i=1}^k J_k^{(i)}(F) - 4I_k(F) > 0. \tag{4.40}$$

We assume F is symmetric, which can be shown to still give optimal results ([1], Lemma 41). This analysis will be guided by the Polymath paper [1]. With F symmetric, we have that $J_k^{(i)}(F) = J_k^{(j)}(F)$ for all $i, \ j$. Therefore we have $\sum_{i=1}^k J_k^{(i)}(F) = kJ_k^{(1)}(F)$.

Lemma 4.12. *With $k = 54$, we can find $F \in \mathcal{S}_k$ such that*

$$kJ_k^{(1)}(F) - 4I_k(F) > 0. \tag{4.41}$$

Proof. Consider general k for the time being. With the assumption that F is symmetric, we take

$$F(t_1, ..., t_k) = \sum_{i=1}^{n} a_i b_i(t_1, ..., t_k), \tag{4.42}$$

where $b_i : [0, \infty)^k \to \mathbb{R}$ are symmetric, smooth, square-integrable functions supported on $\mathcal{R}_k$, and a_i are real constants. Recall that by definition

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

With this form of F, we have

$$F(t_1, ..., t_k)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j b_i(t_1, ..., t_k) b_j(t_1, ..., t_k),$$
$$\Big(\int F(t_1, ...t_k) dt_1\Big)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \int\int b_i(t_1, t_2, ..., t_k) b_j(t_1', t_2, ..., t_k) dt_1 dt_1',$$

therefore we find that

$$I_k(F) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \int_0^1 ... \int_0^1 b_i(t_1, ..., t_k) b_j(t_1, ..., t_k) dt_1...dt_k,$$
$$J_k(F) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \int_0^1 ... \int_0^1 \int_0^1 b_i(t_1, t_2, ..., t_k) b_j(t_1', t_2, ..., t_k) dt_1 dt_1' dt_2...dt_k.$$

We may now write (4.41) as a difference of quadratic forms,

$$\mathbf{a}^T \mathbf{M}_2 \mathbf{a} - 4\mathbf{a}^T \mathbf{M}_1 \mathbf{a} > 0, \tag{4.43}$$

where $\mathbf{a}$ is the column vector $(a_i)_{i=1}^n$, and we have

$$\Big(\mathbf{M}_1\Big)_{i,j} = \int_0^1 ... \int_0^1 b_i(t_1, ..., t_k) b_j(t_1, ..., t_k) dt_1...dt_k, \tag{4.44}$$

$$\Big(\mathbf{M}_2\Big)_{i,j} = k\int_0^1 ... \int_0^1 \int_0^1 b_i(t_1, t_2, ..., t_k) b_j(t_1', t_2, ..., t_k) dt_1 dt_1' dt_2...dt_k. \tag{4.45}$$

It can be shown ([6], Lemma 8.3) using a simple argument of Lagrangian multipliers that the ratio $\frac{\mathbf{a}^T\mathbf{M}_2\mathbf{a}}{\mathbf{a}^T\mathbf{M}_1\mathbf{a}}$ is maximised when $\mathbf{a}$ is an eigenvector of $\mathbf{M}_1^{-1}\mathbf{M}_2$ corresponding to the largest eigenvalue λ. For such $\mathbf{a}$, we have $\mathbf{M}_2\mathbf{a} = \lambda\mathbf{M}_1\mathbf{a}$, therefore $\mathbf{a}^T\mathbf{M}_2\mathbf{a} - \mathbf{a}^T\lambda\mathbf{M}_1\mathbf{a} = 0$. If $\lambda > 4$, this implies the desired relation (4.43). We therefore just need to find $b_1, ..., b_n$ such that $\mathbf{M}_1^{-1}\mathbf{M}_2$ has an eigenvalue larger than 4.

Definition. A symmetric polynomial is a polynomial where any permutation of its arguments yields the same polynomial. For example, $P(t_1, t_2) = t_1^2 + t_2^2$ is a symmetric polynomial.

Definition. We call a sequence of non-increasing non-negative integers $\alpha = (\alpha_1, ..., \alpha_k)$ a signature. For a tuple $a = (a_1, ...a_k) \in \mathbb{Z}_{\geq 0}^k$, we define $s(a)$ to be its non-increasing rearrangement (making $s(a)$ a signature).

We consider the symmetric polynomials P_α, with

$$P_\alpha(t_1, ..., t_k) = \sum_{\substack{b=(b_1,...,b_k)\in\mathbb{Z}_{\geq 0}^k \\ \text{s.t. } s(b)=\alpha}} t_1^{b_1}...t_k^{b_k}. \tag{4.46}$$

For example, we have $P_{(2,1,0,...,0)}(t_1, ..., t_k) = \sum_{1\leq i<j\leq k} t_i^2 t_j + t_i t_j^2$. It can be shown that for $c, c_1, ..., c_k \in \mathbb{N}$, we have

$$\int ... \int_{\mathcal{R}_k} (1 - t_1 - ... - t_k)^c t_1^{c_1}...t_k^{c_k} dt_1...dt_k = \frac{c!c_1!...c_k!}{(c_1 + ... + c_k + k + c)!}. \tag{4.47}$$

Such a result can be found in [1], Lemma 42, and [6], Lemma 8.1. The proof follows from induction, considering the integral over t_1 and then utilising the substitution $v = t_1/(1 - \sum_{i=2}^k t_i)$. This formula allows us to easily evaluate (4.44) and (4.45) when our b_i are chosen to be of the form of these symmetric polynomials.

After experimentation, the Polymath project found that

$$\mathcal{B} = \{(1 - t_1 - t_2 - ... - t_k)^a P_\alpha(t_1, ..., t_k) : \ a \in \mathbb{Z}_{\geq 0}, \ \alpha_i \text{ all even}, a + \alpha_1 + ... + \alpha_k \leq d\}$$

is a good basis of symmetric polynomials of degree $\leq d$ for some fixed degree d. The meaning of "good" here refers to the computational optimisation problem for calculating the largest eigenvalue of $\mathbf{M}_1^{-1}\mathbf{M}_2$: these polynomials form sufficiently a small basis to control the size of the matrices $\mathbf{M}_i$ whilst giving good results. Obviously, taking $d = \infty$ would give the best results, but would also be incomuputable. The degree taken in the Polymath project was in fact $d = 23$. From here, we take

$$b_i = \begin{cases} P(t_1, ..., t_k) \in \mathcal{B} & , (t_1, ..., t_k) \in \mathcal{R}_k \\ 0 & , \text{otherwise} \end{cases}, \tag{4.48}$$

which are symmetric, smooth, square-integrable functions supported on $\mathcal{R}_k$, as desired. These induce the matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, which can be found explicitly by (4.47). Eigenvalue calculations allow us to find that when $k = 54$, the matrix $\mathbf{M}_1^{-1}\mathbf{M}_2$ has an eigenvalue larger than 4. The corresponding eigenvector then allows us to find $F(t_1, ..., t_k)$ that proves Lemma 4.12. □

Remark. The Polymath paper also gives upper bounds on M_k, subsequently defining the limitations of this method.

Corollary 4.13. *As discussed, Lemma 4.12 implies* $M_{54} > 4$. *Therefore, using Lemma 4.11, we know that for any admissible set* $\mathcal{H}$ *containing* 54 *elements there are infinitely many shifts* n *for which at least two of* $n + h_1, ..., n + h_{54}$ *are prime. This gives the first result in Theorem 4.1, by taking the admissible set*

$$\{0, 4, 10, 18, 24, 28, 30, 40, 54, 58, 60, 70, 72, 82, 84, 88, 94, 102, 108, 112, 114, 118, 124, 130, 132, 138, 142, 150, 154, 160, 168, 172, 174, 180, 184, 192, 198, 202, 208, 214, 220, 222, 228, 234, 238, 240, 244, 250, 252, 258, 262, 264, 268, 270\}$$

courtesy of the MIT website "Narrow admissible tuples" [16], a great website that details the smallest diameter admissible sets for many different values of k.

Corollary 4.14. *From this technique it can be found that* $M_5 > 2$. *Subsequently, assuming the Elliot-Halberstam Conjecture that primes have level of distribution* $\theta = 1$, *Lemma 4.11 implies that for any admissible set* $\mathcal{H}$ *of length* 5, *we have infinitely many shifts* n *for which at least two of* $n + h_1, ..., n + h_5$ *are prime. This gives Theorem 4.2, by taking the admissible set* $\{0, 4, 6, 10, 12\}$, *also from [16].*

Remark. The Polymath project managed to obtain bounded gaps of length 246 here instead. They made this extra improvement by shrinking the domain of integration in $J_k(F)$, while allowing for F to have larger support. This extended the class of functions that we consider, and subsequently proved bounded gaps for admissible sets containing 50 elements, corresponding to an optimal result of 246.

Remark. The Polymath project further proved that $\liminf_n p_{n+1} - p_n \leq 6$ under the assumption of the Generalised Elliot-Halberstam Conjecture, which stems from [17], Conjecture 1.

We move to the case now when k is large. This part will largely follow the proof from Koukoulopoulos [14], Proposition 28.8.

Lemma 4.15. *We have the asymptotic lower bound*

$$M_k \geq \log k - 4 \log \log k + O(1).$$

Proof. We again begin by choosing a particular form for F, giving a lower bound for M_k. Specifically we take

$$F(t_1, ..., t_k) = \mathbf{1}_{\mathcal{R}_k}(t_1, ..., t_k) g(kt_1) ... g(kt_k), \tag{4.49}$$

where the function $g : \mathbb{R} \to \mathbb{R}$ is non-negative, square-integrable and supported on $[0, \xi k]$ for some $\xi \in (0, 1)$ which we shall choose later. We assume that g is normalsied so that $\int_0^\infty g(t)^2 dt = 1$. Similarly to the case for small k we have $M_k \geq k J_k^{(1)}(F)/I_k(F)$ by the symmetry of g. As $\|g\|_2 = 1$, we calculate

$$\begin{aligned} I_k(F) &= \int_0^1 ... \int_0^1 \mathbf{1}_{\mathcal{R}_k}(t_1, ..., t_k) g(kt_1)^2 ... g(kt_k)^2 dt_1 ... dt_k \\ &\leq \Big(\int_0^1 g(kt)^2 dt \Big)^k = \frac{1}{k^k} \Big(\int_0^1 g(t)^2 dt \Big)^k = \frac{1}{k^k}. \end{aligned}$$

Therefore $M_k \geq k^{k+1} J_k^{(1)}(F)$, and all we need to do is try and maximise $J_k^{(1)}(F)$ under the above conditions. We have

$$\begin{aligned}
k^{k+1} J_k^{(1)}(F) &= k^{k+1} \int_0^1 ... \int_0^1 \Big(\int_0^1 F(t_1, ..., t_k) dt_1 \Big)^2 dt_2 ... dt_k \\
&= k^{k+1} \int_0^1 ... \int_0^1 \Big(\int_0^1 \mathbf{1}_{\mathcal{R}_k}(t_1, ..., t_k) g(kt_1) ... g(kt_k) dt_1 \Big)^2 dt_2 ... dt_k \\
&= k^{k+1} \int_0^1 ... \int_0^1 g(kt_2)^2 ... g(kt_k)^2 \Big(\int_0^1 \mathbf{1}_{\mathcal{R}_k}(t_1, ..., t_k) g(kt_1) dt_1 \Big)^2 dt_2 ... dt_k \\
&= \int_0^k ... \int_0^k g(t_2)^2 ... g(t_k)^2 \Big(\int_0^k \mathbf{1}_{\mathcal{R}_k}\Big(\frac{t_1}{k}, ..., \frac{t_k}{k}\Big) g(t_1) dt_1 \Big)^2 dt_2 ... dt_k \\
&= \int_{\mathbb{R}} ... \int_{\mathbb{R}} g(t_2)^2 ... g(t_k)^2 \Big(\int_0^{k - t_2 - ... - t_k} g(t_1) dt_1 \Big)^2 dt_2 ... dt_k.
\end{aligned}$$

We note that when $t_2 + ... + t_k \leq (1-\xi)k$, the the upper limit on the inner integral is $\geq \xi k$. As g is supported on $[0, \xi k]$, we find that in this case the inner integral is equal to $\int_0^\infty g(t_1) dt_1$. Due to positivity of g, we can restrict to this case at the cost of a lower bound for $k^{k+1} J_k^{(1)}(F)$. Noting that $M_k \geq k^{k+1} J_k^{(1)}(F)$, we have

$$\begin{aligned}
M_k &\geq \Big(\int_0^\infty g(t_1) dt_1 \Big)^2 dt_2 ... dt_k \int_{t_2 + ... + t_k \leq (1-\xi)k} g(t_2)^2 ... g(t_k)^2 dt_2 ... dt_k \\
&\geq \Big(\int_0^\infty g(t_1) dt_1 \Big)^2 \mathbb{P}(X_2 + ... + X_k \leq (1-\xi)k), \qquad (4.50)
\end{aligned}$$

where $X_2, ..., X_k$ are independent random variables with corresponding density g^2. We have $\mu = \mathbb{E}[X_i] = \int_0^\infty t g(t)^2 dt$. To bound the probability in (4.50) we will use Chebyshev's inequality, which states that for a random variable X we have $\mathbb{P}(|X - \mu| > \epsilon) \leq \mathrm{Var}[X]/\epsilon^2$. Here we assume $\xi < 1 - \mu$, as otherwise the bounds we obtain are trivial. We have

$$\begin{aligned}
\mathbb{P}(X_2 + ... + X_k > (1-\xi)k) &= \mathbb{P}(X_2 + ... + X_k - (k-1)\mu > (1 - \mu - \xi)k + \mu) \\
&\leq \mathbb{P}(|X_2 + ... + X_k - (k-1)\mu| > (1 - \mu - \xi)k) \\
&\leq \frac{\mathrm{Var}[X_2 + ... + X_k]}{(1 - \xi - \mu)^2 k^2} \\
&\leq \frac{(k-1)\mathrm{Var}[X_2]}{(1 - \xi - \mu)^2 k^2} \\
&\leq \frac{\mathrm{Var}[X_2]}{(1 - \xi - \mu)^2 k}, \qquad (4.51)
\end{aligned}$$

where we have used the fact that $\mathrm{Var}[X_2 + ... + X_k] = (k-1)\mathrm{Var}[X_2]$ for independent $X_2, ..., X_k$. By definition, $\mathrm{Var}[X_2] = \int_0^\infty t^2 g(t)^2 dt - \mu^2$, and

$$\begin{aligned}
\int_0^\infty t^2 g(t)^2 dt = \int_0^\infty t^2 g(t)^2 dt &= \int_0^{\xi k} t^2 g(t)^2 dt \\
&\leq \|t \mathbf{1}_{[0,\xi k]}\|_\infty \|t g(t)^2\|_1 = \xi k \mu,
\end{aligned}$$

by Hölder's inequality, giving $\mathrm{Var}[X_2] \le \mu(\xi k - \mu) \le \xi k \mu$. Combining this with (4.51), we have $\mathbb{P}(X_2 + ... + X_k > (1-\xi)k) \le \xi\mu/(1-\xi-\mu)^2$. Now we find that from (4.50) we have

$$M_k \ge \Big(\int_0^\infty g(t)dt\Big)^2\Big(1 - \frac{\xi\mu}{(1-\xi-\mu)^2}\Big), \tag{4.52}$$

for any L^2 function $g \ge 0$ supported on $[0, \xi k]$ with $\int_0^\infty g(t)^2 dt = 1$ and $\int_0^\infty t g(t)^2 dt = \mu < 1 - \xi$. We make the choice

$$g(t) = c\frac{\mathbf{1}_{[0,\xi k]}}{1 + At}, \tag{4.53}$$

for some $A, c > 0$. The condition $\int_0^\infty g(t)^2 dt = 1$ implies that

$$c^{-2} = \frac{\xi k}{1 + A\xi k},$$

Therefore

$$\mu = \frac{1 + A\xi k}{\xi k}\int_0^{\xi k}\frac{t}{(1+At)^2}dt = \frac{1 + A\xi k}{A^2 \xi k}\Big(\log(1 + A\xi k) - 1 + \frac{1}{1 + A\xi k}\Big). \tag{4.54}$$

We can achieve the inequality $\xi < 1 - \mu$ by taking $A = \log k$ and $\xi = 1/(\log k)^3$, as we then have

$$\begin{aligned}\mu &= \frac{1}{\log k}\Big(\log(k/(\log k)^2) + O(1)\Big) = 1 - \frac{2\log\log k}{\log k} + O\Big(\frac{1}{\log k}\Big)\\ &\le 1 - \xi - \frac{1}{\log k}.\end{aligned} \tag{4.55}$$

Combining this with (4.52) and performing some calculations allows us to arrive at

$$M_k \ge \log k - 4\log\log k + O(1).$$

□

So we have an asymptotic lower bound for M_k. By Lemma 4.11, we know that unconditionally we can find infinitely many shifts n such that at least $\lceil M_k/4 \rceil$ of $(n+h_1, ..., n+h_k)$ are prime. For $m \in \mathbb{N}$, take $k = \lceil Cm^4 e^{4m} \rceil$. Then for large m, we have

$$\lceil M_k/4 \rceil \ge \frac{\log(Cm^4 e^{4m}) - 4\log\log(Cm^4 e^{4m}) + O(1)}{4} > m,$$

for a sufficiently large constant C. This implies there are infinitely many shifts n such that at least m of $(n + h_1, ..., n + h_{\lceil Cm^4 e^{4m} \rceil})$.

We now just need to determine an upper bound on the diameter for the smallest admissible set of size $X = \lceil Cm^4 e^{4m} \rceil$. We shall do this by explicitly constructing an admissible tuple with X elements.

Consider the j-th prime above X, $h_j = p_{\pi(X)+j}$. Then $\{h_1, ..., h_X\}$ form an admissible set. Indeed, if they didn't form an admissible set, then they must cover all the residue classes modulo p for some $p < X$ (as otherwise there would be more residue classes than elements to cover them). This would imply that some $h_j \equiv 0 \pmod{p}$ for some $p \leq X$: however h_j is a prime bigger than X and so we have a contradiction. We just need to find the diameter of this set $|h_X - h_1|$. By the prime number theorem, one can show that $p_n \sim n \log n$. Therefore, for large m, we have

$$\begin{aligned} |h_X - h_1| &= |p_{\pi(\lceil Cm^4 e^{4m} \rceil) + \lceil Cm^4 e^{4m} \rceil} - p_{\pi(\lceil Cm^4 e^{4m} \rceil)+1}| \\ &\leq 2Cm^4 e^{4m} (\log(Cm^4 e^{4m})) \ll e^{4m} m^5. \end{aligned}$$

Therefore we conclude that there are infinitely many bounded gaps between m consecutive primes of size $\ll e^{4m} m^5$, which concludes the proof of Theorem 4.1.

In Section 5, we will proceed to look at how this proof can be applied to large gaps between primes, but first, we will look at some potential limitations of these techniques in proving small gaps between primes.

4.4 The Parity Problem

Unfortunately, it is thought that arguments similar to those presented in this section are insufficient to prove that there are infinitely many prime gaps of size ≤ 4, without some new and innovative idea. The difficulty in decreasing this bound stems from what is known as the *parity problem*. We shall proceed to give a heuristic argument that demonstrates this obstacle.

Recall that our proof began by estimating the following quantity

$$S_2^{(m)} = \sum_{\substack{d_1,...,d_k \\ e_1,...,e_k}} \lambda_{d_1,...,d_k} \lambda_{e_1,...,e_k} \sum_{\substack{N \leq n < 2N \\ n \equiv a \pmod{q}}} \mathbf{1}_p(n + h_m).$$

We note that the proof would still go through if we instead considered the Von-Mangoldt function Λ as oppose to the prime indicator function $\mathbf{1}_p$. More generally, arguments stemming from this proof would likely require us to calculate the analogue

$$S_2^m(f) = \sum_{\substack{d_1,...,d_k \\ e_1,...,e_k}} \lambda_{d_1,...,d_k} \lambda_{e_1,...,e_k} \sum_{\substack{N \leq n < 2N \\ n \equiv a \pmod{q}}} f(n + h_m) \tag{4.56}$$

where f is some multiplicative function. We would then proceed by bounding averages of the quantity

$$\Big| \sum_{\substack{N \leq n < 2N \\ n \equiv a \pmod{q}}} f(n + h_m) - \frac{1}{\varphi(q)} \sum_{N \leq n < 2N} f(n + h_m) \Big|, \tag{4.57}$$

similarly to how we obtained (4.05) and proceeded to use the Bombieri-Vinogradov theorem. If this quantity is sufficiently small, we would expect the rest of the proof to go through in a similar way.

Suppose we could pick some incredibly good choice of w_n that were sufficient to prove that there is a twin prime pair in the interval $[N, 2N)$ for all N sufficiently large. Then with

$$\mathcal{A} = \{n : \ n, \ n+2 \text{ are both prime,}\}$$

we would establish that

$$\sum_{N \le n < 2N} \mathbf{1}_{\mathcal{A}}(n) w_n > 0,$$

for all sufficiently large N. Now, if we multiply our weights w_n by the non-negative function $(1 - \lambda(n)\lambda(n+2))$, where $\lambda(n) = (-1)^{\Omega(n)}$ is the Liouville function, equation (4.56) would instead be of the form

$$S_2^m(f) = \sum_{\substack{d_1,\ldots,d_k \\ e_1,\ldots,e_k}} \lambda_{d_1,\ldots,d_k} \lambda_{e_1,\ldots,e_k} \sum_{\substack{N \le n < 2N \\ n \equiv a \pmod{q}}} f(n+h_m)(1 - \lambda(n)\lambda(n+2)).$$

For $q \le x^{1-\epsilon}$, assuming our function f does not correlate with the Liouville function, we would expect sufficient cancellation to deduce

$$\sum_{\substack{N \le n < 2N \\ n \equiv a \pmod{q}}} f(n+h_m)\lambda(n)\lambda(n+2) = o\Big(\sum_{\substack{N \le n < 2N \\ n \equiv a \pmod{q}}} f(n+h_m) \Big),$$

and so our bounds obtained from (4.57) would be altered by a negligible amount. Therefore the proof would still go through, and these non-negative weights $w_n(1 - \lambda(n)\lambda(n+2))$ would therefore satisfy

$$\sum_{N \le n < 2N} \mathbf{1}_{\mathcal{A}}(n) w_n (1 - \lambda(n)\lambda(n+2)) > 0.$$

However, due to the factor of $(1-\lambda(n)\lambda(n+2))$, such weights are actually zero on n where both n and $n+2$ are prime! Therefore we must actually have

$$\sum_{N \le n < 2N} \mathbf{1}_{\mathcal{A}}(n) w_n (1 - \lambda(n)\lambda(n+2)) = 0.$$

This contradiction leads us to the unfortunate conclusion that our twin-prime detecting weights w_n are unlikely to exist, at least in the discussed context. An analogous argument also holds for gaps of size 4.

Despite this argument seeming rather heuristic and case-specific, the parity problem is a general principle in sieve theory. For example, it prevents the aforementioned work of Chen [11] from proving the full Goldbach conjecture, with his result instead allowing for one number to have two prime factors. The underlying factor here is that sieve weights constructed by most standard procedures in sieve theory do not correlate with the Möbius function, meaning they can't differentiate between an integer with an even number of prime factors and an integer with an odd number of prime factors.

We finish this section by remarking that the parity problem has been overcome in certain cases. Most famously, in 1998, Friedlander and Iwaniec gave sufficient conditions to overcome the parity problem in [18], and used their results to show that there are infinitely many primes of the form $a^2 + b^4$ in [19].

5 Large Gaps Between Primes

In this section, we shall give details for the proof of the following result, once again following the proof of Maynard [9].

Theorem 5.1.

$$\limsup_{n\to\infty} \frac{p_{n+1}-p_n}{L(n)} = \infty,\ L(n) = \frac{\log(n)\log_2(n)\log_4(n)}{(\log_3(n))^2}.$$

5.1 Covering Systems

There is one key idea used to prove results concerning large gaps between primes in [8] [9] [10]. The idea is based on covering systems, which were defined by Erdős in the 1950's [20]. A covering system is a finite collection of congruences $\{a_i \pmod{m_i}\}_{i=1}^k$, $m_i \geq 2$, that covers the whole of the integers. That is, any integer $n \in \mathbb{Z}$ satisfies $n \equiv a_j \pmod{m_j}$ for some $j \in \{1, 2, ..., k\}$.

We can view a prime gap of size $U \in \mathbb{N}$ as a string of composite numbers $n+1, ..., , n+U$. Therefore, the problem of finding a large prime gap becomes a problem of shifting the interval $[1, U] \cap \mathbb{Z}$ by n so that all of $n+1, ..., n+U$ are composite. We wish to minimise this shift n to get the best estimate for $\max_{p_j \leq X} p_{j+1} - p_j$. To obtain such shifts, we can use a similar idea to a covering system in tandem with the Chinese Remainder Theorem.

Suppose that we have the collection of congruences $\{a_i \pmod{m_i}\}_{i=1}^k$ that covers $[1, U] \cap \mathbb{Z}$ for $U \in \mathbb{N}$, where all of our m_i are coprime. Assuming m_i are in increasing order, the Chinese Remainder Theorem tells us that there is a unique integer $n \in [m_k, m_k + \prod_{i=1}^k m_i)$ satisfying $n \equiv -a_i \pmod{m_i}$ for all $i \in \{1, ..., k\}$. The cover definition implies that for $h \in [1, U] \cap \mathbb{Z}$ we have $h \equiv a_i \pmod{m_i}$ for some i. Therefore $n + h \equiv 0 \pmod{m_i}$, and so $n + h$ is not prime (here we need the fact that $n + h > m_i$ to ensure m_i is not the only divisor of $n + h$). As this is true for any $h \in [1, U] \cap \mathbb{Z}$, we find that $n+1, ..., n+U$ is a prime gap of size U. Therefore we know that for $X = m_k + \prod_{i=1}^k m_i$, there is a prime $p_j < X$ such that $p_{j+1} - p_j \geq U$, where p_j denotes some j-th prime. We therefore have

$$\max_{n: p_n < X} p_{n+1} - p_n \geq U.$$

We have now simplified the problem of finding a prime gap to instead constructing some cover of the integers $[1, U] \cap \mathbb{Z}$ using congruence classes with coprime moduli. To get the best results concerning large gaps, we wish to shift our covered interval $[1, U] \cap \mathbb{Z}$ the shortest possible distance n down the number line so that all of $n+1, ..., n+U$ are composite numbers. Hence, for a given U, we wish to minimise $X = m_k + \prod_{i=1}^k m_i$ in the above result. This X is largely determined by the value of $\prod_{i=1}^k m_i$, so we want the product of the moduli of our cover to be as small as possible, under the condition that our moduli are all coprime. The obvious solution to this is to utilise the primes! Congruency classes modulo small primes will cover many integers whilst also maintaining our coprime condition and our desire for a small product of moduli. Therefore for $x > 0$, we shall try and cover the largest possible string of integers $[1, U] \cap \mathbb{Z}$ using

$$\{a_p \pmod{p}\}_{p \leq x}.$$

The previous discussion gives the following lemma:

Lemma 5.2. *If the set* $\{a_p \pmod p\}_{p\le x}$ *covers* $[1,U]\cap\mathbb{Z}$*, then*

$$\max_{n:p_n<X} p_{n+1}-p_n \ge U,$$

where $X = p_{\pi(x)} + \prod_{p\le x} p$*.*

Such a construction is often referred to as the Erdős-Rankin method.

Now our challenge is to cover the the largest possible interval $[1,U]\cap\mathbb{Z}$ using $\{a_p \pmod p\}_{p\le x}$ for fixed $x>0$. We will have U a function of x, as we can cover larger strings of integers when we have more residue classes.

Example. $U = x-1$ and $a_p = -1$ will give a valid cover of $[1,U]\cap\mathbb{Z}$, as for any $1\le n\le x-1$, we have $n+1\equiv 0 \mod p$ for some $p\le x$. The cover will guarantee a prime gap of size $x-1$ below $x+P_x \sim e^x$. Recalling that the average gap below n is $\log n$, this result is tells us no more than the Prime Number Theorem.

Remark. Such a cover corresponds to the primorial construction of large gaps. Similarly to the factorial construction of a gap in section 1.2, this construction utilises the primorial, P_x. Notice that $P_x+2,...,P_x+x$ are all composite, and P_x is much smaller than $x!$. Therefore this construction is more efficient than the factorial construction, but it doesn't tell us anymore than the Prime Number Theorem.

5.2 Reformulation using a cover

Theorem 5.3. *For any fixed* $C>0$*, there are infinitely many* x *such that there exists some cover* $\{a_p \pmod p\}_{p\le x}$ *of* $[1,U]\cap\mathbb{Z}$*, where*

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

Lemma 5.4. *Theorem 5.3 implies Theorem 5.1.*

Proof. Assuming Theorem 5.3, Lemma 5.2 tells us that for infinitely many x, we have

$$\max_{n:p_n<X} p_{n+1}-p_n \ge U,$$

where $X = x+\prod_{p\le x} p$. We have $X = e^{x(1+o(1))}$ by the Prime Number Theorem. Therefore

$$\max_{n:p_n<e^{x(1+o(1))}} p_{n+1}-p_n \ge C\frac{x\log x\log_3 x}{(\log_2 x)^2},$$

and if we take $m = e^{x(1+o(1))}$, we have

$$\max_{n:p_n<m} \frac{p_{n+1}-p_n}{2L(m)(1+o(1))} \ge C/2,$$

with L as defined in Theorem 5.1. Note that for m sufficiently large we have $2L(m)(1 + o(1)) \geq L(m)$. For fixed m, the value that achieves the maximum, say $n = k$, satisfies $p_k < m$ and so $k < m$. Also $k < m$ implies $L(k) < L(m)$ for large m, so for infinitely many k we have

$$\frac{p_{k+1} - p_k}{L(k)} \geq C/2,$$

and so

$$\limsup_{k\to\infty} \frac{p_{k+1} - p_k}{L(k)} \geq C/2.$$

As this statement holds for any $C > 0$, we have Theorem 5.1, as required. □

We shall proceed throughout the remainder of this section to give a proof of Theorem 5.3.

5.3 Choosing an Initial Cover

Take any $C_U > 0$, and fix $\epsilon > 0$. Throughout the proof we shall assume that ϵ is sufficiently small. To form our cover, we shall choose a_p differently depending on the size of p. We begin by defining the quantities

$$(5.01) \qquad y = \exp\left((1-\epsilon)\frac{\log x \log_3 x}{\log_2 x}\right), \qquad z = \frac{x}{\log_2 x}, \qquad U = C_U \frac{x \log y}{\log_2 x},$$

noticing that U is the same as in Theorem 5.3 with $C = (1-\epsilon)C_U$. These choices of y and z date back to 1990 in [21], along with the following choices for a_p:

$$a_p = 1, \text{ for } p \leq y,$$
$$a_p = 0, \text{ for } y < p \leq z.$$

Elements not covered by such choices, say $\mathcal{N} \subseteq [1, U] \cap \mathbb{Z}$, are either y-smooth or have some prime factor $p > z$. For sufficiently large x we have $z^2 > U$, therefore elements of $\mathcal{N}$ have at-most one prime factor larger than z. We may subsequently write $\mathcal{N} = \mathcal{R}' \cup \mathcal{R}$, where

$$(5.02) \qquad \begin{aligned} \mathcal{R}' &= \{m \leq U : \text{m is } y\text{-smooth}, (m-1, P_y) = 1\}, \\ \mathcal{R} &= \{mp \leq U : p > z, \text{ m is } y\text{-smooth}, (mp-1, P_y) = 1\}. \end{aligned}$$

We wish to cover $\mathcal{R}'$ and $\mathcal{R}$ using the remaining residue classes $a_p \pmod p$ for $p > z$. We can calculate the size of $\mathcal{R}'$ using estimates on the number of y-smooth integers $\leq U$ from De Bruijn [22]. These allow us to obtain the estimate

$$(5.03) \qquad |\mathcal{R}'| \ll \frac{x}{(\log x)^{1+\epsilon}},$$

details of which can be found in [21], Theorem 5.2. This is where the appearance of $\epsilon > 0$ in y is essential, as it allows us to get an estimate that is just smaller than $x/\log x$. We now turn our attention to $\mathcal{R}$.

We notice that the set $\mathcal{R}$ is largely determined by prime numbers $p > z$. With this motivation, we wish to formulate $\mathcal{R}$ in such a way that allows us to utilise our understanding of prime numbers. Therefore, for any y-smooth m, we define

$$\mathcal{R}_m = \{z < p \leq U/m : (mp - 1, P_y) = 1\}, \tag{5.04}$$

and we observe (where $m\mathcal{R}$ denotes element-wise multiplication of a set) that

$$\mathcal{R} = \bigcup_{m\ y\text{-smooth}} m\mathcal{R}_m.$$

Furthermore, $\mathcal{R}_m$ is nonempty only when m is even due to the condition $(mp-1, P_y) = 1$. We must also have $m < U/z$ so that $mp \leq U$.

If we can cover $\mathcal{R}_m$ for y-smooth m using some residue classes $\{a_p \pmod p\}$, then we can cover its contribution to $\mathcal{R}$ using the cover $\{ma_p \pmod p\}$. Therefore, instead of trying to cover $\mathcal{R}$, we only need to focus on covering $\mathcal{R}_m$. We hope to do this by exploiting our understanding of the primes.

We shall use the following estimations regarding the size of $|\mathcal{R}_m|$ from [9] (Lemmas 3 and 4) without proof. Due to the form of $\mathcal{R}_m$, the proof relies on the Bombieri-Vonogradov theorem (Theorem 3.2), along with a result from sieve theory known as "The fundamental lemma of sieve theory". More details about this can be found in Chapter 18 of Koukoulopoulos, [14].

Lemma 5.5. *For $0 < m < U(1 - 1/\log x)/z$ even, we have*

$$|\mathcal{R}_m| = \frac{2e^{-\gamma}U(1 + o(1))}{m(\log x)(\log y)} \Big(\prod_{p>2} \frac{p(p-2)}{(p-1)^2}\Big)\Big(\prod_{\substack{p \mid m \\ p>2}} \frac{p-1}{p-2}\Big), \tag{5.05}$$

and for any fixed $n \in \mathbb{N}$ and $m < U/z(\log x)^2$, we have

$$|\mathcal{R}_m \cap (nx, U/m - nx)| \sim |\mathcal{R}_m|. \tag{5.06}$$

We also have the estimates

$$\sum_{1 \leq m < U/z(\log_2 x)^2} |\mathcal{R}_m| \ll \frac{C_U x}{\log x}, \tag{5.07}$$

and

$$\sum_{U/z(\log_2 x)^2 \leq m < U/z} |\mathcal{R}_m| = o\Big(\frac{C_U x}{\log x}\Big). \tag{5.08}$$

These bounds allow us to focus our argument on covering only $\mathcal{R}_m$ for $m < U/z(\log_2 x)^2$. We have the following proposition.

Proposition 5.6. *Suppose that for each $m < U/z(\log_2 x)^2$ we can cover $\mathcal{R}_m$ using residue classes $\{a_q \pmod q\}_{q\in\mathcal{I}_m}$ where $\mathcal{I}_m \subset [x/2, x]$, are disjoint. Then we have Theorem 5.3.*

Proof. The elements in $[1, U] \cap \mathbb{Z}$ left uncovered are

$$\mathcal{R}' \cup \Big(\bigcup_{U/z(\log_2 x)^2 \le m \le U/z} m\mathcal{R}_m \Big). \tag{5.09}$$

By (5.03) and Lemma 5.5, we know that there are $o\Big(\frac{(C_U+1)x}{\log x}\Big)$ elements in these sets, and we need to cover them using residue classes $a_p \pmod p$ with $z < p < x/2$. Application of the Prime Number Theorem with $\pi(x) = x/\log x + O(x/(\log x)^2)$, and the definition of z from (5.01) allows us to find that $\pi(x/2) - \pi(z) > o\Big(\frac{(C_U+1)x}{\log x}\Big)$. Therefore we have more residue classes left to choose than elements to cover! We can simply go through the elements of (5.09) one by one, for each element choosing a distinct prime in $(z, x/2)$ and the corresponding residue class $a_p \pmod p$ that covers the element. □

It remains to prove Proposition 5.6. Phrasing the problem in terms of $\mathcal{I}_m$ for each $m < U/z(\log_2 x)^2$ allows us to focus on covering a general $\mathcal{R}_m$ with residue classes of primes in $\mathcal{I}_m$. We will require that $\mathcal{I}_m$ are sufficiently long intervals to give us enough primes for a cover, but they remain short enough so that they can be disjoint in $[x/2, x]$.

5.4 Random Covers and Connection to Small Gaps

Proposition 5.7. *Take $\epsilon, \delta > 0$ sufficiently small constants. For $m < U/z(\log_2 x)^2$, take $\mathcal{I}'_m \subseteq [x/2, x]$ of length at least $\delta|\mathcal{R}_m| \log x$, then we can choose $\{a_q \pmod q\}_{q\in\mathcal{I}'_m}$ that cover $100(1-\epsilon)\%$ of $\mathcal{R}_m$.*

Lemma 5.8. *Proposition 5.7 implies Proposition 5.6, and subsequently completes the overall proof.*

Proof. If Proposition 5.7 holds, then for any $m < U/z(\log_2 x)^2$, we just need to cover the remaining $\epsilon|\mathcal{R}_m|$ primes. We enlarge $\mathcal{I}'_m$ to include $2\epsilon|\mathcal{R}_m| \log x$ more elements. For sufficiently large x, such an interval will contain $\ge \epsilon|\mathcal{R}_m|$ primes by the prime number theorem. These additional primes give us enough residue classes to trivially cover the remaining elements of $\mathcal{R}_m$.

These new enlarged intervals shall be our $\mathcal{I}_m$, which are any subsets of $[x/2, x]$ of length at least $(\delta + 2\epsilon)|\mathcal{R}_m| \log x$. It remains to show that we can take such $\mathcal{I}_m$ to be disjoint. Using (5.07) in Lemma 5.5 we find that

$$\sum_{m<U/z(\log_2 x)^2} (\delta + 2\epsilon)|\mathcal{R}_m| \log x \ll (\delta + 2\epsilon) C_U x. \tag{5.10}$$

As C_U is fixed, we can find ϵ, δ sufficiently small to allow for $\mathcal{I}_m \subseteq [x/2, x]$ to be disjoint. This proves Proposition 5.7. □

Proposition 5.7 seems to make life difficult, but the motivation behind this reformulation is quite beautiful.

Our sets $\mathcal{R}_m$ (5.04) contain primes in a given interval. To prove Proposition 5.6, our hope is that we can find residue classes $\{a_q \pmod q\}$ that cover many of these primes. This is equivalent to finding a_q where many terms in the arithmetic progression $a_q, a_q + q, a_q + 2q, a_q + 3q, ...$ are prime. In our work on small gaps between primes, we managed to a probability measure on $[N, 2N)$ that is concentrated on $n \in [N, 2N)$ where many of $n+h_1, ..., n+h_k$ were prime. Therefore we can draw a parallel between these two problems, and decide to proceed in the probabilistic way, which motivates the reformulation in Proposition 5.7.

To proceed with proving Proposition 5.7, we begin with $\epsilon, \delta > 0$, $m < U/z(\log_2 x)^2$ and $\mathcal{I'}_m \subseteq [x/2, x]$ of length at least $\delta|\mathcal{R}_m| \log x$. For primes $q \in \mathcal{I'}_m$, we shall choose residue classes according to the probability measures $\nu_{m,q}$ on $\mathbb{Z}/q\mathbb{Z}$ (i.e for $a \in \mathbb{Z}/q\mathbb{Z}$, we choose the residue class $a \pmod q$ with probability $\nu_{m,q}(a)$). Such measures will be induced by very similar weights to those in Section 4.

Lemma 5.9. *Suppose we can find probability measures $\{\nu_{m,q}\}_{q\in\mathcal{I'}_m}$ such that*

$$\sum_{q\in\mathcal{I'}_m} \nu_{m,q}(p) \geq -\log \epsilon/2$$

for every $p \in \mathcal{R}_m$, except for some exceptional primes p that comprise $o(|\mathcal{R}_m|)$ of $\mathcal{R}_m$ as x goes to infinity. We will denote this exceptional set $\mathcal{C}_m$. Then we have Proposition 5.7.

Proof. Our measures $\{\nu_{m,q}\}_{q\in\mathcal{I'}_m}$ induce a product (probability) measure on $\mathbb{Z}/\prod_{q\in\mathcal{I'}_m} q\mathbb{Z}$, and so we find that the probability that some $p \in \mathcal{R}_m$ is not covered by any of our residue classes is $\prod_{q\in\mathcal{I'}_m}(1-\nu_{m,q}(p))$. Taking $G = \mathbb{Z}/\prod_{q\in\mathcal{I'}_m} q\mathbb{Z}$, note that there is a bijection between choices of cover $\{a_q \pmod q\}_{q\in\mathcal{I'}_m}$ and $a \in G$ by the Chinese Remainder Theorem, so a random choice of $a \in G$ corresponds to a random cover. We therefore have

$$\begin{aligned}
\mathbb{E}_G\Big[\#\{p \in \mathcal{R}_m : p \not\equiv a_q \pmod q \ \forall q \in \mathcal{I'}_m\}\Big] &= \sum_{p\in\mathcal{R}_m} \prod_{q\in\mathcal{I'}_m} (1 - \nu_{m,q}(p)) \\
&\leq \sum_{p\in\mathcal{R}_m\setminus\mathcal{C}_m} \prod_{q\in\mathcal{I'}_m} \exp(-\nu_{m,q}(p)) + \sum_{p\in\mathcal{E}_m} 1 \\
&\leq \sum_{p\in\mathcal{R}_m\setminus\mathcal{C}_m} \exp\Big(\sum_{q\in\mathcal{I'}_m} -\nu_{m,q}(p)\Big) + o(|\mathcal{R}_m|) \\
&\leq \sum_{p\in\mathcal{R}_m\setminus\mathcal{C}_m} \epsilon/2 + o(|\mathcal{R}_m|) \leq \epsilon|\mathcal{R}_m|,
\end{aligned}$$

for sufficiently large x, where we have used the fact that $1 - x \leq e^{-x}$ and Lemma 5.9, giving Proposition 5.7, as required. □

All that is needed to complete the proof of Theorem 5.1 is to prove the supposition of Lemma 5.9.

Remark. To formulate this in a familiar way (much like (2.05)), Proposition 5.7 is equivalent to showing that there are some weights η_a for $a \in \mathbb{Z}/\prod_{q\in\mathcal{I}'_m} q\mathbb{Z}$ such that

$$\sum_{a\in\mathbb{Z}/\prod_{q\in\mathcal{I}'_m} q\mathbb{Z}} \Big(\sum_{p\in\mathcal{R}_m\setminus\mathcal{C}_m} (\epsilon/2 - \mathbf{1}_{p\not\equiv a \pmod{q} \forall q\in\mathcal{I}'_m}) \Big)\eta_a > 0.$$

We have $\eta_a = \prod_{q\in\mathcal{I}'_m} \nu_{m,q}(a)$ from our product measure, recalling the bijection between possible covers $\{a_q \pmod{q}\}_{q\in\mathcal{I}'_m}$ and residue classes $a \pmod{\prod_{q\in\mathcal{I}'_m} q}$.

5.5 Calculations

To construct sufficient probability measures $\nu_{m,q}$ that satisfy the supposition of Lemma 5.9, we take inspiration from the shape of our weights from Section 4.

First, however, we need to construct an admissible set $\mathcal{H}$. We use an explicit construction that is technically convenient later in the proof. The set shall be multiplied by q in the proof to mimic an arithmetic progression, as previously motivated. Such multiplication preserves the admissible property.

We introduce the parameter $W = \log_4 x$ once again to remove the effect of small primes. This time, we have $P_W = o(\log_2 x)$ by similar calculations to at the beginning of Section 4.1[4]. The size of our admissible set will be k, which will be a large fixed integer depending on our fixed δ and ϵ. We take $h_i = p_{\pi(k)+i} P_W$ (the i-th prime above k, multiplied by P_W) to be the elements of our admissible set $\mathcal{H} = \{h_1, ..., h_k\}$. We confirm that this set is admissible by noting that, if this set covers $\mathbb{Z}/p\mathbb{Z}$ for some prime p then we must have $p \leq k$. But for x sufficiently large we have $P_W > k$ and so for all i we have $h_i \equiv 0 \pmod{p}$ for any $p \leq k$.

We define

$$\nu_{m,q}(a) = \Big(\sum_{\substack{n\leq U/m \\ n\equiv a \pmod{q} \\ (n(nm-1),P_W)=1}} w_{m,q}(n) \Big) \Big(\sum_{\substack{n\leq U/m \\ (n(nm-1),P_W)=1}} w_{m,q}(n) \Big)^{-1}, \tag{5.11}$$

where

$$w_{m,q}(n) = \Big(\sum_{\substack{d_1,\dots,d_k \\ d_i|n+h_i q}} \sum_{\substack{e_1,\dots,e_k \\ e_i|m(n+h_i q)-1}} \lambda_{d_1,\dots,d_k,e_1,\dots,e_k} \Big)^2, \tag{5.12}$$

and note that in this form it is clear that $\nu_{m,q}$ is a probability measure on $\mathbb{Z}/q\mathbb{Z}$, which is induced by the non-negative weights $w_{m,q}(n)$.

The conditions on these sums allow our weights to be concentrated on residue classes containing many elements of $\mathcal{R}_m$. The restriction to n where $(n(nm-1), P_W)$ in (5.11)

[4]Throughout this large gaps proof, W will act like D_0 from Section 4, and P_W will act like W respectively. This notation is favoured here as it is helpful to be explicitly reminded that the latter is a product of small primes.

removes any contribution from n where n or $nm-1$ have small prime factors. This once again removes difficulty arising from small prime factors (recalling the form of $\mathcal{R}_m$ from (5.04)). The conditions $d_i|n+h_iq$ and $e_i|m(n+h_iq)-1$ in (5.12) provide us with the necessary information to concentrate our weights on residue classes that contain many elements of $\mathcal{R}_m$. Note the multiplication of h_i by q is designed to mimic an arithmetic progression modulo q.

We define our function λ to be

$$\lambda_{d_1,...,d_k,e_1,...,e_k} = \sum_{j=1}^{J}\Big(\prod_{l=1}^{k}\mu(d_l)\mu(e_l)F_{l,j}\Big(\frac{\log d_l}{\log x}\Big)G\Big(\frac{\log e_l}{\log y}\Big)\Big), \tag{5.13}$$

where $F_{i,j}, G$ and J will depend only on k. This means we have the inequality $\lambda_{\mathbf{d},\mathbf{e}} \ll_k 1$. Note that we will use $\mathbf{d}$ and $\mathbf{e}$ to represent the tuples $(d_1,...,d_k)$ and $(e_1,...,e_k)$ respectively throughout the remainder of this section.

We will take $F_{i,j}, G : [0,\infty) \to \mathbb{R}$ to be smooth functions, with

$$\sup\{x : G(x) \neq 0\} \leq 1 \qquad \text{and} \qquad \forall j,\ \sup\{\sum_{i=1}^{k} u_i : F_{i,j}(u_i) \neq 0\} \leq 1/10, \tag{5.14}$$

It should be emphasized that the support of $F_{i,j}$ for fixed j is purposely designed to mimic $\mathcal{R}_k$ from Section 4. We shall choose $F_{i,j}$ so that the following function

$$F(t_1,...,t_k) = \sum_{j=1}^{J}\prod_{l=1}^{k}F'_{l,j}(t_l) \tag{5.15}$$

is symmetric.

Recalling that we wish to show the supposition of Lemma 5.9, we need to obtain an explicit estimate for (5.11). The first step here is to estimate the second factor in this equation, which is the normalising constant from our weights.

We introduce the following quantities which will appear in our estimations, for example they are useful in counting solutions to certain congruence conditions induced by our weights (5.12). We take

$$g_{m,q}(p) = \#\{1 \leq n \leq p : \prod_{i=1}^{k}(n+h_iq)(m(n+h_iq)-1) \equiv 0 \pmod{p}\}, \tag{5.16}$$

which we extend to a totally multiplicative arithmetic function $g_{m,q}$. Similarly, we define

$$\varphi_{m,q}(p) = p - g_{m,q}(p), \tag{5.17}$$

which we extend to be a totally multiplicative arithmetic function. Finally, we define the constant

$$\mathcal{G}_{m,q} = \prod_{p\leq y}\Big(1-\frac{g_{m,q}(p)}{p}\Big)\Big(1-\frac{1}{p}\Big)^{-2k}. \tag{5.18}$$

With these quantities defined, we have the following lemma

Lemma 5.10. *Using the above weights, we have*

$$\sum_{\substack{n\le U/m\\(n(nm-1),P_W)=1}} w_{m,q}(n) = \Big(1+o_k(1)\Big)\frac{U\mathcal{G}_{m,q}}{m(\log x)^k(\log y)^k} I_k^1(F)I_k^2(G), \tag{5.19}$$

where

$$I_k^{(1)}(F) = \int\cdots\int_{t_1,...,t_k\ge 0} F(t_1,...,t_k)^2\,dt_1\dots dt_k \ , \ \textit{and} \quad I_k^{(2)}(G) = \Big(\int_0^\infty G'(t)^2dt\Big)^k.$$

Proof. We expand out the left hand side of (5.19) after inserting our weights from (5.12). Changing the order of summation gives

$$\sum_{\substack{d_1,...,d_k\\d'_1,...,d'_k}}\sum_{\substack{e_1,...,e_k\\e'_1,...,e'_k}} \lambda_{\mathbf{d},\mathbf{e}}\lambda_{\mathbf{d}',\mathbf{e}'} \sum_{\substack{n\le U/m\\(n(nm-1),P_W)=1\\ [d_i,d'_i]|n+h_iq\ \forall i\\ [e_i,e'_i]|m(n+h_iq)-1\forall i}} 1. \tag{5.20}$$

We make three observations about this sum.

(i) First of all suppose that $p|d_id'_i$ and $p\le W$. Then by construction of $\mathcal{H}$ we have $p|h_i$, and also the fact $p|d_id'_i$ implies that $p|n+h_iq$. Putting these together we find that $p|n$. But $(n(nm-1),P_W)=1$ so this cannot be the case. This is similarly true for any pair $e_ie'_i$. Therefore we have that each of $d_1d'_1,...,d_kd'_k,e_1e'_1,...,e_ke'_k$ are pairwise coprime to P_W.

(ii) Secondly, if $p|d_id'_i$ and $p|d_jd'_j$ then $p>W$ (by the above condition) and $p|n+h_iq, n+h_jq$, hence $p|(h_i-h_j)q$. This implies either $p|q$ or $p|(p_{\pi(k)+j}-p_{\pi(k)+i})$. This second term is smaller than W for sufficiently large x, and so we must instead have $p|q$. This implies that $p=q$ as p and q are prime. However, $p|d_id'_i$ implies $p\le d_i$. The support condition from (5.14) implies that our weights (5.13) can only be nonzero when $\log d_i/\log x\le 1/10$. Hence we only see terms when $\log p/\log x\le 1/10$ and subsequently $p\le x^{1/10}$. However, $p=q>x/2$ means this is not possible for large x. Therefore no prime can divide $d_id'_i$ and $d_jd'_j$ for $i\ne j$ when x is large, and so these must be coprime. An analogous argument can be used for e_i's. Therefore we have that $d_1d'_1,...,d_kd'_k$ are pairwise coprime and $e_1e'_1,...,e_ke'_k$ are pairwise coprime.

(iii) Finally, we must have that $(d_id'_i,e_je'_j)|mq(h_j-h_i)-1$ for all $1\le i,j\le k$. This is because $p|d_id'_i$ and $p|e_je'_j$ imply $p|n+h_iq, m(n+h_iq)-1$. Therefore $p|m(n+h_iq), m(n+h_iq)-1$ and subsequently $p|mq(h_j-h_i)-1$. Due to the support of G from (5.14), any nonzero contributions must have $e_j,e'_j\le y$. Therefore we may write this condition as $(d_id'_i,e_je'_j)|(mq(h_j-h_i)-1,P_y)$, which shall be convenient later.

Using the Chinese Remainder Theorem and the fact that $(d_id'_i,d_jd'_j)=1$ implies $([d_i,d'_i]$ $,[d_j,d'_j])=1$, we note that the condition on the inner sum of (5.20) restricts $n\le U/m$ further to $n\equiv c_1 \pmod{\prod_{i=1}^k[d_i,d'_i]}$ and $n\equiv c_2 \pmod{\prod_{i=1}^k[e_i,e'_i]}$ for some c_1 and c_2. We note that (ii) implies $\prod_{i=1}^k [d_i, d_i'] = [d_1, d_1', ..., d_k, d_k']$, and an analogous result holds for e_i's and e_i''s. Furthermore, it is not too hard to check that these congruences are consistent by reducing them modulo $mq(h_j - h_i) - 1$ and recalling condition (iii). These force n into to a single residue class modulo $[d_1, d_1', ..., d_k, d_k', e_1, e_1', ..., e_k, e_k'] = [\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}']$. The inner sum of (5.20) is therefore

$$\sum_{\substack{n \le U/m \\ (n(nm-1), P_W)=1 \\ n \equiv c_3 \pmod{[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}']}}} 1. \tag{5.21}$$

With the knowledge from (i) that P_W is coprime to $[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}']$, we only need to consider how many residue classes modulo P_W satisfy the condition $(n(nm-1), P_W) = 1$. This condition may be written $n, nm - 1 \not\equiv 0 \pmod p$ for any $p \le W$. Note that $p \le W$ implies $p | h_i$, and so for $p \le W$ we have $g_{m,q}(p) = \#\{1 \le n \le p : n \equiv 0 \pmod p \text{ or } nm - 1 \equiv 0 \pmod p\}$. Therefore, using (5.17), we find that for any $p \le W$, there are $\varphi_{m,q}(p)$ residue classes modulo p satisfying $n, nm - 1 \not\equiv 0 \pmod p$. The condition $(n(nm-1), P_W) = 1$ subsequently corresponds to $\varphi_{m,q}(P_W)$ possible residue classes for n modulo P_W.

Therefore with $t = \varphi_{m,q}(P_W)$ and $r_1, ..., r_t$ in distinct residue classes modulo $P_W[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}']$, the inner sum (5.21) can be written as

$$\sum_{i=1}^{t} \sum_{\substack{n \le U/m \\ n \equiv r_i \pmod{P_W[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}']}}} 1 = \frac{U\varphi_{m,q}(P_W)}{m[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}']P_W} + O(\varphi_{m,q}(P_W)). \tag{5.22}$$

Hence applying this to (5.20), we obtain

$$\sideset{}{'}\sum_{\substack{d_1,...,d_k \\ d_1',...,d_k'}} \sideset{}{'}\sum_{\substack{e_1,...,e_k \\ e_1',...,e_k'}} \lambda_{\mathbf{d},\mathbf{e}} \lambda_{\mathbf{d}',\mathbf{e}'} \Big(\frac{U\varphi_{m,q}(P_W)}{m[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}']P_W} + O(\varphi_{m,q}(P_W)) \Big), \tag{5.23}$$

where we have used $\sum'$ to denote the restrictions to tuples $\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}'$ that satisfy the conditions (i), (ii) and (iii). We proceed to estimate the error term.

We denote $d = \prod_{i=1}^k d_i$ and define d', e, e' analogously. Now, the form of λ from (5.13) and the support condition for $F_{l,j}$ in (5.14) imply that we only witness tuples $\mathbf{d}$ where

$$\sum_{l=1}^{k} \frac{\log d_l}{\log x} \le 1/10,$$

which implies that $d \le x^{1/10}$. Similarly we have $d' \le x^{1/10}$, and the support condition from (5.14) on G similarly imply that $e, e' \le y^k \ll_k x^\epsilon$. Also recalling that $\lambda_{\mathbf{d},\mathbf{e}} \ll_k 1$, we find that the error term contributes

$$\ll_k \varphi_{m,q}(P_W) \sideset{}{'}\sum_{\substack{d_1,...,d_k \\ d_1',...,d_k'}} \sideset{}{'}\sum_{\substack{e_1,...,e_k \\ e_1',...,e_k'}} |\lambda_{\mathbf{d},\mathbf{e}} \lambda_{\mathbf{d}',\mathbf{e}'}| \ll_k P_W \sum_{d \le x^{1/10}} \tau_k(d) \sum_{d' \le x^{1/10}} \tau_k(d') \sum_{e \ll_k x^\epsilon} \tau_k(e) \sum_{e' \ll_k x^\epsilon} \tau_k(e')$$

$$\ll_k x^\epsilon x^{2/10+\epsilon} x^{3\epsilon} \ll_k x^{1/2}, \tag{5.24}$$

where we have used the estimate (6.15) to bound these terms.

We now substitute our definition of $\lambda_{\mathbf{d},\mathbf{e}}$ from (5.13) into (5.23) to obtain the following main term for the sum of our weights

$$\frac{U\varphi_{m,q}(P_W)}{mP_W}\sum_{j=1}^{J}\sum_{j'=1}^{J}\sideset{}{'}\sum_{\substack{d_1,\dots,d_k\\ d'_1,\dots,d'_k}}\sideset{}{'}\sum_{\substack{e_1,\dots,e_k\\ e'_1,\dots,e'_k}} \times\frac{\prod_{l=1}^{k}\mu(d_l)\mu(e_l)\mu(d'_l)\mu(e'_l)F_{l,j}\left(\frac{\log d_l}{\log x}\right)F_{l,j'}\left(\frac{\log d'_l}{\log x}\right)G\left(\frac{\log e_l}{\log y}\right)G\left(\frac{\log e'_l}{\log y}\right)}{[\mathbf{d},\mathbf{d}',\mathbf{e},\mathbf{e}']}. \tag{5.25}$$

The support of $F_{l,j}$ and G from (5.14) allow us to extend $F_{l,j}$ and G to smooth functions on $\mathbb{R}$ with compact support. Therefore the functions $e^tF_{l,j}$ and e^tG are L^2, and so may be written as Fourier transforms of functions $f_{l,j}$ and g respectively. [5]

By definition of the Fourier transform, this means $F_{l,j}(x) = e^{-x}\int_{\mathbb{R}} e^{-i\xi x}f_{l,j}(\xi)d\xi$, and $G(x) = e^{-x}\int_{\mathbb{R}} e^{-i\xi x}g(\xi)d\xi$. Therefore

$$F_{l,j}\Big(\frac{\log d_l}{\log x}\Big) = (e^{-\log d_l})^{1/\log x}\int_{\mathbb{R}}(e^{-\log d_l})^{i\xi/\log x}f_{l,j}(\xi)d\xi = \int_{\mathbb{R}}\frac{f_{l,j(\xi)}}{d_l^{(1+i\xi)/\log x}}d\xi, \tag{5.26}$$

$$G\Big(\frac{\log e_l}{\log y}\Big) = \int_{\mathbb{R}}\frac{g(\xi)}{e_l^{(1+i\xi)/\log y}}d\xi$$

Substituting these into (5.25), we obtain a main term of size

$$\frac{U\varphi_{m,q}(P_W)}{mP_W}\sum_{j=1}^{J}\sum_{j'=1}^{J}\int_{\mathbb{R}}\cdots\int_{\mathbb{R}}\Bigg(\sideset{}{'}\sum_{\substack{d_1,\dots,d_k\\ d'_1,\dots,d'_k}}\sideset{}{'}\sum_{\substack{e_1,\dots,e_k\\ e'_1,\dots,e'_k}}\frac{1}{[\mathbf{d},\mathbf{d}',\mathbf{e},\mathbf{e}']}\prod_{l=1}^{k}\frac{\mu(d_l)\mu(e_l)\mu(d'_l)\mu(e'_l)}{d_l^{\frac{1+i\xi_l}{\log x}}\,{d'}_l^{\frac{1+i\xi'_l}{\log x}}\,e_l^{\frac{1+i\tau_l}{\log y}}\,{e'}_l^{\frac{1+i\tau'_l}{\log y}}}\Bigg)$$

$$\times\Big(\prod_{l=1}^{k}f_{l,j}(\xi_l)f_{l,j'}(\xi'_l)g(\tau_l)g(\tau'_l)d\xi_l d\xi'_l d\tau_l d\tau'_l\Big). \tag{5.27}$$

Here we have swapped the order of summation and integration, noting that the expression is absolutely convergent. We now focus our attention on the sums over d_i, d'_i, e_i, e'_i within the parentheses.

We proceed using the generalisation of multiplicative arithmetic functions to multivariable arithmetic functions from [23]. We say that $K : \mathbb{N}^{4k} \to \mathbb{C}$ is multiplicative if $K(a_1b_1, \dots, a_{4k}b_{4k}) = K(a_1, \dots, a_{4k})K(b_1, \dots, b_{4k})$ whenever $(a_1 \dots a_{4k}, b_1 \dots b_{4k}) = 1$. In a similar way to multiplicative arithmetic functions, this means that our function K is defined on tuples of prime powers $(p_1^{\alpha_1}, \dots, p_{4k}^{\alpha_{4k}})$. Therefore we have $K(n_1, \dots, n_{4k}) =$

[5] We define the Fourier transform of f to be $\hat{f}(t) = \int_{\mathbb{R}} e^{-it\xi}f(\xi)d\xi$.

$\prod_p K(p^{v_p(n_1)}, ..., p^{v_p(n_{4k})})$, where $v_p(n)$ denotes the p-adic valuation of n. In our circumstance, we take

(5.28)

$$K(\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}') = \frac{1}{[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}']} \prod_{l=1}^{k} \frac{\mu(d_l)\mu(e_l)\mu(d_l')\mu(e_l')}{d_l^{\frac{1+i\xi_l}{\log x}} {d'}_l^{\frac{1+i\xi_l'}{\log x}} e_l^{\frac{1+i\tau_l}{\log y}} {e'}_l^{\frac{1+i\tau_l'}{\log y}}} \Big(\prod_{i=1}^{k} \delta((d_i d_i' e_i e_i', P_W)) \Big)$$
$$\times \Big(\prod_{1 \le i \ne j \le k} \delta((d_i d_i', d_j d_j'))\delta((e_i e_i', e_j e_j')) \Big) \Big(\prod_{1 \le i, j \le k} \mathbf{1}((d_i d_i', e_j e_j') | (mq(h_j - h_i) - 1, P_y)) \Big),$$

where the trailing indicator functions induce the conditions denoted previously by $\sum'$ (described following (5.20)). It is not too difficult to see that this function is multiplicative in the sense described above.

Using a generalised Euler product, the term from (5.17) in the first parentheses is therefore equal to

$$\begin{aligned} \sum_{\substack{d_1,...,d_k \\ d_1',...,d_k'}} \sum_{\substack{e_1,...,e_k \\ e_1',...,e_k'}} K(\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}') &= \prod_p \sum_{\substack{\alpha_1,...,\alpha_k \ge 0 \\ \alpha_1',...,\alpha_k' \ge 0}} \sum_{\substack{\beta_1,...,\beta_k \ge 0 \\ \beta_1',...,\beta_k' \ge 0}} K(p^{\alpha_1}, ..., p^{\beta_k'}) \\ &= \prod_p \sum_{\substack{\alpha_i, \alpha_i' \in \{0,1\} \\ 1 \le i \le k}} \sum_{\substack{\beta_i, \beta_i' \in \{0,1\} \\ 1 \le i \le k}} K(p^{\alpha_1}, ..., p^{\beta_k'}), \end{aligned} \tag{5.29}$$

where the tuple in the right hand sum is $(p^{\alpha_1}, ..., p^{\alpha_k}, p^{\alpha_1'}, ..., p^{\alpha_k'}, p^{\beta_1}, ..., p^{\beta_k}, p^{\beta_1'}, ..., p^{\beta_k'})$. Here we have used the fact that $\alpha_i \ge 2$ gives $\mu(p^{\alpha_i}) = 0$, and similarly for $\alpha_i', \beta_i, \beta_i'$. We define the local factors

$$K_p = \sum_{\substack{\alpha_i, \alpha_i' \in \{0,1\} \\ 1 \le i \le k}} \sum_{\substack{\beta_i, \beta_i' \in \{0,1\} \\ 1 \le i \le k}} K(p^{\alpha_1}, ..., p^{\beta_k'}). \tag{5.30}$$

Expanding out this definition using (5.28) gives

(5.31)

$$K_p = 1 + \frac{1}{p} \sum_{\substack{\alpha_i, \alpha_i' \in \{0,1\} \\ 1 \le i \le k}} \sum_{\substack{\beta_i, \beta_i' \in \{0,1\} \\ 1 \le i \le k \\ (\alpha_1, ..., \beta_k') \ne (0,...,0)}} \Bigg(\frac{(-1)^{\sum_{i=1}^k (\alpha_i + \alpha_i' + \beta_i + \beta_i')}}{\prod_{l=1}^k p^{\alpha_l s_l + \alpha_l' s_l' + \beta_l r_l + \beta_l' r_l'}} \delta(p, P_W) \times$$
$$\Big(\prod_{1 \le i \ne j \le k} \delta((p^{\alpha_i + \alpha_i'}, p^{\alpha_j + \alpha_j'}))\delta((p^{\beta_i + \beta_i'}, p^{\beta_j + \beta_j'})) \Big) \Big(\prod_{1 \le i, j \le k} \mathbf{1}((p^{\alpha_i + \alpha_i'}, p^{\beta_j + \beta_j'}) | (mq(h_j - h_i) - 1, P_y)) \Big) \Bigg),$$

where we have written $s_l = (1 + i\xi_l)/\log x, r_l = (1 + i\tau_l)/\log y$, and defined s_l', r_l' similarly. We remark that

$$K_P \ll 1 + \frac{1}{p} \sum_{\substack{\alpha_i, \alpha_i' \in \{0,1\} \\ 1 \le i \le k}} \sum_{\substack{\beta_i, \beta_i' \in \{0,1\} \\ 1 \le i \le k \\ (\alpha_1, ..., \beta_k') \ne (0,...,0)}} \frac{1}{p^{1/\log x}} \ll 1 + O_k(p^{-1-\frac{1}{\log x}}),$$

therefore, for some integer C_k depending on k, we have

$$\sum_{\substack{d_1,\dots,d_k\\ d'_1,\dots,d'_k}}\sum_{\substack{e_1,\dots,e_k\\ e'_1,\dots,e'_k}} K(\mathbf{d},\mathbf{d}',\mathbf{e},\mathbf{e}') = \prod_p K_p \le \prod_p \Big(1+\frac{C_k}{p^{1+1/\log x}}\Big) \ll \prod_p \Big(1+\frac{1}{p^{1+1/\log x}}\Big)^{C_k}$$

$$\ll \zeta\Big(1+\frac{1}{\log x}\Big)^{O_k(1)} \ll (\log x)^{O_k(1)}, \tag{5.32}$$

where we have used Bernoulli's inequality on the first line and a simple integral upper bound for the zeta-function on the second line. This implies that the term in the parentheses from (5.27) is small, and so we can restrict the integrals at a small cost. Fourier inversion gives $f_{l,j}(\xi) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{t(1+i\xi)}F_{l,j}(t)dt$. Using repeated integration by parts, smoothness, and the support condition for $F_{l,j}$ from (5.14), we find that $f_{l,j}(\xi) \ll_{k,A} 1/(1+|\xi|)^A$ for all $A > 0$. Identical results hold for $g(\tau)$. Therefore we restrict the integral to $|\xi_l|, |\xi'_l|, |\tau_l|, |\tau'_l| < \sqrt{\log x}$ for all l. This gives an error $\ll_k (\log x)^{O_k(1)}\int_{|t|>\sqrt{\log x}}(1+|t|)^{-A}dt \ll_k (\log x)^{-2k}$ compared to the value of the full integral, by taking A sufficiently large.

We now return to (5.31), and estimate the value of K_p for various p. Similar estimates are computed in Lemma 30 of [1]. Estimating K_p will allow us to calculate the size of the main term from (5.27) and complete the proof of Lemma 5.10. We concentrate on different cases depending on the indicator functions in (5.31). Certainly, K_p is equal to one unless $p > W$.

We consider primes $W < p \le y$ where $p \nmid mq(h_j - h_i) - 1$ for all i, j, or primes $p > y$. Looking at the final indicator function in (5.31), we realise that any nonzero terms must satisfy $(p^{\alpha_i+\alpha'_i}, p^{\beta_j+\beta'_j}) = 1$ for all $1 \le i, j \le k$. Furthermore, for all $1 \le i \ne j \le k$, we have $(p^{\alpha_i+\alpha'_i}, p^{\alpha_j+\alpha'_j}) = 1$ and $(p^{\beta_i+\beta'_i}, p^{\beta_j+\beta'_j}) = 1$ imposed from the other indicator functions. The only surviving terms from the inner sum must therefore satisfy both of the following:

(i) Either all $b_i, b'_i = 0$ or all $a_i, a'_i = 0$

(ii) Suppose for some j we have $\alpha_j = 1$, then $\alpha'_j \in \{0, 1\}$ and all other α_i, α'_i are zero. Analogous results are true when $\alpha'_j = 1, \beta_j = 1$, or $\beta'_j = 1$.

Such conditions imply that for these primes, we have

$$K_p = 1 + \frac{1}{p}\sum_{l=1}^{k}\Big(-\frac{1}{p^{s_l}} - \frac{1}{p^{s'_l}} - \frac{1}{p^{r_l}} - \frac{1}{p^{r'_l}} + \frac{1}{p^{s_l+s'_l}} + \frac{1}{p^{r_l+r'_l}}\Big). \tag{5.33}$$

To have a nicer estimate for these K_p, we note that

$$\prod_{l=1}^{k}\frac{(1-p^{-1-s_l})(1-p^{-1-s_l'})(1-p^{-1-r_l})(1-p^{-1-r_l'})}{(1-p^{-1-s_l-s_l'})(1-p^{-1-r_l-r_l'})}=\prod_{l=1}^{k}\Big(1-\frac{1}{p^{1+s_l}}\Big)$$

$$\times\Big(1-\frac{1}{p^{1+s_l'}}\Big)\Big(1+\frac{1}{p^{1+s_l+s_l'}}+O(p^{-2})\Big)\Big(1-\frac{1}{p^{1+r_l}}\Big)\Big(1-\frac{1}{p^{1+r_l'}}\Big)\Big(1+\frac{1}{p^{1+r_l+r_l'}}+O(p^{-2})\Big)$$

$$=\prod_{l=1}^{k}\Big(1+\frac{1}{p}\Big(-\frac{1}{p^{s_l}}-\frac{1}{p^{s_l'}}+\frac{1}{p^{s_l+s_l'}}\Big)+O(p^{-2})\Big)\Big(1+\frac{1}{p}\Big(-\frac{1}{p^{r_l}}-\frac{1}{p^{r_l'}}+\frac{1}{p^{r_l+r_l'}}\Big)+O(p^{-2})\Big)$$

$$=1+\frac{1}{p}\sum_{l=1}^{k}\Big(-\frac{1}{p^{s_l}}-\frac{1}{p^{s_l'}}-\frac{1}{p^{r_l}}-\frac{1}{p^{r_l'}}+\frac{1}{p^{s_l+s_l'}}+\frac{1}{p^{r_l+r_l'}}\Big)+O_k(p^{-2}).$$

As the left hand side here is $O_k(1)$, it follows that for primes $W<p\le y$ where $p\nmid mq(h_j-h_i)-1$ for all i,j, and for primes $p>y$, we have

$$K_p=\Big(1+O_k(p^{-2})\Big)\prod_{l=1}^{k}\frac{(1-p^{-1-s_l})(1-p^{-1-s_l'})(1-p^{-1-r_l})(1-p^{-1-r_l'})}{(1-p^{-1-s_l-s_l'})(1-p^{-1-r_l-r_l'})}. \tag{5.34}$$

The only primes left to consider are primes $W<p\le y$ where $p|mq(h_j-h_i)-1$ for some $1\le i,j\le k$. Note that distinguishing between cases where $p^2|mq(h_j-h_i)-1$ will not affect our argument, so we shall assume[6] that $p^2|mq(h_j-h_i)-1$ in all cases where $p|mq(h_j-h_i)-1$. Condition (ii) from above still holds, but now we will see additional terms where $(p^{\alpha_i+\alpha_i'},p^{\beta_j+\beta_j'})=p$ or p^2, for the i,j with $p|mq(h_j-h_i)-1$. This gives

$$K_p=1+\frac{1}{p}\Bigg(\sum_{l=1}^{k}\Big(-\frac{1}{p^{s_l}}-\frac{1}{p^{s_l'}}-\frac{1}{p^{r_l}}-\frac{1}{p^{r_l'}}+\frac{1}{p^{s_l+s_l'}}+\frac{1}{p^{r_l+r_l'}}\Big)+ \sum_{(i,j):p|mq(h_j-h_i)-1}\ \sum_{\substack{\mathcal{T}\subseteq\{s_i,s_i',r_j,r_j'\}\\ \mathcal{T}\cap\{s_i,s_i'\}\neq\emptyset\\ \mathcal{T}\cap\{r_j,r_j'\}\neq\emptyset}}(-1)^{|\mathcal{T}|}p^{(-\sum_{t\in\mathcal{T}}t)}\Bigg) \tag{5.35}$$

$$=\Bigg(1+\frac{1}{p}\Bigg(\sum_{l=1}^{k}\Big(-\frac{1}{p^{s_l}}-\frac{1}{p^{s_l'}}-\frac{1}{p^{r_l}}-\frac{1}{p^{r_l'}}+\frac{1}{p^{s_l+s_l'}}+\frac{1}{p^{r_l+r_l'}}\Bigg)$$

$$\times\Bigg(\prod_{(i,j):p|mq(h_j-h_i)-1}\Big(1+\frac{1}{p}\sum_{\substack{\mathcal{T}\subseteq\{s_i,s_i',r_j,r_j'\}\\ \mathcal{T}\cap\{s_i,s_i'\}\neq\emptyset\\ \mathcal{T}\cap\{r_j,r_j'\}\neq\emptyset}}(-1)^{|\mathcal{T}|}p^{(-\sum_{t\in\mathcal{T}}t)}\Big)\Bigg)+O_k(p^{-2}).$$

[6] Cases where $p^2|mq(h_j-h_i)-1$ corresponds to allowing the choice $\mathcal{T}=\{s_i,s_i',r_j,r_j'\}$ in (5.35).

Therefore, in this case, K_p is the same as in (5.34), but with an additional factor of

$$\begin{aligned}&\Big(1+O_k(p^{-2})\Big)\prod_{(i,j):p|mq(h_j-h_i)-1}\Big(1+\frac{1}{p}\sum_{\substack{\mathcal{T}\subseteq\{s_i,s_i',r_j,r_j'\}\\ \mathcal{T}\cap\{s_i,s_i'\}\neq\emptyset\\ \mathcal{T}\cap\{r_j,r_j'\}\neq\emptyset}}(-1)^{|\mathcal{T}|}p^{(-\sum_{t\in\mathcal{T}}t)}\Big)\\ =&\Big(1+O_k(p^{-2})\Big)\prod_{(i,j):p|mq(h_j-h_i)-1}\Big(1+\frac{1}{p}\Big(1+O\Big(\frac{\log p\sqrt{\log x}}{\log y}\Big)\Big)\Big)\\ =&\Big(1+\frac{\#\{(i,j):p|mq(h_j-h_i)-1\}}{p}\Big)\Big(1+O_k\Big(\frac{1}{p^2}+\frac{\log p\sqrt{\log x}}{p\log y}\Big)\Big)\end{aligned}\tag{5.36}$$

where we have used the fact that $p^{-s_j}=1-s_j\log p+...=1+O(\log p\sqrt{\log x}/\log y)$ by Taylor expansion, and $|s_j|,|r_j|,|s_j'|,|r_j'|\ll\sqrt{\log x}/\log y$, as we restricted the integral to $|\xi_l|,|\xi_l'|,|\tau_l|,|\tau_l'|<\sqrt{\log x}$.

We wish to estimate the quantity $\#\{(i,j):p|mq(h_j-h_i)-1\}$. Note that for primes p where $W<p\le y$ and $p|mq(h_j-h_i)-1$ for some i,j, we have $h_iq \pmod p$ are all distinct. Otherwise, we must have $h_s\equiv h_t \pmod p$, ($p\nmid q$ as q prime and $p<y<q$) but $p\nmid h_s-h_t$ for large x by definition of our admissible set ($h_i=p_{\pi(k)+i}P_W$, where k will be fixed). Recalling our definition of $g_{m,q}$ from (5.16), we notice that we have k different monomials $n+h_iq \pmod p$, so we can find k different solutions to $\prod_{i=1}^k n+h_iq\equiv 0 \pmod p$. The value is therefore determined by the second condition, and we find that $g_{m,q}(p)=2k-\#\{(i,j):p|mq(h_j-h_i)-1\}$. We may substitute this into the first term of (5.36).

We proceed in showing that the second term in (5.36) is negligible. Recall that we wished to estimate (5.29). The total contribution to this sum from the latter term in (5.36) will come from a product over primes $p>W$ where $p|\prod_{h,h'}(mq(h-h')-1)$. We note that we have $m<U/z(\log_2 x)^2$, so $\prod_{h,h'}(mq(h-h')-1)\ll x^{O_k(1)}$, where $q=O(x)$ is the dominating term. We note that small primes have the largest contribution to our product, so the worst case scenario would be when $\prod_{h,h'}(mq(h-h')-1)$ is simply a product of all primes below $O_k(\log x)$ (which is maximal so that $\prod_{h,h'}(mq(h-h')-1)=x^{O_k(1)}$). Also we have $\log y>(\log x)^{1-\epsilon}$. The contribution is therefore

$$\begin{aligned}&\prod_{\substack{p>W\\ p|\prod_{h,h'}(mq(h-h')-1)}}\Big(1+O_k\Big(\frac{1}{p^2}+\frac{\log p\sqrt{\log x}}{p\log y}\Big)\Big)\\ &\ll\prod_{W\le p\le O_k(\log x)}\Big(1+O_k\Big(\frac{1}{p^2}+\frac{\log p}{p(\log x)^{1/2-\epsilon}}\Big)\Big)\\ &\ll\exp\Big(\sum_{W\le p\le O_k(\log x)}O_k\Big(\frac{1}{p^2}+\frac{\log p}{p(\log x)^{1/2-\epsilon}}\Big)\Big)\\ &\ll\exp\Big(O_k\Big(W^{-1}+\frac{\log\log x}{(\log x)^{1/2+\epsilon}}\Big)\Big)=1+o_k(1).\end{aligned}$$

We may now accurately calculate the term in the first parentheses from (5.27), by using the fact that it is equal to $\prod_p K_p$ (by combining (5.29) and (5.30)). The contribution to $\prod_p K_p$ from all primes $W < p < y$ may simply be written as

(5.37)
$$(1+o_k(1))\prod_{W<p\leq y}\Big(1+\frac{2k-g_{m,q}(p)}{p}\Big)\prod_{l=1}^{k}\frac{(1-p^{-1-s_l})(1-p^{-1-s'_l})(1-p^{-1-r_l})(1-p^{-1-r'_l})}{(1-p^{-1-s_l-s'_l})(1-p^{-1-r_l-r'_l})},$$

as any primes p where $p \nmid \prod_{h,h'}(mq(h-h')-1)$ will have $g_{m,q}(k) = 2k$, so will not contribute to this product. Therefore we find that

$$\prod_p K_p = (1+o_k(1))\prod_{W<p\leq y}\Big(1+\frac{2k-g_{m,q}(p)}{p}\Big)$$
(5.38)
$$\times\prod_{p>W}\prod_{l=1}^{k}\frac{(1-p^{-1-s_l})(1-p^{-1-s'_l})(1-p^{-1-r_l})(1-p^{-1-r'_l})}{(1-p^{-1-s_l-s'_l})(1-p^{-1-r_l-r'_l})}.$$

We wish to extend this last product to all primes p, allowing us to obtain products of Riemann zeta functions, which we understand sufficiently well. With this in mind, we note that $s_l, s'_l, r_l, r'_l = o((\log x)^{-1/2+\epsilon})$. Therefore $p^{-s_l} = 1 + o((\log x)^{-1/2+2\epsilon})$, and

$$\prod_{p\leq W}\frac{(1-p^{-1-s_l})(1-p^{-1-s'_l})}{1-p^{-1-s_l-s'_l}} = \prod_{p\leq W}\Big((1-p^{-1-s_l}-p^{-1-s'_l}+p^{-2-s_l-s'_l})\sum_{\alpha\geq 0}p^{\alpha(-1-s_l-s'_l)}\Big)$$
$$= \prod_{p\leq W}\Big(1-p^{-1}+o((\log x)^{-1/2+2\epsilon}+\sum_{\alpha\geq 0}p^{-\alpha}o((\log x)^{-1/2+\epsilon})\Big)$$
$$= \prod_{p\leq W}\Big(1-p^{-1}+o((\log x)^{-1/2+\epsilon})\Big) = \prod_{p\leq W}\Big(1-p^{-1}\Big)\Big(1+o((\log x)^{-1/2+\epsilon})\Big)$$
(5.39)
$$= (1+o(1))\prod_{p\leq W}\Big(1-\frac{1}{p}\Big),$$

where we can use an exponential-log trick with Mertens' estimate to obtain the final line, noting $W = \log_4 x$. We extend this to find

(5.40)
$$\prod_{p\leq W}\prod_{l=1}^{k}\frac{(1-p^{-1-s_l})(1-p^{-1-s'_l})(1-p^{-1-r_l})(1-p^{-1-r'_l})}{(1-p^{-1-s_l-s'_l})(1-p^{-1-r_l-r'_l})} = (1+o_k(1))\prod_{p\leq W}\Big(1-\frac{1}{p}\Big)^{2k}.$$

We may now extend the product in (5.38) to all primes, arriving at

$$\prod_p K_p = (1+o_k(1))\prod_{p\leq W}\Big(1-\frac{1}{p}\Big)^{-2k}\prod_{W<p\leq y}\Big(1+\frac{2k-g_{m,q}(p)}{p}\Big)$$
$$\times\prod_p\prod_{l=1}^{k}\frac{(1-p^{-1-s_l})(1-p^{-1-s'_l})(1-p^{-1-r_l})(1-p^{-1-r'_l})}{(1-p^{-1-s_l-s'_l})(1-p^{-1-r_l-r'_l})}.$$

We now use the fact that $\zeta(1+s_l) = \prod_p (1-1/p^{-1-s_l})^{-1}$, where $s_l = 1 + i\xi_l/\log x$ (and similarly for s_l', r_l, r_l'), to obtain

$$\prod_p K_p = (1+o_k(1)) \prod_{p \le W} \Big(1 - \frac{1}{p}\Big)^{-2k} \prod_{W < p \le y} \Big(1 + \frac{2k - g_{m,q}(p)}{p}\Big)$$
$$\times \prod_{l=1}^{k} \frac{\zeta(1 + \frac{2+i\xi_l+i\xi_l'}{\log x})\zeta(1+\frac{2+i\tau_l+i\tau_l'}{\log y})}{\zeta(1+\frac{1+i\xi_l}{\log x})\zeta(1+\frac{1+i\xi_l'}{\log x})\zeta(1+\frac{1+i\tau_l}{\log y})\zeta(1+\frac{1+i\tau_l'}{\log y})}. \tag{5.41}$$

We note that the bound $\zeta(1+x) \ll 1/x$ for $x > 1$ real can be extended to complex numbers, as $s = 1$ is a simple pole of $\zeta(s)$. Therefore $|z| = o(1)$ implies $\zeta(1+z) = (1+o(1))/z$. This gives

$$\prod_p K_p = \frac{1+o_k(1)}{(\log x)^k (\log y)^k} \prod_{p \le W} \Big(1 - \frac{1}{p}\Big)^{-2k} \prod_{W < p \le y} \Big(1 + \frac{2k - g_{m,q}(p)}{p}\Big)$$
$$\times \prod_{l=1}^{k} \frac{(1+i\xi_l)(1+i\xi_l')(1+i\tau_l)(1+i\tau_l')}{(2+i\xi_l+i\xi_l')(2+i\tau_l+i\tau_l')}. \tag{5.42}$$

We substitute this into (5.27), to find that the main term for the sum of the weights (5.19) is

$$\frac{U\varphi_{m,q}(P_W)}{m(\log x)^k(\log y)^k P_W} \prod_{p \le W} \Big(1 - \frac{1}{p}\Big)^{-2k} \prod_{W < p \le y} \Big(1 + \frac{2k - g_{m,q}(p)}{p}\Big) \sum_{j=1}^{J} \sum_{j'=1}^{J}$$
(5.43)
$$\int ... \int (1+o_k(1)) \prod_{l=1}^{k} \Big(\frac{(1+i\xi_l)(1+i\xi_l')(1+i\tau_l)(1+i\tau_l')}{(2+i\xi_l+i\xi_l')(2+i\tau_l+i\tau_l')} f_{l,j}(\xi_l) f_{l,j'}(\xi_l') g(\tau_l) g(\tau_l') \Big) d\xi_l d\xi_l' d\tau_l d\tau_l',$$

where the integral is running over $|\xi_l|, |\xi_l'|, |\tau_l|, |\tau_l'| \le \sqrt{\log x}$. Due to the rapid decay of $f_{l,j}, g$, note that the integral is bounded by a function of k. Hence the $o_k(1)$ term contributes $o_k(1)$ to the total, and we can take it out of the integral. Furthermore, we can the extend the integral to run over $\mathbb{R}$ in all variables, at a cost of $o_k(1)$ in an analogous way to how we previously limited the domain of integration. Therefore we have the main term

$$\frac{U\varphi_{m,q}(P_W)}{m(\log x)^k(\log y)^k P_W} \prod_{p \le W} \Big(1 - \frac{1}{p}\Big)^{-2k} \prod_{W < p \le y} \Big(1 + \frac{2k - g_{m,q}(p)}{p}\Big) \sum_{j=1}^{J} \sum_{j'=1}^{J} \prod_{l=1}^{k} (1+o_k(1))$$
(5.44)
$$\int_{\mathbb{R}} ... \int_{\mathbb{R}} \Big(\frac{(1+i\xi_l)(1+i\xi_l')(1+i\tau_l)(1+i\tau_l')}{(2+i\xi_l+i\xi_l')(2+i\tau_l+i\tau_l')} f_{l,j}(\xi_l) f_{l,j'}(\xi_l') g(\tau_l) g(\tau_l') \Big) d\xi_l d\xi_l' d\tau_l d\tau_l',$$

and we note that by differentiating $F_{l,j}(t) = e^{-t} \int_{\mathbb{R}} e^{-i\xi t} f_{l,j}(\xi) d\xi$ to find $F_{l,j}'(t) F_{l,j'}'(t)$, and then integrating over t, we have

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \frac{(1+i\xi_l)(1+i\xi_l')}{2+i\xi_l+i\xi_l'} f_{l,j}(\xi_l) f_{l,j'}(\xi_l') d\xi_l d\xi_l' = \int_0^\infty F_{l,j}'(t) F_{l,j'}'(t) dt,$$

and similarly

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{(1+i\tau_l)(1+i\tau_l')}{2+i\tau_l+i\tau_l'} g(\tau_l)g(\tau_l')d\tau_l d\tau_l' = \int_0^\infty G'(t)^2 dt.$$

We recall the definitions of $I_k^1(F)$ and $I_k^2(G)$ from Lemma 5.10

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

and we also recall that $F(t_1,...,t_k) = \sum_{j=1}^J \prod_{l=1}^k F_{l,j}'(t_l)$ from (5.15). Therefore, combining (5.44) with the fact that the error terms accumulated (i.e. from (5.24)) are sufficiently small, we have that the sum of the weights is

$$\sum_{\substack{n \leq U/m \\ (n(nm-1),P_W)=1}} w_{m,q}(n) = \frac{(1+o_k(1))U\varphi_{m,q}(P_W)}{m(\log x)^k(\log y)^k P_W} \prod_{p \leq W}\Big(1-\frac{1}{p}\Big)^{-2k} \prod_{W<p\leq y}\Big(1-\frac{g_{m,q}(p)-2k}{p}\Big)$$

$$\times \sum_{j=1}^J \sum_{j'=1}^J \prod_{l=1}^k \Big(\int_0^\infty F_{l,j}'(t)F_{l,j'}'(t)dt \int_0^\infty G'(t)^2 dt\Big). \tag{5.45}$$

We have

$$\sum_{j=1}^J \sum_{j'=1}^J \prod_{l=1}^k \Big(\int_0^\infty F_{l,j}'(t)F_{l,j'}'(t)dt \int_0^\infty G'(t)^2 dt\Big) = I_k^2(G)I_k^1(F),$$

and, recalling (5.17), we also have

$$\frac{\varphi(P_W)}{P_W} = \prod_{p\leq W}\Big(1-\frac{g_{m,q}(p)}{p}\Big).$$

This allows us to write

$$\begin{aligned}
&\frac{\varphi(P_W)}{P_W}\prod_{p\leq W}\Big(1-\frac{1}{p}\Big)^{-2k}\prod_{W<p\leq y}\Big(1-\frac{g_{m,q}(p)-2k}{p}\Big) \\
&= \prod_{p\leq y}\Big(1-\frac{1}{p}\Big)^{-2k}\prod_{p\leq W}\Big(1-\frac{g_{m,q}(p)}{p}\Big)\prod_{W<p\leq y}\Big(1-\frac{g_{m,q}(p)}{p}+\frac{2k}{p}\Big)\Big(1-\frac{2k}{p}+O_k\Big(\frac{1}{p^2}\Big)\Big) \\
&= \prod_{p\leq y}\Big(1-\frac{1}{p}\Big)^{-2k}\prod_{p\leq y}\Big(1-\frac{g_{m,q}(p)}{p}\Big)\prod_{W<p\leq y}\Big(1+O_k\Big(\frac{1}{p^2}\Big)\Big) = (1+o(1))\mathcal{G}_{m,q}
\end{aligned}$$

where $\mathcal{G}_{m,q}$ is as defined in (5.18). We find therefore find that that (5.45) implies

$$\sum_{\substack{n \leq U/m \\ (n(nm-1),P_W)=1}} w_{m,q}(n) = \Big(1+o_k(1)\Big)\frac{U\mathcal{G}_{m,q}I_k^{(1)}(F)I_k^{(2)}(G)}{m(\log x)^k(\log y)^k},$$

completing the proof of 5.10. □

Recall that the goal here is to show that $\sum_{q\in\mathcal{I}'_m}\nu_{m,q}(p)\geq -\log\epsilon/2$ for all $p\in\mathcal{R}_m\backslash\mathcal{C}_m$, for $\mathcal{C}_m=o(|\mathcal{R}_m|)$ and $\nu_{m,q}$ as defined in (5.11). Lemma 5.10 gives the second (normalising) factor in this expression, in an analogous way to S_1 in the proof of small gaps between primes. In the following lemma, we will use this to finally compute an estimate for $\sum_{q\in\mathcal{I}'_m}\nu_{m,q}(p)$. Then we will just need to choose $F_{l,j}$ and G to make this value sufficiently large.

Lemma 5.11. *Let $m<U/z(\log x)^2$ be even and let $p_0\in\mathcal{R}_m$ with $h_kx<p_0<U/m-h_kx$. Then*

$$\sum_{q\in\mathcal{I}'_m}\nu_{m,q}(p_0)\gg(1+o_k(1))\frac{k|\mathcal{I}'_m|J_k^{(1)}(F)J_k^{(2)}(G)}{(\log x)|\mathcal{R}_m|I_k^{(1)}(F)I_k^{(2)}(G)},\tag{5.46}$$

where $I_k^{(1)}$ and $I_k^{(2)}$ are as defined in Lemma 5.10, and

$$J_k^{(1)}(F)=\int\cdots\int_{t_1,\dots,t_{k-1}\geq0}\Big(\int_{t_k\geq0}F(t_1,...,t_k)dt_k\Big)^2dt_1\dots dt_{k-1},$$

and

$$J_k^{(2)}(G)=G(0)^2\Big(\int_0^\infty G'(t)^2dt\Big)^{k-1}.$$

Proof. First of all, for simplicity of notation, we write

$$T_{m,q}=\sum_{\substack{n\leq U/m\\(n(nm-1),P_W)=1}}w_{m,q}(n),$$

which is the quantity calculated in Lemma 5.10. By (5.11) we then have

$$\sum_{q\in\mathcal{I}'_m}\nu_{m,q}(p_0)=\sum_{q\in\mathcal{I}'_m}T_{m,q}^{-1}\sum_{\substack{n\leq U/m\\n\equiv p_0\pmod q\\(n(nm-1),P_W)=1}}\Big(\sum_{\substack{d_1,\dots,d_k\\d_i|n+h_iq}}\sum_{\substack{e_1,\dots,e_k\\e_i|m(n+h_iq)-1}}\lambda_{d_1,\dots,d_k,e_1,\dots,e_k}\Big)^2.$$

All terms in this sum are non-negative, and we are trying to find a lower bound. Therefore we can restrict the sum over n to $n=p_0-hq$ for $h\in\mathcal{H}$. This restriction reformulates our problem to involve admissible sets, allowing us to use machinery from the small gaps proof. We note that, for $n=p_0-hq$, the condition on the size of p_0 implies $1\leq p_0-hq\leq U/m$. Furthermore, $p_0\in\mathcal{R}_m$ is prime with $(mp-1,P_y)=1)$, so the fact that $P_W|h$ for all $h\in\mathcal{H}$ gives us $(n(nm-1),P_W)=1$ for $n=p_0-hq$. Therefore, we have

$$\sum_{q\in\mathcal{I}'_m}\nu_{m,q}(p_0)\geq\sum_{h\in\mathcal{H}}\sum_{q\in\mathcal{I}'_m}T_{m,q}^{-1}\Big(\sum_{\substack{d_1,\dots,d_k\\d_i|p_0+(h_i-h)q}}\sum_{\substack{e_1,\dots,e_k\\e_i|m(p_0+(h_i-h)q)-1}}\lambda_{d_1,\dots,d_k,e_1,\dots,e_k}\Big)^2.\tag{5.47}$$

As we are considering $q\in\mathcal{I}'_m$ a prime $\geq x/2$, we must have $(q,P_W)=1$. Therefore we shall consider separately the possible residue classes that q can lie in modulo P_W. This quantity is of interest because we want to use the Bombieri-Vinogradov theorem, similarly to how we did so in bounded gaps. From this, we find that $\sum_{q\in\mathcal{I}'_m}\nu_{m,q}(p_0)$ is bounded below by

(5.48)
$$\sum_{h\in\mathcal{H}}\sum_{\substack{w_0 \pmod{P_W}\\(w_0,P_W)=1}}\sum_{\substack{q\in\mathcal{I}'_m\\ q\equiv w_0 \pmod{P_W}}} T_{m,q}^{-1}\Big(\sum_{\substack{d_1,\dots,d_k\\ d_i|p_0+(h_i-h)q}}\sum_{\substack{e_1,\dots,e_k\\ e_i|m(p_0+(h_i-h)q)-1}}\lambda_{d_1,\dots,d_k,e_1,\dots,e_k}\Big)^2.$$

We wish to lose the dependence on q that arises from $T_{m,q}$ in Lemma 5.10, which will otherwise prevent us from using the Bombieri-Vinogradov theorem. We note that the dependence here arises from the term $\mathcal{G}_{m,q}$, and so we wish to simplify this term. By considering the value of $g_{m,q}$ from (5.16) for the cases

(i) $p=2$ (then $p|m$ as recall m is even for $\mathcal{R}_m$ nonempty),

(ii) $2<p\le W$, $p|m$,

(iii) $p\le W$, $(m,p)=1$,

(iv) $W<p\le y$, $p\nmid\prod_{h',h''\in\mathcal{H}}(mq(h-h')-1)$,

(v) $W<p\le y$, $p|\prod_{h',h''\in\mathcal{H}}(mq(h-h')-1)$,

it can be shown that

$$\mathcal{G}_{m,q}^{-1}\ge(1+o_k(1))\mathcal{G}_m\prod_{\substack{W<p\le y\\ p|\prod_{h',h''\in\mathcal{H}}(mq(h-h')-1)}}\Big(1-\frac{2k}{p}\Big), \tag{5.49}$$

where

$$\mathcal{G}_m=2^{-(2k-1)}\Big(\prod_{\substack{p|m\\p>2}}\frac{p-2}{p-1}\Big)\prod_{2<p\le W}\left(\Big(1-\frac{2}{p}\Big)^{-1}\Big(1-\frac{1}{p}\Big)^{2k}\right). \tag{5.50}$$

We may restrict the product over primes on the right hand size of (5.49) to primes less than $z_0=\log x/\log_2 x$ at a cost of a factor of $(1+o_k(1))$, giving the product

$$\prod_{\substack{W<p\le z_0\\ p|\prod_{h',h''\in\mathcal{H}}(mq(h-h')-1)}}\Big(1-\frac{2k}{p}\Big).$$

We recognise this as an Euler product of a multiplicative function in $k(k-1)/2$ variables, where each variable corresponds to a divisor of $mq(h_i-h_j)-1$ for some i,j. Therefore we have

$$\prod_{\substack{W<p\le z_0\\ p|\prod_{h',h''\in\mathcal{H}}(mq(h-h')-1)}}\Big(1-\frac{2k}{p}\Big)=\prod_p\sum_{\beta_{1,2},\dots,\beta_{k,k-1}}L(p^{\beta_{1,2}},\dots,p^{\beta_{k,k-1}}),$$

where

$$L(p^{\beta_{1,2}},...,p^{\beta_{k,k-1}})=\frac{(-2k)^{\omega(p^{\max\{\beta_{1,2},...,\beta_{k,k-1}\}})}}{p^{\max\{\beta_{1,2},...,\beta_{k,k-1}\}}}\prod_{\substack{1\le i,j\le k\\ i\neq j}}\mathbf{1}(p^{\beta_{i,j}}|P_{z_0}/P_W)\mathbf{1}(p^{\beta_{i,j}}|mq(h_j-h_i)-1).$$

Note that $mq(h-h')-1$ are coprime for distinct h,h', so only a single $\beta_{i,j}$ may be nonzero in any tuple. Therefore

$$\prod_{\substack{W<p\le z_0\\ p|\prod_{h',h''\in\mathcal{H}}(mq(h-h')-1)}}\Big(1-\frac{2k}{p}\Big)=\sum_{\substack{a_{1,2},...,a_{k,k-1}|P_{z_0}/P_W\\ a_{i,j}|mq(h_j-h_i)-1}}\frac{(-2k)^{\omega([\mathbf{a}])}}{[\mathbf{a}]},$$

where we are writing $[\mathbf{a}]$ for the lowest common multiple of $a_{1,2},...,a_{k,k-1}$. Hence from (5.49), we have

$$\mathcal{G}_{m,q}^{-1}\ge(1+o_k(1))\mathcal{G}_m\sum_{\substack{a_{1,2},...,a_{k,k-1}|P_{z_0}/P_W\\ a_{i,j}|mq(h_j-h_i)-1}}\frac{(-2k)^{\omega([\mathbf{a}])}}{[\mathbf{a}]}.$$

Therefore we have removed all dependence on q from the expression, with the exception of $a_{i,j}|mq(h_j-h_i)-1$. We view this as a condition on the residue class of q modulo $a_{i,j}$, which we permit in the use of the Bombieri-Vinogradov theorem. Combining the above expression with Lemma 5.10, we have

$$T_{m,q}^{-1}\ge\Big(1+o_k(1)\Big)\frac{m\mathcal{G}_m(\log x)^k(\log y)^k}{UI_k^{(1)}(F)I_k^{(2)}(G)}\sum_{\substack{a_{1,2},...,a_{k,k-1}|P_{z_0}/P_W\\ a_{i,j}|mq(h_j-h_i)-1}}\frac{(-2k)^{\omega([\mathbf{a}])}}{[\mathbf{a}]}. \tag{5.51}$$

From (5.48) we now have the lower bound

$$\Big(1+o_k(1)\Big)\frac{m\mathcal{G}_m(\log x)^k(\log y)^k}{UI_k^{(1)}(F)I_k^{(2)}(G)}\sum_{h\in\mathcal{H}}\sum_{\substack{w_0\pmod{P_W}\\(w_0,P_W)=1}}\sum_{a_{1,2},...,a_{k,k-1}|P_{z_0}/P_W}\frac{(-2k)^{\omega([\mathbf{a}])}}{[\mathbf{a}]}$$

$$\times\Bigg(\sum_{\substack{q\in\mathcal{I}'_m\\ q\equiv w_0\pmod{P_W}\\ a_{i,j}|mq(h_j-h_i)-1}}\Big(\sum_{\substack{d_1,...,d_k\\ d_i|p_0+(h_i-h)q}}\sum_{\substack{e_1,...,e_k\\ e_i|m(p_0+(h_i-h)q)-1}}\lambda_{d_1,...,d_k,e_1,...,e_k}\Big)^2\Bigg). \tag{5.52}$$

We focus on the second line, i.e. sum over q. We consider the case when $h=h_k$ in the outer sum over $\mathcal{H}$. Cases corresponding to other choices of h are analogous. We note that with $h=h_k$, the sum over d_k's has the condition that $d_k|p_0$. The support conditions from (5.14) imply that $\lambda_{\mathbf{d},\mathbf{e}}$ is zero when $d_k=p_0>x^{1/10}$, therefore we only consider the case when $d_k=1$. Similarly we only consider $d'_k=e_k=e'_k=1$. We now write the sum over q from (5.52) as

$$\sum_{\substack{d_1,...,d_k\\ d'_1,...,d'_k\\ d_k=d'_k=1}}\sum_{\substack{e_1,...,e_k\\ e'_1,...,e'_k\\ e_k=e'_k=1}}\lambda_{\mathbf{d},\mathbf{e}}\lambda_{\mathbf{d}',\mathbf{e}'}\sum_{\substack{q\in\mathcal{I}'_m\\ q\equiv w_0\pmod{P_W}\\ [d_i,d'_i]|p_0+(h_i-h_k)q\ \forall i\neq k\\ [e_i,e'_i]|p_0+m(h_i-h_k)q-1\ \forall i\neq k\\ a_{i,j}|mq(h_j-h_i)-1}}1, \tag{5.53}$$

which is reminiscent of (5.20). We find in a similar way that terms on the inner sum satisfy

$$\text{(5.54)}\qquad \begin{aligned} &\text{(i) } d_1d_1', ..., d_kd_k', P_W \text{ are pairwise coprime,}\\ &\text{(ii) } e_1e_1', ..., e_ke_k', P_W \text{ are pairwise coprime,}\\ &\text{(iii) } a_{1,2}, ..., a_{k,k-1}, m \text{ are pairwise coprime,}\\ &\text{(iv) } (d_id_i', e_je_j')|mp_0(h_i - h_j) + h_k - h_i\ \forall i, j,\\ &\text{(v) } (d_id_i', a_{j,l})|(h_j - h_l)mp_0 + h_i - h_k\ \forall i, j, l,\\ &\text{(vi) } (e_ie_i', a_{j,l})|(h_j - h_l)(1 - p_0m) - h_i + h_k\ \forall i, j, l, \end{aligned}$$

where, for example, condition (iv) can be found in the following way: if $p|d_i$ and $p|e_j$ then $p|p_0 + (h_i - h_k)q$ and $p|mp_0 + m(h_j - h_k)q - 1$. Multiplying $p_0 + (h_i - h_k)q$ by $m(h_j - h_k)$ and $mp_0 + m(h_j - h_k)q - 1$ by $(h_i - h_k)$, we find that $p|mp_0(h_j - h_k) + m(h_i - h_j)(h_j - h_k)q$ and $p|mp_0(h_i - h_k) + m(h_j - h_k)(h_i - h_k)q - h_i + h_k$. Subtracting the first condition from the second gives $p|mp_0(h_i - h_j) + h_k - h_i$, and so we arrive at (iv).

We wish to write this inner sum in (5.53) as a sum over a single residue class. Here we similarly make use of the Chinese Remainder Theorem, allowing us to write the sum as a sum over q in a single (invertible) residue class modulo $P_W[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}', \mathbf{a}]$, assuming that the congruence conditions are consistent. We note that this is guaranteed by the conditions in (5.54), also in an analogous way to as seen in Lemma 5.10.

Let $R = P_W[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}', \mathbf{a}]$. Now for some $a \in (\mathbb{Z}/R\mathbb{Z})^\times$, the inner sum in (5.53) may be written as

$$\text{(5.55)}\qquad \sum_{\substack{n \in \mathcal{I}_m' \\ n \equiv a \pmod R}} \mathbf{1}_p(n).$$

We let $\lceil \mathcal{I}_m' \rceil$ denote the largest element of the interval $\mathcal{I}_m'$, and similarly define $\lfloor \mathcal{I}_m' \rfloor$ as the smallest element of the interval $\mathcal{I}_m'$. Then we may write

$$\begin{aligned} &\sum_{\substack{n \in \mathcal{I}_m' \\ n \equiv a \pmod R}} \mathbf{1}_p(n) = \pi(\lceil \mathcal{I}_m' \rceil; A, R) - \pi(\lfloor \mathcal{I}_m' \rfloor - 1; A, R)\\ \implies &\sum_{\substack{n \in \mathcal{I}_m' \\ n \equiv a \pmod R}} \mathbf{1}_p(n) = \frac{\pi(\lceil \mathcal{I}_m' \rceil)}{\varphi(R)} + \Big(\pi(\lceil \mathcal{I}_m' \rceil; A, R) - \frac{\pi(\lceil \mathcal{I}_m' \rceil)}{\varphi(R)}\Big)\\ &\qquad - \frac{\pi(\lfloor \mathcal{I}_m' \rfloor - 1)}{\varphi(R)} - \Big(\pi(\lfloor \mathcal{I}_m' \rfloor - 1; A, R) - \frac{\pi(\lfloor \mathcal{I}_m' \rfloor - 1)}{\varphi(R)}\Big)\\ \implies &\sum_{\substack{n \in \mathcal{I}_m' \\ n \equiv a \pmod R}} \mathbf{1}_p(n) = \frac{\sum_{n \in \mathcal{I}_m'} \mathbf{1}_p(n)}{\varphi(P_W)\varphi([\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}', \mathbf{a}])} + O\Big(E(x, R)\Big), \end{aligned}$$

where

$$E(x; q) = \sup_{t \le x} \sup_{(a,q)=1} \Big|\pi(t; q, a) - \frac{\pi(t)}{\varphi(q)}\Big|.$$

We note that due to the support of λ from (5.14), and the fact that $a_{i,j}|P_{z_0}/P_W = x^{o(1)}$, we have $R \ll x^{1/5+o_k(1)}$. We may now write (5.53) as

$$\sum_{\substack{d_1,\dots,d_k \\ d'_1,\dots,d'_k \\ d_k=d'_k=1}} \sum_{\substack{e_1,\dots,e_k \\ e'_1,\dots,e'_k \\ e_k=e'_k=1}} \lambda_{\mathbf{d},\mathbf{e}}\lambda_{\mathbf{d}',\mathbf{e}'}\Big(\frac{\sum_{n\in\mathcal{I}'_m}\mathbf{1}_p(n)}{\varphi(P_W)\varphi([\mathbf{d},\mathbf{d}',\mathbf{e},\mathbf{e}',\mathbf{a}])} + O\Big(E(x;R)\Big)\Big). \tag{5.56}$$

We have the facts $|\lambda_{\mathbf{d},\mathbf{e}}| \ll_k 1$, and $E(x;q) \ll x/q$. This second estimate follows from the fact that $\pi(x;q,a)$ is counting primes amongst a set containing x/q elements, and $\pi(x)/\varphi(q) \ll x/\varphi(q)\log x \ll x/q$. Furthermore, similar estimates to those following (4.14) allow us to find that there are at most $\tau_{k^2+4k}(r)$ choices of $a_{1,2},...,a_{k,k-1}$, $d_1,...,d_{k-1}$, $e_1,...,e_{k-1}$, $d'_1,...,d'_{k-1}$ and $e'_1,...,e'_{k-1}$ that satisfy $P_W[\mathbf{d},\mathbf{d}',\ \mathbf{e}\ ,\mathbf{e}',\mathbf{a}] = r$. We note that this specific number of divisors k^2+4k can be replaced by any large function in k if one wishes, with no effect on our argument. Therefore the error term arising in (5.52), when we additionally sum over $\sum_{h\in\mathcal{H}}\sum_{w_0}\sum_{a_{i,j}}$, is

$$\begin{aligned}
&\ll_k \sum_{r\ll x^{1/5+\epsilon}} \tau_{k^2+4k}(r)E(x;r) \\
&\ll_k \Big(\sum_{r\ll x^{1/5+\epsilon}} \tau_{k^2+4k}(r)^2 E(x;r)\Big)^{1/2}\Big(\sum_{r\ll x^{1/5+\epsilon}} E(x;r)\Big)^{1/2} \\
&\ll_k x^{1/2}\Big(\sum_{r\ll x^{1/5+\epsilon}} \frac{\tau_{k^2+4k}(r)^2}{r}\Big)^{1/2}\Big(\sum_{r\ll x^{1/5+\epsilon}} \sup_{t\le x}\sup_{(a,r)=1}\Big|\pi(t;r,a)-\frac{\pi(t)}{\varphi(r)}\Big|\Big)^{1/2} \\
&\ll_k x^{3/5+\epsilon}(\log x)^{O_k(1)}\Big(\sum_{r\ll x^{1/5+\epsilon}} \sup_{t\le x}\sup_{(a,r)=1}\Big|\pi(t;r,a)-\frac{\pi(t)}{\varphi(r)}\Big|\Big)^{1/2} \\
&\ll_{k,A} x^{3/5+\epsilon}(\log x)^{O_k(1)}\Big(\frac{x^{2/5+3\epsilon}}{(\log x)^A}\Big)^{1/2} \ll_k \frac{x}{(\log x)^{2k}},
\end{aligned} \tag{5.57}$$

where we have used (6.18) to estimate the first factor and the Bombieri-Vinogradov theorem to estimate the second factor. Note that here we have used a slightly stronger form of the Bombieri-Vonogradov theorem than Theorem 3.2, but this may be proved in exactly the same way. We have taken the implied constant sufficiently large (depending on k) to give a factor of $(\log x)^{2k}$ in the final denominator.

Remark. This calculation motivates the previous choice of support for the functions $F_{i,j}$ and G from (5.14), as these made d_i's and other variables sufficiently small to make this error was negligible. We will now also see why $\mathcal{I}'_m$ needs to be of length at least $\delta|\mathcal{R}_m|\log x$, which was an assumption of Proposition 5.7.

It follows from (5.05) that if $\mathcal{I}'_m$ is of length $\delta|\mathcal{R}_m|\log x$, then $|\mathcal{I}'_m| \gg x/(\log x)^2$. This allows us to use the prime number theorem as the error term will have negligible effect as x gets large. We know that $\mathcal{I}'_m$ is some interval contained in $[x/2,x]$, so it follows from the prime number theorem that the number of primes in this interval is $(1+o(1))|\mathcal{I}'_m|/\log x$. Therefore, the main term in (5.56) simplifies to

$$\frac{(1+o(1))|\mathcal{I}'_m|}{\log x}\sideset{}{^*}\sum_{\substack{d_1,\dots,d_k\\ d'_1,\dots,d'_k\\ d_k=d'_k=1}}\sideset{}{^*}\sum_{\substack{e_1,\dots,e_k\\ e'_1,\dots,e'_k\\ e_k=e'_k=1}}\frac{\lambda_{\mathbf{d},\mathbf{e}}\lambda_{\mathbf{d}',\mathbf{e}'}}{\varphi(P_W)\varphi([\mathbf{d},\mathbf{d}',\mathbf{e},\mathbf{e}',\mathbf{a}])}, \tag{5.58}$$

where $\sum^*$ denotes that we only sum over tuples satisfying the conditions from (5.54). We now insert our expression for λ,

$$\lambda_{d_1,\dots,d_k,e_1,\dots,e_k}=\sum_{j=1}^{J}\Big(\prod_{l=1}^{k}\mu(d_l)\mu(e_l)F_{l,j}\Big(\frac{\log d_l}{\log x}\Big)G\Big(\frac{\log e_l}{\log y}\Big)\Big),$$

and we use the fact that $d_k=d'_k=1$, $e_k=e'_k=1$ to obtain

$$\frac{(1+o(1))|\mathcal{I}'_m|}{\log x}\sideset{}{^*}\sum_{\substack{d_1,\dots,d_{k-1}\\ d'_1,\dots,d'_{k-1}}}\sideset{}{^*}\sum_{\substack{e_1,\dots,e_{k-1}\\ e'_1,\dots,e'_{k-1}}}\frac{1}{\varphi(P_W)\varphi([\mathbf{d},\mathbf{d}',\mathbf{e},\mathbf{e}',\mathbf{a}])}\sum_{j=1}^{J}\sum_{j'=1}^{J}\Bigg(F_{k,j}(0)F_{k,j'}(0)G(0)^2$$

$$*\prod_{l=1}^{k-1}\mu(d_l)\mu(d'_l)\mu(e_l)\mu(e'_l)F_{l,j}\Big(\frac{\log d_l}{\log x}\Big)F_{l,j'}\Big(\frac{\log d'_l}{\log x}\Big)G\Big(\frac{\log e_l}{\log x}\Big)G\Big(\frac{\log e'_l}{\log x}\Big)\Bigg), \tag{5.59}$$

where $\sum^*$ denotes the same divisibility constraints from (5.54) but with conditions on d_k,d'_k,e_k,e'_k removed. In a similar way to Lemma 5.10, we utilise the Fourier transforms from (5.26). We then write this term as

$$\frac{(1+o(1))|\mathcal{I}'_m|F_{k,j}(0)F_{k,j'}(0)G(0)^2}{\varphi(P_W)\log x}\sum_{j=1}^{J}\sum_{j'=1}^{J}\int_{\mathbb{R}}\cdots\int_{\mathbb{R}}\Bigg(\sideset{}{^*}\sum_{\substack{d_1,\dots,d_{k-1}\\ d'_1,\dots,d'_{k-1}}}\sideset{}{^*}\sum_{\substack{e_1,\dots,e_{k-1}\\ e'_1,\dots,e'_{k-1}}}\frac{1}{\varphi([\mathbf{d},\mathbf{d}',\mathbf{e},\mathbf{e}',\mathbf{a}])}$$

$$\times\prod_{l=1}^{k-1}\frac{\mu(d_l)\mu(d'_l)\mu(e_l)\mu(e'_l)}{d_l^{\frac{1+i\xi_l}{\log x}}d_l'^{\frac{1+i\xi'_l}{\log x}}e_l^{\frac{1+i\tau_l}{\log y}}e_l'^{\frac{1+i\tau'_l}{\log y}}}\Bigg)\Big(\prod_{l=1}^{k-1}f_{l,j}(\xi_l)f_{l,j'}(\xi'_l)g(\tau_l)g(\tau'_l)d\xi_ld\xi'_ld\tau_ld\tau'_l\Big). \tag{5.60}$$

This term is very similar to (5.27), and can be estimated in an analogous way. The φ function in the denominator has negligible effect compared to the lowest common multiple in (5.27), and so the main difference comes from the $\mathbf{a}$ term in the denominator. If $p|a_{i,j}$ for some i,j, we find that now $K_p\ll_k 1/p$. This can be seen as $\mathbf{a}$ contributes a factor of at least p to the denominator, where we previously had $K_p\ll_k 1$ seen in (5.33).

We recall that we introduced $\mathcal{G}_{m,q}$ in Lemma 5.10 to simplify a product over primes that originated from the restrictions over the summand, which are made explicit in the delta functions of (5.28). Our conditions from (5.54) are slightly different, and so the term analogous to $\mathcal{G}_{m,q}$ in Lemma 5.10 is now

$$\mathcal{G}^{(2)}_{m,p_0,h}=\prod_{p\le W}\Big(1-\frac{1}{p}\Big)^{-(2k-2)}\prod_{W<p\le y}\Big(1-\frac{g^{(2)}_{m,p_0,h}(p)}{p}\Big)\Big(1-\frac{1}{p}\Big)^{-(2k-2)}, \tag{5.61}$$

where
(5.62)

$$g^{(2)}_{m,p_0,h}(p) = \#\{1 \le n \le p : \prod_{i=1}^{k} \Big(p_0 + (h_i - h)n\Big)\Big(m(p_0 + (h_i - h)n) - 1\Big) \equiv 0 \pmod p\}.$$

Also (in an analogous way to how (5.42) was obtained) we have factors of $\log x$ and $\log y$ in the denominator coming from estimates of the ζ-function. Here they are to the power $k-1$, as we have taken out all dependence on d_k, d'_k, e_k, e'_k. Finally, the factors of the form $F_{k,j}(0), G(0)$ only depend on k, and so are $\ll_k 1$. Therefore we find that for $a_{i,j}$ not all equal to 1, (5.60) contributes

$$\ll_k \frac{\mathcal{G}^{(2)}_{m,p_0,h_k}|\mathcal{I}'_m|}{\varphi(P_W)(\log x)^k(\log y)^{k-1}} \prod_{\substack{p|a_{i,j}\\ \text{for some } i,j}} \frac{O_k(1)}{p},$$

and so the sum over $a_{i,j}$'s of (5.52) from cases where not all $a_{i,j} = 1$ is

$$\begin{aligned}
&\ll \frac{\mathcal{G}^{(2)}_{m,p_0,h_k}|\mathcal{I}'_m|}{\varphi(P_W)(\log x)^k(\log y)^{k-1}} \sum_{\substack{a_{1,2},\dots,a_{k,k-1}|P_{z_0}/P_W\\ (a_{1,2},\dots,a_{k,k-1})\neq(1,\dots,1)}} \frac{(-2k)^{\omega([\mathbf{a}])}}{[\mathbf{a}]} \prod_{\substack{p|a_{i,j}\\ \text{for some } i,j}} \frac{O_k(1)}{p}\\
&\ll \frac{\mathcal{G}^{(2)}_{m,p_0,h_k}|\mathcal{I}'_m|}{\varphi(P_W)(\log x)^k(\log y)^{k-1}} \Bigg(\Big(\sum_{a_{1,2},\dots,a_{k,k-1}\ge 1} \frac{O_k(1)^{\omega([\mathbf{a}])}}{[\mathbf{a}]^2} \prod_{\substack{i,j\ge 1\\ i\neq j}} \mathbf{1}(a_{i,j}|P_{z_0}/P_W)\Big) - 1\Bigg).
\end{aligned}$$

We recognise the interior of this sum as a multiplicative function from $\mathbb{N}^{k(k-1)} \to \mathbb{C}$, hence we use the generalisation of the Euler product to find that the above is equal to

$$\frac{\mathcal{G}^{(2)}_{m,p_0,h_k}|\mathcal{I}'_m|}{\varphi(P_W)(\log x)^k(\log y)^{k-1}} \Bigg(\prod_{W<p\le z_0} \Big(1 + \frac{O_k(1)}{p^2}\Big) - 1\Bigg) = o_k\Big(\frac{\mathcal{G}^{(2)}_{m,p_0,h_k}|\mathcal{I}'_m|}{\varphi(P_W)(\log x)^k(\log y)^{k-1}}\Big).$$

Now when all $a_{i,j} = 1$ in the inner sum of (5.52), this case corresponds almost identically to Lemma 5.10, as we have $[\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}', \mathbf{a}] = [\mathbf{d}, \mathbf{d}', \mathbf{e}, \mathbf{e}']$. Slight differences that alter our results are that the sum is now over $k-1$ elements in each variable (due to the use of $d_k = e_k = d'_k = e'_k = 1$), which gives different integrals to those of Lemma 5.10. Furthermore, we have a different constant ($\mathcal{G}^{(2)}_{m,p_0,h_k}$) coming from the different conditions of (5.54). All of this means that the inner sum over $a_{i,j}$'s of (5.52) is

$$\begin{aligned}
&\frac{(1+o_k(1))\mathcal{G}^{(2)}_{m,p_0,h_k}|\mathcal{I}'_m|}{\varphi(P_W)(\log x)^k(\log y)^{k-1}} G(0)^2\Big(\int_0^\infty G'(t)^2 dt\Big)^{k-1}\\
&\qquad\times \sum_{j=1}^{J}\sum_{j'=1}^{J} F_{k,j}(0)F_{k,j'}(0) \prod_{l=1}^{k-1} \Big(\int_0^\infty F'_{l,j}(t)F'_{l,j'}(t)dt\Big).
\end{aligned} \tag{5.63}$$

We recall the definitions

$$J_k^{(1)}(F) = \int \cdots \int_{t_1,...,t_{k-1}\geq 0} \Big(\int_{t_k \geq 0} F(t_1, ..., t_k) dt_k \Big)^2 dt_1 \dots dt_{k-1},$$

and

$$J_k^{(2)}(G) = G(0)^2 \Big(\int_0^\infty G'(t)^2 dt \Big)^{k-1}.$$

Note that by the definition of F and its support conditions, we have

$$J_k^{(1)}(F) = \int \cdots \int_{t_1,...,t_{k-1}\geq 0} \Big(\sum_{j=1}^{J} F_{k,j}(0) \prod_{l=1}^{k-1} F'_{l,j}(t_l) \Big)^2 dt_1 ... dt_{k-1},$$

which is precisely the term on the second line of (5.63). We subsequently find that (5.63) is

$$\frac{(1+o_k(1))\mathcal{G}^{(2)}_{m,p_0,h_k} |\mathcal{I}'_m| J_k^{(1)}(F) J_k^{(2)}(G)}{\varphi(P_W)(\log x)^k (\log y)^{k-1}}. \tag{5.64}$$

This finally allows us to calculate (5.52) explicitly, noting that F is symmetric and so $J_k^{(1)}$ and $J_k^{(2)}$ are independent of h. We therefore have that $\sum_{q \in \mathcal{I}'_m} \nu_{m,q}(p_0)$ is bounded below by

$$\Big(1 + o_k(1)\Big) \frac{m \mathcal{G}_m (\log y) |\mathcal{I}'_m| J_k^{(1)}(F) J_k^{(2)}(G)}{\varphi(P_W) U I_k^{(1)}(F) I_k^{(2)}(G)} \sum_{h \in \mathcal{H}} \mathcal{G}^{(2)}_{m,p_0,h} \sum_{\substack{w_0 \pmod{P_W} \\ (w_0, P_W)=1}} 1,$$

where all aforementioned errors have been engulfed into the $o_k(1)$ term. By calculating the inner-most sum and bringing $\mathcal{G}_m$ inside the sum over $h \in \mathcal{H}$, we have

$$\sum_{q \in \mathcal{I}'_m} \nu_{m,q}(p_0) \geq \Big(1 + o_k(1)\Big) \frac{m (\log y) |\mathcal{I}'_m| J_k^{(1)}(F) J_k^{(2)}(G)}{U I_k^{(1)}(F) I_k^{(2)}(G)} \sum_{h \in \mathcal{H}} \mathcal{G}^{(2)}_{m,p_0,h} \mathcal{G}_m. \tag{5.65}$$

This is very similar to Lemma 5.11. We just need to calculate $\sum_{h \in \mathcal{H}} \mathcal{G}^{(2)}_{m,p_0,h} \mathcal{G}_m$. By definitions (5.50) and (5.61), we have

$$\begin{aligned} \mathcal{G}^{(2)}_{m,p_0,h} \mathcal{G}_m =& 2^{-(2k-1)} \prod_{p \leq W} \Big(1 - \frac{1}{p}\Big)^{-(2k-2)} \prod_{2 < p \leq W} \frac{\Big(1 - \frac{1}{p}\Big)^{2k}}{\Big(1 - \frac{2}{p}\Big)} \prod_{\substack{p \mid m \\ p > 2}} \frac{p-2}{p-1} \prod_{W < p \leq y} \frac{\Big(1 - \frac{g^{(2)}_{m,p_0,h}(p)}{p}\Big)}{\Big(1 - \frac{1}{p}\Big)^{-(2k-2)}} \\ =& 2^{-1} \prod_{2 < p < W} \frac{\Big(1 - \frac{1}{p}\Big)^2}{\Big(1 - \frac{2}{p}\Big)} \prod_{\substack{p \mid m \\ p > 2}} \frac{p-2}{p-1} \prod_{W < p \leq y} \frac{\Big(1 - \frac{g^{(2)}_{m,p_0,h}(p)}{p}\Big)}{\Big(1 - \frac{1}{p}\Big)^{-(2k-2)}}, \end{aligned}$$

where we note that this first product converges. Furthermore, in the definition of $g^{(2)}_{m,p_0,h}(p)$ from (5.62), we are counting solutions to a product of $2k-2$ degree 1 polynomials in n. Therefore, we find $g^{(2)}_{m,p_0,h}(p) \leq 2k-2$ for all primes p, and summing over $h \in \mathcal{H}$, we have

$$\begin{aligned}\sum_{h\in\mathcal{H}} \mathcal{G}^{(2)}_{m,p_0,h}\mathcal{G}_m \gg& \sum_{h\in\mathcal{H}}\prod_{\substack{p|m\\p>2}}\frac{p-2}{p-1}\prod_{W<p\leq y}\frac{\left(1-\frac{2k-2}{p}\right)}{\left(1-\frac{1}{p}\right)^{-(2k-2)}}\\ \gg& k(1+o_k(1))\prod_{\substack{p|m\\p>2}}\frac{p-2}{p-1} \gg \frac{Uk(1+o_k(1))}{|\mathcal{R}_m| m\log x\log y},\end{aligned} \tag{5.66}$$

where we have used (5.05) from Lemma 5.5 in the last step, noting that the first product term in the estimate (5.05) is constant. Using this estimate (5.66) in (5.65) gives

$$\sum_{q\in\mathcal{I}'_m}\nu_{m,q}(p_0) \gg (1+o_k(1))\frac{|\mathcal{I}'_m|}{(\log x)|\mathcal{R}_m|}\times\frac{kJ^{(1)}_k(F)J^{(2)}_k(G)}{I^{(1)}_k(F)I^{(2)}_k(G)},$$

as required. □

This result brings us very close to the supposition of Lemma 5.9. The following Lemma will provide an estimate for the second term here, which will be enough to give the desired result.

Lemma 5.12. *There exists smooth functions* $F, G : [0,\infty) \to \mathbb{R}$ *that satisfy* (5.14) *and* (5.15) *such that*

$$\frac{kJ^{(1)}_k(F)J^{(2)}_k(G)}{I^{(1)}_k(F)I^{(2)}_k(G)} \gg \log k.$$

Proof. First we note that with support conditions, the term $kJ^{(1)}_k(F)/I^{(1)}_k(F)$ is almost identical to the term appearing in (4.37). With this motivation, we make a straightforward choice for G. We note that

$$\frac{J^{(2)}_k(G)}{I^{(2)}_k(G)} = \frac{G(0)}{\int_0^\infty G'(t)^2 dt}.$$

We take the smooth function

$$G(t) = \mathbf{1}_{[0,1]}(t)\cdot(t-1)^2,$$

which has the desired support conditions from (5.14). We have

$$\frac{J^{(2)}_k(G)}{I^{(2)}_k(G)} = \frac{3}{4}. \tag{5.67}$$

Now we recall from the proof of Lemma 4.15 that we can find a Riemann integrable function g such that

$$\tilde{F}(t_1,...,t_k) = \mathbf{1}_{\mathcal{R}_k}(t_1,...,t_k)g(kt_1)...g(kt_k), \tag{5.68}$$

that satisfies $kJ_k^{(1)}(\tilde{F})/I_k^{(1)}(\tilde{F}) \geq \log k$ for sufficiently large k. Note that the definition of J_k's differs here slightly, but they are in fact equal for F symmetric. Specifically, we took

$$g(t) = c\frac{\mathbf{1}_{[0,\xi k]}}{1 + At},$$

for $c = \left(\frac{1+A\xi k}{\xi k}\right)^2$, $A = \log k$ and $\xi = 1/(\log k)^3$. To satisfy support conditions, we need our F to be supported on $\{(t_1, ..., t_k) : \sum_{i=1}^k t_i \leq 1/10\}$, whereas $\tilde{F}$ is supported on $\mathcal{R}_k = \{(t_1, ..., t_k) : \sum_{i=1}^k t_i \leq 1\}$. Subsequently we choose $F_{l,j}$ such that $F(t_1, ..., t_k)$ is a smooth approximation to $\tilde{F}(10t_1, ..., 10t_k)$ satisfying

$$\begin{aligned} \frac{kJ_k^{(1)}(F)}{I_k^{(1)}(F)} &\geq \Big(1 - 10\epsilon\Big)\frac{kJ_k^{(1)}(\tilde{F}(10t_1, ..., 10t_k))}{I_k^{(1)}(\tilde{F}(10t_1, ..., 10t_k))} \\ &\geq \Big(\frac{1}{10} - \epsilon\Big)\frac{kJ_k^{(1)}(\tilde{F})}{I_k^{(1)}(\tilde{F})} \geq \Big(\frac{1}{10} - \epsilon\Big)\log k, \end{aligned} \tag{5.69}$$

where we have a factor of $1/10$ due to the fact $J_k^{(1)}$ and $I_k^{(1)}$ integrate over $k+1$ and k variables respectively. We note that such a choice is possible due to L^2 and L^1-dense results regarding linear combinations of products of smooth, compactly supported non-negative functions. With F, G as described here, we combine (5.67) and (5.69) to obtain the desired result. □

Combining the above result with Lemma 5.11 we find that for any $m < U/z(\log x)^2$ even and $p_0 \in \mathcal{R}_m$ with $h_k x < p_0 < U/m - h_k x$, we have

$$\sum_{q \in \mathcal{I}'_m} \nu_{m,q}(p_0) \gg (1 + o_k(1))\frac{|\mathcal{I}'_m| \log k}{(\log x)|\mathcal{R}_m|},$$

Recall that we chose $\mathcal{I}'_m$ of length greater than or equal to $\delta|\mathcal{R}_m| \log x$, therefore we have

$$\sum_{q \in \mathcal{I}'_m} \nu_{m,q}(p_0) \gg (1 + o_k(1))\delta \log k,$$

and taking k fixed sufficiently large, depending on ϵ and δ gives

$$\sum_{q \in \mathcal{I}'_m} \nu_{m,q}(p_0) \leq -\log \epsilon/2,$$

for any $m < U/z(\log x)^2$ even and $p_0 \in \mathcal{R}_m$ with $h_k x < p_0 < U/m - h_k x$. We conclude the proof of the suppositions of Lemma 5.9 by noting that for fixed k, the exceptional set $\mathcal{C}_m = \{p \in \mathcal{R}_m : p \notin (h_k x, U/m - h_k x)\}$ has zero natural density in $\mathcal{R}_m$ (that is, $\mathcal{C}_m = o(|\mathcal{R}_m|)$). This is a direct implication of (5.06), Lemma 5.5, concluding our proof.

6 Appendix

6.1 Estimates

Lemma 6.1. *Take $\lambda_{d_1,\ldots,d_k}$ as described in Section 4. With*

$$u_{r_1,\ldots,r_k} = \Big(\prod_{i=1}^{k} \mu(r_i)\varphi(r_i)\Big) \sum_{\substack{d_1,\ldots,d_k \\ r_i | d_i \forall i}} \frac{\lambda_{d_1,\ldots,d_k}}{\prod_{i=1}^{k} d_i},$$

$u_{\max} := \sup_{r_1,\ldots,r_k} |u_{r_1,\ldots,r_k}|$, and $\lambda_{\max}$ defined analogously, we have the estimate

$$\lambda_{\max} \ll u_{\max}(\log R)^k.$$

Proof. Similarly to (4.11), we can show that on the support of $\lambda_{d_1,\ldots,d_k}$ we have

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

where $u_{r_1,\ldots,r_k}$ is supported on tuples with $(\prod_{i=1}^{k} r_i, W) = 1$, $\prod_{i=1}^{k} r_i$ square-free and $\prod_{i=1}^{k} r_i < R$. Now

$$\lambda_{\max} \leq \sup_{\substack{d_1,\ldots,d_k \\ \prod_{i=1}^{k} d_i \text{ square-free}}} u_{\max}\Big(\prod_{i=1}^{k} d_i\Big) \sum_{\substack{r_1,\ldots,r_k \\ d_i | r_i \ \forall i \\ \prod_{i=1}^{k} r_i \leq R \\ \prod_{i=1}^{k} r_i \text{ square-free}}} \prod_{i=1}^{k} \frac{\mu(r_i)^2}{\varphi(r_i)}.$$

We use the change of variables $d_i t_i = r_i$, noting that $\varphi(d_i t_i) \geq \varphi(d_i)\varphi(t_i)$, and the square-free conditions on $\prod_{i=1}^{k} r_i$ induce similar conditions on $\prod_{i=1}^{k} t_i$ and coprime conditions with $\prod_{i=1}^{k} d_i$. We also use the fact that on square-free d we have $d/\varphi(d) = \sum_{e|d} 1/\varphi(d)$, stemming from the relation $id = 1 * \varphi$. This gives

$$\begin{aligned}
\lambda_{\max} &\leq \sup_{\substack{d_1,\ldots,d_k \\ \prod_{i=1}^{k} d_i \text{ square-free}}} u_{\max}\Big(\prod_{i=1}^{k} \frac{d_i}{\varphi(d_i)}\Big) \sum_{\substack{t_1,\ldots,t_k \\ \prod_{i=1}^{k} t_i \leq R/\prod_{i=1}^{k} d_i \\ \prod_{i=1}^{k} t_i d_i \text{ square-free}}} \prod_{i=1}^{k} \frac{\mu(t_i d_i)^2}{\varphi(t_i)} \\
&\leq u_{\max} \sup_{\substack{d_1,\ldots,d_k \\ \prod_{i=1}^{k} d_i \text{ square-free}}} \Big(\prod_{i=1}^{k} \sum_{e_i | d_i} \frac{1}{\varphi(e_i)}\Big) \sum_{\substack{t_1,\ldots,t_k \\ \prod_{i=1}^{k} t_i \leq R/\prod_{i=1}^{k} d_i \\ (\prod_{i=1}^{k} t_i, \prod_{i=1}^{k} d_i) = 1 \\ \prod_{i=1}^{k} t_i \text{ square-free}}} \prod_{i=1}^{k} \frac{\mu(t_i)^2}{\varphi(t_i)} \\
&\leq u_{\max} \sup_{\substack{d_1,\ldots,d_k \\ \prod_{i=1}^{k} d_i \text{ square-free}}} \Big(\prod_{i=1}^{k} \sum_{e_i | d_i} \frac{1}{\varphi(e_i)}\Big) \sum_{\substack{t_1,\ldots,t_k \\ \prod_{i=1}^{k} t_i \leq R/\prod_{i=1}^{k} d_i \\ (\prod_{i=1}^{k} t_i, \prod_{i=1}^{k} d_i) \\ \prod_{i=1}^{k} t_i \text{ square-free}}} \frac{\mu(\prod_{i=1}^{k} t_i)^2}{\varphi(\prod_{i=1}^{k} t_i)}.
\end{aligned}$$

With $\prod_{i=1}^k d_i$ square-free we have $\prod_{i=1}^k e_i$ square-free, allowing us to write

$$\lambda_{\max} \leq u_{\max} \sup_{d_1,\ldots,d_k} \Big(\sum_{e | \prod_{i=1}^k d_i} \frac{\mu(e)^2}{\varphi(e)} \Big) \sum_{\substack{t \leq R / \prod_{i=1}^k d_i \\ (t, \prod_{i=1}^k d_i) = 1}} \frac{\mu(t)^2 \tau_k(t)}{\varphi(t)}$$

$$\leq u_{\max} \sup_{d_1,\ldots,d_k} \sum_{t \leq R} \frac{\mu(t)^2 \tau_k(t)}{\varphi(t)} \ll u_{\max} (\log R)^k.$$

as required. To arrive at the last line we have used the coprimality conditions to merge these sums, followed by (6.16) to arrive at the final estimate. □

Lemma 6.2. *For $A, \epsilon > 0$, let γ be a multiplicative function with*

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

and

$$\gamma(p) = 1 + O(p^{-\epsilon}),$$

then for f a totally multiplicative function defined by $f(p) = \frac{\gamma(p)}{p - \gamma(p)}$, we have

$$\sum_{n \leq x} f(n) \mu(n)^2 = c_\gamma \log(x) + O_\gamma(1).$$

Where O_γ refers to dependence on constants A, ϵ and

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

Proof. For simplicity, denote $M_{\mu^2 f}(x) := \sum_{n \leq x} f(n) \mu(n)^2$. We first recognise that $c_\gamma = L(1, \mu^2 \cdot f \cdot id * \mu)$ (absolute convergence shall be shown later). Motivated by this, we write

$$\begin{aligned} M_{\mu^2 f}(x) &= \sum_{n \leq x} \frac{1}{n} f(n) \mu(n)^2 n \\ &= \sum_{n \leq x} \frac{1}{n} \sum_{d | n} (\mu^2 \cdot f \cdot id * \mu)(d) \\ &= \sum_{d \leq x} \frac{r(d)}{d} \sum_{m \leq x/d} \frac{1}{m} \\ &= \log(x) \sum_{d \leq x} \frac{r(d)}{d} - \sum_{d \leq x} \frac{\log(d) r(d)}{d} + O\Big(\sum_{d \leq x} \Big| \frac{r(d)}{d} \Big| \Big), \end{aligned} \tag{6.01}$$

where we defined $r(d) = (\mu^2 \cdot f \cdot id * \mu)(d)$.

r is multiplicative, and on prime powers we have

$$r(p^\alpha) = \begin{cases} 1 & \alpha = 0 \\ pf(p) - 1 & \alpha = 1 \\ -pf(p) & \alpha = 2 \\ 0 & \alpha > 2 \end{cases},$$

and our assumptions on γ imply that $|r(p)| = O(p^{-\epsilon})$ and $|r(p^2)| = O(1)$. Without loss of generality we assume that $\epsilon < 1/2$ for the remainder of the proof. For $\sigma > 1/2$ we consider the Euler product

$$\begin{aligned}
\sum_{d \geq 1} \Big| \frac{r(d)}{d^\sigma} \Big| &= \prod_p \Big(1 + \Big| \frac{r(p)}{p^\sigma} \Big| + \Big| \frac{r(p^2)}{p^{2\sigma}} \Big| \Big) \\
&\leq \prod_p \Big(1 + K \Big(\frac{p^{-\epsilon}}{p^\sigma} + \frac{1}{p^{2\sigma}} \Big) \Big) \\
&\leq \prod_p \Big(1 + \frac{\tilde{K}}{p^{\sigma+\epsilon}} \Big) \\
&\leq \prod_p \Big(1 + \frac{1}{p^{\sigma+\epsilon}} \Big)^{\tilde{K}} \text{ by Bernoulli's inequality} \\
&\leq \prod_p \Big(1 + \frac{1}{p^{\sigma+\epsilon}} + \frac{1}{p^{2(\sigma+\epsilon)}} + \frac{1}{p^{3(\sigma+\epsilon)}} + \ldots \Big)^{\tilde{K}} \\
&\leq \zeta(\sigma+\epsilon)^{\tilde{K}},
\end{aligned}$$

and so we have absolute convergence for $\sigma > 1 - \epsilon$ (with $\epsilon < 1/2$). Now due to absolute convergence of $L(1, r)$ and $L'(1, r)$, (6.01) becomes

$$M_{\mu^2 f}(x) = \log(x) \sum_{d \leq x} \frac{r(d)}{d} + O_\gamma(1). \tag{6.02}$$

Here we will utilise Rankin's trick, giving

$$\begin{aligned}
\sum_{d \leq x} \frac{r(d)}{d} &= L(1, r) - \sum_{x < d} \frac{r(d)}{d} \\
&= c_\gamma + O\Big(\sum_{x<d} \Big| \frac{r(d)}{d^{1-\epsilon/2} x^{\epsilon/2}} \Big| \Big) \\
&= c_\gamma + O_\gamma \Big(x^{-\epsilon/2} L(1 - \epsilon/2, r) \Big) \\
&= c_\gamma + O_\gamma \Big(\frac{1}{\log x} \Big),
\end{aligned}$$

and so (6.02) implies the required result

$$M_{\mu^2 f}(x) = c_\gamma \log(x) + O_\gamma(1).$$

□

Corollary 6.3. *For g a completely multiplicative function defined on primes by $g(p) = p-2$, we have*

$$\sum_{\substack{n\le x\\(n,W)=1}} \frac{\mu(n)^2}{g(n)} \ll \log x, \tag{6.03}$$

by taking $\gamma(p) = \frac{p}{p-1}\mathbf{1}_{p\nmid W}(p)$. Also

$$\sum_{n\le x} \frac{\mu(n)^2}{\varphi(n)} = \log x + O(1), \tag{6.04}$$

by taking $\gamma(p) = 1$. Finally we have

$$\sum_{\substack{n\le x\\(n,W)=1}} \frac{\mu(n)^2}{\varphi(n)} = \frac{\varphi(W)\log x}{W} + O(1), \tag{6.05}$$

by taking $\gamma(p) = \mathbf{1}_{p\nmid W}(p)$.

Lemma 6.4. *Assume the conditions on f, γ from Lemma 6.2. Then*

$$\sum_{n\ge 1} \mu(n)^2 f(n)^2 < \infty,$$

and

$$\sum_{n\ge x} \mu(n)^2 f(n)^2 \ll \frac{(\log x)^C}{x},$$

for some constant $C > 0$.

Proof. First we shall show absolute convergence of the Dirichlet series $L(s, \mu^2 f^2)$,

$$\begin{aligned}
\sum_{n\ge 1}\left|\frac{\mu(n)^2 f(n)^2}{n^s}\right| &= \prod_p \Big(1 + \Big|\frac{\gamma(p)^2}{(p-\gamma(p))^2 p^s}\Big|\Big)\\
&\le \prod_p \Big(1 + K\frac{1 + Cp^{-\epsilon}}{|p-1-Cp^{-\epsilon}|^2 p^s}\Big)\\
&\le \prod_p \Big(1 + \frac{\tilde{K}}{p^{2+s}}\Big)\\
&\le \prod_p \Big(1 + \frac{1}{p^{2+s}}\Big)^{\tilde{K}} \text{ by Bernoilli's inequality}\\
&\le \zeta(2+s)^K, \qquad (6.06)
\end{aligned}$$

where $K, \tilde{K}, C$ are positive constants taken sufficiently large. Therefore we have absolute convergence of $L(s, \mu^2 f^2)$ for any $s > -1$. Importantly, we have convergence for $s = 0$, giving the first desired result.

We now apply Rankin's trick. Note that for any $s \in \mathbb{R}_{>}0$, we have $\mathbf{1}_{[1,\infty]}(n) \leq n^s$ and therefore $\mathbf{1}_{[x,\infty]}(n) \leq \left(\frac{n}{x}\right)^s$. Therefore

$$\sum_{n \geq x} \mu(n)^2 f(n)^2 \leq \frac{1}{x^s} \sum_{n \geq 1} \mu(n)^2 f(n)^2 n^s.$$

We take $s = 1 - 1/\log x$, which gives

$$\begin{aligned}\sum_{n \geq x} \mu(n)^2 f(n)^2 &\leq \frac{1}{x^{1-1/\log x}} L\Big(-1 + \frac{1}{\log x}, \mu^2 f^2\Big) \\ &\leq \frac{e}{x} L\Big(-1 + \frac{1}{\log x}, \mu^2 f^2\Big) \ll \frac{(\log x)^C}{x},\end{aligned}$$

for some constant $C > 0$, giving the required result. In the last step we have used (6.06) and the bound $\zeta(1+\sigma) \ll 1/\sigma$. □

Corollary 6.5. *The following results are consequences of Lemma 6.4. For g a completely multiplicative function defined on primes by $g(p) = p - 2$, we have*

$$\sum_{\substack{n \geq 1 \\ (n,W)=1}} \frac{\mu(n)^2}{g(n)^2} < \infty, \tag{6.07}$$

$$\sum_{\substack{n \geq x \\ (n,W)=1}} \frac{\mu(n)^2}{g(n)^2} \ll \frac{(\log x)^C}{x}, \tag{6.08}$$

by taking $\gamma(p) = p/(p-1)\mathbf{1}_{p \nmid W}(p)$. Also,

$$\sum_{n \geq 1} \frac{\mu(n)^2}{\varphi(n)^2} < \infty, \tag{6.09}$$

$$\sum_{n \geq x} \frac{\mu(n)^2}{\varphi(n)^2} \ll \frac{(\log x)^C}{x}, \tag{6.10}$$

by taking $\gamma(p) = 1$. Furthermore,

$$\sum_{\substack{n \geq 1 \\ m|n}} \frac{\mu(n)^2}{\varphi(n)^2} \leq \frac{1}{\varphi(m)} \sum_{n \geq 1} \frac{\mu(n)^2}{\varphi(n)^2} \ll \frac{1}{\varphi(m)}, \tag{6.11}$$

$$\sum_{\substack{n \geq x \\ m|n}} \frac{\mu(n)^2}{\varphi(n)^2} \leq \frac{1}{\varphi(m)} \sum_{n \geq x} \frac{\mu(n)^2}{\varphi(n)^2} \ll \frac{(\log x)^C}{\varphi(m)x}, \tag{6.12}$$

where we have used (6.09) and (6.10) here. These hold for fixed implied constants as m varies.

We proceed to state a number of different estimates, followed immediately by proofs.

$$\frac{1}{\varphi(n)} \ll \frac{\log n}{n}, \tag{6.13}$$

$$\sum_{n\le x} \frac{1}{\varphi(n)} \ll (\log x)^2. \tag{6.14}$$

Proof. From [14], Theorem 3.4 (Mertens' Third Estimate), we have

$$\prod_{p\le x}\Big(1-\frac{1}{p}\Big) = \frac{e^{-\gamma}}{\log x} + O\Big(\frac{1}{(\log x)^2}\Big).$$

The multiplicative relation on φ gives

$$\begin{aligned}\frac{\varphi(n)}{n} &= \prod_{p|n}\Big(1-\frac{1}{p}\Big) \ge \prod_{p\le n}\Big(1-\frac{1}{p}\Big)\\ &\ge \frac{e^{-\gamma}}{\log n} + O\Big(\frac{1}{(\log n)^2}\Big),\end{aligned}$$

therefore we have

$$\frac{1}{\varphi(n)} \le \frac{(\log n)^2}{ne^{-\gamma}\log n + O(n)} \ll \frac{\log n}{n},$$

as required. Furthermore, $\sum_{n\le x} 1/\varphi(n) \ll \log x \sum_{n\le x} 1/n \ll (\log x)^2$, as required. □

$$\sum_{n\le x} \tau_k(n) \ll x(\log x)^k. \tag{6.15}$$

Proof. We proceed by induction on k. For $k = 1$ we have $\tau_1(n) = 1$ hence $\sum_{n\le x}\tau_1(n) \ll x$. For the inductive step, we have

$$\begin{aligned}\sum_{n\le x}\tau_{k+1}(n) &= \sum_{n\le x}\sum_{d|n}\tau_k(n/d) = \sum_{d\le x}\sum_{m\le x/d}\tau_k(m)\\ &\ll \sum_{d\le x}\frac{x}{d}\Big(\log\frac{x}{d}\Big)^k \text{ using inductive hypothesis}\\ &\ll x(\log x)^k\sum_{d\le x}\frac{1}{d} \ll x(\log x)^{k+1}.\end{aligned}$$

□

$$\sum_{n\le x}\frac{\mu(n)^2\tau_k(n)}{\varphi(n)} \ll (\log x)^k. \tag{6.16}$$

Proof. We proceed again by induction on k. For $k = 1$ we have $\sum_{n\le x} \frac{\mu(n)^2\tau_1(n)}{\varphi(n)} \ll \sum_{n\le x} \frac{\mu(n)^2}{\varphi(n)} \ll \log x$ by (6.04). For the inductive step, we have

$$\sum_{n\le x} \frac{\mu(n)^2\tau_{k+1}(n)}{\varphi(n)} = \sum_{n\le x} \frac{\mu(n)^2}{\varphi(n)} \sum_{d|n} \tau_k(d) = \sum_{d\le x} \sum_{m\le x/d} \frac{\mu(md)^2}{\varphi(md)} \tau_k(d).$$

Using $\mu(md)^2 \le \mu(m)^2\mu(d)^2$ and $\varphi(md) \ge \varphi(m)\varphi(d)$, in addition to (6.04), we have

$$\begin{aligned}\sum_{n\le x} \frac{\mu(n)^2\tau_{k+1}(n)}{\varphi(n)} &= \sum_{n\le x} \frac{\mu(n)^2}{\varphi(n)} \sum_{d|n} \tau_k(d) \le \sum_{d\le x} \frac{\mu(d)^2\tau_k(d)}{\varphi(d)} \sum_{m\le x/d} \frac{\mu(m)^2}{\varphi(m)} \\ &\ll \sum_{d\le x} \frac{\mu(d)^2\tau_k(d)}{\varphi(d)} \log(x) \ll (\log x)^{k+1} \text{ using inductive hypothesis.}\end{aligned}$$

□

(6.17) $$\sum_{n\le x} \frac{\mu(n)^2\tau_{3k}^2(n)}{\varphi(n)} \ll (\log x)^{9k^2},$$

Proof. this follows from the fact that $\tau_{3k}(p)^2 = 9k^2 = \tau_{9k^2}(p)$, hence $\mu(n)^2\tau_{3k}(n)^2 = \mu(n)^2\tau_{9k^2}(n)$. Using this relation followed by (6.16) this gives the desired results. □

(6.18) $$\sum_{n\le x} \frac{\tau_k(n)^2}{n} \ll x(\log x)^{k+1},$$

Proof. We have

$$\begin{aligned}\sum_{n\le x} \frac{\tau_k(n)^2}{n} &= \sum_{n\le x} \frac{(1*\tau_{k-1})^2(n)}{n} = \sum_{n\le x} \frac{1}{n}\Big(\sum_{d|n} \tau_{k-1}(d)\Big)^2 \\ &\le \sum_{n\le x} \frac{1}{n}\Big(\sum_{d\le n} \tau_{k-1}(d)\Big)\Big(\sum_{d\le n} \mathbf{1}_{d|n}(d)\Big) \\ &\ll \sum_{n\le x} (\log n)^{k-1}\tau_2(n) \ll (\log x)^{k-1} \sum_{n\le x} \tau_2(n) \\ &\ll x(\log x)^{k+1},\end{aligned}$$

where we have used Cauchy-Schwarz to obtain the second line and (6.15) to obtain the third and fifth lines. □

References

[1] DHJ Polymath. “Variants of the Selberg sieve, and bounded intervals containing many primes”. In: *Research in the Mathematical sciences* 1.1 (2014), pp. 1–83.

[2] Harald Cramér. “On the order of magnitude of the difference between consecutive prime numbers”. eng. In: *Acta Arithmetica* 2.1 (1936), pp. 23–46. URL: http://eudml.org/doc/205441.

[3] Andrew Granville. “Harald Cramér and the distribution of prime numbers”. In: *Scandinavian Actuarial Journal* 1995.1 (1995), pp. 12–28. DOI: 10.1080/03461238.1995.10413946. eprint: https://doi.org/10.1080/03461238.1995.10413946. URL: https://doi.org/10.1080/03461238.1995.10413946.

[4] Daniel A Goldston, János Pintz, and Cem Y Yildirim. “Primes in tuples I”. In: *Annals of Mathematics* (2009), pp. 819–862.

[5] Yitang Zhang. “Bounded gaps between primes”. In: *Annals of Mathematics* (2014), pp. 1121–1174.

[6] James Maynard. “Small gaps between primes”. In: *Annals of mathematics* (2015), pp. 383–413.

[7] R. A. Rankin. “The Difference between Consecutive Prime Numbers”. In: *Journal of the London Mathematical Society* s1-13.4 (1938), pp. 242–247. DOI: https://doi.org/10.1112/jlms/s1-13.4.242. eprint: https://londmathsoc.onlinelibrary.wiley.com/doi/pdf/10.1112/jlms/s1-13.4.242. URL: https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/jlms/s1-13.4.242.

[8] Kevin Ford et al. “Large gaps between consecutive prime numbers”. In: *annals of Mathematics* (2016), pp. 935–974.

[9] James Maynard. “Large gaps between primes”. In: *annals of Mathematics* (2016), pp. 915–933.

[10] James Maynard et al. “Long gaps between primes”. In: *Journal of the American Mathematical Society* (2017).

[11] Jingren Chen. “ON THE REPRESENTATION OF A LARGER EVEN INTEGER AS THE SUM OF A PRIME AND THE PRODUCT OF AT MOST TWO PRIMES”. In: 1973.

[12] Daniel Alan Goldston et al. “Small gaps between primes exist”. In: *Proceedings of the Japan Academy, Series A, Mathematical Sciences* 82.4 (2006), pp. 61–65.

[13] Graduate Mathematics. *The Bombieri-Vinogradov theorem (1/6) Kannan Soundararajan (Stanford) [2015]*. 2015. URL: https://www.youtube.com/watch?v=3poB-ZIPolU&t=3330s&ab_channel=GraduateMathematics (visited on 02/27/2022).

[14] D. Koukoulopoulos. *The Distribution of Prime Numbers*. Graduate Studies in Mathematics. American Mathematical Society, 2019. ISBN: 9781470447540. URL: https://books.google.co.uk/books?id=V1rayAEACAAJ.

[15] H. Halberstam and Hans-Egon Richert. *Sieve methods / H. Halberstam and H. E. Richert*. English. Academic Press London ; New York, 1974, xiv, 364 p. ISBN: 0123182506.

[16] MIT. *Narrow admissible tuples*. 2015. URL: https://math.mit.edu/~primegaps/ (visited on 02/27/2022).

[17] Enrico Bombieri, John B Friedlander, and Henryk Iwaniec. "Primes in arithmetic progressions to large moduli". In: *Acta Mathematica* 156 (1986), pp. 203–251.

[18] John Friedlander and Henryk Iwaniec. "Asymptotic Sieve for Primes". In: *The Annals of Mathematics* 148.3 (Nov. 1998), p. 1041. ISSN: 0003-486X. DOI: 10.2307/121035. URL: http://dx.doi.org/10.2307/121035.

[19] John Friedlander and Henryk Iwaniec. "The Polynomial X 2 + Y 4 Captures Its Primes". In: *The Annals of Mathematics* 148.3 (Nov. 1998), p. 945. ISSN: 0003-486X. DOI: 10.2307/121034. URL: http://dx.doi.org/10.2307/121034.

[20] Pál Erdős. *On integers of the form* $2^k + p$ *and some related problems*. English. Summa Brasil. Math. 2, 113-123 (1950). 1950.

[21] Helmut Maier and Carl Pomerance. "Unusually Large Gaps Between Consecutive Primes". In: *Transactions of the American Mathematical Society* 322.1 (1990), pp. 201–237. ISSN: 00029947. URL: http://www.jstor.org/stable/2001529.

[22] de Ng Dick Bruijn. "On the number of positive integers \$$\leq x$\$$andfreeofprimefactors$\$ $>$ y\$". In: 1951.

[23] László Tóth. *Multiplicative Arithmetic Functions of Several Variables: A Survey*. 2014. arXiv: 1310.7053 [math.NT].