Small gaps between primes

Andreea Mocanu

March 15, 2015

Abstract

In this paper we study the work of James Maynard [10], in which he proves that

 $\lim\inf_{n\to\infty}(p_{n+m}-p_n)<\infty,$

thus establishing that there are infinitely many intervals of finite length that contain a fixed number of primes. We use Sieve theory, in particular Selberg's sieve, and we look at preliminary work in order to better understand this result. To conclude, we discuss the implications of this result.

Contents

1	Introduction	2
2	Sieve theory	3
	2.1 Sieve of Eratosthenes	4
	2.2 Selberg's sieve	8
	2.3 Admissible sets	14
	2.4 Level of distribution of primes	16
3	The Goldston–Pintz–Yıldırım theorem	18
4	The Maynard theorems	27
	4.1 Main proposition	29
	4.1.1 Selberg sieve manipulations	30
	4.1.2 Smooth choice of y	43
	4.2 Optimisation and consequences	51
	4.3 Main results	58
5	Conclusions and further directions	59

1 Introduction

The study of prime numbers has intrigued people since antiquity. It represents the basis of Analytic Number Theory. In 1896 the prime number theorem was proved independently by de la Vallée Poussin and Hadamard, describing the asymptotic behaviour of primes.

Their work was based on that of Riemann and, with the clever use of Complex Analysis (contour integration), they proved that

$$\lim_{x \to \infty} \frac{\pi(x)}{x/\log x} = 1.$$
 (1)

We write this as $\pi(x) \sim x/\log x$, where we define $\pi(x) := \sum_{p \le x} 1$, the number of primes less than or equal to x.

Equivalently, we can weight each prime with a weight $\log p$ to obtain the estimate

$$\vartheta(x) := \sum_{p \le x} \log p \sim x.$$

This implies that the average gap between consecutive primes less than or equal to a given n is approximately $\log n$, i.e. primes become less common as they become larger. People were concerned with investigating gaps between prime numbers more thoroughly, and since then many conjectures have arisen. We focus on small gaps between primes and when we say small we mean in comparison to the average gap. There are spectacular results in this area in the recent years.

In a paper printed in 2005 [7], Goldston *et al.* proved, among other things, that:

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

where p_n is the *n*-th prime, thus showing that there are infinitely many consecutive primes that have an arbitrarily small gap compared to the average gap. In their proof they use Sieve Theory, which is a compilation of techniques used to count a large set of integers which satisfy certain properties (usually involving prime numbers). Furthermore, under the Elliott– Halberstam conjecture regarding the level of distribution of primes, they prove that:

$$\liminf_{n \to \infty} (p_{n+1} - p_n) \le 16,$$

in other words there are infinitely many pairs of prime numbers that differ by 16 or less. In 2013, Yitang Zhang published a paper [16] in which he established the first finite bound on gaps between prime numbers, showing that

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

His method was a refinement of that of Goldston *et al.* and the result was a breakthrough as the proof is unconditional. The polymath project [12] reduced the gap to 4680, by further developing Zhang's techniques.

The aim of this project is to study the work of James Maynard, who, a few months after Zhang, used a generalisation of the methods used in [7] to prove that

$$\liminf_{n \to \infty} (p_{n+m} - p_n) < \infty.$$

In other words, there are infinitely many intervals of finite length that contain m + 1 or more primes.

He also improved Zhang's estimate considerably by obtaining

$$\liminf_{n \to \infty} (p_{n+1} - p_n) \le 600, \tag{2}$$

and, under the Elliott–Halberstam conjecture, that $\liminf_n (p_{n+1} - p_n) \le 12$ and $\liminf_n (p_{n+2} - p_n) \le 600$.

We will first discuss some sieving techniques which are useful in studying gaps between primes. Then we talk about admissible sets and the Elliott– Halberstam conjecture, as well as the level of distribution of primes, at which point we state the Bombieri–Vinogradov theorem. In Section 3 we discuss the methods used in the Goldston *et al.* paper, which will then help us discuss Maynard's techniques and results in Section 4.

2 Sieve theory

Sieve Theory is an important tool used in Number Theory, whose modern version emerged more or less 100 years ago. To this day, there are a handful of sieves that are used more often than others. Our focus will be on the sieve of Eratosthenes, as an introduction, and the Selberg sieve, which we use in our analysis of Maynard's work.

Let us begin with a formal definition of a sieve, as stated in [1]:

Definition 2.1. Let \mathcal{A} be a finite set of objects and let \mathcal{P} be an indexed set of primes such that to each $p \in \mathcal{P}$ we have associated a subset \mathcal{A}_p of \mathcal{A} . We define the set

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

The purpose of *sieve theory* is to estimate the size of this set from above and below.

In general, \mathcal{A} is a set of positive integers and \mathcal{A}_p is a subset of \mathcal{A} consisting of elements lying in specific congruence classes modulo p.

For example, if $\mathcal{A} = (1, N)$, $\mathcal{P} = \{p, p \leq \sqrt{N}\}$ and $\mathcal{A}_p = \{n \in \mathcal{A}, n \equiv 0 \pmod{p}$ and $n+2 \equiv 0 \pmod{p}\}$, then $\mathcal{S}(\mathcal{A}, \mathcal{P})$ is the set of all twin primes $\leq N$ that are not in \mathcal{P} .

We now introduce the sieve of Eratosthenes, which was written down in the form we present by A.M. Legendre in 1808, in the second edition of his book *Théorie des Nombres*.

2.1 Sieve of Eratosthenes

The sieve of Eratosthenes is a simple way of sifting out primes up to a certain upper bound, x.

Informally, we make a list of all the integers $2, 3, \ldots, \lfloor x \rfloor$, where by $\lfloor x \rfloor$ we denote the greatest integer less than or equal to x. We start by calling 2 a prime and crossing all of its multiples off the list. As 3 is uncrossed, we call it a prime, too, and proceed by crossing off all of its multiples. We then pick the next uncrossed number and repeat the algorithm until the next uncrossed number is greater than \sqrt{x} . All the numbers that are uncrossed by the end are prime.

Formally, we wish to study the number

$$\Phi(x, z) := \#\{n \le x : n \text{ is not divisible by any prime} < z\}, \qquad (3)$$

where x, z are positive real numbers.

Let

$$P_z := \prod_{p < z} p.$$

We can use the following result about the Möbius function, $\mu(\cdot)$,

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{otherwise.} \end{cases}$$
(4)

to write

$$\Phi(x,z) = \sum_{n \le x} \sum_{d \mid (n,P_z)} \mu(d) = \sum_{d \mid P(z)} \mu(d) \sum_{\substack{n \le x \\ d \mid n}} 1 = \sum_{d \mid P_z} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor$$
(5)
$$= x \sum_{d \mid P_z} \frac{\mu(d)}{d} + O(2^z) = x \prod_{p < z} \left(1 - \frac{1}{p} \right) + O(2^z),$$

where in the first line we used the fact that $\#\{n \le x : n \equiv 0 \mod d\} = \lfloor x/d \rfloor = x/d + O(1)$. We can relate this result to $\pi(x)$. We have

$$\pi(x) = (\pi(x) - \pi(z)) + \pi(z) \leq \Phi(x, z) + \pi(z) \leq \Phi(x, z) + z.$$
(6)

Since $1 - x \le e^{-x}$ holds for positive x, which can be shown by derivation, it implies that

$$\prod_{p < z} \left(1 - \frac{1}{p} \right) \le \exp\left(-\sum_{p < z} \frac{1}{p} \right).$$

In addition to this, we use the following result:

$$\sum_{p \le z} \frac{1}{p} \ge \log \log z + O(1),$$

which we prove using partial summation from the following theorem of Chebycheff (1.4.5 in [1]):

Theorem 2.1.

$$\sum_{p \le z} \frac{\log p}{p} \ge \log z + O(1).$$

We do the summation by parts:

$$\begin{split} \sum_{p \le z} \frac{1}{p} &= \sum_{p \le z} \frac{\log p}{p} \frac{1}{\log p} = \frac{1}{\log z} \sum_{p \le z} \frac{\log p}{p} + \int_{1}^{z} \frac{\sum_{p \le t} (\log p/p)}{t \log^{2} t} dt \\ &\ge \frac{1}{\log z} (\log z + O(1)) + \int_{1}^{z} \frac{\log t + O(1)}{t \log^{2} t} dt \\ &= 1 + O\left(\frac{1}{\log z}\right) + \int_{1}^{z} \frac{1}{t \log t} dt + O\left(\int_{1}^{z} \frac{1}{t \log^{2} t} dt\right) \\ &= 1 + O\left(\frac{1}{\log z}\right) + O\left(-\frac{1}{\log z}\right) + \log \log z, \end{split}$$

which gives the required result.

Using this, we can obtain an upper bound for $\Phi(x, z)$ and, by substituting in (6), for $\pi(x)$. We choose $z := c \log x$ for a small positive constant c to get the result:

Proposition 2.2.

$$\pi(x) \ll \frac{x}{\log \log x}$$

We now want to study a more general setting. Let \mathcal{A} be a set of natural numbers lest than or equal to x and let \mathcal{P} be a set of primes. To every $p \in \mathcal{P}$ we assign a number $\omega(p)$ of distinct residue classes modulo p.

Remark 1. In Proposition 2.2, $\omega(p) = 1$ and we fix the residue class 0 (mod p).

Let \mathcal{A}_p denote the set of elements of \mathcal{A} belonging to at least one of these distinct residue classes modulo p, let $\mathcal{A}_1 := \mathcal{A}$ and for any squarefree integer d composed of primes in \mathcal{P} we define

$$\mathcal{A}_d := igcap_{p|d} \mathcal{A}_p$$

and

$$\omega(d) := \prod_{p|d} \omega(p).$$

The Chinese remainder theorem ensures $\omega(d)$ is well defined: given two distinct primes p_1 and p_2 and a residue class modulo each of these primes, there exists a unique residue class modulo p_1p_2 . We proceed by induction.

Let z be a positive real number and define

$$P(z) := \prod_{\substack{p \in \mathcal{P} \\ p < z}} p$$

We define $S(\mathcal{A}, \mathcal{P}, z)$ to be the number of elements of the set

$$\mathcal{A}\setminus \bigcup_{p\mid P(z)} \mathcal{A}_p.$$

Assume that there exists an X such that

$$\#\mathcal{A}_d = \frac{\omega(d)}{d}X + R_d$$

for some R_d .

Theorem 2.3 (The Sieve of Eratosthenes). With the above setting, suppose the following conditions hold: 1. $|R_d| = O(\omega(d));$ 2. for some $\kappa \geq 0$,

$$\sum_{p|P(z)} \frac{\omega(p)\log p}{p} \le \kappa \log z + O(1);$$

3. for some positive real number y, $\# \mathcal{A}_d = 0$ for every d > y. Then

$$S(\mathcal{A}, \mathcal{P}, z) = XW(z) + O\left(\left(X + \frac{y}{\log z}\right)(\log z)^{\kappa+1}\exp\left(-\frac{\log y}{\log z}\right)\right),$$

where

$$W(z) := \prod_{\substack{p \in \mathcal{P} \\ p < z}} \left(1 - \frac{\omega(p)}{p} \right).$$

Remark 2. As we assign more residue classes to each prime $(\omega(p) \text{ increases})$, κ becomes bigger and, as a consequence, so does the error term in the third condition.

Remark 3. In Proposition 2.2, $\kappa = 1$ and the second remark becomes

$$\sum_{p < z} \frac{\log p}{p} \le \log z + O(1),$$

which was proven by Mertens (see [2] for alternative proof).

The proof requires the following lemmata, which we quote from [1] without proof:

Lemma 2.4. With the setting and hypotheses of Theorem 2.3, let

$$F(t,z) := \sum_{\substack{d \le t \\ d | \overline{P}(z)}} \omega(d).$$

Then

$$F(t,z) = O\left(t(\log z)^{\kappa} \exp\left(-\frac{\log t}{\log z}\right)\right).$$

Lemma 2.5. With the setting and hypotheses of Theorem 2.3,

$$\sum_{\substack{d \mid P(z) \\ d > y}} \frac{\omega(d)}{d} = O\left((\log z)^{\kappa+1} \exp\left(-\frac{\log y}{\log z}\right) \right).$$

Proof of Theorem 2.3. We want to use the inclusion-exclusion principle from combinatorics, namely

$$\left| \bigcup_{i=1}^{n} B_{i} \right| = \sum_{k=1}^{n} (-1)^{k+1} \left(\sum_{1 \le i_{1} < \dots < i_{k} \le n} |B_{i_{1}} \cap \dots \cap B_{i_{k}}| \right).$$

Using this and the first and third hypotheses of Theorem 2.3, we get

$$S(\mathcal{A}, \mathcal{P}, z) = \sum_{\substack{d \mid P(z) \\ d \le y}} \mu(d) \# \mathcal{A}_d = \sum_{\substack{d \mid P(z) \\ d \le y}} \mu(d) \frac{X\omega(d)}{d} + O(F(y, z))$$
$$= X\left(\sum_{\substack{d \mid P(z) \\ d \mid P(z)}} \mu(d) \frac{\omega(d)}{d} - \sum_{\substack{d \mid P(z) \\ d > y}} \mu(d) \frac{\omega(d)}{d}\right) + O(F(y, z)).$$

We write the first sum inside the main term as its Euler product and we apply Lemma 2.5 to the second one and Lemma 2.4 to the error term to obtain

$$S(\mathcal{A}, \mathcal{P}, z) = XW(z) + O\left(\left(X + \frac{y}{\log z}\right)(\log z)^{\kappa+1}\exp\left(-\frac{\log y}{\log z}\right)\right).$$

2.2 Selberg's sieve

In 1947, Selberg came with a clever contribution in estimating $\Phi(x, z)$, as defined in (3). His idea was to replace the Möbius function in (5) with a quadratic form, which can then be minimised for optimal results. This method was then used with small modifications by Goldston *et al.* in [7] and subsequently by Maynard, in a multi-dimensional setting, to obtain the results mentioned in the Introduction.

Selberg made the observation that for any sequence (λ_d) of real numbers such that $\lambda_1 = 1$, the following holds:

$$\sum_{d|k} \mu(d) \le \left(\sum_{d|k} \lambda_d\right)^2.$$

This is because the left-hand side of the inequality is equal to 1 if k = 1 and to 0 otherwise, while the right-hand side is 1 if k = 1 and greater than or equal to 0 otherwise.

Hence, from (5) we get

$$\Phi(x,z) \leq \sum_{n \leq x} \left(\sum_{d \mid (n,P_z)} \lambda_d \right)^2$$

=
$$\sum_{n \leq x} \left(\sum_{d_1,d_2 \mid (n,P_z)} \lambda_{d_1} \lambda_{d_2} \right)$$

=
$$\sum_{d_1,d_2 \mid P_z} \lambda_{d_1} \lambda_{d_2} \sum_{\substack{n \leq x \\ [d_1,d_2] \mid n}} 1$$

$$\leq x \sum_{d_1,d_2 \mid P_z} \frac{\lambda_{d_1} \lambda_{d_2}}{[d_1,d_2]} + O\left(\sum_{d_1,d_2 \mid P_z} |\lambda_{d_1}| |\lambda_{d_2}| \right),$$

where by [a, b] we've denoted the least common multiple of a and b. Since our sequence (λ_d) was chosen arbitrarily apart from the one condition, we may make the assumption that $\lambda_d = 0$ for d > z and obtain

$$\Phi(x,z) \le x \sum_{d_1,d_2 \le z} \frac{\lambda_{d_1} \lambda_{d_2}}{[d_1,d_2]} + O\left(\sum_{d_1,d_2 \le z} |\lambda_{d_1}| |\lambda_{d_2}|\right).$$
(7)

Our purpose is to estimate the main term so we look at

$$\sum_{d_1,d_2 \le z} \frac{\lambda_{d_1} \lambda_{d_2}}{[d_1,d_2]} \tag{8}$$

as a quadratic form in $(\lambda_d)_{d \leq z}$ and we try to minimize it.

We use the fact that $d_1, d_2 = d_1d_2$, where we've denote by (a, b) the greatest common divisor of a and b, and that $\sum_{\delta|d} \phi(\delta) = d$, where $\phi(\cdot)$ is Euler's totient function, i.e. $\phi(\delta)$ is the number of positive integers less than or equal to δ that are relatively prime to δ . We write (8) as

$$\sum_{d_1,d_2 \le z} \frac{\lambda_{d_1} \lambda_{d_2}}{[d_1,d_2]} = \sum_{d_1,d_2 \le z} \frac{\lambda_{d_1} \lambda_{d_2}}{d_1 d_2} (d_1,d_2) = \sum_{d_1,d_2 \le z} \frac{\lambda_{d_1} \lambda_{d_2}}{d_1 d_2} \sum_{\substack{\delta \mid (d_1,d_2)}} \phi(\delta)$$
$$= \sum_{\delta \le z} \phi(\delta) \sum_{\substack{d_1,d_2 \le z\\\delta \mid (d_1,d_2)}} \frac{\lambda_{d_1} \lambda_{d_2}}{d_1 d_2} = \sum_{\delta \le z} \phi(\delta) \left(\sum_{\substack{d \le z\\\delta \mid d}} \frac{\lambda_d}{d}\right)^2,$$

so we have diagonalised our quadratic form to

$$\sum_{\delta \le z} \phi(\delta) u_{\delta}^2,$$

under the linear transformation

$$u_{\delta} := \sum_{\substack{d \le z \\ \delta \mid d}} \frac{\lambda_d}{d}.$$
(9)

This transformation is invertible and we can use the dual Möbius inversion formula (1.2.3 in [1]) to obtain

$$\frac{\lambda_{\delta}}{\delta} = \sum_{\substack{d \le z\\\delta \mid d}} \mu\left(\frac{d}{\delta}\right) u_d.$$

Bearing in mind our conditions on (λ_d) , we have $u_{\delta} = 0$ for $\delta > z$ and

$$\sum_{d \le z} \mu(d) u_d = \lambda_1 = 1.$$
(10)

Using this, we can write

$$\sum_{\delta \le z} \phi(\delta) u_{\delta}^2 = \sum_{\delta \le z} \phi(\delta) \left(u_{\delta} - \frac{\mu(\delta)}{\phi(\delta)V(z)} \right)^2 + \frac{1}{V(z)},$$

where $V(z) := \sum_{d \leq z} \mu^2(d) / \phi(d)$. We want to minimise this quadratic form under the constraint (10). It is easy to see that we have a minimal value of 1/V(z), which occurs when

$$u_{\delta} = \frac{\mu(\delta)}{\phi(\delta)V(z)} \implies \lambda_{\delta} = \delta \sum_{\delta \mid d} \frac{\mu(d/\delta)\mu(d)}{\phi(d)V(z)}.$$

Using this, we can prove that $|\lambda_d| \leq 1$ for any d (see p.117 in [1] for proof). Substituting our results into (7), we obtain

$$\Phi(x,z) \le \frac{x}{V(z)} + O(z^2). \tag{11}$$

Remark 4. We can use this to obtain a better estimate for $\pi(x)$ than Proposition 2.2, namely:

$$\pi(x) \ll \frac{x}{\log x},$$

which is Chebyscheff's upper bound for $\pi(x)$. The proof is again based on writing

$$\pi(x) \le \Phi(x, z) + z$$

and we use (11) to estimate $\Phi(x, z)$. We have

$$\sum_{d \le z} \frac{\mu(d)^2}{\phi(d)} \ge \sum_{d \le z} \frac{\mu(d)^2}{d} = \sum_{d \le z} \frac{1}{d} - \sum_{d \le z} \frac{1}{d},$$

where the $\sum_{d\leq z}'$ denotes that we are summing over non squarefree d. We now quote without proof the following proposition from [1] (1.3.3):

Proposition 2.6.

$$\sum_{d \le z} \frac{1}{d} = \log z + O(1).$$

We use this and the fact that

$$\sum_{d \le z} \frac{1}{d} \le \frac{1}{4} \sum_{d \le z/4} \frac{1}{d},$$

since if d has a squarefree divisor $k \ge 2$, say, then $d \le z$ implies $d/k^2 \le z/4$ and also

$$\frac{1}{d} \le \frac{1}{4d/k}.$$

We obtain

$$\pi(x) \ll \frac{x}{\log z} + z^2.$$

We then need to choose an appropriate z again, which in this case will be $z := (x/\log x)^{1/2}$. This gives the result.

We now discuss the general Selberg sieve. Let \mathcal{A} be a finite set of natural numbers and let \mathcal{P} be a set of primes. For every $p \in \mathcal{P}$, let \mathcal{A}_p be a subset of \mathcal{A} (notice that, unlike the sieve of Eratosthenes, \mathcal{A}_p does not necessarily depend on fixed residue classes modulo p). Let $\mathcal{A}_1 := \mathcal{A}$ and for any squarefree integer d composed of primes in \mathcal{P} we define

$$\mathcal{A}_d := \bigcap_{p|d} \mathcal{A}_p.$$

Let z be a positive real number and define P(z) and $S(\mathcal{A}, \mathcal{P}, z)$ as before.

Theorem 2.7 (Selberg's sieve). With the above setting, suppose there exists X > 0 and a multiplicative function $f(\cdot)$ which satisfies f(p) > 1 for any prime $p \in \mathcal{P}$, such that for any squarefree integer d composed of primes in \mathcal{P} we have

$$#\mathcal{A}_d = \frac{X}{f(d)} + R_d \tag{12}$$

for some real number R_d . Write

$$f(n) = \sum_{d|n} f_1(d),$$
 (13)

where $f_1(\cdot)$ is a multiplicative function which is uniquely determined by $f(\cdot)$ using the Möbius inversion formula, in other words $f_1(n) = \sum_{d|n} \mu(d) f(n/d)$. Set

$$V(z) := \sum_{\substack{d \le z \\ d \mid P(z)}} \frac{\mu^2(d)}{f_1(d)}.$$

Then

$$S(\mathcal{A}, \mathcal{P}, z) \le \frac{X}{V(z)} + O\left(\sum_{\substack{d_1, d_2 \le z \\ d_1, d_2 | P(z)}} |R_{[d_1, d_2]}|\right).$$

Remark 5. Note the inequality sign in Selberg's sieve, compared to the equality we had in Theorem 2.3. The main terms in the two sieves have the same order of magnitude (both are $O(x/\log z)$), but there is an improvement in the error term.

Remark 6. In (11), in the formula for V(z), we have $f_1(\cdot)$ to be the Euler function, hence $f(n) = \sum_{d|n} \phi(d) = n$ is the identity function.

The proof of this theorem is analogous to that of (11):

Proof of Theorem 2.7. As before, we start with a sequence (λ_d) of real numbers satisfying $\lambda_1 = 1$ and $\lambda_d = 0$ for d > z.

We want to obtain a similar estimate for $S(\mathcal{A}, \mathcal{P}, z)$ as we did for $\Phi(x, z)$. By the inclusion-exclusion principle, we have

$$S(\mathcal{A}, \mathcal{P}, z) = \sum_{\substack{a \in \mathcal{A} \\ a \notin \mathcal{A}_p \forall p | P(z)}} 1 = \sum_{d | P(z)} \mu(d) \sum_{a \in \mathcal{A}_d} 1 = \sum_{a \in \mathcal{A}} \left(\sum_{\substack{d | P(z) \\ a \in \mathcal{A}_d}} \mu(d) \right).$$

We define, for any $a \in \mathcal{A}$,

$$D(a) := \prod_{\substack{p \in \mathcal{P} \\ a \in \mathcal{A}_p}} p,$$

and if $a \notin \mathcal{A}_p$ for any $p \in \mathcal{P}$, then D(a) := 1. We have

$$\sum_{\substack{d|P(z)\\a\in\mathcal{A}_d}}\mu(d) = \sum_{\substack{d|(P(z),D(a))}}\mu(d) \le \left(\sum_{\substack{d|(P(z),D(a))}}\lambda_d\right)^2 = \left(\sum_{\substack{d|P(z)\\a\in\mathcal{A}_d}}\lambda_d\right)^2.$$

Substituting in our formula for $S(\mathcal{A}, \mathcal{P}, z)$, we obtain:

$$S(\mathcal{A}, \mathcal{P}, z) \leq \sum_{a \in \mathcal{A}} \left(\sum_{\substack{d \mid P(z) \\ a \in \mathcal{A}_d}} \lambda_d \right)^2 = \sum_{a \in \mathcal{A}} \left(\sum_{\substack{d_1, d_2 \mid P(z) \\ a \in \mathcal{A}_{[d_1, d_2]}}} \lambda_{d_1} \lambda_{d_2} \right)$$
$$= \sum_{d_1, d_2 \leq z} \lambda_{d_1} \lambda_{d_2} \# \mathcal{A}_{[d_1, d_2]}$$
$$= X \sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 \mid P(z)}} \frac{\lambda_{d_1} \lambda_{d_2}}{f([d_1, d_2])} + O\left(\sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 \mid P(z)}} |\lambda_{d_1}| |\lambda_{d_1}| |R_{[d_1, d_2]}| \right),$$

using (12).

We look at the sum in our main term as a quadratic form in $(\lambda_d)_{d\leq z}$, which we want to diagonalise and then minimise in order to finish the proof. For that, we need to use the fact that for a multiplicative function $f(\cdot)$ and two positive squarefree integers d_1 and d_2 ,

$$f([d_1, d_2])f((d_1, d_2)) = f(d_1)f(d_2),$$
(14)

which is similar to what we have used before. Using this and (13), we obtain

$$\sum_{\substack{d_1,d_2 \leq z \\ d_1,d_2 | P(z)}} \frac{\lambda_{d_1} \lambda_{d_2}}{f([d_1, d_2])} = \sum_{\substack{d_1,d_2 \leq z \\ d_1,d_2 | P(z)}} \frac{\lambda_{d_1} \lambda_{d_2}}{f(d_1)f(d_2)} f((d_1, d_2))$$

$$= \sum_{\substack{d_1,d_2 \leq z \\ d_1,d_2 | P(z)}} \frac{\lambda_{d_1} \lambda_{d_2}}{f(d_1)f(d_2)} \sum_{\substack{\delta | (d_1,d_2)}} f_1(\delta)$$

$$= \sum_{\substack{\delta \leq z \\ \delta | P(z)}} f_1(\delta) \sum_{\substack{d_1,d_2 \leq z \\ \delta | (d_1,d_2)}} \frac{\lambda_{d_1} \lambda_{d_2}}{f(d_1)f(d_2)}$$

$$= \sum_{\substack{\delta \leq z \\ \delta | P(z)}} f_1(\delta) \left(\sum_{\substack{d \leq z \\ \delta | P(z)}} \frac{\lambda_{d}}{f(d)}\right)^2.$$

Hence our quadratic form is reduced to the diagonal form

$$\sum_{\substack{\delta \le z\\\delta \mid P(z)}} f_1(\delta) u_{\delta}^2,$$

under the linear transformation

$$u_{\delta} := \sum_{\substack{d \le z \\ d \mid P(z) \\ \delta \mid d}} \frac{\lambda_d}{f(d)},$$

which is invertible and by the dual Möbius inversion formula we have

$$\frac{\lambda_{\delta}}{f(\delta)} = \sum_{\substack{d \mid P(z)\\\delta \mid d}} \mu\left(\frac{d}{\delta}\right) u_d.$$

Bearing in mind our conditions on (λ_d) , we have $u_{\delta} = 0$ for $\delta > z$ and $\sum_{\substack{d \leq z \\ d \mid P(z)}} \mu(d) u_d = \lambda_1 = 1$. Using this, we can write

$$\sum_{\substack{\delta \le z \\ P(z)}} f_1(\delta) u_{\delta}^2 = \sum_{\substack{\delta \le z \\ \delta | P(z)}} f_1(\delta) \left(u_{\delta} - \frac{\mu(\delta)}{f_1(\delta)V(z)} \right)^2 + \frac{1}{V(z)},$$

from which we deduce that our quadratic form has a minimal value of 1/V(z), which occurs when

$$u_{\delta} = \frac{\mu(\delta)}{f_1(\delta)V(z)},$$

since the coefficients appearing in the quadratic form, $f_1(d)$, are positive by multiplicativity of $f_1(\cdot)$.

2.3 Admissible sets

In Maynard's paper, the main focus is on k-tuples of prime numbers of the form $\{n+h_1, \ldots, n+h_k\}$. He proves that there are infinitely many intervals of finite length that contain k primes.

This is something number theorist have been working on for a long time. In 1849, de Polignac conjectured that for every positive even natural number k there are infinitely many consecutive prime pairs $\{p_n, p_{n+1}\}$ such that $p_{n+1} - p_n = k$. Zhang proved that this holds for some $k < 7 \times 10^7$ and the result is being constantly improved with the help of sieve theory and computer programs. The case k = 2 is the famous twin prime conjecture.

We start by considering a set $\mathcal{H} = \{h_1, \ldots, h_k\}$ of natural numbers. For each prime p we look at the set

$$\Omega(p) = \{n \pmod{p} : n \equiv -h_i \pmod{p} \text{ for some } h_i \in \mathcal{H} \}$$

of residue classes in \mathcal{H} modulo p.

Definition 2.2 (Admissible set). We call a set \mathcal{H} defined as above *admissible* if, for every prime p, $|\Omega(p)| < p$.

Remark 7. We want our set \mathcal{H} to generate k-tuples of primes of the form $\{n + h_1, \ldots, n + h_k\}$. Therefore, if $\Omega(p) = p$ for some p, at least one of the $n + h_i$ would be congruent to 0 modulo p and therefore could not be a prime.

Remark 8. Given a set \mathcal{H} , it is enough to check if it is admissible up to the biggest prime less than or equal to k, since for a prime greater than k it is impossible to cover all residue classes modulo said prime.

Example 2.1. The set $\mathcal{H} = \{1, 3, 5\}$ is not admissible. We need to check for primes less than or equal to 3: $-1 \equiv -3 \equiv -5 \equiv 1 \pmod{2}$, so $\Omega(2) = \{1\}$, however $-1 \equiv 2 \pmod{3}$, $-3 \equiv 0 \pmod{3}$ and $-5 \equiv 1 \pmod{3}$, so $\Omega(3) = \{1, 2, 3\}$.

G. H. Hardy and J. E. Littlewood [3] stated the following conjecture regarding k-tuples of primes:

Conjecture 2.8 (Hardy-Littlewood). Let \mathcal{H} be an admissible set. Then there exist infinitely many k-tuples of primes of the form $\{n+h_1, \ldots, n+h_k\}$ and in fact we can count them asymptotically:

$$\#\{n \le x : n+h_1, \dots, n+h_k \text{ all prime}\} \sim \mathfrak{G}(\mathcal{H})\frac{x}{(\log x)^k},$$

where we define

$$\mathfrak{G}(\mathcal{H}) := \prod_{p} \left(1 - \frac{|\Omega(p)|}{p} \right) \left(1 - \frac{1}{p} \right)^{-k}$$
(15)

and we call it the singular series.

Remark 9. We can see that for k = 1, the conjecture is equivalent to saying that $\#\{n \le x : n+h \text{ prime}\} \sim \mathfrak{G}(\{h\})x/\log x$, and in this case $|\Omega(p)| = 1$ for all p, implying $\mathfrak{G}(\{h\}) = 1$. Thus we get a restatement of the prime number theorem, (1).

Given an admissible set $\mathcal{H} = \{h_1, \ldots, h_k\}$, we can use Selberg's sieving technique to get an upper bound for the set

 $S = \{n : n + h_1, \dots, n + h_k \text{ all prime and } n \in [N, 2N]\},\$

for some large N, which will later become of interest to us. Let

$$P(n) := (n + h_1) \dots (n + h_k).$$
(16)

We want to use the same trick as the one in the Selberg sieve. To estimate the size of our set S we need a function which is equal to 1 if all of the $n+h_1, \ldots, n+h_k$ are prime and non-negative otherwise, to replace the Möbius function. We define a sequence (λ_d) with $\lambda_1 = 1$ and $\lambda_d = 0$ for d > R for some R. Then

$$|S| \leq \sum_{N \leq n \leq 2N} \left(\sum_{d|P(n)} \lambda_d \right)^2,$$

as long as we fix R to be less than N. This is because if this holds, when all of the $n + h_i$'s are prime, the only non-zero term in the inner sum is for d = 1.

We want to extend $\Omega(\cdot)$ multiplicatively (similar to how we extended $\omega(\cdot)$ multiplicatively in Subsection 2.1) as $n \in \Omega(d)$ if and only if $n \in \Omega(p)$ for all $p \mid d$. We can use the Chinese remainder theorem to show that this implies $|\Omega(d)| = \prod_{p \mid d} |\Omega(p)|$. We follow the same steps as in the proof of (11) to get

$$|S| \le x \left(\sum_{d_1, d_2 \le R} \lambda_{d_1} \lambda_{d_2} \frac{|\Omega([d_1, d_2])|}{[d_1, d_2]} \right) + O\left(\sum_{d_1, d_2} |\Omega([d_1, d_2])| |\lambda_{d_1}| |\lambda_{d_2}| \right).$$

Note that, compared to (7), there is an extra factor of $|\Omega([d_1, d_2])|$ appearing both in the error term and in the main term. This is because if $n \in \Omega(d)$ for some d, then so does n + d, so when we split the interval [N, 2N] into $\lfloor x/d \rfloor$ intervals of length d, each interval contains exactly $|\Omega(d)|$ integers $n \in \Omega(d)$.

To minimise the main term, we know from [14] that the best choice for diagonalising λ_d is

$$\lambda_d \approx \mu(d) \left(\frac{\log R/d}{\log R}\right)^k.$$
 (17)

With this choice, we get

$$|S| \le O\left(2^k k! \mathfrak{G}(\mathcal{H}) \frac{N}{(\log N)^k}\right),\,$$

which is a factor of $2^k \cdot k!$ away from the result stated in the conjecture.

2.4 Level of distribution of primes

We will now talk about primes in arithmetic progressions and we will state the Bombieri–Vinogradov theorem.

We define $\pi(x; k, a)$ to be the number of primes $p \leq x$ with $p \equiv a \pmod{k}$. The definition makes sense only when a is coprime to k, as otherwise p would be divisible by a and hence would not be a prime, unless a = p. Note that the case k = 1 is simply $\pi(x)$. The prime number theorem for primes in progression states:

$$\pi(x;k,a) \sim \frac{\mathrm{li}(x)}{\phi(k)},\tag{18}$$

where $\operatorname{li}(x) := \int_2^x \log^{-1} t \, dt$. When k = 1, we get $\pi(x) \sim \operatorname{li}(x)$ and it is known that $\operatorname{li}(x) \sim x/\log x$, thus we get the prime number theorem.

Note that a does not appear in the right-hand side of (18), implying that primes in arithmetic progressions are evenly distributed, i.e. an arithmetic progression of the form $a \pmod{k}$ contains asymptotically as many primes as one of the form $b \pmod{k}$ whenever a and b are coprime to k.

It is proven in [15] that, for any fixed N > 0,

$$\pi(x;k,a) = \frac{\mathrm{li}(x)}{\phi(k)} + O(x \exp\left(-c\sqrt{\log x}\right))$$

holds for all $x \ge 2$ and all integers k, a with (a, k) = 1 and $1 \le k \le (\log x)^N$ and c is an absolute constant.

We now define the level of distribution of primes:

Definition 2.3 (Level of Distribution). Given $\theta > 0$, we say that primes have *level of distribution* θ if, for any A > 0, the following holds:

$$\sum_{k \le x^{\theta}} \max_{(a,k)=1} \left| \pi(x;k,a) - \frac{\pi(x)}{\phi(k)} \right| \ll_A \frac{x}{(\log x)^A}.$$
 (19)

Under the Generalised Riemann hypothesis (GRH), it can be shown that

$$\pi(x;k,a) = \frac{\pi(x)}{\phi(k)} + O(x^{1/2}\log x),$$

so if we allow k to vary up to $x^{1/2}$ and take the maximum over it, this implies that primes have level of distribution θ for $\theta < 1/2$.

Bombieri and Vinogradov validated this claim independently in 1965 without the use of the GRH, by proving the following theorem:

Theorem 2.9 (Bombieri–Vinogradov). For any A > 0 there exists B = B(A) > 0 such that

$$\sum_{k \le \frac{x^{1/2}}{(\log x)^B}} \max_{y \le x} \max_{(a,k)=1} \left| \pi(y;k,a) - \frac{\mathrm{li}(y)}{\phi(k)} \right| \ll_A \frac{x}{(\log x)^A}.$$
 (20)

Remark 10. We allow k to vary up to $x^{1/2}/(\log x)^B$ since we can find $\varepsilon > 0$ such that $x^{1/2-\varepsilon} \ll x^{1/2}/(\log x)^B \le x^{1/2}$.

Remark 11. We are also taking the maximum over all $y \leq x$, which is an improvement to our definition.

Their proof was based estimating certain L-functions in various rectangles with the use of the large sieve method, which we quote from [1]:

Theorem 2.10 (The large sieve). Let \mathcal{A} be a set of natural numbers lest than or equal to x and let \mathcal{P} be a set of primes. For every $p \in \mathcal{P}$ suppose we are given a set $\{w_{1,p}, \ldots, w_{\omega(p),p}\}$ of $\omega(p)$ distinct residue classes modulo p. Let z be a positive real number and define P(z) as we have before. We define $\mathcal{S}(\mathcal{A}, \mathcal{P}, z)$ to be the number of elements of the set

$$\{n \in \mathcal{A} : n \not\equiv w_{i,p} (\text{mod } p) \ \forall 1 \le i \le \omega(p), \forall p \mid P(z) \}.$$

Then

$$\mathcal{S}(\mathcal{A}, \mathcal{P}, z) \le \frac{z^2 + 4\pi x}{L(z)},$$

where

$$L(z) := \sum_{d \le z} \mu^2(d) \prod_{p \mid d} \frac{\omega(p)}{p - \omega(p)}.$$

It was conjectured by Elliott and Halberstam [3] that primes have level of distribution θ for $\theta < 1$, while Friedlander and Granville [5] proved that (19) does not hold if we replace x^{θ} with $x/(\log x)^{B}$ for any fixed B, like the result obtained in Theorem 2.9.

We can also define an equivalent of the function $\vartheta(\cdot)$ for primes in progressions, by

$$\vartheta^*(y;q,a) := \sum_{\substack{y < n \le 2y\\n \equiv a \mod q}} \varpi(n), \tag{21}$$

where we define $\varpi(n)$ to be equal to $\log n$ if n is a prime and 0 otherwise. It can be shown that (20) is equivalent to

$$\sum_{q \le x^{\theta}} \max_{y \le x} \max_{(a,q)=1} \left| \vartheta^*(y;q,a) - \frac{y}{\phi(q)} \right| \ll \frac{x}{(\log x)^A}.$$
 (22)

3 The Goldston–Pintz–Yıldırım theorem

As we mentioned in the Introduction, Goldston *et al.* use a variation of Selberg's sieve to prove the following theorem:

Theorem 3.1 (GPY).

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

This shows that there are infinitely many consecutive primes that have an arbitrarily small gap between them, compared to the average gap implied by (1). We will now highlight some of the ideas in their proof.

Proof of Theorem 3.1. We start by introducing some basic notation.

Let N be a parameter that increases monotonically to infinity and we fix H and R such that

$$H \ll \log N \ll \log R \le \log N.$$

By k and l we denote arbitrary positive integers that are bounded.

Remark 12. We will see later on that the condition that k and l are chosen arbitrarily is crucial, because the correct choice enables us to complete the proof.

Let $\mathcal{H} = \{h_1, \ldots, h_k\}$ be an admissible set defined as before and we impose on it the condition that $\mathcal{H} \subseteq [1, H]$. We next define a sequence $(\lambda_R(d; a))$ in a similar way to (17), by

$$\lambda_R(d;a) = \begin{cases} 0 & \text{if } d > R, \\ \frac{1}{a!}\mu(d) \left(\log \frac{R}{d}\right)^a & \text{if } d \le R. \end{cases}$$

Our first aim is to evaluate the quantity

$$\sum_{N < n \le 2N} \left(\sum_{n \in \Omega(d)} \lambda_R(d; k+l) \right)^2$$
(23)

Note that $n \in \Omega(d) \iff d \mid P(n)$ with P(n) defined as in (16). We let $\Lambda_R(n; \mathcal{H}, a) := \sum_{n \in \Omega(d)} \lambda_R(d; a)$ and we claim that the following holds:

Lemma 3.2. With the above notation, for $R \leq N^{1/2}/(\log N)^C$, where C > 0 depends only on k and l,

$$\sum_{N < n \le 2N} \Lambda_R(n; \mathcal{H}; k+l)^2 = \frac{\mathfrak{G}(\mathcal{H})}{(k+2l)!} \binom{2l}{l} N(\log R)^{k+2l} + O(N(\log N)^{k+2l-1}(\log \log N)^c),$$

with $\mathfrak{G}(\mathcal{H})$ defined as in (15).

Sketch of proof of Lemma 3.2. We can expand the square in (23) and write our quantity as

$$N\mathcal{T} + O\left(\left(\sum_{d} |\Omega(d)| |\lambda_R(d; k+l)|\right)^2\right),$$

with

$$\mathcal{T} = \sum_{d_1, d_2} \frac{|\Omega([d_1, d_2])|}{[d_1, d_2]} \lambda_R(d_1; k+l) \lambda_R(d_1; k+l).$$

To estimate the error term, note that if d is a squarefree integer which can be written as the product of m distinct primes, we have that $\tau_k(d) = k^m$, where $\tau_k(\cdot)$ is the generalised divisor function, $\tau_k(n) = |\{(a_1, \ldots, a_k) : a_1 \ldots a_k = n\}|$. Then, since $|\mathcal{H}| = k$ and we require $|\Omega(p)|$ to be less than k be definition, we have

$$|\Omega(d)| = |\Omega(p_1)| \dots |\Omega(p_m)| \le k^m = \tau_k(d).$$

We have

$$\zeta^k(s) = \sum_{n=1}^{\infty} \frac{\tau_k(n)}{n^s}.$$

We can then use Perron's discontinuous integral and the Leibniz rule for differentiation to get

$$\sum_{n \le x} \tau_k(n) = \frac{x(\log x)^{k-1}}{(k-1)!} + O(x(\log x)^{k-2}),$$

hence we can write our error term as

$$O\left(\left(\sum_{d} \tau_k(d) \log\left(\frac{R}{d}\right)^a\right)^2\right) = O((R(\log R)^{k-1}(\log R)^a)^2)$$
$$= O(R^2(\log R)^c).$$
(24)

To estimate the main term, we use contour integration along cleverly chosen vertical lines. We concentrate on the main ideas and ignore the error terms. First, note that we can write

$$\lambda_R(d;a) = \frac{\mu(d)}{2\pi i} \int_{(1)} \left(\frac{R}{d}\right)^s \frac{ds}{s^{a+1}},\tag{25}$$

as long as $a \ge 1$, where by (α) we denote the vertical line passing through α in the complex plane. We use this to write \mathcal{T} as

$$\mathcal{T} = \frac{1}{(2\pi i)^2} \int_{(1)} \int_{(1)} F(s_1, s_2; \Omega) \frac{R^{s_1 + s_2}}{(s_1 s_2)^{k+l+1}} ds_1 ds_2, \tag{26}$$

with

$$F(s_1, s_2; \Omega) = \sum_{d_1, d_2} \mu(d_1) \mu(d_2) \frac{|\Omega([d_1, d_2])|}{[d_1, d_2] d_1^{s_1} d_2^{s_2}},$$

which can also be written as

$$F(s_1, s_2; \Omega) = \prod_p \left(1 - \frac{|\Omega(p)|}{p} \left(\frac{1}{p^{s_1}} + \frac{1}{p^{s_2}} - \frac{1}{p^{s_1+s_2}} \right) \right),$$

in the region of absolute convergence. We then let

$$G(s_1, s_2; \Omega) := F(s_1, s_2; \Omega) \left(\frac{\zeta(s_1 + 1)\zeta(s_2 + 1)}{\zeta(s_1 + s_2 + 1)} \right)^k,$$

which is regular and bounded for $\Re(s_1), \Re(s_2) > -1/2$.

We get the singular series in the main term from

$$G(0,0;\Omega) = \prod_{p} \left(1 - \frac{|\Omega(p)|}{p}\right) \left(1 - \frac{1}{p}\right)^{-k} = \mathfrak{G}(\mathcal{H}),$$

by definition.

We have $G(s_1, s_2; \Omega) = O(\exp(c'(\log N)^{-2\sigma} \log \log \log N))$, with $\sigma := \min(\Re(s_1), \Re(s_2), 0)$, which we obtain by writing out $\log G(s_1, s_2; \Omega)$ explicitly.

We can use $G(\cdot)$ to estimate \mathcal{T} by shifting both contours in (26). We let $U = \exp(\sqrt{\log N})$ and we shift the s_1 -contour to the vertical line $c_0(\log U)^{-1} + it$ and the s_2 -contour to $c_0(2\log U)^{-1} + it$, where $t \in \mathbb{R}$ and c_0 is a sufficiently small positive constant. Next, we truncate the contours to $|t| \leq U$ and $|t| \leq U/2$ and we denote the results by L_1 and L_2 . With these choices (26) becomes

$$\begin{aligned} \mathcal{T} = & \frac{1}{(2\pi i)^2} \int_{L_1} \int_{L_2} G(s_1, s_2; \Omega) \left(\frac{\zeta(s_1 + s_2 + 1)}{\zeta(s_1 + 1)\zeta(s_2 + 1)} \right)^k \frac{R^{s_1 + s_2}}{(s_1 s_2)^{k + l + 1}} ds_1 ds_2 \\ &+ O(exp(-c\sqrt{\log N})). \end{aligned}$$

We then shift the s_1 -contour to $L3 = -c_0(\log U)^{-1} + it$, for $|t| \le U$ and we encounter a pole of order l + 1 at $s_1 = 0$ and a pole of order k at $s_1 = -s_2$. We use Cauchy's residue theorem to bound Res $s_1 = -s_2$ asymptotically and obtain

$$\mathcal{T} = \frac{1}{2\pi i} \int_{L_2} \left\{ \operatorname{Res}_{s_1=0} \right\} ds_2 + O((\log N)^{k+l-1/2} (\log \log N)^c),$$
(27)

for some constant c. In order to estimate this, we define

$$Z(s_1, s_2) = G(s_1, s_2; \Omega) \left(\frac{(s_1 + s_2)\zeta(s_1 + s_2 + 1)}{s_1\zeta(s_1 + 1)\zeta(s_2 + 1)} \right)^k,$$

which is regular and bounded in a neighbourhood of the point (0,0), since $G(\cdot)$ is regular in that area, while the poles of $\zeta(s_i+1)$ and the zeros given by s_i at $s_i = 0$ cancel each other in the denominator. Cauchy's residue formula gives us

$$\operatorname{Res}_{s_1=0} = \frac{R^{s_2}}{l! s_2^{l+1}} \left(\frac{\partial}{\partial s_1}\right)_{s_1=0}^l \left\{\frac{Z(s_1, s_2)}{(s_1+s_2)^k} R^{s_1}\right\}.$$

We substitute this into (27) and we shift the contour to $L_4 = -c_0(\log U)^{-1} + it$, for $|t| \leq U/2$, which gives us an error term of size $O(\exp(-c\sqrt{\log N}))$ again. We have a new pole at $s_2 = 0$ and we use Cauchy's residue formula to obtain

$$\mathcal{T} = \underset{s_{1}=0}{\operatorname{Res}} \underset{s_{2}=0}{\operatorname{Res}} + O((\log N)^{k+l})$$

= $\frac{1}{(2\pi i)^{2}} \int_{C_{2}} \int_{C_{1}} \frac{Z(s_{1}, s_{2})R^{s_{1}+s_{2}}}{(s_{1}+s_{2})^{k}(s_{1}s_{2})^{l+1}} ds_{1} ds_{2} + O((\log N)^{k+l}),$ (28)

where we denote by C_1 and C_2 the circles $|s_1| = \rho$ and $|s_2| = 2\rho$, respectively, where $\rho > 0$ is chosen to be small. We write $s_1 = s$ and $s_2 = \xi s$ and substitute into the main term of (28) to obtain

$$\mathcal{T} = \frac{1}{(2\pi i)^2} \int_{C_3} \int_{C_1} \frac{Z(s,\xi s) R^{s(\xi+1)}}{(\xi+1)^k \xi^{l+1} s^{k+2l+1}} ds d\xi + O((\log N)^{k+l}),$$

where C_3 is the circle $|\xi = 2|$. We first evaluate

$$I_1 = \frac{1}{2\pi i} \int_{C_1} \frac{Z(s, s\xi) R^{s(\xi+1)}}{s^{k+2l+1}} ds$$

By Cauchy's integral formula, we have

$$I_1 = \frac{1}{(k+2l)!} \left(\frac{\partial}{\partial s}\right)_{s=0}^{k+2l} (Z(s,\xi s)R^{s(\xi+1)}).$$

We note that

$$\left. \frac{\partial^m Z(s,\xi s)}{\partial s^m} \right|_{s=0} = \frac{m!}{2\pi i} \int_{C(0,\delta)} \frac{Z(s,\xi s)}{s^{m+1}} ds = O((\log \log N)^c),$$

where $C(0, \delta)$ is the circle centered at 0 of radius δ and we obtain the asymptotic by choosing $\delta = 1/(\log \log N)$. On the other hand,

$$\left. \frac{\partial^m R^{s(\xi+1)}}{\partial s^m} \right|_{s=0} = (\log R)^m (\xi+1)^m,$$

and we can use Leibniz' rule for differentiation to obtain

$$\mathcal{T} = \frac{Z(0,0)}{2\pi i (k+2l)!} (\log R)^{k+2l} \int_{C_3} \frac{(\xi+1)^{2l}}{\xi^{l+1}} d\xi + O((\log N)^{k+2l-1} (\log \log N)^c).$$

We estimate the integral

$$I_2 = \frac{1}{2\pi i} \int_{C_3} \frac{(\xi+1)^{2l}}{\xi^{l+1}} d\xi = \frac{1}{l!} \times \frac{\partial^l (\xi+1)^{2l}}{\partial \xi^l} \bigg|_{\xi=0} = \binom{2l}{l}.$$

We obtain

$$\mathcal{T} = \frac{\mathfrak{G}(\mathcal{H})}{(k+2l)!} \binom{2l}{l} (\log R)^{k+2l} + O((\log N)^{k+2l-1} (\log \log N)^c)$$
(29)

Observe that the estimate we obtained in (24) gets absorbed into the error term given by (29) when we multiply through by N, because of the initial conditions imposed on R and N. Hence, the proof is complete. \Box

We next want to introduce weights for primes in our calculations. We define the function $\varpi(\cdot)$ to be $\varpi(n) := \log n$ for n prime, and 0 otherwise. We now want to evaluate the quantity

$$\sum_{N < n \le 2N} \varpi(n+h) \Lambda_R(n; \mathcal{H}, k+l)^2,$$
(30)

where h is an arbitrary positive integer less than or equal to H. We claim that the following holds.

Lemma 3.3. Suppose that the following assumption holds: There exists an absolute constant $0 < \theta < 1$ such that, for any fixed A > 0, (22) holds. Then, with the notation from Subsection 2.4, for $R \leq N^{\theta/2}$, we have

$$\sum_{N < n \le 2N} \overline{\varpi}(n+h) \Lambda_R(n; \mathcal{H}, k+l)^2$$

$$= \begin{cases} \frac{\mathfrak{G}(\mathcal{H} \cup \{h\})}{(k+2l)!} \binom{2l}{l} N(\log R)^{k+2l} \\ + O(N(\log N)^{k+2l-1} (\log \log N)^c) & \text{if } h \notin \mathcal{H}, \\ \frac{\mathfrak{G}(\mathcal{H})}{(k+2l+1)!} \binom{2(l+1)}{l+1} N(\log R)^{k+2l+1} \\ + O(N(\log N)^{k+2l} (\log \log N)^c) & \text{if } h \in \mathcal{H}. \end{cases}$$

Sketch of proof of Lemma 3.3. First, note that we may assume that $h \notin \mathcal{H}$ since the factor (n + h) does not affect the computation of $\Lambda_R(n; \mathcal{H}; k + l)$. Then, let $\delta(x) = 1$ if x = 1 and 0 otherwise. As before, we expand the square and write (30) as

$$\sum_{d_1,d_2} \lambda_R(d_1;k+l)\lambda_R(d_2;k+l) \times \sum_{b \in \Omega([d_1,d_2])} \delta((b+h,[d_1,d_2]))\vartheta^*(N;b+h,[d_1,d_2]),$$

with an error term of size $O(R^2(\log N)^c)$, which we can obtain in the same way as (24). Note that we introduced the function $\delta(\cdot)$ in the inner sum because $\vartheta^*(N; b+h, [d_1, d_2]) = 0$ when b+h and $[d_1, d_2]$ are not coprime.

We now use this and (22) to write our quantity as

$$\mathcal{NT}^* + O\left(\frac{N}{(\log N)^{A/3}}\right),$$

where

$$\mathcal{T}^* = \sum_{d_1, d_2} \frac{\lambda_R(d_1; k+l)\lambda_R(d_1; k+l)}{\phi([d_1, d_2])} \times \sum_{b \in \Omega([d_1, d_2])} \delta((b+h, [d_1, d_2])).$$

We obtain the error term by comparing $\Omega(\cdot)$ to the generalised divisor function $\tau_k(\cdot)$ again.

To evaluate the main term, we write the inner sum in the expression of \mathcal{T}^* as

$$\prod_{p \mid [d_1, d_2]} \left(\sum_{b \in \Omega(p)} \delta((b+h, p)) \right) = \prod_{p \mid [d_1, d_2]} (|\Omega^+(p)| - 1),$$

where $\Omega^+(\cdot)$ is defined to be the function $\Omega(\cdot)$ for the set $\mathcal{H} \cup \{h\}$. This is because $\delta((b+h,p)) = 0 \iff -h \in \Omega(p)$, meaning that $\delta((b+h,p)) = 1$ for each residue class from \mathcal{H} except for one.

We use (25) again to get

$$\mathcal{T}^* = \frac{1}{(2\pi i)^2} \int_{(1)} \int_{(1)} F^*(s_1, s_2; \Omega) \frac{R^{s_1 + s_2}}{(s_1 s_2)^{k+l+1}} ds_1 ds_2,$$

with

$$F^*(s_1, s_2; \Omega) = \sum_{d_1, d_2} \mu(d_1) \mu(d_2) \frac{\prod_{p \mid [d_1, d_2]} (|\Omega^+(p)| - 1)}{\phi([d_1, d_2]) d_1^{s_1} d_2^{s_2}},$$

We can write $F^*(\cdot)$ equivalently as

$$F^*(s_1, s_2; \Omega) = \prod_p \left(1 - \frac{|\Omega^+(p)| - 1}{p - 1} \left(\frac{1}{p^{s_1}} + \frac{1}{p^{s_2}} - \frac{1}{p^{s_1 + s_2}} \right) \right),$$

in the same region of absolute convergence as that of F.

We define

$$G^*(s_1, s_2; \Omega) = F^*(s_1, s_2; \Omega) \left(\frac{\zeta(s_1 + 1)\zeta(s_2 + 1)}{\zeta(s_1 + s_2 + 1)}\right)^k,$$

and we consider two cases, according to whether $h \notin \mathcal{H}$ or $h \in \mathcal{H}$.

If $h \notin \mathcal{H}$, for p > H we have $|\Omega^+(p)| = k + 1$. If \mathcal{H}^+ is admissible, we have, as before, $G^*(0,0;\Omega) = \mathfrak{G}(\mathcal{H}^+)$, and we can estimate

$$\mathcal{T}^* = \frac{\mathfrak{G}(\mathcal{H}^+)}{(k+2l)!} \binom{2l}{l} (\log R)^{k+2l} + O((\log N)^{k+2l-1} (\log \log N)^c).$$

On the other hand, if \mathcal{H}^+ is not admissible, we have $\mathfrak{G}(\mathcal{H}^+) = 0$ and the main term in the previous estimate vanishes, while the error term remains the same.

If $h \in \mathcal{H}$, the above calculations hold with the translation $k \mapsto k-1, l \mapsto l+1$. This is because for a prime p > H, $|\Omega^+(p)|$ is equal to k instead of k+1 and (k-1) + (l+1) = k+l. Hence, the proof is complete. \Box

We are now ready to bring everything together. We want to evaluate the expression

$$\sum_{\substack{\mathcal{H} \subseteq [1,H] \\ |\mathcal{H}|=k}} \sum_{N < n \le 2N} \left(\sum_{h \le H} \varpi(n+h) - \log 3N \right) \times \lambda_R(n; \mathcal{H}, k+l)^2,$$
(31)

We want to prove that it is positive for sufficiently large N, using Lemmata 3.2 and 3.3, so we set $R = N^{\theta/2}$. If this turns out to be true, then the inner sum must be positive, meaning that there exists an integer $n \in (N, 2N]$ such that

$$\sum_{h \le H} \varpi(n+h) - \log 3N > 0.$$

This in turns implies that there exists a subinterval of length H in (N, 2N+H] which contains at least two primes. That is because, if there are no primes, then the sum takes a negative value and if there is only one prime p, say, occurring in the interval, then we have

$$\log p - \log 3N \le \log \left(2N + H\right) - \log 3N.$$

We know from our initial conditions that $H \ll \log N < N$, so 2N + H < 3H, making our expression negative. If, however, there were at least two primes in the subinterval, we would have

$$\min_{N < p_r \le 2N+H} (p_{r+1} - p_r) \le H.$$
(32)

In order to prove that (31) is positive, we need to quote the following result from [6]:

$$\sum_{\substack{\mathcal{H} \subseteq [1,H] \\ |\mathcal{H}|=k}} \mathfrak{G}(\mathcal{H}) = (1+o(1))H^k,$$
(33)

as H tends to infinity. Using this and Lemma 3.2, (31) becomes

$$\sum_{\substack{\mathcal{H} \subseteq [1,H] \\ |\mathcal{H}|=k}} \sum_{\substack{N < n \le 2N \\ h \notin \mathcal{H}}} \left\{ \sum_{\substack{h \le H \\ h \notin \mathcal{H}}} + \sum_{\substack{h \le H \\ h \in \mathcal{H}}} \right\} \times \varpi(n+h) \Lambda_R(n;\mathcal{H},k+l)^2$$
$$- \frac{1}{(k+2l)!} \binom{2l}{l} N H^k (\log N) (\log R)^{k+2l} + o(NH^k (\log N)^{k+2l+1}).$$

Using Lemma 3.3 and (33) again, this is asymptotically equal to

$$\begin{aligned} &\frac{1}{(k+2l)!} \binom{2l}{l} N H^{k+1} (\log R)^{k+2l} \\ &+ \frac{k}{(k+2l+1)!} \binom{2(l+1)}{l+1} N H^k (\log R)^{k+2l+1} \\ &- \frac{1}{(k+2l!)} \binom{2l}{l} N H^k (\log N) (\log R)^{k+2l} \\ &= \left\{ H + \frac{k}{k+2l+1} \cdot \frac{2(2l+1)}{l+1} \cdot \log R - \log N \right\} \\ &\times \frac{1}{(k+2l)!} \binom{2l}{l} N H^k (\log R)^{k+2l}. \end{aligned}$$

So (31) is positive as long as

$$\frac{H}{\log N} \ge 1 + \varepsilon - \frac{k}{k+2l+1} \cdot \frac{2(2l+1)}{l+1} \cdot \frac{\theta}{2},$$

for any $\varepsilon > 0$. If we choose $l = \lfloor \sqrt{k} \rfloor$, it is enough to require

$$\frac{H}{\log N} \ge 1 + \varepsilon - 2\theta.$$

By Bombieri–Vinogradov theorem, we can take $\theta = 1/2 - \epsilon$ for any $\epsilon > 0$, so we want

$$\frac{H}{\log N} \ge 1 + \varepsilon - 1 + 2\epsilon.$$

Choose $\epsilon = \varepsilon$ to get that we need

$$\frac{H}{\log N} \ge 3\varepsilon. \tag{34}$$

Going back to (32), we have

$$\min_{N < p_r \le 2N+H} \frac{p_{r+1} - p_r}{\log p_r} < \min_{N < p_r \le 2N+H} \frac{p_{r+1} - p_r}{\log N} \le \frac{H}{\log N},$$
(35)

We know from our initial conditions that $H = O(\log N)$, so we can pick any $\delta > 0$ and pick $H = \delta \log N$. Then, choose $\varepsilon = \delta/3$, so that (34) holds. Also, (35) becomes

$$\min_{N < p_r \le 2N+H} \frac{p_{r+1} - p_r}{\log N} \le \delta,$$

which in turn proves the theorem.

Furthermore, under the Elliott–Halberstam conjecture, Goldston *et al.* use the admissible 6-tuple $\{7, 11, 13, 17, 19, 23\}$ to show

$$\liminf_{n \to \infty} (p_{n+1} - p_n) \le 23 - 7 = 16.$$

4 The Maynard theorems

We now look at Maynard's work, which revolves around the following conjecture:

Conjecture 4.1 (Prime k-tuples conjecture). Let $\mathcal{H} = \{h_1, \ldots, h_k\}$ be admissible. Then there are infinitely many integers n such that all of $n + h_1, \ldots, n + h_k$ are prime.

He proves the following theorems:

Theorem 4.2.

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

holds for any integer m.

This shows that there are infinitely many intervals of finite length that contain at least m + 1 primes.

Theorem 4.3. Let $m \in \mathbb{N}$. Let $r \in \mathbb{N}$ be sufficiently large depending on m, and let $\mathscr{A} = \{a_1, a_2, \ldots, a_r\}$ be a set of r distinct integers. Then

$$\frac{\#\{\{h_1,\ldots,h_m\}\subseteq\mathscr{A}: \text{ for infinitely many } n \text{ all of } n+h_i \text{ are prime}\}}{\#\{\{h_1,\ldots,h_m\}\subseteq\mathscr{A}\}} \gg_m 1.$$

In other words, a positive proportion of admissible m-tuples satisfy the prime m-tuple conjecture for every m.

Theorem 4.4. We have

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

We note that the proof of this last theorem is based on the Bombieri– Vinogradov theorem, namely on the fact that primes have level of distribution θ for every $0 < \theta < 1/2$. If, however, we assume that primes have level of distribution θ for every $\theta < 1$, we can prove the following:

Theorem 4.5. Under the assumption that primes have level of distribution θ for every $\theta < 1$, we have

$$\liminf_{n} (p_{n+1} - p_n) \le 12,$$
$$\liminf_{n} (p_{n+2} - p_n) \le 600.$$

To prove these results, Maynard uses an improved GPY sieve method, by looking at the following setting: fix k > 0 and fix an admissible set $\mathcal{H} = \{h_1, \ldots, h_k\}$ of size k. Consider the quantity

$$S(N,\rho) = \sum_{N \le n < 2N} \left(\sum_{i=1}^{k} \chi_{\mathbb{P}}(n+h_i) - \rho \right) w_n, \tag{36}$$

where $\chi_{\mathbb{P}}(\cdot)$ is the characteristic function of the primes, *i.e.* $\chi_{\mathbb{P}}(n) = 1$ if n is prime and 0 otherwise, $\rho > 0$ and w_n are non-negative weights. As in Theorem 3.1, we want to prove that this quantity is positive. In that case, at least one term in the sum over n must be positive. Since w_n is positive, we must have that at least $\lfloor \rho + 1 \rfloor$ of the $n + h_i$ are prime. So, as we let $N \to \infty$, we get that there are infinitely many bounded length intervals containing $\lfloor \rho + 1 \rfloor$ primes (bounded by $\max_k |h_k|$).

Unlike the weights $\Lambda_R(n; \mathcal{H}, a)$ used in Theorem 3.1, we consider our weights to be of the form

$$w_n = \left(\sum_{d_i|n+h_i \forall i} \lambda_{d_1,\dots,d_k}\right)^2.$$
(37)

We will choose our λ_{d_1,\ldots,d_k} to look like

$$\lambda_{d_1,\dots,d_k} \approx \left(\prod_{i=1}^k \mu(d_i)\right) f(d_1,\dots,d_k),$$

for a suitable smooth function $f(\cdot)$, which is a multi-dimensional generalisation of our choice for $\lambda_R(d; a)$ in the previous section.

We begin by introducing some notation. For convenience, we choose our weights w_n to be 0 unless $n \equiv \nu_0 \pmod{W}$, where ν_0 is a fixed residue class (mod W), and $W = \prod_{p < D_0} p$, where it suffices to choose

$$D_0 = \log \log \log N.$$

Then,

$$\log W = \sum_{p \le D_0} \log p = \vartheta(D_0),$$

which, by calculations in [1] (p.8), is $O(D_0)$. Hence, $W = O(\log \log N)$. Since \mathcal{H} is admissible, we can choose ν_0 such that $\nu_0 + h_i$ is coprime to W, by applying the Chinese remainder theorem to the primes dividing W.

When $n \equiv \nu_0 \pmod{W}$, we define our weights as in (37).

In order to estimate $S(N, \rho)$, we look at the following two sums

$$S_1 = \sum_{\substack{N \le n < 2N \\ n \equiv \nu_0 \pmod{W}}} \left(\sum_{d_i \mid n+h_i \forall i} \lambda_{d_1,\dots,d_k} \right)^2, \tag{38}$$

$$S_2 = \sum_{\substack{N \le n < 2N \\ n \equiv \nu_0 \pmod{W}}} \left(\sum_{i=1}^k \chi_{\mathbb{P}}(n+h_i) \right) \left(\sum_{d_i \mid n+h_i \forall i} \lambda_{d_1,\dots,d_k} \right)^2, \tag{39}$$

as we did with (23) and (30) in Theorem 3.1.

4.1 Main proposition

We aim to prove the following:

Proposition 4.6. Let the primes have level of distribution $\theta > 0$ and let $R = N^{\theta/2-\delta}$ for some small fixed $\delta > 0$. Let λ_{d_1,\ldots,d_k} be defined in terms of a smooth function $F(\cdot)$ by

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

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

$$S_1 = \frac{(1+o(1))\phi(W)^k N(\log R)^k}{W^{k+1}} I_k(F),$$

$$S_2 = \frac{(1+o(1))\phi(W)^k N(\log R)^{k+1}}{W^{k+1}\log N} \sum_{m=1}^k J_k^{(m)}(F)$$

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

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

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

Remark 13. Note that in the expressions for S_1 and S_2 there is no dependency on the actual elements of the admissible set \mathcal{H} , only on its size, whereas in Lemma 3.2 and Lemma 3.3, the singular series appears in the main term, which clearly depends on the h_i 's.

4.1.1 Selberg sieve manipulations

Our aim is to introduce a change of variables to rewrite S_1 and S_2 in a simpler form.

We restrict the support of λ_{d_1,\ldots,d_k} to tuples for which the product $d := \prod_{i=1}^k d_i$ is less than R, similarly to how we had $\lambda_R(d; a) = 0$ for d > R in the previous section. We also demand (d, W) = 1 and $\mu(d)^2 = 1$, implying that d is squarefree and so $(d_i, d_j) = 1$ for all $i \neq j$. We want to prove the following lemma:

Lemma 4.7. Let

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

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

$$S_1 = \frac{N}{W} \sum_{r_1,\dots,r_k} \frac{y_{r_1,\dots,r_k}^2}{\prod_{i=1}^k \phi(r_i)} + O\left(\frac{y_{max}^2 N(\log R)^k}{WD_0}\right).$$
 (40)

Remark 14. This change of variables is a multi-dimensional generalisation of the change of variables (9) we performed on the quadratic form (λ_d) in Subsection 2.2.

Proof. We expand the square and swap the order of summation in (38) to obtain

$$S_1 = \sum_{\substack{d_1, \dots, d_k \\ e_1, \dots, e_k}} \lambda_{d_1, \dots, d_k} \lambda_{e_1, \dots, e_k} \sum_{\substack{N \le n < 2N \\ n \equiv \nu_0 \pmod{W} \\ [d_i, e_i] | n + h_i \forall i}} 1.$$

We can use the Chinese remainder theorem again, this time to write the inner sum as a sum over a single residue class modulo $q = W \prod_{i=1}^{k} [d_i, e_i]$, as long as $W, [d_1, e_1], \ldots, [d_k, e_k]$ are pairwise coprime. In that case, we have that the inner sum is N/q + O(1). If $W, [d_1, e_1], \ldots, [d_k, e_k]$ are not pairwise coprime, then the inner sum has no contribution: suppose some $[d_i, e_i]$ has a common factor $a \neq 1$, say, with W. Then $a \mid n + h_i$ and $n - \nu_0 \equiv 0 \pmod{a}$, implying $a \mid n - \nu_0$. So $a \mid h_i + \nu_0$, contradicting the coprimality condition between W and $h_i + \nu_0$. On the other hand, if some $[d_i, e_i]$ and $[d_j, e_j]$ have a common factor $b \neq 1$, say, then $n + h_i$ and $n + h_j$ are both divisible by b, contradicting the fact that they are distinct primes. Hence, we have

$$S_{1} = \frac{N}{W} \sum_{\substack{d_{1},...,d_{k} \\ e_{1},...,e_{k}}} \frac{\lambda_{d_{1},...,d_{k}} \lambda_{e_{1},...,e_{k}}}{\prod_{i=1}^{k} [d_{i}, e_{i}]} + O\left(\sum_{\substack{d_{1},...,d_{k} \\ e_{1},...,e_{k}}} |\lambda_{d_{1},...,d_{k}} \lambda_{e_{1},...,e_{k}}|\right),$$

where \sum' denotes the coprimality restriction on $W, [d_1, e_1], \ldots, [d_k, e_k]$. We know λ_{d_1,\ldots,d_k} is non-zero only when $\prod_{i=1}^k d_i < R$, so if we denote by $\lambda_{\max} = \sup_{d_1,\ldots,d_k} |\lambda_{d_1,\ldots,d_k}|$, we can estimate the error term as

$$O\left(\lambda_{\max}^2\left(\sum_{d< R} \tau_k(d)\right)^2\right) = O(\lambda_{\max}^2 R^2(\log R)^{2k-2}),\tag{41}$$

by the exact same argument as in Section 3.1.

To deal with the main term, we want to remove the dependencies between the d_i and the e_j variables. We start by using once again the fact that $d_i, e_i = d_i e_i$ and that $\sum_{\delta \mid d} \phi(\delta) = d$ to write

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

so that our main term becomes

$$\frac{N}{W}\sum_{u_1,\dots,u_k} \left(\prod_{i=1}^k \phi(u_i)\right) \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k \\ u_i \mid (d_i,e_i) \forall i}} \frac{\lambda_{d_1,\dots,d_k} \lambda_{e_1,\dots,e_k}}{d_1 (\prod_{i=1}^k d_i) (\prod_{i=1}^k e_i)}$$

We know that λ_{d_1,\dots,d_k} is non-zero only when $(W,\prod_{i=1}^k d_i) = 1$, so we may drop from the summation the requirement that W is coprime to each $[d_i, e_i]$, and we also know that we can assume that the d_i (and the e_i) are all pairwise coprime for the same reason. Thus we are only left with the condition that $(d_i, e_j) = 1$ for all $i \neq j$, which we can remove by multiplying our expression by $\sum_{s_{i,j}|(d_i, e_j)} \mu(s_{i,j})$, since (4) holds. Our main term then becomes

$$\frac{N}{W}\sum_{u_1,\dots,u_k} \left(\prod_{i=1}^k \phi(u_i)\right) \sum_{\substack{s_{1,2},\dots,s_{k,k-1}\\i\neq j}} \left(\prod_{\substack{1\leq i,j\leq k\\i\neq j}} \mu(s_{i,j})\right) \sum_{\substack{d_1,\dots,d_k\\e_1,\dots,e_k\\u_i|(d_i,e_i)\forall i\\s_{i,j}|(d_i,e_j)\forall i\neq j}} \frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{(\prod_{i=1}^k d_i)(\prod_{i=1}^k e_i)}.$$

We can restrict the $s_{i,j}$ to be coprime to u_i and u_j because otherwise one of the pairs $\{d_i, d_j\}, \{e_i, e_j\}$ would have a common factor, making the corresponding λ vanish. By a similar argument, we can restrict our sum so that $s_{i,j}$ is coprime to $s_{i,a}$ and $s_{b,j}$ for all $a \neq j$ and $b \neq i$. We denote the summation over the $s_{i,j}$ with these restrictions by \sum^* .

We now want to introduce a change of variables, which will make it easier for us to diagonalise our quadratic form.

$$y_{r_1,...,r_k} = \left(\prod_{i=1}^k \mu(r_i)\phi(r_i)\right) \sum_{\substack{d_1,...,d_k \\ r_i \mid d_i \forall i}} \frac{\lambda_{d_1,...,d_k}}{\prod_{i=1}^k d_i}.$$
 (42)

This change of variables is in fact invertible, since

$$\sum_{\substack{r_1,\dots,r_k\\d_i|r_i\forall i}} \frac{y_{r_1,\dots,r_k}}{\prod_{i=1}^k \phi(r_i)} = \sum_{\substack{r_1,\dots,r_k\\d_i|r_i\forall i}} \left(\prod_{i=1}^k \mu(r_i)\right) \sum_{\substack{e_1,\dots,e_k\\r_i|e_i\forall i}} \frac{\lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k e_i} \sum_{\substack{r_1,\dots,r_k\\d_i|r_i\forall i\\r_i|e_i\forall i}} \frac{\lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k e_i} \mu(r_i)$$

We take a closer look at the inner sum. We have

$$\sum_{\substack{r_1,\dots,r_k\\d_i|r_i\forall i\\r_i|e_i\forall i}} \prod_{i=1}^k \mu(r_i) = \sum_{\substack{r_1,\dots,r_k\\(r_i/d_i)|(e_i/d_i)\forall i}} \prod_{i=1}^k \mu\left(\frac{r_i}{d_i}\right) \mu(d_i)$$
$$= \prod_{i=1}^k \mu(d_i) \sum_{\substack{m_1,\dots,m_k\\m_i|n_i\forall i}} \prod_{i=1}^k \mu(m_i) = \prod_{i=1}^k \mu(d_i) \prod_{i=1}^k \sum_{m_i|n_i\forall i} \mu(m_i),$$

where we let m_i to be r_i/d_i and n_i to be e_i/d_i . The inner product is equal to 1 if $n_i = 1$ for all *i* and 0 otherwise, or in other words it is equal to 1 when $e_i = d_i$ and 0 otherwise. Thus, we obtain

$$\sum_{\substack{r_1,\dots,r_k\\d_i|r_i\forall i}} \frac{y_{r_1,\dots,r_k}}{\prod_{i=1}^k \phi(r_i)} = \frac{\lambda_{d_1,\dots,d_k}}{\prod_{i=1}^k \mu(d_i)d_i}$$
(43)

since $1/\mu(n) = \mu(n)$ holds for all n.

Since $r_i \mid d_i$ for all *i*, bearing in mind our conditions on the d_i , we have that y_{r_1,\ldots,r_k} are non-zero only when the following conditions hold: the r_i are pairwise coprime, $(r_i, W) = 1$ for all *i* and $\prod_{i=1}^k r_i < R$.

We now want to bound λ_{\max} asymptotically as a function of y_{\max} , where we let $y_{\max} = \sup_{r_1,\ldots,r_k} |y_{r_1,\ldots,r_k}|$. We take $r = \prod_{i=1}^k m_i$ and we use the fact that $|\mu(\cdot)| = \mu^2(\cdot), \ \mu(n)^2 = 1$ if n is squarefree and 0 otherwise and that $n = \sum_{d|n} \phi(d) \iff n = \phi(n) \sum_{e|n} 1/\phi(e)$ to write

$$\begin{split} \lambda_{\max} &\leq \sup_{\substack{d_1, \dots, d_k \\ \prod_{i=1}^k d_i \text{ squarefree}}} y_{\max} \left(\prod_{i=1}^k d_i \right) \sum_{\substack{r_1, \dots, r_k \\ d_i | r_i \forall i \\ \prod_{i=1}^k r_i \text{ squarefree}}} \left(\prod_{i=1}^k \frac{\mu(r_i)^2}{\phi(r_i)} \right) \\ &\leq y_{\max} \sup_{\substack{d_1, \dots, d_k \\ \prod_{i=1}^k d_i \text{ squarefree}}} \left(\prod_{i=1}^k \frac{d_i}{\phi(d_i)} \right) \sum_{\substack{r < R / \prod_{i=1}^k d_i \\ (r, \prod_{i=1}^k d_i) = 1}} \frac{\mu(r)^2 \tau_k(r)}{\phi(r)} \\ &\leq y_{\max} \sup_{d_1, \dots, d_k} \sum_{d \mid \prod_{i=1}^k d_i} \frac{\mu(d)^2}{\phi(d)} \sum_{\substack{r < R / \prod_{i=1}^k d_i \\ (r, \prod_{i=1}^k d_i) = 1}} \frac{\mu(r)^2 \tau_k(r)}{\phi(r)} \\ &\leq y_{\max} \sum_{u < R} \frac{\mu(u)^2 \tau_k(u)}{\phi(u)}, \end{split}$$

where in the last line we took u = dr and used the fact that $\tau_k(dr) \geq \tau_k(r)$ by definition. We now take a closer look at the inner sum without the Möbius function. We know from the previous section that $\sum_{u < R} \tau_k(u) \sim c R(\log R)^{k-1}$, for some constant c. We have

$$\sum_{u < R} \frac{\tau_k(u) \log u}{u} \sim c \ R(\log R)^{k-1} \frac{\log R}{R} + \int_2^R \frac{\log t - 1}{t^2} c \ t(\log t)^{k-1} dt$$
$$\leq c \ (\log R)^k + c \int_2^R \frac{\log t}{t} (\log t)^{k-1} dt$$
$$\leq c \ (\log R)^k + c' (\log R)^{k+1}.$$

We have $\phi(u) = u \prod_{p|u} (1 - p^{-1})$ and so

$$\frac{1}{\phi(u)} = \frac{1}{u} \prod_{p|u} (1 - p^{-1})^{-1} \ll \frac{1}{u} \prod_{p|u} (1 + p^{-1})$$
$$= \frac{1}{u} \sum_{d|u} \frac{1}{d} = \frac{1}{u^2} \sum_{d|u} d \ll \frac{1}{u^2} u \log u,$$

Where in the last step we used the following theorem from [2]:

$$\sigma(u) := \sum_{d|u} d \ll u \log u. \tag{44}$$

So $1/\phi(u) \ll \log u/u$ and

$$\sum_{u < R} \frac{\tau_k(u)}{\phi(u)} \ll \sum_{u < R} \frac{\tau_k(u) \log u}{u} \ll (\log R)^{k+1}.$$

Therefore, the error term given by (41) is $O(y_{\max}^2 R^2 (\log N)^{4k})$.

We use this and we substitute our change of variables (42) into the main term to obtain

$$S_{1} = \frac{N}{W} \sum_{u_{1},...,u_{k}} \left(\prod_{i=1}^{k} \phi(u_{i}) \right) \sum_{\substack{s_{1,2},...,s_{k,k-1} \\ i \neq j}} \left(\prod_{\substack{1 \le i,j \le k \\ i \neq j}} \mu(s_{i,j}) \right) \\ \times \left(\prod_{i=1}^{k} \frac{\mu(a_{i})\mu(b_{i})}{\phi(b_{i})\phi(a_{i})} \right) y_{a_{1},...,a_{k}} y_{b_{1},...,b_{k}} + O(y_{\max}^{2} R^{2}(\log R)^{4k}),$$

where $a_j = u_j \prod_{i \neq j} s_{j,i}$ and $b_j = u_j \prod_{i \neq j} s_{i,j}$. We bear in mind that we have restricted the $s_{i,j}$ to be coprime to all the other terms in the expressions for a_i and b_j . By definition, the y are equal to 0 when the a_i and the b_j are not squarefree, so we can write $\mu(a_j) = \mu(u_j) \prod_{i \neq j} (s_{j,i})$, and the same holds for $\phi(a_i), \mu(b_j)$ and $\phi(b_j)$. Therefore, we can write

$$S_{1} = \frac{N}{W} \sum_{u_{1},\dots,u_{k}} \left(\prod_{i=1}^{k} \frac{\mu(u_{i})^{2}}{\phi(u_{i})} \right) \sum_{s_{1,2},\dots,s_{k,k-1}} \left(\prod_{\substack{1 \le i,j \le k \\ i \ne j}} \frac{\mu(s_{i,j})}{\phi(s_{i,j})^{2}} \right) y_{a_{1},\dots,a_{k}} y_{b_{1},\dots,b_{k}} + O(y_{\max}^{2} R^{2} (\log R)^{4k}).$$
(45)

Bearing in mind the coprimality conditions between the a_i and b_j and W, we have that the $s_{i,j}$ must be coprime to W, otherwise the y vanish. Since

 $W = \prod_{p < D_0} p$, we only need to consider $s_{i,j} = 1$ or $s_{i,j} > D_0$. When the latter holds, the contribution from the main term in (45) is

$$\ll \frac{y_{\max}^2 N}{W} \left(\sum_{\substack{u < R\\(u,W)=1}} \frac{\mu(u)^2}{\phi(u)} \right)^k \left(\sum_{s_{i,j} > D_0} \frac{\mu(s_{i,j})^2}{\phi(s_{i,j})^2} \right) \left(\sum_{s \ge 1} \frac{\mu(s)^2}{\phi(s)^2} \right)^{k^2 - k - 1}, \quad (46)$$

where we note that the last term appears because we have k(k-1) choices for the pair (i, j) inside the innermost product of (45). We now look at each sum separately. The last sum is a convergent series (and so it is O(1)), since we know from [2] that the following holds

$$\frac{1}{2} \le \frac{\sigma(n)\phi(n)}{n^2} \implies \frac{1}{\phi(n)^2} \le 4\frac{\sigma(n)^2}{n^4} \ll \frac{n^2(\log n)^2}{n^4} = \left(\frac{\log n}{n}\right)^2,$$

where in the last step we used (44).

We now estimate the middle factor in (46), using Chen's result again:

$$\sum_{n \ge D_0} \frac{\mu(n)^2}{\phi(n)^2} \le \sum_{n \ge D_0} \frac{1}{\phi(n)^2} \le 4 \sum_{n \le D_0} \frac{\sigma(n)^2}{n^4}.$$
(47)

We know from Ramanujan's Identity (1.2.9 in [11]) with a = b = 1 that

$$\sum_{n=1}^{\infty} \frac{\sigma(n)^2}{n^s} = \frac{\zeta(s)\zeta(s-1)^2\zeta(s-2)}{\zeta(2s-2)}.$$

The right-hand side has a pole of order 1 at s = 3, so by the Tauberian theorem,

$$A(x) := \sum_{n \le x} \sigma(n)^2 \sim cx^3,$$

for some constant c. We can now use summation by parts in (47) to obtain

$$\sum_{D_0 \le n \le M} \frac{\mu(n)^2}{\phi(n)^2} \le \frac{1}{4} \left(\frac{A(M)}{M^4} - \frac{A(D_0)}{D_0^4} \right) + \int_{D_0}^M \frac{A(t)}{t^5} dt$$
$$\sim \frac{1}{4} \left(\frac{c}{M} - \frac{c'}{D_0} \right) - 2\frac{1}{t} \Big|_{D_0}^M.$$

By letting $M \to \infty$, we obtain that the middle factor in (46) is $O(1/D_0)$. Next, we want to prove:

$$\sum_{\substack{u < R\\(u,W)=1}} \frac{\mu(u)^2}{\phi(u)} \ll \frac{\phi(W)}{W} \log R.$$
(48)

First, we note that the $\mu(\cdot)$ in the numerator ensures we are working with squarefree u's. We then look at:

$$A(R) = \sum_{\substack{u < R\\(u,W)=1}} \frac{u}{\phi(u)}.$$

Given that $\phi(u) = u \prod_{p|u} (1 - p^{-1})$ holds for any u, we get

$$A(R) = \sum_{\substack{u < R \\ (u,W)=1}} \prod_{p|u} \left(1 - \frac{1}{p}\right)^{-1} = \prod_{p|W} \left(1 - \frac{1}{p}\right) \sum_{u < R} \prod_{p|u} \left(1 - \frac{1}{p}\right)^{-1}.$$
 (49)

The first product in the right-hand side of (49) is precisely $\phi(W)/W$ and what is left is $\sum_{u < R} u/\phi(u)$. It is proved by Murty (4.4.12 in [11]) that:

$$\sum_{u < R} \frac{u}{\phi(u)} \ll R$$

So we have shown:

$$A(R) \ll \frac{\phi(W)}{W}R.$$
(50)

We now do the summation by parts:

$$\sum_{\substack{u < R \\ (u,W)=1}} \frac{1}{\phi(u)} = \sum_{\substack{u < R \\ (u,W)=1}} \frac{u}{\phi(u)} \frac{1}{u}$$
$$= \frac{1}{R} A(R) + \int_{1}^{R} \frac{A(t)}{t^{2}} dt.$$

Using (50), the first term of the sum becomes O(1) and is dominated by the second one, which is:

$$\ll \int_{1}^{R} \frac{\phi(W)}{W} \frac{t}{t^{2}} dt = \frac{\phi(W)}{W} \int_{1}^{R} \frac{1}{t} dt = \frac{\phi(W)}{W} \log R.$$

Bringing everything together, (46) can be estimated as

$$O\left(\frac{y_{\max}^2\phi(W)^k N(\log R)^k}{W^{k+1}D_0}\right)$$

Going back to our main estimate, we are left to consider the case when $s_{i,j} = 1$ for all $i \neq j$. We have $a_i = b_i$ and the u_i are squarefree, so $\mu(u_i)^2 = 1$. Therefore

$$S_1 = \frac{N}{W_{u_1,\dots,u_k}} \frac{y_{u_1,\dots,u_k}^2}{\prod_{i=1}^k \phi(u_i)} + O\left(\frac{y_{\max}^2 \phi(W)^k N(\log R)^k}{W^{k+1} D_0} + y_{\max}^2 R^2 (\log R)^{4k}\right)$$
(51)

We have $R = N^{\theta/2-\delta}$, so $R^2 = N^{\theta-2\delta} \le N^{1-2\delta}$, since $0 < \theta \le 1$. We also have $\phi(W)/W < 1$ and

$$W = \exp(\vartheta(D_0)) \implies W \ge \exp(c \log \log \log N) = (\log \log N^c),$$

where to obtain the inequality we used that $\theta(n) \leq 2n \log 2$ (for proof see p.8 in [1]). So the first error term dominates. Since $\phi(W)/W < 1$, the lemma is proved.

We now want to estimate for S_2 . We write $S_2 = \sum_{m=1}^k S_2^{(m)}$, where

$$S_{2}^{(m)} = \sum_{\substack{N \le n < 2N \\ n \equiv \nu_{0} \pmod{W}}} \chi_{\mathbb{P}}(n+h_{m}) \left(\sum_{\substack{d_{1}, \dots, d_{k} \\ d_{i}|n+h_{i} \forall i}} \lambda_{d_{1}, \dots, d_{k}}\right)^{2}.$$
 (52)

We want to estimate $S_2^{(m)}$ in a similar way to how we estimated S_1 . We prove the following lemma:

Lemma 4.8. Let

$$y_{r_1,\dots,r_k}^{(m)} = \left(\prod_{i=1}^k \mu(r_i)g(r_i)\right) \sum_{\substack{d_1,\dots,d_k \\ r_i | d_i \forall i \\ d_m = 1}} \frac{\lambda_{d_1,\dots,d_k}}{\prod_{i=1}^k \phi(d_i)},$$
(53)

where $g(\cdot)$ is the totally multiplicative function defined on primes by g(p) = p - 2. Let $y_{\max}^{(m)} = \sup_{r_1, \dots, r_k} |y_{r_1, \dots, r_k}^{(m)}|$. Then for any fixed A > 0 we have

$$S_{2}^{(m)} = \frac{N}{\phi(W) \log N} \sum_{r_{1},...,r_{k}} \frac{(y_{r_{1},...,r_{k}}^{(m)})^{2}}{\prod_{i=1}^{k} g(r_{i})} + O\left(\frac{(y_{\max}^{(m)})^{2} \phi(W)^{k-2} N(\log N)^{k-2}}{W^{k-1} D_{0}}\right) + O\left(\frac{y_{\max}^{2} N}{(\log N)^{A}}\right).$$
(54)

Proof. We expand the square and swap the order of summation in (52) to obtain

$$S_2^{(m)} = \sum_{\substack{d_1, \dots, d_k \\ e_1, \dots, e_k}} \lambda_{d_1, \dots, d_k} \lambda_{e_1, \dots, e_k} \sum_{\substack{N \le n < 2N \\ n \equiv \nu_0 \pmod{W} \\ [d_i, e_i] | n + h_i \forall i}} \chi_{\mathbb{P}}(n+h_m).$$

As before, we can write the inner sum as a sum over a single residue class modulo $q = W \prod_{i=1}^{k} [d_i, e_i]$, as long as $W, [d_1, e_1], \ldots, [d_k, e_k]$ are pairwise coprime. If either one pair of $W, [d_1, e_1], \ldots, [d_k, e_k]$ have a common factor,

then the inner sum has no contribution for the same reasons as in the previous lemma. In that case, we will have a contribution in the inner sum only when $n+h_m$ is prime, i.e. when it will lie in a residue class coprime to the modulus. This happens if and only if $d_m = e_m = 1$. The inner sum will then contribute $X_N/\phi(q) + O(E(N,q))$, where

$$E(N,q) = 1 + \sup_{\substack{(a,q)=1}} \bigg| \sum_{\substack{N \le n < 2N \\ n \equiv a \pmod{q}}} \chi_{\mathbb{P}}(n) - \frac{1}{\phi(q)} \sum_{\substack{N \le n < 2N}} \chi_{\mathbb{P}}(n) \bigg|,$$

is an alternative definition for the level of distribution of primes and

$$X_N = \sum_{N \le n < 2N} \chi_{\mathbb{P}}(n).$$

Therefore we have

$$S_2^{(m)} = \frac{X_N}{\phi(W)} \sum_{\substack{d_1, \dots, d_k \\ e_1, \dots, e_k \\ e_m = d_m = 1}} \frac{\lambda_{d_1, \dots, d_k} \lambda_{e_1, \dots, e_k}}{\prod_{i=1}^k \phi([d_i, e_i])} + O\left(\sum_{\substack{d_1, \dots, d_k \\ e_1, \dots, e_k}} |\lambda_{d_1, \dots, d_k} \lambda_{e_1, \dots, e_k}| E(N, q)\right),$$

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

We now estimate the error term. We know that the λ_{d_1,\ldots,d_k} are nonzero only when $q < WR^2$. Then, given a squarefree integer r, we have at most $\tau_{3k}(r)$ choices of $d_1,\ldots,d_k,e_1,\ldots,e_k$ for which $W\prod_{i=1}^k [d_i,e_i] = r$, since $[d_i,e_i]$ depends on d_i,e_i and (d_i,e_i) . We use this and the fact that $\lambda_{\max} \ll y_{\max} \log R^{k+1}$ to estimate the error term as

$$\ll y_{\max} \log R^{2k+2} \sum_{r < R^2 W} \mu(r)^2 \tau_{3k}(r) E(N, r).$$

Note that the Möbius function ensures we are working with squarefree r. We have the trivial bound $E(N,r) \ll N/\phi(q)$ and we also assumed that the primes have level of distribution θ , so we can use the Cauchy-Schwarz inequality to obtain the estimate the error term as

$$\ll y_{\max}^2 (\log R)^{2k+2} \left(\sum_{r < R^2 W} \mu(r)^2 \tau_{3k}^2(r) \frac{N}{\phi(r)} \right)^{1/2} \left(\sum_{r < R^2 W} \mu(r)^2 E(N, r) \right)^{1/2},$$

which gives an error term of size $O(y_{\max}^2 N/(\log N)^A)$ for any fixed A > 0.

To estimate the main term, we remove the condition $(d_i, e_j) = 1$ by multiplying our expression by $\sum_{s_{i,j}|(d_i, e_j)} \mu(s_{i,j})$, like we did in the previous proof. We can again restrict $s_{i,j}$ to be coprime to $u_i, u_j, s_{i,a}$ and $s_{b,j}$ for all $a \neq j$ and $b \neq i$. We denote by \sum^* the summation with these restrictions. Next, we claim that

$$\frac{1}{\phi([d_i, e_i])} = \frac{1}{\phi(d_i)\phi(e_i)} \sum_{u_i \mid (d_i, e_i)} g(u_i),$$

where $g(\cdot)$ is the totally multiplicative function defined on primes by g(p) = p-2. Using (14), we know that $\phi([d, e]) = \phi(d)\phi(e)/\phi((d, e))$, for squarefree d and e. We are left to show that

$$\phi(c) = \sum_{u|c} g(u),$$

for squarefree c. Since Euler's totient function is multiplicative, it is enough to check that this is true when c is a prime. Indeed, $\phi(p) = p - 1$, while the right-hand side is equal to 1 + (p - 2).

Our main term becomes

$$\frac{X_N}{\phi(W)} \sum_{u_1,\dots,u_k} \left(\prod_{i=1}^k g(u_i) \right)_{s_{1,2},\dots,s_{k,k-1}} \sum_{1 \le i,j \le k} \mu(s_{i,j}) \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k \\ u_i|(d_i,e_i) \forall i \\ s_{i,j}|(d_i,e_j) \forall i \ne j \\ d_m = e_m = 1}} \frac{\lambda_{d_1,\dots,d_k} \lambda_{e_1,\dots,e_k}}{\lambda_{d_1,\dots,d_k} \lambda_{e_1,\dots,e_k}}$$

Let

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

Since $d_m = 1$ and $r_m \mid d_m$, r_m must be equal to 1. We substitute this into our estimate to obtain

$$\frac{X_N}{\phi(W)} \sum_{u_1,\dots,u_k} \left(\prod_{i=1}^k \frac{\mu(u_i)^2}{g(u_i)} \right) \sum_{\substack{s_{1,2},\dots,s_{k,k-1} \\ i \neq j}} \left(\prod_{\substack{1 \le i,j \le k \\ i \neq j}} \frac{\mu(s_{i,j})}{g(s_{i,j})^2} \right) y_{a_1,\dots,a_k}^{(m)} y_{b_1,\dots,b_k}^{(m)},$$

where $a_j = u_j \prod_{i \neq j} s_{j,i}$ and $b_j = u_j \prod_{i \neq j} s_{i,j}$. As before, the y are equal to 0 when the a_i and the b_j are not squarefree, so we can write $\mu(a_j) = \mu(u_j) \prod_{i \neq j} (s_{j,i})$, and the same holds for $\phi(a_i)$, $\mu(b_j)$ and $\phi(b_j)$.

Again, we consider the two cases of $s_{i,j} = 1$ and $s_{i,j} > D_0$. When the latter holds, we get a contribution which is

$$\ll \frac{(y_{\max}^{(m)})^2 N}{\phi(W) \log N} \left(\sum_{\substack{u < R\\(u,W)=1}} \frac{\mu(u)^2}{g(u)} \right)^{k-1} \left(\sum_{s_{i,j} > D_0} \frac{\mu(s_{i,j})^2}{g(s_{i,j})^2} \right) \left(\sum_s \frac{\mu(s)^2}{g(s)^2} \right)^{k^2-k-1}$$
(55)

We have, for squarefree n, $g(n) = n \prod_{p|n} (1 - 2/p)^{-1}$. We consider, again for squarefree n, the Dirichlet series

$$f(s) = \sum_{n \le x} \frac{1}{g(n)} \cdot \frac{1}{n^s} = \prod_p \left(1 + \frac{1}{p^s(p-2)} \right).$$

We then look at the quotient

$$\frac{f(s)}{\zeta(s+1)} = \prod_{p} \left(1 + \frac{1}{p^s(p-2)} \right) \left(1 - \frac{1}{p^{s+1}} \right) = \prod_{p} \left(1 + \frac{p^{s+1} - p^s(p-2) - 1}{p^{2s+1}(p-2)} \right).$$

The corresponding Dirichelt series is

$$\sum_{p} \frac{p^{s+1} - p^s(p-2) - 1}{p^{2s+1}(p-2)} = \sum_{p} \frac{2p^s - 1}{p^{2s+2}(1-2p^{-1})} \ll \sum_{p} \frac{1}{p^{s+2}},$$

which converges for $\Re(s) > -1$. Thus we may write $f(s) = \zeta(s+1)h(s)$, or equivalently $f(s-1) = \zeta(s)h(s-1)$, with h(s-1) regular for $\Re(s) > 0$. The Tauberian theorem gives

$$\sum_{n \le x} \frac{n}{g(n)} \sim c_1 x,$$

and we can apply similar calculations to the ones in the previous lemma to obtain estimates for the last term of (55). We obtain that (55) is

$$\ll \frac{(y_{\max}^{(m)})^2 \phi(W)^{k-2} N(\log R)^{k-1}}{W^{k-1} D_0 \log N},$$

where we have bounded X_N by $N/\log N$, since $X_N = \pi(2N) - \pi(N) \implies X_N = N/\log N + O(N/(\log N)^2)$. When $s_{i,j} = 1$, we obtain

$$S_{2}^{(m)} = \frac{X_{N}}{\phi(W)} \sum_{u_{1},\dots,u_{k}} \frac{(y_{u_{1},\dots,u_{k}}^{(m)})^{2}}{\prod_{i=1}^{k} g(u_{i})} + O\left(\frac{(y_{\max}^{(m)})^{2} \phi(W)^{k-2} N(\log R)^{k-2}}{W^{k-1} D_{0}}\right) + O\left(\frac{y_{\max}^{2} N}{(\log N)^{A}}\right), \quad (56)$$

since $\log R / \log N = \theta / 2 - \delta < 1$. We apply our estimate of X_N to this and obtain an error term of size

$$\ll \frac{(y_{\max}^{(m)})^2 N}{\phi(W)(\log N)^2} \left(\sum_{\substack{u < R\\(u,W)=1}} \frac{\mu(u)^2}{g(u)}\right)^{k-1} \ll \frac{(y_{\max}^{(m)})^2 \phi(W)^{k-2} N(\log R)^{k-3}}{W^{k-1}},$$

where we used the fact that we can estimate the inner sum as $O(\phi(W)/W \times \log R)$ the same way we obtained (48) and that $R = N^{\theta/2-\delta}$ again. This is absorbed by the first error term in (56), which completes the proof. \Box

We now want to relate our variables $y_{r_1,\ldots,r_k}^{(m)}$ to the variables y_{r_1,\ldots,r_k} from Lemma 4.7. We prove the following lemma:

Lemma 4.9. If $r_m = 1$ then

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

Proof. We start by substituting (43) into (53) and we assume that $r_m = 1$. We have

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

We swap the summation between the d and the a variables and obtain

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

We evaluate the sum over the d variables explicitly. We want to prove the following:

$$\sum_{\substack{d_1...d_k\\d_i|a_i,r_i|d_i\\d_m=1}} \prod_{i=1}^k \frac{\mu(d_i)d_i}{\phi(d_i)} = \prod_{i \neq m} \frac{\mu(a_i)r_i}{\phi(a_i)}.$$
(57)

We consider the 1-dim case of (57) and prove that instead. The result follows by induction:

$$\sum_{\substack{d|a\\r|d}} \frac{\mu(d)d}{\phi(d)} = \frac{\mu(a)r}{\phi(a)}.$$
(58)

We see that the condition $r \mid d \mid a$ is equivalent to $d/r \mid a/r$ and we relabel d/r = s. Using this and the multiplicativity of the Möbius and of the Euler totient functions, (58) becomes:

$$\sum_{s|a/r} \frac{\mu(s)\mu(r)sr}{\phi(s)\phi(r)} = \frac{\mu(a)r}{\phi(a)} \iff \sum_{s|a/r} \frac{\mu(s)s}{\phi(s)} \frac{\mu(r)r}{\phi(r)} = \frac{\mu(a)r}{\phi(a)},$$

and by separating the terms inside the sum that do not depend on s we get

(58)
$$\iff \sum_{s|a/r} \frac{\mu(s)s}{\phi(s)} = \frac{\mu(a)r}{\phi(a)} \frac{\phi(r)}{\mu(r)r}.$$

Using the multiplicativity of $\mu(\cdot)$ and $\phi(\cdot)$ again and relabelling a/r = n, we see that we are in fact trying to prove:

$$\sum_{s|n} \frac{\mu(s)s}{\phi(s)} = \frac{\mu(n)}{\phi(n)}.$$
(59)

Let

$$F(n) = \sum_{s|n} \frac{\mu(s)s}{\phi(s)}.$$

This is a multiplicative function by definition, as all the functions inside the sum are multiplicative. Hence, bearing in mind we are working with square-free variables in (57) and (58), it is enough to prove (59) for primes p. Indeed,

$$F(p) = \frac{1}{\phi(1)} - \frac{p}{\phi(p)} = 1 - \frac{p}{p-1} = \frac{-1}{p-1} = \frac{\mu(p)}{\phi(p)},$$

and so (59), (58) and (57) hold. Note that (59) does not hold for non squarefree numbers, as $\mu(p^t) = 0$ for $t \ge 2$, implying $F(p^k) = \mu(p)/\phi(p)$ always. So we have

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

From the support of the y, we know that we can restrict the summation over the a_j so that $(a_j, W) = 1$, so we either have $a_j = r_j$ or $a_j > D_0 r_j$. When $j \neq m$, the contribution from when $a_j > D_0 r_j$ is

$$\ll y_{\max}\left(\prod_{i=1}^{k} g(r_i)r_i\right) \left(\sum_{\substack{a_j > D_0 r_j \\ \phi(a_j)^2}} \frac{\mu(a_j)^2}{\phi(a_j)^2}\right) \left(\sum_{\substack{a_m < R \\ (a_m, W)=1}} \frac{\mu(a_m)^2}{\phi(a_m)}\right) \prod_{\substack{1 \le i \le k \\ i \ne j, m}} \left(\sum_{\substack{r_i \mid a_i \\ \phi(a_i)^2}} \frac{\mu(a_i)^2}{\phi(a_i)^2}\right) \\ \ll \left(\prod_{i=1}^{k} \frac{g(r_i)r_i}{\phi(r_i)^2}\right) \frac{y_{\max}\phi(W)\log R}{WD_0} \ll \frac{y_{\max}\phi(W)\log R}{WD_0},$$

where we got from the first step to the second by using the same bounds as in (46) and, in order to obtain the last estimate, note that the product in the second line is over a finite number of i's, so it is O(1).

Therefore, when $a_j = r_j$ for all $j \neq m$, we have

$$y_{r_1,\dots,r_k}^{(m)} = \left(\prod_{i=1}^k \frac{g(r_i)r_i}{\phi(r_i)^2}\right) \sum_{a_m} \frac{y_{r_1,\dots,r_{m-1},a_m,r_{m+1},\dots,r_k}}{\phi(a_m)} + O\left(\frac{y_{\max}\phi(W)\log R}{WD_0}\right).$$

Note that

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

We know $r := \prod_{i=1}^{k} r_i$ is coprime to W and that the r_i are squarefree, so

$$\prod_{i=1}^{k} \frac{g(r_i)r_i}{\phi(r_i)^2} = \prod_{\substack{p|r\\(r,W)=1}} (1+O(p^{-2})) = 1 + O\left(\sum_{\substack{p|r\\(r,W)=1}} p^{-2} + \sum_{\substack{pq|r\\r,W=1}} (pq)^{-2} + \dots\right),$$

and the error term is in fact $O(1/D_0)$, since we can bound each term by

$$\sum_{n>D_0} \frac{1}{n^2} \sim \int_{D_0}^{\infty} \frac{1}{x^2} dx \ll \frac{1}{D_0}.$$
 (60)

The new error term given by this estimate is absorbed by the old one and the proof is complete. $\hfill \Box$

4.1.2 Smooth choice of y

We now want to choose our y variables so that the ratio S_2/S_1 is maximal, ignoring the error terms. From our expressions (40) and (54), by using Lagrange multipliers, we obtain

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

for some fixed constant λ . We know from our choice of W that the variables y are supported on integers free of small prime factors. Furthermore, we have g(p) = p - 2 and $\phi(p) = p - 1$, so for our variables r_i , $g(r_i) \approx \phi(r_i) \approx r_i$ holds for all i. Therefore, the above reduces to

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

We choose

$$y_{r_1,\dots,r_k} = F\left(\frac{\log r_1}{\log R},\dots,\frac{\log r_k}{\log R}\right),\tag{61}$$

for some smooth function $F : \mathbb{R}^k \to \mathbb{R}$, supported on $\mathcal{R}_k = \{(x_1, \ldots, x_k) \in [0, 1]^k : \sum_{i=1}^k x_i \leq 1\}$. We restrict the support of $F(\cdot)$ to that particular set, since we know that $\prod_i r_i < R$, implying

$$\frac{\log r_1}{\log R} + \ldots + \frac{\log r_1}{\log R} = \frac{\log r_1 + \ldots + \log r_k}{\log R} = \frac{\log (r_1 \ldots r_k)}{\log R} < 1.$$

Bearing in mind that the r_i are all squarefree and their product is coprime to W, we set $y_{r_1,\ldots,r_k} = 0$, when this does not hold. We want to obtain estimates for S_1 and S_2 with the choice of y given by (61). We use the following lemma from [8], which we quote without proof, with slight changes of notation.

Lemma 4.10. Let $\kappa, A_1, A_2, L > 0$. Let $\gamma(\cdot)$ be a multiplicative function satisfying

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

and

$$-L \le \sum_{w \le p \le z} \frac{\gamma(p) \log p}{p} - \kappa \log\left(\frac{z}{w}\right) \le A_2$$

for any $2 \leq w \leq z$. Let $h(\cdot)$ be the totally multiplicative function defined on primes by $h(p) = \gamma(p)/(p - \gamma(p))$. Finally, let $G : [0,1] \to \mathbb{R}$ be a smooth function and let $G_{\max} = \sup_{t \in [0,1]} (|G(t)| + |G'(t)|)$. Then

$$\sum_{d < z} \mu(d)^2 h(d) G\left(\frac{\log d}{\log z}\right) = \mathfrak{G}\frac{(\log z)^{\kappa}}{\Gamma(\kappa)} \int_0^1 G(x) x^{\kappa-1} dx + O_{A_1, A_2, \kappa}(\mathfrak{G}LG_{\max}(\log z)^{\kappa-1}), \qquad (62)$$

where

$$\mathfrak{G} = \prod_{p} \left(1 - \frac{\gamma(p)}{p} \right)^{-1} \left(1 - \frac{1}{p} \right)^{\kappa}$$

Here the constant implied by the 'O' term is independent of G and L.

Remark 15. We remind the reader that $\Gamma(\cdot)$ is the Gamma function, i.e.

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx,$$

which is well defined for $t \in \mathbb{C}$ and $\Re(t) > 0$. For $n \in \mathbb{Z}$, $\Gamma(n) = (n-1)!$ holds.

Remark 16. Note that the series appearing in this lemma is the equivalent of the inverse of the singular series defined in (15), which also appears in the GPY theorem.

We now estimate S_1 . We prove the following lemma:

Lemma 4.11. Let y_{r_1,\ldots,r_k} be defined by (61) in terms of a smooth function F supported on $\mathcal{R}_k = \{(x_1,\ldots,x_k) \in [0,1]^k : \sum_{i=1}^k x_i \leq 1\}$. Let

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

Then we have

$$S_1 = \frac{\phi(W)^k N(\log R)^k}{W^{k+1}} I_k(F) + O\left(\frac{F_{\max}^2 \phi(W)^k N(\log R)^k}{W^{k+1} D_0}\right),$$

where

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

Proof. We substitute our choice of y given by (61) into the expression for S_1 given by (51) to obtain

$$S_{1} = \frac{N}{W} \sum_{\substack{u_{1},\dots,u_{k}\\(u_{i},u_{j})=1\forall i\neq j\\(u_{i},W)=1\forall i}} \left(\prod_{i=1}^{k} \frac{\mu(u_{i})^{2}}{\phi(u_{i})}\right) F\left(\frac{\log u_{1}}{\log R},\dots,\frac{\log u_{k}}{\log R}\right)^{2} + O\left(\frac{F_{\max}^{2}\phi(W)^{k}N(\log R)^{k}}{W^{k+1}D_{0}}\right).$$
(63)

Note that if we have $(u_i, u_j) \neq 1$ and $(u_i, W) = (u_j, W) = 1$, then u_i and u_j must have a common factor greater than D_0 , since $W = \prod_{p < D_0} p$. Thus we can drop the requirement that $u_i \neq u_j$ for all $i \neq j$ at the cost of an error term of size

$$\ll \frac{NF_{\max}^2}{W} \sum_{\substack{p > D_0 \\ p \mid (u_i, u_j) \\ (u_i, W) = 1 \forall i \neq j}} \sum_{\substack{i=1 \\ p \mid (u_i, u_j) \\ (u_i, W) = 1 \forall i \neq j}} \prod_{i=1}^k \frac{\mu(u_i)^2}{\phi(u_i)} \ll \frac{NF_{\max}^2}{W} \sum_{\substack{p > D_0 \\ p > D_0}} \frac{1}{(p-1)^2} \left(\sum_{\substack{u < R \\ (u, W) = 1}} \frac{\mu(u)^2}{\phi(u)}\right)^k,$$

since $\phi(p) = p - 1$ and by summing over a single variable u and then raising that sum to the power k - 1 we are just adding more terms. We use (48) and (60) to obtain an error of size

$$\ll \frac{F_{\max}^2 \phi(W)^k N(\log R)^k}{W^{k+1} D_0}.$$
 (64)

To estimate the main term, we apply Lemma 4.10 k times (once for each variable u_i). We take $\kappa = 1$ each time and we need $\gamma(p)/(p - \gamma(p)) = h(p) = 1/\phi(p) = 1/(p-1)$, so we take

$$\gamma(p) = \begin{cases} 1 & \text{if } p \not\mid W, \\ 0 & \text{otherwise.} \end{cases}$$

This gives $\gamma(p)/p = 0$ or $\gamma(p)/p \le 1/2$, so we can take $A_1 = 1/2$. Also,

$$L \ll \sum_{\substack{w \le p \le z \\ p \nmid W}} \frac{\log p}{p} - \log\left(\frac{z}{w}\right) = \sum_{w \le p \le z} \frac{\log p}{p} - \sum_{\substack{w \le p \le z \\ p \mid W}} \frac{\log p}{p} - \log\left(\frac{z}{w}\right)$$
$$= \log\left(\frac{z}{w}\right) + O(1) + O\left(\sum_{p \mid W} \frac{\log p}{p}\right) - \log\left(\frac{z}{w}\right)$$
$$\ll \sum_{p \le D_0} \frac{\log p}{p} + O(1) \ll \log D_0,$$

by using Merten's theorem repeatedly, which states

$$\sum_{p \le x} \frac{\log p}{p} = \log x + O(1).$$

We have

$$\mathfrak{G} = \prod_{p|W} \left(1 - \frac{1}{p}\right) \prod_{p \notin W} 1 = \frac{\phi(W)}{W}.$$

When we apply the lemma for the first time, we obtain

$$\begin{split} &\sum_{\substack{u_1,\dots,u_k\\(u_i,u_j)=1\forall i\neq j\\(u_i,W)=1\forall i}} \left(\prod_{i=1}^k \frac{\mu(u_i)^2}{\phi(u_i)}\right) F\left(\frac{\log u_1}{\log R},\dots,\frac{\log u_k}{\log R}\right)^2 \\ &= \sum_{\substack{u_2,\dots,u_k\\(u_i,u_j)=1\forall i\neq j\\(u_i,W)=1\forall i}} \left(\prod_{i=2}^k \frac{\mu(u_i)^2}{\phi(u_i)}\right) \sum_{u_1} \frac{\mu(u_1)^2}{\phi(u_1)} F\left(\frac{\log u_1}{\log R},\dots,\frac{\log u_k}{\log R}\right)^2 \\ &= \frac{\phi(W)\log R}{W} \sum_{\substack{u_2,\dots,u_k\\(u_i,u_j)=1\forall i\neq j\\(u_i,W)=1\forall i}} \left(\prod_{i=2}^k \frac{\mu(u_i)^2}{\phi(u_i)}\right) \int_0^1 F\left(t_1,\dots,\frac{\log u_k}{\log R}\right)^2 dt_1 \\ &+ O\left(\frac{F_{\max}^2\phi(W)\log D_0}{W}\left(\sum_{\substack{u_k< R\\(u,W)=1}} \frac{\mu(u)^2}{\phi(u)^2}\right)^{k-1}\right). \end{split}$$

The error term becomes

$$\ll \frac{F_{\max}^2 \phi(W)^k \log D_0(\log R)^{k-1}}{W^k},$$

and for every subsequent applications of the lemma, the error terms we obtain become smaller and smaller. Thus, the main term in (63) becomes

$$\frac{N}{W} \frac{\phi(W)^k (\log R)^k}{W^k} I_k(F) + O\left(\frac{F_{\max}^2 \phi(W)^k N \log D_0(\log R)^{k-1}}{W^{k+1}}\right).$$
(65)

Comparing the error terms in (65) and (64), we see that the former gets absorbed into the latter, since $D_0 \ll \log \log \log N$ and $R = N^{\theta/2}$. Hence, the lemma is proved.

We now estimate S_2 . We prove the following lemma:

Lemma 4.12. Let y_{r_1,\ldots,r_k} , $F(\cdot)$ and F_{\max} be described as in Lemma 4.11. Then we have

$$S_2^{(m)} = \frac{\phi(W)^k N(\log R)^{k+1}}{W^{k+1} \log N} J_k^{(m)}(F) + O\left(\frac{F_{\max}^2 \phi(W)^k N(\log R)^k}{W^{k+1} D_0}\right),$$

where

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

Proof. We begin by estimating $y_{r_1,\ldots,r_k}^{(m)}$ first. We know $y_{r_1,\ldots,r_k}^{(m)} = 0$ when $r_m \neq 1$ or $(\prod_{i=1}^k r_i, W) \neq 1$ or the r_i are not coprime and squarefree. We substitute our choice of y into the expression for $y^{(m)}$ given by Lemma 4.9. We first look at the case when $y_{r_1,\ldots,r_k}^{(m)} \neq 0$. We have

$$y_{r_1,\dots,r_k}^{(m)} = \sum_{\substack{(u,W) \prod_{i=1}^k r_i) = 1}} \frac{\mu(u)^2}{\phi(u)} F\left(\frac{\log r_1}{\log R},\dots,\frac{\log r_{m-1}}{\log R},\frac{\log u}{\log R},\frac{\log r_{m+1}}{\log R},\dots,\frac{\log r_k}{\log R}\right) + O\left(\frac{F_{\max}\phi(W)\log R}{WD_0}\right).$$

Since $\prod_{i=1}^{k} r_i < R$, we have

$$y_{\max}^{(m)} \ll \frac{F_{\max}\phi(W)\log R}{W}.$$
(66)

We now estimate the sum over u. We apply Lemma 4.10 with $\kappa = 1$ and

$$\gamma(p) = \begin{cases} 1 & \text{if } p \not\mid W \prod_{i=1}^{k} r_i, \\ 0 & \text{otherwise.} \end{cases}$$

We can take $A_1 = 1/2$ again and

$$\begin{split} L \ll 1 + \sum_{p \mid W \prod_{i=1}^{k} r_i} \frac{\log p}{p} \ll \sum_{p < \log R} \frac{\log p}{p} + \sum_{\substack{p \mid W \prod_{i=1}^{k} r_i \\ p > \log R}} \frac{\log p}{p} \\ \ll \log \log R + \sum_{\substack{p \mid W \\ p > \log R}} \frac{\log p}{p} + \sum_{\substack{p \mid \prod_{i=1}^{k} r_i \\ p > \log R}} \frac{\log p}{p} \\ \ll \log \log R + \sum_{\substack{p \leq D_0 \\ p > \log R}} \frac{\log p}{p} + \frac{\sum_{p \mid \prod r_i} \log p}{\log R} \\ \ll \log \log R + \log D_0 + \frac{\log \prod_i r_1}{\log R} \ll \log \log R, \end{split}$$

since $R = N^{\theta/2}$. In this case, we have

$$\mathfrak{G} = \frac{\phi(W)}{W} \prod_{i=1}^{k} \frac{\phi(r_i)}{r_i}.$$

Therefore, we have

$$\begin{aligned} y_{r_1,\dots,r_k}^{(m)} = &(\log R) \frac{\phi(W)}{W} \left(\prod_{i=1}^k \frac{\phi(r_i)}{r_i} \right) F_{r_1,\dots,r_k}^{(m)} \\ &+ O\left(\frac{F_{\max}\phi(W)}{W} \prod_{i=1}^k \frac{\phi(r_i)}{r_i} \log \log R \right) + O\left(\frac{F_{\max}\phi(W) \log R}{WD_0} \right), \end{aligned}$$

where

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

and the second error term dominates. We now substitute this into our expression for $S_2^{(m)}$ given by Lemma 4.8 to obtain

$$\begin{split} S_2^{(m)} = & \frac{\phi(W)N(\log R)^2}{W^2 \log N} \sum_{\substack{r_1, \dots, r_k \\ (r_i, W) = 1 \forall i \\ (r_i, r_j) = 1 \forall i \neq j \\ r_m = 1}} \left(\prod_{i=1}^k \frac{\mu(r_i)^2 \phi(r_i)^2}{g(r_i)r_i^2} \right) (F_{r_1, \dots, r_k}^{(m)})^2 \\ &+ O\left(\frac{F_{\max}^2 \phi(W)^k N(\log N)^k}{W^{k+1} D_0} \right), \end{split}$$

where to obtain this error term we substituted (66) into the first error term in (54) and the second error term in (54) gets absorbed into this one. The error terms given by the expression for $y^{(m)}$ also get absorbed into it. The first one is

$$\ll \frac{F_{\max}\phi(W)\log R}{WD_0} \frac{N}{\phi(W)\log N} y_{\max} \left(\sum_{\substack{u < R \\ (u,W)=1}} \frac{\mu(u)^2}{g(u)}\right)^{k-1} \\ \ll \frac{F_{\max}^2\phi(W)N\log R}{W^2D_0} \left(\frac{\phi(W)\log R}{W}\right)^{k-1} = \frac{F_{\max}^2\phi(W)^k N(\log R)^k}{W^{k+1}D_0}.$$

The second error term given by the expression of $y^{(m)}$ is

$$\ll \frac{F_{\max}^2 \phi(W)^2 (\log R)^2}{W^2 D_0^2} \frac{N}{\phi(W) \log N} \left(\sum_{\substack{u < R \\ (u,W)=1}} \frac{\mu(u)^2}{g(u)} \right)^{k-1}$$
$$\ll \frac{F_{\max}^2 \phi(W)^k N (\log R)^{k+1}}{W^{k+1} D_0^2 \log N}.$$

We can remove the condition that $(r_i, r_j) = 1$ as we did in our treatment of S_1 , at the cost of an error term of size

$$\ll \frac{\phi(W)N(\log R)^2 F_{\max}^2}{W^2 \log N} \left(\sum_{p>D_0} \frac{\phi(p)^4}{g(p)^2 p^4} \right) \left(\sum_{\substack{r < R \\ (r,W)=1}} \frac{\mu(r)^2 \phi(r)^2}{g(r)r^2} \right)^{k-1} \\ \ll \frac{F_{\max}^2 \phi(W)^k N(\log R)^k}{W^{k+1} D_0},$$

where we bound the first sum by $1/D_0$ and the second by $(\log R\phi(W)/W)^{k-1}$. We now want to evaluate the quantity

$$S = \sum_{\substack{r_1, \dots, r_{m-1}, r_{m+1}, \dots, r_k \\ (r_i, W) = 1 \forall i}} \left(\prod_{\substack{1 \le i \le k \\ i \ne m}} \frac{\mu(r_i)^2 \phi(r_i)^2}{g(r_i) r_i^2} \right) (F_{r_1, \dots, r_k}^{(m)})^2.$$

We apply Lemma 4.10 again, once for each variable r_i . We take $\kappa = 1$ and we need $h(p) = \gamma(p)/(p - \gamma(p))$, so

$$\frac{\gamma(p)}{p - \gamma(p)} = \frac{\phi(p)^2}{g(p)p^2},$$

which gives

$$\gamma(p) = \frac{\phi(p)^2 p}{g(p)p^2 + \phi(p)^2},$$

so we take

$$\gamma(p) = \begin{cases} 1 - \frac{p^2 - 3p + 1}{p^3 - p^2 - 2p + 1} & \text{if } p \not\mid W, \\ 0 & \text{otherwise.} \end{cases}$$

This gives

$$\mathfrak{G} = \prod_{p} \left(1 - \frac{\gamma(p)}{p} \right)^{-1} \left(1 - \frac{1}{p} \right) = \frac{\phi(W)}{W} \prod_{p \notin W} \left(1 - \frac{\gamma(p)}{p} \right)^{-1} \left(1 - \frac{1}{p} \right).$$

Now, for primes not dividing W, $\gamma(p) = 1 + O(1/p)$, so

$$\log\left(\prod_{p \notin W} \left(1 - \frac{\gamma(p)}{p}\right)^{-1} \left(1 - \frac{1}{p}\right)\right) = \sum_{p > D_0} \log\left(\frac{1 - 1/p}{1 - \gamma(p)/p}\right)$$
$$= \sum_{p > D_0} \log\left(\frac{(p - 1)/p}{1 - 1/p - O(1/p^2)}\right)$$
$$= \sum_{p > D_0} \log\left(1 + O\left(\frac{1}{p^2 - p - 1}\right)\right)$$
$$= \sum_{p > D_0} O\left(\frac{1}{p^2 - p - 1}\right) = O(1/D_0).$$

So, by Taylor expansion,

$$\prod_{p \notin W} \left(1 - \frac{\gamma(p)}{p} \right)^{-1} \left(1 - \frac{1}{p} \right) = e^{O(1/D_0)} = 1 + O(1/D_0),$$

for $D_0 \to \infty$. Hence,

$$\mathfrak{G} = \frac{\phi(W)}{W} \left(1 + O\left(\frac{1}{D_0}\right) \right).$$

Also,

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

So, for the first application of the lemma, we obtain

$$\begin{split} S &= \sum_{\substack{r_2, \dots, r_{m-1}, r_{m+1}, \dots, r_k \\ (r_i, W) = 1 \forall i}} \left(\prod_{\substack{2 \le i \le k \\ i \ne m}} \frac{\mu(r_i)^2 \phi(r_i)^2}{g(r_i) r_i^2} \right) \sum_{r_1} \frac{\mu(r_1)^2 \phi(r_1)^2}{g(r_1) r_1^2} (F_{r_1, \dots, r_k}^{(m)})^2 \\ &= \frac{\phi(W) \log R}{W} \left(1 + O\left(\frac{1}{D_0}\right) \right) \sum_{\substack{r_2, \dots, r_{m-1}, r_{m+1}, \dots, r_k \\ (r_i, W) = 1 \forall i}} \left(\prod_{\substack{2 \le i \le k \\ i \ne m}} \frac{\mu(r_i)^2 \phi(r_i)^2}{g(r_i) r_i^2} \right) \int_0^1 (F_{t, \dots, r_k}^{(m)})^2 dt \\ &+ O\left(\frac{F_{\max}^2 \phi(W) \log D_0}{W} \left(\sum_{\substack{u \le R \\ (u, W) = 1}} \frac{\mu(u)^2 \phi(u)^2}{g(u) u^2} \right)^{k-2} \right), \end{split}$$

and the first error term dominates, becoming

$$\ll \frac{F_{\max}^2 \phi(W)^{k-1} (\log R)^{k-1}}{W^{k-1} D_0}$$

and for every subsequent applications of the lemma, the error terms we obtain become smaller and smaller. Thus, after k-1 applications of the lemma, we obtain

$$S_2^{(m)} = \frac{\phi(W)^k N(\log R)^{k+1}}{W^{k+1}\log N} J_k^{(m)} + O\left(\frac{F_{\max}^2 \phi(W)^k N(\log N)^k}{W^{k+1}D_0}\right),$$

where

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

hence the lemma is proved.

4.2 Optimisation and consequences

Proposition 4.13. Let the primes have level of distribution $\theta > 0$. Let $\delta > 0$ and $\mathcal{H} = \{h_1, \ldots, h_k\}$ be an admissible set. Let $I_k(F)$ and $J_k^{(m)}(F)$ be given as in Proposition 4.6 and let \mathcal{S}_k denote the set of Riemann-integrable functions $F: [0,1]^k \to \mathbb{R}$ supported on $\mathcal{R}_k = \{(x_1, \ldots, x_k) \in [0,1]^k : \sum_{i=1}^k x_i \leq 1\}$ with $I_k(F) \neq 0$ and $j_k^{(m)}(F) \neq 0$ for each m. Let

$$M_k = \sup_{F \in \mathcal{S}_k} \frac{\sum_{m=1}^k J_k^{(m)}(F)}{I_k(F)}, \qquad r_k = \left\lceil \frac{\theta M_k}{2} \right\rceil,$$

where $\lceil x \rceil$ is the smallest integer bigger than or equal to x. Then there are infinitely many integers n such that at least r_k of the $n + h_i$ $(1 \le i \le k)$ are prime. In particular,

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

Proof of Proposition 4.13. We recall that if we can show that (36) is positive for all large N, then there are infinitely many integers n such that at least two of the $n + h_i$ are prime.

Let $R = N^{\theta/2-\epsilon}$ for a small $\epsilon > 0$. By the definition of M_k , we can choose $F_0 \in \mathcal{S}_k$ such that

$$\sum_{m=1}^{k} J_k^{(m)}(F_0) > (M_k - \epsilon) I_k(F_0).$$
(67)

Since $F_0(\cdot)$ is Riemann-integrable, there exists a smooth function $F_1(\cdot)$ such that

$$\sum_{m=1}^{k} J_k^{(m)}(F_1) > (M_k - 2\epsilon) I_k(F_1).$$
(68)

From Proposition 4.6 we know that we can choose λ_{d_1,\ldots,d_k} such that

$$S(N,\rho) = S_2 - \rho S_1$$

= $\frac{(1+o(1))\phi(W)^k N(\log R)^k}{W^{k+1}} \left(\frac{\log R}{\log N} \sum_{j=1}^k J_k^{(m)}(F_1) - \rho I_k(F_1)\right),$

so we can use (68) and the level of distribution to obtain

$$S(N,\rho) > \frac{(1+o(1))\phi(W)^k N(\log R)^k}{W^{k+1}} I_k(F_1) \left((\frac{\theta}{2} - \epsilon)(M_k - 2\epsilon) - \rho \right).$$

So $S(N, \rho) > 0$ for all large N if we choose $\rho = \theta M_k/2 - \delta$ and we pick ϵ to be sufficiently small. We deduce that there are infinitely many integers n for which at least $\lfloor \rho + 1 \rfloor$ of the $n + h_i$ are prime. If δ is sufficiently small, $\lfloor \rho + 1 \rfloor = \lceil \theta M_k/2 \rceil = r_k$, hence the proof is complete. \Box

Proposition 4.14. Let $k \in \mathbb{N}$ and let M_k be as given by Proposition 4.13. Then

- (1) We have $M_5 > 2$.
- (2) We have $M_{105} > 4$.
- (3) If k is sufficiently large, we have $M_k > \log k 2\log \log k 2$.

We discuss the proof of Proposition 4.14 briefly. For parts (1) and (2) we want to find a lower bound for M_k when k is small. To do that, we use the following lemma:

Lemma 4.15. Let $P_j = \sum_{i=1}^k t_i^j$ denote the *j*-th symmetric power sum polynomial. Then we have

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

where

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

is a polynomial of degree b which depends only on b and j and \mathcal{R}_k is defined as in Proposition 4.13.

Proof. We use induction on k to show that

$$\int_{\mathcal{R}_k} \left(1 - \sum_{i=1}^k t_i \right)^a \prod_{i=1}^k t_i^{a_i} dt_1 \dots dt_k = \frac{a! \prod_{i=1}^k a_i!}{(k+a+\sum_{i=1}^k a_i)!}.$$
 (69)

We consider the integration with respect to t_1 first and we make the substitution $\nu = t_1/(1 - \sum_{i=2}^k t_i)$, which gives $dt_1 = d\nu(1 - \sum_{i=2}^k t_i)$. We obtain

$$\int_{0}^{1-\sum_{i=2}^{k}t_{i}} \left(1-\sum_{i=1}^{k}t_{i}\right) \prod_{i=1}^{a} t_{i}^{a_{i}} dt_{1} = \left(\prod_{i=2}^{k}t_{i}^{a_{i}}\right) \left(1-\sum_{i=2}^{k}t_{i}\right)^{a+a_{1}+1} \int_{0}^{1} (1-\nu)^{a} \nu^{a_{1}} d\nu.$$

On the right-hand side of this equation, we recognise the Beta function,

$$B(a,b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx,$$

which can be expressed in terms of the Gamma function as

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Thus, we obtain

$$\int_{0}^{1-\sum_{i=2}^{k}t_{i}} \left(1-\sum_{i=1}^{k}t_{i}\right) \prod_{i=1}^{a} t_{i}^{a_{i}} dt_{1} = \frac{a!a_{1}!}{(a+a_{1}+1)!} \left(\prod_{i=2}^{k}t_{i}^{a_{i}}\right) \left(1-\sum_{i=2}^{k}t_{i}\right)^{a+a_{1}+1} dt_{1} = \frac{a!a_{1}!}{(a+a_{1}+1)!} dt_{1} = \frac{a!a_{1}!}{$$

We see that (69) follows by induction. The binomial theorem gives

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

Therefore,

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

Now, there are $\binom{k}{r}$ of choosing r of the b_1, \ldots, b_k to be non-zero, so

$$\sum_{\substack{b_1, \dots, b_k \\ \sum_{i=1}^k b_i = b}} \prod_{i=1}^k \frac{(jb_i)!}{b_i!} = \sum_{r=1}^b \binom{k}{r} \sum_{\substack{b_1, \dots, b_r \ge 1 \\ sum_{i=1}rb_i = b}} \prod_{i=1}^r \frac{(jb_i)!}{b_i!},$$

thus the lemma is proved.

We now concentrate on the case when P is a polynomial expression in P_1 and P_2 only.

Lemma 4.16. Let $F(\cdot)$ be given in terms of a polynomial P by

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

Let P be given in terms of a polynomial expression in the symmetric power polynomials $P_1 = \sum_{i=1}^k t_i$ and $P_2 = \sum_{i=1}^k t_i^2$ by $P = \sum_{i=1}^d a_i (1-P_1)^{b_i} P_2^{c_1}$ for constants $a_i \in \mathbb{R}$ and non-negative integers b_i, c_i . Then for each $1 \le m \le k$ we have

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

where

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

and where G is the polynomial given by Lemma 4.15.

We omit the proof of this lemma, but note that its importance is that we deduce from it that $I_k(F)$ and $\sum_{i=1}^k J_k^{(m)}(F)$ can be expressed as quadratic forms in the coefficients $\mathbf{a} = (a_1, \ldots, a_d)$ of P. They will be positive definite real quadratic forms and so

$$\frac{\sum_{m=1}^{k} J_k^{(m)}(F)}{I_k(F)} = \frac{\mathbf{a}^T M_2 \mathbf{a}}{\mathbf{a}^T M_1 \mathbf{a}},$$

for two rational symmetric positive definite matrices M_1, M_2 , which we can calculate explicitly in terms of k for any choice of exponents b_i, c_i . To maximise this expression we use the following lemma from linear algebra:

Lemma 4.17. Let M_1, M_2 be real, symmetric, positive definite matrices. Then

$$\frac{\boldsymbol{a}^T M_2 \boldsymbol{a}}{\boldsymbol{a}^T M_1 \boldsymbol{a}}$$

is maximised when **a** is an eigenvector of $M_1^{-1}M_2$ corresponding to the largest eigenvalue of $M_1^{-1}M_2$. The value of the ratio at its maximum is this largest eigenvalue.

Proof. We can normalise the denominator of this ratio, since multiplying **a** does not change its value. Thus, we can use Lagrange multipliers to maximise $\mathbf{a}^T M_2 \mathbf{a}$ subject to the constraint $\mathbf{a}^T M_1 \mathbf{a} = 1$. We write

$$L(\mathbf{a},\lambda) = \mathbf{a}^T M_2 \mathbf{a} - \lambda (\mathbf{a}^T M_1 \mathbf{a} - 1)$$

and, differentiating, we obtain the following system of equations

$$0 = \frac{\partial L}{\partial a_i} = ((2M_2 - 2\lambda M_1)\mathbf{a})_i,$$

where we have used the fact that the matrices M_i are symmetric, and so $\mathbf{a}^T M_i \mathbf{a} = \sum_{j,k} m_{jk} a_j a_k$. Since M_1 is invertible by positive definiteness, this implies that

$$M_1^{-1}M_2\mathbf{a} = \lambda \mathbf{a}.$$
$$\mathbf{a}^T M_2 \mathbf{a}$$

It follows that

$$\frac{\mathbf{a}^T M_2 \mathbf{a}}{\mathbf{a}^T M_1 \mathbf{a}} = \lambda.$$

To obtain our estimates, we let $F(\cdot)$ be defined in terms of a polynomial Pas in Lemma 4.15. Let P be given by a polynomial expression in $P_1 = \sum_{i=1}^k t_i$ and $P_2 = \sum_{i=1}^k t_i^2$ which is a linear combination of all monomials $(1 - P_1)^b P_2^c$ with $b + 2c \leq 11$. There are 42 such monomials and if we take k = 105 we can calculate the rational symmetric matrices M_1 and M_2 corresponding to the coefficients of our two quadratic forms $I_k(F)$ and $\sum_{i=1}^k J_k^{(m)}(F)$. With the help of a computer, Maynard found the largest eigenvalue of $M_1^{-1}M_2$ to be

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

thus establishing part (1). To establish part (2), take

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

and k = 5 to find that

$$M_5 \ge \frac{\sum_{m=1}^k J_k^{(m)}(F)}{I_k(F)} = \frac{1,417,255}{708,216} > 2.$$

To prove (3), we want to find a lower bound for M_k when k is large. We choose $F(\cdot)$ to be of the form

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

for some smooth function $g: [0, \infty] \to \mathbb{R}$, supported on [0, T]. This choice makes $F(\cdot)$ symmetric, thus removing the dependency of $J_k^{(m)}(F)$ on m. So we only need to consider $J_k(F) = J_k^{(1)}(F)$. Furthermore, we note that for large k we can drop the constraints $\sum_{i=1}^k t_i \leq 1$ at the cost of a small error term. Next, we let $\gamma = \int_{u\geq 0} g(u)^2 du$, and look at the case when $g(\cdot)$ satisfies $\gamma > 0$. We have

$$I_k(F) = \int_{\mathcal{R}_k} F(t_1, \dots, t_k)^2 dt_1 \dots dt_k \le \left(\int_0^\infty g(kt)^2 dt\right)^k = k^{-k} \gamma^k.$$

In order to obtain a lower bound for $J_k(F)$, we restrict our integral to $\sum_{i=2}^k t_i < 1 - T/k$ for some positive quantity T. We have

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

and we write the right-hand side as $J_k^*(F) - E_k^*(F)$, where

$$J_k^*(F) = \int_{t_2,\dots,t_k \ge 0} \left(\int_0^{T/k} \left(\prod_{i=1}^k g(kt_i) \right) dt_1 \right)^2 dt_2 \dots dt_k$$
$$= k^{-k-1} \gamma^{k-1} \left(\int_0^\infty g(u) du \right)^2,$$

$$E_k^*(F) = \int_{\substack{t_2,\dots,t_k \ge 0\\\sum_{i=2}^k t_i > 1-T/k}} \left(\int_0^{T/k} \left(\prod_{i=1}^k g(kt_i) \right) dt_1 \right)^2 dt_2 \dots dt_k$$

= $k^{-k-1} \left(\int_0^\infty g(u) du \right)^2 \int_{\substack{u_2,\dots,u_k \ge 0\\\sum_{i=2}^k u_i > k-T}} \left(\prod_{i=2}^k g(u_i) \right)^2 du_2 \dots du_k.$

Under the restriction

$$\mu = \frac{\int_0^\infty u g(u)^2 du}{\int_0^\infty g(u)^2 du} < 1 - \frac{T}{k},$$

we obtain through simple calculations that

$$\frac{kJ_k(F)}{I_k(F)} \ge \frac{\left(\int_0^\infty g(u)du\right)^2}{\int_0^\infty g(u)^2 du} \left(1 - \frac{T}{k(1 - T/k - \mu)^2}\right).$$
(70)

We want to maximise this lower bound, so we use Lagrange multipliers to maximise $\int_0^T g(u) du$ subject to the constraints

$$\int_{o}^{T} g(u)^{2} du = \gamma \text{ and } \int_{0}^{T} u g(u)^{2} du = \mu \gamma.$$

We use the Euler–Lagrange equations to obtain

$$g(t) = \frac{1}{2\alpha + 2\beta t},$$

where α and β are Lagrange multipliers and $0 \le t \le T$. We then restrict our attention to functions $g(\cdot)$ of the form 1/(1 + At) for some constant A > 0 and $t \in [0, T]$. This gives

$$\int_{0}^{T} g(u)du = \frac{\log(1+AT)}{A}, \quad \int_{0}^{T} g(u)^{2}du = \frac{1}{A}\left(1-\frac{1}{1+AT}\right)$$
(71)

and

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

Next, we choose T such that $1 + AT = e^T$ and, substituting (71) into (70), we obtain

$$\frac{kJ_k(F)}{I_k(F)} \ge A\left(1 - \frac{Ae^A}{k(1 - A/(e^A - 1) - e^A/k)^2}\right).$$

We then choose $A = \log k - 2 \log \log k$ to obtain the final result.

4.3 Main results

We are now ready to prove the main theorems of this section.

Proof of Theorem 4.2. We look at admissible sets \mathcal{H} of size k, when k is large. By Theorem 2.9, we know primes have level of distribution $\theta = 1/2 - \epsilon$ for any $\epsilon > 0$. By Propositions 4.14 and 4.13,

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

If we choose $\epsilon = 1/k$, we have $\theta M_k/2 > m$ if $k \ge Cm^2 e^{4m}$ for some absolute constant C which is independent of m and k:

$$\begin{aligned} \frac{\theta M_k}{2} \geq & \frac{k-2}{4k} (\log k - 2\log \log k - 2) > m \iff \\ & \log k - 2\log \log k - 2 > m \frac{4k}{k-2} \iff \\ & \log k > 2\log \log k + 2 + m \frac{4k}{k-2} \iff \\ & k > (\log k)^2 \cdot \exp 2 \cdot \exp \left(\frac{4mk}{(k-2)}\right) \\ & = (\log k)^2 \cdot \exp 2 \cdot \exp(4m) \cdot \exp\left(\frac{8m}{(k-2)}\right). \end{aligned}$$

Now, $k > e^{4m}$ by choice, so $\exp(8m/(k-2)) < \exp(8m/(e^{4m}-2)) \le P$ for some constant P, since k sufficiently large implies we can take m sufficiently large. Therefore, we require

$$k > Q(\log k)^2 \exp\left(4m\right),$$

where $Q := P \exp 2$. This is equivalent to

$$\frac{1}{R} > \frac{(\log k)^2}{k} \exp(4m).$$

Note that the function $f(x) := (\log x)^2 / x$ is decreasing, so for $k \ge Cm^2 e^{4m}$, we need

$$\frac{1}{R} > \frac{(\log C + 2\log m + 4m)^2}{Cm^2}.$$

For m sufficiently large, we can choose C so that this holds at all times.

Therefore, for any admissible set \mathcal{H} of size $k \geq Cm^2 e^{4m}$, at least m + 1of the $n + h_i$ must be prime for infinitely many integers n. If we choose $\mathcal{H} = \{p_{\pi(k)+1}, \ldots, p_{\pi(k)+k}\}$, it will be admissible since there are k elements and none of them is a multiple of a prime less than k. By the prime number theorem, $p_{\pi(k)+1} - p_{\pi(k)+k} \ll k \log k$. If we choose $k = \lceil Cm^2 e^{4m} \rceil$, $k \log k \sim$ $Cm^2 e^{4m} (\log C + 2 \log m + 4m) \ll m^3 e^{4m}$. Hence, the proof is complete. \Box Proof of Theorem 4.4. We look at admissible sets of size k = 105. We know primes have level of distribution $\theta = 1/2 - \epsilon$ for any $\epsilon > 0$. So if we take ϵ sufficiently small, we have $2 > \theta M_{105}/2 > 1$ and $r_{105} = 2$ in Proposition 4.13. So $\liminf_n (p_{n+1} - p_n) \le \max_{1\le i,j\le 105} (h_i - h_j)$ for any admissible set \mathcal{H} of size 105. Thomas Engelsma found such an admissible set with $h_1 = 0$ and $h_{105} = 600$. Hence, the proof is complete. \Box

Proof of Theorem 4.5. If we assume that primes have level of distribution θ for any $\theta = 1 - \epsilon$, then Propositions 4.13 and 4.14 tell us that $r_{105} = 2$ for small enough ϵ . So $\liminf_n (p_{n+2} - p_n) \leq \max_{1 \leq i,j \leq 105} (h_i - h_j)$, and if we use the same admissible set as above, the first part of the proof is complete.

We next take k = 5 and $\mathcal{H} = \{0, 2, 6, 8, 12\}$. We have $\Omega(2) = \{0\}$, $\Omega(3) = \{0, 2\}$ and $\Omega(5) = \{0, 2, 3, 4\}$, hence \mathcal{H} is admissible. We take $\theta = 1 - \epsilon$ again and, by Proposition 4.14, $M_5 > 2$ and so $r_5 = 1$ for ϵ sufficiently small. Thus, Proposition 4.13 gives $\liminf_n (p_{n+1} - p_n) \leq 12$. Hence, the proof is complete.

5 Conclusions and further directions

Before the work of Maynard and Zhang, the strongest unconditional result about small gaps between primes had been the GPY theorem, which does not prove the existence of infinitely many bounded gaps. Recent results take us one step closer to obtaining the twin prime conjecture, as we aim to reduce the gap in (2) to 2.

Many mathematicians are working on this problem, and the polymath project [13] reduces the gap to 246 by combining the techniques developed by Maynard with those in [12]. Furthermore, it is proved in [13] that under the generalised Elliott–Halberstam conjecture 5.1, the gap is reduced to 6.

Conjecture 5.1 (Generalised Elliott–Halberstam conjecture). Let $\epsilon > 0$ and $A \ge 1$ be fixed. Let N, M be quantities such that $x^{\epsilon} \ll N \ll x^{1-\epsilon}$ and $x^{\epsilon} \ll M \ll x^{1-\epsilon}$ with $NM \asymp x$, and let $\alpha, \beta : \mathbb{N} \to \mathbb{R}$ be sequences supported on [N, 2N] and [M, 2M] respectively, such that one has the pointwise bounds

$$|\alpha(n)| \ll \tau(n)^{O(1)} (\log x)^{O(1)}; |\beta(m)| \ll \tau(m)^{O(1)} (\log x)^{O(1)}$$
(72)

for all natural numbers n.m. Suppose also that $\beta(\cdot)$ obeys the Siegel-Walfisz type bound

$$|\Delta(\beta \mathbf{1}_{(\cdot,r)=1}; a(q))| \ll \tau(qr)^{O(1)} M(\log x)^{-A}$$

for any $q, r \geq 1$, any fixed A and any primitive residue class a modulo q. Then, for any $Q \ll x^{\theta}$, we have

$$\sum_{q \le Q} \sup_{a \in (\mathbb{Z}/q\mathbb{Z})^{\times}} |\Delta(\alpha \star \beta; a(q))| \ll x(\log x)^{-A}.$$

Remark 17. We define $\alpha \star \beta(\cdot)$ to be the Dirichlet convolution of $\alpha(\cdot)$ and $\beta(\cdot)$, that is

$$\alpha \star \beta(n) := \sum_{d|n} \alpha(n) \beta\left(\frac{n}{d}\right).$$

We also define, for any function $\alpha : \mathbb{N} \to \mathbb{C}$ with finite support and any residue class a modulo q, the (signed) discrepancy $\Delta(\alpha; a(q))$ to be the quantity

$$\Delta(\alpha; a(q)) := \sum_{n \equiv a \pmod{q}} \alpha(n) - \frac{1}{\phi(q)} \sum_{(n,q)=1} \alpha(n).$$

Remark 18. It is proved in [13] that for any fixed $0 < \theta < 1$ the generalised Elliott–Halberstam conjecture implies the Elliott–Halberstam conjecture.

For a review of the topic of bounded gaps between primes and some of the applications of Maynard's work, see [9].

On a separate note, mathematicians are also interested in estimating 'long' gaps between primes, i.e the quantity $\limsup_{n\to\infty} (p_{n+1} - p_n)$. In their paper [4], Ford *et al.* prove that

$$\max_{p_{n+1} \le X} (p_{n+1} - p_n) \gg \frac{\log X \log \log \log \log \log \log X}{\log \log \log X}$$

for sufficiently large X.

Maynard's results are incredibly strong on their own, but it is his techniques that represent the most powerful tool in [10]. He brought the multidimensional Selberg sieve back into prominence, as it continues to be developed and used not only in the study of small gaps between primes, but in various other classical problems in analytic number theory.

Acknowledgments

It is a pleasure to thank my supervisor, Dr Yiannis Petridis, for his guidance and his constant support.

References

- A. C. Cojocaru and M. Ram Murty, An introduction to sieve methods and their applications, Cambridge University Press, 2005.
- [2] W. Chen, Lecture notes in elementary number theory, https://rutherglen.science.mq.edu.au/wchen/ lnentfolder/lnent.html.
- [3] P. D. T. A. Elliott and H. Halberstam, A conjecture in prime number theory, 1970 Symposia Mathematica, Vol. IV (IN-DAM, Rome, 1968/69) pp. 59–72 Academic Press, London.
- [4] K. Ford, B. Green, S. Konyagin, J. Maynard, T. Tao, *Long gaps between primes*, preprint.
- [5] J. Friedlander and A. Granville, *Limitations to the equidistribution of primes*, I. Ann. of Math. (2), 129(2):363–382, 1989.
- [6] P.X. Gallagher, On the distribution of primes in short intervals, Mathematika 23 (1976), no. 1, 4–9; Corrigendum, ibid, 28 (1981), 86.
- [7] D. Goldston, S. Graham, J. Pintz and C. Yıldırım, Small gaps between primes exist, Proc. Japan Acad. Ser. A Math. Sci. 82 (2006), no. 4, 61–65.
- [8] D. A. Goldston, S. W. Graham, J. Pintz, and C. Y. Yıldırım Small gaps between products of two primes, Proc. Lond. Math. Soc. (3) 98 (2009), no. 3, 741–774; Corrigendum. ibid, 106 (2013), no. 2, 477–480.
- [9] A. Granville, Primes in intervals of bounded length, Bull. Amer. Math. Soc. (N.S.) 52 (2015), no. 2, 171–222.
- [10] J. Maynard, Small gaps between primes, Ann. of Math. (2) 181 (2015), no. 1, 383–413.
- [11] M. Ram Murty, Problems in Analytic Number Theory, Second edition, Springer Science+Business Media, LLC (2008).
- [12] D.H.J. Polymath, New equidistribution estimates of Zhang type, preprint.

- [13] D.H.J. Polymath, Variants of the Selberg sieve, and bounded intervals containing many primes, preprint.
- [14] K. Soundararajan, Small gaps between prime numbers: the work of Goldston-Pintz-Yildirim, Bull. Amer. Math. Soc. (N.S.) 44 (2007), no. 1, 1–18.
- [15] A. Strömbergsson, Analytic number theory Lecture notes based on Davenport's book, http://www2.math.uu.se/ ~astrombe/analtalt08/www_notes.pdf.
- [16] Y. Zhang, Bounded gaps between primes, Ann. of Math. (2) 179 (2014), no. 3, 1121–1174.