BOUNDED INTERVALS CONTAINING PRIMES

Essay for Part III of the Mathematical Tripos

Josha Box BSc
University of Cambridge
Department of Mathematics
Supervised by Dr. Adam Harper

April 19, 2016

Abstract

This essay discusses the method introduced by James Maynard in 2013 to show boundedness of gaps between primes. Also, two subsequent improvements to this method by the Polymath8b group are implemented to prove the best currently known result $\liminf p_{n+1} - p_n \leq 246$ about prime gaps. Moreover, the boundedness of gaps between arbitrary prime m-tuples is proved and Maynard's sieve method is compared to its predecessor, the GPY sieve. Finally, an argument is given that denies such sieve methods to be applied to prove the twin prime conjecture.

Contents

1 Introduction

1.1 The Twin Prime Conjecture

The *Twin Prime Conjecture* states that there are infinitely many primes p such that $p+2$ is also prime. Some say Euclid was the first to pose the conjecture; what we can say for sure is that mathematicians have unsuccessfully attempted to prove the conjecture for hundreds of years. A consequence of the prime number theorem is that the average gap between prime numbers close to n is approximately of size $\log n$, so twin primes appear to be a rare species. This was confirmed by Viggo Brun [2], who proved in 1915 that

$$\sum_{p:\ p+2 \text{ prime}} \frac{1}{p} + \frac{1}{p+2} < \infty,$$

where the sum runs over all twin prime pairs. Moreover, he showed that the number of twin primes below x is bounded above by $Cx/(\log x)^2$ for some constant $C > 0$. This bound still goes to infinity as x increases; and computers indeed continued to find larger and larger twin primes.

In April 2013, a major breakthrough was realised by Yitang Zhang [18], who announced that he had found a number $C \approx 70000000$ such that

$$\liminf_n p_{n+1} - p_n \leq C,$$

where p_n is the n-th prime number, thus establishing bounded gaps between primes for the first time. Before 2013, the best known result was $\liminf_n (p_{n+1} - p_n)/\log n = 0$, obtained by Goldston, Pintz and Yıldırım in 2005 [7]. In November 2013, just half a year after Zhang's announcement, James Maynard severely decreased the bound when he released a paper showing that in fact $\liminf_n p_{n+1} - p_n \leq 600$ [10]. Maynard's proof was significantly simpler than Zhang's; it built on the more elementary techniques of Goldston, Pintz and Yıldırım. Moreover, Maynard's approach allowed him to prove finite gaps between prime m-tuples for every m, i.e.

$$\liminf_n p_{n+m} - p_n < \infty \quad \text{for every } m \in \mathbb{N}.$$

At approximately the same time, Terence Tao independently obtained similar results using the same ideas, which he published on his blog [15]. Subsequently, a group of mathematicians united as the *Polymath8b* group [13], in order to further reduce Maynard's bound. This led to many improvements, amongst which is the current world record

$$\liminf_n p_{n+1} - p_n \leq 246.$$

Peculiarly, however, current methods do not allow this bound to be reduced to any number smaller than 6. Therefore, the twin prime conjecture remains out of reach (for now).

1.2 The Hardy-Littlewood Conjecture

The approach followed by James Maynard was inspired by the following conjecture.

Conjecture 1.1 (Prime k-tuples conjecture)**.** *For any admissible set $\mathcal{H} = \{h_1, \dots, h_k\}$ of integers, there are infinitely many $n \in \mathbb{N}$ such that $n + h_1, \dots, n + h_k$ are all prime.*

A set $\mathcal{H} = \{h_1, \dots, h_k\}$ is called *admissible* if for every prime number ℓ the integers $h_1, \dots, h_k$ do not fill up all residue classes modulo ℓ. If $\mathcal{H}$ is not admissible then for every n there is an

h_i such that $\ell \mid n + h_i$, so being admissible is an obvious condition we need to impose on sets $\mathcal{H}$. The conjecture states that this obvious condition is indeed sufficient to ensure that there are infinitely many $n \in \mathbb{N}$ making all of $n + h_1, \dots, n + h_k$ prime. Also note that taking $\mathcal{H} = \{0, 2\}$ (which is obviously admissible) in the prime k-tuples conjecture yields the twin prime conjecture.

A proof for the prime k-tuples conjecture has not been found for any $k \geq 2$. Suppose that instead we can show for a fixed admissible set $\mathcal{H} = \{h_1, \dots, h_k\}$ of integers that there are infinitely many n such that at least m of $n + h_1, \dots, n + h_k$ are prime. Then we immediately obtain

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

This is the path we will take.

So why do we expect the prime k-tuples conjecture to be true? Let us use the Cramér model to say something about this. We let P_n be independent random Bernoulli$(1/\log n)$ variables for $n \in \mathbb{N}$. Then according to this model, we expect to have a number of

$$\mathbb{E} \sum_{n \leq x} \prod_{m=1}^{k} P_{n+h_m} = \sum_{n \leq x} \prod_{m=1}^{k} \frac{1}{\log(n + h_m)}$$

$n \leq x$ such that all of $n+h_1, \dots, n+h_m$ are prime. We have $\log(n+h) = \log n + h/n - h^2/n^2 + \dots = \log n + \mathcal{O}(1/n)$ and hence the above equals

$$\sum_{n \leq x} \frac{1}{(\log n)^k} \left(1 + \mathcal{O}\left(\frac{\log n}{n}\right)\right) = \frac{x}{(\log x)^k}(1 + o(1)).$$

This shows that

$$\mathbb{E} \sum_{n \leq x} \prod_{m=1}^{k} P_{n+h_m} \sim \frac{x}{(\log x)^k}.$$

Of course, this model is clearly wrong when $\mathcal{H}$ is not admissible. Indeed, we have made an obvious error in our model when assuming the P_n to be independent. In reality, if we want all of $n + h_1, \dots, n + h_{k-1}$ and $n + h_k$ to be prime, then for each prime p, n must lie outside the residue classes $\overline{-h_1}, \dots, \overline{-h_k}$ modulo p. This happens with probability $\prod_p (1 - v_{\mathcal{H}}(p)/p)$, where we let $v_{\mathcal{H}}(p)$ denote the the number of distinct residue classes modulo p occupied by $\mathcal{H}$. However, assuming independence, we computed this probability as $\prod_p (1 - 1/p)^k$. As a result, we have a correction factor

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

Note that for all large p we have $v_{\mathcal{H}}(p) = k$ and it is not difficult to see that the sum of the logarithms of the terms converges for each $k \geq 1$. As a result, $\mathfrak{S}(\mathcal{H}) \neq 0$ if and only if $v_{\mathcal{H}}(p) \neq p$ for all p. This is precisely the statement that $\mathcal{H}$ is admissible! It is therefore reasonable to make the following stronger conjecture, as was done long ago by Hardy and Littlewood [9].

Conjecture 1.2 (Hardy-Littlewood)**.** *If $\mathcal{H}$ is admissible, we have*

$$\#\{n \leq x \mid n + h \textit{ is prime } \ \forall h \in \mathcal{H}\} \sim \mathfrak{S}(\mathcal{H}) \frac{x}{(\log x)^k}.$$

1.3 The GPY sieve

Goldston, Pintz and Yıldırım (GPY) took, based on ideas of Selberg and others, a sieve-theoretic approach towards establishing bounded prime gaps. In this section we attempt to explain and motivate this.

Consider N to be a large number. Each number n in the interval $[N, 2N)$ has by the prime number theorem a probability of about $1/\log N$ to be prime. So when k is large enough (about $\log N$) and n is chosen uniformly at random in $[N, 2N)$ we have

$$\sum_{m=1}^{k} \mathbb{P}(n+m \text{ is prime}) > 1.$$

Then by a variant of the pigeonhole principle, $\mathbb{P}(\text{at least 2 of the } n+m \text{ are prime}) > 0$ and we have found two primes at most k apart. This gives us a prime gap of size $\ll \log N$, which is not as small as we wish. However, the above argument can be improved, since there is no reason we should choose n *uniformly* at random in $[N, 2N)$. Perhaps we can find a probability distribution that gives more weight to primes. Moreover, we can consider more general shifted numbers $n+h$ for $h \in \mathcal{H} = \{h_1, \dots, h_k\}$ instead of $n+1, \dots, n+k$.

Indeed, suppose we can find a sequence of non-negative numbers (w_n) determining a probability density $w_n / \sum_{N \le n < 2N} w_n$, such that for all sufficiently large N we have

$$\sum_{m=1}^{k} \mathbb{P}(n+h_m \text{ is prime}) = \sum_{m=1}^{k} \frac{\sum_{N \le n < 2N} w_n \mathbf{1}_{n+h_m \text{ is prime}}}{\sum_{N \le n < 2N} w_n} > \rho \tag{1}$$

for some $\rho > 1$. Then for infinitely many $n \in \mathbb{N}$ we have that $\#\{1 \le m \le k \mid n+h_m \text{ is prime}\} \ge \lceil \rho \rceil \ge 2$, thus proving bounded gaps between primes. The ratio in (1) is maximal when $w_n = \mathbf{1}_{\text{all } n+h_m \text{ prime}}$. However, if this is the case then showing that $\sum_{N \le n \le 2N} w_n$ is non-zero is as difficult as establishing the prime k-tuples conjecture. Therefore, the *weights* w_n are allowed to be a little less optimal with the hope that this will simplify estimations for the two sums. So we desire to choose "nice" weights w_n that are larger than, but as close as possible to, $\mathbf{1}_{\text{all } n+h_m \text{ prime}}$.

This is essentially a sieve problem. For all but finitely many $n \in [N, 2N)$, $P(n) = \prod_{h \in \mathcal{H}} (n+h)$ is a product of integers $n+h$ all smaller than N^2 and so $\mathbf{1}_{\text{all } n+h_i \text{ prime}} = \mathbf{1}_{p \mid P(n) \Rightarrow p > N}$. In general, given a sequence of integers $\mathcal{A}$ and a set of primes $\mathcal{P}$, an *upper bound sieve method* produces weights $\widetilde{w}_n \ge \mathbf{1}_{p \mid n \Rightarrow p \notin \mathcal{P}}$ of a fixed type such that

$$\sum_{p \mid n \Rightarrow p \notin \mathcal{P}} \mathbf{1}_{n \in \mathcal{A}} \le \sum_{n \in \mathcal{A}} \widetilde{w}_n$$

is almost as tight as possible for weights of that type. Each sieve method deals with a specific type of weight, since it is too general to optimize the above over all weights. In particular, with $\mathcal{P} = \{p \text{ prime} \mid p < N\}$ and $\mathcal{A} = \{P(n) \mid N \le n < 2N\}$ an upper bound sieve method thus gives weights $w_n = \widetilde{w}_{P(n)} \ge \mathbf{1}_{p \mid P(n) \Rightarrow p > N}$ such that the inequality

$$\sum_{N \le n < 2N} \mathbf{1}_{p \mid P(n) \Rightarrow p > N} = \sum_{p \mid n \Rightarrow p \notin \mathcal{P}} \mathbf{1}_{n \in \mathcal{A}} \le \sum_{n \in \mathcal{A}} \widetilde{w}_n = \sum_{N \le n < 2N} w_n \tag{2}$$

is tight, which is what we want. It thus makes sense to apply an upper bound sieve method.

GPY chose the Λ^2 *Selberg sieve*, which means they considered weights of the form

$$\widetilde{w}_n = \left(\sum_{d \mid n} \lambda_d \right)^2, \tag{3}$$

with $\lambda_d \in \mathbb{R}$ constrained by

$$\lambda_1 = 1, \;\; \lambda_d = 0 \;\text{ when some prime factor of } d \text{ is not in } \mathcal{P} \quad \text{and} \quad \lambda_d = 0 \text{ when } d \geq R \tag{4}$$

for some *threshold* R. Conveniently, being a square, $\widetilde{w}_n$ is non-negative. Also, divisor sums are well understood in multiplicative number theory, which makes it possible to compute sums involving these weights (i.e., these weights are "nice"). The condition "$\lambda_d = 0$ when some prime factor of d is not in $\mathcal{P}$" ensures that $\widetilde{w}_n \geq \mathbf{1}_{p|n \Rightarrow p \notin \mathcal{P}}$ and the threshold R is a common feature in all sieves; it is required to keep the error terms in the computations under control. Note that we want to have as much freedom as possible in choosing the λ_d's, so we would like R to be as large as possible. It turns out (see Remark 3.9) that R cannot be chosen larger than $\sqrt{N}$. With $R \leq \sqrt{N}$, we find for all d with a prime factor greater than N that $\lambda_d = 0$ as $d > R$, so we do not need to worry about the condition "$\lambda_d = 0$ when some prime factor of d is not in $\mathcal{P}$" anymore.

The problem is now to optimise (2) subject to (3) and (4). For general sieve problems with the Selberg sieve, the optimal *Selberg weight constituents* λ_d are known to be

$$\lambda_d = \mathbf{1}_{d \leq R}\mu(d)(\log R/d)^k.$$

This is not very surprising, as one can verify that $\sum_{d|m} \mu(d) \left(\log \frac{m}{d}\right)^k$ vanishes when m has more than k prime factors. The above choice of λ_d is a smoothed approximation to $\mu(d)(\log(P(n)/d)^k$. Note that, when $R \approx \sqrt{N}$, the values $\log R$ and $\log P(n)$ for $n \in [N, 2N)$ do differ by approximately a factor $2k$. This is irrelevant, as the ratio in (1) is unchanged when scaling w_n by a constant factor.

These weights appear to be promising, but they do not turn out to give the results we want. Goldston, Pintz and Yıldırım [7] noticed that the problem of maximising the ratio of the two sums in (1) is not exactly the same as the problem we try to solve with the Selberg sieve. Their idea was hence to consider a slightly more general form of the sieve weight constituents

$$\lambda_d = \mathbf{1}_{d \leq R}\mu(d)F(\log R/d),$$

where F is a polynomial, and attempt to find the best possible F. They found an $\ell > 0$ (e.g. $\ell = \sqrt{k}$) such that $F(x) = x^{k+l}$ gives better results than $F(x) = x^k$. With that they could almost prove bounded gaps between primes. Finally, Maynard and Tao took the idea of GPY one step further. They exploited the structure of the numbers $P(n)$ and successfully considered a more general form of the sieve weights

$$w_n = \left(\sum_{d_1|n+h_1, \ldots, d_k|n+h_k} \lambda_{d_1,\ldots,d_k} \right)^2. \tag{5}$$

We are back in the GPY case when the λ's depend only on the product $\prod_i d_i$. It is the extra freedom gained by allowing the λ's to depend on the individual divisors of the $n + h_i$ that makes the difference. We refer to the sieve-theoretic approach using the above weights as the *Maynard-Tao sieve*. These weights are compared to the GPY weights in more detail in Section 5.

1.4 Acknowledgements

This essay is largely based on the papers 'Small gaps between primes' by James Maynard [10] and 'Variants of the Selberg sieve, and bounded intervals containing many primes' by the Polymath8b group [13]. Every mention of these authors in this essay refers to these papers.

The author would like to express his thanks to Dr A.J. Harper for his valuable comments.

2 Establishing the sieve input

2.1 The duos Bombieri-Vinogradov and Elliott-Halberstam

In this subsection the the main arithmetic input of the Maynard-Tao sieve is discussed. We would like to find weights of the form (5) that maximise the ratio of the sums in (1), the first of which is

$$\sum_{m=1}^{k} \sum_{N \le n < 2N} \mathbf{1}_{n+h_m \text{ is prime}} w_n.$$

Estimating the inner sum for each m obviously requires some arithmetic information. As we will see in the proof of Lemma 3.7(ii), it turns out that we need an upper bound for

$$\sum_{q \le R^2} E(N; q, a), \tag{6}$$

where

$$E(N; q, a) := \left| \sum_{\substack{1 \le n < N \\ n \equiv a \bmod q}} \mathbf{1}_{n \text{ is prime}} - \frac{1}{\phi(q)} \sum_{1 \le n < N} \mathbf{1}_{n \text{ is prime}} \right|,$$

with different a coprime to q. We expect $E(N; q, a)$ to be small when q is sufficiently small compared to N, because there does not appear to be a reason why primes would prefer some residue classes $\bmod q$ above others. Recall that we would like to take $R \le \sqrt{N}$ as large as possible so we desire an upper bound for (6) for large R^2. The prime number theorem for arithmetic progressions says that for every $A > 0$ we have $E(N; q, a) \ll \frac{N}{(\log N)^A}$ for *fixed* q coprime to a. The implicit constant here does depend on q however, making this bound useless for (6) as R will depend on N. Yet one would expect a similar upper bound to exist for sufficiently small q depending on N. This has been proved for $q \le (\log N)^B$ for any B (a result known as the *Siegel-Walfisz theorem*), but we will need q to be much larger. We can do a lot better assuming the Generalised Riemann Hypothesis, but not unconditionally. Enter Bombieri [1] and Vinogradov [17], who proved a very strong result concerning the average values of $E(N; q, a)$, which is precisely what we need.

Theorem 2.1 (Bombieri-Vinogradov)**.** *For every positive constant A there exists a B such that*

$$\sum_{q \le \sqrt{x}/(\log x)^B} \max_{(a,q)=1} E(x; q, a) \ll \frac{x}{(\log x)^A}.$$

Note that the upper bound is the same as in the prime number theorem for arithmetic progressions, but we now sum over q. As can be seen from (6), this would allow us to take R (almost) up to $N^{1/4}$, which approaches the upper bound $\sqrt{N}$. Elliott and Halberstam [3] conjectured that in fact we can sum over q nearly all the way up to N, which would allow us to take R nearly up to $\sqrt{N}$ as desired.

Conjecture 2.2 (Elliott-Halberstam)**.** *For every positive constant A and every $\epsilon > 0$ we have*

$$\sum_{q \le x^{1-\epsilon}} \max_{(a,q)=1} E(x; q, a) \ll \frac{x}{(\log x)^A}.$$

Because it plays an important role in our story, we give such statements a name.

Definition 2.3. *We say the primes have* level of distribution θ *when for every* $A > 0$ *we have*

$$\sum_{q \le x^\theta} \max_{(a,q)=1} E(x;q,a) \ll \frac{x}{(\log x)^A}.$$

In the coming sections we consider an arbitrary level of distribution $0 < \theta < 1$, in order to see how our results improve when stronger versions of the Bombieri-Vinogradov are proved. Note that the Bombieri-Vinogradov theorem implies that the primes have level of distribution θ for every $\theta < 1/2$ and the Elliott-Halberstam conjecture asserts the primes have level of distribution θ for every $\theta < 1$. It has been shown by Friedlander and Granville [4] that the primes do *not* have level of distribution 1, so the Elliott-Halberstam conjecture is the best possible.

2.2 Computing sums of arithmetic functions

Let $\tau_k(n)$ be the number of ways to write an integer n as a product of k integers. These numbers will arise at various points as combinatorial coefficients. In order to deal with them, we need the following following lemma.

Lemma 2.4. *We have*

$$\sum_{n<R} \tau_k(n) \ll_k R(\log R)^{k-1} \text{ and } \sum_{n<R} \frac{\tau_k(n)}{n} \ll_k (\log R)^k.$$

Proof. We first prove a general formula. Suppose that $F(n) = \sum_{d|n} f(d)$ and $x \ge 0$. Then

$$\sum_{n \le x} F(n) = \sum_{n \le x} \sum_{d|n} f(d) = \sum_{d \le x} f(d) \sum_{n \le x,\ d|n} 1 = \sum_{d \le x} f(d) \lfloor x/d \rfloor = x \sum_{d \le x} \frac{f(d)}{d} + \mathcal{O}\left(\sum_{d \le x} |f(d)| \right).$$

We will apply this with $F(n) = \tau_k(n)$ and $f(n) = \tau_{k-1}(n)$. Indeed, if we want to write n as a product of k factors, we may first choose a divisor d and then the remaining $k-1$ divisors. Hence $\tau_k(n) = \sum_{d|n} \tau_{k-1}(n/d) = \sum_{d|n} \tau_{k-1}(d)$. We thus obtain

$$\sum_{n<R} \tau_k(n) = R \sum_{d \le R} \frac{\tau_{k-1}(d)}{d} + \mathcal{O}\left(\sum_{d<R} |\tau_{k-1}(d)| \right).$$

We now proceed by induction on k for $k \ge 2$. By the above we have

$$\sum_{n<R} \tau_2(n) = R \sum_{d \le R} \frac{1}{d} + \mathcal{O}\left(\sum_{d<R} \frac{1}{d} \right) \ll R \log R + \log R \ll R \log R.$$

Now assume that $\sum_{n<R} \tau_{k-1}(n) \ll R(\log R)^{k-2}$. Then we get

$$\sum_{n<R} \tau_k(n) = R \sum_{d<R} \frac{\tau_{k-1}(d)}{d} + \mathcal{O}(R \left(\log R\right)^{k-2}).$$

We use partial summation to evaluate the above sum. This yields

$$\begin{aligned} \sum_{d<R} \frac{\tau_{k-1}(d)}{d} &= \frac{1}{R} \sum_{d<R} \tau_{k-1}(d) + \int_1^R \sum_{d \le t} \tau_{k-1}(d) \frac{\mathrm{d}t}{t^2} \ll (\log R)^{k-2} + \int_1^R (\log t)^{k-2} \frac{\mathrm{d}t}{t} \\ &= (\log R)^{k-2} + \frac{1}{k-1} (\log R)^{k-1} \ll (\log R)^{k-1}. \end{aligned}$$

This completes the proof for the first inequality and simultaneously proves the second. $\square$

The following lemma will allow us to approximate sums of arithmetic functions of a certain kind. More general statements appear for example in [8] (lemmata 5.3 and 5.4).

Lemma 2.5. *Suppose that γ is a multiplicative arithmetic function such that for sufficiently large primes p we have $\gamma(p) - 1 \ll 1/p$ as $p \to \infty$ and consider a multiplicative function g defined on primes p by $g(p) = \gamma(p)/(p - \gamma(p))$. Then we have*

(i)

$$\sum_{n \leq x} \mu(n)^2 g(n) = \mathfrak{G} \log x + \mathcal{O}\left(1 + \mathfrak{G} \sum_p \frac{(\gamma(p) - 1)\log p}{p}\right),$$

where $\mathfrak{G} = \prod_p (p-1)/(p - \gamma(p))$ and

(ii)

$$\sum_{n \geq x} \mu(n)^2 g(n)^2 = \mathfrak{H} x^{-1} + \mathcal{O}(1),$$

where

$$\mathfrak{H} = \prod_p \left(1 + \frac{(\gamma(p)^2 - 1)p + \gamma(p) - \gamma(p)^2/p}{(p - \gamma(p))^2}\right).$$

Proof. We may assume that g is strongly multiplicative because of the factor $\mu(n)^2$. Let $f(n) = n\mu(n)^2 g(n)$ and $h(n) = f * \mu(n)$. Then we have $f(n) = \sum_{d|n} h(n)$ and we see that $h(n)$ is multiplicative. Moreover after a simple calculation we obtain for each prime p

$$h(p) = \frac{p\gamma(p) - p + \gamma(p)}{p - \gamma(p)}, \quad h(p^2) = \frac{-p\gamma(p)}{p - \gamma(p)} \quad \text{and} \quad h(p^k) = 0 \text{ for } k \geq 3.$$

Now if $H(s) = \sum_n h(n) n^{-s}$ converges absolutely at s then we have the Euler product $H(s) = \prod_p (1 + h(p)p^{-s} + h(p^2)p^{-2s} + \ldots)$. Similarly, we see that $\sum_n |h(n)| n^{-1}$ converges if and only if

$$\prod_p \left(1 + \frac{|h(p)|}{p} + \frac{|h(p^2)|}{p^2} + \ldots\right) = \prod_p \left(1 + \frac{\gamma(p) - 1}{p - \gamma(p)} + \frac{2\gamma(p)}{p(p - \gamma(p))}\right)$$

converges, in which case they are equal. This last product converges since $\gamma(p) - 1 \ll 1/p$, making the sum of the logarithms of the terms converge. Hence $H(s)$ converges absolutely at $s = 1$ and we have

$$H(1) = \prod_p \left(1 + \frac{\gamma(p) - 1}{p - \gamma(p)}\right) = \prod_p \frac{p - 1}{p - \gamma(p)} = \mathfrak{G}$$

by the Euler product. We now work out that

$$\sum_{n \leq x} \frac{f(n)}{n} = \sum_{n \leq x} \frac{1}{n} \sum_{d|n} h(d) = \sum_{d \leq x} \frac{h(d)}{d} \sum_{n \leq x/d} \frac{1}{n}.$$

The inner sum equals $\log(x/d) + \mathcal{O}(1)$ so that we obtain

$$\sum_{n \leq x} \frac{f(n)}{n} = \mathfrak{G} \log x + H'(1) - \log x \sum_{d \geq x} \frac{h(d)}{d} + \mathcal{O}(H(1)).$$

Next, note that $H'(1) = H(1)(\log H)'(1)$. Using the Euler product, we compute that

$$(\log H)'(1) = \sum_p \frac{(\gamma(p) - 1)\log p}{p}.$$

Therefore, it remains to bound the tail sum. We see that

$$\sum_{n} |h(n)| n^{-3/4} = \prod_{p} \left(1 + \frac{(\gamma(p) - 1)p + \gamma(p)}{p^{3/4}(p - \gamma(p))} + \frac{\gamma(p)}{p^{1/2}(p - \gamma(p))} \right)$$

converges since the sum of the logarithms of the factors does (again because $\gamma(p) - 1 \ll 1/p$). Hence the abscissa of absolute convergence $\sigma_a(H) \leq 3/4$. An elementary result in Dirichlet series (exercise 4 of Chapter 1.2 in [11]) says that $\sum_{n \geq x} |h(n)| n^{-\sigma} \ll x^{\sigma_a - \sigma + \epsilon}$ for every $\epsilon > 0$. In particular, we find with $\sigma = 1$ and $\epsilon = 1/8$ a sufficient bound $\sum_{n \geq x} |h(n)| n^{-1} \ll x^{-1/8}$.

For the second part of the lemma, we define $\widetilde{f}(n) = \mu(n)^2 g(n)^2 n^2$. Then we can write $\widetilde{f}(n) = \sum_{d|n} \widetilde{h}(d)$ and using that $\widetilde{h}(n) = \sum_{d|n} \mu(n/d) \widetilde{f}(d)$ we compute that

$$\widetilde{h}(p) = \frac{p^2 \gamma(p)^2}{(p - \gamma(p))^2} - 1, \;\; \widetilde{h}(p^2) = -\frac{p^2 \gamma(p)^2}{(p - \gamma(p))^2}, \;\; \widetilde{h}(p^k) = 0 \text{ for } k \geq 3.$$

Then

$$\sum_{n \geq D} \mu(n)^2 g(n)^2 = \sum_{n \geq D} \frac{1}{n^2} \sum_{d|n} \widetilde{h}(d) = \sum_{d} \frac{\widetilde{h}(d)}{d^2} \sum_{n \geq D/d} \frac{1}{n^2} = \frac{1}{D} \sum_{d} \frac{\widetilde{h}(d)}{d} + \mathcal{O}(1).$$

It remains to compute the most right-hand sum (and see that it converges). For this, we again use the Euler product:

$$\prod_{p} \left(1 + \frac{|\widetilde{h}(p)|}{p} + \frac{|\widetilde{h}(p)|}{p^2} \right) = \prod_{p} \left(1 + \frac{|\gamma(p) - 1||\gamma(p) + 1| p + 3|\gamma(p)| + |\gamma(p)^2/p|}{|p - \gamma(p)|^2} \right)$$

converges because $\gamma(p) - 1 \ll 1/p$. Hence we have

$$\sum_{d} \frac{\widetilde{h}(d)}{d} = \widetilde{H}(1) = \prod_{p} \left(1 + \frac{(\gamma(p)^2 - 1)p + \gamma(p) - \gamma(p)^2/p}{(p - \gamma(p))^2} \right),$$

as desired. □

Corollary 2.6. *For any integer $W \geq 0$ we have*

$$\sum_{n \leq x} \frac{\mu(n)^2}{\phi(n)} = \log x + \mathcal{O}(1), \qquad \sum_{n \leq x, (n,W)=1} \frac{\mu(n)^2}{\phi(n)} = \frac{\phi(W)}{W} \log x + \mathcal{O}(1),$$

$$\sum_{n \leq x} \frac{\mu(n)^2}{g(n)} \ll \log x, \qquad \sum_{n \leq x, (n,W)=1} \frac{\mu(n)^2}{g(n)} \ll \frac{\phi(W)}{W} \log x,$$

$$\sum_{n \geq x} \frac{\mu(n)^2}{\phi(n)^2} \ll \frac{1}{x}, \qquad \sum_{n \geq x, (n,W)=1} \frac{\mu(n)^2}{\phi(n)^2} \ll \frac{\phi(W)}{W} \frac{1}{x},$$

$$\sum_{n \geq x} \frac{\mu(n)^2}{g(n)^2} \ll \frac{1}{x}, \quad \text{and} \quad \sum_{n \geq x, (n,W)=1} \frac{\mu(n)^2}{g(n)^2} \ll \frac{\phi(W)}{W} \frac{1}{x},$$

where g is a multiplicative arithmetic function defined on the primes by $g(p) = p - 2$.

Proof. We apply (i) and (ii) from the above lemma with $\gamma(p)=1$, $\gamma(p)=\mathbf{1}_{p\nmid W}$, $\gamma(p)=\frac{p}{p-1}$ and $\gamma(p)=\mathbf{1}_{p\nmid W}\frac{p}{p-1}$ respectively. $\square$

Corollary 2.7. *When $g(n)\in\mathbb{R}_{\geq 0}$ for all n, we have, keeping the notation from Lemma 2.5,*

$$\sum_{n\leq x}\mu(n)^2\tau_k(n)g(n)\ll\mathfrak{S}^{k+1}(\log x)^{k+1}\quad\textit{and}\quad\sum_{n\leq x}\mu(n)^2\tau_k(n)^2g(n)\ll\mathfrak{S}^{3k+1}(\log x)^{3k+1}.$$

Proof. We prove the first part by induction. The case $k=0$ is Lemma 2.5. For the induction step, we note that

$$\begin{aligned}\sum_{n\leq x}\mu(n)^2\tau_k(n)g(n)&=\sum_{n\leq x}\mu(n)^2g(n)\sum_{d|n}\tau_{k-1}(d)=\sum_{d\leq x}\mu(d)^2\tau_k(d)g(d)\sum_{\substack{n\leq x/d\\(n,d)=1}}\mu(n)^2g(n)\\&\ll\mathfrak{S}\sum_{d\leq x}\mu(d)^2\tau_{k-1}(d)g(d)\log(x/d)\ll\mathfrak{S}\log x\sum_{d\leq x}\mu(d)^2\tau_{k-1}(d)g(d)\end{aligned}$$

by Lemma 2.4. For the second inequality, we use that for coprime d_1,d_2 we have $\tau_\ell(d_1)\tau_\ell(d_2)\leq\tau_{2\ell}(d_1d_2)$ to write

$$\begin{aligned}\sum_{n\leq x}\mu(n)^2\tau_k(n)^2g(n)&=\sum_{n\leq x}\mu(n)^2g(n)\sum_{d|n}\sum_{d_1d_2=d}\tau_{k-1}(d_1)\tau_{k-1}(d_2)\leq\sum_{n\leq x}\mu(n)^2g(n)\sum_{d|n}\tau_2(d)\tau_{k-1}(d)\\&=\sum_{d\leq x}\mu(d)^2\tau_2(d)\tau_{k-1}(d)g(d)\sum_{\substack{n\leq x/d\\(n,d)=1}}\mu(n)^2g(n).\end{aligned}$$

The right-hand sum is $\ll\log x$ by Lemma 2.5. We show by induction that $\sum_{d\leq x}\mu(d)^2\tau_2(d)\tau_{k-1}(d)g(d)\ll(\mathfrak{S}\log x)^{3k}$. The case $k=1$ is the first inequality in the statement. For $k>1$ we have

$$\begin{aligned}&\sum_{d\leq x}\mu(d)^2\tau_2(d)\tau_{k-1}(d)g(d)=\sum_{d\leq x}\mu(d)^2\tau_2(d)g(d)\sum_{e|d}\tau_{k-2}(e)\\&\quad=\sum_{e\leq x}\mu(e)^2\tau_{k-2}(e)\tau_2(e)g(e)\sum_{\substack{d\leq x/e\\(d,e)=1}}\mu(d)^2\tau_2(d)g(d)\ll(\mathfrak{S}\log x)^3\sum_{e\leq x}\mu(e)^2\tau_{k-2}(e)\tau_2(e)g(e),\end{aligned}$$

where we used for coprime and square-free e,d that $\tau_2(de)=2^{\omega(de)}=2^{\omega(d)}2^{\omega(e)}=\tau_2(d)\tau_2(e)$. $\square$

The powers of $\log x$ occurring in Corollary 2.7 are not optimal, but for our purposes any power of $\log x$ suffices. The following corollary serves as a substitute for Lemma 6.1 in [10].

Corollary 2.8. *Consider a piecewise differentiable function $G:[0,1]\to\mathbb{R}$ and let $G_{\max}=\sup_{t\in[0,1]}(|G(t)|+|G'(t)|)$. Then, keeping the notation from Lemma 2.5, we have*

$$\sum_{d\leq x}\mu(d)^2g(d)G\left(\frac{\log d}{\log z}\right)=\mathfrak{S}\log x\int_0^1G(t)\mathrm{d}t+\mathcal{O}\left(\mathfrak{S}\sum_p\frac{(\gamma(p)-1)\log p}{p}G_{\max}\right).$$

This is now an exercise in partial summation; a proof can be found in [6, Lemma 4].

3 The Maynard-Tao sieve

3.1 Reduction of the problem

In this subsection, the main results are stated and proved modulo two crucial propositions. The content of this subsection is based on Maynard's paper [10].

Let $\mathcal{H} = \{h_1, \ldots, h_k\}$ be an admissible set. First recall that our aim is to find weights of the form as in (5) that maximise the ratio of the two sums in (1). In order to deal with small prime factors, we only consider weights w_n that vanish outside of a prefixed residue class $v_0 \bmod W$, where $W = \prod_{p \le D_0} p$ and $D_0 = \log\log\log N$. Note that $W \le 4^{D_0} \ll (\log\log N)^2$. We choose v_0 such that $h_i + v_0$ is coprime to W for each i. We can do this by the Chinese remainder theorem since $\mathcal{H}$ is admissible. The upshot is that D_0 and W are small enough to be negligible compared to the other variables, but that still $D_0 \to \infty$ so that it exceeds all small prime factors. This is the only point in the story where we need $\mathcal{H}$ to be admissible; indeed, only modulo small primes can a set $\mathcal{H}$ of fixed size occupy all residue classes.

The aim is then to find an upper bound for

$$S_1 = \sum_{\substack{N \le n < 2N, \\ n \equiv v_0 \bmod W}} \left(\sum_{d_i | n+h_i \ \forall i} \lambda_{d_1,\ldots,d_k} \right)^2$$

and a lower bound for $S_2 = \sum_{m=1}^{k} S_2^{(m)}$, where

$$S_2^{(m)} = \sum_{\substack{N \le n < 2N, \\ n \equiv v_0 \bmod W}} \mathbf{1}_{n+h_m \text{ is prime}} \left(\sum_{d_i | n+h_i \ \forall i} \lambda_{d_1,\ldots,d_k} \right)^2 .$$

We choose the smooth weight constituents of the form

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

when $(\prod_{i=1}^{k} d_i, W) = 1$ and $\lambda_{d_1,\ldots,d_k} = 0$ otherwise. Here F is a piecewise differentiable function supported on $\mathcal{R}_k = \{(x_1, \ldots, x_k) \in [0,1]^k \mid \sum_{i=1}^{k} x_i \le 1\}$ and R is the sieve threshold. These weights generalise the GPY sieve weights, as is discussed in Section 5. For F as above, we define

$$I_k(F) = \int_0^\infty \cdots \int_0^\infty F(t_1, \ldots, t_k)^2 \mathrm{d}t_1 \ldots \mathrm{d}t_k,$$

$$J_k^{(m)}(F) = \int_0^\infty \cdots \int_0^\infty \left(\int_0^\infty F(t_1 \ldots, t_k) \mathrm{d}t_m \right)^2 \mathrm{d}t_1 \ldots \mathrm{d}t_{m-1} \mathrm{d}t_{m+1} \ldots \mathrm{d}t_k$$

and

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

The following proposition summarizes the output of the sieve method. The main arithmetic ingredients used to prove it are the prime number theorem and the Bombieri-Vinogradov theorem.

Proposition 3.1. *Suppose that the primes have level of distribution $\theta > 0$ and let $R = N^{\theta/2-\delta}$ for $\delta > 0$ sufficiently small. Then with weight constituents as defined above, we have*

$$\frac{S_2}{S_1} = (1 + o(1)) \left(\frac{\theta}{2} - \delta\right) \frac{\sum_{m=1}^{k} J_k^{(m)}(F)}{I_k(F)}.$$

Consequently, for any admissible set $\mathcal{H} = \{h_1, \dots, h_k\}$ of size k there are infinitely many n such that at least $m := \lceil M_k\theta/2 \rceil$ of $n + h_1, \dots, n + h_k$ are prime. In particular,

$$\liminf_{n} p_{n+m-1} - p_n \leq \max_{i,j} |h_i - h_j|.$$

This proposition is proved in the next subsection. Note that S_2/S_1 is indeed independent of W.

The remaining problem to find functions F providing a lower bound for M_k is purely analytic.

Proposition 3.2. *We have*

(i) $M_5 > 2$,

(ii) $M_{54} > 4$ *and*

(iii) $M_k > \log k - 2 \log \log k - 2$ *for k sufficiently large.*

This proposition is proved in Section 4. Parts (i) and (iii) were first proved by Maynard, whereas part (ii) was first proved by Polymath8b. We proceed to show how we can combine these two propositions to obtain results on prime gaps.

Theorem 3.3. *We have*

(i) $\liminf_n p_{n+1} - p_n \leq 270$,

(ii) $\liminf_n p_{n+1} - p_n \leq 12$ *if we assume the Elliott-Halberstam conjecture and*

(iii) $\liminf_n p_{n+m} - p_n \ll m^3 e^{4m}$ *(unconditionally) for m sufficiently large.*

Proof. For (i), we can take $\theta = 1/2 - \epsilon$ for $\epsilon > 0$ arbitrarily small by the Bombieri-Vinogradov theorem. It has been computed by Engelsma that there exists an admissible set $\mathcal{H}$ of size 54 with diameter $\max_{h,g \in \mathcal{H}} |h - g| = 270$ and that this is the smallest diameter of an admissible set of this size. Thus we are done by Propositions 3.2(i) and 3.1.

Part (ii) follows in the same way using Proposition 3.2(ii) and the fact that $\{0, 2, 6, 8, 12\}$ is an admissible set of size 5 with diameter 12.

For part (iii) we again take $\theta = 1/2 - \epsilon$ for $\epsilon > 0$ arbitrarily small. Then by Proposition 3.2(iii) we find for k sufficiently large

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

When $k \geq Cm^2 e^{4m}$ for C a sufficiently large constant independent of k and m, we get

$$\log k - 2 \log \log k - 2 \geq 4m + 2 \log m + \log C - 2 \log(4m + \log m + \log C) > 4m$$

when m is also sufficiently large. So taking $\epsilon = 1/k$, $k = \lceil Cm^2 e^{4m} \rceil$ and m sufficiently large, we find $M_k\theta/2 > m$. As an admissible set, we take the first k primes greater than k, i.e. $\mathcal{H} = \{p_{\pi(k)+1}, \dots, p_{\pi(k)+k}\}$. None of the elements is divisible by a number $\leq k$ and, having just k elements, $\mathcal{H}$ cannot occupy all residue classes modulo a number greater than k. Thus $\mathcal{H}$ is admissible and since

$$\pi(k \log k) \gg \frac{k \log k}{\log k + \log \log k} \gg k \gg k + \pi(k)$$

we see that $\mathcal{H}$ has diameter $\ll k \log k \ll m^3 e^{4m}$. This finishes the proof by Proposition 3.1. $\square$

The bound in (i) will be reduced to 246 in Section 6.

Now consider an admissible set $\mathcal{H} = \{h_1, \dots, h_k\}$ of size k such that for infinitely many n at least m of $n + h_1, \dots, n + h_k$ are prime. By the pigeonhole principle, one of the finitely many subsets of $\mathcal{H}$ of size m must obey the prime m-tuples conjecture. Hence Theorem 3.3 allows us to prove a partial prime m-tuples conjecture.

Theorem 3.4. *(Partial m-Tuples Theorem) A positive proportion of the admissible m-tuples obeys the prime m-tuples conjecture, i.e. for x sufficiently large in terms of m we have*

$$\frac{\textit{the number of admissible } m\textit{-tuples below } x \textit{ obeying the conjecture}}{\textit{the number of admissible } m\textit{-tuples below } x} \gg_m 1,$$

where a set $\mathcal{H}$ is said to lie below x *when $\mathcal{H} \subset \{1, 2 \dots, \lfloor x \rfloor\}$.*

Proof. We first find a lower bound for the number of admissible sets below x obeying the prime m-tuples conjecture. We take $k = Cm^3 e^{4m}$ so that by the remark above, every admissible set of size k contains a subset of size m that obeys the prime m-tuples conjecture. We may assume that $k \le x$. We proceed by finding a lower bound for the number of admissible k-tuples below x. For each prime $p \le k$, we remove from $\{1, 2, \dots, \lfloor x \rfloor\}$ the entire residue class $\bmod p$ that contains the fewest elements. This class has size at most $\lfloor x \rfloor / p$, so the remaining set $\mathcal{A}$ has size

$$|\mathcal{A}| \ge \lfloor x \rfloor \prod_{p \le k} (1 - 1/p) \gg_m \lfloor x \rfloor.$$

Moreover, since $\mathcal{A}$ does not occupy all residue classes $\bmod p$ for any $p \le k$, every size k subset of $\mathcal{A}$ is admissible. There are $\binom{|\mathcal{A}|}{k}$ such sets and each of them contains at least one (admissible) subset of size m obeying the prime m-tuples conjecture. Every such set of size m obeying the prime m-tuples conjecture is contained in precisely $\binom{|\mathcal{A}|-m}{k-m}$ size k sets in $\mathcal{A}$, so we obtain

$$\binom{|\mathcal{A}|}{k} \binom{|\mathcal{A}| - m}{k - m}^{-1} \asymp_m |\mathcal{A}|(|\mathcal{A}| - 1) \cdots (|\mathcal{A}| - m + 1) \gg_m |\mathcal{A}|^m \gg_m x^m$$

distinct admissible sets below x of size m that obey the prime m-tuples conjecture. On the other hand, there are only $\binom{\lfloor x \rfloor}{m} \asymp_m x^m$ sets of size m below x. $\square$

3.2 Applying the sieve

In this subsection we prove Proposition 3.1 by estimating both S_1 and $S_2^{(m)}$ in a way characteristic of the Selberg sieve. Let $[d_i, e_i]$ denote the least common multiple of two integers d_i and e_i. The main idea is to expand the square and swap the sums to obtain

$$S_1 = \sum_{\substack{d_1, \dots, d_k \\ e_1, \dots, e_k}} \lambda_{d_1, \dots, d_k} \lambda_{e_1, \dots, e_k} \sum_{\substack{N \le n \le 2N \\ n \equiv v_0 \bmod W \\ [d_i, e_i] | n + h_i \forall i}} 1 \tag{7}$$

and

$$S_2^{(m)} = \sum_{\substack{d_1, \dots, d_k \\ e_1, \dots, e_k}} \lambda_{d_1, \dots, d_k} \lambda_{e_1, \dots, e_k} \sum_{\substack{N \le n \le 2N \\ n \equiv v_0 \bmod W \\ [d_i, e_i] | n + h_i \forall i}} \mathbf{1}_{n + h_m \text{ is prime}}. \tag{8}$$

We then aim to diagonalise the quadratic forms that arose.

First recall that we set $R = N^{\theta/2-\delta}$. We will restrict the support of the $\lambda_{d_1,\ldots,d_k}$ to those $(d_1,\ldots,d_k)$ for which $(\prod_{i=1}^k d_i, W) = 1$ (to avoid complications arising from small prime factors), $\mu(\prod_{i=1}^k d_i)^2 = 1$ (for computational convenience) and $\prod_{i=1}^k d_i < R$ holds. The last condition is a k-dimensional analogue of the restricted support of the Selberg sieve. Later F will be chosen to be supported on $\mathcal{R}_k$ in order to satisfy this condition. We first diagonalise the quadratic forms.

Definition 3.5. *We define the* diagonalising weight constituents for S_1 *to be*

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

and for $S_2^{(m)}$ *to be*

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

where g is the strongly multiplicative function defined on primes p by $g(p) = p-2$.

We note here that $y_{r_1,\ldots,r_k} = y^{(m)}_{r_1,\ldots,r_k} = 0$ when $(r_1,\ldots,r_k)$ lies outside of the support of $\lambda_{r_1,\ldots,r_k}$. We see that we can retrieve the λ-variables from the y-variables as well. For $d_1,\ldots,d_k$ with $\mu(\prod_{i=1}^k d_i)^2 = 1$ we have

$$\sum_{\substack{r_1,\ldots,r_k \\ d_i|r_i\forall i}} \frac{y_{r_1,\ldots,r_k}}{\prod_{i=1}^k \phi(r_i)} = \sum_{\substack{r_1,\ldots,r_k \\ d_i|r_i\forall i}} \left(\prod_{i=1}^k \mu(r_i)\right) \sum_{\substack{e_1,\ldots,e_k \\ r_i|e_i\forall i}} \frac{\lambda_{e_1,\ldots,e_k}}{\prod_{i=1}^k e_i} = \sum_{e_1,\ldots,e_k} \frac{\lambda_{e_1,\ldots,e_k}}{\prod_{i=1}^k e_i} \sum_{\substack{r_1,\ldots,r_k \\ d_i|r_i\forall i \\ r_i|e_i\forall i}} \prod_{i=1}^k \mu(r_i) = \frac{\lambda_{d_1,\ldots,d_k}}{\prod_{i=1}^k \mu(d_i)d_i},$$

where in the last equality we need that $\sum_{s|e/d}\mu(s) = \prod_{p|e/d}(1-\mu(p)) = \mathbf{1}_{e=d}$.

In order to deal with the error terms that we will obtain, we first relate the maximum values of the y and λ variables.

Lemma 3.6. *We have*

$$\lambda_{\max} \ll y_{\max}(\log R)^k,$$

where $\lambda_{\max} = \sup_{d_1,\ldots,d_k} |\lambda_{d_1,\ldots,d_k}|$ and $y_{\max} = \sup_{r_1,\ldots,r_k}|y_{r_1,\ldots,r_k}|$.

The proof of this lemma is not very enlightening and will thus be omitted. The following lemma corresponds to lemmata 5.1 and 5.2 of [10].

Lemma 3.7. *We have*

(i)

$$S_1 = \frac{N}{W}\sum_{r_1,\ldots,r_k} \frac{y^2_{r_1,\ldots,r_k}}{\prod_{i=1}^k \phi(r_i)} + \mathcal{O}\left(\frac{y^2_{\max}N(\log R)^k}{WD_0}\right) \quad \textit{and}$$

(ii) for any $A > 0$

$$S_2^{(m)} = \frac{N}{\phi(W)\log N}\sum_{r_1,\ldots,r_k} \frac{(y^{(m)}_{r_1,\ldots,r_k})^2}{\prod_{i=1}^k g(r_i)} + \mathcal{O}\left(\frac{(y^{(m)}_{\max})^2\phi(W)^{k-2}N(\log N)^{k-2}}{W^{k-1}D_0}\right) + \mathcal{O}\left(\frac{y^2_{\max}N}{(\log N)^A}\right),$$

where $y^{(m)}_{\max} = \sup_{r_1,\ldots,r_k}|y^{(m)}_{r_1,\ldots,r_k}|$.

Remark 3.8. *The proof of (ii) requires both the prime number theorem and the distribution level of the primes to deal with the indicator function. This is the main arithmetic input of the sieve.*

Proof. We begin with (ii), where we start off with (8). If $W, [d_1, e_1], \dots, [d_k, e_k]$ are pairwise coprime, the conditions $n \equiv v_0 \bmod W$ and $n \equiv -h_i \bmod [d_i, e_i]\ \ \forall i$ mean that n has to lie in a fixed residue class modulo $q = W \prod_{i=1}^k [d_i, e_i]$ by the Chinese remainder theorem. Also, the indicator function only contributes when n is coprime to q. We will see that this case gives the main contribution. With the level of distribution θ and the prime number theorem in mind, we write when $\mu(q)^2 = 1$

$$\sum_{\substack{N \le n < 2N \\ n \equiv v_0 \bmod W \\ [d_i, e_i] | n + h_i \forall i}} \mathbf{1}_{n + h_m \text{ is prime}} = \frac{X_N}{\phi(q)} + \mathcal{O}(E(N, q)), \tag{9}$$

where $X_N = \pi(2N - 1) - \pi(N - 1)$ and

$$E(N, q) = 1 + \sup_{a:\ (a,q)=1} \left| \sum_{\substack{N \le n < 2N \\ n \equiv a \bmod q}} \mathbf{1}_{n \text{ is prime}} - \frac{1}{\phi(q)} \sum_{N \le n < 2N} \mathbf{1}_{n \text{ is prime}} \right|.$$

We note that the $+1$ in the definition of $E(N, q)$ accounts for the fact that we sum over $N \le n < 2N$ instead of $N - h_m \le n < 2N - h_m$.

What happens to the inner sum in (9) if $\mu(q)^2 \ne 1$? This is where we see the benefit of introducing W. Either for some i we have $(W, [d_i, e_i]) > 1$ or for some i, j we have $([d_i, e_i], [d_j, e_j]) > 1$. In the first case, we get $(W, v_0 + h_i) > 1$ contradicting our choice of v_0. In the second, we find a non-trivial common factor of $n + h_i$ and $n + h_j$ and hence of their difference $h_i - h_j$. As N is large, this factor also divides W putting us in the first case. As a result, the inner sum is zero in both cases. Since the inner sum in (9) is trivially zero when $d_m \ne 1$ or $e_m \ne 1$, we may write

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

where the primed sum means that we sum only over those d_i and e_i with $W, [d_1, e_1], \dots, [d_k, e_k]$ pairwise coprime.

We deal with the big oh term first. The restricted support of $\lambda_{d_1, \dots, d_k}$ implies that we only need to consider square-free $q < R^2 W$. Choosing $d_1, \dots, d_k, e_1, \dots, e_k$ so that $r/W = \prod_{i=1}^k [d_i, e_i]$ can be done in at most $\tau_{3k}(r)$ ways because

$$r/W = \prod_{i=1}^k \frac{d_i}{(d_i, e_i)} \frac{e_i}{(d_i, e_i)} (d_i, e_i)$$

since d_i and e_i are square-free. Using Lemma 3.6 we obtain

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

We have $WR^2 = WN^{\theta - 2\delta} \ll N^{\theta - \delta}$, so assuming the primes have level of distribution θ, we find for every $A > 0$ that $\sum_{r < R^2 W} E(N, r) \ll_A N/(\log N)^A$. By the triangle inequality, we also have the crude bound $E(N,r) \ll N/\phi(r)$. Applying this, Cauchy-Schwarz, Corollary 2.7 and the fact that $R^2W \ll N$, we obtain for any $A>0$ that the previously displayed is at most

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

In the main term we now want to separate the e_i and the d_i variables to make the substitution. To this end, note that $\mathbf{1}_{(d_i,e_i)=1} = \sum_{s_{i,j}|(d_i,e_j)} \mu(s_{i,j})$ and that for square-free d_i and e_i we have

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

where g is the strongly multiplicative function defined on primes p by $g(p)=p-2$. Both formulas are straightforward to verify. The main term of $S_2^{(m)}$ then becomes

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

Note here that we have removed the conditions $(d_i,d_j)=(e_i,e_j)=(d_i,W)=(e_i,W)=1$ because $\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}=0$ when one of these conditions if violated. Similarly, we see that $\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}=0$ when $s_{i,j}$ is not coprime to one of $u_i, u_j, s_{a,j}$ or $s_{i,b}$ for some $a\neq i$ or $b \neq j$. Hence we may restrict the sum over $s_{1,2},\dots,s_{k,k-1}$ to those $s_{i,j}$ coprime to all of $u_i,u_j,s_{a,j}$ and $s_{i,b}$ for all $a\neq i$ and $b\neq j$. We denote this sum by Σ^*.

Writing $a_j = u_j \prod_{i\neq j} s_{j,i}$ and $b_j = u_j \prod_{i\neq j} s_{i,j}$ we can now make the substitution to obtain

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

In the above we have used that $y_{a_1,\dots,a_k}=0$ when $a=\prod_{i=1}^k a_i$ is not square-free to factor the multiplicative arithmetic functions. In the case that for some i,j we have $s_{i,j}\neq 1$, we have $s_{i,j} > D_0$ since $(s_{i,j},W)=1$. Using that $y_{a_1,\dots,a_k}\neq 0$ only if $a_m=1$ and that $X_N \ll N/\log N$, we see that this case contributes at most

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

Using Corollary 2.6 we can upper bound this by

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

We have now shown that

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

At last, we now apply the prime number theorem $X_N = N/\log N + \mathcal{O}(N/(\log N)^2)$ to replace X_N by $N/\log N$ at the cost of an error of size

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

which is smaller than the first error term.

The proof of (i) is a k-dimensional analogue of the proof for diagonalising weights in the Λ^2 Selberg sieve. Moreover, it is very similar to the proof of (ii), so we will only sketch it briefly. We start with (7). Again the inner sum can be rewritten as running over all n in some fixed residue class mod q and so it equals $N/q + \mathcal{O}(1)$. Consequently,

$$S_1 = N \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k}} \frac{\lambda_{d_1,\dots,d_k} \lambda_{e_1,\dots,e_k}}{q} + \mathcal{O}\left(\sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k}} |\lambda_{d_1,\dots,d_k} \lambda_{e_1,\dots,e_k}| \right).$$

The main term can be treated very similarly to the main term of $S_2^{(m)}$. Since $\lambda_{d_1,\dots,d_k} = 0$ when $d = \prod_i d_i \geq R$ and there are $\tau_k(d)$ ways to write d as a product of k integers, the error term is

$$\ll \lambda_{\max}^2 \left(\sum_{d<R} \tau_k(d) \right)^2 \ll y_{\max}^2 R^2 (\log R)^{4k-2}$$

by Corollary 2.7 and Lemma 3.6. This is smaller than the error term given in the statement of the lemma, because $R^2 = N^\alpha$ for some $\alpha < 1$. ☐

Remark 3.9. *In the proofs of both (i) and (ii) we used that $R^2 \ll N$. If instead we had chosen $R^2 = N^{1+\eta}$ for some $\eta \geq 0$ then in our estimation of S_1 we would have picked up an error of size $y_{\max}^2 R^2 (\log R)^{4k-2} = y_{\max}^2 N^{1+\eta} (\log N)^{4k-2}$. However, by Corollary 2.7 the main term of S_1 is*

$$\frac{N}{W} \sum_{r_1,\dots,r_k} \frac{y_{r_1,\dots,r_k}^2}{\prod_{i=1}^k \phi(r_i)} \ll \frac{N y_{\max}^2}{W} \left(\sum_{r<R} \frac{\mu(r)^2}{\phi(r)} \right)^k \ll \frac{N y_{\max}^2 (\log R)^k}{W} \ll y_{\max}^2 N^{1+\eta} (\log N)^{4k-2},$$

so the supposed "error term" would dominate the main term making all our estimations worthless. The same problem would occur with $S_2^{(m)}$. Thus, indeed, as with the 1-dimensional Selberg sieve, we cannot take R to be larger than $\sqrt{N}$. This observation will also be important in Section 6 when we apply the "ϵ-trick".

The following lemma allows us to compare the new variables $y_{r_1,\dots,r_k}$ and $y_{r_1,\dots,r_k}^{(m)}$.

Lemma 3.10. *If $r_m = 1$ we have*

$$y_{r_1,\dots,r_k}^{(m)} = \sum_{a_m} \frac{y_{r_1,\dots,r_{m-1},a_m,r_{m+1},\dots,r_k}}{\phi(a_m)} + \mathcal{O}\left(\frac{y_{\max} \phi(W) \log R}{W D_0} \right).$$

The main idea is to substitute the expression for the λ-variables in terms of the y-variables in the definition of $y_{r_1,\dots,r_k}^{(m)}$. We omit the details.

Motivated by the GPY sieve weights, we now restrict the choice of $\lambda_{d_1,\dots,d_k}$ more by setting

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

where F is a piecewise differentiable function supported on $\mathcal{R}_k$. A benefit of taking a smooth function here is that we can transform sums into integrals using Corollary 2.7; and integrals are nice to work with. Furthermore, we again set $y_{r_1,\dots,r_k} = 0$ when either $\mu(r_1\cdots r_k)^2 = 0$ or $(r_1\cdots r_k, W) > 1$. We define

$$F_{\max} = \sup_{(x_1,\dots,x_k)\in[0,1]^k} |F(t_1,\dots,r_k)| + \sum_{i=1}^{k}\left|\frac{\partial F}{\partial t_i}(t_1,\dots,t_k)\right|.$$

The estimations of S_1 and $S_2^{(m)}$ can now be rewritten using Corollary 2.8.

Lemma 3.11. *With the y-variables defined as above, we have*

(i)

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

(ii)

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

Proof. Again we start with (ii). We find using our choice of y that whenever $y^{(m)}_{r_1,\dots,r_k} \neq 0$ we have

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

We see using Corollary 2.6 that $y^{(m)}_{\max} \ll \phi(W)F_{\max}(\log R)/W$. We now want to apply Corollary 2.8 with $\gamma(p) = 1$ if $p \nmid W\prod_{i=1}^k r_i$ and $\gamma(p) = 0$ otherwise. For this, note that

$$\sum_p \frac{|\gamma(p)-1|\log p}{p} = \sum_{p\mid W\prod_{i=1}^k r_i} \frac{\log p}{p} \ll \sum_{p<\log R}\frac{\log p}{p} + \sum_{p\mid W\prod_{i=1}^k r_i}\frac{\log p}{\log R} \ll \log\log N,$$

using Mertens and the fact that $\sum_{p\mid W\prod_{i=1}^k r_i}\log p \le \sum_{d\mid W\prod_{i=1}^k r_i}\Lambda(d) = \log(W\prod_{i=1}^k r_i) \ll \log R$. We now obtain

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

where

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

Substituting this into our expression for $S_2^{(m)}$, we claim to get

$$S_2^{(m)} = \frac{\phi(W)N(\log R)^2}{W^2 \log N} \sum_{\substack{r_1,\dots,r_k \\ (r_i,W)=1 \forall i \\ (r_i,r_j)=1 \forall i\neq j \\ r_m=1}} \left(\prod_{i=1}^k \frac{\mu(r_i)^2 \phi(r_i)}{g(r_i) r_i} \right) \left(F^{(m)}_{r_1,\dots,r_k} \right)^2 + \mathcal{O}\left(\frac{F_{\max}^2 \phi(W)^k N(\log R)^k}{W^{k+1} D_0} \right). \tag{10}$$

The main term comes from taking the square of the main term of $y^{(m)}_{r_1,\dots,r_k}$. Note that we have restricted the sum because $y^{(m)}_{r_1,\dots,r_k} = 0$ when either $(r_i, W) \neq 0$ for some i, $(r_i, r_j) \neq 0$ for some $i \neq j$ or $r_m \neq 1$. The largest error term obtained from squaring $y^{(m)}_{r_1,\dots,r_m}$ is the product between the main term and the error term. Using that $\phi(r_i) \leq r_i$, this contributes at most

$$\ll \frac{N(\log R)^2 \phi(W) F_{\max}^2}{D_0 W^2 \log N} \sum_{\substack{r_1,\dots,r_k \\ r_m=1}} \frac{\prod_{j=1}^k \mu(r_j)^2}{\prod_{i=1}^k g(r_i)} = \frac{N(\log R)^2 \phi(W) F_{\max}^2}{D_0 W^2 \log N} \left(\sum_{r<R} \frac{\mu(r)^2}{g(r)} \right)^{k-1}.$$

We have inserted the $\mu(r_j)^2$ and the condition $r_m = 1$ here because $y^{(m)}_{r_1,\dots,r_k} = 0$ when $\mu(r_j)^2 = 0$ for some j or $r_m \neq 1$. Using Corollary 2.6 again to evaluate the last sum, we obtain the same error term as in (10). We also have the two big oh terms already present in $S_2^{(m)}$. Using that $y^{(m)}_{\max} \ll \phi(W) F_{\max} \log(R)/W$, we see that the first of those is of the same size as the error term in (10). The second big oh term $y_{\max}^2 N/(\log N)^A \ll F_{\max}^2 N/(\log R)^A$ is seen to be smaller when A is sufficiently large.

Now if $(r_i, W) = (r_j, W) = 1$ but $(r_i, r_j) > 1$, then r_i and r_j must have a common prime factor $p > D_0$. Hence the requirement $(r_i, r_j) = 1$ in the sum in (10) can be removed at the cost of an error of size

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

which can be absorbed in our previous error term. In the above we use Corollary 2.6 after applying the upper bound $\phi(d) \leq d$.

Finally, we apply Corollary 2.8 for each variable r_i with $i \neq m$ in turn with $\gamma(p) = 1 + 1/(p^2 - p - 1)$ if $p \nmid W$ and $\gamma(p) = 0$ otherwise to evaluate

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

Using that

$$\sum_p \frac{|\gamma(p) - 1| \log p}{p} \ll \sum_{p|W} \frac{\log p}{p} \ll \log D_0$$

by Mertens, we finally obtain

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

as desired. Again the proof of (i) is similar to, but slightly simpler than the proof of (ii); it will be omitted. $\square$

Indeed, the error terms are smaller than the main terms (albeit by just a factor D_0), so we have proved Proposition 3.1.

4 Solving the variational problem

4.1 Small k

Our aim is to find functions F that approximate

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

where the supremum runs over the piecewise differentiable functions $F : [0,1]^k \to \mathbb{R}$ supported on $\mathcal{R}_k$ such that $I_k(F) \neq 0$. We first notice that both $\sum_{m=1}^k J_k^{(m)}(F)$ and $I_k(F)$ are symmetric in m, which implies that we can restrict our search to symmetric functions.

Lemma 4.1. *We have*

$$M_k = \sup_F \frac{kJ_k(F)}{I_k(F)},$$

where $J_k(F) := J_k^{(1)}(F)$ and the supremum runs over all symmetric piecewise differential functions F on $\mathcal{R}_k$.

Proof. Suppose F is piecewise differentiable. Then so is $|F|$, and $I_k(F) = I_k(|F|)$ and $J_k^{(m)}(|F|) \geq J_k^{(m)}(F)$. Hence we may choose a sequence F_n of non-negative piecewise differentiable functions such that $\sum_{m=1}^k J_k^{(m)}(F)/I_k(F) \to M_k$. After rescaling, we may assume $I_k(F_n) = 1$ for each n. We note that

$$Q(F) := M_k I_k(F) - \sum_{m=1}^k J_k^{(m)}(F)$$

is by definition a positive semi-definite quadratic form on $\mathcal{S}_k$ of which the induced bilinear form is clearly symmetric. Thus, $Q(F)$ obeys the triangle inequality. We define

$$\overline{F}_n(t_1, \ldots, t_k) = \frac{1}{\sqrt{k!}} \sum_{\sigma \in S_k} F(t_{\sigma(1)}, \ldots, t_{\sigma(k)}).$$

Since $Q(F)$ is symmetric, the triangle inequality gives $Q(\overline{F}_n) \leq Q(F)$ and we have $Q(\overline{F}_n) \to 0$ as well. Also, as each F_n is non-negative $I_k(\overline{F_n}) \geq I_k(F_n/k!) = 1/k!$ so $I_k(\overline{F})$ is bounded away from zero. Thus we find

$$\frac{kJ_k(\overline{F_n})}{I_k(\overline{F_n})} \to M_k$$

as desired. □

Since every continuous function can be well approximated by polynomials, we consider functions of the form

$$F(t_1, \ldots, t_k) = P(t_1, \ldots, t_k) \mathbf{1}_{(t_1, \ldots, t_k) \in \mathcal{R}_k},$$

where P is a polynomial. We know from the theory of symmetric polynomials that each can be expressed as a polynomial in the power sum polynomials $P_j = \sum_{i=1}^k t_i^j$. However we prefer to work with a linear basis, because this allows us to write $J_k(F)$ and $I_k(F)$ as a positive definite quadratic forms: if $P = a_1 f_1 + \dots a_k f_k$ for piecewise differentiable functions f_k and scalars a_k, then

$$I_k(F) = \sum_{1 \le i,j \le d} a_i a_j \int_{\mathcal{R}_k} f_i f_j \mathrm{d}t_1 \dots \mathrm{d}t_k$$

and

$$J_k^{(m)}(F) = \sum_{i \le i,j \le d} a_i a_j \int_0^1 \cdots \int_0^1 \left(\int_0^1 f_i \mathrm{d}t_m \right) \left(\int_0^1 f_j \mathrm{d}t_m \right) \mathrm{d}t_1 \dots \mathrm{d}t_{m-1} \mathrm{d}t_{m+1} \dots \mathrm{d}t_k.$$

If we can evaluate the two integrals above, we have computable symmetric positive-definite matrices M_2 and M_1 such that

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

where $\mathbf{a} = (a_1, \dots, a_d)$. Such expressions have a well-known maximum.

Lemma 4.2. *Suppose that M_1, M_2 are real symmetric positive-definite square matrices of the same size. Then the maximal value of $\frac{\mathbf{a}^T M_2 \mathbf{a}}{\mathbf{a}^T M_1 \mathbf{a}}$ is the largest eigenvalue of $M_1^{-1} M_2$. This maximum is attained when $\mathbf{a}$ is an eigenvector of $M_1^{-1} M_2$.*

Proof. Note that the fraction is unaffected when we change $\mathbf{a}$ by a non-zero scalar. Hence we may assume that $\mathbf{a}^T M_1 \mathbf{a} = 1$ and it remains to maximise $\mathbf{a}^T M_2 \mathbf{a}$ subject to the constraint $\mathbf{a}^T M_1 \mathbf{a} = 1$. This is a straightforward computation with Lagrange multipliers. □

It remains to find a choice of $f_1, \dots, f_d$ that allow us to compute the matrices M_2 and M_1. To this end, let a *signature* α be a non-increasing finite sequence of integers. Then we define

$$P_\alpha = \prod_{s(a_1, \dots, a_k) = \alpha} t_1^{a_1} \cdots t_k^{a_k},$$

where $s(a_1, \dots a_k)$ is the unique signature consisting of the integers $a_1, \dots, a_k$. Clearly, any product of the power sum polynomials can be written as a linear combination of P_α's, so they form a basis. It remains to find elements of this basis that can be well integrated on $\mathcal{R}_k$. The following lemma will provide inspiration.

Lemma 4.3 (Beta function identity)**.** *We have*

$$\int_{\mathcal{R}_k} \left(1 - \sum_{i=1}^k t_i \right)^a \prod_{i=1}^k t_i^{a_i} \mathrm{d}t_1 \dots \mathrm{d}t_k = \frac{a! a_1! \cdots a_k!}{(k + a + \sum_{i=1}^k a_i)!}$$

Proof. The crucial observation here is to use the substitution $v = t_1/(1 - \sum_{i=2}^k t_i)$ to obtain

$$\begin{aligned} \int_0^{1 - \sum_{i=2}^k t_i} \left(1 - \sum_{i=1}^k t_i \right)^a \left(\prod_{i=1}^k t_i^{a_i} \right) \mathrm{d}t_1 &= \left(\prod_{i=2}^k t_i^{a_i} \right) \left(1 - \sum_{i=2}^k t_i \right)^{a + a_i + 1} \int_0^1 (1-v)^a v^{a_1} \mathrm{d}v \\ &= \frac{a! a_1!}{(a + a_1 + 1)!} \left(\prod_{i=2}^k t_i^{a_i} \right) \left(1 - \sum_{i=2}^k t_i \right)^{a + a_i + 1}. \end{aligned}$$

for the inner integral. Applying this recursively we find the result. □

This lemma takes care of the integrals arising from I_k. We can compute the integrals arising from J_k as well, using first the displayed line in the proof to get rid of the squared inner integral and then applying the lemma with $k-1$ instead of k.

So good candidates for integration would be polynomial expressions in $1-P_1$ and P for some relatively simple polynomial P. (Note that P should not be a function of the sum P_1, since that would bring us to the GPY case; see Section 5.) Maynard chose $P = P_2$, the next simplest symmetric polynomial, and found $M_{105} > 4$. The Polymath8b group [13] managed to do a lot better by considering more general polynomials.

Lemma 4.4. *The polynomials $(1-P_1)^a P_\alpha$ with $a \in \mathbb{Z}_{\geq 0}$ and α a signature avoiding 1 form a linear basis for the symmetric polynomials.*

Proof. We begin with the basis consisting of the P_α's. Let the *length* of a signature α be the number of non-zero entries. Suppose that α is a signature of length ℓ containing 1 and let α' be the signature we obtain after removing 1 from α. Then one readily sees that $P_\alpha - P_1 P_{\alpha'}$ is a linear combinations of P_β's with length of β strictly smaller than ℓ. Thus, we recursively find that each P_γ can be expressed as a linear combination of $P_1^a P_\alpha$'s with a a non-negative integer and α a signature avoiding 1. Consequently they span the space and so do the $(1-P_1)^a P_\alpha$'s. Moreover, they are clearly linearly independent. $\square$

Thus, a reasonable choice would be to consider linear combinations of all $(1-P_1)^a P_\alpha$, where $a \geq 0$, α is a signature avoiding 1 and the total degree $a + \alpha_1 + \ldots + \alpha_\ell \leq d$ is bounded by some threshold d. In order to compute M_1 and M_2 efficiently, we choose a $d \in \mathbb{Z}_{>0}$ and make a table of the coefficients $c_{\alpha,\beta,\gamma}$ defined by

$$P_\alpha P_\beta = \sum_\gamma c_{\alpha,\beta,\gamma} P_\gamma.$$

for $\deg P_\alpha P_\beta \leq d$. Then we write the corresponding integrals as a linear combination of integrals over polynomials of the form $(1-P_1)^a P_\gamma$ and we evaluate those using Lemma 4.3 . Lastly, we find the largest eigenvalue of $M_1^{-1} M_2$. In order to speed up the computations, the Polymath8b group considered signatures α with only even entries. Doing this with $d = 23$ and $k = 54$, they found $M_{54} \geq 4.00238$, proving Proposition 3.2(ii).

4.2 Large k

This subsection discusses the proof of Proposition 3.2(iii). The previous arguments only allow lower bounds for M_k to be computed for fixed small k. Therefore, a different method is required to find a lower bound for all sufficiently large k. We present Tao's probabilistic approach [16] to this problem. Maynard [10] approaches the problem in a similar way, but uses a physical interpretation of the involved integrals instead. The main idea is to choose F to be of the form

$$F(t_1, \ldots, t_k) = \mathbf{1}_{\sum_i t_i \leq 1} k^{1/2} g(kt_1) \cdots k^{1/2} g(kt_k),$$

where $g : [0, \infty] \to \mathbb{R}$ is piecewise differentiable, supported on $[0, T]$ for some $T > 0$ and normalised so that $\int_0^\infty g(t)^2 \mathrm{d}t = 1$. This is a convenient choice as it almost makes the k-dimensional integrals a product of k one-dimensional integrals. Such F's are symmetric in $t_1, \ldots t_k$, so $J_k^{(m)}(F)$ is again independent of m. We write $I_k := I_k(F)$ and $J_k := J_k^{(1)}(F)$ for convenience. Note that indeed kJ_k/I_k is invariant under scaling of g. We also define $\mu = \int_0^\infty t g(t) \mathrm{d}t$.

Lemma 4.5. *With the above definitions, we have*

$$M_k \geq \left(\int_0^\infty g(t)\mathrm{d}t\right)^2 \left(1 - \frac{kT}{(k - T - k\mu))^2}\right)$$

provided $\mu < 1 - T/k$.

Proof. We see that

$$I_k \leq \int_{[0,\infty)^n} \prod_{i=1}^k kg(kt_i)^2 \mathrm{d}t_1 \cdots \mathrm{d}t_n = 1$$

by the normalisation assumption. Also, when $1 - \sum_{i=1}^{k-1} t_i \geq T/k$ the inner integral in J_k is

$$\int_0^\infty \mathbf{1}_{\sum_i t_i \leq 1} k^{1/2} g(kt_1) \cdots k^{1/2} g(kt_k) \mathrm{d}t_k = \left(\prod_{i=1}^{k-1} \sqrt{k} g(kt_i)\right) k^{-1/2} \int_0^\infty g(t)\mathrm{d}t$$

since in that case the range of the integral is $[0, T/k]$ and $\sum_{i=1}^k t_i \leq 1$ for $t_k \in [0, T/k]$. Thus we find

$$\begin{aligned} M_k \geq kJ_k &\geq \left(\int_0^\infty g(t)\mathrm{d}t\right)^2 \int_{\sum_{i=1}^{k-1} t_i \leq 1 - T/k} \left(\prod_{i=1}^{k-1} kg(kt_i)^2\right) \mathrm{d}t_1 \cdots \mathrm{d}t_{k-1} \\ &= \left(\int_0^\infty g(t)\mathrm{d}t\right)^2 \int_{\sum_{i=1}^{k-1} t_i' \leq k - T} \left(\prod_{i=1}^{k-1} g(t_i')^2\right) \mathrm{d}t_1' \cdots \mathrm{d}t_{k-1}'. \end{aligned}$$

The key observation at this point is that this can be interpreted as a probability. Let $X_1, \ldots, X_{k-1}$ be non-negative independent random variables on $\mathbb{R}$, each identically distributed with density function $g(t)^2$. Then the above translates into

$$M_k \geq \left(\int_0^\infty g(t)\mathrm{d}t\right)^2 \mathbb{P}(X_1 + \ldots + X_{k-1} \leq k - T).$$

We would like to use Chebyshev's inequality to lower bound this. Chebyshev says that most mass lies around the mean, so to get a lower bound we need the mean of $X_1 + \ldots + X_{k-1}$ to be smaller than $k - T$. As $\mathbb{E}(X_1 + \ldots + X_{k-1}) = (k-1)\mu$ and $\mu < 1 - \frac{T}{k}$, this is indeed the case. Since

$$\mathrm{Var}(X_1 + \ldots + X_{k-1}) = (k-1)\mathrm{Var}X_1 \leq (k-1)\mathbb{E}X_1^2 \leq (k-1)T\mu,$$

Chebyshev's inequality implies

$$M_k \geq \left(\int_0^\infty g(t)\mathrm{d}t\right)^2 \left(1 - \frac{(k-1)T\mu}{(k - T - (k-1)\mu))^2}\right) \geq \left(\int_0^\infty g(t)\mathrm{d}t\right)^2 \left(1 - \frac{kT}{(k - T - k\mu))^2}\right)$$

provided $\mu < 1 - T/k \leq 1$. □

Now it is a matter of maximizing the right-hand side for g and T subject to the constraints $\mu < 1 - T/k$ and $\int_0^\infty g(t)^2 \mathrm{d}t = 1$. Maynard carefully showed in Section 8 of [10] that

$$g(t) = \frac{c}{1 + At}, \quad c^2 = \frac{1 + AT}{T}, \quad 1 + AT = e^A \quad \text{and} \quad T = \log k - 2 \log\log k$$

obey both conditions and give $M_k \geq \log k - 2\log\log k - 2$, as desired.

The Polymath8b group more carefully optimised the above arguments in [13] and found $M_k \geq \log k - C$ for some constant C.

4.3 An upper bound for M_k

In order to see how much room for improvement is left, we find an upper bound for M_k. This result corresponds to Corollary 6.4 of [13].

Proposition 4.6. *For each k, we have $M_k \leq \frac{k}{k-1}\log k$.*

Proof. Let F be a piecewise differentiable function on $\mathcal{R}_k$ and consider a function $G_m : \mathcal{R}_k \to [0,\infty)$ such that $\int_0^\infty G(t_1,\dots,t_k)\mathrm{d}t_m \leq 1$ for every $t_1,\dots,t_{m-1},t_{m+1},\dots,t_k \geq 0$. Using Cauchy-Schwarz, we bring the F^2 inside to find

$$\left(\int_0^\infty F(t_1,\dots,t_k)\mathrm{d}t_m\right)^2 \leq \int_0^\infty \frac{F(t_1,\dots,t_k)^2}{G(t_1,\dots,t_k)}\mathrm{d}t_m \int_0^\infty G(t_1,\dots,t_k)\mathrm{d}t_m \leq \int_0^\infty \frac{F(t_1,\dots,t_k)^2}{G(t_1,\dots,t_k)}\mathrm{d}t_m$$

for each $t_1,\dots,t_{m-1},t_{m+1},\dots,t_k \geq 0$, which implies $J_k^{(m)}(F) \leq \int_{\mathcal{R}_k} F(t_1,\dots,t_k)^2/G(t_1,\dots,t_k)\mathrm{d}t_m$. So if we have such functions G_m for every $1 \leq m \leq k$, then we conclude that

$$\frac{\sum_{m=1}^k J_k^{(m)}(F)}{I_k(F)} \leq \sup_{\mathcal{R}_k} \sum_{m=1}^k \frac{1}{G_m(t_1,\dots,t_k)}.$$

We have this bound for every F, so for M_k as well. It remains to find suitable functions G_m. Here is a good choice: we take

$$G_m = \frac{k-1}{\log k}\frac{1}{1-\sum_{i\neq m} t_i + (k-1)t_m}.$$

The factor $(k-1)/\log k$ is chosen to make $\int_0^{1-\sum_{i\neq m} t_i} G(t_1,\dots,t_k)\mathrm{d}t_m = 1$ and we see that

$$\sum_{m=1}^k \frac{\log k}{k-1}\left(1-\sum_{i\neq m} t_i + (k-1)t_m\right) = \frac{k}{k-1}\log k,$$

thus proving the proposition. □

The possible values of M_k are thus within $\log k - 2\log\log k - 2 \leq M_k \leq \frac{k}{k-1}\log k$. Moreover, for specific values of k, we have computed stronger lower bounds. For $k = 105$, Maynard obtained $M_{105} \geq 4.0020697...$, which is not extremely close to the upper bound $M_{105} \leq 4.6987...$. The subsequent result $M_{54} \geq 4.00238...$ from the Polymath8b group, is much closer to the upper bound which equals 4.06024... for $k = 54$. The Polymath8b group [13] compared the two bounds for more small values of k and found the two to be similarly close in each case. In conclusion, the algorithm described in 4.1 for computing lower bounds for M_k appears to be close to optimal, so there may not be to be much to gain from attempting to optimise it any further.

5 Comparison with the GPY method

Before 2013, the standard approach towards proving bounded gaps used the classical sieve weights of the form $(\sum_{d|(n+h_1)\cdots(n+h_k)} \lambda_d)^2$. In their celebrated paper, Goldston, Pintz and Yıldırım were unfortunately unable to prove bounded gaps between primes. In this section we describe a reason for this and the fact that Maynard's approach was more successful. We first investigate in what way the Maynard weights generalise the GPY weights. As remarked by Maynard [10], it appears that $\lambda_{d_1,\dots,d_k}$ as defined in Proposition 3.1 is approximately defined in terms of the integral of F with respect to each coordinate. The following proposition makes this intuition precise.

Proposition 5.1. *For the weights $\lambda_{d_1,\dots,d_k}$ as defined in Proposition 3.1, there exists a constant C independent of $R, d_1, \dots, d_k$ such that*

$$\lambda_{d_1,\dots,d_k} = (1+o(1))C(\log R)^k \frac{\phi(W)^k}{W^k}\mu\left(\prod_{i=1}^k d_i\right)\int_{\frac{\log(d_1)}{\log R}}^1 \cdots \int_{\frac{\log d_k}{\log R}}^1 F(t_1,\dots,t_k)\mathrm{d}t_1\dots\mathrm{d}t_k.$$

Proof. We start by rewriting the weights as

$$\begin{aligned}\lambda_{d_1,\dots,d_k} &= \mu\left(\prod_{i=1}^k d_i\right)\prod_{i=1}^k \frac{d_i}{\phi(d_i)} \sum_{\substack{r_1,\dots,r_k\\ d_i|r_i\ \forall i\\ (r_i,W)=1\ \forall i}} \frac{\mu\left(\prod_{i=1}^k r_i\right)^2}{\prod_{i=1}^k \phi(r_i)} F\left(\frac{\log(r_1)}{\log R},\dots,\frac{\log(r_k)}{\log R}\right)\\ &= \mu\left(\prod_{i=1}^k d_i\right)\prod_{i=1}^k \frac{d_i}{\phi(d_i)} \sum_{\substack{s_1,\dots,s_k\\ (s_i,d_iW)=1\ \forall i\\ (s_i,s_j)=1\ \forall i\neq j}} \frac{\prod_{i=1}^k \mu(s_i)^2}{\prod_{i=1}^k \phi(s_i)} F\left(\frac{\log(d_1s_1)}{\log R},\dots,\frac{\log(d_ks_k)}{\log R}\right),\end{aligned}$$

where we may sum over the s_i up to R as F is supported on $\mathcal{R}_k$. Also, we exchanged $\prod_i \mu(d_i)$ for $\mu(\prod_i d_i)$ since $\lambda_{d_1,\dots d_k} = 0$ when $d_1,\dots,d_k$ are not all pairwise coprime. Using Corollary 2.8 for s_k with $g = \phi$ and $G(t) = F(\log(d_1s_1)/\log R,\dots,\log(d_{k-1}s_{k-1})/\log R, t+\log d_k/\log R)$, the sum is asymptotic to

$$\begin{aligned}&\log(R) \sum_{\substack{s_1,\dots,s_{k-1}\\ (s_i,d_iW)=1\ \forall i\\ (s_i,s_j)=1\ \forall i\neq j}} \frac{\prod_{i=1}^{k-1}\mu(s_i)^2}{\prod_{i=1}^{k-1}\phi(s_i)} \frac{\phi(Wd_k\prod_{i=1}^{k-1}s_k)}{Wd_k\prod_{i=1}^{k-1}s_k}\int_0^1 F\left(\frac{\log(d_1s_1)}{\log R},\dots,\frac{\log(d_{k-1}s_{k-1})}{\log R}, t+\frac{\log d_k}{\log R}\right)\mathrm{d}t\\ &= \log(R)\frac{\phi(W)}{W}\frac{\phi(d_k)}{d_k} \sum_{\substack{s_1,\dots,s_{k-1}\\ (s_i,d_iW)=1\ \forall i\\ (s_i,s_j)=1\ \forall i\neq j}} \frac{\prod_{i=1}^{k-1}\mu(s_i)^2}{\prod_{i=1}^{k-1}s_i}\int_0^1 F\left(\frac{\log(d_1s_1)}{\log R},\dots,\frac{\log(d_{k-1}s_{k-1})}{\log R}, t+\frac{\log d_k}{\log R}\right)\mathrm{d}t,\end{aligned}$$

where in the end we used that W, d_k and $\prod_{i=1}^{k-1}s_k$ are pairwise coprime by definition of the supports of $\lambda_{d_1,\dots,d_k}$ and the remaining sum. Now we can do a similar thing for s_{k-1}. We apply Corollary 2.8 with a different multiplicative arithmetic function g defined on primes by $g(p) = \gamma(p)/(p-\gamma(p))$ with $\gamma(p) = \mathbf{1}_{(p,d_{k-1}W\prod_{i\leq k-1}s_i)=1}p/(p+1)$. From the support of γ, we find that

$$\mathfrak{G}_\gamma = C_\gamma\frac{\phi(W)\phi(d_{k-1})}{Wd_{k-1}},$$

where C_γ is a constant not depending on any variable except γ. We see that we can apply the above argument for $s_{k-1}, s_{k-2},\dots,s_1$ in turn, but each time for an arithmetic function g defined in terms of a different function γ, to conclude that

$$\lambda_{d_1,\dots,d_k} = (1+o(1))C(\log R)^k\frac{\phi(W)^k}{W^k}\mu\left(\prod_{i=1}^k d_i\right)\int_{[0,1]^k} F\left(t_1+\frac{\log d_1}{\log R},\dots,t_k+\frac{\log d_k}{\log R}\right)\mathrm{d}t_1\dots\mathrm{d}t_k$$

for some constant C. Note that $\prod_i d_i/\phi(d_i)$ cancels out against $\prod_i \phi(d_i)/d_i$ obtained from this procedure. This proves the proposition. $\square$

Note that the quotient S_2/S_1 is unaffected when multiplying each w_n by a constant not depending on n. Also, the quotient S_2/S_1 as computed in Proposition 3.1 already has a factor $(1+o(1))$. Therefore, the *Tao weight constituents*

$$\widetilde{\lambda}_{d_1,\ldots,d_k} := \mu\left(\prod_{i=1}^k d_i\right)\widetilde{F}\left(\frac{\log d_1}{\log R},\ldots,\frac{\log d_k}{\log R}\right),$$

where $\widetilde{F}(t_1,\ldots,t_k)$ is a differentiable function supported on $\mathcal{R}_k$ with piecewise differentiable derivative $\frac{\partial^k}{\partial t_1\cdots\partial t_k}\widetilde{f}$, must give the same results as the $\lambda_{d_1,\ldots,d_k}$. Indeed, these are the weight constituents used by Tao [15] and by the Polymath8b group [13].

Recall that Goldston, Pintz and Yıldırım considered weights of the form $w_n = (\sum_{d|\prod_{h\in\mathcal{H}}(n+h)}\lambda_d)^2$. Considering such weights is equivalent to $\lambda_{d_1,\ldots,d_k}$ being a function of the product $d_1\cdots d_k$. With $d = d_1\cdots d_k$ and $\widetilde{f}(d) = \widetilde{F}(d_1,\ldots,d_k)$, the Tao sieve weights then become

$$\widetilde{\lambda}_{d_1,\ldots,d_k} = \mu(d)\widetilde{f}\left(\frac{\log d}{\log R}\right) = \mu(d) f\left(\log\frac{R}{d}\right),$$

where $f(t) = \widetilde{f}(1 - t/\log R)$. These are indeed the weight constituents (1.3) used by GPY. This shows in what way the smooth Maynard sieve weights generalise the smooth GPY sieve weights. Let us verify that we also obtain the same results.

Lemma 5.2. *If $F(t_1,\ldots,t_k) = G(t_1+\ldots+t_k)$ for some function G, then*

$$I_k(F) = \frac{1}{(k-1)!}\int_0^1 G(t)^2 t^{k-1}\mathrm{d}t \quad \textit{and} \quad J_k^{(m)}(F) = \frac{1}{(k-2)!}\int_0^1\left(\int_t^1 G(v)\mathrm{d}v\right)^2 t^{k-2}\mathrm{d}t.$$

Proof. Using the substitution $v = t_1+\ldots+t_k$ and then partial integration we find

$$\begin{aligned}\int_0^{1-\sum_{i=3}^k t_i}\int_0^{1-\sum_{i=2}^k t_i} G(t_1+\ldots+t_k)^2\mathrm{d}t_1\mathrm{d}t_2 &= \int_0^{1-\sum_{i=3}^k t_i}\int_{\sum_{i=2}^k t_i}^1 G(v)^2\mathrm{d}v\mathrm{d}t_2\\ &= \left[t_2\int_{\sum_{i=2}^k t_i}^1 G(v)^2\mathrm{d}v\right]_0^{1-\sum_{i=3}^k t_i} + \int_0^{1-\sum_{i=3}^k t_i} G\left(\sum_{i=2}^k t_i\right)^2 t_2\mathrm{d}t_2\\ &= \int_0^{1-\sum_{i=3}^k t_i} G\left(\sum_{i=2}^k t_i\right)^2 t_2\mathrm{d}t_2.\end{aligned}$$

Continuing this argument, we recursively find the desired equality for $I_k(F)$. The same works for $J_k(F)$. $\square$

For a piecewise differentiable function $G:\ [0,1]\to\mathbb{R}$ we define $I_k(G) := I_k(F)$ and $J_k(G) := J_k^{(1)}(F)$, where $F:\ \mathcal{R}_k\to\mathbb{R}$ is given by $F(t_1,\ldots,t_k) = G(t_1+\ldots+t_k)$. Note that for F of this form $J_k^{(m)}(F)$ is independent of m.

In order to prove bounded gaps between primes using F of this form, we would like to find a piecewise differentiable function G such that $kJ_k(G)/I_k(G) > 4$ for some k. Goldston, Pintz and Yıldırım considered $f(x) = x^{k+\ell}$, which up to an unimportant constant factor translates by the above into $G(x) = (1-x)^\ell$ (we get $\widetilde{f}(x) = (\log R)^{k+\ell}(1-x)^{k+\ell}$ and then we need to differentiate k times). For this G, the beta function identity and Lemma 5.2 can be used to compute that

$$\frac{kJ_k(G)}{I_k(G)} = \frac{2k(2\ell+1)}{(\ell+1)(k+2\ell+1)}.$$

With $\ell = \sqrt{k}$, say, we thus see that $kJ_k(G)/I_k(G)$ converges to 4 *from below* as $k \to \infty$ and this saves approximately a factor 2 over the $\ell = 0$ case. This proves bounded intervals between primes using the GPY method, provided the primes have level of distribution strictly greater than 1/2. This was indeed a result of GPY [7]: they were just "an epsilon" away from proving bounded prime gaps unconditionally. Unfortunately, however, GPY could never have obtained $kJ_k(G)/I_k(G) > 4$, as Soundararajan [14] noticed. The presented proof was found by Tao [15].

Proposition 5.3. *We have $kJ_k(G)/I_k(G) < 4$ for every $G: [0,1] \to \mathbb{R}$ piecewise differentiable and every $k \geq 2$.*

Proof. We define the function $g(t) = \int_t^1 G(v)\mathrm{d}v$ and note that $g'(t)^2 = G(t)^2$. We will show for any differentiable function $g: [0,1] \to \mathbb{R}$ with $g(1) = 0$ that

$$k \int_0^1 g'(t)^2 \frac{t^{k-2}}{(k-2)!}\mathrm{d}t < 4 \int_0^1 g(t)^2 \frac{t^{k-1}}{(k-1)!}$$

for each integer $k \geq 2$. Also define $f(t) = g(t)t^{k/2-1}$, which transforms the above into

$$\frac{k(k-1)}{4} \int_0^1 f(t)\mathrm{d}t < \int_0^1 (f'(t) - (k/2-1)t^{-1}f(t))^2 t\mathrm{d}t.$$

We note that $\int_0^1 2f(t)f'(t)\mathrm{d}t = \int_0^1 (f(t)^2)'\mathrm{d}t = f(1)^2 - f(0)^2 = 0$. Hence expanding the right-hand square and subtracting $(k/2-1)^2 \int_0^1 f(t)^2 \mathrm{d}t$ from both sides, we obtain

$$\frac{3k-4}{4} \int_0^1 f(t)^2 \mathrm{d}t < \int_0^1 f'(t)^2 t + \frac{(k-2)^2}{4} f(t)^2(t^{-1}-1)\mathrm{d}t.$$

We split the proof into two cases. If y is close to 1, more specifically if $y \geq 1 - \frac{2}{\sqrt{3k-4}}$, we show

$$\frac{3k-4}{4} \int_y^1 f(t)^2 \mathrm{d}t < \int_0^1 f'(t)^2 \mathrm{d}t.$$

If y is far from 1, more specifically if $y \leq 1 - \frac{3k-4}{k^2-k}$, we show that

$$\frac{3k-4}{4} \int_0^y f(t)^2 \mathrm{d}t < \frac{(k-2)^2}{4} \int_0^1 f(t)^2(t^{-1}-1)\mathrm{d}t.$$

If y is close to 1, we see using Cauchy-Schwarz that

$$f(y)^2 = \left| \int_0^y f'(t)\sqrt{t}\frac{1}{\sqrt{t}}\mathrm{d}t \right|^2 \leq \int_0^1 f'(t)^2 t\mathrm{d}t |\log(y)| = \int_0^1 f'(t)^2 \mathrm{d}t \log(1/y)$$

so that the integral

$$\int_y^1 f(t)^2 \mathrm{d}t \leq (y-1)\log y \int_0^1 f'(t)^2 \mathrm{d}t.$$

Since $(y-1)\log y \leq (y-1)^2 \leq \frac{4}{3k-4}$, we obtain the desired inequality. For y far from 1, we have

$$y \leq \frac{(k-2)^2}{(k-2)^2 + 3k - 4}, \quad \text{i.e.} \quad \frac{3k-4}{4} \leq \frac{(k-2)^2}{4}\left(\frac{1}{y} - 1\right).$$

This is sufficient to prove the desired inequality for y far from 1. It remains to prove that the two regions overlap, that is $(3k-4)^3 \leq 4(k^2-k)^2$ for all $k \geq 2$. This is elementary. $\square$

The optimal value of kJ_k/I_k was found by Pintz, Révész and Farkas [12] using Bessel functions.

Unfortunately, the above proof is not very insightful, so let us give a more intuitive argument why the Maynard sieve weights perform better. We again view the weights as a probability density. Then GPY found a probability density on $[N, 2N)$ such that for a randomly chosen $n \in [N, 2N)$ according to this density, $\mathbb{P}(n + h_i \text{ is prime}) \asymp 1/k$ as $k \to \infty$ and $N \to \infty$. It is unfortunate that the constant that bounds $k\mathbb{P}(n + h_i \text{ is prime})$ is precisely the critical value 4, but what we really desire is for $k\mathbb{P}(n + h_i \text{ is prime})$ to increase to infinity as $k \to \infty$.

Recall that the GPY weights were constructed to place most mass on those n where $\prod_{h\in\mathcal{H}}(n + h_i)$ has at most k prime factors, i.e. those where all $n + h_i$ are prime. Since we need w_n to be chosen such that $\sum_{m=1}^{k} \mathbb{P}(n + h_m \text{ is prime})$ is computable, we make a certain error. It would be helpful if these deviations of w_n from $\mathbf{1}_{\text{all } n+h_i \text{ prime}}$ occur when at least many of the $n+h_i$ are prime; after all, we require $\mathbb{P}(n+h_i \text{ is prime})$ to be large for each i. Indeed, this is what Maynard ensured by allowing the sieve weights to depend on the individual $n + h_i$'s rather than their product. In fact, we found a probability density such that $\mathbb{P}(n + h_i \text{ is prime}) \asymp (\log k)/k$.

To illustrate this, recall that the GPY weights resemble the square of $\sum_{d|\prod(n+h_i)} \mu(d)(\log \frac{\prod_{i=1}^{k}(n+h_i)}{d})^k$ which vanishes when not all $n + h_i$ are prime. Similarly it follows from the chosen form of the function $F(t_1, \ldots, t_k) = \mathbf{1}_{\sum t_i \leq 1}\tilde{g}(kt_1)\cdots\tilde{g}(kt_k)$ for the Tao weights for large k, that they resemble the square of

$$\left(\sum_{d_1|n+h_1} \mu(d_1) \log \frac{n + h_1}{d_1}\right) \cdots \left(\sum_{d_k|n+h_k} \mu(d_k) \log \frac{n + h_k}{d_k}\right) \tag{11}$$

which vanishes when not all $n + h_i$ are prime powers. Suppose now that $n + h_2, \ldots, n + h_k$ are prime, but $n + h_1$ is not a prime power. Then $\sum_{d_1|n+h_1} \mu(d_1) \log \frac{n+h_1}{d_1} = 0$ but $\sum_{d_i|n+h_i} \mu(d_i) \log \frac{n+h_i}{d_i} = \log(n + h_i)$ for $i > 1$. As w_n can be thought of as being a smoothed approximation to the square of (11), the first factor of w_n is small. However, as opposed to the GPY case, this is now partially compensated by the larger values at $n+h_2, \ldots, n+h_k$. Moreover, we see that w_n is larger when more of the $n + h_i$ are prime.

Why does this make such a big difference? GPY found $\mathbb{P}(n+h_i \text{ is prime}) \asymp 1/k$ which differs a factor of about k from the optimal value $\mathbb{P}(n + h_i \text{ is prime}) = 1$. Hence the smooth weights w_n approximate $\mathbf{1}_{\text{all } n+p_i \text{ prime}}$ rather roughly and this becomes more rough as k increases. As a result, when k increases we can benefit more from choosing w_n better at those n where it deviates from $\mathbf{1}_{\text{all } n+p_i \text{ prime}}$. This is why we could save a factor $\log k$ that increases with k.

6 Further improvements to the Maynard-Tao sieve

After Maynard's paper came out, the Polymath8b group managed to make a series of improvements to the sieve, proving amongst other results the following theorem.

Theorem 6.1. *We have* $\liminf p_{n+1} - p_n \leq 246$.

The improvements can be divided into two categories:

(1) finding functions that give better approximations to (the equivalent of) M_k and

(2) enlarging the allowed support of F in order to consider a larger set of functions.

Improvements of the first kind initially led to $\liminf p_{n+1} - p_n \leq 270$ and have already been considered in Section 4. In this section, a way to slightly enlarge the support of the allowed functions is discussed, resulting in a proof of Theorem 6.1.

6.1 A trivial way to enlarge the support of F

A more or less trivial way to enlarge the support of F was already noted by Maynard in his paper [10]. He remarks that in the proof of lemmata 3.7(ii) and 3.11(ii) $\lambda_{d_1,\ldots,d_k}$ is only really required to be supported on those $(d_1,\ldots,d_k)$ with $\prod_{i\neq m} d_i \leq R$. This is because we need to set $d_m = 1$. For such d_i we have

$$\left(\prod_i d_i\right)^{k-1} = \prod_m \prod_{i\neq m} d_i \leq R^k,$$

so we can use $R^{k/(k-1)}$ instead of R in our estimates for S_1 and S_2. However, we do need to be careful here: as noted in Remark 3.9 this works only when $2k(\theta/2)/(k-1) < 1$, i.e. when $k > 1/(1-\theta)$. Hence this idea is of little use when assuming the Elliott-Halberstam conjecture. If indeed $k \geq 1/(1-\theta)$, the above reasoning implies that instead of M_k it suffices to consider

$$M_k' = \sup_F \frac{\sum_{m=1}^k \int_0^\infty \cdots \int_0^\infty \left(\int_0^\infty F(t_1,\ldots,t_k)\mathrm{d}t_m\right)^2 \mathrm{d}t_1 \ldots \mathrm{d}t_{m-1}\mathrm{d}t_{m+1}\ldots \mathrm{d}t_k}{\int_0^\infty \cdots \int_0^\infty F(t_1,\ldots,t_k)^2 \mathrm{d}t_1 \ldots \mathrm{d}t_k},$$

where now F runs over the piecewise differentiable functions supported on the slightly larger set

$$\mathcal{R}_k' = \{(t_1,\ldots,t_k) \subset [0,1]^k \mid t_1 + \ldots t_{m-1} + t_{m+1} \ldots + t_k \leq 1 \text{ for each } 1 \leq m \leq k\}.$$

Unfortunately, Maynard also remarks in his paper [10] that this enlarged support gives negligible numerical benefits. We thus need to find a better way to enlarge the support.

6.2 Sieving on an ϵ-enlarged support

Note that a precise asymptotic for $S_2^{(m)}$ as established in Lemma 3.11(ii) is not really necessary; it suffices to find a lower bound. This simple idea allows us to further enlarge the support. The following corresponds to Theorem 3.12(i) from [13]. The proof has been adapted to fit our setup.

Proposition 6.2 (ϵ-trick)**.** *Suppose the primes have level of distribution θ. Then for any $\epsilon > 0$ such that $1+\epsilon < 1/\theta$ and any piecewise differentiable function F supported on $(1+\epsilon)\mathcal{R}_k$ with $\lambda_{d_1,\ldots,d_k}$ defined in terms of F, we have*

$$S_2^{(m)} \geq (1+o(1))\frac{\phi(W)^k N(\log R)^{k+1}}{W^{k+1}\log N} J_{k,\epsilon}^{(m)}(F),$$

where

$$J_{k,\epsilon}^{(m)}(F) := \int_{(1-\epsilon)\mathcal{R}_{k-1}} \left(\int_0^\infty F(t_1,\ldots,t_k)\mathrm{d}t_m\right)^2 \mathrm{d}t_1 \ldots \mathrm{d}t_{m-1}\mathrm{d}t_{m+1}\ldots \mathrm{d}t_k.$$

Remark 6.3. *Note that the enlarged support of F comes at the cost of a smaller integration range $(1-\epsilon)\mathcal{R}_{k-1}$ so it is not a priori clear that the ϵ-trick gives better results. This loss is minimised by applying ideas from the previous subsection to allow the squared inner integral in $J_{k,\epsilon}^{(m)}$ to be unrestricted. Also, note that letting $\epsilon \to 0$ we get the original asymptotic from Lemma 3.11(ii) back, so the ϵ-trick cannot do worse.*

Proof. As in the proof of Lemma 3.7(ii) we have

$$S_2^{(m)} = \sum_n \mathbf{1}_{n+h_m \text{ is prime}} w_n = \sum_n \mathbf{1}_{n+h_m \text{ is prime}} \left(\sum_{\substack{d_1,\ldots,d_k \\ d_m=1}} \lambda_{d_1,\ldots,d_k}\right)^2,$$

where now we define $\lambda_{d_1,\ldots,d_k}$ in terms of a piecewise differentiable function F supported on $(1+\epsilon)\mathcal{R}_k$. We now have no hope of obtaining the same asymptotic as before. However, we can obtain a lower bound if we introduce a new k-dimensional Selberg weight $\widetilde{w}_n$ in terms of $\widetilde{\lambda}_{d_1,\ldots,d_k}$, which we define similar to $\lambda_{d_1,\ldots,d_k}$, but in terms of another function $\widetilde{F}$ supported also on $(1+\epsilon)\mathcal{R}_k$. The upper bound we take is $w_n = \widetilde{w}_n + (w_n - \widetilde{w}_n) \geq w_n - \widetilde{w}_n$. We choose $\widetilde{F}$ to agree with F when $\sum_{i\neq m} t_k > 1-\epsilon$, which will imply that the contributions from $\sum_{i\neq m} t_k > 1-\epsilon$ cancel each other out nicely. More specifically, we get as in the proof of Lemma 3.7(ii) that

$$\sum_n \mathbf{1}_{n+h_m \text{ is prime}}(w_n - \widetilde{w}_n) = \frac{X_N}{\phi(W)} \sideset{}{'}\sum_{\substack{d_1,\ldots,d_k \\ e_1,\ldots,e_k \\ d_m=e_m=1}} \frac{\lambda_{d_1,\ldots,d_k}\lambda_{e_1,\ldots,e_k} - \widetilde{\lambda}_{d_1,\ldots,d_k}\widetilde{\lambda}_{e_1,\ldots,e_k}}{\prod_{i=1}^k \phi([d_i,e_i])}$$
$$+ \mathcal{O}\left(\sum_{\substack{d_1,\ldots,d_k \\ e_1,\ldots,e_k \\ d_m=e_m=1}} |\lambda_{d_1,\ldots,d_k}\lambda_{e_1,\ldots,e_k} - \widetilde{\lambda}_{d_1,\ldots,d_k}\widetilde{\lambda}_{e_1,\ldots,e_k}| E(N,q) \right)$$

In the big oh term, we now only need to consider those $d_1,\ldots,d_k$ and $e_1,\ldots,e_k$ where either $\prod_{i\neq m} d_i \leq R^{1-\epsilon}$ or $\prod_{i\neq m} e_i \leq R^{1-\epsilon}$, since $\lambda_{d_1,\ldots,d_k}\lambda_{e_1,\ldots,e_k}$ and $\widetilde{\lambda}_{d_1,\ldots,d_k}\widetilde{\lambda}_{e_1,\ldots,e_k}$ agree in the other case. Thus, we find for q such that $E(N,q)$ appears with a non-zero coefficient in the error term that

$$q = W \prod_{i\neq m} [d_i,e_i] \leq W \prod_{i\neq m} d_i e_i \leq W R^{1-\epsilon} R^{1+\epsilon} = W R^2.$$

This shows that we do have control over this error term, provided $2(1+\epsilon)\theta/2 < 1$, i.e. $1+\epsilon < 1/\theta$, as was noted in Remark 3.9. The other error terms produced in the proofs of Lemma 3.7(ii) and Lemma 3.11(ii) depend only on $\log R$ so they form no threat. These proofs now work in exactly the same way to give

$$\sum_n \mathbf{1}_{n+h_m \text{ is prime}}(w_n - \widetilde{w}_n) \sim \frac{\phi(W)^k N (\log R)^{k+1}}{W^{k+1}\log N} \int_{(1-\epsilon)\mathcal{R}_{k-1}} \left(\int_0^1 F(t_1,\ldots,t_k)^2 - \widetilde{F}(t_1,\ldots,t_k)^2 \mathrm{d}t_m \right)^2.$$

To finish the proof, we just choose $\widetilde{F}$ to be zero when $\sum_{i\neq m} t_k < 1-\epsilon$. Note that we only require $\widetilde{F}$ to be *piecewise* smooth. □

When estimating S_1 we change nothing, so we just work with the weights w_n. Again by Remark 3.9, however, we need $\epsilon + 1 < 1/\theta$ for this to work. Together with the previous proposition, this proves the following theorem.

Theorem 6.4. *Suppose the primes have level of distribution θ and consider $\epsilon \geq 0$ such that $1+\epsilon < 1/\theta$. Also, define*

$$M_{k,\epsilon} := \sup_F \frac{\sum_{m=1}^k J_{k,\epsilon}^{(m)}(F)}{I_k(F)},$$

where the supremum runs over all piecewise differentiable F supported on $(1+\epsilon)\mathcal{R}_k$. If we can show that $M_k > 2\ell/\theta$ for a positive integer ℓ, then for any admissible set $\mathcal{H} = \{h_1,\ldots,h_k\}$ there are infinitely many n such that at least $\ell+1$ of the the translates $n+h_1,\ldots,n+h_k$ are prime.

Unfortunately, the condition $1+\epsilon < 1/\theta$ means that the ϵ-trick will be useless once we assume the Elliott-Halberstam conjecture. For $\theta < 1$ the ϵ-trick will help us lower the bound.

Proposition 6.5. *We have $M_{50,1/4} > 4$.*

Proof. In order to obtain these lower bounds, we need a slight modification to our approach in Section 4 since we need to integrate over $(1-\epsilon)\mathcal{R}_k$ and $(1+\epsilon)\mathcal{R}_k$ instead of $\mathcal{R}_k$. With the trivial substitution $t_i' = t_i/(1+\epsilon)$ for each $1 \le i \le k$, we can apply Lemma 4.3 to integrate polynomials $(1+\epsilon-P_1)^a P_\alpha$ over $(1+\epsilon)\mathcal{R}_k$. For integrating over $(1-\epsilon)\mathcal{R}_k$ we would like to work with polynomials $(1-\epsilon-P_1)P_\alpha$, however. To fix this, we just convert between the two by noting $(1+\epsilon-P_1)^a = (2\epsilon+1-\epsilon-P_1)^a$ and applying the binomial theorem. Taking polynomials of the form $(1-P_1)^a P_\alpha$ with α consisting only of even integers and total degree bounded by $d = 25$ the Polymath8b group computed, in the same way as in the previous section, for $k = 50$ and $\epsilon = 1/25$ that indeed $M_{50,1/25} \ge 4.00124....$ $\square$

The value $\epsilon = 1/25$ was found by trying different values of ϵ of the kind $\epsilon = 1/m$. Note that $M_{50} \le 50\log(50)/49 = 3.99186... < 4$ so this result could not have been achieved without the ϵ-trick! The admissible set of size 50 with smallest diameter has diameter 246, as was found by Engelsma. This finishes the proof of Theorem 6.1.

7 Limitations arising from the Parity Problem

We have shown that $\liminf p_{n+1} - p_n \le 12$ provided the Elliott-Halberstam conjecture holds. In this section, we give a (heuristic) argument why no similar sieve theoretic approach can without additional techniques ever be applied to prove the twin prime conjecture. This is a manifestation of the *parity problem* of sieves, which roughly says that, individually, general sieve methods are unable to accurately count the size of sets that contain only integers with an even number of prime factors or only integers with an odd number of them.

In order to find the implications of the parity problem in the case of finding primes within short intervals, we examine our approach more closely. We started with an admissible set $\mathcal{H} = \{h_1, \ldots, h_k\}$ and proceeded to show that $\sum_{N \le n < 2N} \#\{i \mid n + h_i \text{ is prime}\} w_n > \sum_{N \le n < 2N} w_n$ for cleverly chosen non-negative weights w_n. This means that for some $N \le n < 2N$, we have $w_n > 0$ and $\#\{i \mid n + h_i \text{ is prime}\} \ge 2$, hence

$$\sum_{N \le n < 2N} \mathbf{1}_A(n) w_n > 0, \tag{12}$$

where $A = \{n \mid \text{ at least two of } n + h_1, \ldots, n + h_k \text{ are prime}\}$.

The first thing we did when proving Lemma 3.7 (i), was to remark that

$$S_1 = \sum_{\substack{d_1,\ldots,d_k \\ e_1,\ldots,e_k}} \lambda_{d_1,\ldots d_k} \lambda_{e_1,\ldots,e_k} \sum_{\substack{N \le n < 2N \\ n \equiv v_0 \bmod W \\ [d_i, e_i] \mid n + h_i}} 1$$

after which one wants to approximate the inner sum. When approximating $S_2^{(m)}$ we obtain the same expression with $\mathbf{1}_{n+h_m \text{ is prime}}$ instead of 1 in the inner sum. All our results rely on the estimations of these inner sums. One can imagine that different sieve methods might rewrite their analogue to S_1 and S_2 differently and instead require estimations for sums of the form

$$\sum_{\substack{N \le n < 2N \\ n \equiv a \bmod q}} f(n + h) \tag{13}$$

for $h \in \mathcal{H}$, $q \leq N^{1-\epsilon}$ and f a multiplicative function. In fact, the Polymath8b group [13] arrives at such sums with f different from an indicator function. In order to estimate those sums, one needs to use a generalised version of the Bombieri-Vinogradov theorem or the Elliott-Halberstam conjecture.

Principle 7.1 (Parity problem for primes in bounded intervals)**.** *The twin prime conjecture cannot be proved using a sieve-theoretic approach with weights w_n that relies solely on estimations of sums of the form (13) and arrives at a conclusion of the form (12).*

Derivation. Suppose that we could use such an approach with weights w_n to prove the twin prime conjecture. Assume without loss of generality that $\mathcal{H} = \{0, 2\}$, so that $A = \{n \mid n,\ n+2 \text{ are prime}\}$. Then consider additional weights $v_n = 1 - \lambda(n)\lambda(n+2)$, where $\lambda(n) = (-1)^{\Omega(n)}$ is the Liouville function. The Liouville function is similar to the Möbius function and, similar to "Möbius cancellation", we have "Liouville cancellation", meaning that $\sum_{n \leq N} \lambda(n) = o(N)$. In fact, this is $\ll N/\log^{10} N$ by the prime number theorem. Moreover, one would expect that

$$\sum_{\substack{N \leq n < 2N \\ n \equiv a \bmod q}} \lambda(n)\lambda(n+2) = o\left(\frac{N}{\phi(q)}\right) \quad \text{and} \quad \sum_{\substack{N \leq n < 2N \\ n \equiv a \bmod q}} f(n+h)\lambda(n)\lambda(n+2) = o\left(\sum_{\substack{N \leq n < 2N \\ n \equiv a \bmod q}} f(n+h)\right)$$

when $q \leq N^{1-\epsilon}$ and f does not "conspire" with the Liouville function. We now consider the sieve with weights $w_n v_n$ instead. Then we get for example

$$S_1 = \sum_{\substack{d_1, \ldots, d_k \\ e_1, \ldots, e_k}} \lambda_{d_1, \ldots d_k} \lambda_{e_1, \ldots, e_k} \sum_{\substack{N \leq n < 2N \\ n \equiv v_0 \bmod W \\ [d_i, e_i] | n + h_i}} 1 \cdot v_n$$

and more generally one would need to approximate sums of the form

$$\sum_{\substack{N \leq n < 2N \\ n \equiv a \bmod q}} f(n+h) v_n.$$

However, by the previous arguments we have

$$\sum_{\substack{N \leq n < 2N \\ n \equiv a \bmod q}} f(n+h) v_n = (1 + o(1)) \sum_{\substack{N \leq n < 2N \\ n \equiv a \bmod q}} f(n+h)$$

and so we obtain the same estimates for our new weights. Consequently, we must have $\sum_{N \leq n < 2N} \mathbf{1}_A(n) w_n v_n > 0$ as well, which is impossible as v_n is constructed to be zero on A. $\square$

In fact, as shown in [13], a similar argument denies such a sieve-theoretic proof for $\liminf_n p_{n+1} - p_n \leq 4$. The parity problem does not completely rule out a proof of the twin prime conjecture using sieve methods, however. Friendlander and Iwaniec [5], for example, circumvented the parity problem and used sieve methods to find an asymptotic for the number of prime values of the polynomial $X^2 + Y^4$. In fact, one parity obstruction was overcome in this essay as well by finding a lower bound for the number of primes in short intervals. Instead of sieving for primes directly (which would be problematic by the parity problem), we sieved for numbers $P(n) = \prod_{n \in \mathcal{H}} (n+h)$ with few prime factors and used a pigeonhole-type of argument to relate this to the existence of primes in short intervals. Thus, one may still cherish the hope of finding a clever idea that defeats the parity obstruction to the twin prime conjecture.

References

[1] E. Bombieri. On the large sieve. Mathematika, 12:201-225, 1965.

[2] V. Brun. Über das Goldbachsche Gesetz und die Anzahl der Primzahlpaare, Arch. f. Math. og Naturv., 34(8):319, 1915.

[3] P. D. T. A. Elliott and H. Halberstam. A conjecture in prime number theory. Symposia Mathematica, Vol. IV (INDAM, Rome, 1968/69), pages 59-72. Academic Press, London, 1970.

[4] J. Friedlander and A. Granville. Limitations to the equi-distribution of primes. I. Ann. of Math. (2), 129(2):363-382, 1989.

[5] J. Friedlander and H. Iwaniec. The polynomial $X^2 + Y^4$ captures its primes. Ann. of Math. (2), 148(3):945-1040, 1998.

[6] D. A. Goldston, S. W. Graham, J. Pintz and C. Y. Yıldırım. Small gaps between products of two primes. Proc. London Math. Soc., 98(3):741-774, 2009.

[7] D. A. Goldston, J. Pintz, and C. Y. Yıldırım. Primes in tuples I. Ann. of Math. (2), 170(2):819-862, 2009.

[8] H. Halberstam and H.E. Richert. Sieve Methods. Academic Press, 1974.

[9] G.H. Hardy and J.E. Littlewood. Some problems of Parititio Numerorum (III): On the expression of a number as a sum of primes. Acta Math., 44:170, 1922.

[10] J. Maynard. Small gaps between primes. Ann. of Math., 181(1):383-413, 2015.

[11] H.L. Montgomery and R.C. Vaughan. Multiplicative Number Theory. Cam. Univ. Press, 2006. ISBN 9781107405820.

[12] J. Pintz, S. Révész and B. Farkas. On the optimal weights function in the Goldston-Pintz-Yıldırım method for finding small gaps between consecutive primes. Numb. Th., Analysis, and Comb., Proceedings of the Turan Memorial Conference, de Gryuter, 2013.

[13] D.H.J. Polymath. Variants of the Selberg sieve, and bounded intervals containing many primes. RMS, 1(12), 2015.

[14] K. Soundararajan. Small gaps between prime numbers: the work of Goldston-Pintz-Yıldırım. Bull. Amer. Math. Soc. 44:1-18, 2007.

[15] T. Tao. Polymath8b: Bounded intervals with many primes, after Maynard. Blog post on `https://terrytao.wordpress.com/`, 2013.

[16] T. Tao. The prime tuples conjecture, sieve theory, and the work of Goldston-Pintz-Yıldırım, Motohashi-Pintz, and Zhang. Blog post on `https://terrytao.wordpress.com/`, 2013.

[17] A. I. Vinogradov. The density hypothesis for Dirichlet L-series. Izv. Akad. Nauk SSSR Ser. Mat., 29: 903-934, 1965. Corrigendum: Izv. Akad. Nauk SSSR Ser. Mat., 30:719-720, 1966.

[18] Y. Zhang. Bounded gaps between primes. Ann. of Math., 179:1121-1174, 2014.