BOUNDED GAPS BETWEEN PRIMES

TONY FENG

MAY 2, 2014

AN ESSAY WRITTEN FOR

PART III OF THE MATHEMATICAL TRIPOS

2013-2014

ACKNOWLEDGMENTS

I thank Professor Timothy Gowers for setting this essay and for his advice during its writing. I would never have had this opportunity without the support of the Churchill Foundation. I am grateful to my friends for their help and companionship, both in mathematics and in life. Finally, I thank my parents for their unwavering love and support.

CONTENTS

1 INTRODUCTION

1.1 TOWARDS SMALL PRIME GAPS

Over the past year, a remarkable story has unfolded in analytic number theory. One of the oldest and most fundamental open problems in the field is to understand the additive structure of the prime numbers. What does the spacing between prime numbers look like? Are there very small or very large gaps, and if so, how frequently do they appear? Despite the many advances that mathematicians have made in studying the large-scale distribution of the primes, delicate additive questions such as these have remained intractable ... until almost exactly one year ago, when a remarkable breakthrough from the unlikeliest of sources shook the mathematical world.

Let p_n denote the nth prime number. According to the prime number theorem, the gap between p_n and p_{n+1} is about $\log n$ on average, so in particular the prime numbers become sparser and sparser as one marches up through the integers. However, we can still ask if extremely small gaps occur infinitely often. The twin prime conjecture, one of the oldest and most famous problems in the subject, predicts that there are infinitely many pairs of primes separated by the smallest possible gap of 2.

Conjecture 1.1.1 (Twin prime conjecture)**.**

$$\liminf_{n\to\infty} p_{n+1} - p_n = 2.$$

This is a just a special case of a far-reaching conjecture of Hardy and Littlewood describing the frequency of prime gaps of any sizes. The Hardy-Littlewood conjecture predicts not only how often twin primes occur, but also how often any finite tuple of the form $(n + h_1, n + h_2, \ldots, n + h_k)$ consists entirely of prime numbers. But even though analytic number theorists have believed for many years that they know the answers to these questions, progress towards *proving* the existence of small gaps between primes has been slow. As recently as 2005, the problem of establishing infinitely many bounded gaps between primes was considered by many mathematicians to be "hopeless" ([7]). For years, mathematicians studied the more modest quantity

$$\Delta_1 := \liminf_{n\to\infty} \frac{p_{n+1} - p_n}{\log n},$$

which compares prime gaps to their expected sizes. The Prime Number Theorem trivially implies that $\Delta_1 \leq 1$. Early work by Hardy-Littlewood and others chipped away at this bound using the generalized Riemann hypothesis, but it was Erdös who made the first unconditional improvement, showing that $\Delta_1 \leq 1 - c$ for some calculable positive constant c. Ricci improved this bound to $15/16$, and Bombieri and Davenport lowered it to $\frac{1}{2}$. From then the bound crawled down, through the efforts of many mathematicians, to 0.4665, then 0.4571, then 0.4542, and so on ... (see [7] for a more complete discussion of the historical progress). It seemed as if mathematicians would be fighting for each incremental gain for the foreseeable future, but in 2005 Goldston, Pintz, and Yildirim released a landmark paper [7] showing that $\Delta_1 = 0$. They also show that bounded gaps between primes follow from the Elliott-Halberstam conjecture.

Theorem 1.1.2 (Goldston-Pintz-Yildirim, 2009)**.** *If the Elliott-Halberstam conjecture is true then*

$$\liminf_{n\to\infty} p_{n+1} - p_n \leq 16.$$

We will give a technical definition of the Elliott-Halberstam conjecture later; for now, we just say that it is an assertion that primes are strongly equidistributed in arithmetic progressions. The Elliott-Halberstam conjecture is believed to be very difficult, but not quite as difficult as the twin prime conjecture. (Many problems in analytic number theory can be described in terms of a sum of arithmetic functions, and their difficulty can be roughly classified by how many factors of the indicator function of the prime numbers appear in a summand. In terms of this classification, the twin prime conjecture features two appearances and the Elliott-Halberstam conjecture only one.)

Following this initial success, mathematicians attacked the problem of establishing that

$$\liminf_{n\to\infty} p_{n+1} - p_n < \infty$$

with renewed vigor. After the experts suffered many setbacks, most people still believed that the problem was well beyond current methods. Then, in 2013, the unlikeliest of figures burst onto the scene to prove them wrong.

The unexpected breakthrough came from Yitang Zhang, a virtual unknown in the mathematics world. After receiving his Ph.D. from Purdue in 1991 (in a subject distinct from number theory), Zhang was unsuccessful in obtaining an academic job. He all but disappeared from the academic community, pursuing odd jobs as an accountant and even a Subway worker to support himself. He finally secured a Lecturer position at the University of New Hampshire in 1999. From then, he worked on several prominent problems in analytic number theory in relative secrecy and isolation, until announcing a fantastic proof of bounded gaps between primes in 2013 ([20]). See [11] for these and more details on Zhang's remarkable story.

Theorem 1.1.3 (Zhang, 2013)**.** *We have*

$$\liminf_{n\to\infty} p_{n+1} - p_n \leq 70,000,000.$$

Zhang's result, an impossible success story in the face of adversity, was all the more impressive in that it improved *technically* on the work of experts, in ways that the experts had thought impossible. In the 1980s, Bombieri, Friedlander, and Iwaniec - three of the most renowned analytic number theorists of the modern era - had collaborated on very difficult results that, when combined with the work of Goldston-Pintz-Yildirim, fell just short of establishing bounded gaps between primes. Zhang, working in virtual isolation, attempted the same route that had foiled the experts, and pushed their methods just far enough to cross the finish line.

Mathematicians quickly converged to discuss Zhang's work. Most prominently, Terence Tao gathered an online community of mathematicians to sharpen and optimize the various components of Zhang's argument. After several months, this massively collaborative online project, dubbed "Polymath 8," pared down Zhang's bound from $70,000,000$ to about 5000. They also found many simplifications to Zhang's proof, but it was still a difficult and intricate argument.

In November 2013, another huge breakthrough rocked the mathematics community. James Maynard, a postdoctoral researcher with a newly minted Ph.D., found a way to dramatically improve the apparatus used by Goldston-Pintz-Yildirim. Whereas Zhang had made technical improvements to the existing arguments, Maynard discovered a fundamental alteration that made them much more efficient. Maynard's work was completely independent of Zhang's, and led to a much shorter and easier proof of a stronger result:

Theorem 1.1.4 (Maynard, 2013). *We have*

$$\liminf_{n\to\infty} p_{n+1} - p_n \leq 600.$$

Furthermore, if the Generalized Elliott-Halberstam conjecture is true then

$$\liminf_{n\to\infty} p_{n+1} - p_n \leq 12.$$

In fact, his work says much more: it automatically proves that there are multiple primes among $\{n + h_1, \ldots, n + h_k\}$ infinitely often, for any choice of $h_1, \ldots, h_k$ that is "admissible." We will give a precise definition of admissibility later, but suffice it to say for now that it means there is no obvious obstruction to the numbers $\{n + h_1, \ldots, n + h_k\}$ all being prime for infinitely many n. For instance, it is obvious that $n + 1$ and $n + 2$ cannot both be prime infinitely often, since one of the two must be divisible by 2.

Theorem 1.1.5 (Maynard, 2013). *For any $m > 0$, there exists k such that for any admissible k-tuple $(h_1, \ldots, h_k)$, there are at least m primes among the integers $\{n + h_1, \ldots, n + h_k\}$ for infinitely many n.*

In particular, this implies that $\liminf p_{n+m} - p_n$ is finite *for any m*! It is even possible to give an explicit bound for this difference.

Theorem 1.1.6 (Maynard, 2013). *For any $m > 0$,*

$$\liminf_{n\to\infty} p_{n+m} - p_n \leq m^3 e^{4m+5}.$$

Maynard's results were simultaneously and independently discovered by Tao, who then initiated another communal project, Polymath 8b, devoted towards improving the result further. As this essay is being written, the Polymath 8b project is still ongoing, but it has entered its final stages. Through collaboration from a wide pool of contributors, Polymath 8b improved on several facets of Maynard's results, with the following consequences in particular.

Theorem 1.1.7 (Polymath 8b, 2014). *We have*

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

Furthermore, if the Generalized Elliott-Halberstam conjecture is true then

$$\liminf_{n\to\infty} p_{n+1} - p_n \leq 6.$$

Theorem 1.1.8 (Polymath 8b, 2014)**.** *For any $m > 0$,*

$$\liminf_{n\to\infty} p_{n+m} \ll e^{3.817m}.$$

In this paper, we will recount this wonderful story. We will outline the arguments of Goldston-Pintz-Yildirm and briefly touch on Zhang's. Our focus, however, is on the work of Maynard-Tao and Polymath 8b, which are based on the earlier arguments but technically supersede them. We will give a complete exposition of Maynard's proof, and sketch Tao's independent approach. Afterwards, we will describe some of the major achievements of Polymath 8b.

We actually prove a slightly more general result. Insteading of considering primes represented by translations $\{n + h_1, \ldots, n + h_k\}$, we consider those represented by general linear forms $\{g_1 n + h_1, \ldots, g_k n + h_k\}$. The work of Maynard and Tao goes through to this more general setting with only minor technical modifications, as has already been observed by the experts [8]. We give the general proof for completenes, as it has not yet appeared in the literature but is already being used in research papers (cf. [19], [5]). For instance, we give a complete proof of the following more general form of the Maynard-Tao theorem cited in [19].

Theorem 1.1.9. *For any $m > 0$, there exists a k such that any admissible k-tuple of linear forms $(g_1 n + h_1, \ldots, g_k n + h_k)$ contains at least m primes for infinitely many n.*

1.2 MODELING THE PRIMES

It is natural to ask *why* we expect small prime gaps. And why should we even care? To number theorists, the fact that such a concise and simple assertion about prime numbers has proved so challenging is already compelling enough. More substantially, however, the Twin Prime Conjecture is significant because it reflects deep heuristics in analytic number theory, which link many the subject's most important questions. We hope that a confirmation of the Twin Prime Conjecture will contribute towards understanding the underlying principles that guide the field.

These heuristics generally say that if a sequence has no obvious structure, then it should be modeled by a random variable. To be more concrete, let us consider modeling the distribution of the prime numbers. Let $\pi(x)$ be the prime number counting function, defined as

$$\pi(x) := \#\{p \leq x \colon p \text{ prime}\}.$$

The prime number theorem says that

$$\pi(x) \sim \mathrm{Li}(x) := \int_2^x \frac{dt}{\log t}.$$

We can model this distribution probabilistically by pretending that each integer n has a probability $\frac{1}{\log n}$ of being prime. More formally, we model the event $\{n \text{ is prime}\}$ by a Bernoulli random variable x_n which is 1 with probability $\frac{1}{\log n}$. It seems that this model is very close to the truth. For instance, the central limit theorem says that we should expect

$$\pi(x) = \mathrm{Li}(x) + O(x^{1/2+\epsilon}) \text{ for any } \epsilon > 0.$$

This turns out to be essentially equivalent to the Riemann hypothesis, a cornerstone conjecture in analytic number theory.

Let us try to see what else we can predict from this model. Suppose that we are interested in the primes lying in a certain arithmetic progression $\{nq + a\colon n \in \mathbb{N}\}$. If $(a,q) > 1$, then obviously there is at most one such prime, but otherwise we have no reasons to believe that the primes should be predisposed towards a particular residue class. Therefore, if we introduce the more general counting function

$$\pi(x;q,a) := \#\{p \le x\colon p \text{ prime }, p \equiv a \pmod{q}\}$$

then we expect that for $(a,q) = 1$,

$$\pi(x;q,a) = \frac{\mathrm{Li}(x)}{\varphi(q)} + O\left(x^{1/2+\epsilon}\right) \text{ for any } \epsilon > 0.$$

This turns out to be essentially equivalent to the generalized Riemann hypothesis.

The moral is that our simple probabilistic model can be used to guess deep facts. We can try to use it to make predictions about small gaps between primes. To do this, we shall have to make our model more precise. We have introduced random variables that model the primality of a given integer. Since we are interested in when several numbers are simultaneously prime, we need to consider the correlations between our random variables $\{x_n\}$.

Sometimes, it is obvious that the primality of two distinct numbers is correlated. For instance, n and $n+1$ cannot both be prime, since one of them must be divisible by 2. Therefore, in our probabilistic model x_n and x_{n+1} cannot be independent. More generally, if we consider a tuple $(h_1, \ldots, h_k)$ we see that $n + h_1, \ldots, n + h_k$ cannot all be prime infinitely often if there is some prime p that is guaranteed to divide at least one of them, i.e. if $h_1, \ldots, h_k$ occupy every residue class modulo p. If a tuple does not have this type of obvious obstruction, we say that it is *admissible*.

Let $\mathcal{H} = (h_1, \ldots, h_k)$ be an admissible tuple. Since we cannot see an obvious obstruction to $n + h_1, \ldots, n + h_k$ all being simultaneously prime, the most natural guess is to model $x_{n+h_1}, \ldots, x_{n+h_k}$ as independent random variables. This is called the *Cramér model*. Since the probability that any $x_{n+h_i} = 1$ is about $\frac{1}{\log n}$, we find that the probability that $n + h_1, \ldots, n + h_k$ are all primes is about $\frac{1}{\log^k n}$, so the Cramér model predicts that

$$\#\{n \le x\colon (n + h_1, \ldots, n + h_k) \text{ all prime }\} \sim \frac{x}{\log^k x}.$$

In particular, the Cramér model predicts that infinitely many such prime tuples exist.

Actually, this model is a little too crude. To see why, imagine the condition of n being prime as a gauntlet of conditions of the form $(n,p) = 1$ for all primes $p \neq n$. Fixing a prime p, we realize that the events that n and $n+h$ are both coprime to p are not quite independent. If they were independent, then the probability of both occurring would be $(1 - p^{-1})^2$, but in fact the residue class of n modulo p determines that of $n + h$, so that if h is not divisible by p then the probability is actually $1 - \frac{2}{p}$.

To capture this idea, let $\nu_{\mathcal{H}}(p)$ denote the number of distinct residue classes modulo p occupied by the member sof $\mathcal{H} = \{h_1, \dots, h_k\}$. (Observe that for all but finitely many primes p we have $\nu_{\mathcal{H}}(p) = k$.) Note that in these terms, $\mathcal{H}$ is admissible if and only if $\nu_{\mathcal{H}}(p) < p$ for all primes p. Then the probability that $n + h_1, \dots, n + h_k$ are all non-zero modulo p is $1 - \frac{\nu_{\mathcal{H}}(p)}{p}$. To obtain the correct asymptotic for prime tuples, we should multiply by a correction factor for each prime p that corrects the independence assumption, which is

$$\left(1 - \frac{\nu_{\mathcal{H}}(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-k}.$$

This leads to the Hardy-Littlewood conjectures. To state them more easily, we denote the total correction factor by

$$\mathfrak{S}(\mathcal{H}) = \prod_{p \text{ prime}} \left(1 - \frac{\nu_{\mathcal{H}}(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-k}.$$

The constant $\mathfrak{S}(\mathcal{H})$ converges because $\nu_{\mathcal{H}}(p) = k$ for all but finitely many p. It is traditionally called the *singular series* of $\mathcal{H}$, because it was originally discovered in series form by Hardy-Littlewood when they first arrived at the conjecture using the circle method (for more on this, see [8]).

Conjecture 1.2.1 (Hardy-Littlewood)**.** *Let $\mathcal{H}$ be an admissible tuple. Then*

$$\#\{n \le x \colon (n + h_1, \dots, n + h_k) \textit{ all prime } \} \sim \mathfrak{S}(\mathcal{H}) \frac{x}{\log^k x}.$$

This refines an earlier conjecture of Dickson that

$$\{n \colon (n + h_1, \dots, n + h_k) \text{ all prime } \} \text{ is infinite.}$$

1.3 OUTLINE OF THE ESSAY

We now outline the contents of this essay.

In §2, we give an overview of sieve methods leading to bounded gaps between primes. The entire chapter is included for motivation only, and no later work depends logically upon it (except that certain notation established in this chapter is used later). In particular, the chapter proves no technical results. We believe, however, that it will be very useful for understanding the work of Maynard and Tao.

We begin with an introduction to sieve theory through the sieve of Eratosthenes, and quickly proceed to describe Selberg's sieve, which is the fundamental sieve used in all work on bounded gaps between primes. We then sketch the ideas of Goldston-Pintz-Yildirim, which form the fundamental framework for the work of Zhang and Maynard and Tao. Maynard's proof is essentially a multi-dimensional generalization of what we describe in §2.3, so reading this section will be useful to understand the structure of his arguments, which can be obscured by all the technical details involved in carefully estimating error terms and so on. We then briefly indicate the nature of Zhang's work, so that the reader will see how it differs from that of Maynard and Tao. Finally, we sketch Maynard's argument, emphasizing its analogy to Goldston-Pintz-Yildirim's.

In §3, we embark on a detailed study of Maynard's proof. This is the technical heart of the essay. The most important technical aspect of the proof is in estimating two sums of sieve weights. It is worth noting that Maynard and Tao approach this very differently. We will mostly follow Maynard, but occasionally use Tao's arguments when we feel that they are much clearer. We also sketch Tao's methods at the end of the chapter.

In §4, we describe the progress achieved by the Polymath 8b project. Since that project is still ongoing, and its arguments are mostly scattered across several internet threads, we summarize only what we feel are the most important results. In particular, we avoid delving into the technical details of what is essentially a functional optimization problem, instead focusing on the ideas. We also describe heuristics that lead experts to believe that no further progress can be made without significant new ideas beyond sieve theory.

1.4 NOTATION

We record here some useful notation employed in the essay, which should be standard in analytic number theory.

- We say $f \sim g$ if $f, g\colon \mathbb{N} \to \mathbb{R}$ are two functions such that $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$.
- We say $f = O(g)$ if $f, g\colon \mathbb{N} \to \mathbb{R}$ are two functions and there exists some constant C such that $|f(n)| \leq C|g(n)|$ for all n.
- We say $f \ll g$ if $f = O(g)$.
- We say that $f = o(g)$ if $f, g\colon \mathbb{N} \to \mathbb{R}$ are two functions such that $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$.
- We say that $f \asymp g$ if $f = O(g)$ and $g = O(f)$.
- We use $f \approx g$ to informally denote that f is approximately equal to g, sometimes with the connotation that f is g plus some error terms.
- We use $\log^k x$ to denote $(\log x)^k$.
- We denote by (a, b) the greatest common divisor of a and b, and by $[a, b]$ the least common multiple of a and b. Unfortunately, the same notation is used to denote tuples or intervals; we hope that context will make clear which is intended.

2 OVERVIEW OF SIEVING FOR PRIMES

2.1 THE SIEVE OF ERATOSTHENES

Sieve theory originated as a method to count interesting sets in terms of simpler sets. It is easiest to illustrate through an example. Suppose we want to count the number of primes in some interval I, say $I = [N, 2N]$. Note that this is equivalent to estimating the prime number function $\pi(X)$, by partitioning into dyadic intervals.

The primes are rather difficult to count, but we can describe them in terms of sets that are simpler to count. For example, we know that there should be about $\frac{N}{2}$ even numbers in I, and $\frac{N}{3}$ numbers divisible by 3, and $\frac{N}{6}$ numbers divisible by 6. We can then use combinatorics to combine this information: by the inclusion-exclusion principle, about $N - \frac{N}{2} - \frac{N}{3} + \frac{N}{6} = \frac{N}{3}$ numbers in I are coprime to 2 and 3. This is already progress towards our problem of counting prime numbers! We now see that we can describe the primes in terms of the sets $\{n \in I \colon d \mid n\}$, which are easy to count, through some applications of the inclusion-exclusion principle. This idea was written down by Eratosthenes, as an algorithm for finding the prime numbers by sifting out integers divisible by d. Its modern sieve form is due to Legendre.

To formalize the sieve, we imagine that we know all primes up to some number $z < N$. Let

$$m = \prod_{p \text{ prime } < z} p.$$

Then the number of primes in I is upper bounded by the number of integers coprime to m.

$$\pi(2N) - \pi(N) \leq \#\{n \in I \colon (n, m) = 1\}.$$

Moreover, this inequality is actually an equality if $\sqrt{2N} < z < N$. We can attempt to count the right hand side sifting out multiples as described above. We introduce a useful notation for packaging the inclusion-exclusion arguments.

Definition 2.1.1. The Möbius μ-function is defined as

$$\mu(n) = \begin{cases} 0 & p^2 \mid n \text{ for some prime p}, \\ (-1)^k & n = p_1 \dots p_k \text{ for distinct primes } p_1, \dots, p_k. \end{cases}$$

The key property of μ is the identity

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

This is easy to see by applying the principle of inclusion-exclusion to the prime divisors of n, or otherwise. Therefore, we may write

$$\#\{n \in I \colon (n, m) = 1\} = \sum_{N \leq n \leq 2N} \sum_{d \mid (n,m)} \mu(d).$$

Swapping the order of summation, we rewrite this sum as

$$\#\{n \in I\colon (n,m) = 1\} = \sum_{d|m} \mu(d) \sum_{\substack{N \le n \le 2N \\ d|n}} 1$$
$$= \sum_{d|m} \mu(d) \left\lfloor \frac{N}{d} \right\rfloor$$

The quantity $\lfloor \frac{X}{d} \rfloor$ is difficult to work with, but it may be estimated by the smooth function $\frac{X}{d}$ with an error of at most 1, so the sum above is

$$\#\{n \in I\colon (n,m) = 1\} = N \sum_{d|m} \frac{\mu(d)}{d} + O\left(\sum_{d|m} \mu(d) \right).$$

Now, it turns out that

$$\sum_{d|m} \frac{\mu(d)}{d} = \prod_{p<z} \left(1 - \frac{1}{p}\right) \asymp \frac{1}{\log z}$$

(see Mertens' Theorem in [13], for instance). Substituting this above, our estimate is

$$\#\{n \in I\colon (n,m) = 1\} \ll \frac{N}{\log z} + O\left(\sum_{d|m} \mu(d) \right).$$

It is clear that larger we can take z, the better estimate we get from the main term. However, we must take care that error term does not swamp the main term. Intuitively, we think that the error term should be small because there should be a lot of cancellation in the sum. But it is hard to capture this cancellation rigorously, and the best we can do without enormous effort is to use the trivial bound $|\mu(d)| \le 1$, ignoring all cancellation, which bounds the error by

$$O\left(\sum_{d|m} \mu(d) \right) = O(2^{\pi(z)})$$

since m is the product of the distinct primes $\le z$. So we are forced to take z on the order of magnitude of $\log N$, which gives the very weak estimate

$$\pi(2N) - \pi(N) \ll \frac{N}{\log \log N}.$$

If we could take $z = N^{\epsilon}$ while controlling the error term in a satisfactory way, then the main term would instead be

$$\pi(2N) - \pi(N) \ll \frac{N}{\log N},$$

which is the weak Prime Number Theorem.

2.2 SELBERG'S SIEVE

From a modern perspective, the approach of counting a set with other physical sets is too crude. It is better to use sums of weighted functions. When the weights are integer linear combinations of indicator functions, then we recover the previous combinatorial methods, but in general this approach is more expressive.

If we are interested in counting a set, then the goal of sieve theory is to approximate the indicator function of that set with weights that are easier to sum. This usually means that we should use weights described by relatively smooth functions. There is an inherent tension in trying to use "smooth" weights to describe a "rough" set like the prime numbers. The challenge is to balance the tradeoff by picking weights that are smooth enough to analyze, but rough enough so that the relevant arithmetic data can be extracted from them.

From this perspective, the problem with the sieve of Eratosthenes is that the weights $\mu(n)$ are too hard to control. They fluctuate haphazardly among the values $\{-1, 0, 1\}$. One philosophical reason for this problem is that we used them to give an *exact* expression for the quantity we were interested in (essentially the prime counting function), which is itself somewhat unpredictable. In order to describe this rough function in terms of a smooth one (as in the Prime Number Theorem), we need to replace our exact weights with ones that behave more tamely.

Suppose, then, that we replace the weights $\mu(n)$ used in the sieve of Eratosthenes with some other set of weights $w(n)$. What conditions do we need $w(n)$ to satisfy? The key property of $\mu(n)$ that we used in order to sift out coprimes numbers was the identity

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

For the purposes of obtaining an *upper* bound on $\#\{n \in I \colon (n, m) = 1\}$, we can replace

$$\sum_{N \leq n \leq 2N} \sum_{d|(n,m)} \mu(d) \qquad \text{with} \qquad \sum_{N \leq n \leq 2N} w((n, m))$$

where the $w(n)$ are any weights satisfying the condition

$$w(n) \geq 1 \text{ if } n = 1 \quad \text{ and } \quad w(n) \geq 0 \text{ for all } n.$$

Selberg proposed the following choice of such weights. To guarantee positivity, $w(n)$ will be a square of the form

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

To guarantee that $w(1) \geq 1$, we simply take $\lambda_1 = 1$. This choice certainly satisfies the conditions that we want.

Choosing weights of this form allows great flexibility in that we can set λ_d arbitrarily (for $d > 1$), and this turns out to be a powerful asset. In these terms, our estimate is

$$\begin{aligned}
\#\{n \in I \colon (n,m) = 1\} &\leq \sum_{N \leq n \leq 2N} \left(\sum_{d|(n,m)} \lambda_d \right)^2 \\
&= \sum_{\substack{d|m \\ e|m}} \lambda_d \lambda_e \sum_{\substack{N \leq n \leq 2N \\ [d,e]|n}} 1 \\
&= \sum_{\substack{d|m \\ e|m}} \lambda_d \lambda_e \left\lfloor \frac{N}{[d,e]} \right\rfloor \\
&= N \sum_{\substack{d|m \\ e|m}} \frac{\lambda_d \lambda_e}{[d,e]} + O\left(\sum_{\substack{d|m \\ e|m}} \lambda_d \lambda_e \right).
\end{aligned}$$

In order to better control the error term, Selberg imposed the constraint $\lambda_n = 0$ for $n \geq R$, for some constant R called the "level of support." This trick is used in essentially all modern sieves. In applications, choosing this R is a delicate balance: choosing a larger R allows us more flexbility in the choice of weights, and hence a better main term, but the cost is that the error term is harder to control.

We now wish to optimize the λ_n subject to our constraints

$$\boxed{\begin{aligned} &\lambda_1 = 1 \\ &\lambda_n = 0 \text{ for } n \geq R. \end{aligned}}$$

Selberg's next insight was to recognize that the coefficient $\sum_{\substack{d|m \\ e|m}} \frac{\lambda_d \lambda_e}{[d,e]}$ of the principal term is a quadratic form that can be explicitly diagonalized. The key is to use arithmetic function identities to remove the coupling between d, e. Recalling the identity $(d,e)[d,e] = de$, we can write

$$\begin{aligned}
Q(\lambda) &= \sum_{\substack{d|m \\ e|m}} \frac{\lambda_d \lambda_e}{[d,e]} \\
&= \sum_{\substack{d|m \\ e|m}} \frac{\lambda_d}{d} \frac{\lambda_e}{e} (d,e).
\end{aligned}$$

Next, we use the identity $m = \sum_{a|m} \varphi(a)$ to remove the coupling involved in taking the greatest common divisor:

$$\begin{aligned}
Q(\lambda) &= \sum_{\substack{d|m \\ e|m}} \frac{\lambda_d}{d} \frac{\lambda_e}{e} \sum_{a|(d,e)} \varphi(a) \\
&= \sum_{a|m} \varphi(a) \sum_{\substack{a|d|m \\ a|e|m}} \frac{\lambda_d}{d} \frac{\lambda_e}{e} \\
&= \sum_{a|m} \varphi(a) \left(\sum_{a|d|m} \frac{\lambda_d}{d} \right)^2.
\end{aligned}$$

At this point, it is natural to introduce the change of variables

$$y_a := \sum_{a|d|m} \frac{\lambda_d}{d}$$

so that Q is explicitly diagonalized as

$$Q(y) = \sum_{a|m} \varphi(a) y_a^2.$$

It is not obvious that this change of variables is actually invertible, but it follows from a dual version of the Möbius inversion theorem. This identity also plays a crucial role in the later work of Goldston-Pintz-Yildirim and Maynard, so we note it down now.

Lemma 2.2.1 (Dual Möbius inversion). *Let $f : \mathbb{N} \to \mathbb{R}$ be a function with finite support. If*

$$F(a) = \sum_{a|d} f(d)$$

then

$$f(d) = \sum_{d|a} \mu\left(\frac{a}{d}\right) F(a).$$

Proof. This is a straightforward application of swapping the order of summation and applying the divisor sum of the Möbius function:

$$\begin{aligned}\sum_{d|a} \mu\left(\frac{a}{d}\right) F(a) &= \sum_{d|a} \mu\left(\frac{a}{d}\right) \sum_{a|e} f(e) \\ &= \sum_{e} f(e) \sum_{d|a|e} \mu\left(\frac{a}{d}\right) \\ &= \sum_{e} f(e) \sum_{a'|\frac{e}{d}} \mu(a') \\ &= f(d).\end{aligned}$$

□

Applying this in our case, we find that

$$\frac{\lambda_d}{d} = \sum_{d|a|m} \mu\left(\frac{a}{d}\right) y_a.$$

In particular, our linear condition $\lambda_1 = 1$ translates into one linear condition

$$\sum_{a|m} \mu(a) y_a = 1$$

and the level of support for the y_a is the same as for the λ_n, so

$$y_a = 0 \text{ for } a \geq R.$$

We may now choose the finitely many nonzero y_a from the hyperplane $\sum \mu(a) y_a = 1$ to minimize $Q(y)$, and it is a straightforward exercise in employing Lagrange multipliers (or in this case, completing the square) to see that the result is

$$y_a = \left(\sum_{\substack{k \leq R \\ k|m}} \frac{\mu(k)^2}{\varphi(k)} \right)^{-1} \frac{\mu(a)}{\varphi(a)}$$

Substituting this back in, we find that

$$\lambda_d = d \left(\sum_{\substack{k \le R \\ k|m}} \frac{\mu(k)^2}{\varphi(k)} \right)^{-1} \sum_{\substack{d|a|m \\ a \le R}} \mu\left(\frac{a}{d}\right) \frac{\mu(a)}{\varphi(a)}$$

By construction m is squarefree, so we can factor out $\varphi(a) = \varphi(d)\varphi\left(\frac{a}{d}\right)$ and $\mu(a) = \mu(d)\mu\left(\frac{a}{d}\right)$ and write

$$\lambda_d = \mu(d) \frac{d}{\varphi(d)} \left(\sum_{\substack{k \le R \\ k|m}} \frac{\mu(k)^2}{\varphi(k)} \right)^{-1} \sum_{\substack{a \le R/d \\ a|m/d}} \frac{\mu(a)^2}{\varphi(a)}.$$

We would like to find a convenient smooth approximation to λ_d, which will be easier to handle. It turns out that

$$\sum_{k \le R} \frac{\mu(k)^2}{\varphi(k)} \approx= \log R + O(1).$$

This estimate is a consequence of later results we prove on sums of arithmetic functions; this special case appears in [13]. If we estimate that the divisibility constraints $k \mid m$ and $a \mid m$ cut down the two sums in our expression for λ_d by approximately the same proportions, and that $\varphi(d) \approx d$ (which is the case for typical numbers without very many distinct prime factors), then we arrive at the choice of smooth weights

$$\lambda_d \approx \mu(d) \frac{\log(R/d)}{\log R}.$$

This inspires the expressions that enter into the more sophisticated weights we will consider later.

It will be useful for motivation to consider how one might apply the Selberg sieve to bound the number of twin primes. Proceeding as above, set m to be the product of primes up to z. An upper bound for the cardinality of the set $\#\{n \in [N, 2N] \colon n \text{ and } n+2 \text{ coprime to } m\}$ is

$$\sum_{n \le x} \left(\sum_{d|(n(n+2),m)} \lambda_d \right)^2$$

where $\lambda_1 = 1$, since this picks out n such that $n(n+2)$ is coprime with m. By a diagonalizing the quadratic form and making a judicious choice of weights, one can prove that the number of twin primes is no larger than the order of magnitude predicted by the Hardy-Littlewood Conjecture.

Theorem 2.2.2. *The number of pairs $(n, n+2)$ such that n and $n+2$ are both prime and $n \le x$ is $O\left(\frac{x}{(\log x)^2}\right)$.*

Proof. See any standard text, e.g. [13] or [2]. □

2.3 THE GOLDSTON-PINTZ-YILDIRIM SIEVE

We now describe the seminal work of Goldston, Pintz, and Yildirim in [7], following the expositions in [8] and [17]. We go into some detail, since the steps here will all be repeated in a more general form, but with the same basic ideas, when we describe Maynard's work in §3. To emphasize the structure of the argument over the technical details, we will ignore all error terms and focus on the shape of the calculations.

2.3.1 *The basic framework*

Thus far, we have discussed methods of bounding *above* the number of primes, or prime tuples. It is a general feature of sieve theory, and what makes proving bounded gaps between primes so difficult, that non-trivial lower bounds are much harder to obtain than upper bounds. This is a reflection of the "parity problem," a heuristic that we will discuss later in §4.5. For this reason, mathematicians were pessimistic about establishing bounded gaps between primes until a breakthrough in 2005 by Goldston, Pintz, and Yildirim. As is common in the literature, we will sometimes abbreviate them as GPY.

Let $\chi_{\mathcal{P}}$ denote the indicator function of the primes:

$$\chi_{\mathcal{P}}(n) = \begin{cases} 1 & n \text{ is prime,} \\ 0 & \text{otherwise.} \end{cases}$$

The basic idea is to find non-negative weights $w(n)$ such that

$$\sum_{n\in[N,2N]} \chi_{\mathcal{P}}(n+h)w(n) \text{ is "large" relative to } \sum_{n\in[N,2N]} w(n).$$

If we can do this, then $n+h$ must be prime for some values of n in this range. In particular, suppose that both of the following inequalities hold:

$$\sum_{n\in[N,2N]} \chi_{\mathcal{P}}(n)w(n) > \frac{1}{2} \sum_{n\in[N,2N]} w(n)$$
$$\sum_{n\in[N,2N]} \chi_{\mathcal{P}}(n+h)w(n) > \frac{1}{2} \sum_{n\in[N,2N]} w(n).$$

Then, summing, we may conclude that

$$\sum_{n\in[x,x+y]} (\chi_{\mathcal{P}}(n) + \chi_{\mathcal{P}}(n+h) - 1)w(n) > 0.$$

Since the $w(n)$ are non-negative, we deduce that at least one n in the interval $[N, 2N]$ has the property that n *and* $n+h$ are prime. In principle, this could allow us to establish the infinitude of prime pairs.

Unfortunately, it turns out that finding weights satisfying the inequalities above is too ambitious a task. However, we can gain a little room by observing that it suffices to prove that for a k-tuple $\mathcal{H} = (h_1, h_2, \ldots, h_k)$, we have

$$\sum_{n\in[N,2N]} \chi_{\mathcal{P}}(n+h_i)w(n) > \frac{1}{k} \sum_{n\in[N,2N]} w(n).$$

For, summing as before, this implies that some two of $\{n+h_1,\dots,n+h_k\}$ are prime for some n in $[N,2N]$.

Therefore, we set

$$S_1 = \sum_{n\in[N,2N]} w(n) \tag{1}$$

and

$$S_2^{(\ell)} = \sum_{n\in[N,2N]} \chi_{\mathcal{P}}(n+h_\ell)w(n) \tag{2}$$

Remark 2.3.1. This notation is slightly confusing because the sums depend on N, but it is well established in the literature. We are interested in establishing asymptotics (in N) for $S_1, S_2^{(\ell)}$.

In fact, our setup is naturally geared towards proving a family of assertions stronger than bounded gaps between primes.

Definition 2.3.2. Let $DHL[k,m]$ denote the assertion that if $(h_1,\dots,h_k)$ is any admissible k-tuple, then for infinitely many integers n there are at least m primes among the set $\{n+h_1,\dots,n+h_k\}$.

The DHL stands for Dickson-Hardy-Littlewood, whose conjecture implies that DHL is true for all k and $m\le k$.

Lemma 2.3.3. *If for all sufficiently large N we have*

$$S_2^{(\ell)} > \frac{\rho}{k}S_1$$

then $DHL[k,\lfloor\rho+1\rfloor]$ is true.

Proof. From the hypothesis, we see that

$$0<\sum_{\ell=1}^{k} S_2^{(\ell)} - \rho S_1 = \sum_{n\in[2N,N]} w(n)\left(\sum_{\ell=1}^{k}\chi_{\mathcal{P}}(n+h_\ell)-\rho\right).$$

Since the $w(n)$ are all non-negative, it must be the case that some summand is positive, which implies that for some $n\in[N,2N]$ at least $\lfloor\rho+1\rfloor$ of the integers $\{n+h_1,\dots,n+h_k\}$ must be prime. □

Corollary 2.3.4. *If $(h_1,\dots,h_k)$ is any admissible k-tuple, then $DHL[k,m]$ implies*

$$\liminf_{n\to\infty} p_{n+m}-p_n \le h_k-h_1.$$

2.3.2 *Choice of weights*

So far, the goal we have set seems simple enough. The subtlety comes in choosing appropriate weights. Since the primes in $[N,2N]$ have density about $\frac{1}{\log N}$, one would expect that

$$\sum_{n\in[N,2N]}\chi_{\mathcal{P}}(n+h_\ell)w(n)\approx\frac{1}{\log N}\sum_{n\in[N,2N]}w(n)$$

for typical choice of weights $w(n)$. We need to improve the $\frac{1}{\log N}$ factor all the way to $\frac{1}{k}$, so we must choose the $w(n)$ to be fairly good approximations to the indicator function of primes.

Goldston, Pintz, and Yildirim follow Selberg's general form of weights:

$$w(n) = \left(\sum_{d|(n+h_1)\dots(n+h_k)} \lambda_d \right)^2$$

where λ_d has level of support R. For convenience, we write $P(n) = (n + h_1)\dots(n+h_k)$. This choice of weights then turns each of the sums $S_1, S_2^{(\ell)}$ into a quadratic form in the λ_d, and the goal is to maximize their ratio. In order to estimate the sums $S_1, S_2^{(\ell)}$, we diagonalize the quadratic forms as before. There are several reasons to do this. At a näive level, the diagonalized sum is simpler easier to evaluate, since so many cross-terms vanish. The more sophisticated reason is that we will eventually be choosing weights λ_n that oscillate in sign, so there will be a lot of cancellation in the sums. When we diagonalize the form, we obtain a sum of positive terms so that the cancellation is automatically accounted for.

$$\begin{aligned} S_1 &= \sum_{n\in[N,2N]} w(n) \\ &= \sum_{n\in[N,2N]} \left(\sum_{d|P(n)} \lambda_d \right)^2. \\ &= \sum_{n\in[N,2N]} \sum_{d|P(n)} \lambda_d \sum_{e|P(n)} \lambda_e \\ &= \sum_{\substack{d\le R\\ e\le R}} \lambda_d\lambda_e \sum_{\substack{n\in[N,2N]\\ d|P(n)\\ e|P(n)}} 1. \end{aligned}$$

Now, $P(n) \equiv 0 \pmod d$ and $P(n) \equiv 0 \pmod e$ if and only if n is equivalent to a certain set of congruence classes modulo $[d,e]$. By the Chinese remainder theorem, the number of residue classes of n mod m such that $P(n) \equiv 0 \pmod m$ is a multiplicative function $f(m)$.

As a sanity check, let's get some intuition for what $f(p)$ looks like. It is the number of roots of the degree k polynomial $P(n)$ in $\mathbb{Z}/p$, i.e. the number of residue classes represented by the members of $\mathcal{H}$. This is just what we called $\nu_{\mathcal{H}}(p)$ earlier. For all but finitely many primes $f(p) = k$, since we have $f(p) < k$ only if two of the h_i are equal modulo p. The condition that $(h_1, \dots, h_k)$ is an admissible tuple is precisely that $f(p) < p$ for all primes p. We will eventually choose our λ_d to be supported on squarefree integers, so in fact the preceding discussion determines f in all cases of interest by multiplicativity.

With this notation, we may write the sum above as

$$\begin{aligned} S_1 &= \sum_{\substack{d\le R\\ e\le R}} \lambda_d\lambda_e \frac{f([d,e])}{[d,e]} y + O\left(\sum_{\substack{d\le R\\ e\le R}} \lambda_d\lambda_e \right) \\ &= N \sum_{\substack{d\le R\\ e\le R}} \frac{\lambda_d f(d)}{d} \frac{\lambda_e f(e)}{e} \frac{(d,e)}{f((d,e))} + O\left(\sum_{\substack{d\le R\\ e\le R}} \lambda_d\lambda_e \right). \end{aligned}$$

The error term can be shown to be negligible with careful bookkeeping. To diagonalize the main term as we did before, we use Möbius inversion to write

$$\frac{(d,e)}{f((d,e))} = \sum_{a|(d,e)} g(a)$$

for some multiplicative function g. What does this g look like? Since it is multiplicative, it suffices to understand $g(p^k)$ for p prime. We can focus on squarefree arguments, since that is what we will ultimately be using. By the Möbius inversion formula, we see that

$$g(p) = -1 + \frac{p}{f(p)} = \frac{p - f(p)}{f(p)}.$$

As $f(p)$ is typically k, $g(p)$ typically $\frac{p-k}{k} \approx \frac{p}{k}$.

We may now rewrite

$$\begin{aligned} S_1 &\sim N \sum_{\substack{d \le R \\ e \le R}} \frac{\lambda_d f(d)}{d} \frac{\lambda_e f(e)}{e} \sum_{\substack{a|d \\ a|e}} g(a) \\ &= N \sum_{a \le R} g(a) \left(\sum_{a|d} \frac{\lambda_d f(d)}{d} \right)^2 \\ &= N \sum_{a \le R} \frac{1}{g(a)} \left(g(a) \sum_{a|d} \frac{\lambda_d f(d)}{d} \right)^2 . \end{aligned}$$

Then introducing the change of variables

$$y_a := g(a) \sum_{a|d} \frac{\lambda_d f(d)}{d},$$

we can rewrite

$$S_1 \sim N \sum \frac{1}{g(a)} y_a^2.$$

The reason for rewriting the sum with coefficients $\frac{1}{g(a)}$ instead of $g(a)$ is that we want the coefficients of the quadratic form $\sum \frac{1}{g(a)} y_a^2$ to vary slowly, which will make the sum easier to estimate. We have already observed that $g(p) \sim \frac{p}{k}$, so $\frac{1}{g(p)}$ decreases slowly.

Now we saw before the choice of the λ_d is equivalent to the choice of y_a, with the same level of support. We want to choose this to maximize the ratio $S_2^{(\ell)}/S_1$. When we studied Selberg's sieve, we saw that we could optimize the y_a after diagonalizing the quadratic form. In the present case, we want to maximize a ratio of two quadratic forms that are not simultaneously diagonalizable, which is more challenging. However, motivated by the form of the Selberg sieve weights (which were selected to minimize S_1), we choose

$$y_a = \mu(a) F\left(\frac{\log a}{\log R}\right)$$

where F is some smooth function to be chosen later. Then

$$S_1 \sim N \sum_{a \le R} \frac{\mu(a)^2}{g(a)} F\left(\frac{\log a}{\log R}\right)^2 .$$

Note that the factor $\mu(a)$ essentially serves to restrict the support of y_a to the squarefree integers.

2.3.3 *Sums of multiplicative functions*

The next step is to exploit our choice of smooth weights to approximate S_1 with a more analytically tractable expression. We do this by partial summation, and for that we need to understand the size of the sums

$$\sum_{a\le x}\frac{1}{g(a)}.$$

We have already remarked that $\frac{1}{g(p)} \approx \frac{k}{p}$, so we expect this sum to be on the order of magnitude of

$$\left(\sum_{n\le x}\frac{1}{n}\right)^k \asymp (\log x)^k.$$

To be more precise, if we define the singular series $\mathfrak{S}(g)$ associated to g to be

$$\mathfrak{S}(g) = \prod_p \left(1+\frac{1}{g(p)}+\frac{1}{g(p^2)}+\dots\right)\left(1-\frac{1}{p}\right)^k$$

then it turns out that

$$\sum_{a\le x}\frac{1}{g(a)} \sim \frac{\mathfrak{S}(g)}{k!}\log^k x.$$

From this, it is an exercise in partial summation to obtain

$$\sum_{a\le R}\frac{1}{g(a)}F\left(\frac{\log a}{\log R}\right)^2 \sim \mathfrak{S}(g)\log^k R\int_0^1 \frac{u^{k-1}}{(k-1)!}F(u)^2\,du.$$

We shall prove these estimates rigorously when we tackle Maynard's theorem later.

2.3.4 *Primes in arithmetic progression*

We follow the same approach to analyze $S_2^{(\ell)}$.

$$\begin{aligned}
S_2^{(\ell)} &= \sum_{n\in[N,2N]} w(n)\chi_{\mathcal{P}}(n+h_\ell)\\
&= \sum_{n\in[N,2N]} \chi_{\mathcal{P}}(n+h_\ell)\left(\sum_{d|P(n)}\lambda_d\right)^2.\\
&= \sum_{n\in[N,2N]} \chi_{\mathcal{P}}(n+h_\ell)\sum_{d|P(n)}\lambda_d\sum_{e|P(n)}\lambda_e\\
&= \sum_{\substack{d\le R\\ e\le R}}\lambda_d\lambda_e\sum_{\substack{n\in[N,2N]\\ d|P(n)\\ e|P(n)}}\chi_{P}(n+h_\ell)
\end{aligned}$$

Now the inner sum is more challenging to evaluate. As before, the first two conditions $d \mid P(n)$ and $e \mid P(n)$ are equivalent to n being in a certain set of equivalence classes modulo $[d,e]$. Therefore,

$$\sum_{\substack{n\in[N,2N]\\ d|P(n)\\ e|P(n)}}\chi_{P}(n+h_\ell)$$

counts the number of primes landing in a certain number of residue classes modulo $[d, e]$ in an interval of length N.

The prime number theorem tells us roughly how many primes will lie in a given interval, and we expect that all residue classes modulo q will see roughly the same number of primes, with the exception of those that have obvious obstructions to being prime (specifically, those $a \pmod q$ where $\gcd(a, q) > 1$). In terms of obtaining a good estimate, the question is how small the error term is.

The Prime Number Theorem says that $\pi(x) = \mathrm{Li}(x) + E(x)$, and the best known bound on the error term is

$$|E(x)| \ll \frac{x}{(\log x)^A}$$

for any constant A (where the implicit constant depends on A). The argument for the Prime Number Theorem generalizes readily to primes in arithmetic progressions to show that if $\pi(x; q, a)$ denotes the prime counting function for the primes in the *primitive* residue class $a \pmod q$, then

$$\pi(x; q, a) = \frac{1}{\phi(q)} \mathrm{Li}(x) + E(x; q, a)$$

where

$$|E(x; q, a)| \ll \frac{x}{(\log x)^A}$$

for any $A > 0$, and the implicit constant depends on q and A.

This result is inadequate in our present situation for several reasons. First, we are actually considering varying the moduli q, which may be quite large relative to x, and the implicit constant depends on q. Second, the error bound is just far too weak.

What is the most optimistic estimate that we can hope for? The Generalized Riemann Hypothesis implies the much stronger bound

$$|E(x; q, a)| \ll x^{\frac{1}{2}} \log x$$

where the implicit constant does not depend on q. This is far beyond our present means to prove, but there is a very useful result of Bombieri and Vinogradov that asserts that this is true *on average*. To state it, we define the maximum error for the modulus q:

$$E(x; q) = \max_{(a,q)=1} |E(x; q, a)|.$$

Theorem 2.3.5 (Bombieri-Vinogradov). *For any positive constant A, there exists a constant B such that*

$$\sum_{q \le Q} \max_{y \le x} |E(y; q)| \ll \frac{x}{(\log x)^A} \text{ with } Q = \frac{x^{\frac{1}{2}}}{(\log x)^B}.$$

We omit the proof in this essay, but it is quite classical and relies on only elementary techniques: see [3] for instance. How good is this? It says that the average error for moduli $q \le x^{\frac{1}{2}} / (\log x)^B$ is

$$\ll \frac{x(\log x)^{-A}}{Q} = x^{\frac{1}{2}} (\log x)^{B-A},$$

which is just as good as what is implied by the Generalized Riemann Hypothesis. Therefore, we can think of the Bombieri-Vinogradov theorem as saying that "the Generalized Riemann Hypothesis is true on average."

This is precisely what we need for our present application, since what shows up in S_2 are terms of the form $\pi(2N; [d,e], a) - \pi(N; [d,e], q)$. If the λ_d have level of support R, then $[d,e]$ has size up to R^2. Therefore, to control the relevant error terms with Bombieri-Vinogradov, we can choose R to be approximately $x^{\frac{1}{4}}$. We remarked earlier that choosing a larger level of support allows more flexibility in the sieve, so we would ideally like to take R to be even larger. We now introduce a family of hypotheses that prescribe how large we can take the level of support.

Definition 2.3.6. We say that the primes have *level of distribution* θ if for any constant A, we have

$$\sum_{q \leq x^\theta} \max_{y \leq x} |E(y;q)| \ll \frac{x}{(\log x)^A}$$

where the implicit constant depends on A.

Following Polymath 8, we introduce a family of hypotheses on the distribution of the primes.

Definition 2.3.7. We denote by $EH[\theta]$ the assertion that the primes have level of distribution θ.

Then the Bombieri-Vinogradov Theorem asserts that $EH[\theta]$ is true for any $\theta < \frac{1}{2}$. The Elliott-Halberstam conjecture asserts that we can take the parameter all the way up to any $\theta < 1$.

Conjecture 2.3.8 (Elliot-Halberstam). *$EH[\theta]$ is true for any $\theta < 1$.*

We abbreviate the Elliott-Halberstam conjecture as EH. The important takeaway is that when $EH[\theta]$ is true, then we may take level of support to be $R = x^{\frac{\theta}{2}}$.

Now let us continue estimating $S_2^{(\ell)}$.

$$\begin{aligned} S_2^{(\ell)} &= \sum_{\substack{d \leq R \\ e \leq R}} \lambda_d \lambda_e \sum_{\substack{n \in [N,2N] \\ d | P(n) \\ e | P(n)}} \chi_p(n + h_\ell) \\ &\sim (\mathrm{Li}(2N) - \mathrm{Li}(N)) \sum_{\substack{d \leq R \\ e \leq R}} \lambda_d \lambda_e \frac{f^{(\ell)}([d,e])}{\varphi([d,e])} \end{aligned}$$

where $f^{(\ell)}([d,e])$ is the number of permissible residue classes for n. To elaborate, we saw that the condition $P(n) \equiv 0 \pmod{m}$ boils down to forcing n into certain residue classes modulo m, and we called $f(m)$ the number of such residue classes. However, some of these are forbidden by the condition $\gcd(n + h_\ell, [d,e]) = 1$, which arises in applying our prime number estimate because we require $n + h_\ell$ to be prime. For a particular prime p, $f^{(\ell)}(p) = f(p) - 1$. We saw that $f(p) = k$ for all but finitely many primes, so $f^{(\ell)}(p) = k - 1$ for all but finitely many primes.

Now the analysis proceeds exactly as before. We introduce an arithmetic function $g^*(a)$, satisfying

$$f^{(\ell)}(d) = \sum_{a|d} g^{(\ell)}(a).$$

The sum can then be rearranged as

$$\begin{aligned}\sum_{\substack{d\leq R\\e\leq R}} \lambda_d\lambda_e \frac{f^{(\ell)}([d,e])}{\varphi([d,e])} &= \sum_{\substack{d\leq R\\e\leq R}} \frac{\lambda_d f^{(\ell)}(d)}{\varphi(d)} \frac{\lambda_e f(e)}{\varphi(e)} \frac{(d,e)}{f^{(\ell)}(d,e)} \\ &= \sum_{a\leq R} g^{(\ell)}(a) \left(\sum_{a|d} \frac{\lambda_d f^{(\ell)}(d)}{\varphi(d)} \right)^2 \\ &= \sum_{a\leq R} \frac{1}{g^{(\ell)}(a)} \left(g^{(\ell)}(a) \sum_{a|d} \frac{\lambda_d f^{(\ell)}(d)}{\varphi(d)} \right)^2 .\end{aligned}$$

Setting

$$y_a^{(\ell)} = g^{(\ell)}(a) \sum_{a|d} \frac{\lambda_d f^{(\ell)}(d)}{\varphi(d)},$$

we obtain

$$S_2^{(\ell)} \sim (\mathrm{Li}(2N) - \mathrm{Li}(N)) \sum_{a\leq R} \frac{1}{g^{(\ell)}(a)} (y_a^{(\ell)})^2.$$

The $y_a^{(\ell)}$ are determined by the λ_d, which are in turn determined by our choice of y_a. Therefore, we should solve for the $y_a^{(\ell)}$ in terms of y_a in order to see the explicit dependence of S_2 on the sieve weights y_a. We will do an analogous calculation later, so for now we just state the result:

$$y_a^{(\ell)} = \frac{a}{\phi(a)} \sum_{m\geq 1} \frac{\mu(ma) y_{ma}}{\phi(m)}.$$

Note that the summands are supported on m such that $(m, a) = 1$ (since y_{ma} is supported on squarefree integers) and $ma \leq R$.

We know how to approximate sums of this form with our results on sums of multiplicative functions, which give

$$y_a^{(\ell)} \sim \log R \int_{\frac{\log a}{\log R}}^1 F(t)\, dt.$$

If we substitute this above in our expression for $S_2^{(\ell)}$, and apply our results on sums of multiplicative functions again, we find that

$$S_2^{(\ell)} \sim (\mathrm{Li}(2N) - \mathrm{Li}(N)) \mathfrak{S}(g) (\log R)^{k+1} \int_0^1 \left(\int_t^1 F(u)\, du \right)^2 \frac{t^{k-2}}{(k-2)!}\, dt.$$

Finally, $(\mathrm{Li}(2N) - \mathrm{Li}(N)) \sim \frac{N}{\log N}$, so we arrive at the estimate

$$S_2^{(\ell)} \sim \frac{N}{\log N} \mathfrak{S}(g) (\log R)^{k+1} \int_0^1 \left(\int_t^1 F(u)\, du \right)^2 \frac{t^{k-2}}{(k-2)!}\, dt.$$

Compare this with

$$S_1 \sim N \mathfrak{S}(g) (\log R)^k \int_0^1 F(t)^2 \frac{t^{k-1}}{(k-1)!}\, dt.$$

2.3.5 *Tying the knots*

Consider the ratio of the approximations to $S_2^{(\ell)}$ and S_1 that we have just computed:

$$\rho(F) = \theta \int_0^1 \left(\int_t^1 F(u)\, du \right)^2 \frac{t^{k-2}}{(k-2)!}\, dt \Big/ \int_0^1 \left(\int_t^1 F(u)\, du \right)^2 \frac{t^{k-2}}{(k-2)!}\, dt \sim \frac{S_2^{(\ell)}}{S_1}.$$

If we can show that $\rho(F) > \frac{1}{k}$ for some suitable choice of F, then we will have

$$\sum_{\ell=1}^{k} S_2^{(\ell)} > S_1.$$

By the arguments in §2.3.1, this implies that there are infinitely many n such that at least two of $\{n+h_1, \dots, n+h_k\}$ are prime; in particular, we get infinitely many bounded gaps between primes. The problem is now essentially an optimization problem, and it turns out that an essentially optimal choice is $F(t) = \frac{(1-t)^j}{j!}$ for some j. With this choice, we use standard beta integral identities to compute the the integrals. The result is

$$S_1 \sim \frac{1}{(k+2j)!}\binom{2j}{j} \qquad \text{and} \qquad S_2^{(\ell)} \sim \frac{1}{(k+2j+1)!}\binom{2j+2}{\ell+1}.$$

We will skip the explicit computations, but the punchline is that if you substitute these formulas in to our expression for $\rho(F)$, then $\rho(F)$ falls *just* short of $\frac{1}{k}$. We have just barely missed proving bounded gaps between primes! In fact, one can show that this is not merely a defect of our choice of F: no function can break the $\frac{1}{k}$ barrier.

However, if one takes $\theta = \frac{1}{2} + \eta$ for any $\eta > 0$, then one *can* achieve $\rho(F) > \frac{1}{k}$ for a sufficiently large j, and $k = (2j+1)^2$. Therefore, any level of distribution $\theta > \frac{1}{2}$ is sufficient to establish bounded gaps between primes.

In particular, assuming the Elliot-Halberstam conjecture, one can show that for a judicious choice of function F, we have $\rho_6(F) > \frac{1}{6}$, which implies $DHL[6,2]$. Checking that the 6-tuple $\{7, 11, 13, 17, 19, 23\}$ is admissible, one obtains Theorem 1.1.2.

2.4 THE WORK OF YITANG ZHANG

We now briefly describe Zhang's breakthrough work, which first established bounded gaps between primes. Zhang completed a program initiated by Motohashi and Pintz suggesting that the the GPY sieve could be slightly modified to depend on a weaker condition on the level of distribution of the primes.

We saw in the GPY argument that *any* level of distribution θ, with $\theta > \frac{1}{2}$, implies bounded gaps between primes. If we examine the argument carefully, we see that we do not quite need the full strength of the $EH[\theta]$ assumption. Zhang's breakthrough comes by proving a relaxed version of $EH[\theta]$, which is still just strong enough for the GPY argument to work, for some $\theta > \frac{1}{2}$.

The key idea is to restrict the sieve to "smooth numbers," or numbers without very large prime factors. Zhang proves such a result but only after one restricts the sum of error terms to smooth moduli. He then shows that the sums S_1 and $S_2^{(\ell)}$ from the GPY sieve are not significantly modified by restricting the sums to smooth n.

More precisely, we restrict our attention to primes lying in some interval $I = (1, N^{\varpi})$, for some small ϖ. We then define $\mathcal{S}_I \subset \mathbb{N}$ to be the subset consisting of all integers whose prime factors come from I. Ultimately, we will choose $\varpi \approx k^{-1/2}$. We then define the error term

$$E(N;q,a) = \left| \sum_{\substack{N \leq n \leq 2N \\ n \equiv a \pmod q}} \chi_{\mathcal{P}}(n) - \frac{1}{\varphi(q)} \sum_{N \leq n \leq 2N} \chi_{\mathcal{P}}(n) \right|.$$

Note that this is exactly the kind of error term that arises in $S_2^{(\ell)}$ in the GPY sieve.

Definition 2.4.1. Let $MPZ[\varpi]$ denote the assertion that for any fixed $A > 0$,

$$\sum_{\substack{q \leq N^{\frac{1}{2}+2\varpi} \\ q \in \mathcal{S}_I}} \sum_{\substack{a \in (\mathbb{Z}/q\mathbb{Z})^\times \\ \mathcal{P}(a) \equiv 0 \pmod q}} E(N;q,a) \ll \frac{N}{\log^A N},$$

where the implicit constant may depend on A.

Note that $MPZ[\varpi]$ is implied by $EH[\frac{1}{2} + 2\varpi]$, and can be viewed as being essentially $EH[\frac{1}{2} + 2\varpi]$ "with the sum restricted to smooth moduli."

Theorem 2.4.2 (Zhang)**.** *$MPZ[\varpi]$ is true for $0 < \varpi \leq \frac{1}{1168}$.*

This result is Zhang's main technical breakthrough. His proof builds on work by Fouvry-Iwaniec [4], Friedlander-Iwaniec [6], and Bombieri-Friedlander-Iwaniec [1], which also achieved higher levels of distribution at a cost of weakening the content of $EH[\theta]$. Remarkably, several experts believed that Theorem 2.4.2 was beyond the reach of these methods.

The proof of Theorem 2.4.2 is beyond the scope of this essay. Zhang's argument draws upon very deep mathematics, such Deligne's work on the Weil conjectures using ℓ-adic cohomology (although the Polymath later simplified the arguments to require only the one-dimensional case of the Weil Conjectures, which is comparatively elementary). In contrast, the theorem of Maynard and Tao that we will present is essentially elementary, depending only on the Bombieri-Vinogradov theorem.

In fact, Motohashi and Pintz had already proved that a smoothed version of the GPY sieve was enough to deduce bounded gaps, in the hopes that somebody would prove a result like Theorem 2.4.2. Building on their work, Zhang showed:

Theorem 2.4.3 (Motohashi-Pintz-Zhang)**.** *If $MPZ[\varpi]$ is true for any $\varpi > 0$, then $DHL[k,2]$ is true for some k.*

The basic idea is to define smoothed versions of the sums S_1 and S_2, where one restricts the sum to $n \in \mathcal{S}_I$, with $I = [1, x^{\varpi}]$.

$$S_1' = \sum_{\substack{N \leq n \leq 2N \\ n \in \mathcal{S}_I}} w(n)$$

and

$$S_2' = \sum_{\substack{N \le n \le 2N \\ n \in \mathcal{S}_I}} \chi_{\mathcal{P}}(n+h_i) w(n)$$

with weights

$$w(n) = \left(\sum_{\substack{d | \mathcal{P}(n) \\ d \in \mathcal{S}_I}} \lambda_d \right)^2 .$$

One that establishes asymptotics for S_1' and S_2' that are similar to those for S_1 and S_2: it turns out that if $k\varpi$ is fairly large, then not much is lost by the smoothing. This calculation is essentially elementary, but quite involved and tangential to the aim of this essay, so we omit it.

2.5 THE MAYNARD-TAO SIEVE

After Zhang's breakthrough, Maynard and Tao independently revisited the basic framework of the GPY sieve and discovered that it could be modified in a simple manner to achieve a better bound on small gaps between primes, and with an easier argument.

The key innovation is a more general choice of weights. We saw that the GPY sieve involves weights

$$w(n) = \left(\sum_{d | (n+h_1) \dots (n+h_k)} \lambda_d \right)^2 .$$

Tao and Maynard use the weights

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

The GPY weights are, of course, a special case of this. Perhaps somewhat surprisingly, it turns out that the extra flexibility in this more general choice of weight is enough to obtain bounded gaps between primes under any level of distribution.

Incredibly, it appears that experts had already essentially considered this choice of weight, and believed that it would not give any additional gains. According to [12], Selberg himself essentially wrote down these weights. Goldston and Pintz had also originally considered very similar multi-dimensional weights, with the same formula but a more restricted support.

With these more general weights, the argument proceeds almost identically to that of GPY. As before, one defines sums

$$S_1 = \sum_{n \in [N, 2N]} w(n) \quad \text{and} \quad S_2^{(\ell)} = \sum_{n \in [N, 2N]} \chi_{\mathcal{P}}(n+h_\ell) w(n) \tag{3}$$

Substituting the choice of weights turns the main terms of the sums into quadratic forms in the $\lambda_{d_1, \dots, d_k}$. After diagonalizing the quadratic forms, one defines $\lambda_{d_1,\ldots,d_k}$ in terms of a smooth function $G : \mathbb{R}^k \to \mathbb{R}$ supported on the unit simplex $\mathcal{R}_k := \{(t_1,\ldots,t_k) \colon t_1 + \ldots + t_k \leq 1\}$ by

$$\lambda_{d_1,\ldots,d_k} \approx \left(\prod_{i=1}^{k} \mu(d_i)\right) G\left(\frac{\log d_1}{\log R}, \ldots, \frac{\log d_k}{\log R}\right)$$

with some restrictions on the support. Then applying the estimates on sums of multiplicative functions, one finds that for any smooth function $F \colon \mathbb{R}^k \to \mathbb{R}$ there is a choice of weight with

$$S_1 \sim C(R,k) \int_0^1 \ldots \int_0^1 F(t_1,\ldots,t_k)^2 \, dt_1 \ldots dt_k$$

and

$$S_2 \sim C(R,k) \frac{\log R}{\log N} \sum_{m=1}^{k} \int_0^1 \ldots \int_0^1 \left(\int_0^1 F(t_1,\ldots,t_k) \, dt_m\right)^2 dt_1 \ldots dt_{m-1} \, dt_{m+1} \ldots dt_k.$$

The problem is then to choose an appropriate function F to maximize the ratio $S_2^{(\ell)}/S_1$. In the GPY sieve, we found that *no* choice of F could achieve a ratio of $\frac{1}{k}$ under Bombieri-Vinogradov, where $R = N^{1/2}$. With Maynard's sieve, however, we find that we have enough flexibility to achieve a ratio of about $\frac{\log k}{k}$, up to constants. This shows that for any m, $DHL[k,m]$ is true for sufficiently large k, and carefully bookkeeping yields:

Theorem 2.5.1.

$$\liminf_{n\to\infty} p_{n+m} - p_n \ll m^3 e^{4m}.$$

In particular, Maynard shows that under Bombieri-Vinogradov, for $k = 105$ there is a choice of F, expressed in terms of the elementary symmetric polynomials in $(t_1,\ldots,t_k)$, such that this ratio is large enough to imply $DHL[105,2]$. The bound

$$\lim_{n\to\infty} p_{n+1} - p_n < 600$$

then follows from finding an explicit 105-tuple with diameter 600. Under the Elliott-Halberstam conjecture, Maynard exhibits an explicit symmetric polynomial F such that the ratio is large enough to imply $DHL[5,2]$. By constructing an admissible 5-tuple with diameter 12, he establishes

$$\lim_{n\to\infty} p_{n+1} - p_n < 12 \text{ under Elliott-Halberstam.}$$

3 SMALL GAPS BETWEEN PRIMES

In this chapter, we embark on a complete and detailed study of the Maynard-Tao sieve. We mostly follow Maynard's treatment [12], but we work in slightly greater generality. At the end, we will also sketch Tao's approach as recorded in Polymath 8b [16].

3.1 THE BASIC SETUP

The Maynard-Tao sieve operates under the framework introduced by Goldston-Pintz-Yildirim. As before, the basic idea is to compare sums of the form

$$\sum w(n) \quad \text{and} \quad \sum w(n)\chi_{\mathcal{P}}(n+h_\ell).$$

However, we introduce a number of technical modifications to make the sieving cleaner. First, we employ the "W-trick" introduced by Green and Tao in their seminal paper [9]. Let D_0 be an integer growing very slowly with N and $W = \prod_{p \text{ prime } < D_0} p$. The precise choice of D_0 does not matter, but we choose $D_0 = \log\log\log N$, so that $W \ll \log\log N$ by the Prime Number Theorem. We will sieve over integers n lying in a given fixed residue class modulo W. Morally, this has the effect of "turning off" the primes less than D_0, which is useful for eliminating irregularities coming from small primes. For our purposes, the utility in turning off the small primes is that it increases the "probability" that two W-tricked integers are coprime.

We work in slightly more generality than Maynard's original paper. Maynard, following Goldston-Pintz-Yildirim, studied primes represented by translations $\{n+h_1, n+h_2, \ldots, n+h_k\}$. We will study primes represented by the more general linear forms $\{g_1 n+h_1, g_2 n+h_2, \ldots, g_k n+h_k\}$. We say that this set of forms is *admissible* if for every prime p, there is some n such that none of $\{g_1 n+h_1, g_2 n+h_2, \ldots, g_k n+h_k\}$ is divisible by p. Note that it is again an easy finite computation to verify that a form is admissible: the admissibility condition is automatically satisfied with respect to any prime $p > k$ not dividing any g_i. In this more general setting, we continue to think of admissibililty as meaning that there is no "obvious" divisibility obstruction to every member of $\{g_1 n+h_1, \ldots, g_k n+h_k\}$ being prime.

In light of this more general problem, we override our previous definition of DHL.

Definition 3.1.1. Let $DHL[k, m]$ denote the assertion that if $(g_1 n+h_1, \ldots, g_k n+h_k)$ is any admissible k-tuple, then for infinitely many integers n there are at least m primes among the set $\{g_1 n+h_1, \ldots, g_n+h_k\}$.

Remark 3.1.2. This differs from the notation used in Polymath 8 (which employs our previous definition of DHL), but all its results are true for our more general definition as well.

We will sieve over a residue class $\nu_0 \pmod W$ such that $g_i\nu_0 + h_i \neq 0 \pmod W$ for all i. Such a residue class exists by admissibility and the Chinese

Remainder Theorem (and its existence is the only point where admissibility is required in the subsequent arguments). We then define the sums

$$S_1 := \sum_{\substack{N \le n \le 2N \\ n \equiv \nu_0 \pmod W}} w(n) \quad \text{and} \quad S_2^{(\ell)} := \sum_{\substack{N \le n \le 2N \\ n \equiv \nu_0 \pmod W}} w(n) \chi_{\mathcal{P}}(g_\ell n + h_\ell)$$

where the $w(n)$ are positive weights to be specified. The goal is to show that $S_2^{(\ell)} > \frac{\rho}{k} S_1$ for some $\rho \ge \frac{1}{k}$ and all ℓ. If we can do this, then we would have

$$\sum_{\ell=1}^{k} S_2^{(\ell)} - \rho S_1 = \sum_{\substack{N \le n \le 2N \\ n \equiv \nu_0 \pmod W}} w(n) \left(\sum_{\ell=1}^{k} \chi(g_\ell n + h_\ell) - \rho \right) > 0.$$

This would imply that at least one summand is positive, and since $w(n)$ is positive, we may conclude that

$$\left(\sum_{\ell=1}^{k} \chi(g_\ell n + h_\ell) - \rho \right) > 0$$

for at least one n. Then it must be the case that at least $\lfloor \rho + 1 \rfloor$ of the integers in $\{g_\ell n + h_\ell \colon \ell = 1, \ldots, k\}$ are prime. This discussion proves:

Lemma 3.1.3. *If $S_2^{(\ell)} > \frac{\rho}{k} S_1$ for all sufficiently large N, then $DHL[k, \lfloor \rho + 1 \rfloor]$ holds.*

Thus far, everything is the same as in GPY. The key new ingredient is a more general choice of weights:

$$w(n) = \left(\sum_{\substack{d_1, \ldots, d_k \\ d_i | g_i n + h_i \forall i}} \lambda_{d_1, \ldots, d_k} \right)^2$$

where the $\lambda_{d_1,\ldots,d_k}$ will be specified later. (Note that the special case where $\lambda_{d_1,\ldots,d_k} =: \lambda_{d_1 d_2 \ldots d_k}$ is independent of the factorization reproduces the original GPY sieve). Throughout the chapter, we assume $EH[\theta]$ for some fixed $\theta \in (0,1)$ and choose a level of support $R = N^{\theta/2-\epsilon}$ for arbitrarily small $\epsilon > 0$. Then we will choose $\lambda_{d_1,\ldots,d_k}$ to be supported on tuples $(d_1, \ldots, d_k)$ for which the product $d = \prod d_i$ satisfies

$$\boxed{\begin{aligned} &(d, W) = 1 \\ &d < R \\ &d \text{ is square-free.} \end{aligned}} \tag{4}$$

In the sequel, we will be performing many mult-index sums with several conditions. We will follow the convention of listing the indices summed on the first line, and then the conditions below.

3.2 DIAGONALIZING THE QUADRATIC FORM

Again, our sums can be interpreted as quadratic forms in the $\lambda_{d_1,\ldots,d_k}$, and we begin by diagonalizing them. This is very similar to what we did for the GPY sieve.

Let us first establish a multivariable analogue of the "dual Möbius inversion" formula, which will be useful for implementing a diagonalizing change of variables.

Lemma 3.2.1. *[Multivariable dual Möbius inversion] Let $s_{d_1,\ldots,d_k}$ be a sequence of real numbers, supported on finitely many integer tuples $(d_1, \ldots, d_k)$. If*

$$y_{a_1,\ldots,a_k} = \sum_{\substack{d_1,\ldots,d_k \\ a_i | d_i \forall i}} s_{d_1,\ldots,d_k}$$

then

$$s_{d_1,\ldots,d_k} = \sum_{\substack{a_1,\ldots,a_k \\ d_i | a_i \forall i}} \prod_{i=1}^{k} \mu\left(\frac{a_i}{d_i}\right) y_{a_1,\ldots,a_k}. \tag{5}$$

Proof. Substituting the definition of $y_{a_1,\ldots,a_k}$ into the right hand side of the formula (5), we obtain

$$\begin{aligned}
\sum_{\substack{a_1,\ldots,a_k \\ d_i | a_i \forall i}} \prod_{i=1}^{k} \mu\left(\frac{a_i}{d_i}\right) y_{a_1,\ldots,a_k} &= \sum_{\substack{a_1,\ldots,a_k \\ d_i | a_i \forall i}} \prod_{i=1}^{k} \mu\left(\frac{a_i}{d_i}\right) y_{a_1,\ldots,a_k} \sum_{\substack{e_1,\ldots,e_k \\ a_i | e_i \forall i}} s_{e_1,\ldots,e_k} \\
&= \sum_{e_1,\ldots,e_k} s_{e_1,\ldots,e_k} \sum_{\substack{a_1,\ldots,a_k \\ d_i | a_i | e_i \forall i}} \prod_{i=1}^{k} \mu\left(\frac{a_i}{d_i}\right).
\end{aligned}$$

Now the inner sum is 0 unless $d_i = e_i$, since

$$\sum_{\substack{a'_1,\ldots,a'_k \\ a'_i | \frac{e_i}{d_i} \forall i}} \prod_{i=1}^{k} \mu(a'_i) = \prod_{i=1}^{k} \sum_{a'_i | \frac{e_i}{d_i} \forall i} \mu(a'_i) = \mathbb{1}(e_i = d_i \forall i).$$

□

Remark 3.2.2. Recall that we wish to choose $\lambda_{d_1,\ldots,d_k}$ supported on tuples $(d_1, \ldots, d_k)$ satisfying the support conditions (4). From the inversion formula, we see that all these are satisfied if we define $\lambda_{d_1,\ldots,d_k}$ in terms of $y_{a_1,\ldots,a_k}$ satisfying the same support conditions (4).

We shall also need the following estimates.

Lemma 3.2.3.

$$\sum_{n \leq X} \frac{\mu(n)^2}{\varphi(n)} \ll \log X.$$

Motivation. There is a useful general principles for establishing asymptotics of the form

$$\mathcal{M}_F(X) := \sum_{n \leq X} F(n) \sim c_F X$$

where F is some multiplicative function. If we write

$$F(n) = \sum_{d|n} f(d)$$

then

$$\begin{aligned}\sum_{n\le X} F(n) &= \sum_{n\le X}\sum_{d|n} f(d)\\ &= \sum_{d\le X} f(d)\left(\frac{X}{d}+O(1)\right)\\ &= X\sum_{d\le X}\frac{f(d)}{d} + O\left(\sum_{d\le X} f(d)\right).\end{aligned}$$

If $\sum_{d\le X}\frac{f(d)}{d}$ converges to some constant c_F and we can control the error term in a satisfactory manner, then we find that $\sum_{n\le X} F(n)$ has "mean value" c_F. This says that $F(n)$ is c_F "on average."

Proof. We are not quite in the situation described above, since the arithmetic function we are summing is $\frac{\mu(n)^2}{\varphi(n)}$, which decreases to 0. So we instead estimate

$$\sum_{n\le X}\frac{\mu(n)^2 n}{\varphi(n)}$$

and then use partial summation to estimate the sum that we are really interested in. Note that since $\varphi(n) \asymp n$ "on average," we do expect the above sum to have a mean value.

In our case, it is technically slicker to apply a slight variant of these ideas. We note the inequality

$$\frac{\mu(n)^2 n}{\varphi(n)} \le \sum_{d|n}\frac{\mu(d)^2}{\varphi(d)}.$$

Indeed, if n is squarefree, we can check that this is an equality by noting that both sides are multiplicative and agree on primes. If n is not squarefree, then the left hand side is 0 and the right hand side is positive.

Therefore, if we let $F(n) = \frac{\mu(n)^2 n}{\varphi(n)}$ we have

$$\mathcal{M}_F(X) = \sum_{n\le X} F(n) \le X\sum_{d\le X}\frac{\mu(d)^2}{d\varphi(d)} + O\left(\sum_{d\le X}\frac{1}{\varphi(d)}\right) \ll X$$

since

$$\sum_{d\le X}\frac{\mu(d)^2}{d\varphi(d)} < \sum_{d=1}^{\infty}\frac{\mu(d)^2}{d\varphi(d)} < \infty$$

by considering the Euler product, for instance.

Now, we write the sum as a Riemann-Stieltjes integral and apply integration by parts:

$$\begin{aligned}\sum_{n\le X}\frac{\mu(n)^2}{\varphi(n)} &= \int_1^X \frac{d\mathcal{M}_F(y)}{y}\\ &= \frac{\mathcal{M}_F(X)}{X} + \int_1^X \mathcal{M}_F(y)\frac{dy}{y^2}\\ &\ll 1 + \int_1^X \frac{dy}{y}\\ &\ll \log X.\end{aligned}$$

□

Lemma 3.2.4. *If $q \le X$ is squarefree, then*

$$\sum_{\substack{n \le X \\ (n,q)=1}} \frac{\mu(n)^2}{\varphi(n)} \ll \frac{\varphi(q)}{q} \log X.$$

Intuitively, one expects that for "smooth" arithmetic functions $f(n)$, the sums

$$\sum_{\substack{n \le X \\ n \equiv a \pmod q}} f(n)$$

should be roughly equal across all primitive residue classes a. Lemma 3.2.4 quantifies this in a special case.

Proof. By term-by-term comparisons of the sums, we have

$$\sum_{n \le qX} \frac{\mu(n)^2}{\varphi(n)} \ge \left(\sum_{d|q} \frac{\mu(d)^2}{\varphi(d)} \right) \sum_{\substack{n \le X \\ (n,q)=1}} \frac{\mu(n)^2}{\varphi(n)}.$$

By Lemma 3.2.3,

$$\sum_{n \le qX} \frac{\mu(n)^2}{\varphi(n)} \ll \log(qX) \ll \log X.$$

Therefore, it suffices to show that

$$\sum_{d|q} \frac{\mu(d)^2}{\varphi(d)} = \frac{q}{\varphi(q)}.$$

Indeed, since q is squarefree we have

$$\sum_{d|q} \frac{\mu(d)^2}{\varphi(d)} = \prod_{p|q} \left(1 + \frac{1}{p-1} \right) = \prod_{p|q} \frac{p}{\varphi(p)} = \frac{q}{\varphi(q)}.$$

□

Now we are ready to diagonalize S_1.

Proposition 3.2.5. *Let S_1 be as defined above and assume that $\lambda_{d_1,\dots,d_k}$ are real numbers satisfying the conditions in (4). Define*

$$y_{a_1,\dots,a_k} = \left(\prod_{i=1}^{k} \varphi(a_i)\mu(a_i) \right) \sum_{\substack{d_1,\dots,d_k \\ a_i | d_i \forall i}} \frac{\lambda_{d_1,\dots,d_k}}{\prod_i d_i}$$

and let

$$y_{\max} = \max_{a_1,\dots,a_k} |y_{a_1,\dots,a_k}|.$$

Then

$$S_1 = \frac{N}{W} \sum_{a_1,\dots,a_k} \left(\prod_{i=1}^{k} \frac{1}{\varphi(u_i)} \right) y_{a_1,\dots,a_k}^2 + O\left(\frac{y_{\max}^2 N \varphi(W)^k (\log R)^k}{W^{k+1} D_0} \right).$$

Proof. This is essentially just a computation, similar in spirit to what we did for the GPY sums (25), although the details of bounding all the error terms are rather involved. The idea is to substitute the formula for the weights to express S_1 as a quadratic form in the $\lambda_{d_1,\dots,d_k}$, and then use arithmetic identities to diagonalize it. As far as those steps are concerned, the procedure is a straightforward multivariable generalization of the argument in §2.3.2, but there is one additional complication in that separating the divisors introduces a little additional coupling in the sum. The difference from this extra coupling is the main error term.

$$\begin{aligned} S_1 &= \sum_{\substack{N\le n\le 2N \\ n\equiv \nu_0 \pmod W}} w(n) \\ &= \sum_{\substack{N\le n\le 2N \\ n\equiv \nu_0 \pmod W}} \left(\sum_{\substack{d_1,\dots,d_k \\ d_i \mid g_i n + h_i \forall i}} \lambda_{d_1,\dots,d_k} \right)^2 \\ &= \sum_{\substack{N\le n\le 2N \\ n\equiv \nu_0 \pmod W}} \left(\sum_{\substack{d_1,\dots,d_k \\ d_i \mid g_i n + h_i \forall i}} \lambda_{d_1,\dots,d_k} \right) \left(\sum_{\substack{e_1,\dots,e_k \\ e_i \mid g_i n + h_i \forall i}} \lambda_{e_1,\dots,e_k} \right) \\ &= \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k}} \lambda_{d_1,\dots,d_k} \lambda_{e_1,\dots,e_k} \sum_{\substack{N\le n\le 2N \\ n\equiv \nu_0 \pmod W \\ [d_i,e_i] \mid g_i n + h_i \forall i}} 1. \end{aligned}$$

Now, the inner sum is a set of congruence conditions on n modulo W and $[d_i, e_i]$ for all i. If these moduli are all coprime, then the inner sum is $\frac{N}{W\prod [d_i,e_i]}$. We claim that if they are not all coprime, then the contribution to the sum is 0. Indeed, by hypothesis $\lambda_{d_1,\dots,d_k}$ is non-zero only if $(d_i, d_j) = 1$ for all i, j and also $(d_i, W) = 1$ for all i, so a common prime factor can only come from d_i and e_j. But if $p \mid d_i$ and e_j, then $p \mid g_j(g_i n + h_i) - g_i(g_j n + h_j) = g_j h_i - g_i h_j$, all of whose prime factors divide W for all sufficiently large N, so the support conditions implies that there is no contribution from these terms. Therefore,

$$S_1 = \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,d_k \\ (d_i,e_j)=1 \forall i\ne j}} \lambda_{d_1,\dots,d_k} \lambda_{e_1,\dots,e_k} \frac{N}{W\prod_{i=1}^k [d_i,e_i]} + O\left(\sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,d_k \\ (d_i,e_j)=1 \forall i\ne j}} |\lambda_{d_1,\dots,d_k} \lambda_{e_1,\dots,e_k}| \right).$$

Let us analyze the error term. Define $\tau_k(n)$ to be the number of tuples $(d_1, \dots, d_k)$ such that $\prod d_i = n$. Then the error term can be bounded as

$$\ll \lambda_{\max}^2 \left(\sum_{n<R} \tau_k(n) \right)^2.$$

Lemma 3.2.6.

$$\sum_{n\le R} \tau_k(n) \ll R(\log R)^{k-1}.$$

Proof. The result is clear for $k = 1$, since $\tau_1(n) = 1$. By induction, we may assume it for $k-1$ and show it for k:

$$\sum_{n\le R} \tau_k(n) = \sum_{d\le R} \sum_{n\le R/d} \tau_{k-1}(n/d) \ll \sum_{d\le R} \frac{R}{d} \log\left(\frac{R}{d}\right)^{k-2} \ll R(\log R)^{k-1}.$$

□

It suffices to use the more conservative bound $\tau_k(n) \ll R(\log R)^k$, since this error term will later be subsumed by another, so we do so just for convenience. So far, we have

$$S_1 = \frac{N}{W} \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k \\ (d_i,e_j)=1 \forall i\neq j}} \frac{\lambda_{d_1,\dots,d_k} \lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k [d_i,e_i]} + O\left(\lambda_{\max}^2 R^2 (\log R)^{2k}\right).$$

Now, the main term is similar to the analogous term in the GPY sieve, but the sum over coprime integers introduces extra coupling to handle. It will turn out that the W-trick increases the density of coprime pairs of integers just enough to make this extra coupling insignificant. For now, we can replace this condition by inserting sums that will sieve out the coprime terms. Recall that

$$\sum_{s_{ij} \mid d_i, e_j} \mu(s_{ij}) = \begin{cases} 1 & (d_i, e_j) = 1, \\ 0 & \text{otherwise} \end{cases}$$

so we can rewrite the main term of S_1 as

$$\frac{N}{W} \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k}} \left(\prod_{i\neq j} \sum_{s_{ij}\mid d_i,e_j} \mu(s_{ij}) \right) \frac{\lambda_{d_1,\dots,d_k} \lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k [d_i,e_i]}.$$

Now we proceed with diagonalizing the sum. Using the identity

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

to decouple the terms, we may rewrite the main term as

$$\begin{aligned} S_1 &\approx \frac{N}{W} \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k}} \left(\prod_{i\neq j} \sum_{s_{ij}\mid d_i,e_j} \mu(s_{ij}) \right) \frac{\lambda_{d_1,\dots,d_k}}{\prod_{i=1}^k d_i} \frac{\lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k e_i} \prod_{i=1}^k (d_i, e_i) \\ &= \frac{N}{W} \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k}} \left(\prod_{i\neq j} \sum_{s_{ij}\mid d_i,e_j} \mu(s_{ij}) \right) \frac{\lambda_{d_1,\dots,d_k}}{\prod_{i=1}^k d_i} \frac{\lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k e_i} \prod_{i=1}^k \left(\sum_{\substack{u_1,\dots,u_k \\ u_i \mid d_i,e_i \forall i}} \varphi(u_i) \right) \\ &= \frac{N}{W} \sum_{u_1,\dots,u_k} \prod_{i=1}^k \varphi(u_i) \left(\prod_{i\neq j} \sum_{s_{ij}} \mu(s_{ij}) \right) \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k \\ u_i\mid d_i,e_i \forall i \\ s_{ij}\mid d_i,e_j \forall i,j}} \frac{\lambda_{d_1,\dots,d_k}}{\prod_{i=1}^k d_i} \frac{\lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k e_i}. \end{aligned}$$

Note that in the above sum, we may assume that u_i and s_{ij} are coprime for $j \neq i$, since $a_i \mid e_i$ and $\lambda_{e_1,\ldots,e_k}$ is supported only on tuples satisfying $(e_i, e_j) = 1$ for all i, j. Therefore, if we set

$$a_i = u_i \prod_{j \neq i} s_{ij} \text{ and } b_i = u_i \prod_{j \neq i} s_{ji}$$

then we may write the main term as

$$S_1 \approx \frac{N}{W} \sum_{u_1,\ldots,u_k} \prod_{i=1}^{k} \varphi(u_i) \left(\prod_{i \neq j} \sum_{s_{ij}} \mu(s_{ij}) \right) \sum_{\substack{d_1,\ldots,d_k \\ a_i | d_i \forall i}} \frac{\lambda_{d_1,\ldots,d_k}}{\prod d_i} \sum_{\substack{e_1,\ldots,e_k \\ b_i | e_i \forall i}} \frac{\lambda_{e_1,\ldots,e_k}}{\prod e_i}.$$

Now we perform a change of variables. Let

$$y_{a_1,\ldots,a_k} = \left(\prod_{i=1}^{k} \mu(a_i) \varphi(a_i) \right) \sum_{\substack{d_1,\ldots,d_k \\ a_i | d_i \forall i}} \frac{\lambda_{d_1,\ldots,d_k}}{\prod_i d_i}.$$

Lemma 3.2.1 shows that this change of variables is invertible, with inverse

$$\frac{\lambda_{d_1,\ldots,d_k}}{\prod_{i=1}^{k} d_i} = \sum_{\substack{a_1,\ldots,a_k \\ d_i | a_i \forall i}} \mu\left(\frac{a_i}{d_i}\right) \mu(a_i) \frac{y_{a_1,\ldots,a_k}}{\prod_{i=1}^{k} \varphi(a_i)} \tag{6}$$

$$= \prod_{i=1}^{k} \mu(d_i) \sum_{\substack{a_1,\ldots,a_k \\ d_i | a_i \forall i}} \frac{y_{a_1,\ldots,a_k}}{\prod_{i=1}^{k} \varphi(a_i)} \tag{7}$$

since we restrict the support to squarefree integers anyway.

Substituting this change of variables above, the main term of S_1 becomes

$$\frac{N}{W} \sum_{u_1,\ldots,u_k} \left(\prod_{i=1}^{k} \varphi(u_i) \right) \left(\prod_{i \neq j} \sum_{s_{ij}} \mu(s_{ij}) \right) \prod_{i=1}^{k} \frac{\mu(a_i)}{\varphi(a_i)} y_{a_1,\ldots,a_k} \prod_{i=1}^{k} \frac{\mu(b_i)}{\varphi(b_i)} y_{b_1,\ldots,b_k}$$
$$= \frac{N}{W} \sum_{u_1,\ldots,u_k} \left(\prod_{i=1}^{k} \frac{\mu(u_i)^2}{\varphi(u_i)} \right) \left(\prod_{i \neq j} \sum_{s_{ij}} \frac{\mu(s_{ij})}{\varphi(s_{ij})^2} \right) y_{a_1,\ldots,a_k} y_{b_1,\ldots,b_k}.$$

When all the s_{ij} are 1, the contribution to the sum is

$$\frac{N}{W} \sum_{u_1,\ldots,u_k} \left(\prod_i \frac{1}{\varphi(u_i)} \right) y_{u_1,\ldots,u_k}^2.$$

This is the main term we want, so we bound the contribution from the other terms. By our restriction on the support of $y_{a_1,\ldots,a_k}$, the only non-zero contribution comes from s_{ij} coprime to W, which by the definition of W forces any prime divisor of s_{ij} to be larger than D_0. Therefore, if $s_{ij} > 1$ then in fact $s_{ij} > D_0$. Summing over all possible such s_{ij}, and factoring out the common terms, we find that the total contribution from all $s_{ij} > 1$ is

$$\ll y_{\max}^2 \frac{N}{W} \underbrace{\left(\sum_{\substack{D_0 < s_{ij} \leq R \\ (s_{ij}, W) = 1}} \frac{\mu(s_{ij})^2}{\varphi(s_{ij})^2} \right)}_{\ll D_0^{-1}} \underbrace{\left(\sum_{\substack{u \leq R \\ (u, W) = 1}} \frac{\mu(u)^2}{\varphi(u)} \right)^k}_{\ll \left(\frac{\varphi(W) \log R}{W} \right)^k} \underbrace{\left(\sum_{\substack{1 \leq s \leq R \\ (s, W) = 1}} \frac{\mu(s)^2}{\varphi(s)^2} \right)^{k^2 - k - 1}}_{\ll 1}$$

where we have used Lemma 3.2.4 to estimate the middle term. So we can bound the above error term as

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

We conclude that

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

To finish, we want to combine the error terms. To do this we must compare $y_{\max}$ and $\lambda_{\max}$, which we do using the inversion formula. From (6) we obtain

$$\begin{aligned}|\lambda_{d_1,\ldots,d_n}| &= \left|\prod_{i=1}^k d_i \sum_{\substack{a_1,\ldots,a_k \\ d_i|a_i\forall i}} \prod \frac{\mu(a_i)^2}{\varphi(a_i)} y_{a_1,\ldots,a_k}\right| \\ &\ll y_{\max}\prod_{i=1}^k \frac{d_i}{\varphi(d_i)}\left(\sum_{n\le R/\prod d_i}\frac{\mu(n)^2}{\varphi(n)}\tau_k(n)\right).\end{aligned}$$

Now, we use another useful arithmetic identity for squarefree integers d:

$$\frac{d}{\varphi(d)} = \sum_{a|d}\frac{1}{\varphi(a)},$$

which can be proved by observing that both sides are multiplicative, and checking them on primes. Since $y_{d_1,\ldots,d_k}$ is only supported on tuples where $\prod_{i=1}^k d_i$ is squarefree, we can substitute this as

$$\begin{aligned}|\lambda_{d_1,\ldots,d_n}| &\ll y_{\max}\left(\sum_{a|\prod d_i}\frac{\mu(a)^2}{\varphi(a)}\right)\left(\sum_{n\le R/\prod d_i}\frac{\mu(n)^2}{\varphi(n)}\tau_k(n)\right) \\ &\ll y_{\max}\sum_{n\le R}\frac{\mu(n)^2}{\varphi(n)}\tau_k(n) \\ &\ll y_{\max}\left(\sum_{n\le R}\frac{\mu(n)^2}{\varphi(n)}\right)^k \\ &\ll y_{\max}(\log R)^k.\end{aligned}$$

Therefore, the second error term is $\ll y_{\max}^2 R^2 (\log R)^{4k}$. Since we take $R^2 = N^{\theta-2\epsilon}$, this is subsumed by the first error term for all sufficiently large N. □

Remark 3.2.7. It is not obvious that the error term in Proposition 3.2.5 is smaller than the main term. We will study the main term carefully in the next section, but for now let us note that

$$\sum_{u_1,\ldots,u_k}\left(\prod_{i=1}^k\frac{1}{\varphi(u_i)}\right) = \left(\sum_{\substack{u\le R\\(u,W)=1}}\frac{1}{\varphi(u_i)}\right)^k \approx \left(\frac{\varphi(W)\log R}{W}\right)^k$$

so if $y^2_{u_1,\ldots,u_k}$ does not decrease very rapidly then we expect the main term to have size about

$$y^2_{\max} \frac{N}{W} \left(\frac{\varphi(W) \log R}{W} \right)^k.$$

This just beats out the error term by a factor of $D_0 = \log\log\log N$. In fact, this rough approximation turns out to be essentially correct.

Next, we embark on a similar study of $S_2^{(\ell)}$.

Proposition 3.2.8. *Let $S_2^{(\ell)}$ be as defined above and assume that $\lambda_{d_1,\ldots,d_k}$ are real numbers satisfying the conditions in (4). Define*

$$y^{(\ell)}_{a_1,\ldots,a_k} = \left(\prod_{i=1}^k \mu(a_i) g(a_i) \right) \sum_{\substack{d_1,\ldots,d_k \\ d_\ell = 1 \\ a_i | d_i \forall i}} \frac{\lambda_{d_1,\ldots,d_k}}{\prod_{i=1}^k \varphi(d_i)}$$

and let

$$y^{(\ell)}_{\max} = \max_{a_1,\ldots,a_k} |y^{(\ell)}_{a_1,\ldots,a_k}|.$$

Then

$$S_2^{(\ell)} = \frac{N}{\varphi(W) \log N} \sum_{\substack{u_1,\ldots,u_k \\ u_\ell = 1}} \left(\prod_{i=1}^k \frac{1}{g(u_i)} \right) (y^{(\ell)}_{u_1,\ldots,u_k})^2 + O\left(\frac{N (y^{(\ell)}_{\max})^2 \varphi(W)^{k-2} (\log R)^{k-2}}{W^{k-1} D_0} \right).$$

Proof. Again, the proof is similar in structure to the analogous argument that we have given in the GPY (one-variable) case. As before, we substitute in the formula for the weights to obtain a quadratic form, and use the $EH[\theta]$ assumption to control the error term. The main term is a quadratic form in the $\lambda_{d_1,\ldots,d_k}$, and we then make a change of variables to diagonalize it. Many of the details in manipulating the sums and error terms are identical to those in the proof of Proposition 3.2.5.

We start out by writing

$$\begin{aligned} S_2^{(\ell)} &= \sum_{\substack{N \le n \le 2N \\ n \equiv v_0 \pmod W}} \chi_{\mathcal{P}}(g_\ell n + h_\ell) w(n) \\ &= \sum_{\substack{N \le n \le 2N \\ n \equiv v_0 \pmod W}} \chi_{\mathcal{P}}(g_\ell n + h_\ell) \left(\sum_{\substack{d_1,\ldots,d_k \\ d_i | g_i n + h_i \forall i}} \lambda_{d_1,\ldots,d_k} \right) \left(\sum_{\substack{e_1,\ldots,e_k \\ e_i | g_i n + h_i \forall i}} \lambda_{e_1,\ldots,e_k} \right) \\ &= \sum_{\substack{d_1,\ldots,d_k \\ e_1,\ldots,e_k}} \lambda_{d_1,\ldots,d_k} \lambda_{d_1,\ldots,e_k} \sum_{\substack{N \le n \le 2N \\ n \equiv v_0 \pmod W \\ [d_i, e_i] | g_i n + h_i \forall i}} \chi_{\mathcal{P}}(g_\ell n + h_\ell) \end{aligned}$$

Again, there is no contribution unless $(d_i, e_j) = 1$ for all i, j and also $(d_i, W) = 1$ for all i. When that is the case, the inner sum counts the number of primes in the interval $[g_\ell N + h_\ell, 2 g_\ell N + h_\ell]$ satisfying several congruence conditions. For all sufficiently large N, the condition that $[d_i, e_i] \mid g_i n + h_i$ is a single congruence condition on n modulo $[d_i, e_i]$ (specifically, $D_0 > g_i$ is enough to guarantee that g_i is coprime to all $[d_i, e_i]$). Therefore, for all sufficiently large N the sum counts primes in the range $[g_\ell N + h_\ell, 2g_\ell N + h_\ell]$ congruent to a specific residue class modulo $g_\ell W \prod_{i=1}^k [d_i, e_i]$.

We must obviously have $d_\ell = e_\ell = 1$, since both divide $g_\ell n + h_\ell$ by assumption, and there are no other restrictions. Recall that, by using admissibility and the Chinese Remainder Theorem, we chose ν_0 so that the residue class of interest is indeed coprime to the modulus. Define $E(N,q)$ to be the maximum difference between the average number of primes per admissible residue class mod q and the actual number:

$$E(N,q) = \sup_{(a,q)=1} \max_{a \le q} \left| \frac{1}{\varphi(q)} \sum_{N \le n \le 2N} \chi_P(n) - \sum_{\substack{N \le n \le 2N \\ n \equiv a \pmod q}} \chi_P(n) \right|.$$

Letting $X_n = \pi(2g_\ell N) - \pi(g_\ell N)$, we can write the inner sum as

$$\sum_{\substack{N \le n \le 2N \\ n \equiv \nu_0 \pmod W \\ [d_i, e_i] | g_i n + h_i \forall i}} \chi_{\mathcal{P}}(g_\ell n + h_\ell) = \frac{X_N}{\varphi(g_\ell W \prod_{i=1}^k [d_i, e_i])} + O(E(g_\ell N, g_\ell W \prod_{i=1}^k [d_i, e_i]))$$

Here we have ignored the translations by h_ℓ in the prime number function, since they make a minute difference that can certainly be absorbed into the error term above.

Substituting this above into our formula for $S_2^{(\ell)}$ (and simplifying $\varphi(g_\ell W) = g_\ell \varphi(W)$), we obtain

$$S_2^{(\ell)} = \frac{X_N}{g_\ell \varphi(W)} \sum_{\substack{d_1, \ldots, d_k \\ e_1, \ldots, e_k \\ (d_i, e_j) = 1 \forall i \ne j \\ e_\ell = d_\ell = 1}} \frac{\lambda_{d_1, \ldots, d_k} \lambda_{e_1, \ldots, e_k}}{\prod_{i=1}^k \varphi([d_i, e_i])} + O\left(\sum_{\substack{d_1, \ldots, d_k \\ e_1, \ldots, e_k}} \lambda_{d_1, \ldots, d_k} \lambda_{e_1, \ldots, e_k} E(N, q) \right)$$

where $q = g_\ell W \prod_{i=1}^k [d_i, e_i]$. The number of tuples $(d_1, \ldots, d_k)$ and $(e_1, \ldots, e_k)$ with $[d_i, e_i] = n$ is bounded above by $\tau_{3k}(n)$, since we can think of the first k integers are describing $\gcd(d_i, e_i)$ for $i = 1, \ldots, k$. Therefore, the error term is

$$\ll \lambda_{\max}^2 \sum_{q \le g_\ell R^2 W} \tau_{3k}(q) E(N, q).$$

The assumption $EH[\theta]$ implies that for any $A > 0$, we have a bound

$$\sum_{q \le R^2 W} E(N, q) \ll \frac{N}{(\log N)^A}.$$

We use Cauchy-Schwarz to bound this separately in terms of the sum over τ_{3k} and the sum over $E(N,q)$. To balance the exponents correctly, we have to distribute $E(N,q)^{1/2}$ in both factors, and in one we use the trivial bound $E(N,q) \ll \frac{N}{\varphi(q)}$ (which follows from the Prime Number Theorem in arithmetic progressions, for instance). Since $R^2W \ll N^\theta$, our assumption $EH[\theta]$ implies that

$$\begin{aligned}\sum_{q\le g_\ell R^2W}\tau_{3k}(q)E(N,q) &\ll \sum_{q\le g_\ell R^2W}\tau_{3k}(q)E(N,q)^{1/2}E(N,q)^{1/2}\\ &\ll \left(\sum_{q\le g_\ell R^2W}\tau_{3k}(q)^2E(N,q)\right)^{1/2}\left(\sum_{q\le g_\ell R^2W}E(N,q)\right)^{1/2}\\ &\ll \left(\sum_{q\le g_\ell R^2W}\tau_{3k}(q)^2\frac{N}{\varphi(q)}\right)^{1/2}\left(\sum_{q\le g_\ell R^2W}E(N,q)\right)^{1/2}\\ &\ll N^{1/2}(\log N)^{6k}\frac{N^{1/2}}{(\log N)^{A/2}}.\end{aligned}$$

We then insert this back into the error term of our expression for $S_2^{(\ell)}$. After replacing A by a larger constant to absorb all the factors of $\log N$, we obtain

$$S_2^{(\ell)} = \frac{X_N}{g_\ell\varphi(W)}\sum_{\substack{d_1,\dots,d_k\\ e_1,\dots,e_k\\ (d_i,e_j)=1\forall i\ne j\\ e_\ell=d_\ell=1}}\frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k\varphi([d_i,e_i])} + O\left(\lambda_{\max}^2\frac{N}{(\log N)^A}\right).$$

Now, we proceed the diagonalize as before. We insert the factor

$$\prod_{i\ne j}\sum_{s_{ij}|d_i,e_j}\mu(s_{ij})$$

to sieve out the condition $(d_i,e_j)=1$ for all $i\ne j$. Then we seek an arithmetic function $g(n)$ such that

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

By Möbius inversion, g is multiplicative and we find that $g(p)=p-2$. Since we choose $\lambda_{d_1,\dots,d_k}$ to be supported on tuples of squarefree integers, we need only define g on primes (not prime powers). With this, the main term for $S_2^{(\ell)}$ becomes

$$\frac{X_N}{g_\ell\varphi(W)}\sum_{\substack{d_1,\dots,d_k\\ e_1,\dots,e_k\\ e_\ell=d_\ell=1}}\left(\prod_{i\ne j}\sum_{s_{ij}|d_i,e_j}\mu(s_{ij})\right)\frac{\lambda_{d_1,\dots,d_k}}{\prod_{i=1}^k\varphi(d_i)}\frac{\lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k\varphi(e_i)}\sum_{u_i|d_i,e_i}g(u_i).$$

As before, we note that the only indices in the sum that contribute are those with $(u_i,s_{ij})=1$ for all i,j by our assumptions on the support of λ. We again define

$$a_i = u_i\prod_{j\ne i}s_{ij}\text{ and }b_i = u_i\prod_{j\ne i}s_{ji}.$$

Swapping the order of summation above, the main term becomes

$$\frac{X_N}{g_\ell\varphi(W)}\sum_{\substack{u_1,\dots,u_k\\ u_\ell=1}}\prod_{i=1}^k g(u_i)\left(\prod_{\substack{i\ne j\\ i,j\ne\ell}}\sum_{s_{ij}|d_i,e_j}\mu(s_{ij})\right)\sum_{\substack{d_1,\dots,d_k\\ e_1,\dots,e_k\\ e_\ell=d_\ell=1\\ a_i|d_i\forall i\\ b_i|e_i\forall i}}\frac{\lambda_{d_1,\dots,d_k}}{\prod\varphi(d_i)}\frac{\lambda_{e_1,\dots,e_k}}{\prod\varphi(e_i)}.$$

Now we introduce the change of variables

$$y^{(\ell)}_{a_1,\dots,a_k} = \left(\prod_{i=1}^{k}\mu(a_i)g(a_i)\right)\sum_{\substack{d_1,\dots,d_k\\ d_\ell=1\\ a_i|d_i\forall i}}\frac{\lambda_{d_1,\dots,d_k}}{\prod_{i=1}^{k}\varphi(d_i)}$$

so that the sum becomes

$$S_2 = \frac{X_N}{g_\ell\varphi(W)}\sum_{\substack{u_1,\dots,u_k\\ u_\ell=1}}\prod_{i=1}^{k}\frac{\mu(u_i)^2}{g(u_i)}\left(\prod_{\substack{i\neq j\\ i,j\neq\ell}}\sum_{s_{ij}|d_i,e_j}\frac{\mu(s_{ij})}{g(s_{ij})^2}\right)y^{(\ell)}_{a_1,\dots,a_k}y^{(\ell)}_{b_1,\dots,b_k}$$
$$+O\left(\lambda^2_{\max}\frac{N}{(\log N)^A}\right).$$

It remains only to consider the contribution from $s_{ij}=1\forall i,j$, and then bound the contribution from the other terms. Arguing as before, if $s_{ij}>1$ then $s_{ij}>D_0$, so the contribution from such terms is at most

$$(y^{(\ell)}_{\max})^2\frac{X_N}{g_\ell\varphi(W)}\underbrace{\sum_{s_{ij}>D_0}\frac{\mu(s_{ij})^2}{g(s_{ij})^2}}_{\ll D_0^{-1}}\underbrace{\left(\sum_{\substack{u\leq R\\(u,W)=1}}\frac{\mu(u)^2}{g(u)}\right)^{k-1}}_{\ll\left(\frac{\varphi(W)\log R}{W}\right)^{k-1}}\underbrace{\left(\sum_{\substack{s\leq R\\(s,W)=1}}\frac{\mu(s)^2}{g(s)^2}\right)^{k^2-3k+1}}_{\ll 1}.$$

Using the Prime Number Theorem to estimate $X_N \ll \frac{N}{\log N}$, we see that after splitting the sum into the cases $s_{ij}=1\forall i,j$ or $s_{ij}>1$ for some i,j we have

$$S^{(\ell)}_2 = \frac{X_N}{g_\ell\varphi(W)}\sum_{\substack{u_1,\dots,u_k\\ u_\ell=1}}\left(\prod_{i=1}^{k}\frac{\mu(u_i)^2}{g(u_i)}\right)(y^{(\ell)}_{u_1,\dots,u_k})^2$$
$$+O\left(\frac{N(y^{(\ell)}_{\max})^2\varphi(W)^{k-2}(\log R)^{k-2}}{W^{k-1}D_0}\right)+O\left(\frac{(\lambda^{(\ell)}_{\max})^2N}{(\log N)^A}\right).$$

Note that by a very similar argument to that in the proof of Proposition 3.2.5, we have again $\lambda^2_{\max}\ll(y^{(\ell)}_{\max})^2(\log R)^{2k}$. Since we can adjust the constant A to be as large as we want, at only the cost of a larger implicit constant, and $W\ll\log\log N$, we can absorb the second error term into the first. Therefore,

$$S^{(\ell)}_2 = \frac{X_N}{g_\ell\varphi(W)}\sum_{\substack{u_1,\dots,u_k\\ u_\ell=1}}\left(\prod_{i=1}^{k}\frac{\mu(u_i)^2}{g(u_i)}\right)(y^{(\ell)}_{u_1,\dots,u_k})^2+O\left(\frac{N(y^{(\ell)}_{\max})^2\varphi(W)^{k-2}(\log R)^{k-2}}{W^{k-1}D_0}\right).$$

Since $\mu(u_i)$ is square-free if $y_{u_1,\dots,u_k}$ is non-zero, we may replace $\mu(u_i)^2$ by 1. Finally, we apply the Prime Number Theorem to estimate $X_N=\pi(2g_\ell N)-\pi(g_\ell N)$ as

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

(we can actually take any constant for the exponent of log in the error term, but 2 suffices) and the error term in the Prime Number Theorem contributes

$$\ll \frac{(y_{\max}^{(\ell)})^2 N}{\varphi(W)(\log N)^2} \underbrace{\left(\sum_{\substack{a<R \\ (a,W)=1}} \frac{\mu(u)^2}{g(u)} \right)^{k-1}}_{\ll \left(\frac{\varphi(W)\log R}{W}\right)^{k-1}},$$

which can be absorbed into the first main term above. Finally, the difference between using $\log N$ and $\log N + \log g_\ell$ also gives a negligible error, so we conclude that

$$S_2^{(\ell)} = \frac{N}{\varphi(W)\log N} \sum_{\substack{u_1,\ldots,u_k \\ u_\ell=1}} \left(\prod_{i=1}^{k} \frac{1}{g(u_i)} \right) (y_{u_1,\ldots,u_k}^{(\ell)})^2 + O\left(\frac{N(y_{\max}^{(\ell)})^2 \varphi(W)^{k-2}(\log R)^{k-2}}{W^{k-1}D_0} \right)$$

as desired. □

In order to compare the sums $S_2^{(\ell)}$ and S_1, it remains to relate $y_{a_1,\ldots,a_k}^{(\ell)}$ and $y_{a_1,\ldots,a_k}$. Since both are defined in terms of $\lambda_{d_1,\ldots,d_k}$, this is in principle a straightforward exercise in applying our dual Möbius inversion formula, but there is again some calculation involved in obtaining a nice approximation

Note that in Proposition 3.2.8 we only have to work with weights $y_{a_1,\ldots,a_k}^{(\ell)}$ satisfying $a_\ell = 1$, so those are the only weights we solve for.

Lemma 3.2.9. *Keeping the notation above, if $a_\ell = 1$ then*

$$y_{a_1,\ldots,a_k}^{(\ell)} = \sum_{e_\ell \le R} \frac{y_{a_1,\ldots,e_\ell,\ldots,a_k}}{\varphi(e_\ell)} + O\left(\frac{y_{\max}\varphi(W)\log R}{WD_0} \right).$$

Proof. We defined

$$y_{a_1,\ldots,a_k}^{(\ell)} = \left(\prod_{i=1}^{k} \mu(a_i) g(a_i) \right) \sum_{\substack{d_1,\ldots,d_k \\ d_\ell=1 \\ a_i|d_i \forall i}} \frac{\lambda_{d_1,\ldots,d_k}}{\prod_{i=1}^{k} \varphi(d_i)}.$$

Also recall from (6) that

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

Substituting this above, we obtain

$$\begin{aligned} y_{a_1,\ldots,a_k}^{(\ell)} &= \left(\prod_{i=1}^{k} \mu(a_i) g(a_i) \right) \sum_{\substack{d_1,\ldots,d_k \\ d_\ell=1 \\ a_i|d_i \forall i}} \frac{\lambda_{d_1,\ldots,d_k}}{\prod_{i=1}^{k} \varphi(d_i)} \\ &= \left(\prod_{i=1}^{k} \mu(a_i) g(a_i) \right) \sum_{\substack{d_1,\ldots,d_k \\ d_\ell=1 \\ a_i|d_i \forall i}} \prod_{i=1}^{k} \frac{d_i \mu(d_i)}{\varphi(d_i)} \sum_{\substack{e_1,\ldots,e_k \\ d_i|e_i \forall i}} \frac{y_{e_1,\ldots,e_k}}{\prod_{i=1}^{k} \varphi(e_i)} \\ &= \left(\prod_{i=1}^{k} \mu(a_i) g(a_i) \right) \sum_{\substack{e_1,\ldots,e_k \\ a_i|e_i \forall i}} \frac{y_{e_1,\ldots,e_k}}{\prod_{i=1}^{k} \varphi(e_i)} \sum_{\substack{d_1,\ldots,d_k \\ d_\ell=1 \\ a_i|d_i|e_i \forall i}} \prod_{i=1}^{k} \frac{d_i \mu(d_i)}{\varphi(d_i)}. \end{aligned}$$

We can evaluate the innermost sum explicitly. Write $d_i = a_i d_i'$ and $e_i = a_i e_i'$. We may assume that $e_1 \dots e_k$ is squarefree (otherwise $y_{e_1,\dots,e_k}$ is zero), so since all the arithmetic functions involved are multiplicative,

$$\frac{d_i \mu(d_i)}{\varphi(d_i)} = \frac{a_i \mu(a_i)}{\varphi(a_i)} \frac{d_i' \mu(d_i')}{\varphi(d_i')}.$$

We can then factorize the sum into an Euler product

$$\prod_{i \neq \ell}^{k} \frac{a_i \mu(a_i)}{\varphi(a_i)} \prod_{p | e_i'} \left(1 + \frac{\mu(p)}{p-1}\right) = \prod_{i=1}^{k} \frac{a_i \mu(a_i)}{\varphi(a_i)} \prod_{p | e_i'} \left(\frac{-1}{p-1}\right)$$
$$= \prod_{i=1}^{k} \frac{a_i \mu(e_i)}{\varphi(e_i)}.$$

Substituting this above, we obtain

$$y^{(\ell)}_{a_1,\dots,a_k} = \left(\prod_{i=1}^{k} \mu(a_i) g(a_i)\right) \sum_{\substack{e_1,\dots,e_k \\ a_i | e_i \forall i}} \frac{y_{e_1,\dots,e_k}}{\prod_{i \neq \ell} \varphi(e_i)} \prod_{i \neq \ell}^{k} \frac{a_i \mu(e_i)}{\varphi(e_i)}.$$

Observe that in the indices a_j with $j \neq \ell$, the coefficient of summation is $\frac{1}{\varphi(e_j)^2} \approx \frac{1}{e_j^2}$, so we expect this contribution to be small and the main term to come from $j = \ell$, where the coefficient of summation is $\frac{1}{\varphi(e_\ell)}$. By the assumptions on the support of $y_{e_1,\dots,e_k}$, either $e_j = a_j$ or $e_j > D_0 a_j$. So let us bound the contribution in the second case for $j \neq \ell$: it is

$$\ll y_{\max} \prod_{i=1}^{k} g(a_i) a_i \underbrace{\left(\sum_{e_j > D_0 a_j} \frac{\mu(e_j)^2}{\varphi(e_j)^2}\right)}_{\ll \frac{1}{D_0}} \underbrace{\left(\sum_{\substack{e_\ell < R \\ (e_\ell, W)=1}} \frac{\mu(e_\ell)^2}{\varphi(e_\ell)}\right)}_{\ll \frac{\varphi(W) \log R}{W}} \underbrace{\left(\prod_{i \neq j, \ell} \sum_{\substack{e_i \leq R \\ (e_i, W)=1}} \frac{\mu(e_i)^2}{\varphi(e_i)^2}\right)}_{\ll 1}$$
$$\ll \frac{y_{\max} \varphi(W) \log R}{W D_0}.$$

The rest of the sum is when $e_j = a_j$ for all $j \neq \ell$, in which case we get a contribution of

$$\left(\prod_{i=1}^{k} \frac{\mu(a_i)^2 a_i g(a_i)}{\varphi(a_i)^2}\right) \sum_{e_\ell} \frac{y_{a_1,\dots,e_\ell,\dots,a_k}}{\varphi(e_\ell)}.$$

We can drop the $\mu(a_i)^2$ since the weights are supported on squarefree indices anyway. We thus find that

$$y^{(\ell)}_{a_1,\dots,a_k} = \left(\prod_{i=1}^{k} \frac{a_i g(a_i)}{\varphi(a_i)^2}\right) \sum_{e_\ell} \frac{y_{a_1,\dots,e_\ell,\dots,a_k}}{\varphi(e_\ell)} + O\left(\frac{y_{\max} \varphi(W) \log R}{W D_0}\right).$$

Finally, observe that

$$\left(\prod_{i=1}^{k} \frac{a_i g(a_i)}{\varphi(a_i)^2}\right) = \prod_{p | a_i} \left(1 - \frac{1}{(p-1)^2}\right) = 1 + O(D_0^{-1})$$

since any prime dividing a_i is greater than D_0. □

3.3 SUMS OF MULTIPLICATIVE FUNCTIONS

We now establish some results on sums of multiplicative functions that will be useful in estimating the asymptotics for S_1 and $S_2^{(\ell)}$ that we obtained in Propositions 3.2.5 and 3.2.8. The general setup is that we want to estimate

$$\sum_{n \le X} f(n)F(n)$$

where f is some multiplicative function and F is a smooth function. If we had a good estimate of

$$\mathcal{M}_f(X) := \sum_{n \le X} f(n)$$

then we could apply Riemann-Stieltjes integration to estimate

$$\sum_{n \le X} f(n)F(n) = \int_1^X d\mathcal{M}_f(y)F(y)$$

Therefore, our first step is to establish some results on approximating sums of the form $\mathcal{M}_f(X)$. Suppose that f is some arithmetic function such that $f(p) \approx \frac{k}{p}$, like the functions $f(n) = \frac{1}{\varphi(n)}$ or $\frac{1}{g(n)}$ that appear in S_1 and S_2 (with $k = 1$). We expect that such a sum will have asymptotics like $\mathcal{M}_f(X) \asymp (\log x)^k$, since it is comparable to the kth power of the harmonic series, and we now work towards making this precise. In [12], Maynard deals with these sums by citing general results related to the Selberg-Delange method. In order to keep our treatment accessible and self-contained, we present an elementary argument from [10] for an alternate result that is good enough for our purposes.

Definition 3.3.1. We define the *Dirichlet series* associated to f,

$$L_f(s) = \sum_{n=1}^{\infty} \frac{f(n)}{n^{-s}}.$$

A priori, this is only a formal object, but with modest bounds on $|f(n)|$ (as arise in most circumstances) it will define a holomorphic function in a right half-plane of the form $\mathrm{Re}\, s > \sigma$. By unique prime factorization and the multiplicativity of f, this series factorizes as an Euler product

$$L_f(s) = \prod_{p \text{ prime}} \left(1 + f(p)p^{-s} + f(p^2)p^{-2s} + \ldots\right).$$

Note that in the case where $f(n) = 1$ for all n, this is the Euler product for the Riemann zeta function. The formal inverse may be written as .

$$L_f(s)^{-1} = \prod_{p \text{ prime}} \left(1 + f(p)p^{-s} + f(p^2)p^{-2s} + \ldots\right)^{-1}.$$

Continuing with the analogy with the Riemann zeta function, we can take the formal logarithmic derivative of the Euler product to obtain

$$\begin{aligned}\frac{d}{ds}\log L_f(s) &= \sum_{p \text{ prime}} \frac{d}{ds}\log\left(1+f(p)p^{-s}+f(p^2)p^{-2s}+\dots\right)\\ &= \sum_{p \text{ prime}} \frac{d}{ds}\sum_{k=1}^{\infty}(-1)^{k+1}\frac{\left(f(p)p^{-s}+f(p^2)p^{-2s}+\dots\right)^k}{k}\\ &= -\sum_{p \text{ prime}} \log p\left(f(p)p^{-s}+2f(p^2)p^{-2s}+\dots\right)\\ &\qquad \sum_{k=1}^{\infty}(-1)^{k+1}\left(f(p)p^{-s}+f(p^2)p^{-2s}+\dots\right)^{k-1}\end{aligned}$$

Remark 3.3.2. Note that Λ_f is supported on prime powers. We will typically want to apply our results in the case where f is supported on squarefree integers. In this case, $f(p^2) = f(p^3) = \dots = 0$, so our equation above simplifies to show that

$$\Lambda_f(p^k) = f(p)^k \log p.$$

Definition 3.3.3. If f is a multiplicative function, then we define the multiplicative function $\Lambda_f(n)$ by the formal identity

$$-\frac{L_f'(s)}{L_f(s)} = \sum_{n=1}^{\infty}\frac{\Lambda_f(n)}{n^s}.$$

Since

$$-L_f'(s) = \sum_{n=1}^{\infty}(f(n)\log n)n^{-s},$$

comparing Dirichlet coefficients in the identity

$$L_f(s)\left(\frac{L_f'(s)}{L_f(s)}\right) = L_f'(s)$$

implies the convolution identity

$$f(n)\log n = (f*\Lambda)(n) := \sum_{d\mid n} f(d)\Lambda\left(\frac{n}{d}\right). \tag{8}$$

Proposition 3.3.4. *Let $\kappa > 0$. Suppose that f is a multiplicative function satisfying*

$$\sum_{n\le x}\Lambda_f(n) = \kappa\log x + O(1)$$

for some $\kappa \ge 0$, and

$$\sum_{n\le x}|f(n)| \ll \log^{\kappa} x.$$

Then

$$\sum_{n\le x} f(n) = \frac{\mathfrak{S}(f)}{\Gamma(\kappa+1)}\log^{\kappa} x + O\left(\log^{\kappa-1} x\right),$$

where

$$\mathfrak{S}(f) = \prod_{p \text{ prime}}\left(1+f(p)p^{-1}+f(p^2)p^{-2}+\dots\right)\left(1-\frac{1}{p}\right)^{\kappa}.$$

Remark 3.3.5. One should think of the condition

$$\sum_{n \le x} \Lambda_f(n) = \kappa \log x + O(1)$$

as saying that $f(p) \approx \frac{\kappa}{p}$ on average. Indeed, suppose that f is supported on squarefree integers. By Remark 3.3.2, the sum can be rewritten as

$$\sum_{p \text{ prime } \le x} \log p \sum_{k=1}^{\infty} f(p)^k.$$

If $f(p) \approx \frac{\kappa}{p}$, then the contribution from all of the higher prime powers is $\ll 1$, since $\frac{\log p}{p^2} \ll p^{2-\epsilon}$, so the sum is

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

It is a general principle that summing a nice function $f(n)$ over primes, weighted by $\log p$, is commensurate with summing over all n. Indeed, one formulation of the Prime Number Theorem gives the estimate on Chebyshev's Ψ function:

$$\Psi(x) := \sum_{n \le x} \Lambda(n) = x + O\left(\frac{x}{(\log x)^2}\right).$$

Rewriting the sum in terms of Riemann-Stieltjes integration and applying integration by parts, we see that

$$\begin{aligned}
\sum_{p \text{ prime } \le x} \frac{\log p}{p} &= \int_1^x d\Psi(t) \frac{1}{t} \\
&= \frac{\Psi(t)}{t}\Big|_1^x + \int_1^x \frac{\Psi(t) dt}{t^2} \\
&= O(1) + \int_1^x \frac{t\,dt}{t^2} + O\left(\int_1^x \frac{dt}{t \log^2 t}\right) \\
&= \log x + O(1).
\end{aligned}$$

So we have shown that if f is supported on squarefree integers and $\lim_{p \to \infty} \frac{f(p)}{p} = \kappa$, then

$$\sum_{n \le x} \Lambda_f(n) = \kappa \log x + O(1).$$

Proof. The first key step is to consider the sum

$$\sum_{n \le x} f(n) \log\left(\frac{x}{n}\right).$$

This is a useful trick to study general arithmetic sums, since it can be viewed as a smoothed version of the sum: the $\log\left(\frac{x}{n}\right)$ term acts as a smoothing factor since it slowly decays to 0 as $n \to x$.

We express this sum in two different ways. On one hand, by Riemann-Stieltjes integration by parts, it is

$$\sum_{n \le x} f(n) \log\left(\frac{x}{n}\right) = \int_1^x d\mathcal{M}_f(t) \log\left(\frac{x}{t}\right) = \int_1^x \mathcal{M}_f(t) \frac{dt}{t}. \tag{9}$$

On the other hand, we can use the convolution identity (8) and the bounds in the hypothesis to calculate:

$$\begin{aligned}\sum_{n\le x} f(n)\log\left(\frac{x}{n}\right) &= \mathcal{M}_f(x)\log x - \sum_{n\le x} f(n)\log n \\ &= \mathcal{M}_f(x)\log x - \sum_{n\le x}\sum_{d|n} f(d)\Lambda\left(\frac{n}{d}\right) \\ &= \mathcal{M}_f(x)\log x - \sum_{d\le x} f(d)\sum_{e\le \frac{x}{d}}\Lambda_f(e) \\ &= \mathcal{M}_f(x)\log x - \sum_{d\le x} f(d)\left(\kappa\log\left(\frac{x}{d}\right)+O(1)\right) \\ &= \mathcal{M}_f(x)\log x - \kappa\sum_{d\le n} f(d)\log\left(\frac{x}{d}\right)+O(\log^\kappa x).\end{aligned}$$

Rearranging, we find that

$$(\kappa+1)\sum_{n\le x} f(n)\log\left(\frac{x}{n}\right) = \mathcal{M}_f(x)\log x + O\left(\log^\kappa x\right).$$

Substituting (9), we may rewrite this as

$$\mathcal{M}_f(x)\log x = \Delta(x) + (\kappa+1)\int_2^x \mathcal{M}_f(t)t^{-1}dt \tag{10}$$

where $\Delta(x) \ll \log^\kappa x$ (we have changed the bounds of integration to avoid any problems with dividing by $\log t$ later on).

Now comes the second key trick: we divide both sides by $x\log^{\kappa+2} x$ and integrate. Since we believe that $\mathcal{M}_f(x) \approx \log^\kappa x$, this is the correct scale for the integrals to converge to constants of interest.

$$\int_2^x \frac{\mathcal{M}_f(t)}{t\log^{\kappa+1} t}dt = \int_2^x \frac{\Delta(t)}{t\log^{\kappa+2} t}dt + (\kappa+1)\int_2^x \frac{dt}{t\log^{\kappa+2} t}\int_2^t \frac{\mathcal{M}_f(u)}{u}du. \tag{11}$$

Exchanging the order of integration on the last term gives

$$\begin{aligned}(\kappa+1)\int_2^x \frac{dt}{t\log^{\kappa+2} t}\int_2^t \frac{\mathcal{M}_f(u)\,du}{u} &= \int_2^x \frac{\mathcal{M}_f(u)\,du}{u}(\kappa+1)\int_u^x \frac{dt}{t\log^{\kappa+2} t} \\ &= \int_2^x \frac{\mathcal{M}_f(u)\,du}{u}\left(\frac{1}{\log^{\kappa+1} u} - \frac{1}{\log^{\kappa+1} x}\right).\end{aligned}$$

We substitute this back into (11), noticing that the left hand side cancels exactly with one term on the right hand side to leave

$$\frac{1}{\log^{\kappa+1} x}\int_2^x \frac{\mathcal{M}_f(u)du}{u} = \int_2^x \frac{\Delta(t)}{t\log^{\kappa+2} t}dt.$$

Finally, we can substitute this into (10) to obtain

$$\mathcal{M}_f(x)\log x = \Delta(x) + \log^{\kappa+1} x\int_2^x \frac{\Delta(t)}{t\log^{\kappa+2} t}dt.$$

Since $\Delta(t) \ll \log^\kappa t$, the integral $\int_2^\infty \frac{\Delta(t)}{t\log^{\kappa+2} t}dt$ converges absolutely, and the cost of extending the integral to $[2,\infty)$ is $\int_x^\infty \frac{\Delta(t)}{t\log^{\kappa+2} t}dt \ll \frac{1}{\log x}$, so we have

$$\mathcal{M}_f(x) = c\log^\kappa x + O(\log^{\kappa-1} x)$$

with $c = \int_2^\infty \frac{\Delta(t)}{t \log^{\kappa+2} t}\, dt$.

Now it remains to show that $c = \mathfrak{S}(f)$. The point is that the asymptotic form tells us the order of the pole of $L_f(s)$ at $s = 0$, and we can then find the constant of proportionality by comparison with the Riemann zeta function, which has a simple pole with residue 1 at $s = 1$. Indeed, the asymptotic that we just derived shows that the series $L_f(s)$ converges absolutely for $\operatorname{Re} s > 0$, and for any such s we may write

$$\begin{aligned} L_f(s) &= \int_1^\infty d\mathcal{M}_f(t) t^{-s} \\ &= \int_1^\infty \mathcal{M}_f(t) d(t^{-s}) \\ &= \int_0^\infty \mathcal{M}_f(e^u) d(e^{-su}) \\ &= \int_0^\infty (cu^\kappa + O(u^{\kappa-1})) e^{-su} du \\ &= \left(cs^{-\kappa} + O(s^{-\kappa-1})\right) \Gamma(\kappa+1). \end{aligned}$$

Since $\zeta(s)$ has a simple pole at $s = 1$, we see that

$$\zeta(s+1)^{-\kappa} L_f(s) \sim c_f \Gamma(\kappa+1).$$

On the other hand, we have an Euler product for $\operatorname{Re} s > 0$:

$$\zeta(s+1)^{-\kappa} L_f(s) = \prod_{p \text{ prime}} \left(1 - p^{-s-1}\right)^\kappa (1 + f(p)p^{-s} + f(p^2)p^{-2s} + \ldots),$$

which converges to $\mathfrak{S}(f)$ as $s \to 0^+$. □

Proposition 3.3.6. *Let $\kappa > 0$. Suppose that f is a multiplicative function satisfying the hypotheses of Proposition 3.3.4. Let $F : [0,1] \to \mathbb{R}$ be a differentiable function and*

$$F_{\max} = \sup_{x \in [0,1]} |F(x)| + |F'(x)|.$$

Then

$$\sum_{n \le x} f(n) F\left(\frac{\log n}{\log x}\right) = \frac{\mathfrak{S}(f) \log^\kappa x}{\Gamma(\kappa)} \int_0^1 u^{\kappa-1} F(u)\, du + O\left(F_{\max} (\log x)^{\kappa-1}\right).$$

Proof. We first rewrite the sum in terms of Riemann-Stieltjes integration:

$$\sum_{n \le x} f(n) F\left(\frac{\log n}{\log x}\right) = \int_1^X d\mathcal{M}_f(t) F\left(\frac{\log t}{\log x}\right).$$

Now we may substitute $\mathcal{M}_f(t) = \frac{\mathfrak{S}(f)}{\Gamma(\kappa+1)} \log^\kappa t + E(t)$, where $E(t) \ll \log^{\kappa-1} t$ by Proposition 3.3.4. We obtain

$$\begin{aligned} \sum_{n \le X} f(n) F\left(\frac{\log n}{\log x}\right) &= \int_1^X d\mathcal{M}_f(t) F\left(\frac{\log t}{\log x}\right) \\ &= \int_1^x d\Big(\frac{\mathfrak{S}(f) \log^\kappa t}{\Gamma(\kappa+1)} + E(t)\Big) F\left(\frac{\log t}{\log x}\right) \\ &= \mathfrak{S}(f) \int_1^x \frac{\kappa \log^{\kappa-1} t\, dt}{\Gamma(\kappa+1) t} F\left(\frac{\log t}{\log x}\right) + \int_1^x dE(t) F\left(\frac{\log t}{\log x}\right). \end{aligned}$$

Performing the substitution $t = x^u$, the first term becomes

$$\mathfrak{S}(f) \int_0^1 \frac{\kappa(u \log x)^{\kappa-1}}{\Gamma(\kappa+1)} F(u) \, \log x \, du,$$

which is the main term we want. To bound the error term, we perform integration by parts, and again make the substitution $t = x^u$:

$$\begin{aligned} \int_1^x dE(t) F\left(\frac{\log t}{\log x}\right) &= E(t)F(t)\big|_1^x + O\left(\int_1^x E(t) F'\left(\frac{\log x}{\log t}\right) \frac{dt}{t \log x}\right) \\ &\ll (u \log x)^{\kappa-1} F(u)\big|_0^1 + (\log x)^{\kappa-1} \int_0^1 u^{\kappa-1} F'(u) \, du \\ &\ll (\log x)^{\kappa-1} F_{\max}. \end{aligned}$$

□

3.4 CHOICE OF WEIGHTS

We are now ready to make our choice of weights. For any smooth function $F : \mathbb{R}^k \to \mathbb{R}$, supported in the simplex $\mathcal{R}_k := \{(t_1, \ldots, t_k) \colon t_1 + \ldots + t_k \leq 1\}$, we define

$$y_{a_1,\ldots,a_k} = F\left(\frac{\log a_1}{\log R}, \ldots, \frac{\log a_k}{\log R}\right) \tag{12}$$

if $a := \prod a_i$ is coprime to W and squarefree; otherwise we set $y_{a_1,\ldots,a_k} = 0$. Note that our condition on the support of F automatically guarantees that $y_{a_1,\ldots,a_k} = 0$ if $a > R$. By the inversion formula, $\lambda_{d_1,\ldots,d_k}$ satisfies the support conditions that we promised in (4).

Remark 3.4.1. The work of Goldston-Pintz-Yildirim is recovered in the special case where $F(t_1, \ldots, t_k) = F(t_1 + \ldots + t_k)$ depends only on the sum of the arguments, since that corresponds to weights

$$w(n) = \left(\sum_{d | (g_1 n + h_1) \ldots (g_k n + h_k)} \lambda_d\right)^2.$$

Proposition 3.4.2. *Let $F : \mathbb{R}^k \to \mathbb{R}$ be a smooth function supported on $\mathcal{R}_k$, and let $y_{a_1,\ldots,a_k}$ be defined in terms of F by* (12). *If*

$$F_{\max} = \sup_{(t_1,\ldots,t_k) \in \mathcal{R}_k} \left(|F(t_1, \ldots, t_k)| + \sum_{i=1}^k \left| \frac{\partial F}{\partial x_i}(t_1, \ldots, t_k) \right| \right)$$

then

$$S_1 \sim \frac{N \varphi(W)^k (\log R)^k}{W^{k+1}} \int_0^1 \ldots \int_0^1 F(t_1, \ldots, t_k)^2 dt.$$

Proof. Substituting (12) into Proposition 3.2.5 gives

$$S_1 = \frac{N}{W} \sum_{\substack{a_1,\ldots,a_k \\ (a_i, a_j) = 1 \forall i \neq j \\ (a_i, W) = 1 \forall i}} \left(\prod_{i=1}^k \frac{1}{\varphi(a_i)}\right) F\left(\frac{\log a_1}{\log R}, \ldots, \frac{\log a_k}{\log R}\right)^2 + O\left(\frac{F_{\max}^2 N \varphi(W)^k (\log R)^k}{W^{k+1} D_0}\right).$$

We seek to remove the coupling condition $(a_i, a_j) = 1$. If (a_i, a_j) have a common factor, then it must be larger than D_0 since $(a_i, W) = 1$. By the same arguments as were used in the proofs of Propositions 3.2.5 and 3.2.8, the error in dropping this condition is $\ll F_{\max}^2 \frac{N\varphi(W)^k(\log R)^k}{W^{k+1}D_0}$, which can be absorbed into the error term above (indeed, recall that the original error term came from estimating the cost of dropping this kind of condition!).

Now S_1 can be estimated by applying Proposition 3.3.6 with respect to each index, with f being the multiplicative function defined by $f(n) = \frac{\mu(n)^2}{\varphi(n)}$ if $(n, W) = 1$ and $f(n) = 0$ otherwise. The hypothesis

$$\sum_{n \le x} |f(n)| \ll \log^\kappa x$$

is satisfied for $\kappa = 1$ by Lemma 3.2.3, and the hypothesis

$$\sum_{n \le x} \Lambda_f(n) = \kappa \log x + O(1)$$

is satisfied by Remark 3.3.5. Note that for this f, the singular series simplifies as

$$\mathfrak{S}(f) = \prod_{p \nmid W} \left(1 + \frac{1}{p-1}\right) \prod_{p \text{ prime}} \left(1 - \frac{1}{p}\right) = \prod_{p \mid W} \frac{p-1}{p} = \frac{\varphi(W)}{W}$$

since the factors cancel for all primes $p \nmid W$. Therefore, Proposition 3.3.6 says that

$$S_1 = \frac{N\varphi(W)^k(\log R)^k}{W^{k+1}} \int_0^1 \cdots \int_0^1 F(t_1, \ldots, t_k)^2 dt + O\left(\frac{F_{\max}^2 \varphi(W)^k N (\log R)^k}{W^{k+1}D_0}\right)$$

and the result follows by observing that for fixed F, the main term dominates due to the $D_0^{-1} = (\log\log\log N)^{-1}$ factor in the error term. □

We now apply a similar analysis of $S_2^{(\ell)}$.

Proposition 3.4.3. *Let $F : \mathbb{R}^k \to \mathbb{R}$ be a function that is differentiable and supported on $\mathcal{R}_k$, and let $y_{a_1,\ldots,a_k}$ be defined in terms of F by (12). If*

$$F_{\max} = \sup_{(t_1,\ldots,t_k) \in \mathcal{R}_k} \left(|F(t_1, \ldots, t_k)| + \sum_{i=1}^k \left| \frac{\partial F}{\partial x_i}(t_1, \ldots, t_k) \right| \right)$$

then

$$S_2^{(\ell)} \sim \frac{N\varphi(W)^k(\log R)^{k+1}}{W^{k+1}\log N} \int_0^1 \cdots \int_0^1 \left(\int_0^1 F(t_1, \ldots, t_k) dt_\ell \right)^2 dt_1 \ldots \widehat{dt_\ell} \ldots dt_k$$

Proof. Without loss of generality, we can prove the proposition in the case $\ell = k$.

We want to substitute our choice of weights into the diagonalized form in Proposition 3.2.8, but we cannot do it immediately since the proposition is phrased in terms of $y_{a_1,\ldots,a_k}^{(k)}$. So we first use the change of variables formula in Lemma 3.2.9, which gives

$$y_{a_1,\ldots,a_{k-1},1}^{(k)} = \sum_{e_k \le R} \frac{y_{a_1,\ldots,a_{k-1},e_k}}{\varphi(e_k)} + O\left(\frac{y_{\max}\varphi(W)\log R}{WD_0}\right).$$

Substituting (12), we find that for $a_1, \ldots, a_{k-1}$ satisfying the conditions for the support of $y_{a_1,\ldots,a_{k-1},1}$,

$$y^{(k)}_{a_1,\ldots,a_{k-1},1} = \sum_{\substack{e_k \le R \\ (e_k, W\prod a_i)=1}} \frac{\mu(e_k)^2}{\varphi(e_k)} F\left(\frac{\log a_1}{\log R}, \ldots, \frac{\log a_{k-1}}{\log R}, \frac{\log e_k}{\log R}\right) + O\left(\frac{F_{\max}\varphi(W)\log R}{WD_0}\right).$$

By this and Lemma 3.2.4, we note that

$$y_{\max} \ll \frac{\varphi(W)F_{\max}\log R}{W}. \tag{13}$$

Now, we see that this change of variables itself takes the form of a sum weighted by multiplicative functions, and we are in position to use Proposition 11 to estimate it. By the same argument as in the previous proof, we may apply Proposition 11 with $f(n) = \mu(n)^2\varphi(n)$ and $\kappa = 1$ to deduce that

$$y^{(\ell)}_{a_1,\ldots,a_{k-1},1} = \left(\prod_{i=1}^{k-1}\frac{\varphi(a_i)}{a_i}\right)\frac{\varphi(W)\log R}{W}\int_0^1 F(a_1,\ldots,a_{k-1},t)\,dt + O\left(\frac{F_{\max}\varphi(W)\log R}{WD_0}\right).$$

We next substitute this into the diagonalized form for S_2 in Proposition 3.2.8, which said that

$$S_2^{(k)} = \frac{N}{\varphi(W)\log N}\sum_{\substack{a_1,\ldots,a_{k-1} \\ (a_i,a_j)=1\forall i,j \\ (a_i,W)=1\forall i}}\left(\prod_{i=1}^{k-1}\frac{\mu(a_i)^2}{g(a_i)}\right)(y^{(k)}_{a_1,\ldots,a_{k-1},1})^2 + O\left(\frac{N(y^{(\ell)}_{\max})^2\varphi(W)^{k-2}(\log R)^{k-2}}{W^{k-1}D_0}\right).$$

By (13), we can bound the error term here by

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

Now let's turn to the main term:

$$S_2^{(k)} \approx \frac{N\varphi(W)(\log R)^2}{W^2\log N}\sum_{\substack{a_1,\ldots,a_{k-1} \\ (a_i,a_j)=1\forall i,j \\ (a_i,W)=1}}\left(\prod_{i=1}^{k-1}\frac{\mu(a_i)^2\varphi(a_i)^2}{g(a_i)a_i^2}\right)\left(\int_0^1 F(a_1,\ldots,a_{k-1},t)\,dt\right)^2.$$

We want to remove the coupling condition $(a_i, a_j) = 1$, so let's estimate the cost of doing so. By the same argument as has been used many times before, we see that any common prime factor must be greater than D_0, so the contribution from all primes that divide a_i and a_j is

$$\ll \frac{F_{\max}^2 N\varphi(W)(\log R)^2}{W^2\log N}\frac{1}{D_0}\left(\frac{\varphi(W)\log R}{W}\right)^{k-1} \ll \frac{F_{\max}^2 N\varphi(W)^k(\log R)^k}{W^{k+1}D_0},$$

which can be absorbed into the error term that we already have.

We can now conclude that

$$S_2^{(k)} = \frac{N\varphi(W)(\log R)^2}{W^2\log N}\sum_{\substack{a_1,\ldots,a_{k-1} \\ (a_i,W)=1}}\left(\prod_{i=1}^{k-1}\frac{\mu(a_i)^2\varphi(a_i)^2}{g(a_i)a_i^2}\right)\left(\int_0^1 F(a_1,\ldots,a_{k-1},t)\,dt\right)^2 + O\left(\frac{NF_{\max}^2\varphi(W)^k(\log R)^\prime}{W^{k+1}D_0}\right).$$

This fits into the framework of Proposition 11 with the multiplicative function $f(n) = \frac{\mu(n)^2\varphi(n)^2}{g(n)n^2}$ if $(n, W\prod a_i) = 1$ and $f(n) = 0$ and $\kappa = 1$. To see this one can prove an analog of of Lemma 3.2.4, but it is fairly clear that $f(n) \approx \frac{\mu(n)^2}{n}$ for typical n. So we can again apply Proposition 11 $k-1$ times to conclude that

$$S_2^{(k)} = \frac{N\varphi(W)^k(\log R)^{k+1}}{W^{k+1}\log N}\int_0^1\cdots\int_0^1\left(\int_0^1 F(t_1,\ldots,t_k)dt_k\right)^2 dt_1\ldots dt_{k-1} + O\left(\frac{NF_{\max}^2\varphi(W)^k(\log R)^k}{W^{k+1}D_0}\right).$$

Again, the main term dominates due to the D_0^{-1} factor in the error term. □

3.5 OPTIMIZATION OF WEIGHTS FOR SMALL k

Recall that our goal is to show that $\sum_{\ell=1}^k S_2^\ell - k\rho S_1 \geq 0$ for some $\rho \geq 1$, which would imply that there are infinitely many n such that at least $\lfloor \rho+1 \rfloor$ of the $g_i n + h_i$ are prime. By our calculations in Proposition 3.4.2 and 3.4.3,

$$S_1 \sim \frac{N\varphi(W)^k(\log R)^k}{W^{k+1}}\int_0^1\cdots\int_0^1 F(t_1,\ldots,t_k)^2 dt.$$

and

$$S_2^{(\ell)} \sim \frac{N\varphi(W)^k(\log R)^{k+1}}{W^{k+1}\log N}\int_0^1\cdots\int_0^1\left(\int_0^1 F(t_1,\ldots,t_k)dt_k\right)^2 dt_1\ldots dt_{k-1}.$$

We have then reduced the problem to one of studying the integrals

$$I_k(F) := \int_0^1\cdots\int_0^1 F(t_1,\ldots,t_k)^2 dt$$

and

$$J_k^{(\ell)}(F) := \int_0^1\cdots\int_0^1\left(\int_0^1 F(t_1,\ldots,t_k)dt_\ell\right)^2 dt_1\ldots\widehat{dt_\ell}\ldots dt_k.$$

Recall that we are choosing $R = N^{\theta/2-\epsilon}$, so

$$\frac{S_2^{(\ell)}}{S_1} \sim \frac{(\theta/2-\epsilon)J_k^{(\ell)}(F)}{I_k(F)}.$$

We want to choose F so as to maximize the preceding ratio.

Definition 3.5.1. We define

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

where the supremum is over smooth functions $F : \mathbb{R}^k \to \mathbb{R}$ supported in $\mathcal{R}_k$.

We can rephrase Lemma 3.1.3 in terms of M_k, which reduces the problem of finding prime tuples to a functional optimization problem.

Corollary 3.5.2. *EH*$[\theta]$ *implies* $DHL[\lfloor \frac{\theta}{2}M_k + 1\rfloor, 2]$.

We can obtain lower bounds on M_k by specializing to specific functions F. In particular, to obtain $DHL[k,2]$ and hence bounded gaps between primes, we need to show that $\lfloor \frac{\theta}{2} M_k + 1 \rfloor > 1$ for some k. Under Bombieri-Vinogradov, we can take $\theta = \frac{1}{2} - \epsilon$, so we just need to establish that $M_k > 4$ for some k. The work of Goldston-Pintz-Yildirim is the special case of this discussion where F is a function of $t_1 + \ldots + t_k$, and in that case the problem was that the ratio never exceeded 4. Therefore, we will have to consider more general functions.

By symmetry, the maximum can be achieved by a symmetric function. Furthermore, any continuous symmetric function on the compact simplex $\mathcal{R}_k$ can be uniformly approximated by symmetric polynomials, so we may as well restrict ourselves to functions of the form

$$F(t_1, \ldots, t_k) = \begin{cases} P(t_1, \ldots, t_k) & (t_1, \ldots, t_k) \in \mathcal{R}_k, \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

where $P(t_1, \ldots, t_k)$ is a symmetric polynomial. Strictly speaking, this function is not smooth, but we can clearly construct smooth approximations for which the ratio M_k will be arbitrarily close to that of F.

Since we know that we have to go beyond the first elementary symmetric polynomial (which is essentially what Goldston-Pintz-Yildirim used), the next simplest possibility is to look for a polynomial that is a combination of the first and second symmetric power polynomials. In general, we let

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

be the jth symmetric power polynomial. We seek an appropriate test function in the space of polynomials spanned by those of the form

$$P := \sum_{i=1}^{d} \alpha_i (1 - P_1)^{a_i} P_2^{b_i}.$$

By imposing upper bounds on the $\{a_i\}$ and $\{b_i\}$, we reduce the search space to a finite-dimensional vector space. Then $J_k^{(\ell)}(F)$ and $I_k(F)$ are both quadratic forms on this vector space, and can hence be explicitly described in terms of matrices: if F is represented by the vector v in some basis, then

$$J_k^{(\ell)}(F) = v^T M_2 v \quad \text{and} \quad I_k(F) = v^T M_1 v.$$

As the following lemma explains, finding the best test function in this finite-dimensional space is then a matter of linear algebra.

Lemma 3.5.3. *Let M_1 and M_2 be real, symmetric, positive-definite matrices. Then*

$$\frac{v^T M_2 v}{v^T M_1 v}$$

is maximized when v is an eigenvector of $M_1^{-1} M_2$ corresponding to its largest eigenvalue, and the maximum is that eigenvalue.

Proof. This is a consequence of the standard linear algebra fact that if T is a non-zero, symmetric operator on a finite-dimensional inner product space, then

$$\max_{||v||=1} ||Tv||$$

is equal to the largest eigenvalue of T, and is maximized when v is the appropriate scalar multiple of a corresponding eigenvector.

Since M_1 is real, symmetric, and positive-definite, we may interpret $\langle w, v\rangle = w^T M_1 v$ as an inner product on $\mathbb{R}^n$. In these terms,

$$v^T M_2 v = v^T M_1 (M_1^{-1} M_2) v = ||M_1^{-1} M_2 v||$$

and applying the previous paragraph to $T := M_1^{-1} M_2$ gives the result. □

In particular, in our situation the entries of the matrices M_1 and M_2 are, in principle, a routine computation to compute. We now sketch how these calculations go. A more formal treatment is given by Maynard in [12].

We are ultimately interested in computing $I_k(F)$ and $J_k(F)$, and with F as in (14) this is

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

where $P = \sum \alpha_i (1 - P_1)^{a_i} P_2^{b_i}$. We build up some formulas to handle these types of integrals. The starting point is the beta identity

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

Lemma 3.5.4. *We have the identity*

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

Proof. Consider the innermost integral with respect to dt_1:

$$\int_0^{1-\sum_2^k t_i} \left(1 - \sum_{i=1}^k t_i\right)^a \prod_{i=1}^k t_i^{b_i}$$

Make a change of variables $u = \frac{t_1}{1 - \sum_{i=2}^k t_i}$, so that this integral becomes

$$\prod_{i=2}^k t_i^{b_i} \left(1 - \sum_{i=2}^k t_i\right)^{a+b_1+1} \int_0^1 (1-u)^a u^{b_1} \, du = \frac{a! b_1!}{(a + b_1 + 1)!} \prod_{i=2}^k t_i^{b_i} \left(1 - \sum_{i=2}^k t_i\right)^{a+b_1+1}.$$

What remains is an integral of the same form over $\mathcal{R}_{k-1}$, so we are done by induction. □

Lemma 3.5.5. *If $P_j(t_1, \dots, t_k)$ denotes the j^{th} symmetric power polynomial in $t_1, \dots, t_k$, then*

$$\int_{\mathcal{R}_k} (1 - P_1)^a P_2^b \, dt_1 \dots dt_k = \frac{a!}{k + a + 2b!} \sum_{b_1 + \dots + b_k = b} \frac{b!}{b_1! \dots b_k!} \prod_{i=1}^k (2b_i)!.$$

Proof. By the multinomial theorem,

$$P_2^b = \sum_{b_1+\ldots+b_k=b} \frac{b!}{b_1!\ldots b_k!} \prod_{i=1}^{k} t^{2b_i}.$$

Therefore, using Lemma 3.5.4

$$\begin{aligned}
\int_{\mathcal{R}_k} (1-P_1)^a P_2^b \, dt_1 \ldots dt_k &= \int_{\mathcal{R}_k} (1-P_1)^a \sum_{b_1+\ldots+b_k=b} \frac{b!}{b_1!\ldots b_k!} \prod_{i=1}^{k} t^{2b_i} \, dt_1 \ldots dt_k \\
&= \sum_{b_1+\ldots+b_k=b} \frac{b!}{b_1!\ldots b_k!} \int_{\mathcal{R}_k} (1-P_1)^a \prod_{i=1}^{k} t^{2b_i} \, dt_1 \ldots dt_k \\
&= a! \sum_{b_1+\ldots+b_k=b} \frac{b!}{b_1!\ldots b_k!} \frac{\prod_{i=1}^k (2b_i)!}{(k+a+\sum_{i=1}^k 2b_i)!}.
\end{aligned}$$

Rearranging the factors leads to the assertion of the Lemma. □

Now, we describe how to express $I_k(F)$, where F is defined by (14) with

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

as a formula in terms of the α_i, a_i, and b_i. We can simply expand

$$\begin{aligned}
I_k(F) &= \int_{\mathcal{R}_k} P^2 \, dt_1 \ldots dt_k \\
&= \int_{\mathcal{R}_k} \sum_{i,j} \alpha_i \alpha_j (1-P_1)^{a_i+a_j} P_2^{b_i+b_j} \, dt_1 \ldots dt_k.
\end{aligned}$$

This is a linear combination of integrals, all of which are described by Lemma 3.5.5. The explicit formulas are spelled out in detail in [12], Lemma 7.1.

Proposition 3.5.6. *Let $k \in \mathbb{N}$ and M_k be as in (3.5.1). Then*

1. $M_5 > 2$.
2. $M_{105} > 4$.

Proof. We simply need to produce functions that witness the above bounds.

For the first one, take $k = 5$ and

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

With this choice, we compute

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

For the second one, let V be the vector space spanned by polynomials of the form $(1-P_1)^b P_2^c$ with $b + 2c \leq 11$. This is has dimension 42, and with $k = 105$ we use the explicit formulas obtained by the method described above to calculate the entries of the 42×42 matrices M_1 and M_2 corresponding to the quadratic forms $J_k^{(\ell)}(F)$ and $I_k(F)$. We then find by explicit computation that the largest eigenvalues of $M_1^{-1}M_2$ is

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

so $M_{105} > 4$ by Lemma 3.5.3. □

Corollary 3.5.7. *$DHL[105,2]$ is true. Under the Elliott-Halberstam conjecture, $DHL[5,2]$ is true.*

Proof. The claims follow from Proposition 3.5.6 and Corollary 3.5.2 with $\theta = \frac{1}{2} - \epsilon$ and $\theta = 1 - \epsilon$, respectively, for sufficiently small ϵ. □

Corollary 3.5.8. *We have*

$$\liminf_{n\to\infty} p_{n+1} - p_n \leq 600.$$

Under the Elliott-Halberstam conjecture, we have

$$\liminf_{n\to\infty} p_{n+1} - p_n \leq 12.$$

Proof. By Corollary 3.5.7, it suffices to produce an admissible 105-tuple with diameter 600 and an admissible 5-tuple with diameter 12. These were found by computer search; the former is ([12], p.6) {0, 10, 12, 24, 28, 30, 34, 42, 48, 52, 54, 64, 70, 72, 78, 82, 90, 94, 100, 112, 114, 118, 120, 124, 132, 138, 148, 154, 168, 174, 178, 180, 184, 190, 192, 202, 204, 208, 220, 222, 232, 234, 250, 252, 258, 262, 264, 268, 280, 288, 294, 300, 310, 322, 324, 328, 330, 334, 342, 352, 358, 360, 364, 372, 378, 384, 390, 394, 400, 402, 408, 412, 418, 420, 430, 432, 442, 444, 450, 454, 462, 468, 472, 478, 484, 490, 492, 498, 504, 510, 528, 532, 534, 538, 544, 558, 562, 570, 574, 580, 582, 588, 594, 598, 600} and the latter is $\{0, 2, 6, 8, 12\}$. □

3.6 WEIGHTS FOR LARGE k

We now seek to obtain a lower bound on M_k that increases to ∞ as $k \to \infty$. By Lemma 3.1.3, this would imply that for *any* $m > 0$, $DHL[k, m]$ is true for sufficiently large k. In particular, we would be able to deduce that

$$\liminf_{n\to\infty} p_{n+m} - p_n < \infty \text{ for any } m > 0.$$

We follow Tao's argument (from the first Polymath 8b thread at [16]), which is essentially a probabilistic rephrasing of Maynard's.

We can obtain a lower bound by specializing to a specific symmetric function F, so that

$$M_k \geq k \frac{J_k^{(k)}(F)}{I_k(F)}.$$

We choose to define F as

$$F(t_1, \ldots, t_k) = \mathbb{1}_{\mathcal{R}_k} \prod_{i=1}^{k} k^{1/2} g(kt_i)$$

where $g : [0, \infty) \to \mathbb{R}$ is a function supported on $[0, T]$ (for some parameter T that will be optimized later) and satisfying

$$\int_0^\infty g(t)^2 \, dt = 1.$$

We think of $g(t)^2$ as defining a probability density for a random variable X. (Again, we technically to approximate F by a sequence of smooth functions to deduce the bound on M_k.)

Now, we massage the quantity $\frac{J_k^{(k)}(F)}{I_k(F)}$ into a more friendly form by bounding the denominator above and the numerator below. We can bound $I_k(F)$ above by removing the cutoff factor in F, and extending the integral from the simplex $\mathcal{R}_k$ to all of $[0,\infty)^k$.

$$I_k(F) = \int_{\mathcal{R}_k} F(t)^2\, dt_1 \dots dt_k \le \prod_{i=1}^{k} \int_0^\infty g(kt_i)^2 k\, dt_i = 1.$$

Therefore, we can conclude that

$$M_k \ge kJ_k^{(k)}(F).$$

We bound $J_k^{(k)}(F)$ below by restricting the integral to the region $t_1 + \dots + t_{k-1} \le 1 - \frac{T}{k}$. When this is satisfied, t_k can take any value between 0 and $\frac{T}{k}$, so

$$M_k \ge \left(\int_0^T g(t)\, dt\right)^2 \int_{t_1+\dots+t_{k-1} \le 1-\frac{T}{k}} \left(\prod_{i=1}^{k-1} kg(kt_i)^2\right) dt_1 \dots dt_{k-1}$$

The function $\prod_{i=1}^{k-1} kg(kt_i)^2$ can be interpreted as the joint density for $\frac{1}{k}(X_0, \dots, X_{k-1})$ where the X_i are independent and identically distributed (i.i.d.) copies of X. Under this probabilistic interpretation, we can write

$$M_k \ge \left(\int_0^T g(t)\, dt\right)^2 \text{Prob}(X_1 + \dots + X_{k-1} \le k - T).$$

To get a reasonable lower bound on this probability (recalling that we are interested in the large k limit), we should choose the random variable X to have mean $\mu < \frac{k-T}{k-1}$. We can then use concentration inequalities to show that there is a high probability that $X_1 + \dots + X_{k-1}$ does not significantly exceed its mean. The concentration inequality we use here is Chebyshev's, although we will see later that we can do a little better with Hoeffding's inequality.

Theorem 3.6.1 (Chebyshev). *If X is a random variable with mean μ and variance σ^2, then*

$$\text{Prob}(|X - \mu| > \alpha\sigma) \le \frac{1}{\alpha^2}.$$

By the support of the probability distribution, we can crudely bound the variance as

$$\sigma = E(X^2) \le TE(X) \le T\mu.$$

Since $X_1, \dots, X_{k-1}$ are independent and identically distributed as X,

$$\mathbb{E}(X_1 + \dots + X_{k-1}) = (k-1)\mu$$

and

$$\text{Var}(X_1 + \dots + X_{k-1}) = (k-1)\sigma^2 \le (k-1)T\mu.$$

Applying Chebyshev, we deduce that

$$\begin{aligned}
\text{Prob}(\sum_{i=1}^{k-1} X_i \leq k-T) &= 1 - \text{Prob}(\sum_{i=1}^{k-1} X_i \geq k-T) \\
&= 1 - \text{Prob}(\sum_{i=1}^{k-1} X_i - (k-1)\mu \geq k - T - (k-1)\mu) \\
&\geq 1 - \frac{(k-1)\sigma^2}{(k-T-(k-1)\mu)^2} \\
&\geq 1 - \frac{(k-1)T\mu}{(k-T-(k-1)\mu)^2}.
\end{aligned}$$

Just to make the final expression nicer, we can bound $k-1$ by k and μ by 1 to conclude that

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

To summarize, we have shown:

Lemma 3.6.2. *If $g : [0,\infty) \to \mathbb{R}$ is a function supported in $[0,T]$ and satisfying:*

$$\int_0^\infty g(t)^2\, dt = 1$$

and

$$\int_0^\infty t g(t)^2\, dt = \mu,$$

with $(k-1)\mu < k-T$, then

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

It remains to choose an appropriate function g. The optimal choice is determined by the constrained maximization problem

$$\max_g \left(\int_0^\infty g(t)\right)^2 \text{ such that } \int_0^\infty g(t)^2\, dt = 1 \text{ and } \int_0^\infty t g(t)^2\, dt = \mu.$$

To guess the right form of the function, we use Lagrange multipliers. We want to find a stationary value for

$$\int_0^T g(t)\, dt - \alpha\left(\int_0^T g(t)^2\, dt - 1\right) - \beta\left(\int_0^T t g(t)^2\, dt - \mu\right).$$

"Differentiating" with respect to $g(t)$ yields

$$1 - 2\alpha g(t) - 2\beta t g(t) = 0.$$

This suggests the choice

$$g(t) = \begin{cases} \frac{c}{1+At} & t \in [0,T], \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

Proposition 3.6.3. $M_k > k - 4\log\log k - O(1)$.

Proof. With g as in (15), we set $A := \log k$ and $T := \frac{k}{\log^3 k}$, and calculate the constants relevant to the estimate in Lemma 3.6.2. First, we have

$$\int_0^T g(t)^2\,dt = \int_0^T \frac{c^2\,dt}{(1+At)^2} = \left(\frac{c^2}{A} - \frac{c^2}{A(1+AT)}\right) = \frac{c^2 T}{1+AT}.$$

Therefore, $c^2 = \frac{1+AT}{T} = \log k + O(1)$ and hence $c = \log^{1/2} k + O(\log^{-1/2} k)$.

Next, we estimate the mean:

$$\begin{aligned}\mu = \int_0^T t g(t)^2\,dt &= \int_0^T \frac{c^2 t\,dt}{(1+At)^2}\\ &= \frac{c^2}{A^2}\left(\log(1+AT) - 1 + \frac{1}{1+AT}\right)\\ &= 1 - \frac{2\log\log k}{\log k} + O\left(\frac{1}{\log k}\right).\end{aligned}$$

Putting this estimate in Lemma 3.6.2, we obtain

$$M_k \geq \left(\int_0^T g(t)\,dt\right)^2 \left(1 + O\left(\frac{1}{\log k}\right)\right).$$

Finally, we evaluate the integral with our choice of g:

$$\int_0^T \frac{c\,dt}{1+At} = c\log(1+AT) = \log^{1/2} k - \frac{2\log\log k}{\log^{1/2} k} + O(\log^{-1/2} k).$$

Squaring and substituting this back in above, we arrive at the conclusion of the proposition. □

Remark 3.6.4. By optimizing A and T more carefully, Maynard proves the slightly stronger result that $M_k > \log k - 2\log\log k - 2$.

Corollary 3.6.5. *For any positive integer m, there is an effectively computable constant C such that if $k \geq Cm^5 e^{4m}$, then $DHL[k, m]$ holds.*

Proof. According to Bombieri-Vinogradov, $EH[\theta]$ is true for any $0 < \theta < \frac{1}{2}$. Lemma 3.5.2 shows that if $M_k > 4m$ then $DHL[k, m]$ holds.

Proposition 3.6.3 guarantees that $M_k > \log k - 4\log\log k - O(1)$, so we seek k large enough so that

$$\log k - 4\log\log k - O(1) > m.$$

We can rewrite this as

$$\log\left(\frac{k}{\log^4 k}\right) > m + O(1),$$

which is equivalent to

$$\frac{k}{\log^4 k} > Ce^m.$$

The choice $k \geq Cm^5 e^{4m}$ works for all sufficiently large k.

□

Corollary 3.6.6. *For any positive integer m, there exists an effectively computable constant C such that*

$$\liminf_{n\to\infty} p_{n+m} - p_n \leq Cm^6 e^{4m}.$$

In particular, the limit is finite.

Proof. We just have to show the existence of small admissible tuples of any size k. We claim that $(p_{\pi(k)+1}, p_{\pi(k)+2}, \ldots, p_{\pi(k)+k})$ is admissible. Indeed, recall that all k-tuples are automatically admissible with respect to any prime greater than k, and no prime less than k divides any element of the tuple, since $p_{\pi(k)+1}$ is the first prime larger than k. Now observe that the Prime Number Theorem implies that the diameter is $p_{\pi(k)+k} - p_{\pi(k)+1} \ll k \log k$, and then the conclusion follows from Corollary 3.6.5.

□

3.7 TAO'S METHOD

As the reader can see, the main technical components of the Maynard-Tao sieve are Propositions 3.4.2 and 3.4.3. We have thus far followed Maynard's combinatorial approach to proving them, which goes back to Selberg's original analysis. It is worth noting that Tao analyzes the sums $S_1, S_2^{(\ell)}$ by a Fourier-analytic method, which is in the spirit of the original work of Goldston, Pintz, and Yildirim. In this section, we will sketch Tao's calculations, assuming some basic analytic theory of Fourier transforms and the Riemann zeta function, specifically that $\zeta(s)$ can be analytically continued with a simple pole at $s = 1$.

The relationship between Maynard's method and Tao's method is roughy the relationship between working in "physical space" and "Fourier space." Tao's approach has the advantage that it illuminates the nature of the integral approximation in terms of the Fourier transform, and explains the constant factors nicely. On the other hand, the Fourier method seems slightly less robust than the elementary combinatorial approach. In particular, for Zhang's work in restricting the GPY sieve to smooth moduli, it seems that one has to work in "physical" space, using the combinatorial approach, rather than Fourier space.

We consider, as before, the sums

$$S_1 := \sum_{\substack{N \leq n \leq 2N \\ n \equiv \nu_0 \pmod W}} w(n) \quad \text{and} \quad S_2^{(\ell)} := \sum_{\substack{N \leq n \leq 2N \\ n \equiv \nu_0 \pmod W}} w(n)\chi_{\mathcal{P}}(g_\ell n + h_\ell) \tag{16}$$

with choice of weights

$$w(n) := \left(\sum_{\substack{d_1,\ldots,d_k \\ d_i | g_i n + h_i \forall i}} \left(\prod_{i=1}^{k} \mu(d_i) \right) f\left(\frac{\log d_1}{\log R}, \ldots, \frac{\log d_k}{\log R} \right) \right)^2$$

where $f : [0, \infty)^k \to \mathbb{R}$ is some smooth function supported on the simplex

$$\mathcal{R}_k := \{(t_1, \ldots, t_k) \in \mathbb{R}^k : t_1 + \ldots + t_k \leq 1\}.$$

We want to derive some smooth asymptotics for these sums. Let's focus on S_1 first. As before, we can write it as a negligible error term plus the main term

$$\frac{N}{W} \sum_{\substack{d_1,\ldots,d_k \\ e_1,\ldots,e_k \\ (d_i,e_j)=1 \forall i,j \\ (d_ie_i,W)=1 \forall i}} \left(\prod_{i=1}^{k} \frac{\mu(d_i)\mu(e_i)}{[d_i,e_i]} \right) f\left(\frac{\log d_1}{\log R}, \ldots, \frac{\log d_k}{\log R} \right) f\left(\frac{\log e_1}{\log R}, \ldots, \frac{\log e_k}{\log R} \right). \tag{17}$$

What we want to do is essentially to take the Fourier transform of this expression. By performing a change of variables in the inverse Fourier transform of f, we can write

$$f(u_1, \ldots, u_k) = \int_{\mathbb{R}^k} e^{-\sum(1+is_j)u_j} \eta(s_1, \ldots, s_k)\, ds_1 \ldots ds_k. \tag{18}$$

Here η is essentially the Fourier transform of f, but composed with a translation. Technically speaking, this shift will be useful later to move the argument of the Riemann zeta function just right of the line $s = 1$ where it as a pole. The motivation is that the technique traditionally used to to evaluate sums like (16) is the inverse Laplace transform (this is what GPY use), and the inverse Laplace transform of $f(u)$ is essentially the Fourier transform of $e^u f(u)$, which gives the shifted formula above.

Using this, we can re-express the main term of S_1 (17) as

$$\frac{N}{W} \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} H(\vec{s}, \vec{t}) \eta(s_1, \ldots, s_k) \eta(t_1, \ldots, t_k) ds_1 \ldots ds_k dt_1 \ldots dt_k \tag{19}$$

where

$$H(\vec{s}, \vec{t}) = \sum_{\substack{d_1,\ldots,d_k \\ e_1,\ldots,e_k \\ (d_i,e_j)=1 \forall i,j \\ (d_ie_i,W)=1 \forall i}} \prod_{j=1}^{k} \frac{\mu(d_j)\mu(e_j)}{[d_i,e_i] d_j^{(1+is_j)/\log R} e_j^{(1+it_j)/\log R}}.$$

Since η is rapidly decreasing (it is a Schwarz function), we expect the integral (19) to be concentrated close to the origin, in which region we can use the asymptotics for $\zeta(s)$ near $s = 1$. To see this, note that we can factor H as an Euler product:

$$H(\vec{s}, \vec{t}) = \prod_{p>w} \left(1 - \sum_{j=1}^{k} \left(p^{-1-\frac{1+is_j}{\log R}} + p^{-1-\frac{1+it_j}{\log R}} - p^{-1-\frac{1+is_j}{\log R} - \frac{1+is_j}{\log R}} \right) \right). \tag{20}$$

We now recall that $\zeta(s) \sim \frac{1}{s-1}$ for s near 1, and is holomorphic and bounded for $\operatorname{Re} s \gg 1$. From this we can deduce

$$|H(\vec{s}, \vec{t})| \leq \prod_{p>w} \left(1 + \frac{3}{p^{1+1/\log R}} \right) \ll \log^3 R$$

by taking logarithms and comparing with $\log \zeta(s)$. This shows that, since η is a Schwarz function, the main contribution to the integral (19) comes from region where we have (say) $|\vec{s}|, |\vec{t}| \leq \sqrt{\log R}$. In this region, we can use the asymptotic on $\zeta(s)$ near $s = 1$ again to approximate the Euler product (20) as

$$H(\vec{s},\vec{t}) = (1+o(1))\left(\frac{W}{\varphi(W)}\right)^k \prod_{j=1}^{k} \frac{\zeta\left(1+\frac{1+is_j}{\log R}+\frac{1+it_j}{\log R}\right)}{\zeta\left(1+\frac{1+is_j}{\log R}\right)\zeta\left(1+\frac{1+it_j}{\log R}\right)} + O(1)$$
$$= (1+o(1))\left(\frac{W}{\varphi(W)}\right)^k \frac{1}{\log^k R}\prod_{j=1}^{k}\frac{(1+is_j)(1+it_j)}{1+is_j+1+it_j}.$$

Using the fact that $\eta(\vec{s})$ is a Schwartz function again, we can absorb the $o(1)$ term into the existing error term, and to restore the integral over all of $\mathbb{R}^k$, so that the main term of S_1 becomes

$$\frac{N}{W}\left(\frac{W}{\varphi(W)}\right)^k \frac{1}{\log^k R}\int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\prod_{i=1}^{k}\frac{(1+is_j)(1+it_j)}{1+is_j+1+it_j}\eta(\vec{s})\eta(\vec{t})ds_1\dots ds_k dt_1\dots dt_k.$$

In order to evaluate this, we differentiate (18) with respect to each variable. Writing $f_{1,\dots,k} := \frac{\partial}{\partial x_k}\cdots\frac{\partial}{\partial x_k}f$, (18) gives

$$f_{1,\dots,k}(u_1,\dots,u_k) = (-1)^k\int_{\mathbb{R}^k}\prod_{j=1}^{k}(1+is_j)e^{-(1+is_j)u_j}\eta(s_1,\dots,s_k)ds_1\dots ds_k.$$

Hence

$$f_{1,\dots,k}(u_1,\dots,u_k)^2 = \int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\prod_{j=1}^{k}(1+is_j)(1+it_j)e^{-(1+is_j)u_j}e^{-(1+it_j)u_j}\eta(\vec{s})\eta(\vec{t})d\vec{s}d\vec{t}.$$

Now integrating over $u_1,\dots,u_k$ and applying Fubini's theorem, we obtain

$$\int_{\mathbb{R}^k} f_{1,\dots,k}(u_1,\dots,u_k)^2 d\vec{u} = \int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\frac{(1+is_j)(1+it_j)}{1+is_j+1+it_j}\eta(\vec{s})\eta(\vec{t})d\vec{s}d\vec{t},$$

and substituting this above shows that

$$S_1 \sim \frac{N}{W}\left(\frac{W}{\varphi(W)}\right)^k\frac{1}{\log^k R}\int_{\mathbb{R}^k} f_{1,\dots,k}(u_1,\dots,u_k)^2 du_1\dots du_k.$$

Up to the constant factors, this is the same as what we found using Maynard's method with $F = f_{1,\dots,k}$. The argument for $S_2^{(\ell)}$ is similar. Without loss of generality, we consider only $S_2^{(1)}$. After using the prime number theorem for arithmetic progressions and $EH[\theta]$ to bound the error, the main term of $S_2^{(1)}$ becomes a sum of the form

$$\frac{N}{\varphi(W)\log N}\sum_{\substack{d_2,\dots,d_k\\ e_2,\dots,e_k\\ (d_i,e_j)=1\forall i,j\\ (d_ie_i,W)=1\forall i}}\prod_{j=2}^{k}\mu(d_j)\mu(e_j)f_0\left(\frac{\log d_2}{\log R},\dots,\frac{\log d_k}{\log R}\right)f_0\left(\frac{\log e_2}{\log R},\dots,\frac{\log e_k}{\log R}\right)$$

where $f_0(t_1,\dots,t_{k-1}) = f(t_1,\dots,t_{k-1},0)$. This sum has the same form as the one that we just evaluated, and the same arguments show that

$$S_2^{(1)} \sim \frac{N}{\varphi(W)\log N}\left(\frac{W}{\varphi(W)}\right)^{k-1}\frac{1}{\log^{k-1}R}\int_{\mathbb{R}^{k-1}} f_{1,\dots,k-1}(u_1,\dots,u_{k-1},0)^2 du_2\dots du_k.$$

4 THE POLYMATH 8B PROJECT

After Maynard and Tao announced their new, more efficient proof of bounded gaps, Terence Tao launched a new massively collaborative project, called Polymath 8b, devoted to sharpening Maynard's results with the goal of decreasing the bounds on small prime gaps as much as possible. The project is still ongoing, but it has just entered the final stage of writing up results.

The Polymath Wiki [16] contains the current world records, as well as links to all of the discussions from which our material is drawn. The Polymath 8b Project is considered to be the content of the (currently ten) pages on Terence Tao's blog linked from [16]. Given that it consists of comments scattered across ten different online threads, it is difficult to present a comprehensive, coherent account of the story. In this chapter, we will focus only on describing the ideas leading to the current state of the art. Unfortunately, this means that we will not include some instructive parts of the Polymath, such as the heuristic discussions and unsuccessful ideas. We will also omit the more technical analysis arguments, and instead focus on explaining the main ideas. Implementation issues were an important concern in obtaining explicit bounds, but we do not discuss them here. In the spirit of keeping the Polymath a purely collaborative forum, we choose not to assign credit for ideas to the individuals who first proposed them.

Recall that Maynard showed the following results.

1. $DHL[105, 2]$ is true. In particular, we have
$$\liminf p_{n+1} - p_n \leq 600.$$
2. Assuming the Elliott-Halberstam conjecture, $DHL[5, 2]$ is true. In particular, under EH we have
$$\liminf p_{n+1} - p_n \leq 12.$$
3. $M_k \geq \log k - 2 \log\log k - 2$, so for any m, $DHL[k, m]$ is true for all sufficiently large k. In particular, there exists some effectively computable constant C such that
$$\liminf p_{n+m} - p_n \leq Cm^3 e^{4m}.$$

Polymath 8b made the following improvements.

1. $DHL[50, 2]$ is true. In particular, we have
$$\liminf p_{n+1} - p_n \leq 246.$$
2. Assuming the *generalized* Elliott-Halberstam conjecture, $DHL[3, 2]$ is true. In particular, we have
$$\liminf p_{n+1} - p_n \leq 6.$$
(No improvements were made using the ordinary Elliott-Halberstam conjecture.)

3. We have the bounds $\log k - \log\log\log k - O(1) \le M_k \le \frac{k}{k-1}\log k$, hence

$$\liminf p_{n+m} - p_n \le e^{3.817m}.$$

   Actually, slightly tighter asymptotics were eventually obtained, but we will not explain them here.

Our goal is to indicate how these improvements are achieved.

4.1 ASYMPTOTICS FOR M_k

Recall that we defined the quantity (3.5.1)

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

which dictates the size of the intervals in which we can find primes. Specifically, under $EH[\theta]$, $m \ge \lfloor \frac{\theta M_k}{2} \rfloor$ implies $DHL[k,m]$.

We showed how to obtain a lower bound for M_k by considering the ratio for a specific choice of smooth function F. In this section, we establish tighter lower bounds as well as *upper bounds* on M_k. This indicates a limit to the size of gaps that we deduce using the method of Maynard and Tao.

4.1.1 *Upper Bound*

Let us consider a toy problem first, corresponding to the one-variable case $k = 1$. We essentially already analyzed this in §3.6. It boils down to the constrained optimization problem

$$\max_F \left(\int_0^T F(t)\,dt \right)^2 \text{ subject to } \int_0^T F(t)^2\,dt = 1 \text{ and } \int_0^T tF(t)^2\,dt \le 1. \tag{21}$$

We used a calculus of variations argument to suggest that the best choice would be a function of form $\frac{1}{1+At}$. Therefore, it is natural to apply the Cauchy-Schwarz inequality as

$$\begin{aligned} \left(\int_0^T F(t)\,dt \right)^2 &\le \int_0^T \frac{dt}{1+At} \int_0^T (1+At)F(t)^2\,dt \\ &= \frac{\log(1+AT)}{A} \int_0^T (1+At)F(t)^2\,dt \end{aligned}$$

since we expect this to be essentially an equality for the optimal F. Using the conditions on F, we find that

$$\left(\int_0^T F(t)\,dt \right)^2 \le \frac{(1+A)\log(1+AT)}{A} \int_0^T F(t)^2\,dt$$

so that (21) is at most $\left(1+\frac{1}{A}\right)\log(1+AT)$. We want to choose the parameter A so as to minimize this expression. It is essentially $\left(1+\frac{1}{A}\right)(\log A + \log T)$, and T will be large compared to A, so the largest term is $\log T$ and the next largest terms are $\frac{\log T}{A}$ and $\log A$. We optimize by balancing these secondary terms, which leads to the choice $A \approx \log T$ (up to factors which are at most $\log\log T$).

Now, let's try to generalize this discussion to the problem of bounding M_k. On the ℓ^{th} integral, we use a Cauchy-Schwarz comparison with the function $\frac{1}{1+Akt_\ell}$ (the extra factor of k is to be consistent with earlier normalizations). Setting $s_\ell = t_1 + \ldots + \widehat{t_\ell} + \ldots + t_k$, we have

$$\left(\int_0^{1-s_\ell} F(\vec{t})\, dt_\ell\right)^2 \leq \int_0^{1-s_\ell} \frac{dt_\ell}{1+Akt_\ell} \int_0^{1-s_\ell} F(\vec{t})^2(1+Akt_\ell)\, dt_\ell$$
$$\leq \frac{\log(1+Ak)}{Ak} \int_0^{1-s_\ell} F(\vec{t})^2(1+Akt_\ell)\, dt_\ell.$$

Integrating over the rest of the variables, we conclude that

$$J_k^{(\ell)} \leq \frac{\log(1+Ak)}{Ak} \int_{\mathcal{R}_k} F(t_1, \ldots, t_k)^2(1+Akt_\ell)\, dt_1 \ldots dt_k.$$

Summing over ℓ and using the trivial bound $\sum t_\ell \leq 1$, we find that

$$M_k \leq \frac{\log(1+Ak)}{Ak}(k+kA) = \log(1+Ak)\left(1+\frac{1}{A}\right).$$

Taking a hint from our analysis of the toy problem, we set $A = \log k$ to prove that

$$M_k \leq \log(1+k\log k)\left(1+\frac{1}{\log k}\right).$$

Since $\log(1+k\log k) \approx \log k + \log\log k$, this shows:

Proposition 4.1.1. *$M_k \leq \log k + \log\log k + O(1)$.*

In fact, we can do a little bit better by essentially the same argument, with a slightly different choice of function in the Cauchy-Schwarz step. Note that there is a little inefficiency in the bounds on the Cauchy-Schwarz factors, and the improvement essentially comes from removing this inefficiency.

Proposition 4.1.2. *$M_k \leq \frac{k}{k-1}\log k$.*

Proof. We apply Cauchy-Schwarz to deduce that

$$\left(\int_0^{1-s_\ell} F(\vec{t})\, dt_\ell\right)^2 \leq \int_0^{1-s_\ell} \frac{dt_\ell}{1-s_\ell+(k-1)t_\ell} \int_0^{1-s_\ell} F(\vec{t})^2(1-s_\ell+(k-1)t_\ell)\, dt_\ell$$

where $s_\ell = t_1 + \ldots + \widehat{t_\ell} + \ldots + t_k$ as above. Note that

$$\int_0^{1-s_\ell} \frac{dt_\ell}{1-s_\ell+(k-1)t_\ell} = \frac{\log(k-1)(1-s_\ell) - \log(1-s_\ell)}{k-1} = \frac{\log k}{k-1}.$$

Substituting this above, and integrating over the rest of $\mathcal{R}_k$, we obtain

$$J_k^{(\ell)} \leq \frac{\log k}{k-1} \int_{\mathcal{R}_k} F(t_1, \ldots, t_k)^2(1-s_\ell+(k-1)t_\ell)\, dt_\ell.$$

Summing over k then gives the result. □

This bound is quite tight for small k. Maynard computes $M_5 \geq 2.001162$ to obtain the conditional prime gap bound of 12; the upper bound gives $M_5 \leq 2.0011797$. For $k = 4$ it gives $M_4 \leq 1.848$, while we know from specialization that $M_4 \geq 1.845$. (Recall that we need $M_k > 2$ to conclude $DHL(k, 2)$ under the Elliott-Halberstam conjecture.) Therefore, we need some ideas to show $DHL[4, 2]$ using Maynard's sieve.

As k grows, the bound becomes looser. We know $M_{59} \geq 3.95608$, but the upper bound gives only $M_{59} \leq 4.148$, so it is unclear if $M_{59} > 4$ (recall that this is what is required to conclude $DHL(k, 2)$ unconditionally, under Bombieri-Vinogradov). The smallest k for which the upper bound gives $M_k < 4$ is $k = 50$, with $M_{50} \leq 3.99\ldots$. This sets a limit on the unconditional prime gap that can be obtained by the Maynard-Tao sieve (although we will mention a few small modifications that lead to improvements; in particular, $DHL[50, 2]$ is true under Bombieri-Vinogradov). It was eventually shown that $M_{54} > 4$, which seems to be the limit.

4.1.2 *Lower bound*

The improved lower bounds result from a more careful analysis of the argument already given. The main new ingredients are more careful parameter tuning and replacing Chebyshev's bound by Hoeffding's inequality, which is another concentration of measure inequality.

Proposition 4.1.3. $M_k \geq \log k - \log\log\log k + O(1)$.

Proof. As before, we specialize a function F of the form

$$F(t_1, \ldots, t_k) = \mathbb{1}_{\mathcal{R}_k} \prod_{i=1}^{k} k^{1/2} g(kt_i)$$

where $g : [0, \infty) \to \mathbb{R}$ is a smooth function. Let

$$m_1 = \int_0^\infty g(t)\, dt \qquad \text{and} \qquad m_2 = \int_0^\infty g(t)^2\, dt.$$

Then $\frac{g(t)^2}{m_2}$ is the density function for a random variable X, and our arguments in §3.6 show that

$$M_k \geq \frac{m_1^2}{m_2} \operatorname{Prob}(X_1 + \ldots + X_{k-1} \leq k - T) \tag{22}$$

where the X_i are i.i.d. copies of X. We choose

$$g(t) = \frac{1}{1 + At} \mathbb{1}_{[0,T]}$$

where $A = \log k$ and $T = \epsilon \frac{k}{A}$; here $0 < \epsilon \leq 1$ is a tuning parameter that we will set later. With this choice, we compute

$$m_1 = \log(1 + AT) = \log(1 + \epsilon k) \qquad \text{and} \qquad m_2 = \frac{T}{1 + AT} \leq \frac{1}{A} = \log k.$$

X has mean

$$\mu = \frac{1}{m_2} \int_0^T t g(t)^2 = \frac{1}{m_2 A^2} \left(\log(1 + AT) - 1 + \frac{1}{1 + AT} \right),$$

which we can estimate as

$$\mu = \left(1 + \frac{1}{\epsilon k}\right) \frac{1}{\log k} \left(\log(1+\epsilon k) - 1 + \frac{1}{1+\epsilon k}\right)$$
$$\leq 1 - \frac{1}{\log k} + O\left(\frac{\log k}{\epsilon k}\right).$$

Since the X_i are independent, this shows that the mean of $X_1 + \ldots + X_{k-1}$ is

$$(k-1)\mu \leq k - \frac{k}{\log k} + O\left(\frac{\log k}{\epsilon}\right).$$

To prepare for the concentration inequality step, we use the preceding estimates to write

$$\begin{aligned}\text{Prob}(\sum_{i=1}^{k-1} X_i \leq k - T) &= 1 - \text{Prob}(\sum_{i=1}^{k-1} X_i > k - T) \\ &\geq 1 - \text{Prob}(\sum_{i=1}^{k-1} X_i - (k-1)\mu \geq \frac{(1+\epsilon)k}{\log k} + O(\frac{\log k}{\epsilon})).\end{aligned}$$

We can then apply Hoeffding's inequality, noting that X_i is supported on $[0, T]$, to deduce that

$$\text{Prob}(X_1 + \ldots + X_{k-1} \leq k - T) \geq 1 - \exp\left(-\frac{c}{\epsilon^2}\right).$$

A good choice of ϵ is obtained by balancing the error terms in (22), and this gives $\epsilon = (\log\log k)^{1/2}$ (up to factors involving more logarithms). With this choice,

$$\text{Prob}(X_1 + \ldots + X_{k-1} \leq k - T) \geq 1 - O\left(\frac{1}{\log k}\right).$$

If we substitute this into (22) along with the estimates above, we obtain

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

□

4.2 ENLARGING THE SIEVE SUPPORT

The work of Maynard and Tao reduces the problem of finding primes in bounded gaps to a variational problem of studying the quantity

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

Indeed, recall that if $M_k > 4$ then $DHL[k,2]$ holds unconditionally, while if $M_k > 2$ then $DHL[k,2]$ holds conditional on the Elliott-Halberstam conjecture.

The main technical innovation of Polymath 8b, which leads to the improved numerical bounds on prime gaps, is in finding more general variational problems that can be used to deduce $DHL[k,2]$. Broadly speaking, the goal is to expand the space of functions allowed. Currently, the restriction to functions having support contained in $\mathcal{R}_k$ comes from the fact that the size of the support is related to the size of the moduli that occur in the sums $S_1, S_2^{(\ell)}$, which dictates the scale of the error terms. Therefore, the fundamental issue is to obtain better control of the error terms.

If we examine Maynard's proof more carefully, we see that the asymptotic for S_1 holds more generally in the dilated simplex $\frac{1}{\theta}\mathcal{R}_k$. Indeed, we restricted the sieve support to $R \approx N^{\theta/2}$ in order to control the error terms for primes in arithmetic progressions in $S_2^{(\ell)}$; but for S_1 analogue of these error terms for all integers can be trivially controlled all the way up to N. Also, for $J_k^{(\ell)}$ we really only need $t_1 + \ldots + t_{\ell-1} + t_{\ell+1} + \cdots \leq 1$, since in the sum we are always considering terms with $t_\ell = 0$. (In fact, this is already remarked in Maynard's paper.) Therefore, we can actually work with the larger region

$$\mathcal{R}_k' = \bigcup_{\ell=1}^{k} \{(t_1, \ldots, t_k) \colon t_1 + \ldots + t_{\ell-1} + t_{\ell+1} + \cdots \leq 1\}.$$

Defining

$$M_k' = \sup_F \frac{\sum_{\ell=1}^k \int_{\mathcal{R}_{k-1}} \left(\int_0^{1/\theta} F(t_1, \ldots, t_k) dt_\ell\right)^2}{\int_{R_k'} F(t_1, \ldots, t_k)^2}$$

where the supremum is over F supported in $\mathcal{R}_k'$, the same argument as before shows that $M_k' > \lfloor \frac{2}{\theta} \rfloor$ implies $DHL[k, 2]$ under $EH[\theta]$.

4.2.1 *Expanding beyond the simplex*

With a little more effort and input, we can extend the support further. Let

$$\mathcal{R}_k(\theta)' = \{(t_1, \ldots, t_k) \in [0, 1/\theta]^k : t_1 + \ldots + \widehat{t_\ell} + \ldots + t_k \leq 1 \text{ for all } \ell\}.$$

Suppose R is a region whose sumset satisfies the containment

$$R + R := \{r + r' \colon r, r' \in R\} \subset 2\mathcal{R}_k(\theta)' \cup \frac{2}{\theta}\mathcal{R}_k.$$

If $F \colon \mathbb{R}^k \to \mathbb{R}$ is a smooth function supported on R, we define

$$I_k''(F) = \int_R F(t_1, \ldots, t_k)^2 \, dt_1 \ldots dt_k$$

and

$$(J_k^{(\ell)})''(F) = \int_{\mathcal{R}_{k-1}} \left(\int_0^{1/\theta} F(t_1, \ldots, t_k) \, dt_\ell\right)^2 dt_1 \ldots dt_k.$$

Let

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

where the supremum is over smooth functions F supported in R and satisfying

$$\int F(t_1, \ldots, t_k) \, dt_\ell = 0 \text{ whenever } t_1 + \ldots + \widehat{t_\ell} + \ldots + t_k > 1.$$

Finally, we require an equidistribution result like the Elliott-Halberstam condition. We let $GEH[\theta]$ denote the assumption that the *Generalized Elliott-Halberstam condition* holds for parameter $\theta \in (0, 1)$. Instead of stating this condition precisely right now, we will jump into the proof and see what kind of additional assumption is needed. For now, we just mention that the Elliott-Halberstam condition says that the prime numbers are evenly distributed in residue classes; the generalized Elliott-Halberstam condition says that a more general class of arithmetic functions are evenly distributed in residue classes.

The conclusion is that under $GEH[\theta]$, if $M_k'' > \frac{2m}{\theta}$ then $DHL[k, m+1]$ holds.

Example. Some candidates for R that Polymath 8b used include the prism

$$\{(t_1, \dots, t_k) \in [0, 1/\theta]^k \colon t_1 + \ldots + t_{k-1} \le 1\}$$

and the symmetric region

$$\{(t_1, \dots, t_k) \in [0, 1/\theta]^k \colon t_1 + \ldots + t_k \le \frac{k}{k-1}.$$

It is clear that both strictly contain the standard unit simplex $\mathcal{R}_k$.

Proof Sketch. As usual, we follow the framework set up by Goldston, Pintz, and Yildirim, which compares S_1 and $\sum_{\ell=1}^{(k)} S_2^{(\ell)}$ defined by

$$S_1 = \sum_{\substack{n \in [N,2N] \\ n \equiv \nu_0 \pmod W}} w(n)$$

$$S_2^{(\ell)} = \sum_{\substack{n \in [N,2N] \\ n \equiv \nu_0 \pmod W}} w(n) \chi_{\mathcal{P}}(g_\ell n + h_\ell).$$

The weights $w(n)$ are squares of divisors sums in terms of Selberg weights $\lambda_{d_1,\ldots,d_k} = \mu(d_1, \dots, d_k) F(\frac{\log d_1}{\log R}, \dots, \frac{\log d_k}{\log R})$, where $R = N^{\frac{\theta}{2} - \epsilon}$. The condition on the support of F are imposed in order to control the error terms involved in counting integers in arithmetic progression (in S_1) or primes in arithmetic progressions (in $S_2^{(\ell)}$).

Under $EH[\theta]$, we can only control the error coming from counting primes in arithmetic progressions for moduli up to $N^\theta \approx R^2$. In $S_2^{(\ell)}$ the moduli involved are (up to fixed constants) $\prod_{i \neq \ell} d_i \prod_{i \neq \ell} e_i$, so we can handle this as long as $t_1 + \ldots + \widehat{t_\ell} + \ldots + t_k \le 1$. The vanishing marginals condition ensures that there is negligible contribution in regions where this condition is not satisfied. So essentially the same arguments as before show that $(J_k^{(\ell)})''(F)$ is asymptotic to $S_2^{(\ell)}$ with F supported in R.

The problem of counting *integers* in arithmetic progressions is much easier, and we can control the ensuing error terms for moduli all the way up to $N \approx R^{2/\theta}$. Therefore, the asymptotics will be essentially the same for F supported in $\frac{2}{\theta}\mathcal{R}_k$. However, we are also allowing the support to lie in $2\mathcal{R}_k(\theta)'$, and this involves moduli larger than N, which present a problem. This is where we require some sort of generalized version of the Elliott-Halberstam condition.

To deal with the large moduli, we use a technical fact F can be approximated by a linear combination of "pure tensors" of the form $F_1(t_1) \ldots F_k(t_k)$ which are all supported in small cubes, such that the quantities $I_k''(F)$ and $(J_k^{(\ell)})''(F)$ are continuous in the approximation. This almost follows directly from the Stone-Weierstrass theorem, but some care has to be taken to ensure that all the support conditions are satisfied; we omit the technical details.

The point is that for each pure tensor, we can factorize F as $F(t_1, \ldots, t_k) = \tilde{F}_k(t_1, \ldots, t_{k-1})F_k(t_k)$, which induces a corresponding factorization $\lambda_{d_1,\ldots,d_k} = \lambda_{d_1,\ldots,d_{k-1}}\lambda_{d_k}$. Writing

$$\widetilde{w}(n) = \left(\sum_{\substack{d_1,\ldots,d_{k-1} \\ d_i | g_i n + h_i \forall i<k}} \lambda_{d_1,\ldots,d_{k-1}} \right)^2$$

we can express

$$S_1 = \sum_{\substack{N \le n \le 2N \\ n \equiv \nu_0 \pmod W}} w(n) = \sum_{\substack{N \le n \le 2N \\ n \equiv \nu_0 \pmod W}} \widetilde{w}(n) \left(\sum_{d_i, e_i | g_i n + h_i} \lambda_{d_i}\lambda_{e_i} \right)^2 .$$

The term $\left(\sum_{d_i,e_i|g_in+h_i} \lambda_{d_i}\lambda_{e_i}\right)^2$ is itself an arithmetic function of $g_in + h_i$, which can be written as a convolution of arithmetic functions $\gamma * 1$. Expressed in this way, the sum S_1 is analogous to $S_2^{(\ell)}$ with $\tilde{w}(n)$ playing the role of $w(n)$ and $\gamma * 1$ playing the role of $\chi_{\mathcal{P}}(n)$. If we had an analogous result to $EH[\theta]$ ensuring that $\gamma * 1$ is well-distributed in residue classes for moduli up to N^θ, then we would be able to control the error terms in the region $(t_1, \ldots, t_{k-1}) \in 2\mathcal{R}'_k(\theta)$. The content of the Generalized Elliott-Halberstam conjecture is precisely to provide the necessary equidistribution results for convolutions of nice enough functions. □

Now we describe the Generalized Elliott-Halberstam conjecture in more detail.

Definition 4.2.1. For any function $\alpha : \mathbb{N} \to \mathbb{C}$ with finite support, and any primitive congruence class $a \pmod q$, we define

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

Thus $E(\alpha; q, a)$ is a measure of how evenly distributed the function α is in congruence classes mod q. Recall that the condition $EH[\theta]$ says that if $A \ge 1$ is any fixed constant, then

$$\sum_{q \le x^\theta} \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} |E(\chi_{\mathcal{P}} \mathbb{1}_{[1,x]}; q, a)| \ll \frac{x}{(\log x)^A}.$$

Remark 4.2.2. Usually, this is phrased in terms of $E(\Lambda; q, a)$ instead, where Λ is an arithmetic function that behaves like $\chi_{\mathcal{P}}$.

Definition 4.2.3. We denote by $GEH[\theta]$ the following assertion. Let M, N be fixed constants such that $x^\epsilon \le M, N \le x^{1-\epsilon}$ and $MN \sim x$. If $\{\alpha(n)\}$ and $\{\beta(n)\}$ are finitely supported sequences at scale M, N, respectively, then for any $A > 0$

$$\sum_{q \le x^\theta} \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} |E(\alpha * \beta; q, a)| \ll \frac{x}{(\log x)^A}.$$

We have not said precisely what it means for α to be at scale M. This essentially means that $|\alpha(n)| \ll \tau(n)^{O(1)} \log^{O(10} n$, but there are some more conditions. See [1] for a precise formulation.

Conjecture 4.2.4 (Generalized Elliott-Halberstam). *$GEH[\theta]$ holds for all $\theta \in (0,1)$.*

We abbreviate this conjecture by GEH. It is true, but not obvious, that $GEH[\theta]$ implies $EH[\theta]$. However, $GEH[\theta]$ is not necessarily expected to be much more difficult than $EH[\theta]$. For instance, $EH[\theta]$ is known only for $\theta \in (0, \frac{1}{2})$ (Bombieri-Vinogradov), and Motohashi verified in [14] that $GEH[\theta]$ is true for all θ in this region as well, by modifying the proof of Bombieri-Vinogradov.

Theorem 4.2.5 (Motohashi, 1976). *$GEH[\theta]$ holds for any $\theta \in (0, \frac{1}{2})$.*

4.2.2 *The ϵ-trick*

A further trick was introduced to generalize the variational problem, which has become known as the "ϵ-trick" in the Polymath 8b discussions.

The starting observation is that don't really need asymptotics for S_1 and $S_2^{(\ell)}$: it suffices to have an upper bound for S_1 and a lower bound for $S_2^{(\ell)}$. This means that we can extend the support of the function F into a range where we *cannot* precisely control the error term, as long as we can bound the error term in the right direction.

Consider the sum $S_2^{(k)}$:

$$S_2^{(k)} = \sum w(n)^2 \chi_{\mathcal{P}}(g_i n + h_i).$$

Here $w(n)$ is defined in terms of a smooth function F. In our previous setups, the support of F is such that the moduli involved in counting primes are small enough to be handled by $EH[\theta]$. More precisely, this is guaranteed by the condition that the "marginal distribution"

$$\int_0^\infty F(t_1, \ldots, t_k)\, dt_k$$

is supported in the region $\mathcal{R}_{k-1} : t_1 + \ldots + t_{k-1} \leq 1$. Suppose we want to extend this support condition to the region

$$t_1 + \ldots + t_{k-1} \leq 1 + \epsilon$$

for some small $\epsilon > 0$. The extra region introduces moduli that are too large to be controlled by $EH[\theta]$, so our asymptotics no longer hold. However, we can still hope to obtain a useful lower bound. Indeed, consider writing

$$\begin{aligned} S_2^{(k)} &= \sum \chi_{\mathcal{P}}(g_i n + h_i)(\widetilde{w}(n)^2 + w(n)^2 - \widetilde{w}(n)^2) \\ &\geq \sum \chi_{\mathcal{P}}(g_i n + h_i)(w(n)^2 - \widetilde{w}(n)^2) \end{aligned}$$

where $\widetilde{w}(n)$ is defined just as $w(n)$ in terms of a smooth function $\widetilde{F}$ which is identical to F in the region $t_1 + \ldots + t_{k-1} > 1 - \frac{\epsilon}{2}$ and vanishes for $t_1 + \ldots + t_{k-1} < 1 - \epsilon$. Then the contribution from the large moduli, corresponding to $t_1 + \ldots + t_{k-1} > 1 - \frac{\epsilon}{2}$ cancels out in the lower bound above, and the contribution from the small moduli can be controlled as before, to give an asymptotic lower bound with integral factor

$$\int_{(1-\epsilon)\mathcal{R}_{k-1}} \left(\int_0^\infty F(t_1, \ldots, t_k) dt_k \right)^2.$$

The upshot of this trick is that we allow ourselves to extend the support of the marginal $\int F(t_1, \dots, t_k)dt_k$ at the cost of decreasing the numerator of the ratio that we wish to maximize.

Let us now formalize the fruits of this discussion. Fix $0 < \epsilon < 1$. For a compactly support smooth function $F : [0, \infty)^k \to \mathbb{R}$ we define

$$J_{k,\epsilon}^{(\ell)}(F) = \int_{(1-\epsilon)\mathcal{R}_k} \left(\int_0^\infty F(t_1, \dots, t_k) dt_\ell \right)^2 dt_1 \dots dt_k.$$

and

$$I_k(F) = \int_{\mathbb{R}^k} F(t_1, \dots, t_k)^2 dt_1 \dots dt_k.$$

Define

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

where the supremum is over smooth functions F supported on a polytope R as in the previous section, and satisfying

$$\int_0^\infty F(t_1, \dots, t_k)\, dt_\ell = 0 \text{ whenever } t_1 + \dots + \widehat{t_\ell} + \dots + t_k > 1 + \epsilon.$$

Then $DHL[k, m+1]$ is true if $M_k > \frac{2m}{\theta}$.

Remark 4.2.6. A simple way to enforce the latter condition on the marginal distribution is to require F to be supported in $(1+\epsilon)\mathcal{R}_k$. This is typically how the ϵ-trick is applied in Polymath 8b.

4.3 improving the unconditional bound

One can think of establishing small prime gaps in terms of two separate steps:

1. bounds on M_k (or variants M_k', $M_k''(\theta)$, or $M_{k,\epsilon}(\theta)$) to establish $DHL[k, 2]$,
2. for a given k, computing an admissible k-tuple of small diameter.

The second problem has already been studied in the literature and essentially optimized in Polymath 8a. Several algorithms are known (see [16] for a list with references), but most of Polymath's narrow admissible tuples were obtained by the "greedy-greedy algorithm." The basic idea is as follows. We begin with a "candidate set" $[s, s+x]$. Note that any k-tuple is admissible for any prime $p > k$, so one only has to check admissibility for primes $p \leq k$. Recall that a k-tuple is admissible if it misses a residue class for each prime $p \leq k$. For each prime less than k, we sieve out a specific residue class mod p from our candidate set. At the end, if there are at least k elements left then any k of them form an admissible k-tuple. The two greedy optimizations are in searching over candidate sets $[s, s+x]$, and then for each prime choosing the residue class that leaves the most survivors in the candidate set.

We now turn to the problem of bounding M_k (or its variants). We have already described how a lower bound can be computed after specializing to any finite-dimensional space of symmetric functions, in terms of the largest eigenvalue of the matrix corresponding to a quadratic form. Maynard established $M_{105} > 4$ by specializing to a specific symmetric function of the form

$\sum \alpha_i (1-P_1)_i^a P_2^{b_i}$, where P_j is the symmetric j-power sum of $t_1, \ldots, t_k$. (Strictly speaking, the function F being used is the polynomial multiplied by the indicator function of the support set. This is not smooth, but can be approximated by smooth functions to arbitrary precision.) A natural way to try and prove $M_k > 4$ for smaller k is to expand the space of functions in which to search.

One could include P_j for $j \geq 2$, or work with other bases of symmetric functions. Polymath 8b restricted its attention to polynomial symmetric functions, since these are easier to work with, and dense in the space of continuous symmetric functions by Weierstrass's Theorem. Polymath 8b implemented a large quadratic program to search through finite-dimensional spaces parametrized by degree in terms of a convenient basis.

To explain what worked, we introduce some useful notation. For a multi-index $\alpha = (\alpha_1, \ldots, \alpha_k)$, we define the symmetric polynomial

$$m(\alpha) = \sum_{\sigma \in S_k} x_{\sigma(1)}^{\alpha_1} \ldots x_{\sigma(k)}^{\alpha_k}.$$

Polymath 8b searched over spaces of symmetric polynomials spanned by $\{(1 - P_1)^a m(\alpha)\}$ of bounded degree. Other bases were tried, but for unexplained reasons this basis performed significantly better. Searching over the specific space spanned by polynomials with degree at most 23 and α involving only even entries, it was found that $M_{54} \geq 4.001115\ldots$ which proves:

Theorem 4.3.1. *$DHL[54, 2]$ is true. In particular,*

$$\liminf p_{n+1} - p_n \leq 270.$$

Using the ϵ-trick, one can do a little bit better. Taking $\epsilon = \frac{1}{25}$, Polymath 8b showed $DHL[50, 2]$ by specializing to the space of polynomials spanned by $(1 - P_1)^a m(\alpha)$ with total degree at most 27, and α having only even entries. The corresponding matrix had size 2526×2526 and took two weeks to compute. The largest eigenvector was found after another two days to be $4.0043\ldots$, which shows that $M_{50,\frac{1}{25}} > 4$.

Theorem 4.3.2. *$DHL[50, 2]$ is true. In particular,*

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

This is currently the world record.

4.4 IMPROVING THE CONDITIONAL BOUND

Polymath 8b was unable to decrease Maynard's original conditional bound

$$\liminf p_{n+1} - p_n \leq 12 \text{ under EH.}$$

However, using the methods mentioned above (enlarging the support and the ϵ-trick), a bound of 8 was achieved under the GEH conjecture. Using a calculus of variations argument, Polymath 8b found (heuristically) a formula for $M_{k,\epsilon}$, with the optimal choice being described in terms of Bessel functions. Computing with the predicted optimal function, it was found that

$$M_{4,0.18}(1) \geq 2.01869.$$

Noting that $\{0, 2, 6, 8\}$ is an admissible tuple, we have the following theorem.

Theorem 4.4.1. *Under GEH, $DHL[4,2]$ is true, and in particular*

$$\liminf p_{n+1} - p_n \leq 8.$$

Trying the same method for $k = 3$, Polymath 8b found

$$M_{3,\epsilon} > 1.959633.$$

Unfortunately, this falls just short of $DHL[3,2]$. However, Polymath 8b eventually established $DHL[3,2]$ by combining the ϵ-trick with further expansion of the support. Since the tuple $\{0,2,6\}$ is admissible, we obtain:

Theorem 4.4.2. *Under GEH, $DHL[3,2]$ is true, and in particular*

$$\liminf p_{n+1} - p_n \leq 6.$$

This is the best that we expect to be able to do without radically new ideas, due to a phenomenon called the *parity problem*, which we will explain in the next section.

The support expansion involved in obtaining Theorem 4.4.2 is rather technical. The basic idea is that the main contribution to the sums S_1 comes from n which are "almost primes," i.e. have few prime factors. Indeed, $S_2^{(\ell)}$ is essentially the contribution from the primes. One can perform an expansion about the number of primes dividing n in order to quantify this intuition; this step is used by Bombieri-Friedlander-Iwaniec in [1], attempting to improve on the Bombieri-Vinogradov theorem. After conditioning on a bounded number of prime divisors, the sums over divisors involved in the weights becomes "smoother" and hence tractable to control. The upshot is that, in terms of the expansion of support described in § 4.2.1, one can further expand to functions supported on a region R such that

$$R + R \subset \{(t_1, \ldots, t_k) \in [0,4]^k : t_1 + \ldots + t_{\ell-1} + t_{\ell+1} + \ldots + t_k \leq 2 \forall \ell\}$$

(noting that we can take $\theta = 1$ under Elliott-Halberstam) and satisfying ϵ-tricked versions of the conditions on the marginal distributions. This is then enough room to obtain $DHL[3,2]$.

4.5 THE PARITY PROBLEM AND LIMITS OF SIEVE THEORY

It may seem that we are now tantalizingly close to the twin prime conjecture, at least if we assume additional ingredients such as the Generalized Elliott-Halberstam conjecture. We have conditionally proved $DHL[3,2]$; all we need is to push this to $DHL[2,2]$. However, there is a fundamental barrier, called the "parity problem" in the literature, which suggests that the distance between $DHL[3,2]$ and $DHL[2,2]$ will be highly nontrivial to bridge. In particular, we do not expect to close this distance by further relaxing the variational problem in the ways that we have explored so far. The parity problem refers to the phenomenon that sieve theory is generally unable to detect the parity of the number of prime factors that an integer has. Tao [18] describes the general phenomenon as follows (paraphrased):

> If A is a set consisting of integers with only an odd number of prime factors (or only an even number of prime factors), then sieve theory cannot prove a non-trivial lower bound on $\#A$. Furthermore, any sieve-theoretic upper bound on $\#A$ will be off by a factor of at least 2.

This principle manifests itself in many problems. For instance, the twin prime conjecture is apparently very difficult, but it is relatively easy to prove that there are infinitely many pimes p such that $p+2$ is either prime or a product of two primes (Chen's Theorem). Goldbach's conjecture is also apparently very difficult, but Chen also proved that every sufficiently large even number is either the sum of two primes, or the sum of a prime and a product of two primes. The basic reason is that sieve theory attempts to count interesting sets of integers (such as the prime numbers in an interval $[N, 2N]$) by manipulating simple sets of integers (such as the multiples of d in $[N, 2N]$), but these simple sets generally contains roughly equal proportions of integers with an odd or even number of factors. Recall, for example, that the sieve of Eratosthenes uses the latter sets as "measuring cups" to measure the former set. The measuring instruments themselves are insensitive to the parity of the number of prime factors, so we expect that the result will be too.

To be more precise, we introduce the Liouville function $\lambda(n)$, defined to be $(-1)^k$ if n has k prime factors counted with multiplicity. It agrees with the Möbius function on the squarefree integers, which comprise a positive proportion of all integers, so the two are essentially equivalent for the purposes of this discussion, but the Liouville function is more convenient. Then $\lambda(n) = \pm 1$, and the Möbius pseudorandomness principle says that $\{\lambda(n)\}$ behaves like independent realizations of a uniform random variable in $\{\pm 1\}$ except when there are obvious obstructions. For instance, an obvious obstruction is that $\lambda(n) = -\lambda(2n)$, so these two realizations are not independent; on the other hand, there is no obvious reason why $\lambda(n)$ and $\lambda(n+2)$ would be correlated, so in heuristics we assume that they are independent.

If the $\lambda(n)$ do behave like realizations of a random variable, then in $\sum \lambda(n)$ we expect there to be a high level of cancellation. In fact, the Prime Number Theorem is equivalent to

$$\sum_{n \leq N} \lambda(n) = o(N).$$

However, we expect to be able to say more. The central limit theorem predicts that a sum of N random signs has mean 0 with variance on the order of $\sqrt{N}$, so we expect a square root error term above. Indeed, it turns out that the Riemann hypothesis is essentially equivalent to

$$\sum_{n \leq N} \lambda(n) = o(N^{\frac{1}{2}+\epsilon}) \text{ for all } \epsilon > 0.$$

If we restrict the sum to a certain residue class $a \pmod q$, then we expect to be able say the same, by the pseudorandomness heurstic. Indeed, the assertion

$$\sum_{\substack{n \leq N \\ n \equiv a \pmod q}} \lambda(n) = o(N^{\frac{1}{2}+\epsilon}) \text{ for all } \epsilon > 0$$

is essentially equivalent to the Generalized Riemann Hypothesis. This high degree of cancellation can be interpreted as saying that λ is "orthogonal" to sets of the form $q\mathbb{Z} + a$ (or more precisely, orthogonal to their indicator functions) for the purpose of obtaining asymptotics.

Now let's see why this pseudorandomness principle leads to the parity problem. Suppose that $A \subset [1, N]$ is some set consisting entirely of integers with only an odd number of factors (without loss of generality). Then λ is identically equal to -1 on the set A. A sieve-theoretic lower bound on $\#A$ comes through finding an identity of the form

$$\mathbb{1}_A(n) \geq \sum_d \nu(d)\mathbb{1}_{d|n}(n) \tag{23}$$

and summing from 1 to N to deduce

$$\#A \geq \sum_d \nu(d)\frac{N}{d} + O\left(\sum \nu(d)\right).$$

The hope is that the main term will dominate, and $\sum_d \frac{\nu(d)}{d}$ will be positive. But consider multiplying (23) by $(1 + \lambda(n))$, and then summing. Since $1 + \lambda$ vanishes on A, this gives

$$0 \geq \sum_d \nu(d)\frac{N}{d} + \sum_d \nu(d) \sum_{\substack{n \leq N \\ d|n}} \lambda(n) + \dots$$

In keeping with the intuition that λ is orthogonal to $d\mathbb{Z}$, we expect $\sum_{\substack{n \leq N \\ d|n}} \lambda(n)$ to be swamped by the first sum, but that implies $\sum_d \frac{\nu(d)}{d} \leq 0$, so our lower bound was trivial after all.

Similarly, if we multiply an upper bound

$$\mathbb{1}_A(n) \leq \sum_d \nu(d)\mathbb{1}_{d|n}(n) \tag{24}$$

by $(1 + \lambda(n))$ and sum, then we obtain asymptotically the same upper bound for $2\#A$ as for $\#A$. The point here is that typical sieve weights do not "see" the λ function, so they cannot distinguish between 1 and $1 + \lambda(n)$.

In particular, if A is the set of primes $p \leq N$ such that $p + 2$ is also prime, then we expect to have difficulty proving any lower bound on A using sieve-theoretic means. That is not to say that it is impossible, but we expect to have to inject some additional ingredients. For instance, the same difficulty applies to the set of all primes less than N, but in this case one can use the additional multiplicative structure of the primes to obtain the Prime Number Theorem (this is the substance of the Erdös-Selberg elementary proof). In this essay, we have explained how to prove the infinitude of bounded gaps by sieve estimates, although we had to inject ingredients such as the Prime Number Theorem and the Bombieri-Vinogradov Theorem. It is also worth noting that we have not managed to produce a specific h such that we can prove the infinitude of prime pairs $(p, p + h)$.

We can use this pseudorandomness principle to explain why we do not expect to be able to achieve the twin prime conjecture using GPY methods, even with input like the generalized Elliott-Halberstam hypothesis. In the GPY setup, we estimate the sums

$$\sum w(n), \quad \sum w(n)\chi_{\mathcal{P}}(n), \quad \sum w(n)\chi_{\mathcal{P}}(n + 2) \tag{25}$$

and try to parlay them into a positive lower bound for

$$\sum w(n)\chi_{\mathcal{P}}(n)\chi_{\mathcal{P}}(n+2). \tag{26}$$

By the Möbius pseudorandomness principle, we expect $\lambda(n)\lambda(n+2)$ to be orthogonal to $w(n)$:

$$\sum \lambda(n)\lambda(n+2)w(n) = o\left(\sum |w(n)|\right).$$

Now $\lambda(n)$ is not orthogonal to $\chi_{\mathcal{P}}(n)$, as clearly $\lambda(n) = -1$ whenever $\chi_{\mathcal{P}}(n) \neq 0$. However, we expect $\lambda(n+2)$ to behave like a random sign independent of $\chi_{\mathcal{P}}(n)$, and we similarly expect $\lambda(n)$ to be independent of $\chi_{\mathcal{P}}(n+2)$. Therefore, the Möbius pseudorandomness conjecture predicts that the sums (25) will be essentially unmodified if we multiply their summands by $1-\lambda(n)\lambda(n+2)$, in which case the same hypothetical manipulations that led to (26) should give a positive lower bound for

$$\sum w(n)\chi_{\mathcal{P}}(n)\chi_{\mathcal{P}}(n+2)(1-\lambda(n)\lambda(n+2)). \tag{27}$$

But the factor $\chi_{\mathcal{P}}(n)\chi_{\mathcal{P}}(n+2)$ is only non-zero when n and $n+2$ are both primes, while $(1-\lambda(n)\lambda(n+2))$ is zero in these cases, so that sum (27) is trivially non-positive.

In fact similar reasoning suggests that we cannot obtain $\liminf p_{n+1} - p_n \leq 4$ using GPY methods. We don't expect to be able to show $DHL[2,2]$ using current methods, but one could imagine to somehow obtain this bound by manipulating $DHL[3,2]$. In particular, we know (conditional on GEH) that there are infinitely many pairs of primes among the tuples $(n, n+2, n+6)$. One might hope to boostrap off this to show that, in fact, either $(n, n+2)$ or $(n, n+6)$ are prime pairs infinitely often. Following GPY, we study the sums

$$\sum w(n), \sum w(n)\chi_{\mathcal{P}}(n), \sum w(n)\chi_{\mathcal{P}}(n+2), \sum w(n)\chi_{\mathcal{P}}(n+6). \tag{28}$$

Suppose that one could somehow obtain sufficiently good estimates on these sums in order to parlay them into a positive lower bound for

$$\sum w(n)\chi_{\mathcal{P}}(n)\chi_{\mathcal{P}}(n+2) \text{ or } \sum w(n)\chi_{\mathcal{P}}(n+2)\chi_{\mathcal{P}}(n+6). \tag{29}$$

Set $a_n = (1-\lambda(n)\lambda(n+2))(1-\lambda(n+2)\lambda(n+6))$. The pseudorandomness conjecture predicts that $a_n - 1$ is orthogonal to the summands of the sums in (28), so we should obtain essentially the same estimates on those sums if we multiply the summands by a_n. However, a_n is designed to be non-positive when $\lambda(n) = \lambda(n+2)$ or $\lambda(n+2) = \lambda(n+6)$, and in particular on the prime pairs that we are trying to detect. Under the pseudorandomness conjecture, we expect to be able to follow the same hypothetical manipulations leading to (29) and deduce a positive lower bound for

$$\sum a_n w(n)\chi_{\mathcal{P}}(n)\chi_{\mathcal{P}}(n+2) \text{ or } \sum a_n w(n)\chi_{\mathcal{P}}(n+2)\chi_{\mathcal{P}}(n+6).$$

As already mentioned, however, the summands are never positive, so this is impossible.

So, while we have made huge strides towards understanding the additive structure of the prime numbers over the past year, we expect that we still have much further to go before proving (for instance) the twin prime conjecture. But if this story has taught us anything, it is that inspiration may strike when it is least expected.

BIBLIOGRAPHY

[1] E. Bombieri, J. B. Friedlander, and H. Iwaniec. Primes in arithmetic progressions to large moduli. III. *J. Amer. Math. Soc.*, 2(2):215–224, 1989. ISSN 0894-0347. doi: 10.2307/1990976. URL http://dx.doi.org/10.2307/1990976.

[2] Alina Carmen Cojocaru and M. Ram Murty. *An introduction to sieve methods and their applications*, volume 66 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2006. ISBN 978-0-521-64275-3; 0-521-61275-6.

[3] Harold Davenport. *Multiplicative number theory*, volume 74 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 2000. ISBN 0-387-95097-4. Revised and with a preface by Hugh L. Montgomery.

[4] E. Fouvry and H. Iwaniec. On a theorem of Bombieri-Vinogradov type. *Mathematika*, 27(2):135–152 (1981), 1980. ISSN 0025-5793. doi: 10.1112/S0025579300010032. URL http://dx.doi.org/10.1112/S0025579300010032.

[5] Tristan Freiberg. A note on the theorem of maynard and tao. *preprint*, 2013. URL http://arxiv.org/pdf/1311.5319.pdf.

[6] John B. Friedlander and Henryk Iwaniec. Incomplete Kloosterman sums and a divisor problem. *Ann. of Math. (2)*, 121(2):319–350, 1985. ISSN 0003-486X. doi: 10.2307/1971175. URL http://dx.doi.org/10.2307/1971175. With an appendix by Bryan J. Birch and Enrico Bombieri.

[7] Daniel A. Goldston, János Pintz, and Cem Y. Yıldırım. Primes in tuples. I. *Ann. of Math. (2)*, 170(2):819–862, 2009. ISSN 0003-486X. doi: 10.4007/annals.2009.170.819. URL http://dx.doi.org/10.4007/annals.2009.170.819.

[8] Andrew Granville. Primes in intervals of bounded length. *preprint*, 2013. URL http://www.dms.umontreal.ca/~andrew/CEBBrochureFinal.pdf.

[9] Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math. (2)*, 167(2):481–547, 2008. ISSN 0003-486X. doi: 10.4007/annals.2008.167.481. URL http://dx.doi.org/10.4007/annals.2008.167.481.

[10] Henryk Iwaniec and Emmanuel Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004. ISBN 0-8218-3633-1.

[11] Erica Klarreich. Unheralded mathematician bridges the prime gap. https://www.simonsfoundation.org/quanta/20130519-unheralded-mathematician-bridges-the-prime-gap/, 2013.

[12] James Maynard. Small gaps between primes. *preprint*, 2013. URL http://arxiv.org/pdf/1311.4600v2.pdf.

[13] Hugh L. Montgomery and Robert C. Vaughan. *Multiplicative number theory. I. Classical theory*, volume 97 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2007. ISBN 978-0-521-84903-6; 0-521-84903-9.

[14] Yoichi Motohashi. An induction principle for the generalization of Bombieri's prime number theorem. *Proc. Japan Acad.*, 52(6):273–275, 1976. ISSN 0021-4280.

[15] Yoichi Motohashi and János Pintz. A smoothed GPY sieve. *Bull. Lond. Math. Soc.*, 40(2):298–310, 2008. ISSN 0024-6093. doi: 10.1112/blms/bdn023. URL http://dx.doi.org/10.1112/blms/bdn023.

[16] D. H. J. Polymath. Bounded gaps between primes. http://michaelnielsen.org/polymath1/index.php?title=Bounded_gaps_between_primes, 2013.

[17] K. Soundararajan. Small gaps between prime numbers: the work of Goldston-Pintz-Yıldırım. *Bull. Amer. Math. Soc. (N.S.)*, 44(1):1–18, 2007. ISSN 0273-0979. doi: 10.1090/S0273-0979-06-01142-6. URL http://dx.doi.org/10.1090/S0273-0979-06-01142-6.

[18] Terence Tao. Open question: the parity problem in sieve theory. http://terrytao.wordpress.com/2007/06/05/open-question-the-parity-problem-in-sieve-theory/, 2007.

[19] Banks W., T. Freiberg, and Turnage-Butterbaugh. C. Consecutive primes in tuples. *preprint*, 2013. URL http://arxiv.org/pdf/1311.7003.pdf.

[20] Yitang Zhang. Bounded gaps between primes. *Ann. of Math. (2)*, 179(3):1121–1174, 2014. ISSN 0003-486X. doi: 10.4007/annals.2014.179.3.7. URL http://dx.doi.org/10.4007/annals.2014.179.3.7.