Small gaps between primes: Maynard's method and further extensions

Yifan Gong

Abstract

This paper explores the core theoretical advancements in the study of bounded gaps between primes, with a primary focus on comparing the Goldston-Pintz-Yıldırım (GPY) method and the breakthrough multidimensional sieve developed by James Maynard. First, we review the theoretical framework of the GPY method and provide a brief conceptual proof, illustrating how its one-dimensional Selberg-style weight effectively detects prime tuples but ultimately hits a theoretical bottleneck under the unconditional level of distribution ($\vartheta = 1/2$) provided by the Bombieri-Vinogradov Theorem.Subsequently, this paper meticulously details why Maynard's method constitutes a substantial improvement over the GPY approach. The breakthrough lies in extending the one-dimensional sieve into a multidimensional weight where variables can be evaluated independently. This multidimensional construction shatters the rigid coupling of variables inherent in the GPY method, granting the crucial extra flexibility needed to optimize the ratio of the sieve sums unconditionally.Furthermore, we analyze the specific functional choices that lead to the dramatic improvements in small gap bounds. We detail how the optimal smooth function F is selected: for large k, through probabilistic interpretations and the Euler-Lagrange equation; and for small k, by utilizing symmetric polynomials to reduce the problem to maximizing the ratio of two positive-definite quadratic forms (an eigenvalue problem). Finally, we outline the rigorous proof mechanics of this method, including the application of the "W-trick" to handle biases from small prime factors, and the asymptotic evaluation of the sums S_1 and S_2 via Möbius inversion and Selberg sieve manipulations, ultimately establishing the unconditional finite bounds on prime gaps and demonstrating further extensions of the method.

1 Introduction

For any natural number m, let H_m denote the quantity

$$H_m := \liminf_{n\to\infty}(p_{m+n} - p_n),$$

where p_n denotes the n^{th} prime. The twin prime conjecture asserts that $H_1 = 2$, that is, we expect that there are infinitely many prime tuples of the form $(p, p+2)$. Generalizing this, the Hardy-Littlewood prime tuples conjecture [6] states that H_m is equal to the diameter of the narrowest admissible $(m+1)$-tuple (see Section 2 for a definition of this term). For instance, this conjecture predicts that $H_2 = 6$, corresponding to the expectation that there are infinitely many prime triplets of the form $(p, p+2, p+6)$ or

$(p, p+4, p+6)$. Historically, progress on bounding H_m was exceptionally slow. Recall the Prime Number Theorem, which states that the prime counting function $\pi(X)$ (the number of primes $p \leqslant X$) satisfies

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

Since there are asymptotically $X/\log X$ primes up to X, the average gap between consecutive primes around p_n is asymptotically $\log p_n$. For a long time, it was unknown whether one could find prime gaps that are significantly smaller than this average, let alone bounded by an absolute constant. A major milestone was reached in 2005 when Goldston, Pintz, and Yıldırım [3] developed the so-called "GPY method." They successfully proved that there are pairs of primes arbitrarily close to each other relative to the average gap. Furthermore, they demonstrated that if one assumes a strong unproven condition about the distribution of primes in arithmetic progressions—namely, the Elliott-Halberstam conjecture (see Claim 2.2)—then the absolute prime gap H_1 is finite. This conjecture essentially postulates that primes have a level of distribution ϑ that can be any value less than 1.

Theorem 1.1 (Goldston, Pintz, and Yıldırım [3]). *Assuming the Elliot-Halberstam conjecture $H_1 \leqslant 16$. Unconditionally, we have*

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

Remark. The "unconditional" result here refers to the findings derived solely(without needing any unproven conjectures) from the Bombieri-Vinogradov Theorem (Theorem 2.3), which establishes that primes have a level of distribution up to $\vartheta = 1/2$. Because this method is limited by this 1/2 barrier, it can only prove that gaps become small relative to the average, but not that they are bounded by a constant.

In 2014, Yitang Zhang [14] achieved a historic breakthrough by establishing the first unconditional absolute constant bound on H_1. He accomplished this by refining the GPY method and unconditionally proving a weakened, yet sufficient variant of the Elliott-Halberstam conjecture, effectively pushing the level of distribution ϑ just beyond the 1/2 barrier for certain smooth moduli.

Theorem 1.2 (Zhang [14]). $H_1 \leqslant 70{,}000{,}000$.

By optimizing many components in Zhang's argument, the Polymath8 project [10] was able to lower this bound to $H_1 \leqslant 4680$. Very shortly after Zhang's breakthrough, a new variant of the GPY method was discovered independently by Maynard [7] and Tao (unpublished). This work, relying only on the Bombieri-Vinogradov Theorem, was able to provide substantial improvements for all H_m, not only for $m = 1$.

Theorem 1.3 (Maynard [7]). *Unconditionally, we have the following bounds:*
(i) $H_1 \leqslant 600$.
(ii) $H_m \leq Cm^3e^{4m}$ *for all* $m \geq 1$ *and an absolute (and effective) constant* C.
Assuming the Elliott-Halberstam conjecture, we have the following:
(iii) $H_1 \leqslant 12$.
(iv) $H_2 \leqslant 600$.

The Polymath8b project [11] extensively optimized the work by Maynard and Tao, yielding the current world records for bounded prime gaps. Unconditionally, they established $H_1 \leqslant 246$ and $H_2 \leqslant 395{,}126$. Furthermore, assuming the Elliott-Halberstam conjecture, they demonstrated that the gaps could be reduced to $H_1 \leqslant 6$ and $H_2 \leqslant 252$. The method developed by Maynard and Tao is not merely a quantitative improvement over the GPY sieve; its multidimensional structure provides an unprecedented level of robustness. While the classical GPY method is highly sensitive to the level of distribution ϑ, Maynard's approach allows for compensation of weaker distribution information by increasing the dimension k of the sieve.

This newfound flexibility opens the door to a variety of further applications beyond simple bounded gaps. For instance, the method can be adapted to show that there exist intervals containing significantly more primes than what is predicted on average by the Prime Number Theorem—a phenomenon known as the existence of "dense clusters" of primes. To maintain the flow of our primary discussion, we defer the rigorous statement and proof of this particular application to Section 8.

1.1 Organization of the paper

The paper is organized as follows. The paper is structured as follows. Following this introduction, Section 1.2 establishes the notational conventions used throughout, with a particular focus on the asymptotic symbols often encountered in analytic number theory. In Section 2, we introduce the foundational theory of prime distribution in arithmetic progressions. This includes defining the level of distribution ϑ as a parameter for understanding and discussing the critical roles of the Elliott-Halberstam conjecture and the Bombieri-Vinogradov Theorem. We also establish the key arithmetic functions and the admissibility criteria for prime tuples required for the sieve methods. Then, in Section 3, we provide a detailed overview of the classical GPY method and its application to proving Theorem 1.1. Building upon this framework, Section 4 introduces the core components and propositions of Maynard's multidimensional sieve. Section 5 offers a comparative analysis between the GPY and Maynard-Tao approaches, highlighting the theoretical flexibility gained through multidimensionality and the role of the level of distribution. The technical optimization of the sieve weights and the underlying combinatorial proofs are presented in Sections 6 and 7, respectively. Finally, Section 8 explores the robustness of Maynard's method through further applications, concluding with the rigorous statement and proof of Theorem 8.1 regarding the existence of dense clusters of primes.

1.2 Notation

In analytic number theory, we frequently encounter complex error terms and asymptotic bounds. To describe the rate of growth of a function as $N \to \infty$, we use the following standard notations:

- **Big-O notation (O):** $f(N) = O(g(N))$ means that there exists an absolute constant $C > 0$ such that $|f(N)| \leqslant C \cdot g(N)$ for all sufficiently large N. It describes an asymptotic upper bound.
- **Vinogradov symbol ($\ll$):** Throughout this paper, we extensively use the notation $f(N) \ll g(N)$. Crucially, this is strictly synonymous with $f(N) = O(g(N))$. While $\ll$ typically means "much less than" in other mathematical disciplines, in analytic

number theory, it simply means "is asymptotically bounded by" (up to a constant factor). We use it interchangeably with Big-O to maintain the readability of long formulas.

- **Small-o notation (o):** $f(N) = o(g(N))$ is a stricter condition. It means that $f(N)$ becomes completely negligible compared to $g(N)$, i.e., $\lim_{N\to\infty} f(N)/g(N) = 0$.

Throughout this paper, ϑ denotes a parameter representing the level of distribution of primes in arithmetic progressions, typically taking values in the interval $(0, 1]$. We shall view k as a fixed integer.
All sums and products will be over the natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$, with the exceptions of sums and products over p, which will be understood to be over the prime numbers $\mathbb{P} = \{2, 3, \ldots\}$.
Throughout the paper, φ will denote the Euler totient function, $\tau_r(n)$ the number of ways of writing n as a product of r natural numbers, and μ the Möbius function. Let $\omega(n)$ denote the number of distinct prime factors of an integer n. We let χ_E denote the indicator function of a set E, so $\chi_E(n) = 1$ if $n \in E$ and $\chi_E(n) = 0$ otherwise. We use (a, b) to denote the greatest common divisor of a and b, and $[a, b]$ denotes the least common multiple of a and b.

2 Uniformity in arithmetic progressions and estimation over arithmetic functions.

Let $n \in \mathbb{N}$ and define the arithmetic function

$$\theta(n) := \begin{cases} \log n & \text{if } n \text{ prime,} \\ 0 & \text{otherwise} \end{cases}$$

and the counting function for primes in an arithmetic progression $a \pmod q$ (where the modulus is $q \in \mathbb{N}$, and a is an integer coprime to q, typically bounded by $1 \leqslant a \leqslant q$) is defined as:

$$\theta(N; q, a) := \sum_{\substack{n \leqslant N \\ n \equiv a \pmod q}} \theta(n).$$

Definition 2.1. *We say the primes possess a level of distribution ϑ if for any $A > 0$ and $\varepsilon > 0$, the following estimate holds for the range of moduli $q \leqslant N^{\vartheta-\varepsilon}$:*

$$\sum_{q \leqslant N^{\vartheta-\varepsilon}} \max_{(a,q)=1} \left| \theta(N; q, a) - \frac{N}{\varphi(q)} \right| \ll \frac{N}{(\log N)^A}$$

where the implicit constant depends on the choice of A and ε.

Claim 2.2 (Elliott-Halberstam Conjecture)**.** *The primes have a level of distribution ϑ for any value in the interval $(0, 1]$.*

That is, the distribution is "uniform" for almost all moduli up to $N^{1-\varepsilon}$. While the full range $(0, 1]$ remains a conjecture, we rely on a cornerstone unconditional result that establishes a guaranteed baseline for this distribution:

Theorem 2.3 (Bombieri-Vinogradov Theorem)**.** *The primes have a level of distribution* $\vartheta = 1/2$.

In other words, the distribution of primes in arithmetic progressions is well-understood for the range of moduli $q \in (0, N^{1/2-\varepsilon}]$.

While theorems like Bombieri-Vinogradov provide powerful average estimates for the distribution of individual primes, proving bounded gaps requires us to simultaneously track multiple prime conditions. This necessitates a shift in focus from the macroscopic distribution modulo q to the microscopic arrangement of prime clusters.

Specifically, to search for prime clusters of the form $n + h_1, \ldots, n + h_k$, we must first ensure that there are no trivial divisibility barriers inherent to the chosen shifts. To formalize this, for any prime p and a tuple $\mathcal{H} = \{h_1, \ldots, h_k\}$, let $\nu_p(\mathcal{H})$ denote the number of distinct residue classes modulo p occupied by the elements of $\mathcal{H}$. Equivalently, $\nu_p(\mathcal{H})$ is the number of distinct solutions $n \pmod p$ to the congruence $\prod_{i=1}^{k}(n + h_i) \equiv 0 \pmod p$. For squarefree integers q, we extend this definition to $\nu_q(\mathcal{H})$ by multiplicativity.

Definition 2.4 (Admissibility)**.** *Let $k \in \mathbb{N}$. A set of shifts $\mathcal{H} = \{h_1, \ldots, h_k\}$ consisting of k distinct non-negative integers is called* ***admissible*** k***-set*** *if, for every prime p, the elements of $\mathcal{H}$ do not cover all residue classes modulo p. In terms of our notation, this requires:*

$$\nu_p(\mathcal{H}) < p \quad \textit{for all primes } p.$$

When a set of shifts $\mathcal{H}$ is admissible, the corresponding sequence $(n + h_1, \ldots, n + h_k)$ evaluated at an integer n is referred to as an ***admissible*** k***-tuple****.*

Intuitively, if $\nu_p(\mathcal{H}) = p$ for some small prime p, it means $\mathcal{H}$ covers every single residue class modulo p. Therefore, for any choice of integer n, at least one of the shifted numbers $n + h_i$ will be a multiple of p, and thus they could never all be prime simultaneously for large n. For example, consider the 3-set $\mathcal{H}_{bad} = \{0, 2, 4\}$ modulo 3, its elements fall into the residue classes $0, 2, 1$. Thus, $\nu_3(\mathcal{H}_{bad}) = 3$, meaning it covers all classes and is not admissible.

We denote by $H(k)$ the diameter of the narrowest admissible k-tuple, defined as:

$$H(k) := \min_{\mathcal{H} \text{ admissible}} (\max \mathcal{H} - \min \mathcal{H}).$$

For instance, the narrowest admissible 5-set is $\mathcal{H}_5 = \{0, 2, 6, 8, 12\}$, which yields $H(5) = 12$ since $\nu_2(\mathcal{H}_5) = 1 < 2$, $\nu_3(\mathcal{H}_5) = 2 < 3$, and $\nu_5(\mathcal{H}_5) = 4 < 5$; for prime $p \geqslant 7$, it automatically guaranties $\nu_p(\mathcal{H}_5) \leqslant 5 < p$. Through extensive computational searches, the diameters for much larger k have been precisely determined. A critical value for later theorems in this paper is the diameter of the narrowest admissible 105-tuple, which is known to be $H(105) = 600$ [11].

We define the von Mangoldt function and apply the von Mangoldt identity

$$\Lambda(n) := \sum_{d|n} \mu(d) \log \frac{n}{d} = \begin{cases} \log p & \text{if } n = p^m \text{ with } p \text{ prime}, \\ 0 & \text{otherwise}. \end{cases}$$

Thus, it allows us to distinguish prime powers from composite numbers. In general, the generalized von Mangoldt function is defined as

$$\Lambda_k(n) := \sum_{d|n} \mu(d) \left(\log \frac{n}{d}\right)^k,$$

it vanishes if n has more than k prime factors (that is, $\Lambda_k(n) = 0$ if $\omega(n) > k$). To see this, consider the Dirichlet generating function of $\Lambda_k(n)$:

$$\sum_{n=1}^{\infty} \frac{\Lambda_k(n)}{n^s} = \left(\sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}\right)\left(\sum_{n=1}^{\infty} \frac{(\log n)^k}{n^s}\right) = \frac{1}{\zeta(s)} \cdot (-1)^k \zeta^{(k)}(s).$$

Using the Euler product representation $\zeta(s) = \prod_p E_p(s)$ where $E_p(s) = (1 - p^{-s})^{-1}$, and applying the generalized Leibniz rule for the k-th derivative of an infinite product, we have:

$$\frac{\zeta^{(k)}(s)}{\zeta(s)} = \sum_{i_1+\cdots+i_m=k} \sum_{p_1,\ldots,p_m} \frac{E_{p_1}^{(i_1)}(s)}{E_{p_1}(s)} \cdots \frac{E_{p_m}^{(i_m)}(s)}{E_{p_m}(s)}.$$

Since each term in the sum is a product of at most k factors (as $\sum i_j = k$ and each $i_j \geq 1$), and each factor $\frac{E_p^{(i)}(s)}{E_p(s)}$ is a Dirichlet series supported only on the powers of a single prime p, the resulting series is supported on integers n with $\omega(n) \leq k$. Therefore, if $\omega(n) > k$, the coefficient $\Lambda_k(n)$ must be zero.

For an admissible k-set $\mathcal{H}$, let

$$\mathcal{P}(n) = (n + h_1) \cdots (n + h_k).$$

By using the generalized von Mangoldt function, we can detect if every $n + h_i$ is a prime power.

Lemma 2.5. *If $\Lambda_k(\mathcal{P}(n)) \neq 0$, then $\mathcal{P}(n)$ has exactly k distinct prime factors for sufficiently large n.*

Proof. We proceed by contradiction. Suppose that $\Lambda_k(\mathcal{P}(n)) \neq 0$ but $\omega(\mathcal{P}(n)) = r < k$. Then, let $p_1, ..., p_r$ be the prime divisors of $\mathcal{P}(n)$, and for each i, we label the specific term $n + h_{j(i)}$ for which the prime power $p_i^{\alpha} \parallel n + h_{j(i)}$ is maximized, so that for any other term $n + h_j$, it is divisible by an equal or lower power of p_i (if there are multiple terms with the maximized prime power, we simply take one of them). By the Pigeonhole principle, we can take a $n + h_J$ different from each $n + h_{j(i)}$. Then, for any prime power p_i^{γ} dividing $n + h_J$, there exists another term $n + h_{j(i)}$ that is divisible by an equal or higher power of p_i. Consequently, p_i^{γ} must divide their difference: $(n + h_J) - (n + h_{j(i)}) = h_J - h_{j(i)}$. Therefore, since this holds for all prime powers dividing $n + h_J$, this yields

$$n + h_J \mid \prod_{\substack{1 \leqslant j \leqslant k \\ j \neq J}} |h_J - h_j|,$$

and so $n \leqslant n + h_J < C$, where C is some constant determined by $\mathcal{H}$. However, since we fixed a k-set while taking $n \to \infty$, it is impossible. □

In practice, the exact function Λ_k is analytically difficult to compute. Sieve theory overcomes this by using a truncated function that mimics the behavior of Λ_k while remaining computable. Provided with $n \in (N, 2N]$ and $R = o(N)$ as $R \to \infty$, we mimic $\Lambda_k(\mathcal{P}(n))$ using the truncated function

$$\Lambda_{k,R}(\mathcal{P}(n)) = \sum_{\substack{d \mid \mathcal{P}(n) \\ d \leqslant R}} \mu(d) \left(\log \frac{R}{d}\right)^k.$$

This summation is restricted to small divisors $d \leqslant R$, ensuring that the resulting sums are mathematically tractable while remaining sensitive to prime clusters.

For any real number α, let $q \in \mathbb{N}$, and define

$$\tau_\alpha(q) := \alpha^{\omega(q)},$$

when α is a positive integer, this agrees with the definition of the divisor functions, i.e., the number of ways to represent q as α positive factors. Clearly, for real α_1 and α_2, and y, we see that

$$\tau_{\alpha_1}(q)\tau_{\alpha_2}(q) = \tau_{\alpha_1\alpha_2}(q), \quad (\tau_\alpha(q))^y = \tau_{\alpha^y}(q).$$

Lemma 2.6 (Lemma 2, [3]). *We have, for any $\alpha \in \mathbb{R}^+$ and $x \geqslant 1$,*

$$\sum_{q \leqslant x} \mu(q)^2 \frac{\tau_\alpha(q)}{q} \leqslant (\alpha + 1 + \log x)^{\alpha+1},$$

and

$$\sum_{q \leqslant x} \mu(q)^2 \tau_\alpha(q) \leqslant x(\alpha + 1 + \log x)^{\alpha+1}.$$

3 Goldston-Pintz-Yıldırım's Approah

To elucidate GPY's approach to Theorem 1.1, we adopt a narrative and analytic strategy closely mirroring that of Granville [4, Section 4].

3.1 The set up

Let $\mathcal{H} = \{h_1, \dots, h_k\}$ be an admissible set of k-sets. For any integer $n \in (N, 2N]$, we consider the corresponding evaluated k-tuple $(n + h_1, \dots, n + h_k)$. To find clusters of primes, our goal is to construct a suitable non-negative weight function $w(n) \geqslant 0$ such that:

$$\sum_{N < n \leqslant 2N} \left(\sum_{i=1}^{k} \theta(n + h_i) - r \log 3N \right) w(n) > 0 \tag{1}$$

for some positive real parameter $r > 0$.

For $n \in (N, 2N]$, if $n + h_i$ is a prime, we have the strict upper bound $\theta(n + h_i) < \log(2N + h_k) < \log 3N$ for sufficiently large N. If the shifted component is composite, $\theta(n + h_i) = 0$. Thus, if the tuple at n contains exactly m primes, the inner sum satisfies:

$$\sum_{i=1}^{k} \theta(n + h_i) < m \log 3N.$$

Since the weight $w(n)$ is non-negative by definition, the only way the entire sum over all $n \in (N, 2N]$ can be strictly positive is if there exists at least one n where the expression inside the large parentheses is strictly positive. This requires:

$$m \log 3N > \sum_{i=1}^{k} \theta(n + h_i) > r \log 3N \implies m > r.$$

Since the number of primes m must be an integer, this inequality guaranties that there are at least $m = \lfloor r \rfloor + 1$ primes in the tuple $(n+h_1, \dots, n+h_k)$. This immediately yields that there exists some $n \in (N, 2N]$ such that there are at least m prime components in the k-tuple, in particular,

$$\liminf_{n\to\infty}(p_{n+m-1} - p_n) \leqslant \max \mathcal{H} - \min \mathcal{H}.$$

(In the classical GPY argument aiming for H_1, one simply attempts to achieve this positivity condition with $r = 1$, ensuring at least 2 primes).

The most direct idea is to take weight $w(n)$ equal to the generalized von Mangoldt function $(\Lambda_k(\mathcal{P}(n)))$ introduced in Section 2, which perfectly isolates the case where each $n + h_i$ is almost prime. However, substituting Λ_k directly into our sum is analytically intractable because it requires summing over all divisors without restriction.

To overcome this, we follow the classical sieve approach and mimic the behavior of Λ_k using a smoothly truncated weight $\Lambda_{k,R}$. We define the non-negative weight:

$$w(n) = \left(\sum_{\substack{d|\mathcal{P}(n) \\ d \leqslant R}} \lambda(d) \right)^2 \tag{2}$$

where the coefficients $\lambda(d)$ are explicitly defined as:

$$\lambda(d) := \mu(d) \frac{1}{u!} \left(\log \frac{R}{d} \right)^u$$

for a truncation level $R < N^{\frac{1}{3}}$ and some integer $u = k + l$ with $l \geqslant 0$ as a non-negative parameter. The extra parameter l provides the crucial degrees of freedom needed to optimize the sieve.

3.2 Evaluating the sum.

We expand the sum (1) using (2) with $r = 1$ and obtain

$$\sum_{d_1,d_2 \leqslant R} \lambda(d_1)\lambda(d_2) \left(\sum_{i=1}^{k} \sum_{\substack{N<n\leqslant 2N \\ [d_1,d_2]|\mathcal{P}(n)}} \theta(n+h_i) - \log 3N \sum_{\substack{N<n\leqslant 2N \\ [d_1,d_2]|\mathcal{P}(n)}} 1 \right) \tag{3}$$

Here, the condition that both d_1 and d_2 divide $\mathcal{P}(n)$ is mathematically equivalent to saying that their least common multiple, $q = [d_1, d_2]$, divides $\mathcal{P}(n)$. Note that for q to contribute to this sum, we must have q square-free. This reduces a double divisibility condition into a single congruence condition modulo q, allowing us to evaluate the inner sums.

Notice that for a prime $p \mid \mathcal{P}(n)$, we have $n \equiv -h_i \pmod{p}$ for some $h_i \in \mathcal{H}$. The quantity $\nu_p(\mathcal{H})$ counts the number of distinct solutions for n modulo p. For a square-free q, the Chinese Remainder Theorem tells us that the total number of solutions modulo q is simply the multiplicative extension $\nu_q(\mathcal{H})$.

Evaluating the second inner sum in (3): For the simpler sum $\sum 1$, we are just counting integers. The number of integers $n \in (N, 2N]$ belonging to a specific residue class modulo q is simply $\frac{N}{q} + O(1)$. Since there are $\nu_q(\mathcal{H})$ such valid classes, the second inner sum evaluates to:

$$(\log 3N)\, \nu_q(\mathcal{H}) \left(\frac{N}{q} + O(1)\right).$$

Evaluating the first inner sum in (3)**:** The sum involving $\theta(n+h_i)$ is more delicate because $\theta(n+h_i)$ is zero unless $n+h_i$ is prime. If $n+h_i$ shares a common prime factor with q, it cannot be prime (ignoring negligible cases where $n+h_i = p$). Therefore, we must restrict our sum to residue classes modulo q where $(n+h_i, q) = 1$.

For each prime $p \mid q$, the condition $p \mid \mathcal{P}(n)$ originally gave $\nu_p(\mathcal{H})$ valid residue classes. However, the coprime condition $(n+h_i, p) = 1$ explicitly forbids the class $n \equiv -h_i \pmod p$. Thus, exactly one residue class is excluded, leaving $\nu_p(\mathcal{H}) - 1$ valid classes per prime p. We denote this modified multiplicative function as $\nu_q(\mathcal{H})^*$.

For each of the k terms in the outer sum $\sum_{i=1}^{k}$, there are $\nu_q(\mathcal{H})^*$ valid residue classes modulo q. The sum of θ over each such coprime class in the interval is approximated by the expected average $\frac{X_N}{\varphi(q)}$ plus an error term $E(N,q)$. Summing over all k indices gives:

$$k\left(\nu_q(\mathcal{H})^*\right)\left(\frac{X_N}{\varphi(q)} + O(E(N,q))\right)$$

where $X_N = \sum_{N \leqslant n \leqslant 2N} \theta(n)$, and

$$E(N,q) = 1 + \sup_{(a,q)=1} \left| \sum_{\substack{N \leqslant n \leqslant 2N \\ n \equiv a \pmod q}} \theta(n) - \frac{X_N}{\varphi(q)} \right|.$$

Combining the results: Putting both evaluations back into (3), we have

$$\sum_{d_1, d_2 \leqslant R} \lambda(d_1)\lambda(d_2) \left(k\left(\nu_q(\mathcal{H})^*\right)\left(\frac{X_N}{\varphi(q)} + O(E(N,q))\right) - \log 3N\, \nu_q(\mathcal{H})\left(\frac{N}{q} + O(1)\right)\right). \tag{4}$$

3.3 Bounding the error terms.

We now evaluate the error in (4), and note that the latter can be absorbed by the first error term. Trivially $\nu_q(\mathcal{H}) \leq k^{\omega(q)} = \tau_k(q)$, the total error is then bounded by

$$\ll (\log R)^{2u} k \sum_{q \leqslant R^2} \mu(q)^2 \tau_k(q) \tau_3(q) E(N,q).$$

The limit $q \leqslant R^2$ arises because $q = [d_1, d_2]$ with $d_1, d_2 \leqslant R$. The term $\tau_3(q)$ accounts for the number of ways to form a specific q from divisors d_1 and d_2 (bounded by the number of ways to write q as a product of 3 factors).

To handle this sum, we split $E(N,q)$ into $1+(E(N,q)-1)$ and apply the Cauchy-Schwarz inequality to separate the $E(N,q)-1$ term. Combining this with the bounds from Lemma 2.6, the error becomes:

$$\ll kR^2(3\log R)^{3k+2u+1} + k(\log R)^{2u} \sqrt{\sum_{q \leqslant R^2} \frac{\mu(q)^2 \tau_{9k^2}(q)}{q}} \sqrt{\sum_{q \leqslant R^2} q(E(N,q)-1)^2}$$

by the trivial bound for $E(N,q) - 1 \leqslant \frac{2N}{q}\log N$, we have the bound for the latter term

$$\ll k(\log R)^{2u}\sqrt{(\log R^2)^{9k^2}}\sqrt{2N\log N}\sqrt{\sum_{q\leqslant R^2}(E(N,q)-1)}$$

$$\ll kN(\log N)^{\frac{9k^2+1-A}{2}}$$

To rigorously bound this, we must apply the Bombieri-Vinogradov Theorem (Theorem 2.3). However, a technical subtlety arises: our error term $E(N,q)$ is defined over the dyadic interval $(N,2N]$ using X_N, whereas Theorem 2.3 is formulated over $(0,N]$ using N. We bridge this by noting that by the Prime Number Theorem, $X_N = \theta(2N) - \theta(N) = N + O(N/(\log N)^C)$ for any $C > 0$. Consequently, the discrepancy over $(N,2N]$ can be effectively bounded by applying the Bombieri-Vinogradov theorem twice (once for $x = 2N$ and once for $x = N$).

Furthermore, to legally apply Theorem 2.3, our maximum modulus must satisfy $q \leqslant R^2 \leqslant N^{1/2-\varepsilon}$. Therefore, we explicitly choose our truncation level to be $R = N^{\frac{1}{4}-\delta}$ for some small fixed $\delta > 0$ (so that $2\delta = \varepsilon$).

Because the Bombieri-Vinogradov theorem grants us a saving of $O(N/(\log N)^A)$ for any arbitrarily large constant $A > 0$, we can choose A to be sufficiently large to comfortably dominate the $\frac{9k^2+1}{2}$ exponent of the logarithm, provided k is relatively small (e.g., $k \leqslant \sqrt{(\log N)/18}$). Hence, the total error term is bounded by $O(N/(\log N)^{A'})$ for some new large constant A', which becomes completely negligible compared to the main term and is thus absorbed into $o(N)$.

3.4 Perron's Formula

To asymptotically evaluate the main terms from the sieve, we encounter sums of multiplicative functions over divisors, such as $\sum_{d\leqslant R}\frac{\lambda(d)\nu_d(\mathcal{H})}{d}$. In analytic number theory, the standard machinery for estimating such discrete sums is to encode the sequence into a Dirichlet generating function and then extract the partial sums using Perron's Formula via contour integration in the complex plane.

For the simplicity of demonstration, we consider the case where $l = 0$ so that $u = k$. Our sum becomes:

$$\sum_{d\leqslant R}\frac{\lambda(d)\nu_d(\mathcal{H})}{d} = \frac{1}{k!}\sum_{d\leqslant R}\frac{\mu(d)\nu_d(\mathcal{H})}{d}\left(\log\frac{R}{d}\right)^k.$$

Let (c) denote the vertical contour $s = c + it$ with $-\infty < t < \infty$. For $c > 0$, Perron's formula states:

$$\frac{1}{2\pi i}\int_{(c)}\frac{x^s}{s^{k+1}}ds = \begin{cases} 0 & \text{if } 0 \leqslant x \leqslant 1, \\ \frac{1}{k!}(\log x)^k & \text{if } x \geqslant 1. \end{cases}$$

Applying this to the sums of multiplicative functions over divisors, we can rewrite it as:

$$\frac{1}{k!}\sum_{d\leqslant R}\frac{\mu(d)\nu_d(\mathcal{H})}{d}\left(\log\frac{R}{d}\right)^k = \frac{1}{2\pi i}\int_{(1)}F(s)\frac{R^s}{s^{k+1}}ds$$

where $F(s)$ is the Dirichlet series:

$$F(s) = \sum_{d=1}^{\infty}\frac{\mu(d)\nu_d(\mathcal{H})}{d^{1+s}} = \prod_p\left(1 - \frac{\nu_p(\mathcal{H})}{p^{1+s}}\right)$$

We can extract the behavior of $F(s)$ by comparing it to the Riemann zeta function $\zeta(1+s)$. We write:

$$G_{\mathcal{H}}(s) = \prod_p \left(1 - \frac{\nu_p(\mathcal{H})}{p^{1+s}}\right)\left(1 - \frac{1}{p^{1+s}}\right)^{-k}$$

$$= \prod_p \left(1 + \frac{k - \nu_p(\mathcal{H})}{p^{1+s}} + O_k\left(\frac{1}{p^{2+2\sigma}}\right)\right)$$

Since $\nu_p(\mathcal{H}) = k$ for all primes $p > \max(\mathcal{H})$, the coefficient of $p^{-(1+s)}$ vanishes for large p. Thus, the Euler product for $G_{\mathcal{H}}(s)$ converges absolutely, making it analytic and uniformly bounded for $\sigma > -\frac{1}{2} + \delta$ for any $\delta > 0$. We then have the factorization:

$$F(s) = \frac{G_{\mathcal{H}}(s)}{\zeta(1+s)^k},$$

with the critical value at $s = 0$ being the singular series:

$$G_{\mathcal{H}}(0) = \mathfrak{S}(\mathcal{H}) = \prod_p \left(1 - \frac{1}{p}\right)^{-k}\left(1 - \frac{\nu_p(\mathcal{H})}{p}\right).$$

The integrand now has a simple pole at $s = 0$ (coming from the zeta function and the cancelation with the denominator). By shifting the contour of integration to a new path $\mathcal{L} : s = -\frac{c}{\log(1+|t|)} + it$ (moving it slightly to the left of the imaginary axis but remaining within the zero-free region of the zeta function), we sweep across this simple pole. By Cauchy's Residue Theorem, the main term of our integral is precisely the residue at this simple pole, which evaluates to $G_{\mathcal{H}}(0) = \mathfrak{S}(\mathcal{H})$. The remaining integral over the contour $\mathcal{L}$ decays rapidly and is absorbed into the error term.

To rigorously evaluate the actual double sum involving d_1 and d_2, we extend this machinery to two complex variables. We construct the corresponding two-variable Dirichlet series:

$$F(s_1, s_2) = \sum_{d_1, d_2 \geqslant 1} \frac{\mu(d_1)\mu(d_2)\nu_q(\mathcal{H})}{d_1^{s_1} d_2^{s_2} q}.$$

Because the summand is a multiplicative function and the Möbius function restricts it to squarefree integers, we can evaluate its Euler product prime-by-prime. For any given prime p, we only have four squarefree cases to consider regarding its divisibility of d_1 and d_2: $p \nmid d_1, d_2$; $p \mid d_1, p \nmid d_2$; $p \nmid d_1, p \mid d_2$; and $p \mid d_1, p \mid d_2$. Evaluating the summand for these four cases yields the p-th local factor:

$$1 - \frac{\nu_p(\mathcal{H})}{p^{1+s_1}} - \frac{\nu_p(\mathcal{H})}{p^{1+s_2}} + \frac{\nu_p(\mathcal{H})}{p^{1+s_1+s_2}}.$$

Recall that for all sufficiently large primes p, we have $\nu_p(\mathcal{H}) = k$. To understand the analytic behavior of $F(s_1, s_2)$, we isolate its poles by comparing this local factor to the Euler factors of the Riemann zeta function. By matching the first-order terms, we can extract the singular behavior and factorize the Dirichlet series as:

$$F(s_1, s_2) = \frac{\zeta(1+s_1+s_2)^k}{\zeta(1+s_1)^k \zeta(1+s_2)^k} G(s_1, s_2),$$

where $G(s_1, s_2)$ is an Euler product that converges absolutely in a wider region past the imaginary axes.

Applying Perron's formula twice (once for each variable), the discrete double sum transforms into a continuous double contour integral:

$$\frac{1}{(2\pi i)^2}\int_{(1)}\int_{(c)}\frac{\zeta(1+s_1+s_2)^k}{\zeta(1+s_1)^k\zeta(1+s_2)^k}G(s_1,s_2)\frac{R^{s_1+s_2}}{s_1^{k+1}s_2^{k+1}}ds_1ds_2.$$

This integral representation reveals the intricate coupling of the GPY method: the terms $\zeta(1+s_1)^k\zeta(1+s_2)^k$ provide the necessary poles to extract main terms, while the numerator $\zeta(1+s_1+s_2)^k$ acts as a zero that heavily influences the cross-terms. By carefully shifting these contours, evaluating the multidimensional residues, and subsequently optimizing the parameters k and l, this rigorous complex analytic machinery establishes the bounds in Theorem 1.1.

4 Maynard's Multidimensional Sieve: Key Ingredients

The new ingredient of Maynard's approach is the use of a more general form of sieve weight $w_n \geqslant 0$. Let $\mathcal{H}$ be a k-admissible set, considering the sum

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

where $\rho > 0$, and $S > 0$. Similar to Section 2, this yields that there exists some $n \in (N, 2N]$ such that there exists at least $\lfloor\rho\rfloor + 1$ prime components in the k-tuple. The idea here is to apply the following "W-trick" and k-dimensional weight (see more details in Section 5). Let $W = \prod_{p\leqslant w} p$ with $w = \log\log\log N$. Recall from Section 2 that because our k-set $\mathcal{H}$ is admissible, it does not cover all residue classes modulo any prime. Specifically, for every prime $p \mid W$, there is at least one residue class that avoids $\mathcal{H}$. Then, by the Chinese Remainder Theorem, we can choose $\nu_0 \pmod W$ such that for all n satisfying $n \equiv \nu_0 \pmod W$, we have $(n+h_i, W) = 1$ for each i. We now transform the sum S into $S_1 - \rho S_2$ where

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

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

Proposition 4.1. *Let the primes have an exponent of distribution $\vartheta > 0$, and let $R = N^{\frac{\vartheta}{2}-\delta}$ for some small fixed $\delta > 0$. Let $\lambda_{d_1,\ldots,d_k}$ be defined in terms of a fixed smooth function F by*

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

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

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

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

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

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

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

Intuitively, one can think of $I_k(F)$ as the "total mass" or "variance" of our weights, while $J_k^{(m)}(F)$ represents the "success rate" or "expected contribution" of detecting a prime specifically at the m-th position of our tuple (from S_2). The goal of the entire method is thus reduced to an optimization problem: finding a shape for the function F that maximizes the ratio of the total expected prime contributions to the total weight mass, denoted by M_k.

Let $\mathcal{S}_k$ denote the set of Riemann-integrable functions $F : [0,1]^k \to \mathbb{R}$ supported on $\mathcal{R}_k = \{(x_1, \dots, x_k) \in [0,1]^k : \sum_{i=1}^{k} x_i \leqslant 1\}$ with $I_k(F) \neq 0$ and $J_k^{(m)}(F) \neq 0$ for each m. Let

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

Since for any $F_0 \in \mathcal{S}_k$, we can find a F_1 smooth function satisfying the properties in Proposition 4.1 with $\|F_1 - F_0\|_2 < \varepsilon$. Then the sum is

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

$$\geqslant \frac{\varphi(W)^k N (\log R)^k I_k(F_1)}{W^{k+1}} \left(\left(\frac{\vartheta}{2} - \delta \right) (M_k - 2\delta) - \rho + o(1) \right).$$

Take $\rho < \frac{\vartheta M_k}{2}$, then $S > 0$ for all large N, provided δ is sufficiently small.

Proposition 4.2. *Let* $k \in \mathbb{N}$. *Then*
(1) $M_5 > 2$.
(2) $M_{105} > 4$.
(3) If k *is sufficiently large,* $M_k > \log k - 2 \log \log k - 2$.

By Proposition 4.2 and the bounds for $H(k)$, we can establish Theorem 1.3.

5 Explanation of the refinements

5.1 The "W-trick"

This technique was discovered in 2004 by Terence Tao and Ben Green [5] to avoid biases (mod p) $\leqslant w$ and to consider only selected residue classes modulo W. To understand this intuitively, consider that prime numbers are not randomly distributed among all integers—they obviously never fall on multiples of 2, 3, or 5 (except for the primes themselves). If we search for primes without accounting for this, our sieve wastes "energy" by considering residue classes that are guaranteed to be composite. In particular, such a useful device can be applied to GPY's argument where $\mathfrak{S}(\mathcal{H})$ has a factor $\nu_p(\mathcal{H})$ that needs to be taken care of when p is small ($\nu_p(\mathcal{H}) = k$, if $p > h_k$) [13]. By pre-selecting an integer $W = \prod_{p \leqslant w} p$ (for a slowly growing $w = \log\log\log N$) and restricting our search strictly to $n \equiv \nu_0 \pmod{W}$, we manually bypass these small primes.

In Maynard's construction, the "W-trick" is extremely useful because It removes the complicated singular series $\mathfrak{S}(\mathcal{H})$ that caused difficulties in the original GPY argument. Furthermore, if we restrict r to be squarefree and coprime to W, arithmetic functions like Euler's totient $\varphi(r)$ and the multiplicative function $g(r)$ (defined by $g(p) = p - 2$ on p primes by multiplicity) become almost identical to r itself (i.e., $\frac{r}{\varphi(r)} = 1 + o(1), \frac{r}{g(r)} = 1 + o(1)$). This brilliant simplification ensures that when we evaluate the ratio S_2/S_1, we only need to focus on optimizing the continuous function M_k, cleanly separating the analysis from messy arithmetic functions.

5.2 1 dimensional v-s. multidimensional

The fundamental limitation of the GPY method lies in its 1-dimensional sieve weight: $d \mid \prod_{i=1}^{k}(n + h_i)$. This structure treats all prime factors of the tuple symmetrically. It assigns a single variable d to the entire product, analogous to forcing a multidimensional function to be in the form of $F(t_1, \ldots, t_k) = g(t_1 + \cdots + t_k)$. This rigid form restricts the maximum possible value of M_k to at most 4. This rigid symmetry artificially constrains the optimization space. It has been proven that under this symmetric constraint, the ratio M_k is bounded strictly by $M_k \leqslant 4$. Recalling our golden inequality $\rho < \frac{\vartheta M_k}{2}$, and knowing that the Bombieri-Vinogradov theorem only gives $\vartheta = 1/2$ unconditionally, the 1-dimensional GPY method is absolutely capped at $\rho < 1$. Thus, it can never unconditionally prove the existence of intervals with 2 or more primes (which requires $\rho \geqslant 1$).

Maynard's breakthrough was to shatter this symmetry by introducing a multidimensional sieve weight. By assigning a separate divisor variable $d_i \mid (n+h_i)$ to each individual component of the tuple, the coordinates $t_1, \ldots, t_k$ become flexible and dependent. This multidimensional construction grants the extra degrees of freedom needed to push $M_k > 4$ (and eventually grow to infinity with k), thereby breaking the $\rho = 1$ barrier completely and unconditionally.

6 Choice of $F \in \mathcal{S}_k$

6.1 Choice of smooth function for large k.

We choose F to be the form

$$F(t_1, \dots, t_k) = \begin{cases} \prod_{i=1}^k k^{\frac{1}{2}} g(kt_i), & \text{if } \sum_{i=1}^k t_i \leqslant 1, \\ 0, & \text{otherwise,} \end{cases}$$

for the smooth function $g : [0, \infty) \to \mathbb{R}$ supported on $[0, T]$, with T to be chosen later, normalized so that $\|g\|_2 = 1$. Then, F is symmetric, and $J_k^{(m)}(F)$ is independent of m. For simplicity, write $J_k = J_k^{(1)}(F)$, $I_k = I_k(F)$. We have

$$I_k = \underset{\mathcal{R}_k}{\int \cdots \int} F(t_1, \dots, t_k)^2 dt_1 \dots dt_k \leqslant k^k \left(\int_0^\infty g(kt)^2 dt \right)^k \leqslant 1.$$

It follows that

$$M_k \geqslant k \underset{t_2 + \cdots + t_k \leqslant 1 - \frac{T}{k}}{\int \cdots \int} \left(\int_0^\infty F(t_1, \dots, t_k) dt_1 \right)^2 dt_2 \dots dt_k$$

$$= k \underset{t_2 + \cdots + t_k \leqslant 1 - \frac{T}{k}}{\int \cdots \int} \left(\prod_{i=2}^k (k^{\frac{1}{2}} g(kt_i)) k^{-\frac{1}{2}} \int_0^\infty g(t) dt \right)^2 dt_2 \dots dt_k$$

$$= \left(\int_0^\infty g(t) dt \right)^2 \underset{t_2 + \cdots + t_k \leqslant 1 - \frac{T}{k}}{\int \cdots \int} \left(\prod_{i=2}^k k^{\frac{1}{2}} g(kt_i) \right)^2 dt_2 \dots dt_k. \tag{5}$$

Evaluating a $(k-1)$-dimensional integral directly when k is large is extremely difficult. However, Tao [12] beautifully observed that we can interpret this integral using probability theory. Notice that the integrand $\prod (k^{\frac{1}{2}} g(kt_i))^2$ acts exactly like a joint probability density function.

Let us define independent, identically distributed (i.i.d.) random variables $X_1, \dots, X_{k-1}$ whose probability density function is given by $g(t)^2 dt$. The massive multiple integral (5) is then exactly equal to the probability that the sum of these random variables falls within a certain range. Specifically, it yields:

$$M_k \geqslant \left(\int_0^\infty g(t) dt \right)^2 P(X_1 + \cdots + X_{k-1} \leqslant k - T).$$

Let $Z = X_1 + \cdots + X_{k-1}$. We estimate the lower bound for the above probability as follows. First, we define the mean of our underlying distribution as:

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

By linearity, the expected value of the sum is $\mathbb{E}[Z] = (k-1)\mu$.

To bound the tail probability, we observe that $P(Z \leqslant k - T) = 1 - P(Z > k - T)$. We can bound the complement by centering the variable:

$$P(Z > k - T) \leqslant P\big(|Z - \mathbb{E}[Z]| > k - T - (k-1)\mu\big).$$

Since the support of g is $[0, T]$, we can bound the variance of each random variable by its second moment: $\mathrm{Var}(X_i) \leqslant \mathbb{E}[X_i^2] \leqslant T\mathbb{E}[X_i] = T\mu$. Applying Chebyshev's inequality to the sum Z (and assuming $k - T > (k-1)\mu$ for our eventual choice of g), we establish the lower bound:

$$P(Z \leqslant k - T) \geqslant 1 - \frac{(k-1)T\mu}{(k - T - (k-1)\mu)^2}.$$

By relaxing the parameters slightly for simplicity, this yields the clearer bound:

$$P(Z \leqslant k - T) \geqslant 1 - \frac{kT\mu}{(k - T - k\mu)^2}.$$

Since $k - T > k\mu$, we have $\mu \leqslant 1$, this leads to

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

To maximize our lower bound, we wish to maximize the integral $\int_0^T g(t)dt$ subject to the constraints $\int_0^T g(t)^2 dt = 1$ and $\int_0^T tg(t)^2 dt = \mu$. We incorporate these constraints using Lagrange multipliers α and β, constructing the functional:

$$\mathcal{L}(g) = \int_0^T g(t)dt - \alpha\left(\int_0^T g(t)^2 dt - 1\right) - \beta\left(\int_0^T tg(t)^2 dt - \mu\right).$$

By the Euler-Lagrange equation, a necessary condition for a local extreme is that the functional derivative with respect to g must vanish. Therefore, the Euler-Lagrange equation simplifies dramatically to just setting the partial derivative of the integrand with respect to g to zero for all $t \in [0, T]$:

$$\frac{\partial}{\partial g}\Big(g(t) - \alpha g(t)^2 - \beta t g(t)^2\Big) = 1 - 2\alpha g(t) - 2\beta t g(t) = 0.$$

Solving for $g(t)$, we obtain the explicit form:

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

Note that the ratio kJ_k/I_k is not affected by multiplying g by a positive constant. Therefore, we can factor out the constant terms and restrict our attention to functions of the simpler form $g(t) = 1/(1 + At)$ for $t \in [0, T]$ and some constant $A > 0$. We quote the choice of $A = \log k - 2\log\log k > 0$, and $1 + AT = e^A$ for the choice of T by Maynard. This ensures $k - T > k\mu$ for k large and $M_k \geqslant \log k - 2\log\log k - 2$.

6.2 Choice of weight for small k

To maximize the ratio with small k we need to consider when such F is optimal.

Lemma 6.1. *Let $\mathscr{L}_k$ denote the linear operator defined by*

$$\mathscr{L}_k F(u_1, \dots, u_k) = \sum_{m=1}^{k} \int_0^{1-\sum_{i\neq m} u_i} F(u_1, \dots, u_{m-1}, t_m, u_{m+1}, \dots, u_k) dt_m$$

whenever $(u_1, \dots, u_k) \in \mathcal{R}_k$, and zero otherwise. Then F maximizes the ratio $\sum_{m=1}^{k} J_k^{(m)}(F)/I_k(F)$ if F is the eigenfunction for $\mathscr{L}_k$, and the corresponding eigenvalue is the value of the ratio at F.

Proof. We define the inner product on $\mathcal{S}_k$ as

$$\langle F, G\rangle = \int \cdots \int_{\mathcal{R}_k} F(t_1, \dots, t_k) G(t_1, \dots, t_k) dt_1 \dots dt_k.$$

Then $I_k(F) = \int \cdots \int F^2 dt = \langle F, F\rangle$, and by Fubini's Theorem, we find that

$$\sum_{m=1}^{k} J_k^{(m)}(F) = \langle F, \mathscr{L}_k F\rangle,$$

so that

$$R(F) = \frac{\sum_{m=1}^{k} J_k^{(m)}(F)}{I_k(F)} = \frac{\langle F, \mathscr{L}_k F\rangle}{\langle F, F\rangle}.$$

Note that such $R(F)$ is maximized if

$$\delta R(F) = \frac{\delta\langle F, \mathscr{L}_k F\rangle\langle F, F\rangle - \langle F, \mathscr{L}_k F\rangle \delta\langle F, F\rangle}{\langle F, F\rangle^2} = 0$$

where $\delta\langle F, \mathscr{L}_k F\rangle = \lim_{\epsilon\to 0}(\langle F+\epsilon\delta F, \mathscr{L}_k(F+\epsilon\delta F)\rangle - \langle F, \mathscr{L}_k F\rangle)/\epsilon$ with δF is a sufficiently small perturbation. Since $\mathscr{L}_k$ is linear and symmetric, and the inner product is symmetric, this gives

$$\delta\langle F, \mathscr{L}_k F\rangle = \langle \delta F, \mathscr{L}_k F\rangle + \langle F, \mathscr{L}_k \delta F\rangle = 2\langle \delta F, \mathscr{L}_k F\rangle.$$

It follows that the numerator of $\delta R(F)$ is

$$2\langle \delta F, \mathscr{L}_k F\rangle\langle F, F\rangle - \langle F, \mathscr{L}_k F\rangle 2\langle \delta F, F\rangle = 0,$$

and by dividing $2\langle F, F\rangle$, we have

$$\langle \delta F, \mathscr{L}_k F\rangle - \frac{\langle F, \mathscr{L}_k F\rangle}{\langle F, F\rangle}\langle \delta F, F\rangle = 0.$$

If we plug $R(F)$ into the previous equation, it yields

$$\langle \delta F, \mathscr{L}_k F - R(F)F\rangle = 0.$$

Therefore, $R(F)$ is maximized if $\mathscr{L}_k(F) = R(F)F$; in other words, F is the eigenfunction of $\mathscr{L}_k$ with $R(F)$ being the eigenvalue, so that $M_k = \sup_{F\in\mathcal{S}_k} R(F)$. □

However, solving the operator equation $\mathscr{L}_k(F) = \lambda F$ exactly for an arbitrary k is extremely difficult with current analytical techniques (it remains unsolved for $k > 2$). Since we cannot find the exact infinite-dimensional optimal function F, we compromise. We restrict our search to a finite-dimensional subspace of polynomials because any function in L^2 can be approximated by polynomials. Hence, we consider the optimal function F of the form

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

Note, if F satisfies $\mathscr{L}_k F = \lambda F$, then for any permutation of $t_1, \dots, t_k$, $F(\sigma(t_1), \dots, \sigma(t_k))$ is also an eigenfunction. Therefore, if we take the average of the sum of F with every permutation of $t_1, \dots, t_k$, we obtain a symmetric function that is also an eigenfunction. Therefore, we assume that the optimal function is symmetric. Recall that in algebra, all symmetric polynomials can be written as a polynomial expression in the first k (number of variables) power sum symmetric polynomials $P_j = \sum_{i=1}^k x_i^j$.

The trick is that we do not need to go through every expression with first k power sum symmetric polynomials. Maynard cleverly restricted the search space to a specific tractable form:

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

where $a_i \in \mathbb{R}$ are coefficients to be determined, and b_i, c_i are fixed non-negative integers. For such a form of P, it is enough to obtain an eigenvalue $R(F) \approx 4.002 > 4$ with $k = 105$.

But why this specific form? Because integrating polynomials of this type over a simplex $\mathcal{R}_k$ can be explicitly computed using the Beta function identity. This reduces our impossible calculus problem to a solvable linear algebra problem. Both the numerator $J_k^{(m)}(P)$ and the denominator $I_k(P)$ become positive-definite quadratic forms in terms of the coefficient vector $a = (a_1, \dots, a_d)$:

$$\frac{\sum_{m=1}^k J_k^{(m)}(P)}{I_k(P)} = \frac{a^T A_2 a}{a^T A_1 a}$$

where A_1 and A_2 are computable symmetric matrices.

Lemma 6.2. *Such a ratio is maximized when a is an eigenvector of $A_1^{-1} A_2$ corresponding to the largest eigenvalue of $A_1^{-1} A_2$. The value of the ratio at its maximum is this largest eigenvalue.*

Proof. Since multiplying a by a scalar doesn't affect the ratio. WLOG, we assume that $a^T A_1 a$, and apply the Lagrangian equation to maximize $a^T A_2 a$ subject to the constraint $a^T A_1 a = 1$, it yields

$$L(a, \lambda) = a^T A_2 a - \lambda (a^T A_1 a - 1)$$

with $\lambda > 0$ and

$$0 = \frac{\partial L}{\partial a} = 2 A_2 a - 2 \lambda A_1 a = 0.$$

It follows that

$$(A_1^{-1} A_2) a = \lambda a, \quad \frac{a^T A_2 a}{a^T A_1 a} = \frac{a^T \lambda A_1 a}{a^T A_1 a} = \lambda.$$

□

7 Sieve asymptotics

7.1 Selberg sieve manipulations

We restrict the support of $\lambda_{d_1,\dots,d_k}$ to tuples for which $d = \prod_{i=1}^k d_i \leqslant R$ holds with $\mu(d)^2 = 1$ and $(d, W) = 1$. Now, we expand the sum S_1 and S_2 with this choice of $\lambda_{d_1,\dots,d_k}$ and obtain

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

We take N large with $W > h_k$. If $([d_i, e_i], [d_j, e_j]) > 1$ for some $i < j$, then $(n+h_i, n+h_j) > 1$ and $([d_i, e_i], [d_j, e_j]) \leqslant h_j - h_i \leqslant h_k$, which is impossible. Hence, with the support of $\lambda_{d_1,\dots,d_k}$, we consider $W, [d_1, e_1], \dots, [d_k, e_k]$ to be pairwise coprime; otherwise, the sum is 0. Then, by CRT, the inner sum is $N/q + O(1)$ where $q = W\prod_{i=1}^k [d_i, e_i]$. This gives

$$S_1 = \frac{N}{W} \sum_{\substack{d_1,\dots,d_k\\ e_1,\dots,e_k}}{}' \frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{\prod_{i=1}^k [d_i, e_i]} + O\Big(\sum_{\substack{d_1,\dots,d_k\\ e_1,\dots,e_k}}{}' |\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}|\Big)$$

where $\sum'$ is used to denote the restriction for $[d_1, e_1], \dots, [d_k, e_k]$ to be pairwise coprime. Put $\lambda_{\max} = \sup_{d_1,\dots,d_k} |\lambda_{d_1,\dots,d_k}|$, the error is then

$$\ll \lambda_{\max}^2 \Big(\sum_{d\leqslant R} \tau_k(d)\Big)^2 \ll \lambda_{\max}^2 R^2 (\log R)^{2k}.$$

For the main term, we apply the identity

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

It follows that

$$\frac{N}{W} \sum_{u_1,\dots,u_k} \Big(\prod_{i=1}^k \varphi(u_i)\Big) \sum_{\substack{d_1,\dots,d_k\\ e_1,\dots,e_k\\ u_i|d_i,e_i\forall i}}{}' \frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{(\prod_{i=1}^k d_i)(\prod_{i=1}^k e_i)}.$$

The requirement for $[d_1, e_1], \dots, [d_k, e_k]$ to be pairwise disjoint can be written as $(d_i, e_j) = 1$ for all $i \neq j$ since $\lambda_{d_1,\dots,d_k} = 0$ if $(d_i, d_j) \neq 1$ for some $i \neq j$. By applying the identity

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

this transforms the main term to

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

The sum over $s_{i,j}$ remains the same when restricting $(s_{i,j}, u_i) = (s_{i,j}, u_j) = 1$ and $(s_{i,j}, s_{i,a}) = (s_{i,j}, s_{b,j}) = 1$ for all $a \neq j$, $b \neq i$, and is denoted by $\sum^*$. Then, Maynard makes a change of variable by letting

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

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

By applying Möbius inversion, we have

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

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

Thus, any choice of $y_{r_1,\ldots,r_k}$ supported on $r_1, \ldots, r_k$, with the product $r = \prod_{i=1}^{k} r_i$ square-free and satisfying $r < R$ and $(r, W) = 1$, will provide a suitable choice of $d_1, \ldots, d_k$. Now, since $d/\varphi(d) = \sum_{e|d} 1/\varphi(e)$ for square-free d, we find by taking $r' = \prod_{i=1}^{k} r_i/d_i$ that $\lambda_{\max} \ll y_{\max} (\log R)^k$.
Hence, the error term $O(\lambda_{\max}^2 R^2 (\log R)^{2k})$ is of size $O(y_{\max}^2 R^2 (\log R)^{4k})$. As for the main term, write $a_j = u_j \prod_{i \neq j} s_{j,i}$ and $b_j = u_j \prod_{i \neq j} s_{i,j}$, then $(a_i, a_j) = 1$, $\mu(a_i)^2 = 1$ for all $i \neq j$ and similarly for all b_i's. This gives us

$$\frac{N}{W} \sum_{u_1,\ldots,u_k} \Big(\prod_{i=1}^{k} \frac{\mu(u_i)^2}{\varphi(u_i)}\Big) \sum_{s_{1,2},\ldots,s_{k,k-1}}{}^{*} \Big(\prod_{\substack{1 \leqslant i,j \leqslant k \\ i \neq j}} \frac{\mu(s_{i,j})}{\varphi(s_{i,j})^2}\Big) y_{a_1,\ldots,a_k} y_{b_1,\ldots,b_k}.$$

In this way, the contribution when $s_{i,j} > D_0$ is

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

When $s_{i,j} = 1$ with $i \neq j$, the rest of the terms are

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

For S_2, we write $S_2 = \sum_{m=1}^{k} S_2^{(m)}$, where

$$S_2^{(m)} = \sum_{\substack{N \leqslant n < 2N \\ n \equiv \nu_0 \pmod{W}}} \chi_{\mathbb{P}}(n + h_m) w_n.$$

Similarly to what we do in previous analyzes and GPY's argument, by putting

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

where g is a totally multiplicative function defined on primes by $g(p) = p-2$, for any fixed $A > 0$, we have

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

Now we relate our new variables $y^{(m)}_{r_1,\ldots,r_k}$ to the $y_{r_1,\ldots,r_k}$ variables from S_1.

Lemma 7.1 (Lemma 5.3, [7])**.** *If* $r_m = 1$*, then*

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

7.2 Smooth choice of y

Our goal is to choose $y_{r_1,\ldots,r_k}$ so that the ratio of the main terms of S_2 and S_1 is maximized. We apply the Lagrangian Euler equation to maximize S_2 subject to the constraint of S_1, and apply Lemma 7.1 so that

$$\lambda \frac{y_{r_1,\ldots,r_k}}{\prod_{i=1}^{k}\varphi(r_i)} = \sum_{m=1}^{k} \frac{y^{(m)}_{r_1,\ldots,r_{m-1},1,r_{m+1},\ldots,r_k}}{\prod_{i\neq m} g(r_i)\varphi(r_m)}$$

by taking the partial derivative with respect to $y_{r_1,\ldots,r_k}$ where λ is a fixed constant. Note,

$$\frac{\partial S_2}{\partial y_{r_1,\ldots,r_k}} = \frac{\partial S_2}{\partial y^{(m)}_{r_1,\ldots,1,r_{m+1},\ldots,r_k}} \frac{\partial y^{(m)}_{r_1,\ldots,r_k}}{\partial y_{r_1,\ldots,r_k}} = \sum_{m=1}^{k} \frac{2y^{(m)}_{r_1,\ldots,r_{m-1},1,\ldots,r_k}}{\prod_{i\neq m} g(r_i)} \frac{1}{\varphi(r_m)}.$$

Hence,

$$\lambda y_{r_1,\ldots,r_k} = \Big(\prod_{i=1}^{k} \frac{\varphi(r_i)}{g(r_i)}\Big) \sum_{m=1}^{k} \frac{g(r_m)}{\varphi(r_m)} y^{(m)}_{r_1,\ldots,r_{m-1},1,r_{m+1},\ldots,r_k}.$$

Since the variables $y_{r_1,\ldots,r_k}$ are supported on integers free of small prime factors, we have $\varphi(r) \approx g(r) \approx r$ for most such r, and so the above relation simplifies to:

$$\lambda y_{r_1,\ldots,r_k} = \sum_{m=1}^{k} y^{(m)}_{r_1,\ldots,r_{m-1},1,r_{m+1},\ldots,r_k}.$$

At this stage, we are faced with a massive discrete optimization problem. Finding the exact discrete values of $y_{r_1,\ldots,r_k}$ that maximize our target ratio is analytically intractable due to the arithmetic fluctuations of the divisors. However, standard sieve theory heuristics suggest that the optimal weights should not fluctuate wildly; rather, they should vary smoothly with respect to the logarithmic size of the divisors.

To leverage the powerful tools of calculus, specifically to transition from discrete summations over r_i to continuous Riemann integrals. Maynard restricts the search space by sampling y from a fixed smooth function F. We define:

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

where $F : \mathbb{R}^k \to \mathbb{R}$ is a smooth function supported on the simplex $\mathcal{R}_k = \{(x_1, \ldots, x_k) \in [0,1)^k : \sum_{i=1}^k x_i \leqslant 1\}$. We also set $y_{r_1,\ldots,r_k} = 0$ if $r = \prod_{i=1}^k r_i$ is not square-free ($\mu(r) = 0$) or if $(r, W) > 1$. Let $F_{\max} = \sup_{(t_1,\ldots,t_k)\in[0,1]^k} |F(t_1,\ldots,t_k)| + \sum_{i=1}^k \left|\frac{\partial F}{\partial t_i}(t_1,\ldots,t_k)\right|$. Now, we substitute this choice of y into our expressions for S_1 and S_2, respectively. This gives

$$S_1 = \frac{N}{W} \sum_{\substack{u_1,\ldots,u_k \\ (u_i,u_j)=1\forall i\neq j \\ (u_i,W)=1}} \Big(\prod_{i=1}^k \frac{\mu(u_i)^2}{\varphi(u_i)}\Big) F\Big(\frac{\log u_1}{\log R}, \ldots, \frac{\log u_k}{\log R}\Big)^2 + O\Big(\frac{F_{\max}^2 \varphi(W)^k N (\log R)^k}{W^{k+1} D_0}\Big).$$

Note, we can drop the requirement that $(u_i, u_j) = 1$, this contributes to an error

$$\ll \frac{F_{\max}^2 N}{W} \sum_{p>D_0} \sum_{\substack{u_1,\ldots,u_k<R \\ p|u_i,u_j \\ (u_i,W)=1\forall i}} \prod_{i=1}^k \frac{\mu(u_i)^2}{\varphi(u_i)}$$

$$\ll \frac{F_{\max}^2 N}{W} \sum_{p>D_0} \frac{1}{(p-1)^2} \Big(\sum_{\substack{u<R \\ (u,W)=1}} \frac{\mu(u)^2}{\varphi(u)}\Big)^k$$

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

Thus, we are left to evaluate the sum

$$\sum_{\substack{u_1,\ldots,u_k \\ (u_i,W)=1\forall i}} \Big(\prod_{i=1}^k \frac{\mu(u_i)^2}{\varphi(u_i)}\Big) F\Big(\frac{\log u_1}{\log R}, \ldots, \frac{\log u_k}{\log R}\Big)^2.$$

Lemma 7.2. *Let $A_1, A_2, L > 0$. Let γ be a multiplicative function satisfying*

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

and

$$-L \leqslant \sum_{w\leqslant p<z} \frac{\gamma(p)\log p}{p} - \log\frac{z}{w} \leqslant A_2$$

for any $2 \leqslant w \leqslant z$. Let g be the multiplicative function defined by

$$g(d) = \prod_{p|d} \frac{\gamma(p)}{p - \gamma(p)}.$$

Assume also that $G : [0,1] \to \mathbb{R}$ be smooth, and let $G_{\max} = \sup_{t\in[0,1]}(|G(t)| + |G'(t)|)$. Then

$$\sum_{d<z} \mu(d)^2 g(d) G\Big(\frac{\log d}{\log z}\Big) = \mathfrak{S}\log z \int_0^1 G(x)dx + O_{A_1,A_2}(\mathfrak{S} L G_{\max}),$$

where

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

This is [2, Lemma 4], with $\kappa = 1$ and slight changes to the notation.

We can now apply this lemma k times, dealing with the sum over each u_i in turn. For each application, we take

$$\gamma(p) = \begin{cases} 1, & p \nmid W, \\ 0, & \text{otherwise.} \end{cases}$$

then

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

it gives that

$$\sum_{\substack{u_1,\ldots,u_k \\ (u_i,W)=1\forall i}} \Big(\prod_{i=1}^k \frac{\mu(u_i)^2}{\varphi(u_i)}\Big) F\Big(\frac{\log u_1}{\log R}, \ldots, \frac{\log u_k}{\log R}\Big)^2 = \frac{\varphi(W)^k (\log R)^k}{W^k} I_k(F)$$

$$+ O\Big(\frac{F_{\max}^2 \varphi(W)^k (\log D_0)(\log R)^{k-1}}{W^k}\Big)$$

where

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

Similarly, we can apply this lemma to $y^{(m)}_{r_1,\ldots,r_k}$ and then to S_2; this gives

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

where

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

Therefore, it completes the proof of the first half of Theorem 1.3.

8 Further Applications: Dense Clusters

One of the most significant takeaways of Maynard's approach is its versatility in handling subsets where equidistribution results are limited. By choosing a sufficiently large dimension k, we can establish the existence of intervals containing many primes even in these constrained settings. The primary goal of this section is to prove the following result regarding dense clusters:

Theorem 8.1 (Maynard [8])**.** *For any $N, m \geqslant 1$, there are $\gg N \exp(-\sqrt{\log N})$ integers $n_0 \in [N, 2N]$ such that*

$$\pi(n_0 + m) - \pi(n_0) \gg \log m.$$

This theorem demonstrates that there are many intervals of length m containing considerably more primes than the average count ($\approx m/\log N$) predicted by the Prime Number Theorem.

8.1 Set up

Consider the modified sum

$$S = \sum_{n\in\mathcal{A}(x)} \Big(\sum_{i=1}^{k} \chi_{\mathbb{P}}(n+h_i) - m - k\chi_{\mathcal{B}}(n) \Big) w_n$$

for some sets of integers $\mathcal{A}, \mathcal{B}$ and sets of primes $\mathbb{P}$. If $S > 0$, this means that there exists a $n \in \mathcal{A}(x) := \{n \in \mathcal{A} : x < n \leqslant 2x\}$ such that there are at least $m+1$ of $n+h_i$'s that are primes and that $n \notin \mathcal{B}$.

Instead of taking the "W-trick" as before, we take $W = \prod_{p\leqslant 2k^2, p|B} p$ for a B integer to be chosen, and this B functions to reduce the effect from $k\chi_{\mathcal{B}}(n)$. The reason we use k in the "W-trick" is that we are taking B as a restriction in the uniform estimates of the prime distribution, and k functions to control the effect of this error without a stronger assumption. This yields a term of a singular series, and apart from this, the idea basically follows from the previous proof of Maynard's approach.

8.2 Equidistribution needed to prove Theorem 1.4

Lemma 8.2. *If $z \geqslant 2$, then of all L-functions formed with primitive characters to moduli $q \leqslant z$, there is at most one function that has a real zero for $\sigma = 1 - c_1/\log z$.*

This is [9, Lemma 9]. Let $q_0 \leqslant \exp(2c_1\sqrt{\log x})$ be the real non-principal character χ modulo q_0 for which $L(s,\chi)$ has a real zero with $\sigma > 1 - c_2(\log x)^{-1/2}$. And for $q \neq q_0$, we have the bound [1, Chapter 20]

$$\varphi(q)^{-1} {\sum_{\chi}}^{*} |\psi(x,\chi)| \ll x\exp(-3c_1\sqrt{\log x})$$

where the summation is over all primitive χ mod q and $\psi(x,\chi) = \sum_{n\leqslant x} \Lambda(n)\chi(n)$. Choose B to be the largest prime factor of q_0 if there exists a real zero; otherwise, choose it to be zero. Following a standard proof of the BV Theorem, we have

$$\sum_{\substack{q\leqslant x^{\frac{1}{2}-\varepsilon} \\ (q,B)=1}} \Big|\pi(x;q,a) - \frac{\pi(x)}{\varphi(q)}\Big| \ll x\exp(-c_1\sqrt{\log x}) + \sum_{\substack{q\leqslant \exp(2c_1\sqrt{\log x}) \\ (q,B)=1}} {\sum_{\chi}}^{*} \frac{|\psi'(x;\chi)|}{\varphi(q)}.$$

In such equidistribution, we are able to compute the size of the error explicitly without k involved. In other words, we can choose k freely for the sake of limiting the equidistribution results.

References

[1] Harold Davenport. *Multiplicative number theory*, volume 74. Springer Science & Business Media, 3rd edition, 2000.

[2] Daniel A Goldston, Sidney W Graham, János Pintz, and Cem Y Yıldırım. Small gaps between products of two primes. *Proceedings of the London Mathematical Society*, 98(3):741–774, 2009. arXiv preprint math/0609615.

[3] Daniel A Goldston, János Pintz, and Cem Y Yıldırım. Primes in tuples i. *Annals of Mathematics*, 170(2):819–862, 2009. arXiv preprint math/0508185.

[4] Andrew Granville. Primes in intervals of bounded length. *Bulletin of the American Mathematical Society*, 52(2):171–222, 2015.

[5] Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics*, 167(2):481–547, 2008.

[6] Godfrey Harold Hardy and John Edensor Littlewood. Some problems of 'partitio numerorum'; iii: On the expression of a number as a sum of primes. *Acta Mathematica*, 44(1):1–70, 1923.

[7] James Maynard. Small gaps between primes. *Annals of Mathematics*, 181(1):383–413, 2015. arXiv preprint arXiv:1311.4600.

[8] James Maynard. Dense clusters of primes in subsets. *Compositio Mathematica*, 152(7):1517–1554, 2016. arXiv preprint arXiv:1405.2593.

[9] Arthur Page. On the number of primes in an arithmetic progression. *Proceedings of the London Mathematical Society*, 2(1):116–141, 1935.

[10] D. H. J. Polymath. New equidistribution estimates of zhang type. *Algebra & Number Theory*, 8(9):2067–2199, 2014. arXiv preprint arXiv:1402.0811.

[11] D. H. J. Polymath. Variants of the selberg sieve, and bounded intervals containing many primes. *Research in the Mathematical Sciences*, 1(12), 2014. arXiv preprint arXiv:1407.4897.

[12] Terence Tao. Polymath8b: Bounded intervals with many primes, after maynard. Terence Tao's Blog, 2013.

[13] Terence Tao. The prime tuples conjecture, sieve theory, and the work of goldston-pintz-yildirim, motohashi-pintz, and zhang. Terence Tao's Blog, 2013.

[14] Yitang Zhang. Bounded gaps between primes. *Annals of Mathematics*, 179(3):1121–1174, 2014.