Gaps between primes

Travor Liu

16 October 2025

- Why do you want to climb Mt. Everest, Sir? - Because it's there.

George L. Mallory

Contents

Preface

This document is a translation of my Zhihu article series on prime gaps, written in 2022 while I was an undergraduate student at University College London.

The first article discusses a corollary of the prime number theorem

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

The second article relates the prime gap and the distribution of zeros of the Riemann zeta function $\zeta(s)$: If Θ is the supremum of the real parts of zeros of $\zeta(s)$ in the critical strip, then

$$p_{n+1}-p_n \ll p_n^{\Theta} \log p_n.$$

The third article introduces sieve methods into the picture and proves an early result of P. Erdös [8]

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

The rest of the series is dedicated to bounded gaps between primes. Three articles are devoted to developing the groundbreaking sieve of Goldston, Pintz, and Yıldırım [11], eventually showing that

$$\liminf_{n\to\infty}(p_{n+1}-p_n) \leq C \tag{0.1}$$

for some $C < \infty$ provided that the level of distribution θ of primes in arithmetic progression is larger than $\frac{1}{2}$ (The Bombieri–Vinogradov theorem gives $\theta = \frac{1}{2}$). This is then followed by an article addressing the limitations of their method.

Subsequently, we discuss the breakthrough work of Yitang Zhang [23], who overcame the difficulties in the GPY sieve and showed that one can unconditionally take $C = 7 \times 10^7$ in (0.1). Zhang's treatment of the error terms is too technical to be presented in the series, so the article only details the developments of the main term in Zhang's sieve.

The last two articles are dedicated to the work of James Maynard [14], in which he showed that one can take $C = 600$ in (0.1). In addition, he showed that

$$\liminf_{n\to\infty}(p_{n+m}-p_n) \ll m^3 e^{4m}. \tag{0.2}$$

The first of these develops Maynard's version of the GPY sieve and converts it into a variational problem. The second article discusses Maynard's solution to the variational problem and deduces (0.2) and his estimate for C.

Travor Liu
Stanford, California
September 18, 2025

Notation

p denotes a prime number.

$s = \sigma + it$ refers to a complex number with real part σ and imaginary part t.

$\rho = \beta + i\gamma$ refers to a zero of the Riemann zeta function $\zeta(s)$ in the critical strip with real part β and imaginary part γ.

$n \equiv a(q)$ means $n \equiv a \pmod q$.

$\pi(x; q, a) = \sum_{\substack{p \le x \\ n \equiv a(q)}} 1$ and $\pi(x) = \pi(x; 1, 1)$.

$\mathrm{li}(x)$ is the logarithmic integral defined by the principal value integral

$$\lim_{\varepsilon \to 0^+} \left(\int_0^{1-\varepsilon} + \int_{1+\varepsilon}^{x} \right) \frac{\mathrm{d}u}{\log u}.$$

$\Lambda(n)$ is the von Mangoldt function equal to $\log p$ if $n = p^k$ and zero otherwise.

$\psi(x; q, a) = \sum_{\substack{n \le x \\ n \equiv a(q)}} \Lambda(n)$ and $\psi(x) = \psi(x; 1, 1)$.

1 A beginning from the prime number theorem

In 2013, Yitang Zhang [23] caused a sensation in the mathematical world by establishing the existence of infinitely many pairs of primes whose difference is bounded by 70 million. After Zhang, this bound is continually reduced. As of today, the best bound is 246 obtained by the Polymath project [15] in December 2014. In this series of articles, we will introduce some important results in the study of prime gaps.

In this article, we use tools from analytic number theory to discuss the most basic properties of the prime gap.

1.1 Partial sum of prime gaps

Denote by p_n the n'th prime. Then by definition,

$$p_{N+1} = p_1 + (p_{N+1} - p_1) = p_1 + \sum_{1 \le n \le N} (p_{n+1} - p_n),$$

so when $N = \lfloor x \rfloor$, one has

$$S(x) = \sum_{n \le x} (p_{n+1} - p_n) = p_{\lfloor x \rfloor + 1} - p_1 \tag{1.1}$$

To better study the properties of (1.1), we need to estimate the size of $p_{\lfloor x \rfloor + 1}$ with respect to x. Let $\pi(x)$ be the number of primes within x. Then the prime number theorem states that

$$\pi(x) \sim \frac{x}{\log x}. \tag{1.2}$$

Set $x = p_n$, so this becomes $\pi(x) = n \sim p_n / \log p_n$. Taking logarithms, we get $\log n \sim \log p_n$. Substituting this back into (1.2), we obtain

$$p_n \sim n \log n. \tag{1.3}$$

Remark. By using a stronger version of the prime number theorem, we can improve (1.3) to $p_n = n(\log n + \log\log n + O(1))$.

Plugging (1.3) into (1.1), we deduce that

$$S(x) \sim x \log x. \tag{1.4}$$

For sequences a_n and b_n, we say that b_n is an **average order** of a_n is $\sum_{n \le x} a_n \sim \sum_{n \le x} b_n$ as $x \to \infty$. According Stirling's formula, we know $\sum_{n \le x} \log n \sim x \log x$, so $\log n$ is an average order of $p_{n+1} - p_n$. This information motivates us to compare the magnitude of the prime gap with the natural logarithm.

1.2 Prime gap and natural logarithm

Suppose a, b are constants such that $a \log n \leq p_{n+1} - p_n \leq b \log n$ for all large n. Then

$$[a + o(1)]x \log x \leq S(x) \leq [b + o(1)]x \log x. \tag{1.5}$$

Now, plugging (1.4) into (1.5), we conclude that $a \leq 1 \leq b$. This means for all $\varepsilon > 0$, there exists infinitely many n such that $p_{n+1} - p_n > (1 - \varepsilon) \log n$ and infinitely many other n such that $p_{n+1} - p_n < (1 + \varepsilon) \log n$. Combining this with $\log n \sim \log p_n$ and the language of $\limsup$ and $\liminf$, we obtain the following inequalities:

$$\liminf_{n \to \infty} \frac{p_{n+1} - p_n}{\log p_n} \leq 1, \tag{1.6}$$

$$\limsup_{n \to \infty} \frac{p_{n+1} - p_n}{\log p_n} \geq 1. \tag{1.7}$$

1.3 Conclusion

In this article, we began by discussing the partial sum $S(x)$ of prime gaps. Using the prime number theorem, we proved $S(x) \sim x \log x$, allowing us to deduce (1.6) and (1.7). Specifically, (1.6) indicates that the gap between consecutive primes can be as small as their logarithms infinitely many times, while (1.7) indicates that the gap can be as large as their logarithms indefinitely. Hence, the prime number theorem marked the commencement of two types of investigations into the prime gap:

Small gaps between primes: Can we find infinitely many n such that $p_{n+1} - p_n \leq f(n)$? The state-of-the-art result in this direction is $f(n) \leq 246$ due to Polymath.

Large gaps between primes: Can we find infinitely many n such that $p_{n+1} - p_n \geq F(n)$? The best record up to now is

$$F(n) \gg \frac{\log n \log\log n \log\log\log\log n}{\log\log\log n}$$

due to Ford–Green–Konyagin–Maynard–Tao [9] in 2017.

In the subsequent articles, we focus on the small gaps between primes.

June 23, 2022

2 Prime gaps and the zeros of $\zeta(s)$

In 1936, Harald Cramer [5] proved using complex-analytic methods that under the Riemann hypothesis,

$$p_{n+1} - p_n = O(\sqrt{p_n} \log p_n).$$

In this article, we prove a generalization:

Theorem 2.1. *When $\Theta \geq \frac{1}{2}$ is the supremum of the abscissa of the zeros of $\zeta(s)$, one has*

$$p_{n+1} - p_n = O(p_n^{\Theta} \log p_n).$$

2.1 Method of investigation

Let $\pi(x)$ be the number of primes within x. Then $\pi(y) > \pi(x)$ if and only if $(x, y]$ contains a prime. This observation allows us to estimate the upper bound of prime gaps.

The relationship with the zeros of $\zeta(s)$ is established by Riemann's explicit formula:

$$\pi(x) = \mathrm{li}(x) - \lim_{T \to +\infty} \sum_{\substack{\rho \\ |\gamma| \leq T}} \mathrm{li}(x^{\rho}) + O\left(\frac{\sqrt{x}}{\log x}\right). \tag{2.1}$$

Because the sum over zeros in (2.1) is conditionally convergent, Cramer's derivation relies on his earlier extensive study [4] of the sum $\sum_{\rho} e^{\rho z}$ in 1919. Today, we have more advanced tools to prevent us from directly manipulating conditionally convergent series, so we can get a simplified proof for Theorem 2.1.

2.2 From infinite series to finite sums

For various conveniences, in analytic number theory, the partial sum $\psi(x)$ of the von Mangoldt function $\Lambda(n)$ is used in place of $\pi(x)$ when it comes to the distribution of primes, for it only differs from

$$\vartheta(x) = \sum_{p \leq x} \log p$$

by an error of $\ll \sqrt{x}$. Therefore, if we can find some $f(x)$ growing faster than $\sqrt{x}$ such that for $L = L(x)$,

$$\psi(x, L) = \sum_{x - L < n \leq x + L} \Lambda(n) > f(x), \tag{2.2}$$

then the interval $(x - L, x + L]$ will contain a prime. By Perron's formula [6, p. 109], one has for $2 \leq T \leq x$ that

$$\psi(x) = x - \sum_{|\Im \rho| \leq T} \frac{x^{\rho}}{\rho} + O\left(\frac{x \log^2 x}{T}\right). \tag{2.3}$$

As the formula (2.3) only involves a finite sum, we can manipulate terms freely without worrying about convergence issues.

Let $N(T)$ be the number of zeros of $\zeta(s)$ with $0 \leq \beta \leq 1$ and $0 < \gamma \leq T$. Then by the Riemann–von Mangoldt formula [6, p. 98], we have

$$N(T) = \frac{T}{2\pi} \log \frac{T}{2\pi} - \frac{T}{2\pi} + O(\log T). \tag{2.4}$$

Combining (2.4) with partial summation, one has

$$\sum_{|\Im\rho| \leq T} \frac{1}{|\rho|} = O(\log^2 T).$$

Setting $T = x^{1-\Theta}$, we see that (2.3) becomes

$$\psi(x) = x + O(x^{\Theta} \log^2 x).$$

Therefore, when $L \leq x$, there exists absolute $A > 0$ such that

$$\psi(x, L) > 2L - Ax^{\Theta} \log^2 x.$$

Set $L = Ax^{\Theta} \log^2 x$, so that the right-hand side is $\gg x^{\Theta} \log^2 x$. By partial summation, we see that there exists some $C > 0$ such that the number of primes in $(x - Cx^{\Theta} \log^2 x, x + Cx^{\Theta} \log^2 x]$ is $\gg x^{\Theta} \log^2 x$.

Since $x \pm Cx^{\Theta} \log^2 x \asymp x$, we also deduce that

Theorem 2.2. *When $\Theta \geq \frac{1}{2}$ is the supremum of the abscissa of the zeros of $\zeta(s)$, one has*

$$p_{n+1} - p_n = O(p_n^{\Theta} \log^2 p_n).$$

Thus, we see that a direct asymptotic evaluation of (2.2) only gives a result off from Theorem 2.1 by a logarithm. However, we can fill the gap by introducing a weight $w(n) \geq 0$ such that

$$\psi(x, L) \geq \sum_n w(n)\Lambda(n).$$

2.3 Choice of weights

We are now in a situation similar to one possible development of the large sieve. By taking ideas from [6, p. 155], we introduce the linear weight:

$$w(t) = \max\left(1 - \frac{|x-t|}{L}, 0\right),$$

so we have

$$\begin{aligned}\sum_n w(n)\Lambda(n) &= \int_{x-L}^{x+L} w(t)\mathrm{d}\psi(t) = -\int_{x-L}^{x+L} w'(t)\psi(t)\mathrm{d}t \\ &= \frac{1}{L}\int_x^{x+L} \psi(t)\mathrm{d}t + \frac{1}{L}\int_x^{x-L} \psi(t)\mathrm{d}t.\end{aligned}$$

Now, defining

$$\psi_1(x) = \int_0^x \psi(t)\mathrm{d}t, \tag{2.5}$$

we simplify this to

$$\psi(x, L) \geq \frac{1}{L}\psi_1(x+L) + \frac{1}{L}\psi_1(x-L) - \frac{2}{L}\psi_1(x). \tag{2.6}$$

In the rest of this article, we estimate the right-hand side of (2.6) by evaluating ψ_1 asymptotically.

2.4 Asymptotic formula for $\psi_1(x)$

By integration by parts, one has

$$\psi_1(x) = x\psi(x) - \int_0^x t\mathrm{d}\psi(t) = \sum_{n\leq x}(x-n)\Lambda(n). \tag{2.7}$$

Using the observation that for $k > 0$,

$$\frac{1}{2\pi i}\int_{k-i\infty}^{k+i\infty} \frac{y^{s+1}}{s(s+1)}\mathrm{d}s = \max(y-1, 0),$$

we can rewrite (2.7) into a contour integral:

$$\psi_1(x) = \frac{-1}{2\pi i}\int_{2-i\infty}^{2+i\infty} \frac{x^{s+1}}{s(s+1)}\frac{\zeta'}{\zeta}(s)\mathrm{d}s. \tag{2.8}$$

Moving the contour to $\sigma = -1$ and applying standard estimates for ζ'/ζ (see, for example, [6, p. 108]), we obtain

$$\psi_1(x) = \frac{x^2}{2} - \sum_{\rho}\frac{x^{\rho+1}}{\rho(\rho+1)} + O(1). \tag{2.9}$$

By partial summation with (2.4), one finds that the sum over ρ converges absolutely. Motivated by the arguments in the previous section, we still truncate the sum, so it follows from

$$\sum_{|\gamma|>T}\frac{1}{\rho(\rho+1)} \ll \sum_{\gamma>T}\frac{1}{\gamma^2} = \int_T^{+\infty}\frac{\mathrm{d}N(u)}{u^2} \ll \frac{\log T}{T}$$

that for $2 \leq T \leq x$, (2.9) becomes

$$\psi_1(x) = \frac{x^2}{2} - \sum_{|\gamma|\leq T}\frac{x^{\rho+1}}{\rho(\rho+1)} + O\left(\frac{x^{\Theta+1}\log x}{T}\right) \tag{2.10}$$

Remark. We get an error better than directly integrating (2.3).

2.5 Proof of Theorem 2.1

Plugging (2.10) into (2.6), one has

$$\begin{aligned}\psi(x,L) \geq &\frac{(x+L)^2+(x-L)^2-2x^2}{2L} \\ &-\frac{1}{L}\sum_{|\gamma|\leq T}\frac{(x+L)^{\rho+1}+(x-L)^{\rho+1}-2x^{\rho+1}}{\rho(\rho+1)} \\ &+O\left(\frac{x^{\Theta+1}\log x}{LT}\right).\end{aligned}$$

Notice that

$$\frac{y^{\rho+1}}{\rho(\rho+1)}=\int_0^y \mathrm{d}t\int_0^t u^{\rho-1}\mathrm{d}u=\int_0^y (y-u)u^{\rho-1}\mathrm{d}u,$$

so the blue part becomes

$$\begin{aligned}&\frac{(x+L)^{\rho+1}+(x-L)^{\rho+1}-2x^{\rho+1}}{\rho(\rho+1)} \\ &=\int_{x-L}^{x+L}(L-|x-u|)u^{\rho-1}\mathrm{d}u=O(L^2x^{\Theta-1}).\end{aligned}$$

Plugging these back, we get

$$\psi(x,L)\geq L-A_1(LTx^{-1})x^{\Theta}\log x-A_2(LTx^{-1})^{-1}x^{\Theta}\log x.$$

Now, setting $T=xL^{-1}$, $A_3>A_1+A_2$, and $L=A_3x^{\Theta}\log x$, we deduce that

$$\sum_{x-L<p\leq x+L}\log p\gg\psi(x,L)\gg x^{\Theta}\log x,$$

which indicates that for some $C>0$ and all large x, the interval $(x,x+Cx^{\Theta}\log x]$ always contains a prime, completing the proof of Theorem 2.1

2.6 Conclusion

In this article, we connect the problem of prime gaps with the zeros of $\zeta(s)$ using explicit formulas. By using the truncated explicit formula (2.3), we deduce Theorem 2.2. Finally, by introducing weights to the estimation of $\psi(x,L)$, we improved Theorem 2.2 to Theorem 2.1.

From the derivations, we can also find out the limitations of the method. Because $\psi(x)$ and $\vartheta(x)$ differ by an error of $\asymp\sqrt{x}$, the approach in the present article is incapable of going beyond $p_{n+1}-p_n=O(\sqrt{p_n})$. Nevertheless, the idea of introducing weights in the process of refining Theorem 2.2 plays a crucial role in the study of prime gaps. Stay tuned for the new articles!

June 29, 2022

3 Sieve method and small gaps

In §1, we proved using the prime number theorem that

$$\liminf_{n\to\infty} \frac{p_{n+1}-p_n}{\log p_n} \leq 1. \tag{3.1}$$

The first improvement to (3.1) was given by Hardy and Littlewood in 1926. In 1926, they applied the circle method and successfully replaced the right-hand side of (3.1) with $\frac{2}{3}$ under GRH. In 1940, Rankin [16] replaced $\frac{2}{3}$ with $\frac{3}{5}$ under GRH. In the same year, Erdös [8] used an elementary method to show unconditionally that the right-hand side of (3.1) can be replaced with $1-\eta$ for some $\eta > 0$. In this article, we will walk through Erdös's approach.

3.1 Main idea

According to the definition of limit infimum, if (3.1) cannot be improved, then for all $\delta > 0$, there is some $n_0(\delta)$ such that

$$n > n_0 \Rightarrow p_{n+1} - p_n > (1-\delta)\log p_n.$$

On the other hand, there exist infinitely many n for which

$$p_{n+1} - p_n < (1+\delta)\log p_n.$$

If we can derive a contradiction from this information, then we can deduce

$$\liminf_{n\to\infty} \frac{p_{n+1}-p_n}{\log p_n} \neq 1$$

and thus completing the proof.

3.2 Differencing and summing

Denote by $q_1, q_2, \ldots, q_t$ the primes in $(x, 2x]$. Then evidently, $= q_t - q_1 \leq x$. We can also write x as a sum of prime gaps:

$$S = q_t - q_{t-1} + q_{t-1} - q_{t-2} + \cdots + q_2 - q_1 = \sum_{1\leq k<t} (q_{k+1} - q_k) \tag{3.2}$$

Let T_1 be the number of k's such that

$$(1-\delta)\log q_k \leq q_{k+1} - q_k \leq (1+\delta)\log q_k$$

and T_2 be the number of k's such that this inequality is false. Then by (3.2), we have

$$S \geq T_1(1-\delta)\log x + T_2(1+\delta)\log x. \tag{3.3}$$

If we can show that the right-hand side of (3.3) is $> x$ for sufficiently large x, then we can reach a contradiction to conclude. To fulfill this objective, we need to estimate T_1 and T_2

3.3 Treatments for T_2

Let $\pi(y)$ be the number of primes $\leq y$. Then, according to the prime number theorem,

$$T_2 = \pi(2x) - \pi(x) - T_1 = [1 + o(1)]\frac{x}{\log x} - T_1.$$

Plugging into (3.3), we get

$$S > -2T_1\delta \log x + [1 + \delta + o(1)]x, \tag{3.4}$$

completing the easy step of our derivation.

3.4 Estimation of T_1

In analytic number theory, a typical strategy to estimate a single sum is to convert it into a double sum and then interchange the order of summation. Define $I = [(1-\delta)\log x, (1+\delta)\log 2x]$, so

$$T_1 \leq \sum_{\substack{1\leq k<t \\ q_{k+1}-q_k \in I}} 1 = \sum_{m\in I} \sum_{\substack{1\leq k<t \\ q_{k+1}-q_k = m}} 1. \tag{3.5}$$

3.5 Preliminary handling of the blue term

Because we are looking for an upper bound, we can relax the conditions in the summation to simplify the task. In the context of (3.5), we can weaken $q_{k+1} - q_k = m$ to $q_k + m$ being a prime, so

$$\sum_{\substack{1\leq k<t \\ q_{k+1}-q_k = m}} 1 \leq \sum_{\substack{1\leq k<t \\ q_k+m \text{ prime}}} 1 = \sum_{\substack{x<p\leq 2x \\ p+m \text{ prime}}} 1. \tag{3.6}$$

For an integer $N > 1$, it follows from the pigeonhole principle that N is a prime if and only if it has no prime factor $\leq \sqrt{N}$, so when $2 \leq z \leq \sqrt{x}$, one has

$$\sum_{\substack{x<p\leq 2x \\ p+m \text{ prime}}} 1 \leq \sum_{\substack{x<n\leq 2x \\ p<z \Rightarrow p\nmid n(n+m)}} 1. \tag{3.7}$$

3.6 Application of sieve methods

Let

$$\mathcal{A} = \{n(n+m) : x < n \leq 2x\}, \quad \mathcal{A}_d = \{a \in \mathcal{A} : d|a\},$$

and $\mathcal{P}$ denotes the set of primes. Then under the standard sieve notation, the right-hand side of (3.7) is precisely $S(\mathcal{A}, \mathcal{P}, z)$. Therefore, we have transformed the prime gap problem into a sieve problem.

Let ν_d be the number of solutions to $n(n+m) \equiv 0 \pmod{d}$ in $\mathbb{Z}/d\mathbb{Z}$. Then

$$|\mathcal{A}_d| = \frac{\nu_d}{d} x + O(1)$$

and for prime p,

$$\nu_p = \begin{cases} 1 & p \mid m \\ 2 & p \nmid m \end{cases}$$

Consequently, by the fundamental lemma of sieve theory [12, Theorem 2.2], there exists some $A > 0$ such that for $z = x^A$,

$$\begin{aligned} S(\mathcal{A}, \mathcal{P}, z) &\ll x \prod_{2<p<z} \left(1 - \frac{\nu_p}{p}\right) = x \prod_{\substack{2<p<z \\ p\mid m}} \frac{p-1}{p-2} \prod_{2<p<z} \left(1 - \frac{2}{p}\right) \\ &\ll \frac{x}{\log^2 x} \prod_{\substack{p\mid m \\ p>2}} \left(1 + \frac{1}{p-2}\right) \ll \frac{x}{\log^2 x} \prod_{p\mid m} \left(1 + \frac{1}{p}\right). \end{aligned}$$

Plugging this into (3.7), (3.6), and (3.5), we obtain

$$\begin{aligned} \frac{\log^2 x}{x} T_1 &\ll \sum_{m\in I} \prod_{p\mid m} \left(1 + \frac{1}{p}\right) \le \sum_{m\in I} \sum_{d\mid m} \frac{1}{d} \\ &= \sum_{d \le (1+\delta)\log 2x} \frac{1}{d} \sum_{\substack{m\in I \\ d\mid m}} 1 \ll \sum_{d \le (1+\delta)\log 2x} \frac{\delta \log x}{d^2} + \sum_{d \le (1+\delta)\log 2x} \frac{1}{d} \\ &\ll \delta \log x + \log\log x \ll \delta \log x. \end{aligned}$$

Having completed the estimation of T_1, we proceed to the final computations.

3.7 Lower bound for S

Plugging our conclusions in the previous section into (3.4), we see that for some large $A > 0$, one has

$$S > (1 + \delta - A\delta^2)x = [1 + \delta(1 - A\delta)]x.$$

Thus, if $\delta < A^{-1}$, the right hand side will be $> x$, creating a contradiction to $S \le x$. There for there exists some $\eta > 0$ such that

$$\liminf_{n\to\infty} \frac{p_{n+1} - p_n}{\log p_n} \le 1 - \delta \tag{3.8}$$

3.8 Conclusion

In this article, we improved the PNT bound (3.1) to (3.8) by introducing sieve methods. In 1954, Ricci showed that (3.8) holds for $\eta \geq \frac{1}{16}$. In 1965, by replacing the GRH assumption with the Bombieri–Vinogradov theorem in the Hardy–Littlewood argument, Bombieri and Davenport [2] showed that $\eta \geq \frac{1}{2}$. During the second half of the 20th century, these bounds were improved by Pilt'ai, Uchiyama, Huxley, Maier, and others. Eventually in 2009, Goldston, Pintz, and Yıldırım [11] settled the question by establishing

$$\liminf_{n\to\infty} \frac{p_{n+1} - p_n}{\log p_n} = 0 \tag{3.9}$$

via a deft sieve design.

Not only did the method of Goldston, Pintz, and Yıldırım produce (3.9), but it also demonstrated that under a hypothesis on the regularity of the distribution of primes in arithmetic progressions, there exist infinitely many pairs of primes whose distance is bounded by 16. Their methods will be expounded in the next few articles of the series. Please stay tuned for updates!

July 6, 2022

4 Primes in tuples and the GPY sieve

In the previous articles, we have investigated the prime gap $p_{n+1} - p_n$ via both analytic and elementary methods. In the next few articles, we focus on a work that made extensive use of both elementary and analytic methods — the GPY sieve, which is named after Goldston, Pintz, and Yıldırım who authored the 2009 paper [11]. Their work indicated that under a certain assumption on the distribution of primes in arithmetic progressions, there exist infinitely many pairs of primes with bounded distance.

In the present article, we discuss the motivation of the GPY sieve from a historical perspective. The technicals will be deferred to future articles. From now on, let us turn our focus back to the 20th century.

4.1 Prime k-tuple conjecture

Let $\mathcal{H} = \{h_1, h_2, \ldots, h_k\}$ be a set of integers. Then we can formulate a conjecture as follows:

Conjecture 4.1 (Naive prime k-tuple)**.** *For all $\mathcal{H}$, there exist infinitely many n such that each $n + h_i$ is prime for $1 \leq i \leq k$.*

The term "naive" is added because we can easily come up with a counterexample. Take $\mathcal{H} = \{1, 2, \cdots, k-1, k\}$. Then it is clear that for each $n \in \mathbb{Z}$, at least one of $n + j$ is divisible by k, so they cannot all be primes. Consequently, we need to impose some restrictions on $\mathcal{H}$ to make the conjecture more plausible.

Admissible k-tuple By generalizing our previous counterexample, we see that Conjecture 4.1 is false as long as we can find some prime p such that for every n, p divides some $n + h_i$. In other words, let

$$Q(n) = (n + h_1)(n + h_2) \cdots (n + h_k).$$

Then the conjecture is false if $Q(n)$ is always divisible by a fixed prime p.

Now, let ν_p be the number of $n \in \mathbb{Z}/p\mathbb{Z}$ such that

$$Q(n) \equiv 0 \pmod{p}.$$

Then the condition above is equivalent to $\nu_p = p$ for some p. For $\mathcal{H}$ to satisfy Conjecture 4.1, it is thus necessary that $\forall p, \nu_p < p$. Therefore, we say $\mathcal{H}$ is an **admissible k-tuple** if $\forall p, \nu_p < p$.

Remark. $\mathcal{H}$ is admissible as long as $\nu_p < p$ for all $p \leq k$.

Based on the analyses above, Hardy and Littlewood [13, p. 61] conjectured the following:

Conjecture 4.2 (Hardy–Littlewood prime k-tuple). *Let $\mathcal{H}$ be an admissible k-tuple. Then there exist infinitely many n such that each $n + h_i$ is a prime for $1 \leq i \leq k$. Moreover, as $x \to \infty$,*

$$\#\{n \leq x : n + h_i \text{ prime}, 1 \leq i \leq k\} \sim H \frac{x}{\log^k x},$$

in which

$$H = \prod_p \left(1 - \frac{\nu_p}{p}\right)\left(1 - \frac{1}{p}\right)^k.$$

If $\mathcal{H}$ is not admissible, then $H = 0$, so if Conjecture 4.2 is valid, then Conjecture 4.1 holds if and only if $\mathcal{H}$ is admissible.

4.2 Prime k-tuples, twin primes, and prime gap

By sieve methods, one can easily show that for each k there is some C_k such that for each admissible k-tuple $\mathcal{H}$, there exists infinitely many n such that each $n + h_i$ is a product of at most C_k primes. In particular, when $k = 2$, it follows from the method of Jingrun Chen [3] that

Theorem 4.1 (Chen, 1973). *For each even h, there exist infinitely many primes p such that $p + h$ is either a prime or a product of two primes.*

However, the prime k-tuple conjecture has more inspirations. If for a fixed admissible $\mathcal{H}$, we can find infinitely many n such that at least two members of $n + h_1, n + h_2, \ldots, n + h_k$ are primes, then we find infinitely many pairs of primes with bounded distance:

$$\liminf_{n\to\infty}(p_{n+1} - p_n) \leq \max_{1\leq i<j\leq k} |h_i - h_j|. \tag{4.1}$$

It is this inequality that makes Goldston, Pintz, and Yılıdırım construct their seminal sieve. Having analyzed the principles, we turn to computations.

4.3 Weighted sums

Abstractly, for sets $X, Y \subset \mathbb{Z}$, to show that they have a non-empty intersection, one direct approach is to prove $|X \cap Y| > 0$. To achieve this, it is also helpful to introduce weights. Let $w : X \to \mathbb{R}$ be such that

$$w(n) \begin{cases} > 0 & n \in Y, \\ \leq 0 & n \notin Y. \end{cases}$$

Then

$$S = \sum_{n\in X} w(n) > 0$$

is a sufficient condition to $|X \cap Y| > 0$.

Now, we apply this philosophy to the problem of prime gaps. Let $\chi_{\mathbb{P}}$ be the characteristic function for primes. Then, based on the analyses in the previous section, we define

$$S(N) = \sum_{1 \le n \le N} \left(\sum_{1 \le i \le k} \chi_{\mathbb{P}}(n + h_i) - 1 \right) \tag{4.2}$$

Therefore, to establish (4.1), it suffices to show that $\lim_{N \to +\infty} S(N) = +\infty$.

Nevertheless, according to the prime number theorem, $\sum_{n \le N} \chi(n) \sim N/logN$, making (4.2) negative for large N, so (4.2) cannot help us investigate prime gaps. Regardless, Goldston, Pintz, and Yıldırım did not give up and decided to take in some ideas from Selberg.

4.4 Weighted Selberg sieve

In 1947, Selberg developed a powerful sieve [19] based on the non-negativity of squares. By incorporating Selberg's idea into the picture, we see that when $\{\lambda_d\}$ is a real sequence such that $\lambda_1 = 1$, the following

$$S'(N) = \sum_{n \le N} \left(\sum_{1 \le i \le k} \chi_{\mathbb{P}}(n + h_i) - 1 \right) \left(\sum_{d | Q(n)} \lambda_d \right)^2 \tag{4.3}$$

diverging to positive infinity can also serve as a sufficient condition to (4.1). When n is small, there would be some computational complications concerning the h_i's, so Goldston, Pintz, and Yıldırım replaced the range of the outer summation with a dyadic interval:

$$S''(N) = \sum_{N < n \le 2N} \left(\sum_{1 \le i \le k} \chi_{\mathbb{P}}(n + h_i) - 1 \right) \left(\sum_{d | Q(n)} \lambda_d \right)^2 . \tag{4.4}$$

Thus, (4.1) will directly follow from $S''(N) > 0$.

4.5 Conclusion

In this article, we began by generalizing the twin primes conjecture to prime k-tuple conjectures. Realizing the counterexamples to our naive conjecture Conjecture 4.1, we introduced the notion of admissible tuples and formulated the Hardy–Littlewood conjecture Conjecture 4.2. Although we are unable to prove the conjecture using sieves, through the inequality (4.1), a partial form of this conjecture can lead to significant progress in the prime gap problem.

With this realization in mind, we developed weighted sums and combined ideas from Selberg, resulting in the GPY sieve (4.4).

Although Goldston, Pintz, and Yıldırım's initial choice of λ_d fails to establish bounded gaps between primes unconditionally, analyzing their work is still valuable. Only through in-depth analysis of the GPY sieve can we fathom their limitations and properly appreciate the works of Yitang Zhang and James Maynard. Due to space reasons, these discussions will be presented in subsequent articles. Please stay tuned for updates!

July 24, 2022

5 Elementary transforms and equidistributions

In the previous article, we formulated the GPY sieve using the Hardy–Littlewood k-tuple conjecture.

$$S = \sum_{N<n\le 2N} \left(\sum_{1\le i\le k} \chi_{\mathbb{P}}(n+h_i) - 1 \right) \left(\sum_{d|Q(n)} \lambda_d \right)^2 \tag{5.1}$$

and showed that the existence of infinitely many pairs of primes with bounded gaps will follow from $S > 0$ for all large N. In this article, we elaborate on the computation of S:

$$U = \sum_{N<n\le 2N} \left(\sum_{d|Q(n)} \lambda_d \right)^2 , \tag{5.2}$$

$$V_i = \sum_{N<n\le 2N} \chi_{\mathbb{P}}(n+h_i) \left(\sum_{d|Q(n)} \lambda_d \right)^2 . \tag{5.3}$$

Consequently, if we can compute the asymptotic expansion for U and each V_i, then (5.1) can be studied by $\sum_{1\le i\le k} V_i - U$.

5.1 Preliminary expansion of U

By interchanging the order of summation, (5.2) immediately becomes

$$U = \sum_{d_1,d_2} \lambda_{d_1}\lambda_{d_2} \sum_{\substack{N<n\le 2N \\ [d_1,d_2]|Q(n)}} 1.$$

Let ν_d denote the number of solutions to $Q(n) \equiv 0 \pmod{d}$ in $\mathbb{Z}/d\mathbb{Z}$. Then in each subinterval of $(N, 2N]$ of length d, there are exactly ν_d many n's such that $d|Q(n)$, so

$$U = \sum_{d_1,d_2} \lambda_{d_1}\lambda_{d_2}\nu_{[d_1,d_2]} \left\{ \frac{N}{[d_1,d_2]} + O(1) \right\} = NM_U + E_U. \tag{5.4}$$

Writing $g_U(d) = \nu_d/d$, M_U becomes

$$M_U = \sum_{d_1,d_2} g_U([d_1,d_2])\lambda_{d_1}\lambda_{d_2}. \tag{5.5}$$

Similar to Selberg's sieve, we can assume a priori that $|\lambda_d| \le 1$ and λ_d is supported on square-free integers in $[1, R]$, so we have

$$|E_U| \le \sum_{d_1,d_2\le R} \mu^2(d_1)\mu^2(d_2)\nu_{[d_1,d_2]} \le \sum_{d\le R^2} \mu^2(d)3^{\omega(d)}\nu_d.$$

By the Chinese remainder theorem, ν_d is a multiplicative function of d, so it follows from Rankin's trick that this is

$$\leq R^2 \sum_{p|d \Rightarrow p \leq R^2} \frac{\mu^2(d) 3^{\omega(d)} \nu_d}{d} = R^2 \prod_{p \leq R^2} \left(1 + \frac{3\nu_p}{p}\right).$$

Since $Q(n)$ is a polynomial of degree k, it has at most k roots in $\mathbb{Z}/p\mathbb{Z}$, which indicates that $\nu_p \leq k$ and

$$E_U \ll R^2 \prod_{p \leq R^2} \left(1 + \frac{3k}{p}\right) \leq R^2 \exp\left\{\sum_{p \leq R^2} \frac{3k}{p}\right\} \ll R^2 (\log R)^{3k}. \tag{5.6}$$

Comparing (5.6) to (5.4), we see that R cannot exceed the square root of N for otherwise the error term E_U may exceed the main term NM_U. Thus, the only remaining task for U is the computation of M_U. For now, we transfer our focus to V_i.

5.2 Preliminary expansion of V_i

After interchanging the order of summation, we need to compute

$$\sum_{\substack{N < n \leq 2N \\ Q(n) \equiv 0 (d)}} \chi_{\mathbb{P}}(n + h_i) = \sum_{\substack{1 \leq a \leq d \\ Q(a) \equiv 0(d)}} \sum_{\substack{N - h_i < p \leq 2N - h_i \\ p \equiv a + h_i (d)}} 1. \tag{5.7}$$

If $a + h_i$ is not coprime to d, then the purple term vanishes for all large N, so we impose an extra condition $(a + h_i, d) = 1$ outside. Hence, by the prime number theorem in arithmetic progressions,

$$\begin{aligned} &\sum_{\substack{N < n \leq 2N \\ Q(n) \equiv 0(d)}} \chi_{\mathbb{P}}(n + h_i) \\ &\qquad = \sum_{\substack{1 \leq a \leq d \\ Q(a) \equiv 0(d) \\ (a + h_i, d) = 1}} [\pi(2N; d, a + h_i) - \pi(N; d, a + h_i) + O(1)] \\ &\qquad = \sum_{\substack{1 \leq a \leq d \\ Q(a) \equiv 0(d) \\ (a + h_i, d) = 1}} \left[\frac{1}{\varphi(d)} \int_N^{2N} \frac{\mathrm{d}u}{\log u} + O\{E_1(N, d)\}\right], \end{aligned} \tag{5.8}$$

in which

$$E(x, d) = \max_{(a,d)=1} \left|\pi(x; q, a) - \frac{1}{\varphi(q)} \int_2^x \frac{\mathrm{d}u}{\log u}\right|$$

and $E_1(N, d) = E(N, d) + E(2N, d)$. Since the dependence on $a + h_i$ has been eliminated in (5.8), we can now compute the green sum directly. Set

$$b_i(d) = \#\{1 \leq a \leq d : d | Q(a), (a + h_i, d) = 1\}.$$

Then $b_i(d)$ is multiplicative and $b_i(p) = \nu_p - 1$, so

$$b_i(d) = b(d) = \prod_{p|d} (\nu_p - 1). \tag{5.9}$$

Therefore, (5.8) becomes

$$\sum_{\substack{N<n\leq 2N \\ Q(n)\equiv 0(d)}} \chi_{\mathbb{P}}(n+h_i) = \frac{b(d)}{\varphi(d)} \int_N^{2N} \frac{\mathrm{d}u}{\log u} + O\{b(d)E_1(N,d)\}.$$

Plugging this back into (5.3), we get

$$V_i = M_V \int_N^{2N} \frac{\mathrm{d}u}{\log u} + O(E_V), \tag{5.10}$$

in which when $g_V(d) = b(d)/\varphi(d)$, we have

$$M_V = \sum_{d_1,d_2} g_V([d_1,d_2])\lambda_{d_1}\lambda_{d_2}. \tag{5.11}$$

By Cauchy–Schwarz, E_V becomes

$$\begin{aligned} E_V &\ll \sum_{d\leq R^2} \mu^2(d)3^{\omega(d)}b(d)E_1(N,d) \\ &\leq \left(\sum_{p|d\Rightarrow p\leq R^2} \mu^2(d)9^{\omega(d)}b^2(d)E_1(N,d) \right)^{1/2} \\ &\quad\times \left(\sum_{d\leq R^2} \mu^2(d)E_1(N,d) \right)^{1/2}. \end{aligned} \tag{5.12}$$

From $\pi(x;q,d) \ll x/q$, we know $E_1(x,d) \ll x/d$, so the blue part becomes

$$\begin{aligned} &\sum_{p|d\Rightarrow p\leq R^2} \mu^2(d)9^{\omega(d)}b^2(d)E_1(N,d) \ll N \sum_{p|d\Rightarrow d\leq R^2} \frac{\mu^2(d)9^{\omega(d)}b^2(d)}{d} \\ &= N \prod_{p\leq R^2} \left(1 + \frac{9b^2(p)}{p}\right) = N \prod_{p\leq R^2} \left(1 + \frac{9(\nu_p-1)^2}{p}\right) \\ &\leq N \prod_{p\leq R^2} \left(1 + \frac{9(k-1)^2}{p}\right) \ll N(\log R)^{9(k-1)^2}. \end{aligned} \tag{5.13}$$

As for the brown part, expanding gives

$$\sum_{d\leq R^2} \mu^2(d)E_1(N,d) \leq \sum_{d\leq R^2} E(N,d) + \sum_{d\leq R^2} E(2N,d). \tag{5.14}$$

To estimate the remaining sums on the right, we need a concept known as the level of distribution of primes.

5.3 Equidistribution of primes in arithmetic progressions

From the naive prime number theorem for arithmetic progressions, we know for fixed q and $(a,q)=1$,

$$\pi(x;q,a) \sim \mathrm{li}(x)/\varphi(q),$$

but in many situations, we need to sum over $\pi(x;q,a)$ over q. As a result, we encounter error terms of the form

$$\mathcal{E}(x,Q) = \sum_{q\le Q} \max_{(a,q)=1} \left| \pi(x;q,a) - \frac{\mathrm{li}(x)}{\varphi(q)} \right|.$$

By the Siegel–Walfisz theorem [6, p. 133], we know that for $Q=(\log x)^A$, one has

$$\mathcal{E}(x,Q) \ll_A \frac{x}{\log^A x}. \tag{5.15}$$

This later allowed I. M. Vinogradov [22] to solve the ternary Goldbach problem. In 1948, A. Rényi [18] established the existence of $\theta>0$ such that (5.15) continues to hold for $Q=x^{\theta-\varepsilon}$, allowing him to demonstrate that every large even integer is a sum of a prime and a product of $\le C$ primes for some fixed $C>0$ (a.k.a. proposition $\{1,C\}$). In 1965, A. I. Vinogradov [21] and E. Bombieri [1] independently showed that (5.15) holds when $\theta=\frac{1}{2}$, improving $\{1,C\}$ to $\{1,3\}$. In 1970, Elliot and Halberstam [7] conjectured that (5.15) is valid even if $\theta=1$. As a result, we introduce the following proposition:

Proposition 5.1 ($EH(\theta)$)**.** *For all $\varepsilon>0$ and $Q=x^{\theta-\varepsilon}$, (5.15) holds.*

When this proposition holds, we say that the primes have **level of distribution** θ.

Therefore, the Bombieri–Vinogradov theorem is equivalent to $EH(\frac{1}{2})$ and the Elliot–Halberstam conjecture is the same as $EH(1)$. For the sake of generality, we carry out subsequent computations with an unspecified θ and the assumption of $EH(\theta)$.

According to (5.14), we see that for $R=N^{\frac{\theta}{2}-\varepsilon}$, one has

$$\sum_{d\le R^2} \mu^2(d)E_1(N,d) \ll \frac{N}{\log^A N}.$$

Plugging this with (5.13) into (5.12), we get $E_V \ll_A N\log^{-A} N$, so the error terms are no longer a concern.

5.4 Asymptotic formula for S

Plugging (5.4) and (5.10) into (5.1), we deduce from

$$\int_N^{2N} \frac{\mathrm{d}u}{\log u} \sim \frac{1}{\log N}\int_N^{2N} \mathrm{d}u = \frac{N}{\log N}$$

that

$$S \sim \frac{N}{\log N}(kM_V - M_U \log N). \tag{5.16}$$

Our goal is accomplished as long as the red component is positive for large N. Thus, we should focus on the quadratic forms M_V, M_U.

5.5 Diagonalization of M

According to (5.5) and (5.11), both M_U and M_V can be computed in the same manner, so we let (M, g) denote any one of (M_U, g_U) and (M_V, g_V). Then it follows from the multiplicativity that

$$M = \sum_{d_1,d_2} \frac{1}{g((d_1,d_2))}\lambda_{d_1} g(d_1)\lambda_{d_2} g(d_2).$$

As in the derivation of Selberg's sieve ([12, Chapter 3] or [10, §7.1]), define a multiplicative function $h(d)$ by

$$h(p) = \frac{g(p)}{1-g(p)},$$

so one has

$$\begin{aligned} M &= \sum_{d_1,d_2} g(d_1)\lambda_{d_1} g(d_2)\lambda_{d_2} \sum_{m|(d_1,d_2)} \frac{1}{h(m)} \\ &= \sum_{m\le R} \frac{1}{h(m)} \sum_{\substack{d_1,d_2\le R \\ m|(d_1,d_2)}} g(d_1)\lambda_{d_1} g(d_2)\lambda_{d_2}. \end{aligned}$$

Therefore, when we define the following quantities:

$$h_U(d) = \prod_{p|d} \frac{g_U(p)}{1-g_U(p)} = \prod_{p|d} \frac{\nu_p}{p-\nu_p}, \tag{5.17}$$

$$h_V(d) = \prod_{p|d} \frac{g_V(p)}{1-g_V(p)} = \prod_{p|d} \frac{\nu_p - 1}{p-\nu_p}, \tag{5.18}$$

$$\alpha_m = \sum_{\substack{d\le R \\ m|d}} g_U(d)\lambda_d = g_U(m) \sum_{\substack{n\le R/m \\ (n,m)=1}} g_U(n)\lambda_{nm}, \tag{5.19}$$

$$\beta_m = \sum_{\substack{d\le R \\ m|d}} g_V(d)\lambda_d = g_V(m) \sum_{\substack{n\le R/m \\ (n,m)=1}} g_V(n)\lambda_{nm}, \tag{5.20}$$

we can rewrite M_U in (5.5) and M_V in (5.11) into diagonal forms:

$$M_U = \sum_{m\le R} \frac{\alpha_m^2}{h_U(m)}, \quad M_V = \sum_{m\le R} \frac{\beta_m^2}{h_V(m)}. \tag{5.21}$$

5.6 Conclusion

In this article, we began the investigation from (5.1). By interchanging the order of summation on quadratic forms, we transformed (5.1) to (5.16). By introducing the prime number theorem for arithmetic progressions in the estimation of E_V, we effectively demonstrated the role of the level of distribution in the development of the GPY sieve. Finally, by defining the auxiliary function $h(d)$, we reduce the quadratic forms M_U, M_V in the main term into diagonal forms (5.21).

Now, we have finished the derivations of the elementary part of the GPY sieve that is independent of the choice of λ_d. In the next article, we will introduce a special choice of λ_d and apply complex-analytic methods to obtain asymptotic formulas for α_m and β_m, deriving the analytic part of the GPY sieve. Please stay tuned for updates!

Aug 4, 2022

6 Contour integration and the GPY theorem

In the previous article, we applied elementary methods to transform the GPY sieve (5.1) into

$$S \sim \frac{N}{\log N}(kM_V - M_U \log N), \tag{6.1}$$

in which

$$M_U = \sum_{m \le R} \frac{\alpha_m^2}{h_U(m)}, \quad M_V = \sum_{m \le R} \frac{\beta_m^2}{h_V(m)}, \tag{6.2}$$

$$\alpha_m = \sum_{\substack{d \le R \\ m \mid d}} g_U(d)\lambda_d = g_U(m) \sum_{\substack{n \le R/m \\ (n,m)=1}} g_U(n)\lambda_{nm}, \tag{6.3}$$

$$\beta_m = \sum_{\substack{d \le R \\ m \mid d}} g_V(d)\lambda_d = g_V(m) \sum_{\substack{n \le R/m \\ (n,m)=1}} g_V(n)\lambda_{nm}. \tag{6.4}$$

In this article, we will obtain asymptotic formulas for these quantities. Thus, we need to specify the sieve parameter λ_d.

6.1 GPY's choice of λ_d

Since $Q(n)$ is a polynomial of degree k, one naturally believes that the GPY sieve is a k-dimensional sieve problem. As a result, it is plausible that we can achieve the best result by plugging in the optimal λ_d for the k-dimensional Selberg upper bound sieve:

$$\lambda_d = \mu(d) \left(\frac{\log R/d}{\log R} \right)^k. \tag{6.5}$$

Remark. The actual optimal λ_d in a Selberg sieve problem depends on g and h but is asymptotic to (6.5) [12, Lemma 5.4].

However, the computations of Goldston, Pintz, and Yıldırım suggest that under (6.5) we only have $S \le 0$ even assuming $EH(\theta)$ at $\theta = 1$.

As a result, the authors decided to attack the problem using a sieve of a different dimension. Specifically, they set λ_d to be the optimal parameter for a $(k+\ell)$-dimensional Selberg upper bound sieve:

$$\lambda_d = \mu(d) \left(\frac{\log R/d}{\log R} \right)^{k+\ell}. \tag{6.6}$$

Thus, we proceed to the expansion of α_m, β_m using (6.6).

6.2 Asymptotic expansion of α_m, β_m

According to (6.3) and (6.4), α_m and β_m have very similar structures, so we only elaborate on the computation for α_m, and the reader can use an almost identical argument to treat β_m.

By contour integration, one has

$$\frac{m!}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{x^s}{s^{m+1}}\mathrm{d}s=\begin{cases}(\log x)^m & x>1,\\ 0 & 0<x\le 1.\end{cases} \tag{6.7}$$

Plugging this into (6.3), we get

$$\alpha_m=\frac{\mu(m)g_U(m)}{(\log R)^{k+\ell}}\frac{(k+\ell)!}{2\pi i}\int_{c-i\infty}^{c+i\infty}\underbrace{\sum_{\substack{n\ge 1\\(n,m)=1}}\frac{\mu(n)g_U(n)}{n^s}}_{F_U(s)}\frac{x^s}{s^{k+\ell+1}}\mathrm{d}s, \tag{6.8}$$

where $x=R/m$. By the Euler product formula for Dirichlet series with multiplicative coefficients, $F_U(s)$ can be rewritten into

$$F_U(s)=\prod_{p\nmid m}\left(1-\frac{g_U(p)}{p^s}\right)=\prod_{p\nmid m}\left(1-\frac{\nu_p}{p^{s+1}}\right).$$

Since $\nu_p\le p$, $F_U(s)$ converges absolutely when $\sigma=\Re(s)>0$. To obtain asymptotics for (6.8), we need to analytically continue $F_U(s)$ to a larger domain containing $\sigma=0$.

Analytic continuation of $F_U(s)$ Since $g_U(n)$ does not have good analytic properties, we introduce a certain power of the Riemann ζ-function to offset the poles of $F_U(s)$.

Since $\nu_p<k$ if and only if $Q(n)$ has a repeated root in $\mathbb{Z}/p\mathbb{Z}$, so defining

$$\Delta=\prod_{1\le i<j\le k}|h_i-h_j|,$$

we see that $\nu_p<k$ if and only if $p|\Delta$. Because Δ only depends on $\mathcal{H}$, we see that all but finitely many p satisfies $\nu_p=k$, so when we factor $F_U(s)$ as

$$F_U(s)=\zeta^{-k}(s+1)G_U(s), \tag{6.9}$$

the infinite product

$$G_U(s)=\prod_{p|m}\left(1-\frac{\nu_p}{p^{s+1}}\right)^{-1}\prod_p\left(1-\frac{\nu_p}{p^{s+1}}\right)\left(1-\frac{1}{p^{s+1}}\right)^{-k} \tag{6.10}$$

will be analytic in a region larger than $\sigma > 0$. Using the power series expansion of the logarithm, we see that when $s = \sigma + it$,

$$\log\left(1 - \frac{\nu_p}{p^{s+1}}\right)\left(1 - \frac{1}{p^{s+1}}\right)^{-k} = \frac{\nu_p - k}{p^{s+1}} + O\left(\frac{1}{p^{2\sigma+2}}\right).$$

Notice that $\nu_p = k$ for all large p, so the product for $G_U(s)$ converges absolutely for $\sigma > -\frac{1}{2}$. Therefore, we can move our path of integration to somewhere slightly to the left of $\sigma = 0$. Nevertheless, as in the proof of the prime number theorem, we need to obtain upper bounds for $F_U(s)$ to determine the adequate path of integration.

Upper bounds for $F_U(s)$ According to (6.9), we realize that to get bound $F_U(s)$, we require information from $\zeta(s+1)$. According to the classical theory of ζ-function, there is some $c_0 > 0$ such that in the region

$$\sigma \geq -c_0/\log|t|, |t| \geq 4, \tag{6.11}$$

$\zeta(s+1)$ has an analytic logarithm with the property

$$|\log \zeta(s+1)| \leq \log\log|t| + O(1).$$

Therefore, in the region described by (6.11), one has

$$|F_U(s)| \ll |G_U(s)| \log^k |t|. \tag{6.12}$$

According to (6.9), the task is reduced to bounding the blue product. Set $\delta = \max(-\sigma, 0)$. Then

$$\begin{aligned}
\left|\prod_{p|m}\left(1 - \frac{\nu_p}{p^{s+1}}\right)^{-1}\right| &\ll \prod_{p|m}\left(1 + \frac{kp^\delta}{p}\right) \leq \exp\left\{\sum_{p|m}\frac{kp^\delta}{p}\right\} \\
&\leq \exp\left\{kW^\delta \sum_{p \leq W}\frac{1}{p} + kW^{\delta-1}\frac{\log m}{\log W}\right\} \\
&\ll \exp(kW^\delta \log\log W + kW^{\delta-\frac{1}{2}}\log m).
\end{aligned}$$

Therefore, under the choice $W = \log R$, one has

$$|G_U(s)| \ll \exp(2kW^\delta \log\log W), \tag{6.13}$$

which, plugging into (6.12), implies

$$|F_U(s)| \ll (\log|t|)^k \exp(2kW^\delta \log\log W)$$

is valid in the region (6.11).

Using this information, we continue the computation of (6.8).

Deformation of the path of integration Since the integral in (6.8) involves infinity, we introduce a truncation parameter $2 \le T \le R^{100}$ to turn the path into a line segment, giving us more flexibility.

By the bound (6.13), we know

$$\int_{c-i\infty}^{c+\infty} - \int_{c-iT}^{c+iT} \ll x^c c^{-k} (\log W)^{2k} \int_T^\infty \frac{\mathrm{d}t}{t^{k+\ell+1}} \ll \frac{x^c c^{-k}}{T^{k+\ell}} (\log W)^{2k},$$

so setting $c = 1/\log x$ suggests that the right-hand side is

$$\ll T^{-k-\ell} (\log x)^k (\log W)^{2k}.$$

On the other hand, when $\delta_0 = c_0/\log T$ and $T = R^{c_0}$, it follows from

$$\int_{c-iT}^{-\delta_0 - iT} + \int_{-\delta_0+iT}^{c+iT} \ll \frac{(\log x)^k}{T^{k+\ell}} (\log\log R)^{2k(\log R)^{\delta_0}}$$

and

$$\int_{-\delta_0-iT}^{-\delta_0+iT} \ll x^{-\delta_0} (\log\log R)^{2k(\log R)^{\delta_0}}$$

that

$$\begin{aligned} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} F_U(s) \frac{x^s}{s^{k+\ell+1}} \mathrm{d}s = \frac{1}{2\pi i} \oint_{(0+)} F_U(s) \frac{x^s}{s^{k+\ell+1}} \mathrm{d}s \\ + O((\log\log R)^{2ek}). \end{aligned} \tag{6.14}$$

Now, the remaining task is to compute the residue integral in (6.14).

Evaluation of the residue When $s \to 0$, $\zeta(s+1) \sim s^{-1}$, so $D_U(s) = s^k F_U(s)$ is analytic near $s = 0$. Therefore, for some small $r > 0$, one has

$$\begin{aligned} \frac{1}{2\pi i} \oint_{(0+)} F_U(s) \frac{x^s}{s^{k+\ell+1}} \mathrm{d}s = \frac{D_U(0)}{\ell!} (\log x)^\ell \\ + \sum_{1 \le q \le \ell} \frac{(\log x)^{\ell-q}}{(\ell-q)!} \frac{1}{2\pi i} \oint_{|s|=r} D_U(s) \frac{\mathrm{d}s}{s^{q+1}}. \end{aligned}$$

For the red integral, notice that when $r = 1/\log W$, one has

$$\oint_{|s|=r} D_U(s) \frac{\mathrm{d}s}{s^{q+1}} \ll (\log\log R)^{\ell+5k} \ll_\varepsilon (\log R)^\varepsilon.$$

According to (6.10), we also have

$$\begin{aligned} D_U(0) = G_U(0) &= \prod_{p|m} \frac{\nu_p}{p-\nu_p} \underbrace{\prod_p \left(1 - \frac{\nu_p}{p}\right) \left(1 - \frac{1}{p}\right)^{-k}}_{H} \\ &= \prod_{p|m} \frac{p}{\nu_p} \prod_{p|m} \frac{\nu_p}{p-\nu_p} H = \frac{h_U(m)}{g_U(m)} H. \end{aligned} \tag{6.15}$$

Therefore, combining the equations between (6.14) and (6.15) with (6.8), we deduce that

$$\alpha_m = \mu(m)\frac{h_U(m)H}{(\log R)^{k+\ell}}\frac{(k+\ell)!}{\ell!}\left(\log\frac{R}{m}\right)^{\ell} + O\{(\log R)^{\varepsilon-k-1}\}. \tag{6.16}$$

Similarly, for β_m, one has

$$\beta_m = \mu(m)\frac{h_V(m)H}{(\log R)^{k+\ell}}\frac{(k+\ell)!}{(\ell+1)!}\left(\log\frac{R}{m}\right)^{\ell+1} + O\{(\log R)^{\varepsilon-k}\}. \tag{6.17}$$

6.3 Asymptotic formulas for M_U, M_V

As before, we only give out the details for the computation of M_U from (6.16). The reader can fill in the details for M_V by adapting a similar argument using (6.17).

Plugging (6.16) into (6.2), we get

$$M_U \sim \frac{H^2}{(\log R)^{2k+2\ell}}\left[\frac{(k+\ell)!}{\ell!}\right]^2 (2\ell)!\sum_{m\le R}\mu^2(m)h_U(m)\frac{1}{(2\ell)!}\left(\log\frac{R}{m}\right)^{2\ell}. \tag{6.18}$$

For the green part, it follows from (6.7) that for $c>0$, one has

$$\sum_{m\le R}\mu^2(m)h_U(m)\frac{1}{(2\ell)!}\left(\log\frac{R}{m}\right)^{2\ell} = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} I_U(s)\frac{R^s}{s^{2\ell+1}}\mathrm{d}s, \tag{6.19}$$

where $I_U(s)$ is given by the Dirichlet series

$$I_U(s) = \sum_{m\ge 1}\frac{\mu^2(m)h_U(m)}{m^s} = \zeta^k(s+1)\underbrace{\prod_p\left(1+\frac{h_U(p)}{p^s}\right)\left(1-\frac{1}{p^{s+1}}\right)^k}_{J_U(s)}. \tag{6.20}$$

By reasoning similar to that in the investigation of $G_U(s)$, we conclude that $J_U(s)$ is absolutely convergent for $\sigma > -\frac{1}{2}$. In addition, when s lies in the region described by (6.11),

$$I_U(s) \ll (\log|t|)^k|J_U(s)| = O\{(\log|t|)^k\}. \tag{6.21}$$

Thus, when $2\le T\le R^{100}$, $c = 1/\log R$, and $\delta_0 = c_0/\log T$, we can use (6.21) to rewrite (6.19) into

$$\begin{aligned}
\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} I_U(s)\frac{R^s}{s^{2\ell+1}}\mathrm{d}s &= \frac{1}{2\pi i}\oint_{c-iT}^{c+iT} I_U(s)\frac{R^s}{s^{2\ell+1}}\mathrm{d}s + O\{T^{-2\ell}(\log R)^k\}\\
&= \frac{1}{2\pi i}\oint_{(0+)} I_U(s)\frac{R^s}{s^{2\ell+1}}\mathrm{d}s + O\{T^{-2\ell}(\log R)^k\}\\
&\quad + O\left(\int_{c-iT}^{-\delta_0-iT} + \int_{-\delta_0-iT}^{-\delta_0+iT} + \int_{-\delta_0+iT}^{c+iT}\right).
\end{aligned}$$

By an argument similar to that in the treatments for α_m, the integrals in the O-term are

$$\ll T^{-2\ell}(\log R)^k + R^{-\delta_0}.$$

Hence, we can set $\log T = \sqrt{\log R}$ to conclude that

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} I_U(s)\frac{R^s}{s^{2\ell+1}}\mathrm{d}s = \frac{1}{2\pi i}\oint_{(0+)} I_U(s)\frac{R^s}{s^{2\ell+1}}\mathrm{d}s + O(e^{-\frac{c_0}{2}\sqrt{\log R}}). \tag{6.22}$$

Now, set $K_U(s) = s^k I_U(s)$, so $K_U(s)$ is analytic near $s = 0$ and $K_U(s) = K_U(0) + O(|s|)$, allowing us to compute the brown integral as follows:

$$\begin{aligned}
\frac{1}{2\pi i}\oint_{(0+)} I_U(s)\frac{R^s}{s^{2\ell+1}}\mathrm{d}s &= \frac{1}{2\pi i}\oint_{(0+)} K_U(s)\frac{R^s}{s^{2\ell+k+1}}\mathrm{d}s \\
&= \frac{K_U(0)}{2\pi i}\oint_{|s|=1/\log R}\frac{R^s}{s^{2\ell+k+1}}\mathrm{d}s + O\left(\oint_{|s|=1/\log R}\frac{|\mathrm{d}s|}{|s|^{2\ell+k}}\right) \\
&= \frac{K_U(0)}{(k+2\ell)!}(\log R)^{k+2\ell} + O\{(\log R)^{k+2\ell-1}\}.
\end{aligned} \tag{6.23}$$

According to (6.20), we know

$$\begin{aligned}
K_U(0) = J_U(0) &= \prod_p (1 + h_U(p))\left(1-\frac{1}{p}\right)^k \\
&= \prod_p \left(1-\frac{\nu_p}{p}\right)^{-1}\left(1-\frac{1}{p}\right)^k = \frac{1}{H}
\end{aligned} \tag{6.24}$$

Finally, plugging the equations between (6.19) and (6.24) into (6.18), we see that M_U satisfies

$$M_U \sim \frac{H}{(\log R)^k}\underbrace{\left[\frac{(k+\ell)!}{\ell!}\right]^2\frac{(2\ell)!}{(k+2\ell)!}}_{\gamma(k,\ell)}. \tag{6.25}$$

Using a similar argument, M_V is asymptotic to

$$M_V \sim \frac{H}{(\log R)^{k-1}}\frac{2(2\ell+1)}{(2\ell+k+1)(\ell+1)}\gamma(k,\ell). \tag{6.26}$$

We have now obtained all the asymptotic formulas required by the GPY sieve. It is time to assemble them to study the prime gaps.

6.4 The GPY theorem

Plugging (6.25) and (6.26) into (6.1), we have

$$S \sim \frac{N}{\log N}\frac{H\gamma(k,\ell)}{(\log R)^k}\left\{\frac{2k(2\ell+1)}{(2\ell+k+1)(\ell+1)}\log R - \log N\right\}. \tag{6.27}$$

In the previous article, we have set $R = N^{\frac{\theta}{2}-\varepsilon}$, so one has

$$S = N\frac{H\gamma(k,\ell)}{(\log R)^k}\{P(k,\ell,\theta) - 1 + O(\varepsilon)\},$$

in which

$$P(k,\ell,\theta) := \frac{k(2\ell+1)}{(2\ell+k+1)(\ell+1)}\theta.$$

When $\ell = 0$, we have

$$P(k,0,\theta) = \frac{k}{k+1}\theta < \theta \leq 1,$$

so plugging the k-dimensional sieve parameter into the GPY sieve does not help us establish bounded gaps between primes. Consequently, the parameter ℓ in the work of Goldston, Pintz, and Yıldırım is indispensable.

Notice that

$$\begin{aligned} P(k,\ell,\theta) &= \frac{k(2\ell+1)}{(2\ell+k+1)(2\ell+2)}2\theta \\ &= 2\theta\left(1 - \frac{2\ell+1}{2\ell+k+1}\right)\left(1 - \frac{1}{2\ell+2}\right), \end{aligned}$$

so we have

$$P(k,\ell,\theta) < \lim_{\ell\to\infty}\lim_{k\to\infty} P(k,\ell,\theta) = 2\theta.$$

According to the definition of limit, we know that when $\theta > \frac{1}{2}$, there exists k, ℓ such that $P(k,\ell,\theta) > 1$, so $S > 0$ for large N. Therefore, we obtained the first breakthrough concerning the bounded gaps between primes:

Theorem 6.1 (Goldston, Pintz, and Yıldırım)**.** *If the primes have level of distribution $\theta > \frac{1}{2}$, then there exists some $C(\theta) \geq 2$ such that*

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

6.5 Conclusion

In this article, we estimated $\alpha_m, \beta_m, M_U, M_V$ via contour integration, thereby obtaining the analytic form for the GPY sieve (6.27).

The original derivation of the GPY sieve in their paper [11] did not diagonalize M_U, M_V, so the authors had to interact with double complex integrals. Although Friedlander and Iwaniec diagonalized M_U, M_V in their book *Opera de*

Cribro [10, §7.13], they did not specify λ_d at the beginning but instead determined λ_d from a choice of α_m, which is not intuitive. Therefore, the derivation of the GPY sieve given in the present series is simpler than that of the original paper and more motivating than that in Friedlander and Iwaniec's book.

One may think the obstruction $\theta > \frac{1}{2}$ is caused by the specific choice (6.6). In the next article, we will derive the GPY sieve under a more general choice λ_d

$$\lambda_d = \mu(d) P\left(\frac{\log R/d}{\log R}\right)$$

to further explore the connection between the GPY sieve and the level of distribution θ. Please stay tuned for updates!

Aug 6, 2022

7 Limitation of the GPY sieve

In the last three articles, we have constructed the GPY sieve

$$S = \sum_{N<n\leq 2N} \left(\sum_{1\leq i\leq k} \chi_{\mathbb{P}}(n+h_i) - 1 \right) \left(\sum_{d|Q(n), d\leq R} \lambda_d \right)^2, \tag{7.1}$$

and, by choosing λ_d to be the optimal parameter for the $(k+\ell)$-dimensional Selberg sieve

$$\lambda_d = \mu(d) \left(\frac{\log R/d}{\log R} \right)^{k+\ell}, \tag{7.2}$$

successfully established the existence of infinitely many pairs of primes with bounded distance, provided that the primes are distributed at a level $\theta > \frac{1}{2}$.

In this article, we consider a generalization of the GPY sieve by replacing (7.2) with

$$\lambda_d = \mu(d) P \left(\frac{\log R/d}{\log R} \right), \tag{7.3}$$

where $P(x)$ is a real-valued, 1-bounded smooth function on $[0,1]$ such that $P(1) = 1$ and $P(x) = O(x^k)$ as $x \to 0$. This construction allows us to explore further the relationship between the GPY sieve and the condition $\theta > \frac{1}{2}$. The first person to consider such a generalization is Kannan Soundararajan [20]. Because the main purpose of his paper was to survey the GPY sieve, he only stated the result of plugging (7.3) into (7.1) without proof, so the present article can serve as a supplement to his survey paper.

7.1 A generalized GPY sieve

In §5, we have converted (7.1) into a problem of inequality:

$$kM_V(k,N,\theta) > M_U(k,N,\theta)\log N \quad \forall N \geq N_0, \tag{7.4}$$

where M_U and M_V are quadratic forms of λ_d defined in (5.5) and (5.11).

In §6, we plugged (7.2) into M_U, M_V, which indicated that for each $\theta > \frac{1}{2}$, we can find $k \in \mathbb{N}$ for which (7.4) holds. Thus, our subsequent task is to carry out the computations in (6) with (7.3) instead. As a first step, we compute the asymptotic formulas for α_m, β_m defined in (6.3) and (6.4).

Remark. By Stone–Weierstrass, we assume P is a polynomial.

Asymptotic expansion of α_m, β_m By our assumptions on $P(x)$, there exists a sequence of finite support $\{a_\ell\}_{\ell\geq 0}$ such that

$$P(x) = \sum_{\ell\geq 0} a_\ell x^{k+\ell}, \tag{7.5}$$

so by reusing the computations in §6.2, we have

$$\alpha_m \sim \mu(m)\frac{h_U(m)H}{(\log R)^k}\sum_{\ell\geq 0} a_\ell \frac{(k+\ell)!}{\ell!}\left(\frac{\log R/m}{\log R}\right)^\ell .$$

Using the differentiation rules for power functions, the right-hand side becomes

$$\alpha_m \sim \mu(m)\frac{h_U(m)H}{(\log R)^k} P^{(k)}\left(\frac{\log R/m}{\log R}\right), \tag{7.6}$$

and by similar reasoning,

$$\beta_m \sim \mu(m)\frac{h_V(m)H}{(\log R)^{k-1}} P^{(k-1)}\left(\frac{\log R/m}{\log R}\right). \tag{7.7}$$

Asymptotic formulas for M_U, M_V Let $\{b_m\}_{m\geq 0}$ be a sequence of finite support such that

$$[P^{(k)}(x)]^2 = \sum_{m\geq 0} b_m x^m. \tag{7.8}$$

Then by combining (7.6) with the methods in §6.3, we deduce that

$$\begin{aligned} M_U &\sim \frac{H}{(\log R)^k}\sum_{m\in\mathcal{B}} b_m \cdot \frac{m!}{(k+m)!} \\ &= \frac{H}{(\log R)^k}\sum_{m\in\mathcal{B}} b_m \cdot \frac{\Gamma(m-1)}{\Gamma(k+m-1)} \\ &= \frac{H}{(\log R)^k}\frac{1}{(k-1)!}\sum_{m\in\mathcal{B}} b_m \frac{\Gamma(k)\Gamma(m-1)}{\Gamma(k+m-1)}. \end{aligned}$$

Finally, using the beta-gamma relation, one finds that

$$M_U \sim \frac{H}{(\log R)^k}\int_0^1 \frac{x^{k-1}}{(k-1)!}[P^{(k)}(1-x)]^2 \mathrm{d}x. \tag{7.9}$$

Similarly, by using (7.7), we have

$$M_V \sim \frac{H}{(\log R)^{k-1}}\int_0^1 \frac{x^{k-2}}{(k-2)!}[P^{(k-1)}(1-x)]^2 \mathrm{d}x. \tag{7.10}$$

With these asymptotic formulas ready, we move on to analyzing the implications of (7.3) in number theory.

7.2 Bounded gaps and integral inequality

Plugging (7.9), (7.10), and $R = N^{\frac{\theta}{2}-\varepsilon}$ into (7.4), we see that (7.4) holds if and only if

$$\int_0^1 \frac{x^{k-2}}{(k-2)!}[P^{(k-1)}(1-x)]^2\mathrm{d}x > \frac{2}{\theta k}\int_0^1 \frac{x^{k-1}}{(k-1)!}[P^{(k)}(1-x)]^2\mathrm{d}x$$

Now, set $Q(x) = P^{(k-1)}(x)$. Then $Q(x)$ is a nonconstant polynomial with $Q(0) = 0$. If Q is given, then P can be determined by repeated integrations. Thus, the existence of P for which (7.4) holds under (7.3) is equivalent to the existence of Q such that

$$\int_0^1 x^{k-2}[Q(1-x)]^2\mathrm{d}x > \frac{2}{\theta k(k-1)}\int_0^1 x^{k-1}[Q'(1-x)]^2\mathrm{d}x. \tag{7.11}$$

The work of Goldston, Pintz, and Yıldırım [11] indicates that when $\theta > \frac{1}{2}$, (7.11) holds for some k under the choice $Q(x) = \frac{(k+\ell)!}{\ell!}x^\ell$, so an interesting question would be whether there is a choice of P that allows (7.11) to hold for some k and some $\theta \le \frac{1}{2}$. If the answer is affirmative, then we can deduce bounded gaps directly from the unconditional Bombieri–Vinogradov theorem.

Unfortunately, $\theta > \frac{1}{2}$ is a necessity to (7.11). In [20], Soundararajan wrote

“ If we set $Q(y) = P^{(k-1)}(y)$, then Q is a polynomial, not identically zero, with $Q(0) = 0$; for such polynomials Q we claim that the unfortunate inequality

$$\int_0^1 \frac{x^{k-2}}{(k-2)!}Q(1-x)^2\mathrm{d}x < \frac{4}{k}\int_0^1 \frac{x^{k-1}}{(k-1)!}Q'(1-x)^2\mathrm{d}x$$

holds. The reader can try her hand at proving this. ”

Now, we give a detailed proof of this inequality.

7.3 Proof of Soundararajan’s inequality

By definition of Q, we know

$$Q(1-x) = Q(1-x) - Q(1-1) = \int_x^1 Q'(1-u)\mathrm{d}u,$$

so by Cauchy–Schwarz, we know for $\alpha > 1$ that

$$\begin{aligned}[Q(1-x)]^2 &\le \int_x^1 u^\alpha[Q'(1-u)]^2\mathrm{d}u\int_x^1 t^{-\alpha}\mathrm{d}u\\ &= \int_x^1 u^\alpha[Q'(1-u)]^2\mathrm{d}u\frac{1-x^{1-\alpha}}{1-\alpha}.\end{aligned}$$

Plugging this into the left-hand side of (7.11), we get

$$\int_0^1 x^{k-2}[Q(1-x)]^2 \mathrm{d}x \le \int_0^1 u^\alpha [Q'(1-u)]^2 \int_0^u x^{k-2} \cdot \frac{1-x^{1-\alpha}}{1-\alpha} \mathrm{d}x\mathrm{d}u$$
$$= \int_0^1 u^{k-1}[Q'(1-u)]^2 \underbrace{\frac{1}{\alpha-1}\left[\frac{u}{k-\alpha} - \frac{u^\alpha}{k-1}\right]}_{F(u)} \mathrm{d}u.$$

Therefore, the remaining task is to show that for $u \in [0,1]$,

$$F(u) \le \frac{4}{k(k-1)}.$$

Maximum of F Differentiating gives

$$(\alpha-1)F'(u) = \frac{1}{k-\alpha} - \frac{\alpha u^{\alpha-1}}{k-1}, \quad F''(u) = -\frac{\alpha(\alpha-1)u^{\alpha-2}}{(\alpha-1)(k-1)} < 0,$$

so F attains its maximum at all zeros of F'. After case-by-case analysis, we conclude that

$$F_{max} = \begin{cases} [\alpha(k-\alpha)]^{-1} & k > \alpha+1, \\ [(k-\alpha)(k-1)]^{-1} & k \le \alpha+1. \end{cases}$$

Intuitively, this maximum is minimized when α and $k-\alpha$ are close to each other. Plugging in $\alpha = \frac{1}{2}(k+1)$, we have

$$\begin{cases} k > \alpha+1 \Rightarrow F_{max} = 4/(k+1)(k-1) < 4/k(k-1), \\ k \le \alpha+1 \Rightarrow k = 2 \Rightarrow F_{max} = 2 = 4/2(2-1). \end{cases}$$

Finally, using the continuity of F and $F(0) = 0$, we obtain the strict inequality in Soundararajan's paper:

Theorem 7.1 (Soundararajan, 2007)**.** *Let $Q : [0,1] \to \mathbb{R}$ be nonconstant and continuously differentiable such that $Q(0) = 0$. Then for $k \ge 2$,*

$$\int_0^1 x^{k-2}[Q(1-x)]^2 \mathrm{d}x < \frac{4}{k(k-1)} \int_0^1 x^{k-1}[Q'(1-x)]^2 \mathrm{d}x.$$

Combining this theorem with (7.11), we see that $\theta > \frac{1}{2}$ is sufficient and necessary for the GPY sieve to establish bounded gaps between primes.

7.4 Conclusion

In this article, we generalized the GPY sieve by replacing the choice of λ_d with (7.3), thereby converting the prime gap problem into an inequality (7.11) concerning $Q(x) = P^{(k-1)}(x)$. Finally, by differential calculus and Cauchy–Schwarz inequality, we proved that $\theta > \frac{1}{2}$ is a necessary and sufficient condition for (7.11) to hold for some Q.

So far, we realize that the prototypical GPY sieve is incapable of demonstrating bounded gaps between primes unconditionally. Our analysis also indicated two directions to address this limitation:

1. Improving Bombieri–Vinogradov theorem: This is exactly how Yitang Zhang [23] got $\liminf_{n\to\infty}(p_{n+1} - p_n) < 7 \times 10^7$.
2. Changing the structure of λ_d: By making λ_d depend on more variables, James Maynard [14] obtained $\liminf_{n\to\infty}(p_{n+1} - p_n) \leq 600$.

These two results will be expounded in subsequent articles. Please stay tuned for updates!

Aug 28, 2022

8 The 70 million bound of Zhang

Through the analysis of the previous article, we found that even if we replace (7.2) with the more general (7.3), the propotypical GPY sieve is capable of producing bounded gaps between primes only under the assumption $\theta > \frac{1}{2}$ and have also indicated two possible ways to address this limitation. Today, we introduce the first approach that Yitang Zhang [23] took. He proved in 2014 that

Theorem 8.1 (Zhang, 2014)**.** *Denote by p_n the n'th prime. Then*

$$\liminf_{n\to\infty}(p_{n+1} - p_n) < 7 \times 10^7. \tag{8.1}$$

8.1 Smoothed GPY sieve

In the prototypical GPY sieve, we only required λ_d to vanish on $d > R$. Zhang, building on this, required $\lambda_d = 0$ when d has a large prime factor. Since an integer free of prime factors $> z$ is **z-smooth**, Zhang's modified sieve is also known as the smoothed GPY sieve. Now, let $P(z)$ be the product of primes $\leq z$. Then Zhang's sieve can be written as

$$S = \sum_{N<n\leq 2N} \left(\sum_{1\leq i\leq k} \chi_{\mathbb{P}}(n+h_i) - 1 \right) \left(\sum_{d|(P(z),Q(n))} \lambda_d \right)^2, \tag{8.2}$$

in which $h_i, \chi_{\mathbb{P}}(n), Q(n)$ are defined in §4 and λ_d still takes (7.2), the parameter corresponding to the $(k+\ell)$-dimensional sieve. By computations in §5, we see that when

$$g_i(d) = \frac{\mu^2(d)}{d} \prod_{p|d}(\nu_p + 1 - i), \quad h_i(d) = \mu^2(d) \prod_{p|d} \frac{g_i(p)}{1-g_i(p)}, \tag{8.3}$$

$$\alpha_i(m,z) = \sum_{\substack{d\leq R/m \\ d|P(z) \\ (d,m)=1}} g_i(md)\lambda_{md}, \quad M_i = \sum_{\substack{m\leq R \\ m|P(z)}} \frac{\alpha_i^2(m,z)}{h_i(m)}, \tag{8.4}$$

the sieve (8.2) can be rewritten into

$$S = [1+o(1)]\frac{N}{\log N}(kM_2 - M_1 \log N) + O(R^2 \log^{3k} R) + O(\mathcal{E}), \tag{8.5}$$

where

$$\mathcal{E} = \sum_{1\leq i\leq k} \sum_{\substack{d\leq R^2 \\ d|P(z)}} 3^{\omega(d)} \sum_{\substack{1\leq c\leq d \\ Q(c-h_i)\equiv 0(d) \\ (c,d)=1}} |E_1(N;d,c)|, \tag{8.6}$$

$$E_1(N;d,c) = \sum_{\substack{N<p\le 2N\\ p\equiv c(d)}} 1 - \frac{1}{\varphi(d)}\int_N^{2N} \frac{\mathrm{d}u}{\log u}. \tag{8.7}$$

To deduce (8.1) from the smoothed GPY sieve, Zhang showed that under the smoothness condition, the range of Bombieri–Vinogradov theorem can be enlarged.

Theorem 8.2 (Zhang)**.** *When* $\varpi = \frac{1}{1168}$, $R = N^{\frac{1}{4}+\varpi}$, $z = N^{\varpi}$, (8.6) *satisfies* $\mathcal{E} \ll N(\log N)^{-A}$.

Remark. Because the derivation of this result invokes deep results in algebraic geometry with which the author is not familiar, the proof is omitted.

According to the blue term in (8.5), (8.1) will follow if one can find an appropriate k and λ_d such that $kM_2 > M_1 \log N$. In the ordinary GPY sieve, we can compute the asymptotic formulas for M_i directly. Still, due to complications coming from smoothness, we can only do the next best thing: finding an upper bound for M_1 and a lower bound for M_2.

8.2 Preliminary treatments for M_i

Define

$$\alpha_i(m) = \sum_{\substack{d\le R/m\\ (d,m)=1}} g_i(md)\lambda_{md}, \quad M_i^* = \sum_{m\le R} \frac{\alpha_i^2(m)}{h_i(m)}. \tag{8.8}$$

Then these are terms arising from the prototypical GPY sieve, so it follows from computations in §6 that

$$\alpha_i(m) \sim \mu(m)\frac{h_i(m)H}{(\log R)^{k+\ell}}\frac{(k+\ell)!}{(\ell+i-1)!}\left(\log\frac{R}{m}\right)^{\ell+i-1}, \tag{8.9}$$

$$M_1^* \sim \frac{\gamma(k,\ell)H}{(\log R)^k}, \quad M_2^* \sim \frac{2(2\ell+1)}{(2\ell+k+1)(\ell+1)}\frac{\gamma(k,\ell)H}{(\log R)^{k-1}}, \tag{8.10}$$

in which H and $\gamma(k,\ell)$ are defined in (6.15) and (6.25). Because M_i^* and M_i have very similar structures, it is reasonable that M_i should be well approximated by (8.10). Be definition of $P(z)$ and the configuration in Theorem 8.2, we know

$$m > N^{1/4} \wedge m|P(z) \Rightarrow d|P(z) \Rightarrow \alpha_i(m,z) = \alpha_i(m),$$

so one has the following decomposition:

$$|M_i - M_i^*| \le \underbrace{\sum_{m\le N^{1/4}} \frac{\alpha_i^2(m)}{h_i(m)}}_{T_{1i}} + \underbrace{\sum_{\substack{m\le N^{1/4}\\ m|P(z)}} \frac{\alpha_i^2(m,z)}{h_i(m)}}_{T_{2i}} + \underbrace{\sum_{\substack{N^{1/4}<m\le R\\ m\nmid P(z)}} \frac{\alpha_i^2(m)}{h_i(m)}}_{T_{3i}}. \tag{8.11}$$

Therefore, to give an upper bound for (8.11), it suffices to estimate T_{1i}, T_{2i}, T_{3i} individually. We only give out details for the treatments in the $i = 1$ case. The reader should be able to supply the details for the $i = 2$ case.

Remark. Since this article aims to emphasize the main idea, we have deliberately omitted the error terms arising from the asymptotic formulas. A full justification is possible by adapting the methods in §6.

8.3 Estimates of T_{11}, T_{21}, T_{31}

Plugging (8.9) into T_{11}, we have

$$T_{11} \sim \frac{\gamma(k,\ell)H^2}{(\log R)^{2k+2\ell}} \frac{(k+2\ell)!}{(2\ell)!} \sum_{m \le N^{1/4}} h_1(m) \left(\log \frac{R}{m}\right)^{2\ell}. \tag{8.12}$$

By Perron's formula and standard properties of $\zeta(s)$, one has

$$\sum_{m \le x} h_1(m) \sim \frac{(\log x)^k}{k!} \prod_p (1 + h_1(p)) \left(1 - \frac{1}{p}\right)^k = \frac{1}{H} \frac{(\log x)^k}{k!}. \tag{8.13}$$

Therefore, by partial summation, one obtains

$$\begin{aligned} \sum_{m \le N^{1/4}} h_1(m) \left(\log \frac{R}{m}\right)^{2\ell} &\sim \frac{1}{H} \underbrace{\int_1^{N^{1/4}} \left(\log \frac{R}{x}\right)^{2\ell} \frac{(\log x)^{k-1}}{(k-1)!x} \mathrm{d}x}_{u = \log x / \log R} \\ &= \frac{(\log R)^{2\ell+k}}{(k-1)!H} \int_0^{(1+4\varpi)^{-1}} (1-u)^{2\ell} u^{k-1} \mathrm{d}u \\ &< \frac{(\log R)^{2\ell+k}}{(k-1)!H} \int_0^{(1+4\varpi)^{-1}} u^{k-1} \mathrm{d}u. \end{aligned}$$

Now, define

$$\delta_1 = k \int_0^{(1+4\varpi)^{-1}} u^{k-1} \mathrm{d}u = (1+4\varpi)^{-k}, \tag{8.14}$$

so (8.12) becomes

$$T_{11} \le [1 + o(1)] \delta_1 \binom{k+2\ell}{k} \frac{\gamma(k,\ell)H}{(\log R)^k}. \tag{8.15}$$

For T_{21}, we first estimate $\alpha_1(m,z)$. Let P' be the product of primes in $(z, R]$. Then

$$\begin{aligned} \alpha_1(m,z) &= \sum_{\substack{d \le R/m \\ (m,d)=1}} g_1(md) \lambda_{md} \sum_{q \mid (d,P')} \mu(q) \\ &= \sum_{q \mid P'} \mu(q) \sum_{\substack{t \le R/mq \\ (t,mq)=1}} g_1(mqt) \lambda_{mqt} = \sum_{\substack{q \le R/m \\ q \mid P'}} \mu(q) \alpha_1(mq). \end{aligned}$$

When $q \le R$ divides P', it follows from

$$z^{\omega(q)} < q \le R$$

that

$$\omega(q) < \frac{\log R}{\log z} = \frac{1/4 + \varpi}{\varpi} = 293,$$

so we have

$$h_1(q) = g_1(q) \prod_{p|q} \left(1 - \frac{\nu_p}{p}\right)^{-1} = g_1(q) \left\{1 + O\left(\frac{1}{z}\right)\right\}. \tag{8.16}$$

Now, by (8.9), we know

$$|\alpha_1(m, z)| \sim \frac{H}{(\log R)^{k+\ell}} \frac{(k+\ell)!}{\ell!} \sum_{\substack{q \le R/m \\ q|P'}} h_1(mq) \left(\log \frac{R}{mq}\right)^{\ell}.$$

Since $m|P(z)$, m and q are coprime, so $h_1(mq) = h_1(m)g_1(q)$. Combining this with (8.16), we deduce that

$$\begin{aligned}
|\alpha_1(m, z)| &\sim \frac{h_1(m)H}{(\log R)^{k+\ell}} \frac{(k+\ell)!}{\ell!} \sum_{\substack{q \le R/m \\ q|P'}} g_1(q) \left(\log \frac{R}{mq}\right)^{\ell} \\
&\le \frac{h_1(m)H}{(\log R)^{k+\ell}} \frac{(k+\ell)!}{\ell!} \left(\log \frac{R}{m}\right)^{\ell} \sum_{\substack{q \le R \\ q|P'}} g_1(q).
\end{aligned}$$

For the green term, it follows from $\omega(q) \le 292$ that

$$\begin{aligned}
\sum_{\substack{q \le R \\ q|P'}} g_1(q) &= \sum_{0 \le v \le 292} \sum_{\substack{q|P' \\ \omega(q)=v}} g_1(q) \le \sum_{0 \le v \le 292} \frac{1}{v!} \left(\sum_{z<p\le R} g_1(p)\right)^{v} \\
&\le \sum_{0 \le v \le 292} \frac{1}{v!} \left(\sum_{z<v\le R} \frac{k}{p}\right)^{v} \sim \sum_{0 \le v \le 292} \frac{(k \log 293)^v}{v!} =: \delta_2.
\end{aligned} \tag{8.17}$$

Comparing this to (8.9), we conclude that

$$|\alpha_i(m, z)| \le [\delta_1 + o(1)]\alpha_i(m).$$

Plugging this into T_{21}, we get

$$T_{21} \le [1 + o(1)]\delta_2^2 \underbrace{\sum_{d \le N^{1/4}} \frac{\alpha_1^2(m)}{h_1(m)}}_{T_{11}} \le [1 + o(1)]\delta_1\delta_2^2 \binom{k+2\ell}{k} \frac{\gamma(k,\ell)H}{(\log R)^k}. \tag{8.18}$$

For T_{31}, plugging (8.9) in gives

$$T_{31} \leq \frac{\gamma(k,\ell)H^2}{(\log R)^{2k}} \frac{(k+2\ell)!}{(2\ell)!} \sum_{\substack{m\leq R\\ m\nmid P(z)}} h_1(m). \tag{8.19}$$

For the purple part, because $m \nmid P(z) \Rightarrow (m, P') > 1$, it is bounded by

$$\begin{aligned}\sum_{\substack{m\leq R\\ m\nmid P(z)}} h_1(m) &\leq \sum_{m\leq R} h_1(m) \sum_{p|(m,P')} 1 = \sum_{z<p\leq R} \sum_{\substack{m\leq R\\ p|m}} h_1(m)\\ &= \sum_{z<p\leq R} h_1(p) \sum_{\substack{t\leq N^{1/4}\\ (t,p)=1}} h_1(t).\end{aligned}$$

For large prime p, $h_1(p) \leq \frac{k}{p-k} \sim \frac{k}{p}$, so combining with (8.13), one has

$$\sum_{\substack{m\leq R\\ m\nmid P(z)}} h_1(m) \leq [1+o(1)]k\log 293 \sum_{t\leq N^{1/4}} h_1(t) \sim \frac{\delta_1 k\log 293}{k!}\frac{(\log R)^k}{H}.$$

Finally, plugging this into (8.19) gives

$$T_{31} \leq [1+o(1)]k\log 293 \binom{k+2\ell}{k} \frac{\gamma(k,\ell)H}{(\log R)^k}. \tag{8.20}$$

Having estimated T_{11}, T_{21}, T_{31}, it is time to synthesize these results to produce an upper bound for M_1.

8.4 Bounds for M_1 and M_2

Plugging (8.15), (8.18), and (8.20) into (8.11), we realize that under

$$\kappa_1 = \delta_1(1+\delta_2^2+k\log 293)\binom{k+2\ell}{k}, \tag{8.21}$$

we have

$$M_1 \leq [1+\kappa_1+o(1)]\frac{\gamma(k,\ell)H}{(\log R)^k} =: M_1'. \tag{8.22}$$

By an argument similar to that in §8.3, we can obtain upper bounds for T_{12}, T_{22}, T_{32} to obtain a lower bound for M_2. That is, under

$$\kappa_2 = \delta_1(1+4\varpi)(1+\delta_2^2+k\log 293)\binom{k+2\ell+1}{k-1}, \tag{8.23}$$

we have

$$M_2 \geq [1+o(1)]\frac{1-\kappa_2}{1+\kappa_1}\frac{2(2\ell+1)\log R}{(2\ell+k+1)(\ell+1)}M_1'. \tag{8.24}$$

8.5 Bounded gaps between primes

Now, let's see where the number 7×10^7 comes from.

According to (8.22) and (8.24), we see that the blue term in (8.5) satisfies

$$kM_2 - M_1 \log N \geq [s - 1 + o(1)]M_1' \log N,$$

where

$$s = \frac{1-\kappa_2}{1+\kappa_1} \frac{k(2\ell+1)(1+4\varpi)}{(2\ell+k+1)(2\ell+2)}. \tag{8.25}$$

Hence, we can win by choosing k, ℓ that makes $s > 1$. By Stirling's approximation, Zhang showed that when

$$k = 3.5 \times 10^6, \quad \ell = 180, \quad \varpi = \frac{1}{1168},$$

one has $0 < \kappa_1 < e^{-1200}, 0 < \kappa_2 < e^{20}\kappa_1$, so by numerical computation,

$$s > \frac{1-\kappa_2}{1+\kappa_1} \times 1.0005 > \frac{1-e^{-1980}}{1+e^{-1200}} \times (1+e^{-8}) > 1.$$

According to the scheme in §4, the remaining task is to find an admissible tuple $\mathcal{H} = \{h_1, h_2, \ldots, h_k\}$ of size $k = 3.5 \times 10^6$ so that

$$\liminf_{n\to\infty}(p_{n+1} - p_n) \leq \max_{1\leq i<j\leq k} |h_i - h_j| \tag{8.26}$$

Admissible tuple of size k Let $h_1 < h_2 \cdots < h_k$ be primes $\geq k$. Then for $p > k$, $\nu_p \leq \deg Q = k < p$. Because

$$p \leq k \Rightarrow Q(0) = h_1 h_2 \cdots h_k \not\equiv 0 \pmod{p},$$

we see that $\nu_p < p$ for $p \leq k$ as well. Hence, $\mathcal{H} = \{h_1, \ldots, h_k\}$ is admissible.

This reasoning tells us that when $V > k$ is a number satisfying

$$\pi(V) - \pi(k),$$

we can choose $h_1, \ldots, h_k$ such that

$$\max_{1\leq i<j\leq k} |h_i - h_j| = h_k - h_1 < V.$$

Because it is inefficient to count the number of primes directly, we invoke a quantitative form of the prime number theorem due to Rosser and Schoenfeld [17, p. 69]: for $x \geq 60$,

$$\frac{x}{\log x} < \pi(x) < \frac{x}{\log x}\left(1 + \frac{2}{\log x}\right).$$

As a result, when $V = 7 \times 10^7$, one has

$$\begin{aligned}\pi(V) - \pi(k) &> \frac{7 \times 10^7}{7\log 10 + \log 7} - \frac{3.5 \times 10^6}{6\log 10 + \log 3.5} \\ &\times \left(1 + \frac{2}{6\log 10 + \log 3.5}\right) \\ &> 3.8 \times 10^6 - 2.4 \times 10^5 \times 1.2 \\ &> 3.5 \times 10^6 = k.\end{aligned}$$

Finally, combining this with (8.26) and (8.5), we obtain Theorem 8.1.

8.6 Conclusion

In this article, we made unconditional the result of Goldston, Pintz, and Yıldırım by smoothing the GPY sieve. Although we did not obtain asymptotic formulas for M_1 and M_2 under the smoothness assumption, we obtained a nice upper bound for M_1 and a lower bound for M_2 by relating them to the corresponding terms M_1^*, M_2^* in the unsmoothed sieve. Finally, we determined a possible k for $s > 1$ to hold, and by numerical computation with the Rosser–Schoenfeld prime number theorem, we deduce the inequality (8.1).

Although Zhang's paper appeared in publication in 2014, it had already shocked the mathematical community in April 2013. However, if Zhang were late for a few more months, his name would not be as well-known as today. Because in November 2013, James Maynard announced something much stronger:

$$\liminf_{n\to\infty}(p_{n+1} - p_n) \le 600. \tag{8.27}$$

According to his paper [14], Maynard began his investigation into prime gaps before Zhang's announcement of (8.1) and used an entirely different approach. By replacing λ_d with $\lambda_{d_1,d_2,\ldots,d_k}$ that takes in vectors, Maynard deduced his bound (8.27) via only the Bombieri–Vinogradov theorem ($\theta = \frac{1}{2}$). His methods will be expounded in the subsequent articles.

Sept 7, 2022

9 Maynard's dimensional reduction strike I

After Yitang Zhang, many mathematicians became interested in the problem of prime gaps. In the autumn of 2013, the Polymath8 project initiated by Terence Tao and others improved Zhang's 70 million bound to

$$\liminf_{n\to\infty}(p_{n+1}-p_n) \le 4680. \tag{9.1}$$

In November of the same year, by making structural refinements to the GPY sieve, James Maynard [14] replaced the right-hand side of (9.1) with 600. In addition, he showed that there exists some fixed $C > 0$ such that for all $m \in \mathbb{N}$,

$$\liminf_{n\to\infty}(p_{n+m}-p_n) < Cm^3e^{4m}.$$

Zhang obtained his 70 million bound by showing that the primes are equidistributed in arithmetic progressions with smooth moduli at the level of $\theta = \frac{1}{2} + \frac{1}{584}$. In contrast, Maynard's 600 bound is only a consequence of the classical Bombieri–Vinogradov theorem (i.e. $\theta = \frac{1}{2}$ for all moduli). This is because the latter took the second route indicated in §7.4. Now, we will see how Maynard achieved this refinement.

9.1 GPY sieve and the dimensional reduction strike

In the study of Goldston–Pintz–Yıldırım and Zhang, the GPY sieve took the form of

$$S = \sum_{n\in I}\left(\sum_{1\le m\le k}\chi_{\mathbb{P}}(n+h_m)-\rho\right)\underbrace{\left(\sum_{d\in\mathcal{D}_n}\lambda_d\right)^2}_{w_n^2}, \tag{9.2}$$

where I is some interval of integers, λ_d is the Selberg sieve parameter for the $(k+\ell)$-dimensional sieve, and $\rho = 1$. The original sieve problem is naturally a k-dimensional problem, so the heart of the GPY–Zhang approach is to first convert a low-dimensional problem to a high-dimensional version and then solve the problem via a higher-dimensional sieve, which we call the "dimensional increment strike." On the contrary, Maynard's idea is more like a "dimensional reduction strike":[1] He replaced w_n in (9.2) with

$$w_n = \sum_{\substack{d_1,\ldots,d_k \\ d_i|(n+h_i)\forall i}} \lambda_{d_1,\ldots,d_k}, \tag{9.3}$$

allowing us to divide a conquer a k-dimensional sieve problem by k one-dimensional sieves. It is this reason that allows Maynard to deduce bounded gaps between primes without the knowledge beyond Bombieri–Vinogradov.

[1] These terms are borrowed from the science fiction *Three-Body Problem* by Cixin Liu.

Remark. In the GPY–Zhang sieve, we required λ_d to vanish when $d > R$ or d is not squarefree, so we impose a similar constraint for the new parameter in (9.3): i.e. $\lambda_{d_1,\ldots,d_k}$ vanishes when $d_1 d_2 \cdots d_k$ is not squarefree or $d_1 d_2 \cdots d_k > R$.

9.2 Construction of the dimensional reduction sieve

Combining (9.2) with (9.3), we realize that to obtain asymptotics for S, it suffices to compute $S_1, S_2^{(m)}$:

$$S_1 = \sum_{n \in I} w_n^2, \quad S_2^{(m)} = \sum_{n \in I} \chi_{\mathbb{P}}(n + h_m) w_n^2. \tag{9.4}$$

In the GPY–Zhang sieve, $I = (N, 2N]$, but in Maynard's version, for technical convenience, we want each $n \in I$ to have the property that $n + h_m$ is free of small prime factors. Specifically, we want n to satisfy

$$D_0 = \log\log\log N, \quad W = \prod_{p \le D_0} p, \quad (n + h_m, W) = 1 \quad \forall n \in I \quad \forall m. \tag{9.5}$$

Remark. We leave D_0 unspecified until we compute the main term of $S_2^{(m)}$.

Because we also require $\mathcal{H} = \{h_1, \ldots, h_k\}$ to be admissible, for each $p \le D_0$ there is some $0 \le a_p < p$ such that

$$Q(a_p) = (a_p + h_1) \cdots (a_p + h_k) \not\equiv 0 \pmod{p} \quad \forall p \le D_0. \tag{9.6}$$

By the Chinese remainder theorem, there is some $0 \le v_0 < W$ such that

$$v_0 \equiv a_p \pmod{p} \quad \forall p \le D_0,$$

so we set

$$I = \{n \in (N, 2N] : n \equiv v_0 \pmod{W}\}$$

in (9.4) to perform subsequent computations.

9.3 Preliminary treatments for S_1

In analytic number theory, miracles happen after interchanging the order of summation. To transform S_1 into an approachable form, we expand w_n^2, so

$$S_1 = \sideset{}{'}\sum_{\substack{d_1,\ldots,d_k \\ e_1,\ldots,e_k}} \lambda_{d_1,\ldots,d_k} \lambda_{e_1,\ldots,e_k} \sum_{\substack{N < n \le 2N \\ n \equiv v_0 (W) \\ [d_i, e_i] \mid (n + h_i) \forall i}} 1, \tag{9.7}$$

in which $\sum'$ sums over terms under the additional requirement that $[d_1, e_1], \ldots, [d_k, e_k], W$ are pairwise coprime. This is because the blue term will otherwise vanish. From the coprimality condition, we can use the Chinese remainder theorem to determine a unique $0 \le v_1 < W[d_1, e_1] \cdots [d_k, e_k]$ such that the range of the blue sum is equivalent to

$$n \equiv v_1 \pmod{W[d_1, e_1] \cdots [d_k, e_k]},$$

allowing us to convert (9.7) into

$$S_1 = \frac{N}{W} \underbrace{\sum_{\substack{d_1,\ldots,d_k \\ e_1,\ldots,e_k}}{}' \frac{\lambda_{d_1,\ldots,d_k}\lambda_{e_1,\ldots,e_k}}{[d_1,e_1]\cdots[d_k,e_k]}}_{Q_1} + E_1. \tag{9.8}$$

In particular, if $|\lambda_{d_1,\ldots,d_k}| \le \lambda_{max}$ and $\tau_k(n)$ is the number of ways to write n as a product of k integers, then the error term E_1 satisfies

$$E_1 \ll \lambda_{max}^2 \left(\sum_{d \le R^2} \mu^2(d)\tau_k(d) \right)^2 \ll \lambda_{max}^2 R^2 (\log R)^{2k}, \tag{9.9}$$

so our remaining task is to transform the quadratic form Q_1 in (9.8).

9.4 Transformation of Q_1

Although our "dimensional reduction" sieve is very different from the original Selberg sieve, we still hope to handle the gigantic term in (9.8). By $[a,b](a,b) = ab$ and the convolution properties of $\varphi(n)$, one has

$$\begin{aligned} Q_1 &= \sum_{\substack{d_1,\ldots,d_k \\ e_1,\ldots,e_k}}{}' \frac{\lambda_{d_1,\ldots,d_k}\lambda_{e_1,\ldots,e_k}}{d_1\cdots d_k e_1\cdots e_k}(d_1,e_1)\cdots(d_k,e_k) \\ &= \sum_{\substack{d_1,\ldots,d_k \\ e_1,\ldots,e_k}}{}' \frac{\lambda_{d_1,\ldots,d_k}\lambda_{e_1,\ldots,e_k}}{d_1\cdots d_k e_1\cdots e_k} \sum_{\substack{u_1,\ldots,u_k \\ u_i|(d_i,e_i)\forall i}} \prod_{1\le i\le k} \varphi(u_i) \\ &= \sum_{u_1,\ldots,u_k} \prod_i \varphi(u_i) \sum_{\substack{d_1,\ldots,d_k \\ e_1,\ldots,e_k \\ u_i|(d_i,e_i)\forall i}}{}' \frac{\lambda_{d_1,\ldots,d_k}\lambda_{e_1,\ldots,e_k}}{d_1\cdots d_k e_1\cdots e_k}. \end{aligned} \tag{9.10}$$

We have required earlier that $\lambda_{d_1,\ldots,d_k}$ to vanish when $d_1 d_2 \ldots d_k$ is squarefree, which is equivalent to saying $d_1,\ldots,d_k$ are individually squarefree and pairwise coprime, so the additional constraints in $\sum'$ are equivalent to $(d_i,e_j)=1$ for all $i \ne j$. Therefore, it follows from Möbius inversion that

$$\begin{aligned} \sum_{\substack{d_1,\ldots,d_k \\ e_1,\ldots,e_k \\ u_i|(d_i,e_i)\forall i}}{}' &= \sum_{\substack{d_1,\ldots,d_k \\ e_1,\ldots,e_k \\ u_i|(d_i,e_i)\forall i}} \sum_{\substack{s_{1,2},\cdots,s_{k,k-1} \\ s_{i,j}|(d_i,e_j)\forall i\ne j}} \prod_{\substack{1\le i,j\le k \\ i\ne j}} \mu(s_{i,j}) \\ &= \sum_{s_{1,2},\cdots,s_{k,k-1}} \prod_{\substack{1\le i,j\le k \\ i\ne j}} \mu(s_{i,j}) \sum_{\substack{d_1,\ldots,d_k \\ e_1,\ldots,e_k \\ u_i|(d_i,e_i)\forall i \\ s_{i,j}|(d_i,e_j)\forall i\ne j}} . \end{aligned} \tag{9.11}$$

Plugging this into (9.10), we get

$$Q_1 = \sum_{u_1,\dots,u_k} \prod_i \varphi(u_i) \sum_{s_{1,2},\cdots,s_{k,k-1}} \prod_{\substack{i,j\\ i\neq j}} \mu(s_{i,j}) \sum_{\substack{d_1,\dots,d_k\\ e_1,\dots,e_k\\ u_i|(d_i,e_i)\forall i\\ s_{i,j}|(d_i,e_j)\forall i\neq j}} \frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{d_1\cdots d_k e_1\cdots e_k}$$
$$= \sum_{u_1,\dots,u_k} \prod_i \varphi(u_i) \sum_{\substack{s_{1,2},\dots,s_{k,k-1}\\ (s_{i,j},W)=1\forall i\neq j}} \prod_{\substack{i,j\\ i\neq j}} \mu(s_{i,j}) \sum_{\substack{d_1,\dots,d_k\\ e_1,\dots,e_k\\ a_i|d_i, b_i|e_i\forall i}} \frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{d_1\cdots d_k e_1\cdots e_k}, \tag{9.12}$$

in which $a_j = u_j \prod_{\substack{i\\ i\neq j}} s_{j,i}$ and $b_j = u_j \prod_{\substack{i\\ i\neq j}} s_{i,j}$. To prevent (9.12) from becoming more formidable, we make some simplifications in the red sum.

Diagonalization and the parameter $y_{r_1,\dots,r_k}$ Define

$$\alpha_{r_1,\dots,r_k} = \sum_{\substack{d_1,\dots,d_k\\ r_i|d_i\forall i}} \frac{\lambda_{d_1,\dots,d_k}}{d_1\cdots d_k}. \tag{9.13}$$

Then by Möbius inversion, one has

$$\frac{\lambda_{d_1,\dots,d_k}}{d_1\cdots d_k} = \sum_{\substack{r_1,\dots,r_k\\ d_i|r_i\forall i}} \prod_i \mu\left(\frac{r_i}{d_i}\right) \alpha_{r_1,\dots,r_k}$$
$$= \prod_i \mu(d_i) \sum_{\substack{r_1,\dots,r_k\\ d_i|r_i\forall i}} \prod_i \mu(r_i)\alpha_{r_1,\dots,r_k}.$$

Building on (9.13), we introduce an extra definition for the sake of cleanliness:

$$y_{r_1,\dots,r_k} = \mu(r_1)\varphi(r_1)\cdots\mu(r_k)\varphi(r_k)\alpha_{r_1,\dots,r_k},$$

so we have

$$y_{r_1,\dots,r_k} = \prod_i \mu(r_i)\varphi(r_i) \sum_{\substack{d_1,\dots,d_k\\ r_i|d_i\forall i}} \frac{\lambda_{d_1,\dots,d_k}}{d_1\cdots d_k} \tag{9.14}$$

and

$$\lambda_{d_1,\dots,d_k} = \prod_i d_i\mu(d_i) \sum_{\substack{r_1,\dots,r_k\\ d_i|r_i\forall i}} \frac{y_{r_1,\dots,r_k}}{\prod_i \varphi(r_i)}. \tag{9.15}$$

Remark. From (9.14), we see that $y_{r_1,\dots,r_k}$ vanishes when $r_1r_2\cdots r_k$ exceeds R, has square factor, or not coprime to W. Additionally, we see that redefining $\lambda_{d_1,\dots,d_k}$ by (9.15) with the aforementioned constraints on the support of $y_{r_1,\dots,r_k}$ also ensures the support of $\lambda_{d_1,\dots,d_k}$ to fulfilled the conditions mentioned in an earlier remark.

Now, plugging (9.14) into (9.12) gives

$$Q_1 = \sum_{u_1,\ldots,u_k} \prod_i \frac{\mu^2(u_i)}{\varphi(u_i)} \sum_{\substack{s_{1,2},\ldots,s_{k,k-1} \\ (s_{i,j},W)=1 \forall i\neq j}} \prod_{\substack{i,j \\ i\neq j}} \frac{\mu(s_{i,j})}{\varphi^2(s_{i,j})} y_{a_1,\ldots,a_k} y_{b_1,\ldots,b_k}. \tag{9.16}$$

According to (9.5), the condition $(s_{ij}, W) = 1$ implies either $s_{i,j} = 1$ or $s_{i,j} > D_0$. Let Q_1' be the subcollection of terms in Q_1 with $s_{i,j} > D_0$. Then by (9.16), we know when $|y_{r_1,\ldots,r_k}| \leq y_{max}$, one always has

$$\begin{aligned} Q_1' &\ll y_{max}^2 \left(\sum_{\substack{u\leq R \\ (u,W)=1}} \frac{\mu^2(u)}{\varphi(u)} \right)^k \left(\sum_{s\geq 1} \frac{\mu^2(s)}{\varphi(s)} \right)^{k(k-1)-1} \sum_{s'>D_0} \frac{\mu^2(s')}{\varphi(s')} \\ &\ll y_{max}^2 \left(\frac{\varphi(W)}{W} \log R \right)^k \frac{1}{D_0} \ll \frac{y_{max}^2 \varphi^k(W)(\log R)^k}{W^k D_0}. \end{aligned}$$

Therefore, (9.16) is reduced to

$$Q_1 = \sum_{u_1,\ldots,u_k} \frac{y_{u_1,\ldots,u_k}^2}{\prod_i \varphi(u_i)} + O\left\{ \frac{y_{max}^2 \varphi^k(W)(\log R)^k}{W^k D_0} \right\}. \tag{9.17}$$

Now is the time to assemble all the results we obtained thus far.

9.5 Asymptotic formula for S_1

By (9.15), we can express λ_{max} in terms of y_{max}:

$$\begin{aligned} \lambda_{max} &\ll y_{max} \prod_i \frac{d_i}{\varphi(d_i)} \sum_{\substack{t_1,\ldots,t_k\geq 1 \\ \prod_i t_i \leq R/\prod_i d_i \\ (t_i,d_i)=1 \\ (t_i,t_j)=1 \forall i\neq j}} \prod_i \frac{\mu^2(t_i)}{\varphi(t_i)} \\ &= y_{max} \prod_{p|\prod_i d_i} \left(1 + \frac{1}{p-1}\right) \sum_{\substack{t\leq R/\prod_i d_i \\ (t,\prod_i d_i)=1}} \frac{\mu^2(t)\tau_k(t)}{\varphi(t)} \\ &\leq y_{max} \sum_{r|\prod_i d_i} \frac{\mu^2(r)}{\varphi(r)} \sum_{\substack{t\leq R/r \\ (t,r)=1}} \frac{\mu^2(t)\tau_k(t)}{\varphi(t)} \\ &\leq y_{max} \sum_{u\leq R} \frac{\mu^2(u)\tau_k(u)}{\varphi(u)} \ll y_{max}(\log R)^k. \end{aligned}$$

Plugging this into (9.9), we have

$$E_1 \ll y_{max}^2 R^2 (\log R)^{4k} \ll \frac{y_{max}^2 N \varphi^k(W)(\log R)^k}{W^{k+1} D_0}.$$

Combining this with (9.8) and (9.17), we obtain

$$S_1 = \underbrace{\frac{N}{W}\sum_{u_1,\dots,u_k}\frac{y^2_{u_1,\dots,u_k}}{\prod_i \varphi(u_i)}}_{T_1} + O\left\{\frac{y^2_{max}N\varphi^k(W)(\log R)^k}{W^{k+1}D_0}\right\}, \tag{9.18}$$

which is precisely the formula in [14, Lemma 5.1]:

Let

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

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

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

9.6 Preliminary treatments for $S_2^{(m)}$

In the GPY–Zhang sieve, the treatments for the corresponding S_1 and $S_2^{(m)}$ are very similar, so we skipped the derivation for $S_2^{(m)}$ in §6 and §8. However, in Maynard's "dimensional reduction" sieve, the difference between the treatments for S_1 and $S_2^{(m)}$ is worth an expanded account. According to Dirichlet's theorem, an arithmetic progression contains infinitely many primes if and only if the first term is coprime to the common difference, so interchanging the order of summation in (9.4) gives

$$S_2^{(m)} = \sum_{\substack{d_1,\dots,d_k\\ e_1,\dots,e_k\\ d_m=e_m=1}}{}' \lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}\sum_{\substack{N<n\le 2N\\ n\equiv v_0(W)\\ [d_i,e_i]|(n+h_i)\forall i}}\chi_{\mathbb{P}}(n+h_m). \tag{9.19}$$

To continue expanding the green part, we invoke the prime number theorem on arithmetic progressions, so when we define

$$E(N,q) = 1 + \max_{(a,q)=1}\left|\sum_{\substack{N<n\le 2N\\ (a,q)=1}}\chi_{\mathbb{P}}(n) - \frac{1}{\varphi(q)}\int_N^{2N}\frac{\mathrm{d}u}{\log u}\right|,$$

one has

$$S_2^{(m)} = \frac{1}{\varphi(W)}\int_N^{2N}\frac{\mathrm{d}u}{\log u}\underbrace{\sum_{\substack{d_1,\dots,d_k\\ e_1,\dots,e_k\\ d_m=e_m=1}}{}'\frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{\prod_i\varphi([d_i,e_i])}}_{Q_2^{(m)}} + O(E_2), \tag{9.20}$$

in which under $EH(\theta)$ for $\frac{1}{2} \le \theta < 1$ and $R = N^{\frac{\theta}{2}-\varepsilon}$, E_2 satisfies

$$E_2 \ll y_{max}^2 (\log R)^{2k} \sum_{d \le R^2} \mu^2(d) \tau_{3k}(d) E(N,d) \ll_A \frac{y_{max}^2 N}{(\log N)^A}. \tag{9.21}$$

Thus, it remains to handle $Q_2^{(m)}$.

9.7 Diagonalization of $Q_2^{(m)}$

Let $g(u)$ be the multiplicative function satisfying

$$\varphi(m) = \sum_{u|m} g(u) \Rightarrow g(p) = p - 2.$$

Then by reasoning in the computation of (9.11) and (9.11), we have

$$Q_2^{(m)} = \sum_{\substack{u_1,\ldots,u_k \\ u_m=1}} \prod_i g(u_i) \sum_{\substack{s_{1,2},\ldots,s_{k,k-1} \\ (s_{i,j},W)=1 \forall i \neq j}} \prod_{\substack{i,j \\ i \neq j}} \mu(s_{i,j}) \sum_{\substack{d_1,\ldots,d_k \\ e_1,\ldots,e_k \\ d_m=e_m=1 \\ a_i|d_i, b_i|e_i \forall i}} \frac{\lambda_{d_1,\ldots,d_k} \lambda_{e_1,\ldots,e_k}}{\prod_i \varphi(d_i)\varphi(e_i)}, \tag{9.22}$$

in which a_i and b_i are defined as in §9.4.

By the Euler product formula for multiplicative functions,

$$\begin{aligned}
\sum_{\substack{n \le x \\ (n,q)=1}} \frac{\mu^2(n)}{g(n)} &\le \prod_{\substack{p \le x \\ p \nmid q}} \left(1 + \frac{1}{p-2}\right) \le \prod_{\substack{p \le x \\ p | q}} \left(1 + \frac{1}{p-2}\right)^{-1} \exp\left\{\sum_{p \le x} \frac{1}{p-2}\right\} \\
&\ll \prod_{\substack{p \le x \\ p|q}} \frac{p-2}{p-1} \log x = \prod_{p \le x} \frac{p-1}{p} \prod_{\substack{p \le x \\ p|q}} \frac{p(p-2)}{(p-1)^2} \log x \\
&\ll \prod_{p|q} \left(1 - \frac{1}{p}\right) \log x = \frac{\varphi(q)}{q} \log x.
\end{aligned}$$

As a result, under the definition

$$y_{r_1,\ldots,r_k}^{(m)} = \prod_i \mu(r_i) g(r_i) \sum_{\substack{d_1,\ldots,d_k \\ d_m=1 \\ r_i|d_i \forall i}} \frac{\lambda_{d_1,\ldots,d_k}}{\prod_i \varphi(d_i)}, \tag{9.23}$$

we see that the subcollection $Q_2'^{(m)}$ of terms in $Q_2^{(m)}$ with $s_{i,j} \neq 1$ satisfies

$$\begin{aligned} Q_2'^{(m)} &\ll (y_{max}^{(m)})^2 \left(\sum_{\substack{u \le R \\ (u,W)=1}} \frac{\mu^2(u)}{g(u)} \right)^{k-1} \left(\sum_{s \ge 1} \frac{\mu^2(s)}{g(s)} \right)^{k(k-1)-1} \sum_{s' > D_0} \frac{\mu^2(s')}{g(s')} \\ &\ll (y_{max}^{(m)})^2 \left(\frac{\varphi(W)}{W} \log R \right)^{k-1} \frac{1}{D_0} \ll \frac{(y_{max}^{(m)})^2 \varphi^{k-1}(W)(\log R)^{k-1}}{W^{k-1} D_0}. \end{aligned}$$

Combining this with (9.22) and (9.23), we deduce that

$$Q_2^{(m)} = \sum_{\substack{u_1,\dots,u_k \\ u_m=1}} \frac{(y_{u_1,\dots,u_k}^{(m)})^2}{\prod_i g(u_i)} + O\left\{ \frac{(y_{max}^{(m)})^2 \varphi^{k-1}(W)(\log R)^{k-1}}{W^{k-1} D_0} \right\}. \tag{9.24}$$

9.8 Asymptotic formula for $S_2^{(m)}$

By an easy integration by parts, one has

$$\int_N^{2N} \frac{\mathrm{d}u}{\log u} = \frac{N}{\log N} + O\left(\frac{N}{\log^2 N} \right),$$

and by (9.5), we know $1/\log N \ll 1/D_0$. Additionally, because

$$\sum_{\substack{u_1,\dots,u_k \\ u_m=1}} \frac{(y_{u_1,\dots,u_k}^{(m)})^2}{\prod_i g(u_i)} \ll (y_{max}^{(m)})^2 \left(\sum_{\substack{u \le R \\ (u,W)=1}} \frac{\mu^2(u)}{g(u)} \right)^{k-1} \ll \frac{(y_{max}^{(m)})^2 \varphi(W)^{k-1} (\log R)^{k-1}}{W^{k-1}},$$

we see that (9.20) becomes

$$\begin{aligned} S_2^{(m)} &= \frac{N}{\varphi(W) \log N} \underbrace{\sum_{\substack{u_1,\dots,u_k \\ u_m=1}} \frac{(y_{u_1,\dots,u_k}^{(m)})^2}{\prod_i g(u_i)}}_{T_2^{(m)}} \\ &\quad + O\left\{ \frac{(y_{max}^{(m)})^2 N \varphi^{k-2}(W)(\log N)^{k-2}}{W^{k-1} D_0} \right\} + O_A\left(\frac{y_{max}^2 N}{\log^A N} \right). \end{aligned} \tag{9.25}$$

This is exactly the formula in [14, Lemma 5.2]:

Let

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

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

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

Remark. In Maynard's original paper, the condition $u_m = 1$ is dropped, but this does not make any difference because from (9.23) it is clear that $y^{(m)}_{u_1,\ldots,u_k}$ vanishes when $u_m \neq 1$.

9.9 Asymptotic formula for $y^{(m)}_{r_1,\ldots,r_k}$

In the GPY sieve, we obtained asymptotics by first specifying λ and then computing the corresponding y and $y^{(m)}$. However, when we were deriving Soundararajan's generalized GPY sieve in §7, we found that expressing λ and $y^{(m)}$ in terms of y could save us a lot of energy, so we combine (9.15) and (9.23) to express $y^{(m)}_{r_1,\ldots,r_k}$ in terms of $y_{r_1,\ldots,r_k}$, yielding

$$\begin{aligned} y^{(m)}_{r_1,\ldots,r_k} &= \prod_i \mu(r_i) g(r_i) \sum_{\substack{d_1,\ldots,d_k \\ d_m=1 \\ r_i \mid d_i \forall i}} \prod_i \frac{\mu(d_i) d_i}{\varphi(d_i)} \sum_{\substack{a_1,\ldots,a_k \\ d_i \mid a_i \forall i}} \frac{y_{a_1,\ldots,a_k}}{\prod_i \varphi(a_i)} \\ &= \prod_i \mu(r_i) g(r_i) \sum_{a_1,\ldots,a_k} \frac{y_{a_1,\ldots,a_k}}{\prod_i \varphi(a_i)} \sum_{\substack{d_1,\ldots,d_k \\ d_m=1 \\ r_i \mid d_i \mid a_i \forall i}} \prod_i \frac{\mu(d_i) d_i}{\varphi(d_i)}. \end{aligned} \tag{9.26}$$

Using the properties of multiplicative functions, we know

$$\begin{aligned} \sum_{\substack{d \\ r \mid d \mid a}} \frac{\mu(d) d}{\varphi(d)} &= \frac{\mu(r) r}{\varphi(r)} \sum_{\substack{t \mid a \\ (t,r)=1}} \frac{\mu(t) t}{\varphi(t)} = \frac{\mu(r) r}{\varphi(r)} \prod_{\substack{p \mid a \\ p \nmid r}} \Big(1 - \frac{p}{p-1}\Big) \\ &= \frac{\mu(r) r}{\varphi(r)} \prod_{\substack{p \mid a \\ p \nmid r}} \frac{\mu(p)}{p-1} = \frac{\mu(r) r}{\varphi(r)} \frac{\mu(a)/\varphi(a)}{\mu(r)/\varphi(r)} = \frac{\mu(a) r}{\varphi(a)}. \end{aligned}$$

Plugging this into (9.26), we get

$$y^{(m)}_{r_1,\ldots,r_k} = \prod_i \mu(r_i) g(r_i) \sum_{\substack{a_1,\ldots,a_k \\ r_i \mid a_i \forall i}} \frac{y_{a_1,\ldots,a_k}}{\prod_i \varphi(a_i)} \prod_{i \neq m} \frac{\mu(a_i) r_i}{\varphi(a_i)}. \tag{9.27}$$

Remark. We assume $r_m = 1$ throughout the section.

By (9.14) and our constraints on the support of $\lambda_{d_1,\dots,d_k}$, we know (9.27) is supported on $(r_i, W) = 1$. In addition, the nonzero contribution in the sum comes from a_i satisfying $(a_i, W) = 1$, so if $a_i \neq r_i$, then $a_i > D_0 r_i$. Consequently, the subcollection of terms with $a_j \neq r_j \forall j \neq m$ in $y^{(m)}_{r_1,\dots,r_k}$ satisfies

$$\ll y_{max} \prod_i g(r_i) r_i \sum_{\substack{a_m \le R \\ (a_m,W)=1}} \frac{\mu^2(a_m)}{\varphi(a_m)} \prod_{i \neq j} \sum_{\substack{a_i \le R \\ (a_i,W)=1 \\ r_i | a_i}} \frac{\mu^2(a_i)}{\varphi(a_i)^2} \sum_{\substack{a_j > D_0 r_j \\ r_j | a_j}} \frac{\mu^2(a_j)}{\varphi(a_j)^2}$$

$$\ll y_{max} \frac{\varphi(W)}{W} \log R \prod_i g(r_i) r_i \prod_{i \neq j} \frac{\mu^2(r_i)}{\varphi(r_i)^2} \sum_{\substack{u \le R \\ (u, W r_i)=1}} \frac{\mu^2(u)}{\varphi(u)^2} \frac{\mu^2(r_j)}{\varphi(r_j)} \sum_{u > D_0} \frac{\mu^2(u)}{\varphi(u)^2}$$

$$\ll y_{max} \prod_i \frac{g(r_i)}{\varphi(r_i)} \frac{\varphi(W) \log R}{W D_0} \ll \frac{y_{max} \varphi(W) \log R}{D_0}.$$

Combining this with (9.27), we deduce that

$$y^{(m)}_{r_1,\dots,r_k} = \prod_i \frac{r_i g(r_i)}{\varphi(r_i)^2} \sum_{a_m} \frac{y_{r_1,\dots,r_{m-1},a_m,r_{m+1},\dots,r_k}}{\varphi(a_m)} + O\left(\frac{y_{max} \varphi(W) \log R}{D_0}\right). \tag{9.28}$$

Define $\beta(n) = n g(n)/\varphi(n)^2$. Then for a large prime p,

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

so it follows from $(r_1 r_2 \cdots r_k, W) = 1$ that

$$1 \le \prod_i \frac{r_i g(r_i)}{\varphi(r_i)^2} \le \prod_{p > D_0} \left\{1 + O\left(\frac{1}{p^2}\right)\right\} = 1 + O\left(\frac{1}{D_0}\right).$$

Additionally, because

$$\sum_{a_m} \frac{y_{r_1,\dots,r_{m-1},a_m,r_{m+1},\dots,r_k}}{\varphi(a_m)} \ll y_{max} \sum_{\substack{a_m \le R \\ (a_m,W)=1}} \frac{\mu^2(a_m)}{\varphi(a_m)} \ll \frac{y_{max} \varphi(W) \log R}{W},$$

we see that (9.28) becomes

$$y^{(m)}_{r_1,\dots,r_k} = \sum_{a_m} \frac{y_{r_1,\dots,r_{m-1},a_m,r_{m+1},\dots,r_k}}{\varphi(a_m)} + O\left(\frac{y_{max} \varphi(W) \log R}{D_0}\right), \tag{9.29}$$

which is precisely [14, Lemma 5.3]:

If $r_m = 1$ then

$$y^{(m)}_{r_1,\dots,r_k} = \sum_{a_m} \frac{y_{r_1,\dots,r_{m-1},a_m,r_{m+1},\dots,r_k}}{\varphi(a_m)} + O\Big(\frac{y_{max} \varphi(W) \log R}{W D_0}\Big).$$

We have now expressed Maynard's sieve in terms of $y_{r_1,\ldots,r_k}$. To continue, we need to specify $y_{r_1,\ldots,r_k}$ so that T_1 and $T_2^{(m)}$ become analytic expressions.

9.10 Analytic expressions for $T_1, T_2^{(m)}$

Similar to how we picked λ_d in §7, we let $F : [0,1]^k \to \mathbb{R}$ be smooth and vanishing if $x_1 + \cdots + x_k > 1$. When $r_1 r_2 \cdots r_k$ is squarefree and corpime to W, define $y_{r_1,\ldots,r_k}$ to be

$$y_{r_1,\ldots,r_k} = F\left(\frac{\log r_1}{\log R}, \ldots, \frac{\log r_k}{\log R}\right), \tag{9.30}$$

so T_1 in (9.18) becomes

$$T_1 = \sum_{\substack{u_1,\ldots,u_k \\ (u_i,W)=1 \\ (u_i,u_j)=1 \forall i\neq j}} \prod_i \frac{\mu^2(u_i)}{\varphi(u_i)} F^2\left(\frac{\log u_1}{\log R}, \ldots, \frac{\log u_k}{\log R}\right). \tag{9.31}$$

To continue the analyses of (9.31) and $T_2^{(m)}$, we require a general result that can easily produce asymptotic formulas.

An asymptotic lemma Let $h(n)$ be a multiplicative function satisfying the following asymptotic conditions for primes:

$$\sum_{p\leq x} h(p)\log p = \log x + O(L), \quad \sum_p h^2(p)\log p < \infty. \tag{9.32}$$

Then we are interested in the sum

$$H(x) = \sum_{n\leq x} \mu^2(n)h(n). \tag{9.33}$$

Similar to how Chebyshev originally studied primes, we attach a logarithmic weight:

$$\begin{aligned}
I(x) &= \sum_{n\le x}\mu^2(n)h(n)\log n = \sum_{p\le x}\log p \sum_{\substack{n\le x\\ p|n}}\mu^2(n)h(n)\\
&= \sum_{p\le x}h(p)\log p \sum_{\substack{t\le x/p\\ p\nmid t}}\mu^2(t)h(t)\\
&= \sum_{p\le x}h(p)\log p \sum_{t\le x/p}\mu^2(t)h(t) - \sum_{p\le x}h(p)\log p\sum_{\substack{t\le x/p\\ p|t}}\mu^2(m)h(m)\\
&= \sum_{p\le x}h(p)\log p \sum_{t\le x/p}\mu^2(t)h(t) - \sum_{p\le x}h^2(p)\log p\sum_{\substack{m\le x/p^2\\ p\nmid m}}\mu^2(m)h(m)\\
&= \sum_{t\le x}\mu^2(t)h(t)\sum_{p\le x/t}h(p)\log p + O\{H(x)\}\\
&= \sum_{t\le x}\mu^2(t)h(t)\left\{\log\frac{x}{t} + O(L)\right\} + O\{H(x)\}\\
&= H(x)\log x - I(x) + O\{LH(x)\}.
\end{aligned}$$

Therefore, we obtain

$$I(x) = \frac{1}{2}H(x)\log x + O\{LH(x)\},$$

which indicates that

$$I_1(x) = \int_1^x H(t)\frac{\mathrm{d}t}{t} = \sum_{n\le x}\mu^2(n)h(n)\log\frac{x}{n} = \frac{1}{2}H(x)\log x + O\{LH(x)\}, \tag{9.34}$$

so we have

$$H(x) = \frac{2I_1(x)}{\log x}\left\{1 + O\left(\frac{L}{\log x}\right)\right\}. \tag{9.35}$$

Differentiating the left-hand side of (9.34) gives

$$\frac{I_1'}{I_1}(x) = \frac{2}{x\log x} + O\left(\frac{L}{x\log^2 x}\right).$$

Integrating, we see that there is some $C > 0$ such that

$$I_1(x) = \frac{1}{2}C\log^2 x + O(LC\log x),$$

and combining this with (9.35) gives

$$H(x) = C\log x + O(LC). \tag{9.36}$$

To determine C, we study the properties of the Dirichlet series $F(s)$ associated with $\mu^2(n)h(n)$. By the Euler product formula, as $s \to 0^+$, one has

$$\begin{aligned} F(s) &= \prod_p \left(1 + \frac{h(p)}{p^s}\right)\left(1 - \frac{1}{p^{s+1}}\right) \prod_p \left(1 - \frac{1}{p^{s+1}}\right)^{-1} \\ &= \prod_p \left(1 + \frac{h(p)}{p^s}\right)\left(1 - \frac{1}{p^{s+1}}\right) \zeta(s+1) \\ &\sim \frac{1}{s} \underbrace{\prod_p (1 + h(p))(1 - p^{-1})}_{\mathfrak{S}}. \end{aligned}$$

By partial summation on (9.32), it is easily verified that the product $\mathfrak{S}$ converges. On the other hand, by the integral formula relating $F(s)$ and $H(x)$, we know

$$F(s) = s \int_1^\infty \frac{H(t)}{t^{s+1}} \mathrm{d}t \sim Cs \int_1^\infty \frac{\log t}{t^{s+1}} \mathrm{d}t = \frac{C}{s},$$

so (9.36) becomes

$$H(x) = \mathfrak{S} \log x + O(\mathfrak{S} L). \tag{9.37}$$

In the study of $T_1, T_2^{(m)}$, we need to study sums of the form

$$H_G(x) = \sum_{n \le x} \mu^2(n) h(n) G\left(\frac{\log n}{\log x}\right).$$

Applying partial summation and plugging in (9.37), we get

$$\begin{aligned} H_G(x) &= \int_1^x G\left(\frac{\log t}{\log x}\right) \mathrm{d}H(t) = \int_0^1 G(u) \mathrm{d}H(x^u) \\ &= \mathfrak{S} \log x \int_0^1 G(u) \mathrm{d}u + \int_0^1 G(u) \mathrm{d}O(L\mathfrak{S}). \end{aligned}$$

Performing integration by parts on the remaining component, we deduce the result:

Lemma 9.1 (Asymptotic lemma)**.** *If $h(n)$ is a multiplicative function satisfying* (9.32) *and $G : [0,1] \to \mathbb{C}$ is continuously differentiable, then*

$$\sum_{n \le x} \mu^2(n) h(n) G\left(\frac{\log n}{\log x}\right) = \mathfrak{S} \log x \int_0^1 G(u) \mathrm{d}u + O\{L \mathfrak{S} G_{max}\},$$

where

$$G_{max} = \max_{0 \le x \le 1} |G(x)| + \max_{0 \le x \le 1} |G'(x)|$$

and

$$\mathfrak{S} = \prod_p (1 + h(p))(1 - p^{-1}).$$

Remark. Our proof is adapted from [10, §A.2]

We are now in a position to determine the analytic expressions for T_1 and $T_2^{(m)}$

Transformation of T_1 To apply the asymptotic lemma, we first define

$$F_{max} = \max_{t_1,\ldots,t_k} |F(t_1,\ldots,t_k)| + \sum_{1\le m\le k} \max_{t_1,\ldots,t_k} \left| \frac{\partial F}{\partial t_i}(t_1,\ldots,t_k) \right|. \tag{9.38}$$

Because $(u_i, W) = 1 \forall i$, dropping the pairwise coprime condition $(u_i, u_j) = 1 \forall i \neq j$ in (9.31) creates an error of

$$\ll F_{max}^2 \sum_{p>D_0} \sum_{\substack{u_1,\ldots,u_k \\ (u_i,W)=1\forall i \\ p|(u_i,u_j)\exists i,j\wedge i\neq j}} \prod_i \frac{\mu^2(u)}{\varphi(u)} \ll F_{max}^2 \sum_{p>D_0} \frac{1}{(p-1)^2} \left(\sum_{\substack{u\le R \\ (u,W)=1}} \frac{\mu^2(u)}{\varphi(u)} \right)^k$$

$$\ll \frac{F_{max}^2}{D_0} \left(\sum_{\substack{u\le R \\ (u,W)=1}} \frac{\mu^2(u)}{\varphi(u)} \right)^k \ll \frac{F_{max}^2 \varphi(W)^k (\log R)^k}{W^k D_0}.$$

Set $h(n) = 1/\varphi(n)$ for $(n, W) = 1$ and zero otherwise. Then $h(n)$ satisfies (9.32) with

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

Moreover, because $\mathfrak{S} = \varphi(W)/W$, it follows from the asymptotic lemma and the error estimates above that

$$T_1 = \frac{\varphi(W)^k (\log R)^k}{W^k} \underbrace{\int_0^1 \cdots \int_0^1 F^2(t_1,\ldots,t_k) \mathrm{d}t_1 \cdots \mathrm{d}t_k}_{I_k(F)}$$
$$+ O\left\{ \frac{F_{max}^2 \varphi(W)^k (\log R)^{k-1} \log D_0}{W^k} \right\} + O\left\{ \frac{F_{max}^2 \varphi(W)^k (\log R)^k}{W^k D_0} \right\}.$$

Combining this with (9.5) and (9.18), one obtains an analytic expression for S_1:

$$S_1 = \frac{\varphi(W)^k N (\log R)^k}{W^{k+1}} I_k(F) + O\left\{ \frac{F_{max}^2 \varphi(W)^k N (\log R)^k}{W^{k+1} D_0} \right\}. \tag{9.39}$$

This completes the proof of [14, Lemma 6.2]:

Let $y_{r_1,\ldots,r_k}$ be given in terms of a smooth function F by [(9.30)], with F supported on $\mathcal{R}_k = \{(x_1,\ldots,x_k) \in [0,1]^k : \sum_{i=1}^k x_i \le 1\}$. Let

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

Then we have

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

where

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

Transformation of $T_2^{(m)}$ To compute $T_2^{(m)}$, we first determine $y^{(m)}_{r_1,\ldots,r_k}$. Applying the asymptotic lemma to (9.28), we see that when $r_m = 1$, one has

$$y^{(m)}_{r_1,\ldots,r_m} = (\log R) \frac{\varphi(W)}{W} \prod_i \frac{\varphi(r_i)}{r_i} \underbrace{\int_0^1 F\left(\frac{\log r_1}{\log R}, \ldots, t_m, \ldots, \frac{\log r_k}{\log R}\right) \mathrm{d}t_m}_{F^{(m)}_{r_1,\ldots,r_k}} + O\left(\frac{F_{max}\varphi(W)\log R}{W D_0}\right).$$

Plugging this into $T_2^{(m)}$ of (9.25), we have

$$T_2^{(m)} = \frac{\varphi(W)^2 (\log R)^2}{W^2} \sum_{\substack{u_1,\ldots,u_k \\ u_m=1 \\ (u_i,W)=1\forall i \\ (u_i,u_j)=1\forall i\neq j}} \prod_i \frac{\mu^2(u_i)\varphi(u_i)^2}{g(u_i)u_i^2} F^{(m)}_{u_1,\ldots,u_k} + O\left\{\frac{F_{max}^2 \varphi(W)^{k+1}(\log R)^{k+1}}{W^{k+1} D_0}\right\}. \tag{9.40}$$

Similar to our computations in T_1, we see that we can drop $(u_i, u_j) = 1 \forall i \neq j$ at the expense of

$$\ll \frac{F_{max}^2 \varphi(W)^2 (\log R)^2}{W^2} \sum_{p > D_0} \frac{\varphi(p)^4}{g(p)^2 p^4} \left(\sum_{\substack{u \leq R \\ (u,W)=1}} \frac{\mu^2(u)\varphi(u)^2}{g(u)u^2} \right)^{k-1}$$
$$\ll \frac{F_{max}^{k+1} \varphi(W)^{k+1} (\log R)^{k+1}}{W^{k+1} D_0}.$$

Therefore, applying the asymptotic lemma to (9.40) gives

$$T_2^{(m)} = \frac{\varphi(W)^{k+1}(\log R)^{k+1}}{W^{k+1}} J_k(F) + O\left\{\frac{F_{max}^2 \varphi(W)^{k+1}(\log R)^{k+1}}{W^{k+1} D_0}\right\}, \tag{9.41}$$

in which

$$J_k^{(m)}(F) = \int_0^1 \cdots \int_0^1 \left(\int_0^1 F(t_1, \ldots, t_k) \mathrm{d}t_m\right)^2 \mathrm{d}t_1 \cdots \mathrm{d}t_{m-1} \mathrm{d}t_{m+1} \cdots \mathrm{d}t_k. \tag{9.42}$$

Plugging these into (9.25), we get

$$S_2^{(m)} = \frac{\varphi(W)^k N(\log R)^{k+1}}{W^{k+1}\log N} J_k^{(m)}(F) + O\left\{\frac{F_{max}^2 \varphi(W)^k N(\log N)^k}{W^{k+1}D_0}\right\}. \quad (9.43)$$

This completes the proof of [14, Lemma 6.3]:

Let $y_{r_1,\ldots,r_k}$, F and F_{max} be as described in [14, Lemma 6.2] Then we have

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

where

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

In addition, plugging (9.30) into (9.15), we get an expression of $\lambda_{d_1,\ldots,d_k}$:

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

This explains why [14, Proposition 4.1] has such a formidable appearance:

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

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

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

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

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

Having converted everything into analytic expressions, we investigate the lower bound of (9.2).

9.11 Maynard's variational problem

Plugging (9.39) and (9.43) into (9.2), we get

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

Let $\mathcal{S}_k$ be the space of continuously differentiable functions $F : [0,1]^k \to \mathbb{R}^k$ supported on $t_1 + \cdots + t_k \leq 1$. Then defining

$$M_k = \sup_{F \in \mathcal{S}_k} \frac{\sum_m J_k^{(m)}(F)}{I_k(F)}, \tag{9.44}$$

we have

$$S > \frac{\varphi(W)^{k+1} N (\log R)^k}{W^k} I_k(F) \left\{ \frac{\theta}{2} M_k - \rho + O(\varepsilon) \right\}. \tag{9.45}$$

Therefore, $S > 0$ will follow from $\theta M_k/2 > \rho$. Combining this with (9.2), $S > 0$ implies there exists infinitely many n for which at least $\lfloor \rho + 1 \rfloor$ members among $n + h_1, \dots, n + h_k$ are primes. This proves [14, Proposition 4.2]:

Let the primes have a level of distribution $\theta > 0$. Let $\delta > 0$ and $\mathcal{H} = \{h_1, \dots, h_k\}$ be an admissible set. Let $I_k(F)$ and $J_k^{(m)}(F)$ be given as in [14, Proposition 4.1], and let $\mathcal{S}_k$ denote the set of [continuously differentiable] functions $F : [0,1]^k \to \mathbb{R}$ supported on $\mathcal{R}_k = \{(x_1, \dots, x_k) \in [0,1]^k : \sum_{i=1}^k x_i \leq 1\}$ with $I_k(F) \neq 0$ and $J_k^{(m)}(F) \neq 0$ for each m. Let

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

Then there are infinitely many integers n such that at least r_k of the $n + h_i$ ($1 \leq i \leq k$) are prime. In particular, $\liminf_n (p_{n+r_k-1} - p_n) \leq \max_{1 \leq i,j \leq k} (h_i - h_j)$.

So far, we have completed a full derivation of the arithmetic aspect of Maynard's work. From (9.45), we see that the problem of small gaps between primes is a matter of optimizing the functional $M_k = M_k(F)$. If we can find k such that $M_k > 4 = 2/(1/2)$, then we will deduce bounded gaps between primes by only invoking the Bombieri–Vinogradov theorem.

9.12 Conclusion

In this article, we applied a variety of techniques from number theory and obtained an analytic expression for Maynard's "dimensional reduction" sieve, eventually converting a problem of prime number theory into a variational problem. In the next article, we will introduce Maynard's solution to this variational problem. Please stay tuned for updates!

Oct 17, 2022

10 Maynard's dimensional reduction strike II

In the last article, we converted the problem of prime gap into a variational problem. Define

$$I_k = \int \cdots \int_{[0,1]^k} F^2(t_1, \ldots, t_k) \mathrm{d}t_1 \cdots \mathrm{d}t_k \tag{10.1}$$

and

$$J_k^{(m)} = \int \cdots \int_{[0,1]^{k-1}} \left(\int_0^1 F(t_1, \ldots, t_k) \mathrm{d}t_m \right)^2 \prod_{\substack{1 \le i \le k \\ i \ne m}} \mathrm{d}t_i. \tag{10.2}$$

Then we want to find $F = F_k$ supported on $t_1 + \cdots + t_k \le 1$ such that

$$M_k = \frac{\sum_{1 \le m \le k} J_k^{(m)}}{I_k} \tag{10.3}$$

attains its maximum. Specifically, when θ is the level of distribution of primes, there exists $C_k \ge 2$ such that there are infinitely many n's such that the interval $[n, n + C_k]$ contains $r_k = \lceil \theta M_k / 2 \rceil$ primes. Therefore, the analytic properties of (10.3) have significant consequences in number theory. In this article, we follow Maynard's steps to attack this variational problem, thereby proving his main results:

Theorem 10.1. $\liminf_{n \to \infty} (p_{n+m} - p_n) \le Cm^3 e^{4m}$ *for some absolute* $C > 0$ *and all* $m \in \mathbb{N}$.

Theorem 10.2. $\liminf_{n \to \infty} (p_{n+1} - p_n) \le 600$.

10.1 Optimization procedure for large k

According to the GPY sieve formula (9.2), permuting $h_1, \ldots, h_k$ does not affect the asymptotic formula, so we naturally assume F to be symmetric about $t_1, \ldots, t_k$. Combining this assumption with (**??**), we see that $J_k^{(m)} = J_k^1 =: J_k$, so $M_k = kJ_k / I_k$. Therefore, optimizing (10.3) is the same as finding upper bounds for I_k and lower bounds for J_k.

To simplify the task, Maynard assumed F to take the form

$$F(t_1, \ldots, t_k) = g(kt_1) g(kt_2) \cdots g(kt_k) \tag{10.4}$$

on its support $t_1 + \cdots + t_k \le 1$, where g is some smooth function on $[0, T]$. Plugging this into (10.1) and (10.2), we get

$$I_k \le \left(\int_0^{T/k} g^2(kt) \mathrm{d}t \right)^k = k^{-k} \gamma^k, \quad \gamma = \int_0^T g^2(u) \mathrm{d}u \tag{10.5}$$

and

$$J_k \geq \underset{\substack{t_2,\ldots,t_k \geq 0 \\ \sum_{2\leq i\leq k} t_i \leq 1-T/k}}{\int\cdots\int} \left(\int_0^{T/k} g(kt_1)\mathrm{d}t_1\right)^2 \prod_{2\leq i\leq k} g^2(kt_i)\mathrm{d}t_2\cdots\mathrm{d}t_k$$

$$= k^{-k-1}\left(\int_0^T g(u)\mathrm{d}u\right)^2 \underset{\substack{u_2,\ldots,u_k\in[0,T] \\ \sum_{2\leq i\leq k} u_i \leq k-T}}{\int\cdots\int} \prod_{2\leq i\leq k} g^2(u_i)\mathrm{d}u_2\cdots\mathrm{d}u_k. \tag{10.6}$$

Let J_k' be the version of J_k without the blue condition. Then

$$J_k' = k^{-k-1}\left(\int_0^\infty g(u)\mathrm{d}u\right)^2\left(\int_0^\infty g^2(u)\mathrm{d}u\right)^{k-1} = k^{-k-1}\gamma^{k-1}\left(\int_0^\infty g(u)\mathrm{d}u\right)^2, \tag{10.7}$$

Compared to (10.2), (10.7) is simpler in structure, so we hope the error

$$E_k = J_k' - J_k = \frac{J_k'}{\gamma^{k-1}} \underset{\substack{u_2,\ldots,u_k\in[0,T] \\ \sum_{2\leq i\leq k} u_i > k-T}}{\int\cdots\int} \prod_{2\leq i\leq k} g^2(u_i)\mathrm{d}u_2\cdots\mathrm{d}u_k \tag{10.8}$$

is small. This means $g(u)$ must be very small for large u. To achieve this, we invoke some intuitions from physics.

10.2 Simplex and center of mass

Remark. Let $P_0, P_2, \ldots, P_n$ be points in a Euclidean space in general position (i.e., no P_j is a convex linear combination of others). Then the set of points of the form

$$\sum_{j=0}^n t_jP_j, \quad t_j\in[0,1], \quad \sum_{j=0}^n t_j = 1$$

is called an n-simplex.

Since squares of real numbers are non-negative, we can regard J_k and J_k' as the mass of certain geometric objects. Indeed, the former denotes the mass of a $(k-1)$-simplex whose vertices consist of the origin and $k-1$ standard basis vectors, and the latter corresponds to that of a $(k-1)$-hypercube, and the density of these objects is given by

$$\rho(u_2,\ldots,u_k) = k^{-k-1}\left(\int_0^T g(u)\mathrm{d}u\right)^2 \prod_{2\leq i\leq k} g^2(u_i) \tag{10.9}$$

From a physical point of view, J_k' well approximates J_k if the center of mass of the $(k-1)$-hypercube lies within the hypercube $[0, 1-\frac{T}{k}]^{k-1}$ embedded in the

$(k-1)$-simplex $\sum_{2\le i\le k} u_i \le 1-\frac{T}{k}$. Therefore, we introduce the inequality

$$\mu = \frac{\int_0^T ug^2(u)\mathrm{d}u}{\int_0^T g^2(u)\mathrm{d}u} < 1 - \frac{T}{k}. \tag{10.10}$$

Under this assumption, we estimate E_k in (10.8).

Since $k-T=(k-1)-(T-1)$, we have

$$\sum_{2\le i\le k} u_i > k-T \iff \frac{1}{k-1}\sum_{2\le i\le k} u_i > 1-\frac{T-1}{k-1}.$$

Moreover, because $k>T$ and $k\ge 2$, so $\frac{T-1}{k-1} > \frac{T}{k}$. Hence, setting the right-hand side as $\eta+\mu$, we get

$$\eta = \left(1-\frac{T-1}{k-1}\right) - \mu > 0 \tag{10.11}$$

and

$$\left(\frac{1}{k-1}\sum_{2\le i\le k} u_i - \mu\right)^2 \ge \begin{cases} \eta^2 & \sum_{2\le i\le k} u_i > k-T \\ 0 & \sum_{2\le i\le k} u_i \le k-T \end{cases}. \tag{10.12}$$

Plugging (10.12) into the multiple integral in (10.8), one has

$$\begin{aligned}\eta^2 \underset{\substack{u_2,\dots,u_k\in[0,T] \\ \sum_{2\le i\le k} u_i > k-T}}{\int\cdots\int} &\le \int_{[0,T]^{k-1}} \left(\frac{1}{k-1}\sum_{2\le i\le k} u_i - \mu\right)^2 \prod_{2\le m\le k} g^2(u_m)\mathrm{d}u_2\cdots\mathrm{d}u_k \\ &= \int_{[0,T]^{k-1}} \Bigg(\frac{2}{(k-1)^2}\sum_{2\le i<j\le k} u_iu_j \\ &- \frac{2\mu}{k-1}\sum_{2\le i\le k} u_i + \mu^2 + \frac{1}{(k-1)^2}\sum_{2\le i\le k} u_i^2\Bigg) \\ &\times \prod_{2\le m\le k} g^2(u_m)\mathrm{d}u_2\cdots\mathrm{d}u_k. \end{aligned} \tag{10.13}$$

By (10.10) and symmetry, we know

$$\int_{[0,T]^{k-1}} u_iu_j \prod_{2\le m\le k} g^2(u_m)\mathrm{d}u_2\cdots\mathrm{d}u_k = \mu^2\gamma^{k-1},$$

$$\int_{[0,T]^{k-1}} u_i \prod_{2\le m\le k} g^2(u_m)\mathrm{d}u_2\cdots\mathrm{d}u_k = \mu\gamma^{k-1},$$

and

$$\int_{[0,T]^{k-1}} u_i^2 \prod_{2\le m\le k} g^2(u_m)\mathrm{d}u_2\cdots\mathrm{d}u_k \le \mu T\gamma^{k-1}.$$

Plugging these back into (10.13), we deduce that

$$E_k \leq \frac{J_k'}{\gamma^{k-1}} \cdot \eta^{-2}\gamma^{k-1}\left(\frac{\mu T}{k-1} - \frac{\mu^2}{k-1}\right) \leq \frac{J_k'\mu T}{\eta^2(k-1)},$$

so

$$J_k \geq J_k'\left(1 - \frac{\mu T}{\eta^2(k-1)}\right).$$

Combining this with (10.3), (10.5), (10.6), and (10.7), we obtain the following lower bound:

$$M_k \geq \gamma^{-1}\left(\int_0^T g(u)\mathrm{d}u\right)^2\left(1 - \frac{\mu T}{\eta^2(k-1)}\right). \tag{10.14}$$

10.3 The optimal choice of g

According to (10.14), the optimization problem with respect to a multivariable function $F(t_1, \ldots, t_k)$ has now become a simpler optimization problem with respect to $g(u)$:

$$\max_g \left|\int_0^T g(u)\mathrm{d}u\right| \quad s.t. \quad \gamma = \int_0^T g^2(u)\mathrm{d}u, \quad \mu\gamma = \int_0^T ug^2(u)\mathrm{d}u. \tag{10.15}$$

By the principle of Lagrange multipliers, we construct the functional

$$\begin{aligned} S(g) &= \int_0^T g(u)\mathrm{d}u - \alpha\left(\int_0^T g^2(u)\mathrm{d}u - \gamma\right) - \beta\left(\int_0^T ug^2(u)\mathrm{d}u - \mu\gamma\right) \\ &= \int_0^T \underbrace{\left(g(u) - (\alpha+\beta u)g^2(u) + \frac{\alpha\gamma+\beta\mu\gamma}{T}\right)}_{L(u,g)}\mathrm{d}u. \end{aligned}$$

Now, by the Euler–Lagrange equation and $\frac{\partial L}{\partial g'} = 0$, we see that $S(g)$ attains extremum under the choice

$$g(u) = \frac{1}{2\alpha + 2\beta u} = \frac{(2\alpha)^{-1}}{1 + Au}, \quad A = \beta/\alpha > 0, \tag{10.16}$$

so we have

$$\int_0^T g(u)\mathrm{d}u = \frac{(2\alpha)^{-1}}{A}\log(1+AT), \quad \gamma = \frac{(2\alpha)^{-2}}{A}\left(1 - \frac{1}{1+AT}\right), \tag{10.17}$$

and

$$\mu\gamma = \frac{(2\alpha)^{-2}}{A^2}\left[\log(1+AT) - \left(1 - \frac{1}{1+AT}\right)\right]. \tag{10.18}$$

These imply that

$$\gamma^{-1}\left(\int_0^T g(u)\mathrm{d}u\right)^2 \geq \frac{\log^2(1+AT)}{A}, \quad \mu = \frac{1}{A}\left(\frac{\log(1+AT)}{1-e^{-\log(1+AT)}} - 1\right). \tag{10.19}$$

Now, we set $T = \frac{e^A-1}{A}$, so

$$\begin{aligned}\mu &= \frac{1}{A}\left(\frac{A}{1-e^{-A}} - 1\right) = 1 - \frac{1}{A} + O(e^{-A}) \\ &= 1 - \frac{e^A}{Ae^A} + O(e^{-A}) \leq 1 - \frac{T}{e^A} + O(e^{-A}) \\ &= 1 - \frac{T}{e^A}[1 + O(Ae^{-A})] \leq 1 - \frac{T}{2e^A}.\end{aligned}$$

for large A. To ensure (10.10) holds, we require $e^A = o(k)$. Now, by (10.11), we also know

$$\eta = \frac{1}{A} - \frac{T-1}{k-1} + O(e^{-A}) = \frac{1-e^A/k}{A}\{1 + O(Ae^{-A})\},$$

which implies

$$\begin{aligned}\frac{\mu T}{\eta^2(k-1)} &= \frac{(1-A^{-1})(e^A-1)/A}{A^{-2}(1-e^A/k)^2(k-1)}\{1+O(Ae^{-A})\} \\ &= \frac{Ae^A}{k}\left\{1 + O\left(\frac{A}{e^A} + \frac{e^A}{k} + \frac{1}{A}\right)\right\}.\end{aligned} \tag{10.20}$$

Now, set $e^A = k/\log^2 k$, so $Ae^{-A} \sim k^{-1}\log^3 k$, $k^{-1}e^A \sim (\log k)^{-2}$, and $A^{-1} \sim (\log k)^{-1}$, so (10.20) becomes

$$\frac{\mu T}{\eta^2(k-1)} \leq \frac{1}{\log k}\left\{1 + O\left(\frac{1}{\log k}\right)\right\} = \frac{1}{\log k} + O\left(\frac{1}{\log^2 k}\right).$$

Combining this with (10.19) and plugging them into (10.14), one deduces that

$$\begin{aligned}M_k &\geq A\left(1 - \frac{\mu T}{\eta^2(k-1)}\right) \\ &= (\log k - 2\log\log k)\left\{1 - \frac{1}{\log k} + O\left(\frac{1}{\log^2 k}\right)\right\} \\ &= \log k - 2\log\log k - 1 + o(1).\end{aligned} \tag{10.21}$$

Now, let us discuss the number-theoretic significance of (10.21).

10.4 Proof of Theorem 10.1

By (10.21), we have for large k that

$$r_k \geq \frac{\theta}{2} M_k \geq \frac{\theta}{2} \log \frac{k}{(\log k)^2} + O(1).$$

Since the main term on the right-hand side is increasing, for each $m \in \mathbb{N}$ one can find $k \in \mathbb{N}$ such that $r_k \geq m+1$. Therefore, we can find $D_m = C_k \geq 2$ such that there exist infinitely many intervals of length D_m containing $m+1$ primes, so

$$\liminf_{n\to\infty} (p_{n+m} - p_n) \leq D_m < \infty. \tag{10.22}$$

We can go further to explore the growth of D_m with respect to m. Set $k = \lceil Bm^2 e^{2m/\theta} \rceil$. Then for large B, m, one has

$$\begin{aligned} r_k &\geq \frac{\theta}{2} \log \frac{Bm^2 e^{2m/\theta}}{(2\log m + 2m/\theta)^2} 1 + O(1) \\ &= m + \frac{\theta}{2} \log B + O(1) > m + 1. \end{aligned}$$

Based on our study of admissible tuples in §8.5, let $h_1, h_2, \ldots, h_k$ be the first k primes greater than k. Then $\mathcal{H} = \{h_1, \ldots, h_k\}$ is admissible. Therefore, we can set

$$C_k = \min_{\{h_1,\ldots,h_k\} \text{ admissible}} \max_{1\leq i<j\leq k} |h_i - h_j| \leq p_{\pi(k)+k} - p_k, \tag{10.23}$$

By the prime number theorem, $p_n \sim n \log n$, so $C_k \ll k \log k$. Plugging in our expressions for k, we obtain a quantitative version of (10.22):

$$\liminf_{n\to\infty} (p_{n+m} - p_n) \leq Cm^3 e^{2m/\theta}. \tag{10.24}$$

By the Bombieri–Vinogradov theorem, we can take $\theta = \frac{1}{2}$. Plugging this into (10.24) concludes the proof of Theorem 10.1, which can be regarded as a generalization of Zhang's theorem.

10.5 Optimization procedure for small k

If we can find k such that

$$M_k > 2/\theta, \tag{10.25}$$

then we have

$$\liminf_{n\to\infty} (p_{n+1} - p_n) \leq C_k. \tag{10.26}$$

From (10.23), we know C_k is increasing with respect to k, so we want to find the smallest k for which (10.25) holds, so we cannot just compute asymptotic lower bounds.

Since continuous functions can be uniformly approximated by polynomials, for small k, Maynard set F to be a symmetric polynomial directly. Since F is supported on $t_1 + \cdots + t_k \leq 1$, a reasonable design is

$$F_{b,c}(t_1, \ldots, t_k) = \left(1 - \sum_{1 \leq i \leq k} t_i\right)^b \left(\sum_{1 \leq i \leq k} t_i^2\right)^c,$$

where $b, c \in \mathbb{Z}_{\geq 0}$. For flexibility, Maynard considered linear combinations of $F_{b,c}$. That is, when $t_1 + \cdots + t_k \leq$,

$$F(t_1, \ldots, t_k) = \sum_{1 \leq j \leq d} a_j F_{b_j, c_j}(t_1, \ldots, t_k). \tag{10.27}$$

Plugging (10.27) into I_k and $J_k^{(m)}$, we see that when $\boldsymbol{a} = (a_1, a_2, \ldots, a_d)^T$ is a column vector, there exists positive definite A_1, A_2 for which

$$M_k = \frac{kJ_k}{I_k} = \frac{\boldsymbol{a}^T A_2 \boldsymbol{a}}{\boldsymbol{a}^T A_1 \boldsymbol{a}}. \tag{10.28}$$

Therefore, we have effectively converted a variational problem to an optimization problem concerning the ratio of quadratic forms.

10.6 Optimization of the quadratic form

Since the ratio (10.28) is invariant under dilation, we can normalize the denominator so we are now faced with a multivariable constraint optimization problem.

$$\max_{\boldsymbol{a}} \boldsymbol{a}^T A_2 \boldsymbol{a} \quad s.t. \quad \boldsymbol{a}^T A_1 \boldsymbol{a} = 1, \tag{10.29}$$

which is approachable using Lagrange multipliers. Define

$$L = \boldsymbol{a}^T A_1 \boldsymbol{a} - \lambda(\boldsymbol{a}^T A_2 \boldsymbol{a} - 1).$$

Then the gradient calculation gives

$$0 = \frac{\partial L}{\partial \boldsymbol{a}} = (2A_2 - 2\lambda A_1)\boldsymbol{a} \iff {A_1}^{-1} A_2 \boldsymbol{a} = \lambda \boldsymbol{a}.$$

Therefore, L attains extremum if and only if $\boldsymbol{a}$ is an eigenvector of $A_1^{-1}A_2$, so the solution to (10.29) is exactly the largest eigenvalue of $A_1^{-1}A_2$:

$$M_k = \boldsymbol{a}^T A_1(\lambda \boldsymbol{a}) = \lambda.$$

10.7 Proof of Theorem 10.2

By running Mathematica code, Maynard found that when $k = 105$, the largest eigenvalue of $A_1^{-1}A_2$ is $\lambda \approx 4.02 > 4$, so when $\theta = \frac{1}{2}$, one has $M_{105} > 2/\theta$.

Hence, the remaining task is to find an admissible $\mathcal{H}$ of size 105. According to the Andrew Sutherland's lookup table[2], we can take

$\mathcal{H} = \{0, 10, 12, 24, 28, 30, 34, 42, 48, 52, 54, 64, 70, 72, 78, 82, 90, 94, 100, 112, 114, 118, 120, 124, 132, 138, 148, 154, 168, 174, 178, 180, 184, 190, 192, 202, 204, 208, 220, 222, 232, 234, 250, 252, 258, 262, 264, 268, 280, 288, 294, 300, 310, 322, 324, 328, 330, 334, 342, 352, 358, 360, 364, 372, 378, 384, 390, 394, 400, 402, 408, 412, 418, 420, 430, 432, 442, 444, 450, 454, 462, 468, 472, 478, 484, 490, 492, 498, 504, 510, 528, 532, 534, 538, 544, 558, 562, 570, 574, 580, 582, 588, 594, 598, 600\}$.

Therefore, $C_{105} = 600$, so we deduce (10.2):

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

which is a significant improvement to Zhang's bound.

10.8 Conclusion

In this article, we began our discussion from Maynard's variational problem and presented two different approaches for large and small k, eventually producing generalizations and improvements of Zhang's theorem. The publication of Maynard's result effectively caused Polymath8 to relaunch. By combining Maynard's method with Zhang's, the project eventually improved Theorem 10.1 and Theorem 10.2 to

$$\liminf_{n\to\infty}(p_{n+m} - p_n) \leq Ce^{\left(4-\frac{28}{157}\right)m}, \quad \liminf_{n\to\infty}(p_{n+1} - p_n) \leq 246. \tag{10.30}$$

For curious readers, please see Tao's blog post [3].

Dec 17, 2022

[2] `https://math.mit.edu/~primegaps/`

[3] `https://terrytao.wordpress.com/2013/11/19/polymath8b-bounded-intervals-with-many-primes-after-maynard/`

References

[1] E. Bombieri, *On the large sieve*, Mathematika **12** (1965), no. 2, 201–225 (en).

[2] E. Bombieri and H. Davenport, *Small Differences Between Prime Numbers*, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences **293** (1966), no. 1432, 1–18, Publisher: The Royal Society.

[3] Jing-run Chen, *On the representation of a larger even integer as the sum of a prime and the product of at most two primes*, Scientia Sinica **16** (1973), no. 2, 157–176.

[4] Harald Cramér, *Studien über die Nullstellen der Riemannschen Zetafunktion*, Mathematische Zeitschrift **4** (1919), no. 1, 104–130, ISBN: 1432-1823.

[5] ———, *On the order of magnitude of the difference between consecutive prime numbers*, Acta Arithmetica **2** (1936), no. 1, 23–46 (en).

[6] Harold Davenport, *Multiplicative Number Theory*, Graduate Texts in Mathematics, vol. 74, Springer New York, New York, NY, 1980.

[7] P. D. T. A. Elliott and H. Halberstam, *A conjecture in prime number theory*, Symposia Mathematica, vol. 4, Academic Press, London, 1970, pp. 59–72.

[8] P. Erdös, *The difference of consecutive primes*, Duke Mathematical Journal **6** (1940), no. 2, 438–441.

[9] Kevin Ford, Ben Green, Sergei Konyagin, James Maynard, and Terence Tao, *Long gaps between primes*, Journal of the American Mathematical Society **31** (2017), no. 1, 65–105 (en).

[10] J. B. Friedlander and Henryk Iwaniec, *Opera de cribro*, Colloquium publications, no. v. 57, American Mathematical Society, Providence, R.I, 2010.

[11] Daniel A. Goldston, János Pintz, and Cem Yıldırım, *Primes in tuples I*, Annals of Mathematics **170** (2009), no. 2, 819–862.

[12] H. Halberstam and H.-E. Richert, *Sieve methods*, L.M.S. monographs, no. 4, Academic Press, London ; New York, 1974.

[13] G. H. Hardy and J. E. Littlewood, *Some problems of 'Partitio numerorum'; III: On the expression of a number as a sum of primes*, Acta Mathematica **44** (1923), 1–70 (en).

[14] James Maynard, *Small gaps between primes*, Annals of Mathematics **181** (2015), no. 1, 383–413 (en).

[15] D. H. J. Polymath, *Variants of the Selberg sieve, and bounded intervals containing many primes*, Research in the Mathematical Sciences **1** (2014), no. 1, 12 (en).

[16] R. A. Rankin, *The difference between consecutive prime numbers. II*, Mathematical Proceedings of the Cambridge Philosophical Society **36** (1940), no. 3, 255–266, Publisher: Cambridge University Press.

[17] J. Barkley Rosser and Lowell Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois Journal of Mathematics **6** (1962), no. 1, 64–94.

[18] A. Rényi, *Über die Darstellung der geraden Zahlen als Summe einer Primzahl und einer Fast-Primzahl*, Izvestiya Akademii Nauk SSSR. Seriya Matematicheskaya **12** (1948), 57–78 (Russian), Publisher: Izdatel'stvo Nauka, Moskva.

[19] Atle Selberg, *On an Elementary Method in the Theory of Primes*, Norske Vid. Selsk. Forh. Trondhjem. **19** (1947), 64–67.

[20] K. Soundararajan, *Small gaps between prime numbers: The work of Goldston-Pintz-Yildirim*, Bulletin of the American Mathematical Society **44** (2007), no. 1, 1–18.

[21] A. I. Vinogradov, *Über die Dichtehypothese für die Dirichletschen L-Reihen*, Izvestiya Akademii Nauk SSSR. Seriya Matematicheskaya **29** (1965), 903–934 (Russian), Publisher: Izdatel'stvo Nauka, Moskva.

[22] I. M. Vinogradov, *Representation of an odd number as a sum of three primes*, Comptes Rendus (Doklady) de l'Académie des Sciences de l'USSR **15** (1937), 291–294.

[23] Yitang Zhang, *Bounded gaps between primes*, Annals of Mathematics **179** (2014), no. 3, 1121–1174 (en).