The Goldston-Pintz-Yıldırım sieve and some applications

Elchin Hasanalizade
elchin@kth.se

Master's Thesis in Mathematics at KTH
Presented December 14, 2012
Supervisor: Pär Kurlberg

CONTENTS

1. SHORT GAPS BETWEEN PRIMES

1.1 Introduction

Let $p_1 = 2 < p_2 = 3 < p_3 = 5 < \cdots$ be the sequence of all prime numbers, and let $\pi(x) := \sum_{p_n \leqslant x} 1$ be the prime counting function. The prime number theorem implies that $\pi(2x) - \pi(x) \sim x/\log x$ as $x \to \infty$, and in this sense the "average" gap between consecutive primes of size around x is about $\log x$. Of course, such gaps are sometimes less than $\log x$, and one might well ask whether it can happen infinitely often that the gap between consecutive primes is much shorter than average. Indeed, in 2005[1], D. Goldston, J. Pintz and C. Yıldırım [**15**] succeeded in proving that, given any $\epsilon > 0$, there exist arbitrarily large x for which the interval $(x, x + \epsilon \log x]$ contains two or more primes, thereby laying to rest an 80 year-old conjecture of Hardy and Littlewood[2]. In words, the gap between consecutive primes is infinitely often arbitrarily smaller than average.

The work of Goldston-Pintz-Yıldırım represented a major breakthrough in multiplicative number theory, and was the culmination of decades of work by various authors. Prior to the groundbreaking work of Goldston-Pintz-Yıldırım, the most significant progress on the problem of short gaps between primes had been due to Bombieri and Davenport [**2**], who proved that there exist arbitrarily large x for which the interval $(x, x+\eta \log x]$ contains two or more primes, provided $\eta \geqslant (2 + \sqrt{3})/8 = 0.46650\ldots$. More important than their value for η was the fact that they introduced Bombieri's work [**1**] on the large sieve to the problem of gaps between primes. What is now commonly referred to as the Bombieri-Vinogradov theorem, one of the greatest achievements of 20th century number theory, has been an essential feature of all progress in this area, including the work of Goldston-Pintz-Yıldırım.

Of course, the aforementioned conjecture of Hardy and Littlewood, which is now a theorem thanks to Goldston-Pintz-Yıldırım, only hints at the most famous of all conjectures concerning gaps between primes, namely the twin prime

[1] Published in the *Annals of Mathematics* in 2009, "Primes in tuples I" appeared on the preprint archive in August 2005: `http://arxiv.org/abs/math/0508185v1`.

[2] This conjecture was made in the unpublished manuscript [**23**] (see [**2**]).

conjecture. The twin prime conjecture asserts that the gap between consecutive primes is ifinitely often as small as it possibly can be, that is, $p_{n+1} - p_n = 2$ for infinitely many n. Goldston-Pintz-Yıldırım[**15**] were able to use their method to prove, conditionally, that $p_{n+1} - p_n \leqslant 16$ for infinitely many n. The condition here is that the Elliott-Halberstam [**4**] conjecture is true. This conjecture concerns the "level of distribution" of the primes. We will precisely define this notion in the next chapter (see Section 2.2), but loosely speaking, to say that the primes have level of distribution θ is to say that, for large x, the primes are very well distributed among the arithmetic progressions modulo q, for q up to (almost) x^θ. The Bombieri-Vinogradov theorem states that $\theta = 1/2$ is an admissible level of distribution for the primes, and the Elliot-Halberstam conjecture goes beyond this to assert that the primes have level of distribution $\theta = 1$.

Naturally, the announcement of the results of Goldston-Pintz-Yıldırım generated a lot of excitement, and at the time it was hoped that their method would be applicable to many other interesting open questions in multiplicative number theory[3]. Though some remarkable results have indeed been proved using the method of Goldston-Pintz-Yıldırım, this initial optimism has perhaps not been fully borne out. The purpose of this thesis is, therefore, to gather together some of the results that can be seen as applications of the method of Goldston-Pintz-Yıldırım, and to add to this list a novel result concerning the divisor function at consecutive integers (see Theorem 1.8).

1.2 Organization of the thesis

In the next section, we will state the main result of Goldston-Pintz-Yıldırım, as well as some of its extensions. We will then give some applications of their method, including our result on the divisor function at consecutive integers. We will end this chapter with a brief historical survey of the incremental progress towards the establishment of the conjecture of Hardy and Littlewood on short gaps between primes, and beyond.

In Chapter 2, we will begin with a discussion of the prime k-tuple conjecture, including a heuristic that lends credence to it. We will then give an exposition of the proof of the main result of Goldston-Pintz-Yıldırım (Theorem 1.1), highlighting some of the most important ideas involved. We will also give a more-or-less self-contained proof of this theorem.

[3] For instance, see Angel Kumchev's notes from the December 2005 ARCC workshop on gaps between primes: `http://www.aimath.org/WWN/primegaps/WorkshopProblems.pdf`.

In Chapter 3, we will very briefly discuss the result (Theorem 1.4) of Goldston-Pintz-Yıldırım to the effect that there are infinitely often bounded gaps between almost primes. This will be the starting point for the proof of our result on the divisor function at consecutive integers (Theorem 1.8). We will then discuss the Erdős-Mirsky conjecture on the divisor function at consecutive integers, before proving Theorem 1.8.

1.3 The main results

In the introduction we stated that the "average" or gap between consecutive primes around x is about $\log x$, by virtue of the prime number theorem. An alternative way of phrasing this is to say that we "expect" the nth prime gap, $p_{n+1} - p_n$, to be about $\log p_n$. For the prime number theorem implies that $n \sim p_n / \log p_n$ as $n \to \infty$, whence

$$\frac{1}{N} \cdot \sum_{n \leqslant N} \frac{p_{n+1} - p_n}{\log p_n} \sim 1 \quad \text{as} \quad N \to \infty,$$

and consequently we also have

$$\Delta := \liminf_{n \to \infty} \frac{p_{n+1} - p_n}{\log p_n} \leqslant 1. \tag{1.3.1}$$

We can now state the main result of Goldston-Pintz-Yıldırım[**15**] as it is usually given, namely, it states that $\Delta = 0$. This is easily seen to be equivalent to the statement that there exist arbitrarily large x for which $(x, x + \epsilon \log x]$ contains (at least) two primes, for any given $\epsilon > 0$.

Theorem 1.1 (Goldston-Pintz-Yıldırım (2005))**.** *Let $p_1 = 2$, $p_2 = 3$, $p_3 = 5, \ldots$ be the sequence of all primes. We have*

$$\liminf_{n \to \infty} \frac{p_{n+1} - p_n}{\log p_n} = 0. \tag{1.3.2}$$

Theorem 1.2 (Goldston-Pintz-Yıldırım (2005))**.** *Let $p_1 = 2$, $p_2 = 3$, $p_3 = 5, \ldots$ be the sequence of all primes. If the primes have level of distribution $\theta > 1/2$, then there exists a constant $c(\theta)$, depending only on θ, such that*

$$\liminf_{n \to \infty} (p_{n+1} - p_n) \leqslant c(\theta). \tag{1.3.3}$$

Moreover, if $\theta > 20/21$ *(in particular, if the Elliott-Halberstam conjecture is true), then we can take* $c(\theta) = 16$.

In Chapter 2 we will prove both of the above theorems, except in Theorem 1.2, in the case $\theta > 20/21$, we will only prove that $c(\theta) = 20$ is admissible.

Goldston-Pintz-Yıldırım also generalized (1.3.1) for primes in arithmetic progressions $a \bmod q$ with a fixed q and any a with $(a, q) = 1$ and obtained analogue of (1.3.3) for E_2 numbers (i.e. numbers which are product of exactly two distinct primes).

Theorem 1.3 (Goldston-Pintz-Yıldırım (2006))**.** *Let a coprime pair of integers* a *and* q *be given, and let* $p'_1 < p'_2 < \cdots$ *be the sequence of all primes in the arithmetic progression* $a \bmod q$. *Then*

$$\liminf_{n\to\infty} \frac{p'_{n+1} - p'_n}{\log p'_n} = 0. \tag{1.3.4}$$

Theorem 1.4 (Goldston-Graham-Pintz-Yıldırım (2009))**.** *Let* $q_1 = 6$, $q_2 = 10$, $q_3 = 14, \ldots$ *be the sequence of all numbers, which are products of exactly two distinct primes. Then*

$$\liminf_{n\to\infty} (q_{n+1} - q_n) \leqslant 26. \tag{1.3.5}$$

An old conjecture of Chowla asserts that there exist infinitely many pairs of consecutive primes p_k, p_{k+1} such that $p_k \equiv p_{k+1} \equiv a \bmod q$. In 2000, Shiu [**37**] proved this conjecture. Combining the ideas of Shiu and Goldston-Pintz-Yıldırım, Freiberg [**9**] improved their result in another direction

Theorem 1.5 (Freiberg (2010))**.** *Let* $p_1 = 2$, $p_2 = 3$, $p_3 = 5, \ldots$ *be the sequence of all primes. Let a coprime pair of integers* a *and* q *be given, and fix any* $\epsilon > 0$. *Then there exist infinitely many* n *such that*

$$p_{n+1} - p_n < \epsilon \log p_n \quad \text{and} \quad p_n \equiv p_{n+1} \equiv a \bmod q.$$

Applying the Goldston-Pintz-Yıldırım construction to the set of powers of a fixed integer, Friedlander and Iwaniec [**10**] obtained the following result

Theorem 1.6 (Friedlander-Iwaniec (2010))**.** *Let* $a \geq 2$ *be an integer. There are infinitely many* $n \in \mathbb{N}$ *that have two distinct representations*

$$n = p + a^m = p' + a^{m'}, \tag{1.3.6}$$

where p, p' are primes and $m, m' \in \mathbb{N}$.

One of the interesting applications of the Goldston-Pintz-Yıldırım method is the solution of Erdős-Mirsky conjecture on the divisor function at consecutive integers. Erdős conjectured that $d(n) = d(n+1)$ infinitely often. In 1984 Heath-Brown [**24**] succeeded in proving this conjecture. Using the method, which yielded the existence of small gaps between primes and bounded gaps for E_2 numbers, Goldston-Graham-Pintz-Yıldırım [**19**] were able to show the following stronger variant of the Erdős-Mirsky conjecture:

Theorem 1.7 (Goldston-Graham-Pintz-Yıldırım (2010))**.** *For any positive integer A with $24|A$, there are infinitely many integers n with*

$$d(n) = d(n+1) = A. \tag{1.3.7}$$

Let $\mathbf{E}$ denote the set of limit points of the sequence $\{d(n)/d(n+1) : n \in \mathbb{N}\}$, and let $\mathbf{L}$ denote the set of limit points of $\{\log(d(n)/d(n+1)) : n \in \mathbb{N}\}$. Erdős conjectured [**6**] that $\mathbf{E} = [0, \infty]$, or equivalently, $\mathbf{L} = [-\infty, \infty]$.

We can now formulate the main result of this thesis

Theorem 1.8. *(a) For any number $x \geqslant 0$ we have*

$$|\mathbf{L} \cap [0, x]| \geqslant \frac{x}{3} \quad \textit{and} \quad |\mathbf{L} \cap [-x, 0]| \geqslant \frac{x}{3}, \tag{1.3.8}$$

where $|\cdot|$ denotes the Lebesgue measure. Moreover, there exists a number $A \geqslant 0$ such that, for any number $x \geqslant A$, we have

$$|\mathbf{L} \cap [0, x]| \geqslant \frac{x - A}{2} \quad \textit{and} \quad |\mathbf{L} \cap [-x, 0]| \geqslant \frac{x - A}{2}. \tag{1.3.9}$$

(b) There exists a number $B \geqslant 0$ such that, for any number $x > 0$, we have

$$|\mathbf{E} \cap (0, x]| \geqslant \frac{x}{B+1}. \tag{1.3.10}$$

It should be mentioned that Kan and Shan [**30**],[**31**] proved that for any real number $\alpha > 0$, either α or 2α is in $\mathbf{E}$, and deduced from this that $|\mathbf{E} \cap [0, x]| \geqslant x/3$ for $x > 0$.

1.4 A survey of results

Hardy and Littlewood [**23**] were the first who investigated

Tab. 1.1:

Year	Author	Δ_1
1926	Hardy,Littlewood [**23**]	$\leq \frac{2}{3}$
1940	Erdös [**5**]	< 1
1954	Ricci [**35**]	$\leq \frac{15}{16}$
1965	Wang Yuan, Hsiek, Yu [**40**]	$\leq \frac{29}{32}$
1966	Bombieri, Davenport [**2**]	$\leq \frac{2+\sqrt{3}}{8} = 0.4665...$
1972	Pilt'jaĭ [**34**]	$\leq \frac{2\sqrt{2}-1}{4} = 0.4571...$
1975	Uchiyama [**39**]	$\leq \frac{9-\sqrt{3}}{16} = 0.4542...$
1977	Huxley [**26**]	$\leq 0.4425...$
1984	Huxley [**27**]	$\leq 0.4394...$
1988	Maier [**32**]	$\leq 0.2484...$
2005	Goldston,Pintz,Yildirim [**15**]	0

$$\Delta_\nu := \liminf_{n\to\infty} \frac{p_{n+\nu} - p_n}{\log p_n}, \tag{1.4.1}$$

where p_n is the nth prime. The prime number theorem trivially gives $\Delta_\nu \leq \nu$. Using the circle method and under assumption of the Generalized Riemann Hypothesis (GRH) for Dirichlet L-functions, they proved

$$\Delta_1 \leq \frac{2}{3}. \tag{1.4.2}$$

Many authors improved this bound (see Table 1).

In 1940, the first step toward showing unconditionally $\Delta_1 = 0$ was taken by Erdős [**5**], who obtained the bound $\Delta_1 < 1 - c$, where c is an unspecified explicitly calculable constant.

In 1965 Bombieri and Davenport [**2**] proved unconditionally that

$$\Delta_\nu \leq \nu - \frac{1}{2}. \tag{1.4.3}$$

Combining their method with that of Erdős, for the case $\nu = 1$ they obtained

$$\Delta_1 \leq 0.4665... \tag{1.4.4}$$

This was possible by replacement of GRH in Hardy-Littlewood method by the Bombieri-Vinogradov theorem, one of the most powerful tools in analytic number theory.

Later Huxley [**26**] refined (1.4.4) to

$$\Delta_1 \leq 0.44254... \tag{1.4.5}$$

and for general $\nu \geq 2$

$$\Delta_\nu \leq \nu - \frac{5}{8} + O\left(\frac{1}{\nu}\right). \tag{1.4.6}$$

In 1988 Maier [**32**], using his matrix method, improved Huxley's result (1.4.4) to $\Delta_1 \leq e^{-\gamma} \cdot 0.4425.. = 0.2484...$, where $\gamma = \lim_{n\to\infty}(\sum_{k=1}^{n} \frac{1}{k} - \log n)$ is Euler's constant.

Finally, in 2005 Goldston, Pintz and Yildirim proved that for any $\epsilon > 0$, there exist infinitely many n such that $p_{n+1} - p_n < \epsilon \log p_n$, i.e. $\Delta_1 = 0$. Moreover, for general $\nu \geq 1$ Goldston-Pintz-Yıldırım[**15**, Theorem3] proved the following

Theorem 1.9 (Goldston-Pintz-Yıldırım (2009))**.** *Suppose that the primes have level of distribution θ. The for $\nu \geq 2$*

$$\Delta_\nu \leq (\sqrt{\nu} - \sqrt{2\theta})^2, \tag{1.4.7}$$

and unconditionally for $\nu \geq 1$ we have

$$\Delta_\nu \leq (\sqrt{\nu} - 1)^2. \tag{1.4.8}$$

By combining Maier's method with their own, Goldston-Pintz-Yıldırım refined (1.4.8) to

$$\Delta_\nu \leq e^{-\gamma} \cdot (\sqrt{\nu} - 1)^2. \tag{1.4.9}$$

In [**16**] the result $\Delta_1 = 0$ was strengthened to

$$\liminf_{n\to\infty} \frac{p_{n+1} - p_n}{(\log p_n)^{1/2}(\log\log p_n)^2} < \infty. \tag{1.4.10}$$

In fact, Goldston-Pintz-Yıldırım [**16**] obtained the following general result for very sparse sequences

Theorem 1.10 (Goldston-Pintz-Yıldırım (2011))**.** *Let $\mathcal{A} \subseteq \mathbb{N}$ be an arbitrary sequence of integers satisfying*

$$\mathcal{A}(N) = |\{n; n \leq N, n \in \mathcal{A}\}| > C\sqrt{\log N}(\log\log N)^2 \quad \textit{for} \quad N > N_0, \tag{1.4.11}$$

where C is an appropriate absolute constant. Then infinitely many elements of $\mathcal{A}-\mathcal{A}$ can be written as the difference of two primes, that is,

$$|(\mathcal{P}-\mathcal{P})\cap(\mathcal{A}-\mathcal{A})|=\infty, \tag{1.4.12}$$

where $\mathcal{P}$ denotes the set of primes. (Here we use the notation $\mathcal{A}-\mathcal{B}=\{a-b: a\in\mathcal{A}, b\in\mathcal{B}\}$)

Some sequences for which this method works are:

$$\mathcal{A}=\{k^m\}_{m=1}^{\infty},\quad k\geq 2\quad \text{fixed}\quad (k\in\mathbb{N}),$$

$$\mathcal{A}=\{k^{x^2+y^2}\}_{x,y=1}^{\infty},\quad k\geq 2\quad \text{fixed}\quad (k\in\mathbb{N}).$$

Let S denote the set of limit points of the sequence $\{\frac{d_n}{\log n}\}$, where $d_n = p_{n+1}-p_n$. It is conjectured that $S=\mathbb{R}^+\cup\{\infty\}$. Goldston-Pintz-Yıldırım's result implies that $0\in S$ (as was shown by Westzynthius $\infty\in S$). In 1988 Hildebrand and Maier proved that the sequence $\{\frac{d_n}{\log n}\}$ has arbitrarily large finite limit points. In fact, they proved the following general statement

Theorem 1.11 (Hildebrand-Maier (1988))**.** *Let k be a positive integer, and let $S^{(k)}$ be the set of limit points in $\mathbb{R}^k$ of the sequence*

$$\left(\frac{d_n}{\log n},\ldots,\frac{d_{n+k-1}}{\log n}\right)\quad (n=1,2,3,...)$$

Then we have, for any sufficiently large number T,

$$|S^{(k)}\cap[0,T]^k|\geq c(k)T^k,$$

where $|\cdot|$ denotes the Lebesgue measure in $\mathbb{R}^k$ and $c(k)$ is a positive constant depending only on k.

Goldston-Pintz-Yıldırım [**20**] have recently proved, by adding a new idea to their method, that short gaps between primes make up a positive proportion of all gaps between primes, in the sense of the following theorem.

Theorem 1.12 (Goldston-Pintz-Yıldırım (2011))**.** *There exists a constant $c>0$ such that, for any given $\eta>0$, we have*

$$\frac{1}{\pi(x)}\sum_{\substack{p_n\leqslant x\\ p_{n+1}-p_n<\eta\log p_n}} 1 \gtrsim e^{-c/\eta^6},\quad x\to\infty. \tag{1.4.13}$$

(The notation $f(x) \gtrsim g(x)$ stands for $\limsup_{x\to\infty} g(x)/f(x) \leqslant 1$.)

Even under the strongest assumptions the Goldston-Pintz-Yıldırım method has limitations. It fails to prove that there are three or more primes in admissible k-tuples with large enough k. In this connection the following result of Maynard [**33**] is of interest

Theorem 1.13 (Maynard(2012))**.** *Let $k \geq 1$ be an integer and assume that the primes have level of distribution $1/2 < \theta < 0.99$. Let*

$$r(\theta, k) = \frac{240k^2}{(2\theta - 1)^2}.$$

Then there exists constant $C(\theta, k)$ such that are infinitely many integers n for which the interval $[n, n + C(\theta, k)]$ contains two primes and k integers, each with at most $r(\theta, k)$ prime factors.

2. THE METHOD OF GOLDSTON-PINTZ-YıLDıRıM

Notation

Throughout this thesis, we use the notation $F \ll G$, $G \gg F$ and $F = O(G)$ as shorthand for $|F| \leqslant c|G|$ for some constant c. Unless stated to the contrary, c shall denote an absolute constant. When we write $F \ll_A G$, $G \gg_A F$ or $F = O_A(G)$, we mean that the implied constant c depends on A. We may sometimes write $F \asymp G$ to denote that $F \ll G \ll F$.

2.1 The prime k-tuple conjecture

One of the greatest unsolved problems in mathematics is to show that there exist infinitely many twin primes, that is pairs of prime numbers that differ by 2 (like 5 and 7, 11 and 13, etc.). This a special case of the k-tuple conjecture, which we will now explain.

Let us consider k distinct linear forms

$$\mathcal{H} = \{g_1x + h_1, g_2x + h_2, \ldots, g_kx + h_k\}, \quad g_ix + h_i \in \mathbb{Z}[x], \quad g_i \geq 1. \qquad (2.1.1)$$

If, for a given $x = n \in \mathbb{N}$, $g_in + h_i$ is a prime for each i,we say that (2.1.1) is a prime k-tuple. It is natural to ask how often (2.1.1) is a prime tuple for $n \in \mathbb{N}$. Consider, for example, the tuple $(n, n+1)$. For $n > 2$ one of the numbers is an even. Likewise, the tuple $(n, n+2, n+4)$ can't be a prime tuple since for $n \geq 1$ one of the components is the multiple of 3. On the other hand, we expect that there are infinitely many prime tuples of the form $(n, n+2)$. This is the twin prime conjecture. In general, (2.1.1) can be a prime for at most finitely many n if $\Omega_p(\mathcal{H}) = \mathbb{Z}/p\mathbb{Z}$ for some prime p, where

$$\Omega_p(\mathcal{H}) := \{n \bmod p : \prod_{i=1}^{k}(g_in + h_i) \equiv 0 \bmod p\}.$$

In other words, (2.1.1) is a prime k-tuple if

$$|\Omega_p(\mathcal{H})| < p \qquad \text{for all primes } p. \tag{2.1.2}$$

We say that $\mathcal{H}$ is an admissible k-tuple if (2.1.2) holds. Hardy and Littlewood [**22**] conjectured that admissible tuples will infinitely often be prime tuples.

The prime number theorem states that the number of primes less than N is around $\frac{N}{\log N}$. So if we choose a random integer m from the interval $[1, N]$, the probability that m is a prime is around to $\frac{1}{\log N}$. In other words, the primes behave like independent random variables $X(n)$ with $X(n) = 1$ (n is prime) with probability $\mathbb{P}(X(n) = 1) = \frac{1}{\log n}$ and $X(n) = 0$ (n is composite) with probability $\mathbb{P}(X(n) = 1) = 1 - \frac{1}{\log n}$. This is known as the Cramer's model. These ideas can be used, for example, to predict the probability that, given a prime number p_n, the next prime lies somewhere between $p_n + \alpha \log p_n$ and $p_n + \beta \log p_n$ with $0 \leq \alpha < \beta$. Thus we want $p_n + 1, \ldots, p_n + h - 1$ to be composite, and $p_n + h$ to be prime for some integer $h \in [\alpha \log p_n, \beta \log p_n]$. According to our heuristics, this occurs with probability

$$\begin{aligned}
&\sum_{\alpha \log n \leq h \leq \beta \log n} \prod_{i=1}^{h-1} \left(1 - \frac{1}{\log(p_n + j)}\right) \frac{1}{\log(p_n + h)} \\
&\sim \sum_{\alpha \log n \leq h \leq \beta \log n} \left(1 - \frac{1}{\log n}\right)^{h-1} \frac{1}{\log n}
\end{aligned}$$

since, as $p_n \sim n \log n$ and $i < h \ll \log n$, $\log(p_n + i) \sim \log n$. This gives

$$\sum_{\alpha \log n \leq h \leq \beta \log n} \left(1 - \frac{1}{\log n}\right)^{h-1} \frac{1}{\log n} \sim \sum_{\alpha < h/\log n < \beta} e^{-h/\log n} \frac{1}{\log n} \sim \int_{\alpha}^{\beta} e^{-t}\, \mathrm{d}t,$$

for large n since the middle sum looks like a Riemann sum approximation to the right hand side integral.

Now consider

$$g_1 n + h_1, g_2 n + h_2, \ldots, g_k n + h_k, \qquad 1 \leq g_i n + h_i \leq N.$$

By our heuristics, for each n the number of $g_i n + h_i$ should be prime with probability $\frac{1}{\log N}$. If the probabilities that each term is a prime are independent then the whole set should a be prime with a probability $\frac{1}{(\log N)^k}$. But this is not true in general and we need an extra factor to correct our naive estimation. The probability that each element of (2.1.2) is coprime to p is $1 - \frac{|\Omega_p(\mathcal{H})|}{p}$ in contrast to $(1-\frac{1}{p})^k$, which would be true if the events $p \mid g_i n + h_i$, $p \mid g_j n + h_j$ were pairwise independent. So the correction factor for p is $\left(1 - \frac{|\Omega_p(\mathcal{H})|}{p}\right)\left(1-\frac{1}{p}\right)^{-k}$ and we have to multiply $\frac{1}{(\log N)^k}$ with the product of all correction factors $\bmod p$. Denote this product by

$$\mathfrak{S}(\mathcal{H}) := \prod_p \left(1 - \frac{|\Omega_p(\mathcal{H})|}{p}\right)\left(1-\frac{1}{p}\right)^{-k} \tag{2.1.3}$$

This is a so-called singular series[1]. If $\mathcal{H}$ is admissible then $\mathfrak{S}(\mathcal{H}) \neq 0$. Indeed, for large p $|\Omega_p(\mathcal{H})| = k$, and for such p, $\prod_{p>k}\left(1-\frac{k}{p}\right)\left(1-\frac{1}{p}\right)^{-k} = \prod_{p>k}\left(1 + O_k\left(\frac{1}{p^2}\right)\right)$ converges and $\mathfrak{S}(\mathcal{H}) \gg_k 1$. Now we can state the prime k-tuple conjecture in the quantitative form:

Conjecture 1 (Hardy–Littlewood [23]). *Let* $\mathcal{H} = (g_1x+h_1, g_2x+h_2, \ldots, g_kx+h_k)$ *be an admissible* k*-tuple. Then*

$$\#\{n \le x : g_1n+h_1, g_2n+h_2, \ldots, g_kn+h_k \quad \textit{are prime}\} \sim \mathfrak{S}(\mathcal{H})\frac{x}{(\log x)^k}, \quad x \to \infty.$$

2.2 The level of distribution of the primes

The level of distribution of the primes plays an important rôle in the proof of theorems 1.1 and 1.2, and indeed most of the significant results on short gaps between primes. We will discuss the level of distribution of the primes in this section, and precisely define this notion.

The level of distribution of the primes concerns the distribution of primes in arithmetic progressions. Given a modulus $q \geqslant 1$ and an integer $a \geqslant 1$,

$$\pi(N; q, a) := \sum_{\substack{p \leqslant N \\ p \equiv a \bmod q}} 1$$

is the prime counting function for the primes in the arithmetic progression a mod q. For obvious reasons, this progression contains more than one prime only if q and a are coprime, and there is no apparent reason for suspecting that the primes might be biased towards any particular arithmetic progression a mod q

[1] Hardy and Littlewood originally defined this as a series.

with $(q,a)=1$. In other words, we expect the primes to be equidistributed among the possible arithmetic progressions to the modulus q. Thus, we expect that

$$\pi(N;q,a) \sim \frac{\pi(N)}{\phi(q)}, \quad N \to \infty,$$

where $\phi(q)$ — Euler's totient function of q — counts the number of congruence classes a mod q with $(q,a)=1$.

The prime number theorem for arithmetic progressions states that this expectation is indeed the truth. We state it in the following form: for any $A>0$ we have, for fixed q and a,

$$\sum_{\substack{p\leqslant N \\ p\equiv a \bmod q}} \log p = \frac{N}{\phi(q)} \cdot \left(1 + O\left(\frac{1}{(\log N)^A}\right)\right), \tag{2.2.1}$$

provided $(q,a)=1$. Recall that the prime number theorem for arithmetic progressions asserts that $\pi(N;q,a) = \frac{\mathrm{Li}(N)}{\phi(q)} + O\left(Ne^{-c(\log N)^{1/2}}\right)$ which by partial summation, is equivalent to the more elegant expression $\vartheta(N;q,a) := \sum_{\substack{p\leqslant N \\ p\equiv a \bmod q}} \log p = \frac{N}{\phi(q)} + O\left(Ne^{-c(\log N)^{1/2}}\right)$, where c is a constant (not always the same one). The Siegel-Walfisz theorem [**3**, §22] states that (2.2.1) holds uniformly for $q \geqslant 1$ and a with $(q,a)=1$. This, however, is only non-trivial for $q \ll (\log x)^A$.

The Grand Riemann Hypothesis (GRH) implies

$$\sum_{\substack{p\leqslant x \\ p\equiv a \bmod q}} \log p = \frac{x}{\phi(q)} + O\left(x^{\frac{1}{2}}(\log x)^2\right), \tag{2.2.2}$$

where the constant is absolute and this is non-trivial for $q \ll x^{\frac{1}{2}}(\log x)^{-2}$ [**29**].

Although GRH is unattackable by current methods of the prime number theory, there is a good approximation for it.

Let

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

and consider the function

$$\vartheta(N;q,a) = \sum_{\substack{n\leqslant N \\ n\equiv a \bmod q}} \vartheta(n).$$

Theorem (Bombieri-Vinogradov) *For any $A > 0$ there exist a $B = B(A)$ such that, for $Q = N^{\frac{1}{2}}(\log N)^{-B}$*

$$\sum_{q \leqslant Q} \max_{\substack{a \\ (a,q)=1}} |\vartheta(N; q, a) - \frac{N}{\phi(q)}| \ll_A \frac{N}{(\log N)^A}. \tag{2.2.3}$$

(For a proof see [**1**]). If (2.2.3) holds for any $A > 0$ and any $\epsilon > 0$ with $Q = N^{\theta - \epsilon}$, then we say that the primes have *level of distribution θ* . Thus by Bombieri-Vinogradov theorem the primes have level of distribution 1/2. Elliott and Halberstam [**4**] conjectured that the primes have level of distribution 1.

Let $\mathcal{H} = \{0, 2m\}$. This is clearly admissible for every k, and let $\Lambda(n)$ denote the von Mangoldt function which equals $\log p$ if $n = p^s$, $s \geq 1$, and zero otherwise. The Hardy-Littlewood conjecture asserts that

$$\sum_{n \leq N} \Lambda(n)\Lambda(n + 2m) = \mathfrak{S}(\mathcal{H})N + o\left(N\right), \quad N \to \infty. \tag{2.2.4}$$

In the above sum, we substitute $\Lambda(n)$ with the truncated von Mangoldt function

$$\Lambda_R(n) = \sum_{\substack{d|n \\ d \leqslant R}} \mu(d) \log\left(\frac{R}{d}\right).$$

For $R \geq n$ obviously $\Lambda_R(n) = \Lambda(n)$. Then

$$\begin{aligned} \sum_{n \leq N} \Lambda_R(n)\Lambda(n + 2m) &= \sum_{n \leq N} \left(\sum_{\substack{d|n \\ d \leqslant R}} \mu(d) \log\left(\frac{R}{d}\right) \right) \Lambda(n + 2m) \\ &= \sum_{d \leq R} \mu(d) \log\left(\frac{R}{d}\right) \sum_{\substack{n \leqslant N+2m \\ n \equiv 2m \bmod d}} \Lambda(n) \\ &= \sum_{d \leq R} \mu(d) \log\left(\frac{R}{d}\right) \Psi(N + 2m; d, 2m), \end{aligned} \tag{2.2.5}$$

where

$$\Psi(N; q, a) = \sum_{\substack{n \leqslant N \\ n \equiv a \bmod q}} \Lambda(n).$$

The contribution from the terms with $(d, 2m) > 1$ is $O\left(R \log N\right)$; thus the right hand side of (2.2.5) is equal to

$$
\begin{aligned}
&(N+2m) \sum_{\substack{d\leqslant R\\(d,2m)=1}} \frac{\mu(d)}{\phi(d)} \log\left(\frac{R}{d}\right)\\
&+\sum_{\substack{d\leqslant R\\(d,2m)=1}} \mu(d)\log\left(\frac{R}{d}\right)\left(\Psi(N+2m;d,2m)-\frac{N+2m}{\phi(d)}\right)+O\left(R\log N\right).
\end{aligned}
\tag{2.2.6}
$$

Estimation of the second sum in (2.2.6) depends on the level of distribution of the primes. For large R (hence, for large θ) the sum (2.2.5) is quite near to the expected asymptotic value (2.2.4) (see [**12**]).

2.3 The Goldston-Pintz-Yıldırım sieve

Consider the following weighted sum

$$
\mathcal{S} := \sum_{N<n\leqslant 2N}\left(\sum_{h_i\in\mathcal{H}} \vartheta(n+h_i)-\log(3N)\right)\Lambda_R(n;\mathcal{H},\ell)^2, \tag{2.3.1}
$$

where $\Lambda_R(n;\mathcal{H},\ell)$ is a non-negative weight to be defined later. If $\mathcal{S}>0$ then there exist at least two integers $h_i,h_j\in\mathcal{H}$, such that $n+h_i, n+h_j$ are primes. Thus we can conclude that $\liminf_{n\to\infty}(p_{n+1}-p_n)\leq \max_{1\leq i<j\leq k}|h_j-h_i|$.

The main idea consists in appropriate choice for the $\Lambda_R(n;\mathcal{H},\ell)$ such that it detects n for which $\{n+h_1,n+h_2,\dots,n+h_k\}$ contains at least two distinct primes. One candidate for this role is the k-th generalized von Mangoldt function

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

which is 0 if $\omega(n)>k$ (of course, it detects also prime powers, but their effect is negligible). This can be verified using the recurrence formula $\Lambda_k(n)=\Lambda_{k-1}(n)\log n+\Lambda_{k-1}(n)*\Lambda(n)$, where $*$ denotes the Dirichlet convolution. Thus our prime tuple detecting function is

$$
\Lambda_k(n;\mathcal{H}) := \frac{1}{k!}\Lambda_k(P_{\mathcal{H}}(n)). \tag{2.3.2}
$$

where $P_{\mathcal{H}}(n)=\prod_{i=1}^k(n+h_i)$ is the polynomial corresponding to the k-tuple $\mathcal{H}=\{x+h_1,x+h_2,\dots,x+h_k\}$ and the factor $\frac{1}{k!}$ is for simplification of estimates.

Instead of (2.3.2), we consider the truncated divisor sum

$$\Lambda_R(n;\mathcal{H}) = \frac{1}{k!}\sum_{\substack{d|P_{\mathcal{H}}(n)\\ d\leqslant R}} \mu(d)\left(\log\frac{R}{d}\right)^k, \tag{2.3.3}$$

which is smooth approximation to (2.3.2) and show the same behaviour as $\Lambda_k(n;\mathcal{H})$ on some averages.

This approximation is not adequate to prove $\Delta = 0$ (with the aid of (2.3.3) Goldston and Yıldırım [**14**] obtained only $\Delta \le \frac{\sqrt{3}-1}{2}$).

Rather than approximate only prime tuples, we consider tuples with many primes in components, that is $\omega(P_{\mathcal{H}}(n)) \le k+\ell$, where $0 \leqslant \ell \leqslant k$ and specify

$$\Lambda_R(n;\mathcal{H},\ell) = \frac{1}{(k+\ell)!}\sum_{\substack{d|P_{\mathcal{H}}(n)\\ d\leqslant R}} \mu(d)\left(\log\frac{R}{d}\right)^{k+\ell}. \tag{2.3.4}$$

The following two lemmas collect the main results on $\Lambda_R(n;\mathcal{H},\ell)$ [**13**, Lemma 1, Lemma 2]:

Lemma 1. *Let $\mathcal{H} = \{h_1, h_2, \ldots, h_k\} \subseteq [1,H]\cap\mathbb{Z}$ with $h_i \neq h_j$ for $i \neq j$; k, ℓ are arbitrarily, but fixed positive integers, $|\mathcal{H}| = k$. Provided $H \ll \log N \ll \log R \leqslant \log N$ and $R \le N^{1/2}/(\log N)^C$ hold with a sufficiently large $C > 0$ depending only on k and ℓ, we have*

$$\begin{aligned}&\sum_{N<n\le 2N} \Lambda_R(n;\mathcal{H},\ell)^2\\ &= \frac{\mathfrak{S}(\mathcal{H})}{(k+2\ell)!}\binom{2\ell}{\ell}N(\log R)^{k+2\ell} + O(N(\log N)^{k+2\ell-1}(\log\log N)^c).\end{aligned} \tag{2.3.5}$$

Lemma 2. *Assume $H \ll \log N \ll \log R \leqslant \log N$ hold and primes have level of distribution θ. Then for $R \leqslant N^{\theta/2}$*

$$\sum_{N<n\le 2N} \vartheta(n+h)\Lambda_R(n;\mathcal{H},k+\ell)^2 = \begin{cases} \frac{\mathfrak{S}(\mathcal{H}\cup\{h\})}{(k+2\ell)!}\binom{2\ell}{\ell}N(\log R)^{k+2\ell} & \\ +O(N(\log N)^{k+2\ell-1}(\log\log N)^c) & \text{if } h \notin \mathcal{H}\\ & \\ \frac{\mathfrak{S}(\mathcal{H})}{(k+2\ell+1)!}\binom{2(\ell+1)}{\ell+1}N(\log R)^{k+2\ell+1} & \\ +O(N(\log N)^{k+2\ell}(\log\log N)^c) & \text{if } h \in \mathcal{H}.\end{cases} \tag{2.3.6}$$

Proof of Theorem 1.2. Suppose $\theta = 1/2+\delta$, $\delta > 0$. Fix an $\epsilon' \in (0, \delta/4)$ so that $\theta/2 - \epsilon' \geqslant (1+\delta)/4$. Fix integers $k = k(\theta) = k(\delta)$ and $\ell = [\sqrt{k}]$ such that

$$\frac{2(2\ell+1)}{\ell+1} \cdot \frac{k}{k+2\ell+1}\left(\frac{\theta}{2} - \epsilon'\right) = 1+\epsilon,$$

for some $\epsilon > 0$. Then set $R = N^{\theta/2-\epsilon'}$. Let $\mathcal{H} = \{h_1, \ldots, h_k\}$, where the h_i are any integers such that $\mathcal{H}$ is admissible.

Now applying (2.3.5) and (2.3.6) we get as $N \to \infty$

$$\begin{aligned}
\mathcal{S} :&= \sum_{N<n\leqslant 2N}\left(\sum_{h_i\in\mathcal{H}} \vartheta(n+h_i) - \log(3N)\right)\Lambda_R(n;\mathcal{H},\ell)^2 \\
&= k\binom{2\ell+2}{\ell+1}\frac{(\log R)^{k+2\ell+1}}{(k+2\ell+1)!}(\mathfrak{S}(\mathcal{H})+o(1))N \\
&- \log 3N\binom{2\ell}{\ell}\frac{(\log R)^{k+2\ell}}{(k+2\ell)!}(\mathfrak{S}(\mathcal{H})+o(1))N \\
&= \Big(\frac{2k}{k+2\ell+1}\frac{2\ell}{\ell+1}\log R - \{1+o(1)\}\log 3N\Big)\frac{(\log R)^{k+2\ell}}{(k+2\ell)!}\binom{2\ell}{\ell}\mathfrak{S}(\mathcal{H})N \\
&= (\epsilon - o(1))\frac{(\log R)^{k+2\ell}}{(k+2\ell)!}\binom{2\ell}{\ell}\mathfrak{S}(\mathcal{H})N\log N. \qquad (2.3.7)
\end{aligned}$$

The term inside the brackets is greater than a positive constant. Assuming $\theta > 20/21$, we can take $\ell = 1$ and $k = 7$. Since the 7-tuple $\{0, 2, 6, 8, 12, 18, 20\}$ (or $\{0, 2, 8, 12, 14, 18, 20\}$) is admissible, we have $\liminf_{n\to\infty}(p_{n+1} - p_n) \leq 20$. □

Proof of Theorem 1.1 . Now consider the modified weighted sum

$$\mathcal{S}_1 := \sum_{n=N+1}^{2N}\left(\sum_{1\leqslant h_0\leqslant H}\vartheta(n+h_0) - \log 3N\right)\sum_{\substack{\mathcal{H}\subseteq\{1,\ldots,H\}\\ |\mathcal{H}|=k}}\Lambda_R(n;\mathcal{H},\ell)^2. \qquad (2.3.8)$$

For any fixed n, the inner sum is negative unless there exist at least two integers $1 \leq h_i < h_j \leq H$, such that $n+h_i, n+h_j$ are primes. Fix $\epsilon > 0$ first of all and let N be a parameter. Set $h = \epsilon \log N$. Set $\epsilon' = \epsilon/100$, which we may suppose is less than $1/4$. Set $R = N^{1/4-\epsilon'}$. Fix integers $k = k(\epsilon)$ and $\ell = [\sqrt{k}]$ such that

$$\frac{2(2\ell+1)}{\ell+1}\cdot\frac{k}{k+2\ell+1}\left(\frac{1}{4} - \epsilon'\right) \geqslant 1 - \frac{\epsilon}{2}.$$

To evaluate (2.3.8), we need a result of Gallagher [**11**], namely that

$$\sum_{\substack{\mathcal{H}\subseteq\{1,\dots,H\}\\|\mathcal{H}|=k}} \mathfrak{S}(\mathcal{H}) = (1+o(1))H^k. \tag{2.3.9}$$

Applying (2.3.5), (2.3.6) and (2.3.9), we obtain

$$\begin{aligned}
\mathcal{S} := &\sum_{\substack{\mathcal{H}\subseteq\{1,\dots,H\}\\|\mathcal{H}|=k}} \Bigg(k\frac{2}{(k+2\ell+1)!}\binom{2\ell+1}{\ell}\mathfrak{S}(\mathcal{H})N(\log R)^{k+2\ell+1} \\
&+ \sum_{\substack{1\leqslant h_0\leqslant H\\ h_0\notin\mathcal{H}}} \frac{1}{(k+2\ell)!}\binom{2\ell}{\ell}\mathfrak{S}(\mathcal{H}\cup\{h_0\})N(\log R)^{k+2\ell} \\
&- \log 3N\frac{1}{(k+2\ell)!}\binom{2\ell}{\ell}\mathfrak{S}(\mathcal{H})N(\log R)^{k+2\ell} + O(N(\log N)^{k+2\ell}(\log\log N)^c)\Bigg) \\
&= \left(\frac{2k}{k+2\ell+1}\frac{2\ell+1}{\ell+1}\log R + H - \log 3N\right)\frac{1}{(k+2\ell)!}\binom{2\ell}{\ell}NH^k(\log R)^{k+2\ell} \\
&+ o\left(NH^k(\log N)^{k+2\ell+1}\right), \quad N\to\infty.
\end{aligned} \tag{2.3.10}$$

Hence, there are exist at least two primes in the interval $(n, n+H]$, $N < n \leqslant 2N$. Therefore, $\liminf_{n\to\infty} \frac{p_{n+1}-p_n}{\log p_n} = 0$, as desired. □

2.4 Proof of the key estimates

In this section we shall outline the proofs of Lemma 1 and Lemma 2. First of all we extend to arbitrary squarefree moduli d the definition of $\Omega_p(\mathcal{H})$. Let $\mathbb{Z}/d\mathbb{Z}$ be the ring of integers mod d and define

$$\Omega_d(\mathcal{H}) = \{n\in\mathbb{Z}/d\mathbb{Z} : \quad P_{\mathcal{H}}(n)\equiv 0 \pmod d\},$$

where $P_{\mathcal{H}}(n) = \prod_{i=1}^k (n+h_i)$ is the polynomial corresponding to the k-tuple $\mathcal{H} = \{x+h_1, x+h_2, \dots, x+h_k\}$. We have to state some auxilary propositions.

Proposition 1. (a) *We have*

$$\sum_{\substack{N<n\leqslant 2N\\ d\mid P_{\mathcal{H}}(n)}} 1 = |\Omega_d(\mathcal{H})|\left(\frac{N}{d} + O(1)\right), \tag{2.4.1}$$

(b) *For every $A>0$ and $d\leqslant(\log N)^A$, we have, given h_0 and squarefree d*

$$\sum_{\substack{N<n\leqslant 2N\\ d|P_{\mathcal{H}}(n)}} \vartheta(n+h_0) = |\Omega_d^*(\mathcal{H}^+)| \left(\frac{N}{\phi(d)} + O_A\left(\frac{N}{(\log N)^A}\right)\right), \tag{2.4.2}$$

where $\mathcal{H}^+ = \mathcal{H} \cup h_0$ *and* $|\Omega_p^*(\mathcal{H}^+)| = |\Omega_p(\mathcal{H}^+)| - 1$ *for each prime* p*. Note that when* $h_0 \in \mathcal{H}$ *then* $\mathcal{H}^+ = \mathcal{H}$*.*

Proof. (a) For each $m \bmod d$, $d|P_{\mathcal{H}}(m)$ we get $[\frac{N}{d}]$ values of $n \equiv m \bmod d$, $n \in (N, 2N]$. The number of such m by definition is $|\Omega_d(\mathcal{H})|$.

(b) Nonzero terms of sum (2.4.2) comes from n such that $n + h_0$ is prime. Then $\sum_{\substack{N<n\leqslant 2N\\ d|P_{\mathcal{H}}(n)}} \vartheta(n+h_0) = \sum_{\substack{N+h_0<n\leqslant 2N+h_0\\ (n,d)=1\\ d|P_{\mathcal{H}}(n-h_0)}} \vartheta(n)$. This sum over n can be replaced by $\sum_{\substack{v=1\\ (d,v)=1\\ d|P_{\mathcal{H}}(v-h_0)}}^{d} (\vartheta(2N+h_0; d, v) - \vartheta(N+h_0; d, v))$. Now let $f(d) = \sum_{\substack{v=1\\ (d,v)=1\\ d|P_{\mathcal{H}}(v-h_0)}}^{d} 1$. This is multiplicative function of d (by the Chinese Remainder theorem) and $f(p)$ is the number of solutions of $P_{\mathcal{H}}(v-h_0) \equiv 0 \bmod p$ with $1 \le v \le p-1$. If $h_0 \in \mathcal{H}$, say $h_0 = h_i$, then we exclude the possible solution $v \equiv h_0 - h_i$ since it contradicts the condition $(p, v) = 1$. Thus $f(p) = |\Omega_p(\mathcal{H})| - 1$ and so $f(p) \le k-1$ (note that $|\Omega_p(\mathcal{H})| < k$ if and only if $p| \prod_{1\le i<j\le k} |h_j - h_i|$). On the other hand, when $p \nmid \prod_{1\le i<j\le k} |h_j - h_i|$, we have $f(p) = k-1$. If $h_0 \notin \mathcal{H}$, then $f(p) = |\Omega_p(\mathcal{H}^+)| - 1$ and so $f(p) \le k$ and when $p \nmid \prod_{0\le i<j\le k} |h_j - h_i|$ we have $f(p) = k$.

By the virtue of the Siegel-Walfisz theorem, we can replace $\vartheta(2N+h_0; d, v) - \vartheta(N+h_0; d, v)$ by $\frac{N}{\phi(d)}$. Hence (2.4.2) holds. □

Proposition 2. *For any* $t \in \mathbb{N}$ *and* $x \geqslant 1$

$$\sum_{d\leqslant x} \frac{t^{\omega(d)}}{d} \leqslant (\log x + 1)^t, \tag{2.4.3}$$

$$\sum_{d\leqslant x} t^{\omega(d)} \leqslant x(\log x + 1)^t, \tag{2.4.4}$$

where summation is over squarefree integers.

Proof. For the first inequality, we have

$$\sum_{d_1\dots d_t\leqslant x}\frac{\mu^2(d_1\dots d_t)}{d_1\dots d_t}\leqslant\left(\sum_{n\leqslant x}\frac{1}{n}\right)^t\leqslant(\log x+1)^t.$$

For (2.4.4), we note that the left-hand side is

$$\sum_{d\leqslant x}t^{\omega(d)}=\sum_{d\leqslant x}\frac{t^{\omega(d)}}{d}d\leqslant x\sum_{d\leqslant x}\frac{t^{\omega(d)}}{d},$$

and we apply first inequality. □

Proposition 3. *Assume that primes have level of distribution $\theta>1/2$ and let $t\in\mathbb{Z}^+$. Denote by*

$$E^*(N,d)=\max_{x\leqslant N}\max_{\substack{a\\(a,d)=1}}|\sum_{\substack{x<n\leqslant 2x\\ n\equiv a \bmod d}}\theta(n)-\frac{x}{\phi(d)}|.$$

Then for any positive constant C and any $\epsilon>0$, we have

$$\sum_{d<N^{\theta-\epsilon}}t^{\omega(d)}E^*(N,d)\ll_{C,h,\epsilon}N(\log N)^{-C}, \tag{2.4.5}$$

where summation is over squarefree integers.

Proof. For $E^*(N,d)$ we have the bound $E^*(N,d)\ll N(\log N)/d$ since $\vartheta(N;d,a)=\sum_{\substack{n\leqslant N\\ n\equiv a \bmod d}}\vartheta(n)<\log N\sum_{\substack{n\leqslant N\\ n\equiv a \bmod d}}1\ll N(\log N)/d$. Using the Cauchy-Schwartz inequality we get

$$\begin{aligned}\sum_{d<N^{\theta-\epsilon}}t^{\omega(d)}E^*(N,d)&\leqslant\left(N\log N\sum_{d<N^{\theta-\epsilon}}\frac{t^{2\omega(d)}}{d}\right)^{1/2}\left(\sum_{d<N^{\theta-\epsilon}}E^*(N,d)\right)^{1/2}\\&\ll_{h,\epsilon,A}N(\log N)^{(t^2-A+1)/2}.\end{aligned}$$

In the last line we have used (2.4.3) (A is positive integer from the definition of the level of distribution of the primes). Setting $A=h^2+1-2C$, we obtain (2.4.5). □

Proposition 4 (Gallagher). *For each fixed $k \in \mathbb{N}$*

$$\sum_{\substack{0\le h_1,h_2,\dots,h_k\le H\\ h_1,h_2,\dots,h_k\\ \textit{distinct}}} \mathfrak{S}(\mathcal{H}) \sim H^k \tag{2.4.6}$$

Proof. Take $y = \frac{1}{2}\log H$. First note that $|\Omega_p(\mathcal{H})| = k$ if $p \nmid Q$, where $Q = \prod_{i<j} |h_j - h_i|$. Note that Q is the absolute value of the Vandermonde's determinant and by the Hadamard's inequality $Q \le k^{k/2}H^k$. The number of prime factors of Q is $O\left(\frac{\log Q}{\log\log Q}\right) = O\left(\frac{\log H}{\log\log H}\right)$ since in the worst case if $Q = \prod_{i\le r} p_i$ and then by the prime number theorem we have $\omega(Q) = r = \pi(p_r)$, $\log Q = \sum_{i\le r}\log p_i = \vartheta(p_r)$ and $\omega(Q) \sim \frac{\vartheta(p_r)}{\log p_r} \sim \frac{\log Q}{\log\log Q}$. For any $h_1, h_2, \dots, h_k$ by the binomial theorem we have

$$\begin{aligned}\prod_{p>y}\left(1-\frac{|\Omega_p(\mathcal{H})|}{p}\right)\left(1-\frac{1}{p}\right)^{-k} &= \prod_{p\mid H, p>y}\left(1+O\left(\frac{1}{p}\right)\right)\prod_{p\nmid H,p>y}\left(1+O\left(\frac{1}{p^2}\right)\right)\\ &= 1+O\left(\frac{\log H}{y\log\log H}\right) = 1+O\left(\frac{1}{\log\log H}\right).\end{aligned} \tag{2.4.7}$$

Using (2.4.7), the left hand side of (2.4.6) can be written as the product MN, where

$$M = \left(1+O\left(\frac{1}{\log\log H}\right)\right)\left(1-\frac{1}{p}\right)^{-k}, \quad N = \sum_{\substack{0\le h_1,h_2,\dots,h_k\le H\\ h_1,h_2,\dots,h_k\\ \text{distinct}}}\prod_{p\le y}\left(1-\frac{|\Omega_p(\mathcal{H})|}{p}\right).$$

By simple combinatorial reasoning we have $N = N' + O\left(H^{k-1}\right)$, where N' corresponds to the sum without the condition that $h_1, h_2, \dots, h_k$ are distinct. Set $P = \prod_{p\le y} p$ and note that by the prime number theorem $P = e^{y+o(y)} = h^{1/2+o(1)}$. Now, in the expression for N the product is

$$\frac{1}{P}\prod_{p\mid P}(p - |\Omega_p(\mathcal{H})|) = \frac{|\{n \bmod P : (n+h_i, P) = 1 \quad \text{for each } i\}|}{P}. \tag{2.4.8}$$

Indeed, first note that $|\Omega'_p(\mathcal{H})| := p - |\Omega_p(\mathcal{H})| = |\{n \bmod p : (n+h_1)\dots(n+h_k) \not\equiv 0 \bmod p\}|$. By the Chinese remainder theorem for distinct primes p and q

we have an isomophism

$$\Omega'_p(\mathcal{H}) \times \Omega'_q(\mathcal{H}) \to \Omega'_{pq}(\mathcal{H}) \quad (a \bmod p, b \bmod q) \mapsto n \bmod pq,$$

where $n \equiv a \bmod p$ and $n \equiv a \bmod q$, and (2.4.8) follows by noting that $(n + h_1)\dots(n+h_k) \not\equiv 0 \bmod p$ if and only if $(n+h_i, p) = 1$ for each i.

Therefore

$$\begin{aligned} N' &= \sum_{0\le h_1,h_2,\dots,h_k\le H} \frac{1}{P}\sum_{n=0}^{P-1}\prod_{i=1}^{k}\sum_{d_i|(n+h_i,P)}\mu(d_i) \\ &= \frac{1}{P}\sum_{n=0}^{P-1}\sum_{d_i,\dots,d_k|P}\mu(d_1)\dots\mu(d_k)\prod_{i=1}^{k}\left(\frac{H}{d_i}+O\left(1\right)\right) \\ &= H^k\sum_{d_i,\dots,d_k|P}\frac{\mu(d_1)\dots\mu(d_k)}{d_1\dots d_k} + O\left(H^{k-1}P\sum_{d_i,\dots,d_k|P}\frac{1}{d_1\dots d_k}\right) \\ &= H^k\prod_{p\le y}\left(1-\frac{1}{p}\right)^k + O\left(H^{k-1/2+o(1)}\right), \end{aligned}$$

since $d_i|P$ $(1\le i\le k)$ and denominators of the fractions in the error term can contain primes up to the power $k-1$. Combined with the expression for M, this proves (2.4.6). □

We are now ready to give the proofs of Lemma 1 and Lemma 2.

Proof of Lemma 1. Using (2.4.1), we find

$$\begin{aligned} &\sum_{n=N+1}^{2N}\Lambda_R(n;\mathcal{H},\ell)^2 \\ &= \frac{1}{(k+\ell)!^2}\sum_{d_1,d_2\le R}\mu(d_1)\mu(d_2)\left(\log\frac{R}{d_1}\right)^{k+\ell}\left(\log\frac{R}{d_2}\right)^{k+\ell}\sum_{\substack{N<n\leqslant 2N\\ [d_1,d_2]|P_{\mathcal{H}}(n)}}1 \\ &= \frac{1}{(k+\ell)!^2}\sum_{d_1,d_2\le R}\mu(d_1)\mu(d_2)|\Omega_{[d_1,d_2]}(\mathcal{H})|\frac{N}{[d_1,d_2]}\left(\log\frac{R}{d_1}\right)^{k+\ell}\left(\log\frac{R}{d_2}\right)^{k+\ell} \\ &\qquad + O\left(T\right), \qquad (2.4.9) \end{aligned}$$

where $T = \sum_{d_1,d_2<R}|\lambda_{d_1,\ell}\lambda_{d_2,\ell}||\Omega_{[d_1,d_2]}(\mathcal{H})|$, $\lambda_{d_i,\ell} = \frac{\mu(d_i)}{(k+\ell)!}\left(\log\frac{R}{d_i}\right)^{k+\ell}$, $(i = 1.2)$ and

$[d_1, d_2]$ denotes the least common multiple of d_1 and d_2. By (2.4.4)

$$\begin{aligned} T &\ll (\log R)^{2(k+\ell)} \sum_{d_1,d_2 \leqslant R} k^{\omega([d_1,d_2])} \\ &\ll (\log R)^{4k} \sum_{r<R^2} (3k)^{\omega(r)} \ll R^2 (\log R)^{7k} \ll N^{1-\epsilon}, \end{aligned}$$

provided $R < N^{1/2-\epsilon}$. We have also used the fact that $|\Omega_d(\mathcal{H})| \leqslant k^{\omega(d)}$, which follows from the multiplicativity property of $|\Omega_d(\mathcal{H})|$.

Using Perron's formula [**3**, §17]

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

with (c) the vertical line in the complex plane passing through c, we can rewrite the main term of (2.4.9) in the following form

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

where

$$\begin{aligned} F(s_1, s_2; \Omega) &= \sum_{d_1,d_2} \mu(d_1)\mu(d_2)\frac{|\Omega([d_1,d_2])|}{[d_1,d_2] d_1^{s_1} d_2^{s_2}} \\ &= \prod_p \left(1 - \frac{|\Omega(p)|}{p}\left(\frac{1}{p^{s_1}} + \frac{1}{p^{s_2}} - \frac{1}{p^{s_1+s_2}}\right)\right), \end{aligned} \tag{2.4.10}$$

Define the function

$$G(s_1, s_2; \Omega) = F(s_1, s_2; \Omega)\left(\frac{\zeta(s_1+1)\zeta(s_2+1)}{\zeta(s_1+s_2+1)}\right)^k, \tag{2.4.11}$$

which is absolutely convergent in the region $\operatorname{Re} s_1, \operatorname{Re} s_2 > -c$. Note that using the Euler product expansion of the right side of (2.4.11), we have

$$G(s_1, s_2; \Omega) \ll \exp(c(\log N)^{-2\sigma} \log\log\log N) \tag{2.4.12}$$

with $\sigma := \min(\operatorname{Re} s_1, \operatorname{Re} s_2, 0) \ge c$. We'll use this bound in truncation of the infinite integral.

Now we get

$$\mathcal{I} = \frac{1}{(2\pi i)^2}\int_{(1)}\int_{(1)} G(s_1, s_2; \Omega)\left(\frac{\zeta(s_1+s_2+1)}{\zeta(s_1+1)\zeta(s_2+1)}\right)^k \frac{R^{s_1+s_2}}{(s_1 s_2)^{k+\ell+1}} ds_1 ds_2. \tag{2.4.13}$$

Take $U = \exp(\sqrt{\log N})$. Define the following contours:

$$\begin{aligned}
T_1 &= \{c_0(\log U)^{-1} + it : t \in \mathbb{R}\} \\
T_1' &= \{c_0(\log U)^{-1} + it : |t| \le U\} \\
T_2 &= \{c_0(2\log U)^{-1} + it : t \in \mathbb{R}\} \\
T_2' &= \{c_0(2\log U)^{-1} + it : |t| \le U/2\} \\
\mathcal{T}_1 &= \{-c_0(\log U)^{-1} + it : |t| \le U\} \\
\mathcal{T}_2 &= \{-c_0(2\log U)^{-1} + it : |t| \le U/2\},
\end{aligned}$$

where $c_0 > 0$ is a sufficiently small constant. We write the integrand in (2.4.13) as

$$\frac{H(s_1, s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1 s_2)^{\ell+1}},$$

where

$$H(s_1, s_2) = G(s_1, s_2; \Omega)\left(\frac{(s_1+s_2)\zeta(s_1+s_2+1)}{s_1\zeta(s_1+1)s_2\zeta(s_2+1)}\right)^k \tag{2.4.14}$$

is regular in a neighborhood of $(0,0)$. Now, we recall some standard facts about the Riemann zeta-function. There exists a small constant $c_1 > 0$ such that $\zeta(\sigma + it) \neq 0$ in the region

$$\sigma \ge 1 - \frac{4c_1}{\log(|t|+3)} \tag{2.4.15}$$

for all t. Moreover,

$$\zeta(\sigma+it) - \frac{1}{\sigma - 1 + it} \ll \log(|t|+3), \tag{2.4.16}$$

$$\frac{1}{\zeta(\sigma+it)} \ll \log(|t|+3) \tag{2.4.17}$$

in this region. From (2.4.11), (2.4.15) and (2.4.16), we get the following estimates if s_1, s_2 and s_1+s_2 lie in the region (2.4.14) then

$$H(s_1,s_2) \ll (\log\log N)^c (\log(|s_1|+3))^{2k} (\log(|s_2|+3))^{2k} \tag{2.4.18}$$

and

$$H(s_1,s_2) \ll (\log\log N)^c \quad if \quad |s_1|,|s_2| \ll 1. \tag{2.4.19}$$

We will show now that

$$\int_{T_2}\int_{T_1\setminus T_1'} \frac{H(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{\ell+1}} ds_1 ds_2 \ll \exp(-c\sqrt{\log N}). \tag{2.4.20}$$

The same bound holds if the domain of integration is $T_1 \times T_2 \setminus T_2'$. If $(s_1,s_2) \in T_1 \times T_2$ then

$$(\log\log N)^c \frac{R^{s_1+s_2}}{(s_1+s_2)^k} \ll (\log\log N)^c (\log U)^k R^{\frac{3c_0}{2\log U}} \ll \exp((\frac{3c_0}{2}+c)\sqrt{\log N})$$

and so by (2.4.17)

$$\begin{aligned} &\int_{T_2}\int_{T_1\setminus T_1'} \frac{H(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{\ell+1}} ds_1 ds_2 \\ &\ll \exp((\frac{3c_0}{2}+c)\sqrt{\log N}) \int_{T_2}\int_{T_1\setminus T_1'} \frac{(\log(|s_1|+3))^{2k}}{|s_1|^{\ell+1}} \cdot \frac{(\log(|s_2|+3))^{2k}}{|s_2|^{\ell+1}} ds_1 ds_2. \end{aligned} \tag{2.4.21}$$

Now, since $\ell \geq 1$

$$\int_{T_1\setminus T_1'} \frac{(\log(|s_1|+3))^{2k}}{|s_1|^{\ell+1}} ds_1 \ll \int_U^\infty \frac{(\log(t+3))^{2k}}{t^{\ell+1}} dt \ll \frac{(\log U)^{2k}}{U}$$

and

$$\begin{aligned} \int_{T_2} \frac{(\log(|s_2|+3))^{2k}}{|s_2|^{\ell+1}} ds_2 &\ll \left\{ \int_{\frac{c_0}{2\log U}}^{\frac{c_0}{2\log U}+i} + \int_{\frac{c_0}{2\log U}+i}^{\frac{c_0}{2\log U}+i\infty} \right\} \frac{(\log(|s_2|+3))^{2k}}{|s_2|^{\ell+1}} ds_2 \\ &\ll (\log U)^{\ell+1} + \int_1^\infty \frac{(\log(t+3))^{2k}}{t^{\ell+1}} dt \\ &\ll (\log U)^{\ell+1}. \end{aligned}$$

By Fubini's theorem, the double integral in the right-hand side of (2.4.20) can be written as the product of the two integrals we have just estimated and from this (2.4.19) follows.

We shift the s_1- and s_2-contours to the lines T_1 and T_2 respectively, then we truncate them to get T_1' and T_2'. Thus

$$
\begin{aligned}
\mathcal{I} &= \frac{1}{(2\pi i)^2}\int_{T_2}\int_{T_1}\frac{H(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{\ell+1}}ds_1ds_2 \\
&= \frac{1}{(2\pi i)^2}\Big\{\int_{T_2'}\int_{T_1'}+\int_{T_2\setminus T_2'}\int_{T_1'}+\int_{T_2}\int_{T_1\setminus T_1'}\Big\}\frac{H(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{\ell+1}}ds_1ds_2 \\
&= \frac{1}{(2\pi i)^2}\int_{T_2'}\int_{T_1'}\frac{H(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{\ell+1}}ds_1ds_2+O\left(\exp(-c\sqrt{\log N})\right) \qquad (2.4.22)
\end{aligned}
$$

by (2.4.19).

Now we shift the s_1-contour to the line $\mathcal{T}_1$. We have singularities at $s_1 = 0$ and $s_1 = -s_2$, which are poles of orders $\ell+1$ and k, respectively (since $\lim_{s_1\to 0}(\zeta(1+s_1))^k {s_1}^k = 1$). By Cauchy's residue theorem we get

$$
\begin{aligned}
&\int_{T_2'}\int_{T_1'}\frac{H(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{\ell+1}}ds_1ds_2 \\
&\quad= \int_{T_2'}\Big\{\int_{\mathcal{C}_1}-\int_{\mathcal{T}_1\cup K_1}\Big\}\frac{H(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{\ell+1}}ds_1ds_2 \\
&\quad= 2\pi i\int_{T_2'}\Big\{\operatorname*{Res}_{s_1=0}+\operatorname*{Res}_{s_1=-s_2}\Big\}\frac{H(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{\ell+1}}ds_1ds_2 \\
&\qquad -\int_{T_2'}\int_{\mathcal{T}_1\cup K_1}\frac{H(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{\ell+1}}ds_1ds_2, \qquad (2.4.23)
\end{aligned}
$$

where $K_1 = \{\sigma \pm iU : |\sigma| \le c_0(\log U)^{-1}+it\}$ and $\mathcal{C}_1 = T_1'\cup\mathcal{T}_1\cup K_1$. Similarly to (2.4.19), one can obtain that

$$
\int_{T_2'}\int_{\mathcal{T}_1\cup K_1}\frac{H(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{\ell+1}}ds_1ds_2 \ll \exp(-c\sqrt{\log N}). \qquad (2.4.24)
$$

We'll show that the residue at $s_1 = -s_2$ may be neglected. For this we rewrite the residue in terms of the integral over the circle $C_1 : |s_1+s_2| = (\log N)^{-1}$

$$
\operatorname*{Res}_{s_1=-s_2}\left\{\frac{H(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{\ell+1}}\right\} = \frac{1}{2\pi i}\int_{C_1}\frac{H(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{\ell+1}}ds_1. \qquad (2.4.25)
$$

By (2.4.11), (2.4.13) and (2.4.16), we have $G(s_1,s_2;\Omega) \ll (\log\log N)^c$, $\zeta(s_1+s_2+1) \ll \log N$, $R^{s_1+s_2} = O(1)$. Also $(s_1\zeta(s_1+1))^{-1} \ll (|s_2|+1)^{-1}\log(|s_2|+3)$

since $|s_2| \ll |s_1| \ll |s_2|$. Altogether this gives us

$$\operatorname*{Res}_{s_1=-s_2}\left\{\frac{H(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{\ell+1}}\right\} \ll (\log N)^{k-1}(\log\log N)^c\left(\frac{\log(|s_2|+2)}{|s_2|+1}\right)^{2k}|s_2|^{-2\ell-2}. \tag{2.4.26}$$

Inserting this into (2.4.25) we have

$$\mathcal{I} = \frac{1}{(2\pi i)}\int_{T_2'}\operatorname*{Res}_{s_1=0}\left\{\frac{H(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{\ell+1}}\right\}ds_2 + O\left((\log N)^{k+\ell-1/2}(\log\log N)^c\right). \tag{2.4.27}$$

Note that $H(s_1,s_2)$ is holomorphic near $(0,0)$ and

$$\operatorname*{Res}_{s_1=0}\left\{\frac{H(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{\ell+1}}\right\} = \frac{R^{s_2}}{\ell!s_2^{\ell+1}}\left(\frac{\partial}{\partial s_1}\right)^{\ell}_{s_1=0}\left\{\frac{H(s_1,s_2)}{(s_1+s_2)^k}R^{s_1}\right\}, \tag{2.4.28}$$

since the pole has order $\ell+1$. We insert this into (2.4.27) and apply the similar argument: that is, we shift now the s_2-contour to the line $\mathcal{T}_2$. Again, it can be shown that the new integral is $O\left(\exp(-c\sqrt{\log N})\right)$ and all we get is the residue at $s_2=0$. Therefore

$$\mathcal{I} = \operatorname*{Res}_{s_2=0}\operatorname*{Res}_{s_1=0}\left\{\frac{H(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{\ell+1}}\right\} + O\left((\log N)^{k+\ell}\right). \tag{2.4.29}$$

Take some sufficiently small $r>0$ and let $C_2 : |s_1| = r$, $C_3 : |s_2| = 2r$. Then we can rewrite (2.4.29) in the form

$$\mathcal{I} = \frac{1}{(2\pi i)^2}\int_{C_3}\int_{C_2}\frac{H(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{\ell+1}}ds_1ds_2 + O\left((\log N)^{k+\ell}\right). \tag{2.4.30}$$

We now introduce new variables t,λ such that $s_1=t$, $s_2=t\lambda$ with contours $C_2' : |t| = r$, $C_3' : |\lambda| = 2$. Then the double integral is equal to

$$\mathcal{I} = \frac{1}{(2\pi i)^2}\int_{C_3'}\int_{C_2'}\frac{H(t,t\lambda)R^{t(\lambda+1)}}{(\lambda+1)^k\lambda^{\ell+1}t^{k+2\ell+1}}dtd\lambda$$

Thus we get

$$\mathcal{I} = \frac{H(0,0)}{2\pi i(k+2\ell)!}(\log R)^{k+2\ell}\int_{C_3'}\frac{(\lambda+1)^{2\ell}}{\lambda^{\ell+1}}d\lambda + O\left((\log N)^{k+\ell-1}(\log\log N)^c\right),$$

where we used (2.4.12) and regularity of $H(s_1,s_2)$ around $(0,0)$.

But $\frac{1}{2\pi i}\int_{C_3'}\frac{(\lambda+1)^{2\ell}}{\lambda^{\ell+1}}d\lambda = \operatorname*{Res}_{\lambda=0}\frac{(\lambda+1)^{2\ell}}{\lambda^{\ell+1}} = \binom{2\ell}{\ell}$ and by (2.4.10), (2.4.11), (2.4.14) and Euler's product formula $\zeta(s) = \prod_p (1-p^{-s})^{-1}$ we have $H(0,0) = \mathfrak{S}(\mathcal{H})$. This completes the proof. □

Proof of Lemma 2 . Let $h_0 \notin \mathcal{H}$ and $\mathcal{H}^+ = \mathcal{H}\cup h_0$. Applying (2.4.2) we see

$$
\begin{aligned}
&\sum_{n=N+1}^{2N} \Lambda_R(n;\mathcal{H}^+,\ell)^2\vartheta(n+h_0)\\
&= \frac{1}{(k+\ell)!^2}\sum_{d_1,d_2\le R}\mu(d_1)\mu(d_2)\left(\log\frac{R}{d_1}\right)^{k+\ell}\left(\log\frac{R}{d_2}\right)^{k+\ell}\sum_{\substack{N<n\leqslant 2N\\ [d_1,d_2]|P_{\mathcal{H}}(n)}}\vartheta(n+h_0)\\
&= \frac{1}{(k+\ell)!^2}\sum_{d_1,d_2\le R}\mu(d_1)\mu(d_2)|\Omega^*_{[d_1,d_2]}(\mathcal{H}^+)|\frac{N}{\phi([d_1,d_2])}\left(\log\frac{R}{d_1}\right)^{k+\ell}\left(\log\frac{R}{d_2}\right)^{k+\ell}\\
&\qquad + O(T_1)\\
&= \frac{N}{(2\pi i)^2}\int_{(1)}\int_{(1)}\prod_p\left(1-\frac{(|\Omega_p(\mathcal{H}^+)|-1)}{(p-1)}\left(\frac{1}{p^{s_1}}+\frac{1}{p^{s_2}}-\frac{1}{p^{s_1+s_2}}\right)\right)\frac{R^{s_1+s_2}}{(s_1s_2)^{k+\ell+1}}ds_1ds_2\\
&\qquad + O(T_1)\\
&= \frac{N}{(2\pi i)^2}\int_{(1)}\int_{(1)} G^*(s_1,s_2;\Omega)\left(\frac{\zeta(s_1+s_2+1)}{\zeta(s_1+1)\zeta(s_2+1)}\right)^{k-1}\frac{R^{s_1+s_2}}{(s_1s_2)^{k+\ell+1}}ds_1ds_2\\
&\qquad + O(T_1), \qquad (2.4.31)
\end{aligned}
$$

where

$$
G^*(s_1,s_2;\Omega) = \prod_p \frac{\left(1-\frac{(|\Omega_p(\mathcal{H})^+|-1)}{(p-1)}\left(\frac{1}{p^{s_1}}+\frac{1}{p^{s_2}}-\frac{1}{p^{s_1+s_2}}\right)\right)\left(1-\frac{1}{p^{1+s_1+s_2}}\right)^{k-1}}{\left(1-\frac{1}{p^{1+s_1}}\right)^{k-1}\left(1-\frac{1}{p^{1+s_2}}\right)^{k-1}},
$$

$$
T_1 = \sum_{d_1,d_2}|\lambda_{d_1,\ell}\lambda_{d_2,\ell}||\Omega^*_{[d_1,d_2]}|E^*(N,[d_1,d_2])
$$

and $\lambda_{d_i,\ell} = \frac{\mu(d_i)}{(k+\ell)!}\left(\log\frac{R}{d_i}\right)^{k+\ell}$, $(i = 1.2)$ Note that $G^*(s_1,s_2;\Omega)$ is absolutely convergent when $\operatorname{Re}s_1, \operatorname{Re}s_2 > -1/2$. The result follows from Lemma 1 (with the translation $k\to k-1$, $\ell\to\ell+1$) since $G^*(0,0) = \mathfrak{S}(\mathcal{H}^+)$. We have to justify our estimation by evaluating error term T_1 and showing that it is negligible.

By (2.4.5)

$$T_1 \ll (\log R)^{2(k+\ell)} \sum_{r<R^2} (3k)^{\omega(r)} E^*(N,r) \ll N/\log N, \tag{2.4.32}$$

provided $R < N^{\theta/2-\epsilon}$.

If $h_0 \in \mathcal{H}$ then we can aplly above evaluation with $k-1$ in place of k and $\ell+1$ in place of ℓ. $\square$

3. ALMOST PRIMES IN TUPLES AND THE DIVISOR FUNCTION AT CONSECUTIVE INTEGERS

Using the same technique, which was successfully applied for short gaps between primes, Goldston, Graham, Pintz and Yıldırım [**17**] proved the existence of small gaps between E_2 numbers (i.e. numbers which are products of exactly two distinct primes). Instead of $\vartheta(n)$ we consider the function $\vartheta * \vartheta(n) := \sum_{d|n} \vartheta(d)\vartheta(\frac{n}{d})$. Note, that $\vartheta * \vartheta(n) \neq 0$ if and only if n is a product of two primes or n is a square of a prime. If $n \in (N, 2N]$ then $\vartheta * \vartheta(n) \leq \frac{(\log 3N)^2}{2}$. Further we need the following natural analogue of Bombieri-Vinogradov theorem for E_2 numbers and analogue of Lemma 2 for the function $\vartheta * \vartheta(n)$

Theorem. *Let $\tilde{E}(N; q, a)$ be defined by*

$$\sum_{\substack{N<n\leqslant 2N \\ n\equiv a \bmod q}} \vartheta * \vartheta(n) = \frac{N}{\phi(q)} \left(\log N + C_0 - 2\sum_{p|q} \frac{\log p}{p} \right) + \tilde{E}(N; q, a),$$

where C_0 is the absolute constant. Then for every $A > 0$, there exists $B > 0$ such that if $Q \leqslant N^{1/2} \log^{-B} N$

$$\sum_{q\leqslant Q} \max_{x\leqslant N} \max_{\substack{a \\ (a,q)=1}} |\tilde{E}(x; q, a)| \ll_A N(\log N)^{-A}.$$

Lemma 3. *Suppose that $\mathcal{H} = \{h_1, h_2, \ldots, h_k\} \subseteq [1, H] \cap \mathbb{Z}$ with $h_i \neq h_j$ for $i \neq j$; k, ℓ, $0 < \ell \leq k$ are arbitrarily, but fixed positive integers, $|\mathcal{H}| = k$ and $\{x+h_1, x+h_2, \ldots, x+h_k\}$ is an admissible k-tuple. Suppose that the primes have level of distribution θ and above theorem is satisfied with $Q \leq N^{\theta-\epsilon}$; $R \leq N^{(\theta-\epsilon)/2}$. If $h_0 \in \mathcal{H}$ then*

$$\sum_{N<n\leqslant 2N} \vartheta * \vartheta(n + h_0)\Lambda_R(n; \mathcal{H}, \ell)^2 = \left\{ \binom{2\ell+2}{\ell+1} (N \log N)\mathfrak{S}(\mathcal{H}) \frac{(\log R)^{k+2\ell+1}}{(k+2\ell+1)!} \right.$$

$$\left. +2T(k, \ell)N\mathfrak{S}(\mathcal{H}) \frac{(\log R)^{k+2\ell+2}}{(k+2\ell+2)!} \right\} \{1 + O(\log\log N/\log N)\}$$

where

$$T(k,\ell) = -2\binom{2\ell+3}{\ell+1} + \binom{2\ell+2}{\ell+1},$$

$$\Lambda_R(n;\mathcal{H},\ell) = \sum_{d|P_{\mathcal{H}}(n)} \lambda_{d,\ell},$$

$$\lambda_{d,\ell} = \mu(d)\frac{f(d)}{f_1(d)}\frac{\mathfrak{S}(\mathcal{H})}{(\ell)!}\sum_{\substack{r<\frac{R}{d}\\(r,d)=1}}\frac{\mu^2(r)}{f_1(r)}(\log R/rd)^{\ell},$$

$$f(d) = \frac{d}{|\Omega_d(\mathcal{H})|}, \quad f_1(d) = f * \mu(d)$$

and the implied constant depends at most on k.

Now, consider the weighted sum

$$\mathcal{S}_2 = \sum_{N<n\leqslant 2N}\left(\sum_{h_i\in\mathcal{H}}\vartheta*\vartheta(n+h_i) - \frac{(\log 3N)^2}{2}\right)\Lambda_R(n;\mathcal{H},\ell)^2. \tag{3.1}$$

For each n, the inner sum is negative unless there are two values $h_i, h_j \in \mathcal{H}$ such that $n+h_i, n+h_j$ are products of two primes. From Lemmas 1 and 3, one can obtain

$$\begin{aligned}\mathcal{S}_2 \gtrsim &\left[\binom{2\ell+2}{\ell+1}\frac{k}{(k+2\ell+1)!}\frac{\theta}{2} + 2\left\{\binom{2\ell+2}{\ell+1} - 2\binom{2\ell+3}{\ell+1}\right\}\frac{k}{(k+2\ell+2)!}\frac{\theta^2}{4}\right.\\ &\left.- \frac{1}{2}\binom{2\ell}{\ell}\frac{k}{(k+2\ell)!}\right]\mathfrak{S}(\mathcal{H})N(\log N)^2(\log R)^{k+2\ell} = \left\{\frac{2\ell+1}{\ell+1}\frac{k}{k+2\ell+1}\theta\right.\\ &\left.- \frac{6\ell^2+11\ell+4}{(\ell+1)(\ell+2)}\frac{k}{(k+2\ell+1)(k+2\ell+2)}\theta^2 - \frac{1}{2}\right\}\\ &\times\binom{2\ell}{\ell}\frac{1}{(k+2\ell)!}\mathfrak{S}(\mathcal{H})N(\log N)^2(\log R)^{k+2\ell}\end{aligned}$$

Assuming the level of distribution of primes is $\theta = 1/2-\epsilon$, $R = N^{(\theta-\epsilon)/2}$, we want to find the minimal $k \geq 1$ such that there exists $\ell \geq 0$ with

$$\frac{2\ell+1}{\ell+1}\frac{k}{k+2\ell+1} - \frac{6\ell^2+11\ell+4}{2(\ell+1)(\ell+2)}\frac{k}{(k+2\ell+1)(k+2\ell+2)} > 1.$$

The answer is $k=9, \ell=1$. Indeed

$$\begin{aligned}\ell=0: &\quad \frac{2}{k+2} < 0 \quad \textit{is} \quad \textit{impossible},\\ \ell=1: &\quad 2k^2 - 11k - 48 > 0 \Rightarrow k > 8.36,\end{aligned}$$

$$\begin{aligned}
\ell = 2: &\quad 8k^2 - 37k - 360 > 0 \Rightarrow k > 9.4,\\
\ell = 3: &\quad 30k^2 - 131k - 2240 > 0 \Rightarrow k > 11.09,\\
\ell \geq 4: &\quad k^2 > 4\ell^2 + 6\ell + 2 \geq 90 \Rightarrow k > 9.48,
\end{aligned}$$

because $1 < \frac{2\ell+1}{\ell+1} < 2$, $\frac{3\ell+4}{\ell+2} > 2$ for $\ell \geq 4$. Since $\mathcal{H} = \{0, 2, 6, 8, 12, 18, 20, 26, 30\}$ is an admissible set the limit infimum of gaps between products of two primes (i.e. E_2 numbers or prime squares) is bounded by 30. With a more elaborative weight Goldston-Graham-Pintz-Yıldırım [**17**] were able to prove

Theorem 1.4. *Let q_n denote the n^{th} number that is a product of exactly two primes. Then assuming the level of distribution of primes is $\theta = 1/2 - \epsilon$, we have*

$$\liminf_{n\to\infty} (q_{n+1} - q_n) \leqslant 26.$$

As an application of their work in [**18**], Goldston, Graham, Pintz and Yıldırım [**19**] have shown, among several other interesting results, that $d(n) = d(n+1) = 24$ for infinitely many integers n, where $d(n)$ is the number of divisors of n. That $d(n) = d(n+1)$ for infinitely many n is the well-known Erdős-Mirsky conjecture [**7**]. In 1983 Spiro [**38**] showed that $d(n) = d(n+5040)$ for infinitely many $n \in \mathbb{N}$. We sketch the key idea behind the Spiro's proof. Consider the set of 8 primes

$$\{p_1, \dots, p_8\} = \{11, 17, 23, 29, 41, 47, 53, 59\}$$

and denote by L the least common multiple of the pairwise differences of these 8 primes:

$$L = l.c.m.\{p_j - p_i : 1 \leq i < j \leq 8\}.$$

Define the polynomial

$$F(n) = \prod_{i=1}^{8} (p_i n + 1).$$

By a result from sieve theory [**21**, Theorem 10.5], there exist infinitely many n such that $\omega(F(n)) \leq 34$. Moreover, one can require $F(n)$ to be, in addition, squarefree and $(F(n), L\prod_{k=1}^{8} p_k) = 1$. Indeed, a restriction of the type $p(F(n)) \geq x^\delta$, $\delta = const > 0$, where $p(n)$ denotes the least prime factor of n, is implicit in the proof of the sieve-theoretic result, and using this property, one sees that the contribution of the non-squarefree integers is negligible. Since $\sum_{i=1}^{8} i = 36 > 34$ and $F(n)$ is squarefree, it follows that for each of these values of n, two of the linear factors of $F(n)$, say p_in+1 and p_kn+1, have the same number of prime factors, are squarefree and are relatively prime to $L\prod_{j=1}^{8} p_j$. In fact, by the pigeon hole principle there exist a fixed pair (i,k) with $1 \le i < k \le 8$ such that above conditions hold for infinitely many n. Now, it is easy to verify that, for each such $n \in \mathbb{N}$, the integers

$$n_1 = \frac{L}{p_k - p_i} p_i(p_kn+1), \quad n_2 = \frac{L}{p_k - p_i} p_k(p_in+1) = n_1 + L$$

have the same number of divisors. Since $L = 5040$, it follows that there are infinitely many integers n_1 with $d(n_1) = d(n_1 + 5040)$.

In 1984, Heath-Brown [**24**], using Spiro's argument, proved the conjecture of Erdős and Mirsky. The key idea behind his proof was to try replace set of primes in Spiro's construction by a set of positive integers $\{a_i\}_{i=1}^{N}$ (for arbitrarily large N) having the properties

$$a_j - a_i | a_j \quad (1 \le i < j \le N)$$

and

$$d(a_j)d(\frac{a_i}{|a_j - a_i|}) = d(a_i)d(\frac{a_j}{|a_j - a_i|}) \qquad (1 \le i < j \le N).$$

Moreover, he showed that for large x

$$D(x) = \#\{n \le x : d(n) = d(n+1)\} \ge \frac{x}{(\log x)^7}.$$

Hildebrand [**25**] improved the lower bound to $\frac{x}{(\log\log x)^3}$. Using a heuristic argument, Bateman and Spiro claimed that $D(x) \sim cx(\log\log x)^{-1/2}$ for some constant $c > 0$.

The Erdős-Mirsky conjecture is equivalent to the statement that $\frac{d(n)}{d(n+1)} = 1$ holds for infinitely many n. More generally, one can ask which numbers occur as limit points of the sequence $\{d(n)/d(n+1)\}_{n=1}^{\infty}$. Let $\mathbf{E}$ denote the set of limit points of the sequence $\{d(n)/d(n+1)\}$, and let $\mathbf{L}$ denote the set of limit points of $\{\log(d(n)/d(n+1))\}$. Then the Erdős-Mirsky conjecture implies that $1 \in \mathbf{E}$. Erdős conjectured [**6**] that $\mathbf{E} = [0, \infty]$, or equivalently, $\mathbf{L} = [-\infty, \infty]$. Erdős, Pomerance and Sarközy [**8**] proved that for any $\alpha \in \mathbb{R}^+$, at least one of the 7 numbers $2^i\alpha$, $i \in \{0, \pm1, \pm2, \pm3\}$, belongs to $\mathbf{E}$. This result was improved by Kan and Shan [**30**],[**31**], who showed that for any real $\alpha > 0$, either α or 2α belongs to $\mathbf{E}$. On the other hand, it can be shown under assumption of the prime k-tuple conjecture that for any $r \in \mathbb{Q}^+$ there exist infinitely many $n \in \mathbb{N}$ such that $\frac{d(n)}{d(n+1)} = r$. For this we need the following

Conjecture ([**36**, C2])**.** Let $a, b, c \in \mathbb{N}$ such that $(a,b) = (a,c) = (b,c) = 1$ and $2|abc$. Then there exist infinitely many pairs of primes (p,q) such that $ap - bq = c$.

Proof. Since $(a,b) = (a,c) = (b,c) = 1$ and $2|abc$, there exist natural numbers r and s such that $ar - bs = c$. Set $f_1(x) = bx + r$, $f_2(x) = ax + s$, therefore $f_1(x)f_2(x) = abx^2 + (ar + bs)x + rs$.

If there were a prime p such that $p|f_1(x)f_2(x)$ for any integer x, then (for $x = 0$) $p|rs$ and (for $x = \pm 1$) $p|ab \pm (ar + bs)$. Hence $p|2ab$ and $p|2(ar + bs)$. If $p = 2$, then from $p|rs$ we had either $2|r$ or $2|s$. If $2|r$, we can not have $2|s$ since then $2|ar \pm bs$, so $2|ab$ and $2|c$ contrary to $(ab, c) = 1$. So if $2|r$, s is odd and $2|ab + (ar + bs)$ implies $2|(a+1)b$. That is a is odd or b is even. If b is even then, according to $ar - bs = c$, c is even unlike $(b,c) = 1$. So b is odd, a is odd then also $c = ar - bs$ is odd unlike $2|abc$. Therefore r can not be even, hence s is even. Repeating again above argument, we see that s also can not be even.

Hence we have $p \neq 2$. As $p|rs$, $p|ar^2 + brs$. From this $p|ar^2$ and $p|ar$. Similarly we get implications $p|ars + bs^2$, $p|bs^2$, $p|bs$. Thus $p|ar - bs$ which is impossible since $(ab, c) = 1$.

Therefore $f_1(x)$ and $f_2(x)$ satisfy all conditions of the prime k-tuple conjecture. So there exist infinitely many $x \in \mathbb{N}$ such that $f_1(x) = p$ and $f_2(x) = q$ are prime numbers. Then $bx + r = p$, $ax + s = q$ which gives $ap - bq = ar - bs = c$. $\square$

Now take $r = \frac{k}{l}$ for some $k, l \in \mathbb{N}$, $(k,l) = 1$ and set $a = 3^{l-1}$, $b = 2^{k-1}$, $c = 1$ in the above conjecture. Then there exist infinitely many primes p, q such that $3^{l-1}p - 2^{k-1}q = 1$ and for $n = 2^{k-1}q$ we have

$$\frac{d(n)}{d(n+1)} = \frac{d(2^{k-1}q)}{d(2^{k-1}q+1)} = \frac{d(2^{k-1}q)}{d(3^{l-1}p)} = \frac{k}{l}.$$

Hildebrand [**25**, Theorem 3] proved that for $x > 0$,

$$|\mathbf{L} \cap [0, x]| \geqslant \frac{x}{36}, \quad |\mathbf{L} \cap [-x, 0]| \geqslant \frac{x}{36}. \tag{3.2}$$

We will now show how the results in [**19**] can be used to improve upon this.

A triple of linear forms is called admissible if for every prime p, there is at least one m mod p such that $L_1(m)L_2(m)L_3(m) \not\equiv 0 \bmod p$. Unconditionally we have

Lemma 1 ([**19**, Corollary 2.1])**.** *Let $L_i(x) := a_i x + b_i$, $i = 1, 2, 3$ be an admissible triple of linear forms, and let r_1, r_2, r_3 be coprime integers with $(r_i, a_i) = 1$ for each i and $(r_i, a_i b_j - a_j b_i) = 1$ for each $i \neq j$. Then there exists $1 \leqslant i < j \leqslant 3$ such that there are infinitely many integers n for which $L_k(n)$ equals r_k times an E_2 number that is coprime to all primes $\leqslant C$, for $k = i, j$.*

In the lemma, C can be any constant. Hildebrand deduced (3.2) from the fact that among any 7 integers $a_1, \ldots, a_7$, there exists $i < j$ such that $d(n)/d(n+1) = a_i/a_j$ for infinitely many n. We can replace $x/36$ by $x/3$, in view of

Corollary 2. *Let a_1, a_2, a_3 be positive integers. For some $i < j$, there are infinitely many integers n such that $d(n)/d(n+1) = a_i/a_j$.*

Proof. Define a triple of linear forms

$$L_1(x) := 9x + 1, \quad L_2(x) := 8x + 1, \quad L_3(x) := 6x + 1,$$

and note that

$$8L_1(x) + 1 = 9L_2(x), \quad 2L_1(x) + 1 = 3L_3(x), \quad 3L_2(x) + 1 = 4L_3(x).$$

Choose any positive integers a_1, a_2, a_3, and let

$$r_1 := 5^{a_1 - 1}, \quad r_2 := 3 \cdot 7^{a_2 - 1}, \quad r_3 := 11^{a_3 - 1}.$$

We check that the hypotheses of Lemma 1 are satisfied. First of all the triple is admissible because if $m \equiv 0 \bmod p$ then $L_1(m)L_2(m)L_3(m) \equiv 1 \bmod p$, for all p. (It suffices to check this for $p = 2, 3$.) We have $(r_i, r_j) = 1$ for $i \neq j$, and $(r_1, 9) = (r_1, 9 \cdot 1 - 8 \cdot 1) = (r_1, 9 \cdot 1 - 6 \cdot 1) = 1$, $(r_2, 8) = (r_2, 8 \cdot 1 - 9 \cdot 1) = (r_2, 8 \cdot 1 - 6 \cdot 1) = 1$, and $(r_3, 6) = (r_3, 6 \cdot 1 - 9 \cdot 1) = (r_3, 6 \cdot 1 - 8 \cdot 1) = 1$.

We put $C = 11$ in the lemma. Then for some $i < j$, there exist infinitely many integers m for which $L_k(m)$ equals r_k times an E_2 number, all of whose prime factors are > 11, for $k = i, j$. If the forms are $L_1(x)$ and $L_2(x)$, then for infinitely many m, we have E_2 numbers A_1 and A_2, such that $(2 \cdot 3 \cdot 5 \cdot 7 \cdot 11, A_1 A_2) = 1$ and

$$\frac{d(8L_1(m))}{d(8L_1(m) + 1)} = \frac{d(8L_1(m))}{d(9L_2(m))} = \frac{d(2^3 r_1 A_1)}{d(3^2 r_2 A_2)} = \frac{d(2^3)d(5^{a_1 - 1})d(A_1)}{d(3^3)d(7^{a_2 - 1})d(A_2)} = \frac{a_1}{a_2}.$$

Similarly, if $L_1(x)$ and $L_3(x)$ are the relevant forms, then we have E_2 numbers

A_1, A_3 such that

$$\frac{d(2L_1(m))}{d(2L_1(m)+1)} = \frac{d(2L_1(m))}{d(3L_3(m))} = \frac{d(2r_1A_1)}{d(3r_3A_3)} = \frac{d(2)d(5^{a_1-1})d(A_1)}{d(3)d(11^{a_3-1})d(A_3)} = \frac{a_1}{a_3}.$$

Finally, if the forms are $L_2(x)$ and $L_3(x)$, then

$$\frac{d(3L_2(m))}{d(3L_2(m)+1)} = \frac{d(3L_2(m))}{d(4L_3(m))} = \frac{d(3r_2A_2)}{d(2^2r_3A_3)} = \frac{d(3^2)d(7^{a_2-1})d(A_2)}{d(2^2)d(11^{a_3-1})d(A_3)} = \frac{a_2}{a_3}.$$

□

Let $q_1 = \frac{b_1}{b_2}$, $q_2 = \frac{b_3}{b_4}$ be positive rational numbers. If we take $a_1 = b_1b_3$, $a_2 = b_2b_3$, $a_3 = b_2b_4$, then from Corollary 2 we get $\frac{d(n)}{d(n+1)} \in \{q_1, q_2, q_1q_2\}$ for every $q_1, q_2 \in \mathbb{Q}^+$. Since rational numbers are dense in $\mathbb{R}$ and every irrational number can be approximated by rationals, we have that for every $r_1, r_2 \in \mathbb{R}^+$ either $r_1 \in \mathbf{E}$ or $r_2 \in \mathbf{E}$ or $r_1r_2 \in \mathbf{E}$.

We are now ready to improve Hildenbrand's result.

Proof of Theorem 1.8. (a) Let

$$\mathbf{L}' = \{\log\frac{r}{s} : r, s \in \mathbb{N}; \frac{d(n+1)}{d(n)} = \frac{r}{s} \text{ for infinitely many } \quad n \in \mathbb{N}\} \tag{3.3}$$

It is obvious that $\overline{\mathbf{L}'} \subset \mathbf{L}$. Corollary 2 shows, that for any positive integers a_1, a_2, a_3 there exist indices $i < j$ such that $\log\frac{a_j}{a_i} \in \mathbf{L}$. From this follows that given any positive real numbers u_1, u_2, u_3 we get

$$u_j - u_i \in \overline{\mathbf{L}'} \subset \mathbf{L} \text{ for some } i < j \tag{3.4}$$

Applying (3.4) with $u_i = iu,\ (i = 1, 2, 3)$, we get that for all $u > 0$

$$u \in \mathbf{L} \bigcup \frac{\mathbf{L}}{2}$$

Now, using subadditivity and positive homogeneity properties of Lebesgue measure, for $x > 0$ we have

$$x = |[0, x] \bigcap \{\mathbf{L} \bigcup \frac{\mathbf{L}}{2}\}| \le |\mathbf{L} \bigcap [0, x]| + |\frac{\mathbf{L}}{2} \bigcap [0, x]| \le \frac{3}{2}|\mathbf{L} \bigcap [0, 2x]|$$

and therefore

$$|\mathbf{L} \bigcap [0, x]| \ge \frac{x}{3} \quad (x > 0).$$

A similar argument with $u_i = (4-i)u$ yields

$$|\mathbf{L}\bigcap[-x,0]| \geq \frac{x}{3} \quad (x>0).$$

Hence (1.3.8) holds.

If $\mathbf{L} = \mathbb{R}$ we're done. Otherwise there exist $A > 0$ such that $A \notin \mathbf{L}$. By Corollary 2 for every $x \in \mathbb{R}$ either A or x or $A + x \in \mathbf{L}$. Hence under our assumption x or $A + x \in \mathbf{L}$. Then

$$x = |\mathbf{L}\bigcap\{[0,A]\bigcup[A,A+x]\}| \leq |\mathbf{L}\bigcap[0,A]|+|\mathbf{L}\bigcap[A,A+x]| \leq 2|\mathbf{L}\bigcap[0,A+x]|$$

and therefore $|\mathbf{L}\bigcap[0,x]| \geq \frac{x-A}{2}$ for $x \geq A$.

(b) If $\mathbf{E} = \mathbb{R}^+$ we're done. Otherwise there exist $B > 0$ such that $B \notin \mathbf{E}$. By Corollary 2 for every $x \in (0,\infty)$ either B or x or $Bx \in \mathbf{E}$. Hence under our assumption x or $Bx \in \mathbf{E}$. Then

$$x = |[0,x]\bigcap\{\mathbf{E}\bigcup\frac{\mathbf{E}}{B}\}| \leq |\mathbf{E}\bigcap[0,x]| + |\frac{\mathbf{E}}{B}\bigcap[0,x]|.$$

If $B > 1$ then using subadditivity and homogeneity properies of Lebesgue measure, we have

$$x \leq |\mathbf{E}\bigcap[0,x]| + \frac{1}{B}|\mathbf{E}\bigcap[0,Bx]| \leq \frac{B+1}{B}|\mathbf{E}\bigcap[0,Bx]|.$$

Hence

$$|\mathbf{E}\bigcap[0,x]| \geq \frac{x}{B+1}.$$

If $0 < B < 1$, then $B = \frac{1}{c}$ for $c > 1$ and

$$\begin{aligned} x &= |[0,x]\bigcap\{\mathbf{E}\bigcup c\mathbf{E}\}| \leq |\mathbf{E}\bigcap[0,x]| + |c\mathbf{E}\bigcap[0,x]| \\ &\leq |\mathbf{E}\bigcap[0,x]| + c|\mathbf{E}\bigcap[0,Bx]| \leq (1+c)|\mathbf{E}\bigcap[0,x]|. \end{aligned}$$

Therefore

$$|\mathbf{E}\bigcap[0,x]| \geq \frac{x}{1+c}.$$

$\square$

BIBLIOGRAPHY

[1] E. Bombieri, 'On the large sieve', *Mathematika* **12** (1965), 201–225. MR0197425 (33:5590)

[2] E. Bombieri, H. Davenport, 'Small differences between prime numbers' *Proc. Roy. Soc. Ser. A,* **293** (1966), 1–18. MR0199165 (33 #7314)

[3] H. Davenport, *Multiplicative number theory*, 3rd edn (Revised and with a preface by H. L. Montgomery; Springer-Verlag, New York, 2000). MR1790423 (2001f:11001)

[4] P. D. T. A. Elliott, H. Halberstam, *A conjecture in prime number theory*, Symposia Mathematica, Vol. IV (INDAM, Rome, (1970), 59–72 MR0276195 (43 #1943)

[5] P. Erdős, 'The difference of consecutive primes', *Duke Math. J.* **6** (1940), 438–441. MR0001759 (1,292h)

[6] P. Erdős, 'Some problems on number theory', *Analytic and elementary number theory (Marseille, 1983)*, Publications Mathématiques d'Orsay **86** (Univ. Paris XI, Orsay, 1986), 53–67. MR844584 (87i:11006)

[7] P. Erdős and L. Mirsky, 'The distribution of values of the divisor function $d(n)$', *Proc. London Math. Soc. (3)* **2** (1952), 257–271. MR0049932 (14:249e)

[8] P. Erdős, C. Pomerance and A. Sarközy, 'On locally repeated values of certain arithmetic functions. II', *Acta Math. Hungar.* **49** (1987), 251–259. MR869681 (88c:11008)

[9] T. Freiberg, 'Strings of congruent primes in short intervals', *J. London Math. Soc. (2)* **84** (2011), 344–364. MR2835334 (2012j:11181)

[10] J. Friedlander and H. Iwaniec, *Opera de Cribro*, AMS, Vol. 57, Providence, 2010. MR2647984 (2011d:11227)

[11] P.X. Gallagher, *On the distribution of primes in short intervals*, Mathematika **23** (1976) 4–9, *Corrigendum*, Mathematika **28** (1981) 86.

[12] D. A. Goldston, ‘On Bombieri and Davenport’s theorem concerning small gaps between primes’, *Mathematika* **39** (1992), 10–17. MR1176465 (93h:11102)

[13] D. A. Goldston, Y. Motohashi, J. Pintz, and C. Y. Yıldırım, ‘Small gaps between primes exist’, *Proc. Japan Acad. Ser. A Math. Sci.* **82** (2006), 61–65. MR2222213 (2007a:11135)

[14] D. A. Goldston, J. Pintz, and C. Y. Yıldırım, ‘The path to recent progress on small gaps between primes’, *Clay Math. Proc.* **7** (2007), 129–139. MR2362197 (2008j:11122)

[15] D. A. Goldston, J. Pintz, and C. Y. Yıldırım, ‘Primes in tuples. I’ *Ann. of Math. (2)* **170**(2) (2009), 819–862. MR2552109 (2011c:11146)

[16] D. A. Goldston, J. Pintz, and C. Y. Yıldırım, ‘Primes in tuples. II’ *Acta Math.* **204**(1) (2010), 1–47. MR2600432 (2011f:11121)

[17] D. A. Goldston, S. Graham, J. Pintz, and C. Y. Yıldırım, ’Small gaps between primes or almost primes’, *Trans. Amer. Math. Soc. (10)* **361** (2009), 5285–5330. MR2515812 (2010d:11108)

[18] D. A. Goldston, S. Graham, J. Pintz, and C. Y. Yıldırım, ‘Small gaps between products of two primes’, *Proc. Lond. Math. Soc. (3)* **98** (2009), 741–774. MR2500871 (2010a:11179)

[19] D. A. Goldston, S. Graham, J. Pintz, and C. Y. Yıldırım, ‘Small gaps between almost primes, the parity problem, and some conjectures of Erdős on consecutive integers’, *Int. Math. Res. Not. IMRN* (2010), Article ID rnq124, 12 pages. doi:10.1093/imrn/rnq124

[20] D. A. Goldston, S. Graham, J. Pintz, and C. Y. Yıldırım, ‘Positive proportion of small gaps between consecutive primes’, *Publ. Math. Debrecen* **79** (2011), 433–444. MR2907978

[21] H. Halberstam and H.- E. Rihert, *Sieve Methods*, Academic Press (London, 1974).

[22] 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 Math. **44** (1923), 1–70.

[23] G. H. Hardy and J. E. Littlewood, *Some problems of 'Partitio Numerorum': VII.* Unpublished manuscript.

[24] D. R. Heath-Brown, 'The divisor function at consecutive integers', *Mathematika* **31** (1984), 141–149. MR762186 (86c:11071)

[25] A. J. Hildebrand, 'The divisor function at consecutive integers', *Pacific J. Math.* **129** (1987), 307–319. MR909033 (88k:11062)

[26] M.N. Huxley, 'An application of the Fouvry-Iwaniec theorem', *Acta Arith.* **43**, (4), (1984), 441–443. MR756293 (85k:11043)

[27] M.N. Huxley, 'Small differences between consecutive primes. II', *Mathematika* **24**, (2), (1977), 142–152. MR0466042 (57 #5925)

[28] A.E. Ingham, *The distribution of prime numbers*, Stechert-Hafner, Inc., New York, 1964 MR0184920 (32 #2391)

[29] H. Iwaniec, E. Kowalski *Analytic number theory*, AMS Colloquium Publications, Providence, RI, 2004. MR2061214 (2005h:11005)

[30] J. Kan and Z. Shan, 'On the divisor function $d(n)$', *Mathematika* **43** (1996), 320–322. MR1433278 (98b:11101)

[31] J. Kan and Z. Shan, 'On the divisor function $d(n)$ II', *Mathematika* **46** (1999), 419–423. MR1832632 (2003c:11122)

[32] H. Maier, 'Small differences between prime numbers', *Michigan Math. J.* **35**, (3), (1988), 323–344. MR978303 (90e:11126)

[33] J. Maynard, 'Bounded length intervals containing two primes and an almost-prime', (2012), Pre-print, available at http://arxiv.org/abs/1205.5020v1

[34] G. Z. Pilt'jaĭ, 'The magnitude of the difference between consecutive primes', *Studies in number theory* **4**, (1972), 73–79. MR0392867 (52 #13680)

[35] G. Ricci, 'Sull'andamento della differenza di numeri primi consecutivi', *Riv. Mat. Univ. Parma* **5**, (1954), 3–54. MR0067139 (16,675e)

[36] A. Schinzel and W. Sierpiński, 'Sur certaines hypothèses concernant les nombres premiers'. *Acta Arith. IV* , (1958), 185–208, MR0106202 (21 #4936)

[37] D. K. L. Shiu , 'Strings of congruent primes', *J. London Math. Soc. (2)* **61** (2000), 359–373. MR1760689 (2001f:11155)

[38] C.A. Spiro, *Ph.D. Thesis* (1981), 187 pp. MR2631715

[39] S. Uchiyama, 'On the difference between consecutive prime numbers', *Acta Arith.* **27**, (1975), 153–157. MR0366839 (51 #3085)

[40] Y. Wang, S. K. Hsiek, K. Yu, 'Two results on the distribution of prime numbers', *J. China Univ. Sci. Techn.* **1**, (1965), 32–38.