TURUN YLIOPISTO
UNIVERSITY OF TURKU

# APPROXIMATIONS TO LANDAU'S PROBLEMS ON PRIME NUMBERS

Jori Merikoski

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. AI OSA – TOM. 644 | ASTRONOMICA – CHEMICA – PHYSICA – MATHEMATICA | TURKU 2021

TURUN YLIOPISTO
UNIVERSITY OF TURKU

# APPROXIMATIONS TO LANDAU'S PROBLEMS ON PRIME NUMBERS

Jori Merikoski

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. AI OSA – TOM. 644 | ASTRONOMICA – CHEMICA – PHYSICA – MATHEMATICA | TURKU 2021

## University of Turku

Faculty of Science
Department of Mathematics and Statistics
Mathematics
Doctoral Programme in Mathematics and Computer Sciences


## Supervised by

Docent Kaisa Matomäki
University of Turku

## Reviewed by

Professor Roger Baker
Brigham Young University

Professor Kevin Ford
University of Illinois

## Opponent

Professor James Maynard
University of Oxford




ISBN 978-951-29-8410-7 (PRINT)
ISBN 978-951-29-8411-4 (PDF)
ISSN 0082-7002 (Print)
ISSN 2343-3175 (Online)
Painosalama, Turku, Finland 2021

UNIVERSITY OF TURKU
Faculty of Science
Department of Mathematics and Statistics
Mathematics
MERIKOSKI, JORI: Approximations to Landau's problems on prime numbers
Doctoral dissertation, 229 pp.
Doctoral Programme in Mathematics and Computer Sciences
April 2021

ABSTRACT

In this thesis we study the distribution of prime numbers by considering approximations to fundamental open conjectures about them. The conjectures in question are the so-called Landau's problems, named after Edmund Landau who in 1912 listed four basic questions about prime numbers at the International Congress of Mathematicians. All of the four problems remain unsolved and are widely considered to be beyond the reach of current mathematics.

To study prime numbers we use sieve methods to break the problem into parts that can be attacked with various analytical tools. Sieves can be viewed as combinatorial machines that take as inputs arithmetic information about a given set and produce as outputs lower and upper bounds for the number of primes or almost-primes in the set.

The thesis consists of four articles. In the first article we consider the problem of showing that all short intervals contain numbers with large prime factors. We use Harman's sieve method and Perron's formula to reduce the problem to certain mean value estimates for Dirichlet polynomials. We obtain a suitable factorization for these Dirichlet polynomials by using a much refined version of the argument of Heath-Brown and Jia from their work on the same problem. After this the Dirichlet polynomial mean values can be bounded by applying the method of Matomäki and Radziwiłł.

In the second article we study the set of limit points of the sequence of normalized prime gaps. Improving on previous results of Baker, Banks, Freiberg, Maynard, and Pintz, we show that at least one third of non-negative real numbers are limit points of the sequence of normalized prime gaps. To attack this problem we combine the Maynard-Tao sieve with Chen's sieve.

In the third and the fourth articles we consider two different approximations to the conjecture that there are infinitely many primes of the form $n^2 + 1$. In the third article we study the largest square divisor of shifted primes and in the fourth article we study the largest prime factor of $n^2 + 1$. In both cases we apply Harman's sieve method. The arithmetic information in the third and the fourth articles is obtained by a square moduli version of Zhang's bilinear equidistribution estimate and by the Deshouillers-Iwaniec bound for sums of Kloosterman sums, respectively. The third article improves the results of Matomäki, while the fourth article improves the results of Deshouillers and Iwaniec, and de la Bretèche and Drappeau.

KEYWORDS: prime numbers, Landau's problems, sieve methods, analytic number theory, Dirichlet polynomials, Kloosterman sums

TURUN YLIOPISTO
Matemaattis-luonnontieteellinen tiedekunta
Matematiikan ja tilastotieteen laitos
Matematiikka
MERIKOSKI, JORI: Approximations to Landau's problems on prime numbers
Väitöskirja, 229 s.
Matemaattis-tietotekninen tohtoriohjelma
Huhtikuu 2021


TIIVISTELMÄ

Tässä väitöskirjassa tutkimme alkulukujen jakaumaa tarkastelemalla aproksimaatiota niitä koskeviin keskeisiin avoimiin konjektuureihin. Nämä konjektuurit tunnetaan nimellä Landaun ongelmat. Tämä nimi tulee Edmund Landausta, joka vuonna 1912 kongressissa International Congress of Mathematicians nosti esiin neljä keskeistä ongelmaa alkulukuja koskien. Kaikki neljä ongelmaa ovat edelleen ratkaisematta ja niiden ajatellaan yleisesti olevan tämänhetkisen matematiikan ulottumattomissa.

Tutkiaksemme alkulukuja käytämme seulamenetelmiä paloittelemaan ongelman osiin, joita voimme tarkastella käyttäen erilaisia analyyttisiä menetelmiä. Seuloja voidaan pitää kombinatorisina koneina, joihin syötetään aritmeettista informaatiota annetusta joukosta ja jotka tuottavat ala- ja ylärajoja kyseisessä joukossa olevien alkulukujen tai melkein-alkulukujen lukumäärälle.

Tämä väitöskirja koostuu neljästä artikkelista. Ensimmäisessä artikkelissa osoitamme, että kaikilta lyhyiltä väleiltä löytyy luku, jolla on suuri alkutekijä. Käytämme Harmanin seulamenetelmää ja Perronin kaavaa palauttamaan ongelman tiettyihin Dirichlet'n polynomien keskiarvohin. Osoitamme näille Dirichlet'n polynomeille tarvittavan tekijöihinjaon käyttäen paljon tarkennetua versiota Heath-Brownin ja Jian argumentista koskien samaa onglemaa. Tätä hyödyntäen pystymme osoittamaan näille Dirichet'n polynomien keskiarvoille ylärajan käyttäen Matomäen-Radziwiłłin menetelmää.

Toisessa artikkelissa tutkimme normalisoitujen alkulukujen etäisyyksien jonon kasautumispisteiden joukkoa. Osoitamme, että ainakin yksi kolmasosa kaikista ei-negatiivisista reaaliluvuista ovat normalisoitujen alkulukujen etäisyyksien jonon kasautumispisteitä, mikä parantaa Bakerin, Banksin, Freibergin, Maynardin, ja Pintzin aiempia tuloksia. Ongelman tutkimiseen käytämme yhdistelmää Maynardin-Taon seulasta ja Chenin seulasta.

Kolmannessa ja neljännessä artikkelissa tarkastelemme kahta eri aproksimaatiota konjektuuriin, jonka mukaan on olemassa äärettömän paljon muotoa $n^2 + 1$ olevia alkulukuja. Kolmannessa artikkelissa tutkimme alkulukujen siirtojen suurinta neliötekijää ja neljännessä artikkelissa tutkimme lukujen $n^2 + 1$ suurimpia alkutekijöitä. Kolmas artikkeli parantaa Matomäen tulosta, kun taas neljäs artikkeli parantaa Deshouillersin-Iwaniecin ja de la Bretèchen-Drappeaun tulosta. Molemmissa tapauksissa käytämme Harmanin seulamenetelmää. Aritmeettinen informaatio saadaan kolmannessa artikkelissa neliömoduli-versiosta Zhangin bilineaarisesta tasanjakautumisarviosta ja neljännessä artikkelissa Deshouillersin-Iwaniecin arvioista summille Kloostermanin summista.



ASIASANAT: alkuluvut, Landaun ongelmat, seulamenetelmät, analyyttinen lukuteoria, Dirichlet'n polynomit, Kloostermanin summat

# Preface

*Be bold, but wary! Keep up your merry hearts, and ride to meet your fortune!*
— J.R.R. Tolkien, The Lord of the Rings

There is always an inner conflict within the heart of the working number theorist. On one hand we are very fortunate that there are many fundamental questions about prime numbers which remain unsolved, such as Landau's problems which are the central theme of this thesis. On the other hand, we are faced with the fact that these conjectures are so overwhelmingly difficult that they will likely not be solved within our lifetime, the reality of which becomes the more apparent the further we progress in learning the art. What point is there then to put so much effort into the unattainable?

This is the position I found myself in as an undergraduate student, having been captured by the allure of the prime numbers many years earlier. It is very easy to be disheartened while pursuing prime numbers, which I must admit I was. Yet here I am, finishing a doctoral thesis on prime numbers and the conjectures I feel most deeply about. I was profoundly inspired by two recent breakthroughs in number theory, namely, the results of Zhang and Maynard on bounded gaps between primes and the result of Matomäki and Radziwiłł on multiplicative functions in short intervals. Seeing other people charge forward fearlessly made me take heart and follow them into battle.

The key to success in research lies not in answers but in asking the right questions (a cliché but a very accurate one). In the case of fundamental questions on prime numbers, this often amounts to finding an easier approximate version of the original problem, which still captures some key features. All of the articles in this thesis follow this approach. The guiding principle in my work has been the dichotomy of sieve methods and arithmetic information, and the goal I set for myself at the start was to learn as much as possible on both topics. I take great pride in how far I have come while realizing that I have touched only a fraction of the substance.

During the work of this thesis I was supported by the UTUGS graduate school as well as grants from the Finnish cultural foundation and the Magnus Ehrnrooth foundation. My visit to ETH Zürich in the Autumn semester 2018 was funded the Emil Aaltonen foundation. Part of the thesis was also completed while I was working on projects funded by the Academy of Finland (project no. 319180) and the Emil Aaltonen foundation. I am thankful to all of the above for funding which has allowed me to focus on mathematics full-time.

I would like to express my deepest gratitude to my supervisor Prof. Kaisa Matomäki, for inspiring discussions, support and guidance, encouragement, and generosity over the course of the past four years. I have been frequently amazed by the swiftness and the detail in comments to early versions of my manuscripts. I could not have wished for a better mentor.

I am deeply grateful to Prof. Emmanuel Kowalski for supervision during my visit to ETH Zürich. Our many discussions have been most inspiring for me. I also wish to thank all the people I have met during conferences and visits for memorable moments and helpful exchanges. I wish to thank my colleagues at the University of Turku for many interesting discussions.

I am also grateful to all my friends. I feel privileged to have such amazing people around me. Special thanks go to the ‘round table’ group from Jyväskylän Lyseon lukio, for the past decade of friendship and many more to follow.

I cannot thank enough my parents for raising and inspiring me. They have always supported and encouraged me, no matter what I have set out to do. I am also grateful to my sister for fun memories and support.

Lastly, the greatest thanks go to my amazing wife Katri for love and support over the past years, as well as for reading this thesis and sprucing up the language. I would not have gotten so far without her. May our years together be long and full of happiness.

# Contents

# List of Original Publications

This thesis consists of the following articles of which I am the sole author.

[I] Large prime factors on short intervals. *Math. Proc. Cambridge Philos. Soc.* 170(1):1-50, 2021.

[II] Limit points of normalized prime gaps. *J. Lond. Math. Soc.* 102(1):99-124, 2020.

[III] On the largest square divisor of shifted primes. *Acta Arith.* 196(4): 349-386, 2020.

[IV] On the largest prime factor of $n^2 + 1$. *To appear in J. Eur. Math. Soc.*

# Notations

We have collected here a list of various notations used. These should mostly coincide with the notations used in the articles [I-IV].

## Sets

- $\mathbb{Z}$ — the set of integers $\{\dots, -2, -1, 0, 1, 2, \dots\}$.
- $\mathbb{P}$ — the set of prime numbers $\{2, 3, 5, 7, 11, \dots\}$.
- $\mathbb{Z}/q\mathbb{Z}$ — the set of integers modulo $q$ for any integer $q \geq 1$, $\{0, 1, \dots, q-1\}$.
- Square-free — an integer which is not divisible by any square of a prime.
- $z$-smooth — an integer with no prime factors greater than $z$.
- $\mathbb{L}$ — the set of limit points of the sequence of normalized prime gaps (see Section 4).
- $[a, b]$, $(a, b)$ — closed and open intervals from $a$ to $b$, respectively.
- $\mathcal{H}$ — admissible tuple (see Section 2.2).

## Letters

- $a, b, c, d, h, i, j, k, \ell, m, n, q, r, s$ — integers.
- $p, p_n$ — prime numbers.
- $\epsilon, \eta$ — small constants, may vary from place to place.
- $C$ — a large constant, may vary from place to place.

## Functions

- $\mathbf{1}_S$ — indicator function of $S$, where $S$ is either a set or some statement.
- $\tau(n)$ — the divisor function, counts the number of divisors of $n$.
- $\Omega(n)$ — the number of prime factors of $n$ counted with multiplicity.
- $\varphi(n)$ — the Euler tontient function, number of integers in $[1,n]$ coprime to $n$.
- $\Lambda(n)$ — the von Mangoldt function, equal to $\log p$ if $n = p^k, k \geq 1$, and 0 otherwise.
- $\lambda(n)$ — the Liouville function, equal to $(-1)^{\Omega(n)}$.
- $\mu(n)$ — the Möbius function, the restriction of $\lambda(n)$ to square-free integers.
- $P^+(n)$ — the largest prime factor of $n$.
- $\rho(n)$ — the number of non-congruent solutions to $\nu^2 + 1 \equiv 0 \ (n)$.
- $\pi(x)$ — the number of primes up to $x$.
- $P(z)$ — the product of primes $p < z$ (see Section 2.1).
- $S(\mathcal{A}, z)$ — the sifting functions (see Section 2.1).
- $\omega(u)$ — the Buchstab function (see Section 2.1).
- $\nu_{\mathcal{H}}(n)$ — the Maynard-Tao sieve weights (see Section 2.2).
- $\lambda_d^+, \lambda_d^-, F_{\text{lin}}(s), f_{\text{lin}}(s)$ — linear sieve notations (see Section 2.3).
- $e(x)$ — $e^{2\pi i x}$.
- $e_q(x)$ — $e^{2\pi i x/q}$, an additive character modulo $q$.
- $S(a, b; c)$ — the Kloosterman sums (see Section 6.2).
- $(a, b)$ — the greatest common divisor of $a$ and $b$.
- $\alpha(m), \beta(n)$ — arbitrary bounded coefficients.

## Summation

- $\sum_{n \sim N}$ — shorthand for $\sum_{N < n \leq 2N}$.
- $\sum_{n\ (q)}$ — sum over residue classes modulo $q$, $\sum_{n=0}^{q-1}$.
- $n \equiv a \ (q)$ — congruence modulo $q$, means that $q$ divides $n - a$.

## Asymptotics

- $f(x) \sim g(x)$ — we have $f(x)/g(x) \to 1$ as $x \to \infty$.
- $f \ll g$, $f = O(g)$ — for functions $f$ and $g$ with $g$ positive there is some absolute constant $C$ such that $|f| \leq Cg$.
- $f \asymp g$ — both $f \ll g$ and $g \ll f$ hold.
- $f(x) = o(g(x))$ — we have $f(x)/g(x) \to 0$ as $x \to \infty$.

# 1 Introduction

In 1912 at the International Congress of Mathematicians Edmund Landau famously listed four fundamental problems about prime numbers that he considered to be 'unattackable at the present state of mathematics'.

1. (Goldbach's problem). Is every even integer greater than 2 a sum of two primes?

2. (Twin prime conjecture). Are there infinitely many prime numbers $p$ such that $p + 2$ is also a prime?

3. (Legendre's conjecture). Is there a prime between $n^2$ and $(n + 1)^2$ for every integer $n \geq 1$?

4. Are there infinitely many prime numbers of the form $n^2 + 1$?

These are known as *Landau's problems*. While the problems remain unsolved, they have inspired a substantial proportion of developments in analytic number theory. The four original articles in this thesis are all closely related to these basic problems (this is accidental but at the same time not a coincidence, since questions such as these originally kindled my passion for mathematics).

It is a common strategy to consider an approximate version of the problem when faced with an intractable question such as one of the above. For example, instead of proving that a given set contains infinitely many prime numbers, we could try to show that it contains infinitely many integers $n$ with a large prime factor $p \geq n^\theta$ for some fixed $0 < \theta \leq 1$ as large as possible. The parameter $\theta$ provides a measure of how well we understand the original problem, with $\theta = 1$ corresponding to the full problem. Admittedly, the previous statement is somewhat naive since we expect a jump in difficulty between $\theta = 1 - \epsilon$ for any $\epsilon > 0$ and $\theta = 1$.

The first article [I] of this thesis concerns an approximation to the third problem in Laundau's list. Since $(n+1)^2 = n^2 + 2n + 1$, Legendre's conjecture follows if we can show that for all sufficiently large $x > x_0$ there is a prime number in the short interval $[x, x + 2\sqrt{x}]$, provided that we can verify the conjecture numerically up to

$x_0$. We can apply analytic methods to count the number of primes in short intervals. By the Prime number theorem the number of primes up to $x$ is

$$\pi(x) := |\{p \leq x : p \in \mathbb{P}\}| \sim \int_2^x \frac{du}{\log u} = \frac{x}{\log x}(1 + O((1/\log x))) \quad \text{as} \quad x \to \infty, \tag{1.1}$$

that is, at height $u$ the density of primes is approximately $1/\log u$. Assuming the Lindelöf Hypothesis this asymptotic holds also for shorter intervals in the form

$$|\{x \leq p \leq x + x^{1/2+\epsilon} : p \in \mathbb{P}\}| \sim \int_x^{x+x^{1/2+\epsilon}} \frac{du}{\log u} \quad \text{as} \quad x \to \infty \tag{1.2}$$

for any fixed $\epsilon > 0$, which falls just short of Legendre's conjecture. By a famous heuristic of Cramér the corresponding asymptotic is conjectured to hold also for much shorter intervals of length $(\log x)^{2+\epsilon}$ for any $\epsilon > 0$ (cf. [27] for more details), which would more than suffice to solve the problem.

The best unconditional results for primes on short intervals are an asymptotic formula of the type (1.2) for the length $x^{7/12}$ (cf. [30], [36]), and a correct-order lower bound for intervals of length $x^{0.525}$ (cf. [6]). The length $x^{1/2}$ is a natural barrier for the current analytic methods. In the first article [I] we show that for all sufficiently large $x$ the interval $[x, x+x^{1/2}\log^{1.39} x]$ contains an integer with a prime factor $p > x^{18/19}$, improving on results of Heath-Brown and Jia [34], and Jia and Liu [41]. The methods used to prove this will be discussed in Section 3.

The second article [II] is closely related to Landau's second problem. To generalize the question, we can ask if for any given even integer $2k$ there are infinitely many primes $p$ such that $p + 2k$ is also a prime. It should be noted that the first and the second of Landau's problems are thought to be 'morally equivalent', that is, we expect that solution to either one will quickly provide a solution to the other (at least if in Goldbach's problem we want to prove the claim merely for all sufficiently large even integers). To see this, note that in both cases we are trying to count solutions to a linear equation among pairs of primes $(p, q)$ ($p + q = 2k$ for Goldbach and $p - q = 2$ for twin primes).

In [II] we approximate the second problem by requiring only that the difference $p - q$ is approximately equal to some given quantity. Additionally, we restrict the difference to be a gap between consecutive primes. More precisely, recall that by the Prime number theorem (1.1) the average distance from a prime $p$ to the next is asymptotically $\log p$. Therefore, it makes sense to ask how much of the set of non-negative real numbers is filled by the sequence normalized prime gaps $d_n := (p_{n+1} - p_n)/\log p_n$, where $p_n$ denotes the $n$th prime number. That is, what can we say about the measure of the set of $t \geq 0$ for which there is a subsequence of primes $\{p_{n_j}\}_{j=1}^{\infty}$ such that

$$p_{n_j+1} - p_{n_j} = (t + o(1)) \log p_{n_j} \quad \text{as} \quad j \to \infty. \tag{1.3}$$

Let $\mathbb{L}$ denote the set of limit points of the sequence of normalized prime gaps $\{d_n\}_{n=1}^{\infty}$. Then a conjecture of Erdős states that $\mathbb{L} = [0, \infty]$. Improving on results of Banks, Freiberg, and Maynard [11], and Pintz [52], the main result of [II] is that for every $T > 0$ we have

$$\mu(\mathbb{L} \cap [0, T]) \geq T/3,$$

where $\mu$ denotes the Lebesque measure on $\mathbb{R}$. In other words, at least one third of non-negative real numbers $t$ can be approximated by prime gaps in the sense of (1.3). We will discuss the details in Section 4.

In the last two articles [III], [IV] of this thesis we consider two different approximations to Landau's fourth problem. Detailed expositions of these articles appear in Sections 5 and 6, respectively. In [III] we try to show that there are infinitely many primes $p$ such that $p - 1$ is divisible by a large square of an integer $d^2 \geq p^{\theta}$ for as large $\theta \in (0, 1)$ as possible. Recall that the original problem asks for primes $p$ such that $p - 1$ is equal to a perfect square $n^2$. The main result in [III] states that there are infinitely many primes $p$ such that $p-1$ is divisible by a square $d^2 > p^{1/2+1/2000}$, improving on a result of Matomäki [45] who obtained this with the exponent $1/2-\epsilon$ for any $\epsilon > 0$. As with the problem of primes in short intervals, the threshold $\theta = 1/2$ is a natural limit from an analytic perspective. Therefore, even though the numerical improvement of 1/2000 in [II] is tiny, there is a big qualitative difference.

Complementary to this, in the last article [IV] we consider the largest prime factor of $n^2 + 1$. We show that there are infinitely many integers $n$ such that $n^2 + 1$ has a prime factor $p > n^{1.279}$. This improves results of Deshouillers and Iwaniec [17], and de la Bretèche and Drappeau [16].

In all of the articles we employ state-of-the-art sieve methods to attack the problems. Since there is a large overlap between the articles in this respect, we present an account of the different sieve methods in Section 2. In the subsequent four sections of the thesis we will mainly focus on the arithmetic and analytic aspects of the articles, with some additional details about the sieve arguments.

# 2 Sieve methods

In this section we will discuss the various sieves employed in the thesis. The articles [I], [III], and [IV] rely on Harman's sieve method, while in [II] we use a combination of the Maynard-Tao sieve and Chen's sieve. At the end of this section there are some brief philosophical remarks about sieves.

## 2.1 Harman's sieve method

In the articles [I], [III], and [IV] we apply Harman's sieve method (see Harman's book [29] for an excellent introduction). Harman's sieve method originally appeared in [28], and according to [29, Chapter 3.8] this was inspired by the work of Heath-Brown and Iwaniec [33]. Here we refer to is as a 'sieve method' rather than a 'sieve' because it is not a precisely stated theorem but a collection of techniques and a philosophy. This is because in each application the exact requirements are often very different. For example, the details in [I], [III], and [IV] are notably distinct and it is hard to imagine a general theorem which would encompass all three. On the downside this means that the computations tend to be quite involved, especially if we aim for a numerically optimal result. The benefit is that we have a very powerful, general, and flexible tool to study primes. We now give a brief exposition of the main ideas.

Suppose that we want to estimate a sum over primes

$$S(\mathcal{A}, 2\sqrt{x}) := \sum_{x<p\leq 2x} a_p \tag{2.1}$$

weighted by some non-negative sequence $\mathcal{A} = (a_n)_{x<n\leq 2x}$. Often $a_n$ is the characteristic function of some interesting set, for example, $a_n = \mathbf{1}_{\mathbb{P}}(2k-n)$ for the Goldbach problem, $a_n = \mathbf{1}_{\mathbb{P}}(n+2)$ in the case of twin primes, $a_n = \mathbf{1}_{[x,x+2\sqrt{x}]}(n)$ for Legendre's conjecture, and $a_n = \mathbf{1}_{\square}(n-1)$ for Landau's fourth problem. More generally, for any $z > 1$ we define

$$P(z) := \prod_{p<z} p$$

and

$$S(\mathcal{A}, z) := \sum_{\substack{x < n \leq 2x \\ (n, P(z))=1}} a_n.$$

The goal of sieve methods is to find upper and lower bounds for the sums over almost-primes $S(\mathcal{A}, z)$, with the eventual hope of understanding (2.1).

To compute (2.1) we will try to compare it with a simpler sum

$$S(\mathcal{B}, 2\sqrt{x}) := \sum_{x < p \leq 2x} b_p,$$

where $b_n$ are non-negative weights chosen so that heuristically we expect $S(\mathcal{A}, 2\sqrt{x}) \sim S(\mathcal{B}, 2\sqrt{x})$. Often we simply choose $b_n \equiv \delta$ for some appropriate constant $\delta = \delta(\mathcal{A})$, so that by the Prime number theorem $S(\mathcal{B}, 2\sqrt{x}) \sim \delta x/\log x$. In any case, we are comparing an unknown sum along primes with weights $a_p$ to a sum that we already know how to compute.

To get an asymptotic formula for $S(\mathcal{A}, 2\sqrt{x})$ we can use the well-known identity of Vaughan [58], which is a simplified version of Vinogradov's earlier work. Let $\alpha(n)$ and $\beta(n)$ denote arbitrary coefficients which are bounded in absolute value by the divisor function $\tau(n)$. Vaughan's identity is a combinatorial decomposition for primes which roughly states that we can obtain the desired asymptotic $S(\mathcal{A}, 2\sqrt{x}) \sim S(\mathcal{B}, 2\sqrt{x})$ provided that we have Type I information

$$\sum_{d \leq D} \sum_{n \sim x/d} \alpha(d) a_{dn} = \sum_{d \leq D} \sum_{n \sim x/d} \alpha(d) b_{dn} + O(S(\mathcal{B}, 2\sqrt{x})/\log^C x) \tag{2.2}$$

and Type II information

$$\sum_{m \sim M} \sum_{n \sim N} \alpha(m)\beta(n) a_{mn} = \sum_{m \sim M} \sum_{n \sim N} \alpha(m)\beta(n) b_{mn} + O(S(\mathcal{B}, 2\sqrt{x})/\log^C x) \tag{2.3}$$

for sufficiently wide ranges of $D$, $M$, and $N$ with $MN = x$. In the error term we typically need to save just some power of logarithm (for example, $C = 10$). For any $\gamma \in (0, 1/2)$ the ranges $D \leq \max\{x^{1-\gamma}, x^{2\gamma}\}$ and $x^{\gamma} \leq M, N \leq x^{1-\gamma}$ would suffice [29, Chapter 2]. Note that the wider the range of $D$ the narrower we can make the range of $M$ and $N$, and vice versa. The parameter $D$ is also known as the level of distribution of $\mathcal{A}$.

We briefly explain why such a decomposition might be desirable. In the Type I sums we already have an unweighted variable $n$, so that (2.2) can usually be obtained either by trivial arguments or by Fourier analytic techniques (for example, Poisson summation). In the Type II sums the important feature is that we can control the lengths $M$ and $N$, which allows us to apply the Cauchy-Schwarz inequality at a suitable moment. A classic example of this is Vinogradov's proof that every sufficiently large odd number is a sum of three primes (see [15], for instance). We will see further instances of this in sections 3, 5, and 6. In fact, such an argument is so commonplace in analytic number theory that 'to Cauchy-Schwarz' is a widely recognized verb.

Suppose now that we fall short of the requirements for Vaughan's identity, that is, that the ranges of $D$, $M$, and $N$ where we can handle (2.2) and (2.3) are too narrow. All is not lost since by Harman's sieve method we can still try to prove lower and upper bounds for $S(\mathcal{A}, \sqrt{x})$ (recall that $a_n$ are non-negative). It turns out that to show that $S(\mathcal{A}, \sqrt{x}) \gg S(\mathcal{B}, \sqrt{x})$ we often do not need a lot of Type II information. Furthermore, Harman's sieve method is better than Vaughan's identity for pedagogic reasons as it very concretely reveals why primes should possess a decomposition into sums of Type I and Type II. At least for me, the first time I saw Vaughan's identity it was nothing short of a magic trick, which is pedagogically frustrating.

A helpful device used in Harman's sieve method is Buchstab's identity, which really is just the inclusion-exclusion principle with a convenient notation. For any integer $d \geq 1$ and for any sequence $\mathcal{C} = \{c_n\}_{x<n\leq 2x}$, define $\mathcal{C}_d := \{c_{dn}\}_{x/d<n\leq 2x/d}$, so that

$$S(\mathcal{C}_d, z) = \sum_{\substack{n\sim x/d \\ (n,P(z))=1}} c_{dn}.$$

Then Buchstab's identity states that for any $1 \leq y < z < x$

$$S(\mathcal{C}, z) = S(\mathcal{C}, y) - \sum_{y\leq p<z} S(\mathcal{C}_p, p). \tag{2.4}$$

The sieve method itself is perhaps best explained by a simple working example. We consider the situation in [45], where Type I information (2.2) is available for

$$D \ll x^{1/2}$$

and the Type II sums (2.3) can be handled in the range

$$x^{\theta} \ll M, N \ll x^{1-\theta} \quad \text{for} \quad \theta := 3/8 + 2\epsilon$$

for any fixed $\epsilon > 0$. This is clearly not sufficient for Vaughan's identity but we can still show the lower bound

$$S(\mathcal{A}, 2\sqrt{x}) \geq (2/3 + o(1))S(\mathcal{B}, 2\sqrt{x}).$$

Let $z := x^{1/4-4\epsilon}$. The exponent is determined by the width of the Type II range which is $(1-\theta)-\theta$. We begin by applying Buchstab's identity (2.4) twice

$$
\begin{aligned}
S(\mathcal{A},2\sqrt{x}) &= S(\mathcal{A},z) - \sum_{z\leq p<2\sqrt{x}} S(\mathcal{A}_p,p)\\
&= S(\mathcal{A},z) - \sum_{z\leq p<x^\theta} S(\mathcal{A}_p,z) - \sum_{x^\theta\leq p<2\sqrt{x}} S(\mathcal{A}_p,p) + \sum_{\substack{z\leq p_2<p_1<x^\theta\\ p_1p_2^2<4x}} S(\mathcal{A}_{p_1p_2},p_2)\\
&=: S_1(\mathcal{A})-S_2(\mathcal{A})-S_3(\mathcal{A})+S_4(\mathcal{A}) \ \geq \ S_1(\mathcal{A})-S_2(\mathcal{A})-S_3(\mathcal{A}). \qquad (2.5)
\end{aligned}
$$

Notice that it is crucial here that the sequence $a_n$ is non-negative so that we can use positivity to drop the fourth sum. Similarly, we obtain a decomposition

$$S(\mathcal{B},2\sqrt{x}) = S_1(\mathcal{B})-S_2(\mathcal{B})-S_3(\mathcal{B})+S_4(\mathcal{B}), \qquad (2.6)$$

where $S_j(\mathcal{B})$ are as in (2.5) but with $a_n$ replaced by $b_n$. We now claim that by (2.2) and (2.3) we get

$$S_j(\mathcal{A}) = S_j(\mathcal{B}) + O(S(\mathcal{B},2\sqrt{x})/\log^C x) \quad \text{for} \quad j\in\{1,2,3\}. \qquad (2.7)$$

We prove this below, but let us first show how to apply this. Assuming (2.7) we obtain from (2.5) and (2.6)

$$
\begin{aligned}
S(\mathcal{A},2\sqrt{x}) &\geq (1+o(1))(S_1(\mathcal{B})-S_2(\mathcal{B})-S_3(\mathcal{B}))\\
&= (1+o(1))S(\mathcal{B},2\sqrt{x}) - (1+o(1))S_4(\mathcal{B}),
\end{aligned}
$$

where by the Prime number theorem (1.1) (using $b_n\equiv\delta$ and substituting in the last step $n_j = x^{\alpha_j}$)

$$
\begin{aligned}
S_4(\mathcal{B}) &= \sum_{\substack{z\leq p_2<p_1<x^\theta\\ p_1p_2^2<4x}} S(\mathcal{B}_{p_1p_2},p_2) = \delta \sum_{\substack{z\leq p_2<p_1<x^\theta\\ p_1p_2^2<4x}} \sum_{x/p_1p_2<p\leq 2x/p_1p_2} 1\\
&\sim \delta x \sum_{\substack{z\leq p_2<p_1<x^\theta\\ p_1p_2^2<4x}} \frac{1}{p_1p_2\log(x/p_1p_2)}\\
&\sim \delta x \sum_{\substack{z\leq n_2<n_1<x^\theta\\ n_1n_2^2<4x}} \frac{1}{n_1(\log n_1)n_2(\log n_2)\log(x/n_1n_2)}\\
&= (1+O(\epsilon))S(\mathcal{B},2\sqrt{x}) \int_{1/4}^{\theta}\int_{1/4}^{\min\{\alpha,(1-\alpha_1)/2\}} \frac{d\alpha_2 d\alpha_1}{\alpha_1\alpha_2(1-\alpha_1-\alpha_2)}. \qquad (2.8)
\end{aligned}
$$

A calculation reveals that the double integral is $\leq 1/3$, so that we get the desired lower bound $S(\mathcal{A},2\sqrt{x}) \geq (2/3+o(1))S(\mathcal{B},2\sqrt{x})$.

*Proof of the claim* (2.7). Recall that we are assuming that (2.2) and (2.3) hold in the ranges $D \ll x^{1/2}$ and $x^\theta \ll M, N \ll x^{1-\theta}$. Let $\mathcal{C} = (c_n) \in \{\mathcal{A}, \mathcal{B}\}$. For $j = 3$ we note that

$$S_3(\mathcal{C}) = \sum_{x^\theta \leq p < 2\sqrt{x}} S(\mathcal{C}_p, p) = \sum_{x^\theta \leq p < 2\sqrt{x}} \sum_{x/p < q \leq 2x/p} c_{pq},$$

so that the asymptotic formula (2.7) follows from the Type II information (2.3) once we split the sums dyadically and remove the cross-conditions between $p$ and $q$ by Perron's formula (cf. [29, Chapter 3.2], for example).

For $j \in \{1, 2\}$ we have to evaluate sums of the form

$$\sum_{m \sim M} \alpha(m) S(\mathcal{C}_m, z) = \sum_{m \sim M} \alpha(m) \sum_{\substack{x/m < n \leq 2x/m \\ (n, P(z)) = 1}} c_{mn},$$

where $M \leq x^\theta$ and $z = x^{1/4 - 4\epsilon}$. We use the Möbius function to expand these sums to get

$$\begin{aligned} \sum_{m \sim M} \alpha(m) S(\mathcal{C}_m, z) &= \sum_{m \sim M} \alpha(m) \sum_{x/m < n < 2x/m} c_{mn} \sum_{d | (n, P(z))} \mu(d) \\ &= \sum_{m \sim M} \alpha(m) \sum_{d | P(z)} \mu(d) \sum_{x/dm < n < 2x/dm} c_{dmn} = S_{\mathrm{I}}(\mathcal{C}) + S_{\mathrm{II}}(\mathcal{C}), \end{aligned}$$

where we have split the sum depending on the size of $d$ into

$$S_{\mathrm{I}}(\mathcal{C}) := \sum_{m \sim M} \alpha(m) \sum_{\substack{d | P(z) \\ dM \leq x^\theta}} \mu(d) \sum_{x/dm < n < 2x/dm} c_{dmn} \qquad \text{and}$$

$$S_{\mathrm{II}}(\mathcal{C}) := \sum_{m \sim M} \alpha(m) \sum_{\substack{d | P(z) \\ dM > x^\theta}} \mu(d) \sum_{x/dm < n < 2x/dm} c_{dmn}.$$

For $S_{\mathrm{I}}(\mathcal{C})$ we get an asymptotic formula directly from (2.2) since $dm \ll D$. For $S_{\mathrm{II}}(\mathcal{C})$ we write $d = p_1 \cdots p_k$ with $p_1 > \cdots > p_k$ to get

$$S_{\mathrm{II}}(\mathcal{C}) = \sum_{k \ll \log x} (-1)^k \sum_{\substack{p_k < \cdots < p_1 < z \\ p_1 \cdots p_k M > x^\theta}} \sum_{m \sim M} \alpha(m) \sum_{x/dm < n < 2x/dm} c_{p_1 \cdots p_k mn}.$$

Since $p_j < z$, by the greedy algorithm we find a unique $\ell \leq k$ such that

$$x^\theta < p_1 \cdots p_\ell M \leq x^\theta z = x^{1-\theta} \qquad \text{and} \qquad p_1 \cdots p_{\ell-1} M \leq x^\theta.$$

Hence, denoting $m' := m p_1 \cdots p_\ell$ and $n' := n p_{\ell+1} \cdots p_k$, splitting the ranges dyadically, and removing the cross-conditions between $m'$ and $n'$ such as $p_{\ell+1} < p_\ell$ by

Perron's formula [29, Chapter 3.2], we can decompose $S_{\mathrm{II}}(\mathcal{C})$ into $\ll \log^4 x$ sums of the form

$$\sum_{m' \sim M'} \sum_{n' \sim N'} \alpha_{\ell,k}(m')\beta_{\ell,k}(n')c_{m'n'}$$

with $M'N' \asymp x$ and $x^\theta \ll M', N' \ll x^{1-\theta}$, so that we get an asymptotic formula for $S_{\mathrm{II}}(\mathcal{C})$ by (2.3). Thus, (2.7) holds for $j \in \{1, 2, 3\}$, completing the proof.

A key parameter in the above argument is the width of the Type II information, which in the above example is $(1-\theta)-\theta = 1/4-4\epsilon$. As mentioned, this determines the size of $z = x^{1-2\theta}$ in the Buchstab decompositions. This was chosen so that the sums $S_{\mathrm{II}}(\mathcal{A})$ could be computed. Note also that the sieve argument is continuous with respect to the quality of the arithmetic information. We can obtain a non-trivial lower bound for $S(\mathcal{A}, 2\sqrt{x})$ with even narrower Type II information, but then we might need to iterate the above argument. That is, in the sum corresponding to $S_4(\mathcal{C})$ that we simply discarded, we could apply Buchstab's identity again twice to generate even more Type II sums and reduce the deficit. In general, we iterate Buchstab's identity to get a finite decomposition

$$S(\mathcal{C}, 2\sqrt{x}) = \sum_j \epsilon_j S_j(\mathcal{C})$$

for $\mathcal{C} \in \{\mathcal{A}, \mathcal{B}\}$ with $\epsilon_j \in \{1, -1\}$, where $S_j(\mathcal{C})$ are sums over almost-primes. To produce a lower bound, for all $j$ with $\epsilon_j = -1$ we need an asymptotic formula. We often also get an asymptotic formula for some of the positive terms, but suppose that there is a set $\mathcal{J}$ of indices that we cannot handle and $\epsilon_j = 1$ for $j \in \mathcal{J}$. Then we get

$$\begin{aligned} S(\mathcal{A}, 2\sqrt{x}) &= \sum_j \epsilon_j S_j(\mathcal{A}) \geq \sum_{j \notin \mathcal{J}} \epsilon_j S_j(\mathcal{A}) \sim \sum_{j \notin \mathcal{J}} \epsilon_j S_j(\mathcal{B}) \\ &= S(\mathcal{B}, 2\sqrt{x}) - \sum_{j \in \mathcal{J}} S_j(\mathcal{B}). \end{aligned}$$

Therefore, to get a non-trivial lower bound we need for $\epsilon > 0$

$$\sum_{j \in \mathcal{J}} S_j(\mathcal{B}) \leq (1 + o(1) - \epsilon) S(\mathcal{B}, 2\sqrt{x}).$$

At this point we need to discuss the dirty part of the work, that is, translating sums over almost-primes into integrals which can be bounded numerically. To count the almost-primes, let $\omega(u)$ denote the Buchstab function (cf. [29, Chapter 1] for the properties below). Then using the Prime number theorem we can prove that for $y^\epsilon < z < y$

$$\sum_{y < n \leq 2y} \mathbf{1}_{(n, P(z))=1} = (1 + o(1))\omega\left(\frac{\log y}{\log z}\right) \frac{y}{\log z}. \tag{2.9}$$

Note that for $1 < u \le 2$ we have $\omega(u) = 1/u$. In the numerical computations we need the following simple bounds (cf. [40, Lemma 20])

$$\omega(u) \begin{cases} = 0, & u < 1 \\ = 1/u, & 1 \le u < 2 \\ = (1 + \log(u-1))/u, & 2 \le u < 3 \\ \le 0.5644, & 3 \le u < 4 \\ \le 0.5617, & u \ge 4 \\ \ge 0.5607, & u \ge 2.47. \end{cases} \tag{2.10}$$

Recall that $\omega(u)$ approaches exponentially the constant $e^{-\gamma} = 0.56158\ldots$ ($\gamma$ being the Euler-Mascheroni constant), so that this is quite a good approximation.

Assume that the range $\mathcal{U} \subset [x^\epsilon, x^{1-\epsilon}]^k$ is sufficiently well-behaved, for example, an intersection of sets of the type $\{\boldsymbol{u} : u_i < u_j\}$ or $\{\boldsymbol{u} : V < f(u_1, \ldots, u_k) < W\}$ for some polynomial $f$ and some fixed $V, W$. If $b_n \equiv \delta$, then by a similar argument as in (2.8) using (2.9)

$$\sum_{(q_1,\ldots,q_k)\in\mathcal{U}} S(\mathcal{B}_{q_1,\ldots,q_k}, q_k) = (1 + o(1)) S(\mathcal{B}, 2\sqrt{x}) \int \omega(\boldsymbol{\beta}) \frac{d\beta_1 \cdots d\beta_k}{\beta_1 \cdots \beta_{k-1}\beta_k^2}, \tag{2.11}$$

where the integral is over the range $\{\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k) : (x^{\beta_1}, \ldots, x^{\beta_k}) \in \mathcal{U}\}$, and

$$\omega(\boldsymbol{\beta}) := \omega\left(\frac{1 - \beta_1 - \cdots - \beta_k}{\beta_k}\right).$$

With this we get

$$\sum_{j\in\mathcal{J}} S_j(\mathcal{B}) \sim \left(\sum_{j\in\mathcal{J}} \Omega_j\right) S(\mathcal{B}, 2\sqrt{x}),$$

for certain integrals $\Omega_j$, so that the sieve argument is successful if $\sum_{j\in\mathcal{J}} \Omega_j < 1$. In the articles the integrals $\Omega_j$ are estimated numerically by using a Python code to compute a rigorous upper bound.

## 2.2 The Maynard-Tao sieve

The main tool used in the article [II] is the Maynard-Tao sieve [48], which is a refinement of the GPY sieve (named after its creators Goldston, Pintz, and Yıldırım) [26]. In our case [II] we are able to use this (the version in [11], to be precise) as a 'black box', but we give here a short outline of the main ideas. To approximate the twin prime conjecture we can ask if there are infinitely many prime pairs such that their difference is bounded by some absolute constant. Zhang [62] famously proved that this is indeed the case, showing that

$$\liminf_{n\to\infty} (p_{n+1} - p_n) \leq 7 \cdot 10^7. \tag{2.12}$$

Zhang used the GPY sieve, and his improvement was to obtain a suitable version of the Bombieri-Vinogradov theorem on primes in arithmetic progressions which extended the range of moduli past the so-called square-root barrier (compare (2.16) and (5.2) for the precise statements).

Soon after this Maynard [48] (and Tao independently) found an improvement on the GPY sieve which showed bounded gaps between primes using simply the Bombieri-Vinogradov theorem, and with the constant $7 \cdot 10^7$ in (2.12) replaced by 600. The current world-record of the constant is 246 which was obtained by the collaborative Polymath project to optimize the arguments [54].

We now describe the main idea of the Maynard-Tao sieve. We say that a $K$-tuple

$$\mathcal{H} := \{h_1, \cdots, h_K\}$$

of distinct integers is admissible if for every prime $p$ it avoids at least one residue class, that is, $|\mathcal{H} \bmod p| \leq p-1$. This means that the polynomial $n \mapsto (n+h_1)\cdots(n+h_K)$ has no fixed prime divisors, so that in principle it is possible that for infinitely many $n$ the set $n+\mathcal{H}$ consists solely of primes. Let $\nu_{\mathcal{H}}(n)$ be a non-negative weight function, and suppose that we can show for all large $x$ that

$$S_2 - S_1 := \sum_{x<n\leq 2x} \Big( \sum_{h\in\mathcal{H}} \mathbf{1}_{\mathbb{P}}(n+h) - 1 \Big) \nu_{\mathcal{H}}(n) > 0. \tag{2.13}$$

Then, since $\nu_{\mathcal{H}}(n) \geq 0$, there must be an integer $n \in (x, 2x]$ such that $n+\mathcal{H}$ contains at least two primes, which implies that (2.12) holds with the constant

$$\max_{i,j\leq K} |h_j - h_i|.$$

The main problem then is the choice of the weight function $\nu_{\mathcal{H}}(n)$ so that we can show (2.13). Intuitively we want to make $S_2$ as large as possible. Therefore, we want to construct $\nu_{\mathcal{H}}(n)$ so that it is concentrated on $n$ where there are many primes in $n+\mathcal{H}$, under the restriction that we need to be able to compute the sums $S_1$ and

$$\Sigma_h := \sum_{x<n\leq 2x} \mathbf{1}_{\mathbb{P}}(n+h)\nu_{\mathcal{H}}(n). \tag{2.14}$$

The first move is to restrict the weight $\nu_{\mathcal{H}}(n)$ to a suitable arithmetic progression $n \equiv b \ (W)$ for $W = \prod_{p\leq z} p$ with some small $z$, in such a way that there are no small prime divisors of $(n+h_1)\cdots(n+h_K)$. This is possible since $\mathcal{H}$ is admissible. We can choose $z = \log\log\log x$ (so that $W \ll \log\log x$), for instance.

To find weights $\nu_{\mathcal{H}}(n)$ that are concentrated on primes we take inspiration from Selberg's upper-bound sieve (see [23, Chapter 7], for instance). To show an upper bound for a sum along primes such as (2.1), we need a point-wise upper bound $\mathbf{1}_{\mathbb{P}}(n) \leq \nu(n)$. In particular, we require $\nu(n) \geq 0$. A very elegant way to do this is to choose $\nu(n)$ to be a square

$$\nu(n) := \Big( \sum_{\substack{d|n \\ d \leq D^{1/2}}} \lambda_d \Big)^2$$

with $\lambda_1 = 1$ (so that $\nu(p) = 1$ for all primes $p > D^{1/2}$), and then optimize the rest of the weights $\lambda_d \in \mathbb{R}$. Here $D < x$ is the level of distribution for $\mathcal{A}$ and it is determined by the range for which we can prove (2.2). To motivate the choice of the weights, recall that

$$\mathbf{1}_{\mathbb{P}}(n) = \sum_{d|(n, P(\sqrt{2n}))} \mu(d),$$

so that a reasonable guess would be $\lambda_d = \mu(d)F(d)$ for some fairly simple function $F$. It is possible to perform the optimization of $\lambda_d$ precisely (see [23, Chapter 7.1], for instance), but an asymptotically equivalent choice is

$$\lambda_d = \mu(d) \Big( \frac{\log \sqrt{D}/d}{\log \sqrt{D}} \Big)^{\kappa},$$

where $\kappa$ is the so-called sieve dimension [23, Chapter 7.2.3]. That is, $F$ is a smooth function decaying from 1 to 0 on $[1, \sqrt{D}]$.

We do not explicitly know the support of the Selberg sieve weights $\nu(n)$, but since this produces very good upper bounds for (2.1) we know that $\nu(n)$ is not very far off from $\mathbf{1}_{\mathbb{P}}(n)$. Motivated by this the GPY sieve [26] uses weights of the form

$$\nu_{\mathcal{H}}^{(\mathrm{GPY})}(n) = \Big( \sum_{\substack{d|(n+h_1)\cdots(n+h_K) \\ d \leq \sqrt{D}}} \lambda_d \Big)^2$$

with $\lambda_d = \mu(d)F(d)$ for some suitable smooth function $F$. The refinement in the Maynard-Tao sieve [48] is to allow the weights $\lambda_d$ to depend independently on the divisors of $n + h_j$, so that we have

$$\nu_{\mathcal{H}}(n) := \Big( \sum_{\substack{d_j|(n+h_j)\,\forall j \\ d_1\cdots d_K \leq \sqrt{D}}} \lambda_{d_1,\ldots,d_K} \Big)^2, \tag{2.15}$$

with $\lambda_{d_1,\dots,d_K}$ essentially of the form

$$\mu(d_1)\cdots\mu(d_K)F(d_1,\dots,d_K)$$

for some smooth function $F$ to be optimized. In addition, if $(d_1\cdots d_K,W)\neq 1$, then we set $\lambda_{d_1,\cdots,d_K}=0$.

To compute the $\Sigma_h$ defined in (2.14), we rely on the Bombieri-Vinogradov theorem which says that primes are evenly distributed among primitive residue classes on average over moduli, that is, for any $\epsilon>0$

$$\sum_{d\le x^{1/2-\epsilon}} \max_{(a,d)=1} \Big| \sum_{\substack{p\le x\\ p\equiv a\,(d)}} 1 - \frac{1}{\varphi(d)}\sum_{p\le x} 1\Big| \ll_C x\log^{-C} x. \tag{2.16}$$

This gives the Type I information for the current problem, with $D=x^{1/2-\epsilon}$. With the choice of $\nu_{\mathcal{H}}$ we have (denoting $p:=n+h_k$)

$$\begin{aligned}
\Sigma_{h_k} &= \sum_{\substack{x<n\le 2x\\ n\equiv b\,(W)}} \mathbf{1}_{\mathbb{P}}(n+h_k)\Bigg(\sum_{\substack{d_j\,|(n+h_j)\,\forall j\\ d_1\cdots d_K\le\sqrt{D}\\ (d_1\cdots d_K,W)=1\\ d_k=1}} \lambda_{d_1,\dots,d_K}\Bigg)^2\\
&= \sum_{\substack{d_j\,|(n+h_j)\,\forall j\\ d_1\cdots d_K\le\sqrt{D}\\ (d_1\cdots d_K,W)=1\\ d_k=1}} \lambda_{d_1,\dots,d_K} \sum_{\substack{e_j\,|(n+h_j)\,\forall j\\ e_1\cdots e_K\le\sqrt{D}\\ (e_1\cdots e_K,W)=1\\ e_k=1}} \lambda_{e_1,\dots,e_K} \sum_{\substack{x+h_k<p\le 2x+h_k\\ p\equiv h_k+b\,(W)\\ p\equiv h_k-h_j\,([d_j,e_j])}} 1.
\end{aligned}$$

The different divisibility conditions in the last sum may be combined by the Chinese remainder theorem into a single divisibility condition modulo $W[d_1,e_1]\cdots[d_K,e_K]\le WD$ to get

$$\Sigma_{h_k} = \sum_{\substack{d_j\,|(n+h_j)\,\forall j\\ d_1\cdots d_K\le\sqrt{D}\\ (d_1\cdots d_K,W)=1\\ d_k=1}} \lambda_{d_1,\dots,d_K} \sum_{\substack{e_j\,|(n+h_j)\,\forall j\\ e_1\cdots e_K\le\sqrt{D}\\ (e_1\cdots e_K,W)=1\\ e_k=1}} \lambda_{e_1,\dots,e_K} \sum_{\substack{x+h_k<p\le 2x+h_k\\ p\equiv a\,(W[d_1,e_1]\cdots[d_K,e_K])}} 1$$

for some residue class $a$ depending on $h_j$, $b$, $W$, $d_j$, and $e_j$. Thus, if $D\le x^{1/2-\epsilon}$, the sums $\Sigma_h$ can be evaluated by using (2.16) (after applying Cauchy-Schwarz to remove the weights $|\lambda_{d_1,\dots,d_K}|$ and $|\lambda_{e_1,\dots,e_K}|$ in the error term) to get

$$\Sigma_{h_k} \sim \frac{x}{\log x} \sum_{\substack{d_j\,|(n+h_j)\,\forall j\\ d_1\cdots d_K\le\sqrt{D}\\ (d_1\cdots d_K,W)=1\\ d_k=1}} \sum_{\substack{e_j\,|(n+h_j)\,\forall j\\ e_1\cdots e_K\le\sqrt{D}\\ (e_1\cdots e_K,W)=1\\ e_k=1}} \frac{\lambda_{d_1,\dots,d_K}\lambda_{e_1,\dots,e_K}}{\varphi(W[d_1,e_1]\cdots[d_K,e_K])}.$$

The sum here can be evaluated explicitly in terms of $F(d_1, \dots, d_K)$.

The sum $S_1$ in (2.13) can like-wise be evaluated by expanding the square and simply computing the sum over $n$. The end result is essentially that for all $h \in \mathcal{H}$ we have $\Sigma_h \gg S_1$ [48, Proposition 4.1], so that by taking $K$ large enough we obtain (2.12) for some constant depending on the choice of $F$.

We will not discuss here the optimization of the smooth weight $F(d_1, \dots, d_K)$ since we can take it for granted in our article [II].

## 2.3 The linear sieve

In [II] we will use Chen's sieve and the Maynard-Tao sieve in tandem. For Chen's sieve we need to recall the well-known linear sieve (see for example [23, Chapter 12] or [29, Chapter 4]). Recall that in Harman's sieve method we required both Type I and Type II information. We can also ask what sort of estimates are available if only Type I information is given. The linear sieve answers this question. Historically this happened the other way around, that is, the linear sieve of Iwaniec and Rosser was developed first and Harman's sieve method extends this (see [29, Chapter 3.8]).

A well-known phenomenon known as the parity problem asserts that we cannot obtain non-trivial lower bounds for sums over primes (2.1) if we only have Type I information (2.2). In short, even with optimal Type I information ($D = x^{1-\epsilon}$) we cannot distinguish between numbers with even number of prime factors from those with odd number of prime factors. We refer to [23, Chapter 16] for a precise description of the parity phenomenon obtained by Bombieri. Note that in Harman's sieve method it is the Type II information which crucially allows us to break the parity barricade. However, good upper bounds are available which require only Type I information, and for sums of almost-primes we can produce also lower bounds.

More generally, we can ask for lower and upper bounds for the sum $S(\mathcal{A}, z)$ provided that we have Type I information (2.2) up to some level $D$. This amounts to optimizing a pair of sieve weights $\lambda_d^+$ and $\lambda_d^-$ such that for all $n \geq 1$

$$\sum_{\substack{d \mid (n, P(z)) \\ d \leq D}} \lambda_d^- \leq \mathbf{1}_{(n,P(z))=1} = \sum_{d \mid (n,P(z))} \mu(d) \leq \sum_{\substack{d \mid (n, P(z)) \\ d \leq D}} \lambda_d^+ .$$

In practice $\lambda_d^\pm$ will be $\mu(d)$ with support restricted to some well-chosen sets $\mathcal{D}^\pm$. Unlike with Harman's sieve, with minimal assumptions this optimization problem can be solved exactly.

More precisely, we assume that Type I information is in the form

$$\sum_{d \leq D} \sum_{n \sim x/d} \alpha(d) a_{dn} = X \sum_{d \leq D} \alpha(d) g(d) + O(X \log^{-C} x) \qquad (2.17)$$

for some quantity $X = X(\mathcal{A}, x)$ and for some fixed multiplicative function $g(d)$ satisfying $0 \leq g(p) < 1$ for all primes $p$. Furthermore, in the linear sieve we assume that for some constant $L > 0$ we have for all $2 \leq w < z$

$$\prod_{w \leq p < z} (1 - g(p))^{-1} \leq \frac{\log z}{\log w}\left(1 + \frac{L}{\log w}\right). \tag{2.18}$$

This assumption simply means that the sieve dimension is at most 1, see [23, Chapter 11] for the more general beta-sieve.

To state the bounds we need to define the linear sieve functions (see [23, Chapter 12], for instance). Let $F_{\text{lin}}(s), f_{\text{lin}}(s)$ be the continuous solution to the system of delay-differential equations

$$\begin{cases} (sF_{\text{lin}}(s))' = f_{\text{lin}}(s-1) \\ (sf_{\text{lin}}(s))' = F_{\text{lin}}(s-1) \end{cases}$$

with the initial condition

$$\begin{cases} sF_{\text{lin}}(s) = 2e^{\gamma}, & \text{if } 1 \leq s \leq 3 \\ sf_{\text{lin}}(s) = 0, & \text{if } s \leq 2. \end{cases}.$$

Here $\gamma$ is the Euler-Mascheroni constant. Note that for $2 \leq s \leq 4$

$$f_{\text{lin}}(s) = \frac{2e^{\gamma}\log(s-1)}{s}.$$

Suppose then that $z = D^{1/s}$ for some $s \geq 1$, and that (2.17) and (2.18) hold. Then the linear sieve bounds are [23, Chapters 11 and 12]

$$S(\mathcal{A}, z) \leq (F_{\text{lin}}(s) + o(1) + \mathcal{O}(\log^{-1/6} D))X \prod_{p \leq z}(1 - g(p)) \qquad \text{and} \tag{2.19}$$

$$S(\mathcal{A}, z) \geq (f_{\text{lin}}(s) - o(1) - \mathcal{O}(\log^{-1/6} D))X \prod_{p \leq z}(1 - g(p)) \tag{2.20}$$

In general these bounds are optimal, that is, given (2.17) up to level $D$ we cannot hope for better bounds [23, Chapter 12.3]. For example, using the upper bound sieve with $D = z = x^{1/2-\epsilon}$ for twin primes (that is, $a_n = \mathbf{1}_{\mathbb{P}}(n+2)$), with Type I information provided by (2.16), we get

$$\pi_2(x) := \sum_{p \leq x} \mathbf{1}_{\mathbb{P}}(n+2) \leq (4 + o(1)) \cdot 2 \prod_{p \geq 3}\left(1 - \frac{1}{(p-1)^2}\right)\frac{x}{\log^2 x}, \tag{2.21}$$

which is larger than the expected count by a factor of 4 (this bound can also be obtained using Selberg's sieve).

## 2.4 Chen's Sieve

The linear sieve is neutral with respect to Buchstab's identity (2.4). In fact, one way to think about the linear sieve is as the limit of iterating Buchstab's identity on some initial sieve weights [23, Chapter 11.1]. In Section 2.1 we saw that if Type II information is available then the linear sieve bound can be improved. In certain applications we can use Chen's sieve which is based on the 'switching trick' to improve on the linear sieve bounds. Chen [13] famously used this to prove that there are infinitely many primes $p$ such that $p+2$ is the product of at most two prime factors. The reason we are unable to solve the full twin prime conjecture is due to the parity problem.

In our application [II] it turns out that the bound (2.21) is just barely not good enough. Indeed, a bound of this type with the constant 4 replaced by $4-\epsilon$ for any $\epsilon > 0$ would suffice. We now sketch how to apply Chen's sieve to get 3.99 in place of 4. The current best constant here is 3.3996 by Wu [61, Theorem 3], which relies on many further developments. In the actual application we will have to combine Chen's sieve with the Maynard-Tao sieve weights, but this is of little additional effort.

We begin with the combinatorial bound

$$\begin{aligned}\mathbf{1}_{(n,P(Z))=1} \leq \mathbf{1}_{(n,P(Y))=1} &- \frac{1}{2}\sum_{Y \leq p < Z} \mathbf{1}_{p|n}\mathbf{1}_{(n,P(Y))=1} \\ &+ \frac{1}{2}\sum_{Y\leq p<q<r<Z}\sum_{(s,P(q))=1} \mathbf{1}_{n=pqrs},\end{aligned} \tag{2.22}$$

which holds for any $Y = x^{\alpha}$, $Z = x^{\beta}$, $0 < \alpha < \beta < 1/4$ and $n \in [x, 2x]$. For $(n, P(Y)) > 1$ this is obvious as all terms vanish, so let $(n, P(Y)) = 1$ and denote $k = \sum_{Y \leq p < Z} \mathbf{1}_{p|n}$. If $k = 0$, then both sides of (2.22) are equal to one. For $k \geq 1$ the left-hand side is zero. If $k = 1$, then the right-hand side is $1-1/2+0 = 1/2 > 0$. For $k \geq 2$ the right-hand side is $1 - k/2 + (k-2)/2 = 0$, since in the last sum $p$ and $q$ are fixed and there are $k-2$ ways to choose $r$, which proves the claim.

The fact that (2.22) is sufficient for the job was first noted by Pan [50] (according to Wu [61]). Using the upper bound (2.22) we get

$$\pi_2(2x) - \pi_2(x) \leq \sum_{\substack{x<n\leq 2x \\ (n,P(Z)=1)}} \mathbf{1}_{\mathbb{P}}(n+2) \leq S_1 - S_2/2 + S_3/2$$

where for $a_n := \mathbf{1}_{\mathbb{P}}(n+2)$ we have

$$S_1 := \sum_{x<n\leq 2x} a_n \mathbf{1}_{(n,P(Y))=1}, \qquad S_2 := \sum_{Y\leq p<Z}\sum_{\substack{x<n\leq 2x \\ p|n}} a_n \mathbf{1}_{(n,P(Y))=1}, \qquad \text{and}$$

$$S_3 := \sum_{x<n\leq 2x} a_n \sum_{Y\leq p<q<r<Z}\sum_{(s,P(q))=1} \mathbf{1}_{n=pqrs}.$$

Let

$$\mathbf{M}(x) := 2 \prod_{p \geq 3} \left(1 - \frac{1}{(p-1)^2}\right) \frac{x}{\log^2 x}$$

denote the expected main term. The assumptions (2.17) and (2.18) in the linear sieve are easily verified with $g(d) = 1/\varphi(d)$. For $S_1$ we have the level of distribution $D = x^{1/2-\epsilon}$ by the Bombieri-Vinogradov theorem (2.16) and we get by applying (2.19)

$$S_1 \leq \frac{F_{lin}(1/(2\alpha)) + O(\epsilon)}{\alpha e^\gamma} \mathbf{M}(x). \tag{2.23}$$

For $S_2$ we apply the linear sieve lower bound (2.20) with the level of distribution $D/p$ to get (denoting $p = x^t$)

$$S_2 \geq \frac{1 - O(\epsilon)}{\alpha e^\gamma} \int_\alpha^\beta f_{lin}\left(\frac{1/2 - t}{\alpha}\right) \frac{dt}{t} \mathbf{M}(x). \tag{2.24}$$

The key idea appears in the upper bound for $S_3$. If we simply applied the linear sieve upper bound to the variable $s$ with $(s, P(q)) = 1$, we cannot beat the original bound (2.21). Recall that $a_n = \mathbf{1}_{\mathbb{P}}(n+2)$, which means that we can instead apply the linear sieve to the variable $p' = (n+2)$, once we write

$$S_3 := \sum_{x+2 < p' \leq 2x+2} a'_{p'} \qquad \text{for} \qquad a'_n = \sum_{Y \leq p < q < r < Z} \sum_{(s, P(q)) = 1} \mathbf{1}_{n = pqrs+2}.$$

This is often referred to as the switching trick. Note that the Bombieri-Vinogradov theorem (2.16) holds also for almost-primes such as $pqrs$ in the above, so that for the new sequence $a'_n$ we get the same level of distribution $D = x^{1/2-\epsilon}$. Applying the linear sieve upper bound (2.19) with $z = D$ (or Selberg's sieve) we get by (2.9) (denoting $p = x^{u_1}, q = x^{u_2}, r = x^{u_3}$)

$$S_3 \leq (4 + O(\epsilon)) \int_{\alpha < u_1 < u_2 < u_3 < \beta} \omega\left(\frac{1 - u_1 - u_2 - u_3}{u_2}\right) \frac{du_1 du_2 du_3}{u_1 u_2^2 u_3} \mathbf{M}(x), \tag{2.25}$$

where $\omega$ is the Buchtab function, so that the upper bound for $S_3$ is off by a factor of 4.

Combining the bounds (2.23), (2.24), and (2.25), choosing $\alpha = 1/7$ and $\beta = 3/14$ (which are not necessarily optimal), and computing the integrals numerically we get

$$\begin{aligned} \pi_2(2x) - \pi_2(x) &\leq S_1 - S_2/2 + S_3/2 \\ &\leq (4.19 - 0.279 + 0.076 + o(1)) \mathbf{M}(x) \\ &< (3.99 + o(1)) \mathbf{M}(x), \end{aligned}$$

as claimed.

As mentioned earlier, Chen used the switching trick to prove that there are infinitely many primes $p$ such that $\Omega(p+2) \leq 2$. This can be shown by a similar argument as above, where instead of (2.22) we use

$$\mathbf{1}_{\Omega_2(n)\leq 2} \geq 1 - \frac{1}{2}\sum_{p\leq x^{1/3}} \mathbf{1}_{p|n} - \frac{1}{2}\sum_{p_1\leq x^{1/3}<p_2<p_3} \mathbf{1}_{n=p_1p_2p_3} - \frac{1}{2}\sum_{p\leq x^{1/3}} \mathbf{1}_{p^2|n},$$

which holds for $x^{2/3} < n \leq x$. Here we multiply both sides by $\mathbf{1}_{(n,P(z))=1}$ with $z = x^{1/8}$, bound the fourth term trivially (using $p \geq z$), and use switching for the third term to get a correct-order lower bound

$$\sum_{p\leq x} \mathbf{1}_{\Omega(p+2)\leq 2}\mathbf{1}_{(p+2,P(z))=1} \gg \frac{x}{\log^2 x}.$$

We refer to [57] for a detailed version of this argument.

## 2.5 Sieves and arithmetic information

In all of the four articles in this thesis the argument breaks down into two parts that are in some sense orthogonal — the sieve and the arithmetic information. Personally, I see sieves as combinatorial machines that take arithmetic information about the given sequence as inputs, and produce upper and lower bounds for primes or almost-primes.

The quality of the arithmetic information is measured by the level of distribution (size of $D$ in (2.2)) as well as the location and the width of the Type II information (ranges of $M$ and $N$ in (2.3)). An important point to remember is that sieve bounds are continuous with respect to the quality of the information, be it Harman's sieve method or the linear sieve.

We have now given an account of the different sieves that appear in this thesis, and in this respect the four articles overlap significantly. However, the heart of the matter is in obtaining the arithmetic information (apart from [II] which is really all about the sieves), and in this aspect each of the articles apply very different means. In the following sections we will describe the main ideas of each article separately, with occasional comments on some differences in the sieve arguments.

# 3 Large prime factors on short intervals

In the first article [I], motivated by the problem of prime numbers in short intervals, we ask what is the smallest $\gamma$ such that for all sufficiently large $x$ the interval $[x, x+y]$ contains a number with a prime factor $p > x^{1-\gamma}$. This problem has received much attention in the cases $y = x^{1/2+\epsilon}$ for an arbitrarily small but fixed $\epsilon > 0$ and $y = x^{1/2}$.

For $y = x^{1/2+\epsilon}$ the current best result is the exponent $\gamma = 1/26 = 0.0384\ldots$ by Jia and Liu [41], which improved on the previous results [7, 8, 9, 10, 31, 34, 42]. In comparison, with $y = x^{1/2}$ the best result is $\gamma = 0.2572$ by Baker and Harman [5], which illustrates that there is a large jump in degree of difficulty around $x^{1/2}$. The methods used differ greatly between the two cases.

The main purpose of the article [I] is to try to bridge the gap between the above-mentioned results as much as possible. More precisely, we try to reduce $y$ closer to $x^{1/2}$ from $y = x^{1/2+\epsilon}$ by combining the methods in [41] with the argument of Matomäki and Radziwiłł [47]. This problem was suggested to me by my supervisor Prof. Kaisa Matomäki when I first started my doctoral studies, with the purpose of me getting some hands-on experience with Harman's sieve method as well as with the new developments in Dirichlet polynomial methods by Matomäki and Radziwiłł [47].

To explain the main idea in the article, it is useful to compare the problem with $x^\delta$-smooth numbers on short intervals. Recall that a number is $z$-smooth if none of its prime factors exceed $z$. A folklore conjecture states that for any $\delta > 0$ for all large enough $x > x(\delta)$ the intervals $[x, x + x^{1/2}]$ contain an $x^\delta$-smooth number. Prior to [47] the best results towards this was by Matomäki [46] with intervals of length $x^{1/2} \log^{7/3+\epsilon} x$. Matomäki and Radziwiłł [47, Corollary 1] succeeded to prove that for some sufficiently large constant $C(\delta)$ the intervals $[x, x + C(\delta)x^{1/2}]$ contain $x^\delta$-smooth numbers. This is one of the many consequences of their general theorem on multiplicative functions.

Due to some technical but in retrospect quite natural obstacles, I was not able to handle intervals of order $Cx^{1/2}$ for the problem of the largest prime factor on short intervals. The main result in [I] has a logarithmic factor in place of $C$, which is still much smaller than the factor $x^\epsilon$ in [41] but with a quantitatively similar exponent $\gamma$.

**Theorem 1.** *For all sufficiently large $x$ the interval $[x, x + x^{1/2}\log^{1.39} x]$ contains a number with a prime factor $p > x^{1-\gamma}$ for $\gamma = 1/19$.*

The exponent $\gamma$ is determined by optimizing Harman's sieve method. The logarithmic factor is required because we have to work with a sparse sequence due to a certain very strong factorization property, as we will explain below.

Compared to our problem, the question of smooth numbers in short intervals is simpler in the crucial aspect that it can be re-formulated in a bilinear form. More precisely, if $f(n)$ denotes the characteristic function of $x^\delta$-smooth numbers, then we simply need a non-trivial lower bound for the sum

$$\sum_{\substack{x \leq n_1 n_2 \leq x + C(\delta)\sqrt{x} \\ \sqrt{x} \leq n_1 \leq 2\sqrt{x}}} f(n_1) f(n_2),$$

which follows from [47, Theorem 2]. It is vital that both of the variables $n_1$ and $n_2$ are very closely of the same size (within a multiplicative constant).

In the case of the largest prime factor this is not possible directly, since we require a lower bound for

$$\sum_{\substack{x \leq pk \leq x+y \\ p \sim x^{1-\gamma}}} 1 \tag{3.1}$$

with $\gamma > 0$ as small as possible. However, we can apply Harman's sieve method (see Section 2.1) to produce a bilinear sum (Type II information)

$$\sum_{\substack{x \leq mnk \leq x+y \\ m \sim M \\ n \sim N}} \alpha(m)\beta(n)$$

with $MN = x^{1-\gamma}$ and $x^{1/2-\gamma} \ll M, N \ll x^{1/2}$. Unfortunately, we still do not quite get a suitable bilinear sum since for intervals of the type $[x, x + Cx^{1/2}]$ we need two variables $n_1, n_2$ which are both within a constant from $\sqrt{x}$.

To solve this issue we rely on a key idea of Jutila [42]. Since we only want a lower bound, we may request that the free variable $k$ in (3.1) is a product of many small prime factors, say, $k = p_1 \cdots p_K$. We can then combine the variables

$$n_1 := mp_1 \cdots p_\ell \qquad \text{and} \qquad n_2 := np_{\ell+1} \cdots p_K$$

to try to guarantee that $n_1, n_2 \approx \sqrt{x}$. In the next section we recall how this argument proceeds for intervals of length $y = x^{1/2+\epsilon}$. In the subsequent section we recall the Matomäki-Radziwiłł argument, and in the last section we explain how to combine these two. Recall that for the sieve we also need the corresponding Type I information, but since this is much simpler to handle we focus only on the Type II sums here.

*Remark* 1. The exponent 1.39 in Theorem 1 is determined by optimizing the construction described in Section 3.3. More precisely, let $\beta := 1.388\dots$ denote the minimum of the function

$$r \mapsto \frac{\log(1-\log(r-1)) - \log(-\log(r-1)) + \log 2}{2\log r} - \frac{1}{2}$$

for $1 < r < 2$, which is obtained at $r := 1.625\dots$ Then the exponent 1.39 in Theorem 1 can be replaced by $\beta + \delta$ for any $\delta > 0$.

## 3.1 Intervals of length $y = x^{1/2+\epsilon}$

In this section we explain how the Type II sums are handled for intervals of length $y := x^{1/2+\epsilon}$ in [34] and [41] (see also [29, Chapter 5]). As mentioned above, we restrict the free variable $k$ in (3.1) to be a product of $K$ small primes. More precisely, for the given $\gamma$ and $\epsilon > 0$, let $K = \lceil 4/\epsilon \rceil$ and $P := x^{\gamma/K}$. We then wish to show a non-trivial lower bound for

$$\sum_{\substack{x < pp_1\cdots p_K \leq x+y \\ p_j \sim P}} 1.$$

By using Harman's sieve method the main issue is to evaluate the Type II sums to get

$$\sum_{\substack{x < mnp_1\cdots p_K \leq x+y \\ m\sim M,\, n\sim N \\ p_j \sim P}} \alpha(m)\beta(n) = \frac{y}{x} \sum_{\substack{x < mnp_1\cdots p_K \leq 2x \\ m\sim M,\, n\sim N \\ p_j \sim P}} \alpha(m)\beta(n) + O(y/\log^C x),$$

where $M, N \in [x^{1/2-\gamma}, x^{1/2}]$ with $MN = x^{1-\gamma}$ and $\alpha, \beta$ are bounded coefficients. By Perron's formula (see [29, Lemma 1.1], for example) the left-hand side equals (for any $T \gg 1$)

$$\frac{1}{2\pi i}\int_{1-iT}^{1+iT} A(s)B(s)P(s)^K \frac{(x+y)^s - x^s}{s} ds + O_\eta\left(\frac{x^{1+\eta}}{T}\right),$$

where

$$A(s) = \sum_{m\sim UM} \alpha(m) m^{-s}, \quad B(s) = \sum_{n\sim N} \beta(n) n^{-s}, \quad \text{and} \quad P(s) = \sum_{p\sim P} p^{-s}.$$

We get the correct main term from the small frequencies $t$, and the error term is morally bounded by

$$yI := y\int_{T_0}^{x/y} |A(1+it)B(1+it)P(1+it)^K| dt, \tag{3.2}$$

where $T_0 = \log^{100C} x$ with $C$ some sufficiently large constant (see [29, Section 5.2] for details). Therefore, we need to show $I \ll \log^{-C} x$.

A well-known bound is the Mean value theorem for Dirichlet polynomials, which states that for any complex coefficients $c_n$ (see [39, Chapter 9], for instance)

$$\int_0^T \Big| \sum_{n \leq N} c_n n^{it} \Big|^2 dt \ll (T+N) \sum_{n \leq N} |c_n|^2. \tag{3.3}$$

By Vinogradov's zero free region we have for $\log^{100C} x \leq t \leq x/y$

$$P(1+it) \ll \log^{-2C} x. \tag{3.4}$$

Hence, after pulling out one factor of $P(1+it)$, in order to show that $I \ll \log^{-C} x$ we need

$$\int_0^{x/y} |A(1+it)B(1+it)P(1+it)^{K-1}| dt \ll \log^C x.$$

The key point is that since $K \geq 4/\epsilon$ and $M, N \geq x^{1/2-\gamma}$, there exists some $L \leq K-1$ such that

$$MP^L,\ NP^{K-L-1} \gg x^{1/2-\epsilon}. \tag{3.5}$$

Hence, by Cauchy-Schwarz and by (3.3) we get

$$\begin{aligned}
&\int_0^{x^{1/2-\epsilon}} |A(1+it)B(1+it)P(1+it)^{K-1}|\, dt \\
&\leq \left( \int_{T_0}^{x^{1/2-\epsilon}} |A(1+it)P(1+it)^L|^2\, dt \right)^{1/2} \left( \int_{T_0}^{x^{1/2-\epsilon}} |B(1+it)P(1+it)^{K-L-1}|^2\, dt \right)^{1/2} \\
&\ll \left( \frac{x^{1/2-\epsilon}}{MP^L} + 1 \right)^{1/2} \left( \frac{x^{1/2-\epsilon}}{NP^{K-L-1}} + 1 \right)^{1/2} \log^C x \ll \log^C x,
\end{aligned} \tag{3.6}$$

which is sufficient.

Note that for intervals of length $y = hx^{1/2}$ we need to have (3.5) with $x^{1/2}/h$ in place of $x^{1/2-\epsilon}$. That is, the shorter the interval the more precise the factorization (3.5) has to be, and in our application we essentially need both factors to be within a factor of $\log^\delta x$ of $x^{1/2}$ for some small $\delta > 0$. We will explain how to achieve this in Section 3.3. However, the high level of accuracy required means that in Theorem 1 we have to include the factor of $\log^{1.39} x$. Note that the shorter the interval $y$ is the longer the integration range in $I$ is. To handle longer integrals, we recall the argument of Matomäki and Radziwiłł in the next section.

## 3.2 The Matomäki-Radziwiłł argument

There is another problem with the argument in the previous section if we wish to reduce the length of the interval from $x^{1/2+\epsilon}$ much closer to $x^{1/2}$, for example, to $y = x^{1/2}\log^{10\delta} x$ for any small $\delta > 0$. Namely, when we pull out one factor of $P(1+it)$ for which we use the Vinogradov bound (3.4), the length of the remaining Dirichlet polynomial is diminished too much compared to the length $x/y$ of the integration, so that we cannot obtain the factorization (3.5) with $x^{1/2-\epsilon}$ replaced by $x/y$. If we try to modify the set-up so that we have a shorter polynomial factor, say, $\sum_{p\sim P'} p^{-s}$ with $P' = \log^{\delta} x$, then the problem is that we do not know a corresponding point-wise upper bound (3.4) for such a short polynomial.

Matomäki and Radziwiłł [47] overcome this issue by iterating the argument. To illustrate the main principles, suppose that we want a bound of the form

$$I := \int_{T_0}^{x^{1/2}\log^{-10\delta} x} |F(1+it)|^2\, dt \ll \log^{-\delta} x \tag{3.7}$$

where $F(s)$ is a Dirichlet polynomial of length $\asymp x^{1/2}$ and of the shape

$$F(s) := \sum_{\substack{j \\ 2^j \in [Q_3, Q_3^{1+\eta}]}} Q_1(s)Q_2(s)Q_{3,j}(s)A_j(s)$$

for some ranges $Q_1$, $Q_2$, and $Q_3$ with

$$Q_1(s) := \sum_{q_1\sim Q_1} q_1^{-s}, \qquad Q_2(s) := \sum_{q_2\sim Q_2} q_2^{-s}, \qquad Q_{3,j}(s) := \sum_{q_3\sim 2^j} q_3^{-s},$$
$$\text{and} \qquad A_j(s) := \sum_{m\sim x^{1/2}/(Q_1Q_2 2^j)} \alpha(m)m^{-s},$$

where $\alpha(m)$ are some bounded coefficients and $q_j$ denote prime numbers. In our problem there is flexibility to choose the ranges $Q_j$, and we pick

$$Q_1 := \log^{10\delta} x, \qquad Q_2 := Q_1^H, \quad H := \left\lceil \frac{(\log\log x)^{1/2}}{10\delta} \right\rceil, \qquad \text{and} \tag{3.8}$$

$$Q_3 := \exp(2\lfloor \log^{9/10} x \rfloor). \tag{3.9}$$

Here $Q_3$ is large enough so that (3.4) applies. Also crucial here is that $Q_2$ is large but not too large compared to $Q_1$. In comparison to the previous section, we now need the primes $q_3 \in [Q_3, Q_3^{1+\eta}]$ to vary over longer than dyadic intervals so that the numbers $q_1q_2q_3m$ in $F(s)$ represent a proportion $\log^{-o(1)} x$ of numbers near $x^{1/2}$, so that (3.7) is better than the 'trivial bound' (3.3) by $\log^{-\delta+o(1)} x$. We will denote

$$B(s) := \sum_{\substack{j \\ 2^j \in [Q_3, Q_3^{1+\eta}]}} Q_{3,j}(s)A_j(s),$$

so that $F(s) = Q_1(s)Q_2(s)B(s)$

To bound $I$ we begin by splitting the integration into three parts $\mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{U}$, where

$$\mathcal{T}_1 := \{t : |Q_1(1+it)| \leq Q_1^{-1/4+2\epsilon}\},$$
$$\mathcal{T}_2 := \{t : |Q_1(1+it)| > Q_1^{-1/4+2\epsilon}, |Q_2(1+it)| \leq Q_2^{-1/4+\epsilon}\},$$

and $\mathcal{U} := [T_0, x\log^{-10\delta} x] \setminus (\mathcal{T}_1 \cup \mathcal{T}_2)$.

**Integral over** $\mathcal{T}_1$. We simply pull out the factor $|Q_1(1+it)|^2 \leq (Q_1^{-1/4+2\epsilon})^2 \leq \log^{-\delta} x$ to get

$$\begin{aligned}\int_{\mathcal{T}_1} &|Q_1(1+it)Q_2(1+it)B(1+it)|^2\, dt \\ &\leq (\log^{-\delta} x) \int_0^{x^{1/2}\log^{-10\delta} x} |Q_2(1+it)B(1+it)|^2\, dt \ll \log^{-\delta} x \end{aligned} \tag{3.10}$$

by the Mean value theorem (3.3).

**Integral over** $\mathcal{T}_2$. This is where the key idea occurs. If we simply pull $|Q_2(1+it)|^2$ in front, then the remaining polynomial is too short for an application of (3.3). However, we can now use the lower bound $|Q_1(1+it)| > Q_1^{-1/4+2\epsilon}$ to restore the length of the polynomial. That is, using $1 \leq (|Q_1(1+it)|Q_1^{1/4-2\epsilon})^{2H}$ and $|Q_2(1+it)| \leq Q_2^{-1/4+\epsilon}$ we get

$$\begin{aligned}\int_{\mathcal{T}_2} &|Q_1(1+it)Q_2(1+it)B(1+it)|^2\, dt \\ &\leq (Q_2^{-1/4+\epsilon})^2 (Q_1^{1/4-2\epsilon})^{2H} \int_0^{x^{1/2}\log^{-10\delta} x} |Q_1(1+it)^{H+1}B(1+it)|^2\, dt \ll \log^{-\delta} x\end{aligned}$$

by (3.3). It is important that $H$ is not too big so that we do not get too many collisions from the mean square of the coefficients of $Q_1(1+it)^H$.

**Integral over** $\mathcal{U}$. We use the triangle inequality and Cauchy-Schwarz to the $j$-summation to get

$$\begin{aligned}\int_{\mathcal{U}} &|F(1+it)|^2 dt \\ &\ll (\log^2 x) \max_{\substack{j \\ 2^j \in [Q_3, Q_3^{1+\eta}]}} \int_{\mathcal{U}} |Q_1(1+it)Q_2(1+it)Q_{3,j}(1+it)A_j(1+it)|^2\, dt,\end{aligned}$$

and bound the contribution from each $j$ individually. We say that a finite set of points $\mathcal{T}$ is well-spaced if for all distinct $t_1, t_2 \in \mathcal{T}$ we have $|t_1 - t_2| \geq 1$. We may pick a well-spaced set $\mathcal{T} \subseteq \mathcal{U}$ such that

$$\begin{aligned}\int_{\mathcal{U}} &|Q_1(1+it)Q_2(1+it)Q_{3,j}(1+it)A_j(1+it)|^2\,dt \\ &\ll \sum_{t\in\mathcal{T}} |Q_1(1+it)Q_2(1+it)Q_{3,j}(1+it)A_j(1+it)|^2 \\ &\ll (\log^{-C} x)\sum_{t\in\mathcal{T}} |Q_1(1+it)Q_2(1+it)A_j(1+it)|^2\end{aligned}$$

by using (3.4) on $Q_{3,j}(1+it)$. Here the polynomial $Q_2(s)$ is sufficiently long so that we may use large-value results (see [47, Lemma 8]) with $|Q_2(1+it)| \geq Q_2^{-1/4+\epsilon}$ to show that $|\mathcal{T}| \ll T^{1/2-\epsilon}$. The desired bound then follows from the Halász-Motgomery inequality (see [39, Theorem 9.6], for instance), which states that for any coefficients $c_n$

$$\sum_{t\in\mathcal{T}}\Big|\sum_{n\leq N} c_n n^{-it}\Big|^2 \ll (N + T^{1/2}|\mathcal{T}|)\sum_{n\leq N}|c_n|^2,$$

completing the proof of $I \ll \log^{-\delta} x$.

To handle even shorter intervals $y$ one can iterate the above argument (each iteration diminishes $y/x^{1/2}$ morally by one iteration of log). In our case we are able handle only intervals of type $y = x^{1/2}\log^C x$ for some $C > 0$ due to reasons explained in the next section. That is, we are interested in the length only up to a factor of $\log^{\delta} x$, so the above argument suffices for us.

## 3.3 A factorization problem

In this section we try to combine the arguments of the two previous sections and see why in Theorem 1 we need the power of logarithm $\log^{1.39} x$ for the proof to succeed. The main problem here is the well-known ineffectiveness of Fourier analytic techniques (Dirichlet polynomial mean values) to handle sparse sets. To demonstrate this, consider a sequence $c_n$, $n \sim N$, which is the characteristic function of some well-behaved set of density $\rho$ around $N$. Then we expect

$$\frac{1}{N}\sum_{\substack{mn\in[N^2,N^2+N]\\ m,n\sim N}} c_m c_n \asymp \rho^2.$$

However, by Perron's formula this sum is related to the following mean value, and the bound (3.3) for this is

$$\int_0^N \Big|\sum_{n\sim N} c_n n^{-1-it}\Big|^2\,dt \ll \frac{1}{N}\sum_{n\sim N}|c_n|^2 = \rho.$$

That is, we have already lost a square root of the density. This is because the diagonal term in the Mean value theorem (3.3) corresponds to square root cancellation on average over $t$.

To get a sufficient factorization property such as in Section 3.1 but for shorter intervals we need to work with a set of density $\log^{-A} x$ for some $A > 0$. To gain back the loss, we essentially need to work with intervals of length $x^{1/2}\log^{A/2} x$. This is because by using the Matomäki-Radziwiłł argument we can win only a factor of $\log^{-\delta} x$ for some small $\delta > 0$. Optimizing the argument leads to the exponent 1.39 in Theorem 1.

We now explain the basic set-up in [I], ignoring some technical details. For the factorization property we need small prime factors $p_j \in I_j$ for some suitable intervals $I_j$. Let $r := 1.625\ldots$ and $\beta := 1.388\ldots$ as in Remark 1, let $\theta := r-1$, define $\omega := x^{\gamma(r-1)}$, and for some large constant $K$ set

$$I_j := \begin{cases} (\omega^{(1-\epsilon)r^{-j}}, \omega^{r^{-j}}], & j = 1,2,\ldots,K \\ (\omega^{\theta r^{-j}}, \omega^{r^{-j}}], & j = K+1,\ldots,J, \end{cases}$$

where

$$J := \left\lceil \frac{1}{\log r} \log\left(\frac{\log \omega}{\log K}\right)\right\rceil.$$

Note that $I_j$ are disjoint since $r > (1+\sqrt{5})/2$, and we have $\omega^{r^{-J}} \in [K^{1/r}, K]$. Remark also that if $p_j \in I_j$, then

$$p_1 \cdots p_J = x^{O(\epsilon+1/K)} x^{\gamma(r-1)(r^{-1}+r^{-2}+\cdots)+o(1)} = x^{\gamma+o(1)}$$

if $\epsilon$ and $1/K$ are very small. Because the number of prime factors required is $J \sim (\log r)^{-1}\log\log x$, we lose a power of $\log x$ in the density, and $r$ is chosen to minimize this loss. Morally, restricting one prime factor $p_j \in I_j$ costs us some constant $\kappa$ in the density, so that restricting $J$ prime factors costs us $\kappa^J = (\log x)^{\log\kappa/\log r + o(1)}$.

To keep the density as large as possible we need $I_j$ to be longer than dyadic intervals. To bump up the density we also choose $I := [1, x^\epsilon]$ and let $c, c' \in I$ run freely, so that for Theorem 1 we need a lower bound for

$$\sum_{x \le pcc' q_1 q_2 q_3 p_1 \cdots p_J \le x+y} 1$$

with $y = x\log^{\beta+10\delta} x$ for any $\delta > 0$. In the summation the variables are restricted to $c, c' \in I$, $p_j \in I_j$, and the primes $q_j$ are as in Section 3.2 (we could also use the primes $p_j$ in the Matomäki-Radziwiłł argument, but it is simpler to include the $q_j$ separately). Note that the free prime variable is $p = x^{1-\gamma+o(1)}$ since we can make $\epsilon$ arbitrarily small. The set-up in [I] is slightly more complex, but these are the main components.

Similarly as in Section 3.1, the error term in the Type II sums is bounded in terms of

$$\int_{T_0}^{x/y} |F(1+it)|dt$$

for

$$F(s) := \sum_{mncc'q_1q_2q_3p_1\cdots p_J \sim x} \alpha(m)\beta(n)(mncc'q_1q_2q_3p_1\cdots p_J)^{-s}$$

with the support of $\alpha$ and $\beta$ restricted to $m, n \in [x^{1/2-\gamma+\epsilon}, x^{1/2-\epsilon}]$. By a far more difficult factoring algorithm (see [I, Lemmata 17 and 18]) we are able to bound this by

$$\sum_{\pi \sqcup \tau = \{1,2,\dots,J\}} \int_{T_0}^{x/y} |F_\pi(1+it)G_\tau(1+it)|dt$$
$$\leq \sum_{\pi \sqcup \tau = \{1,2,\dots,J\}} \left( \int_{T_0}^{x/y} |F_\pi(1+it)|^2 dt \right)^{1/2} \left( \int_{T_0}^{x/y} |G_\tau(1+it)|^2 dt \right)^{1/2}, \quad (3.11)$$

where $F_\pi(s)$ includes the primes $p_j$ with $j \in \pi$, the variables $m$ and $c$, and the primes $q_1, q_2, q_3$, and $G_\tau(s)$ includes the primes $p_j$ with $j \in \tau$ and the variables $n$ and $c'$. We need to consider all partitions $\pi \sqcup \tau = \{1, 2, \dots, J\}$ since the primes $p_j$ are of different sizes, and the key point in [I, Lemma 17] is that this factorization is possible without amassing cross-conditions between primes $p_j$ with $j \in \pi$ and primes $p_j$ with $j \in \tau$. There are numerous conditions within each $F_\pi$ and $G_\tau$ separately, to guarantee that we have a unique choice of factorization

$$mncc'q_1q_2q_3p_1\cdots p_J = \Big(mcq_1q_2q_3 \prod_{j\in\pi} p_j\Big)\Big(nc' \prod_{j\in\tau} p_j\Big)$$

so that $F(s)$ partitions into $\sum_{\pi\sqcup\tau} F_\pi(s)G_\tau(s)$.

The integrals in (3.11) can now be bounded via the technique of Matomäki-Radziwiłł [47] similarly as in the previous section. To compute the power 1.39 of logarithm in Theorem 1 we need a version of the Mean value theorem (3.3) suitable for (3.11), which is obtained with great technical difficulty by [I, Lemma 20]. Finally, the exponent $\gamma = 1/19$ which determines the size of the large prime factor is decided by optimizing Harman's sieve method. Note that we have Type II information in the range $[x^{1/2-\gamma+\epsilon}, x^{1/2-\epsilon}]$ so that the quality of the arithmetic information is proportional to $\gamma$. Another layer of difficulty in [I] is imposed by the restriction that we cannot lose any powers of $\log x$ at any point of the argument. For example, when applying Perron's formula to remove an unwanted cross-condition we typically lose one factor of $\log x$. To overcome this issue we had to develop essentially a loss-free version of Harman's sieve method, and we refer to [I] for the details this.

# 4 Limit points of normalized prime gaps

By the Prime number theorem the average gap from a prime $p$ to the next is asymptotically $\log p$. We would like to understand the distribution of the prime gaps. For instance, the twin prime conjecture asks if the prime gap is 2 infinitely often. In the article [II] we study the set of real numbers $\alpha$ such that the prime gap is infinitely often $(\alpha + o(1)) \log p$.

The origin of the second article [II] is quite a typical story in research, in the sense that what I ended up proving is completely different from the original goal. I was studying an entirely different problem and was quite stuck. During a casual discussion with Prof. James Maynard at a conference in Bristol in 2018 he suggested that I should have a look at their paper [11] on the limit point problem, to see if that would be of any help (of which I am very grateful). After studying the article I tried unsuccessfully to solve another problem of similar nature, but by accident I discovered a trick which allowed me to improve on the best result at that time on limit points by Pintz [52].

To state the result precisely, let $\mathbb{L}$ denote the set of limit points of the sequence of normalized prime gaps $\{d_n\}_{n=1}^{\infty}$, where $d_n := (p_{n+1} - p_n)/\log p_n$. A conjecture of Erdős [21] states that $\mathbb{L} = [0, \infty]$. This is expected by Cramér's heuristic model (see [27], for instance), which suggests the much more precise asymptotic Poisson distribution for prime gaps

$$\frac{1}{N}\Big\{ n \leq N : \, (p_{n+1} - p_n)/\log p_n \in [a, b] \Big\} \sim \int_a^b e^{-u} \, du, \qquad N \to \infty. \tag{4.1}$$

Curiously, only two points are known to definitely belong to $\mathbb{L}$, namely 0 and $\infty$. That $\infty \in \mathbb{L}$ was proved by Westzynthius [60], while $0 \in \mathbb{L}$ is the famous result of Goldston, Pintz, and Yıldırım [26] obtained via their GPY sieve.

The Lebesque measure of $\mathbb{L}$ has been much studied [21, 25, 51, 56]. Note that $\mathbb{L}$ is a closed set and as such also a measurable set. Let $\mu$ denote the Lebesque measure on $\mathbb{R}$. Prior to [II] the best result was by Pintz [52], who showed that

$$\mu(\mathbb{L} \cap [0, T]) \geq (1/4 - o(1))T, \quad T \to \infty, \tag{4.2}$$

by refining the method of Banks, Freiberg, and Maynard [11]. This has been generalized to other normalizations of prime gaps by Baker and Freiberg [4]. The main result in [II] gives an improvement on the constant 1/4 as well as removes the $o(1)$-term.

**Theorem 2.** *For all $T > 0$ we have*

$$\mu(\mathbb{L} \cap [0,T]) \geq T/3.$$

This theorem is a corollary of [II, Theorem 1], which is as follows.

**Theorem 3.** *Let $\beta_1 \leq \beta_2 \leq \beta_3 \leq \beta_4$ be any real numbers. Then*

$$\mathbb{L} \cap \{\beta_j - \beta_i : 1 \leq i < j \leq 4\} \neq \emptyset.$$

A similar result was obtained by Pintz [52] with 4 replaced by 5, which allowed him to show (4.2) via a combinatorial argument as in [11]. To see how to deduce Theorem 2 form Theorem 3, suppose that $\kappa \geq 2$ is the smallest integer such that there exists a tuple $\beta_1 < \cdots < \beta_\kappa$ with

$$\mathbb{L} \cap \{\beta_j - \beta_i : 1 \leq i < j \leq \kappa\} = \emptyset, \tag{4.3}$$

and fix such $\beta_1 < \cdots < \beta_\kappa$. By Theorem 3 we see that $\kappa \leq 3$. Then by the definition of $\kappa$ for any $x > \beta_\kappa$ we must have $x - \beta_j \in \mathbb{L}$ for some $j \leq \kappa$, that is,

$$x \in \bigcup_{j=1}^{\kappa} (\mathbb{L} + \beta_j).$$

Hence, for any $T > \beta_\kappa$ the interval $(\beta_\kappa, T]$ is covered by $\kappa \leq 3$ translates of $\mathbb{L}$. This implies that

$$\mu(\mathbb{L} \cap [0,T]) \geq (1/\kappa - o(1))T \geq (1/3 - o(1))T.$$

To remove the $o(1)$-term we want to consider a tuple $\beta_1 < \cdots < \beta_\kappa$ which is in some appropriate sense the minimal tuple such that (4.3) holds (see [II, proof of Proposition 4] for details).

In the next section we recall the method of Banks, Freiberg, and Maynard [11], and the idea of Pintz [52]. In the subsequent section we explain the trick used in [II] to amplify the result.

## 4.1 The BFM-method and a refinement of Pintz

The strategy of Banks, Freiberg, and Maynard builds on the Maynard-Tao sieve (see Section 2.2). Let $k \geq 2$ and suppose that we want to prove a version of Theorem 3 for $k$ in place of 4, that is, that for all $\beta_1 \leq \cdots \leq \beta_k$

$$\mathbb{L} \cap \{\beta_j - \beta_i : 1 \leq i < j \leq k\} \neq \emptyset. \tag{4.4}$$

To do this we use the Maynard-Tao sieve with an admissible $K$-tuple of the form

$$\mathcal{H} = \mathcal{H}_1 \cup \cdots \cup \mathcal{H}_k.$$

Here $K = kL$ is some sufficiently large constant, and $\mathcal{H}_j$ are admissible $L$-tuples satisfying

$$h \in \mathcal{H}_j \Rightarrow h = (\beta_j + o(1)) \log x.$$

To prove (4.4) we now need to show that for all large $x$ there exists an $n \in [x, 2x]$ such that for some $1 \leq i < j \leq k$ both $n + \mathcal{H}_i$ and $n + \mathcal{H}_j$ contain prime numbers $p^{(i)}, p^{(j)}$, respectively. Indeed, then for some $h_i \in \mathcal{H}_i$ and $h_j \in \mathcal{H}_j$

$$p^{(j)} - p^{(i)} = h_j - h_i = (\beta_j - \beta_i + o(1)) \log x,$$

so that $\beta_j - \beta_i + o(1)$ is a normalized difference of primes. Furthermore, we need to guarantee that these primes are consecutive. However, let us ignore this for the moment, we will explain how this affects the argument at the end of this section.

Recall that by the Maynard-Tao sieve we can detect $n \in [x, 2x]$ such that $n + \mathcal{H}$ contains at least two primes. Therefore, we just need to ensure that these two primes belong to different $n + \mathcal{H}_j$. This holds for $n \in [x, 2x]$ if

$$\sum_{h \in \mathcal{H}} \mathbf{1}_{\mathbb{P}}(n+h) - 1 - \sum_{j=1}^{k} \sum_{\substack{h, h' \in \mathcal{H}_j \\ h < h'}} \mathbf{1}_{\mathbb{P}}(n+h) \mathbf{1}_{\mathbb{P}}(n+h') > 0. \tag{4.5}$$

Here the third term ensures that the primes detected are not in a single set $n + \mathcal{H}_j$. If $\nu_{\mathcal{H}}(n)$ are the Maynard-Tao sieve weights (2.15), then this follows if we have

$$\sum_{x < n \leq 2x} \Bigg( \sum_{h \in \mathcal{H}} \mathbf{1}_{\mathbb{P}}(n+h) - 1 - \sum_{j=1}^{k} \sum_{\substack{h, h' \in \mathcal{H}_j \\ h < h'}} \mathbf{1}_{\mathbb{P}}(n+h) \mathbf{1}_{\mathbb{P}}(n+h') \Bigg) \nu_{\mathcal{H}}(n) > 0.$$

Hence, compared to Section 2.2, the only new ingredient needed is an upper bound for the weighted sums over prime pairs for $h, h' \in \mathcal{H}_j,\ h < h'$

$$S(h, h') := \sum_{x < n \leq 2x} \mathbf{1}_{\mathbb{P}}(n+h) \mathbf{1}_{\mathbb{P}}(n+h')) \nu_{\mathcal{H}}(n). \tag{4.6}$$

We can choose $\nu_{\mathcal{H}}(n) = (\sum_{d_j | (n+h_j),\, d_1 \cdots d_K \leq \sqrt{D}} \lambda_{d_1, \ldots, d_K})^2$ with $D = x^{\delta}$ for some very small $\delta > 0$. In practise this means that for the weighted sum $S(h, h')$ we get by (2.16) the level of distribution $x^{1/2 - \epsilon}/D = x^{1/2 - \epsilon - \delta}$, which is essentially the same as for the unweighted counterpart. To show this one has to expand the square in $\nu_{\mathcal{H}}(n)$ and then apply the Bombieri-Vinogradov theorem (2.16).

Similar to (2.21), in [11] Selberg's sieve is used to give a bound for $S(h,h')$ which is larger than the expected main term by a factor of $4$. This it turns out is sufficient for (4.4) with $k = 9$. To prove this for $k = 5$ Pintz [52] observes that instead of (4.5) it suffices to consider a weighted version

$$\sum_{x<n\leq 2x}\Bigg(\sum_{h\in\mathcal{H}}\mathbf{1}_{\mathbb{P}}(n+h)-\mu'-\mu\sum_{j=1}^{k}\sum_{\substack{h,h'\in\mathcal{H}_j\\ h<h'}}\mathbf{1}_{\mathbb{P}}(n+h)\mathbf{1}_{\mathbb{P}}(n+h')\Bigg)\nu_{\mathcal{H}}(n)$$

with $\mu > 0$ chosen optimally and

$$\mu' := \max_{v\in\mathbb{N}}\left(v-\mu\binom{v}{2}\right).$$

To see why this helps note that in (4.5) the last of the three terms can give a very large negative contribution if $n+\mathcal{H}_j$ contains many primes for some given $j \leq k$. To mollify this we choose a small weight $\mu$, at the cost of increasing $\mu'$. The trick we explain in the next section further helps to diminish the contribution of $n$ for which $(n+\mathcal{H}_j)\cap\mathbb{P}$ is very large, allowing us to get $k = 4$.

We still need to explain how to guarantee that the primes detected are consecutive. Recall first that by the well-known construction of Erdős and Rankin [20, 55], for any $C > 0$ we can restrict $n$ to an arithmetic progression $n \equiv b\,(W)$ for some $W \sim x^{\epsilon}$ such that the interval $[n, n + C\log x]$ contains no prime numbers. By a small modification as in [11] we can construct $b$ such that for $n \equiv b\,(W)$ the only possible primes in $[n, n + C\log x]$ must lie in the set $n+\mathcal{H}$. Thus, if we restrict to $n \equiv b\,(W)$, the primes detected in the above will necessarily be consecutive. However, since $W = x^{\epsilon}$, this causes an issue with the Bombieri-Vinogradov theorem (2.16) since the moduli are now restricted to multiples of $W$. In [11] this is solved by the fact that there is at most one prime $q$ for which there might be an exceptional character modulo $q$, and in that case we may demand that $q \nmid W$, and restrict the weights $\lambda_{d_1,\dots,d_K}$ to $q \nmid d_1\cdots d_K$. We can essentially take this part for granted in [II] so we will not discuss this further here.

## 4.2 The trick

We now explain the main new ideas in [II]. By a slight generalization of the previous section we can show the following. Suppose that for some constant $A$ we have a bound for prime pairs ($h,h'\in\mathcal{H},\ h<h'$)

$$S(h,h') \leq (A+o(1))\cdot(\text{expected main term}). \tag{4.7}$$

Let $M := \lceil Aa\rceil + 1$ for some integer $a \geq 1$, and suppose that $\mathcal{H} = \mathcal{H}_1\cup\cdots\cup\mathcal{H}_M$ is an admissible $K$-tuple with $K$ a sufficiently large multiple of $M$. Then for some

$n \in [x, 2x]$ for at least $a+1$ distinct indices $j$ the set $n+\mathcal{H}_j$ contains a prime number. This result has already appeared in [4, 11].

At first sight (using $a = 1$) it appears that to improve $k = 5$ to $k = 4$ in (4.4) we would need to improve the constant from $A = 4$ to $A = 3$. However, it turns out that any $A < 4$ is enough, say, $A = 3.99$. To see this suppose that

$$\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2 \cup \mathcal{H}_3 \cup \mathcal{H}_4$$

is as in the previous section with the added structure that for each $j \in \{1, 2, 3, 4\}$

$$\mathcal{H}_j = \bigcup_{\ell=1}^{100} \mathcal{H}_{j,\ell}.$$

Then applying the above result with $a = 100$ and $M = \lceil 3.99 \cdot 100 \rceil + 1 = 400$ we see that there are at least $a + 1 = 101$ distinct pairs $(j, \ell)$ such that $n + \mathcal{H}_{j,\ell}$ contains a prime. Hence, by the pigeon-hole principle, there must be $1 \leq i < j \leq 4$ such that both $n + \mathcal{H}_i$ and $n + \mathcal{H}_j$ contain a prime, as claimed. As in the previous section, we may use the modified Erdős-Rankin construction to get consecutive primes.

The main technical task in [II] is then obtaining (4.7) with $A = 3.99$. It turns out that the presence of the Maynard-Tao sieve weight $\nu_{\mathcal{H}}(n)$ does not cause much difficulties (since $D \leq x^{\delta}$), and we are able to achieve this by Chen's sieve (see Section 2.4).

## 4.3 Parity barrier for the limit point problem

Note that to obtain (4.4) with $k = 3$ we would need (4.7) with $A = 3 - \epsilon$ for any $\epsilon > 0$. To show that $\mathbb{L} = [0, \infty]$ we would require $A = 2 - \epsilon$, which is equivalent to breaking the so-called parity barrier. Notice that by [23, Chapter 15] we expect $A = 2 - \epsilon$ to be as difficult as obtaining a non-trivial lower bound for $S(h, h')$.

It is reasonable to ask why the parity problem comes up when we try to show $\mathbb{L} = [0, \infty]$. Perhaps it is just a figment of the method that could be removed by some elaborate set-up. However, the following argument suggests that there is a real issue.

Instead of asking for twin primes, we can ask if $\lambda(p + 2)$ takes both values $\pm 1$ infinitely many times as $p$ varies over primes, where $\lambda(n)$ denotes the Liouville function. This question is unsolved, and clearly subject to a parity problem similar to the twin prime conjecture. We claim that this problem is morally equivalent to showing that $\mathbb{L} = [0, \infty]$.

To see this let $\mathcal{H}_1$ and $\mathcal{H}_2$ be sets of integers with

$$h \in \mathcal{H}_1 \Rightarrow h = o(\log x) \quad \text{and} \quad h \in \mathcal{H}_2 \Rightarrow h = (t + o(1)) \log x.$$

To show that $t \in \mathbb{L}$ we require that for arbitrarily large $x$ there exists $n \in [x, 2x]$ such that both $n + \mathcal{H}_1$ and $n + \mathcal{H}_2$ contain a prime. Note that consecutiveness is again obtained via the modified Erdős-Rankin construction.

Similarly, let $\mathcal{L}_1$ and $\mathcal{L}_2$ be sets of affine functions of the form

$$\mathcal{L}_j = \{L_a : \, L_a(n) = an - 2, \; \Omega(a) \equiv j \; (2)\}.$$

Suppose then that we show that for all large $x$ there exists $n \in [x, 2x]$ and some $L_{a_1} \in \mathcal{L}_1$, $L_{a_2} \in \mathcal{L}_2$ such that $L_{a_1}(n) = p_1 \in \mathbb{P}$ and $L_{a_2}(n) = p_2 \in \mathbb{P}$. Then $\lambda(p+2)$ takes both values infinitely often, since

$$\Omega(p_1 + 2) = \Omega(a_1) + \Omega(n) \not\equiv \Omega(a_2) + \Omega(n) = \Omega(p_2 + 2) \pmod 2$$

implies that $\lambda(p_1 + 2) \neq \lambda(p_2 + 2)$.

We expect these two problems to be roughly of the same level of difficulty. This is because with respect to the arithmetic information, given by the Bombieri-Vinogradov theorem (2.16), there is very little difference between shifts $n + h$ and affine functions $L(n)$. It is still possible that there is some way to exploit the averaging over the shifts $h \in \mathcal{H}_j$ but this average is very short $o(\log x)$, so it is hard to see how this would help. Likewise, it is possible that the average over $L \in \mathcal{L}_j$ could help with the second problem.

## 4.4 A correction

Unfortunately there is an error in the published version of the article [II] in the proofs of Lemmata 15 and 16. Namely, when we write

$$r_d = \sum_{\substack{d_1,\dots,d_K \\ d'_1,\dots d'_K \\ d_j = d'_j = d_\ell = d'_\ell = 1}} \lambda_{d_1,\dots,d_K} \lambda_{d'_1,\dots,d'_K} \Bigg( \sum_{\substack{N < n \leq 2N \\ n \equiv b\,(W) \\ n \equiv -h_\ell\,(d) \\ n \equiv -h_i\,([d_i, d'_i])}} 1_{\mathbb{P}}(n + h_j) - g(d) \sum_{\substack{N < n \leq 2N \\ n \equiv b\,(W) \\ n \equiv -h_i\,([d_i, d'_i])}} 1_{\mathbb{P}}(n + h_j) \Bigg)$$

in the proof of Lemma 15, then the first sum in the brackets is empty if $(d, d_i) > 1$ for some $i$ but the second sum is not empty. Note that this problem does not happen in the proof of Lemma 17 (or in the proof of [11, Lemma 6(iii)]) where the Selberg sieve is used, thanks to the fact that the Selberg sieve wieghts are readily of the same form as the Maynard-Tao sieve weights. This issue can be fixed. As is so often the case, the fundamental lemma of the sieve comes to the rescue. The idea is to handle small prime factors with Selberg type sieve weights (in the spirit of the fundamental lemma of the sieve), so that in the linear sieve $g(d)$ and $r_d$ will be supported on numbers $d$ with no small prime factors, so that the contribution from the part where $(d, d_i) > 1$ is negligible. The details of this correction can be found in Section 6 of the updated arXiv version of the article (arXiV:1811.03008).

# 5 Large square divisors of shifted primes

Recall that Landau's fourth problem, which asks for primes $p$ such that $p-1$ is a perfect square. To approximate this, we can ask what the largest $\theta \in (0,1)$ is such that there are infinitely many primes $p$ with a large square divisor $d^2|(p-1)$, $d^2 > p^\theta$. To rewrite this problem, we seek primes in arithmetic progressions $p \equiv 1\,(d^2)$ to large square moduli. By Dirichlet's theorem for any fixed modulus $q$ and residue class $a$ coprime to $q$ there are infinitely many primes $p \equiv a\,(q)$. The generalization of the Prime number theorem tells us that the primes are equidistributed among primitive residue classes, that is, for any fixed modulus $q$ and $(a,q)=1$ (see [15, Chapter 22], for instance)

$$\sum_{\substack{p\leq x\\ p\equiv a\,(q)}} 1 \sim \frac{1}{\varphi(q)}\sum_{p\leq x} 1. \tag{5.1}$$

By the Bombieri-Vinogradov theorem (2.16) we know that this holds on average over moduli $q \leq x^{1/2-\epsilon}$.

For our problem we essentially want a version of (2.16) with the moduli restricted to squares $q = d^2$. The result of Baier and Zhao [1, 2] gives this for moduli up to $d^2 \sim x^{4/9-\epsilon}$ for any $\epsilon > 0$, solving the above-mentioned square divisor problem for $\theta = 4/9 - \epsilon$. By combining this with Harman's sieve method, Matomäki [45] improved this to $\theta = 1/2 - \epsilon$, which corresponds to the Bombieri-Vinogradov range. Note that Matomäki's result gives a lower bound for primes $p \equiv 1\,(d^2)$ for almost all $d^2 \sim x^\theta$ instead of an asymptotic. It should be noted that in both of these results we may also restrict $d$ to be a prime. Finally, Baker [3] has shown the asymptotic version analogous to (2.16) with square moduli up to $d^2 \leq x^{1/2-\epsilon}$.

Note that even if we assume the Generalized Riemann hypothesis, which implies (5.1) with the error term $O(x^{1/2+\epsilon})$ uniformly in $a$ and $q$ for $\epsilon > 0$ arbitrarily small, the expnent $\theta = 1/2 - \epsilon$ is the best result that was previously known for the square divisor problem. In the third article [III] the main goal is to break past the $\theta = 1/2$ barrier. The starting point is the result of Zhang [62] which extended the Bombieri-

Vinogradov range for smooth moduli, showing that for any fixed $a$

$$\sum_{\substack{d \leq x^{1/2+2\varpi} \\ d\, x^{\delta}\text{-smooth and square-free}}} \Bigg| \sum_{\substack{p \leq x \\ p \equiv a\,(d)}} 1 - \frac{1}{\varphi(d)} \sum_{p \leq x} 1 \Bigg| \ll_C x \log^{-C} x \tag{5.2}$$

with $2\varpi = 1/584$ for some fixed small $\delta > 0$. The exponent $2\varpi = 7/300$ was obtained by the collaborative Polymath project [53]. This was sharpened to $2\varpi = 1/42$ by Maynard [49], with a less restrictive assumption for the moduli. These results in turn were inspired, among others, by the previous works of Bombieri, Fouvry, Friedlander, and Iwaniec [12, 22, 24].

Zhang's result came in 2014 while I was still an undergraduate student, but the prospect of studying the method and using it in other problems stuck with me, until in the second year of my doctoral studies I realized that I could apply this to the square divisor problem. My thesis supervisor Prof. Kaisa Matomäki had studied the problem [45], but this is in fact a complete coincidence. It should also be noted that most of the work in [III] was conducted while I was visiting ETH Zürich under guidance of Prof. Emmanuel Kowalski, to whom I am very grateful for hospitality and numerous inspiring discussions.

The main result in [III] breaks the $\theta = 1/2$ barrier for the first time. Since it is of no additional effort, we prove the result for general shifts of primes.

**Theorem 4.** *Fix $a \neq 0$ and let $\theta = 1/2 + 1/2000$. Then there are infinitely many primes $p$ such that $d^2 | (p-a)$ for some square $d^2 \geq p^{\theta}$.*

As mentioned before, the plan is to use arguments similar to Zhang [62]. That is, we assume that $d$ is composed of small prime factors $d = p_1 \cdots p_K$ with $p_j \asymp x^{\delta}$ for some small $\delta$. Similarly to the first article [I], the benefit is that we may obtain a suitable factorization $d = d_1 d_2$ with accuracy $x^{\delta}$, which is in fact quite a common theme in analytic number theory. However, the precise way this is exploited is completely different to [I]. Similarly to Matomäki's work [45], we will use Harman's sieve method (see Section 2.1) to tackle the problem. In the next section we describe the set-up and some details on the sieve argument. In the subsequent section we explain Zhang's argument in our setting.

## 5.1 The set-up and the sieve

As mentioned, we will restrict the moduli $d$ for factorization purposes. More precisely, let $D := x^{\theta}$ for $\theta = 1/2 + 1/2000$, let $\delta = 1/K$ for some large integer $K$, $P = D^{1/K}$, and define the intervals

$$I_j := (2^{j-1} P^{1/2}, 2^j P^{1/2}] \quad \text{for} \quad j \in \{1, 2, \ldots, K\}.$$

Set

$$\mathcal{D} := \{p_1^2 \cdots p_K^2 : p_j \in I_j \quad \text{for} \quad j \in \{1, 2, \dots, K\}\}, \tag{5.3}$$

so that any $d^2 \in \mathcal{D}$ satisfies $d^2 \asymp D$. For $d^2 \in \mathcal{D}$ and $z < x$, denote

$$S(\mathcal{A}^d, z) := \sum_{\substack{x < n \leq 2x \\ n \equiv a \, (d^2) \\ (n, P(z)) = 1}} 1.$$

To prove Theorem 4 we show that for almost all moduli $d^2 \in \mathcal{D}$ we have

$$S(\mathcal{A}^d, 2\sqrt{x}) \gg \frac{x}{\varphi(d^2) \log x}, \tag{5.4}$$

which is even stronger than the claim in Theorem 4. To apply Harman's sieve method, we only need the arithmetic information for almost all moduli $d^2 \in \mathcal{D}$. As usual, the arduous task is to handle the Type II sums, which is done in [III, Proposition 4].

**Proposition 5.** (Type II information). *Let $\mathcal{D}$ be as in (5.3), $2\varpi = 1/2000$, $\sigma = 1/19.5$, and $MN = x$ with*

$$M, N \in [x^{1/2-\sigma}, x^{1/2+\sigma}] \setminus [x^{1/2-2\varpi-\delta}, x^{1/2+2\varpi+\delta}].$$

*Let $\alpha(m)$ and $\beta(n)$ be bounded functions. Then for some $\eta > 0$*

$$\sum_{d^2 \in \mathcal{D}} \Bigg| \sum_{\substack{mn \equiv a \, (d^2) \\ m \sim M, \, n \sim N}} \alpha(m)\beta(n) - \frac{1}{\varphi(d^2)} \sum_{\substack{(mn, d^2) = 1 \\ m \sim M, \, n \sim N}} \alpha(m)\beta(n) \Bigg| \ll \frac{x^{1-\eta}}{\sqrt{D}}.$$

The width parameter $\sigma$ is determined by the condition

$$19\sigma + 90\varpi + 71\delta < 1,$$

which holds for $\varpi = 1/4000$ and $\sigma = 19.5$ if $\delta > 0$ is small enough. It turns out that this gives just enough room for Harman's sieve method, although this requires some elbow grease to optimize the Buchstab decompositions. The fact that this range is so narrow is entirely due to the fact that the moduli are restricted to a sparse set.

Compared to the sieves in Sections 2.1 and 3, the new feature here is the gap $[x^{1/2-2\varpi-\delta}, x^{1/2+2\varpi+\delta}]$ in the Type II range near $x^{1/2}$. However, since this gap is proportional to $\varpi$ we expect that the effect of this is negligible if we make $\varpi$ very small. After all, we expect that sieves are continuous with respect to the quality of the information. To make the argument rigorous in order to get the lower bound

(5.4), it turns out that we need to cover this gap at least with Type I information [III, Proposition 5]. To see this note that by applying Buchstab's identity we get

$$S(\mathcal{A}^d, 2\sqrt{x}) = S(\mathcal{A}^d, z) - \sum_{z \leq p < 2\sqrt{x}} S(\mathcal{A}^d_p, p).$$

Here in the second sum we cannot handle the terms where $p \geq x^{1/2-2\varpi-\delta}$ as a Type II sum, and we cannot simply discard them because they are negative. However, if we have Type I information up to $x^{1/2+2\varpi+\delta}$, then this issue is resolved by another application of Buchstab's identity, and the error term resulting from the gap is indeed proportional to $\varpi$ (see [III, Section 2.1 for details]).

## 5.2 Bilinear equidistribution estimate of Zhang type for square moduli

In this section we sketch the proof of Proposition 5. To motivate this, suppose that we wish to bound a trilinear sum of the form

$$\sum_{m \sim M} \sum_{n \sim N} \sum_{d \sim D} \alpha_m \beta_n \gamma_d \Phi(m, n, d)$$

for some bounded coefficients $\alpha_m, \beta_n, \gamma_d$, where we expect cancellation from some oscillating function $\Phi(m, n, d)$. To do this we may apply Cauchy-Schwarz to smoothen the coefficient $\alpha_m$, but there are two ways to do this, resulting in either

$$M^{1/2} D^{1/2} \Bigg( \sum_{n_1, n_2 \sim N} \beta_{n_1} \bar{\beta}_{n_2} \sum_{d \sim D} \sum_{m \sim M} \Phi(m, n_1, d) \overline{\Phi(m, n_2, d)} \Bigg)^{1/2} \tag{5.5}$$

or

$$M^{1/2} \Bigg( \sum_{n_1, n_2 \sim N} \beta_{n_1} \bar{\beta}_{n_2} \sum_{d_1, d_2 \sim D} \gamma_{d_1} \bar{\gamma}_{d_2} \sum_{m \sim M} \Phi(m, n_1, d_1) \overline{\Phi(m, n_2, d_2)} \Bigg)^{1/2}. \tag{5.6}$$

In the first case the diagonal part is $n_1 = n_2$, and in the second case it is $d_1 = d_2, n_1 = n_2$. In the diagonal part we cannot exhibit any cancellation in the sum over $m$, so that the savings must come from the fact that the diagonal is a sparse subset of the parameters. This means that in the diagonal contribution we get a better bound in the second case (5.6). More precisely, the diagonal contributions for (5.5) and (5.6) are trivially bounded by $MDN^{1/2}$ and $MD^{1/2}N^{1/2}$, respectively. These should be compared to the trivial bound $MDN$ for the original sum.

In the off-diagonal part we can show cancellation in the sum over $m$, but now in the second case (5.6) the oscillating function $\Phi(m, n_1, d_1)\overline{\Phi(m, n_2, d_2)}$ is much more complicated than in (5.5). Hence, we expect that for the off-diagonal part the second option is worse.

Thus, we are faced with a trade-off between the diagonal and the off-diagonal contributions. If we assume that the sum over $d$ factorizes as

$$\sum_{d\sim D}\gamma_d=\sum_{r\sim R}\sum_{q\sim Q}\gamma'_{r,q},$$

we can find a middle-ground between the two cases by keeping the sum over $q$ inside and the sum over $r$ outside of the application of Cauchy-Schwarz. In practice the factorization $D=RQ$ is always decided so that the bound for the diagonal contribution is just sufficient, so that the oscillating factor $\Phi(m,n_1,rq_1)\overline{\Phi(m,n_2,rq_2)}$ remains as simple as possible.

Since $d^2\in\mathcal{D}$ is of the form $d^2=p_1^2\cdots p_K^2$ for $p_j\in I_j$, we can obtain such a factorization with accuracy $x^\delta$, so that for Proposition 5 we need to bound a sum of the form (writing $r\in\mathcal{R}:=\{p_1\cdots p_\ell\}$, and $q\in\mathcal{Q}:=\{p_{\ell+1}\cdots p_k\}$)

$$S(M,N):=\sum_{r^2\in\mathcal{R}}\sum_{q^2\in\mathcal{Q}}c_{rq}\sum_{m\sim M,\ n\sim N}\alpha(m)\beta(n)\Phi(m,n,rq),$$

where $|c_{rq}|\leq 1$ and $\Phi(m,n,d)=\mathbf{1}_{mn\equiv a\,(d^2)}-\mathbf{1}_{(mn,d^2=1)}/\varphi(d^2)$. The choice of this factorization is dynamic in the sense that it depends on the sizes of $M$ and $N$. More precisely, if $N<M$, then it turns out that we need $r^2q$ to be a bit smaller than $N$ to control the diagonal contribution.

By applying Cauchy-Schwarz we get

$$\begin{aligned}
&S(M,N)\ll M^{1/2}R^{1/4}\cdot\\
&\Bigg(\sum_{r^2\in\mathcal{R}}\sum_{m\sim M}\Bigg|\sum_{q^2\in\mathcal{Q}}c_{rq}\sum_{n\sim N}\beta(n)(\mathbf{1}_{mn\equiv a\,(r^2q^2)}-\mathbf{1}_{(mn,r^2q^2=1)}/\varphi(r^2q^2))\Bigg|^2\Bigg)^{1/2}\\
&=:M^{1/2}R^{1/4}(W-V_1-V_2-U)^{1/2}
\end{aligned}$$

by expanding the square to get four terms. To bound $S(M,N)$ we need to show that the sums $W$, $V_1$, $V_2$, and $U$ are all asymptotically equal to each other, so that the main terms cancel in the sum $W-V_1-V_2+U$. This is commonly known as Linnik's dispersion method.

By far the hardest task is to compute the sum

$$W:=\sum_{r^2\in\mathcal{R}}\sum_{q_1^2,q_2^2\in\mathcal{Q}}c_{rq_1}\bar{c}_{rq_2}\sum_{n_1,n_2\sim N}\beta(n_1)\overline{\beta(n_2)}\sum_{\substack{m\sim M\\ mn_1\equiv a\,(r^2q_1^2)\\ mn_2\equiv a\,(r^2q_2^2)}}1.$$

In the diagonal part ($n_1=n_2,q_1=q_2$) of $W$ we use a trivial bound, which is sufficient by the choice of the factorization. Suppose for simplicity that in the remaining part we have $(q_1,q_2)=1$. From the two congruences in the sum over $m$ we infer that $n_2 = n_1 + \ell r^2$ for some $|\ell| \ll L := N/R$, so that the sum is of the form

$$\sum_{r^2 \in \mathcal{R}} \sum_{q_1^2, q_2^2 \in \mathcal{Q}} c_{rq_1} \bar{c}_{rq_2} \sum_{n \sim N} \sum_{0 < |\ell| \ll L} \beta(n) \overline{\beta(n + \ell r^2)} \sum_{\substack{m \sim M \\ m \equiv \gamma \, (r^2 q_1^2 q_2^2)}} 1$$

for some residue class $\gamma = \gamma(a, r, q_1, q_2, n, \ell)$. We can use the Poisson summation formula to sum over $m$ to extract the correct main term from the frequency $h = 0$, with the resulting error term essentially bounded by

$$\sum_{r^2 \in \mathcal{R}} \sum_{q_1^2, q_2^2 \in \mathcal{Q}} c_{rq_1} \bar{c}_{rq_2} \sum_{0 < |\ell| \ll L} \sum_{n \sim N} \beta(n) \overline{\beta(n + \ell r^2)} \frac{1}{H} \sum_{0 < |h| \leq H} f_h e_{r^2 q_1^2 q_2^2}(h \gamma(a, r, q_1, q_2, n, \ell)),$$

where $H := r^2 q_1^2 q_2^2 / M$ and $f_h$ are some bounded coefficients. Here $e_q(x) := e^{2\pi i x/q}$ is an additive character modulo $q$. To make this rigorous we need to replace the condition $m \sim M$ by a smoothed version at the Cauchy-Schwarz step, but let us ignore such technical details here.

We now wish to apply Cauchy-Schwarz yet again to smoothen the coefficient of $n$. To optimize this step we have to re-factorize $q_1^2 = u^2 v^2$ suitably, which we skip here to simplify the presentation. By applying Cauchy-Schwarz with sums over $n$, $r$, $\ell$ outside we produce a sum of the form

$$\sum_{0 < |\ell| \ll L} \sum_{r^2 \in \mathcal{R}} \sum_{\substack{q_1^2, q_2^2 \in \mathcal{Q} \\ s_1^2, s_2^2 \in \mathcal{Q}}} \frac{1}{H^2} \sum_{h_1, h_2} c(\ell, r, q_1, s_1, q_2, h_1, h_2) \sum_{n \sim N} e_{r^2 q_1^2 q_2^2}(h_1 \gamma(n)) e_{r^2 s_1^2 s_2^2}(-h_2 \tilde{\gamma}(n)),$$

where $\gamma(n) = \gamma(a, r, q_1, q_2, n, \ell)$ and $\tilde{\gamma}(n) = \gamma(a, r, s_1, s_2, n, \ell)$. Again, the diagonal part is bounded trivially, and in the off-diagonal part we assume for simplicity that $(q_1, s_1) = 1$ and $(q_2, s_2) = 1$. Then, tracking the functions $\gamma(n)$ and $\tilde{\gamma}(n)$ more carefully, we see that the innermost sum is of the form

$$\Sigma(N) := \sum_{n \sim N} e_{d_1^2 d_2^2}(f(n)) = \sum_{n \sim N} e_{d_1^2}(c_1/n) e_{d_2^2}(c_2/(n + \tau))$$

for $d_1 = r q_1 s_1$ and $d_2 = q_2 s_2$, some constants $c_1, c_2, \tau$, and with the inverses $1/n$ and $1/(n + \tau)$ computed modulo $d_1^2$ and $d_2^2$, respectively. Here the sum over $n$ is restricted to $(n, d_1) = (n + \tau, d_2) = 1$.

Recall that by the Poisson summation formula for any integer $q \geq 1$ and any $q$-periodic function $g(n)$ we morally have

$$\sum_{n \sim N} g(n) \sim \frac{N}{q} \sum_{|t| \leq q/N} c_t \sum_{n \in \mathbb{Z}/q\mathbb{Z}} g(n) e_q(tn) \tag{5.7}$$

for some bounded smooth coefficients $c_t$ (see [39, Chapter 12.2], for instance). This is also known as completing the sum over $n$. Hence, the sum $\Sigma(N)$ can be bounded provided that we can bound the completed sums

$$\sum_{n \in \mathbb{Z}/d_1^2 d_2^2 \mathbb{Z}} e_{d_1^2 d_2^2}(f(n) + tn)$$

for $t \in \mathbb{Z}/d_1^2 d_2^2 \mathbb{Z}$. By the Chinese remainder theorem this can be bounded if we can bound exponential sums of the form

$$S(f, p^j) := \sum_{n \in \mathbb{Z}/p^j \mathbb{Z}} e_{p^j}(f(n) + tn)$$

for $p | d_1 d_2$ with $j \in \{1, 2\}$. Note that it is possible that the rational function $f(n)+tn$ is divisible by some prime $p$ but not by $p^2$, so that we need a bound also for $j = 1$.

For $j = 1$ we have the famous Weil bound [59] (proved by techniques from algebraic geometry) which gives the square root cancellation $S(f, p) \ll \sqrt{p}$. For $j = 2$ we can use bounds of Cochrane and Zheng [14] to show the square root bound $S(f, p^2) \ll p$. Combining these we essentially get $\Sigma(N) \ll d_1 d_2$, which is non-trivial if $N \gg d_1 d_2$. The non-trivial range can be extended by Heath-Brown's $q$-van der Corput method [32], which gives $\Sigma(N) \ll N^{1/2}(d_1 d_2)^{1/3}$, which is better than the previous bound for $N \ll (d_1 d_2)^{4/3}$, and is non-trivial up to $N \gg (d_1 d_2)^{1/3}$. Applying Heath-Brown's $q$-van der Corput method is one of the many new ideas in the Polymath paper [53] and here we also need the fact that the modulus can be factorized suitably. Combining everything we finally get Proposition 5. We refer to [53, Section 4] for a more thorough discussion of the fascinating topic of algebraic exponential sums (in the case of $j = 1$).

# 6 Large prime factors of $n^2 + 1$

In the article [IV] we approximate Landau's fourth problem by studying the largest prime factor of $n^2 + 1$. This problem has a long history as it was first studied by Chebyshev in the late 19th century. Let $P^+(k)$ denote the largest prime factor of $k$. Chebyshev proved that

$$\limsup_{n\to\infty} P^+(n^2+1)/n \to \infty \quad \text{as} \quad n \to \infty.$$

This result was published posthumously in a memoir by Markov. The result was sharpened by Nagell and Erdős (see the introduction in [35] for exact references for the above).

The modern history of the problem begins with Hooley [35], who used a sieve argument and the Weil bound for Kloosterman sums to prove that $P^+(n^2 + 1) > n^{1.10014\dots}$ infinitely often. The exponent was improved to $1.202468\dots$ by Deshouillers and Iwaniec [17] as an application of their bounds for sums of Kloosterman sums [18]. In addition, they showed that under Selberg's eigenvalue conjecture this could be improved to $1.2247\dots$ The best unconditional result prior to [IV] was the result of de la Bretèche and Drappeau [16] who obtained this with the exponent $1.2182$. To show this they combined the method of [17] with the current best bound towards Selberg's eigenvalue conjecture by Kim and Sarnak [43]. Note that improving this exponent to 2 corresponds to the original problem of primes $p = n^2 + 1$.

My interest in this problem started when I came across the Deshouillers-Iwaniec paper [17]. After reading this I followed what is perhaps the most important strategy in research — always try the first thing that comes to mind. I had just been working with Harman's sieve method, so naturally this is where my mind went, since all the previous works relied only on the linear sieve. Of course I doubted that this would work, since surely someone must have tried this before. Fortunately for me this had not been done, and using Harman's sieve method with the existing Type II estimates leads to a small improvement on [16]. During the Summer of 2019 I was able to also prove stronger Type II information, which led to a much better exponent. The main result in [IV] is as follows.

**Theorem 6.** *For infinitely many integers $n$ the largest prime factor of $n^2+1$ is at least $n^{1.279}$. Assuming Selberg's eigenvalue conjecture the exponent $1.279$ may be improved to $1.312$.*

The improvement comes from applying Harman's sieve method (see Section 2.1), which allows us to make use of Type II information. As mentioned above, to further improve the result I also proved a new Type II estimate in [IV]. In the next section we describe an idea of Chebyshev for detecting large prime factors, which underlies all of the progress on this problem. We also give some details about the sieve argument of [IV]. In Section 6.2 we give a short introduction to 'Kloosterma-nia' techniques, named so because of the vast number of results and applications. In Section 6.3 we show how this is applied in [IV] to handle the Type II sums for this problem.

## 6.1 Chebyshev's device and the sieve argument

Chebyshev's device starts with the elementary identity for the von Mangoldt function

$$\sum_{d|n} \Lambda(n) = \log n.$$

Using this we get

$$\sum_m \Lambda(m) \sum_{\substack{\ell\sim x\\ \ell^2+1\equiv 0\,(m)}} 1 = \sum_{\ell\sim x}\sum_{m|\ell^2+1}\Lambda(m) = \sum_{\ell\sim x}\log(\ell^2+1) = (2+o(1))x\log x.$$

Suppose that for every $n\in[x,2x]$ the largest prime factor of $n^2+1$ is at most $x^{\varpi}$ for some $\varpi<2$. Then by the above

$$\sum_{p\leq x^{\varpi}} \log p \sum_{\substack{\ell\sim x\\ \ell^2+1\equiv 0\,(p)}} 1 = \sum_p \log p \sum_{\substack{\ell\sim x\\ \ell^2+1\equiv 0\,(p)}} 1 = (2+o(1))x\log x.$$

Therefore, if we can show that

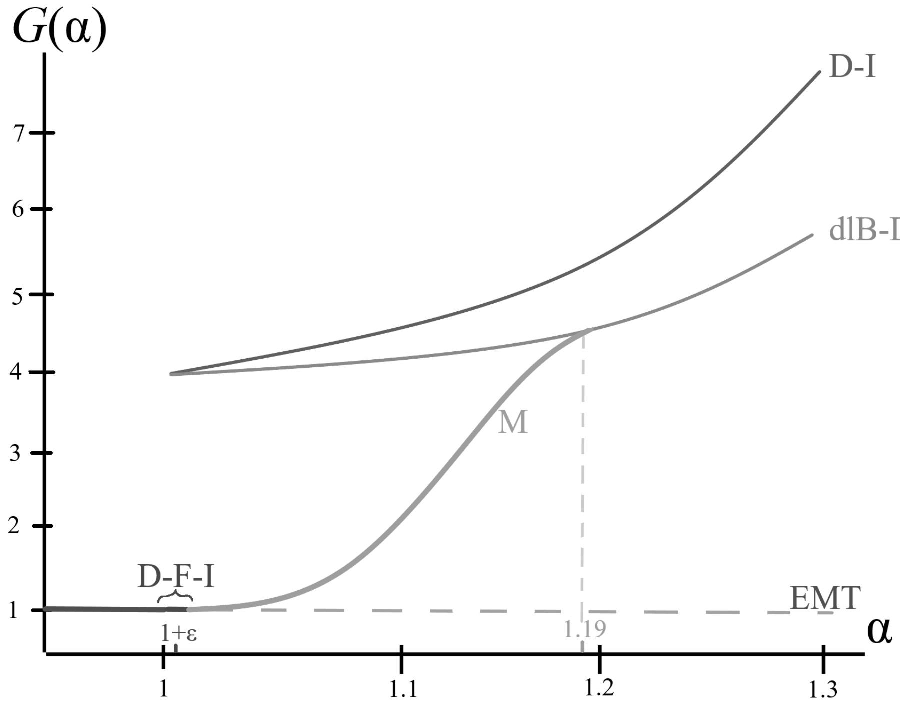
$$\sum_{p\leq x^{\varpi}} \log p \sum_{\substack{\ell\sim x\\ \ell^2+1\equiv 0\,(p)}} 1 \leq (2-\epsilon+o(1))x\log x,$$

we have a contradiction, so that the largest prime factor must be at least $x^{\varpi}$.

It is easy to give the correct asymptotic formula $x\log x$ for the sum over $p<x$ (see [17, Section 2]), so that we need to show

$$\sum_{x<p\leq x^{\varpi}} \log p \sum_{\substack{\ell\sim x\\ \ell^2+1\equiv 0\,(p)}} 1 \leq (1-\epsilon+o(1))x\log x. \tag{6.1}$$

Figure 6.1: Here we see the functions $G(\alpha)$ obtained in the different works. D-I refers to [17], dlB-D to [16], D-F-I to [19] (implicit in the article), M to [IV], and EMT to the expected main term. Notice the discontinuity at $\alpha = 1$ in [16] and [17].




In reality we have to replace the condition $\ell \sim x$ by a smooth function but let us ignore this here.

By splitting the sums dyadically to $p \sim P$ for $P = x^{\alpha} \leq x^{\varpi}$, to show (6.1) we need to show upper bounds of the form

$$S(x, P) := \sum_{p \sim P} \log p \sum_{\substack{\ell \sim x \\ \ell^2+1 \equiv 0\,(p)}} 1 \leq (G(\alpha) + o(1)) x \log 2$$

for some function $G(\alpha)$ with

$$\int_1^{\varpi} G(\alpha) d\alpha \leq 1 - \epsilon.$$

The function $G(\alpha)$ measures how far off we are from the expected main term for $S(x, P)$ with $P = x^{\alpha}$. That is, we have transformed the original problem of the largest prime factor into a sieve upper bound problem.

As mentioned, it is easy to get $G(\alpha) = 1$ for $\alpha \leq 1$. In Figure 6.1 we compare the bound of [IV] to the previous results. Note that this graph is not precise, but it illustrates the main features. Deshouillers and Iwaniec [17] obtained

$$G(\alpha) := \frac{4 \cdot (1-2\theta)\alpha}{1-2\theta\alpha}$$

via the linear sieve with $\theta = 1/4$. De la Bretèche and Drappeau [16] got the same $G(\alpha)$ but with $\theta = 7/64$. Notice how in both results for $\alpha = 1+\epsilon$ we have $G(\alpha) = 4 + O(\epsilon)$, that is, there is a jump from $G(\alpha) = 1$ to $G(\alpha) \geq 4$ at $\alpha = 1$. This already suggests that something is missing, since sieve bounds are supposed to be continuous, especially since the bound $G(1+\epsilon) = 1 + O(\epsilon)$ is implicit in the work of Duke, Friedlander, and Iwaniec [19]. This means that the function $G(\alpha)$ is continuous at least a tiny bit past $\alpha = 1$.

In [IV] we resolve this by using Harman's sieve method in the intermediate range $1 \leq \alpha \leq 1.19$ to get a continuous incline from $G(1) = 1$, joining to the bound of [16] at $\alpha = 1.19$. After $1.19$ we cannot improve $G(\alpha)$ compared to [16]. This is because we have Type II information available only up to $\alpha \leq 1.19$. For technical reasons the bound in [IV] is not quite continuous, and while with a lot of work this could be made so, this has little effect to the exponent $\varpi$. We do not present the exact formula for our $G(\alpha)$ as it is fairly complicated, being defined piece-wise as a sum of various Buchstab integrals. The Type I information used is exactly the same as in [16], and we will not discuss this part other than to say that the proofs are very similar to our Type II estimate.

In the previous sections we have seen only a lower bound version of Harman's sieve method. By similar techniques we can also prove sieve upper bounds, and in the presence of Type II information we can improve on the linear sieve upper bound. One way to see this is by recalling that the linear sieve is neutral with respect to Buchstab's identity. Hence, if we apply Buchstab's identity to generate Type II sums for which we now have an asymptotic formula, we improve the upper bound of the linear sieve.

For any integer $m$ let $\rho(m)$ denote the number of non-congruent solutions to $\nu^2 + 1 \equiv 0 \ (m)$. Note that $\rho$ is a multiplicative function, with $\rho(p) = 2$ for primes $p \equiv 1\,(4)$, and $\rho(p) = 0$ if $p \equiv 3\,(4)$. The Type II information is provided by [IV, Proposition 2] (again, one has to smoothen the condition $\ell \sim x$). See Figure 6.2 for an illustration of the viable ranges.

**Proposition 7.** (Type II information). *Let $\theta = 7/64$. Let $P = x^\alpha$ for some $\alpha \geq 1$, and let $MN = P$ for $M, N \geq 1$. Let $\alpha(m)$ and $\beta(n)$ be divisor bounded coefficients such that $\beta(n)$ is supported on square-free integers. Then*

$$\sum_{\substack{m \sim M \\ n \sim N}} \alpha(m)\beta(n) \log mn \sum_{\substack{\ell \sim x \\ \ell^2+1 \equiv 0\,(mn)}} 1 = x \sum_{\substack{m \sim M \\ n \sim N}} \frac{\alpha(m)\beta(n)\rho(mn)}{mn} \log mn + O(x^{1-\eta}),$$

*if one of the following holds:*
**(i)**

$$x^{\alpha-1+\eta} \ll N \ll x^{(2-2\theta-\alpha)/3-\eta} = x^{(57-32\alpha)/96-\eta}.$$

**(ii)**(Duke-Friedlander-Iwaniec+de la Bretèche-Drappeau) $\beta(n)$ *is supported on primes and*

$$x^{2(\alpha-1)+\eta} \ll N \ll x^{(4-(3+2\theta)\alpha)/(3-6\theta)} = x^{(128-103\alpha)/75-\eta}.$$

The range (i) is our own while the range (ii) is implicit in the work [16] based on [19]. The first range is better if $\alpha > 2671/2496 = 1.07\dots$, and non-trivial if $\alpha \leq 153/128 = 1.19\dots$ A key difference to the previous sections is the location of the Type II information, since $N$ is much smaller than $M$. This causes some changes in the sieve argument but nothing too drastic. The fact that $\beta(n)$ is supported on square-free numbers is not an issue, since we can guarantee this in the sieve argument.

We will explain how this is proved in the last section. For this we need to recall estimates from the theory of Kloosterman sums, which we will do in the next section alongside explaining the connection to the present problem.

## 6.2 Sums of Kloosterman sums and quadratic congruences

For weighted sums along linear equations (for example, sum over $n_1, n_2$ with $n_2 - n_1 = D$) we can apply Fourier analytic methods to compute the sum. For many second-order equations it is beneficial to use other type of harmonics which come from classical automorphic forms, and in practice involve the Kloosterman sums. We refer to the books of Iwaniec [37, 38] for a detailed introduction to the topic. The first sections in [18] also contain a summary of the basic theory. We can take much of the results as a black box, so that this superficial exposition should be sufficient.

For any integers $a, b, c$ with $c \neq 0$ the Kloosterman sum is defined as

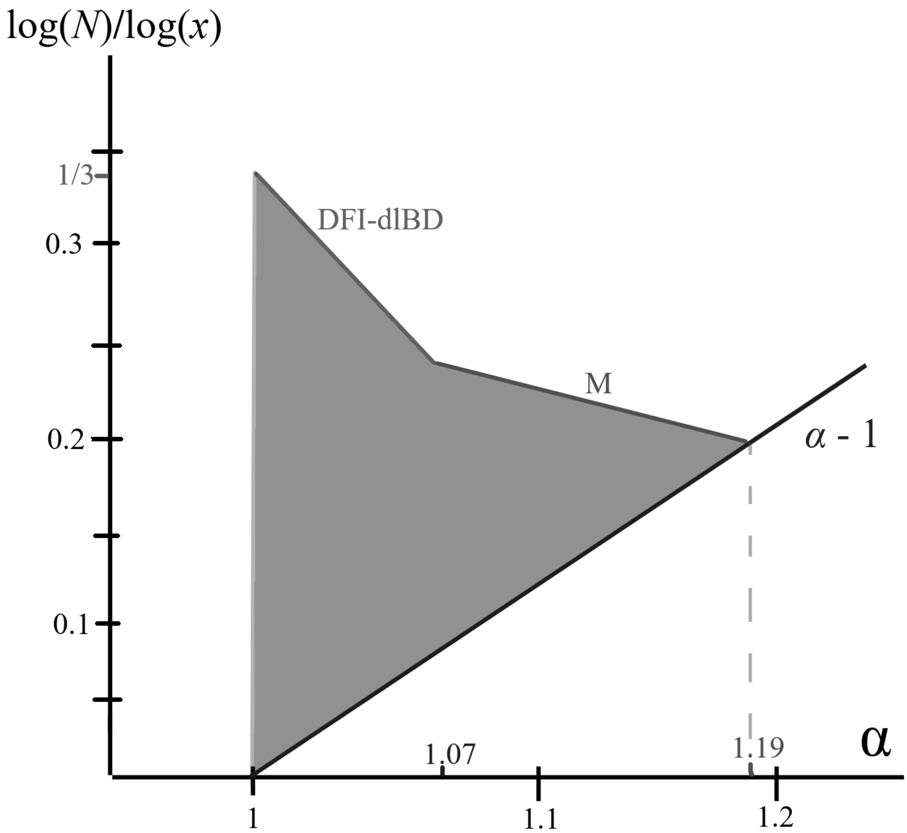
$$S(a,b;c) := \sum_{\substack{n\ (c) \\ (n,c)=1}} e_c(an + b\overline{n}).$$

Here $\overline{n}$ denotes the multiplicative inverse of $n$ modulo $c$. Kloosterman originally arrived at this definition while working on the representation of integers by a diagonal quadratic form $a_1 n_1^2 + a_2 n_2^2 + a_3 n_3^2 + a_4 n_4^2$ with fixed $a_1, a_1, a_3, a_4$ [44].

Kloosterman sums also appear naturally when we consider weighted sums along the determinant equation

$$\det \begin{pmatrix} n_1 & n_2 \\ n_3 & n_4 \end{pmatrix} = n_1 n_4 - n_2 n_3 = D$$

Figure 6.2: We can handle Type II sums if $\log N/\log x$ is in the range depicted. For small $\alpha$ it is beneficial to use the bound from [16, 19]. For large $\alpha$ we have proved a new estimate. Note that the range dissapears at $\alpha = 1.19$, and compare this with Figure 6.1.



for some fixed $D$. Our case with the sum over $\ell^2+1\equiv 0\,(p)$ can also be written in a similar form. Here $\ell^2+1=pk$ for some $k$, so that we are summing over $\ell$ and $k$ subject to

$$\det\begin{pmatrix}\ell & p\\ k & \ell\end{pmatrix}=\ell^2-pk=-1.$$

This representation is used most beautifully in the paper of Duke, Friedlander, and Iwaniec [19], where the argument connects such sums directly to automorphic forms.

For our Type II sums we use the layman's path to Kloosterman sums paved by Hooley [35]. To compute the sum over $\ell$ we split the sum into distinct residue classes and use Poisson summation (5.7) to obtain morally (the condition $\ell\sim x$ needs to be replaced by a smooth weight)

$$\sum_{\substack{\ell\sim x\\ \ell^2+1\equiv 0\,(q)}}1=\sum_{\substack{\nu\,(q)\\ \nu^2+1\equiv 0\,(q)}}\sum_{\substack{\ell\sim x\\ \ell\equiv\nu\,(q)}}1 \tag{6.2}$$

$$=x\frac{\rho(q)}{q}+\frac{1}{H}\sum_{0<|h|\le H}f_h\sum_{\substack{\nu\,(q)\\ \nu^2+1\equiv 0\,(q)}}e_q(-h\nu)+O_C(x^{-C}). \tag{6.3}$$

In the above we have $H=q/x$, and $f_h$ are certain smooth bounded coefficients. The first term on the right-hand side is the expected main term, so that we need to bound the remaining error term (on average over $q$).

To relate this to Kloosterman sums we require the following classic lemma of Gauss (see [17, Lemma 2], for instance).

**Lemma 8.** *If the equation $\nu^2+1\equiv 0\ (q)$ has a solution, then q has a representation as a sum of two squares*

$$q=r^2+s^2,\quad (r,s)=1,\quad r,s>0.$$

*Furthermore, there is a one-to-one correspondence between such representations and the solutions to $\nu^2+1\equiv 0\ (q)$, and we have (denoting the inverse of r modulo s by $\overline{r}$).*

$$\frac{\nu}{q}\equiv\frac{r}{s(r^2+s^2)}-\frac{\overline{r}}{s}\mod 1.$$

Using this lemma we can replace $e_q(-h\nu)$ by $e_s(h\overline{r})$ since

$$e\Big(-\frac{hr}{s(r^2+s^2)}\Big)=1+O\Big(\frac{Hr}{sq}\Big)=1+O\Big(\frac{r}{sx}\Big),$$

so that

$$\frac{1}{H}\sum_{0<|h|\leq H} f_h \sum_{\substack{\nu\,(q)\\ \nu^2+1\equiv 0\,(q)}} e_q(-h\nu) = \frac{1}{H}\sum_{0<|h|\leq H} f_h \sum_{\substack{r^2+s^2=q\\ r,s>0\\ (r,s)=1}} e_s(h\overline{r})\Big(1+O\Big(\frac{r}{sx}\Big)\Big)$$

$$= \frac{1}{H}\sum_{0<|h|\leq H} f_h \sum_{\substack{r^2+s^2=q\\ r,s>0\\ (r,s)=1}} e_s(h\overline{r}) + O\Big(\frac{\tau(q)r}{sx}\Big).$$

Thus, if we sum the error term in (6.2) weighted with some coefficients $\gamma(q)$ we get a sum of the form

$$\frac{1}{H}\sum_{0<|h|\leq H} f_h \sum_{\substack{r,s>0\\ (r,s)=1}} \gamma(r^2+s^2) e_s(h\overline{r}).$$

Assuming that we can use Poisson summation (5.7) to complete the sum over $r$ we end up with Kloosterman sums $S(t,h;s)$, where $t$ is the frequency variable from the second Poisson summation.

In practice the coefficients $\gamma$ are quite complicated, and we arrive at weighted averages of Kloosterman sums of the form

$$\sum_{\substack{m,n,\varrho,s\\ (s,\varrho)=1}} A_m B_{n,\varrho} S(m\overline{\varrho},n;s) \tag{6.4}$$

for Type I sums, and

$$\sum_{\substack{m,n,\rho,s\\ (s,\rho)=1}} A_{m,\varrho} B_{n,\varrho} S(m\overline{\varrho},n;s) \tag{6.5}$$

for Type II sums, for some bounded weights $A_m$ and $B_n$. A key aspect here is that the summation over $s$ is smoothly weighted, running over some range, say, $s\sim S$.

The classical Weil bound [59] gives $|S(a,b,p)| \leq 2\sqrt{p}$ for $\gcd(a,b,p) = 1$, which is in general optimal. This implies the point-wise bound $|S(a,b,c)| \ll c^{1/2+\epsilon}$ for $\gcd(a,b,c) = 1$, which is what Hooley [35] used to bound (6.4). However, we expect much additional cancellation from the signs of the Kloosterman sums (which are real valued) in the sum over $s$

$$\sum_{\substack{s\sim S\\ (s,\varrho)=1}} S(m\overline{\varrho},n;s).$$

Powerful methods from the theory of automorphic forms (namely, the Kuznetsov trace formula, which is a kind of Poisson summation formula for summing over

$s$) allowed Deshouillers and Iwaniec [17, 18] to show cancellation in (6.4). Their argument makes use of the averages over $m, n, \rho$ as well. The methods of [18] also apply to the more complicated sums (6.5) with somewhat weaker bounds. Both of these can be improved assuming Selberg's eigenvalue conjecture [16], and the Kim-Sarnak [43] result provides an approximation to this, which allowed de la Bretèche and Drappeau [16] to improve on [17].

## 6.3 Type II sums

In this section we sketch the proof of Proposition 7(i). To simplify the presentation we assume that the sum is restricted to $(m, n) = 1$. Using (6.2) we extract the correct main term, so that to control the error term we need the bound

$$\Sigma(M,N) := \frac{1}{H} \sum_{1 \le |h| \le H} f_h \sum_{\substack{m \sim M \\ n \sim N \\ (m,n)=1}} \alpha(m)\beta(n) \sum_{\substack{\nu\ (mn) \\ \nu^2+1 \equiv 0\ (mn)}} e_{mn}(-h\nu) \ll x^{1-\eta}.$$

Notice that the trivial bound for this sum is $MN = P = x^{\alpha}$ for $\alpha > 1$, and we need a bound $\ll x^{1-\eta}$. That is, for large $\alpha$ we need to save a large power of $x$. This is the main reason why we can handle the Type II sums only in the range $\alpha \le 1.19$. The fact that this range is even so large is a testament to the power of the automorphic methods.

Before introducing Kloosterman sums we want to apply Cauchy-Schwarz to smoothen the coefficient $\alpha(m)$ (akin to the argument in Section 5.2). We would like to simplify the argument greatly by pulling the sum over $\nu\,(mn)$ outside while still keeping the sum over $n$ inside. To this end, define

$$Q := \prod_{\substack{p \le 2N \\ p \nmid m \\ p \equiv 1,2\,(4)}} p,$$

so that $n|Q$, since $\beta(n)$ is supported on square-free numbers. By the Chinese remainder theorem every solution to $\nu^2 + 1 \equiv 0\,(mn)$ lifts up to exactly $\rho(Q)/\rho(n)$ solutions modulo $mQ$, so that

$$\sum_{\substack{\nu\ (mn) \\ \nu^2+1 \equiv 0\ (mn)}} e_{mn}(-h\nu) = \frac{\rho(n)}{\rho(Q)} \sum_{\substack{\nu\ (mQ) \\ \nu^2+1 \equiv 0\ (mQ)}} e_{mn}(-h\nu).$$

Therefore, we have

$$\Sigma(M,N) = \sum_{m \sim M} \alpha(m) \frac{1}{\rho(Q)} \sum_{\substack{\nu\ (mQ) \\ \nu^2+1 \equiv 0\ (mQ)}} \frac{1}{H} \sum_{1 \le |h| \le H} f_h \sum_{\substack{n \sim N \\ (m,n)=1}} \beta(n)\rho(n) e_{mn}(-h\nu).$$

We would like to take a moment to point out how counter-intuitive this step is, which is why it took me several months to find it. After all, when we pass on to the Kloosterman sums the most crucial parameter is the size of the modulus $mn$, and we have just made it exponentially larger!

To maintain the suspense, we keep calm and carry on applying Cauchy-Schwarz to find

$$\Sigma(M,N) \ll M^{1/2}\Bigg(\sum_{m\sim M}\frac{1}{H^2}\sum_{1\le|h_1|,|h_2|\le H} f_{h_1}\overline{f_{h_2}} \sum_{\substack{n_1,n_2\sim N\\(m,n_1n_2)=1}} \beta(n_1)\overline{\beta(n_2)} \frac{\rho(n_1)\rho(n_2)}{\rho(Q)}\sum_{\nu^2+1\equiv 0\ (mQ)} e_{mn_1}(-h_1\nu)e_{mn_2}(h_2\nu)\Bigg)^{1/2}.$$

As usual, in the diagonal part ($h_1 = h_2, n_1 = n_2$) we are forced to use a trivial bound, which sets the limit $N \gg x^{\alpha-1+\eta}$ in Proposition 7. For simplicity suppose that in the remaining part we have $(n_1,n_2) = 1$. Then $e_{mn_1}(-h_1\nu)e_{mn_2}(h_2\nu) = e_{mn_1n_2}((h_2n_1 - h_1n_2)\nu)$, and we get

$$\begin{aligned}\frac{\rho(n_1)\rho(n_2)}{\rho(Q)}&\sum_{\nu^2+1\equiv 0\ (mQ)} e_{mn_1n_2}((h_2n_1-h_1n_2)\nu)\\ &= \frac{\rho(n_1n_2)}{\rho(Q)}\sum_{\nu^2+1\equiv 0\ (mQ)} e_{mn_1n_2}((h_2n_1-h_1n_2)\nu)\\ &= \sum_{\nu^2+1\equiv 0\ (mn_1n_2)} e_{mn_1n_2}((h_2n_1-h_1n_2)\nu)\end{aligned}$$

by applying the Chinese remainder theorem to collapse the sum. Hence, the off-diagonal part is

$$\frac{1}{H^2}\sum_{1\le|h_1|,|h_2|\le H} f_{h_1}\overline{f_{h_2}}\sum_{\substack{n_1,n_2\sim N\\(n_1,n_2)=1}}\beta(n_1)\overline{\beta(n_2)} \sum_{\substack{m\sim M\\(m,n_1n_2)=1}}\sum_{\nu^2+1\equiv 0\ (mn_1n_2)} e_{mn_1n_2}((h_2n_1-h_1n_2)\nu).$$

The argument from here on follows the outline of the previous section, that is, we introduce Kloosterman sums, which results in sums of the form (6.5) (with $\varrho = n_1n_2$, $n = h_1n_2 - h_2n_1$, and $q = r^2 + s^2 \equiv 0\,(n_1n_2)$ for $q \sim MN^2$), and then bound these using the estimates in [18]. This works provided that $N \ll x^{(2-2\theta-\alpha)/3-\eta}$, completing the sketch of the proof of Proposition 7.

# References


[1] S. Baier and L. Zhao. Bombieri-Vinogradov type theorems for sparse sets of moduli. *Acta Arith.*, 125(2):187–201, 2006.

[2] S. Baier and L. Zhao. An improvement for the large sieve for square moduli. *J. Number Theory*, 128(1):154–174, 2008.

[3] R. Baker. Primes in arithmetic progressions to spaced moduli. III. *Acta Arith.*, 179(2):125–132, 2017.

[4] R. Baker and T. Freiberg. Limit points and long gaps between primes. *Q. J. Math.*, 67(2):233–260, 2016.

[5] R. Baker and G. Harman. Numbers with a large prime factor. II. In *Analytic number theory*, pages 1–14. Cambridge Univ. Press, Cambridge, 2009.

[6] R. C. Baker, G. Harman, and J. Pintz. The difference between consecutive primes. II. *Proc. London Math. Soc. (3)*, 83(3):532–562, 2001.

[7] A. Balog. Numbers with a large prime factor. *Studia Sci. Math. Hungar.*, 15(1-3):139–146, 1980.

[8] A. Balog. Numbers with a large prime factor. II. In *Topics in classical number theory, Vol. I, II (Budapest, 1981)*, volume 34 of *Colloq. Math. Soc. János Bolyai*, pages 49–67. North-Holland, Amsterdam, 1984.

[9] A. Balog, G. Harman, and J. Pintz. Numbers with a large prime factor. III. *Quart. J. Math. Oxford Ser. (2)*, 34(134):133–140, 1983.

[10] A. Balog, G. Harman, and J. Pintz. Numbers with a large prime factor. IV. *J. London Math. Soc. (2)*, 28(2):218–226, 1983.

[11] W. D. Banks, T. Freiberg, and J. Maynard. On limit points of the sequence of normalized prime gaps. *Proc. Lond. Math. Soc. (3)*, 113(4):515–539, 2016.

[12] E. Bombieri, J. B. Friedlander, and H. Iwaniec. Primes in arithmetic progressions to large moduli. *Acta Math.*, 156(3-4):203–251, 1986.

[13] J. R. Chen. On the Goldbach's problem and the sieve methods. *Sci. Sinica*, 21(6):701–739, 1978.

[14] T. Cochrane and Z. Zheng. Exponential sums with rational function entries. *Acta Arith.*, 95(1):67–95, 2000.

[15] H. Davenport. *Multiplicative number theory*, volume 74 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 2000. Revised and with a preface by Hugh L. Montgomery.

[16] R. de la Bretèche and S. Drappeau. Niveau de répartition des polynômes quadratiques et crible majorant pour les entiers friables. *J. Eur. Math. Soc. (JEMS)*, 22(5):1577–1624, 2020.

[17] J.-M. Deshouillers and H. Iwaniec. On the greatest prime factor of $n^2+1$. *Ann. Inst. Fourier (Grenoble)*, 32(4):1–11 (1983), 1982.

[18] J.-M. Deshouillers and H. Iwaniec. Kloosterman sums and Fourier coefficients of cusp forms. *Invent. Math.*, 70(2):219–288, 1982/83.

[19] W. Duke, J. B. Friedlander, and H. Iwaniec. Equidistribution of roots of a quadratic congruence to prime moduli. *Ann. of Math. (2)*, 141(2):423–441, 1995.

[20] P. Erdős. On the difference of consecutive primes. *Q. J. Math. Oxford Ser.*, 1935.

[21] P. Erdős. Some problems on the distribution of prime numbers. *C. I. M. E. Teoria dei numeri, Math. Congr. Varenna, 1954.*, 1955.

[22] E. Fouvry and H. Iwaniec. Primes in arithmetic progressions. *Acta Arith.*, 42(2):197–218, 1983.

[23] J. Friedlander and H. Iwaniec. *Opera de cribro*, volume 57 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2010.

[24] J. B. Friedlander and H. Iwaniec. Incomplete Kloosterman sums and a divisor problem. *Ann. of Math. (2)*, 121(2):319–350, 1985. With an appendix by Bryan J. Birch and Enrico Bombieri.

[25] D. A. Goldston and A. H. Ledoan. Limit points of the sequence of normalized differences between consecutive prime numbers. In *Analytic number theory*, pages 115–125. Springer, Cham, 2015.

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

[27] A. Granville. Harald Cramér and the distribution of prime numbers. Number 1, pages 12–28. 1995. Harald Cramér Symposium (Stockholm, 1993).

[28] G. Harman. On the distribution of $\alpha p$ modulo one. *J. London Math. Soc. (2)*, 27(1):9–18, 1983.

[29] G. Harman. *Prime-detecting sieves*, volume 33 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2007.

[30] D. R. Heath-Brown. The number of primes in a short interval. *J. Reine Angew. Math.*, 389:22–63, 1988.

[31] D. R. Heath-Brown. The largest prime factor of the integers in an interval. *Sci. China Ser. A*, 39(5):449–476, 1996.

[32] D. R. Heath-Brown. The largest prime factor of $X^3 + 2$. *Proc. London Math. Soc. (3)*, 82(3):554–596, 2001.

[33] D. R. Heath-Brown and H. Iwaniec. On the difference between consecutive primes. *Invent. Math.*, 55(1):49–69, 1979.

[34] D. R. Heath-Brown and C. Jia. The largest prime factor of the integers in an interval. II. *J. Reine Angew. Math.*, 498:35–59, 1998.

[35] C. Hooley. On the greatest prime factor of a quadratic polynomial. *Acta Math.*, 117:281–299, 1967.

[36] M. N. Huxley. On the difference between consecutive primes. *Invent. Math.*, 15:164–170, 1972.

[37] H. Iwaniec. *Topics in classical automorphic forms*, volume 17 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1997.

[38] H. Iwaniec. *Spectral methods of automorphic forms*, volume 53 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI; Revista Matemática Iberoamericana, Madrid, second edition, 2002.

[39] H. Iwaniec and E. Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004.

[40] C. Jia. Almost all short intervals containing prime numbers. *Acta Arith.*, 76(1):21–84, 1996.

[41] C. Jia and M.-C. Liu. On the largest prime factor of integers. *Acta Arith.*, 95(1):17–48, 2000.

[42] M. Jutila. On numbers with a large prime factor. *J. Indian Math. Soc. (N.S.)*, 37:43–53 (1974), 1973.

[43] H. H. Kim. Functoriality for the exterior square of $\mathrm{GL}_4$ and the symmetric fourth of $\mathrm{GL}_2$. *J. Amer. Math. Soc.*, 16(1):139–183, 2003. With appendix 1 by Dinakar Ramakrishnan and appendix 2 by Kim and Peter Sarnak.

[44] H. D. Kloosterman. On the representation of numbers in the form $ax^2 + by^2 + cz^2 + dt^2$. *Acta Math.*, 49(3-4):407–464, 1927.

[45] K. Matomäki. A note on primes of the form $p = aq^2 + 1$. *Acta Arith.*, 137(2):133–137, 2009.

[46] K. Matomäki. Another note on smooth numbers in short intervals. *Int. J. Number Theory*, 12(2):323–340, 2016.

[47] K. Matomäki and M. Radziwiłł. Multiplicative functions in short intervals. *Ann. of Math. (2)*, 183(3):1015–1056, 2016.

[48] J. Maynard. Small gaps between primes. *Ann. of Math. (2)*, 181(1):383–413, 2015.

[49] J. Maynard. Primes in arithmetic progressions to large moduli I: Fixed residue classes. URL: `https://arxiv.org/abs/2006.06572`. *Preprint*, 2020.

[50] C. B. Pan. On the upper bound of the number of ways to represent an even integer as a sum of two primes. *Sci. Sinica*, 23(11):1368–1377, 1980.

[51] J. Pintz. Polignac numbers, conjectures of Erdös on gaps between primes, arithmetic progressions in primes, and the bounded gap conjecture. In *From arithmetic to zeta-functions*, pages 367–384. Springer, [Cham], 2016.

[52] J. Pintz. A note on the distribution of normalized prime gaps. *Acta Arith.*, 184(4):413–418, 2018.

[53] D. H. J. Polymath. New equidistribution estimates of Zhang type. *Algebra Number Theory*, 8(9):2067–2199, 2014.

[54] D. H. J. Polymath. Variants of the Selberg sieve, and bounded intervals containing many primes. *Res. Math. Sci.*, 1:Art. 12, 83, 2014.

[55] R. A. Rankin. The Difference between Consecutive Prime Numbers. *J. London Math. Soc.*, 11(4):242–245, 1936.

[56] G. Ricci. Recherches sur l'allure de la suite $\{p_{n+1} - p_n/\log p_n\}$. In *Colloque sur la Théorie des Nombres, Bruxelles, 1955*, pages 93–106. Georges Thone, Liège; Masson and Cie, Paris, 1956.

[57] T. Tao. 254a, supplement 5: The linear sieve and Chen's theorem (optional), 2015. URL (Accessed 25.8.2020): `https://terrytao.wordpress.com/2015/01/29/254a-supplement-5-the-linear-sieve-and-chens-theorem-optional/`.

[58] R.-C. Vaughan. Sommes trigonométriques sur les nombres premiers. *C. R. Acad. Sci. Paris Sér. A-B*, 285(16):A981–A983, 1977.

[59] A. Weil. On some exponential sums. *Proc. Nat. Acad. Sci. U.S.A.*, 34:204–207, 1948.

[60] E. Westzynthius. Über die Verteilung der Zahlen, die zu den n ersten Primzahlen teilerfremd sind. *Commentat. Phys.-Math.*, 5(25):1–37, 1931.

[61] J. Wu. Chen's double sieve, Goldbach's conjecture and the twin prime problem. *Acta Arith.*, 114(3):215–273, 2004.

[62] Y. Zhang. Bounded gaps between primes. *Ann. of Math. (2)*, 179(3):1121–1174, 2014.

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama Oy, Turku, Finland 2021

ISBN 978-951-29-8410-7 (PRINT)
ISBN 978-951-29-8411-4 (PDF)
ISSN 0082-7002 (Print)
ISSN 2343-3175 (Online)