BOUNDED GAPS BETWEEN PRIMES

ANDREW GRANVILLE

ABSTRACT. Recently, Yitang Zhang proved the existence of a finite bound B such that there are infinitely many pairs p_n, p_{n+1} of consecutive primes for which $p_{n+1} - p_n \leqslant B$. This can be seen as a massive breakthrough on the subject of twin primes and other delicate questions about prime numbers that had previously seemed intractable. In this article we will discuss Zhang's extraordinary work, putting it in its context in analytic number theory, and sketch a proof of his theorem.

Zhang even proved the result with $B = 70\,000\,000$. A co-operative team, *polymath8*, collaborating only on-line, has been able to lower the value of B to 4680, and it seems plausible that these techniques can be pushed somewhat further, though the limit of these methods seem, for now, to be $B = 12$.

CONTENTS

1991 *Mathematics Subject Classification.* 11P32.

To Yiliang Zhang, for showing that one can, no matter what.

sec:intro

1. Introduction

1.1. **Intriguing questions about primes.** Early on in our mathematical education we get used to the two basic rules of arithmetic, addition and multiplication. When we define a prime number, simply in terms of the number's multiplicative properties, we discover a strange and magical sequence of numbers. On the one hand, so easily defined, on the other, so difficult to get a firm grasp of, since they are defined in terms of what they are not (i.e. that they *cannot* be factored into two smaller integers)).

When one writes down the sequence of prime numbers:

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, \ldots$$

one sees that they occur frequently, but it took a rather clever construction of the ancient Greeks to even establish that there really are infinitely many. Looking further at a list of primes, some patterns begin to emerge; for example, one sees that they often come in pairs:

$$3 \text{ and } 5,\ 5 \text{ and } 7,\ 11 \text{ and } 13,\ 17 \text{ and } 19,\ 29 \text{ and } 31,\ 41 \text{ and } 43,\ 59 \text{ and } 61, \ldots$$

One might guess that there are infinitely many such prime pairs. But this is an open, elusive question, the *twin prime conjecture.* Until recently there was little theoretical evidence for it. All that one could say is that there was an enormous amount of computational evidence that these pairs never quit; and that this conjecture (and various more refined versions) fit into an enormous network of *conjecture*, which build a beautiful elegant structure of all sorts of prime patterns; and if the twin prime conjecture were to be false then the whole edifice would crumble.

The twin prime conjecture is certainly intriguing to both amateur and professional mathematicians alike, though one might argue that it is an artificial question, since it asks for a very delicate additive property of a sequence defined by its multiplicative properties. Indeed, number theorists had struggled, until very recently, to identify an approach to this question that seemed likely to make any significant headway. In this article we will discuss these latest shocking developments. In the first few sections we will take a leisurely stroll through the historical and mathematical background, so as to give the reader a sense of the great theorem that has been recently proved, and also from a perspective that will prepare the reader for the details of the proof.

1.2. **Other patterns.** Looking at the list of primes above we see other patterns that begin to emerge, for example, one can find four primes which have all the same digits, except the last one:

$$11, 13, 17 \ \text{ and } 19, \ \text{which is repeated with } 101, 103, 107 \ \text{ and } 109,$$

and one can find many more such examples – are there infinitely many? More simply how about prime pairs with difference 4,

$$3 \text{ and } 7,\ 7 \text{ and } 11,\ 13 \text{ and } 17,\ 19 \text{ and } 23,\ 37 \text{ and } 41,\ 43 \text{ and } 47,\ 67 \text{ and } 71, \ldots;$$

or difference 10,

$$3 \text{ and } 13,\ 7 \text{ and } 17,\ 13 \text{ and } 23,\ 19 \text{ and } 29,\ 31 \text{ and } 41,\ 37 \text{ and } 47,\ 43 \text{ and } 53, \ldots?$$

Are there infinitely many such pairs? Such questions were probably asked back to antiquity, but the first clear mention of twin primes in the literature appears in a paper of de Polignac from 1849. In his honour we now call any integer h, for which there are infinitely many prime pairs $p, p+h$, a *de Polignac number.*[1]

Then there are the *Sophie Germain pairs*, primes p and $q := 2p+1$, which prove useful in several simple algebraic constructions:[2]

$$2 \text{ and } 5,\ 3 \text{ and } 7,\ 5 \text{ and } 11,\ 11 \text{ and } 23,\ 23 \text{ and } 47,\ 29 \text{ and } 59,\ 41 \text{ and } 83, \dots;$$

Now we have spotted all sorts of patterns, we need to ask ourselves whether there is a way of predicting which patterns can occur and which do not. Let's start by looking at the possible differences between primes: It is obvious that there are not infinitely many prime pairs of difference 1, because one of any two consecutive integers must be even, and hence can only be prime if it equals 2. Thus there is just the one pair, 2 and 3, of primes with difference 1. One can make a similar argument for prime pairs with odd difference. Hence if h is an integer for which there are infinitely many prime pairs of the form $p,\ q = p + h$ then h must be even. We have seen many examples, above, for each of $h = 2,\ h = 4$ and $h = 10$, and the reader can similarly construct lists of examples for $h = 6$ and for $h = 8$, and indeed for any other even h that takes her or his fancy. This leads us to bet on the *generalized twin prime conjecture*, which states that for any even integer $2k$ there are infinitely many prime pairs $p,\ q = p + 2k$.

What about prime triples? or quadruples? We saw two examples of prime quadruples of the form $10n+1,\ 10n+3,\ 10n+7,\ 10n+9$, and believe that there are infinitely many. What about other patterns? Evidently any pattern that includes an odd difference cannot succeed. Are there any other obstructions? The simplest pattern that avoids an odd difference is $n, n+2, n+4$. One finds the one example $3,\ 5,\ 7$ of such a prime triple, but no others. Further examination makes it clear why not: One of the three numbers is always divisible by 3. This is very similar to what happened with $n, n+1$; and one can verify that, similarly, one of $n, n+6, n+12, n+18, n+24$ is always divisible by 5. The general obstruction can be described as follows:

For a given set of distinct integers $a_1 < a_2 < \ldots < a_k$ we say that prime p is an *obstruction* if p divides at least one of $n+a_1, \ldots, n+a_k$, for every integer n. In other words, p divides

$$\mathcal{P}(n) = (n+a_1)(n+a_2)\ldots(n+a_k)$$

for every integer n; which can be classified by the condition that the set $a_1, a_2, \ldots, a_k \pmod p$ includes all of the residue classes mod p. If no prime is an obstruction then we say that $x+a_1, \ldots, x+a_k$ is an *admissible* set of forms.[3].

[1] Pintz makes a slightly definition: That is, that p and $p+h$ should be consecutive primes.

[2] These are useful because, in this case, the group of reduced residues mod q is a cyclic group of order $q-1 = 2p$, and therefore isomorphic to $C_2 \times C_p$ if $p > 2$. Therefore every element in the group has order 1 (that is, $1 \pmod q$), 2 (that is, $-1 \pmod q$), p (the squares mod q) or $2p = q-1$. Hence $g \pmod q$ generates the group of reduced residues if and only if g is not a square mod q and $g \not\equiv -1 \pmod q$.

[3] Notice that $a_1, a_2, \ldots, a_k \pmod p$ can occupy no more than k residue classes mod p and so, if $p > k$ then p cannot be an obstruction.

Number theorists have long made the optimistic conjecture if there is no such "obvious" obstruction to a set of linear forms being infinitely often prime, then they are infinitely often simultaneously prime. That is:

Conjecture: *If $x + a_1, \dots, x + a_k$ is an admissible set of forms then there are infinitely many integers n such that $n + a_1, \dots, n + a_k$ are all prime numbers.*

In this case, we call $n + a_1, \dots, n + a_k$ a *k-tuple* of prime numbers.

To date, this has not been proven for any $k > 1$ though, following Zhang's work, we are starting to get close for $k = 2$. Indeed, Zhang proves a weak variant of this conjecture, as we shall see.

The above conjecture can be extended, as is, to all sets of k linear forms with integer coefficients in one variable, so long as we extend the notion of admissibility to also exclude the obstruction that two of the linear forms have different signs for all, but finitely many, n, since a negative integer cannot be prime (for example, n and $2 - n$); some people call this the "obstruction at the 'prime', -1". We can also extend the conjecture to more than one variable (for example the set of forms $m, m + n, m + 4n$):

The prime k-tuplets conjecture: *If a set of k linear forms in n variables is admissible then there are infinitely many sets of n integers such that when we substitute these integers into the forms we get a k-tuple of prime numbers.*

There has been substantial recent progress on this conjecture. The famous breakthrough was Green and Tao's theorem for the k-tuple of linear forms in the two variables a and d:

$$a,\ a + d,\ a + 2d, \dots,\ a + (k - 1)d.$$

Along with Ziegler, they went on to prove the prime k-tuplets conjecture for any admissible set of linear forms, provided no two satisfy a linear equation over the integers. What a remarkable theorem! Unfortunately these exceptions include many of the questions we are most interested in; for example, $p,\ q = p + 2$ satisfy the linear equation $q - p = 2$; and $p,\ q = 2p + 1$ satisfy the linear equation $q - 2p = 1$).

Finally, we also believe that the conjecture holds if we consider any admissible set of k irreducible polynomials with integer coefficients, with any number of variables. For example we believe that $n^2 + 1$ is infinitely often prime, and that there are infinitely many prime triples $m,\ n,\ m^2 + 2n^2$.

We will end this section by stating Zhang's main theorem and a few of the more beguiling consequences:

Zhang's main theorem: *There exists an integer k such that the following is true: If $x + a_1, \dots, x + a_k$ is an admissible set of forms then there are infinitely many integers n such that* at least two of *$n + a_1, \dots, n + a_k$ are prime numbers.*

Note that the result states that only two of the $n + a_i$ are prime, not all (as would be required in the prime k-tuplets conjecture). Zhang proved this result for a fairly large value of k, that is $k = 3500000$, which has been reduced to $k = 632$ by the polymath8 team. Of course if one could take $k = 2$ then we would have the twin prime conjecture, but the most optimistic plan at the moment, along the lines of Zhang's proof, would yield $k = 5$.

To deduce that there are bounded gaps between primes from Zhang's Theorem we need only show the existence of an admissible set with k elements. This is not difficult, simply by letting the a_i be the first k primes $> k$.[4] Hence we have proved:

Corollary: [Bounded gaps between primes] *There exists a bound B such that there are infinitely many integers pairs of prime numbers $p < q < p + B$.*

Finding the smallest B for a given k is a challenging question. The prime number theorem together with our construction above suggests that $B \leqslant k(\log k + C)$ for some constant C, but it is interesting to get better bounds.

Our Corollary further implies

Corollary: *There is an integer $h, 0 < h \leqslant B$ such that there are infinitely many pairs of primes $p, p + h$.*

That is, some positive integer $\leqslant B$ is a de Polignac number. In fact one can go a little further using Zhang's main theorem:

Corollary: *Let k be as in Zhang's Theorem, and let A be* any *admissible set of k integers. There is an integer $h \in (A - A)^+ := \{a - b : \ a > b \in A\}$ such that there are infinitely many pairs of primes $p, p + h$.*

Finally we can deduce from this

Corollary: *A positive proportion of integers are de Polignac numbers*

Proof. If $A \subset \{0, \dots, B\}$ is an admissible set then $mA := \{ma : a \in A\}$ is admissible for every integer $m \geqslant 1$. Given large x let $M = [x/B]$. By Zhang's Theorem there exists a pair $a_m < b_m \in A$ such that $m(b_m - a_m)$ is a de Polgnac number. Since there are at most $B/2$ differences $d = b - a$ with $a < b \in A$ there must be some difference which is the value of $b_m - a_m$ for at least $2M/B$ values of $m \leqslant M$. This gives rise to $\geqslant 2M/B \geqslant x/B^2$ distinct de Polignac numbers of the form $md \leqslant x$. □

Our construction above implies that the proportion is at least $1/k^2(\log k + C)^2$.

[4] This is admissible since none of the a_i is $0 \pmod p$ for any $p \leqslant k$, and the $p > k$ were handled in the previous footnote.

1.3. **The simplest analytic approach.** There are 14 odd primes up to 50, that is 14 out of the 25 odd integers up to 50, so one can deduce that several pairs differ by 2. We might hope to take this kind of density approach more generally: If A is a sequence of integers of density 1/2 (in all of the integers) then we can easily deduce that there are many pairs of elements of A that differ by *no more than 2*. One might guess that there are pairs that differ by exactly 2, but this is by no means guaranteed, as the example $A := \{n \in \mathbb{Z} : \ n \equiv 1 \text{ or } 2 \pmod 4\}$ shows. Moreover, to use this kind of reasoning to hunt for twin primes, we presumably need a lower bound on the density of primes as one looks at larger and larger primes. This was something that intrigued the young Gauss who, by examining Chernik's table of primes up to one million, surmised that "the density of primes at around x is roughly $1/\log x$" (and this was subsequently verified, as a consequence of the *prime number theorem*). Therefore we are guaranteed that there are infinitely many pairs of primes $p < q$ with $q - p \leqslant \log p$, which is not quite as small a gap as we are hoping for! Nonetheless this raises the question: Fix $c > 0$. Can we prove that

There are infinitely many pairs of primes $p < q$ with $q < p + c\log p$?

This follows for all $c \geqslant 1$ by the prime number theorem, but it is not easy to prove such a result for any particular value of $c < 1$. The first such results were proved conditionally assuming the Generalized Riemann Hypothesis. This is, in itself, surprising: The Generalized Riemann Hypothesis was formulated to better understand the distribution of primes in arithmetic progressions, so why would it appear in an argument about short gaps between primes? It is far from obvious by the argument used, and yet this connection has deepened and broadened as the literature developed. We will discuss primes in arithmetic progressions in detail in the next section.

The first unconditional (though inexplicit) such result, bounding gaps between primes, was proved by Erdős in 1940 using the small sieve (we will obtain any $c > e^{-\gamma} \approx 0.5614$ by such a method in section 3.2). In 1966, Bombieri and Davenport [2] substituted the Bombieri-Vinogradov theorem for the Generalized Riemann Hypothesis in earlier, conditional arguments, to prove this unconditionally for any $c \geqslant \frac{1}{2}$; and in 1988 Maier [25] observed that one can easily modify this to obtain any $c \geqslant \frac{1}{2}e^{-\gamma}$. The Bombieri-Vinogradov Theorem is also a result about primes in arithmetic progressions, as we will discuss later. Maier further improved this, by combining the approaches of Erdős and of Bombieri and Davenport, to some bound a little smaller than $\frac{1}{4}$, with substantial effort.

The first big breakthrough occurred in 2005 when Goldston, Pintz and Yildirim [15] were able to show that there are infinitely many pairs of primes $p < q$ with $q < p + c\log p$, for *any* given $c > 0$. Indeed they extended their methods to show that, for any $\epsilon > 0$, there are infinitely many pairs of primes $p < q$ for which

$$q - p < (\log p)^{1/2+\epsilon}.$$

It is their method which forms the basis of the discussion in this paper. Like Bombieri and Davenport, they showed that one can could better understand small gaps between primes, by obtaining strong estimates on primes in arithmetic progressions, as in the

Bombieri-Vinogradov Theorem. Even more, if one assumes a strong, but widely believed, conjecture about the equi-distribution of primes in arithmetic progressions, which extends the Bombieri-Vinogradov Theorem, then one can show that there are infinitely many pairs of primes $p < q$ which differ by no more than 16 (that is, $p < q \leqslant p + 16$)! What an extraordinary statement, and one that we will briefly discuss: We know that if $p < q \leqslant p + 16$ then $q - p = 2,\ 4,\ 6,\ 8,\ 10,\ 12,\ 14$ or 16, and so at least one of these difference occurs infinitely often. That is, there exists a positive, even integer $2k \leqslant 16$ such that there are infinitely pairs of primes $p,\ p + 2k$. Very recently this has been refined further by James Maynard, improving the upper bound to 12, by a variant of the original argument.

After Goldston, Pintz and Yildirim, most of the experts tried and failed to obtain enough of an improvement of the Bombieri-Vinogradov Theorem to deduce the existence of some finite bound B such that there are infinitely many pairs of primes that differ by no more than B. To improve the Bombieri-Vinogradov Theorem is no mean feat and people have longed discussed "barriers" to obtaining such improvements. In fact a technique had been developed by Fouvry fouvry [10], and by Bombieri, Friedlander and Iwaniec bfi [3], but these were neither powerful enough nor general enough to work in this circumstance.

Enter Yitang Zhang, an unlikely figure to go so much further than the experts, and to find exactly the right improvement and refinement of the Bombieri-Vinogradov Theorem to establish the existence of the elusive bound B such that there are infinitely many pairs of primes that differ by no more than B. By all accounts, Zhang was a brilliant student in Beijing from 1978 to the mid-80s, finishing with a master's degree, and then working on the Jacobian conjecture for his Ph.D. at Purdue, graduating in 1992. He did not proceed to a job in academia, working in odd jobs, such as in a sandwich shop, at a motel and as a delivery worker. Finally in 1999 he got a job at the University of New Hampshire as a lecturer, with a high teaching load, working with many of the less qualified undergraduate students. From time-to-time a lecturer devotes their energy to working on proving great results, but few have done so with such aplomb as Zhang. Not only did he prove a great result, but he did so by improving *technically* on the experts, having important key ideas that they missed and developing a highly ingenious and elegant construction concerning exponential sums. Then, so as not to be rejected out of hand, he wrote his difficult paper up in such a clear manner that it could not be denied. Albert Einstein worked in a patent office, Yitang Zhang in a Subway sandwich shop; both found time, despite the unrelated calls on their time and energy, to think the deepest thoughts in science. Moreover Zhang did so at the relatively advanced age of 50 (or more). Truly *extraordinary.*

2. The distribution of primes, divisors and prime k-tuplets

2.1. **The prime number theorem.** As we mentioned in the previous section, Gauss observed, at the age of 16, that "the density of primes at around x is roughly $1/\log x$", which leads quite naturally to the conjecture that

$$\#\{\text{primes } p \leqslant x\} \approx \int_2^x \frac{dt}{\log t} \sim \frac{x}{\log x} \quad \text{as } x \to \infty.$$

(We use the symbol $A(x) \sim B(x)$ for two functions A and B of x, to mean that $A(x)/B(x) \to 1$ as $x \to \infty$.) This was proved in 1896, the *prime number theorem*, and the integral provides a considerably more precise approximation to the number of primes $\leqslant x$, than $x/\log x$. However, this integral is rather cumbersome to work with, and so it is natural to instead weight each prime with $\log p$; that is we work with

$$\theta(x) := \sum_{\substack{p \text{ prime} \\ p \leqslant x}} \log p$$

and the prime number theorem implies[5] that

$$\theta(x) \sim x \quad \text{as } x \to \infty. \tag{2.1}$$

pnt2

eveHeuristic

2.2. **A sieving heuristic to guess at the prime number theorem.** How many integers up to x have no prime factors $\leqslant y$? If $y \geqslant \sqrt{x}$ then this counts 1 and all of the primes between y and x, so an accurate answer would yield the prime number theorem.

The usual heuristic is to start by observing that there are $x/2 + O(1)$ integers up to x that are not divisible by 2. A proportion $\frac{2}{3}$rds of these remaining integers are not divisible by 3; then a proportion $\frac{4}{5}$ths of the remaining integers are not divisible by 5, etc. Hence we guess that the number of integers $\leqslant x$ which are free of prime factors $\leqslant y$, is roughly

$$\prod_{p \leqslant y} \left(1 - \frac{1}{p}\right) \cdot x.$$

Evaluating the product here is tricky but was accomplished by Mertens: If $y \to \infty$ then

$$\prod_{p \leqslant y} \left(1 - \frac{1}{p}\right) \sim \frac{e^{-\gamma}}{\log y}.$$

Here γ is the Euler-Mascheroni constant, defined as $\lim_{N \to \infty} \frac{1}{1} + \frac{1}{2} + \ldots + \frac{1}{N} - \log N$. There is no obvious explanation as to why this constant, defined in a very different context, appears here.

If $\sqrt{x} < y = o(x)$ (that is, for any fixed $\epsilon > 0$ we have $y \leqslant \epsilon x$ once x is sufficiently large) then we know from the prime number theorem that there are $\sim x/\log x$ integers left unsieved, whereas the prediction from our heuristic varies considerably as y varies in this range. This shows that the heuristic is wrong for large y. Taking $y = \sqrt{x}$ it

[5] This is really stating things backwards since, in proving the prime number theorem, it is significantly easier to include the $\log p$ weight, and then deduce estimates for the number of primes by partial summation.

predicts too many primes by a factor of $2e^{-\gamma}$; taking $y = x/\log x$ it predicts too few primes by a factor of $e^{-\gamma}$. In fact this heuristic gives an accurate estimate provided $y = x^{o(1)}$. We will exploit the difference between this heuristic and the correct count, to show that there are smaller than average gaps between primes in section MaierTrick 3.2.

2.3. **The prime number theorem for arithmetic progressions, I.** Any prime divisor of (a, q) is an obstruction to the primality of values of the polynomial $qx + a$, and these are the only such obstructions. The prime k-tuplets conjecture therefore implies that if $(a, q) = 1$ then there are infinitely many primes of the form $qn + a$. This was first proved by Dirichlet in 1837. Once proved one might ask for a more quantitative result. If we look at the primes in the arithmetic progressions $\pmod{10}$:

$$\begin{array}{l} 11,\ 31,\ 41,\ 61,\ 71,\ 101 \\ 3,\ 13,\ 23,\ 43,\ 53,\ 73,\ 83,\ 103 \\ 7,\ 17,\ 37,\ 47,\ 67,\ 97,\ 107 \\ 19,\ 29,\ 59,\ 79,\ 89,\ 109 \end{array}$$

then there seem to be roughly equal numbers in each, and this pattern persists as we look further out. Let $\phi(q)$ denote the number of $a \pmod q$ for which $(a, q) = 1$, and so we expect that

$$\theta(x; q, a) := \sum_{\substack{p \text{ prime} \\ p \leqslant x \\ p \equiv a \pmod q}} \log p \sim \frac{x}{\phi(q)} \quad \text{as } x \to \infty.$$

This is the *prime number theorem for arithmetic progressions* and was first proved by suitably modifying the proof of the prime number theorem.

The function $\phi(q)$ was studied by Euler, who showed that it is *multiplicative*, that is

$$\phi(q) = \prod_{p^e \| q} \phi(p^e)$$

(where $p^e \| q$ means that p^e is the highest power of prime p dividing q) and that $\phi(p^e) = p^e - p^{e-1}$ for all $e \geqslant 1$.

2.4. **Dirichlet's divisor trick.** Another multiplicative function of importance is the divisor function

$$\tau(n) := \sum_{d|n} 1$$

where the sum is over the positive integers d that divide n. It is not difficult to verify that $\tau(p^e) = e + 1$.

If n is squarefree and has k prime factors then $\tau(n) = 2^k$, so we see that $\tau(n)$ varies greatly depending on the arithmetic structure of n. Nonetheless one might ask for the average of $\tau(n)$, that is the average number of divisors of a positive integer $\leqslant x$. A first approach yields that

$$\sum_{n\leqslant x}\tau(n)=\sum_{n\leqslant x}\sum_{d|n}1=\sum_{d|n}\sum_{\substack{n\leqslant x\\ d|n}}1=\sum_{d\leqslant x}\left[\frac{x}{d}\right],$$

since the positive integers up to x that are divisible by d can be written as dm with $m\leqslant x/d$, and so there are $[x/d]$ such integers, where $[t]$ denotes the largest integer $\leqslant t$. It evident that $[t]=t+O(1)$, where $O(1)$ signifies that there is a correction here of at most a bounded multiple of 1. If we substitute this approximation in above, we obtain

$$\frac{1}{x}\sum_{n\leqslant x}\tau(n)=\frac{1}{x}\sum_{d\leqslant x}\left(\frac{x}{d}+O(1)\right)=\sum_{d\leqslant x}\frac{1}{d}+O\left(\frac{1}{x}\sum_{d\leqslant x}1\right)$$

One can approximate $\sum_{d\leqslant x}\frac{1}{d}$ by $\int_1^x dt/t=\log x$. Indeed the difference tends to a limit, the Euler-Mascheroni constant $\gamma:=\lim_{N\to\infty}\frac{1}{1}+\frac{1}{2}+\ldots+\frac{1}{N}-\log N$. Hence we have proved that the integers up to x have $\log x+O(1)$ divisors, on average, which is quite remarkable for such a wildly fluctuating function.

Dirichlet studied this argument and noticed that when we approximate $[x/d]$ by $x/d+O(1)$ for large d, say for those d in $(x/2,x]$, then this is not really a very good approximation, and gives a large cumulative error term, $O(x)$. However we know that $[x/d]=1$ exactly, for each of these d, and so we can estimate this sum by $x/2+O(1)$, which is much more precise. Dirichlet realized that the correct way to formulate this observation is to write $n=dm$, where d and m are integers. When d is small then we should fix d, and count the number of such m, with $m\leqslant x/d$ (as we did above); but when m is small, then we should fix m, and count the number of d with $d\leqslant x/m$. In this way our sums are all over long intervals, which allows us to get an accurate approximation of their value. In fact we can exploit the symmetry here to simply "break the sum" at $x^{1/2}$. Hence Dirichlet proceeded as follows:

$$\begin{aligned}\sum_{n\leqslant x}\tau(n)&=\sum_{n\leqslant x}\sum_{dm=n}1=\sum_{d\leqslant\sqrt{x}}\sum_{\substack{n\leqslant x\\ d|n}}1+\sum_{m<\sqrt{x}}\sum_{\substack{n\leqslant x\\ m|n}}1-\sum_{d\leqslant\sqrt{x}}\sum_{m<\sqrt{x}}1\\ &=\sum_{d\leqslant\sqrt{x}}\left(\frac{x}{d}+O(1)\right)+\sum_{m<\sqrt{x}}\left(\frac{x}{m}+O(1)\right)-x+O(\sqrt{x}).\end{aligned}$$

One can do even better with these sums than above, showing that $\sum_{n\leqslant N}1/n=\log N+\gamma+O(1/N)$. Hence we can deduce that

$$\frac{1}{x}\sum_{n\leqslant x}\tau(n)=\log x+2\gamma-1+O\left(\frac{1}{\sqrt{x}}\right),$$

an extraordinary improvement upon the earlier error term.

In the calculations in this article, this same idea is essential. We will take some functions, that are difficult to sum, and rewrite them as a sum of products of other functions, that are easier to sum, and find a way to sum them over long enough intervals for our methods to take effect. So we should define the convolution of two functions f and g as $f*g$ where

$$(f*g)(n):=\sum_{ab=n}f(a)g(b),$$

for every integer $n \geqslant 1$, where the sum is over all pairs of positive integers a, b whose product is n. Hence $\tau = 1 * 1$, where 1 is the function with $1(n) = 1$ for every $n \geqslant 1$.

Let $\delta_1(n) = 1$ if $n = 1$, and $\delta_1(n) = 0$ otherwise. Another important multiplicative function is the Mobius function $\mu(n)$, since $1 * \mu = \delta_1$. From this one can verify that $\mu(p) = -1$ and $\mu(p^e) = 0$ for all $e \geqslant 2$, for all primes p.

We define $L(n) := \log n$, and we let $\Lambda(n) = \log p$ if n is a power of prime p, and $\Lambda(n) = 0$ otherwise. By factoring n, we see that $L = 1 * \Lambda$. We therefore deduce that $\Lambda = (\mu * 1) * \Lambda = \mu * (1 * \Lambda) = \mu * L$; that is

$$\Lambda(n) = \sum_{ab=n} \mu(a) \log b \ = \begin{cases} \log p & \text{if } n = p^m, \text{ where } p \text{ is prime}, m \geqslant 1; \\ 0 & \text{otherwise.} \end{cases} . \qquad (2.2)$$ VMidentity

We can approach the prime number theorem via this identity by summing over all $n \leqslant x$ to get

$$\sum_{n \leqslant x} \Lambda(n) = \sum_{ab \leqslant x} \mu(a) \log b.$$

The left-hand side equals $\theta(x)$ plus a contribution from prime powers p^e with $e \geqslant 2$, and it is easily shown that this contribution is small (in fact $O(\sqrt{x})$). The right hand side is the convolution of an awkward function μ and something very smooth and easy to sum, L. Indeed, it is easy to see that $\sum_{b \leqslant B} \log b = \log B!$ and we can estimate this very precisely using Stirling's formula. One can infer (see GS [18] for details) that the prime number theorem is equivalent to proving that

$$\frac{1}{x} \sum_{n \leqslant x} \mu(n) \to 0 \quad \text{as } x \to \infty.$$

In our work here we will need a more convoluted identity that (VMidentity 2.2) to prove our estimates for primes in arithmetic progressions. There are several possible suitable identities, the simplest of which is due to Vaughan vaughan [35]:

$$\textit{Vaughan's identity}: \qquad \Lambda_{\geqslant V} = \mu_{<U} * L - \mu_{<U} * \Lambda_{<V} * 1 + \mu_{\geqslant U} * \Lambda_{\geqslant V} * 1 \qquad (2.3)$$ Vaughident

where $g_{>W}(n) = g(n)$ if $n > W$ and $g(n) = 0$ otherwise; and $g = g_{\leqslant W} + g_{>W}$. To verify this identity, we manipulate the algebra of convolutions:

$$\begin{aligned} \Lambda_{\geqslant V} &= \Lambda - \Lambda_{<V} = (\mu * L) - \Lambda_{<V} * (1 * \mu) \\ &= \mu_{<U} * L + \mu_{\geqslant U} * L - \mu_{<U} * \Lambda_{<V} * 1 - \mu_{\geqslant U} * \Lambda_{<V} * 1 \\ &= \mu_{<U} * L - \mu_{<U} * \Lambda_{<V} * 1 + \mu_{\geqslant U} * (\Lambda * 1 - \Lambda_{<V} * 1), \end{aligned}$$

Primektuples

2.5. **A quantitative prime k-tuplets conjecture.** We are going to develop a heuristic to guesstimate the number of pairs of twin primes $p, p+2$ up to x. We start with Gauss's statement that "the density of primes at around x is roughly $1/\log x$. Hence the probability that p is prime is $1/\log x$, and the probability that $p+2$ is prime is $1/\log x$ so, assuming that these events are independent, the probability that p and $p+2$

are simultaneously prime is

$$\frac{1}{\log x}\cdot\frac{1}{\log x} \;=\frac{1}{(\log x)^2};$$

and so we might expect about $x/(\log x)^2$ pairs of twin primes $p, p+2 \leqslant x$. But there is a problem with this reasoning, since we are implicitly assuming that the events "p is prime for an arbitrary integer $p \leqslant x$", and "$p+2$ is prime for an arbitrary integer $p \leqslant x$", can be considered to be independent. This is obviously false since, for example, if p is even then $p+2$ must also be. [6] So, to correct for the non-independence, we consider the ratio of the probability that both p and $p+2$ are not divisible by q, to the probabiliity that p and p' are not divisible by q, for each small prime q.

Now the probability that q divides an arbitrary integer p is $1/q$; and hence the probability that p is not divisible by q is $1-1/q$. Therefore the probability that both of two independently chosen integers are not divisible by q, is $(1-1/q)^2$.

The probability that q does not divide either p or $p+2$, equals the probability that $p \not\equiv 0$ or $-2 \pmod q$. If $q>2$ then p can be in any one of $q-2$ residue classes mod q, which occurs, for a randomly chosen $p \pmod q$, with probability $1-2/q$. If $q=2$ then p can be in any just one residue class mod 2, which occurs with probability $1/2$. Hence the "correction factor" for divisibility by 2 is

$$\frac{(1-\frac{1}{2})}{(1-\frac{1}{2})^2}=2,$$

whereas the "correction factor" for divisibility by any prime $q>2$ is

$$\frac{(1-\frac{2}{q})}{(1-\frac{1}{q})^2}.$$

Now divisibility by different small primes in independent, as we vary over values of n, by the Chinese Remainder Theorem, and so we might expect to multiply together all of these correction factors, corresponding to each "small" prime q. The question then becomes, what does "small" mean? In fact, it doesn't matter much because the product of the correction factors over larger primes is very close to 1, and hence we can simply extend the correction to be a product over all primes q. (More precisely, the infinite product over all q, converges.) Hence we define the *twin prime constant* to be

$$C := 2\prod_{\substack{q \text{ prime}\\ q\geqslant 3}}\frac{(1-\frac{2}{q})}{(1-\frac{1}{q})^2}\approx 1.3203236316,$$

and we conjecture that the number of prime pairs $p, p+2 \leqslant x$ is

$$\sim C\frac{x}{(\log x)^2}.$$

[6] Also note that the same reasoning would tell us that there are $\sim x/(\log x)^2$ prime pairs $p, p+1 \leqslant x$.

Computational evidence suggests that this is a pretty good guess. The analogous argument implies the conjecture that the number of prime pairs $p, p+2k \leqslant x$ is

$$\sim C \prod_{\substack{p|k \\ p\geqslant 3}} \left(\frac{p-1}{p-2}\right) \frac{x}{(\log x)^2}.$$

This argument is easily modified to make an analogous prediction for any k-tuple: Given $a_1, \ldots, a_k$, let $\Omega(p)$ be the set of distinct residues given by $a_1, \ldots, a_k \pmod p$, and then let $\omega(p) = |\Omega(p)|$. None of the $n + a_i$ is divisible by p if and only if n is in any one of $p - \omega(p)$ residue classes mod p, and therefore the correction factor for prime p is

$$\frac{(1 - \frac{\omega(p)}{p})}{(1 - \frac{1}{p})^k}.$$

Hence we predict that the number of prime k-tuplets $n + a_1, \ldots, n + a_k \leqslant x$ is,

$$\sim C(a) \frac{x}{(\log x)^k} \quad \text{where} \quad C(a) := \prod_p \frac{(1 - \frac{\omega(p)}{p})}{(1 - \frac{1}{p})^k}.$$

An analogous conjecture, via similar reasoning, can be made for the frequency of prime k-tuplets of polynomial values in several variables. What is remarkable is that computational evidence suggests that these conjectures do approach the truth, though this rests on a rather shaky theoretical framework. A more convincing theoretical framework (though rather more difficult) was given by Hardy and Littlewood [19] – see section 3.3.

Recogktuple

2.6. **Recognizing prime k-tuples.** The identity (2.2) allows us to distinguish prime powers from composite numbers in an arithmetic way. Such identities not only recognize primes, but can be used to identify integers with no more than k prime factors. For example

$$\Lambda_2(n) := \sum_{d|n} \mu(d)(\log n/d)^2 \;=\; \begin{cases} (2m-1)(\log p)^2 & \text{if } n = p^m; \\ 2\log p \log q & \text{if } n = p^a q^b,\ p \neq q; \\ 0 & \text{otherwise.} \end{cases}$$

In general

$$\Lambda_k(n) := \sum_{d|n} \mu(d)(\log n/d)^k$$

equals 0 if $\nu(n) > k$ (where $\nu(m)$ denotes the number of distinct prime factors of m). We will be working with (a variant of) the expression

$$\Lambda_k(\mathcal{P}(n)).$$

We have seen that if this is non-zero then $\mathcal{P}(n)$ has $\leqslant k$ distinct prime factors. We will next show that if $0 < a_1 < \ldots < a_k$ and $n \geqslant a_1 \ldots a_k$ then $\mathcal{P}(n)$ must have *exactly* k distinct prime factors. In that case if the k prime factors of $\mathcal{P}(n)$ are $p_1, \ldots, p_k$, then

$$\Lambda_k(\mathcal{P}(n)) = k!(\log p_1) \ldots (\log p_k).$$

Now, suppose that $\mathcal{P}(n)$ has $r \leqslant k-1$ distinct prime factors, call them $p_1, \dots, p_r$. For each p_i select $j = j(i)$ for which the power of p_i dividing $n + a_j$ is maximized. Evidently there exists some $J,\ 1 \leqslant J \leqslant k$ which is not a $j(i)$. Therefore if $p_i^{e_i} \| n + a_J$ then

$$p_i^{e_i} | (n + a_J) - (n + a_{j(i)}) = (a_J - a_{j(i)}),\ \text{which divides} \prod_{\substack{1 \leqslant j \leqslant k \\ j \neq J}} (a_J - a_j).$$

Hence

$$n + a_J = \text{lcm}_i\ p_i^{e_i}\ \text{divides} \prod_{\substack{1 \leqslant j \leqslant k \\ j \neq J}} (a_J - a_j),$$

and so $n < n + a_J \leqslant \prod_j a_j \leqslant n$, by hypothesis, which is impossible.

The expression for $\Lambda(n)$ in (VMidentity 2.2) can be re-written as

$$\Lambda(n) = \sum_{d|n} \mu(d) \log n/d,\ \text{and even}\ = \sum_{d|n} \mu(d) \log R/d,$$

for any R, provided $n > 1$. Selberg has shown that the truncation

$$\sum_{\substack{d|n \\ d \leqslant R}} \mu(d) \log R/d$$

is also "sensitive to primes"; and can be considerably easier to work with in various analytic arguments. In our case, we will work with the function

$$\sum_{\substack{d|\mathcal{P}(n) \\ d \leqslant R}} \mu(d) (\log R/d)^k,$$

which is analogously "sensitive" to prime k-tuplets, and easier to work with than the full sum for $\Lambda_k(\mathcal{P}(n))$.

3. Uniformity in arithmetic progressions

3.1. **When primes are first equi-distributed in arithmetic progressions.** There is an important further issue when considering primes in arithmetic progressions: In many applications it is important to know when we are first guaranteed that the primes are more-or-less equi-distributed amongst the arithmetic progressions $a \pmod q$ with $(a,q)=1$; that is

$$\theta(x;q,a) \sim \frac{x}{\phi(q)} \text{ for all } (a,q)=1. \tag{3.1}$$

PNTaps

To be clear, here we want this to hold when x is a function of q, as $q \to \infty$.

If one does extensive calculations then one finds that, for any $\epsilon > 0$, if q is sufficiently large and $x \geqslant q^{1+\epsilon}$ then the primes up to x are equi-distributed amongst the arithmetic progressions $a \pmod q$ with $(a,q)=1$, that is (3.1) holds. This is not only unproved at the moment, also no one really has a plausible plan of how to show such a result. However the slightly weaker statement that (3.1) holds for any $x \geqslant q^{2+\epsilon}$, can be shown to be true, assuming the Generalized Riemann Hypothesis. This gives us a clear plan for proving such a result, but one which has seen little progress in the last century!

The best unconditional results known are much weaker than we have hoped for, equidistribution only being proved once $x \geqslant e^{q^\epsilon}$. This is the *Siegel-Walfisz Theorem*, and it can be stated in several (equivalent) ways with an error term: For any $B>0$ we have

$$\theta(x;q,a) = \frac{x}{\phi(q)} + O\left(\frac{x}{(\log x)^B}\right) \text{ for all } (a,q)=1. \tag{3.2}$$

SW1

Or: for any $A>0$ there exists $B>0$ such that if $q<(\log x)^A$ then

$$\theta(x;q,a) = \frac{x}{\phi(q)}\left\{1 + O\left(\frac{1}{(\log x)^B}\right)\right\} \text{ for all } (a,q)=1. \tag{3.3}$$

SW2

That x needs to be so large compared to q limited the number of applications of this result.

The great breakthough of the second-half of the twentieth century came in appreciating that for many applications, it is not so important that we know that equidistribution holds for *every* a with $(a,q)=1$, and *every* q up to some Q, but rather that this holds for *most* such q (with $Q = x^{1/2-\epsilon}$). It takes some juggling of variables to state the Bombieri-Vinogradov Theorem: We are interested, for each modulus q, in the size of the largest error term

$$\max_{\substack{a \mod q\\(a,q)=1}} \left|\theta(x;q,a) - \frac{x}{\phi(q)}\right|,$$

or even

$$\max_{y\leqslant x} \max_{\substack{a \mod q\\(a,q)=1}} \left|\theta(y;q,a) - \frac{y}{\phi(q)}\right|.$$

The bounds $0 \leqslant \theta(x;q,a) \ll \frac{x}{q}\log x$ are trivial, the upper bound obtained by bounding the possible contribution from each term of the arithmetic progression. (Throughout the symbol "$\ll$", as in "$f(x) \ll g(x)$" means "there exists a constant $c>0$ such that

$f(x) \leqslant cg(x)$.") We would like to improve on the "trivial" upper bound, perhaps by a power of $\log x$, but we are unable to do so for all q. However, it turns out that we can prove that if there are exceptional q, then they are few and far between, and the Bombieri-Vinogradov Theorem expresses this in a useful form. The first thing we do is add up the above quantities over all $q \leqslant Q < x$. The "trivial" upper bound is then

$$\ll \sum_{q \leqslant Q} \frac{x}{q} \log x \ll x(\log x)^2.$$

The Bombieri-Vinogradov states that we can beat this trivial bound by an arbitrary power of $\log x$, provided Q is a little smaller than $\sqrt{x}$:

The Bombieri-Vinogradov Theorem. *For any given $A > 0$ there exists a constant $B = B(A)$, such that*

$$\sum_{q \leqslant Q} \max_{\substack{a \mod q \\ (a,q)=1}} \left| \theta(x; q, a) - \frac{x}{\phi(q)} \right| \ll_A \frac{x}{(\log x)^A}$$

where $Q = x^{1/2}/(\log x)^B$.

In fact one can take $B = 2A + 5$; and one can also replace the summand here by the expression above with the extra sum over y (though we will not need to do this here).

It is believed that this kind of estimate holds with Q significantly larger than $\sqrt{x}$; indeed Elliott and Halberstam conjectured [8] that one can take $Q = x^c$ for any constant $c < 1$:

The Elliott-Halberstam conjecture *For any given $A > 0$ and η, $0 < \eta < \frac{1}{2}$, we have*

$$\sum_{q \leqslant Q} \max_{\substack{a \mod q \\ (a,q)=1}} \left| \theta(x; q, a) - \frac{x}{\phi(q)} \right| \ll \frac{x}{(\log x)^A}$$

where $Q = x^{1/2+\eta}$.

However, it was shown in [13] that one *cannot* go so far as to take $Q = x/(\log x)^B$.

This conjecture was the starting point for the work of Goldston, Pintz and Yıldırım [15], as well as of Zhang [38]. This starting point was a beautiful argument from [15], that we will spell out in the next section, which yields the following result.

gpy-thm **Theorem 3.1** (Goldston-Pintz-Yıldırım). [15] *Let $k \geqslant 2$, $l \geqslant 1$ be integers, and $0 < \eta < 1/2$, such that*

$$1 + 2\eta > \left(1 + \frac{1}{2l+1}\right)\left(1 + \frac{2l+1}{k}\right). \tag{3.4}$$ thetal

If the Elliott-Halberstam conjecture holds with $Q = x^{1/2+\eta}$ then the following is true: If $x + a_1, \dots, x + a_k$ is an admissible set of forms then there are infinitely many integers n such that at least two of $n + a_1, \dots, n + a_k$ are prime numbers.

The conclusion here is exactly the statement of Zhang's main theorem.

For now the Elliott-Halberstam conjecture seems too difficult to prove, but progress has been made when restricting to one particular residue class: Fix integer $a \neq 0$. We believe that for any fixed η, $0 < \eta < \frac{1}{2}$, one has

$$\sum_{\substack{q \leqslant Q \\ (q,a)=1}} \left| \theta(x;q,a) - \frac{x}{\phi(q)} \right| \ll \frac{x}{(\log x)^A}$$

where $Q = x^{1/2+\eta}$. The key to progress has been to notice that if one can“factor” the key terms here into a sum of convolutions then it is easier to make progress, much as we saw with Dirichlet and the divisor problem. In this case the key convolution is (2.2) and Vaughan's identity (2.3). A second type of factorization that takes place concerns the modulus: it is much easier to proceed if we can factor the modulus q as, say dr where d and r are roughly some pre-specified sizes. The simplest class of integers q for which this sort of thing is true is the *y-smooth integers*, those integers whose prime factors are all $\leqslant y$. For example if we are given a y-smooth integer q and we want $q = dr$ with d not much smaller than D, then we select d to be the largest divisor of q that is $\leqslant D$ and we see that $D/y < d \leqslant D$. This is precisely the class of moduli that Zhang considered:[7]

Yitang Zhang's Theorem *There exist constants $\eta, \delta > 0$ such that for any given integer a, we have*

$$\sum_{\substack{q \leqslant Q \\ (q,a)=1 \\ q \text{ is } y-\text{smooth} \\ q \text{ squarefree}}} \left| \theta(x;q,a) - \frac{x}{\phi(q)} \right| \ll_A \frac{x}{(\log x)^A} \tag{3.5}$$

where $Q = x^{1/2+\eta}$ and $y = x^{\delta}$.

Zhang [38] proved his Theorem for $\eta/2 = \delta = \frac{1}{1168}$, and the argument works provided $414\eta + 172\delta < 1$. We will prove this result, by a somewhat simpler proof, provided $162\eta + 90\delta < 1$. We expect this estimate holds for every $\eta \in [0, 1/2]$ and every $\delta \in (0, 1]$, but just proving it for any positive pair $\eta, \delta > 0$ is an extraordinary breakthrough that has an enormous effect on number theory, since it is such an applicable result (and technique). This is the technical result that truly lies at the heart of Zhang's result about bounded gaps between primes, and sketching a proof of this is the focus of the second half of this article. starting section 5.

3.2. **A first result on gaps between primes.** We will now exploit the difference between the heuristic, presented in section 2.2, for the prime number theorem, and the correct count.

Let $m = \prod_{p \leqslant y} p$, $N = m^2$ and $x = mN$, so that $y \sim \log m = \frac{1}{3} \log x$ by the prime number theorem, (2.1). We consider the primes in the short intervals

$$[mn + 1, mn + J] \text{ for } N \leqslant n < 2N$$

[7] We will prove this with $\psi(x;q,a) := \sum_{n \leqslant x,\ n \leqslant a \pmod q} \Lambda(n)$ in place of $\theta(x;q,a)$. It is not difficult to show that the difference between the two sums is $\ll x^{1/2+o(1)}$.

with $J = y\log y$. Note that all of the short intervals are $\subset (x, 2x]$. The total number of primes in all of these short intervals is

$$\sum_{n=N+1}^{2N} \pi(mn+J) - \pi(mn+1) = \sum_{j=1}^{J} \pi(2x;m,j) - \pi(x;m,j) \sim \sum_{\substack{1\leqslant j\leqslant J\\ (j,m)=1}} \frac{x}{\phi(m)\log x}$$

assuming (PNTaps 3.1). Hence, since the maximum is always at least the average,

$$\max_{n\in(N,2N]} \pi(mn+J) - \pi(mn+1) \geqslant \frac{J}{\log x} \cdot \frac{\#\{1 \leqslant j \leqslant J:\ (j,m)=1\}}{(\phi(m)/m)J}$$
$$\sim e^{\gamma} \frac{J}{\log x}.$$

using the prime number theorem, and Merten's Theorem, as in section SieveHeuristic 2.2. Therefore we have proved that there is in an interval of length J, between x and $2x$, which has at least $\frac{J}{e^{-\gamma}\log x}$ primes, and so there must be two that differ by $\lesssim e^{-\gamma}\log x$.

HLheuristic

3.3. **Hardy and Littlewood's heuristic for the twin prime conjecture.** The rather elegant and natural heuristic for the quantitative twin prime conjecture, which we described in section Primektuples 2.5, was not the original way in which Hardy and Littlewood made this extraordinary prediction. The genesis of their technique lies in the *circle method.*, that they developed together with Ramanujan. The idea is that one can distinguish the integer 0 from all other integers, since

$$\int_0^1 e(nt)dt = \begin{cases} 1 & \text{if } n = 0; \\ 0 & \text{otherwise,} \end{cases} \tag{3.6}$$

expintegra

where, for any real number t, we write $e(t) := e^{2\pi i t}$. Notice that this is literally an integral around the unit circle. Therefore to determine whether the two given primes p and q differ by 2, we simply determine

$$\int_0^1 e((p-q-2)t)\ dt.$$

If we sum this up over all $p, q \leqslant x$, we find that the number of twin primes $p, p+2 \leqslant x$ equals, exactly,

$$\sum_{\substack{p,q\leqslant x\\ p,q \text{ primes}}} \int_0^1 e((p-q-2)t)\ dt = \int_0^1 |P(t)|^2 e(-2t)\ dt, \quad \text{where } \ P(t) := \sum_{\substack{p\leqslant x\\ p \text{ prime}}} e(pt).$$

In the circle method one next distinguishes between those parts of the integral which are large (the *major arcs*), and those that are small (the *minor arcs*). Typically the major arcs are small arcs around those t that are rationals with small denominators. Here the width of the arc is about $1/x$, and we wish to understand the contribution at $t = a/m$, where $(a, m) = 1$. Note then that

$$P(a/m) = \sum_{\substack{b \pmod m\\ (b,m)=1}} e_m(ab)\pi(x;m,b).$$

where $e_m(b) = e(\frac{b}{m}) = e^{2\pi i b/m}$. We note the easily proved identity

$$\sum_{r \pmod m,\ (r,m)=1} e_m(rk) = \phi((k,m))\mu(m/(m,k)).$$

Assuming the prime number theorem for arithmetic progressions with a good error term we therefore see that

$$P(a/m) \approx \frac{x}{\phi(m)\log x} \sum_{\substack{b \pmod m \\ (b,m)=1}} e_m(ab) = \frac{\mu(m)}{\phi(m)} \frac{x}{\log x}.$$

Hence in total we predict that the number of prime pairs $p, p+2 \leqslant x$ is roughly

$$\approx \frac{1}{x} \sum_{m \leqslant M} \sum_{a:\ (a,m)=1} e_m(-2a) \left| \frac{\mu(m)}{\phi(m)} \frac{x}{\log x} \right|^2 \approx \frac{x}{(\log x)^2} \sum_{m \geqslant 1} \frac{\mu(m)^2}{\phi(m)^2} \cdot \phi((2,m))\mu(m/(2,m))$$

$$= \frac{x}{(\log x)^2} \left(1 + \frac{1}{\phi(2)}\right) \prod_{p>2} \left(1 - \frac{1}{\phi(p)^2}\right) = C \frac{x}{(\log x)^2},$$

as in section quantPrimektuples ??. Moreover the analogous argument yields the more general conjecture for prime pairs $p, p+h$.

Why doesn't this argument lead to a proof of the twin prime conjecture? For the moment we have little idea how to show that the minor arcs contribute very little. Given that we do not know how to find cancelation amongst the minor arcs, we would need to show that the integrand is typically very small on the minor arcs, meaning that there is usually a lot of cancelation in the sums $P(t)$. For now this is an important open problem. Nonetheless, it is this kind of argument that has led to Helfgott's recent proof HH [21] that every odd integer $\geqslant 3$ is the sum of no more than three primes.

4. Goldston-Pintz-Yildirim's argument

We now give a version of the combinatorial argument of Goldston-Pintz-Yıldırım [15], which was the inspiration for proving that there are bounded gaps between primes:

4.1. The set up.

Let $\mathcal{H} = (a_1 < a_2 < \ldots < a_k)$ be an admissible k-tuple, and take $x > a_k$. Our goal is to select a function ν for which $\nu(n) \geqslant 0$ for all n, such that

$$\sum_{x<n\leqslant 2x} \nu(n)(\sum_{i=1}^{k} \theta(n+a_i) - \log 3x) > 0. \tag{4.1}$$

If we can do this then there must exist an integer n such that

$$\nu(n)(\sum_{i=1}^{k} \theta(n+a_i) - \log 3x) > 0.$$

In that case $\nu(n) \neq 0$ so that $\nu(n) > 0$, and therefore

$$\sum_{i=1}^{k} \theta(n+a_i) > \log 3x.$$

However each $n + a_i \leqslant 2x + a_k < 2x + x$ and so each $\theta(n+a_i) < \log 3x$. This implies that at least two of the $\theta(n+a_i)$ are non-zero, that is, at least two of $n+a_1, \ldots, n+a_k$ are prime.

A simple idea, but the difficulty comes in selecting the function $\nu(n)$ with these properties for which we can evaluate the sum. In [15] they had the further idea that they could select $\nu(n)$ so that it would be sensitive to when each $n + a_i$ is prime, or "almost prime", and so they relied on the type of construction that we discussed in section 2.6. In order that $\nu(n) > 0$ one can simply take it to be a square. Hence we select

$$\nu(n) := \Bigg(\sideset{}{'}\sum_{d|\mathcal{P}(n)} \lambda(d) \Bigg)^2$$

where

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

when $d \in \mathcal{D}$, and $\lambda(d) = 0$ otherwise, for some positive integer $m = k + \ell$, where $\mathcal{D}$ is a subset of the squarefree integers in $\{1, \ldots, R\}$, and we select $R < x^{1/3}$. In the argument of [15], $\mathcal{D}$ includes all of the squarefree integers in $\{1, \ldots, R\}$, whereas Zhang uses only the y-smooth ones. Our formulation works in both cases.

4.2. Evaluating the sums, I.

Now, expanding the above sum gives

$$\sideset{}{'}\sum_{\substack{d_1,d_2\\ D:=[d_1,d_2]}} \lambda(d_1)\lambda(d_2) \left(\sum_{i=1}^{k} \sum_{\substack{x<n\leqslant 2x\\ D|\mathcal{P}(n)}} \theta(n+a_i) - \log 3x \sum_{\substack{x<n\leqslant 2x\\ D|\mathcal{P}(n)}} 1 \right). \tag{4.2}$$

Let $\Omega(D)$ be the set of congruence classes $m \pmod D$ for which $D|P(m)$; and let $\Omega_i(D)$ be the set of congruence classes $m \in \Omega(D)$ with $(D, m+a_i) = 1$. Hence the parentheses in the above line equals

$$\sum_{i=1}^{k} \sum_{m\in\Omega_i(D)} \sum_{\substack{x<n\leqslant 2x \\ n\equiv m \pmod D}} \theta(n+a_i) - \log 3x \sum_{m\in\Omega(D)} \sum_{\substack{x<n\leqslant 2x \\ n\equiv m \pmod D}} 1. \tag{4.3}$$

gpy3

The final sum evidently equals $x/D + O(1)$; the error term much smaller than the main term. We will come back to these error terms a little later. For the first sums we expect PNTaps (3.1) holds, so that each

$$\theta(2x; D, m+a_i) - \theta(x; D, m+a_i) \sim \frac{x}{\phi(D)}.$$

We again neglect, for now, the error terms, and will substitute these two estimates into the previous line. First though, note that the sets $\Omega(D)$ and $\Omega_i(D)$ may be constructed using the Chinese Remainder Theorem from the sets with D prime. Therefore if $\omega(D) := |\Omega(D)|$ then $\omega(.)$ is a multiplicative function. Moreover each $|\Omega_i(p)| = \omega(p) - 1$, which we denote by $\omega^*(p)$, and each $|\Omega_i(D)| = \omega^*(D)$, extending ω^* to be a multiplicative function. Putting this altogether we obtain here a main term of

$$k\omega^*(D)\frac{x}{\phi(D)} - (\log 3x)\omega(D)\frac{x}{D} = x\left(k\frac{\omega^*(D)}{\phi(D)} - (\log 3x)\frac{\omega(D)}{D}\right).$$

This is typically negative which is why we cannot simply take our weights to all be positive. Substituting this in above we obtain, in total, the sums

$$x\left(k \sideset{}{'}\sum_{\substack{d_1,d_2 \\ D:=[d_1,d_2]}} \lambda(d_1)\lambda(d_2)\frac{\omega^*(D)}{\phi(D)} - (\log 3x) \sideset{}{'}\sum_{\substack{d_1,d_2 \\ D:=[d_1,d_2]}} \lambda(d_1)\lambda(d_2)\frac{\omega(D)}{D}\right). \tag{4.4}$$

gpy4

We shall explain a little later how these were evaluated in gpy [15]. First though let's return to the error terms:

4.3. **Bounding the error terms.**

The first one above, from counting integers in an arithmetic progression, yields in total,

$$\ll \sum_{d_1,d_2\leqslant R} |\lambda(d_1)||\lambda(d_2)| \log 3x \leqslant R^2 \log 3x \leqslant x^{2/3}\log 3x,$$

since each $|\lambda(d)| \leqslant 1$ by definition. For the second one we will need our bound on primes in arithmetic progression: For any integer b we have

$$\sideset{}{'}\sum_{\substack{D\leqslant Q \\ (D,b)=1}} \left|\theta(X; D, b) - \frac{X}{\phi(D)}\right| \ll_A \frac{X}{(\log X)^A} \tag{4.5}$$

PNTassump

where the constant depends only on A. Here $Q = x^{1/2+\eta}$ and the restriction $\sum'$ is vacuous if we assume the Elliott-Halberstam conjecture, and means that D is y-smooth if we are using Zhang's estimate.

Using the same bounds $|\lambda(d)| \leqslant 1$, we have the upper bound on the second term of

$$\leqslant \sideset{}{'}\sum_{\substack{d_1,d_2\\ D:=[d_1,d_2]}} \sum_{i=1}^{k} \sum_{m\in\Omega_i(D)} \left|\theta(2x;D,m+a_i) - \theta(x;D,m+a_i) - \frac{x}{\phi(D)}\right|.$$

Let $O_i(D) = \Omega_i(D) + a_i$ (which may also be constructed from the $O_i(p)$ using the Chinese Remainder Theorem). Note that $|O_i(D)| = \omega_i(D) \leqslant (k-1)^{\omega(D)}$ where, here, $\omega(D)$ denotes the number of distinct prime factors of D. Each D that appears is squarefree and is $\leqslant R^2$, and can occur for at most $3^{\omega(D)}$ pairs d_1, d_2. Since $\tau(D) = 2^{\omega(D)}$ we deduce that, for $A = \log(3(k-1))/\log 2$, the above is

$$\leqslant \sum_{i=1}^{k} \sum_{X=x \text{ or } 2x} \sideset{}{'}\sum_{D\leqslant Q} \tau(D)^A \cdot \frac{1}{\omega_i(D)} \sum_{b\in O_i(D)} \left|\theta(X;D,b) - \frac{X}{\phi(D)}\right| \tag{4.6}$$

gpy5

where $Q = R^2$.

Now let m be the lcm of the integers $D \leqslant Q$, counted in the sum. Notice that $O_i(m)$ reduced mod D, gives $\omega_i(m/D)$ copies of $O_i(D)$, and hence

$$\frac{1}{\omega_i(D)} \sum_{b\in O_i(D)} \left|\theta(X;D,b) - \frac{X}{\phi(D)}\right| = \frac{1}{\omega_i(m)} \sum_{b\in O_i(m)} \left|\theta(X;D,b) - \frac{X}{\phi(D)}\right|,$$

so that the quantity in (4.6) equals

$$\sum_{i=1}^{k} \sum_{X=x \text{ or } 2x} \frac{1}{\omega_i(m)} \sum_{b\in O_i(m)} \left\{ \sideset{}{'}\sum_{D\leqslant Q} \tau(D)^A \left|\theta(X;D,b) - \frac{X}{\phi(D)}\right| \right\}. \tag{4.7}$$

gpy6

Now fix k, X and b. To bound the sum over D we need to remove the $\tau(D)^A$ term, which we do by Cauchying. It will help to notice the trivial bounds $0 \leqslant \theta(X;D,b) \ll (X\log X)/D$, so that $D|\theta(X;D,b) - \frac{X}{\phi(D)}| \ll X\log X$. Hence

$$\begin{aligned}\left(\sideset{}{'}\sum_{D\leqslant Q} \tau(D)^A \left|\theta(X;D,b) - \frac{X}{\phi(D)}\right|\right)^2 &\leqslant \sum_{D\leqslant Q} \frac{\tau(D)^{2A}}{D} \cdot \sideset{}{'}\sum_{D\leqslant Q} D\left|\theta(X;D,b) - \frac{X}{\phi(D)}\right|^2 \\ &\leqslant X(\log X)^B \sideset{}{'}\sum_{D\leqslant Q} \left|\theta(X;D,b) - \frac{X}{\phi(D)}\right|\end{aligned}$$

and this is $\ll_C X^2/(\log X)^C$ for any C, by (4.5). Hence the quantity in (4.6) is

$$\ll_A k\frac{X}{(\log X)^A},$$

for any $A > 0$, which is acceptable.

4.4. **Perron's formula.**

There are two methods to calculate the main terms, one more analytic ([15]), the other, ([34], [16]), more combinatorial. We shall outline both.

It is possible to obtain an asymptotic estimate for the mean value of multiplicative functions g for which $g(p)$ is "close" to some given integer k, for all sufficiently large p.

The Selberg-Delange theorem tells us that

$$\sum_{n\leqslant x}\frac{g(n)}{n}\sim\prod_{p\leqslant x}\left(1+\frac{g(p)}{p}+\frac{g(p^2)}{p^2}+\dots\right)\left(1-\frac{1}{p}\right)^k\cdot\frac{(\log x)^k}{k!}.$$

When $g(p)$ is sufficiently close to some k that the Euler product converges, we can replace the product up to x, by the product over *all* primes p in the line above. This makes this formula easy to manipulate; in particular, by partial summation, we obtain

$$\sum_{n\leqslant x}\frac{g(n)}{n}\cdot\frac{(\log(x/n))^\ell}{\ell!}\sim C(g)\cdot\frac{(\log x)^{k+\ell}}{(k+\ell)!} \tag{4.8}$$ $\boxed{\texttt{SD+}}$

for $k\geqslant 1,\ \ell\geqslant 0$ using the beta integral $\int_0^1(1-v)^\ell v^{k-1}dv=(k-1)!\ell!/(k+\ell)!$, where

$$C(g):=\prod_{p\text{ prime}}\left(1+\frac{g(p)}{p}+\frac{g(p^2)}{p^2}+\dots\right)\left(1-\frac{1}{p}\right)^k.$$

4.5. **The combinatorial approach.** We will suppose for now that the $\Lambda(d)$ remain unchosen. We need to evaluate the sums

$$\sideset{}{'}\sum_{\substack{d_1,d_2\\ D:=[d_1,d_2]}}\lambda(d_1)\lambda(d_2)\frac{\omega^*(D)}{\phi(D)}\quad\text{and}\quad\sideset{}{'}\sum_{\substack{d_1,d_2\\ D:=[d_1,d_2]}}\lambda(d_1)\lambda(d_2)\frac{\omega(D)}{D}\ .$$

As shown by Soundararajan [34]sound, we may evaluate these much like Selberg does in his upper bound sieve. The main idea is a change of variable: Let ϕ_ω be the multiplicative function (defined here, only on squarefree integers) for which $\phi_\omega(p)=p-\omega(p)$, and then

$$y(r):=\mu(r)\frac{\phi_\omega(r)}{\omega(r)}\sideset{}{'}\sum_{n:\ r|n}\frac{\lambda(n)\omega(n)}{n};$$

and one can verify this is invertible with

$$\lambda(d)=\mu(d)\frac{d}{\omega(d)}\sideset{}{'}\sum_{n:\ d|n}\frac{y(n)\omega(n)}{\phi_\omega(n)}$$

Now

$$\begin{aligned}\sideset{}{'}\sum_{\substack{d_1,d_2\\ D:=[d_1,d_2]}}\lambda(d_1)\lambda(d_2)\frac{\omega(D)}{D}&=\sideset{}{'}\sum_{\substack{d_1,d_2\\ D:=[d_1,d_2]}}\frac{\omega(D)}{D}\mu(d_1)\frac{d_1}{\omega(d_1)}\sideset{}{'}\sum_{r:\ d_1|r}\frac{y(r)\omega(r)}{\phi_\omega(r)}\mu(d_2)\frac{d_2}{\omega(d_2)}\sideset{}{'}\sum_{s:\ d_2|s}\frac{y(s)\omega(s)}{\phi_\omega(s)}\\ &=\sum_{r,s}\frac{y(r)\omega(r)}{\phi_\omega(r)}\frac{y(s)\omega(s)}{\phi_\omega(s)}\sideset{}{'}\sum_{\substack{d_1,d_2\\ d_1|r,\ d_2|s}}\mu(d_1)\mu(d_2)\frac{(d_1,d_2)}{\omega((d_1,d_2))}\end{aligned}$$

By writing $d_j=e_jf_j$ where $e_j|(r,s)$ and $f_1|r/(r,s),\ f_2|s/(r,s)$, we see that the sum over f_j equals 0 unless $r/(r,s)=s/(r,s)=1$; that is $r=s$. Hence the above is

$$=\sum_{r,}\frac{y(r)^2\omega(r)^2}{\phi_\omega(r)^2}\sideset{}{'}\sum_{d_1,d_2|r}\mu(d_1)\mu(d_2)\frac{(d_1,d_2)}{\omega((d_1,d_2))}$$

Letting $g = (d_1, d_2)$ and writing $d_1 = ge_1,\ d_2 = ge_2$, so $ge_1e_2|r$, we see that the sum over e_2 is 0 unless $r = ge_1$. The above becomes

$$\sideset{}{'}\sum_{\substack{d_1,d_2\\ D:=[d_1,d_2]}} \lambda(d_1)\lambda(d_2)\frac{\omega(D)}{D} = \sum_r \frac{y(r)^2\omega(r)^2}{\phi_\omega(r)^2}\sum_{g|r}\frac{g}{\omega(g)}\mu(r/g) = \sum_r \frac{y(r)^2\omega(r)}{\phi_\omega(r)}. \tag{4.9} \qquad \boxed{\texttt{solve1}}$$

One can similarly show that

$$\sideset{}{'}\sum_{\substack{d_1,d_2\\ D:=[d_1,d_2]}} \lambda(d_1)\lambda(d_2)\frac{\omega^*(D)}{\phi(D)} = \sum_r \frac{y^*(r)^2\omega^*(r)}{\phi_\omega(r)} \tag{4.10} \qquad \boxed{\texttt{solve2}}$$

where

$$y^*(r) = r\sideset{}{'}\sum_{n:\ r|n}\frac{y(n)}{\phi(n)}.$$

We select

$$y(r) = y_\ell(r) := \begin{cases} C(a)\frac{(\log(R/r))^\ell}{\ell!} & \text{if } r \text{ is squarefree, and } r \leqslant R;\\ 0 & \text{otherwise,}\end{cases}$$

in the notation of section $\overset{\texttt{Primektuples}}{2.5}$. By $(\overset{\texttt{SD+}}{4.8})$ this implies that

$$y^*(r) \sim y_{\ell+1}(r);$$

$$\lambda(d) = \begin{cases} \{1+o(1)\}\mu(d)\frac{(\log(R/d))^{k+\ell}}{(k+\ell)!} & \text{if } d \text{ is squarefree, and } d \leqslant R;\\ 0 & \text{otherwise.}\end{cases}$$

Moreover $(\overset{\texttt{SD+}}{4.8})$ also implies that

$$(\overset{\texttt{solve1}}{4.9}) \sim C(a)\binom{2\ell}{\ell}\cdot\frac{(\log R)^{k+2\ell}}{(k+2\ell)!}$$

and

$$(\overset{\texttt{solve2}}{4.10}) \sim C(a)\binom{2\ell+2}{\ell+1}\cdot\frac{(\log R)^{k+2\ell+1}}{(k+2\ell+1)!}$$

4.6. **Finding a positive difference; the proof of Theorem $\overset{\texttt{gpy-thm}}{3.1}$.** Now inserting these last two estimates into $(\overset{\texttt{gpy4}}{4.4})$ we obtain

$$x\left(\{1+o(1)\}\frac{k}{(k+2\ell+1)!}\binom{2\ell+2}{\ell+1}\cdot\frac{C(a)}{(\log R)^{k+2\ell-1}} - \{1+o(1)\}(\log 3x)\frac{1}{(k+2\ell)!}\binom{2\ell}{\ell}\cdot\frac{C(a)}{(\log R)^{k+2\ell}}\right)$$

$$\geqslant \frac{C(a)x\log 3x}{4(\log R)^{k+2\ell}}\frac{k}{(k+2\ell+1)!}\binom{2\ell+2}{\ell+1}\left(\frac{2\log Q}{\log 3x} - \left(1+\frac{1}{2\ell+1}\right)\left(1+\frac{2\ell+1}{k}\right)+o(1)\right)$$

as $Q = R^2$. This is > 0 if $(\overset{\texttt{thetal}}{3.4})$ holds, and so we deduce Theorem $\overset{\texttt{gpy-thm}}{3.1}$.

4.7. **The challenge in completing the proof of Zhang's Theorem.** We modify the proof in the last section suitably. In the arguments above we replace y and y^*, by z and z^*, where we select

$$z(r) = z_\ell(r) := \begin{cases} C(a)\frac{(\log(R/r))^\ell}{\ell!} & \text{if } r \text{ is squarefree, } y\text{-smooth and } r \leqslant R; \\ 0 & \text{otherwise,} \end{cases}$$

We bound ($\overset{\texttt{solve1}}{\text{4.9}}$)(with z in place of y) from above, trivially, as follows:

$$\sum_r \frac{z(r)^2\omega(r)}{\phi_\omega(r)} \leqslant \sum_r \frac{y(r)^2\omega(r)}{\phi_\omega(r)} \sim C(a)\binom{2\ell}{\ell} \cdot \frac{(\log R)^{k+2\ell}}{(k+2\ell)!},$$

from the calculation in the previous section.

To bound ($\overset{\texttt{solve2}}{\text{4.10}}$)(with z in place of y) from below, is more subtle. Notice that each term is $\geqslant 0$, so we have a lower bound by restricting attention to only those $r \in [R/y, R]$ which are y-smooth. Now if $y(n) \neq 0$ and $r|n$ then $n/r \leqslant R/r \leqslant y$, and so n is y-smooth; hence

$$z^*(r) = r \sum_{\substack{n:\ r|n \\ n \text{ is } y\text{-smooth}}} \frac{z(n)}{\phi(n)} = r \sum_{\substack{n:\ r|n \\ n \text{ is } y\text{-smooth}}} \frac{y(n)}{\phi(n)} = r \sideset{}{'}\sum_{n:\ r|n} \frac{y(n)}{\phi(n)} = y^*(r).$$

Therefore

$$\begin{aligned}\sum_r \frac{z^*(r)^2\omega^*(r)}{\phi_\omega(r)} &\geqslant \sum_{R/y\leqslant r\leqslant R} \frac{z^*(r)^2\omega^*(r)}{\phi_\omega(r)} = \sum_{\substack{R/y\leqslant r\leqslant R \\ r \text{ is } y\text{-smooth}}} \frac{y^*(r)^2\omega^*(r)}{\phi_\omega(r)} \\ &\geqslant \sum_{R/y\leqslant r\leqslant R} \frac{y^*(r)^2\omega^*(r)}{\phi_\omega(r)} \left(1 - \sum_{\substack{p|r \\ p>y}} 1\right) \\ &\geqslant \sum_{r\leqslant R} \frac{y^*(r)^2\omega^*(r)}{\phi_\omega(r)} - \sum_{r\leqslant R/y} \frac{y^*(r)^2\omega^*(r)}{\phi_\omega(r)} - \sum_{p>y} \sum_{\substack{r\leqslant R \\ p|r}} \frac{y^*(r)^2\omega^*(r)}{\phi_\omega(r)}\end{aligned}$$

Now, by ($\overset{\texttt{SD+}}{\text{4.8}}$), we have

$$\sum_{\substack{r\leqslant R \\ p|r}} \frac{y^*(r)^2\omega^*(r)}{\phi_\omega(r)} \sim \frac{\omega(p)-1}{p-1} \cdot C(a)\binom{2\ell+2}{\ell+1} \cdot \frac{(\log R/p)^{k+2\ell+1}}{(k+2\ell+1)!}$$

Summing this over $y < p \leqslant R$, and as $\omega(p) \leqslant k$ and $R/p \leqslant R/y$ we deduce that

$$\sum_{p>y} \sum_{\substack{r\leqslant R \\ p|r}} \frac{y^*(r)^2\omega^*(r)}{\phi_\omega(r)} \lesssim (k-1)\log(1/\delta)(1-\delta)^{k+2\ell+1} \cdot C(a)\binom{2\ell+2}{\ell+1} \cdot \frac{(\log R)^{k+2\ell+1}}{(k+2\ell+1)!}$$

If one proceeds as in the proof of (4.8) (i.e. by partial summation) one obtains

$$\sum_{r\leqslant R/y}\frac{y^*(r)^2\omega^*(r)}{\phi_\omega(r)}\sim\frac{\int_0^{1-\delta}(1-v)^{2\ell}v^{k-1}dv}{\int_0^1(1-v)^{2\ell}v^{k-1}dv}\cdot\sum_{r\leqslant R}\frac{y^*(r)^2\omega^*(r)}{\phi_\omega(r)}$$
$$\lesssim\frac{(k+2\ell)!}{(k-1)!(2\ell)!}(1-\delta)^{k-1}\cdot C(a)\binom{2\ell+2}{\ell+1}\cdot\frac{(\log R)^{k+2\ell+1}}{(k+2\ell+1)!}$$

Assuming that $\ell\asymp\sqrt{k}$, we deduce that

$$\sum_r\frac{z^*(r)^2\omega^*(r)}{\phi_\omega(r)}\gtrsim\{1+O(k^{2\ell+1}(1-\delta)^k)\}C(a)\binom{2\ell+2}{\ell+1}\cdot\frac{(\log R)^{k+2\ell+1}}{(k+2\ell+1)!}$$

Proceeding as in the previous section (with z in place of y) and taking $Q=x^{\frac{1}{2}+\eta}$ with $L=2\ell+1\asymp\sqrt{k}$, we are successful provided

$$1+2\eta>1+\frac{2}{L}+O(1/k+k^L(1-\delta)^k)+o(1),$$

which works for $\delta=(2L\log k)/k$ and $\eta=2/L$.

4.8. **Numerics.** Later we will show that we may work here under the assumption that $162\eta+90\delta<1$. The above inequalities hold (more-or-less) with $L=863, k=L^2$ and $\eta=1/(L-1)$. Hence we should be able to take $k\leqslant 750,000$ and $B\approx 10^7$.

Remark 4.1. These arguments actually give quantitative information: One can deduce ([29],[26]) that if $\mathcal{H}$ is an admissible k-tuple and x is sufficiently large, then there are $\gg x/\log^k x$ values of $n\in[x,2x]$ such that $n+\mathcal{H}$ contains two primes. In justifying our weights we claimed that they are "sensitive" to all of the elements of $n+\mathcal{H}$ being prime: To be more explicit, one can further prove that all of the elements of $n+\mathcal{H}$ have no prime factors less than x^c (for some fixed $c>0$), as well as two of them being prime.

GeneralBV

5. Distribution in arithmetic progressions

Our goal, in the rest of the article, is to prove PNTassump (4.5). In this section we will see how this question fits into a more general framework, as developed by Bombieri, Friedlander and Iwaniec bfi [3], so that the results here should allow us to deduce analogous results for interesting arithmetic sequences other than the primes.

5.1. **General sequence in arithmetic progressions with large common differences.** One can ask whether *any* given sequence $(\beta_n)_{n\geqslant 1} \in \mathbb{C}$ is well-distributed in arithmetic progressions. To this end we might ask that it is well-distributed in a range analogous to SW1 (3.2). Therefore we say that β satisfies a *Siegel-Walfisz condition* if, for any fixed $A > 0$, and whenever $(a, q) = 1$, we have

$$\left| \sum_{\substack{n\leqslant x \\ n\equiv a \pmod q}} \beta_n - \frac{1}{\phi(q)} \sum_{\substack{n\leqslant x \\ (n,q)=1}} \beta_n \right| \ll_A \frac{\|\beta\| x^{\frac12}}{(\log x)^A} ,$$

with $\|\beta\| = \|\beta\|_2$ where, as usual,

$$\|\beta\|_p := \left(\sum_{n\leqslant x} |\beta_n|^p \right)^{\frac1p} .$$

It is necessary to have a term like $\|\beta\|$ on the right-hand side to account for the size of the terms of the sequence β. [8] Note that this estimate is trivial if $q \geqslant (\log x)^{2A}$ (after Cauchying), so is only of interest for x very large compared to q.

Using the large sieve, Bombieri, Friedlander and Iwaniec bfi [3] were able to prove two extraordinary results. In the first they showed that if β satisfies a Siegel-Walfisz condition,[9] then it is well-distributed for *almost all* arithmetic progressions $a \pmod q$, for *almost all* $q \leqslant x/(\log x)^B$:

Theorem 5.1. *Suppose that the sequence of complex numbers $\beta_n, n \leqslant x$ satisfies a Siegel-Walfisz condition. For any $A > 0$ there exists $B = B(A) > 0$ such that*

$$\sum_{q\leqslant Q} \sum_{a:\ (a,q)=1} \left| \sum_{n\equiv a \pmod q} \beta_n - \frac{1}{\phi(q)} \sum_{(n,q)=1} \beta_n \right|^2 \ll \|\beta\|^2 \frac{x}{(\log x)^A}$$

where $Q = x/(\log x)^B$.

The analogous result for $\Lambda(n)$ is known as the *Barban-Davenport-Halberstam theorem* and in this case one can even obtain an asymptotic.

[8] Analogously, we might have used $\|\beta\|_r x^{1-\frac1r}$ for any $r > 1$ in place of $\|\beta\| x^{\frac12}$. This bound is trivial for $q \geqslant (\log x)^{Ar/(r-1)}$ by Holdering (instead of Cauchying).

[9] Their condition appears to be weaker than that assumed here, but is actually equivalent by Lemma SWcoprime 6.2.

In the second result they show that rather general convolutions are well-distributed for *all* arithmetic progressions $a \pmod q$, for *almost all* $q \leqslant x^{1/2}/(\log x)^B$:

BFI2 **Theorem 5.2.** *Suppose that we have two sequences of complex numbers α_m, $M < m \leqslant 2M$, and $\beta_n, N < n \leqslant 2N$, where β_n satisfies the Siegel-Walfisz condition. For any $A > 0$ there exists $B = B(A) > 0$ such that*

$$\sum_{q \leqslant Q} \max_{a:\ (a,q)=1} \left| \sum_{n \equiv a \pmod q} (\alpha * \beta)(n) - \frac{1}{\phi(q)} \sum_{(n,q)=1} (\alpha * \beta)(n) \right| \ll \|\alpha\| \|\beta\| \frac{x^{1/2}}{(\log x)^A}$$

where $Q = x^{1/2}/(\log x)^B$, provided $x = MN$ with $x^\epsilon \ll M, N \ll x^{1-\epsilon}$.

In fact their proof works provided $N \geqslant \exp((\log x)^\epsilon)$ and $M \geqslant (\log x)^{2B+4}$.

This allowed them to give a proof of the Bombieri-Vinogradov theorem for primes that seems to be less dependent on very specific properties of the primes (as we will see in the next subsection). The subject, though, had long been stuck with the bound $x^{1/2}$ on the moduli.[10]

Bombieri, Friedlander and Iwaniec bfi [3] made the following conjecture.[11] They noted that in many applications, it suffices to work with a fixed (as is true in the application here).

Conjecture 5.3. *Suppose that we have two sequences of complex numbers α_m, $M < m \leqslant 2M$, and $\beta_n, N < n \leqslant 2N$, where β_n satisfies the Siegel-Walfisz condition. For any $A, \epsilon > 0$, and every integer a, we have*

$$\sum_{\substack{q \leqslant Q \\ (q,a)=1}} \left| \sum_{n \equiv a \pmod q} (\alpha * \beta)(n) - \frac{1}{\phi(q)} \sum_{(n,q)=1} (\alpha * \beta)(n) \right| \ll \|\alpha\| \|\beta\| \frac{x^{1/2}}{(\log x)^A}$$

where $Q = x^{1-\epsilon}$, provided $x = MN$ with $x^\epsilon \ll M, N \ll x^{1-\epsilon}$.

The extraordinary work of Zhang breaks through the $\sqrt{x}$ barrier in some generality, working with moduli slightly larger than $x^{1/2}$. In this case the moduli are y-smooth, with $y = x^\delta$; here we say that q is *y-smooth* if all of its prime factors are $\leqslant y$, that is $P(q) \leqslant y$, where we write $P(q)$ for q's largest prime factor.

We say that $\alpha * \beta$ satisfies the *average sieving condition* if for each fixed $A > 0$, we have

$$\sum_{q < Q} \sum_{\substack{x < mn \leqslant x + x/(\log x)^A \\ mn \equiv a \pmod q}} |\alpha_m||\beta_n| \ll_A \|\alpha\| \|\beta\| \frac{x^{1/2}}{(\log x)^A} (\log x)^{O(1)}.$$

[10] There had been some partial progress with moduli $> x^{1/2}$, as in bfi-2 [4], but no upper bounds which "win" by an arbitrary power of $\log x$ (which is what is essential to applications).

[11] They actually conjectured that one can take $Q = x/(\log x)^B$. They also conjectured that if one assumes the Siegel-Walfisz condition with $\|\beta\|_s N^{1-\frac{1}{s}}$ in place of $\|\beta\| N^{\frac{1}{2}}$ then we may replace $\|\alpha\|\|\beta\| x^{1/2}$ in the upper bound here by $\|\alpha\| M^{1-\frac{1}{r}} \|\beta\| N^{1-\frac{1}{s}}$.

for any $Q < x^{2/3}$. We say that $\alpha * \beta$ satisfies the *necessary sieving condition* if both $\alpha * \beta$ and $\alpha^4 * \beta^4$ satisfy the average sieving condition. It is not difficult to show that these conditions hold if, for instance, $|\alpha(n)|, |\beta(n)| \ll (\tau(n)\log x)^{O(1)}$ for all n.

The key result is as follows:

Vdyadicrange **Theorem 5.4.** *There exist constants $\eta, \delta > 0$ with the following property. Suppose that we have two sequences of complex numbers α_m, $M < m \leqslant 2M$, and $\beta_n, N < n \leqslant 2N$, where β satisfies the Siegel-Walfisz condition, and that $\alpha * \beta$ satisfies the necessary sieving condition. For any $A > 0$, for any integer a,*

$$\sum_{\substack{q\leqslant Q\\ P(q)\leqslant x^{\delta}\\ (q,a)=1\\ q \text{ squarefree}}} \left| \sum_{n\equiv a \pmod q} (\alpha*\beta)(n) - \frac{1}{\phi(q)} \sum_{(n,q)=1} (\alpha*\beta)(n) \right| \ll_A \|\alpha\|\|\beta\| \frac{x^{1/2}}{(\log x)^A}$$

where $Q = x^{1/2+\eta}$, provided $x^{1/3} \ll N \leqslant M \ll x^{2/3}$.

BVwiderange **Corollary 5.5.** *There exist constants $\eta, \delta > 0$ with the following property. Suppose that we have two sequences of complex numbers α_m, β_n $x^{1/3} < m, n \leqslant x^{2/3}$, which both uniformly satisfy the Siegel-Walfisz condition, and that $\alpha * \beta$ satisfies the necessary sieving condition. For any $A > 0$, for any integer a,*

$$\sum_{\substack{q\leqslant Q\\ P(q)\leqslant x^{\delta}\\ (q,a)=1\\ q \text{ squarefree}}} \left| \sum_{\substack{n\leqslant x\\ n\equiv a \pmod q}} (\alpha*\beta)(n) - \frac{1}{\phi(q)} \sum_{\substack{n\leqslant x\\ (n,q)=1}} (\alpha*\beta)(n) \right| \ll_A \|\alpha\|\|\beta\| \frac{x^{1/2}}{(\log x)^A}$$

where $Q = x^{1/2+\eta}$.

Proof. of Corollary BVwiderange *5.5* Theorem BVdyadicrange 5.4 gives the result when the support for both α and β are within dyadic intervals. Here we deduce the result over wider ranges of m and n with $mn \leqslant x$ for some given x.

Let $T = (\log x)^A$, and R be the smallest integer with $(1 + 1/T)^R > x$. Let $S_{i,j}$ be the set of pairs (m, n) with $(1+1/T)^i \leqslant m < (1+1/T)^{i+1}$, $(1+1/T)^j \leqslant n < (1+1/T)^{j+1}$. Notice that if $i + j \leqslant R - 3$ and $(m, n) \in S_{i,j}$ then $mn \leqslant (1 + 1/T)^{i+j+2} \leqslant (1 + 1/T)^{R-1} \leqslant x$. Finally let S_0 be the set of pairs (m, n) with $mn \leqslant x$, that are not included in any of the $S_{i,j}$ with $i+j \leqslant R-3$. If $(m, n) \in S_0$ then $mn \geqslant (1+1/T)^{i+j} \geqslant (1+1/T)^{R-2} \geqslant x(1-3/T)$.

Now, by the triangle inequality, the sum over all pairs m, n is bounded by the sum, for each such set S, over the sums for $(m, n) \in S$. For any S of the form $S_{i,j}$ we use Theorem BVdyadicrange 5.4 with A replaced by $3A + 2$. For $S = S_0$ we get the bound from the hypothesis that $\alpha * \beta$ satisfies the average sieving condition. The result follows from summing these bounds. □

5.2. Vaughan's identity, and the deduction of the main theorems for primes.

We will bound each term that arises from Vaughan's identity, (2.3), rewritten as,

$$\Lambda = \Lambda_{<V} + \mu_{<U} * L - \mu_{<U} * \Lambda_{<V} * 1 + \mu_{\geqslant U} * \Lambda_{\geqslant V} * 1.$$

To start with, note that

$$\sum_{q<Q} \sum_{n\equiv a_q \pmod q} \Lambda_{<V}(n) \leqslant \sum_{q<Q} \left(\frac{V}{q}+1\right) \log V \ll V \log^2 x + Q \log x$$

which is an acceptable error term when we let $U=V=x^{1/3}$, with $Q<x^{2/3-o(1)}$.

Next we estimate the second term in Vaughan's identity:

$$\begin{aligned}\sum_{\substack{x<n\leqslant 2x\\ n\equiv a \pmod q}} (\mu_{<U} * L)(n) &= \sum_{\substack{u<U\\ (u,q)=1}} \mu(u) \sum_{\substack{x/u<m\leqslant 2x/u\\ m\equiv a/u \pmod q}} L(m)\\ &= \sum_{\substack{u<U\\ (u,q)=1}} \mu(u) \left(\frac{x}{uq}(\log\frac{4x}{u}-1)+O(\log x)\right).\end{aligned}$$

By averaging over all arithmetic progressions a mod q with $(a,q)=1$, we obtain the same estimate for $1/\phi(q)$ times the same sum over n with $(n,q)=1$. Therefore the difference is

$$\sum_{\substack{x<n\leqslant 2x\\ n\equiv a \pmod q}} (\mu_{<U} * L)(n) - \frac{1}{\phi(q)} \sum_{\substack{x<n\leqslant 2x\\ (n,q)=1}} (\mu_{<U} * L)(n) \ll \sum_{\substack{u<U\\ (u,q)=1}} \log x \ll U \log x.$$

Now summing over all $q\leqslant Q$, yields a contribution of $\ll UQ\log x \ll x/(\log x)^A$ for any A.

We will further write

$$\mu_{<U} * \Lambda_{<V} * 1 = \mu_{<U} * \Lambda_{<V} * 1_{<UV} + (\mu * \Lambda)_{<U} * 1_{\geqslant UV},$$

and we now deal with the second part, much as the above, noting that $|(\mu*\Lambda)_{<U}(u)| \leqslant |(1*\Lambda)_{<U}(u)| \leqslant \log u \leqslant \log x$:

$$\begin{aligned}\sum_{\substack{x<n\leqslant 2x\\ n\equiv a \pmod q}} ((\mu*\Lambda)_{<U} * 1_{\geqslant UV})(n) &= \sum_{\substack{u<U\\ (u,q)=1}} (\mu*\Lambda)_{<U}(u) \sum_{\substack{\max\{x/u,UV\}<m\leqslant 2x/u\\ m\equiv a/u \pmod q}} 1\\ &= \sum_{\substack{u<U\\ (u,q)=1}} (\mu*\Lambda)_{<U}(u) \left(\frac{1}{q}(\frac{2x}{u}-\max\{x/u,UV\})+O(\log x)\right),\end{aligned}$$

from which we deduce, by averaging over all arithmetic progressions a mod q with $(a,q)=1$,

$$\sum_{\substack{x<n\leqslant 2x\\ n\equiv a \pmod q}} ((\mu*\Lambda)_{<U}*1_{\geqslant UV})(n) - \frac{1}{\phi(q)} \sum_{\substack{x<n\leqslant 2x\\ (n,q)=1}} ((\mu*\Lambda)_{<U}*1_{\geqslant UV})(n) \ll \sum_{\substack{u<U\\ (u,q)=1}} \log x \ll U \log x.$$

Now summing over all $q\leqslant Q$, yields a contribution of $\ll UQ\log x \ll x/(\log x)^A$ for any A.

We are now left to work with two sums of convolutions:

$$\sum_{\substack{mn \asymp x \\ mn \equiv a \pmod q}} (\mu_{<U} * \Lambda_{<V})(m) 1_{<UV}(n) \quad \text{and} \quad \sum_{\substack{mn \asymp x \\ mn \equiv a \pmod q}} (\Lambda_{\geqslant V} * 1)(m) \mu_{\geqslant U}(n),$$

where $x^{1/3} \ll m, n \ll x^{2/3}$, and each convolution takes the form $\alpha(m)\beta(n)$ where $|\alpha(m)| \leqslant \log m,\ |\beta(n)| \leqslant 1$, α and β satisfy the Siegel-Walfisz criterion,[12] and $\alpha * \beta$ satisfies the necessary sieving condition (since . We can therefore apply Corollary BVwiderange 5.5 to each such sum, and the result follows.

[12] We need to change things a bit since SW is not known for the convolution. Some version can be deduced though with upper bound in terms of the 2-norms of the two original sequences, rather than the 2-norm of the convolution.

6. Preliminary reductions

We now show, through several straightforward manipulations, how we can reduce Theorem 5.4 to proving the following result. As before, $P(q)$ denotes the largest prime factor of q, and now, $p(q)$ denotes the smallest prime factor of q.

Theorem 6.1. *Fix constants $\eta, \delta, A > 0$. Suppose that we have two sequences of complex numbers α_m, $M < m \leqslant 2M$, and $\beta_n, N < n \leqslant 2N$, where β_n satisfies the Siegel-Walfisz condition, and $\alpha * \beta$ satisfies the necessary sieving condition, where $x^{1/3} \ll N \leqslant M \ll x^{2/3}$, with $x = MN$. Suppose also that $N/(yx^\epsilon) < R \leqslant N/x^\epsilon$ and $x^{1/2}/(\log x)^B < QR \leqslant x^{1/2+\eta}$, where $y := x^\delta$. For any $A > 0$, for any integers a, b, b' with $p(abb') > y$, we have*

$$\sum_{\substack{q\in[Q,2Q] \\ D_0<p(q)\leqslant P(q)\leqslant y}} \sum_{\substack{r\in[R,2R], \\ P(r)\leqslant y \\ qr \ \text{squarefree}}} \left| \sum_{\substack{n\equiv a \pmod r \\ n\equiv b \pmod q}} (\alpha*\beta)(n) - \sum_{\substack{n\equiv a \pmod r \\ n\equiv b' \pmod q}} (\alpha*\beta)(n) \right| \ll_A \|\alpha\|\|\beta\| \frac{x^{1/2}}{(\log x)^A}, \tag{6.1}$$

where $D_0 = x^{\epsilon/\log\log x}$.

We will prove this result for any $\eta, \delta > 0$ satisfying $162\eta + 90\delta < 1$.

The proof of this result, and indeed of all the results in the literature of this type, use Linnik's dispersion method. The idea is to express the fact that n belongs to an arithmetic progression using Fourier analysis; summing up over n gives us a main term plus a sum of exponential sums, and then the challenge is to bound each of these exponential sums. In this case we do so by using long-established bounds for exponential sums over finite fields. After some preliminary reductions in this section we will proceed to develop the necessary theory of exponential sums in the following two sections, and then see how these may be used to resolve our problem in the final section. Although this proof is a little technical, it is not especially deep (indeed considerably less deep than previous developments in this area), thanks to the polymath8 project.

Proof. that Theorem 6.1 implies Theorem 5.4. The sum in Theorem 5.4 is over all moduli $d \leqslant x^{1/2+\eta}$ with $P(d) \leqslant y$, with $(d, a) = 1$. The Bombieri-Vinogradov theorem (Theorem 5.2), gives the desired estimate for all $d \leqslant x^{1/2}/(\log x)^B$, so we may restrict our attention to the remaining d. Moreover we may split this range into dyadic intervals, so we may assume that $D < d \leqslant 2D$ where $x^{1/2}/(\log x)^B < D \leqslant x^{1/2+\eta}$. As in the hypothesis, we have that d is squarefree, with $P(d) \leqslant y$.

We now show that we may assume that $(a, d) = 1$ for all such d: Let $m = \prod_{p\leqslant y} p$, and $r = m/(a, m)$. Select an integer b with $b \equiv a \pmod r$ and $b \equiv 1 \pmod{(a, m)}$, which is possible by the Chinese Remainder Theorem. Hence if $(d, a) = 1$ then $(d, b) = 1$ and $b \equiv a \pmod d$, so proving the above estimate for b implies the above estimate for a. The one difference is that $(b, d) = 1$ for all the d in our range.

Next we show that we may restrict our attention to those d with $\nu(d) \leqslant C \log\log x$, that is that have $\leqslant C \log\log x$ prime factors. By Cauchying twice, the square of

$$\sum_{\substack{D<d\leqslant 2D \\ P(d)\leqslant x^\delta \\ \nu(d)>C\log\log x \\ d \text{ squarefree}}} \sum_{n\equiv a \pmod d} |(\alpha * \beta)(n)|, \tag{6.2}$$

Cautwice

is

$$\leqslant \sum_{\substack{D<d\leqslant 2D \\ \nu(d)>C\log\log x}} 1 \cdot \sum_{\substack{D<d\leqslant 2D \\ P(d)\leqslant x^\delta}} \frac{x}{D} \sum_{n\equiv a \pmod d} |(\alpha * \beta)(n)|^2.$$

To bound the first term here we use the Hardy-Ramanujan result that

$$\sum_{\substack{n\leqslant x \\ \nu(n)=k}} 1 \ll \frac{x}{\log x} \frac{(\log\log x + O(1))^{k-1}}{(k-1)!}.$$

To bound the second term we note that $|(\alpha*\beta)(n)|^2 \leqslant \tau(n)(|\alpha|^2 * |\beta|^2)(n)$ by Cauchying, so that

$$\left(\sum_{D<d\leqslant 2D} \sum_{n\equiv a \pmod d} |(\alpha * \beta)(n)|^2 \right)^2 \leqslant \sum_{D<d\leqslant 2D} \sum_{n\equiv a \pmod d} \tau(n)^3 \cdot \sum_{D<d\leqslant 2D} \sum_{n\equiv a \pmod d} (|\alpha|^4 * |\beta|^4)(n);$$

which implies that

$$\sum_{D<d\leqslant 2D} \sum_{n\equiv a \pmod d} |(\alpha * \beta)(n)|^2 \ll \|\alpha\|_8^2 \|\beta\|_8^2 \; x^{3/4} (\log x)^{O(1)}.$$

by using the average sieving condition for $\alpha^4 * \beta^4$ with $A = 0$. hence the quantity in Cautwice (6.2) is

$$\ll \left(\frac{D}{(\log x)^{C(\log C - 1)+1}} \cdot \frac{x}{D} \cdot \|\alpha\|_8^2 \|\beta\|_8^2 \; x^{3/4} (\log x)^{O(1)} \right)^{1/2} \ll \frac{\|\alpha\|_8 \|\beta\|_8 \; x^{7/8}}{(\log x)^A},$$

by taking C sufficiently large. Now $\|\alpha\| \|\beta\| \; x^{1/2} \leqslant \|\alpha\|_8 \|\beta\|_8 \; x^{7/8}$ by Holder's inequality, and we should really state our result in terms of these 8-norms. But for now we will assume that $\|\alpha\|_8 \|\beta\|_8 \; x^{7/8} \ll \|\alpha\| \|\beta\| \; x^{1/2} (\log x)^{O(1)}$ so we can express our result in terms of 2-norms.[13]

The reason for restricting the values of d as in the last paragraph is that it allows us to factor d in a convenient way. If $d = p_1 p_2 \dots p_m$ with $p_1 < p_2 < \dots < p_m$ then select r of the form $p_1 p_2 \dots p_\ell$ as large as possible with $r \leqslant N/x^\epsilon$. Evidently $r > N/(yx^\epsilon) > x^{1/4}$. Note also that $p_\ell > D_0$, else $r \leqslant D_0^\ell \leqslant D_0^{\nu(d)} \leqslant D_0^{C\log\log x} < x^{C\epsilon} < x^{1/4}$ if ϵ were chosen

[13] If we Cauchy instead by taking $|(\alpha * \beta)(n)|^2 \leqslant (1 * |\alpha|^2)(n)(1 * |\beta|^2)(n)$ then

$$\|\alpha * \beta\|^4 \leqslant \left(\sum_n (1 * |\alpha|^2)(n)(1 * |\beta|^2)(n) \right)^2 \leqslant \sum_n \tau(n)(1 * |\alpha|^4)(n) \cdot \sum_n \tau(n)(1 * |\beta|^4)(n).$$

The first term in this product is $\sum_a |\alpha(a)|^4 \sum_{n:\, a|n} \tau(n) \ll \sum_a |\alpha(a)|^4 \tau(a) \cdot N \log N$. We eventually show that $\|\alpha * \beta\| \leqslant \|\alpha\|_8 \|\beta\|_8 \; x^{3/8} (\log x)^{5/4}$.

sufficiently small. Writing $d = qr$ we see that $p(q) > p_\ell > D_0$. Hence there exists R and Q as in the hypothesis of Theorem ReducedRange 6.1 with

$$q \in [Q, 2Q],\ D_0 < p(q) \leqslant P(q) \leqslant y \text{ and } r \in [R, 2R],\ P(r) \leqslant y.$$

We will apply the factorization, with $\gamma = \alpha * \beta$

$$\sum_{n \equiv a \pmod{qr}} \gamma(n) - \frac{1}{\phi(qr)} \sum_{(n,qr)=1} \gamma(n) =$$

$$\sum_{\substack{n \equiv a \pmod q \\ n \equiv a \pmod r}} \gamma(n) - \frac{1}{\phi(q)} \sum_{\substack{(n,q)=1 \\ n \equiv a \pmod r}} \gamma(n) + \frac{1}{\phi(q)} \left(\sum_{\substack{(n,q)=1 \\ n \equiv a \pmod r}} \gamma(n) - \frac{1}{\phi(r)} \sum_{\substack{(n,q)=1 \\ (n,r)=1}} \gamma(n) \right)$$

For the first terms we apply Theorem ReducedRange 6.1 with $b = a$, for each $b' \pmod q$ with $(b', q) = 1$, and average, to obtain by the triangle inequality

$$\sum_{\substack{q \in [Q,2Q] \\ D_0 < p(q) \leqslant P(q) \leqslant y}} \sum_{\substack{r \in [R,2R], \\ P(r) \leqslant y}} \left| \sum_{\substack{n \equiv a \pmod q \\ n \equiv a \pmod r}} \gamma(n) - \frac{1}{\phi(q)} \sum_{\substack{(n,q)=1 \\ n \equiv a \pmod r}} \gamma(n) \right| \ll_A \|\alpha\| \|\beta\| \frac{x^{1/2}}{(\log x)^A}, \tag{6.3}$$

straw-3

For the second terms we take absolute values and sum over q and r separately to obtain the upper bound

$$\leqslant \sum_{q \leqslant x^{1/2}} \frac{1}{\phi(q)} \sum_{r \leqslant x^{1/2-\epsilon}} \left| \sum_{\substack{(n,q)=1 \\ n \equiv a \pmod r}} \gamma(n) - \frac{1}{\phi(r)} \sum_{\substack{(n,q)=1 \\ (n,r)=1}} \gamma(n) \right|.$$

Now in Lemma SWcoprime 6.2 below, we show that $\beta_n 1_{(n,q)=1}$ satisfies a Siegel-Walfisz condition, since β_n does. By Theorem BFI2 5.2 (with α and β replaced by $\alpha_n 1_{(n,q)=1}$ and $\beta_n 1_{(n,q)=1}$, respectively), we deduce that $\gamma_n 1_{(n,q)=1}$ satisfies a Bombieri-Vinogradov Theorem. Substituting this into the last equation gives

$$\ll_A \sum_{q \leqslant x^{1/2}} \frac{1}{\phi(q)} \|\alpha\| \|\beta\| \frac{x^{1/2}}{(\log x)^{A+1}}$$

and the result follows. □

SWcoprime

Lemma 6.2. *If β_n satisfies a Siegel-Walfisz condition then for any $m \geqslant 1$ we have*

$$\left| \sum_{\substack{n \equiv a \pmod q \\ (n,m)=1}} \beta_n - \frac{1}{\phi(q)} \sum_{n:\ (n,mq)=1} \beta_n \right| \ll \tau(m) \|\beta\| \frac{N^{\frac{1}{2}}}{(\log N)^C}.$$

Proof. of Lemma SWcoprime *6.2* We may assume that $q \leqslant (\log N)^{2C}$ else, by Cauchying,

$$\left| \sum_{\substack{n \equiv a \pmod q \\ (n,m)=1}} \beta_n \right|^2 \leqslant \sum_{n \equiv a \pmod q} 1 \cdot \sum_n |\beta_n|^2 \leqslant \frac{N}{q} \|\beta\|^2;$$

And then, by averaging this over all a with $(a, q) = 1$, one deduces the result provided $q > (\log N)^{2C}$.

Now for an arbitrary m we decompose the sum as

$$\sum_{\substack{n \equiv a \pmod q \\ (n,m)=1}} \beta_n = \sum_{d|m} \mu(d) \sum_{\substack{n \equiv a \pmod q \\ d|n}} \beta_n$$

and, Cauchying, the square of the sum here, over $d \geqslant (\log N)^{2C}$, is

$$\leqslant \tau(m) \sum_{\substack{d|m \\ d \geqslant (\log N)^{2C}}} \|\beta\|^2 \frac{N}{dq} \leqslant \tau(m)^2 \|\beta\|^2 \frac{N}{(\log N)^{2C}}.$$

For the smaller d we use the identity

$$\sum_{\substack{n \equiv a \pmod q \\ n \equiv 0 \pmod d}} \beta_n = \sum_{r|d} \mu(r) \sum_{b:\ (b,r)=1} \sum_{\substack{n \equiv a \pmod q \\ n \equiv b \pmod r}} \beta_n.$$

Applying the Siegel-Walfisz condition for each such modulus qr (with C replaced by $6C$) we obtain an upper bound

$$\sum_{d < (\log N)^{2C}} \sum_{r|d} \phi(r) \|\beta\| \frac{N^{\frac{1}{2}}}{(\log N)^{6C}} \ll \|\beta\| \frac{N^{\frac{1}{2}}}{(\log N)^{C}}.$$

□

7. Complete exponential sums

exp-sec

In section HLheuristic 3.3 we developed some notation for exponentials, for example $e_q(a) = e(\frac{a}{q})$. For a rational number a/b we have to be a little more careful in defining $e_q(a/b)$: If b has a factor in common with q then define $e_q(a/b) = 0$. If $(b, q) = 1$, select $c \pmod q$ so that $bc \equiv a \pmod q$ and then define $e_q(a/b) = e_q(c)$, and note that this is well-defined.

In this section we will obtain upper bounds for $\sum_n e_q(f(n))$ where $f(x)$ is a rational function, and the sum is over all $n \in (\mathbb{Z}/q\mathbb{Z})^*$ for which the denominator of $f(n)$ is coprime to q. By a rational function we mean that $f(x) = P(x)/Q(x)$ for some $P, Q \in \mathbb{Z}[x]$ and we define $\deg f = \max(\deg P, \deg Q)$. We will then derive such bounds, for squarefree q, from bounds for primes p, using the following consequence of the Chinese remainder theorem yields: If $q_1, \dots, q_k$ are pairwise coprime natural numbers, then for any integer a and $q := q_1 \dots q_k$ we have

$$e_q(a) = \prod_{j=1}^{k} e_{q_j}\left(\frac{a}{(q/q_j)}\right). \qquad (7.1)$$

CRTgeneral

In particular this implies that

$$\sum_{n\in\mathbb{Z}/q\mathbb{Z}} e_q(f(n)) = \prod_{p|q} \sum_{n\in\mathbb{Z}/p\mathbb{Z}} e_p\left(\frac{f(n)}{(q/p)}\right). \qquad (7.2)$$

CRTexpsum

7.1. Two special cases.

If $f(x) = ax + b$ then

$$\sum_x e_q(ax+b) = e_q(b)\sum_{j=0}^{q-1} e\left(\frac{aj}{q}\right) = \begin{cases} q\ e_q(b) & \text{if } q \text{ divides } a; \\ 0 & \text{otherwise ,} \end{cases}$$

the discrete analogue of (expintegral 3.6). If $f(x) = c/(x+d)$ with $c \not\equiv 0 \pmod p$, then we make the change of variable $x = c/y - d$, which is a bijection from $x \in \mathbb{F}_p\backslash\{-d\} \to y \in \mathbb{F}_p\backslash\{0\}$, so that

$$\sum_{x\in\mathbb{Z}/p\mathbb{Z}} e_p\left(\frac{c}{x+d}\right) = \sum_{y\neq 0} e_p(y) = -1. \qquad (7.3)$$

inversesum

This can be combined with (CRTexpsum 7.2) to deduce the following (see [zhang 38, Proposition 11]):

dork

Lemma 7.1. *Let d_1, d_2 be natural numbers with $[d_1, d_2]$ square-free, and let c_1, c_2, l_1, l_2 be integers. Then*

$$\left|\sum_{n\in\mathbb{Z}/[d_1,d_2]\mathbb{Z}} e_{d_1}\left(\frac{c_1}{n+l_1}\right) e_{d_2}\left(\frac{c_2}{n+l_2}\right)\right| \leqslant (c_1, d_1')(c_2, d_2')(d_1, d_2)$$

where $d_i' := d_i/(d_1, d_2)$ for $i = 1, 2$.

Proof. We will prove the p-component of this bound for each prime divisor p of $[d_1, d_2]$, and then deduce the full result using the Chinese Remainder Theorem, as in (CRTexpsum 7.2), as the right-hand side of our bound is a multiplicative function.

The bound is trivial if (c_1, d_1'), (c_2, d_2'), or (d_1, d_2) is equal to p, since there are no more than p terms in the sum, so we may assume without loss of generality that $d_1 = p$, $d_2 = 1$, and c_1 is coprime to p. The result then follows immediately from (inversesum 7.3). □

Notice that this bound is probably improvable, since we have not exploited any possible cancelation in the sums for the primes that divide (d_1, d_2).

7.2. **The deeper theory of exponential sums.** In general there is some significant cancelation in exponential sums, and we now discuss those deeper results (due primarily to Weil) that we need. In fact one can as easily state rather general results, but *we will only use those results when f takes the form*

$$\frac{a}{x} + cx, \quad \text{or} \quad \frac{a}{x} + \frac{b}{x+\ell} + cx, \quad \text{or} \quad \frac{a}{x(x+k)} + \frac{b}{(x+\ell)(x+\ell+k)} + cx \qquad (7.4)$$ needExpSum

for any given integers a, b, c, k, ℓ, with $k, \ell \not\equiv 0 \pmod p$.

The Weil conjectures for curves weil [37] (proven for arbitrary varities by Deligne in WeilII [7]), imply "square root cancellation" for various natural exponential sums over finite fields (note that $\mathbb{Z}/p\mathbb{Z}$ is isomorphic to the finite field $\mathbb{F}_p$).

prime-exp **Lemma 7.2.** *If p is prime p and $f(x)$ is a rational function in $\mathbb{F}_p[X]$ of degree d, with $1 \leqslant d < p$, then*

$$\left| \sum_{x \in \mathbb{F}_p} e_p\left(f(x)\right) \right| \ll d\sqrt{p}. \qquad (7.5)$$ pqp

This bound follows from the Weil conjectures applied to $y^p - y = P(x)/Q(x)$ in $\mathbb{F}_p$. An elementary proof based on Stepanov's method may also be found in cochrane [6].

Setting $d < p$ is natural, in that, for examples like $f(x) \ = g(x)^p - g(x) + c$, we see that $f(n) \pmod p$ is constant by Fermat's little theorem, so there would be no cancelation in the exponential sum.

We do not need to obtain the full square root cancellation in (pqp 7.5) in our work here: Any bound of the form p^c for some fixed $c < \frac{2}{3}$ would suffice in our argument. This gives hope that there may be a more elementary argument.

We next extend Lemma prime-exp 7.2 to square-free moduli:

For q an integer and $f(x)$ a rational function, define (q, f) to be the largest integer m dividing q for which $f(x) \equiv 0 \pmod m$. It is not difficult to show that if $f(x)$ is a rational function for which $f'(x) \equiv 0 \pmod p$ then $f(x) \equiv g(x)^p \pmod p$ for some rational function $g(x)$.[14] Hence if $p > \deg f$ then $f(x) \equiv c^p \equiv c \pmod p$, for some constant c. This generalizes to: If $f'(x) \equiv 0 \pmod q$, when q is squarefree and $\deg f < p$ for every prime p dividing q, then $f(x) \equiv c \pmod q$.

[14] By induction on $\deg P + \deg Q$ where $f = P/Q$: If $\deg P \geqslant \deg Q$ then we show that $p | \deg P - \deg Q$ so $P/Q - h^p$ has lower degree. Otherwise replace f by $1/f$.

weil **Proposition 7.3.** *Let q be a squarefree positive integer, and let $f \in \mathbb{Z}(X)$ be a rational function of degree d. There exists a constant $A = A_d$, for which*

$$\left| \sum_{n \in \mathbb{Z}/q\mathbb{Z}} e_q(f(n)) \right| \leqslant \tau(q)^A q^{1/2} \frac{(f', q)}{(f'', q)^{1/2}}.$$

For $f(t) := at + b/t$ we get the bound $\ll \tau(q)^A q^{1/2}(a, b, q)/(2b, q)^{1/2}$, slightly weaker than Weil's Kloosterman sum bound.

Proof. We will prove the result for $q = p$ prime, and then the result follows in general, by (CRTexpsum 7.2), as the right hand side of the result is a multiplicative function in q.

Note that the sum has p terms, each of absolute value 1, so the sum has absolute value $\leqslant p$, by the triangle inequality. Therefore we may henceforth assume that $p > \deg f$, since the result follows for the finitely many primes $p \leqslant \deg f$, simply by taking A sufficiently large. It also follows when $p|f'$ since then $p|f''$ and so the upper bound is $p^{1/2}(f', p)/(f'', p)^{1/2} = p$.

Hence we may assume that $p \nmid f'$. If $p \nmid f''$ then the result follows from Lemma prime-exp 7.2. If $p|f''$ then, as we noted above, $f'(x) \equiv c \pmod p$ for some integer c. But then $g(x) = f(x) - cx$ satsifies $g'(x) \equiv 0 \pmod p$ and so there exists an integer d for which $g(x) \equiv d \pmod p$; that is $f(x) \equiv cx + d \pmod p$. But then the sum $= 0$ and the result follows. □

incompsec

8. Incomplete exponential sums

In the previous section we bounded "complete" exponential sums $\sum_n e(f(n)/q)$ in which the summation variable n ranges over the whole cyclic group $\mathbb{Z}/q\mathbb{Z}$ or, equivalently, the integers in an interval of length q. For arithmetic applications we typically need to obtain non-trivial bounds when n varies over a shorter interval, an "incomplete sum". To do this we use the bounds obtained for the complete sums, by invoking what is, in effect, the *discrete Fourier transform*:

$$\hat{f}(h) := \sum_{b \pmod q} f(b) e_q(hb), \tag{8.1}$$

ftq-def

for any function f of period q. One begins with the trivial observation that

$$\frac{1}{q} \sum_{b \pmod q} e_q((m-a)b) = \begin{cases} 1 & \text{if } m \equiv a \pmod q, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, summing $f(a)$, times the characteristic function $I(a)$ for the interval I, we obtain

$$\begin{aligned} \sum_{m \in I} f(m) = \sum_m I(m) f(m) &= \sum_{m \pmod q} I(m) \sum_{a \pmod q} f(a) \cdot \frac{1}{q} \sum_{b \pmod q} e_q((m-a)b) \\ &= \frac{1}{q} \sum_{b \pmod q} \Bigg(\sum_{m \pmod q} I(m) e_q(mb) \Bigg) \Bigg(\sum_{a \pmod q} f(a) e_q(-ab) \Bigg) \\ &= \frac{1}{q} \sum_{b \pmod q} \hat{I}(b) \hat{f}(-b), \end{aligned}$$

which can be viewed as an example of Plancherel's formula. Typically we might expect to have a "main term" given by $b = 0$; that is

$$\frac{1}{q}\, \hat{I}(0) \hat{f}(0) = |I| \cdot \frac{1}{q} \sum_{a \pmod q} f(a),$$

the length of the interval, times the average of f. In order to prove this is dominant we will need to have some control of the other terms. The Fourier transform of the characteristic function for an interval does have some considerable cancellation: If the interval is $[x, x+M)$ and $1 \leqslant |b| \leqslant q/2$ then

$$\hat{I}(b) = \sum_{j=0}^{M-1} e_q(b(x+j)) = e_q(bx) \cdot \frac{e_q(bM) - 1}{e_q(b) - 1}.$$

The numerator has absolute value $\leqslant 2$ and, using the Taylor expansion, the denominator has absolute value $\asymp |b|/q$. Hence

$$|\hat{I}(b)| \ll \min\{M, q/|b|\},$$

and so we deduce that

$$\begin{aligned}\left|\sum_{m\in I} f(m) - |I|\,\frac{\hat{f}(0)}{q}\right| &\ll \frac{M}{q}\sum_{1\leqslant |b|\leqslant q/M} |\hat{f}(b)| + \sum_{q/M<|b|\leqslant q/2} \frac{|\hat{f}(b)|}{b}\\ &\ll \sum_{\substack{0\leqslant j\leqslant J\\ H_j:=2^jq/M}} \frac{1}{H_j}\sum_{1\leqslant |h|\leqslant H_j} |\hat{f}(b)|\end{aligned}$$

where J is the largest integer for which $2^J < M$. We deduce that

$$\left|\sum_{m\in I} f(m) - |I|\,\frac{\hat{f}(0)}{q}\right| \ll \log M \cdot \max_{b\not\equiv 0 \pmod q} |\hat{f}(b)| \tag{8.2}$$

FouBound

For the example $f = \sum_i c_i 1_{m\equiv a_i \pmod q}$ where I is the interval of length $(M, 2M]$, we obtain the bound

$$\left|\sum_i c_i \left(\sum_{\substack{m\asymp M\\ m\equiv a_i \pmod q}} 1 - \frac{M}{q}\right)\right| \ll \sum_{\substack{0\leqslant j\leqslant J\\ H_j:=2^jq/M}} \frac{1}{H_j}\sum_{1\leqslant |h|\leqslant H_j} \left|\sum_i c_i e_q(a_ih)\right|. \tag{8.3}$$

ExponExpan

We will insert the estimates of Proposition 7.3 into (8.2) and (8.3) to obtain "square-root cancellation" for incomplete exponential sums of the form $|\sum_n e_q(f(n))|$ for various moduli q, with the sum over n in an interval of length $N < q$ (as in [38]). However, Graham and Ringrose [17] showed that we can improve the (analogous) incomplete character sum bounds for smaller N when q is smooth, and we do so here, following [30], for incomplete exponential sums.

inc **Proposition 8.1.** *Let q be a square-free integer, and let $f = \frac{P}{Q}$ with $P, Q \in \mathbb{Z}[X]$ and $\deg(P) < \deg(Q) \ll 1$. Suppose that $(q, f) = 1$, and write $\sum_n$ for $\sum_{n\asymp N}$.*

(i) We have the bound

$$\left|\sum_n e_q(f(n))\right| \ll \left(\frac{N}{q} + \log q\right) \tau(q)^A q^{1/2}. \tag{8.4}$$

vdc-0

(ii) If $q = q_1q_2$ and $N < q$ then

$$\left|\sum_n e_q(f(n))\right| \ll \left(q_1^{1/2} + q_2^{1/4}\right) \tau(q)^A (\log q) N^{1/2}. \tag{8.5}$$

vdc-1

(iii) If q is y-smooth and $N < q$ then

$$\left|\sum_n e_q(f(n))\right| \ll \tau(q)^A (qy)^{1/6} (\log q) N^{1/2}.$$

Proof. We may assume that q has no prime factors $\leqslant \deg f$, else one can factor $q = q_0q'$ where q_0 is the product of all the prime factors of q that are $\leqslant \deg f$, split the summation over n into residue classes mod q_0, and then apply the result mod q' to each of the subsums. This then implies that $(q, f'+h) = (q, f'') = 1$ else, as $p > \deg f$, we see that $f \equiv c$ or $ct + d \pmod p$, but this is impossible by the hypothesis.

Now by Proposition weil 7.3 we have, for $F(n) = e_q(f(n))$, that

$$\hat{F}(h) = \sum_{b \pmod q} e_q(f(b) + hb) \leqslant \tau(q)^A q^{1/2}$$

for every h. Therefore (FouBound 8.2) yields part (i).

For part (ii) we may assume

$$q_1 \leqslant N \leqslant q_2$$

else if $N < q_1$ we have the trivial bound $\leqslant N < (q_1 N)^{1/2}$, and if $N > q_2$ then (i) implies the result since $q^{1/2} = (q_1 q_2)^{1/2} < (q_1 N)^{1/2}$.

The main idea will be to reduce our incomplete exponential sum mod q, to a sum of incomplete exponential sums mod q_2. Now

$$e_q(f(n + kq_1)) = e_{q_1}(f(n)/q_2)\ e_{q_2}(f(n + kq_1)/q_1)$$

so that, by a simple change of variable, we have

$$\sum_n e_q(f(n)) = \sum_n e_q(f(n + kq_1))) = \sum_n\ e_{q_1}(f(n)/q_2)\ e_{q_2}(f(n + kq_1)/q_1).$$

Now, if we sum this over all $k, 1 \leqslant k \leqslant K := \lfloor N/q_1 \rfloor$, then we have

$$K \sum_n e_q(f(n)) = \sum_n e_{q_1}(f(n)/q_2)\ \sum_{k=1}^{K} e_{q_2}(f(n + kq_1)/q_1),$$

and so

$$\begin{aligned} \left| K \sum_n e_q(f(n)) \right|^2 &\leqslant \left(\sum_n \left| \sum_{k=1}^{K} e_{q_2}(f(n + kq_1)/q_1) \right| \right)^2 \\ &\ll N \sum_n \left| \sum_{k=1}^{K} e_{q_2}(f(n + kq_1)/q_1) \right|^2 = N \sum_{1 \leqslant k, k' \leqslant K} \sum_n e_{q_2}(g_{k,k'}(n)), \end{aligned}$$

where $g_{k,k'}(n) := (f(n + kq_1) - f(n + k'q_1))/q_1 \pmod{q_2}$ if $n + kq_1,\ n + k'q_1 \in I$, and $g_{k,k'}(n) := 0$ otherwise. If $k = k'$ then $g_{k,k}(n) = 0$, and so these terms contribute $\leqslant KN^2$.

We now prove that $(q_2, g_{k,k'}) = (q_2, k - k')$: Suppose that p divides $(q_2, g_{k,k'})$, so that $p \nmid q_1$. Now $f(n + kq_1) \equiv f(n + k'q_1) \pmod p$ for all n, and so if $p \nmid (k - k')$ then $p | f(a) - f(0)$ for all a. Now if $f(a) \equiv c \pmod p$ for every $a \pmod p$, and $p > \deg f$ then $f(x) \equiv c \pmod p$, contradicting the hypothesis. On the other hand if $p | (k - k')$ then $p | g_{k,k'}$.

Now part (i) yields a bound (taking q there to be $q_2/(q_2, k-k')$) for the above which is

$$\ll KN^2 + N\tau(q_2)^A \sum_{1\leqslant k\neq k'\leqslant K} \left(\frac{N(q_2,k-k')^{1/2}}{q_2^{1/2}} + \frac{q_2^{1/2}\log q_2}{(q_2,k-k')^{1/2}} \right)$$

$$\ll K^2 N\left(q_1 + q_2^{1/2}\tau(q_2)^A \left(\frac{1}{K}\sum_{j=1}^{K} (j,q_2)^{1/2} + \log q_2 \right)\right),$$

as $N \leqslant q_2$ (so that $N/q_2^{1/2} \leqslant q_2^{1/2}$) and since each j appears as a difference $|k-k'|$ at most $2K$ times. The result of part (ii) follows since

$$\left(\frac{1}{K}\sum_{j=1}^{K} (j,q_2)^{1/2}\right)^2 \leqslant \frac{1}{K}\sum_{j=1}^{K}(j,q_2) \leqslant \frac{1}{K}\sum_{j=1}^{K}\sum_{d|q_2,\ d|j} d \leqslant \frac{1}{K}\sum_{d|q_2} d\cdot\frac{K}{d} \leqslant \tau(q_2).$$

For part (iii) we observe that if q is y-smooth then it has divisors in any interval of multiplicative length y. In particular we can select q_1 in the interval $q^{1/3}y^{-2/3} < q_1 \leqslant (qy)^{1/3}$ so that $q^{2/3}y^{-1/3} < q_2 \leqslant (qy)^{2/3}$, and hence part (ii) implies our result. □

8.1. **Some specific incomplete sums.** In our particular application, we need only the following special case of the above proposition, which is a strengthening of zhang [38, Lemma 11]:

dons **Corollary 8.2.** *Let d_1, d_2 be square-free integers, with $(c_1,d_1) = (c_2,d_2) = 1$, and let $h := [d_1,d_2]/(d_1,d_2)$. For any $a \pmod q$, we have*

$$\left| \sum_{n\equiv a \pmod q} e_{d_1}\left(\frac{c_1}{n+l_1}\right) e_{d_2}\left(\frac{c_2}{n+l_2}\right) \right| \ll \left(\frac{[d_1,d_2]}{(q,[d_1,d_2])}\right)^{1/2+o(1)} + \frac{N}{[q,h]}.$$

If d_1 and d_2 are also y-smooth then

$$\left| \sum_{n\equiv a \pmod q} e_{d_1}\left(\frac{c_1}{n+l_1}\right) e_{d_2}\left(\frac{c_2}{n+l_2}\right) \right| \ll y^{1/6}\left(\frac{[d_1,d_2]}{(q,[d_1,d_2])}\right)^{1/6+o(1)} \left(\frac{N}{q}\right)^{1/2} + \frac{N}{[q,h]}.$$

Proof. Writing $n = a + mq$ and $q = rq_0$, the sum is now over an interval of values of m of length $M = N/q + O(1)$. The first exponential in the summand becomes

$$e_{d_1}\left(\frac{c_1}{n+l_1}\right) = e_{q_1}\left(\frac{c_1}{(a+l_1)d_1/q_1}\right) e_{d_1/q_1}\left(\frac{c_1 Q/q_1}{m+(a+l_1)Q}\right)$$

where $q_1 = (q,d_1)$ and $q = q_1 r_1$, with $Qq \equiv 1 \pmod{d_1/q_1}$. Note that the first term here is fixed as m varies. An analogous identity is true for the second term. Hence we can write

$$\left| \sum_{n\equiv a \pmod q} e_{d_1}\left(\frac{c_1}{n+l_1}\right) e_{d_2}\left(\frac{c_2}{n+l_2}\right) \right| = \left| \sum_m e_{d_1/q_1}\left(\frac{c_1'}{m+l_1'}\right) e_{d_2/q_2}\left(\frac{c_2'}{m+l_2'}\right)\right|$$

with $(c_i', d_i/q_i) = (c_i, d_i/q_i) = 1$ for each i.

Now on each subinterval of length $[d_1/q_1, d_2/q_2]$ $(= [d_1,d_2]/(q,[d_1,d_2]))$ we have a complete exponential sum which is of size $\leqslant (d_1/q_1, d_2/q_2)$ $(= (d_1,d_2)/(q,d_1,d_2))$ by Lemma 7.1. Therefore this yields an upper bound, in total, of

$$\ll \frac{N/q}{[d_1,d_2]/(q,[d_1,d_2])}(d_1,d_2)/(q,d_1,d_2) = \frac{N}{[q,h]}$$

where $h := [d_1,d_2]/(d_1,d_2)$.

The remaining part of the sum is an incomplete sum modulo $[d_1,d_2]/(q,[d_1,d_2])$ of length no longer than the modulus. The first result now follows immediately from Proposition 8.1(i). The second result follows Proposition 8.1(iii). □

dons2 **Corollary 8.3.** *Suppose that grq_1 and grq_2 are squarefree integers, and that $q|g$. For any $a \pmod q$ we have*

$$\left| \sum_{n\equiv a \pmod q} e_r\left(\frac{c_r}{n}\right) e_g\left(\frac{c_g}{n}\right) e_{q_1}\left(\frac{c_1}{n}\right) e_{q_2}\left(\frac{c_2}{n+l}\right) \right| \ll (rq_1q_2(g/q))^{1/2+o(1)} + \frac{(c_r,r)}{r}\frac{N}{q}.$$

Moreover, if r, g, q_1, q_2 are all y-smooth then we also have the upper bound

$$\ll y^{1/6}\,(rq_1q_2(g/q))^{1/6+o(1)}\left(\frac{N}{q}\right)^{1/2} + \frac{(c_r,r)}{r}\frac{N}{q}.$$

Proof. We can combine any two such exponentials $e_r\left(\frac{a}{n}\right)e_s\left(\frac{b}{n}\right) = e_{rs}\left(\frac{c}{n}\right)$ with $(r,s)=1$ by taking $c = s(a/s)_r + r(a/r)_s$ (with $(b)_q$ the least residue of $b \pmod q$), and so $(c,rs) = (a,r)(b,s)$. To apply the previous corollary we need to replace every $e_d(c/n)$ by $e_{d/(c,d)}(c/(c,d)n)$, and therefore the summand becomes, say, $e_{d_1}\left(\frac{c_1}{n}\right)e_{d_2}\left(\frac{c_2}{n+l}\right)$, where $d_1 = \frac{r}{(c_r,r)}\frac{g}{(c_g,g)}\frac{q_1}{(c_1,q_1)}$ and $d_2 = \frac{q_2}{(c_2,q_2)}$. We use the inequality $\frac{[d_1,d_2]}{(q,[d_1,d_2])} \leqslant \frac{d_1d_2}{(q,d_1d_2)} \leqslant rq_1q_2(g/q)$. We also note that h $(= \frac{[d_1,d_2]}{(d_1,d_2)})$ is divisible by $\frac{r}{(c_r,r)}$ and that $(q,r)=1$ (as $q|g$ and $(g,r)=1$), so that $[q,h] = q\frac{h}{(q,h)} \geqslant q\frac{r}{(c_r,r)}$. □

nential2-sec

8.2. **More complicated exponential sums.** In this section we will prove a couple of rather complicated exponential sum estimates that will be needed in the final proof. We begin by defining the following exponential. Suppose that $k, h, r, g, \ell_1, \ell_2, a, b_1, b_2$ are given integers, such that $rg\ell_1\ell_2$ is squarefree and coprime with ab_1b_2. Then define $\Phi_k(h,n;r,g,\ell_1,\ell_2) = \Phi(n)$ by

$$\Phi(n) := e_r\left(\frac{ah}{ng\ell_1\ell_2}\right) e_g\left(\frac{b_1h}{nr\ell_1\ell_2}\right) e_{\ell_1}\left(\frac{b_1h}{nrg\ell_2}\right) e_{\ell_2}\left(\frac{b_2h}{(n+kr)rg\ell_1}\right), \qquad (8.6)$$ phi-def

when $n + kr \equiv b_2n/b_1 \pmod g$, and $\Phi_k(h,n;r,\ell_1,\ell_2) := 0$ otherwise. Notice that $\Phi(n)$ can be rewritten (inconveniently) as an exponential of the form $e_q(t)$ for some integer t, where $q = rg\ell_1\ell_2$.

We are interested in bounding the following exponential sum:

$$S_{k,r}(h,j,g,\ell_1,\ell_2,m_1,m_2) = \sum_n \Phi_k(h,n;r,g,\ell_1,\ell_2)\overline{\Phi_k(j,n;r,g,m_1,m_2)}. \qquad (8.7)$$ S-def

Note that if $(b_2-b_1,g)\nmid k$ then there are no solutions n to $n+kr\equiv b_2n/b_1 \pmod g$, and so $S_{k,r}(h,j,g,\ell_1,\ell_2,m_1,m_2)=0$.

We begin with an estimate on this exponential sum zhang [38, (12.5)], followed by one that appears in polymath8 [30]:

expsumLemma **Proposition 8.4.** *Assume that that $rg\ell_1\ell_2$ and rgm_1m_2 are both squarefree and coprime with ab_1b_2, and that $(b_2-b_1,g)|k$. Let $q_0:=g/(b_2-b_1,g)$. If $r\asymp R$ and $\ell_1,\ell_2,m_1,m_2\asymp Q/g$ then*

$$|S_{k,r}(h,j,g,\ell_1,\ell_2,m_1,m_2)|\ll (R(g/q_0))^{1/2}\,(Q/g)^2x^{o(1)}+\frac{(\Delta,r)}{R}\frac{N}{q_0}, \tag{8.8}$$ exse

where $\Delta:=hm_1m_2-j\ell_1\ell_2$. If $m_1=\ell_1$ and $gr\ell_1\ell_2m_2$ is y-smooth, then we can take $\Delta=hm_2-j\ell_2$ and get the bound

$$|S_{k,r}(h,j,\ell_1,\ell_2,\ell_1,m_2)|\ll (R(g/q_0))^{1/6}\,y^{1/6}\,(Q/g)^{1/2}\left(\frac{N}{q_0}\right)^{1/2}x^{o(1)}+\frac{(\Delta,r)}{R}\frac{N}{q_0}. \tag{8.9}$$ exse-2

Proof. If $(b_2-b_1,g)|kr$ and $n+kr\equiv b_2n/b_1 \pmod g$, then n belongs to a single congruence class mod q_0, call it $t \pmod{q_0}$.

We begin by simplifying the expression for $\Phi_k(h,n;r,g,\ell_1,\ell_2)\overline{\Phi_k(j,n;r,g,m_1,m_2)}$. when $n+kr\equiv b_2n/b_1 \pmod g$: The exponent for e_r is

$$\frac{ah}{ng\ell_1\ell_2}-\frac{aj}{ngm_1m_2}=\frac{a\Delta}{ng\ell_1\ell_2m_1m_2}.$$

One makes a similar calculation for e_r. We create an exponential mod $[\ell_1,m_1]$ from the exponentials mod ℓ_1 and mod m_1, and therefore we have exponent

$$\frac{b_1h}{nrg\ell_2}\cdot\frac{[\ell_1,m_1]}{\ell_1}-\frac{b_1j}{nrgm_2}\cdot\frac{[\ell_1,m_1]}{m_1}=\frac{b_1\Delta/(\ell_1,m_1)}{nrg\ell_2m_2}.$$

We perform the analogous calculation mod $[\ell_2,m_2]$.

Hence we have shown that $S_{k,r}$ is the sum, over n in our interval for which $n+kr\equiv b_2n/b_1 \pmod g$, of

$$e_r\left(\frac{a\Delta}{ng\ell_1\ell_2m_1m_2}\right)e_g\left(\frac{b_1\Delta}{nr\ell_1\ell_2m_1m_2}\right)e_{[\ell_1,m_1]}\left(\frac{b_1\Delta/(\ell_1,m_1)}{nrg\ell_2m_2}\right)e_{[\ell_2,m_2]}\left(\frac{b_2\Delta/(\ell_2,m_2)}{(n+kr)rg\ell_1m_1}\right).$$

The first result then follows from Corollary dons2 8.3 (i), and the crude bound $[\ell_1,m_1][\ell_2,m_2]\leqslant \ell_1m_1\ell_2m_2$. (We can deduce a similar result from Corollary dons2 8.3 (ii).) The second result similarly follows from Corollary dons2 8.3 (ii) since $(\ell_1,r)=1$ and from the crude bound $[\ell_1,\ell_1][\ell_2,m_2]\leqslant \ell_1\ell_2m_2$. (Again, a similar result from Corollary dons2 8.3 (i).)

□

typei-ii-sec

9. The Grand Finale

Our goal is to establish Theorem ReducedRange 6.1:

Theorem ReducedRange **6.1** *Fix constants $\eta, \delta, A > 0$. Suppose that we have two sequences of complex numbers α_m, $M < m \leqslant 2M$, and $\beta_n, N < n \leqslant 2N$, where β_n satisfies the Siegel-Walfisz condition, and that $\alpha * \beta$ satisfies the necessary sieving condition, where $x^{1/3} \ll N \leqslant M \ll x^{2/3}$, with $x = MN$. Suppose also that $N/(yx^{\epsilon}) < R \leqslant N/x^{\epsilon}$ and $x^{1/2}/(\log x)^B < QR \leqslant x^{1/2+\eta}$, where $y := x^{\delta}$. For any $A > 0$, for any integers a, b, b' with $p(abb') > y$, we have*

$$\sum_{\substack{q \in [Q,2Q] \\ D_0 < p(q) \leqslant P(q) \leqslant y}} \sum_{\substack{r \in [R,2R], \\ P(r) \leqslant y \\ qr \text{ squarefree}}} \left| \sum_{\substack{n \equiv a \pmod r \\ n \equiv b \pmod q}} (\alpha * \beta)(n) - \sum_{\substack{n \equiv a \pmod r \\ n \equiv b' \pmod q}} (\alpha * \beta)(n) \right| \ll_A \|\alpha\| \|\beta\| \frac{x^{1/2}}{(\log x)^A}, \tag{9.1}$$

straw-2

where $D_0 = x^{\epsilon/\log\log x}$. In fact it suffices to take $162\eta + 90\delta < 1$.

We chose r to be slightly less than N to ensure that the constraint $n \equiv a \pmod r$ still incorporates some non-trivial averaging in the α weight, which allows one to profitably use the dispersion method of Linnik. We chose q to be free of small prime factors, so that two such q's are likely to be coprime.

Throughout the argument below, the restrictions on m, n, q, r from the hypothesis will be taken as given.

In the left-hand side of (straw-2 9.1) we replace the absolute value in the (q, r) term by a complex number $c_{q,r}$ of absolute value 1, and then each $(\alpha * \beta)(\ell) = \sum_{mn=\ell} \alpha(m)\beta(n)$ to obtain, after a little re-arranging:

$$\sum_r \sum_m \alpha(m) \left(\sum_q \sum_{n:\ mn \equiv a \pmod r} c_{q,r} \beta(n) (1_{mn \equiv b \pmod q} - 1_{mn \equiv b' \pmod q}) \right).$$

By the Cauchy-Schwarz inequality the square of this is

$$\leqslant \sum_r \sum_m |\alpha(m)|^2 \leqslant R \|\alpha\|^2$$

times

$$\sum_r \sum_m \left| \sum_q \sum_{n:\ mn \equiv a \pmod r} c_{q,r} \beta(n) (1_{mn \equiv b \pmod q} - 1_{mn \equiv b' \pmod q}) \right|^2. \tag{9.2}$$

sq

When we expand the sum, we obtain the sum of four terms of the form

$$\pm\sum_r\sum_m\sum_{q_1,q_2}\sum_{\substack{n_1,n_2\\ mn_1\equiv mn_2\equiv a \pmod r}} c_{q_1,r}\overline{c_{q_2,r}}\beta(n_1)\overline{\beta(n_2)}1_{mn_1\equiv b_1 \pmod{q_1}}1_{mn_2\equiv b_2 \pmod{q_2}}$$
$$=\pm\sum_r\sum_{q_1,q_2}\sum_{\substack{n_1,n_2\\ n_1\equiv n_2 \pmod r}} c_{q_1,r}\overline{c_{q_2,r}}\beta(n_1)\overline{\beta(n_2)}\cdot\sum_m 1_{\substack{m\equiv b_1/n_1 \pmod{q_1}\\ m\equiv b_2/n_2 \pmod{q_2}\\ m\equiv a/n_1 \pmod r}}$$

where we get "+" when $b_1 = b_2 = b$ or b', and "−" otherwise, since $(mn, qr) = 1$. Notice that the last sum is 0 unless $b_1/n_1 \equiv b_2/n_2 \pmod{(q_1, q_2)}$; and that this criterion is irrelevant if $(q_1, q_2) = 1$.

9.1. **The main terms.** When the last sum (over m) is non-zero then we "expect" it to be $M/r[q_1, q_2]$. In our range this can be < 1, which makes no sense for an individual sum, but we expect this to be about right "on average". The key idea is to deal with the deviation from this average using exponential sums. This is the "dispersion method". First though, let us deal with the "expected" main term:

$$\pm\sum_r\sum_{q_1,q_2}\sum_{\substack{n_1,n_2\\ n_2\equiv n_1 \pmod r\\ n_2\equiv (b_2/b_1)n_1 \pmod{(q_1,q_2)}}} c_{q_1,r}\overline{c_{q_2,r}}\beta(n_1)\overline{\beta(n_2)}\cdot\frac{M}{r[q_1,q_2]}.$$

We pull out the term with $(q_1, q_2) = 1$ to obtain

$$\pm\sum_r\sum_{\substack{q_1,q_2\\ (q_1,q_2)=1}}\sum_{\substack{n_1,n_2\\ n_1\equiv n_2 \pmod r}} c_{q_1,r}\overline{c_{q_2,r}}\beta(n_1)\overline{\beta(n_2)}\cdot\frac{M}{rq_1q_2},$$

which is independent of the values of b_1, b_2 and hence cancels, when we sum over the four terms.

Otherwise $g := (q_1, q_2) \geqslant D_0$. If $g \leqslant N/R$ then there are approximately N/gR values of n_2 for each n_1. Hence by Cauchy-Schwarz

$$\sum_{\substack{n_1,n_2\\ n_2\equiv n_1 \pmod r\\ n_2\equiv (b_2/b_1)n_1 \pmod g}} |\beta(n_1)\beta(n_2)| \ll \frac{N}{gR}\|\beta\|^2$$

Therefore the total contribution above is

$$\ll\sum_r\sum_{D_0<g\leqslant N/R}\sum_{\substack{q_1,q_2\\ g|q_1,\ g|q_2}}\frac{N}{gR}\|\beta\|^2\cdot\frac{gM}{rq_1q_2}\ll\frac{x\|\beta\|^2}{R}\sum_{D_0<g\leqslant N/R}\frac{1}{g^2}\ll\frac{x\|\beta\|^2}{RD_0}.$$

For larger g, the sum above becomes

$$\ll\sum_r\sum_{N/R<g\leqslant Q}\sum_{\substack{n_1,n_2\\ n_2\equiv n_1 \pmod r\\ n_2\equiv (b_2/b_1)n_1 \pmod g}} |\beta(n_1)\beta(n_2)|\cdot\frac{M}{rg}$$

Since there is at most one value of n_2 for each n_1 we can use Cauchy-Schwarz to show that the sum over n_1, n_2 is $\leqslant \|\beta\|^2$. So the total contribution is $\|\beta\|^2 M \log x$, which is far smaller then the previous contribution, as $M \leqslant x^{1-\epsilon}/R$, by hypothesis. Hence the contribution of the expected main terms is in total

$$\left(R\|\alpha\|^2 \cdot \frac{x\|\beta\|^2}{RD_0}\right)^{1/2} \ll \|\alpha\|\|\beta\| \frac{x^{1/2}}{(\log x)^A}.$$

9.2. **Crude error terms for large** g**.** If $g > G := Q^2R/M$ then $M/r[q_1, q_2] \geqslant 1$, and the count of the number of m values is as above with an error term of $O(1)$. We will simply sum up these crude error terms. (In fact one can take this error term for any g). Now if we Cauchy the sum over β we get in total $\ll (\frac{N}{Rg} + 1)\|\beta\|^2$. The sums over the q_i's divisible by g, contribute Q/g each, and the sum over r, contributes R, so over all $g \geqslant G$ the error term is

$$\ll \sum_{G<g\leqslant Q} R\frac{Q^2}{g^2}(\frac{N}{Rg} + 1)\|\beta\|^2 \ll Q^2(\frac{N}{G^2} + \frac{R}{G})\|\beta\|^2.$$

Taking $G = Q^2R/M$ this is

$$\ll \left(\frac{x}{D^2} + 1\right)\frac{x\|\beta\|^2}{N} \ll \frac{x(\log x)^{2B}\|\beta\|^2}{N}$$

where $D = QR$.[15] Hence the contribution here is

$$\ll \|\alpha\|\|\beta\|x^{1/2}(\log x)^B(R/N)^{1/2}$$

which is certainly acceptable given our choice of R.

9.3. **Exponential sums.** After removing these contributions, we are left with four terms, each of which is bounded by a sum of the form

$$\sum_{r\asymp R}\sum_{g\leqslant G}\sum_{\substack{\ell_1,\ell_2\asymp Q/g\\(\ell_1,\ell_2)=1}}\left|\sum_{\substack{n_1,n_2\asymp N\\ n_1\equiv n_2 \pmod r\\ b_1/n_1\equiv b_2/n_2 \pmod g}}\beta(n_1)\overline{\beta(n_2)}\cdot\left(\sum_{\substack{m\asymp M\\ m\equiv m_0(n_1,n_2) \pmod{rg\ell_1\ell_2}}} 1 - \frac{M}{rg\ell_1\ell_2}\right)\right|$$

writing $q_1 = g\ell_1,\ q_2 = g\ell_2$, where $m_0 = m_0(n_1, n_2)$ is that residue class mod $g\ell_1\ell_2r$ ExponExpan which is $\equiv b_1/n_1 \pmod{g\ell_1}$, $\equiv b_2/n_2 \pmod{g\ell_2}$, $\equiv a/n_1 \pmod r$. Using (8.3) this is

$$\ll \sum_{r\asymp R}\sum_{g\leqslant G}\sum_{\substack{\ell_1,\ell_2\asymp Q/g\\(\ell_1,\ell_2)=1}}\sum_{\substack{0\leqslant i\leqslant J\\ H_i:=2^jG/g}}\frac{1}{H_i}\sum_{1\leqslant|h|\leqslant H_i}\left|\sum_{\substack{n_1,n_2\asymp N\\ n_1\equiv n_2 \pmod r\\ n_2\equiv(b_2/b_1)n_1 \pmod g}}\beta(n_1)\overline{\beta(n_2)}e_{rg\ell_1\ell_2}(m_0(n_1,n_2)h)\right|.$$

Writing $n_1 = n$ and $n_2 = n + kr$ for some $k,\ |k| \leqslant N/R$, this equals

$$\ll \sum_{g\leqslant G}\sum_{\substack{0\leqslant i\leqslant J\\ H_i:=2^iG/g}}\sum_{r\asymp R}\sum_{\substack{\ell_1,\ell_2\asymp Q/g\\(\ell_1,\ell_2)=1}}\cdot$$

[15] We can divide G by $(N/R)^{1/2}/(\log x)^C$, and still have a good error term.

$$\frac{1}{H_i}\sum_{1\leqslant|h|\leqslant H_i}\left|\sum_{\substack{k\leqslant N/R\\(b_2-b_1,g)|k}}\ \sum_{\substack{n\asymp N\\(b_2-b_1)n\equiv b_1kr \pmod g}}\beta(n)\overline{\beta(n+kr)}\Phi_k(h,n;r,g,\ell_1,\ell_2)\right|. \qquad (9.3)$$ Esum

bearing in mind the definition (phi-def 8.6), so that there is only a term if $(b_2-b_1,g)|k$. We will see two techniques for dealing with these sums, both of which begin by using the Cauchy-Schwarz inequality to eliminate the $\beta(n)$ factors, so reducing things to incomplete exponential sum estimates, which we handle by using the estimates from Section incompsec 8.

9.4. **Technique # 1.** We replace the absolute value above by a complex number c_{h,ℓ_1,ℓ_2} of absolute value 1, so that the sum

$$\sum_{\substack{\ell_1,\ell_2\asymp Q/g\\(\ell_1,\ell_2)=1}}\frac{1}{H_i}\sum_{1\leqslant|h|\leqslant H_i}\left|\sum_n \beta(n)\overline{\beta(n+kr)}\Phi_k(h,n;r,g,\ell_1,\ell_2)\right|$$

equals

$$\sum_{\substack{n\asymp N\\(b_2-b_1)n\equiv b_1kr \pmod g}}\beta(n)\overline{\beta(n+kr)}\sum_{\substack{\ell_1,\ell_2\asymp Q/g\\(\ell_1,\ell_2)=1}}\frac{1}{H_i}\sum_{1\leqslant|h|\leqslant H_i}c_{h,\ell_1,\ell_2}\Phi_k(h,n;r,g,\ell_1,\ell_2). \qquad (9.4)$$ PreCauchy

Applying the Cauchy-Schwarz inequality, the square of this is less than or equal to[16]

$$\sum_n|\beta(n)\beta(n+kr)|^2\leqslant\sum_n|\beta(n)|^4=\|\beta\|_4^4$$

(applying the Cauchy-Schwarz inequality again), times

$$\sum_n\left|\sum_{\substack{\ell_1,\ell_2\asymp Q/g\\(\ell_1,\ell_2)=1}}\frac{1}{H_i}\sum_{1\leqslant|h|\leqslant H_i}c_{h,\ell_1,\ell_2}\Phi_k(h,n;r,g,\ell_1,\ell_2)\right|^2$$

$$\leqslant\frac{1}{H_i^2}\sum_{1\leqslant|h|,|j|\leqslant H_i}\ \sum_{\substack{\ell_1,\ell_2,m_1,m_2\asymp Q/g\\(\ell_1,\ell_2)=(m_1,m_2)=1}}|S_{k,r}(h,j,g,\ell_1,\ell_2,m_1,m_2)|,$$

by expanding and then taking absolute values for each fixed $h,j,\ell_1,\ell_2,m_1,m_2$, where the exponential sum $S_{k,r}(h,j,\ell_1,\ell_2,m_1,m_2)$ is defined in (S-def 8.7). By Proposition expsumLemma 8.4(i), this is $x^{o(1)}$ times

$$\ll(R(b_2-b_1,g))^{1/2}(Q/g)^6+\frac{N}{R}\frac{(b_2-b_1,g)}{g}\frac{1}{H_i^2}\sum_{1\leqslant|h|,|j|\leqslant H_i}\ \sum_{\substack{\ell_1,\ell_2,m_1,m_2\asymp Q/g\\(\ell_1,\ell_2)=(m_1,m_2)=1}}(hm_1m_2-j\ell_1\ell_2,r).$$

Now, in the sums in the second term let $u=hm_1m_2,\ v=j\ell_1\ell_2$ so that $1\leqslant|u|,|v|\ll H_i(Q/g)^2$ and the pair is represented at most $\tau_3(u)\tau_3(v)=x^{o(1)}$ times. Therefore the

[16] If we apply Holder's inequality with exponents $6,6,6,2$, we can replace $\|\beta\|_4^4$ in this upper bound by $(N/g')^{1/3}\|\beta\|_6^4$; and more generally $(N/g')^{1-2/m}\|\beta\|_{2m}^4$, where $g'=g/(b_2-b_1,g)$.

difference $w = u - v$ satisifies $|w| \ll H_i(Q/g)^2$ is represented at most $x^{o(1)}H_i(Q/g)^2$ times. Now

$$\sum_{|w|\leqslant W}(w,r) \leqslant \sum_{|w|\leqslant W}\sum_{d|(w,r)} d \leqslant \sum_{d|r} d\left(\frac{2W}{d}+1\right) \leqslant (2W+r)\tau(r).$$

Hence the above is $x^{o(1)}$ times

$$\ll (R(b_2-b_1,g))^{1/2}(Q/g)^6 + \frac{N}{R}\frac{(b_2-b_1,g)}{g}(Q/g)^2((Q/g)^2+R/H_i).$$

Collecting this information together, and summing over r and k, yields an upper bound on Esum (9.3) of

$$\ll \|\beta\|_4^2 N x^{o(1)} \sum_{g\leqslant G}\sum_{0\leqslant i\leqslant J}\left(\frac{R^{1/4}Q^3}{g^3(b_2-b_1,g)^{3/4}} + \frac{N^{1/2}Q^2}{R^{1/2}}\frac{1}{g^{5/2}(b_2-b_1,g)^{1/2}} + \frac{(MN)^{1/2}}{R^{1/2}}\frac{1}{g(b_2-b_1,g)^{1/2}}\frac{1}{2^{i/2}}\right)$$

$$\ll \|\beta\|_4^2 N x^{o(1)}\left(R^{1/4}Q^3+\frac{N^{1/2}Q^2}{R^{1/2}}+\frac{(MN)^{1/2}}{R^{1/2}}\right)$$

Finally we assume that $\|\beta\|_4^2N^{1/2} \ll \|\beta\|_2^2x^{o(1)}$ (note that $\|\beta\|_2^2 \leqslant \|\beta\|_4^2N^{1/2}$ by Cauchying), and therefore the total contribution is

$$\ll \|\alpha\|^2\|\beta\|^2x^{o(1)}\left(N^{1/2}\frac{(QR)^3}{R^{7/4}}+\frac{N(QR)^2}{R^{3/2}}+R^{1/2}N^{1/2}x^{1/2}\right)$$

$$\ll \|\alpha\|^2\|\beta\|^2x^{o(1)}\left(\frac{x^{3/2+3\eta}}{N^{5/4}}x^{\frac{7}{4}(\delta+\epsilon)}+\frac{x^{1+2\eta}}{N^{1/2}}x^{\frac{3}{2}(\delta+\epsilon)}+Nx^{\frac{1}{2}-\frac{1}{2}\epsilon}\right) \tag{9.5}$$

FinalBound

using the inequalities $N/x^{\delta+\epsilon} < R \leqslant N/x^{\epsilon}$, $x^{1/2-o(1)} \leqslant QR \leqslant x^{1/2+\eta}$. Now since $N \ll x^{1/2}$, the last term is $\ll x^{1-\epsilon/2}$. We will bound FinalBound1 (9.5) in section FinalSection 9.6.

We remark that had we used Proposition expsumLemma 8.4(ii) in place of Proposition expsumLemma 8.4(i), then the first term in FinalBound1 (9.5) would have been

$$x^{\frac{7}{6}+\frac{7}{3}\eta+\frac{7}{4}\delta+\frac{5}{4}\epsilon}/N^{1/2} \ \text{ in place of } x^{\frac{3}{2}+3\eta+\frac{7}{4}(\delta+\epsilon)}/N^{5/4}.$$

This yields a suitable bound in a wider range for N, but not for all $N \gg x^{1/3}$ so, either way, we need another argument for smaller N.

9.5. **Technique # 2.**

We also employ a variation on this theme, including ℓ_1 in the outside summation in PreCauchy (9.4) when we apply Cauchy-Schwarz. Hence the square of our quantity is

$$\ll \frac{Q}{g}\|\beta\|_4^4\frac{1}{H_i^2}\sum_{1\leqslant|h|,|j|\leqslant H_i}\sum_{\ell_1\asymp Q/g}\sum_{\substack{\ell_2,m_2\asymp Q/g\\(\ell_2m_2,\ell_1)=1}}|S_{k,r}(h,j,g,\ell_1,\ell_2,\ell_1,m_2)|.$$

By Proposition expsumLemma 8.4(ii), and the assumption that $\|\beta\|_4^2N^{1/2} \ll \|\beta\|_2^2x^{o(1)}$,

$$\ll \|\beta\|^4x^{o(1)}\left((Ry)^{1/6}\frac{Q^{9/2}}{N^{1/2}}\frac{(b_2-b_1,g)^{2/3}}{g^5}+\frac{Q^2}{R}\frac{(b_2-b_1,g)}{g^3}\frac{1}{H_i^2}\sum_{1\leqslant|h|,|j|\leqslant H_i}\sum_{\ell_2,m_2\asymp Q/g}(hm_2-j\ell_2,r)\right)$$

Proceeding as above, and since $\tau(u)\tau(v)\tau(r) = x^{o(1)}$, we obtain

$$\ll \|\beta\|^4 x^{o(1)} \left((Ry)^{1/6} \frac{Q^{9/2}}{N^{1/2}} \frac{(b_2-b_1,g)^{2/3}}{g^5} + \frac{Q^2}{R} \frac{(b_2-b_1,g)}{g^3} \left(\frac{Q^2}{g^2} + \frac{M}{2^i Q} \right) \right)$$

as $H_i = 2^i RQ^2/gM$.

Collecting this information together, and summing over r and k, yields an upper bound on (9.3) of

$$\ll \|\beta\|^2 N x^{o(1)} \sum_{g \leqslant G} \sum_{0 \leqslant i \leqslant J} \left((Ry)^{1/12} \frac{Q^{9/4}}{N^{1/4}} + \frac{Q^2}{R^{1/2}} + \frac{(MQ)^{1/2}}{2^{i/2} R^{1/2}} \right) \frac{1}{g^{3/2}(b_2-b_1,g)^{1/2}}$$

$$\ll \|\beta\|^2 N x^{o(1)} \left((Ry)^{1/12} \frac{Q^{9/4}}{N^{1/4}} + \frac{Q^2}{R^{1/2}} + \frac{(MQ)^{1/2}}{R^{1/2}} \right).$$

Therefore the total contribution is

$$\ll \|\alpha\|^2 \|\beta\|^2 x^{o(1)} \left(\frac{y^{1/12}(QR)^{9/4} N^{3/4}}{R^{7/6}} + \frac{N(QR)^2}{R^{3/2}} + N^{1/2}(MNQR)^{1/2} \right).$$

$$\ll \|\alpha\|^2 \|\beta\|^2 x^{o(1)} \left(\frac{x^{\frac{9}{8}+\frac{9}{4}\eta+\frac{5}{4}\delta+\frac{7}{6}\epsilon}}{N^{5/12}} + \frac{x^{1+2\eta+\frac{3}{2}(\delta+\epsilon)}}{N^{1/2}} + N^{1/2} x^{\frac{3}{4}+\frac{1}{2}\eta} \right).$$

using the inequalities $N/x^{\delta+\epsilon} < R \leqslant N/x^{\epsilon}$, $x^{1/2-o(1)} \leqslant QR \leqslant x^{1/2+\eta}$,

$$\ll \|\alpha\|^2 \|\beta\|^2 x^{o(1)} \left(x^{\frac{71}{72}+\frac{9}{4}\eta+\frac{5}{4}\delta+\frac{7}{6}\epsilon} + x^{\frac{5}{6}+2\eta+\frac{3}{2}(\delta+\epsilon)} + N^{1/2} x^{\frac{3}{4}+\frac{1}{2}\eta} \right). \qquad (9.6)$$ FinalBound

as $N \gg x^{1/3}$. The third term is $\ll x^{1-\epsilon/2}$ provided $N \leqslant x^{\frac{1}{2}-\eta-\epsilon}$.

We remark that had we used Proposition 8.4(i) in place of Proposition 8.4(ii), then the first term above would have been

$$x^{\frac{11}{8}+\frac{11}{4}\eta+\frac{3}{2}\delta+\frac{3}{2}\epsilon}/N \ \text{ in place of } x^{\frac{9}{8}+\frac{9}{4}\eta+\frac{5}{4}\delta+\frac{7}{6}\epsilon}/N^{5/12}.$$

This yields a suitable bound only for N somewhat bigger than $x^{3/8}$, not for all $N \gg x^{1/3}$, whereas the argument we have used allows N to be this small.

FinalSection

9.6. **Bounds in different ranges.** In (9.5) and (9.6), we want the quantity in brackets to be $\ll x^{1-\epsilon}$. We use (9.5) in the range $x^{\frac{1}{2}-\eta-\epsilon} < N \ll x^{\frac{1}{2}}$, so that it is

$$\ll x^{\frac{7}{8}+\frac{17}{4}\eta+\frac{17}{4}\epsilon+\frac{7}{4}\delta} + x^{\frac{3}{4}+\frac{5}{2}\eta+\frac{5}{2}\epsilon+\frac{3}{2}\delta} + x^{1-\epsilon/2}.$$

We use (9.6) in the range $x^{\frac{1}{3}} \ll N \leqslant x^{\frac{1}{2}-\eta-2\epsilon}$, so that it is

$$\ll x^{\frac{71}{72}+\frac{9}{4}\eta+\frac{5}{4}\delta+\frac{7}{6}\epsilon} + x^{\frac{5}{6}+2\eta+\frac{3}{2}(\delta+\epsilon)} + x^{1-\epsilon/2}.$$

These are all $\ll x^{1-\epsilon/2}$, for a sufficiently small choice of $\epsilon > 0$, as long as

$$162\eta + 90\delta < 1.$$

CorrectNorms

9.7. **Correcting the norms.** We made some unnecessary assumption of the norms in the arguments above. In fact we used 4-norms and 8-norms. Simply using the inequalities, for γ supported in $[M, 2M]$, that $\|\gamma\|_2 \leqslant \|\gamma\|_4 M^{1/4} \leqslant \|\gamma\|_8 M^{3/8}$, we can correct Theorem 6.1 by replacing $\|\alpha\|\|\beta\|x^{1/2}$ by $\|\alpha\|_8\|\beta\|_8 x^{7/8}$.

9.8. **Better results.** In polymath8 [30] the authors obtain better results using somewhat deeper techniques.

One key observation is that y-smoothness was used in the above argument to construct a divisor r of a given integer d in a prespecified interval of multiplicative length y. In fact one can make do just with this property and, to improve our exponential sum estimates, that r also has a divisor in a prespecified interval of multiplicative length y. By going to such a larger class of moduli q they improve the restriction to

$$84\eta + 48\delta < 1.$$

Following Zhang they also gained bounds on certain higher order convolutions (of the shape $\alpha * 1 * 1 * 1$), though here needing some deeper exponential sum estimates, and were then able to improve the restriction to (slightly better than)

$$43\eta + 27\delta < 1.$$

10. Weaker hypotheses

In section 7 we stated that we only need the estimate (7.5) for the exponential sums in (7.4). It is worth noting (7.5) may be weakened to the upper bound $\ll p^\theta$, for any given $\theta \in (\frac{1}{2}, \frac{2}{3})$, and we can still obtain the same result:

From Proposition 8.1 onwards we replace the exponent $\frac{1}{2}$ by θ, and $\frac{1}{6}$ by $\frac{\theta}{3(1+\theta)}$. Eventually this leads us in technique # 1, to replacing the first term in (9.5) and the line above, by

$$N^{1/2}\frac{(QR)^{2+2\theta}}{R^{\frac{3}{2}\theta+1}} \leqslant \frac{x^{1+\theta+(2+2\theta)\eta+(\frac{3}{2}\theta+1)(\delta+\epsilon)}}{N^{\frac{(3\theta+1)}{2}}}$$

which is, for $N > x^{\frac{1}{2}-\eta-\epsilon}$,

$$\ll x^{\frac{3+\theta}{4}+\frac{5+7\theta}{2}\eta+(\frac{3}{2}\theta+1)\delta+(3\theta+\frac{3}{2})\epsilon}.$$

Similarly, in technique # 2, we replace the first term in (9.6) and the line above, by

$$N^{3/4}y^{\frac{\theta}{4(1+\theta)}}\frac{(QR)^{2+\frac{3\theta}{4(1+\theta)}}}{R^{1+\frac{\theta}{2(1+\theta)}}} \leqslant \frac{x^{(2+\frac{3\theta}{4(1+\theta)})(\frac{1}{2}+\eta)+(1+\frac{3\theta}{4(1+\theta)})\delta+(1+\frac{\theta}{2(1+\theta)})\epsilon}}{N^{\frac{1}{4}+\frac{\theta}{2(1+\theta)}}}$$

which is $< x^{1-\epsilon}$, for $N > x^{\frac{1}{3}}$, provided

$$(11\theta+8)\eta+(7\theta+4)\delta < \frac{2-3\theta}{6}$$

so we deduce such a theorem provided $\theta < \frac{2}{3}$.

References

[1] E. Bombieri, *On the large sieve*, Mathematatika **12** (1965), 201–225.

[2] E. Bombieri, H. Davenport *Small difference between prime numbers*, Proc. Roy. Soc. Ser. A **293** (1966), 1-18.

[3] E. Bombieri, J. Friedlander, H. Iwaniec*Primes in arithmetic progressions to large moduli*, Acta Math. **156** (1986), no. 3-4, 203–251.

[4] E. Bombieri, J. Friedlander, H. Iwaniec, *Primes in arithmetic progressions to large moduli. II*, Math. Ann. **277** (1987), no. 3, 361–393.

[5] E. Bombieri, J. Friedlander, H. Iwaniec, *Primes in arithmetic progressions to large moduli. III*, J. Amer. Math. Soc. **2** (1989), no. 2, 215–224.

[6] T. Cochrane, C. Pinner, *Using Stepanov's method for exponential sums involving rational functions*, J. Number Theory **116** (2006), no. 2, 270–292.

[7] P. Deligne, *La conjecture de Weil. II*, Publications Mathématiques de l'IHÉS **52** (1980), 137–252.

[8] P. D. T. A. Elliott, H. Halberstam, *A conjecture in prime number theory*, Symp. Math. **4** (1968), 59–72.

[9] E. Fouvry, *A new form of the error term in the linear sieve*, Acta Arith., **37** (1980), 307–320.

[10] E. Fouvry, *Autour du théorème de Bombieri-Vinogradov*, Acta Math. **152** (1984), no. 3-4, 219–244.

[11] E. Fouvry, H. Iwaniec, *On a theorem of Bombieri-Vinogradov type*, Mathematika **27** (1980), no. 2, 135–152 (1981).

[12] E. Fouvry, H. Iwaniec, *Primes in arithmetic progressions*, Acta Arith. **42** (1983), no. 2, 197–218.

[13] J. Friedlander, A. Granville, *Limitations to the equi-distribution of primes. I*, Ann. of Math. **129** (1989), 363-382.

[14] J. Friedlander, H. Iwaniec, *Incomplete Kloosterman sums and a divisor problem*, With an appendix by Bryan J. Birch and Enrico Bombieri. Ann. of Math. (2) **121** (1985), no. 2, 319–350.

[15] D. Goldston, J. Pintz, C. Yıldırım, *Primes in tuples. I*, Ann. of Math. **170** (2009), no. 2, 819–862.

[16] D. Goldston, S. Graham, J. Pintz, C. Yıldırım, *Small gaps between primes or almost primes*, Trans. Amer. Math. Soc. **361** (2009), no. 10, 5285–5330.

[17] S. W. Graham, C. J. Ringrose, *Lower bounds for least quadratic nonresidues*, Analytic number theory (Allerton Park, IL, 1989), 269–309, Progr. Math., 85, Birkhäuser Boston, Boston, MA, 1990.

[18] A. Granville, K. Soundararajan, *Multiplicative number theory; the pretentious approach*, to appear.

[19] G. H. Hardy, J. E. Littlewood, *Some problems of "Partitio Numerorum", III: On the expression of a number as a sum of primes*, Acta Math. **44** (1923), 1–70.

[20] D. R. Heath-Brown, *Prime numbers in short intervals and a generalized Vaughan identity*, Canad. J. Math. 34 (1982), no. 6, 1365–1377.

[21] H. A. Helfgott, *Major arcs for Goldbach's theorem*, to appear.

[22] D. Hensley, I. Richards, *On the incompatibility of two conjectures concerning primes*, Analytic number theory (Proc. Sympos. Pure Math., Vol. XXIV, St. Louis Univ., St. Louis, Mo., 1972), pp. 123–127. Amer. Math. Soc., Providence, R.I., 1973.

[23] D. Hensley, I. Richards, *Primes in intervals*, Acta Arith. **25** (1973/74), 375–391.

[24] H. D. Kloosterman, *On the representation of numbers in the form $ax^2 + by^2 + cz^2 + dt^2$*, Acta Mathematica **49** (1926), pp. 407–464.

[25] H. Maier, *Small differences between prime numbers*, Michigan Math. J. **35** (1988), 323344.

[26] J. Maynard, *Bounded length intervals containing two primes and an almost-prime II*, preprint.

[27] L. J. Mordell, *On a sum analogous to a Gauss's sum*, Quart. J. Math. Oxford Ser. **3** (1932), 161–167.

[28] Y. Motohashi, J. Pintz, *A smoothed GPY sieve*, Bull. Lond. Math. Soc. **40** (2008), no. 2, 298–310.

[29] J. Pintz, *Polignac Numbers, Conjectures of Erdős on Gaps between Primes, Arithmetic Progressions in Primes, and the Bounded Gap Conjecture*, preprint.

[30] D.H.J. Polymath, *A new bound for small gaps between primes*, preprint.

[31] A. Schinzel, *Remarks on the paper "Sur certaines hypothéses concernant les nombres premiers"*, Acta Arith. **7** (1961/1962) 1–8.

[32] A. Selberg, *On elementary methods in prime number-theory and their limitations*, in Proc. 11th Scand. Math. Cong. Trondheim (1949), Collected Works, Vol. I, 388397, Springer-Verlag, Berlin-Göttingen-Heidelberg, 1989.

[33] P. Shiu, *A Brun-Titchmarsh theorem for multiplicative functions*, J. Reine Angew. Math. 313(1980), 161–170.

[34] K. Soundararajan, *Small gaps between prime numbers: the work of Goldston-Pintz-Yildirim*, Bull. Amer. Math. Soc. (N.S.) **44** (2007), no. 1, 1–18.

[35] R. C. Vaughan, *Sommes trigonométriques sur les nombres premiers*, C. R. Acad. Sci. Paris Sér. A **285** (1977), 981–983.

[36] A. I. Vinogradov, *The density hypothesis for Dirichlet L-series*, Izv. Akad. Nauk SSSR Ser. Mat. **29** (1965), 903–934.

[37] A. Weil, *Numbers of solutions of equations in finite fields*, Bulletin of the American Mathematical Society **55** (1949), 497–508.

[38] Y. Zhang, *Bounded gaps between primes*, to appear, Annals of Mathematics.

Département de mathématiques et de statistiques, Université de Montréal, Montréal QC H3C 3J7, Canada.

E-mail address: andrew@dms.umontreal.ca