Coordinate Distribution of Gaussian Primes

John Friedlander, Henryk Iwaniec

Contents
1 Introduction and statement of results 1
2 Interlude: an easier result 5
3 The congruence sums 5
4 Estimation of the remainder 8
5 A model for \( A \equiv (a_n) \) 9
6 Sums over primes 11
7 Bilinear forms in the Gaussian Domain 14
8 The diagonal terms 17
9 In the off-diagonal area 18
10 Separation of variables 20
11 Estimation of \( \mathcal{R}_h''(m_1, m_2) \) 22
12 Small determinant 23
13 Estimation of \( \mathcal{R}_h'(m_1, m_2) \) on average 24
14 Preparation of the main terms 26
15 Small moduli 29
16 Large moduli 30
17 Proof of SMT. Conclusion 31
18 Derivation of MT 32
19 Derivation of APT 32
Appendix 33

1 Introduction and statement of results

The modern history of prime number theory might well be said to begin with the statement of Fermat to the effect that the primes of the form \( 4n + 1 \) can be written as the sum of two squares. The first recorded proof is due to Euler. We think of these today as being the primes which occur as the norms of the unramified splitting primes \( a + 2bi \) in the Gaussian field \( \mathbb{Q}(i) \) and we shall
refer to them as Gaussian primes. Following the proof of the prime number theorem, we have the well-known asymptotic formula for the number of these:

$$
\psi(x; 4, 1) = \sum_{n \leq x \atop n \equiv 1 \pmod{4}} \Lambda(n) = \sum_{n \leq x \atop n = a^2 + (2b)^2} \Lambda(n) \sim \frac{1}{2} x,
$$

where we are going to restrict to $a$ and $b$ being positive.

Beginning in the 1990’s one began seeing how to count the frequency of subsets of these primes for which one of the squares has an additional interesting arithmetic property. The first result to note in this connection was the work [FoIw] of E. Fouvry and H. Iwaniec in which the asymptotic formula was obtained for the case wherein one of the squares was the square of a prime (actually their result was rather more general). Subsequently, in [FrIw1], the current authors obtained the asymptotic in the setting where one of the squares was the square of a square and thus for the number of primes which could be written as the sum of a square plus a fourth power. This result had an additional interest in first successfully establishing the asymptotic formula for a thin set of prime values of a polynomial, that is, one having density $\ll x^{1-\delta}$ for some positive $\delta$.

Following a gap of some fifteen to twenty years, there have now been a number of newer developments along these lines of research. R. Heath-Brown and X. Li [HL] have shown that, in the statement of [FrIw1], one can replace the fourth power of an integer by the fourth power of a prime and still establish for these the relevant asymptotic formula. Very recently, K. Pratt [Pr] has succeeded with the thin set obtained when one of the squares is the square of an integer which is missing three prescribed digits from its decimal expansion. P. Lam, D. Schindler and S. Xiao [LSX] have succeeded in extending the original work [FoIw], replacing the Gaussian integers and Gaussian primes by the corresponding values of an arbitrary irreducible positive definite binary quadratic form.

In all of these highly interesting works one is concerned with the specialization to a particular subset those values taken on by one of the two coordinates. In this work we shall be motivated by the question wherein we ask something special about both of them.

We are going to count the primes $\pi = a + 2bi$ in the ring $\mathbb{Z}[i]$ which have their coordinates $a$, $b$ restricted to special integers. Ideally, we would like to reach $\pi = a + 2bi$ with $a$ and $b$ both primes, but we are too old to reach these by currently developed technology. However, we still have enough strength for catching $\pi = a + 2bi$ with $a$ prime and $b$ almost-prime.

We accomplish the goal by estimating sums of type

$$
G(x) = \sum_{4k^2 + \ell^2 \leq x} \sum \beta_k \gamma_\ell \Lambda(4k^2 + \ell^2)
$$

with coefficients $\beta_k, \gamma_\ell$ which live on primes and almost-primes. In most parts of our considerations these coefficients can be quite general, but sometimes we have to specialize.

Let $\Lambda_r = \mu * (\log)^r$ denote the von Mangoldt function of order $r \geq 1$ and $\Lambda = \Lambda_1$. The $\Lambda_r(n)$ vanish unless $r$ has at most $r$ distinct prime factors and, in any case, $0 \leq \Lambda_r(n) \leq (\log n)^r$. In the Appendix we shall give some heuristic arguments leading to the determination of an asymptotic formula for

$$
G_r(x) = \sum_{4k^2 + \ell^2 \leq x} \sum \Lambda_r(k) \Lambda(\ell) \Lambda(4k^2 + \ell^2).
$$

**Conjecture.** We have

$$
G_r(x) \sim crx (\log \sqrt{x})^{r-1}
$$

(1.2)
\[ c = \prod_{p \equiv 1(4)} \left( 1 - \frac{3}{p} \right) \left( 1 - \frac{1}{p} \right)^{-3} \prod_{p \equiv 3(4)} \left( 1 - \frac{1}{p^2} \right)^{-1}. \]

The case \( r = 1 \) is most challenging, because it requires breaking the parity barrier of sieve theory.

**Conjecture 1.1** (Gaussian Primes Conjecture).

\[
\sum \sum_{4k^2 + \ell^2 \leq x} \Lambda(k)\Lambda(\ell)\Lambda(4k^2 + \ell^2) \sim cx.
\]

We are able to estimate \( G_r(x) \) positively for \( r \geq 7 \).

**Theorem 1.1** (\( G_7 \)). We have

\[
\sum \sum_{4k^2 + \ell^2 \leq x} \Lambda_7(k)\Lambda(\ell)\Lambda(4k^2 + \ell^2) \approx x \left( \log x \right)^6.
\]

**Remarks 1.1.** If \( n \) is not squarefree or \( n \) has a small prime factor then \( \Lambda_r(n) \) contributes to \( G_r(x) \) a negligible amount, so we are really catching primes \( 4k^2 + \ell^2 \) with \( \ell \) prime and \( k \) having at most \( r \) prime factors, all distinct.

In fact we shall estimate a more restricted sum.

**Theorem 1.2** (Almost Primes Theorem). Let \( \beta_k = 1 \) if \( k \) has at most 7 prime factors, all of which are larger than \( k^{1/49} \), and \( \beta_k = 0 \) otherwise. Then

\[
\sum \sum_{4k^2 + \ell^2 \leq x} \beta_k \Lambda(\ell)\Lambda(4k^2 + \ell^2) \approx x \left( \log x \right)^{-1}.
\]

**Remarks 1.2.** The lower bound of (1.4) follows from the lower bound of (1.5), because \( \beta_k \left( \log k \right)^7 \ll \Lambda_7(k) \). The upper bounds can be derived directly by application of any crude sieve method so we skip the proof.

We shall establish an asymptotic formula for \( G(x) \) with relatively small error term where \( \beta \) is the convolution \( 1 \ast \lambda \) with \( \lambda \) supported on a relatively short segment. Put

\[
\beta_k = \sum_{h|k} \lambda_h
\]

with \( |\lambda_h| \leq 1 \) for \( h \) squarefree, \( h \leq y \) say, \( \lambda_h = 0 \) otherwise. Obviously we have in mind the sieve weights \( \lambda_h \) of level \( y \). Having the weights \( \lambda_h \) at our disposal we can build the \( \beta_k \) having our favorite property. There are numerous possibilities to play with these weights.

The second coordinate \( \ell \) is counted with weight \( \gamma_\ell \) about which we do not need to know much. However, after serious attempts to handle \( \gamma_\ell \) in great generality we gave up this ambition, because of tremendous complications in resolving the main term in certain bilinear forms over the Gaussian domain. We are going to assume that

\[
|\gamma_\ell| \leq \log \ell, \text{ if } \ell \text{ is an odd prime,}
\]
and $\gamma_\ell = 0$, otherwise. Moreover, we need the asymptotic formula

$$
\sum_{\ell \leq x, \ell \equiv a(q)} \gamma_\ell = \frac{x}{\phi(q)} + O(x(\log x)^{-B})
$$

to hold for every $q \geq 1$, $(a, q) = 1$, $x \geq 2$ and any $B \geq 2$, the implied constant depending only on $B$.

**Remarks 1.3.** For $\gamma_\ell = \log \ell$ the formula (1.8) is just the Siegel-Walfisz theorem. Have in mind that our assumption (1.8) is meaningful for $q < (\log x)^A$ with any $A \geq 2$, but has no value for much larger moduli. By resizing, one is allowed to change (1.7) and (1.8) by a fixed positive constant independent of the residue classes $a \pmod{q}$.

**Theorem 1.3 (Main Theorem).** Suppose $\beta_h$ are given by (1.6) with $|\lambda_h| \leq 1$ for $h$ squarefree,

$$
h \leq y = x^\theta, \quad 0 < \theta < \frac{1}{12}
$$

and $\lambda_h = 0$, otherwise. Suppose $\gamma_\ell$ satisfies (1.7) and (1.8). Then we have

$$
\sum_{4k^2 + \ell^2 \leq x} \beta_k \gamma_\ell \Lambda(4k^2 + \ell^2) = \kappa V x + O(x(\log x)^{-A})
$$

with any $A \geq 2$, the implied constant depending only on $A$, where

$$
V = \sum_{h \leq y} \lambda_h g(h)
$$

and $g(h)$ is the multiplicative function with $g(p) = 1/(p-2)$ if $p \equiv 1 \pmod{4}$ and $g(p) = 1/p$ if $p \not\equiv 1 \pmod{4}$. Moreover

$$
\kappa = \prod_p \left( 1 - \frac{\chi(p)}{(p-1)(p-\chi(p))} \right), \quad \chi \pmod{4}.
$$

Before getting to the Main Theorem let us express some principles of its proof. First of all, our arguments borrow substantial parts from the works [FoIw] and [FrIw2], but as we impose restrictions on both coordinates of the Gaussian integers $\ell + 2ki$ some fresh ideas occur. We consider the sequence $A = (a_n)$ of numbers

$$
a_n = \sum_{4k^2 + \ell^2 = n} \beta_k \gamma_\ell
$$

and count them over primes. There will be a lot of Fourier analysis performed so it helps to start with a smoothed counting.

Let $f(t)$ be a function supported on $\frac{1}{2}x \leq t \leq x$, twice differentiable and such that

$$
|t^j f^{(j)}(t)| \leq 1, \quad j = 0, 1, 2.
$$

We are going to evaluate asymptotically the sum

$$
S(x) = \sum_n a_n f(n) \Lambda(n).
$$
Theorem 1.4 (Smoothed Main Theorem). Suppose $\beta_k$ and $\gamma_\ell$ satisfy the conditions of MT. Then

\begin{equation}
S(x) = \kappa V \int f(t) \, dt + O(x(\log x)^{-A})
\end{equation}

with any $A \geq 2$, the implied constant depending only on $A$.

It is not difficult to derive MT from SMT; see a brief explanation in Chapter 18.

Classical ideas for estimating sums of type (1.15) begin by partitioning into a sum of sums

\begin{equation}
A_d(x) = \sum_{n \equiv 0 \pmod{d}} a_n f(n)
\end{equation}

which we call “congruence sums”, and double sums

\begin{equation}
B = \sum_m \sum_n u_m v_n a_{mn} f(mn)
\end{equation}

with suitable coefficients $u_m, v_n$, which we call “bilinear forms”. There are plenty of possibilities, see Chapters 17 and 18 of [FrIw2]. For our purpose we choose Theorem 18.5 of [FrIw2], which is derived by finessing Bombieri’s asymptotic sieve.

The congruence sums are treated in Sections 3, 4 with an application of the large sieve type inequality for roots of the quadratic congruence $\nu^2 + 1 \equiv 0 \pmod{d}$, see Lemma 3.1. The bilinear forms are treated in Sections 7–16. These bilinear forms are modified in various directions to create special features, as required for the application of distinct tools.

2 Interlude: an easier result

If one stares at our sum

\begin{equation}
G_r(x) = \sum_{4k^2 + \ell^2 \leq x} \Lambda_r(k)\Lambda(\ell)\Lambda(4k^2 + \ell^2)
\end{equation}

it seems only natural to ask what happens when we consider the visually similar sum

\begin{equation}
H_r(x) = \sum_{4k^2 + \ell^2 \leq x} \Lambda(k)\Lambda(\ell)\Lambda_r(4k^2 + \ell^2).
\end{equation}

Actually, this is a much easier problem and we can obtain the correct order of magnitude as soon as $r \geq 3$. In the Appendix we give a very short proof of the following result.

Proposition 2.1. We have

\begin{equation}
H_3(x) \asymp x(\log x)^2.
\end{equation}

3 The congruence sums

In this section we extract the main term from the congruence sum $A_d(x)$ and provide a Fourier series expansion for the error term. Then we estimate the absolute remainder (the sum of absolute values of the error terms) in Section 4.
We have

\begin{equation}
A_d(x) = \sum_{4k^2 + \ell^2 \equiv 0(d)} \beta_k \gamma_\ell f(4k^2 + \ell^2)
= \sum_h \sum_\ell \lambda_h \gamma_\ell \sum_{4b^2 h^2 + \ell^2 \equiv 0(d)} f(4b^2 h^2 + \ell^2).
\end{equation}

The summation is void if \(d\) is even so we always assume that \(d\) is odd. Taking advantage of \(\ell\) being an odd prime we insert the restriction \((\ell, d) = 1\) up to an error term \(O(\rho(d)d^{-1}\sqrt{x}\log x)\), where \(\rho(d)\) is the number of roots of

\begin{equation}
\nu^2 + 1 \equiv 0 \pmod{d}.
\end{equation}

Keep in mind that \(\rho(d)\) is multiplicative with \(\rho(p) = 1 + \chi(p)\) where \(\chi\) is the non-principal character modulo 4. Consequently \((h, d) = 1\). Now we split the inner sum over \(b\) into residue classes \(b \equiv \nu \ell \hat{2} h \pmod{d}\) and divide the congruence by \(\ell\) getting

\begin{equation}
\sum_{\nu^2 + 1 \equiv 0(d)} \sum_{b \equiv \nu \ell \hat{2} h (d)} f(4b^2 h^2 + \ell^2).
\end{equation}

Recall the popular notation \(\overline{a} \pmod{d}\) which stands for the multiplicative inverse of \(a \pmod{d}\); \(a \overline{a} \equiv 1 \pmod{d}\) if \((a, d) = 1\). Do not confuse it with complex number conjugation. Working with \((3.3)\) we no longer need the restriction \((\ell, d) = 1\) so we drop it up to the same error term which we committed when installing it.

First we evaluate \((3.3)\) quickly by

\[
\sum_{\nu^2 + 1 \equiv 0(d)} \left( \frac{1}{2dh} \int_0^\infty f(t^2 + \ell^2) dt + O(1) \right) = \frac{\rho(d)}{2dh} I(\ell) + O(\rho(d))
\]

where

\begin{equation}
\begin{split}
I(\ell) = \int_0^\infty f(t^2 + \ell^2) dt.
\end{split}
\end{equation}

Hence our congruence sum satisfies the approximation

\begin{equation}
A_d(x) = \frac{\rho(d)}{2d} V_d W(x) + O(\rho(d)y\sqrt{x})
\end{equation}

with

\begin{equation}
V_d = \sum_{(h, d) = 1} \lambda_h / h
\end{equation}

and

\begin{equation}
W(x) = \sum_\ell \gamma_\ell I(\ell).
\end{equation}
If we use the assumption (1.8) (the PNT for \(\gamma_\ell\) with \(q = 1\)) we get

\[(3.8) \quad W(x) = \frac{\pi}{4} \int f(t) \, dt + O(x(\log x)^{-B}).\]

However, to maintain transparency we shall keep the original expression (3.7) until (1.8) is really needed.

The elementary formula (3.5) suffices for \(d\) odd, uniformly in the range \(d \leq y^{-1}\sqrt{x}(\log x)^{-A}\). By the large sieve for characters \(\chi \pmod{d}\) we can get good results on average over \(d < \sqrt{x}(\log x)^{-A}\). However we can do even better by applying Poisson’s formula to (3.3). We extend the summation over \(b > 0, b \equiv \nu\ell h \pmod{d}\) to all \(b \equiv \nu\ell h \pmod{d}\), thus counting every term twice, except for \(b = 0\) in which case \(d = 1\). We find that (3.3) is equal to

\[
\sum \frac{1}{2dh} \sum_{s} e \left( \frac{\nu s \ell h}{d} \right) F_{\ell} \left( \frac{s}{2dh} \right) - \varepsilon_{d} f(\ell^{2}) = \frac{1}{2dh} \sum \rho_{\nu d \ell h}(d) F_{\ell} \left( \frac{s}{2dh} \right) - \varepsilon_{d} f(\ell^{2})
\]

where \(\varepsilon_{1} = 1, \varepsilon_{d} = 0\) if \(d \neq 1\),

\[(3.9) \quad F_{\ell}(v) = \frac{1}{2} \int_{-\infty}^{\infty} f(t^{2} + \ell^{2})e(-vt) \, dt
\]

and

\[(3.10) \quad \rho_{e}(d) = \sum_{\nu^{2} + 1 \equiv 0(d)} e \left( \frac{\nu c}{d} \right)
\]

is the Weyl harmonic from the theory of equidistribution of the roots of (3.2). Hence, for \(d\) odd we have

\[(3.11) \quad A_{d}(x) = \frac{\rho(d)}{2d} V_{d}W(x) + r_{d}(x) + O \left( \frac{\rho(d)}{d} y\sqrt{x} \log x \right),
\]

where

\[(3.12) \quad r_{d}(x) = \sum_{(h, d) = 1} \lambda_{h} \gamma_{\ell}(dh)^{-1} \sum_{s > 0} \rho_{\nu \ell h}(d) F_{\ell} \left( \frac{s}{2dh} \right).
\]

Here the main term comes from the zero frequency \(s = 0\) and \(r_{d}(x)\) can be considered to be an error term because it will turn out to have small effect due to cancellation in the Weyl harmonics. The last term in (3.11) is negligible.

There is a considerable cancellation of the terms in (3.12) due to the spacing of the fractions \(\nu/d\) modulo 1 as \(\nu\) runs over the roots of (5.2). This property of \(\nu/d\) leads to the following large sieve type inequality.

**Lemma 3.1.** Let \(h \geq 1\). For any complex numbers \(\alpha_{n}\) we have

\[(3.13) \quad \sum_{X < d \leq 2X} \sum_{(d, h) = 1} \sum_{|n| \leq N} |\alpha_{n}|^{2} \left( \frac{\nu nh}{d} \right)^{2} \leq 400(hX + N) \sum_{n \leq N} |\alpha_{n}|^{2}.
\]

**Proof.** See section 20.2 of [FrIw2]. \(\square\)
4 Estimation of the remainder

We need a bound for the remainder

\[ R(x, D) = \sum_{d \leq D \atop d \text{ odd}} |r_d(x)| \]

where \( r_d(x) \) is given by the Fourier series (3.12). Since we shall not take advantage of the summation over \( h \) we partition (3.12) into

\[ r_d(x) = \frac{1}{d} \sum_{(2h, d) = 1} \lambda_h h^{-1} r_d(x; h) \]

where

\[ r_d(x; h) = \sum_{\ell} \gamma_{\ell} \sum_{s > 0} \rho_s \ell^2 F(\frac{s}{2dh}) \]

and we estimate the partial remainders

\[ R_h(X) = \sum_{X < d \leq 2X \atop (d, 2h) = 1} |r_d(x; h)| \]

separately for every \( h \leq y \) and \( 1 \leq 2X \leq D \). We have

\[ |R(x, D)| \leq \sum_{h \leq y} |\lambda_h| h^{-1} \sum_X R_h(X) X^{-1} \]

where \( X = D/2, D/4, D/8, \ldots \).

In order to apply (3.13) we build a single variable \( n = s \ell \) out of the two variables \( s, \ell \) which we need to separate from the modulus \( d \). We accomplish the separation by the change of the variable \( t \) in the Fourier integral (3.9) into \( t \sqrt{s}/s \) getting

\[ F(\frac{s}{2dh}) = \frac{\sqrt{x}}{s} \int_0^{\infty} f(\ell^2 + xt^2/s^2) \cos(\pi t \sqrt{x}/dh) \, dt. \]

The trivial bound

\[ F(\frac{s}{2dh}) \ll \sqrt{x} \]

cannot be improved if \( s \ll dh/\sqrt{x} \asymp hX/\sqrt{x} = S \), say. If \( s \) is larger we can gain by twice integrating (4.6) by parts. We obtain another expression

\[ F(\frac{s}{2dh}) = -2\sqrt{x} \left( \frac{dh}{\pi s} \right)^2 \int_0^{\infty} \left( f' + \frac{2xt^2}{s^2} f'' \right) \cos(\pi t \sqrt{x}/dh) \, dt \]

where the derivatives \( f', f'' \) are evaluated at \( \ell^2 + xt^2/s^2 \). Now estimating (4.8) trivially we get

\[ F(\frac{s}{2dh}) \ll \sqrt{x} (S/s)^2. \]
Let $S_0 \geq 1$. The part of (4.3) with $S_0 \leq s < 2S_0$ is estimated by

$$\sqrt{x} \int_0^{2S_0} \left| \sum_{n \leq N} \alpha_n(t) \rho_n x^{21}(d) \right| dt$$

with $N = 2\sqrt{x}$, where $\alpha_n(t)$ does not depend on $d$,

$$\alpha_n(t) = \sum_{\ell s \leq n < 2\ell s} \gamma \ell s^{-1} f(x^2 + xt^2/s^2) \ll S_0^{-1} \log n.$$  

Summing (4.11) over $X < d \leq 2X$ with $(d, 2h) = 1$ we derive by (3.13) (apply the Cauchy-Schwarz inequality) that the partial remainder (4.4) restricted by $S_0 \leq s < 2S_0$ is bounded by

$$(4.12) \quad (xX)^{1/2} (hX + \sqrt{xS_0})^{1/2} (\sqrt{xS_0})^{1/2} \log(xS_0).$$

We have derived (4.12) using the formula (4.6). Similarly, if we use the formula (4.8), then we get the bound (4.12) with an extra factor $(S/S_0)^{2}$. Combining both bounds we see that, with optimal cutoff point, the worst result comes from $S_0 \asymp S = hX/\sqrt{x}$. Hence, we conclude that

$$(4.13) \quad R_h(X) \ll hX^{3/2} \log x.$$  

Finally, inserting (4.13) into (4.5) we obtain

**Proposition 4.1.** We have

$$(4.14) \quad R(x, D) \ll y(Dx)^{1/2} \log x.$$

**Remark 4.1.** The bound (4.14) is useful if $y^2 D \ll x(\log x)^{-A}$.

### 5 A model for $A = (a_n)$

By means of multiplicative functions we construct a sequence for which the main terms of the congruence sums agree with those for $A_d(x)$. We consider $B = (b_n)$ with the numbers

$$b_n = \psi(n) \sum_{(2h, n) = 1} \lambda_h \phi(h)/h$$

where the multiplicative functions $\psi(n)$ and $\phi(h)$ are given by

$$\psi(2^\alpha) = 1, \quad \phi(2^\alpha) = 1$$

and

$$\psi(p^\alpha) = \rho(p) \left(1 - \frac{1}{p} \right) \left(1 - \frac{\rho(p)}{p} \right)^{-1}, \quad \phi(p^\alpha) = \left(1 - \frac{\rho(p)}{p} \right)^{-1},$$

if $p \neq 2$ and $\alpha \geq 1$. Recall that $\rho(p) = 1 + \chi(p)$ is the number of roots of $\nu^2 + 1 \equiv 0 (\text{mod } p)$. 

9
Let \( w(y) \) be a smooth function supported on \( 0 < y < 1 \) with
\[
(5.4) \quad \int_0^1 w(y) \, dy = 1.
\]
We are going to evaluate asymptotically the sum
\[
(5.5) \quad B_d(x) = \sum_{n \equiv 0 \, (\text{mod } d)} b_n w(n/x).
\]
Note that \( B_d(x) = 0 \) if \( d \) is even.

**Proposition 5.1.** For \( d \) odd we have
\[
(5.6) \quad B_d(x) = x \frac{\rho(d)}{2d} V_d + O \left( \frac{\rho(d)}{\sqrt{d}} \prod_{p \mid d} \left( 1 + \frac{1}{\sqrt{p}} \right) \sqrt{x \log x} \right)
\]
where \( H \) is the constant
\[
(5.7) \quad H = \prod_p \left( 1 - \frac{\rho(p)}{p} \right) \left( 1 - \frac{1}{p} \right)^{-1}.
\]
The implied constant in (5.6) depends only on the crop function \( w \).

**Proof.** We execute the summation via \( L \)-functions rather than by Poisson’s formula. We have
\[
(5.8) \quad B_d(x) = \sum_{(h,d) = 1} \lambda_h \frac{\phi(h)}{h} \sum_{(n, 2h) = 1} \psi(dn) w(dn/x).
\]
The corresponding Dirichlet series is equal to
\[
L(s) = \sum_{(n, 2h) = 1} \psi(dn)(dn)^{-s} = \frac{\psi(d)}{d^s} \prod_{p \mid d} \left( 1 - \frac{1}{p^s} \right)^{-1} \prod_{p \mid 2dh} \left( 1 + \frac{\psi(p)}{p^s} \left( 1 - \frac{1}{p^s} \right)^{-1} \right)
\]
\[
= \frac{\psi(d)}{d^s} \zeta(s) \prod_{p \mid 2h} \left( 1 - \frac{1}{p^s} \right) \prod_{p \mid 2dh} \left( 1 + \frac{\chi(p)}{p^s} \left( 1 - \frac{\rho(p)}{p} \right)^{-1} \right) \prod_{p \mid dh} \left( 1 + \frac{\psi(p)}{p^s} \left( 1 - \frac{\rho(p)}{p} \right)^{-1} \right).
\]
Now we borrow \( L(s, \chi)/\zeta(2s) \) and return it in the form of its Euler product, getting
\[
L(s) = \frac{\zeta(s) L(s, \chi)}{\zeta(2s)} P(s) \frac{\psi(d)}{d^s} \prod_{p \mid 2h} \left( 1 - \frac{1}{p^s} \right) \prod_{p \mid dh} \left( 1 + \frac{\chi(p)}{p^s} \left( 1 - \frac{\rho(p)}{p} \right)^{-1} \right)^{-1}
\]
where
\[
P(s) = \prod_p \left( 1 - \frac{\chi(p)}{p^s} \right) \left( 1 - \frac{1}{p^{2s}} \right)^{-1} \left( 1 + \frac{\chi(p)}{p^s} \left( 1 - \frac{\rho(p)}{p} \right)^{-1} \right).
\]
For \( p \neq 2 \) the local factor of \( P(s) \) is 
\[ 1 + \frac{\chi(p)}{(p - 1 + \chi(p))}(p^s + \chi(p)) \]
so the product converges for \( \text{Re} \, s > 0 \). We compute the residue of \( L(s) \) at \( s = 1 \)

\[ \text{res}_{s=1} L(s) = P \frac{\psi(d)}{d} \prod_{p | 2h} \left( 1 - \frac{1}{p} \right) \prod_{p | dh} \left( 1 - \frac{\rho(p)}{p} \right) \left( 1 - \frac{1}{p} \right)^{-1} = \frac{P \rho(d)}{2d \phi(h)} \]

where \( P = P(1)L(1, \chi)/\zeta(2) \). Checking the local factors we find

\[ P = \prod_{p} \left( 1 - \frac{\rho(p)}{p} \right)^{-1} \left( 1 - \frac{1}{p} \right) = \frac{1}{H}. \]

Finally (5.6) follows by contour integration with the error term obtained by trivial estimations on the line \( \text{Re} \, s = \frac{1}{2} \).

**Remarks 5.1.** The main term of (5.6) agrees with that of (3.11) after normalization. Checking the local factors of \( H \) in (5.7) and \( \kappa \) in (1.12) against \( L(1, \chi) \) we see that

\[ \kappa = HL(1, \chi) = \frac{\pi}{4} H. \]

### 6 Sums over primes

Theorem 18.5 of [FrIw2] gives an inequality between a sum over primes, sums of congruence sums and a bilinear form. We can use this inequality as it stands, but we get faster results with a slightly different inequality (which is actually derived in [FrIw2], but not stated explicitly).

**Proposition 6.1.** Let \( 1 < z \leq \sqrt{x} \). For any complex numbers \( c_n \) we have

\[ \sum_{xz^{-2} < n \leq x} c_n \Lambda(n) \leq \sum_{d \leq z} \mu(d) C'_d(x) + (\log x) \sum_{d \leq xz^{-1}} |C_d(x)| \]

\[ + 2(\log x) \sum_{n} \left| \sum_{mn \leq x \atop z < m \leq z^2} \mu(m)c_{mn} \right| \]

where

\[ C'_d(x) = \sum_{n \leq x \atop n \equiv 0(d)} c_n \log n, \quad C_d(x) = \sum_{n \leq x \atop n \equiv 0(d)} c_n. \]

**Remarks 6.1.** The double sum over \( m, n \) is a bilinear form. The key feature of this form is that the inner sum is weighted by the clean Möbius function \( \mu(m) \); it is not contaminated by some incomplete Dirichlet convolutions presented by similar identities in the literature. Moreover, we sum \( \mu(d)C'_d(x) \) with the Möbius factor \( \mu(d) \) rather than with absolute values. This slight (not vital) difference will simplify our work.
We apply (6.1) with \( z = x^\delta \), \( 0 < \delta \leq \frac{1}{8} \), for the sequence of numbers

\[
c_n = a_n f(n) - H \frac{W(x)}{x} \psi(n) b_n
\]

where \( A = (a_n) \) is our target sequence (1.13) and \( B = (b_n) \) is its model (5.1). Note that \( c_n = 0 \), unless \( n < x, n \) odd. The congruence sums of \( C = (c_n) \) have no main term; compare (3.11) with (5.6).

On the left-hand side of (6.1) we get (up to \( O(x z^3) \))

\[
\sum_n a_n f(n) \Lambda(n) - H \frac{W(x)}{x} \sum_n b_n \Lambda(n) w \left( \frac{n}{x} \right)
\]

\[
= S(x) - H \frac{W(x)}{x} \sum_h \lambda_h \phi(h) \sum_{(n, 2h) = 1} \psi(n) \Lambda(n) w \left( \frac{n}{x} \right)
\]

\[
= S(x) - H \frac{W(x)}{x} V \sum_{p \equiv 1(4)} 2w \left( \frac{p}{x} \right) \log p + O(\sqrt{x} \log x)
\]

\[
= S(x) - HVW(x) + O(x(\log x)^{1/2})
\]

by the PNT, where \( A \) is any number \( \geq 2 \).

On the right-hand side of (6.1) we get three sums. The first sum is

\[
R' = \sum_{d \leq z \atop d \text{ odd}} \mu(d) \left( \sum_{n \equiv 0(d)} a_n f(n) \log n - H \frac{W(x)}{x} \sum_{n \equiv 0(d)} b_n w \left( \frac{n}{x} \right) \log n \right).
\]

The second sum is

\[
R = \sum_{d \leq xz^{-1} \atop d \text{ odd}} \left| A_d(x) - H \frac{W(x)}{x} B_d(x) \right|.
\]

The third sum is the bilinear form

\[
B = \sum_n \left| \sum_{z < m \leq z^2} \mu(m) a_m f(mn) - H \frac{W(x)}{x} \sum_{z < m \leq z^2} \mu(m) b_m w \left( \frac{mn}{x} \right) \right|.
\]

We estimate \( R' \) by applying two elementary approximations to the main terms, namely (3.5) with \( f(t) \) replaced by \( f(t) \log t \) and (5.6) with \( w(y) \) replaced by \( w(y) \log xy \). We obtain

\[
R' = \sum_{d \leq z \atop d \text{ odd}} \mu(d) \frac{\rho(d)}{2d} V_d \sum_{\ell} \gamma_{\ell} \int_0^{\infty} f(t^2 + \ell^2) \int_0^1 w(y) \log \left( \frac{t^2 + \ell^2}{xy} \right) dy dt + O(yz \sqrt{x} \log x).
\]

Note that the extra logarithmic factors \( \log t \) and \( \log xy \) in the crop functions make the resulting main term different. They do not match exactly, yet they are close. If \( f(t) \) is supported in a relatively short interval centered at \( cx \) with the constant

\[
c = \exp \left( \int w(y) \log y \, dy \right)
\]

12
then the above main terms cancel out up to a sufficiently small error term, showing that \( R' \) is negligible. But we do not need to make such a restriction for \( f(t) \), because we may exploit cancellation from the summation over \( d \). Indeed, by the PNT we get

\[
\sum_{d \leq z \atop \text{d odd}} \mu(d) \frac{\rho(d)}{2d} V_d = \sum_h \lambda_h \sum_{d \leq z \atop (d,2h)=1} \mu(d) \frac{\rho(d)}{2d} \ll (\log z)^{-A}.
\]

Hence

(6.4) \( R' \ll x (\log x)^{-A} \)

with any \( A \geq 2 \), the implied constant depending only on \( A \).

In the second sum \( R \) the main terms match exactly, they cancel out and the remaining terms are estimated in (4.14), (5.6), respectively. We get

(6.5) \( R \ll z^{-\frac{1}{2}} y x (\log x) \).

In the bilinear form \( B \) we also get cancellation due to sign changes of the Möbius function \( \mu(m) \).

It is difficult to see that \( \mu(m) \) does not correlate with the original sequence \( a_{mn} \), but this is clear for the model sequence \( b_{mn} \). We have

\[
\sum_{z \leq m \leq z^2} \mu(m) b_{mn} w\left(\frac{mn}{x}\right) = \sum_{(2h,n)=1} \lambda_h \phi(h) \sum_{z \leq m \leq z^2 \atop (m,2h)=1} \mu(m) \psi(mn) w\left(\frac{mn}{x}\right).
\]

By the PNT we find that the last sum over \( m \) is \( \ll n^{-\frac{1}{2}} x (\log x)^{-A-3} \). Next, summing over \( h \leq y \) and \( n < x z^{-1} \) we lose a factor \( (\log x)^2 \). Hence the total contribution of the model sequence to the bilinear form \( B \) is \( \ll x (\log x)^{-A-1} \) so we are left with

(6.6) \[ B(x,z) = \sum_n \left| \sum_{z \leq m \leq z^2} \mu(m) a_{mn} f(mn) \right|. \]

Adding up the above estimates we conclude this section with the following result which does not contain the model sequence.

**Proposition 6.2.** Let

(6.7) \( y^2 (\log x)^{2A+4} \leq z \leq x^\frac{1}{2} \).

Then

(6.8) \[ |S(x) - HVW(x)| \leq 2B(x,z) \log x + O\left(x (\log x)^{-A}\right). \]

If we assume (1.8), then \( W(x) \) satisfies (3.8) so (6.8) becomes

(6.9) \[ |S(x) - \kappa V \int f(t) \, dt| \leq 2B(x,z) \log x + O\left(x (\log x)^{-A}\right). \]

To complete the proof of (1.16) it remains to show that

(6.10) \[ B(x,z) \ll x (\log x)^{-A-1} \]

subject to the condition (6.7).
7 Bilinear forms in the Gaussian Domain

It remains to estimate the bilinear form (4.11). We need the bound

\[ B(x, z) \ll x (\log x)^{-A} \]

with any \( A \geq 2 \). In this section we make several simplifications before launching the essential arguments.

First we split the segment \( z < m \leq z^2 \) into dyadic intervals \( M < m \leq 2M \). Assume for simplicity that \( \log z / \log 2 \) is an integer so we cover the segment exactly with \( 2 \log z / \log 2 \) dyadic intervals. We get

\[ B(x, z) \leq \sum_M \sum_n \left| \sum_{m \sim M} \mu(m) a_{mn} f(mn) \right| \]

where \( M \) runs over the numbers \( z, 2z, 4z, \ldots \). Next we transfer the common factor \( c = (m, n) \) from \( m \) to \( n \) getting

\[ B(x, z) \leq \sum_M \sum_n \sum_{c^2 | n} \left| \sum_{m \sim M/c} \mu(m) a_{mn} f(mn) \right| \]

The contribution of terms with \( c > C \) is estimated trivially by

\[ \sum_h |\lambda_h| h^{-1} \sum_{c > C} \rho(c) c^{-2} x \log x \ll C^{-1} x (\log x)^2. \]

This bound satisfies (7.1) if \( C = (\log x)^{A+3} \). Now we ignore the condition \( c^2 | n \) for \( c \leq C \) getting

\[ B(x, z) \leq B^*(M)(\log x)^{A+4} + O(x (\log x)^{-A-1}) \]

where

\[ B^*(M) = \sum_n \left| \sum_{m \sim M} \mu(m) a_{mn} f(mn) \right| \]

for some \( M \) with \( z/C \leq M < z^2 \). Note that the support of \( f(t) \) implies that \( n \) runs over the segment \( N/4 < n < N \) with \( MN = x \).

Next we write (see (1.6) and (1.9))

\[ a_n = \sum_h \lambda_h a_n(h) \]

where

\[ a_n(h) = \sum_{4k^2 + \ell^2 = n} \lambda_{k^2 + \ell^2} \]

Hence

\[ B^*(M) \leq \sum_h |\lambda_h| B^*_h(M) \]

14
where

\[
B_h^*(M) = \sum_n \sum_{m \sim M} \mu(m)a_{mn}(h) f(mn).
\]

Note that we have introduced the restriction \((m, 2h) = 1\), which is permitted because it is redundant. Indeed, if \(e = (m, 2h) \neq 1\), then \(e|\ell^2, e|\ell, e^2|mn, e^2|m\), contradiction!

Typically, for bilinear forms of this nature, one applies Cauchy’s inequality and interchanges the order of summation. However, in our case \(a_{mn}(h)\) has multiplicity which would become more difficult to treat after application of Cauchy’s inequality. Our next step is to express the variables in terms of Gaussian integers so that there is no multiplicity, after which Cauchy’s inequality can be applied without leading to such complications.

In the following the gothic letters \(a, b, m, n, \ldots\) denote Gaussian integers and the corresponding Latin letters \(a, b, m, n, \ldots\) denote the norms; \(a = \overline{a}, b = \overline{b}\). \(m = \overline{m}, n = \overline{n}, \ldots\) By the unique factorization in \(\mathbb{Z}[i]\) we obtain \(B_h^*(M) \leq B_h^*(M)\) where

\[
B_h^*(M) = \sum_n \sum_{m \sim M} \mu(m)|\xi(mn)|f(mn).
\]

Here we put

\[
(7.2) \quad \xi(a) = \gamma \Re a.
\]

Note that \(m = \overline{m}\) is squarefree odd so this inner sum runs over Gaussian integers \(m\) with \((m, \overline{m}) = 1\) (called primitive). In this case the M"obius function \(\mu(m)\) in rational integers agrees with the M"obius function \(\mu(m)\) in Gaussian integers. For notational convenience we shall be writing \(m \sim M\) to say that \(m = \overline{m}\sim M\).

The condition \((m, n) = 1\) was needed for performing the unique factorization in \(\mathbb{Z}[i]\). After that, the resulting condition \((m, n) = 1\) is a hindrance so we are going to remove it using a similar argument by which we inserted it, but now in the Gaussian domain.

We start from the formula

\[
\sum_{bc = m, c \in \mathbb{Z}[i]} \mu(b) = \begin{cases} 4\mu(m) & \text{if } (m, q) = 1 \\ 0 & \text{otherwise} \end{cases}
\]

which holds for any \(m, q\) in \(\mathbb{Z}[i]\), \(mq \neq 0\) (the factor 4 accounts for four units). Hence the inner sum in \(B_h^*(M)\) is bounded by

\[
\sum_{c|m, c \in \mathbb{Z}[i]} \sum_{(cm, 2h) = 1, cm \sim M} \mu(m)|\xi(cm)|f(cm) \sum_{m, \ell, (m, 2h) = 1} \mu(m)|\xi(mn)|f(mn)
\]

and

\[
B_h^*(M) \leq \sum_n \sum_{c|m, c \in \mathbb{Z}[i]} \sum_{(m, 2h) = 1} \mu(m)|\xi(mn)|f(mn).
\]
We keep the terms with $c \leq C_1 = (\log x)^{2A+8}$ and estimate the remaining terms with larger $c$ trivially getting

\[ B(x, z) \leq B(M)(\log x)^{3A+12} + O\left((\log x)^{-A-1}\right) \]

for some $M$ with $z/C_1 \leq M < z^2$, where

\[ B(M) = \sum_h |\lambda_h| B_h(M) \]

and

\[ B_h(M) = \sum_n \left| \sum_{m \sim M, (m, 2h)=1} \mu(m)\xi(mn)f(mn) \right|. \]

Now we need to show that

\[ B(M) \ll x(\log x)^{-4A-13} \]

for some $M$ with

\[ z(\log x)^{-3A-11} \leq M < z^2. \]

Some properties of $n$ in the outer sum of (7.5) are hidden but can be inferred from the equation $mn = 4bh^2 + \ell^2$ and the support of $f(mn)$ being $x/2 < mn < x$. In particular, the inequality $\ell < \sqrt{x}$ is redundant information in every expression containing the crop function $f$. From now on the dyadic segment $m \sim M$ never changes so sometimes we skip writing $m \sim M$ or $m \sim M$, but never forget it.

Now we are ready to apply Cauchy’s inequality as follows:

\[ B^2(M) \ll C(M)N \log y \]

where

\[ C(M) = \sum_h |\lambda_h| hC_h(M) \]

and

\[ C_h(M) = \sum_n \left| \sum_{m \sim M, (m, 2h)=1} \mu(m)\xi(mn)f(mn) \right|^2. \]

Note that we borrowed a factor $h$ into $C(M)$. Now we need to show that

\[ C(M) \ll NM^2(\log x)^{-8A-27}. \]
Squaring out and interchanging the order of summation we write

\[ C_h(M) = \sum_{m_1 \sim M, m_2 \sim M} \mu(m_1) \mu(m_2) D_h(m_1, m_2) \]

with

\[ D_h(m_1, m_2) = \sum_n \xi(m_1 n) \bar{\xi}(m_2 n) f(m_1 n) \bar{f}(m_2 n) \]

where the summation runs over all Gaussian integers \( n \) satisfying

\[ \text{Im} m_1 n \equiv \text{Im} m_2 n \equiv 0 \pmod{2 h}. \]

Opening the Gaussian domain we see that

\[ D_h(m_1, m_2) = \sum_{\ell_1} \sum_{\ell_2} \gamma_{\ell_1} \gamma_{\ell_2} f(m_1 n) \bar{f}(m_2 n) \]

where the summation runs over the solutions of the system

\[
\begin{align*}
    m_1 n &= \ell_1 + 2 h b_1 i \\
    m_2 \bar{n} &= \ell_2 + 2 h b_2 i
\end{align*}
\]

in \( n, \ell_1, \ell_2 \) and \( b_1, b_2 \). Since \( b_1, b_2 \) run over rational integers unrestricted, equivalently we can express this system by two congruences

\[ m_1 n \equiv \bar{m}_1 \bar{n} \pmod{4 h} \]
\[ m_2 n \equiv \bar{m}_2 \bar{n} \pmod{4 h} \]

and two equations

\[ m_1 n + \bar{m}_1 \bar{n} = 2 \ell_1 \]
\[ m_2 n + \bar{m}_2 \bar{n} = 2 \ell_2. \]

Put

\[ \Delta = \Delta(m_1, m_2) = \frac{i}{2} (m_1 \bar{m}_2 - \bar{m}_1 m_2) = \text{Im} \, \bar{m}_1 m_2 \]

so \( \Delta \) is a rational integer, relatively small;

\[ |\Delta| < 4 M < 4 \pi^2. \]

8 The diagonal terms

First we give a quick estimation of \( D_h(m_1, m_2) \) in the singular case \( \Delta = \Delta(m_1, m_2) = 0 \). We get

\[ m_1 \bar{m}_2 = \bar{m}_1 m_2, \quad m_1 | m_2 \text{ and } m_2 | m_1, \quad m_2 = \varepsilon m_1 \text{ with } \varepsilon = \pm 1, \pm i. \]

From the system (7.17) we obtain

\[ \ell_1 m_2 - \ell_2 m_1 = -i \Delta \bar{n}. \]
In the singular case this yields \( \ell_1 = \ell_2 \), so \( \varepsilon = 1 \). Therefore we have \( m_1 = m_2 = m \) and \( \ell_1 = \ell_2 = \ell \), say. In this case \( D_h(m, m) \) is bounded by the number of solutions in \( b \) and \( \ell \) of

\[
\ell + 2hbi \equiv 0 \pmod{m}, \quad |\ell + 2hbi| < \sqrt{x}.
\]

Here \( |b| < \sqrt{x}/2h \) and \( \ell < \sqrt{x} \), \( \ell^2 + 4h^2b^2 \equiv 0 \pmod{m} \). Hence we conclude that

(8.2) \( D_h(m, m) \ll x\rho(m)/mh. \)

The contribution of \( D_h(m, m) \) to \( C_h(M) \) is estimated by

(8.3) \[
\frac{x}{h} \sum_{m \sim M} \rho(m) \leq \frac{x}{h} \sum_{m \sim M} \frac{\rho(m)^2}{m} \ll \frac{x}{h} (\log M)^2.
\]

Hence the contribution of \( D_h(m, m) \) to \( C(M) \) is

\( \ll yx(\log x)^2 \ll NM^2(\log x)^{-8A-27} \) by (7.9) as required by (7.11), provided

(8.4) \( y(\log x)^{11A+40} \ll z. \)

9 In the off-diagonal area

From now on we assume that \( \Delta = \Delta(m_1, m_2) \neq 0 \). Now the system of equations (7.17) has a unique solution in the complex number \( n \) given by

(9.1) \( in = \ell_1\overline{m_2} - \ell_2\overline{m_1}. \)

Since \( n \) must be a Gaussian integer this means \( \ell_1, \ell_2 \) satisfy

(9.2) \( \ell_1 m_2 \equiv \ell_2 m_1 \pmod{\Delta}. \)

For \( n \) given by (9.1) the congruences (7.10) become

\[
\begin{align*}
m_1(\ell_1\overline{m_2} - \ell_2\overline{m_1}) + \overline{m_1}(\ell_1 m_2 - \ell_2 m_1) & \equiv 0 \pmod{4\Delta h}, \\
m_2(\ell_1\overline{m_2} - \ell_2\overline{m_1}) + \overline{m_2}(\ell_1 m_2 - \ell_2 m_1) & \equiv 0 \pmod{4\Delta h}.
\end{align*}
\]

We write these congruences in the form similar to (9.2):

(9.3) \( \ell_1(m_1\overline{m_2} + \overline{m_1}m_2) \equiv 2\ell_2 m_1 \pmod{4\Delta h}, \)

(9.4) \( \ell_2(m_1\overline{m_2} + \overline{m_1}m_2) \equiv 2\ell_1 m_2 \pmod{4\Delta h}. \)

In other words the summation in (7.13) runs over the odd prime numbers \( \ell_1, \ell_2 \) satisfying the congruences (9.2), (9.3), (9.4), and \( n = n\overline{m} \) is determined by (9.1).

The congruences (9.2), (9.3), (9.4) imply several conditions on \( m_1, m_2 \). It will be easier to see these conditions after pulling out the common factor \( d = (m_1, m_2) \). We put (temporarily)

\[
m_1 = a_1d, \quad m_2 = a_2d \quad \text{with} \quad (a_1, a_2) = 1.
\]

Note that \( d \) is primitive and \( (d, 2ha_1a_2) = 1 \), because \( m_1, m_2 \) are primitive squarefree, co-prime with \( 2h \). Put

\[
D = \Delta(a_1, a_2) = \Delta(m_1, m_2)/d, \quad d = \overline{d}d.
\]

18
Note that \((a_1 a_2, 2D) = 1\). Dividing \((9.2)\) by \(d\) and conjugating we get
\[
el_1 \bar{\alpha}_2 \equiv e_2 \bar{\alpha}_1 \pmod{dD},
\]
and dividing \((9.3), (9.4)\) by \(d\) we get
\[
el_1 (a_1 \bar{\alpha}_2 + \bar{\alpha}_1 a_2) \equiv 2e_2 a_1 \pmod{4Dh}
\]
\[
el_2 (a_1 \bar{\alpha}_2 + \bar{\alpha}_1 a_2) \equiv 2e_1 a_2 \pmod{2Dh}.
\]
Recall that \(a_1 = a_1 \bar{\alpha}_1\) and \(a_2 = a_2 \bar{\alpha}_2\). Since \(a_1 \bar{\alpha}_2 + \bar{\alpha}_1 a_2 = 2a_1 \bar{\alpha}_2 + 2iD\) we can write \((9.6), (9.7)\) in the form
\[
el_1 (a_1 \bar{\alpha}_2 + iD) \equiv e_2 a_1 \pmod{2Dh}
\]
\[
el_2 (a_1 \bar{\alpha}_2 + iD) \equiv e_1 a_2 \pmod{2Dh}.
\]
Multiplying these congruences by sides and dividing by \(e_1 e_2\) we get
\[
D^2 \equiv 0 \pmod{2Dh},
\]
hence
\[
D \equiv 0 \pmod{2h} \quad \text{and} \quad \Delta \equiv 0 \pmod{2h}.
\]
Having the condition \((9.10)\) it is now clear that \((9.8)\) is equivalent to \((9.9)\). Indeed \((9.8)\) implies
\[
a_1 e_2 (a_1 \bar{\alpha}_2 + iD) \equiv e_2 a_1 \equiv e_1 (a_1 \bar{\alpha}_2 + iD)^2 = e_1 (a_1 \bar{\alpha}_2 + 2iD a_1 \bar{\alpha}_2 - D^2) = a_1 e_1 a_2 - e_1 D^2 \equiv a_1 e_1 a_2 \pmod{2Dh}
\]
which yields \((9.9)\). Conversely \((9.9)\) implies \((9.8)\) by similar arguments.

Now we are left with \((9.5)\) and \((9.9)\). These two congruences determine \(e_2/e_1\) uniquely modulo the least common multiple of \(dD\) and \(2Dh\) which is \(2dDh\). Since \(e_2/e_1\) is rational it is determined uniquely modulo the least common multiple of \(2Dh\) and \(2Dh\) which is \(2dDh\). Therefore we can write the two congruences \((9.6)\) and \((9.9)\) for \(e_1, e_2\) as one congruence
\[
el_2 \equiv \omega e_1 \pmod{2Dh}
\]
where \(\omega\) is the unique rational reduced residue class modulo \(2Dh = 2\Delta h\) such that
\[
\omega a_1 \equiv a_1 \bar{\alpha}_2 \pmod{dD}
\]
and
\[
\omega a_1 \equiv a_1 \bar{\alpha}_2 + iD \pmod{2Dh}.
\]
By \((7.15)\) and \((9.11)\) we can write
\[
D_h(m_1, m_2) = \sum_{e_2 \equiv \omega e_1 \pmod{2\Delta h}} \gamma e_1 \bar{e}_2 f(m_2 n) \overline{f}(m_1 n)
\]
with \(n\) given by \((9.11)\), that is \(n\) is a quadratic form in \(e_1, e_2\);
\[
n = n(e_1, e_2) = |e_1 m_2 - e_2 m_1|^2 \Delta^{-2}.
\]
By the distribution of primes $\ell_1$, $\ell_2$ in arithmetic progressions we expect that the main term of (9.14) should be

\begin{equation}
E_h(m_1, m_2) = \frac{1}{\varphi(2\Delta h)} \sum_{\ell_1} \sum_{\ell_2} \gamma_{\ell_1} \gamma_{\ell_2} f(m_2n) \overline{f}(m_1n)
\end{equation}

which does not depend on $\omega$. Subtracting $E_h(m_1, m_2)$ from $D_h(m_1, m_2)$ we get

\begin{equation}
R_h(m_1, m_2) = \sum_{\ell_2 \equiv \omega \ell_1 (2\Delta h)} \gamma_{\ell_1} \gamma_{\ell_2} f(m_2n) \overline{f}(m_1n) - \frac{1}{\varphi(2\Delta h)} \sum_{\ell_1} \sum_{\ell_2} \gamma_{\ell_1} \gamma_{\ell_2} f(m_2n) \overline{f}(m_1n)
\end{equation}

which is regarded as an error term.

We need to sum $E_h(m_1, m_2)$ and $R_h(m_1, m_2)$ over $m_1$, $m_2$ as in (10.12) and over $h$ as in (7.3) restricted by $\Delta(m_1, m_2) \equiv 0 \pmod{2h}$, see (9.10). Therefore our moduli $2\Delta h$ run over multiples of $4h^2$.

### 10 Separation of variables

In Section 12 we shall estimate the error terms by means of the large sieve. To this end we need to separate the variables $\ell_1$, $\ell_2$ from $m_1$, $m_2$, because $m_1$, $m_2$ are constituents of the moduli $\Delta(m_1, m_2)h$. Although in most cases the determinant $\Delta(m_1, m_2)$ is as large as $M$, it can take smaller values which require special attention. Our technique of separation of variables addresses this issue.

We are going through the Fourier transform of

\begin{equation}
f(x_1, x_2) = f(m_2n(x_1, x_2)) = \int f(x_1, x_2) e(\alpha_1 x_1 + \alpha_2 x_2) d\alpha_1 d\alpha_2
\end{equation}

where

\begin{equation}g(\alpha_1, \alpha_2) = \int f(x_1, x_2) e(-\alpha_1 x_1 - \alpha_2 x_2) dx_1 dx_2.
\end{equation}

Recall that $n(x_1, x_2)$ is the quadratic form given by (9.15). By the linear change of variables $(x_1, x_2) = (x, y)$ given by

$$x_1 = x \text{ Im } \frac{m_1}{m_2} + y \text{ Re } \frac{m_1}{m_2}, \quad x_2 = y,$$

we diagonalize $n(x_1, x_2) = (x^2 + y^2)/m_2$ getting

$$g(\alpha_1, \alpha_2) = I \int f(x^2 + y^2) \overline{f} \left(\frac{m_1}{m_2}(x^2 + y^2)\right) e(-\alpha_1 x - (\alpha_2 + \alpha_1 R)y) dx dy$$

where we denote temporarily $I = \text{ Im } m_1/m_2$ and $R = \text{ Re } m_1/m_2$. Note that $I = -\Delta(m_1, m_2)/m_2 \neq 0$ so $|I| > M^{-1}$. Moreover $I^2 + R^2 = m_1/m_2 \approx 1$ so if $I$ is small, then $|R| \approx 1$.

Because $f(x^2 + y^2) \overline{f}((x^2 + y^2)m_1/m_2)$ is radial, so is its Fourier transform. Precisely, it holds in general that

\begin{equation}I \int f(x^2 + y^2) e(-ax - by) dx dy = F(a^2 + b^2)
\end{equation}
where $F(s)$ is the Hankel transform of $f(t)$,

$$F(s) = \pi \int_0^\infty J_0(2\pi \sqrt{st}) f(t) \, dt.$$  

(10.4)

Here $J_0(z)$ is the Bessel function

$$J_0(z) = \frac{1}{\pi} \int_0^\pi \cos(z \cos \alpha) d\alpha.$$

For the proof of (10.3) apply polar coordinates.

In our case $s = a^2 + b^2 = (\alpha_1 I)^2 + (\alpha_2 + \alpha_1 R)^2$ is the quadratic form

$$s(\alpha_1, \alpha_2) = |\alpha_2 + \alpha_1 m_1/m_2|^2 = \alpha_2^2 + 2\alpha_1 \alpha_2 \text{Re} \frac{m_1}{m_2} + \alpha_1^2 \frac{m_1}{m_2},$$

(10.5)

$$F(s) = \pi \int_0^\infty J_0(2\pi \sqrt{st}) f(t) f(tm_1/m_2) \, dt,$$

(10.6)

g(\alpha_1, \alpha_2) = IF(s(\alpha_1, \alpha_2))
and

$$f(m_2 n) f(m_1 n) = I \iint F(s(\alpha_1, \alpha_2)) e(\alpha_1 \ell_1 + \alpha_2 \ell_2) \, d\alpha_1 \, d\alpha_2.$$  

(10.7)

Going through the Fourier transform we lost sight on the ranges of $\ell_1, \ell_2$ so let us record that

$$|\ell_1, \ell_2| < \sqrt{x}.$$  

(10.9)

This information is redundant when the original function (10.1) is present.

Estimating directly and after integrating by parts two times of (10.6) we find that

$$F(s) \ll x(1 + sx)^{-2}.$$  

(10.10)

Hence $F(s)$ is very small if $s > x^{-1}(\log x)^{2C}$ so the integration (10.8) runs effectively over the set (ellipse)

$$S = \{(\alpha_1, \alpha_2) \in \mathbb{R}^2 : s(\alpha_1, \alpha_2) = (\alpha_1 I)^2 + (\alpha_2 + \alpha_1 R)^2 \leq x^{-1}(\log x)^{2C}\}$$  

(10.11)

whose volume (the Lebesgue measure) is equal to

$$|S| = \frac{\pi (\log x)^{2C}}{|I|x}.$$  

(10.12)

Note that the trivial integration shows that (10.8) is bounded

$$|I| \iint_{\mathbb{R}^2} |F(s(\alpha_1, \alpha_2))| \, d\alpha_1 \, d\alpha_2 = \iint_{\mathbb{R}^2} |F(\alpha_1^2 + \alpha_2^2)| \, d\alpha_1 \, d\alpha_2$$

$$\ll x \iint_{\mathbb{R}^2} (1 + (\alpha_1^2 + \alpha_2^2)x)^{-2} \, d\alpha_1 \, d\alpha_2 = \iint_{\mathbb{R}^2} (1 + \alpha_1^2 + \alpha_2^2)^{-2} \, d\alpha_1 \, d\alpha_2 \ll 1.$$  

(21)
Similarly we find that the integral over $\mathbb{R}^2 \setminus S$ is small:

$$
|I| \int_{\mathbb{R}^2 \setminus S} |F(s(\alpha_1, \alpha_2))| \, d\alpha_1 \, d\alpha_2 < |I|x(\log x)^{-C} \int_{\mathbb{R}^2} (1 + s(\alpha_1, \alpha_2)x)^{-\frac{3}{2}} \, d\alpha_1 \, d\alpha_2
$$

$$
= (\log x)^{-C} \int_{\mathbb{R}^2} (1 + \alpha_1^2 + \alpha_2^2)^{-\frac{3}{2}} \, d\alpha_1 \, d\alpha_2 \ll (\log x)^{-C}.
$$

Therefore, we lost essentially nothing by the separation of the variables $\ell_1, \ell_2$ through the Fourier transform (10.8). We get

$$
(10.13) \quad f(m_2n)f(m_1n) = I \int_{S} F(s(\alpha_1, \alpha_2))e(\alpha_1 \ell_1 + \alpha_2 \ell_2) \, d\alpha_1 \, d\alpha_2 + O((\log x)^{-C}).
$$

### 11 Estimation of $R''_h(m_1, m_2)$

Recall that the error term $R_h(m_1, m_2)$ is given by (9.17). Introducing (10.13) into (9.17) we get

$$
(11.1) \quad R_h(m_1, m_2) = R'_h(m_1, m_2) + R''_h(m_1, m_2)
$$

where

$$
(11.2) \quad R''_h(m_1, m_2) \ll \frac{x}{\varphi(\Delta h)}(\log x)^{-C},
$$

$$
(11.3) \quad R'_h(m_1, m_2) = I \int_{S} F(s(\alpha_1, \alpha_2))H(\alpha_1, \alpha_2) \, d\alpha_1 \, d\alpha_2
$$

and

$$
(11.4) \quad H(\alpha_1, \alpha_2) = \sum_{\ell_1, \ell_2 \equiv \omega_\ell (2\Delta h)} \gamma_{\ell_1, \ell_2} e(\alpha_1 \ell_1 + \alpha_2 \ell_2) - \frac{1}{\varphi(2\Delta h)} \sum_{\ell_1} \sum_{\ell_2} \gamma_{\ell_1, \ell_2} e(\alpha_1 \ell_1 + \alpha_2 \ell_2).
$$

The total contribution of $R''_h(m_1, m_2)$ to $C_h(M)$ is (see (7.12) and (9.10))

$$
\sum_{(m_1, m_2, 2h) = 1} \sum_{0 \neq \Delta(m_1, m_2) \equiv 0(2h)} \mu(m_1)\mu(m_2)R''_h(m_1, m_2).
$$

The determinant $\Delta = \Delta(m_1, m_2)$ occurs with a multiplicity which is bounded by $8M$ so the above contribution is bounded by

$$
Mx(\log x)^{-C} \sum_{1 \leq \Delta < 4M} 1/\varphi(\Delta h) \ll h^{-2}Mx(\log x)^2-C.
$$

Inserting this bound into (7.9) we find that the total contribution of $R''_h(m_1, m_2)$ to $C(M)$, say $C''(M)$, satisfies

$$
(11.5) \quad C''(M) \ll Mx(\log x)^{3-C}.
$$

This bound satisfies our requirement (7.11) if we take $C$ to be a sufficiently large constant, specifically $C \geq 8A + 30$. 

22
12 Small determinant

The estimation $\mathcal{R}_h'(m_1, m_2)$ is quite delicate because the determinant $\Delta = \Delta(m_1, m_2) = \text{Im} m_1 m_2$ can be small, in which case the separation of the variables $\ell_1, \ell_2$ by means of the Fourier transform (see (11.3) and (11.4)) cannot be treated in a straightforward fashion. The set $S$ has relatively large measure, see (10.12), and there is a lot of room for $\alpha_1$. Recall that $s(\alpha_1, \alpha_2)$ is the quadratic form in $\alpha_1, \alpha_2$ and

$$s(\alpha_1, \alpha_2) = (\alpha_1 I)^2 + (\alpha_2 + \alpha_1 R)^2 \lesssim (\eta I)^2$$

with

$$I = \text{Im}(m_1/m_2) = -\Delta/m_2 \asymp |\Delta|/M, \quad R = \text{Re}(m_1/m_2) \asymp 1$$

and

$$\eta = (\log x)^C/\sqrt{x}|I|, \quad |\alpha_1| \leq \eta, \quad |\alpha_2 + \alpha_1 R| < \eta|I|.$$  

We detect the congruence $\ell_2 \equiv \omega \ell_1 \pmod{2\Delta h}$ in (11.4) by means of Dirichlet characters $\chi (\text{mod } 2\Delta h)$ getting

$$|H(\alpha_1, \alpha_2)| \leq \frac{1}{\varphi(2\Delta h)} \sum_{x \neq x_0} \left| \sum_{\ell} \gamma_\ell \chi(\ell) e(\alpha_1 \ell) \right| \left| \sum_{\ell} \gamma_\ell \chi(\ell) e(-\alpha_2 \ell) \right|.$$  

Hence, by the Cauchy-Schwarz inequality

$$\mathcal{R}_h'(m_1, m_2) = I \iint_S F H \ll |I| x \iint_S |H|$$

$$\leq |I| \frac{x}{\varphi(2\Delta h)} \left( \sum_x \iint_S \left| \sum_{\ell} \gamma_\ell \chi(\ell) e(\alpha_1 \ell) \right|^2 \right)^{\frac{1}{2}} \left( \sum_x \iint_S \left| \sum_{\ell} \gamma_\ell \chi(\ell) e(-\alpha_2 \ell) \right|^2 \right)^{\frac{1}{2}}.$$  

Note that we have included $\chi = \chi_0$. From the first sum of integrals we get

$$\sum_x \iint_S \left| \sum_{\ell} \right|^2 \leq 2\eta |I| \sum_x \int_{|\alpha| < \eta} \left| \sum_{\ell} \gamma_\ell \chi(\ell) e(\alpha \ell) \right|^2 d\alpha.$$  

Now we enlarge the integral by introducing a majorant weight function $w(\alpha)$ whose Fourier transform $\hat{w}(v)$ has compact support. For this job we choose

$$w(\alpha) = 4 \left( \frac{\sin \pi \alpha/2\eta}{\pi \alpha/2\eta} \right)^2, \quad \hat{w}(v) = 8\eta \max(1 - 2\eta|v|, 0).$$  

We get the bound

$$16\eta^2 |I| \varphi(2\Delta h) \sum_{\ell_1 \equiv \ell_2(2\Delta h)} \sum_{|\ell_1 - \ell_2| < 1/2\eta} |\gamma_1 \gamma_2| \ll \eta |I| \sqrt{x} = (\log x)^C.$$  

From the second sum of integrals we get

$$\sum_x \iint_S \left| \sum_{\ell} \right|^2 \leq |S| \varphi(2\Delta h) \sum_{\ell_1 \equiv \ell_2(2\Delta h)} \sum_{|\ell_1 - \ell_2| < 1/2\eta} |\gamma_1 \gamma_2| \ll (\log x)^{2C}/|I|.$$  

23
Multiplying both estimates we conclude that

\[(12.3) \quad \mathcal{R}_h'(m_1, m_2) \ll \frac{|I|\frac{1}{2}}{\phi(\Delta h)} x \log(x)^{3C/2}.
\]

This bound is better than (11.2) for \(\mathcal{R}_h''(m_1, m_2)\) if

\[(12.4) \quad |I| \leq (\log x)^{-5C}.
\]

Therefore we are done in this case.

13 Estimation of \(\mathcal{R}_h'(m_1, m_2)\) on average

In most cases

\[(13.1) \quad I = \text{Im} \frac{m_1}{m_2} = -\frac{\Delta}{m_2} \geq \frac{\Delta}{M}
\]

is not smaller than (12.3). Assuming \(I\) does not satisfy (12.3) we give a better treatment of \(\mathcal{R}_h'(m_1, m_2)\) using the Siegel-Walfisz condition and the large sieve inequality.

We begin by removing the twists by additive characters from the multiplicative character sum (12.2). To this end we apply partial summation losing factors \(1 + 2\pi |\alpha_1| \sqrt{x}\) and \(1 + 2\pi |\alpha_2| \sqrt{x}\).

Specifically, we apply the expression

\[(13.2) \quad e(\alpha \ell) = 1 + 2\pi i \int_0^\ell e(\alpha t) \, dt
\]

to the sums over \(\ell\) in (12.2) getting

\[|H(\alpha_1, \alpha_2)| \leq (1 + 2\pi |\alpha_1| \sqrt{x})(1 + 2\pi |\alpha_2| \sqrt{x}) G(t_1, t_2)
\]

with

\[G(t_1, t_2) = \frac{1}{\phi(\Delta h)} \sum_{\chi \neq \chi_0} \left| \sum_{t_1 < \ell < \sqrt{x}} \gamma_\ell \chi(\ell) \right| \left| \sum_{t_2 < \ell < \sqrt{x}} \gamma_\ell \chi(\ell) \right|
\]

for some \(0 < t_1, t_2 < \sqrt{x}\). The loss is not large because, for \((\alpha_1, \alpha_2)\) in \(S\),

\[(1 + 2\pi |\alpha_1| \sqrt{x})(1 + 2\pi |\alpha_2| \sqrt{x}) \ll I^{-2}(\log x)^{3C} \leq (\log x)^{12C}.
\]

Integrating this over \(S\) against \(F(s(\alpha_1, \alpha_2)) \ll x\) we conclude by (11.3) that

\[(13.3) \quad \mathcal{R}_h'(m_1, m_2) \ll G(t_1, t_2)(\log x)^{14C}.
\]

Remarks 13.1. The cropping parameters \(t_1, t_2\) come from integration in the expression (13.2). We could carry such integration to the very end of our arguments and only then choose the worst values \(t_1, t_2\) which are independent of the preceding variables \(m_1, m_2, h\). To simplify the presentation we accept (13.3) having \(t_1, t_2\) independent of \(m_1, m_2, h\). By \(2G(t_1, t_2) \leq G(t_1, t_1) + G(t_2, t_2)\) we arrive at

\[(13.4) \quad \mathcal{R}_h'(m_1, m_2) \ll (\log x)^{14C} \frac{1}{\phi(2\Delta h)} \sum_{\chi \neq \chi_0} |\mathcal{L}(\chi)|^2
\]
with
\begin{equation}
\mathcal{L}(\chi) = \sum_{t < t < \sqrt{x}} \chi(t) \gamma_t
\end{equation}
for \( t = t_1 \) or \( t = t_2 \).

We need to sum \( \mathcal{R}'_h(m_1, m_2) \) over \( m_1, m_2 \) as in (7.12) and over \( h \) as in (7.9) subject to the condition \( \Delta = \Delta(m_1, m_2) \equiv 0 \pmod{2h} \), see (9.10). The total contribution of \( \mathcal{R}'_h(m_1, m_2) \) to \( \mathcal{C}(M) \) is bounded by \( \mathcal{R}(M)(\log x)^{14C} \) where

\[
\mathcal{R}(M) = \sum_{h} |\lambda_h|h \sum_{(m_1, m_2) = 1} \frac{|\mu(m_1)\mu(m_2)|}{\varphi(2\Delta h)} \sum_{\chi \pmod{2\Delta h}} |\mathcal{L}(\chi)|^2.
\]

Recall that \( m_1 \sim M, m_2 \sim M \) and \( m_1, m_2 \) are primitive. The determinant \( \Delta \) occurs with certain multiplicity which is bounded by \( 8M \), so

\[
\mathcal{R}(M) \ll M \sum_{h < y} \frac{|\lambda_h|}{\varphi(h)} \sum_{h < 4M} \frac{1}{\varphi(r)} \sum_{\chi \pmod{rh^2}} |\mathcal{L}(\chi)|^2.
\]

Each character \( \chi \neq \chi_0 \) is induced by a unique primitive character \( \chi_1 \pmod{q} \) with \( q \neq 1, q|r^2 \) and \( \chi(t) = \chi_1(t) \) for primes \( t > rh^2 \). Hence

\[
\mathcal{R}(M) \ll M \sum_{1 < q \leq Q} c(q) \sum_{\chi_1 \pmod{q}} |\mathcal{L}(\chi_1)|^2
\]

where \( Q = 8My \) and

\[
c(q) \ll \sum_{h < y} \frac{|\lambda_h|}{\varphi(h)} \sum_{r \leq 8M} \frac{1}{\varphi(r)} \ll \log M \sum_{h < y} \frac{|\lambda_h|}{\varphi(h)} (q, h^2)
\]

\[
\ll \tau(q)^2 q^{-1} \min(\sqrt{q}, y)(\log M)^2.
\]

Hence

\[
\mathcal{R}(M) \ll M(\log M)^2 \sum_{1 < q \leq Q} q^{-1} \min(\sqrt{q}, y) \sum_{\chi_1 \pmod{q}} |\mathcal{L}(\chi_1)|^2.
\]

Using the S-W condition for small \( q \) and the large sieve inequality for larger \( q \) we get

\begin{equation}
\mathcal{R}(M) \ll Mx(\log x)^{-B}
\end{equation}

with any \( B \geq 2 \), provided \( Q \min(\sqrt{Q}, y) < x^{\frac{1}{2} - \varepsilon} \). Hence (13.6) holds if

\begin{equation}
yz \leq x^{\frac{1}{2} - \varepsilon}.
\end{equation}

Finally, the total contribution of the error terms \( \mathcal{R}_h(m_1, m_2) \) to \( \mathcal{C}(M) \) is bounded by

\begin{equation}
\mathcal{R}(M)(\log x)^{14C} \ll Mx(\log x)^{14C - B}.
\end{equation}

This bound satisfies our requirement (7.11) if we take \( B \) large.

Every bound obtained so far satisfies our requirements subject to the conditions (6.7) and (13.7). It remains to estimate the contribution of the main terms \( \mathcal{E}_h(m_1, m_2) \) to \( \mathcal{C}_h(M) \) on average over \( h \), see (9.10), (7.9), (7.11). It turns out that the main term is a harder piece than the error terms!
14 Preparation of the main terms

Recall that the main terms $E_h(m_1, m_2)$ are defined by (9.16) and we need to estimate the sums

\begin{equation}
F_h(M) = h \sum_{(m_1, m_2, 2h) = 1 \atop 0 \neq \Delta(m_1, m_2) \equiv 0(2h)} \mu(m_1) \mu(m_2) E_h(m_1, m_2),
\end{equation}

\begin{equation}
\mathcal{F}(M) = \sum_{h \leq \mu} |\lambda_h| F_h(M).
\end{equation}

Our goal is to show that

\begin{equation}
\mathcal{F}(M) \ll NM^2 (\log M)^{-B}
\end{equation}

with any $B \geq 2$, which bound is fine for the requirement (7.11).

In this section we make preparations for the application of tools in the next two sections. First it helps to execute the summation over $\ell_1, \ell_2$ in (9.16). To this end we exploit our assumption (1.8) for $q = 1$, that is the PNT for the $\gamma_\ell$’s.

Let us check that the restrictions (10.9) are redundant. Indeed, from the support of $f$ and $n$ given by (9.15) we get

\begin{equation}
m_2n = \ell_2^2 + \left( \frac{\ell_1 - \ell_2 R}{I} \right)^2 < x,
\end{equation}

hence $\ell_2 < \sqrt{x}$. Interchanging $\ell_1, \ell_2$ and $m_1, m_2$ we get a similar formula for $m_1n$, hence $\ell_1 < \sqrt{x}$.

We show that the partial derivatives of $f(x_1, x_2)$ defined by (10.1) satisfy

\begin{equation}
x_1 \frac{\partial}{\partial x_1} f(x_1, x_2) \ll 1, \quad x_2 \frac{\partial}{\partial x_2} f(x_1, x_2) \ll 1.
\end{equation}

To this end, we compute as follows:

\[
\frac{\partial}{\partial x_1} f(m_2n(x_1, x_2)) = \frac{\partial}{\partial x_1} f \left( \frac{x_2^2 + \left( \frac{x_1 - x_2 R}{I} \right)^2}{x_2^2 + \left( \frac{x_1 - x_2 R}{I} \right)^2} \right) = \frac{2 x_1 - x_2 R}{I} f' \left( \frac{x_2^2 + \left( \frac{x_1 - x_2 R}{I} \right)^2}{x_2^2 + \left( \frac{x_1 - x_2 R}{I} \right)^2} \right) \ll \sqrt{x} x^{-1}.
\]

Hence

\[
x_1 \frac{\partial}{\partial x_1} f(m_2n(x_1, x_2)) \ll \frac{x_1}{\sqrt{x}} \ll 1.
\]

Similarly for $f(m_1n(x_1, x_2))$ and for the partial derivatives with respect to $x_2$. Hence, (14.4) holds.

Using the Prime Number Theorem by partial summation (9.16) yields

\[
\varphi(2\Delta h) E_h(m_1, m_2) = \int \int f(x_1, x_2) \, dx_1 \, dx_2 + O \left( x (\log x)^{-B} \right)
\]

26
with any $B \geq 2$. Here the integral is just the Fourier transform $g(\alpha_1, \alpha_2)$ at $(\alpha_1, \alpha_2) = (0, 0)$, see (10.2). Then (10.7) and (10.6) yield
\[
g(0, 0) = \mathcal{F}(s(0, 0)) = \mathcal{F}(0),
\]
\[
F(0) = \pi \int_0^\infty J_0(t)f(t)\mathcal{F}(tm_1/m_2) dt = \pi m_2 \int_0^\infty f(tm_2)\mathcal{F}(tm_1) dt,
\]
\[
m_2 I = m_2 \text{Im} \frac{m_1}{m_2} = \text{Im} \frac{m_1}{m_2} = -\Delta.
\]
Combining these results we obtain
\[
(14.5) \quad \mathcal{E}_h(m_1, m_2) = -\frac{\pi \Delta}{\varphi(2\Delta h)} \int_0^\infty f(tm_2)\mathcal{F}(tm_1) dt + O\left(\frac{x}{\varphi(\Delta)(\log x)^B}\right),
\]
Inserting (14.3) into (14.1) we get (note that $\varphi(2\Delta h) = 2h\varphi(\Delta)$)
\[
(14.6) \quad \mathcal{F}_h(M) = -\frac{\pi}{2} \int_0^\infty \mathcal{K}_h(t) dt + O\left(\frac{x}{(\log x)^B} \sum \sum |\mu(m_1)\mu(m_2)|/\varphi(\Delta)\right)
\]
where
\[
(14.7) \quad \mathcal{K}_h(t) = \sum \sum \mu(m_1)\mu(m_2)f(tm_1)\mathcal{F}(tm_2)\Delta/\varphi(\Delta).
\]
The error term in (14.6) on average over $h \leq y$ satisfies the bound (14.3) so we are done with it. The integral in (14.6) is over the segment $N/4 < t < N$ so we need to show that
\[
(14.8) \quad \mathcal{K}(t) = \sum_{h \leq y} |\lambda_h \mathcal{K}_h(t)| \ll M^2(\log M)^{-B}
\]
for any $N/4 < t < N$ (recall $MN = x$) and any $B \geq 3$. Writing
\[
\frac{\Delta}{\varphi(\Delta)} = \prod_{p | \Delta} \left(1 - \frac{1}{p}\right)^{-1} = \frac{2h}{\varphi(2h)} \sum_{d | \Delta} \frac{\mu^2(d)}{\varphi(d)}
\]
we get
\[
(14.9) \quad \mathcal{K}_h(t) = \frac{2h}{\varphi(2h)} \sum_{(d, 2h) = 1} \frac{\mu^2(d)}{\varphi(d)} \sum \sum \mu(m_1)\mu(m_2)f(tm_1)\mathcal{F}(tm_2).
\]
The inner sum over $m_1, m_2$ is bounded by $(8M)^2/dh$. Hence the contribution of $d > D$ is $\ll M^2/D\varphi(h)$. Summing over $h \leq y$, this does not exceed the bound (14.3), unless
\[
(14.10) \quad d \leq (\log M)^{B+1}.
\]
Assuming \((14.10)\), we can drop the restriction \(\Delta(m_1, m_2) \neq 0\). If \(\Delta(m_1, m_2) = 0\), then \(m_1 = m_2\), so these added terms contribute to \((14.9)\) at most \(O(M \log M)\) and to \((14.8)\) at most \(O(y M \log M)\) which is admissible if
\[
y \leq M (\log M)^{-B-1}.
\]

Writing \(m_1 = u_1 + iv_1\) and \(m_2 = u_2 + iv_2\) the congruence \(\Delta(m_1, m_2) \equiv 0(2dh)\) means \(u_1v_2 \equiv u_2v_1 \pmod{2dh}\). Hence \((2dh, v_1) = (2dh, v_2) = b\), say, because \((u_1, v_1) = (u_2, v_2) = 1\). Put \(2dh = bc\), \(v_1 = bw_1\), \(v_2 = bw_2\) so \((w_1 w_2, c) = 1\) and the congruence become \(u_1 w_2 = u_2 w_1 \pmod{c}\), or equivalently
\[
u_1 \overline{w}_1 = u_2 \overline{w}_2 \pmod{c}
\]
where \(\overline{w}\) \((\text{mod } c)\) denotes the multiplicative inverse (not the complex conjugate). Hence \((14.9)\) becomes (up to an admissible error term)
\[
K_h(t) = \frac{2h}{\varphi(2h)} \sum_{(d,2h)=1} \sum_{2dh=bc} \frac{\mu^2(d)}{\varphi(d)} \sum_{(m_1, m_2, 2h)=1} \sum_{(w_1, w_2, c)=1} \mu(m_1) \mu(m_2) f(tm_1) \overline{f}(tm_2)
\]
where \(m_1 = u_1 + bw_1\) and \(m_2 = u_2 + bw_2\). The inner sum over \(m_1, m_2\) is bounded by \(O(M^2/b^2c)\). Hence the contribution of \(b > b_0\) is bounded by \(O(\tau(h) M^2 / \varphi(h) b_0)\) which is negligible for \(b_0 = L B^2\).

From now on we assume that
\[
b \leq (\log M)^{B+2}.
\]

The condition \((m_1, m_2, 2h) = 1\) in the inner sum of \((14.13)\) is equivalent to \((m_1, m_2, c/(c, d)) = 1\).

This is a harmless, but inconvenient condition. We are going to remove it by a cute trick. Let \(T^*\) denote the sum over \(m_1, m_2\) with the condition \((m_1, m_2, c/(c, d)) = 1\) and \(T\) the sum without this condition. We show that
\[
0 \leq T^* \leq T.
\]

**Proof.** Recall that the congruence \(u_1 \overline{w}_1 = u_2 \overline{w}_2 \pmod{c}\) implies \(m_1 \overline{m}_2 = \overline{m}_1 m_2 \pmod{c}\). Hence the condition \((m_1, m_2, c/(c, d)) = 1\) is equivalent to \((m_2, c/(c, d)) = 1\), because \(m_1, m_2\) are odd primitive. Hence
\[
T^* = \sum_{(m_1, m_2, c/(c, d))=1} \sum_{m_1} \sum_{m_2 \equiv u_2 + bw_2 \pmod{c}, (w_2, c) = (m_2, c/(c, d)) = 1} \mu(m_1) f(tm_1) e \left( \frac{a}{c} u_1 \overline{w}_1 \right)
\]
\[
\times \left( \sum_{m_2 = u_2 + bw_2 \pmod{c}, (w_2, c) = (m_2, c/(c, d)) = 1} \mu(m_2) \overline{f}(tm_2) e \left( \frac{a}{c} u_2 \overline{w}_2 \right) \right).
\]

By Cauchy’s inequality \(T^* \leq T^* \sqrt{T^{*2}}\), hence \((14.13)\) holds. 

By the above considerations we derive the following inequality

\begin{equation}
K_h(t) \leq \frac{2h}{\varphi(2h)} \sum_{d,2h=bc} \varphi(c) T(b, c)
\end{equation}

where

\begin{equation}
T(b, c) = \frac{1}{\varphi(c)} \sum_{a \pmod{c}} \left| \sum_{m=\frac{u+ibw}{(w,c)=1}} \mu(m)f(tm)e\left(\frac{a u \overline{m}}{c}\right) \right|^2
\end{equation}

and the sums \( \sum' \sum' \) are restricted by the conditions (14.10) and (14.14). Moreover we dropped out of (14.16) a few parts which we already showed to be admissible for the goal (14.8). Here we have \( 2h\varphi(c)/\varphi(2h)\varphi(d) = b\varphi(c)/d\varphi(bc) \leq b/d\varphi(b) \) so

\begin{equation}
K_h(t) \leq \sum' \sum' \frac{b}{\varphi(b)} \left( \sum_{2dh=bc} d^{-1} \right) T(b, c)
\end{equation}

where

\begin{equation}
Q = 2y(\log M)^{B+1}.
\end{equation}

Hence

\begin{equation}
K(t) \leq (\log M) \sum_{bc \leq Q} \sum' \frac{b}{\varphi(b)} T(b, c).
\end{equation}

Writing \( a/c \) in the lowest terms we get

\begin{equation}
K(t) \leq (\log M) \sum_{bq \leq Q} \sum' \frac{b}{\varphi(b)\varphi(r)} T(b, q, r)
\end{equation}

where

\begin{equation}
T(b, q, r) = \frac{1}{\varphi(q)} \sum_{a \pmod{q}} \left| \sum_{m=\frac{u+ibw}{(w,qr)=1}} \mu(m)f(tm)e\left(\frac{a u \overline{m}}{q}\right) \right|^2.
\end{equation}

15 Small moduli

We can estimate the sum over \( m = u + ibw \) in (14.22) using the Siegel-Walfisz type theorem in the Gaussian domain. See Lemma 5 of [Fog] or Lemma 16.1 of [FrIw1] and the references therein.
Lemma 15.1. Let \( \ell \geq 1 \) and \( \omega \in \mathbb{Z}[i] \). For \( x \geq 2 \) we have
\[
(15.1) \quad \sum_{\substack{m \equiv \omega \ (\text{mod } \ell) \cr m \leq x}} \mu(m) \ll x(\log x)^{-B_1}
\]
with any \( B_1 \geq 1 \), the implied constant depending only on \( B_1 \).

Remark 15.1. The bound (15.1) is trivial (it has no value) if \( \ell > (\log x)^B \).

We relax the condition \((w, r) = 1\) by M"obius formula and apply (15.1) as follows
\[
\left| \sum_{m = u + bw \atop (w, qr) = 1} \mu(m)f(tm)e \left( \frac{a}{q} \frac{um}{w} \right) \right| \leq \sum_{k \mid r, k \leq K \atop (w, q) = 1} \left| \sum_{m = u + bw \atop (w, qr) = 1} \mu(m)f(tm)e \left( \frac{a}{q} \frac{um}{w} \right) \right| + O \left( \frac{\tau(r)M}{bK} \right)
\]
\[
\leq \sum_{k \mid r, k \leq K} \sum_{a \equiv \alpha (\text{mod } bkq) \atop \beta \equiv 0 (bk)} \left| \sum_{m \equiv \alpha + i\beta \ (\text{mod } bkq)} \mu(m)f(tm) \right| + O \left( \frac{\tau(r)M}{bK} \right)
\]
\[
\ll \tau(r)bKq^2M(\log M)^{-8B} + \tau(r)M/bK = 2\tau(r)qM(\log M)^{-4B}
\]
for \( K \) with \( bKq = (\log M)^{4B} \). Hence
\[
(15.2) \quad T(b, q, r) \ll (\tau(r)qM)^2(\log M)^{-8B},
\]
and the partial sum of (14.21) with \( q \leq Q_0 \), say \( K(q \leq Q_0) \), satisfies
\[
(15.3) \quad K(q \leq Q_0) \ll Q_0^3M^2(\log M)^{7-7B}.
\]
This bound satisfies (14.8) if
\[
(15.4) \quad Q_0 = (\log M)^{2B-3}.
\]

16 Large moduli

It remains to estimate the partial sums of (14.21) with \( Q_1 < q \leq 2Q_1 \), say \( K(q \sim Q_1) \), for \( Q_0 \leq Q_1 \leq Q/2 \), i.e.
\[
(16.1) \quad (\log M)^{2B-3} \leq Q_1 \leq y(\log M)^{B+1}.
\]

In this range we no longer need help from the M"obius function \( \mu(m) \); the cancellation is due to the variation of \( e \left( \frac{a}{q} \frac{um}{w} \right) \). We need saving a bit larger than the size of the conductor \( q \) so the saving from averaging over the classes \( a \ (\text{mod } q) \) (making the Ramanujan sum) is not enough. But even a little extra averaging extracted from \( q \) would do the job by means of the large sieve inequality. However,
we do not have any multiplicative structure of $q$ from which to borrow a little extra averaging so we throw the whole range $q \sim Q_1$ into the game.

For $m = u + ibw$ with $m \sim M$ the first coordinate $u$ runs over the segment $|u| \leq \sqrt{2M}$ which is sufficiently long for exploiting the large sieve inequality effectively. Because we do not need help from the second coordinate $v = bw$, $(w, qr) = 1$, we can simplify the matter by estimating (14.22) as follows

$$T(b, q, r) \leq \frac{4\sqrt{M}}{bq} \sum_{|u| < \sqrt{2M}b^{-1}} \sum_{a(q)}^* \sum_{m = u + ibw} \mu(m) f(tm)e\left(\frac{au}{q}\right).$$

Summing over $q \sim Q_1$ we get by the large sieve inequality

$$(16.2) \quad \sum_{q \sim Q_1} \frac{1}{q} \sum_{a(q)}^* \left| \sum_m \right|^2 \ll (Q_1 + \sqrt{M/Q_1})\sqrt{M} \leq 2M/Q_0$$

provided $QQ_0 \leq 2\sqrt{M}$, i.e.

$$(16.3) \quad y \leq \sqrt{M} (\log M)^{-3B}.$$ 

Recall that $M$ satisfies (15.7). Hence

$$\sum_{q \sim Q_1} T(b, q, r) \ll M^2/bQ_0$$

and

$$(16.4) \quad \sum_{q \sim Q_1} K(q \sim Q_1) \ll Q_0^{-1}(M \log M)^2 = M^2(\log M)^{5-2B}.$$ 

This is sufficient for (14.8), if $B \geq 3$.

**17 Proof of SMT. Conclusion**

Putting together the results of Sections 6–16 we complete the proof of (6.10) and of SMT (see (6.8)) under the following conditions:

$$y^2(\log x)^{2A+4} \leq z \leq x^{\frac{4}{3}}, \text{ see (6.7)},$$

$$y(\log x)^{11A+40} \leq z, \text{ see (8.4)},$$

$$yz \leq x^{\frac{4}{3} - \varepsilon}, \text{ see (13.7)},$$

$$y < z^{\frac{1}{3}}(\log x)^{-3B-2A}, \text{ see (16.3)}.$$ 

The choice $z = x^{\frac{4}{3}}$ and $y = x^\theta$ with any $\theta < \frac{1}{12}$ is good. This completes the proof of (1.16).
18 Derivation of MT

It is not hard to derive MT from SMT simply by subdividing the range $1 \leq t \leq x$ into dyadic segments

$$ T < t \leq 2T, \quad T = 2^{-a} x, \quad a = 1, 2, \ldots $$

and smoothing at the end points over two short intervals

$$ T < t < T(1 + \delta), \quad 2T(1 - \delta) < t < 2T. $$

The total contribution of $n$’s in the short intervals is estimated trivially by $O(\delta x (\log x)^4)$ which is absorbed by the error term in (1.10) if

$$ \delta = (\log x)^{-A - 4}. $$

The resulting smooth function $f(t)$ supported in a given dyadic segment is

$$ f(t) = 1, \text{ except for } t \text{ in the short intervals adjacent to the end points where } t \approx \delta - j. $$

Because we require only $j = 0, 1, 2$, the condition (1.14) can be secured by resizing $f(t)$ by a factor $\delta^2$. This factor does not ruin (1.16), because we can use (1.16) with $A$ replaced by $3A + 8$.

19 Derivation of APT

We derive the Almost Primes Theorem from the Main Theorem by applying the Almost-Prime Sieve from Chapter 25 of [FrIw2] to the sequence $C = (c_k), 1 \leq k \leq K = \sqrt{x}$, with

$$ c_k = \sum_{4k^2 + \ell^2 \leq x} \Lambda(\ell)\Lambda(4k^2 + \ell^2). $$

We have

$$ \sum_k c_k = X + O(x (\log x)^{-A}) $$

with $X = \kappa x$. For any $1 \leq h \leq y$, $h$ squarefree, we set the error terms

$$ r_h = \sum_{k \equiv 0 \pmod{h}} c_k - g(h)X $$

and we derive by (1.10) with some $\lambda_h = \pm 1$ that

$$ \sum_{h \leq y} |r_h| = \sum_{h \leq y} \lambda_h r_h $$

$$ = \sum_{4k^2 + \ell^2 \leq x} \beta_k \Lambda(\ell)\Lambda(4k^2 + \ell^2) - X \sum_{h \leq y} \lambda_h g(h) \ll x (\log x)^{-A}. $$

In other words, speaking the language of sieve theory, our sequence $C = (c_k)$ has the absolute level of distribution $y$ and the density function $g(h)$ satisfies the linear sieve condition (5.38) of [FrIw2]. Therefore, Theorem 25.1 of [FrIw2] is applicable giving

$$ \sum_{(k,P(z)) = 1 \atop \nu(k) \leq r} c_k \asymp x (\log x)^{-1}, $$

32
with \( z = y^\frac{1}{4} \), subject to the condition (25.25) of [FrIw2]. In our situation this condition reads \( y > K^{\varepsilon + 1/\Delta_r} \), that is \( \Delta_r > 1/2\theta \). Since \( \Delta_r > r + 1 - \log 4/\log 3 \) (see (25.24) of [FrIw2]) and \( \theta \) is any number \( < \frac{1}{12} \), we are fine with \( r = 7 \). This completes the proof of the Almost Primes Theorem and hence of Theorem \( G_7 \).

Appendix

We now give a proof of Proposition 2.1. As will be seen, the argument uses nothing of what has gone before and is much simpler than the main theorems of the paper.

Proof. We are going to apply the sieve to study the sequence \( A = (a_n) \), with

\[
 a_n = \sum_{4k^2 + \ell^2 = n} \Lambda(k)\Lambda(\ell).
\]

Note that, for notational convenience, we restrict \( k, \ell \), rather than \( 4k^2 + \ell^2 \) and we use \( x \) rather than \( \sqrt{x} \). If \( d \) is odd we have

\[
 A_d = \sum_{n \equiv 0 (\text{mod } d)} a_n \\
 = \sum_{\nu^2 + 1 \equiv 0 (d)} \sum_{\ell \equiv 2\nu k (d)} \Lambda(k)\Lambda(\ell) \\
 = \sum_{\nu^2 + 1 \equiv 0 (d)} \sum_{a(d)}^* \psi(x; d, 2\nu a) \psi(x; d, a) + O((\log x)^6)
\]

where, as we recall, \( E(x; d, a) \) is the error term in the prime number theorem for that arithmetic progression and \( \rho(d) \) is the number of roots of \( \nu^2 + 1 \equiv 0 (\text{mod } d) \). Put

\[
 r_d = A_d - \frac{\rho(d)}{\varphi(d)} \psi(x)^2.
\]

Then

\[
 |r_d| \leq \rho(d) \sum_{a(d)}^* |E(x; d, a)|^2 + O((\log x)^6)
\]

\[
 \leq \frac{\rho(d)}{\varphi(d)} \sum_{a(d)}^* |E(x; d, a)| + (\log x)^6.
\]

33
Hence, the remainder of level $D$ is estimated as follows:

$$R(D) = \sum_{d \leq D} |r_d| \ll x(\log x) \left( \sum_{d \leq D} \sum_{\alpha(d)}^\ast |E(x,d,\alpha)|^2 \right)^{1/2} + D(\log x)^6 \ll x^2(\log x)^{-A}$$

by the Barban-Davenport-Halberstam Theorem (see (9.75) of [FrIw2]), where $A$ is any positive number and $D = x^2(\log x)^{-B}$ with some $B = B(A)$. Therefore the sequence $A = (a_n)$ is supported on $n \leq N = 5x^2$, it satisfies the linear sieve conditions and it has level of distribution $D \asymp N^{1/2}(\log N)^{-B}$. Now, just about any sieve, such as for example Theorem 6.9 of [FrIw2], gives the upper bound claimed in the proposition. Since also $\Delta_3 > 4 - \log 4/\log 3 > 2$, it follows from Theorem 25.1 of [FrIw2] that the lower bound in the proposition holds and specifically

$$\left( A1 \right) \sum_{\omega(n) \leq 3} a_n \asymp x^2(\log x)^{-1},$$

which implies the proposition.

We conclude the paper with heuristics supporting the formula (1.12). If $r \geq 2$ we use Bombieri’s sieve in Theorem 3.5 of [FrIw2] showing that (1.2) holds with the constant

$$\left( A2 \right) c = \kappa \prod_p (1 - g(p)) \left( 1 - \frac{1}{p} \right)^{-1} = \kappa \prod_{p \equiv 1(4)} \left( 1 - \frac{1}{p-2} \right) \left( 1 - \frac{1}{p} \right)^{-1}.$$

Recall that $\kappa$ is given by (1.12), hence $c$ is given by (1.3).

Of course, this result is conditional subject to the assumption that the sequence $C = (c_k)$ has exponent of distribution as large as 1, meaning (1.10) holds for $y = x^\theta$ with any $\theta < \frac{1}{2}$.

If $r = 1$, we write

$$\Lambda(k) = \sum_{h|k} \lambda_h, \quad \lambda_h = -\mu(h) \log h,$$

and apply (1.10). For $r = 1$ Bombieri’s sieve gives no help so we simply ignore that (1.10) is applicable unconditionally only for $h < y$, because we believe that for larger $h$ the Möbius function does not correlate with anything “different” on its way. We arrive at (GPC) with the constant

$$\kappa \sum_h \mu(h)(- \log h) g(h) = c.$$

### References

[Fog] E. Fogels, On the distribution of prime ideals, *Acta Arith.* 7 (1961/62), 255–264.

[FoIw] E. Fouvry and H. Iwaniec, Gaussian primes, *Acta Arith.* 79 (1997), 249–287.

[FrIw1] J.B. Friedlander and H. Iwaniec, The polynomial $X^2 + Y^4$ captures its primes, *Ann. of Math.* 148 (1998), 945–1040.
[FrIw2] J.B. Friedlander and H. Iwaniec, Opera de Cribro, Amer. Math. Soc. Colloq. Pub. 57 AMS (Providence), 2010.

[FrIw3] J.B. Friedlander and H. Iwaniec, Sums over determinants, (preprint)

[HL] D.R. Heath-Brown and X. Li, Prime values of $a^2 + p^4$, Invent. Math. 208 (2017), 441–499.

[LSX] P. C-H. Lam, D. Schindler and S.Y. Xiao, On prime values of binary quadratic forms with a thin variable, (preprint).

[Pr] K. Pratt, Primes from sums of two squares and missing digits, (preprint).

Department of Mathematics, University of Toronto
Toronto, Ontario M5S 2E4, Canada

Department of Mathematics, Rutgers University
Piscataway, NJ 08903, USA