J. Eur. Math. Soc. 24, 737–772 (2022) DOI 10.4171/JEMS/1083

John Friedlander · Henryk Iwaniec

Coordinate distribution of Gaussian primes

Received November 17, 2018; revised July 6, 2020

Abstract. We study the problem of writing Gaussian primes as the sum of two squares, both of which are interesting arithmetically, in particular, when one is the square of a prime and the other the square of an almost-prime.

Keywords. Primes, squares, bilinear forms

Contents

John Friedlander: Department of Mathematics, University of Toronto, Toronto ON M5S 2E4, Canada; frdlndr@math.toronto.edu

Henryk Iwaniec: Department of Mathematics, Rutgers University, Piscataway NJ 08903, USA; iwaniec@math.rutgers.edu

Mathematics Subject Classification (2020): 11N32, 11N36

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 $4m + 1$ 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 following well-known asymptotic formula for the number of these:

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

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

Beginning in the 1990s 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 [2] 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 [3], 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 δ.

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 [6] have shown that, in the statement of [3], 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 [8] 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 [7] have succeeded in extending the original work [2], 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

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

with coefficients β_k, γ_ℓ 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 $r \geqslant 1$ and $\Lambda = \Lambda_1$. The $\Lambda_r(n)$ vanish unless n has at most r distinct prime factors and, in any case, we have $0 \leqslant \Lambda_r(n) \leqslant (\log n)^r$. In the Appendix we shall give some heuristic arguments leading to the determination of an asymptotic formula for

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

Conjecture. *We have*

$$G_r(x) \sim c\, r\, x(\log \sqrt{x})^{r-1} \tag{1.1}$$

with

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

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

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

$$\sum\sum_{4k^2+\ell^2 \leqslant x} \Lambda(k)\Lambda(\ell)\Lambda(4k^2+\ell^2) \sim cx. \tag{GPC}$$

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

Theorem 1.1 (G_7). *We have*

$$\sum\sum_{4k^2+\ell^2 \leqslant x} \Lambda_7(k)\Lambda(\ell)\Lambda(4k^2+\ell^2) \asymp x(\log x)^6. \tag{1.3}$$

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 ℓ 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 seven prime factors, all of which are larger than $k^{\frac{1}{49}}$, and $\beta_k = 0$ otherwise. Then*

$$\sum\sum_{4k^2+\ell^2 \leqslant x} \beta_k\Lambda(\ell)\Lambda(4k^2+\ell^2) \asymp x(\log x)^{-1}. \tag{1.4}$$

Remarks 1.2. The lower bound of (1.3) follows from the lower bound of (1.4), because $\beta_k(\log k)^7 \ll \Lambda_7(k)$. 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 β is the convolution $1 * \lambda$ with λ supported on a relatively short segment. Put

$$\beta_k = \sum_{h|k} \lambda_h \tag{1.5}$$

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

The second coordinate ℓ is counted with weight γ_ℓ about which we do not need to know much. However, after serious attempts to handle γ_ℓ 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

$$|\gamma_\ell| \leqslant \log \ell \quad \text{if } \ell \text{ is an odd prime,} \tag{1.6}$$

and $\gamma_\ell = 0$, otherwise. Moreover, we need the asymptotic formula

$$\sum_{\ell \leqslant x,\, \ell \equiv a(q)} \gamma_\ell = \frac{x}{\phi(q)} + O(x(\log x)^{-B}) \tag{1.7}$$

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

Remarks 1.3. For $\gamma_\ell = \log \ell$ formula (1.7) is just the Siegel–Walfisz theorem. Have in mind that our assumption (1.7) is meaningful for $q < (\log x)^A$ with any $A \geqslant 2$, but has no value for much larger moduli. By resizing, one is allowed to multiply (1.6) and (1.7) by a fixed positive constant independent of the residue classes $a \pmod q$.

Theorem 1.3 (Main Theorem). *Suppose β_h are given by (1.5) with $|\lambda_h| \leqslant 1$ for h squarefree,*

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

and $\lambda_h = 0$, otherwise. Suppose γ_ℓ satisfies (1.6) and (1.7). Then we have

$$\sum\sum_{4k^2+\ell^2 \leqslant x} \beta_k \gamma_\ell \Lambda(4k^2 + \ell^2) = \kappa V x + O(x(\log x)^{-A}) \tag{1.8}$$

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

$$V = \sum_{h \leqslant y} \lambda_h g(h) \tag{1.9}$$

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

$$\kappa = \prod_p \left(1 - \frac{\chi(p)}{(p-1)(p-\chi(p))}\right) \quad \textit{with } \chi \pmod 4. \tag{1.10}$$

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 [2] and [4], but as we impose restrictions on both coordinates of the Gaussian integers $\ell + 2ki$ some fresh ideas occur. We consider the sequence $\mathcal{A} = (a_n)$ of numbers

$$a_n = \sum_{4k^2+\ell^2=n} \beta_k \gamma_\ell \tag{1.11}$$

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 $\frac{1}{2}x \leqslant t \leqslant x$, twice differentiable and such that

$$|t^j f^{(j)}(t)| \leqslant 1, \quad j = 0, 1, 2. \tag{1.12}$$

We are going to evaluate asymptotically the sum

$$S(x) = \sum_n a_n f(n)\Lambda(n). \tag{1.13}$$

Theorem 1.4 (Smoothed Main Theorem). *Suppose β_k and γ_ℓ satisfy the conditions of Theorem 1.3. Then*

$$S(x) = \kappa V \int f(t)\,dt + O(x(\log x)^{-A}) \tag{1.14}$$

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

It is not difficult to derive Theorem 1.3 from Theorem 1.4; see a brief explanation in Section 18.

Classical ideas for estimating sums of type (1.13) begin by partitioning into a sum of sums

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

which we call "congruence sums", and double sums

$$B = \sum_m \sum_n u_m v_n a_{mn} f(mn)$$

with suitable coefficients u_m, v_n, which we call "bilinear forms". There are plenty of possibilities, see [4, Chapters 17 and 18]. For our purpose we choose [4, Theorem 18.5], which is derived by finessing Bombieri's asymptotic sieve.

The congruence sums are treated in Sections 3 and 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. One problem of independent interest to which they give rise (see Section 16) is further developed in [5].

2. Interlude: An easier result

If one stares at our sum

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

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

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

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

Proposition 2.1. *We have*

$$H_3(x) \asymp x(\log x)^2.$$

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

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

The summation is void if d is even so we always assume that d is odd. Taking advantage of ℓ 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

$$\nu^2 + 1 \equiv 0 \pmod{d}. \tag{3.1}$$

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

$$\sum_{\nu^2+1\equiv 0(d)} \sum_{b \equiv \nu\ell\overline{2h}(d)} f(4b^2h^2+\ell^2). \tag{3.2}$$

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 (3.2), 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 (3.2) 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

$$I(\ell) = \int_0^\infty f(t^2+\ell^2)\,dt.$$

Hence our congruence sum satisfies the approximation

$$A_d(x) = \frac{\rho(d)}{2d} V_d W(x) + O(\rho(d) y \sqrt{x}) \tag{3.3}$$

with

$$V_d = \sum_{(h,d)=1} \frac{\lambda_h}{h}$$

and

$$W(x) = \sum_{\ell} \gamma_\ell I(\ell). \tag{3.4}$$

If we use assumption (1.7) (the PNT for γ_ℓ with $q = 1$), we get

$$W(x) = \frac{\pi}{4} \int f(t)\, dt + O(x(\log x)^{-B}). \tag{3.5}$$

However, to maintain transparency we shall keep the original expression (3.4) until (1.7) is really needed.

The elementary formula (3.3) suffices for odd integers d, uniformly in the range $d \leqslant y^{-1}\sqrt{x}(\log x)^{-A}$. 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 (3.2). We extend the summation over $b > 0$, $b \equiv \nu\ell\overline{2h} \pmod d$ to all $b \equiv \nu\ell\overline{2h} \pmod d$, thus counting every term twice, except for $b = 0$ in which case $d = 1$. We find that (3.2) is equal to

$$\begin{aligned} &\sum_{\nu^2+1\equiv 0(d)} \frac{1}{2dh} \sum_s e\left(\frac{\nu s\ell\overline{2h}}{d}\right) F_\ell\left(\frac{s}{2dh}\right) - \varepsilon_d f(\ell^2) \\ &\quad = \frac{1}{2dh} \sum_s \rho_{s\ell\overline{2h}}(d) F_\ell\left(\frac{s}{2dh}\right) - \varepsilon_d f(\ell^2), \end{aligned}$$

where $\varepsilon_1 = 1$, $\varepsilon_d = 0$ if $d \neq 1$,

$$F_\ell(v) = \frac{1}{2} \int_{-\infty}^{\infty} f(t^2 + \ell^2) e(-vt)\, dt \tag{3.6}$$

and

$$\rho_c(d) = \sum_{\nu^2+1\equiv 0(d)} e\left(\frac{\nu c}{d}\right)$$

is the Weyl harmonic from the theory of equidistribution of the roots of (3.1). Hence, for d odd we have

$$A_d(x) = \frac{\rho(d)}{2d} V_d W(x) + r_d(x) + O\left(\frac{\rho(d)}{d} y\sqrt{x}\log x\right), \tag{3.7}$$

where

$$r_d(x) = \sum_{(h,d)=1}\sum \lambda_h \gamma_\ell (dh)^{-1} \sum_{s>0} \rho_{s\ell\overline{2h}}(d) F_\ell\left(\frac{s}{2dh}\right). \tag{3.8}$$

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 (3.7) is negligible.

There is a considerable cancellation of the terms in (3.8) due to the spacing of the fractions ν/d modulo 1 as ν runs over the roots of (3.1). This property of $\frac{\nu}{d}$ leads to the following inequality of large sieve type.

Lemma 3.1. *Let $h \geqslant 1$. For any complex numbers α_n we have*

$$\sum_{\substack{X<d\leqslant 2X\\(d,h)=1}} \sum_{\nu^2+1\equiv 0(d)} \Big|\sum_{n\leqslant N} \alpha_n e\Big(\frac{\nu n\bar{h}}{d}\Big)\Big|^2 \leqslant 400(hX+N)\sum_{n\leqslant N}|\alpha_n|^2. \tag{3.9}$$

Proof. See [4, Section 20.2]. ∎

4. Estimation of the remainder

We need a bound for the remainder

$$R(x,D)=\sum_{\substack{d\leqslant D\\ d\text{ odd}}}|r_d(x)|,$$

where $r_d(x)$ is given by the Fourier series (3.8). Since we shall not take advantage of the summation over h, we partition (3.8) into

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

where

$$r_d(x;h)=\sum_{\ell}\gamma_\ell\sum_{s>0}\rho_{s\ell\overline{2h}}(d)F_\ell\Big(\frac{s}{2dh}\Big), \tag{4.1}$$

and we estimate the partial remainders

$$R_h(X)=\sum_{\substack{X<d\leqslant 2X\\(d,2h)=1}}|r_d(x;h)| \tag{4.2}$$

separately for every $h\leqslant y$ and $1\leqslant 2X\leqslant D$. We have

$$|R(x,D)|\leqslant\sum_{h\leqslant y}|\lambda_h|h^{-1}\sum_X R_h(X)X^{-1}, \tag{4.3}$$

where $X=\frac{D}{2},\frac{D}{4},\frac{D}{8},\dots$.

In order to apply (3.9) we build a single variable $n=s\ell$ out of the two variables s and ℓ which we need to separate from the modulus d. We accomplish the separation by the change of the variable t in the Fourier integral (3.6) into $\frac{t}{s}\sqrt{x}$ getting

$$F_\ell\Big(\frac{s}{2dh}\Big)=\frac{\sqrt{x}}{s}\int_0^\infty f\Big(\ell^2+\frac{xt^2}{s^2}\Big)\cos\Big(\frac{\pi t\sqrt{x}}{dh}\Big)\,dt. \tag{4.4}$$

The trivial bound

$$F_\ell\Big(\frac{s}{2dh}\Big)\ll\sqrt{x}$$

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 (4.4) by parts. We obtain another expression

$$F_\ell\left(\frac{s}{2dh}\right) = -\frac{2\sqrt{x}}{s}\left(\frac{dh}{\pi s}\right)^2 \int_0^\infty \left(f' + \frac{2xt^2}{s^2} f''\right) \cos\left(\frac{\pi t\sqrt{x}}{dh}\right) dt, \tag{4.5}$$

where the derivatives f', f'' are evaluated at $\ell^2 + \frac{xt^2}{s^2}$. Now estimating (4.5) trivially, we get

$$F_\ell\left(\frac{s}{2dh}\right) \ll \sqrt{x}\left(\frac{S}{s}\right)^2.$$

Let $S_0 \geq 1$. The part of (4.1) with $S_0 \leqslant s < 2S_0$ is estimated by

$$\sqrt{x} \int_0^{2S_0} \Big| \sum_{n \leqslant N} \alpha_n(t) \rho_{n\overline{2h}}(d) \Big| \, dt \tag{4.6}$$

with $N = 2\sqrt{x} S_0$, where $\alpha_n(t)$ does not depend on d,

$$\alpha_n(t) = \sum_{\substack{\ell s = n \\ S_0 \leqslant s < 2S_0}} \gamma_\ell s^{-1} f\left(\ell^2 + \frac{xt^2}{s^2}\right) \ll S_0^{-1} \log n.$$

Summing (4.6) over $X < d \leqslant 2X$ with $(d, 2h) = 1$, we derive by estimate (3.9) (apply the Cauchy–Schwarz inequality) that the partial remainder (4.2) restricted by $S_0 \leqslant s < 2S_0$ is bounded by

$$(xX)^{\frac{1}{2}} (hX + \sqrt{x} S_0)^{\frac{1}{2}} (\sqrt{x} S_0)^{\frac{1}{2}} \log(x S_0). \tag{4.7}$$

We have derived (4.7) using formula (4.4). Similarly, if we use formula (4.5), then we get the bound (4.7) with an extra factor $(\frac{S}{S_0})^2$. Combining both bounds, we see that, with optimal cutoff point, the worst result comes from $S_0 \asymp S = \frac{hX}{\sqrt{x}}$. Hence, we conclude that

$$R_h(X) \ll h x^{\frac{1}{2}} X^{\frac{3}{2}} \log x. \tag{4.8}$$

Finally, inserting (4.8) into (4.3), we obtain

Proposition 4.1. *We have*

$$R(x, D) \ll y (Dx)^{\frac{1}{2}} \log x. \tag{4.9}$$

Remark 4.1. The bound (4.9) is useful if $y^2 D \ll x (\log x)^{-A}$.

5. A model for $\mathcal{A} = (a_n)$

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 $\mathcal{B} = (b_n)$ with the numbers

$$b_n = \psi(n) \sum_{(2h,n)=1} \frac{\lambda_h \phi(h)}{h}, \tag{5.1}$$

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

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

and

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

if $p \neq 2$ and $\alpha \geqslant 1$. Recall that $\rho(p) = 1 + \chi(p)$ is the number of roots of the congruence $\nu^2 + 1 \equiv 0 \pmod p$.

Let $w(y)$ be a smooth function supported on $0 < y < 1$ with

$$\int_0^1 w(y)\,dy = 1.$$

We are going to evaluate asymptotically the sum

$$B_d(x) = \sum_{n \equiv 0 \pmod d} b_n w\Big(\frac{n}{x}\Big).$$

Note that $B_d(x) = 0$ if d is even.

Proposition 5.1. *For d odd we have*

$$B_d(x) = \frac{x}{H}\frac{\rho(d)}{2d}V_d + O\Big(\frac{\rho(d)}{\sqrt{d}}\prod_{p|d}\Big(1 + \frac{1}{\sqrt{p}}\Big)\sqrt{x}\log x\Big), \tag{5.2}$$

where H is the constant

$$H = \prod_p \Big(1 - \frac{\rho(p)}{p}\Big)\Big(1 - \frac{1}{p}\Big)^{-1}. \tag{5.3}$$

The implied constant in (5.2) depends only on the crop function w.

Proof. We execute the summation via L-functions rather than by Poisson's formula. We have

$$B_d(x) = \sum_{(h,d)=1} \lambda_h \frac{\phi(h)}{h} \sum_{(n,2h)=1} \psi(dn) w\Big(\frac{dn}{x}\Big).$$

The corresponding Dirichlet series is equal to

$$\begin{aligned} L(s) &= \sum_{(n,2h)=1} \psi(dn)(dn)^{-s} \\ &= \frac{\psi(d)}{d^s}\prod_{p|d}\Big(1 - \frac{1}{p^s}\Big)^{-1}\prod_{p\nmid 2dh}\Big(1 + \frac{\psi(p)}{p^s}\Big(1 - \frac{1}{p^s}\Big)^{-1}\Big) \\ &= \frac{\psi(d)}{d^s}\zeta(s)\prod_{p|2h}\Big(1 - \frac{1}{p^s}\Big)\prod_{p\nmid 2dh}\Big(1 + \frac{\psi(p) - 1}{p^s}\Big) \\ &= \frac{\psi(d)}{d^s}\zeta(s)\prod_{p|2h}\Big(1 - \frac{1}{p^s}\Big)\prod_{p\nmid dh}\Big(1 + \frac{\chi(p)}{p^s}\Big(1 - \frac{\rho(p)}{p}\Big)^{-1}\Big). \end{aligned}$$

Now we borrow $\frac{L(s,\chi)}{\zeta(2s)}$ and return it in the form of its Euler product, getting

$$L(s)=\frac{\zeta(s)L(s,\chi)}{\zeta(2s)}P(s)\frac{\psi(d)}{d^s}\prod_{p|2h}\Big(1-\frac{1}{p^s}\Big)\prod_{p|dh}\Big(1+\frac{\chi(p)}{p^s}\Big(1-\frac{\rho(p)}{p}\Big)^{-1}\Big)^{-1},$$

where

$$P(s)=\prod_p\Big(1-\frac{\chi(p)}{p^s}\Big)\Big(1-\frac{1}{p^{2s}}\Big)^{-1}\Big(1+\frac{\chi(p)}{p^s}\Big(1-\frac{\rho(p)}{p}\Big)^{-1}\Big).$$

For $p\neq 2$ the local factor of $P(s)$ is $1+\chi(p)/(p-1+\chi(p))(p^s+\chi(p))$ so the product converges for $\operatorname{Re}s>0$. We compute the residue of $L(s)$ at $s=1$

$$\operatorname*{res}_{s=1}L(s)=P\frac{\psi(d)}{d}\prod_{p|2h}\Big(1-\frac{1}{p}\Big)\prod_{p|dh}\Big(1-\frac{\rho(p)}{p}\Big)\Big(1-\frac{1}{p}\Big)^{-1}=\frac{P\rho(d)}{2d\phi(h)},$$

where $P=\frac{P(1)L(1,\chi)}{\zeta(2)}$. Checking the local factors, we find

$$P=\prod_p\Big(1-\frac{\rho(p)}{p}\Big)^{-1}\Big(1-\frac{1}{p}\Big)=\frac{1}{H}.$$

Finally, (5.2) follows by contour integration with the error term obtained by trivial estimations on the line $\operatorname{Re}s=\frac{1}{2}$. ∎

Remarks 5.1. The main term of formula (5.2) agrees with that of (3.7) after normalization. Checking the local factors of H in (5.3) and κ in (1.10) against $L(1,\chi)$, we see that

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

6. Sums over primes

Theorem 18.5 of [4] 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 [4], but not stated explicitly).

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

$$\Big|\sum_{xz^{-2}<n\leqslant x}c_n\Lambda(n)\Big|\leqslant\Big|\sum_{d\leqslant z}\mu(d)\mathcal{C}'_d(x)\Big|+(\log x)\sum_{d\leqslant xz^{-1}}|\mathcal{C}_d(x)|$$
$$+2(\log x)\sum_n\Big|\sum_{\substack{mn\leqslant x\\ z<m\leqslant z^2}}\mu(m)c_{mn}\Big|,\tag{6.1}$$

where

$$\mathcal{C}'_d(x)=\sum_{\substack{n\leqslant x\\ n\equiv 0(d)}}c_n\log n,\quad \mathcal{C}_d(x)=\sum_{\substack{n\leqslant x\\ n\equiv 0(d)}}c_n.$$

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 $\mu(d)\mathcal{C}'_d(x)$ 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 $z = x^\delta$, $0 < \delta \leqslant \frac{1}{8}$, for the sequence of numbers

$$c_n = a_n f(n) - H\frac{W(x)}{x} w\Big(\frac{n}{x}\Big) b_n,$$

where $\mathcal{A} = (a_n)$ is our target sequence (1.11) and $\mathcal{B} = (b_n)$ is its model (5.1). Note that $c_n = 0$, unless $n < x$, n odd. The congruence sums of $\mathcal{C} = (c_n)$ have no main term; compare (3.7) with (5.2).

On the left-hand side of (6.1) we get (up to $O(xz^3)$)

$$\begin{aligned}
&\sum_n a_n f(n)\Lambda(n) - H\frac{W(x)}{x}\sum_n b_n\Lambda(n) w\Big(\frac{n}{x}\Big)\\
&\quad = S(x) - H\frac{W(x)}{x}\sum_h \lambda_h\frac{\phi(h)}{h}\sum_{(n,2h)=1}\psi(n)\Lambda(n) w\Big(\frac{n}{x}\Big)\\
&\quad = S(x) - H\frac{W(x)}{x} V\sum_{p\equiv 1(4)} 2w\Big(\frac{p}{x}\Big)\log p + O(\sqrt{x}\log x)\\
&\quad = S(x) - HVW(x) + O(x(\log x)^{-A})
\end{aligned}$$

by the PNT, where A is any number $\geqslant 2$.

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

$$R' = \sum_{\substack{d\leqslant z\\ d \text{ odd}}}\mu(d)\Big(\sum_{n\equiv 0(d)} a_n f(n)\log n - H\frac{W(x)}{x}\sum_{n\equiv 0(d)} b_n w\Big(\frac{n}{x}\Big)\log n\Big).$$

The second sum is

$$R = \sum_{\substack{d\leqslant xz^{-1}\\ d \text{ odd}}}\Big|A_d(x) - H\frac{W(x)}{x}B_d(x)\Big|.$$

The third sum is the bilinear form

$$B = \sum_n\Big|\sum_{z<m\leqslant z^2}\mu(m)a_{mn}f(mn) - H\frac{W(x)}{x}\sum_{z<m\leqslant z^2}\mu(m)b_{mn}w\Big(\frac{mn}{x}\Big)\Big|.$$

We estimate R' by applying two elementary approximations to the main terms, namely formula (3.3) with $f(t)$ replaced by $f(t)\log t$ and formula (5.2) with $w(y)$ replaced by $w(y)\log xy$. We obtain

$$\begin{aligned}
R' = &\sum_{\substack{d\leqslant z\\ d \text{ odd}}}\mu(d)\frac{\rho(d)}{2d}V_d\sum_\ell\gamma_\ell\int_0^\infty f(t^2+\ell^2)\int_0^1 w(y)\log\Big(\frac{t^2+\ell^2}{xy}\Big)\,dy\,dt\\
&+ O(yz\sqrt{x}\log x).
\end{aligned}$$

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

$$c = \exp\Big(\int w(y) \log y \, dy\Big),$$

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

$$\sum_{\substack{d \leqslant z \\ d \text{ odd}}} \mu(d) \frac{\rho(d)}{2d} V_d = \sum_h \frac{\lambda_h}{h} \sum_{\substack{d \leqslant z \\ (d,2h)=1}} \mu(d) \frac{\rho(d)}{2d} \ll (\log z)^{-A}.$$

Hence

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

with any $A \geqslant 2$, 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 (4.9), (5.2), respectively. We get

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

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 the function $\mu(m)$ does not correlate with the original sequence a_{mn}, but this is clear for the model sequence b_{mn}. We have

$$\sum_{z \leqslant m \leqslant z^2} \mu(m) b_{mn} w\Big(\frac{mn}{x}\Big) = \sum_{(2h,n)=1} \lambda_h \frac{\phi(h)}{h} \sum_{\substack{z < m \leqslant z^2 \\ (m,2h)=1}} \mu(m) \psi(mn) w\Big(\frac{mn}{x}\Big).$$

By the PNT we find that the last sum over m is $\ll n^{-1} x(\log x)^{-A-3}$. Next, summing over $h \leqslant y$ and $n < xz^{-1}$ 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

$$B(x,z) = \sum_n \Big| \sum_{z < m \leqslant z^2} \mu(m) a_{mn} f(mn) \Big|. \tag{6.2}$$

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

Proposition 6.2. *Let*

$$y^2 (\log x)^{2A+4} \leqslant z \leqslant x^{\frac{1}{8}}. \tag{6.3}$$

Then

$$|S(x) - HVW(x)| \leqslant 2B(x,z) \log x + O(x(\log x)^{-A}). \tag{6.4}$$

If we assume (1.7), then $W(x)$ satisfies (3.5) so (6.4) becomes

$$|S(x) - \kappa V \int f(t)\, dt| \leqslant 2B(x,z) \log x + O(x(\log x)^{-A}).$$

To complete the proof of (1.14), it remains to show that

$$B(x,z) \ll x(\log x)^{-A-1} \tag{6.5}$$

subject to condition (6.3).

7. Bilinear forms in the Gaussian domain

It remains to estimate the bilinear form (6.2). We need the bound

$$B(x,z) \ll x(\log x)^{-A-1} \tag{7.1}$$

with any $A \geqslant 2$. In this section we make several simplifications before launching the essential arguments.

First we split the segment $z < m \leqslant z^2$ into dyadic intervals $M < m \leqslant 2M$. Assume for simplicity that $\frac{\log z}{\log 2}$ is an integer so we cover the segment exactly with $\frac{2\log z}{\log 2}$ dyadic intervals. We get

$$B(x,z) \leqslant \sum_M \sum_n \Big| \sum_{m\sim M} \mu(m) a_{mn} f(mn) \Big|,$$

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

$$B(x,z) \leqslant \sum_M \sum_n \sum_{c^2|n} \Big| \sum_{\substack{m\sim \frac{M}{c} \\ (m,n)=1}} \mu(m) a_{mn} f(mn) \Big|.$$

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 $c^2 \mid n$ for $c \leqslant C$ getting

$$B(x,z) \leqslant B^*(M)(\log x)^{A+4} + O(x(\log x)^{-A-1}),$$

where

$$B^*(M) = \sum_n \Big| \sum_{\substack{m\sim M \\ (m,n)=1}} \mu(m) a_{mn} f(mn) \Big|$$

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

Next we write (see (1.5) and (1.11))

$$a_n = \sum_h \lambda_h a_n(h),$$

where

$$a_n(h) = \sum_{\substack{4k^2+\ell^2=n \\ h|k}} \gamma_\ell.$$

Hence

$$B^*(M) \leqslant \sum_h |\lambda_h| B_h^*(M),$$

where

$$B_h^*(M) = \sum_n \Bigg| \sum_{\substack{m\sim M \\ (m,2hn)=1}} \mu(m) a_{mn}(h) f(mn) \Bigg|.$$

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 $e \mid \ell^2$, $e \mid \ell$, $e^2 \mid mn$, $e^2 \mid m$, 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 $\mathfrak{a}, \mathfrak{b}, \mathfrak{m}, \mathfrak{n}, \ldots$ denote Gaussian integers and the corresponding latin letters $a, b, m, n, \ldots$ denote the norms; $a = \mathfrak{a}\bar{\mathfrak{a}}$, $b = \mathfrak{b}\bar{\mathfrak{b}}$, $m = \mathfrak{m}\bar{\mathfrak{m}}$, $n = \mathfrak{n}\bar{\mathfrak{n}}, \ldots$. By the unique factorization in $\mathbb{Z}[i]$ we obtain $B_h^*(M) \leqslant \mathcal{B}_h^*(M)$, where

$$\mathcal{B}_h^*(M) = \sum_{\mathfrak{n}} \Bigg| \sum_{\substack{(\mathfrak{m},2hn)=1,\ m\sim M \\ \operatorname{Im} \mathfrak{m}\mathfrak{n} \equiv 0 (2h)}} \mu(m) \xi(\mathfrak{m}\mathfrak{n}) f(mn) \Bigg|.$$

Here we put

$$\xi(\mathfrak{a}) = \gamma_{\operatorname{Re} \mathfrak{a}}.$$

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

The condition $(m, n) = 1$ was needed for performing the unique factorization in $\mathbb{Z}[i]$. After that, the resulting condition $(\mathfrak{m}, \mathfrak{n}) = 1$ 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

$$\sum_{\mathfrak{b}\mathfrak{c}=\mathfrak{m}, \mathfrak{c}|\mathfrak{q}} \mu(\mathfrak{b}) = \begin{cases} 4\mu(\mathfrak{m}) & \text{if } (\mathfrak{m}, \mathfrak{q}) = 1, \\ 0 & \text{otherwise,} \end{cases}$$

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

$$\sum_{\mathfrak{c}|n^\infty} \Bigg| \sum_{\substack{(\mathfrak{c}\mathfrak{m},2h)=1,\, cm\sim M \\ \operatorname{Im} \mathfrak{c}\mathfrak{m}\mathfrak{n} \equiv 0 (2h)}} \mu(m) \xi(\mathfrak{c}\mathfrak{m}\mathfrak{n}) f(cmn) \Bigg|$$

and

$$\mathcal{B}_h^*(M) \leqslant \sum_{\mathfrak{n}} \sum_{\mathfrak{c}|n^\infty} \Bigg| \sum_{\substack{\mathfrak{m}\sim \frac{M}{c},\, (\mathfrak{m},2h)=1 \\ \operatorname{Im} \mathfrak{m}\mathfrak{n}\equiv 0(2h)}} \mu(m)\xi(\mathfrak{m}\mathfrak{n}) f(mn) \Bigg|.$$

Note that there is no condition $(m, c) = 1$.

We keep the terms with $c \leqslant C_1 = (\log x)^{2A+8}$ and estimate the remaining terms with larger c trivially getting

$$B(x,z) \leqslant \mathcal{B}(M)(\log x)^{3A+12} + O(x(\log x)^{-A-1})$$

for some M with $\frac{z}{CC_1} \leqslant M < z^2$, where

$$\mathcal{B}(M) = \sum_h |\lambda_h| \mathcal{B}_h(M)$$

and

$$\mathcal{B}_h(M) = \sum_{\mathfrak{n}} \Bigg| \sum_{\substack{\mathfrak{m}\sim M,\, (\mathfrak{m},2h)=1 \\ \operatorname{Im} \mathfrak{m}\mathfrak{n}\equiv 0(2h)}} \mu(m)\xi(\mathfrak{m}\mathfrak{n}) f(mn) \Bigg|. \tag{7.2}$$

Now we need to show that

$$\mathcal{B}(M) \ll x(\log x)^{-4A-13}$$

for some M with

$$z(\log x)^{-3A-11} \leqslant M < z^2. \tag{7.3}$$

Some properties of $\mathfrak{n}$ in the outer sum of (7.2) are hidden but can be inferred from the equation $mn = 4b^2h^2 + \ell^2$ and the support of $f(mn)$ being $\frac{x}{2} < mn < x$. 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 $\mathfrak{m} \sim M$, but never forget it.

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

$$\mathcal{B}^2(M) \ll \mathcal{C}(M) N \log y,$$

where

$$\mathcal{C}(M) = \sum_h |\lambda_h| h \mathcal{C}_h(M) \tag{7.4}$$

and

$$\mathcal{C}_h(M) = \sum_{\mathfrak{n}} \Bigg| \sum_{\substack{\mathfrak{m}\sim M,\, (\mathfrak{m},2h)=1 \\ \operatorname{Im} \mathfrak{m}\mathfrak{n}\equiv 0(2h)}} \mu(m)\xi(\mathfrak{m}\mathfrak{n}) f(mn) \Bigg|^2.$$

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

$$\mathcal{C}(M) \ll NM^2(\log x)^{-8A-27}. \tag{7.5}$$

Squaring out and interchanging the order of summation, we write

$$\mathcal{C}_h(M) = \sum_{\substack{(\mathfrak{m}_1\mathfrak{m}_2,2h)=1 \\ \mathfrak{m}_1\sim M,\, \mathfrak{m}_2\sim M}} \sum \mu(m_1)\mu(m_2) \mathcal{D}_h(\mathfrak{m}_1, \mathfrak{m}_2) \tag{7.6}$$

with

$$\mathcal{D}_h(\mathfrak{m}_1, \mathfrak{m}_2) = \sum_{\mathfrak{n}} \xi(\mathfrak{m}_1\mathfrak{n})\bar{\xi}(\mathfrak{m}_2\mathfrak{n}) f(m_1 n)\overline{f}(m_2 n), \tag{7.7}$$

where the summation runs over all Gaussian integers $\mathfrak{n}$ satisfying

$$\operatorname{Im} \mathfrak{m}_1\mathfrak{n} \equiv \operatorname{Im} \mathfrak{m}_2\mathfrak{n} \equiv 0 \ (\operatorname{mod} 2h).$$

Opening the Gaussian domain, we see that

$$\mathcal{D}_h(\mathfrak{m}_1, \mathfrak{m}_2) = \sum_{\ell_1}\sum_{\ell_2} \gamma_{\ell_1}\overline{\gamma}_{\ell_2} f(m_1 n)\overline{f}(m_2 n), \tag{7.8}$$

where the summation runs over the solutions of the system

$$\mathfrak{m}_1\mathfrak{n} = \ell_1 + 2hb_1 i,$$
$$\mathfrak{m}_2\overline{\mathfrak{n}} = \ell_2 + 2hb_2 i$$

in $\mathfrak{n}$, ℓ_1, ℓ_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

$$\begin{aligned} \mathfrak{m}_1\mathfrak{n} &\equiv \overline{\mathfrak{m}}_1\overline{\mathfrak{n}} \ (\operatorname{mod} 4h), \\ \mathfrak{m}_2\mathfrak{n} &\equiv \overline{\mathfrak{m}}_2\overline{\mathfrak{n}} \ (\operatorname{mod} 4h) \end{aligned} \tag{7.9}$$

and two equations

$$\begin{aligned} \mathfrak{m}_1\mathfrak{n} + \overline{\mathfrak{m}}_1\overline{\mathfrak{n}} &= 2\ell_1, \\ \mathfrak{m}_2\mathfrak{n} + \overline{\mathfrak{m}}_2\overline{\mathfrak{n}} &= 2\ell_2. \end{aligned} \tag{7.10}$$

Put

$$\Delta = \Delta(\mathfrak{m}_1, \mathfrak{m}_2) = \frac{i}{2}(\mathfrak{m}_1\overline{\mathfrak{m}}_2 - \overline{\mathfrak{m}}_1\mathfrak{m}_2) = \operatorname{Im} \overline{\mathfrak{m}}_1\mathfrak{m}_2$$

so Δ is a rational integer, relatively small;

$$|\Delta| < 4M < 4z^2.$$

8. The diagonal terms

First we give a quick estimation of $\mathcal{D}_h(\mathfrak{m}_1, \mathfrak{m}_2)$ in the singular case $\Delta = \Delta(\mathfrak{m}_1, \mathfrak{m}_2) = 0$. We get $\mathfrak{m}_1\overline{\mathfrak{m}}_2 = \overline{\mathfrak{m}}_1\mathfrak{m}_2$, $\mathfrak{m}_1 \mid \mathfrak{m}_2$ and $\mathfrak{m}_2 \mid \mathfrak{m}_1$, $\mathfrak{m}_2 = \varepsilon\mathfrak{m}_1$ with $\varepsilon = \pm 1, \pm i$. From system (7.10) we obtain

$$\ell_1\mathfrak{m}_2 - \ell_2\mathfrak{m}_1 = -i\Delta\overline{\mathfrak{n}}.$$

In the singular case this yields $\varepsilon\ell_1 = \ell_2$, so $\varepsilon = 1$. Therefore we have $\mathfrak{m}_1 = \mathfrak{m}_2 = \mathfrak{m}$ and $\ell_1 = \ell_2 = \ell$, say. In this case $\mathcal{D}_h(\mathfrak{m}, \mathfrak{m})$ is bounded by the number of solutions in b and ℓ of

$$\ell + 2hbi \equiv 0 \ (\operatorname{mod} \mathfrak{m}), \quad |\ell + 2hbi| < \sqrt{x}.$$

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

$$\mathcal{D}_h(\mathfrak{m}, \mathfrak{m}) \ll \frac{x\rho(m)}{mh}.$$

The contribution of $\mathcal{D}_h(\mathfrak{m},\mathfrak{m})$ to $\mathcal{C}_h(M)$ is estimated by

$$\frac{x}{h}\sum_{\mathfrak{m}\sim M}\frac{\rho(m)}{m}\leqslant\frac{x}{h}\sum_{m\sim M}\frac{\rho(m)^2}{m}\ll\frac{x}{h}(\log M)^2.$$

Hence the contribution of $\mathcal{D}_h(\mathfrak{m},\mathfrak{m})$ to $\mathcal{C}(M)$ is $\ll yx(\log x)^2\ll NM^2(\log x)^{-8A-27}$ by (7.4) as required by (7.5), provided

$$y(\log x)^{11A+40}\leqslant z. \tag{8.1}$$

9. In the off-diagonal area

From now on we assume that $\Delta=\Delta(\mathfrak{m}_1,\mathfrak{m}_2)\neq 0$. Now the system of equations (7.10) has a unique solution in the complex number $\mathfrak{n}$ given by

$$i\Delta\mathfrak{n}=\ell_1\overline{\mathfrak{m}}_2-\ell_2\overline{\mathfrak{m}}_1. \tag{9.1}$$

Since $\mathfrak{n}$ must be a Gaussian integer, this means ℓ_1, ℓ_2 satisfy

$$\ell_1\mathfrak{m}_2\equiv\ell_2\mathfrak{m}_1\ (\mathrm{mod}\ \Delta). \tag{9.2}$$

For $\mathfrak{n}$ given by (9.1) the congruences (7.9) become

$$\mathfrak{m}_1(\ell_1\overline{\mathfrak{m}}_2-\ell_2\overline{\mathfrak{m}}_1)+\overline{\mathfrak{m}}_1(\ell_1\mathfrak{m}_2-\ell_2\mathfrak{m}_1)\equiv 0\ (\mathrm{mod}\ 4\Delta h),$$
$$\mathfrak{m}_2(\ell_1\overline{\mathfrak{m}}_2-\ell_2\overline{\mathfrak{m}}_1)+\overline{\mathfrak{m}}_2(\ell_1\mathfrak{m}_2-\ell_2\mathfrak{m}_1)\equiv 0\ (\mathrm{mod}\ 4\Delta h).$$

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

$$\ell_1(\mathfrak{m}_1\overline{\mathfrak{m}}_2+\overline{\mathfrak{m}}_1\mathfrak{m}_2)\equiv 2\ell_2 m_1\ (\mathrm{mod}\ 4\Delta h), \tag{9.3}$$
$$\ell_2(\mathfrak{m}_1\overline{\mathfrak{m}}_2+\overline{\mathfrak{m}}_1\mathfrak{m}_2)\equiv 2\ell_1 m_2\ (\mathrm{mod}\ 4\Delta h). \tag{9.4}$$

In other words the summation in (7.7) runs over the odd prime numbers ℓ_1, ℓ_2 satisfying the congruences in (9.2), (9.3), (9.4), and $n=\mathfrak{n}\overline{\mathfrak{n}}$ is determined by (9.1).

The congruences (9.2), (9.3), (9.4) imply several conditions on $\mathfrak{m}_1$, $\mathfrak{m}_2$. It will be easier to see these conditions after pulling out the common factor $\mathfrak{d}=(\mathfrak{m}_1,\mathfrak{m}_2)$. We put (temporarily)

$$\mathfrak{m}_1=\mathfrak{a}_1\mathfrak{d},\ \mathfrak{m}_2=\mathfrak{a}_2\mathfrak{d}\quad\text{with }(\mathfrak{a}_1,\mathfrak{a}_2)=1.$$

Note that $\mathfrak{d}$ is primitive and $(\mathfrak{d},2h\mathfrak{a}_1\mathfrak{a}_2)=1$, because $\mathfrak{m}_1$, $\mathfrak{m}_2$ are primitive squarefree, co-prime with $2h$. Put

$$D=\Delta(\mathfrak{a}_1,\mathfrak{a}_2)=\frac{\Delta(\mathfrak{m}_1,\mathfrak{m}_2)}{d},\qquad d=\mathfrak{d}\overline{\mathfrak{d}}.$$

Note that $(a_1a_2,2D)=1$. Dividing (9.2) by $\mathfrak{d}$ and conjugating, we get

$$\ell_1\overline{\mathfrak{a}}_2\equiv\ell_2\overline{\mathfrak{a}}_1\ (\mathrm{mod}\ \mathfrak{d}D), \tag{9.5}$$

and dividing (9.3), (9.4) by $d=\mathfrak{d}\overline{\mathfrak{d}}$, we get

$$\ell_1(\mathfrak{a}_1\overline{\mathfrak{a}}_2+\overline{\mathfrak{a}}_1\mathfrak{a}_2)\equiv 2\ell_2 a_1\ (\mathrm{mod}\ 4Dh) \tag{9.6}$$
$$\ell_2(\mathfrak{a}_1\overline{\mathfrak{a}}_2+\overline{\mathfrak{a}}_1\mathfrak{a}_2)\equiv 2\ell_1 a_2\ (\mathrm{mod}\ 4Dh). \tag{9.7}$$

Recall that $a_1 = \mathfrak{a}_1\bar{\mathfrak{a}}_1$ and $a_2 = \mathfrak{a}_2\bar{\mathfrak{a}}_2$. Since $\mathfrak{a}_1\bar{\mathfrak{a}}_2 + \bar{\mathfrak{a}}_1\mathfrak{a}_2 = 2\mathfrak{a}_1\bar{\mathfrak{a}}_2 + 2iD$, we can write (9.6) and (9.7) in the form

$$\ell_1(\mathfrak{a}_1\bar{\mathfrak{a}}_2 + iD) \equiv \ell_2 a_1 \pmod{2Dh} \tag{9.8}$$

$$\ell_2(\mathfrak{a}_1\bar{\mathfrak{a}}_2 + iD) \equiv \ell_1 a_2 \pmod{2Dh}. \tag{9.9}$$

Multiplying these congruences by sides and dividing by $\ell_1\ell_2$, we get $D^2 \equiv 0 \pmod{2Dh}$, hence

$$D \equiv 0 \pmod{2h} \quad \text{and} \quad \Delta \equiv 0 \pmod{2h}. \tag{9.10}$$

Having condition (9.10) it is now clear that (9.8) is equivalent to (9.9). Indeed, (9.8) implies

$$\begin{aligned} a_1\ell_2(\mathfrak{a}_1\bar{\mathfrak{a}}_2 + iD) &\equiv \ell_1(\mathfrak{a}_1\bar{\mathfrak{a}}_2 + iD)^2 \\ &= \ell_1(\mathfrak{a}_1^2\bar{\mathfrak{a}}_2^2 + 2iD\mathfrak{a}_1\bar{\mathfrak{a}}_2 - D^2) \\ &= a_1\ell_1 a_2 - \ell_1 D^2 \\ &\equiv a_1\ell_1 a_2 \pmod{2Dh}, \end{aligned}$$

which yields (9.9). Conversely (9.9) implies (9.8) by similar arguments.

We are left with (9.5) and (9.9). These two congruences determine ℓ_2/ℓ_1 uniquely modulo the least common multiple of $\mathfrak{d}D$ and $2Dh$ which is $2\mathfrak{d}Dh$. Since ℓ_2/ℓ_1 is rational, it is determined uniquely modulo the least common multiple of $2\bar{\mathfrak{d}}Dh$ and $2\mathfrak{d}Dh$ which is $2dDh$. Therefore we can write the two congruences (9.5) and (9.9) for ℓ_1, ℓ_2 as one congruence

$$\ell_2 \equiv \omega\ell_1 \pmod{2dDh}, \tag{9.11}$$

where ω is the unique rational reduced residue class modulo $2dDh = 2\Delta h$ such that

$$\omega a_1 \equiv \mathfrak{a}_1\bar{\mathfrak{a}}_2 \pmod{\mathfrak{d}D}$$

and

$$\omega a_1 \equiv \mathfrak{a}_1\bar{\mathfrak{a}}_2 + iD \pmod{2Dh}.$$

By (7.8) and (9.11) we can write

$$\mathcal{D}_h(\mathfrak{m}_1, \mathfrak{m}_2) = \sum_{\ell_2 \equiv \omega\ell_1} \sum_{\pmod{2\Delta h}} \gamma_{\ell_1}\bar{\gamma}_{\ell_2} f(m_2 n)\bar{f}(m_1 n) \tag{9.12}$$

with n given by (9.1), that is n is a quadratic form in ℓ_1, ℓ_2;

$$n = n(\ell_1, \ell_2) = |\ell_1\mathfrak{m}_2 - \ell_2\mathfrak{m}_1|^2 \Delta^{-2}. \tag{9.13}$$

By the distribution of primes ℓ_1, ℓ_2 in arithmetic progressions we expect that the main term of (9.12) should be

$$\mathcal{E}_h(\mathfrak{m}_1, \mathfrak{m}_2) = \frac{1}{\varphi(2\Delta h)} \sum_{\ell_1} \sum_{\ell_2} \gamma_{\ell_1}\bar{\gamma}_{\ell_2} f(m_2 n)\bar{f}(m_1 n) \tag{9.14}$$

which does not depend on ω. Subtracting $\mathcal{E}_h(\mathfrak{m}_1, \mathfrak{m}_2)$ from $\mathcal{D}_h(\mathfrak{m}_1, \mathfrak{m}_2)$ we get

$$\begin{aligned} \mathcal{R}_h(\mathfrak{m}_1, \mathfrak{m}_2) = &\sum_{\ell_2 \equiv \omega \ell_1 (2\Delta h)}\sum \gamma_{\ell_1} \overline{\gamma}_{\ell_2} f(m_2 n) \overline{f}(m_1 n) \\ &- \frac{1}{\varphi(2\Delta h)} \sum_{\ell_1} \sum_{\ell_2} \gamma_{\ell_1} \overline{\gamma}_{\ell_2} f(m_2 n) \overline{f}(m_1 n) \end{aligned} \tag{9.15}$$

which is regarded as an error term.

We need to sum $\mathcal{E}_h(\mathfrak{m}_1, \mathfrak{m}_2)$ and $\mathcal{R}_h(\mathfrak{m}_1, \mathfrak{m}_2)$ over $\mathfrak{m}_1$, $\mathfrak{m}_2$ as in (7.6) and over h as in (7.4) restricted by $\Delta(\mathfrak{m}_1, \mathfrak{m}_2) \equiv 0 \pmod{2h}$, 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 ℓ_1, ℓ_2 from $\mathfrak{m}_1$, $\mathfrak{m}_2$, because $\mathfrak{m}_1$, $\mathfrak{m}_2$ are constituents of the moduli $\Delta(\mathfrak{m}_1, \mathfrak{m}_2)h$. Although in most cases the determinant $\Delta(\mathfrak{m}_1, \mathfrak{m}_2)$ 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

$$\begin{aligned} f(x_1, x_2) &= f(m_2 n(x_1, x_2)) \overline{f}(m_1 n(x_1, x_2)) \\ &= \iint g(\alpha_1, \alpha_2) e(\alpha_1 x_1 + \alpha_2 x_2)\, d\alpha_1\, d\alpha_2, \end{aligned} \tag{10.1}$$

where

$$g(\alpha_1, \alpha_2) = \iint f(x_1, x_2) e(-\alpha_1 x_1 - \alpha_2 x_2)\, dx_1\, dx_2. \tag{10.2}$$

Recall that $n(x_1, x_2)$ is the quadratic form given by (9.13). By the linear change of variables $(x_1, x_2) = (x, y)$ given by

$$x_1 = x \operatorname{Im} \frac{\mathfrak{m}_1}{\mathfrak{m}_2} + y \operatorname{Re} \frac{\mathfrak{m}_1}{\mathfrak{m}_2}, \quad x_2 = y,$$

we diagonalize $n(x_1, x_2) = \frac{1}{m_2}(x^2 + y^2)$ getting

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

where we denote temporarily $I = \operatorname{Im} \frac{\mathfrak{m}_1}{\mathfrak{m}_2}$ and $R = \operatorname{Re} \frac{\mathfrak{m}_1}{\mathfrak{m}_2}$. Note that

$$I = -\frac{\Delta(\mathfrak{m}_1, \mathfrak{m}_2)}{m_2} \neq 0$$

so $|I| > M^{-1}$. Moreover,

$$I^2 + R^2 = \frac{m_1}{m_2} \asymp 1$$

so if I is small, then $|R| \asymp 1$.

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

$$\iint f(x^2+y^2)e(-ax-by)\,dx\,dy = F(a^2+b^2), \tag{10.3}$$

where $F(s)$ is the Hankel transform of $f(t)$,

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

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

$$s(\alpha_1,\alpha_2) = \left|\alpha_2+\alpha_1\frac{\mathfrak{m}_1}{\mathfrak{m}_2}\right|^2 = \alpha_2^2 + 2\alpha_1\alpha_2 \operatorname{Re}\frac{\mathfrak{m}_1}{\mathfrak{m}_2} + \alpha_1^2\frac{m_1}{m_2}, \tag{10.4}$$

$$F(s) = \pi\int_0^\infty J_0(2\pi\sqrt{st})f(t)\overline{f}\left(\frac{tm_1}{m_2}\right)dt, \tag{10.5}$$

$$g(\alpha_1,\alpha_2) = IF(s(\alpha_1,\alpha_2)) \tag{10.6}$$

and

$$f(m_2 n)\overline{f}(m_1 n) = I\iint F(s(\alpha_1,\alpha_2))e(\alpha_1\ell_1+\alpha_2\ell_2)\,d\alpha_1\,d\alpha_2. \tag{10.7}$$

Going through the Fourier transform, we lost sight on the ranges of ℓ_1, ℓ_2 so let us record that

$$\ell_1, \ell_2 < \sqrt{x}. \tag{10.8}$$

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

Estimating directly and after integrating by parts two times of (10.5), we find that

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

Hence $F(s)$ is very small if $s > x^{-1}(\log x)^{2C}$ so the integration (10.7) runs effectively over the set (ellipse)

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

whose volume (the Lebesgue measure) is equal to

$$|S| = \frac{\pi(\log x)^{2C}}{|I|x}. \tag{10.9}$$

Note that the trivial integration shows that (10.7) is bounded

$$
\begin{aligned}
|I| \iint_{\mathbb{R}^2} |F(s(\alpha_1,\alpha_2))|\, d\alpha_1\, d\alpha_2 &= \iint_{\mathbb{R}^2} |F(\alpha_1^2+\alpha_2^2)|\, d\alpha_1\, d\alpha_2 \\
&\ll x \iint_{\mathbb{R}^2} (1+(\alpha_1^2+\alpha_2^2)x)^{-2}\, d\alpha_1\, d\alpha_2 \\
&= \iint_{\mathbb{R}^2} (1+\alpha_1^2+\alpha_2^2)^{-2}\, d\alpha_1\, d\alpha_2 \ll 1.
\end{aligned}
$$

Similarly we find that the integral over $\mathbb{R}^2 \setminus S$ is small;

$$
\begin{aligned}
|I| \iint_{\mathbb{R}^2\setminus S} |F(s(\alpha_1,\alpha_2))|\, d\alpha_1\, d\alpha_2 &\ll |I| x(\log x)^{-C} \iint_{\mathbb{R}^2} (1+s(\alpha_1,\alpha_2)x)^{-\frac{3}{2}}\, d\alpha_1\, d\alpha_2 \\
&= (\log x)^{-C} \iint_{\mathbb{R}^2} (1+\alpha_1^2+\alpha_2^2)^{-\frac{3}{2}}\, d\alpha_1\, d\alpha_2 \\
&\ll (\log x)^{-C}.
\end{aligned}
$$

Therefore, we lost essentially nothing by the separation of the variables ℓ_1, ℓ_2 through the Fourier transform (10.7). We get

$$
f(m_2 n)\overline{f}(m_1 n) = I \iint_S F(s(\alpha_1,\alpha_2)) e(\alpha_1\ell_1+\alpha_2\ell_2)\, d\alpha_1 d\alpha_2 + O((\log x)^{-C}). \quad (10.10)
$$

11. Estimation of $\mathcal{R}_h''(\mathfrak{m}_1, \mathfrak{m}_2)$

Recall that the error term $\mathcal{R}_h(\mathfrak{m}_1, \mathfrak{m}_2)$ is given by (9.15). Introducing (10.10) into (9.15), we get

$$
\mathcal{R}_h(\mathfrak{m}_1, \mathfrak{m}_2) = \mathcal{R}_h'(\mathfrak{m}_1, \mathfrak{m}_2) + \mathcal{R}_h''(\mathfrak{m}_1, \mathfrak{m}_2)
$$

where

$$
\mathcal{R}_h''(\mathfrak{m}_1, \mathfrak{m}_2) \ll \frac{x}{\varphi(\Delta h)} (\log x)^{-C}, \quad (11.1)
$$

$$
\mathcal{R}_h'(\mathfrak{m}_1, \mathfrak{m}_2) = I \iint_S F(s(\alpha_1,\alpha_2)) H(\alpha_1,\alpha_2)\, d\alpha_1\, d\alpha_2 \quad (11.2)
$$

and

$$
\begin{aligned}
H(\alpha_1,\alpha_2) = &\sum_{\ell_2 \equiv \omega\ell_1 (2\Delta h)} \gamma_{\ell_1}\overline{\gamma}_{\ell_2} e(\alpha_1\ell_1+\alpha_2\ell_2) \\
&- \frac{1}{\varphi(2\Delta h)} \sum_{\ell_1}\sum_{\ell_2} \gamma_{\ell_1}\overline{\gamma}_{\ell_2} e(\alpha_1\ell_1+\alpha_2\ell_2).
\end{aligned} \quad (11.3)
$$

The total contribution of $\mathcal{R}_h''(\mathfrak{m}_1, \mathfrak{m}_2)$ to $\mathcal{C}_h(M)$ is (see (7.6) and (9.10))

$$
\sum_{\substack{(\mathfrak{m}_1\mathfrak{m}_2, 2h)=1 \\ 0\neq\Delta(\mathfrak{m}_1,\mathfrak{m}_2)\equiv 0 (2h)}} \sum \mu(m_1)\mu(m_2)\mathcal{R}_h''(\mathfrak{m}_1, \mathfrak{m}_2).
$$

The determinant $\Delta = \Delta(\mathfrak{m}_1, \mathfrak{m}_2)$ occurs with a multiplicity which is bounded by $8M$ so the above contribution is bounded by

$$Mx(\log x)^{-C} \sum_{\substack{1 \leqslant \Delta < 4M \\ \Delta \equiv 0 (2h)}} \frac{1}{\varphi(\Delta h)} \ll h^{-2} Mx(\log x)^{2-C}.$$

Inserting this bound into (7.4), we find that the total contribution of $\mathcal{R}''_h(\mathfrak{m}_1, \mathfrak{m}_2)$ to $\mathcal{C}(M)$, say $\mathcal{C}''(M)$, satisfies

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

This bound satisfies our requirement (7.5) if we take C to be a sufficiently large constant, specifically $C \geqslant 8A + 30$.

12. Small determinant

The estimation $\mathcal{R}'_h(\mathfrak{m}_1, \mathfrak{m}_2)$ is quite delicate because the determinant $\Delta = \Delta(\mathfrak{m}_1, \mathfrak{m}_2) = \operatorname{Im} \overline{\mathfrak{m}}_1 \mathfrak{m}_2$ can be small, in which case the separation of the variables ℓ_1, ℓ_2 by means of the Fourier transform (see (11.2) and (11.3)) cannot be treated in a straightforward fashion. The set S has relatively large measure, see (10.9), and there is a lot of room for α_1. Recall that $s(\alpha_1, \alpha_2)$ is the quadratic form in α_1, α_2 and

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

with

$$I = \operatorname{Im}\left(\frac{\mathfrak{m}_1}{\mathfrak{m}_2}\right) = -\frac{\Delta}{m_2} \asymp \frac{|\Delta|}{M}, \quad R = \operatorname{Re}\left(\frac{\mathfrak{m}_1}{\mathfrak{m}_2}\right) \asymp 1$$

and

$$\eta = \frac{(\log x)^C}{\sqrt{x}|I|}, \quad \text{so } |\alpha_1| \leqslant \eta, \ |\alpha_2 + \alpha_1 R| < \eta |I|.$$

We detect the congruence $\ell_2 \equiv \omega \ell_1 \pmod{2\Delta h}$ in (11.3) by means of Dirichlet characters $\chi \pmod{2\Delta h}$ getting

$$|H(\alpha_1, \alpha_2)| \leqslant \frac{1}{\varphi(2\Delta h)} \sum_{\chi \neq \chi_0} \left| \sum_{\ell} \gamma_\ell \chi(\ell) e(\alpha_1 \ell) \right| \left| \sum_{\ell} \gamma_\ell \chi(\ell) e(-\alpha_2 \ell) \right|. \tag{12.1}$$

Hence, by the Cauchy–Schwarz inequality

$$\begin{aligned} \mathcal{R}'_h(\mathfrak{m}_1, \mathfrak{m}_2) &= I \iint_S FH \ll |I| x \iint_S |H| \\ &\leqslant |I| \frac{x}{\varphi(2\Delta h)} \left(\sum_\chi \iint_S \left| \sum_\ell \gamma_\ell \chi(\ell) e(\alpha_1 \ell) \right|^2 \right)^{\frac{1}{2}} \\ &\quad \times \left(\sum_\chi \iint_S \left| \sum_\ell \gamma_\ell \chi(\ell) e(-\alpha_2 \ell) \right|^2 \right)^{\frac{1}{2}}. \end{aligned}$$

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

$$\sum_\chi \iint_S \Big|\sum_\ell\Big|^2 \leqslant 2\eta|I| \sum_\chi \int_{|\alpha|<\eta} \Big|\sum_\ell \gamma_\ell \chi(\ell) e(\alpha\ell)\Big|^2 d\alpha.$$

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

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

We get the bound

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

From the second sum of integrals we get

$$\sum_\chi \iint_S \Big|\sum_\ell\Big|^2 \leqslant |S|\varphi(2\Delta h) \sum_{\ell_1 \equiv \ell_2 (2\Delta h)}\sum |\gamma_{\ell_1}\gamma_{\ell_2}| \ll \frac{(\log x)^{2C}}{|I|}.$$

Multiplying both estimates we conclude that

$$\mathcal{R}'_h(\mathfrak{m}_1, \mathfrak{m}_2) \ll \frac{|I|^{\frac{1}{2}}}{\varphi(\Delta h)} x (\log x)^{\frac{3C}{2}}.$$

This bound is better than (11.1) for $R''_h(\mathfrak{m}_1, \mathfrak{m}_2)$ if

$$|I| \leqslant (\log x)^{-5C}. \tag{12.2}$$

Therefore we are done in this case.

13. Estimation of $\mathcal{R}'_h(\mathfrak{m}_1, \mathfrak{m}_2)$ on average

In most cases

$$I = \operatorname{Im} \frac{\mathfrak{m}_1}{\mathfrak{m}_2} = -\frac{\Delta}{m_2} \asymp \frac{|\Delta|}{M}$$

is not smaller than (12.2). Assuming I does not satisfy (12.2), we give a better treatment of $\mathcal{R}'_h(\mathfrak{m}_1, \mathfrak{m}_2)$ using the Siegel–Walfisz condition and the large sieve inequality.

We begin by removing the twists by additive characters from the multiplicative character sum (12.1). 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

$$e(\alpha\ell) = 1 + 2\pi i\alpha \int_0^\ell e(\alpha t)\, dt \tag{13.1}$$

to the sums over ℓ in (12.1) getting

$$|H(\alpha_1,\alpha_2)| \leqslant (1+2\pi|\alpha_1|\sqrt{x})(1+2\pi|\alpha_2|\sqrt{x})G(t_1,t_2)$$

with

$$G(t_1,t_2) = \frac{1}{\varphi(\Delta h)} \sum_{\chi\neq\chi_0} \Big|\sum_{t_1<\ell<\sqrt{x}} \gamma_\ell\chi(\ell)\Big|\Big|\sum_{t_2<\ell<\sqrt{x}} \gamma_\ell\chi(\ell)\Big|$$

for some $0 < t_1, t_2 < \sqrt{x}$. The loss is not large because, for (α_1,α_2) in S,

$$(1+2\pi|\alpha_1|\sqrt{x})(1+2\pi|\alpha_2|\sqrt{x}) \ll I^{-2}(\log x)^{2C} \leqslant (\log x)^{12C}.$$

Integrating this over S against $F(s(\alpha_1,\alpha_2)) \ll x$, we conclude by (11.2) that

$$\mathcal{R}'_h(\mathfrak{m}_1,\mathfrak{m}_2) \ll G(t_1,t_2)(\log x)^{14C}. \tag{13.2}$$

Remarks 13.1. The cropping parameters t_1 and t_2 come from integration in the expression (13.1). 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 $\mathfrak{m}_1$, $\mathfrak{m}_2$, h. To simplify the presentation we accept (13.2) having t_1, t_2 independent of $\mathfrak{m}_1$, $\mathfrak{m}_2$, h. By $2G(t_1,t_2) \leqslant G(t_1,t_1) + G(t_2,t_2)$, we arrive at

$$\mathcal{R}'_h(\mathfrak{m}_1,\mathfrak{m}_2) \ll (\log x)^{14C} \frac{1}{\varphi(2\Delta h)} \sum_{\chi\neq\chi_0} |\mathcal{L}(\chi)|^2$$

with

$$\mathcal{L}(\chi) = \sum_{t<\ell<\sqrt{x}} \chi(\ell)\gamma_\ell$$

for $t = t_1$ or $t = t_2$.

We need to sum $\mathcal{R}'_h(\mathfrak{m}_1,\mathfrak{m}_2)$ over $\mathfrak{m}_1$, $\mathfrak{m}_2$ as in (7.6) and over h as in (7.4) subject to the condition $\Delta = \Delta(\mathfrak{m}_1,\mathfrak{m}_2) \equiv 0 \pmod{2h}$, see (9.10). The total contribution of $\mathcal{R}'_h(\mathfrak{m}_1,\mathfrak{m}_2)$ to $\mathcal{C}(M)$ is bounded by $\mathcal{R}(M)(\log x)^{14C}$, where

$$\mathcal{R}(M) = \sum_h |\lambda_h| h \sum_{\substack{(\mathfrak{m}_1,\mathfrak{m}_2,2h)=1 \\ 0\neq\Delta(\mathfrak{m}_1,\mathfrak{m}_2)\equiv 0(2h)}}\sum \frac{|\mu(m_1)\mu(m_2)|}{\varphi(2\Delta h)} \sum_{\substack{\chi \pmod{2\Delta h} \\ \chi\neq\chi_0}} |\mathcal{L}(\chi)|^2.$$

Recall that $m_1 \sim M$, $m_2 \sim M$ and $\mathfrak{m}_1$, $\mathfrak{m}_2$ are primitive. The determinant Δ occurs with certain multiplicity which is bounded by $8M$, so

$$\mathcal{R}(M) \ll M \sum_{h<y} \frac{|\lambda_h|}{\varphi(h)} \sum_{hr<4M} \frac{1}{\varphi(r)} \sum_{\substack{\chi \pmod{rh^2} \\ \chi\neq\chi_0}} |\mathcal{L}(\chi)|^2.$$

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

$$\mathcal{R}(M) \ll M \sum_{1<q\leqslant Q} c(q) \sideset{}{^*}\sum_{\chi_1 \pmod{q}} |\mathcal{L}(\chi_1)|^2,$$

where $Q = 8My$ and

$$c(q) \ll \sum_{h<y} \frac{|\lambda_h|}{\varphi(h)} \sum_{\substack{r<8M\\ rh^2\equiv 0(q)}} \frac{1}{\varphi(r)} \ll \frac{\log M}{\varphi(q)} \sum_{h<y} \frac{|\lambda_h|}{\varphi(h)}(q,h^2)$$
$$\ll \tau(q)^2 q^{-1} \min(\sqrt{q}, y)(\log M)^2.$$

Hence

$$\mathcal{R}(M) \ll M(\log M)^2 \sum_{1<q\leqslant Q} q^{\varepsilon-1} \min(\sqrt{q}, y) \sideset{}{^*}\sum_{\chi_1 \,(\mathrm{mod}\, q)} |\mathcal{L}(\chi_1)|^2.$$

Using the Siegel–Walfisz condition for small q and the large sieve inequality for larger q, we get

$$\mathcal{R}(M) \ll Mx(\log x)^{-B} \tag{13.3}$$

with any $B \geqslant 2$, provided $Q \min(\sqrt{Q}, y) < x^{\frac{1}{2}-\varepsilon}$. Hence (13.3) holds if

$$yz \leqslant x^{\frac{1}{4}-\varepsilon}. \tag{13.4}$$

Finally, the total contribution of the error terms $\mathcal{R}_h(\mathfrak{m}_1, \mathfrak{m}_2)$ to $\mathcal{C}(M)$ is bounded by

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

This bound satisfies our requirement (7.5) if we take B large.

Every bound obtained so far satisfies our requirements subject to conditions (6.3) and (13.4). It remains to estimate the contribution of the main terms $\mathcal{E}_h(\mathfrak{m}_1, \mathfrak{m}_2)$ to $\mathcal{C}_h(M)$ on average over h, see (9.14), (7.4), (7.5). 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 $\mathcal{E}_h(\mathfrak{m}_1, \mathfrak{m}_2)$ are defined by (9.14) and we need to estimate the sums

$$\mathcal{F}_h(M) = h \sum_{\substack{(\mathfrak{m}_1\mathfrak{m}_2, 2h)=1\\ 0\neq\Delta(\mathfrak{m}_1,\mathfrak{m}_2)\equiv 0(2h)}}\sum \mu(m_1)\mu(m_2)\mathcal{E}_h(\mathfrak{m}_1, \mathfrak{m}_2) \tag{14.1}$$

and

$$\mathcal{F}(M) = \sum_{h\leqslant y} |\lambda_h| \mathcal{F}_h(M).$$

Our goal is to show that

$$\mathcal{F}(M) \ll NM^2(\log M)^{-B} \tag{14.2}$$

with any $B \geqslant 2$, which bound is fine for the requirement (7.5).

In this section we make preparations for the application of tools in the next two sections. First it helps to execute the summation over ℓ_1, ℓ_2 in (9.14). To this end, we exploit our assumption (1.7) for $q = 1$, that is the PNT for the coefficients γ_ℓ.

Let us check that the restrictions (10.8) are redundant. Indeed, from the support of f and n given by (9.13) we get

$$m_2 n = \ell_2^2 + \left(\frac{\ell_1 - \ell_2 R}{I}\right)^2 < x,$$

hence $\ell_2 < \sqrt{x}$. Interchanging ℓ_1, ℓ_2 and $\mathfrak{m}_1, \mathfrak{m}_2$, we get a similar formula for $m_1 n$, hence $\ell_1 < \sqrt{x}$.

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

$$x_1 \frac{\partial}{\partial x_1} f(x_1, x_2) \ll 1, \quad x_2 \frac{\partial}{\partial x_2} f(x_1, x_2) \ll 1. \tag{14.3}$$

To this end, we compute as follows:

$$\begin{aligned}\frac{\partial}{\partial x_1} f(m_2 n(x_1, x_2)) &= \frac{\partial}{\partial x_1} f\left(x_2^2 + \left(\frac{x_1 - x_2 R}{I}\right)^2\right) \\ &= 2\frac{x_1 - x_2 R}{I} f'\left(x_2^2 + \left(\frac{x_1 - x_2 R}{I}\right)^2\right) \ll \sqrt{x} x^{-1}.\end{aligned}$$

Hence

$$x_1 \frac{\partial}{\partial x_1} f(m_2 n(x_1, x_2)) \ll \frac{x_1}{\sqrt{x}} \ll 1.$$

Similarly for $f(m_1 n(x_1, x_2))$ and for the partial derivatives with respect to x_2. Hence, (14.3) holds.

Using the Prime Number Theorem by partial summation (9.14) yields

$$\varphi(2\Delta h)\mathcal{E}_h(\mathfrak{m}_1, \mathfrak{m}_2) = \iint f(x_1, x_2)\, dx_1 dx_2 + O(x(\log x)^{-B})$$

with any $B \geqslant 2$. Here the integral is just the Fourier transform $g(\alpha_1, \alpha_2)$ at the point $(\alpha_1, \alpha_2) = (0, 0)$, see (10.2). Then (10.6) and (10.5) yield

$$\begin{aligned} g(0,0) &= IF(s(0,0)) = IF(0), \\ F(0) &= \pi \int_0^\infty J_0(0) f(t) \overline{f}\left(\frac{t m_1}{m_2}\right) dt = \pi m_2 \int_0^\infty f(t m_2) \overline{f}(t m_1)\, dt, \\ m_2 I &= m_2 \operatorname{Im} \frac{\mathfrak{m}_1}{\mathfrak{m}_2} = \operatorname{Im} \mathfrak{m}_1 \overline{\mathfrak{m}}_2 = -\Delta. \end{aligned}$$

Combining these results, we obtain

$$\mathcal{E}_h(\mathfrak{m}_1, \mathfrak{m}_2) = -\frac{\pi\Delta}{\varphi(2\Delta h)} \int_0^\infty f(t m_2) \overline{f}(t m_1)\, dt + O\left(\frac{x}{\varphi(\Delta h)} (\log x)^{-B}\right). \tag{14.4}$$

Inserting (14.4) into (14.1), we get (note that $\varphi(2\Delta h) = 2h\varphi(\Delta)$)

$$\mathcal{F}_h(M) = -\frac{\pi}{2} \int_0^\infty \mathcal{K}_h(t)\, dt + O\left(\frac{x}{(\log x)^B} \sum_{\substack{(\mathfrak{m}_1 \mathfrak{m}_2, 2h)=1 \\ 0 \neq \Delta(\mathfrak{m}_1, \mathfrak{m}_2) \equiv 0 (2h)}} \sum \frac{|\mu(m_1)\mu(m_2)|}{\varphi(\Delta)}\right), \tag{14.5}$$

where

$$\mathcal{K}_h(t) = \sum\sum_{\substack{(\mathfrak{m}_1\mathfrak{m}_2,2h)=1 \\ 0\neq\Delta(\mathfrak{m}_1,\mathfrak{m}_2)\equiv 0(2h)}} \mu(m_1)\mu(m_2)f(tm_1)\overline{f}(tm_2)\Delta/\varphi(\Delta).$$

The error term in (14.5) on average over $h \leqslant y$ satisfies the bound (14.2) so we are done with it. The integral in (14.5) is over the segment $\frac{N}{4} < t < N$ so we need to show that

$$\mathcal{K}(t) = \sum_{h\leqslant y} |\lambda_h \mathcal{K}_h(t)| \ll M^2(\log M)^{-B} \tag{14.6}$$

for any $\frac{N}{4} < t < N$ (recall $MN = x$) and any $B \geqslant 3$. Writing

$$\frac{\Delta}{\varphi(\Delta)} = \prod_{p|\Delta}\Big(1-\frac{1}{p}\Big)^{-1} = \frac{2h}{\varphi(2h)} \sum_{\substack{d|\Delta \\ (d,2h)=1}} \frac{\mu^2(d)}{\varphi(d)},$$

we get

$$\mathcal{K}_h(t) = \frac{2h}{\varphi(2h)} \sum_{(d,2h)=1} \frac{\mu^2(d)}{\varphi(d)} \sum\sum_{\substack{(\mathfrak{m}_1\mathfrak{m}_2,2h)=1 \\ 0\neq\Delta(\mathfrak{m}_1,\mathfrak{m}_2)\equiv 0(2dh)}} \mu(m_1)\mu(m_2)f(tm_1)\overline{f}(tm_2). \tag{14.7}$$

The inner sum over $\mathfrak{m}_1$, $\mathfrak{m}_2$ is bounded by $\frac{(8M)^2}{dh}$. Hence the contribution of $d > D$ is $\ll \frac{M^2}{D\varphi(h)}$. Summing over $h \leqslant y$, this does not exceed the bound (14.6), unless

$$d \leqslant (\log M)^{B+1}. \tag{14.8}$$

Assuming (14.8), we can drop the restriction $\Delta(\mathfrak{m}_1,\mathfrak{m}_2) \neq 0$. If $\Delta(\mathfrak{m}_1,\mathfrak{m}_2) = 0$, then $m_1 = m_2$, so these added terms contribute to (14.7) at most $O(M \log M)$ and to (14.6) at most $O(yM \log M)$ which is admissible if

$$y \leqslant M(\log M)^{-B-1}.$$

Writing $\mathfrak{m}_1 = u_1 + iv_1$ and $\mathfrak{m}_2 = u_2 + iv_2$ the congruence $\Delta(\mathfrak{m}_1,\mathfrak{m}_2) \equiv 0(2dh)$ means $u_1v_2 \equiv u_2v_1 \pmod{2dh}$. Hence we have $(2dh, v_1) = (2dh, v_2) = b$, say, because $(u_1, v_1) = (u_2, v_2) = 1$. Put $2dh = bc$, $v_1 = bw_1$, $v_2 = bw_2$ so $(w_1w_2, c) = 1$ and the congruence become $u_1w_2 \equiv u_2w_1 \pmod c$, or equivalently

$$u_1\overline{w}_1 \equiv u_2\overline{w}_2 \pmod c,$$

where $\overline{w} \pmod c$ denotes the multiplicative inverse (not the complex conjugate). Hence (14.7) becomes (up to an admissible error term)

$$\mathcal{K}_h(t) = \frac{2h}{\varphi(2h)} \sum\sum_{\substack{(d,2h)=1 \\ 2dh=bc}} \frac{\mu^2(d)}{\varphi(d)} \sum\sum_{\substack{(\mathfrak{m}_1\mathfrak{m}_2,2h)=1 \\ (w_1w_2,c)=1 \\ u_1\overline{w}_1\equiv u_2\overline{w}_2(c)}} \mu(m_1)\mu(m_2)f(tm_1)\overline{f}(tm_2), \tag{14.9}$$

where $\mathfrak{m}_1 = u_1 + ibw_1$ and $\mathfrak{m}_2 = u_2 + ibw_2$. The inner sum over $\mathfrak{m}_1$, $\mathfrak{m}_2$ 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

$$b \leqslant (\log M)^{B+2}. \tag{14.10}$$

The condition $(\mathfrak{m}_1\mathfrak{m}_2, 2h) = 1$ in the inner sum of (14.9) is equivalent to

$$\Big(\mathfrak{m}_1\mathfrak{m}_2, \frac{c}{(c,d)}\Big) = 1.$$

This is a harmless, but inconvenient condition. We are going to remove it by a cute trick. Let T^* denote the sum over $\mathfrak{m}_1$, $\mathfrak{m}_2$ with the condition $(\mathfrak{m}_1\mathfrak{m}_2, \frac{c}{(c,d)}) = 1$ and T the sum without this condition. We show that

$$0 \leqslant T^* \leqslant T. \tag{14.11}$$

Proof. Recall that the congruence $u_1\overline{w}_1 \equiv u_2\overline{w}_2 \pmod c$ implies

$$\mathfrak{m}_1\overline{\mathfrak{m}}_2 \equiv \overline{\mathfrak{m}}_1\mathfrak{m}_2 \pmod c.$$

Hence the condition $(\mathfrak{m}_1\mathfrak{m}_2, \frac{c}{(c,d)}) = 1$ is equivalent to $(\mathfrak{m}_2, \frac{c}{(c,d)}) = 1$, because $\mathfrak{m}_1$, $\mathfrak{m}_2$ are odd primitive. Hence

$$\begin{aligned} T^* &= \sum\sum_{(\mathfrak{m}_1\mathfrak{m}_2, \frac{c}{(c,d)})=1} = \sum_{\mathfrak{m}_1}\sum_{(\mathfrak{m}_2, \frac{c}{(c,d)})=1} \\ &= \frac{1}{c}\sum_{a \pmod c}\Bigg(\sum_{\substack{\mathfrak{m}_1 = u_1 + ibw_1 \\ (w_1,c)=1}} \mu(m_1) f(tm_1) e\Big(\frac{a}{c}u_1\overline{w}_1\Big)\Bigg) \\ &\quad\times\Bigg(\sum_{\substack{\mathfrak{m}_2 = u_2 + ibw_2 \\ (w_2,c)=(\mathfrak{m}_2, \frac{c}{(c,d)})=1}} \mu(m_1)\overline{f}(tm_2) e\Big(-\frac{a}{c}u_2\overline{w}_2\Big)\Bigg). \end{aligned}$$

By Cauchy's inequality, $T^* \leqslant T^{\frac{1}{2}}(T^*)^{\frac{1}{2}}$, hence (14.11) holds. ■

By the above considerations we derive the following inequality:

$$\mathcal{K}_h(t) \leqslant \frac{2h}{\varphi(2h)}\sum{}'\sum_{\substack{(d,2h)=1 \\ 2dh=bc}}{}' \frac{\varphi(c)}{c\varphi(d)} T(b,c), \tag{14.12}$$

where

$$T(b,c) = \frac{1}{\varphi(c)}\sum_{a \pmod c}\Bigg|\sum_{\substack{\mathfrak{m}=u+ibw \\ (w,c)=1}} \mu(m) f(tm) e\Big(\frac{a}{c}u\overline{w}\Big)\Bigg|^2$$

and the sums $\sum'\sum'$ are restricted by the conditions in (14.8) and (14.10). Moreover, we dropped out of (14.12) a few parts which we already showed to be admissible for the goal (14.6). Here we have $2h\varphi(c)/\varphi(2h)c\varphi(d) = b\varphi(c)/d\varphi(bc) \leqslant b/d\varphi(b)$ so

$$\mathcal{K}_h(t) \leqslant \sum_{bc \leqslant Q}{}'\sum \frac{b}{\varphi(b)}\Big(\sum_{2dh=bc}{}' d^{-1}\Big) T(b,c),$$

where $Q = 2y(\log M)^{B+1}$. Hence

$$\mathcal{K}(t) \leqslant (\log M) \sum_{bc\leqslant Q}{}' \sum \frac{b}{\varphi(b)} T(b, c).$$

Writing $\frac{a}{c}$ in the lowest terms, we get

$$\mathcal{K}(t) \leqslant (\log M) \sum{}' \sum_{bqr\leqslant Q} \sum \frac{b}{\varphi(b)\varphi(r)} T(b, q, r), \tag{14.13}$$

where

$$T(b, q, r) = \frac{1}{\varphi(q)} \sum_{a \pmod q}{}^{*} \Bigg| \sum_{\substack{\mathfrak{m}=u+ibw \\ (w,qr)=1}} \mu(m) f(tm) e\left(\frac{a}{q} u\overline{w}\right) \Bigg|^2. \tag{14.14}$$

15. Small moduli

We can estimate the sum over $\mathfrak{m} = u + ibw$ in (14.14) using the following Siegel–Walfisz theorem in the Gaussian domain. See [1, Lemma 5] or [3, Lemma 16.1] and the references therein.

Lemma 15.1. *Let $\ell \geqslant 1$ and $\omega \in \mathbb{Z}[i]$. For $x \geqslant 2$ we have*

$$\sum_{\substack{\mathfrak{m}\equiv\omega \pmod \ell \\ m\leqslant x}} \mu(\mathfrak{m}) \ll x(\log x)^{-B_1} \tag{15.1}$$

with any $B_1 \geqslant 1$, 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 Möbius formula and apply (15.1) as follows:

$$\begin{aligned}
\Bigg| \sum_{\substack{\mathfrak{m}=u+ibw \\ (w,qr)=1}} \mu(m) f(tm) e\left(\frac{a}{q} u\overline{w}\right) \Bigg| &\leqslant \sum_{\substack{k\mid r \\ (k,q)=1}} \Bigg| \sum_{\substack{\mathfrak{m}=u+ibkw \\ (w,q)=1}} \mu(m) f(tm) e\left(\frac{a}{q} u\overline{kw}\right) \Bigg| \\
&= \sum_{\substack{k\mid r,\, k\leqslant K \\ (k,q)=1}} \Bigg| \sum_{\mathfrak{m}} \Bigg| + O\left(\tau(r)\frac{M}{bK}\right) \\
&\leqslant \sum_{\substack{k\mid r \\ k\leqslant K}} \sum_{\substack{\alpha,\beta \pmod{bkq} \\ \beta\equiv 0 (bk)}} \sum \Bigg| \sum_{\mathfrak{m}\equiv\alpha+i\beta \pmod{bkq}} \mu(m) f(tm) \Bigg| \\
&\quad + O\left(\tau(r)\frac{M}{bK}\right) \\
&\ll \tau(r) bKq^2 M(\log M)^{-8B} + \frac{\tau(r)M}{bK} \\
&= 2\tau(r) qM(\log M)^{-4B}
\end{aligned}$$

for K with $bKq = (\log M)^{4B}$. Hence

$$T(b,q,r) \ll (\tau(r)qM)^2 (\log M)^{-8B},$$

and the partial sum of (14.13) with $q \leqslant Q_0$, say $\mathcal{K}(q \leqslant Q_0)$, satisfies

$$\mathcal{K}(q \leqslant Q_0) \ll Q_0^3 M^2 (\log M)^{7-7B}.$$

This bound satisfies (14.6) if

$$Q_0 = (\log M)^{2B-3}.$$

16. Large moduli

It remains to estimate the partial sums of (14.13) with $Q_1 < q \leqslant 2Q_1$, say $\mathcal{K}(q \sim Q_1)$, for $Q_0 \leqslant Q_1 \leqslant \frac{Q}{2}$, i.e.

$$(\log M)^{2B-3} \leqslant Q_1 \leqslant y(\log M)^{B+1}.$$

In this range we no longer need help from the Möbius function $\mu(m)$; the cancellation is due to the variation of $e(\frac{a}{q}u\overline{w})$. We need saving a bit larger than the size of the conductor q so the saving from averaging over the classes $a \pmod q$ (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 elements $\mathfrak{m}$ of the form $\mathfrak{m} = u + ibw$ with $m \sim M$ the first coordinate u runs over the segment $|u| \leqslant \sqrt{2M}$ 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[1] by estimating (14.14) as follows:

$$T(b,q,r) \leqslant \frac{4\sqrt{M}}{b\varphi(q)} \sum_{|w|<\sqrt{2M}b^{-1}} \sideset{}{^*}\sum_{a(q)} \Big| \sum_{\mathfrak{m}=u+ibw} \mu(m) f(tm) e\Big(\frac{au}{q}\Big)\Big|^2.$$

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

$$\sum_{q\sim Q_1} \frac{1}{\varphi(q)} \sideset{}{^*}\sum_{a(q)} \Big|\sum_{\mathfrak{m}}\Big|^2 \ll \Big(Q_1 + \frac{\sqrt{M}}{Q_1}\Big)\sqrt{M} \leqslant \frac{2M}{Q_0}$$

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

$$y \leqslant \sqrt{M}(\log M)^{-3B}. \tag{16.1}$$

[1]This aspect of the estimation leads to a problem of independent interest and has been further developed in [5]. However, it does not on its own lead to a sharpening of the main results of this paper.

Recall that M satisfies (7.3). Hence

$$\sum_{q\sim Q_1} T(b,q,r) \ll \frac{M^2}{b^2 Q_0}$$

and

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

This is sufficient for (14.6) if $B \geqslant 3$.

17. Proof of Theorem 1.4. Conclusion

Putting together the results of Sections 6–16, we complete the proof of (6.5) and of Theorem 1.4 (see (6.4)) under the following conditions:

$$\begin{aligned} y^2(\log x)^{2A+4} &\leqslant z \leqslant x^{\frac{1}{8}} && \text{see (6.3)},\\ y(\log x)^{11A+40} &\leqslant z && \text{see (8.1)},\\ yz &\leqslant x^{\frac{1}{4}-\varepsilon} && \text{see (13.4)},\\ y &< z^{\frac{1}{2}}(\log x)^{-3B-2A} && \text{see (16.1)}. \end{aligned}$$

The choice $z = x^{\frac{1}{6}}$ and $y = x^{\theta}$ with any $\theta < \frac{1}{12}$ is good. This completes the proof of (1.14).

18. Derivation of Theorem 1.3

It is not hard to derive Theorem 1.3 from Theorem 1.4 simply by subdividing the range $1 \leqslant t \leqslant x$ into dyadic segments

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

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 (1.8) if

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

The resulting smooth function $f(t)$ supported in a given dyadic segment is $f(t) = 1$, except for t in the short intervals adjacent to the end points where $t^j f^{(j)}(t) \ll \delta^{-j}$. Because we require only $j = 0, 1, 2$, condition (1.12) can be secured by resizing $f(t)$ by a factor δ^2. This factor does not ruin (1.14), because we can use (1.14) with A replaced by $3A + 8$.

19. Derivation of Theorem 1.2

We derive the Almost Primes Theorem 1.2 from the Main Theorem 1.3 by applying the Almost-Prime Sieve from [4, Chapter 25] to the sequence $\mathcal{C} = (c_k)$, $1 \leqslant k \leqslant K = \sqrt{x}$, with

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

We have

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

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

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

and we derive by (1.8) with some $\lambda_h = \pm 1$ that

$$\begin{aligned}\sum_{h \leqslant y} |r_h| &= \sum_{h \leqslant y} \lambda_h r_h \\ &= \sum\sum_{4k^2+\ell^2 \leqslant x} \beta_k \Lambda(\ell)\Lambda(4k^2+\ell^2) - X \sum_{h \leqslant y} \lambda_h g(h) \ll x(\log x)^{-A}.\end{aligned}$$

In other words, speaking the language of sieve theory, our sequence $\mathcal{C} = (c_k)$ has the absolute level of distribution y and the density function $g(h)$ satisfies [4, linear sieve condition (5.38)]. Therefore, [4, Theorem 25.1] is applicable giving

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

with $z = y^{\frac{1}{4}}$, subject to [4, condition (25.25)]. In our situation this condition reads

$$y > K^{\varepsilon + \frac{1}{\Delta_r}},$$

that is $\Delta_r > \frac{1}{2\theta}$. Since $\Delta_r > r + 1 - \frac{\log 4}{\log 3}$ (see [4, condition (25.24)]) and θ is any number $< \frac{1}{12}$, we are fine with $r = 7$. This completes the proof of Theorem 1.2 and hence of Theorem 1.1.

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 $\mathcal{A} = (a_n)$, with

$$a_n = \sum\sum_{\substack{4k^2+\ell^2=n \\ 1 \leqslant k,\ell \leqslant x}} \Lambda(k)\Lambda(\ell).$$

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

$$\begin{aligned} A_d &= \sum_{n\equiv 0 \pmod d} a_n \\ &= \sum_{\nu^2+1\equiv 0(d)} \sum_{\substack{\ell\equiv 2\nu k(d)\\(\ell k,d)=1}} \Lambda(k)\Lambda(\ell) \\ &= \sum_{\nu^2+1\equiv 0(d)} \sum_{a(d)}{}^{*} \psi(x;d,2\nu a)\psi(x;d,a) + O((\log x)^6) \\ &= \sum_{\nu^2+1\equiv 0(d)} \sum_{a(d)}{}^{*} \left(\frac{\psi(x)}{\varphi(d)} + E(x;d,2\nu a)\right)\left(\frac{\psi(x)}{\varphi(d)} + E(x;d,a)\right) + O((\log x)^6) \\ &= \frac{\rho(d)}{\varphi(d)}\psi(x)^2 + \sum_{\nu^2+1\equiv 0(d)} \sum_{a(d)}{}^{*} E(x;d,2\nu a)E(x;d,a) + O((\log x)^6), \end{aligned}$$

where, as we recall, $E(x;d,a)$ is the error term in the prime number theorem for that arithmetic progression (note that, up to the existing error term, the cross terms disappear after summation over the residue classes $a \pmod d$ $(a,d)=1$) and $\rho(d)$ is the number of roots of $\nu^2+1\equiv 0 \pmod d$. Put

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

Then

$$\begin{aligned} |r_d| &\leqslant \rho(d)\sum_{a(d)}{}^{*} |E(x;d,a)|^2 + O((\log x)^6) \\ &\ll \frac{\rho(d)}{\varphi(d)}\, x \sum_{a(d)}{}^{*} |E(x;d,a)| + (\log x)^6. \end{aligned}$$

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

$$\begin{aligned} R(D) &= \sum_{d\leqslant D} |r_d| \\ &\ll x(\log x)\left(\sum_{d\leqslant D}\sum_{a(d)}{}^{*} |E(x;d,a)|^2\right)^{\frac12} + D(\log x)^6 \\ &\ll x^2(\log x)^{-A} \end{aligned}$$

by the Barban–Davenport–Halberstam Theorem (see [4, (9.75)]), where A is any positive number and $D = x^2(\log x)^{-B}$ with some $B = B(A)$. Therefore the sequence $\mathcal{A} = (a_n)$ is supported on $n \leqslant N = 5x^2$, it satisfies the linear sieve conditions and it has level of distribution $D \asymp N^{\frac12}(\log N)^{-B}$. Now, just about any sieve, such as for example [4, Theorem 6.9], gives the upper bound claimed in the proposition. Since also

$$\Delta_3 > 4 - \frac{\log 4}{\log 3} > 2,$$

it follows from [4, Theorem 25.1] that the lower bound in the proposition holds and specifically

$$\sum_{\substack{\omega(n)\leqslant 3\\(n,P(D^{\frac14}))=1}} a_n \asymp x^2(\log x)^{-1},$$

which implies the proposition. ∎

We conclude the paper with heuristics supporting formula (1.1). If $r \geqslant 2$, we use Bombieri's sieve in [4, Theorem 3.5] showing that (1.1) holds with the constant

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

Recall that κ is given by (1.10), hence c is given by (1.2). Of course, this result is conditional subject to the assumption that the sequence $\mathcal{C} = (c_k)$ has exponent of distribution as large as 1, meaning (1.8) holds for $y = x^\theta$ with any $\theta < \frac12$.

If $r = 1$, we write

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

and apply (1.8). For $r = 1$ Bombieri's sieve gives no help so we simply ignore that (1.8) 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

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

Acknowledgments. Henryk Iwaniec would like to thank the University of Toronto for hospitality and support for his visit in May 2018 during which the work on this paper was begun. We want to thank the referee for a careful reading and for catching imperfections here and there.

Funding. Research of John Friedlander was supported in part by NSERC Grant A5123. Research of Henryk Iwaniec was supported in part by NSF Grant DMS-1406981.

References

[1] Fogels, E.: On the distribution of prime ideals. Acta Arith. **7**, 255–269 (1961/62) Zbl 0107.26603 MR 161849

[2] Fouvry, E., Iwaniec, H.: Gaussian primes. Acta Arith. **79**, 249–287 (1997) Zbl 0881.11070 MR 1438827

[3] Friedlander, J., Iwaniec, H.: The polynomial $X^2 + Y^4$ captures its primes. Ann. of Math. (2) **148**, 945–1040 (1998) Zbl 0926.11068 MR 1670065

[4] Friedlander, J., Iwaniec, H.: Opera de Cribro. Amer. Math. Soc. Colloq. Publ. 57, American Mathematical Society, Providence (2010) Zbl 1226.11099 MR 2647984

[5] Friedlander, J., Iwaniec H.: Sums over vanishing determinants. Proceedings of the Second Symposium on Analytic Number Theory (Cetraro, 2019), Rivista di Mat. Univ. Parma (to appear), arXiv:1905.03215

[6] Heath-Brown, D. R., Li, X.: Prime values of $a^2 + p^4$. Invent. Math. **208**, 441–499 (2017) Zbl 1405.11124 MR 3639597

[7] Lam, P. C.-H., Schindler, D., Xiao, S. Y.: On prime values of binary quadratic forms with a thin variable. J. Lond. Math. Soc. (2) **102**, 749–772 (2020) Zbl 07288975 MR 4171433

[8] Pratt, K.: Primes from sums of two squares and missing digits. Proc. Lond. Math. Soc. (3) **120**, 770–830 (2020) Zbl 1453.11120 MR 4037831