CRITICAL ZEROS OF LACUNARY L-FUNCTIONS

J.B. CONREY AND H. IWANIEC

Acknowledgments. This project started years ago during numerous visits of the second author to the American Institute of Mathematics and two visits of the first author to Rutgers University. We thank both institutions for these opportunities and their support. Then the work continued when the second author enjoyed a Senior Fellow position at the Institute for Theoretical Studies – ETH Zürich in August 2014 – July 2015. He is happy to acknowledge the superb working conditions and generous support received from ETH-ITS while the project was in progress and completed. We are also grateful for support from the NSF grants DMS 1406981 and DMS 1101574. We thank Corentin Perret for technical help.

1. Introduction

We consider an L-function given by the Euler product

$$L(s) = \prod_p (1 - \lambda(p)p^{-s} + \kappa(p)p^{-2s})^{-1} \tag{1.1}$$

with $|\lambda(p)| \leqslant 2$ and $|\kappa(p)| \leqslant 1$, so the product converges absolutely in the half-plane $\operatorname{Re} s > 1$. Hence $L(s)$ has the absolutely converging Dirichlet series expansion

$$L(s) = \sum_n \lambda(n)n^{-s} \qquad \text{if } \operatorname{Re} s > 1, \tag{1.2}$$

with multiplicative coefficients $\lambda(n)$ which are bounded by the divisor function $\tau(n)$. Moreover, we assume that $L(s)$ admits analytic continuation to the whole complex s-plane and it is holomorphic, expect possibly for a simple pole at $s = 1$. Furthermore, $L(s)$ satisfies a standard functional equation which we write in the following form

$$L(s) = X(s)\overline{L}(1 - s) \tag{1.3}$$

where $\overline{L}(s)$ stands for the L-function with Dirichlet series coefficients complex conjugated and $X(s)$ is called the root factor. Note that $|X(s)| = 1$ if $\operatorname{Re} s = 1/2$. One may consider the equation (1.3) as a definition of $X(s)$. Typically $X(s)$ turns out to be an exponential function times the ratio of one or two gamma functions. We do not need to specify the root factor. For our purpose it suffices to assume that $X(s)$ is holomorphic in the strip $0 \leqslant \operatorname{Re} s < 1$ and it satisfies

$$X(s + z) = X(s)(Q|s|)^{-2z} \left\{1 + O\left(|z|(|s| + |z|)^{-1}\right)\right\} \tag{1.4}$$

if $\operatorname{Re} s = 1/2$ and $-1/4 \leqslant \operatorname{Re} z \leqslant 0$, where the implied constant depends only on the parameters (shifts) in the involved gamma functions. In specific cases (1.4) follows by Stirling's formula.

We say that $L(s)$ is "lacunary" if its coefficients vanish or are quite small frequently. We measure this phenomenon by postulating the following estimate

$$\sum_{Q^4 < n \leqslant Q^{4A}} |\lambda(n)|n^{-1} \leqslant \varepsilon A \tag{1.5}$$

to hold with some $\varepsilon = \varepsilon(Q) > 0$ for all $A \geqslant 1$. Here we think of $\varepsilon = \varepsilon(Q)$ being arbitrarily small as Q gets large (Q^2 is closely related to the conductor of $L(s)$). However, even a fixed $\varepsilon > 0$ but sufficiently small would suffice for nice applications. If (1.5) holds, then we say that the $L(s)$ is "ε-lacunary".

Remarks. The lacunarity condition with ε relatively small reveals that the coefficients $\lambda(n)$ of $L(s)$ appear less often than the primes numbers in the segments $[Q^4, Q^{4A}]$. When it comes to perform some mollification, this property means that one applies sieve of small dimension. But it is hard to believe that such L-functions do exist in reality, therefore our undertaking here is mainly for learning the phenomena and exercising delicate techniques. For instance we estimate the lacunary bilinear form (22.4) without losing

vital savings from sifting effects along the lines (22.6)–(22.10). We are mostly interested in special L-functions, nevertheless we set the above introduction in some generality, because it exposes the "exceptional characters" at work more clearly than the roundabout argument with "exceptional zeros".

Our primary source of lacunary L-functions is the quadratic field $K = \mathbb{Q}(\sqrt{D})$ of discriminant D. Let $\psi : \mathcal{C}l(K) \to \mathbb{C}$ be a character of the ideal class group of K. There are $h(D) = |\mathcal{C}l(K)|$ such characters. For each of these we have the L-function

$$L(s,\psi) = \sum_{\mathfrak{a}} \psi(\mathfrak{a})(N\mathfrak{a})^{-s} = \prod_{\mathfrak{p}} (1 - \psi(\mathfrak{p})(N\mathfrak{p})^{-s})^{-1}$$

which satisfies our conditions with

$$\lambda(n) = \sum_{N\mathfrak{a}=n} \psi(\mathfrak{a}) \tag{1.6}$$

and

$$Q = \sqrt{|D|}/2\pi. \tag{1.7}$$

In this case, the root number (the sign of the functional equation) is 1 and the root factor $X(s)$ is equal to $Q^{1-2s}\Gamma(1-s)/\Gamma(s)$ if $D < 0$ and $Q^{1-2s}\Gamma^2(\frac{1-s}{2})/\Gamma^2(\frac{s}{2})$ if $D > 0$.

The coefficients of $L(s,\psi)$ are bounded by the coefficients of $L(s,\psi_0)$ for the trivial character $\psi_0 = 1$. By the factorization $L(s,\psi_0) = \zeta(s)L(s,\chi)$, where $\chi \pmod{|D|}$ is the Dirichlet real character (given by the Kronecker symbol associated with the field $K = \mathbb{Q}(\sqrt{D})$) one sees that the coefficients of $L(s,\psi_0)$ are

$$\lambda_0(n) = (1 \star \chi)(n) = \sum_{d|n} \chi(d). \tag{1.8}$$

Clearly, $|\lambda(n)| \leqslant \lambda_0(n) \leqslant \tau(n)$. Moreover, we have (cf. (22.109) of [IK04])

$$\sum_{n\leqslant N} \lambda_0(n)n^{-1} = L(1,\chi)(\log N + \gamma) + L'(1,\chi) + O\left(|D|^{1/4}N^{-1/2}\log 2N\right)$$

which formula implies the following bound

$$\sum_{Q^4<n\leqslant N} \lambda_0(n)n^{-1} \leqslant L(1,\chi)\log N \tag{1.9}$$

provided $|D| = (2\pi Q)^2$ is sufficiently large.

Definition 1.1. We say that the character $\chi \pmod{|D|}$ is "ε-exceptional" if

$$L(1,\chi)\log|D| \leqslant \varepsilon. \tag{1.10}$$

Then, the corresponding discriminant D of the field $K = \mathbb{Q}(\sqrt{D})$ is called "ε-exceptional".

By (1.9) and (1.10) one infers (1.5). Therefore, if the discriminant D is "ε-exceptional", then every $L(s,\psi)$ is ε-lacunary.

We recall some worthy shortcut notations which are used in analytic number theory. First, if f, g are complex-valued functions, then the relation $f \ll g$ means that $|f| \leqslant cg$ holds for all the relevant arguments with certain (implied) constant $c > 0$. Next, the relation $f \asymp g$ means that $f \ll g$ and $g \ll f$ hold. Then $O(g)$ stands for a function (or a quantity) which is $\ll g$. Note that the above relations can hold only if $g \geqslant 0$. For example the statement $\sin x \ll \sin x$ is false. If t, T are real numbers, then the notation $t \sim T$ stands for the inequality $T < t \leqslant 2T$. Occasionally, we shall use the same symbol to denote different things, but the reader should not be confused, because the proper meaning will be clear from the context.

2. Statement of Results

Let $N(T)$ denote the number of zeros $\rho = \beta + i\gamma$ of $L(s)$ (counted with multiplicity) in the rectangle $0 < \beta < 1$, $\gamma \sim T$. By contour integration using the functional equation one derives the formula

$$N(T) = \frac{T}{\pi} \log QT + O(T) \tag{2.1}$$

for any $T \geqslant Q$. Here the dominant term emerges from variation of the argument of the root factor and the implied constant depends only on the gamma parameters.

Let $N_0(T)$ denote the number of zeros $\rho = 1/2 + i\gamma$ of $L(s)$ (counted with multiplicity) in the segment $\gamma \sim T$. Any natural L-function should satisfy the Riemann Hypothesis so $N_0(T)$ should be equal to $N(T)$. For L-functions of degree 1 or 2 it is known that a positive proportion of zeros are on the critical line $\operatorname{Re} s = 1/2$, that is $N_0(T) \asymp N(T)$ for all T with $\log T$ sufficiently larger than $\log Q$. It seems possible to show that if $L(s)$ is ε-lacunary (see the condition (1.5)), then

$$N_0(T) = \{1 + O(\varepsilon^{1/2})\}N(T) + O(T)$$

for all T with $Q^{1/\sqrt{\varepsilon}} \leqslant T \leqslant Q^{1/\varepsilon}$.

For transparency we work out only the case of functions $L(s) = L(s, \psi)$ which are attached to the characters ψ on ideal classes of the quadratic field $K = \mathbb{Q}(\sqrt{D})$. In greater generality as described in Section 1 the arguments should be very similar. Our main result is the following

Theorem 2.1. *Let $N_{00}(T)$ denote the number of simple zeros $\rho = 1/2 + i\gamma$ of $L(s, \psi)$ with $\gamma \sim T$ and $N(T)$ the number of all zeros $\rho = \beta + i\gamma$ of $L(s, \psi)$ with $0 < \beta < 1$, $\gamma \sim T$ counted with multiplicity. We have*

$$N_{00}(T) = N(T) + O\left(T \log |D| + (L(1, \chi) \log T)^{1/4} T \log T\right) \tag{2.2}$$

where the implied constant is absolute. Putting $\varepsilon = \varepsilon(D) = L(1, \chi) \log |D|$ we get

$$N_{00}(T) = \{1 + O(\delta)\}N(T) \tag{2.3}$$

with any $\delta \geqslant \varepsilon^{1/5}$ and every T with $|D|^{1/\delta} \leqslant T \leqslant |D|^{\delta^4/\varepsilon}$, where the implied constant is absolute.

Remarks. First of all, the approximate formulas (2.2) and (2.3) are unconditional, but of course, (2.3) is meaningful only if $\varepsilon = \varepsilon(D)$ is sufficiently small. This does not hold in reality since the Riemann Hypothesis implies the lower bound $L(1,\chi) \gg 1/\log\log|D|$. But so far the best known bound is $L(1,\chi) \gg |D|^{-\theta}$ with any $\theta > 0$, the result due to C.L. Siegel which is not effective (the implied constant depends on θ and it cannot be computed numerically if $\theta < 1/2$). Therefore, it is still interesting to speculate on the effect of the assumption that $\varepsilon(D) \to 0$ as D varies over some infinite sequence of discriminants no matter how sparse it is.

Definition 2.2. An infinite sequence of discriminants D is called "exceptional" if

$$\varepsilon(D) = L(1,\chi)\log|D| \to 0. \tag{2.4}$$

Corollary 2.3. *As $\varepsilon = \varepsilon(D) \to 0$ over an exceptional sequence, then for every $\psi \in \widehat{\mathcal{C}l}(K)$ the critical simple zeros of $L(s,\psi)$ of height $\sim T$ comprise a 100% of all the zeros of height $\sim T$ for any T with $|D|^{-\log\varepsilon} \leqslant T \leqslant |D|^{-1/\varepsilon\log\varepsilon}$.*

This result sounds more impressive when applied for the trivial ideal class group character. In this case $L(s,\psi_0)$ factors into the Riemann zeta function $\zeta(s)$ and the Dirichlet L-function $L(s,\chi)$. For each factor separately we know the true values (asymptotically correct estimates) for the full numbers of zeros (counted with multiplicity) in the rectangle $s = \sigma + it$ with $0 < \sigma < 1$, $t \sim T$; these are

$$\frac{T}{2\pi}\log T + O(T), \qquad \frac{T}{2\pi}\log|D|T + O(T)$$

respectively. The sum of these values agrees with $N(T)$ (see (2.1)), consequently Theorem 2.1 implies

Corollary 2.4. *Let D run over an exceptional sequence of discriminants so $\varepsilon = \varepsilon(D) = L(1,\chi)\log|D| \to 0$. Choose any T with $|D|^{-\log\varepsilon} \leqslant T \leqslant |D|^{-1/\varepsilon\log\varepsilon}$. Then the number of critical zeros of $\zeta(s)$ of height $\sim T$ which are simple and different from these of $L(s,\chi)$ approaches asymptotically the number of all zeros of $\zeta(s)$ of height $\sim T$.*

If $L(s)$ is a lacunary L-function of degree two, then the twisted L-function

$$L(s;\lambda\chi') = \sum_{n=1}^{\infty}\lambda(n)\chi'(n)n^{-s}$$

by any fixed Dirichlet character χ' is also lacunary of degree two. The arguments presented in this paper for $\chi' = 1$ and $\lambda = 1 \star \chi$, where χ is the real character to exceptional conductor $|D|$, extend easily to $L(s;\lambda\chi') = L(s,\chi')L(s,\chi\chi')$. In particular Corollary 2.4 generalizes to any Dirichlet L-function $L(s,\chi')$ in place of $\zeta(s)$.

If one is willing to assume that

$$L(1,\chi) \ll (\log|D|)^{-2015}$$

for an infinite sequence of discriminants D, then the same results would be achieved much faster by substantially simpler arguments (ignoring the sifting effects of the mollifier in various places). However the arguments are not powerful enough to cover the L-functions of degree larger than two, even if the lacunarity condition is assumed to be extremely strong.

3. Levinson's Method

There are two well established methods for counting zeros of L-functions on the critical line – the Selberg method [Sel42] and the Levinson method [Lev74]. They are diametrically opposite to each other. Selberg's method relays on observing the sign changes of a suitably normalized and mollified L-function as its argument runs over a segment of the critical line. There is no risk of getting negative bound for the counting number, but the method is not perfect for intrinsic reasons; for one that the zeros are not supposed to be almost evenly spaced. Therefore, it needs a sensitive design for counting the zeros adequately (asymptotically precise) when passing through the segments. Yet, it may be the case that under the lacunarity condition the zeros do pretend to be more or less evenly spaced, contrary to the Pair Correlation Conjecture of Montgomery. This question was addressed by R. Heath-Brown during the AIM conference in Seattle of August 1996 (unpublished).

The method of Levinson is risky, because it may produce a negative bound for the counting number of critical zeros if the relevant estimates are crude. On the other hand it opens a possibility for accounting a 100% of the critical zeros if the mollification is nearly perfect. This is indeed the scenario for lacunary L-functions. A far reaching version of Levinson's method has been developed in [Con89], see also the Appendix in [CIS13]. In this section we are going to adopt Proposition A of [CIS13] to our particular context. We shall also borrow numerous arguments developed in [CI02] for handling the off-diagonal terms in Sections 13-17.

Thanks to the lacunarity of $L(s)$ we do not care about delicate choices of the parameters involved in Levinson's original setup. We shall also take numerous advantages of the lacunarity for technical simplifications. In particular the root factor does not play a role (no hassling with cross-terms). Although, the off-diagonal terms do appear, their contribution is nominal, because the lacunarity strikes twice independently. However, by no means one can neglect the off-diagonal terms quickly. For simplicity we shall sacrifice some surplus of the gain, but of course, not everything (see, for example, how we derived the bound (22.9) from the expression (22.6)).

We start with the linear combination of $L(s)$ and its derivative:

$$G(s) = L(s) + L'(s)/\log N \tag{3.1}$$

with $N \geqslant 2$ (the level) to be chosen later, see (4.14). To $G(s)$ we attach a mollifier which is given by a Dirichlet polynomial

$$M(s) = \sum_{m \leqslant M} v(m) m^{-s} \tag{3.2}$$

with coefficients $v(m)$ to be determined later subject to $v(1) = 1$, $|v(m)| \leqslant \tau(m)$. For now we assume that the mollifier $M(s)$ has length $M \leqslant T^{1/2}$, but we shall see that shorter mollifiers do their designated job (to produce sifting effects) pretty well, again due to the lacunarity properties. Putting

$$F(s) = G(s)M(s) - 1 \tag{3.3}$$

we have Levinson's inequality (see Proposition A of [CIS13])

$$\begin{aligned} N_{00}(T) &\geqslant N(T) - \frac{1}{\pi a}\int_T^{2T} \log\left|1 + F\left(\frac{1}{2} - a + it\right)\right| dt + O(T) \\ &\geqslant N(T) - \frac{1}{\pi a} I_a(T) + O(T) \end{aligned} \tag{3.4}$$

where

$$I_a(T) = \int_T^{2T} \left|F\left(\frac{1}{2} - a + it\right)\right| dt.$$

This holds for any $a > 0$ and $T > Q^4$ with absolute implied constant. Since we do not care loosing an absolute constant factor it is possible to replace $I_a(T)$ by

$$I(T) = \int_T^{2T} \left|F\left(\frac{1}{2} + it\right)\right| dt. \tag{3.5}$$

Lemma 3.1. *Let $T \geqslant M^2 \geqslant Q^8$. For $0 < a \leqslant 1/2$ we have*

$$I_a(T) \leqslant T^{4a}(I(T) + O(T^{7/8})) \tag{3.6}$$

where the implied constant is absolute.

Proof. Put

$$H(z) = \frac{2a(4-a^2)}{(z^2-a^2)(z^4-4)} T^{4z}.$$

It is clear that

$$\begin{aligned} \frac{1}{2\pi}\int_{-\infty}^{\infty} |H(iv)|dv = \frac{a}{\pi}\int_{-\infty}^{\infty} \frac{4-a^2}{4+v^2}\frac{dv}{a^2+v^2} &< 1, \\ \frac{1}{2\pi}\int_{-\infty}^{\infty} |H(iv)|(v^2+1)dv &< 4, \\ \int_{-\infty}^{\infty} |H(-1+iv)|(v^2+1)dv &\ll T^{-4}. \end{aligned}$$

Since $H(z)$ has simple pole at $z = -a$ with residue T^{-4a} we get

$$T^{-4a}F(s-a) = \frac{1}{2\pi i}\int_{(0)} F(s+z)H(z)dz - \frac{1}{2\pi i}\int_{(-1)} F(s+z)H(z)dz$$

if $\operatorname{Re} s = 1/2$. Hence $T^{-4a}I_a(T) \leqslant V - W$, say, where

$$\begin{aligned} V &= \frac{1}{2\pi}\int_{-\infty}^{\infty} |H(iv)| \int_T^{2T} \left|F\left(\frac{1}{2} + it + iv\right)\right| dtdv, \\ W &= \frac{1}{2\pi}\int_{-\infty}^{\infty} |H(-1+iv)| \int_T^{2T} \left|F\left(-\frac{1}{2} + it + iv\right)\right| dtdv. \end{aligned}$$

By the convexity bound for $L(s)$ we derive

$$F(s) \ll (MQ|s|)^{1/2}(\log MQ|s|)^2, \qquad \text{if } \operatorname{Re} s = 1/2.$$

Hence

$$\int_T^{T+v} \left|F\left(\frac{1}{2}+it\right)\right| dt \ll (v^2+1)(MQT)^{1/2}(\log T)^2.$$

The same bound holds with T replaced by $2T$. Hence we get

$$V \leqslant \int_T^{2T} \left|F\left(\frac{1}{2}+it\right)\right| dt + O(T^{7/8}).$$

Next, by the functional equation for $L(s)$ and the trivial estimation on the line $\operatorname{Re} s = 3/2$ we derive

$$F(s) \ll (MQ|s|)^2, \qquad \text{if } \operatorname{Re} s = -1/2.$$

Hence

$$W \ll (MQT)^2 \int |H(-1+iv)|(v^2+1)dv \ll 1.$$

This completes the proof of (3.6). □

After inserting (3.6) into (3.4) it is clear that the best choice of the shift is $a = 1/4 \log T$ giving

Lemma 3.2. *Let $T \geqslant M^2 \geqslant Q^8$. Then*

$$N_{00}(T) > N(T) - 4I(T)\log T + O(T) \tag{3.7}$$

where $I(T)$ is given by (3.5) *with $F(s) = G(s)M(s) - 1$ and the implied constant is absolute.*

We are going to show that $I(T)/T$ is small, provided the mollifying factor $M(s)$ is chosen properly. Naturally this suggests that $M(s)$ should pretend to be the inverse of $L(s)$, or slightly better of $G(s)$, but due to the lacunarity it does not matter which one is on the target. Writing

$$L(s)^{-1} = \prod_p (1 - \lambda(p)p^{-s} + \chi(p)p^{-2s}) = \sum_m \rho(m)m^{-s}, \tag{3.8}$$

we get the multiplicative function $\rho(m)$ with

$$\rho(p) = -\lambda(p),\ \rho(p^2) = \chi(p),\ \rho(p^\alpha) = 0 \text{ if } \alpha > 2. \tag{3.9}$$

We take

$$M(s) = \sum_{m \leqslant M} \rho(m)g(m)m^{-s} \tag{3.10}$$

where $g(m)$ is a nice cropping function. For instance

$$g(m) = \left(1 - \frac{\log m}{\log M}\right)^r \tag{3.11}$$

with r a sufficiently large integer will do the job ($r = 32$ is fine). The large degree of vanishing at the end point $m = M$ is necessary for our technique of producing some sifting effects.

Note that $\rho(m)$ is supported on cubefree numbers,

$$|\rho(m)| \leqslant \lambda_0(m) \qquad \text{for all } m \tag{3.12}$$

$$\rho(m) = \mu(m)\lambda(m) \qquad \text{if } m \text{ is squarefree.} \tag{3.13}$$

Remarks. Certain parts of the forthcoming sums are supported on the mollifier terms $m \equiv 0 \pmod{|D|}$, specifically the off-diagonal constituents (16.9). We could easily eliminate these parts right now by restricting (3.10) to $m \not\equiv 0 \pmod{|D|}$. This incomplete mollifier does the job as good as the full one, because it is easy to estimate the missing terms by $|D|^{-1/2}T(\log T)^{2015}$ directly using Cauchy-Schwarz inequality and the mean value estimates for $|G(s)|^2$ and $|M(s)|^2$ (giving up the lacunarity features and the sifting effects). This alteration can be implemented any time so we postpone the issue to the comments in the last section.

4. A Partition of $G(s)$

To apply the mollifier $M(s)$ to $G(s)$ and observe its sifting effects we need to expand $G(s)$ into Dirichlet polynomials. To this end we fix two smooth functions $a(x), b(x)$ on $\mathbb{R}$ with

$$a(x) + b(x) = 1 - x, \tag{4.1}$$

$a(x)$ supported on $x \leqslant \alpha$ and $b(x)$ supported on $x \geqslant \beta$, where $0 < \beta < \alpha < 1$ are fixed numbers (see Figure 1). In applications we shall choose α, β greater but close to $\frac{1}{2}$.

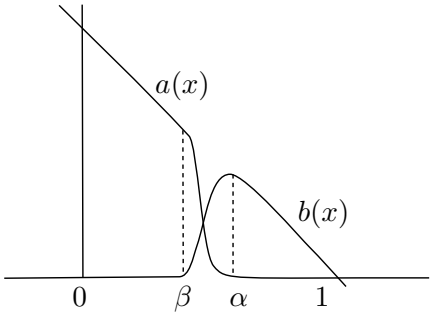

Figure 1.

Given $N \geqslant 2$ consider the function

$$f(z) = z^2 \int_0^\infty a\left(\frac{\log y}{\log N}\right) y^{z-1} dy. \tag{4.2}$$

The integral converges absolutely in $\operatorname{Re} z > 0$. Integrating by parts we get

$$f(z) \log N = \int_\beta^\alpha a''(u) N^{uz} du. \tag{4.3}$$

This expression shows that $f(z)$ is an entire function and its power series expansion begins with

$$f(z) = (\log N)^{-1} + z + \dots \tag{4.4}$$

By Mellin inversion, (4.2) yields

$$a\left(\frac{\log y}{\log N}\right) = \frac{1}{2\pi i}\int_{(1)} f(z)y^{-z}z^{-2}dz, \quad \text{if } y>0. \tag{4.5}$$

Moving to the line $\operatorname{Re} z = -1$ we get by (4.4) and (4.1)

$$b\left(\frac{\log y}{\log N}\right) = \frac{-1}{2\pi i}\int_{(-1)} f(z)y^{-z}z^{-2}dz, \quad \text{if } y>0. \tag{4.6}$$

Now consider the Dirichlet polynomial

$$A(s) = \sum_n a\left(\frac{\log n}{\log N}\right)\lambda(n)n^{-s} \tag{4.7}$$

which has length N^α. Let s be in the critical strip $0 < \operatorname{Re} s < 1$. By (4.5) we derive

$$\begin{aligned} A(s) &= \frac{1}{2\pi i}\int_{(1)} L(s+z)f(z)z^{-2}dz \\ &= \frac{1}{2\pi i}\int_{(-1)} L(s+z)f(z)z^{-2}dz + G(s) + Rf(1-s)(1-s)^{-2} \end{aligned}$$

where the second term $G(s) = L(s) + L'(s)/\log N$ comes from the double pole at $z=0$ and the third term comes from a possible simple pole at $z = 1-s$. In the integral over the line $\operatorname{Re} z = -1$ we apply the functional equation $L(s+z) = X(s+z)\overline{L}(1-s-z)$ with

$$X(s+z) = X(s)(Q|s|)^{-2z}\{1+z^2\eta(s,z)\}, \tag{4.8}$$

say, see (1.3) and (1.4). This integral splits accordingly

$$\frac{1}{2\pi i}\int_{(-1)} L(s+z)f(z)z^{-2}dz = -X(s)B(s) - X(s)R(s)$$

where

$$B(s) = \frac{-1}{2\pi i}\int_{(-1)} \overline{L}(1-s-z)(Q|s|)^{-2z}f(z)z^{-2}dz \tag{4.9}$$

and

$$R(s) = \frac{1}{2\pi i}\int_{(-\varepsilon)} \overline{L}(1-s-z)(Q|s|)^{-2z}f(z)\eta(s,z)dz. \tag{4.10}$$

Note that in $R(s)$ we moved back the integration to the line $\operatorname{Re} z = -\varepsilon$ with $\varepsilon = 1/\log Q|s|$ without passing poles, because $\eta(s,z)$ is holomorphic in z (not in s). Now, by (4.6), the integral (4.9) expands into the series

$$B(s) = \sum_n b\left(\frac{\log Q^2|s|^2/n}{\log N}\right)\overline{\lambda}(n)n^{s-1}. \tag{4.11}$$

We have proved the following formula

Proposition 4.1. *For s in the critical strip $0 < \operatorname{Re} s < 1$ we have*

$$G(s) = A(s) + X(s)B(s) + X(s)R(s) + Rf(1-s)(1-s)^{-2} \tag{4.12}$$

where $A(s), B(s)$ are given by (4.7), (4.11), *respectively, $R(s)$ is the integral* (4.10) *and R denotes the residue of the L-function; $R = L(1,\chi)$ if ψ is trivial and $R = 0$ otherwise.*

If $\operatorname{Re} s = 1/2$, then $R(s)$ can be easily estimated by

$$R(s) \ll Q^{1/2}|s|^{-\frac{1}{12}}. \tag{4.13}$$

This follows by $L(1-s-z) \ll (|s+z|Q)^{1/2}$ (the convexity bound), $f(z)\log N \ll 1$ (see (4.3)) and $\eta(s,z) \ll |z|^{-1}(|s|+|z|)^{-1}$ (see (1.4)).

Assume $T \geqslant Q^8$ and choose the level of Levinson's form (3.1)

$$N = Q^2T^2. \tag{4.14}$$

From now on we let s be in the segment

$$s = \frac{1}{2} + it,\ T \leqslant t \leqslant 2T. \tag{4.15}$$

Then the root factor satisfies $|X(s)| = 1$ and for the polar term we derive by (4.3)

$$f(1-s)(1-s)^{-2} \ll N^{\alpha/2}T^{-2} \leqslant N^{1/2}T^{-2} = QT^{-1}.$$

The residue of $L(s)$ is $R \ll \log Q$. Therefore, Proposition 4.1 yields

Corollary 4.2. *For s in the segment* (4.15) *we have*

$$G(s) = A(s) + X(s)B(s) + O(T^{-1/48}) \tag{4.16}$$

where the implied constant is absolute.

Next we look into coefficients of $B(s)$, they depend on $|s|$ mildly if s is in the segment (4.15). Precisely we have

$$b\left(\frac{\log Q^2|s|^2/n}{\log N}\right) = b\left(1 - \frac{\log n}{\log N} + \delta\right)$$

where

$$\delta = \delta(s) = \frac{2\log|s|/T}{\log N},\ 0 \leqslant \delta \leqslant \frac{\log 4}{\log N}. \tag{4.17}$$

The small shift in the function $b(1-x+\delta)$ is dispensable, it can be isolated by Taylor's expansion

$$b(1-x+\delta) = b(1-x) + \delta b'(1-x) + \frac{1}{2}\delta^2 b''(1-x) + \Delta(s,x),$$

say, with the last term $\Delta(s,x)$ being sufficiently small for easy direct handling (the sifting effect can be ignored). Nevertheless we opt working with one function $b(1-x+\delta)$ contaminated by the shift δ rather than with several derivatives without the shift.

For s on the line $\operatorname{Re} s = 1/2$ we can write the complex conjugate of $B(s)$, see (4.11), in the following fashion

$$\overline{B}(s) = \sum_n b^*\left(\frac{\log n}{\log N}\right)\lambda(n)n^{-s} \tag{4.18}$$

where

$$b^*(x) = b(1-x+\delta). \tag{4.19}$$

Note that $b^*(x)$ is supported on $x \leqslant 1-\beta+\delta$ so $B(s)$ runs over $1 \leqslant n \leqslant 4N^{1-\beta}$. We have

$$b^*(x) = x - \delta, \quad \text{if } x \leqslant 1-\alpha+\delta, \tag{4.20}$$

so the sum (4.18) begins with

$$\sum_{n \leqslant N^{1-\alpha}} \left(\frac{\log n}{\log N} - \delta\right) \lambda(n) n^{-s}. \tag{4.21}$$

On the other hand the sum (4.7) begins with

$$\sum_{n \leqslant N^{\beta}} \left(1 - \frac{\log n}{\log N}\right) \lambda(n) n^{-s}. \tag{4.22}$$

5. Estimating $I(s)$

Multiplying (4.16) by the mollifier (3.10) we obtain the inequality

$$|F(s)| \leqslant |C(s)| + |C^*(s)| + O(|M(s)|T^{-1/48}) \tag{5.1}$$

with $C(s) = A(s)M(s) - 1$ and $C^*(s) = \overline{B}(s)M(s)$ for s in the segment (4.15). The Dirichlet polynomials

$$C(s) = \sum_{1<\ell<MN^{\alpha}} c(\ell)\ell^{-s}, \; C^*(s) = \sum_{1 \leqslant \ell < 4MN^{1-\beta}} c^*(\ell)\ell^{-s}$$

have coefficients given by convolutions; specifically

$$c(\ell) = \sum_{mn=\ell} \rho(m)\lambda(n)g(m)h(n), \qquad h(n) = a\left(\frac{\log n}{\log N}\right), \tag{5.2}$$

$$c^*(\ell) = \sum_{mn=\ell} \rho(m)\lambda(n)g(m)h^*(n), \qquad h^*(n) = b^*\left(\frac{\log n}{\log N}\right), \tag{5.3}$$

and $g(m)$ is the cropping function which we have chosen in (3.11). Inserting the inequality (5.1) into the integral (3.5) we obtain

$$I(T) \leqslant J(T) + J^*(T) + O(T^{47/48}(\log T)^2) \tag{5.4}$$

where

$$J(T) = \int_T^{2T} \left|C\left(\frac{1}{2}+it\right)\right| dt, \quad J^*(T) = \int_T^{2T} \left|C^*\left(\frac{1}{2}+it\right)\right| dt$$

and the error term comes from the classical mean-value theorem for Dirichlet polynomials.

Next we apply the Cauchy-Schwarz inequality and we expand the range of integration getting $J(T)^2 \leqslant T\mathcal{K}(T)$ where

$$\mathcal{K}(T) = \int \Phi\left(\frac{t}{T}\right) \left|C\left(\frac{1}{2}+it\right)\right|^2 dt \tag{5.5}$$

and $\Phi(u)$ is any non-negative smooth function, compactly supported on $\mathbb{R}^+$ with $\Phi(u) \geqslant 1$ in the interval $1 \leqslant u \leqslant 2$. Similarly $J^*(T)^2 \leqslant T\mathcal{K}^*(T)$ where

$$\mathcal{K}^*(T) = \int_T^{2T} \left| C^*\left(\frac{1}{2} + it\right)\right|^2 dt, \tag{5.6}$$

but without smoothing, because it would give no advantage. Therefore we have shown the following inequality

$$I(T) \leqslant (T\mathcal{K}(T))^{\frac{1}{2}} + (T\mathcal{K}^*(T))^{\frac{1}{2}} + O(T^{47/48}(\log T)^2). \tag{5.7}$$

6. The Diagonals

We take β somewhat larger than $\frac{1}{2}$ to make sure that the sum $C^*(s)$ is shorter than T. Specifically $C^*(s)$ has length $\leqslant 4MN^{1-\beta}$ so it is enough to assume that

$$MN^{1-\beta} \leqslant T(\log T)^{-11}. \tag{6.1}$$

Then the classical mean-value theorem for Dirichlet polynomials shows that $\mathcal{K}^*(T)$ is equal to the contribution of the diagonal terms up to a small error term. This claim requires some explanation, because the coefficients of $C^*(s)$ depend on $|s|$. We have

$$\mathcal{K}^*(T) = \sum_{\ell_1}\sum_{\ell_2} (\ell_1\ell_2)^{-\frac{1}{2}} \int_T^{2T} c^*(\ell_1)c^*(\ell_2)(\ell_1/\ell_2)^{it} dt$$

and $c^*(\ell_1)c^*(\ell_2)$ is equal to

$$\sum_{\substack{m_1n_1=\ell_1\\ m_2n_2=\ell_2}}\sum \rho(m_1)\rho(m_2)\lambda(n_1)\lambda(n_2)g(m_1)g(m_2)b\left(s; \frac{\log n_1}{\log N}, \frac{\log n_2}{\log N}\right)$$

with $b(s; x_1, x_2) = b(\delta(s) + 1 - x_1)b(\delta(s) + 1 - x_2)$ and $\delta(s)$ given by (4.17). The diagonal terms $\ell_1 = \ell_2$ contribute $T(\mathcal{K}_0^* + c^*(1)^2)$, where

$$\mathcal{K}_0^* = \sum_{1<\ell<4MN^{1-\beta}} c^*(\ell)^2\ell^{-1} \tag{6.2}$$

and $c^*(\ell)$ are taken with the $\delta = \delta(s)$ at some fixed point s in the segment (4.15) by the mean-value theorem. Note we have extracted the first term $c^*(1) = h^*(1) = b^*(0) = -\delta \ll 1/\log N$.

If $\ell_1 \neq \ell_2$ we integrate by parts. Since

$$\frac{\partial}{\partial t} b(s; x_1, x_2) \ll \frac{\partial}{\partial t}\delta(s) \asymp t/(t^2+1)\log N \asymp 1/T\log N,$$

we estimate the contribution of the off-diagonal terms by

$$\sum_{\ell_1\neq\ell_2<4MN^{1-\beta}}\sum \tau^2(\ell_1)\tau^2(\ell_2)(\ell_1\ell_2)^{-\frac{1}{2}}|\log \ell_1/\ell_2|^{-1} \ll MN^{1-\beta}(\log MN)^{10}$$

(the small shift $\delta(s)$ does not influence the bound). By our assumption (6.1) we conclude that

$$\mathcal{K}^*(T) = T\mathcal{K}_0^* + O(T/\log T). \tag{6.3}$$

In case of $\mathcal{K}(T)$ there is no perturbation by the shift $\delta(s)$ so the diagonal contribution is exactly $\hat{\Phi}(0)T\mathcal{K}_0$, where

$$\mathcal{K}_0 = \sum_{1<\ell<MN^\alpha} c(\ell)^2\ell^{-1}. \tag{6.4}$$

Note there is no term with $l = 1$. However the polynomial $C(s)$ has length $MN^\alpha > MN^\beta > MN^{1/2} = MQT > T$ so there is a significant contribution of the off-diagonal terms, say $T\mathcal{K}^{\neq}(T)$, where

$$\mathcal{K}^{\neq}(T) = \sum\sum_{1<\ell_1\neq\ell_2<MN^\alpha} \Psi(T\log \ell_1/\ell_2)c(\ell_1)c(\ell_2)(\ell_1\ell_2)^{-\frac{1}{2}} \tag{6.5}$$

and $\Psi(2\pi v)$ stands for the Fourier transform of $\Phi(u)$. We have

$$\mathcal{K}(T) = \Psi(0)T\mathcal{K}_0 + T\mathcal{K}^{\neq}(T). \tag{6.6}$$

Our goal is to show that the diagonal sums $\mathcal{K}_0$ and $\mathcal{K}_0^*$ are small. The partial sums

$$S(X,Y) = \sum_{X<\ell\leqslant Y} c(\ell)^2\ell^{-1} \tag{6.7}$$

with $1 \leqslant X < Y$ need different handling in various ranges. The corresponding partial sums

$$S^*(X,Y) = \sum_{X<\ell\leqslant Y} c^*(\ell)^2\ell^{-1} \tag{6.8}$$

are similar to $S(X,Y)$ so we shall treat $S(X,Y)$ in details and only occasionally we shall make comments to illuminate small differences. The final estimates for $S(X,Y)$ and $S^*(X,Y)$ will be the same.

The off-diagonal sum $\mathcal{K}^{\neq}(T)$ requires a lot more sophisticated analysis which we postpone to the last ten sections.

7. Estimation of the First Diagonal Terms

The lacunarity of $\lambda(\ell)$ with small ℓ is not frequent and the best available bound is the trivial one $|\lambda(\ell)| \leqslant \tau(n)$. However every convolution coefficient $c(\ell), \ell \neq 0$, is small by the sifting effect. We shall see that in the range $\ell \leqslant M$ the coefficients $c(\ell)$ are supported on almost primes ($c(\ell)$ vanishes if ℓ has many distinct prime divisors).

In this section only we introduce the von Mangoldt functions $\Lambda_j(n)$ of degree $j = 0, 1, 2, \dots, r$, which are derived from the Euler product (1.1) and scaled down by factors $(\log M)^{-j}$. Sorry we use the same notation as for the classical von Mangoldt functions derived from $\zeta(s)$, hopefully without confusion.

If $\ell \leqslant M$, then (5.2) becomes

$$c(\ell) = \sum_{mn=\ell} \rho(m)\lambda(n)\left(1-\frac{\log m}{\log M}\right)^r\left(1-\frac{\log n}{\log N}\right) \tag{7.1}$$

because the restrictions $m \leqslant M$, $n \leqslant N$ are redundant. Moreover (5.3) becomes

$$c^*(\ell) = \sum_{mn=\ell} \rho(m)\lambda(n) \left(1 - \frac{\log m}{\log M}\right)^r \left(\frac{\log n}{\log N} - \delta\right). \tag{7.2}$$

Clearly (7.1) and (7.2) are very similar. The generating Dirichlet series for the unrestricted convolution coefficients (7.1) is equal to

$$\begin{aligned} Z_r(s) &= \left(\sum_n \lambda(n)\left(1 - \frac{\log n}{\log N}\right) n^{-s}\right)\left(\sum_m \rho(m)\left(1 - \frac{\log m}{\log M}\right)^r m^{-s}\right) \\ &= \left(L(s) + \frac{L'(s)}{\log N}\right)\left(\frac{M^s}{L(s)}\right)^{(r)} M^{-s}(\log M)^{-r} \end{aligned}$$

and that for (7.2) is equal to

$$Z_r^*(s) = -\left(\delta L(s) + \frac{L'(s)}{\log N}\right)\left(\frac{M^s}{L(s)}\right)^{(r)} M^{-s}(\log M)^{-r}.$$

For example, for $r = 0$ we get

$$Z_0(s) = 1 + \frac{L'(s)}{L(s)\log N} = 1 - \gamma \sum_\ell \Lambda_1(\ell)\ell^{-s}$$

with $\gamma = \log M/\log N$ (scaling adjustment factor). Therefore $c_1(\ell) = \gamma\Lambda_1(\ell)$ is supported on prime powers.

For any $r \geqslant 0$ we have the formula

$$\left(\frac{M^s}{L(s)}\right)^{(r)} = M^s(\log M)^r \sum_{0\leqslant j\leqslant r} \binom{r}{j} (\log M)^{-j} \left(\frac{1}{L(s)}\right)^{(j)}$$

and $(1/L(s))^{(j)}$ is equal to

$$\frac{j!}{L(s)} \sum_{a_1+2a_2+\dots=j} \frac{(a_1+a_2+\dots)!}{a_1!a_2!\dots}\left(\frac{-L'(s)}{1!L(s)}\right)^{a_1}\left(\frac{-L''(s)}{2!L(s)}\right)^{a_2}\dots$$

Write $\binom{r}{j} = r!/j!b!$ with $b = r - j$ and $\log = \log M$. These formulas yield

$$\begin{aligned} Z_r(s) = \left(1 + \frac{\gamma L'}{L\log}\right) \sum_{b+a_1+2a_2+\dots=r} & \frac{r!(a_1+a_2+\dots)!}{b!a_1!a_2!\dots} \\ & \left(\frac{-L'}{1!L\log}\right)^{a_1}\left(\frac{-L''}{2!L\log}\right)^{a_2}\dots \end{aligned}$$

Comparing the coefficients in Dirichlet series expansions on both sides we conclude

Lemma 7.1. *If $1 \leqslant \ell \leqslant M$, then $c(\ell)$ can be written as the sum*

$$\sum_{a_1+2a_2+\dots\leqslant r} \phi(a_1, a_2, \dots)(\Lambda_0 - \gamma\Lambda_1) \star (\underbrace{\Lambda_1 \star \dots \star \Lambda_1}_{a_1 \text{ times}}) \star (\underbrace{\Lambda_2 \star \dots \star \Lambda_2}_{a_2 \text{ times}}) \star \dots$$

with suitable coefficients $\phi(a_1, a_2, \dots)$, $\phi(0, 0, \dots) = 1$.

Using the obvious estimate

$$\sum_{\ell\leqslant Y}|\Lambda_1(\ell)|^2\ell^{-1}\ll\left(\frac{\log Y}{\log M}\right)^2$$

and recurrence formulas for Λ_k we get

$$V_k(Y)=\sum_{\ell\leqslant Y}|\Lambda_k(\ell)|^2\ell^{-1}\ll\left(\frac{\log Y}{\log M}\right)^{2k}.$$

Hence, putting $\Lambda_{(k)}=\Lambda_{k_1}\star\cdots\star\Lambda_{k_n}$ for $(k)=(k_1,\dots,k_n)$ we derive

$$\sum_{\ell\leqslant Y}|\Lambda_{(k)}(\ell)|^2\ell^{-1}\leqslant V_{k_1}(Y)\dots V_{k_n}(Y)\ll\left(\frac{\log Y}{\log M}\right)^{2(k_1+\cdots+k_n)}$$

where the implied constant depends only on (k).

Since $c(\ell)$ for $1<\ell\leqslant M$ is a linear combination of $\Lambda_k(\ell)$ with $(k)\neq(0,0,\dots)$ by Lemma 7.1 we derive by the above estimates

Lemma 7.2. *If $2\leqslant Y\leqslant M$, then*

$$S(1,Y)\ll(\log Y/\log M)^2 \tag{7.3}$$

where the implied constant depends only on r.

By the above arguments it is clear that the same bound (7.3) holds for $S^*(1,Y)$.

The bound (7.3) is valid for $Y\leqslant M$, but it is good only if Y is relatively smaller. Suppose $Q^8\leqslant X<Y\leqslant M$. Now the lacunarity of $\lambda(\ell)$ kicks in and we are going to exploit it on top of the mollifier sifting effects. Lemma 7.1 shows that $c(\ell)$ vanishes if ℓ has more than r distinct prime divisors. Writing uniquely $\ell=dk$, where $(d,k)=1$, k squarefree, d powerful, we get $c(\ell)\ll\lambda_0(k)\ll 1$ (recall that $\rho(m)$ is supported on cubefree numbers). Hence,

$$\begin{aligned}S(X,Y)&\ll\sum_{X<dk\leqslant Y}\lambda_0(k)(dk)^{-1}\\&\ll\sum_d d^{-1}\sum_{\sqrt{X}<k\leqslant Y}\lambda_0(k)k^{-1}+\sum_{d\geqslant\sqrt{X}}d^{-1}\sum_{k\leqslant Y}\lambda_0(k)k^{-1}\\&\ll\sum_{\sqrt{X}<k\leqslant Y}\lambda_0(k)k^{-1}+X^{-\frac14}\log Y.\end{aligned}$$

Applying (1.9) we derive the following estimate

Lemma 7.3. *If $Q^8\leqslant X<Y\leqslant M$, then*

$$S(X,Y)\ll L(1,\chi)\log Y \tag{7.4}$$

where the implied constant depends only on r.

The same bound (7.4) holds for $S^*(X,Y)$.

8. Reducing to the Squarefree Diagonal Terms

If $\ell > M$, then the convolution coefficients

$$c(\ell) = \sum_{mn=\ell} \rho(m)\lambda(n)g(m)h(n) \tag{8.1}$$

may not be supported on almost primes, because $g(m), h(n)$ are no longer polynomials in $\log m/\log M$ and $\log n/\log N$, respectively. Therefore the previous arguments fail. We shall estimate $S(X,Y)$ with $X \geqslant M$ in different ways. But first we reduce the sum $S(X,Y)$ to

$$S^\flat(X,Y) = \sum_{\substack{X<\ell\leqslant Y\\(\ell,q)=1}} \mu(\ell)^2 c(\ell)^2 \ell^{-1} \tag{8.2}$$

where q is a fixed squarefree number to be chosen later.

Throughout d runs over numbers such that $p \mid d \Rightarrow p^2 \mid dq^2$. Writing uniquely $\ell = dk$ with k squarefree, $(k, dq) = 1$ we get

$$c(\ell) = \sum_{uv=d} \rho(u)\lambda(v)c_{uv}(k)$$

where

$$c_{uv}(k) = \sum_{mn=k} \rho(m)\lambda(n)g(um)h(vn). \tag{8.3}$$

Hence

$$c(\ell)^2 \leqslant \tau(d)^6 \sum_{uv=d} c_{uv}(k)^2$$

and

$$S(X,Y) \leqslant \sum_d \frac{\tau(d)^6}{d} \sum_{uv=d} S^\flat_{uv}\left(\frac{X}{d}, \frac{Y}{d}\right)$$

where $S^\flat_{uv}(X,Y)$ stands for the sum (8.2) with $c(\ell)$ replaced by $c_{uv}(\ell)$. The contribution of large d, say $d > U = (\log Y)^{50}$, is negligible. Precisely, by trivial estimations, we derive the following:

$$|c(\ell)| \ll \tau_4(\ell),$$

$$S^\flat_{uv}\left(\frac{X}{d}, \frac{Y}{d}\right) \ll (\log Y)^{16},$$

$$\sum_{d>U} \frac{\tau(d)^6}{d} \sum_{uv=d} S^\flat_{uv}\left(\frac{X}{d}, \frac{Y}{d}\right) \ll U^{-1}(\log Y)^{48}.$$

Hence

$$S(X,Y) \leqslant \sum_{d\leqslant U} \frac{\tau(d)^6}{d} \sum_{uv=d} S^\flat_{uv}\left(\frac{X}{d}, \frac{Y}{d}\right) + O((\log Y)^{-2}). \tag{8.4}$$

The coefficients $c_{uv}(\ell)$ in $S^\flat_{uv}(X/d, Y/d)$ have slightly shifted crop functions;

$$g(um) = \left(1 - \frac{\log um}{\log M}\right)^r = \left(1 - \frac{\log u}{\log M}\right)^r \left(1 - \frac{\log m}{\log M/u}\right)^r$$

if $1 \leqslant um \leqslant M$, and

$$h(vn) = a\left(\frac{\log vn}{\log N}\right) = a\left(\delta + \frac{\log n}{\log N}\right) \text{ with } \delta = \frac{\log v}{\log N}.$$

In the next two sections we shall get estimates for $S^\flat(X,Y)$ which apply to every $S^\flat_{uv}(X/d, Y/d)$ with $uv = d \leqslant U = (\log Y)^{50}$. The small change of the crop functions (g by rescaling $M \to M/u$ and h by the shift $a(x) \to a(\delta + x)$) does not require any significant changes in the used arguments.

9. Estimating $S^\flat(X,Y)$

For $\ell = mn$ squarefree we have $\rho(m)\lambda(n) = \mu(m)\lambda(\ell)$ and $c(\ell) = \lambda(\ell)\theta(\ell)$ where $\theta = \mu g \star h$ is a kind of a sieve weight,

$$\theta(\ell) = \sum_{m|\ell} \mu(m) g(m) h(\ell/m). \tag{9.1}$$

This factorization separates the lacunarity feature of $\lambda(\ell)$ from the sifting feature of $\theta(\ell)$. We have $c(\ell)^2 \leqslant |\lambda(\ell)|^{\frac{1}{2}} |\theta(\ell)| \tau(\ell)^2$ and by Cauchy's inequality

$$S^\flat(X,Y) \leqslant \left(\sum_{X < \ell \leqslant Y} |\lambda(\ell)| \ell^{-1}\right)^{\frac{1}{2}} \left(\sum_{X < \ell \leqslant Y} \tau(\ell)^4 \mu(q\ell)^2 \theta(\ell)^2 \ell^{-1}\right)^{\frac{1}{2}}. \tag{9.2}$$

The first sum is bounded by $L(1,\chi)\log Y$ if $Y > X \geqslant Q^4$, see (1.9). The second sum is bounded by

$$T(X,Y) = \sum_{X < \ell \leqslant Y} \phi(\ell)\theta(\ell)^2 \tag{9.3}$$

where $\phi(\ell)$ is the completely multiplicative function such that

$$\phi(p) = 0 \quad \text{if } p \mid q, \qquad \phi(p) = r/p \quad \text{if } p \nmid q, \tag{9.4}$$

with $r = 16$. We shall estimate $T(X,Y)$ for any $r \geqslant 1$ which agrees with the exponent in the crop function (3.11) of the mollifier (3.10). Our goal is the following estimate (we assume q is divisible by every $p \leqslant r^2$ so $\phi(p) < 1/\sqrt{p}$).

Lemma 9.1. *If $Q^4 \leqslant M \leqslant X < Y \leqslant N$, then*

$$T(X,Y) \ll (\log Y/\log M)^r \tag{9.5}$$

where the implied constant depends on r.

By (9.5) and (9.2) we get

Corollary 9.2. *If $Q^4 \leqslant M \leqslant X < Y \leqslant N$, then*

$$S^\flat(X,Y) \ll (L(1,\chi)\log Y)^{\frac{1}{2}} (\log Y/\log M)^{16}. \tag{9.6}$$

Remarks. If we assumed the stronger lacunarity property that $L(1,\chi) \ll (\log|D|)^{-r-6}$, then the trivial bound $T(X,Y) \ll (\log Y)^{r+4}$ would have sufficed. However, we are willing to assume only that $\varepsilon(D) = L(1,\chi)\log|D| \to 0$, so our job is much harder.

The same arguments work for $S^{\flat}_{uv}(X/d, Y/d)$ with $uv = d \leqslant U = (\log Y)^{50}$ giving the same bound (9.5). Hence (8.4) yields.

Corollary 9.3. *If $Q^4 \leqslant M \leqslant X < Y \leqslant N$, then*

$$S(X,Y) \ll (L(1,\chi)\log Y)^{\frac{1}{2}}(\log Y/\log M)^{16} + (\log Y)^{-2}. \tag{9.7}$$

10. Sums of the Möbius function

Typically for estimating sums involving the Möbius function one applies analytic methods by contour integration in the zero-free region of $\zeta(s)$. We opt more elementary path which goes through the Prime Number Theorem in the following form

$$\sum_{m\geqslant X} \mu(m)m^{-1} \ll \exp(-c\sqrt{\log x}) \tag{10.1}$$

where c is a positive constant. In the sequel c stands for a positive constant different every time. By (10.1) one derives

$$\sum_{\substack{m\geqslant X\\(m,k)=1}} \mu(m)m^{-1} \ll \sigma_{-1}(k)\exp(-c\sqrt{\log x}). \tag{10.2}$$

Then (10.2) yields the same bound for the sum twisted by the divisor functions

$$\sum_{\substack{m\geqslant X\\(m,k)=1}} \mu(m)\tau_r(m)m^{-1} \ll \sigma_{-1}(k)\exp(-c\sqrt{\log x}). \tag{10.3}$$

Hence, if $f(x)$ is a function on $\mathbb{R}^+$ with $|f(x)| + x|f'(x)| \ll (\log x)^A$ then we get (by partial summation)

$$\sum_{\substack{m\geqslant X\\(m,k)=1}} \mu(m)\tau_r(m)f(m)m^{-1} \ll \sigma_{-1}(k)\exp(-c\sqrt{\log x}) \tag{10.4}$$

where the implied constant depends on A. In particular

$$\sum_{\substack{m\geqslant X\\(m,k)=1}} \mu(m)\tau_r(m)(\log m)^a \ll \sigma_{-1}(k)\exp(-c\sqrt{\log x}). \tag{10.5}$$

Moreover, for the complete sum we have

$$\sum_{(m,k)=1} \mu(m)\tau_r(m)(\log m)^a = 0 \qquad \text{if } 0 \leqslant a < r. \tag{10.6}$$

Indeed, the complete sum (10.6) is the a-th derivative (at $s = 1$) of

$$\sum_{(m,k)=1} \mu(m)\tau_r(m)m^{-s} = \prod_{p\nmid k}(1 - rp^{-s}) = \zeta(s)^{-r}\eta_{rk}(s)$$

say, where $\eta_{rk}(s)$ is holomorphic in $\operatorname{Re} s > 1/2$. Since $\zeta(s)^{-r}$ has zero as $s = 1$ of order r the formula (10.6) follows.

We shall also need the following formula

Lemma 10.1. *Let $\phi(\ell)$ be the completely multiplicative function defined by (9.4). Suppose every prime $p \leqslant r^2$ divides q. Then*

$$\sum_{\ell \leqslant X} \phi(\ell) = P_r(\log x) + O(X^{-1/5r}) \tag{10.7}$$

where $P_r(X)$ is a polynomial of degree r and the implied constant depends on r.

Proof. The generating Dirichlet series of $\phi(\ell)\ell$ is given by

$$Z(s) = \prod_{p \nmid q} \left(1 - \frac{r}{p^s}\right)^{-1} = \zeta(s)^r \prod_{p \mid q} \left(1 - \frac{1}{p^s}\right)^r \prod_{p \nmid q} \left(1 - \frac{1}{p^s}\right)^r \left(1 - \frac{r}{p^s}\right)^{-1}.$$

Here the last infinite product over $p \nmid q$ converges absolutely in $\operatorname{Re} s > 1/2$. Hence (10.7) follows by standard contour integration and the convexity bound $\zeta(s) \ll |s|^{\frac{1}{8}} \log 4|s|$ on the line $\operatorname{Re} s = 3/4$. Specifically, the main term $P_r(\log x)$ in (10.7) is the residue of $Z(s+1)s^{-1}x^s$ at $s = 0$. □

11. Estimation of $T(X, Y)$

Squaring out (9.1) we get

$$\begin{aligned} \theta(\ell)^2 &= \sum_{[m_1,m_2]\mid \ell} \sum \mu(m_1)\mu(m_2) g(m_1) g(m_2) h\left(\frac{\ell}{m_1}\right) h\left(\frac{\ell}{m_2}\right) \\ &= \sum_{dm_1m_2 \mid \ell} \sum \sum \mu(d)\mu(dm_1m_2) g(dm_1) g(dm_2) h\left(\frac{\ell}{dm_1}\right) h\left(\frac{\ell}{dm_2}\right). \end{aligned}$$

Note that $dm_1 < M$ and $dm_2 < M$ by the support of the mollifier. Introducing this into (9.3) we get

$$T(X,Y) = \sum_{X < dm_1m_2\ell \leqslant Y} \sum \sum \sum \sum \begin{array}{l} \mu(d)\mu(dm_1m_2) g(dm_1) g(dm_2) \\ h(\ell m_1) h(\ell m_2) \phi(dm_1m_2) \phi(\ell). \end{array}$$

Note that $\phi(dm_1m_2) = \tau_r(d)\tau_r(m_1)\tau_r(m_2)/dm_1m_2$ if $(dm_1m_2, q) = 1$ and it vanishes otherwise.

If d is close to M, say $M\Delta^{-3} < d < M$, then $m < \Delta^3$ and $g(dm) < (3\log\Delta/\log M)^r$ for $m = m_1m_2$. Hence the contribution of these "boundary" terms to $T(X, Y)$ is bounded trivially by

$$\begin{aligned} &\sum_{M\Delta^{-3} < d < M} \frac{\tau_r(d)}{d} \left(\frac{3\log\Delta}{\log M}\right)^{2r} \left(\sum_{m < \Delta^3} \frac{\tau_r(m)}{m}\right)^2 \sum_{\ell \leqslant Y} \phi(\ell) \\ &\ll (\log M)^{r-1} (\log\Delta) \left(\frac{\log\Delta}{\log M}\right)^{2r} (\log\Delta)^{2r} (\log Y)^r = \left(\frac{\log Y}{\log M}\right)^r \end{aligned}$$

if we choose Δ such that $(\log\Delta)^{4r+1} = \log M$. This bound meets the goal (9.5). The above lines show how important it is to have the crop function $g(m)$ vanishing at the end point M of degree as large as the degree of the divisor function $\tau_r(d)$.

Now, when $d \leqslant M\Delta^{-3}$ with $(\log \Delta)^{4r+1} = \log M$, there is enough room for the Möbius function in $T(X,Y)$ to produce significant cancellation. First, if $m_1 > \Delta$ or $m_2 > \Delta$, then (10.4) shows that the contribution of such terms to $T(X,Y)$ is estimated by

$$\exp(-c\sqrt{\log \Delta}) \left(\sum_{d<M} \tau_r(d) d^{-1}\right)^2 \left(\sum_{\ell<Y} \tau_r(d)\right) \ll \left(\frac{\log Y}{\log M}\right)^r.$$

This bound meets the goal (9.5).

It remains to estimate the partial sum of $T(X,Y)$ over the segment $X < dm_1m_2\ell \leqslant Y$ restricted by the following conditions

$$m_1 \leqslant \Delta,\ m_2 \leqslant \Delta,\ d \leqslant M\Delta^{-3}. \tag{11.1}$$

We assume $M \leqslant X < Y \leqslant N$. Then (11.1) implies $\ell > \Delta$ so we have enough space to execute the summation over ℓ. By Lemma 10.1 using partial summation we get

$$\begin{aligned}\sum_{X<dm_1m_2\ell\leqslant Y} \phi(\ell) h(\ell m_1) h(\ell m_2) &= \int_{X/d}^{Y/d} h\left(\frac{y}{m_1}\right) h\left(\frac{y}{m_2}\right) dP_r\left(\log \frac{y}{m_1m_2}\right) \\ &\quad + O(\Delta^{-1/5r}).\end{aligned}$$

The contribution of the error term to $T(X,Y)$ is $\ll \Delta^{-1/5r}(\log M)^r(\log \Delta)^{2r}$ which is much smaller than required. Collecting the above results we get

$$T(X,Y) = \sum_{d<M\Delta^{-3}} \mu(dq) \frac{\tau_r(d)}{d} \int_{X/d}^{Y/d} T_d(y) \frac{dy}{y} + O\left(\left(\frac{\log Y}{\log M}\right)^r\right) \tag{11.2}$$

with

$$\begin{aligned}T_d(y) = \sum\sum_{m_1m_2<\Delta} &\mu(dqm_1m_2) \frac{\tau_r(m_1m_2)}{m_1m_2} g(dm_1) g(dm_2) \\ &h\left(\frac{y}{m_1}\right) h\left(\frac{y}{m_2}\right) P_r'\left(\log \frac{y}{m_1m_2}\right).\end{aligned}$$

Have in mind that the polynomial $P_r'(X)$ has degree $r-1$,

$$P_r'(X) = \sum_{a<r} c(a) X^a.$$

The crop function $g(dm)$ of the mollifier in the above range is the polynomial in $\log m/\log M$;

$$g(dm) = \left(1 - \frac{\log dm}{\log M}\right)^r = \sum_{0\leqslant j_1\leqslant r} \binom{r}{j} \left(1 - \frac{\log d}{\log M}\right)^{r-j} \left(\frac{-\log m}{\log M}\right)^j.$$

However

$$h\left(\frac{y}{m}\right) = a\left(\frac{\log y/m}{\log N}\right) = a\left(\frac{\log y}{\log N} - \frac{\log m}{\log N}\right)$$

is not, but it can be approximated by a polynomial using the Taylor expansion

$$h\left(\frac{y}{m}\right) = \sum_{e<E} \frac{1}{e!} a^{(e)}\left(\frac{\log y}{\log N}\right) \left(\frac{-\log m}{\log N}\right)^e + O\left(\left(\frac{\log \Delta}{\log N}\right)^E\right).$$

Choosing E sufficiently large in terms of r the error term becomes negligible. Finally we have

$$P_r'\left(\log\frac{y}{m_1m_2}\right)=\sum_{\alpha+\alpha_1+\alpha_2<r}c(\alpha,\alpha_1,\alpha_2)(\log y)^\alpha(\log m_1)^{\alpha_1}(\log m_2)^{\alpha_2},$$

where $c(\alpha,\alpha_1,\alpha_2)=(-1)^{\alpha_1+\alpha_2}c(\alpha+\alpha_1+\alpha_2)(\alpha+\alpha_1+\alpha_2)!/\alpha!\alpha_1!\alpha_2!$. By the above expansions we see that $T_d(y)$ is (up to negligible error terms) a linear combination of sums of type

$$\sum_{m_1}\sum_{m_2}\quad\mu(dqm_1m_2)\frac{\tau_r(m_1m_2)}{m_1m_2}\left(\frac{\log m_1}{\log M}\right)^{j_1}\left(\frac{\log m_2}{\log M}\right)^{j_2}\\ \left(\frac{\log m_1}{\log N}\right)^{e_1}\left(\frac{\log m_2}{\log N}\right)^{e_2}(\log y)^\alpha(\log m_1)^{\alpha_1}(\log m_2)^{\alpha_2}$$

with $j_1,j_2\leqslant r$, $e_1,e_2<E$ and $\alpha+\alpha_1+\alpha_2\leqslant r-1$, where the summation is restricted by $m_1\leqslant\Delta$, $m_2\leqslant\Delta$. These restrictions can be dropped up to error term bounded by $(\log y)^\alpha\exp(-c\sqrt{\log\Delta})$, see (10.5), which is negligible. The complete sum vanishes, see (10.6), unless $j_1+e_1+\alpha_1\geqslant r$ and $j_2+e_2+\alpha_2\geqslant r$, in which case it is bounded by

$$\begin{aligned}&(\log y)^\alpha(\log M)^{-j_1-j_2}(\log N)^{-e_1-e_2}\\ \ll\ &(\log y)^\alpha(\log M)^{e_1+e_2+\alpha_1+\alpha_2-2r}(\log N)^{-e_1-e_2}\\ \ll\ &(\log y)^\alpha(\log M)^{\alpha_1+\alpha_2-2r}\\ \ll\ &(\log y)^\alpha(\log M)^{-\alpha-r-1}\\ \ll\ &(\log y)^{r-1}(\log M)^{-2r}.\end{aligned}$$

Inserting this bound into (11.2) we derive (9.5).

12. Conclusion

We have all parts (except for the off-diagonal terms) ready to conclude the proof of the main Theorem 2.1. Take the mollifier (3.10) of length $M=T^{1/400}\geqslant Q^8$. Choose the breaking points in the partition (4.1) at $\alpha=\frac12+\frac{1}{100}$, $\beta=\frac12+\frac{1}{200}$. Then the level of Levinson's function (3.1) satisfies $T^2<N<T^3$, see (4.14). The diagonal sum (6.4) is estimated by

$$\mathcal{K}_0=S(1,MN^\alpha)\leqslant S(1,N)=S(1,Q^8)+S(Q^8,M)+S(M,N).$$

Applying (7.3), (7.4) and (9.7) we get

$$\mathcal{K}_0\ll\left(\frac{\log Q}{\log T}\right)^2+L(1,\chi)\log T+(L(1,\chi)\log T)^{\frac12}.$$

The same bound holds for the diagonal sum $\mathcal{K}_0^*$ given by (6.2). Hence (5.7) yields

$$I(T)\leqslant T|\mathcal{K}^{\neq}|^{\frac12}+O\left(T\frac{\log Q}{\log T}+T(L(1,\chi)\log T)^{\frac12}+T(L(1,\chi)\log T)^{\frac14}\right)$$

where the implied constant is absolute. Inserting this into (3.7) we get

$$N_{00}(T) > N(T) - 4|\mathcal{K}^{\neq}|^{\frac{1}{2}}T\log T + O\left(T\log Q + T(\log T)(L(1,\chi)\log T)^{\frac{1}{4}}\right). \tag{12.1}$$

Note that the condition $T \geqslant Q^{3200}$ is no longer required, because the estimate (12.1) holds trivially otherwise. It remains to estimate the contribution $\mathcal{K}^{\neq}(T)$ of the off-diagonal terms, see (6.5).

13. An Introduction to the Off-diagonal Terms

Our goal is to show that the contribution of $\mathcal{K}^{\neq}(T)$ to $\mathcal{K}(T)$ is quite small, comparable to $\mathcal{K}_0$, so that it can be omitted in (12.1). We shall only consider the L-function for the trivial ideal class group character $\psi = \psi_0$, in which case

$$L(s) = \zeta(s)L(s,\chi) = \sum_n \lambda(n)n^{-s} \tag{13.1}$$

with $\lambda = 1 \star \chi$. The other cases are similar, in fact simpler, because the main term of $\mathcal{K}^{\neq}(T)$ vanishes.

Before starting advanced arguments we recall the situation in fresh notation to recycle a lot of alphabet which was used so far. We have

$$\mathcal{K}^{\neq}(T) = \sum_{u,v<M}\sum \frac{\rho(u)\rho(v)}{\sqrt{uv}}g(u)g(v)I\left(\frac{u}{v}\right) + O\left(\frac{1}{T}\right) \tag{13.2}$$

where

$$I\left(\frac{u}{v}\right) = \sum_{um\neq vn}\sum \Psi\left(T\log\frac{um}{vn}\right)\frac{\lambda(m)\lambda(n)}{\sqrt{mn}}h(m)h(n) \tag{13.3}$$

and $\Psi(2\pi z)$ denotes the Fourier transform of $\Phi(t)$. The error term $O(1/T)$ in (13.2) is an easy estimate for the contribution of terms $um = 1$ or $vn = 1$ which are added in (13.3).

In (5.5) we said that $\Phi(t)$ was smooth and compactly supported on $\mathbb{R}^+$. Clearly we can modify it here by requesting the symmetry $\Phi(t) = \Phi(-t)$. Then we have $\Psi(z) = \Psi(-z)$ and $I(u/v) = I(v/u)$. Note that

$$\Psi(z) = \int_{-\infty}^{\infty}\Phi(t)\cos(tz)dt \tag{13.4}$$

has fast decaying derivatives, specifically we shall often use the bound

$$\Psi^{(j)}(z) \ll (1+|z|)^{-A},\ j = 0,1,2, \tag{13.5}$$

for real z with any $A \geqslant 4$ where the implied constant depends on A.

Pulling out the greatest common factor of u, v in (13.2) we write

$$\mathcal{K}^{\neq}(T) = \sum_{e<M}\sum_{\substack{u,v<M/e\\(u,v=1)}}\sum \frac{\rho(eu)\rho(ev)}{e\sqrt{uv}}g(eu)g(ev)I\left(\frac{u}{v}\right) + O\left(\frac{1}{T}\right) \tag{13.6}$$

with $I(u/v)$ given by (13.3) without change. Given $(u,v)=1$ we split $I(u/v)$ into

$$I\left(\frac{u}{v}\right) = 2\sum_{h=1}^{\infty} I_h\left(\frac{u}{v}\right) \tag{13.7}$$

where

$$I_h\left(\frac{u}{v}\right) = \sum\sum_{um-vn=h} \Psi\left(T\log\frac{um}{vn}\right)\frac{\lambda(m)\lambda(n)}{\sqrt{mn}}h(m)h(n) \tag{13.8}$$

are additive convolution type sums.

Remarks. Since $\Phi(z)$ decays rapidly $T\log\frac{um}{vn} = T\log\left(1+\frac{h}{vn}\right)$ is essentially bounded so $vn \gg hT \geqslant T$, $um \gg hT \geqslant T$ and um, vn are close to each other, $um/vn = 1+O(1/T)$.

The notation begins to be cumbersome so in the next three sections we are going to present self-contained results about additive convolution sums which will be applicable to (13.8).

14. General Convolution Sums

This is a stand-alone section. Here and in the next three sections our notation is independent of that used in the previous ones. After proving Lemma 17.1 we shall abandon this temporary notation.

Suppose we are given two sequences $\mathcal{A}=(a_m)$, $\mathcal{A}^*=(a_n^*)$, which enjoy some features of the Fourier coefficients of automorphic forms. Our goal is to evaluate the sum

$$\mathcal{B}(h) = \sum_{m-n=h} a_m a_n^* g(m)g^*(n) \tag{14.1}$$

for $h \geqslant 1$, where $g(x), g^*(x)$ are smooth functions, compactly supported on $\mathbb{R}^+$. Sums of such type were treated in Section 4 of [CI02] in a great generality using ideas of Kloosterman's circle method. Now we need (14.1) in a little bit more general setting, in which case the arguments in Section 4 of [CI02] still apply. Since the required modifications are essentially in the notation we shall state the results without repeating proofs. If the sequences $\mathcal{A}, \mathcal{A}^*$ consist of Hecke eigenvalues for a cusp form there are several results in the literature which are useful for out applications. In particular the formulas of [KMV02] in Appendices A, B come close to what we require with respect to the parameters out of which to built the mollifier. However the shifted convolution for $\lambda = 1 \star \chi$ is not covered in [KMV02]. This paper [KMV02] contains numerous fundamental ideals and gives great details so we recommend to the reader to glance it as a supplement to our arguments below.

All we need about the sequence $\mathcal{A}$ (and $\mathcal{A}^*$) is a kind of Voronoi formula for twisted sums

$$S(\alpha) = \sum_m a_m g(m) e(\alpha m) \tag{14.2}$$

at rational points $\alpha = a/c$ for every $c \geqslant 1$ and $(a,c) = 1$. Naturally, one expects that $S(a/c)$ are quite well approximated by

$$\psi(a,c)\int g(x)dx \tag{14.3}$$

where $\psi(a,c)$ is a nice function which depends on $a \pmod c$ and it satisfies

$$|\psi(a,c)| \leqslant \frac{A}{c}. \tag{14.4}$$

Here the parameter A, and two other parameters B, C in forthcoming conditions, will be specified in later applications subject to $A \geqslant 1,\ B \geqslant 1,\ C \geqslant 2$. Note that $\psi(a,c)$ does not depend on the test function $g(x)$, therefore the approximation (14.3) to the sum $S(a/c)$ is a functional.

Based on (14.3) one should predict that $\mathcal{B}(h)$ is quite well approximated by

$$B(h) = \mathfrak{S}(h)\int g(x+h)g^*(x)dx \tag{14.5}$$

where

$$\mathfrak{S}(h) = \sum_{c=1}^{\infty}\sideset{}{^*}\sum_{a \pmod c} e\left(\frac{ah}{c}\right)\psi(-a,c)\psi^*(a,c). \tag{14.6}$$

Indeed we shall see that under suitable conditions the prediction is pretty accurate. We assume that the Fourier transform of $g(x)$ satisfies

$$\int |\hat g(\alpha)|d\alpha \leqslant B, \qquad \int |\alpha||\hat g(\alpha)|^2 d\alpha \leqslant B^2. \tag{14.7}$$

Moreover the same estimates hold for the Fourier transform of $g^*(x)$. We write

$$S\left(\frac{a}{c}\right) = \psi(a,c)\int g(x)dx + T(a,c), \tag{14.8}$$

where the error term $T(a,c)$ does, of course, depend on $g(x)$ as a functional. It is not sufficient to assume a good upper bound for $T(a,c)$; one has to control the variation of its argument and get a considerable cancellation when summing over the classes $a \pmod c$, $(a,c) = 1$ (this is the very essence of Kloosterman's circle method). We postulate that every $T(a,c)$ has the Fourier series expansion of the following type

$$T(a,c) = \sum_{m=1}^{\infty}\psi_m(a)e\left(\frac{\overline{a}}{c}\ell_m\right)\int g(x)k_m(x)dx \tag{14.9}$$

where $\overline{a}$ denotes the multiplicative inverse of a modulo c, $a\overline{a} \equiv 1 \pmod c$. Here the frequencies ℓ_m are integers which are allowed to depend on c, but not on a. Moreover the kernel functions $k_m(x)$ may depend on c, but not on a. Finally, the coefficients $\psi_m(a)$ are also allowed to depend on c in an arbitrary fashion, but the dependence on a must be mild. Specifically, we

assume that there is a fixed integer $q \geqslant 1$ such that $\psi_m(a)$ is periodic in a modulo (c, q), and

$$|\psi_m(a)| \leqslant \frac{A}{c}\tau(m). \tag{14.10}$$

We also assume that the Fourier transform of $g_m(x) = g(x)k_m(x)$ satisfies

$$|\hat{g}_m(\alpha)| \leqslant cCBm^{-5/4} \tag{14.11}$$

for every $1 \leqslant c \leqslant C$ and every α with $|\alpha|cC \leqslant 1$.

Now we are ready to state the following result (go through Section 4 of [CI02] line by line for constructing a definite proof).

Proposition 14.1. *Assume the conditions* (14.4), (14.7), (14.10), (14.11) *for the sequence* $\mathcal{A} = (a_m)$ *and the corresponding conditions for the sequence* $\mathcal{A}^* = (a_n^*)$. *Then for every* $h \geqslant 1$ *we have*

$$\mathcal{B}(h) = B(h) + R(h) \tag{14.12}$$

where $\mathcal{B}(h)$ *is the convolution sum* (14.1), $B(h)$ *is the predicted main term* (14.5) *and* $R(h)$ *is an error term which satisfies*

$$R(h) \ll \tau(h)A^2C^{-1}\int |g(x+h)g^*(x)|dx + \tau(h)qA^2B^2C^{\frac{3}{2}}(\log C)^2 \tag{14.13}$$

with the implied constant being absolute.

15. Special Convolution Sums

We are interested in the sequence $\lambda = 1 \star \chi$ as in (1.8) where χ is the real primitive character of conductor $|D|$. By Proposition 3.3 of [CI02] we have the following Voronoi type formula

$$\sum_{m=1}^{\infty} \lambda(m)e\left(\frac{a}{c}m\right)g(m) = \rho(a,c)L(1,\chi)\int g(x)dx + T(a,c) \tag{15.1}$$

for any $c \geqslant 1$, $(a, c) = 1$, where

$$T(a,c) = 2\pi i\chi_1(a)\chi_2(c)\frac{\sqrt{(c,D)}}{c\sqrt{|D|}} \sum_{m=1}^{\infty}(\chi_1 \star \chi_2)(m)e\left(\overline{aD/(c,D)}\frac{m}{c}\right) \int g(x)J_0\left(4\pi\sqrt{(c,D)mx}/c\sqrt{|D|}\right)dx,$$

$\chi_1 \pmod{(c,D)}$ and $\chi_2 \pmod{|D|/(c,D)}$ are the real characters such that $\chi_1\chi_2 = \chi$. In the main term we have

$$\rho(a,c) = \begin{cases} \chi(c)/c & \text{if } D \nmid c, \\ \chi(a)\tau(\chi)/c & \text{if } D \mid c \end{cases} \tag{15.2}$$

where $\tau(\chi)$ denotes the Gauss sum. Actually Proposition 3.3 of [CI02] requires D to be odd and negative, so

$$D \text{ is squarefree }, \ D < 0, \ D \equiv 1 \pmod 4. \tag{15.3}$$

Therefore, in the next three sections we shall be working under these conditions. The other cases are very much similar and the final estimates are the same so we skip them.

For every positive integer u we derive from (15.1) the following formula

$$\sum_{m=1}^{\infty} \lambda(m) e\left(\frac{a}{c}um\right) g(um) = \rho\left(\frac{au}{(c,u)}, \frac{c}{(c,u)}\right) \frac{L(1,\chi)}{u} \int g(x)dx + T(au/(c,u), c/(c,u)).$$

(replace $a, c, g(x)$ in (15.1) by $au/(c,u),\ c/(c,u),\ g(ux)$ respectively).

Now we can apply Proposition 14.1 for the sequences $\mathcal{A} = (a_m)$, $\mathcal{A}^* = (a_n^*)$ with $a_m = \lambda(m/u)$, $a_n^* = \lambda(n/v)$, where u, v are given positive integers (subject to the popular convention that an arithmetic function is set its value to zero at non-integers arguments). Suppose $g(x)$ and $g^*(x)$ are smooth functions supported in a dyadic segment $[X, 2X]$ with $X \geqslant 2$ whose derivatives satisfy

$$|x^j g^{(j)}(x)| \leqslant 1,\ j = 0, 1, 2. \tag{15.4}$$

Then one can show (see the arguments in Section 4 of [CI02]) that (14.7) holds with $B \ll 1$ and (14.11) holds for $C = 2(u+v)\sqrt{|D|X}$ with $B \ll (u+v)|D|^{\frac{3}{2}}$. Moreover (14.4) holds with $A \ll \sqrt{|D|}L(1,\chi) \ll \sqrt{|D|}\log|D|$ and (14.11) holds for $q = |D|$ with $A \ll u + v$. Therefore Proposition 14.1 yields

Proposition 15.1. *Let $g(x), g^*(x)$ be smooth functions supported in $[X, 2X]$ with $X \geqslant 2$ whose derivatives satisfy* (15.4). *Then for positive integers u, v, h we have*

$$\sum_{um-vn=h} \lambda(m)\lambda(n) g(um) g^*(vn) = \tag{15.5}$$

$$\mathfrak{S}(h)(uv)^{-1} L^2(1,\chi) \int g(x+h) g^*(x) dx + O\left(\tau(h)(uvD)^6 X^{\frac{3}{4}} (\log X)^2\right)$$

where

$$\mathfrak{S}(h) = \sum_{c=1}^{\infty} \sideset{}{^*}\sum_{a \,(\mathrm{mod}\ c)} e\left(\frac{ah}{c}\right) \rho\left(\frac{-au}{(c,u)}, \frac{c}{(c,u)}\right) \rho\left(\frac{av}{(c,v)}, \frac{c}{(c,v)}\right) \tag{15.6}$$

with $\rho(a,c)$ given by (15.2) *and the implied constant being absolute.*

Remarks. We have not assumed that u, v are co-prime. But, of course, if $(u,v) \nmid h$, then the convolution sum on the left side of (15.5) is void so the series $\mathfrak{S}(h)$ on the right side of (15.5) must vanish as well. This could be verified directly if you will, but not so easily. The result is a generalization of a special case of Theorem 4.4 of [CI02]. Note that the exponent 3/4 in the error term comes from an application of Weil's bound for Kloosterman sums. The exponent 7/8 resulting from a weaker elementary bound due to Kloosterman would be also sufficient for our purpose.

For technical simplifications we can impose some local restrictions on the variables u, v in the formula (13.2). These numbers will be in the support of the coefficients $\rho(u), \rho(v)$ of the mollifier, see (3.8) and (3.9). Therefore we can assume that u and v are cubefree with no multiple ramified prime factors; this means $p \mid D \Rightarrow p^2 \nmid u$ and $p^2 \nmid v$.

16. Computing the Series $\mathfrak{S}(h)$

To ease the computations we assume that u, v are coprime;

$$(u, v) = 1, \tag{16.1}$$

so

$$(D^2, uv) = (D, uv). \tag{16.2}$$

Let $1(x)$ denote the characteristic function of integers. Then (15.2) yields $\rho(a, c)c = \chi(c) + \chi(a)\tau(\chi)1(c/D)$ and (15.6) becomes

$$\mathfrak{S}(h) = \sum_{c=1}^{\infty} \frac{(c, uv)}{c^2} \sum_{a \,(\mathrm{mod}\ c)}^{*} e\left(\frac{ah}{c}\right) \{\dots\}\{\dots\}$$

where

$$\begin{aligned}
\{\dots\}\{\dots\} &= \left\{\chi\left(\frac{c}{(c,u)}\right) - \chi\left(\frac{-au}{(c,u)}\right)\tau(\chi)1\left(\frac{c}{(c,u)D}\right)\right\} \\
&\quad \left\{\chi\left(\frac{c}{(c,v)}\right) + \chi\left(\frac{av}{(c,v)}\right)\tau(\chi)1\left(\frac{c}{(c,v)D}\right)\right\} \\
&= \chi\left(\frac{c^2}{(c,uv)}\right) - \chi\left(\frac{a^2uv}{(c,uv)}\right)1\left(\frac{c}{(c,uv)D}\right)D \\
&\quad + \left[\chi\left(\frac{acv}{(c,uv)}\right)1\left(\frac{c}{(c,v)D}\right) - \chi\left(\frac{acu}{(c,uv)}\right)1\left(\frac{c}{(c,u)D}\right)\right]\tau(\chi)
\end{aligned}$$

because $\chi(-1)) = -1$, $\tau(\chi)^2 = D$ and $1(c/(c,u)D)1(c/(c,v)D) = 1(c/(c,uv)D)$. Note that except for the first term $\chi(c^2/(c,uv))$ the other three terms vanish unless $D \mid c$ in which case $\chi(a^2) = 1$. By this observation we get

$$\begin{aligned}
\{\dots\}\{\dots\} = \quad & \chi\left(\frac{c^2}{(c,uv)}\right) - \chi\left(\frac{uv}{(c,uv)}\right)1\left(\frac{c}{(c,uv)D}\right)D \\
& + (\chi(v) - \chi(u))\chi\left(\frac{ac}{(c,uv)}\right)1\left(\frac{c}{D}\right)\tau(\chi).
\end{aligned}$$

Next we introduce the Ramanujan sum

$$r_h(c) = \sum_{a \,(\mathrm{mod}\ c)}^{*} e\left(\frac{ah}{c}\right) = \sum_{d \mid (c,h)} d\mu(c/d) \tag{16.3}$$

and if $D \mid c$ we introduce the hybrid of Gauss-Ramanujan sum

$$r_h(c, \chi) = \sum_{a \,(\mathrm{mod}\ c)}^{*} \chi(a) e\left(\frac{ah}{c}\right). \tag{16.4}$$

Put

$$uv = w, \qquad \text{so } w \text{ is cubefree.} \tag{16.5}$$

The symbols $1(c/(c,w)d)$ and $1(c/D)$ above mean that $(D,w)D \mid c$ (see the condition (16.2)) and $D \mid c$, respectively. Therefore $\mathfrak{S}(h)$ splits into three parts

$$\mathfrak{S}(h) = \mathfrak{S}^*(h) - \mathfrak{S}'(h) + (\chi(v) - \chi(u))\mathfrak{S}(h,\chi) \tag{16.6}$$

where

$$\mathfrak{S}^*(h) = \sum_{(c,D)=1} \chi((c,w))r_h(c)(c,w)c^{-2}, \tag{16.7}$$

$$\mathfrak{S}'(h) = D \sum_{(D,w)D\mid c} \chi(w/(c,w))r_h(c)(c,w)c^{-2}, \tag{16.8}$$

$$\mathfrak{S}(h,\chi) = \tau(\chi) \sum_{D\mid c} \chi(c/(c,w))r_h(c,\chi)(c,w)c^{-2}. \tag{16.9}$$

Note that $\mathfrak{S}(h,\chi)$ vanishes, unless $D \mid w$, and the third part of (16.6) vanishes, unless $D \mid u$ or $D \mid v$. These are pretty strong conditions on u, v which we can easily go around in applications. Therefore, from now on we assume that

$$D \nmid u \text{ and } D \nmid v \tag{16.10}$$

so the third part of (16.6) does not need to be considered (see Section 24).

By the formula (16.3) we can write the first and the second parts of (16.6) as the convolutions $1 \star \gamma^*$ and $1 \star \gamma'$, say, with

$$\gamma^*(d) = \frac{1}{d} \sum_{(cd,D)=1} \chi((cd,w))(cd,w)\mu(c)c^{-2} \tag{16.11}$$

and

$$\gamma'(d) = \frac{D}{d} \sum_{(D,w)D\mid cd} \chi(w/(cd,w))(cd,w)\mu(c)c^{-2}. \tag{16.12}$$

Finally, assuming the conditions (16.10) we conclude that

$$\mathfrak{S}(h) = \mathfrak{S}^*(h) - \mathfrak{S}'(h) = (1 \star \gamma^*)(h) - (1 \star \gamma')(h). \tag{16.13}$$

Lemma 16.1. *We have $\gamma^*(d) = 0$ unless $(d,D) = 1$ in which case*

$$\gamma^*(d) = \frac{(d,w)}{\zeta(2)d}\chi((d,w))\xi(w/(d,w)) \tag{16.14}$$

where

$$\xi(n) = \prod_{p\mid n} \left(1 + \frac{\chi(p)}{p}\right)^{-1}. \tag{16.15}$$

Proof. First note that the product (16.15) for $n = w/(d,w)$ runs over the set $\mathcal{P}$ of primes $p \mid w$ such that

$$p \parallel w \Rightarrow p \nmid d \text{ and } p^2 \mid w \Rightarrow p^2 \nmid d. \tag{16.16}$$

Clearly the sum (16.11) is void if $(d, D) \neq 1$. If $(d, D) = 1$, then $\gamma^*(d) = \chi((d, w))(d, w)d^{-1}\Sigma$, where

$$\begin{aligned}\Sigma &= \sum_{(c,D)=1} \chi\left(\frac{(cd, w)}{(d, w)}\right) \frac{(cd, w)}{(d, w)} \frac{\mu(c)}{c^2} \\ &= \prod_{p\in\mathcal{P}} \left(1 - \frac{\chi(p)}{p}\right) \prod_{p\notin\mathcal{P}} \left(1 - \frac{1}{p^2}\right) \\ &= \zeta(2)^{-1} \prod_{p\in\mathcal{P}} \left(1 - \frac{\chi(p)}{p}\right) \left(1 - \frac{1}{p^2}\right)^{-1}.\end{aligned} \tag{16.17}$$

This yields the formula (16.14). □

Note that (16.14) gives the upper bound (not to be used)

$$|\gamma^*(d)| \leqslant \frac{(d, w)}{d} \prod_{p|w} \left(1 + \frac{1}{p}\right). \tag{16.18}$$

Lemma 16.2. *We have $\gamma'(d) = 0$, unless $(D, w) \mid d$ in which case*

$$\gamma'(d) = \frac{\mu(q)}{\zeta_q(2)} \frac{(d_1, w_1)(d_1, D)^2}{dD} \chi\xi\left(\frac{w_1}{(d_1, w_1)}\right) \tag{16.19}$$

where $w_1 = w/(D, w)$, $d_1 = d/(D, w)$, $q = D/(D, d_1)$, $\zeta_q(s)$ denotes the Riemann zeta function with missing local factors at $p \mid q$ and $\xi(n)$ is given by (16.15).

Proof. By (16.2) it follows that $(D, w_1) = 1$. Clearly the sum (16.12) is void if (D, w) does not divide d, because c is squarefree. If $d = (D, w)d_1$, then the formula becomes (16.12)

$$\gamma'(d) = \frac{D}{d} \sum_{D|cd_1} \chi(w_1/(cd_1, w_1))(cd_1, w_1)\mu(c)c^{-2}.$$

Here $c = c_1 D/(D, d_1)$ and $(cd_1, w_1) = (cd_1, w_1) = (d_1, w_1)(c_1, w_1/(d_1, w_1))$ giving

$$\gamma'(d) = \mu\left(\frac{D}{(D, d_1)}\right) \frac{(D, d_1)^2}{dD} (d_1, w_1) \chi\left(\frac{w_1}{(d_1, w_1)}\right) \Sigma$$

where

$$\begin{aligned}\Sigma &= \sum_{(c_1, D/(D,d_1))=1} \chi\left(\left(c_1, \frac{w_1}{(d_1, w_1)}\right)\right) \frac{\mu(c_1)}{c_1^2} \\ &= \prod_{p\nmid \frac{D}{(D,d_1)} \frac{w_1}{(d_1,w_1)}} \left(1 - \frac{1}{p^2}\right) \prod_{p\mid \frac{w_1}{(d_1,w_1)}} \left(1 - \frac{\chi(p)}{p}\right) \\ &= \prod_{p\nmid \frac{D}{(D,d_1)}} \left(1 - \frac{1}{p^2}\right) \prod_{p\mid \frac{w_1}{(d_1,w_1)}} \left(1 + \frac{\chi(p)}{p}\right)^{-1}.\end{aligned}$$

Hence it is easy to check the formula (16.19). □

Note that (16.19) gives the upper bound (not to be used)

$$|\gamma'(d)| \leqslant \frac{(d_1, w_1)}{(D, w)} \frac{(D, d_1)^2}{|D| d_1} \prod_{p | w_1} \left(1 + \frac{1}{p}\right). \tag{16.20}$$

Remarks. The formula (16.19) for $\gamma'(d)$ is very similar to the formula (16.14) for $\gamma^*(d)$, but it is a bit more involved with respect to the ramified prime places. However, both expressions share the same essential features which are relevant to the forthcoming analysis of the series $\mathfrak{S}^*(h)$, $\mathfrak{S}'(h)$ and $k(y) = k^*(y) - k'(y)$, see (17.6). Therefore, we are going to work with $\gamma^*(d), \mathfrak{S}^*(h), k^*(y)$ in considerable details and skip the analysis of $\gamma'(d)$, $\mathfrak{S}'(h)$, $k'(y)$ assuming that the final estimates are the same in both cases. Actually, the case of $\gamma'(d)$ in some extreme situations can be treated somewhat faster, because a crude upper bound for $\gamma'(d)$ is good enough. For example, suppose d has no ramified prime factors, i.e.

$$(d, D) = 1. \tag{16.21}$$

Then $\gamma'(d) = 0$, unless $(D, w) = 1$, in which case $d = d_1$, $w = w_1$, $D' = 1$ and (16.19) becomes

$$\gamma'(d) = \frac{\mu(D)}{\zeta_D(2)} \frac{(d, w)}{dD} \chi\left(\frac{w}{(d, w)}\right) \xi\left(\frac{w}{(d, w)}\right). \tag{16.22}$$

Here we gained the factor $1/|D|$ by comparison to (16.14) which is so small that one can cover the range $(d, D) = 1$ exploiting neither the lacunarity of $L(s) = \zeta(s)L(s, \chi)$ nor the sifting effects of the action of the mollifier $M(s)$.

In other extreme example suppose that every ramified prime divides d but not w, i.e.

$$D \mid d, \ (D, w) = 1. \tag{16.23}$$

Then (16.19) reduces to (16.14), precisely we have

$$\gamma'(d) = \chi(w)\gamma^*(d/|D|). \tag{16.24}$$

17. Back to the Off-Diagonal Terms

We are now ready to evaluate the components $I(u/v)$ of the sum $\mathcal{K}^{\neq}(T)$, see (13.2) and (13.1). First by Proposition 15.1 we derive

Lemma 17.1. *Let h, u, v be positive integers with $(u, v) = 1$, $u, v < M$. Then*

$$I_h\left(\frac{u}{v}\right) = \frac{\mathfrak{S}(h)}{\sqrt{uv}} L^2(1, \chi) \int_0^\infty \Psi\left(\frac{hT}{x}\right) h\left(\frac{x}{u}\right) h\left(\frac{x}{v}\right) \frac{dx}{x} + O\left(h^2 T^{-\frac{1}{9}}\right). \tag{17.1}$$

Proof. The left side of (15.5) with

$$g(x) = h\left(\frac{x}{u}\right)\left(\frac{u}{x}\right)^{\frac{1}{2}}, \ g^*(y) = \Psi\left(T \log \frac{y + h}{y}\right) h\left(\frac{y}{v}\right)\left(\frac{v}{y}\right)^{\frac{1}{2}}$$

becomes $I_h(u/v)$ and the main term on the right side of (15.5) becomes

$$\frac{\mathfrak{S}(h)}{\sqrt{uv}}L^2(1,\chi)\int_0^\infty \Psi\left(T\log\frac{x+h}{x}\right)h\left(\frac{x+h}{u}\right)h\left(\frac{x}{v}\right)\frac{dx}{\sqrt{x(x+h)}}. \tag{17.2}$$

However, the conditions of Proposition 15.1 are not exactly satisfied by the above choice of the test functions $g(x), g^*(y)$. To meet the conditions (15.4) we apply a smooth partition of the summation variables with constituents supported in semi-dyadic segments $[X,\sqrt{2}X]$, $[Y,\sqrt{2}Y]$ so that our partial sums run over m,n with $X < um < \sqrt{2}X$, $Y < vn < \sqrt{2}Y$ for some $1/2 \leqslant X, Y \leqslant MN^\alpha$. If the segments are equal or are adjacent to each other, then we are dealing with two test functions supported in the same dyadic segment. Moreover the derivatives satisfy $x^j g^{(j)}(x) \ll (u/X)^{\frac{1}{2}}$ and $y^j g^{*(j)}(y) \ll (v/Y)^{\frac{1}{2}}$. Therefore (15.5) yields the main term (17.2) for such partial sums with an error term bounded by

$$\tau(h)(uvD)^6(uv)^{\frac{1}{2}}X^{-\frac{1}{4}}(\log 4X)^2. \tag{17.3}$$

We have chosen α slightly larger than $1/2$, $\alpha = 51/100$, and M a relatively small power of T, $M = T^{1/400}$. Hence our segments end at $2X, 2Y < 2MN^\alpha = 2M(QT)^{2\alpha} < |D|MT^{2\alpha} < T^{45/44}$, provided $|D| < T^{1/4000}$. We have already said in the Remarks following (13.8) that $\Psi(z)$ decays rapidly, see (13.5). Therefore, the contribution to $I_h(u/v)$ of the partial sums over the segments $[X,\sqrt{2}X]$, $[Y,\sqrt{2}Y]$ is negligible, expect for

$$hT^{43/44} < X \leqslant \sqrt{2}Y \leqslant 2X \leqslant T^{45/44}$$

in which cases (17.3) is much smaller than the error term in (17.1). Furthermore, since h is quite small, $h < T^{1/22}$, we can clear the main term (17.2) by means of the following approximations;

$$\begin{aligned}
\Psi\left(T\log\frac{x+h}{x}\right) &= \Psi\left(\frac{hT}{x}\right) + O(Th^2x^{-2}),\\
h\left(\frac{x+h}{u}\right) &= h\left(\frac{x}{u}\right) + O(h/x),\\
(x+h)^{-\frac{1}{2}} &= x^{-\frac{1}{2}}(1+O(h/x)).
\end{aligned}$$

The error terms in the above approximations are negligible. This completes the proof of Lemma 17.1. □

Remarks. Having derived the formula (17.1) from the results in Sections 14, 15 we no longer need these sections, in particular the test functions g, g^* used over there can be forgotten. From now g stands again for the crop function in the mollifier (3.10).

Introducing (17.1) into (13.7) we obtain

$$I\left(\frac{u}{v}\right) = \frac{L^2(1,\chi)}{\sqrt{uv}}\int_0^\infty k\left(\frac{T}{x}\right)h\left(\frac{x}{u}\right)h\left(\frac{x}{v}\right)\frac{dx}{x} + O(T^{-\frac{1}{9}}) \tag{17.4}$$

where

$$k(y) = 2\sum_{h>0}\mathfrak{S}(h)\Psi(hy). \tag{17.5}$$

Then, introducing $\mathfrak{S} = 1 \star \gamma$ with $\gamma(d) = \gamma^*(d) - \gamma'(d)$, see (16.13), we get

$$k(y) = \sum_d \gamma(d)\phi(dy) \tag{17.6}$$

where

$$\phi(z) = \sum_k \Psi(kz). \tag{17.7}$$

Do not forget that $\gamma(d)$ depends on $w = uv$, see (16.14) and (16.19).

Finally we insert (17.4) into (13.6) to get a complete formula for the off-diagonal contribution

Lemma 17.2. *We have*

$$\begin{aligned} \mathcal{K}^{\neq}(T) &= L^2(1,\chi)\sum_e \sum_{(u,v)=1}\sum \frac{\rho(eu)\rho(ev)}{euv} g(eu)g(ev)J(u,v) \\ &\quad + O(T^{-\frac{1}{9}}(\log T)^6) \end{aligned} \tag{17.8}$$

where

$$J(u,v) = \int_0^\infty k\left(\frac{T}{x}\right) h\left(\frac{x}{u}\right) h\left(\frac{x}{v}\right) \frac{dx}{x}. \tag{17.9}$$

Remarks. We should have restricted u, v in (17.8) by the conditions (16.10) which we ignored, because if $D \mid u$ or $D \mid v$, then some trivial estimations yield a small contribution. We shall address this issue in Section 24.

It is convenient to treat the two parts $\gamma^*(d), \gamma'(d)$ separately, so we write (17.8) in the following fashion

$$\mathcal{K}^{\neq}(T) = L^2(1,\chi)(\mathcal{K}^*(T) - \mathcal{K}'(T)) + O(T^{-\frac{1}{9}}(\log T)^6). \tag{17.10}$$

Here $\mathcal{K}^*$ stands for this multiple sum

$$\mathcal{K}^*(T) = \sum_e \sum_{(u,v)=1}\sum \frac{\rho(eu)\rho(ev)}{euv} g(eu)g(ev)J^*(u,v) \tag{17.11}$$

where $J^*(u,v)$ is defined by the integral (17.9) with the kernel $k(y)$ replaced by

$$k^*(y) = \sum_d \gamma^*(d)\phi(dy). \tag{17.12}$$

The second part $\mathcal{K}'(T)$ is defined in the same fashion but with $\gamma'(d)$ in place of $\gamma^*(d)$.

It is not surprising that the off-diagonal contribution $\mathcal{K}^{\neq}(T)$ gains the factor $L^2(1,\chi)$, because the lacunarity of $L(s) = \zeta(s)L(s,\chi)$ strikes independently two times. However, it is not enough gain to treat $\mathcal{K}^*$ and $\mathcal{K}'(T)$ crudely. We need to exploit important features of the mollifier which creates some sifting effects. The job would be quick if one used the Riemann hypothesis for $L(s)$, but of course, this is prohibited. We shall estimate $\mathcal{K}^*(T)$ unconditionally by delicate elementary arguments.

Remarks. The coefficients $\rho(m)g(m)$ of the mollifier (3.10) are supported on cubefree numbers. Yet, it is technically messy to keep track of the square factors in the off-diagonal part, they play no essential role because we do not mind loosing absolute constants. One can pull out the square factors in the same fashion as we have shown for the diagonal terms in Section 8. This operation requires small changes in the range of variables of test functions, nevertheless the notation suffers. We leave for the prudent reader to fill up details while we are working on (from now on) under the assumption that the sum (17.11) is restricted to eu and ev being squarefree. This convenient assumption makes

$$\rho(eu) = \mu(eu)\lambda(eu),\ \rho(ev) = \mu(ev)\lambda(ev). \tag{17.13}$$

18. Computing the Series $\phi(z)$

We begin by providing crude estimates for $\phi(z)$. It is easy to see directly from (17.7) and (13.5) that $z\phi(z) \ll 1$. Moreover, by Poisson's formula

$$\phi(z) = -\Psi(0) + \frac{1}{z}\sum_m \Phi\left(\frac{m}{z}\right) \ll 1 \tag{18.1}$$

because Φ (the Fourier transform of Ψ) is compactly supported with $\Phi(0) = 0$. Together we conclude

$$\phi(z) \ll (1+z)^{-1}, \quad \text{if } z > 0. \tag{18.2}$$

Hence the series (17.11) converges absolutely.

Remarks. Applying the Euler-McLaurin formula to (18.1) one obtains the exact expression

$$z\phi(z) = \int_0^\infty \{\xi z\}\Phi'(\xi)d\xi \tag{18.3}$$

which yields (18.2) at once by the inequality $0 \leqslant \{x\} \leqslant \min(1, x)$. The bound (18.2) cannot be improved if z is small, see the second term in (18.7).

Another way of computing $\phi(z)$ goes by contour integration. By (17.7) we get

$$\phi(z) = \frac{1}{2\pi i}\int_{(2)} \tilde{\Psi}(s)\zeta(s)z^{-s}ds \tag{18.4}$$

where $\tilde{\Psi}(s)$ is the Mellin transform of $\Psi(y)$, which in turn is the Fourier transform of $\Phi(x)$, see (13.4). By Mellin's inversion followed by partial integration we get the formula

$$s(s+1)\tilde{\Psi}(s) = \int_0^\infty \Psi''(z)z^{s+1}dz. \tag{18.5}$$

This yields analytic continuation of $s(s+1)\tilde{\Psi}(s)$ to the half-plane $\operatorname{Re} s > -2$. For $s = -1$ we find that

$$\int_0^\infty \Psi''(z)dz = -\Psi'(0) = 0$$

by (13.4), so $\tilde{\Psi}(s)$ has no pole at $s=-1$. For $s=0$ we find that

$$\int_0^\infty \Psi''(z)zdz = -\int_0^\infty \Psi'(z)dz = \Psi(0) = \int \Phi(x)dx,$$

so $\tilde{\Psi}(s)$ has a simple pole at $s=0$ with residue $\Psi(0)$. Note that

$$\tilde{\Psi}(1) = \int_0^\infty \Psi(z)dz = 2\pi\Phi(0) = 0,$$

so the simple pole of $\zeta(s)$ is cancelled by the zero of $\tilde{\Psi}(s)$ at $s=1$ in (18.4). Hence we get

$$\phi(z) = \frac{1}{2\pi i}\int_{(\varepsilon)} \tilde{\Psi}(s)\zeta(s)z^{-s}ds \tag{18.6}$$

with any $\varepsilon > 0$. It will be a more friendly expression for $\phi(z)$ if the pole at $s=0$ is removed. To this end we write (18.6) in the following form

$$\phi(z) = \phi_0(z) - \frac{1}{2}\Psi(0)(1-z)^+ \tag{18.7}$$

where

$$\phi_0(z) = \frac{1}{2\pi i}\int_{(\varepsilon)} \theta(s)z^{-s}ds \tag{18.8}$$

and

$$\theta(s) = \tilde{\Psi}(s)\zeta(s) + \Psi(0)/2s(s+1). \tag{18.9}$$

Note that $\theta(s)$ is holomorphic in $\operatorname{Re} s > -1$ and it satisfies

$$(s+1)\theta(s) \ll (|s|+1)^{-1}, \quad \text{if } -1 < \operatorname{Re} s \leqslant A \tag{18.10}$$

because $\zeta(0) = -\frac{1}{2}$ and $s(s+1)\tilde{\Psi}(s) \ll (|s|+1)^{-2}$ in vertical strips. By (18.8) we derive

$$\phi_0(z) \ll z(1+z)^{-A}, \quad \text{if } z > 0, \tag{18.11}$$

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

19. Computing the Series $k^*(y)$

First we show a formula in a bit more general case. Recall that $\gamma^*(d)$ depends on w.

Lemma 19.1. *If w is squarefree, then*

$$\sum_d \gamma^*(d)f(dy) = \frac{\xi(w)}{\zeta(2)}\sum_{c|w}\chi(c)\sum_{(d,D)=1} f(cdy)d^{-1} \tag{19.1}$$

for any $f(y)$, provided the series $\sum f(dy)d^{-1}$ converges absolutely.

Proof. The sequence $\gamma^*(d)$ has been computed in Lemma 16.1. By (16.14) we proceed as follows

$$\begin{aligned}\zeta(s)\sum_d \gamma^*(d)f(dy) &= \sum_{(d,D)=1}\frac{(d,w)}{d}\chi((d,w))\xi\left(\frac{w}{(d,w)}\right)f(dy)\\ &= \sum_{a|w}\chi(a)\xi\left(\frac{w}{a}\right)\sum_{(d,Dw/a)=1} f(ady)d^{-1}\\ &= \xi(w)\sum_{\substack{ab|w\\(ab,D)=1}}\frac{\chi(a)\mu(b)}{\xi(a)b}\sum_{(d,D)=1} f(abdy)d^{-1}.\end{aligned}$$

Given $c \mid w$ with $(c, D) = 1$, the above sum over $ab = c$ is equal to

$$\prod_{p|c}\left(\frac{\chi(p)}{\xi(p)}-\frac{1}{p}\right)=\prod_{p|c}\chi(p)=\chi(c).$$

This completes the proof of (19.1). □

Corollary 19.2. *Suppose $w = uv$ is squarefree. For every $y > 0$ we have*

$$k^*(y)=\frac{\xi(w)}{\zeta(2)}\sum_{c|w}\chi(c)\sum_{(d,D)=1}\phi(cdy)d^{-1}. \tag{19.2}$$

It is difficult to execute the summation over d in (19.2) if $y \asymp 1$, so we shall use the formula (19.2) as it is in its raw format. Nevertheless, regardless applications, we continue developing this formula further since the reader may like to see the shape of the function $k^*(y)$ from various perspectives such as (19.11). First, applying the partition (18.7) to (17.12) we derive

$$k^*(y)=\frac{1}{2\pi i}\int_{(\varepsilon)}\theta(s)\mathfrak{z}^*(s)y^{-s}-\frac{1}{2}\Psi(0)\sum_{dy<1}\gamma^*(d)(1-dy), \tag{19.3}$$

where $\mathfrak{z}^*(s)$ is the zeta function of the sequence $\gamma^*(d)$;

$$\mathfrak{z}^*(s)=\sum_d \gamma^*(d)d^{-s}. \tag{19.4}$$

Next, by Lemma 19.1, if w is squarefree, then

$$\mathfrak{z}^*(s)=\zeta(2)^{-1}\zeta_D(s+1)\xi(w)\prod_{p|w}\left(1+\frac{\chi(p)}{p^2}\right). \tag{19.5}$$

This expression shows that $\mathfrak{z}^*(s)$ is analytic in the whole s-plane and it has only a simple pole at $s = 0$ with the residue $r = \lambda(w)\xi(w)\varphi(D)/\zeta(2)D$. Introducing (19.5) into (19.3) and moving the integration from $\operatorname{Re} s = \varepsilon$ to $\operatorname{Re} s = -1$ we get

$$k^*(y)=r\theta(0)-\frac{\Psi(0)\xi(w)}{2\zeta(2)}\sum_{c|w}\chi(c)\sum_{\substack{cdy<1\\(d,D)=1}}(1-cdy)d^{-1}+O(yw\tau(D)). \tag{19.6}$$

Furthermore, for any $X > 0$ we have

$$\sum_{\substack{d<X\\(d,D)=1}} (1-d/X)d^{-1} = \frac{\varphi(D)}{D}\,(\log X + \gamma - 1 - \alpha(D)) + O(\tau(D)/X), \tag{19.7}$$

where $\gamma = 0.57\ldots$ is the Euler constant and

$$\alpha(D) = \sum_{p|D} \frac{\log p}{p-1} \leqslant \log|D|. \tag{19.8}$$

Hence, the double sum in (19.6) is equal to

$$\frac{\varphi(D)}{D} \sum_{c|w} \chi(c)(-\log cy + \gamma - 1 - \alpha(D)) + O(y\tau(D)\mathfrak{S}_1(w)). \tag{19.9}$$

If $\chi(w) = 1$, which is our case by the mollifier support, then

$$\sum_{c|w} \chi(c) \log \sqrt{w}/c = 0. \tag{19.10}$$

To see this we switch c to its complementary divisor w/c. Hence the sum over c/w in (19.9) becomes $\lambda(w)(-\log y\sqrt{w} + \gamma - 1 - \alpha(D))$. Combining the above results we conclude the following approximate formula which is useful only if yw is small.

Lemma 19.3. *Suppose $w = uv$ is squarefree and $\chi(w) = 1$. For every $y > 0$ we have*

$$k^*(y) = \lambda(w)\xi(w)\frac{\varphi(D)}{2\zeta(2)D}\{\Psi(0)\log y\sqrt{w} + \alpha(D) + \alpha_0\} + O(yw\tau(D)) \tag{19.11}$$

where α_0 and the implied constant depend only on the fixed test function Φ.

20. Evaluation of $J^*(u, v)$

Recall that $J^*(u, v)$ is the integral

$$J^*(u,v) = \int_0^\infty k^*\left(\frac{T}{x}\right) h\left(\frac{x}{u}\right) h\left(\frac{x}{v}\right) \frac{dx}{x} \tag{20.1}$$

with the kernel $k^*(y)$ given by (17.12). Applying (19.2) we derive

$$J^*(u,v) = \frac{\xi(w)}{\zeta(2)} \sum_{c|w} \chi(c) \sum_{(d,D)=1} d^{-1} \int_0^\infty \phi\left(\frac{dT}{x}\right) h\left(\frac{cx}{u}\right) h\left(\frac{cx}{v}\right) \frac{dx}{x}. \tag{20.2}$$

In the sequel we shall use the following abbreviations:

$$\gamma = \frac{\log c}{\log N},\ \gamma_1 = \frac{\log u}{\log N},\ \gamma_2 = \frac{\log r}{\log N},\ \omega_1 = \frac{\log c/u}{\log N},\ \omega_2 = \frac{\log c/v}{\log N}.$$

Therefore, $\gamma = \gamma_1 + \omega_1 = \gamma_2 + \omega_2$. Note that all these five numbers are bounded in absolute value by

$$\mu = \frac{\log M}{\log N} < \beta - \frac{1}{2}$$

which is a small number. Moreover we shall be frequently changing the variable x to $t = \log x/\log N$, so

$$x = N^t,\ x^{-1}dx = (\log N)dt.$$

Now recall that in (5.2) we set $h(x) = a(t)$ and in (4.1) we requested $a(t)$ to be a smooth function on $\mathbb{R}$ with

$$\begin{aligned} a(t) &= 1-t, && \text{if } t \leqslant \beta,\\ 0 < a(t) &< 1-\beta, && \text{if } \beta < t < \alpha,\\ a(t) &= 0, && \text{if } t \geqslant \alpha, \end{aligned}$$

where the transition points $\alpha > \beta > 1/2$ are close to $1/2$. Moreover, recall that $N = Q^2T^2 = |D|(T/2\pi)^2$ and $\log|D|/\log T$ is very small so $\log N$ is close to $2\log T$.

We break the integration at $x = TM^2 = X$, say, and write respectively

$$J^*(u,v) = \frac{\xi(w)}{\zeta(2)}\left(J_1(u,v) + J_2(u,v)\right). \tag{20.3}$$

In the first part we have $h(cx/u)h(cx/v) = (1-t-\omega_1)(1-t-\omega_2) = (1-t)^2 - (\omega_1+\omega_2)(1-t) + \omega_1\omega_2$. Note that

$$\sum_{c|w}\chi(c) = \lambda(w),\ \sum_{c|w}\chi(c)(\omega_1+\omega_2) = 0,\ \sum_{c|w}\chi(c)\omega_1\omega_2 = \lambda(u,v)(\log N)^{-2}$$

where

$$\lambda(u,v) = \sum_{c|uv}\chi(c)\log\frac{c}{u}\log\frac{c}{v}. \tag{20.4}$$

The vanishing of the middle sum above follows from (19.10). We shall compute $\lambda(u,v)$ and other alike arithmetic functions in the next section. Now we have

$$\begin{aligned} J_1(u,v) &= \sum_{(d,D)=1} d^{-1}\int_0^X \phi\left(\frac{dT}{x}\right)\left(\lambda(w)(1-t)^2 + \lambda(u,v)(\log N)^{-2}\right)\frac{dx}{x}\\ &= \lambda(w)A_0(\log N)^2 + \lambda(u,v)A_1, \end{aligned} \tag{20.5}$$

with $A_0 \ll \varphi(D)/|D|$ and $A_1 \ll \varphi(D)/|D|$ which are independent of u, v.

In the second part $J_2(u,v)$ the integration starts at $x = X$ so there is a room for the variable d. We apply (18.7) and (19.7) getting

$$\begin{aligned} &\sum_{(d,D)=1} d^{-1}\phi\left(\frac{dT}{x}\right) = \sum_{(d,D)=1} d^{-1}\left(\phi_0\left(\frac{dT}{x}\right) - \frac{1}{2}\Psi(0)\left(1-\frac{dT}{x}\right)^t\right)\\ =\ &\frac{\varphi(D)}{D}\int_0^\infty \phi_0(z)\frac{dz}{z} - \Psi(0)\frac{\varphi(D)}{2D}\left(\log\frac{x}{T} + \gamma - 1 + \alpha(D)\right)\\ &+O(\tau(D)T/x)\\ =\ &A\log N - B\log x + O(\tau(D)T/x), \end{aligned}$$

say, with

$$A\log N = \frac{\varphi(D)}{D}\int_0^\infty \phi_0(z)z^{-1}dz + \Psi(0)\frac{\varphi(D)}{2D}(\log T - \gamma + 1 - \alpha(D)) \tag{20.6}$$

and

$$B = \Psi(0)\varphi(D)/2D. \tag{20.7}$$

We have bounds

$$A \ll \varphi(D)/|D|, \quad B \ll \varphi(D)/|D| \tag{20.8}$$

and we need nothing else to know about A, B. The error term $O(\tau(D)T/x)$ contributes to $J^\infty(u,v)$ at most

$$\tau(w)\tau(D)T\int_X^\infty x^{-2}dx = \tau(w)\tau(D)M^{-2} < T^{-1/400}.$$

By the above estimates we get $J_2(u,v) = J_{20}(u,v)+O(T^{-1/400})$ with $J_{20}(u,v)$ equal to

$$\begin{aligned}&\sum_{c|w}\chi(c)\int_X^\infty (A\log N - B\log x)\, h\left(\frac{cx}{u}\right)h\left(\frac{cx}{v}\right)\frac{dx}{x}\\ =\ &\sum_{c|w}\chi(c)\int_{Xc/\sqrt{w}}^\infty \left(A\log N - B\log\frac{x\sqrt{w}}{c}\right)h\left(x\sqrt{u/v}\right)h\left(x\sqrt{v/u}\right)\frac{dx}{x}.\end{aligned}$$

If the integration starts from X we get an elegant quantity (see (19.10))

$$K(u/v) = \int_X^\infty (A - B\log x/\log N)\, h(x\sqrt{u/v})h(x\sqrt{v/u})x^{-1}dx \tag{20.9}$$

and

$$J_{20}(u,v) = \lambda(w)K(u/v)\log N. \tag{20.10}$$

Estimating trivially we get $K(u/v) \ll \log N$. This bound has true order of magnitude, but it is not useful, because we shall need a clear view on the dependence on u/v. The remaining part is equal to

$$\begin{aligned}J_{22}(u,v) = \sum_{c|w}\chi(c)\int_{Xc/\sqrt{w}}^X \quad &\left(A\log N - B\log\frac{x\sqrt{w}}{c}\right)\\ &\left[\left(1-\frac{\log x}{\log N}\right)^2 - \left(\frac{\log u/v}{2\log N}\right)^2\right]\frac{dx}{x}.\end{aligned}$$

Put

$$\nu = \log X/\log N = 2\mu + \log T/\log N = 2\mu + \frac{1}{2}\left(1+\frac{\log Q}{\log T}\right)^{-1},$$

$$\delta = \frac{1}{2}(\omega_1+\omega_2) = \frac{1}{2}\log(c/\sqrt{w})/\log N.$$

In this notation we have

$$J_{22}(u,v) = (\log N)^2\sum_{c|w}\chi(c)P(\delta) \tag{20.11}$$

where $P(\delta)$ is the polynomial in δ of degree five given by

$$\begin{aligned} P(\delta) &= \int_{\nu+\delta}^{\nu} (A - Bt + B\delta)\left((1-t)^2 - \frac{1}{4}(\gamma_1-\gamma_2)^2\right) \\ &= \frac{3B}{8}(\gamma_1-\gamma_2)^2\delta^2 + \frac{1-\nu}{2}(2A-3B+B\nu)\delta^2 - \frac{B}{12}\delta^4 \\ &\quad + \text{odd degree monomials.} \end{aligned} \tag{20.12}$$

Since

$$\sum_{c|w} \chi(c)\delta^j = 0, \quad \text{if } 2 \nmid j,$$

we do not need the odd degree monomials. We get

$$J_{22}(u,v) = \lambda_2(uv)A_2 + \lambda_4(uv)A_4(\log N)^{-2} \tag{20.13}$$

where

$$\lambda_j(w) = \sum_{c|w} \chi(c)(\log c/\sqrt{w})^j \tag{20.14}$$

and $A_2 \ll \varphi(D)/|D|$, $A_4 \ll \varphi(D)/|D|$ do not depend on u, v. Gathering the above results we arrive at the following representation of $J^*(u,v)$.

Lemma 20.1. *Suppose $w = uv$ is squarefree and $\chi(w) = 1$. Then*

$$\begin{aligned} J^*(u,v)\zeta(2)/\xi(w) &= \lambda(w)K(u/v)\log N + \lambda(w)A_0(\log N)^2 \\ &\quad + \lambda(u,v)A_1 + \lambda_2(w)A_2 + \lambda_4(w)A_4(\log N)^{-2} \\ &\quad + O(T^{-1/400}), \end{aligned} \tag{20.15}$$

where $K(u/v)$, $\lambda(u,v)$, $\lambda_2(w)$, $\lambda_4(w)$ are given by (20.9), (20.4), (20.14), *respectively. Moreover A_0, A_1, A_2, A_4 do not depend on u, v and they are $\ll \varphi(D)/|D|$.*

Remarks. It is essential that $K(u/v)$ depends on the ratio u/v rather than on u, v respectively. After computing $\lambda(u,v), \lambda_2(w), \lambda_4(w)$ in the next section we shall see that all the terms in (20.15) look alike and each one has the order of magnitude $\lambda(w)(\log N)^2$ (except for the negligible error term). The formula (20.15) displays the behaviour in terms of u, v as needed, but it is long, so we wish to say that our arrangements could have been quicker if we applied Taylor's expansion of $a(t)$. This would bring polynomials in δ of arbitrary degree; consequently we would have struggled with the uniformity in the resulting series coefficients, which is a formidable task. The fact that we are dealing here with $P(\delta)$ of degree five is due to the linearity of $a(t)$ in the long segment $t \leqslant \beta$ with β slightly larger than $1/2$.

21. Computing the λ-functions

Recall that $\lambda(u,v)$ and $\lambda_j(w)$ are defined by convolutions of χ against powers of logarithms. In this sections we arrange these as convolutions of 1 against the von Mangoldt functions

$$\Lambda_j = \mu \star (\log)^j.$$

First writing $(\log c/u)(\log c/v) = (\log c/\sqrt{w})^2 - (\log\sqrt{u/v})^2$ in (20.4) we find that

$$\lambda(u,v) = \lambda_2(w) - \lambda(w)\left(\frac{1}{2}\log\frac{u}{v}\right)^2. \tag{21.1}$$

Next, writing

$$\lambda_j(uv) = \sum_{c|u}\sum_{d|v}\chi(cd)\left(\log\frac{c}{\sqrt{u}} + \log\frac{d}{\sqrt{v}}\right)^j$$

we find that

$$\lambda_j(uv) = \sum_{a+b=j}\binom{j}{a}\lambda_a(u)\lambda_b(v). \tag{21.2}$$

Observe that $\lambda_a(u) = 0$ if a is odd, so a, b run in (21.2) over even numbers. For example we get

$$\lambda_2(uv) = \lambda(v)\lambda_2(u) + \lambda(u)\lambda_2(v) \tag{21.3}$$

$$\lambda_4(uv) = \lambda(v)\lambda_4(u) + 6\lambda_2(u)\lambda_2(v) + \lambda(u)\lambda_4(v). \tag{21.4}$$

We shall attach the λ-functions to the mollifier factors $\rho(eu)\rho(ev)$ which vanish if $\lambda(p) = 0$ for some $p \mid euv$, see (17.13). Therefore, for computing $\lambda_j(u)$ we can assume that $\lambda(u) \neq 0$, in which case

$$\lambda(q) = \tau(q/(q,D)), \text{ if } q \mid u. \tag{21.5}$$

We compute $\lambda_j(u)$ as follows

$$\begin{aligned} 2^j\lambda_j(u) &= \sum_{c|u}\chi(c)\left(\log c - \log\frac{u}{c}\right)^j \\ &= \sum_{a+b=j}\binom{j}{a}(-1)^b(\chi\log^a)\star(\log^b). \end{aligned}$$

Here we write $(\chi\log^a)\star(\log^b) = \chi(1\star\Lambda_a)\star(1\star\Lambda_b) = \chi\star 1\star\chi\Lambda_a\star\Lambda_b = \lambda\star\chi\Lambda_a\star\Lambda_b$, and then

$$\sum_{\ell mn=u}\lambda(\ell)\chi(m)\Lambda_a(m)\Lambda_b(n) = \lambda(u)\sum_{mn|u}\chi(m)\frac{\tau(n,D)}{\tau(mn)}\Lambda_a(m)\Lambda_b(n).$$

Adding the above expressions we obtain

Lemma 21.1. *Suppose u is squarefree with $\lambda(u) \neq 0$. Then we have*

$$\lambda_j(u) = \lambda(u)\sum_{q|u}\Lambda_j^*(q) \tag{21.6}$$

where

$$\Lambda_j^*(q) = \frac{\tau((q,D))}{2^j\tau(q)}\sum_{mn=q}\chi(m)\sum_{a+b=j}\binom{j}{a}(-1)^b\Lambda_a(m)\Lambda_b(n). \tag{21.7}$$

We do not need to know $\Lambda_j^*(q)$ exactly, the following estimate is good enough

$$|\Lambda_j^*(q)| \leqslant 2^{-j}\Lambda_j(q). \tag{21.8}$$

Hence $\Lambda_j^*(q)$ is supported on numbers having at most j distinct prime divisors. Moreover we get

$$\sum_{q\leqslant x} |\Lambda^*(q)|q^{-1} \ll (\log x)^j. \tag{21.9}$$

Note that (21.6) holds for $\lambda_0(u) = \lambda(u)$ with $\Lambda_0^*(q) = \Lambda_0(q)$.

Inserting (21.6) into (21.2) we obtain the following result

Lemma 21.2. *Suppose $w = uv$ is squarefree with $\lambda(w) \neq 0$. Then we have*

$$\lambda_j(w) = \lambda(w) \sum_{a+b=j} \binom{j}{a} \sum_{q|u}\sum_{r|v} \Lambda_a^*(q)\Lambda_b^*(r). \tag{21.10}$$

The second part of (21.1) can be written in the same fashion, exactly we have

$$\Big(\log\frac{u}{v}\Big)^2 = \sum_{q|u}\Lambda_2(q) - 2\sum_{q|u}\sum_{r|v}\Lambda(q)\Lambda(r) + \sum_{r|v}\Lambda_2(r). \tag{21.11}$$

By (21.3) and (21.6) we get

$$\lambda_2(w) = \lambda(w)\left(\sum_{q|u}\Lambda_2^*(q) + \sum_{r|v}\Lambda_2^*(r)\right) \tag{21.12}$$

and

Lemma 21.3. *Suppose $w = uv$ is squarefree with $\lambda(w) \neq 0$. Then we have*

$$\begin{aligned}\lambda(u,v) = \lambda(w)\Bigg\{&\sum_{q|u}\Big(\Lambda_2^*(q) - \frac{1}{4}\Lambda_2(q)\Big) \quad + \quad \sum_{r|v}\Big(\Lambda_2^*(r) - \frac{1}{4}\Lambda_2(r)\Big)\\ &+ \quad \frac{1}{2}\sum_{q|u}\sum_{r|v}\Lambda(q)\Lambda(r)\Bigg\}.\end{aligned} \tag{21.13}$$

We conclude this section by combining the results into a compact implicit form

Lemma 21.4. *Suppose $w = uv$ is squarefree with $\lambda(w) \neq 0$. Then*

$$\begin{aligned}J^*(u,v) = \quad & \lambda(w)\xi(w)(\log N)^2/\zeta(2)\\ &\left\{\frac{K(u,v)}{\log N} + \sum_{a+b\leqslant 4} c(a,b)(\log N)^{-a-b}\sum_{q|u}\sum_{r|v}\Lambda_a^*(q)\Lambda_b^*(r)\right\}\end{aligned} \tag{21.14}$$

where $K(u,v)$ is given by (20.9) *and the coefficients $c(a,b)$ do not depend on u, v and they satisfy $c(a,b) \ll \varphi(D)/|D|$. Moreover $\Lambda_a^*(q), \Lambda_b^*(r)$ given by* (21.7) *are supported on numbers having at most a, b prime factors, respectively and they satisfy the bound* (21.8)*, a fortiori* (21.9)*.*

22. Estimating E_{ab}

According to (21.14) the formula (17.11) splits into

$$\mathcal{K}^*(T) = \left\{F(\log N)^{-1} + \sum_{a+b\leqslant 4} c(a,b)(\log N)^{-a-b}E_{ab}\right\}(\log N)^2\zeta(2)^{-1}$$

$$+O(T^{-1/401}) \tag{22.1}$$

where

$$E = \sum_e \sum_{(u,v)=1}\sum \frac{\rho(eu)\rho(ev)}{euv} g(eu)g(ev)\lambda(uv)\xi(uv)K(u/v), \tag{22.2}$$

and

$$E_{ab} = \sum_e \sum_{(u,v)=1}\sum \frac{\rho(eu)\rho(ev)}{euv} g(eu)g(ev)\lambda(uv)\xi(uv)\sum_{q|u}\Lambda_a^*(q)\sum_{r|v}\Lambda_b^*(r). \tag{22.3}$$

Many arguments for estimating E and E_{ab} are reminiscent of these applied to the diagonal terms in early sections. All cases are similar, but E needs extra attention.

First we do E_{00}, because it is a notationally simpler model for every E_{ab}. In this case (22.3) reduces to

$$E_{00} = \sum_e \sum_{(u,v)=1}\sum \frac{\rho(eu)\rho(ev)}{euv} g(eu)g(ev)\lambda(uv)\xi(uv). \tag{22.4}$$

Observe that the total contribution of E_{00} to the off-diagonal part $\mathcal{K}^{\neq}(T)$ (see (17.10) and (17.11)) is equal to

$$V_{00} = c(0,0)\zeta(2)^{-1}(L(1,\chi)\log N)^2 E_{00}. \tag{22.5}$$

If E_{00} is shown to be bounded (as expected due to the sifting effects) the small factor $L(1,\chi)\log N$ (due to the lacunarity effect) yields the vital saving two times. However, it is hard to show that $E_{00} \ll 1$, because twisting the Möbius function with the character χ does not exactly work that way in our exceptional scenario. An attempt to execute the summation in E_{00} by routine arguments fails as one cannot keep track in the conductor aspect. Therefore we take a roundabout path. We shall loose some portion of the saving factor $(L(1,\chi)\log N)^2$, but fortunately not the whole saving.

We start by reconstructing E_{00} from the following expression

$$W = \left(\sum_{N<\ell\leqslant N^2} - \sum_{N^2<\ell\leqslant N^3}\right)\frac{\tilde\lambda(\ell)}{\ell}\left(\sum_{m|\ell}\mu(m)\frac{g(m)}{\xi(m)}\right)^2 \tag{22.6}$$

where $\tilde\lambda(\ell)$ is the completely multiplicative function with

$$\tilde\lambda(p) = (\lambda(p)\xi(p))^2 = (1+\chi(p))^2(1+\chi(p)/p)^{-2}. \tag{22.7}$$

Opening the square we get

$$W = \sum_e \sum_{(u,v)=1}\sum \frac{\mu(eu)\mu(ev)}{euv}\frac{g(eu)g(ev)}{\xi(ev)\xi(ev)}\tilde\lambda(euv)\left(\sum_k - \sum_k\right)\frac{\tilde\lambda(k)}{k}$$

where k runs over the segments $N/euv < k \leqslant N^2/euv$ and $N^2/euv < k \leqslant N^3/euv$, respectively. Note that the above triple sum preceding the sums over k matches that in (22.4) (check this using (17.13) and our choice (22.7)).

The generating Dirichlet series of $\tilde{\lambda}(k)$ is

$$\tilde{L}(s) = \sum \tilde{\lambda}(k)k^{-s} = \prod_p (1 - \tilde{\lambda}(p)p^{-s})^{-1} = L(s)^2\tilde{R}(s)$$

where the "correcting" factor $\tilde{R}(s)$ is given by the Euler product

$$\tilde{R}(s) = \prod_p (1 - \tilde{\lambda}(p)p^{-s})^{-1}(1 - p^{-s})^2(1 - \chi(p)p^{-s})^2$$

which converges absolutely in $\operatorname{Re} s > \frac{1}{2}$. For $s = 1$ we have

$$\tilde{R}(1) = \frac{1}{\zeta(2)} \prod_{p|D} \left(1 + \frac{1}{p}\right) \asymp \frac{\varphi(D)}{|D|}.$$

Hence, by standard contour integration, we derive

$$\begin{aligned}\left(\sum_k - \sum_k\right) \frac{\tilde{\lambda}(k)}{k} &= \operatorname{res}_{s=0} \tilde{L}(s+1)s^{-1}(N^{2s} - N^s - N^{3s} + N^{2s})(euv)^{-s} \\ &\quad + O(N^{-1/4}) = -\tilde{R}(1)(L(1,\chi)\log N)^2 + O(N^{-1/4}).\end{aligned}$$

Hence

$$W = -\tilde{R}(1)(L(1,\chi)\log N)^2 E_{00} + O(T^{-1/4}). \tag{22.8}$$

It looks like we have lost the entire saving factor $(L(1,\chi)\log N)^2$. Not true, because W is small due to the lacunarity of $\tilde{\lambda}(\ell)$. To estimate W we proceed along the lines in Section 9 getting

$$W \ll \sum_{N<\ell\leqslant N^3} \tilde{\lambda}(\ell)\ell^{-1}\Big(\sum_{m|\ell}\Big)^2 \ll \Big(\sum_\ell \lambda(\ell)\ell^{-1}\Big)^{\frac{1}{2}}\Big(\sum_\ell \tau(\ell)^4\ell^{-1}\Big(\sum_{m|\ell}\Big)^2\Big)^{\frac{1}{2}}.$$

Hence, by the same arguments as in the proof of Corollary (9.2) we get

$$W \ll (L(1,\chi)\log N)^{\frac{1}{2}}. \tag{22.9}$$

The almost constant multiplicative function $1/\xi(m)$ in (22.6) makes no difference to the arguments ($1/\xi(p) = 1 + \chi(p)/p$). Combining (22.8) with (22.9) we conclude that

$$E_{00} \ll \frac{|D|}{\varphi(D)} (L(1,\chi)\log N)^{-\frac{3}{2}}. \tag{22.10}$$

Finally, introducing (22.10) into (22.5) we get (recall $c(0,0) \ll \varphi(D)/|D|$)

$$V_{00} \ll (L(1,\chi)\log N)^{\frac{1}{2}}. \tag{22.11}$$

The other sums E_{ab} can be reduced to E_{00} by scaling the crop function of the mollifier. Specifically we write

$$E_{ab} = \sum_{\substack{(q,r)=1\\ q,r<M}} \sum \Lambda_a^*(q)\Lambda_b^*(r)\frac{\rho(q)\rho(r)}{qr}\lambda(qr)\xi(qr)E_{00}(q,r)$$

where $E_{00}(q,r)$ stands for E_{00} with the crop functions $g(eu), g(ev)$ replaced by $g(equ), g(erv)$ and the summation variables are restricted by the co-primality condition: $(euv, qr) = 1$. This co-primality condition does not really spoil the previous treatment of E_{00} and the fact that the scaled down crop functions $g_q(m) = g(qm)$, $g_r(n) = g(rn)$ are not equal does not matter neither (the subsequent application of Cauchy's inequality resolves this discrepancy). However we need to address the scaling effect. Writing

$$g_q(m) = g(qm) = \left(\frac{\log M/q}{\log M}\right)^r \left(\frac{\log^+ M/qm}{\log M/q}\right)^r$$

it boils down to changing the support range M into M/q and correcting the relevant estimates by factor $((\log M/q)/\log M)^r$. Well, not in full range, because Lemma 9.1 used for estimating W requires $Q^4 < M$, which condition translates into $Q^4 < M/q$. But in the range $M/q \leqslant Q^4$ we can apply Lemma 7.2 (precisely its relevant analogue). Adding the resulting estimates, we derive the following bound

$$W_{ab}(q) \ll (L(1,\chi)\log N)^{\frac{1}{2}} \left(\frac{\log M/q}{\log M}\frac{\log N}{\log M/q}\right)^{\frac{r}{2}} \ll (L(1,\chi)\log N)^{\frac{1}{2}}.$$

The same bound holds for $W_{ab}(r)$ (here $W_{ab}(q)$ and $W_{ab}(r)$ denote the sums of type W with the crop function g replaced by g_q and g_r, respectively). Moreover we have

$$\left(\sum_{q<M} \Lambda_a^*(q) q^{-1}\right)\left(\sum_{r<M} \Lambda_b^*(r) r^{-1}\right) \ll (\log M)^{a+b}.$$

Hence

$$E_{ab} \ll \frac{|D|}{\varphi(D)} (\log M)^{a+b} (L(1,\chi)\log N)^{\frac{3}{2}} \tag{22.12}$$

and the total contribution of E_{ab} to the off-diagonal part $\mathcal{K}^{\neq}(T)$, say V_{ab}, (see (17.10), (17.11), (22.1)) satisfies

$$V_{ab} \ll (L(1,\chi)\log N)^{\frac{1}{2}}. \tag{22.13}$$

23. Estimating E

The case of E can be regarded as a generalization of E_{00} in which the extra kernel function $K(u/v)$ is introduced. We reconstruct E from the following expression

(23.1)

$$W = \left(\sum_{N<\ell\leqslant N^2} - \sum_{N^2<\ell\leqslant N^3}\right) \frac{\tilde\lambda(\ell)}{\ell} \sum_{m_1|\ell}\sum_{m_2|\ell} \mu(m_1)\mu(m_2) \frac{g(m_1)}{\xi(m_1)}\frac{g(m_2)}{\xi(m_2)} K\left(\frac{m_1}{m_2}\right).$$

The same arguments which produced (22.8) now yield

$$W = -\tilde R(1)\,(L(1,\chi)\log N)^2\, E + O(T^{-1/4}). \tag{23.2}$$

Before estimating W we need to separate m_1, m_2 in $K(m_1/m_2)$. This can be accomplished quickly by changing the variable of integration x in (20.9)

into $x/\sqrt{m_1m_2}$ giving

$$K\left(\frac{m_1}{m_2}\right)=\int_{X\sqrt{m_1m_2}}^{N}\left(A-B\frac{\log x/\sqrt{m_1m_2}}{\log N}\right)h\left(\frac{x}{m_1}\right)h\left(\frac{x}{m_2}\right)\frac{dx}{x}.$$

Recall that $N=Q^2T^2$ and $X=M^2T$. Since $h(x)=a(t)$ is linear in $t=\log x/\log N$ for $t<\beta$ we can write

$$K\left(\frac{m_1}{m_2}\right)=\int_X^N \left(A-B\frac{\log x}{\log N}+\frac{B}{2}\frac{\log m_1}{\log N}+\frac{B}{2}\frac{\log m_2}{\log N}\right)$$
$$h\left(\frac{x}{m_1}\right)h\left(\frac{x}{m_2}\right)\frac{dx}{x}+P\left(\frac{\log m_1}{\log N},\frac{\log m_2}{\log N}\right)\log M \tag{23.3}$$

where $P(x_1,x_2)$ is a polynomial of degree five with coefficients $\ll \varphi(D)/|D|$ (they are linear forms in A,B). From the above expressions the following convolution sums emerge (a kind of sifting weights);

$$\theta_a(\ell)=\sum_{m|\ell}g(m)\frac{\mu(m)}{\xi(m)}\left(\frac{\log m}{\log N}\right)^a \quad (0\leqslant a\leqslant 5), \tag{23.4}$$

$$\theta_a(\ell;x)=\sum_{m|\ell}g(m)\frac{\mu(m)}{\xi(m)}\left(\frac{\log m}{\log N}\right)^a h\left(\frac{x}{m}\right) \quad (0\leqslant a\leqslant 1). \tag{23.5}$$

Hence the double sum over the divisors of ℓ in (23.1) splits into the integral

$$\int_X^N\left[\left(A-B\frac{\log x}{\log N}\right)\theta_0(\ell;x)+B\theta_1(\ell;x)\right]\theta_0(\ell;x)\frac{dx}{x} \tag{23.6}$$

and 36 terms of type $c(a,b)\theta_a(\ell)\theta_b(\ell)\log M$ for $0\leqslant a,b\leqslant 5$ with coefficients $c(a,b)\ll\varphi(D)/|D|$. Hence (23.1) yields

$$W\ll\sum_{N<\ell<N^3}\tilde{\lambda}(\ell)\ell^{-1}(\theta_a(\ell)^2+\theta_b(\ell;x)^2)\frac{\varphi(D)}{|D|}\log N$$

for some $0\leqslant a\leqslant 5$, $0\leqslant b\leqslant 1$ and $X<x<N$. By the arguments in Section 11 we show that (compare it with Lemma 9.1)

$$\sum_{N<\ell<N^3}\tau(\ell)^4\ell^{-1}\left(\theta_a(\ell)^2+\theta_b(\ell;x)^2\right)\ll 1. \tag{23.7}$$

Hence we derive in the same way as (22.9) that

$$W\ll(L(1,\chi)\log N)^{\frac{1}{2}}\,\frac{\varphi(D)}{|D|}\log N. \tag{23.8}$$

On the other hand we have the formula (23.2), comparing these we get

$$E\ll(L(1,\chi)\log N)^{-\frac{3}{2}}\log N. \tag{23.9}$$

Finally it shows (see (17.10) and (22.1)) that the total contribution of E to the off-diagonal $\mathcal{K}^{\neq}(T)$, say V, satisfies

$$V\ll(L(1,\chi)\log N)^{\frac{1}{2}}. \tag{23.10}$$

24. Comments about Completing the Proof

After having completed the treatment of the diagonal terms we wrapped up the results in the lower bound (12.1) for $N_{00}(T)$ in which the off-diagonal contribution $\mathcal{K}^{\neq}(T)$ is postponed for handling in the rest of the paper. We are now ready to finish the job by compiling the relevant results.

According to the formula (16.6) we partitioned $\mathcal{K}^{\neq}(T)$ into two similar parts $L^2(1,\chi)\mathcal{K}^*(T) - L^2(1,\chi)\mathcal{K}'(T)$, see (17.10) for exact formula. The third part was eliminated earlier by making the assumption (16.10). This means that the mollifier misses terms which are supported on multiples of D. However the contribution of the missing terms can be treated by undemanding arguments. For example one can show by straightforward estimations that the missing part in the integral $I(T)$ (see (3.5)) is bounded by $|D|^{-\frac{1}{4}}T(\log T)^4$ which is negligible.

Next $\mathcal{K}^*(T)$ was split further into a number of pieces of similar shape (see (22.1)) and the contribution to $\mathcal{K}^{\neq}(T)$ of every piece separately was shown to satisfy the same bound (22.11), (22.13), (23.10). This bound does not exceed the existing error term in the lower bound (12.1). Finally it remains to cover $\mathcal{K}'(T)$. We have gone quite far towards $\mathcal{K}'(T)$ by computing its constituents (see Lemma 16.2) until the case appeared merging the lines of $\mathcal{K}^*(T)$. Without repeating many of the same arguments we claim that $\mathcal{K}'(T)$ contributes no more than $\mathcal{K}^*(T)$.

References

[CI02] J.B. Conrey and H. Iwaniec. Spacing of zeros of Hecke L-functions and the class number problem. *Acta Arith.*, 103:259–312, 2002.

[CIS13] J.B. Conrey, H. Iwaniec, and K. Soundararajan. Critical zeros of Dirichlet L-functions. *J. Reine Angew. Math.*, 681:175–198, 2013.

[Con89] J.B. Conrey. More than two fifths of the zeros of the Riemann zeta function are on the critical line. *J. Reine Angew. Math.*, 339:1–26, 1989.

[IK04] H. Iwaniec and E. Kowalski. *Analytic Number Theory*. Colloquium Publications. American Mathematical Society, 2004.

[KMV02] E. Kowalski, P. Michel, and J. Vanderkam. Rankin-Selberg L-functions in the level aspect. *Duke Math. J.*, 114(1):123–191, 2002.

[Lev74] N. Levinson. More than one third of zeros of Riemann's zeta function are on $\sigma = \frac{1}{2}$. *Advances in Math.*, 13:383–436, 1974.

[Sel42] A. Selberg. On the zeros of Riemann's zeta function on the cirtical line. *Arch. Math. Naturvid.*, 45(9):101–114, 1942.

American Institute of Mathematics, 360 Portage Ave, Palo Alto, CA 94306, USA.

E-mail address: conrey@aimath.org

The ETH Institute for Theoretical Studies, Clausiusstrasse 47, 8092 Zürich, Switzerland,

Department of Mathematics, Rutgers University, New Brunswick, NJ 08903, USA.

E-mail address: iwaniec@math.rutgers.edu