THE RANK OF THE JACOBIAN OF MODULAR CURVES: ANALYTIC METHODS

BY EMMANUEL KOWALSKI

Preface

The interaction of analytic and algebraic methods in number theory is as old as Euler, and assumes many guises. Of course, the basic algebraic structures are ever present in any modern mathematical theory, and analytic number theory is no exception, but to algebraic geometry in particular it is indebted for the tremendous advances in understanding of exponential sums over finite fields, since André Weil's proof of the Riemann hypothesis for curves and subsequent deduction of the optimal bound for Kloosterman sums to prime moduli.

On the other hand, algebraic number theory has often used input from L-functions; not only as a source of results, although few deep theorems in this area are proved without some appeal to Tchebotarev's density theorem, but also as a source of inspiration, ideas and problems.

One particular subject in arithmetic algebraic geometry which is now expected to benefit from analytic methods is the study of the rank of the Mordell-Weil group of an elliptic curve, or more generally of an abelian variety, over a number field. The beautiful conjecture of Birch and Swinnerton-Dyer asserts that this deep arithmetic invariant can be recovered from the order of vanishing of the L-function of the abelian variety at the center of the critical strip.

This conjecture naturally opens two lines of investigation: to try to prove it, but here one is, in general, hampered by the necessary prerequisite of proving analytic continuation of the L-function up to this critical point, before any attempt can be made; or to take it for granted and use the information and insight it gives into the nature of the rank as a means of exploring further its behavior. This is justified by the trust put into the truth of the conjecture.

Indeed, the first approach has been quite successful: in some cases, most notably large classes of elliptic curves over $\mathbf{Q}$, analytic continuation is known, and partial results towards the conjecture have been obtained when the rank is 0 or 1.

On the other hand, even when this is so, the second approach has had the aesthetic disadvantage that most studies of the rank, whether based on the assumption of the full statement of the conjecture or on known cases of it, have also assumed other analytic facts about the L-function, most notably that it satisfies the Generalized Riemann Hypothesis. This is somewhat unsatisfactory, inasmuch as this appears to be a much harder problem than even the Birch and Swinnerton-Dyer conjecture, although zeros of the L-function are of course very relevant to the problem.

The contribution of this thesis is to show that analytic methods and techniques can indeed provide sharp, unconditional answers to some of the questions thus raised. This demonstrates that the implicit promise of the conjecture of Birch and Swinnerton-Dyer, of furnishing an effective way of answering questions about the rank through its analytic interpretation, can be kept without additional assumptions.

The main results have been obtained in collaboration with Philippe Michel, and some auxiliary propositions had been proved earlier in the course of other work with

William D. Duke.

This volume is organized in six chapters. The first contains an introduction to the theory of abelian varieties and the Birch and Swinnerton-Dyer conjecture which is the motivating problem, and ends with the precise statements of the two principal theorems. The second chapter takes up the analytic side of the story. It recalls the results of Eichler-Shimura and Gross-Zagier which make the link between the algebraic geometry and modular forms, and provides an informal, but quite detailed, sketch of the proofs of the theorems. The extent of this first part, which is not original, stems from the fact that whereas the motivating problem lies in arithmetic algebraic geometry, an almost complete translation to a problem of analytic number theory is made, and this problem has intrinsic interest. Perchance, readers of both backgrounds will want to look at this document, and a goal of the text is to give to all an understanding of the other side of the story.

The preliminaries over, at last, the process of proving is engaged with a stiff upper-lip. The third chapter contains a result about the "almost-orthogonality" of the symmetric squares of modular forms which is crucial later for both results, and the fourth deals with another aspect of this kind of orthogonality principle. Then the last two chapters take each theorem in turn. There are a number of similarities in the principles and in some of the steps of both proofs, but since they seem to hold the same virtues of attraction and worth, the ordering is rather arbitrary and to accommodate the random-minded reader whose interest would lie in only one of the two, a certain amount of redundancy has been introduced, or cross-references sometimes inserted.

A conclusion comes, not surprisingly, to conclude all this with some reflections about the meaning of the results and possible developments.

Table of Contents

Notations

We introduce here some basic notations.

- For any set X, $|X|$ denotes its cardinality, a natural number if X is finite, and $+\infty$ if it is infinite.
- If A is an algebraic variety defined over a field k and K is an extension of k, $A(K)$ is the set of points of A with coordinates in K, also called K-rational points.
- $\mathbf{F}_q$ denotes a field with q elements ($q = p^n$ is the power of a prime number p), with $\mathbf{F}_p = \mathbf{Z}/p\mathbf{Z}$ for a prime p.
- If $k/\mathbf{Q}$ is a number field, $\mathcal{O}_k$ will denote its ring of integers. The norm of an ideal $\mathfrak{a} \subset \mathcal{O}_k$ is written $N\mathfrak{a}$.
- In any group G, the subgroup generated by a subset $H \subset G$ is denoted by $<H>$; if $H = \{h\}$ is a singleton, we also write simply $<h>$.
- For any ring A, $A^\times$ is the group of units of A.
- For any two objects X and Y of any category, $X \simeq Y$ means that X and Y are isomorphic.
- The notation χ_D designates the Kronecker symbol of a quadratic field with discriminant D.
- If σ is an automorphism of a field K, the action of σ on $x \in K$ is denoted by $\sigma(x)$ as well as by x^σ.

The following pertain to analytic number theory.

- The notation $n \sim N$, for real numbers n and N, means $N < n \leqslant 2N$.
- For any integer $q \geqslant 1$, ε_q is the trivial Dirichlet character modulo q, and $\zeta_q(s) = L(s, \varepsilon_q)$ is the corresponding L-function, the Riemann Zeta function with the Euler factors for $p \mid q$ removed.
- We denote by τ, φ, μ, Λ the classical arithmetic functions, namely (respectively) the divisor function, the Euler function, the Möbius function and the Van Mangoldt function, so

$$\sum_{n\geqslant 1} \tau(n)n^{-s} = \zeta(s)^2, \quad \sum_{n\geqslant 1} \varphi(n)n^{-s} = \frac{\zeta(s-1)}{\zeta(s)}$$

$$\sum_{n\geqslant 1} \mu(n)n^{-s} = \zeta(s)^{-1}, \quad \sum_{n\geqslant 1} \Lambda(n)n^{-s} = -\frac{\zeta'(s)}{\zeta(s)}.$$

- We also introduce the multiplicative functions N and M defined by

$$N(r) = \prod_{p|r} p, \quad M(r) = \prod_{p||r} p$$

(so N is the squarefree kernel).

- A summation over a family of Dirichlet characters of the form

$$\sum_{\chi}^{*} \alpha_\chi$$

means that χ runs only through the primitive characters in the family. Similarly, a summation of the form

$$\sum_{n \bmod r}^{*} \alpha_n$$

means that x runs only through invertible classes modulo r, and when a fixed modulus q is clear from the context

$$\sum_{n \leqslant N}^{*} \alpha_n$$

means that the summation is restricted to integers coprime with q.

- For $z \in \mathbf{C}$, $e(z) = e^{2\pi i z}$.

- For integers $m \geqslant 0$, $n \geqslant 0$ and $c \geqslant 1$, the Kloosterman sum $S(m, n; c)$ is

$$S(m, n; c) = \sum_{x \bmod c}^{*} e\Big(\frac{mx + n\overline{x}}{c}\Big)$$

where $\overline{x}$ is the inverse of x modulo c. For $n = 0$, this is the Ramanujan sum which has another expression as

$$S(m, 0; c) = \sum_{d|(m,c)} d\mu\Big(\frac{c}{d}\Big).$$

in particular for $(m, c) = 1$, we have $S(m, 0; c) = \mu(c)$.

- For any $s \in \mathbf{R}$, J_s, K_s and Y_s are the Bessel functions usually denoted this way, except when (as in [G-R], where N_s replaces Y_s) some other notation is used. Among many formulae and representations, we recall:

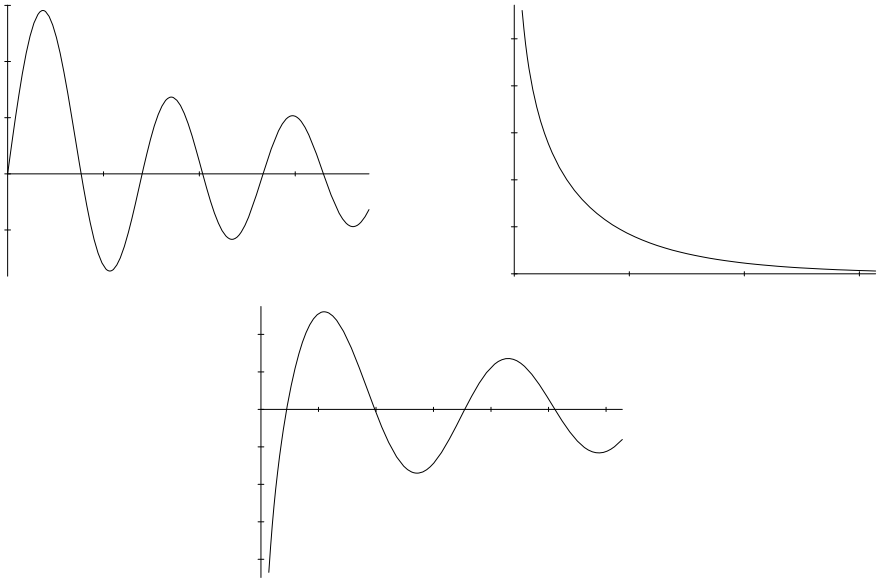
$$J_n(x) = \Big(\frac{x}{2}\Big)^n \sum_{k \geqslant 0} \frac{(-1)^k}{k!(k+n)!}\Big(\frac{x}{2}\Big)^{2k} \quad \text{for } n \in \mathbf{Z}, \text{ [G-R, 8.440]} \tag{1}$$

$$= \frac{1}{\pi}\int_0^{\pi} \cos(n\theta - x\sin\theta)d\theta \quad \text{for } n \geqslant 0, \text{ [G-R, 8.411.1]} \tag{2}$$

$$Y_0(x) = -\frac{2}{\pi}\int_1^{+\infty} \cos\Big(\frac{x}{2}(u + u^{-1})\Big)\frac{du}{u} \quad \text{[G-R, 3.714.2]} \tag{3}$$

$$K_0(x) = \int_1^{+\infty} \exp\Big(-\frac{x}{2}(u + u^{-1})\Big)\frac{du}{u} \quad \text{[G-R, 8.432.1]} \tag{4}$$

J_1, K_0 and Y_n are the Bessel functions that will actually occur in the text, for real arguments $x \geqslant 0$. Notice how the integral representations of K_0 and Y_0 reveal a similarity, not coincidental, with Kloosterman sums. Here are plots of the Bessel functions J_1 (upper left), K_0 (upper right) and Y_0.



- The notation

$$\frac{1}{2i\pi}\int_{(\sigma)} f(s)ds$$

designates the integral over the line $\mathrm{Re}\,(s) = \sigma$ in the complex plane. When s is used as a complex variable, the notation $\sigma = \mathrm{Re}\,(s)$, $t = \mathrm{Im}\,(s)$ is sometimes used without comments or reminder.

- We sometimes write $\log_2 x = \log\log x$.

Finally, here is the standing convention concerning the use of Vinogradov's symbol $\ll$ and Landau's O: the implied constant in both cases is meant to be absolute; this is usually repeated when it appears in the formal statements of theorems, lemmas, propositions, etc... In case there are other parameters involved, say ε, Δ, we (usually) indicate the dependency of the constants by the subscript notations $\ll_{\varepsilon,\Delta}$ or $O_{\varepsilon,\Delta}(\ )$. Sometimes it would be cumbersome to write down explicitly all dependencies, and they are then meant to be understandable according to the context. The following rules are strictly followed, when a parameter appears in the right-hand side of an inequality without being present on the left-hand side, and without other precisions in the text: ε and ϵ denote positive real numbers, with the meaning that for all $\varepsilon > 0$, or all $\epsilon > 0$, the inequality holds (with an implied constant depending on ε, ϵ), while uppercase letters A, B... (resp. δ) denote a large enough (resp. small enough) positive constant which is signaled to exist such that the inequality holds.

The reader with a more algebraic turn of mind is encouraged to show good will towards analytic number theorists and interpret such inequalities in the most reasonable way (provided it is correct and proves the result which is sought...)

Chapter 1
Abelian varieties, Jacobian varieties, modular curves and two theorems

Je crois que je l'ai su tout de suite : je partirais sur le Zeta
ce serait mon navire Argo, celui qui me conduirait à travers
la mer jusqu'au lieu dont j'avais rêvé, à Rodrigues,
pour ma quête d'un trésor sans fin.
J.M.G Le Clézio, *"Le chercheur d'or"*

The motivating context of this work is the Birch and Swinnerton-Dyer conjecture for abelian varieties over number fields, and therefore belongs to arithmetic algebraic geometry, and similarly the final results, stated in that context, are more likely to interest the experts in this subject. The methods which will be used to prove them, however, are those of analytic number theory, and the proofs may therefore be of more interest to analytic number theorists. This introductory chapter aims to give a simple account of abelian varieties and to state the Birch and Swinnerton-Dyer conjecture in the case which is considered here, for readers with little knowledge of algebraic geometry. It ends by the statements of the main theorems of this thesis. The next chapter will then move to the side of analytic number theory.

1.1 Abelian varieties

There are probably few people interested in number theory nowadays who have not been exposed to some aspect of the theory of elliptic curves by expository articles or books ([Si1] being the standard reference). Abelian varieties are natural generalizations of elliptic curves. The main difference in trying to present their definition and properties is that, whereas elliptic curves can be fairly easily given by explicit polynomial equations, in characteristic different from 2 and 3 by the simple Weierstrass form

$$E : Y^2 = X^3 + aX + b, \tag{1.1}$$

(where a and b are such that $\Delta(E) = -16(4a^3 + 27b^2)$ is not zero), there does not exist a similar way of writing down concretely all abelian varieties of a given dimension as the solution set of simple explicit polynomial equations. The definition is therefore necessarily more abstract. The starting point is the property (already used, in concrete special cases, by Diophantus) that the set of solutions of such a Weierstrass equation, with the addition of a point at infinity (in other words, the set of points on the elliptic curve E in the projective plane), is a commutative group, the group law being described by the beautiful geometric condition that three points P, Q and R on E add up to zero, $P + Q + R = 0$, if and only if P, Q and R are collinear. Abelian varieties are obtained by taking the group law as the focus of attention.

Definition 1. Let $k \subset \mathbf{C}$ be a subfield of the complex numbers. An abelian variety A defined over k is a smooth, irreducible, projective variety[1] together with a fixed k-rational point $0 \in A(k)$ and an algebraic group law, also defined over k, namely algebraic morphisms

$$+ \ : \ A \times A \to A$$

$$i \ : \ A \to A$$

(both defined over k) satisfying the usual axioms for a group, with $i(a)$ the inverse (opposite) of $a \in A$ and 0 as identity element.

Remark The epithet "abelian" carries a strong suggestion of commutativity, but no such assumption is made in the definition. It is actually the case that the group law of an abelian variety is necessarily commutative (this is essentially a consequence of the projectivity; indeed, there is no lack of non-projective and non-commutative algebraic groups, such as $GL(n)$, $SL(n)$,...), but the name itself is quite independent of this fact.

Concretely, recall that a projective algebraic variety A defined over k is a subset $A \subset \mathbf{P}^N$ for some integer $N \geqslant 1$, for which there are finitely many homogeneous polynomials (f_i) in $k[X_0, X_1, \ldots, X_N]$ such that A is the set common zeros $x = (x_0, \ldots, x_n)$ of the f_i's:

$$x \in A \iff f_i(x) = 0, \text{ for all } i;$$

and A is furthermore smooth if the Jacobian matrix[2] $(\frac{\partial f_i}{\partial X_j})_{i,j}$ is of maximal rank at every point $x \in A$. To be an abelian variety, A must have the additional group structure described in the definition. We only recall that an algebraic map is one given by rational functions, at least "locally" (in the Zariski topology; this means that a given expression for the group law will only hold when some polynomials do not vanish; if it does, then some other formula will be valid, and both will coincide when none does!). The condition of existence of the identity element $0 \in A(k)$ is not completely innocuous when k is not algebraically closed, say when $k = \mathbf{Q}$. The curve (considered by Selmer)

$$S \ : \ 3X^3 + 4Y^3 + 5Z^3 = 0$$

in $\mathbf{P}^2$ satisfies all conditions to be an elliptic curve, except that there is no solution $(X, Y, Z) \in \mathbf{Q}^3$.

With the geometric definition of the group law given above, the opposite of a point $P = (x, y)$ being $-P = (x, -y)$, and the identity element 0 being the point at infinity (which is simply $[0 : 1 : 0]$ in homogeneous coordinates in $\mathbf{P}^2$, and therefore is defined over $\mathbf{Q}$, hence a fortiori over any field $k \subset \mathbf{C}$), the curve E given by the Weierstrass equation (1.1) made homogeneous, namely

$$Y^2Z = X^3 + aXZ^2 + bZ^3,$$

is an elliptic curve, provided $\Delta \neq 0$. It is defined over k if the coefficients a and b are in k (see [Si1, III]). The non-vanishing of the discriminant Δ is equivalent here to the smoothness.

---

[1] The "naïve" language of algebraic geometry is used here, so A is identified with its complex points.

[2] Not to be mistaken with the Jacobian variety to be defined later.

As a concrete example in higher dimension, we quote, a particular case of equations found on page 3.174 of [Mu2], the following affine equations

$$\begin{cases} 4x + 8x^2 + 32x^3 + 160x^4 + 160xy^2 + 8y^2 + 16yt = 1 + 8z^2 \\ 24x^2 + 64x^3 + 288x^4 + 1536x^5 + 16xz + 192xyt + 960x^2y^2 \\ \qquad +96x^2z + 640x^3z + 64xz^2 + 4z + 8z^2 + 8t^2 = 1 + 8y^2 + 32y^2z \end{cases} \tag{1.2}$$

which provide an embedding in $\mathbf{C}^4$ of a dense affine subset of an abelian variety of dimension 2.

Other examples are given by observing that a product $A \times B$ of abelian varieties is another abelian variety. The next section will explain a general method which associates a certain abelian variety to any algebraic curve (its Jacobian variety), thus creating a very strong link between both theories.

We introduce some further vocabulary in connection with abelian varieties.

Definition 2. A morphism $f : A \to B$ of abelian varieties (defined over k) is a morphism (defined over k) of algebraic varieties which is also a group homomorphism. Morphisms can be composed, so there exists a category of abelian varieties over k.

An isogeny $f : A \to B$ is a morphism of abelian varieties with finite kernel $\ker(f) = f^{-1}(0)$; one says that A and B are isogenous. The relation "there exists an isogeny from between A to B" is an equivalence relation.

To get a feeling about abelian varieties, it is useful to look at them first as complex manifolds. Let A be an abelian variety. By definition $A \subset \mathbf{P}^N_{\mathbf{C}}$ is a smooth algebraic subvariety of some projective space and from the structure of complex manifold of the projective space it acquires one itself. We denote by A^{an} this complex manifold to carefully emphasize that it now carries the "usual" topology of the complex projective space, instead of the algebraic Zariski topology. In particular, A^{an} is compact, since A was closed and $\mathbf{P}^n_{\mathbf{C}}$ is compact. The group structure, being algebraic, is also analytic and therefore A^{an} is a compact, connected and commutative Lie group. Lie groups of this type are quite easy to classify: it is a basic result that there exists a lattice $\Lambda \subset \mathbf{C}^d$ – that is a free $\mathbf{Z}$-module of rank d – and an analytic isomorphism of complex Lie groups (a fortiori of complex manifolds)

$$A^{an} \xrightarrow{\sim} \mathbf{C}^d/\Lambda.$$

The integer d, naturally, is the dimension of the abelian variety d, it is the same as the dimension which is defined algebraically for any algebraic variety.

So the complex points of A are analytically the same as a complex torus. If we are only interested in the group structure of A, this solves the problem: A is isomorphic, as a group, to $\mathbf{S}^1 \times \ldots \times \mathbf{S}^1$, the product of $2d$ circles. Indeed, every lattice $\Lambda \subset \mathbf{C}^d$ has a basis $(\omega_1, \ldots, \omega_{2d})$, where the ω_i are vectors in $\mathbf{C}^d$ which are $\mathbf{R}$-linearly independent, such that

$$\Lambda = \mathbf{Z}\omega_1 \oplus \ldots \oplus \mathbf{Z}\omega_{2d}.$$

All this, of course, generalizes the well-known fact that every elliptic curve E over $\mathbf{C}$ "is" a torus, the quotient of $\mathbf{C}$ by a lattice $\Lambda = \mathbf{Z}\omega_1 \oplus \mathbf{Z}\omega_2$, with ω_1 and ω_2 two $\mathbf{R}$-linearly independent complex numbers.

A very important difference occurs, however, when considering the converse assertion. Given a lattice $\Lambda \subset \mathbf{C}^d$, it is natural to ask if there is a corresponding abelian variety. This requires in particular that the quotient $\mathbf{C}^d/\Lambda$ has an algebraic structure; more precisely it is enough that there exists an holomorphic embedding $f : \mathbf{C}^d/\Lambda \to \mathbf{P}^n_{\mathbf{C}}$ (of complex manifolds) of the torus inside some projective space.[3] The existence of such a morphism f is by no means obvious. It implies that there exist (many) meromorphic functions on $\mathbf{C}^d/\Lambda$, by taking the composition of rational functions, which are meromorphic functions on $\mathbf{P}^n_{\mathbf{C}}$, with f. This means that there must exist meromorphic functions on $\mathbf{C}^d$ which are periodic with a group of period exactly equal to Λ. This existence is a difficult problem of complex analysis.

When $d = 1$, this is always true, and indeed it is classically shown that every torus $\mathbf{C}/\Lambda$ is an elliptic curve by constructing the Weierstrass $\wp$ function of the lattice, which is the meromorphic function on $\mathbf{C}$ given by

$$\wp_\Lambda(z) = \frac{1}{z^2} + \sum_{\substack{\lambda \in \Lambda \\ \lambda \neq 0}} \Big(\frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2} \Big)$$

and then showing that the map $z \mapsto (\wp_\Lambda(z), \wp'_\Lambda(z))$ provides an isomorphism of complex manifolds (here, Riemann surfaces, since $d = 1$) between the torus $\mathbf{C}/\Lambda$ and the elliptic curve E given by the Weierstrass equation

$$y^2 = 4x^3 - 60G_4 x - 140G_6$$

where we have defined, for $k \geqslant 1$

$$G_{2k} = \sum_{\substack{\lambda \in \Lambda \\ \lambda \neq 0}} \lambda^{-2k} \in \mathbf{C}.$$

The corresponding assertion, however, fails completely when $d \geqslant 2$. Indeed, already for $d = 2$, one can construct complex tori $\mathbf{C}^2/\Lambda$ on which the only meromorphic functions are the constants (in a certain sense, this is actually true for almost all higher dimensional tori, see [Mu1, page 36]). The necessary and sufficient conditions for a given torus to be algebraic, and hence to "be" an abelian variety, were already found by Riemann.

Theorem 1. *(Riemann) Let $\Lambda \subset \mathbf{C}^d$ be a lattice. Then the torus $\mathbf{C}^d/\Lambda$ has a structure of a projective variety if and only if there exists a positive definite hermitian form H on $\mathbf{C}^d$ such that $E = \operatorname{Im} H$ is integral-valued when restricted to Λ.*

This condition appears, roughly, as one tries to construct meromorphic functions on the torus as quotients of entire functions f on $\mathbf{C}^d$ which are "automorphic" with respect to the action of Λ on $\mathbf{C}^d$, i.e. for which $f(z + \lambda)$ is related to $f(z)$ by some simple transformation rule for $\lambda \in \Lambda$. Those are basically theta functions and some kind of integral-valued positive definite quadratic form is necessary to define them by series.

---

[3]The sufficiency of this condition follows from Chow's theorem, according to which every closed analytic subvariety of a projective space is actually algebraic.

A bilinear form H as in the Theorem is called a polarization of the complex torus, or of the abelian variety.

Remark This theorem applies to the case $d = 1$ and proves again that all quotients $\mathbf{C}/\Lambda$ are algebraic: let ω_1 and ω_2 be a basis of Λ, and put

$$H(z, w) = \frac{|\omega_1|^{-2}}{\mathrm{Im}\,(\omega_2/\omega_1)} z\overline{w};$$

then it is immediately checked that the bilinear form H is a polarization for the torus $\mathbf{C}/\Lambda$.

This analytic side of the theory of abelian varieties is very useful for gaining insights into their geometric properties. Translating a given problem into one about complex tori, for which an answer might be more easily derived, can help to guess what the solution is in general. For instance, algebraic morphisms between two abelian varieties A_1 and A_2 over $\mathbf{C}$, with $A_i^{an} = \mathbf{C}^{d_i}/\Lambda_i$, correspond to $\mathbf{C}$-linear maps $f : \mathbf{C}^{d_1} \to \mathbf{C}^{d_2}$ such that $f(\Lambda_1) \subset \Lambda_2$, etc...[4]

The analytic theory can also be useful for arithmetic. For example, an easy consequence of the group structure of an abelian variety A is the determination of the structure of its torsion points. Let $A[n]$ be the group of n-torsion points of A ($n \geqslant 1$ an integer),

$$A[n] = \{x \in A \mid nx = 0\},$$

then by the above "uniformization" of A as a complex torus it follows that $A[n] \simeq (\mathbf{Z}/n\mathbf{Z})^{2d}$. Now, if A is defined over a subfield $k \subset \mathbf{C}$, the algebraicity of the group law implies that there is an action of the absolute Galois group of k on $A[n]$, by action on the coordinates, and since $A[n]$ is finite, it follows that points in $A[n]$ have only finitely many conjugates, so they are algebraic over k. If k is a number field, this group $A[n]$, with its action of the Galois group of k, is a very important arithmetic object, one of the main tools in the study of the arithmetic of abelian varieties. The analytic theory thus provides its basic structure as a group, which is in no way obvious from the algebraic viewpoint. It can also be proved algebraically – such proofs are of course required when considering similar questions for abelian varieties over fields of non-zero characteristic, such as finite fields.[5]

Coming back now to the arithmetic side of the theory, let A be an abelian variety, of dimension d, defined over a number field k. As is the case with all algebraic varieties over k, a basic, and often deep and fascinating, problem of diophantine geometry is to understand the set of points of A with coordinates in k, denoted $A(k)$: do points exist, and if yes, how many are there, how are they distributed, and sundry other questions arise, and have arisen since the very beginning of mathematics. For an abelian variety, by the very definition there must be at least one rational point, the identity element $0 \in A(k)$, and moreover, the set $A(k)$ has actually a structure of abelian group with 0 as identity. It could be however that $A(k)$ is reduced to 0, and this certainly is often the case.

---

[4] From now on, the use of the superscript ^{an} will be relaxed; it is hoped that it will be clear in the context whether an abelian variety is considered in its complex analytic or its algebraic guise.

[5] We have not really defined what this means.

From the analytic theory, we know that $A(k)$ must be a subgroup of a product of $2d$ circles, but this is very weak information in itself. The first basic theorem asserts that there is a much stronger restriction.

Theorem 2. *(Mordell-Weil) The group $A(k)$ is a finitely generated abelian group.*

This was first conjectured (apparently: the statement is quite obscure) by Poincaré, then proved by Mordell, both for $d = 1$ (elliptic curves) and $k = \mathbf{Q}$, before Weil proved the (essentially) general case in his thesis. The group $A(k)$ is called the Mordell-Weil group of A. For a proof of the theorem when $d = 1$, see [Si1, VIII]; the general case is very similar in principle, although the algebraic geometry necessary for $d \geqslant 2$ is quite harder ([C-S, ch. 5 and 6]).

Using the structure theory of finitely generated abelian groups, we can thus write uniquely

$$A(k) = A(k)_{tors} \oplus \mathbf{Z}^r$$

where $A(k)_{tors}$ is the torsion subgroup of $A(k)$, which is finite, and $r \geqslant 0$ is an integer, the rank of the Mordell-Weil group. By definition, r is also called simply the rank of the abelian variety, and denoted by $\operatorname{rank} A$. It is the arithmetic invariant with which this thesis is concerned, in a special case. Note that r is zero if and only if $A(k)$ is finite.

Both the rank and the torsion subgroup are very subtle arithmetic invariants. Computing them explicitly, as one can gather by trying to play with some concrete equations, is extremely hard. An instance of this diophantine problem arises from the following classical question: which positive squarefree integers are the area of a right triangle with sides of rational length? Such numbers are called congruent numbers. As the area of the famous right triangle with sides 3, 4 and 5, $n = 6$ is congruent, but it is very difficult to find out whether a given n is. It is elementary (though not obvious) that n is congruent if and only if the elliptic curve $E_n : y^2 = x^3 - n^2x$ over $\mathbf{Q}$ has infinitely many rational points, namely if its rank is at least one, and ad-hoc solutions for specific values of n can be traced back to Fermat ($n = 1$, which is not congruent), Euler ($n = 7$, which is) and beyond to Arab and Greek mathematicians (see [Kob] for a detailed treatment, including Tunnell's spectacular characterization of congruent numbers, which gives an easy way of finding out whether a given integer is a congruent number[6]).

The Mordell-Weil theorem, in a way, provides an upper-bound on the size of the group of rational points. The proof, however, yields little more information that can be directly used to obtain any form of tighter control of it. More than seventy years after the proof of this result, the most important, and most natural, questions about the Mordell-Weil group remain unsolved. Let us take some time to digress on some exemplary problems and the progress that has been made in a few millennia of mathematical research.

• The case of the torsion subgroup is (maybe) simpler. How large is the torsion subgroup $A(k)_{tors}$ for given A? Or how large can it be for all A of a given dimension, defined over a given field? In the case of elliptic curves, there is a simple algorithm to compute $E(k)_{tors}$, using integrality properties of the coordinates of torsion points. Over

[6]This solution is "almost" complete: in one direction, it still depends on unsolved cases of the Birch and Swinnerton-Dyer conjecture.

$\mathbf{Q}$, it was moreover conjectured already by Beppo Levi in 1908, and later by Nagell, still later again by Ogg, that only a finite number of torsion subgroups were possible (with an explicit list of all of them). This was proved by Mazur in 1977. For arbitrary number fields, but still for elliptic curves, Merel [Me] finally proved in 1995 that there is a bound $N(f)$ such that for all elliptic curves E/k, where k is a number field of degree at most f over $\mathbf{Q}$, the bound $|E(\mathbf{Q})_{tors}| \leqslant N(f)$ holds. In higher dimensions, almost nothing is known, although one expects that similar uniform finiteness holds.

The rank r, on the other hand, remains mostly *terra incognita*.

• Is there an algorithm to compute the rank of an abelian variety given explicitly by equations? Or even for elliptic curves?

The proof of the Mordell-Weil theorem is in large part effective, so that an upper-bound can in principle be deduced for r from other "more accessible" invariants, but it does not yield an effective algorithm to compute, in guaranteed finite time, a basis of the free part of $A(k)$. The problem is that there is no limit known on the "size" of the points in a basis.[7] Therefore (except if enough independent rational points are obtained to reach the upper-bound, which is extremely unlikely, given its size in comparison with what is expected to be true in general), there is no means of knowing when to stop looking for possible "bigger" solutions, without some other information. This would become available if another invariant of A, known as the Tate-Shafarevitch group of A, was known to be finite, but this is another major unsolved problem. However, the techniques related to it yield an algorithm which is guaranteed to terminate and return the rank of any (explicitly given) abelian variety which has finite Tate-Shafarevitch group, so presumably works with any abelian variety (for all this, see [Si1, X]).

The size of the solutions can be notoriously very big: as an example, consider the elliptic curve $E : y^2 = x^3 + 877x$. Then Bremner and Cassels proved that E has rank 1, and that a generator $P = (x, y)$ has

$$x = \Big(\frac{612776083187947368101}{7884153586063900210}\Big)^2.$$

• Can the rank r be arbitrarily large for a given k, especially for elliptic curves E over $\mathbf{Q}$?

A positive answer in this special case would of course also settle the general case. While for some time it was thought (at least by some mathematicians) that the rank of elliptic curves over $\mathbf{Q}$ should be absolutely bounded, the opposite is now more often expected. Néron, then Mestre and others, have had great success in constructing elliptic curves with rather large rank. The current records are that there exist infinitely many (non-isomorphic) elliptic curves with rank at least 14 (due to Kihara), and one of rank at least 22 (due to Fermigier). Let us quote this example from [Si1, page 234]: the curve

$$y^2 - 246xy + 36599029y = x^3 - 81199x^2 - 19339780x - 36239244$$

has rank at least 12 over $\mathbf{Q}$.

• In general, given a family of abelian varieties, how does the rank vary among members of this family?

---

[7] Measured, for instance (in the simplest case of $\mathbf{Q}$), by taking the maximum of the numerator and denominator of the affine coordinates of the solutions in some open affine subspace in the projective space containing A.

This question is rather imprecise, as we have not defined what is meant by a family. The congruent number curves $y^2 = x^3 - nx$ defined above can be considered this way. Usually, the family will be some infinite set, and a typical setup is to define some positive real parameter, say q, such that there are only finitely many elements of the set with parameter less than q, and then ask how the average rank among those finitely many varieties, say $r(q)$, varies when q gets large. There is also a very precise notion of an algebraic family of abelian varieties which makes it possible to frame the problem in very precise terms, but it doesn't cover all the interesting cases.

Those are in a way the most "basic" questions. Of course by making abelian varieties interact with other objects of arithmetic, more will arise. Let A/k be defined over a number field; then in addition to the Mordell-Weil group $A(k)$, there is a group $A(K)$ for any extension field K/k. If the extension is finite, this is simply the Mordell-Weil group of A when considered as defined over K only, and as such it remains of finite type. How does the rank of the groups $A(K)$ vary when K runs over some family of number fields? How fast is it increasing?

Considering the difficulty of getting hold of the rank of abelian varieties, it is remarkable that there should exist a completely different way of recovering it, which – if true – would yield tremendous insight into its properties, and into the arithmetic of abelian varieties in general. This is the content of the Birch and Swinnerton-Dyer conjecture, which may be interpreted as a form of local-global principle.

In number theory, and in diophantine geometry especially, the term "global" refers to properties of objects over a number field (fixed), say $\mathbf{Q}$ to simplify the discussion (see [Maz] for a recent survey of this kind of ideas), whereas "local" refers to the fact that many objects and properties of interest can be reduced modulo p for all prime numbers p, and then studied in this (usually simpler) context. Considering all primes together may then help in the study of the global object. For instance, an integer solution $x = (x_0, \dots, x_n)$ of an homogeneous polynomial equation $f(x) = 0$ ($f \in \mathbf{Z}[X_0, \dots, X_n]$) gives a solution modulo p, $\overline{x} = (\overline{x_0}, \dots, \overline{x_n})$ of the reduced equation $\overline{f}(\overline{x}) = 0$, $\overline{f} \in \mathbf{F}_p[X_0, \dots, X_n]$. Thus, a necessary condition for the existence of a global solution x to this equation is the existence of one modulo p for every p (and also of one in $\mathbf{R}$: this corresponds to the infinite place ∞ of $\mathbf{Q}$). The converse to this statement is false in general (as an example, again, the Selmer curve $3X^3 + 4Y^3 + 5Z^3 = 0$ has points modulo p for all p and points in $\mathbf{R}$, but doesn't have rational points). When some form of it holds, however, it is called a local-global principle, or a Hasse principle. The best known example is the Hasse-Minkowski theorem according to which, over any number field k, for any quadratic form

$$Q = \sum_{i,j} a(i,j) X_i X_j$$

with coefficients $a(i,j) \in \mathcal{O}_k$, there exists an $x = (x_i)$ with $Q(x) = 0$ if and only if for every prime ideal $\mathfrak{p} \in \mathcal{O}$ (and for every archimedean completion k_∞ of k), there exists a local solution $x_\mathfrak{p} = (x_{\mathfrak{p},i})$ modulo $\mathfrak{p}$ (respectively, a solution $x_\infty \in k_\infty$). This gives easily, for instance, a complete algorithmic solution to the problem of knowing whether a quadric over k has a rational point.

The Birch and Swinnerton-Dyer conjecture indicates a much subtler relationship between the local and global arithmetic properties of abelian varieties over number fields. The local properties it involves are those encoded in the Hasse-Weil zeta function

(also called simply the L-function). Let A/k be an abelian variety. The L-function of A is an holomorphic function defined by an Euler product

$$L(A,s) = \prod_{\mathfrak{p}} L_{\mathfrak{p}}(A, N\mathfrak{p}^{-s})^{-1}$$

where $\mathfrak{p}$ runs over all prime ideals $\mathfrak{p} \in \mathcal{O}_k$ and $L_{\mathfrak{p}}$ is a monic polynomial, with integer coefficients, of degree equal to $2 \dim A$ (for all but finitely many $\mathfrak{p}$, at most $2 \dim A$ otherwise). This polynomial can be defined from a generating series for the number of points of A in the finite extensions of the residue field $\mathcal{O}_k/\mathfrak{p}$, or more directly from the characteristic polynomial of a Frobenius endomorphism acting on some finite dimensional $\mathbf{Q}_\lambda$-vector space, where $\lambda \neq \mathfrak{p}$ is another prime ideal in $\mathcal{O}_k$. A full description will be given in the next section in the case of the Jacobian of an algebraic curve which is the case of interest in the rest of the thesis.

Conjecture 1. *(Birch and Swinnerton-Dyer) Let A/k be an abelian variety defined over a number field k. Then the rank of the Mordell-Weil group $A(k)$ of A is equal to the order of vanishing of the L-function of A at $s = 1$:*

$$\operatorname{rank} A = \operatorname{ord}_{s=1} L(A,s). \tag{1.3}$$

This is a beautiful conjecture,[8] but not one which allows us to indulge very long in contemplating its beauties, because it presents some immediate difficulties. The trouble is that the L-function is defined by the infinite Euler product which is only convergent for $\operatorname{Re}(s) > \frac{3}{2}$. The special point $s = 1$ is not in this region, and the only way to make sense of this statement is to assume beforehand that $L(A,s)$ can be analytically continued at least to some open subset of the complex plane containing 1. This is not a major difficulty in trying to understand the meaning of the conjecture, since it is a standard assumption that $L(A,s)$ actually admits an analytic continuation to an entire function, but it is much more troublesome when trying to *prove* it, because this analytic continuation is known in a few cases only. Those are, basically, the elliptic curves over $\mathbf{Q}$ to which the methods of Wiles [Wil] and Taylor-Wiles apply (notably, all semi-stable elliptic curves, and in any case a significant proportion of all elliptic curves over $\mathbf{Q}$), abelian varieties with complex multiplication (Deuring, Shimura, Taniyama…) and the Jacobians of modular curves, by Eichler-Shimura theory (see the next section and the next chapter about the last, which will be the main focus of attention).

Let us now assume that $L(A,s)$ is entire, in the more precise form postulating also the existence of a standard functional equation relating the value of the L-function at s and at $2 - s$. This then shows that the critical strip containing all non-trivial zeros[9] of $L(A,s)$ is the strip $\frac{1}{2} \leqslant \operatorname{Re}(s) \leqslant \frac{3}{2}$, and especially that $s = 1$ is the real point on the critical line $\operatorname{Re}(s) = 1$. According to the Generalized Riemann Hypothesis for $L(A,s)$, all non-trivial zeros of $L(A,s)$ should lie on this line. Of course, it is to be expected

[8] As a matter of fact, this is only a weak form of the Birch and Swinnerton-Dyer conjecture, which has been refined to include in particular an exact formula for the leading term of the Taylor expansion of $L(A,s)$ at $s = 1$, involving other arithmetic invariants of A such as the order – presumed to be finite… – of the Tate-Shafarevitch group of A, etc…

[9] The trivial zeros are those accounted for by the poles of the Gamma factors which have to be inserted in the functional equation; they are the integers $-n$, with $n \geqslant 0$.

that the point involved in an equality such as (1.3) should be on the critical line. Since abelian varieties of positive rank do exist, the corresponding equality with $s = s_0$ in lieu of $s = 1$ would violate the Riemann Hypothesis if $\mathrm{Re}\,(s_0) \neq 1$. Thus the Birch and Swinnerton-Dyer conjecture is another indication that the rank is a very delicate invariant: whether an L-function vanishes at some point of the critical line is no trivial matter.

We now ask a few questions, which could be asked of any mathematical conjecture. First, what is the evidence available that it is true? Of course, there are some obvious compatibilities which show that the conjecture can not be disregarded because of some trivial matter: both sides, for instance (!) are integers greater than or equal to zero, and both are additive under products; also both are invariant under isogeny: if A and B are defined over k and $f : A \to B$ is a k-isogeny it is shown quite easily that $\mathrm{rank}\, A = \mathrm{rank}\, B$, and without too much difficulty that the L-functions are actually identical, $L(A, s) = L(B, s)$ (the converse of this last statement, known as Faltings's isogeny theorem, is also true, but is much deeper). More to the point, there is actually significant evidence now, at least for the elliptic curves over $\mathbf{Q}$, for which the analytic continuation is true. It started with the Coates-Wiles theorem, according to which an elliptic curve E over $\mathbf{Q}$ with complex multiplication, such that $L(E, 1) \neq 0$, has rank 0 (that is, its Mordell-Weil group $E(\mathbf{Q})$ is finite). Now, the works of Rubin, Gross-Zagier and Kolyvagin, all put together, show that for modular elliptic curves the equality (1.3) is true if the order of vanishing of $L(E, s)$ at 1 is at most 1. But no clue seems to be available concerning the higher ranks $r \geqslant 2$.

A second natural question: is this conjecture useful, or a mere thing of beauty, to be gazed and wondered at? Obviously, this will depend on the definition of "useful" that is adopted. Let's see if it sheds some light on the different problems about the rank that we mentioned earlier. Concerning the question of finding an algorithm, it would seem to bring a solution: simply compute the values of the successive derivatives of $L(A, s)$ at $s = 1$ until one is non-zero. However, in practical terms, this solution requires the ability to compute to a high degree of accuracy the value of the L-function and its derivatives at $s = 1$. When the curve is given by a Weierstrass equation with large coefficients, this is by no means an easy matter: the numbers involved quickly become too large (for the known curves with high ranks, for instance, it can not be checked today that their L-function vanishes to a high order).

Similarly, the conjecture yields no insight, even at the most heuristic or intuitive level, concerning the existence of elliptic curves over $\mathbf{Q}$ with arbitrarily large rank. Experts in the analytic theory of L-functions have no reason to believe in either possibility.

Much more promising is the third problem of studying the behavior of the rank in families. Indeed, the study of the vanishing or non-vanishing of families of L-functions on average at various points is a standard subject in analytic number theory, and many methods and results have been developed in this direction. The application of some of these to abelian varieties was initiated by Mestre [Mes], but the location of the special point at the center of the critical strip complicates the analysis enormously and Mestre, as Brumer [Br1] and other investigators of this subject after him, had to appeal to the Generalized Riemann Hypothesis, in addition to the other assumptions, to obtain interesting results. Of course, this is considered a very safe assumption, as assumptions go, but it seems it must delay unconditional proofs along those lines until some quite distant future.

At this point, which might be considered not highly satisfactory, it is good to remember that despite sometimes stubborn indications to the contrary, equality is a symmetric relation and $a = b$ might also mean $b = a$. And it turns out that the most spectacular application of the Birch and Swinnerton-Dyer conjecture to date doesn't rely on it as a way of investigating the rank of an abelian variety, but as one of predicting the existence of L-functions with rather large order of vanishing at the center of the critical strip! This is the content of Goldfeld's work [Gol], which culminated, using the theorem of Gross-Zagier, in the first non-trivial effective lower-bound for the class number $h(-D)$ of the imaginary quadratic field $\mathbf{Q}(\sqrt{-D})$, $D \geqslant 1$ a squarefree integer. Roughly speaking, Goldfeld showed, using beautiful analytic techniques, that if there existed *one* L-function, of degree 2, satisfying some standard assumptions, and vanishing at order $r \geqslant 3$ at the center of its critical strip, then there exists an explicit and absolute constant $C > 0$ (depending only on the specific L-function used) such that

$$h(-D) > C(\log D)\alpha(D) \tag{1.4}$$

(where $\alpha(D)$ is very small (almost constant), precisely Oesterlé has shown that one can take

$$\alpha(D) = \prod_{p|D} \Big(1 - \frac{[2\sqrt{p}]}{p+1}\Big)$$

so $\alpha(D) \ll_\varepsilon (\log D)^\varepsilon$).

Without the Birch and Swinnerton-Dyer conjecture, it might have been thought that the existence of such L-functions was very unlikely, but because of it and because some elliptic curves of rank larger or equal to 3 were already known, Goldfeld only needed it to be checked for any particular one of those high-rank curves to prove this lower bound unconditionally. Gross and Zagier, using their theorem about the case of rank 1 of the Birch and Swinnerton-Dyer conjecture, finally proved the required property of some of the rank 3 curves.

After Oesterlé succeeded in computing a possible value of the constant C, the exact bound was found to be sufficient to solve the class number 3 problem of Gauss, namely finding all imaginary quadratic fields with class number equal to 3 (the class number 1 problem had already been solved previously by Heegner, whose proof was mistakenly thought to be in error, then Stark and Baker independently, and the class number 2 problem by Baker and Stark).

In this thesis, we will show that some of the results proved by Mestre and Brumer using the Birch and Swinnerton-Dyer conjecture and the Generalized Riemann Hypothesis are within reach of current methods of analytic number theory, and in particular that the progress towards the Birch and Swinnerton-Dyer conjecture makes it possible to give lower bounds for the rank of some special abelian varieties which are unconditional and sharp. Thus, the implicit promise of this conjecture, that it can be used effectively to study the rank of abelian varieties, can be kept today without additional assumptions.

1.2 The Jacobian of an algebraic curve

The abelian varieties which we will study in detail belong to the special class of the Jacobians of algebraic curves. Historically, it is actually through the Jacobians that

abelian varieties were first considered by Abel and Jacobi, and it was not before the twentieth century that the general abstract definition above was formulated.

It is again quite easy to describe the analytic aspects of the theory, over $\mathbf{C}$. Elliptic curves were first implicitly studied when Euler, Fagnano, Legendre, Jacobi, and many others were studying the functions defined by so called "elliptic integrals", for instance

$$K(\lambda) = \int_0^1 \frac{dx}{\sqrt{(1-x^2)(1-\lambda^2 x^2)}}$$

which is related to the elliptic curve $y^2 = (1-x^2)(1-k^2x^2)$. Abel was the first to attempt to develop a similar theory for integrals related to more general Riemann surfaces, or algebraic curves of higher genus.

Let C be a compact Riemann surface of genus g: topologically, C is represented by the usual picture of a doughnut with g holes, see the picture below. We always assume here that $g \geqslant 1$ (the only compact Riemann surface with $g = 0$ is the projective line $\mathbf{P}^1_{\mathbf{C}}$; its Jacobian, inasmuch as it exists, is reduced to 0). The case $g = 1$ corresponds to a torus, and a curve of genus one over $\mathbf{C}$ is the same as an elliptic curve over $\mathbf{C}$. The Jacobian appears when trying to compute integrals of holomorphic differential forms along paths of C. We denote by $\Omega^1(C)$ the vector space of holomorphic 1-forms on C. This is a finite dimensional vector space of dimension exactly equal to g, as has been known since Riemann.[10] For instance, if E is the Riemann surface associated to an elliptic curve with Weierstrass equation (1.1), $\Omega^1(E)$ is isomorphic to $\mathbf{C}$, and a generator is the holomorphic differential $\omega = \frac{dx}{y}$ (notice the similarity with the integrand in the elliptic integral $K(k)$).

Instead of trying to integrate one form $\omega \in \Omega^1(C)$ at a time, Abel found that it was best to consider all of them simultaneously. Fix a base point $x_0 \in C$, and a basis $(\omega_1, \ldots, \omega_g)$ of the space $\Omega^1(C)$. We want to build a map

$$\left\{ \begin{array}{ccc} C & \to & \mathbf{C}^g \\ x & \mapsto & (\int_{x_0}^x \omega_1, \ldots, \int_{x_0}^x \omega_g) \end{array} \right. \tag{1.5}$$

but of course this is not well-defined as stated, because the value of any of the integrals will depend on the path of integration chosen between x_0 and x. The difference between the integrals along two different paths, however, can be written as an integral along a loop based at x_0:

$$\int_{\gamma_1} \omega_j = \int_{\gamma_2} \omega_j + \int_{\gamma_1 \gamma_2^{-1}} \omega_j,$$

where γ_i $(i = 1, 2)$ is a path $\gamma_i : [0,1] \mapsto C$ ending at x $(\gamma_i(1) = x)$, and the composition and inverse in $\gamma_1\gamma_2^{-1}$ refer to the group law in the fundamental group of C, i.e this loop is the concatenation of γ_1 (going from x_0 to x) and the reverse of γ_2 (going back from x to x_0).

Consequently, if we let Λ denote the subgroup of $\mathbf{C}^g$ generated by the integrals along all loops γ based at x_0, $(\int_\gamma \omega_1, \ldots, \int_\gamma \omega_g)$, then the abortive "map" (1.5) does become

[10] This can be taken as a rigorous definition of g. It is actually very convenient in the algebraic setting when the topological picture might not be available.

well-defined if we push the image to the quotient by Λ:

$$\left\{ \begin{array}{ccc} C & \to & \mathbf{C}^g/\Lambda \\ x & \mapsto & (\int_{x_0}^x \omega_1, \ldots, \int_{x_0}^x \omega_g) \end{array} \right. \tag{1.6}$$

(taking the classes modulo Λ of the integrals). The Jacobian $J(C)$ of C is this quotient $\mathbf{C}^g/\Lambda$, and the map is called the Abel-Jacobi map of C (based at x_0).

As a matter of fact, this first definition remains unsatisfactory, because it depends on a particular choice of a base-point $x_0 \in C$, and of a basis of $\Omega^1(C)$. A more intrinsic definition is possible, but we need to define the first integral homology group of C. Let $C_1(C, \mathbf{Z})$ be the group of 1-chains in C: by definition, this is the free abelian group whose generators are the paths in C, i.e. the continuous maps $\lambda : [0,1] \to C$. Elements in $C_1(C, \mathbf{Z})$ are finite formal sums of the type

$$\sum_{i=1}^n n_i \lambda_i$$

where each λ_i is a path. Among chains, we distinguish the subgroup $Z_1(C, \mathbf{Z})$ of 1-cycles, which are the chains without boundary. The boundary homomorphism ∂ is defined for a path λ by $\partial(\lambda) = \lambda(1) - \lambda(0)$ – where the sum lives in $C_0(C, \mathbf{Z})$, the free abelian group generated[11] by points of C – and extended by additivity, and $Z_1(C, \mathbf{Z})$ is the kernel of ∂. If the path λ is a loop, so $\lambda(1) = \lambda(0)$, for example, it is a 1-cycle.

Lastly, there is the subgroup of $C_1(C, \mathbf{Z})$ of 1-boundaries. This is the image of the boundary map from 2-chains, which are defined analogously as the elements of the free group generated by maps $\square : [0,1] \times [0,1] \to C$; by "going along the boundary" of the square, one obtains a (formal) sum of four paths in C, and $B_1(C, \mathbf{Z})$ is the image of all 2-chains by this morphism. Clearly, the boundary of a square is a loop, so there is an inclusion $B_1(C, \mathbf{Z}) \subset Z_1(C, \mathbf{Z})$. The first integral homology group $H_1(C, \mathbf{Z})$ is the quotient group

$$H_1(C, \mathbf{Z}) = Z_1(C, \mathbf{Z})/B_1(C, \mathbf{Z}).$$

As an aside, we mention quickly the relation between this group and the – maybe – more familiar fundamental group $\pi_1(C, x_0)$ of loops in C issuing from $x_0 \in C$, with concatenation acting as group law, taken modulo homotopies of loops. Quite obviously there is a surjective map $\pi_1(C, x_0) \to H_1(C, \mathbf{Z})$, and one can show (Hurewicz theorem) that the kernel of this map is the commutator subgroup $[\pi_1(C, x_0), \pi_1(C, x_0)]$: in group theoretic language, $H_1(C, \mathbf{Z})$ is the abelianization of $\pi_1(C, x_0)$ – here the base-point is no longer of importance because of the commutativity.

To think of homology classes, consider a simple loop: it defines a homology class. When is it zero? The loop itself is a boundary if one can, in a way, "fill in" continuously the "interior" of the loop, painting something homeomorphic to a square, as it were. For instance, on an annulus $A_{a,b} = \{z \in \mathbf{C} \mid a < |z| < b\}$, any circle of radius r, $a < r < b$ will define a non-trivial homology class. But two circles with different radius define the same (or opposite) class, since one can map a square to cover the annulus between them. Actually, $H_1(A_{a,b}, \mathbf{Z}) \simeq \mathbf{Z}$, with the class of any such circle as generator.

[11] This is also known as the group of divisors of C, see below.

Now, coming back to the Jacobian, let $V = \Omega^1(C)^*$, the dual vector space of $\Omega^1(C)$. The main point of the introduction of $H_1(C, \mathbf{Z})$ is that it is possible to integrate a holomorphic 1-form against an homology class $\lambda \in H_1(C, \mathbf{Z})$, extending the integration of differential forms along a path. Putting aside matters of differentiability which might cause trouble but don't, this follows easily from Stokes's theorem: it is enough to show that if a path λ is the boundary of a square $\square$, then $\int_\lambda \omega = 0$ for any $\omega \in \Omega^1(C)$. But one has

$$\int_\lambda \omega = \int_{\partial\square} \omega = \int_\square d\omega = 0,$$

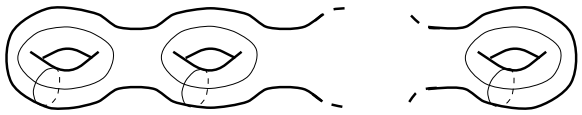
since locally $\omega = f(z)dz$, with f holomorphic, hence $d\omega = \overline{\partial} f(z) d\overline{z} \wedge dz = 0$.

Thus we can define a homomorphism

$$\begin{cases} H_1(C, \mathbf{Z}) & \to & V \\ \lambda & \mapsto & (\omega \mapsto \displaystyle\int_\lambda \omega) \end{cases}$$

and let $\Lambda \subset V$ be the subgroup of V generated by its image. The Jacobian of C is defined to be $J(C) = V/\Lambda$.

The basic homology theory of surfaces shows that $H_1(C, \mathbf{Z})$ is a free abelian group of rank $2g$. The figure below shows a possible basis $\lambda_1, \ldots, \lambda_{2g}$: around each of the g holes there are two loops.



A Riemann surface and an homology basis of cycles

The map $H_1(C, \mathbf{Z}) \to V$ above is shown to be injective, and this implies that $\Lambda \subset V$ is a lattice, so $J(C)$ is a complex torus of dimension g. What is more, it is in fact an abelian variety. The reason for this can be seen using Riemann's condition of Theorem 1: indeed, there exists a natural choice of a polarization $H : V \times V \to \mathbf{C}$ satisfying the conditions there. This is dual to the bilinear form $\Omega^1(C) \times \Omega^1(C) \to \mathbf{C}$ defined by integration $i \int_C \omega \wedge \overline{\omega'}$. Clearly this form H is hermitian, and positive definite since

$$i \int_C \omega \wedge \overline{\omega} > 0$$

for all $\omega \neq 0$ in $\Omega^1(C)$. The necessary integrality property $H(\Lambda \times \Lambda) \subset \mathbf{Z}$ comes from the compatibility of this bilinear form (and of the injective map $H_1(C, \mathbf{Z}) \to V$) with the intersection pairing in homology. This is the form defined, on a basis as in the picture, by simply counting (with orientation) the number of intersection points of two loops, so it is clearly integral valued. In fact, with a suitable enumeration of the basis $(\lambda_1, \ldots, \lambda_{2g})$ of $H_1(C, \mathbf{Z})$, this produces a form $\Lambda \times \Lambda \to \mathbf{Z}$ with matrix

$$\begin{pmatrix} 0 & -\mathrm{Id}_g \\ \mathrm{Id}_g & 0 \end{pmatrix}$$

This definition $J(C) = \Omega^1(C)^*/H_1(C, \mathbf{Z})$, in particular, determines immediately the cotangent space $T_0^* J(C)$ of $J(C)$ at the identity element:

$$T_0^* J(C) \simeq \Omega^1(C). \tag{1.7}$$

As an exercise, take for C an elliptic curve $\mathbf{C}/\Lambda$, and show that there exists a canonical isomorphism $C \simeq J(C)$.

Riemann's theorem shows that $J(C)$ is an abelian variety, in particular an algebraic variety. However, this description over $\mathbf{C}$ is analytic in nature: although it is somewhat explicit, it works with complex integrals and such apparently transcendental constructs. If C/k is an algebraic curve defined over a subfield $k \subset \mathbf{C}$, we know that $J(C)$ can be embedded in some projective space as the solution set of some polynomial equations, and it is natural to wonder to which field the coefficients of those polynomials belong, and in particular: is $J(C)$, by any chance, also defined over the field k? It is of course of particular importance if one wishes to study the arithmetic of the Jacobian of some curve to know that there is such a thing! If the Jacobian could only be defined over a transcendental extension of $\mathbf{Q}$, there would be no notion, for instance, of rational points on it. André Weil was the first to be confronted to this, as he was trying to extend Mordell's theorem to the Jacobians of algebraic curves. It was felt more acutely even when trying to prove the Riemann Hypothesis for curves over finite fields: it was necessary to have a Jacobian variety associated to those curves, with properties similar to the usual ones over $\mathbf{C}$ (largely unstated yet), even though the analytic construction makes no sense in such a context. It was part of Weil's achievement that he succeeded in solving this problem by finding an appropriate, purely algebraic construction of the Jacobian variety.[12] We summarize this.

Theorem 3. *Let C/k be an algebraic curve defined over a subfield k of $\mathbf{C}$. Then the Jacobian variety $J(C)$ of C is an abelian variety defined over k. If C has a k-rational point x, then the Abel-Jacobi map $C \to J(C)$ sending x to 0 is an algebraic morphism defined over k.*

If k is a number field, then $J(C)$ has good reduction at every prime ideal $\mathfrak{p} \subset \mathcal{O}_k$ where C has good reduction.

The key to this theorem is to find an algebraic description of the Jacobian, and in fact one had already been discovered by Abel and Jacobi. Again, a few definitions are required. Let C be a smooth projective algebraic curve (over $\mathbf{C}$, or any algebraically closed field). The group $\mathrm{Div}(C)$ of divisors of C is, by definition, the free abelian group generated by points $x \in C$. An element $D \in \mathrm{Div}(C)$ is thus a formal linear combination $\sum_{x \in C} n_x x$ of points $x \in C$, with $n_x \in \mathbf{Z}$ and only finitely many non-zero n_x. There is a degree homomorphism $\deg : \mathrm{Div}(C) \to \mathbf{Z}$ defined by $\deg(D) = \sum_x n_x$ for D as above. We let $\mathrm{Div}^0(C)$ be its kernel. It is generated by the divisors of the form $x - x_0$ for any fixed $x_0 \in C$, as a simple computation reveals.

Divisors arise naturally when looking at the zeros and poles of a non-zero rational function and their multiplicity. Let $f \in \mathbf{C}(C)$ be a rational function on C, $f \neq 0$. The divisor of f, which is denoted $\mathrm{div}(f)$, is defined to be

$$\mathrm{div}(f) = \sum_{x \in C} \mathrm{ord}_x(f) x$$

(zeros of f minus poles, with multiplicity). Divisors of rational functions are called principal. It is well-known that "there are as many poles as zeros", or in other words

[12] Weil's first proof only showed that the Jacobian was defined over a finite extension of the field of definition of the curve; later Chow, then Weil by other methods, showed that it was defined over the same field.

$\deg \mathrm{div}(f) = 0$ for all non-zero rational functions. What about the converse? Is it true that if $D \in \mathrm{Div}(C)$ is of degree 0, there exists f with $\mathrm{div}(f) = D$? The answer is "No" in general[13], but there is a simple additional condition which ensures that a degree zero divisor is principal, and it is related to the Jacobian.

Fixing a base point $x \in C$, we have an Abel-Jacobi map $\iota : C \to J(C)$, which sends x to $0 \in J(C)$. Since $J(C)$ is a group, it is tempting to try to "add" points of C to get points in $J(C)$ via this map, and of course $\mathrm{Div}(C)$ is just the group where it makes sense (at least, formal sense) to add points of C, so ι extends to a group homomorphism $\iota : \mathrm{Div}(C) \to J(C)$. Now the Abel-Jacobi theorem states that ι is surjective, and that a degree zero divisor $D \in \mathrm{Div}^0(C)$ is principal if and only if $\iota(D) = 0$.

In the special case of an elliptic curve E, when $J(E) \simeq E$, and ι is the identity if 0 is chosen as base-point, this recovers one of the first and most classical results of the theory of elliptic functions ([Si1, VI-2.2]): if f is an elliptic function for a lattice $\Lambda \subset \mathbf{C}$, then

$$\sum_{x \in \mathbf{C}/\Lambda} \mathrm{ord}_x(f) x \in \Lambda,$$

where this time the sum is taken in $\mathbf{C}$. The proof goes by integrating the holomorphic function $z\frac{f'(z)}{f(z)}$ along the boundary of suitable fundamental domain for $\mathbf{C}/\Lambda$; to prove that $\iota(D) = 0$ for D principal on C of higher genus, a similar argument can be used, but it is necessary to use a fundamental domain in the hyperbolic plane, a hyperbolic polygon.

As a consequence of the Abel-Jacobi theorem, there is a bijection

$$J(C) \simeq \mathrm{Pic}^0(C)$$

where, by definition, the degree zero Picard group $\mathrm{Pic}^0(C)$ of C is the quotient of $\mathrm{Div}^0(C)$ by the group of principal divisors. With this identification, the Abel-Jacobi map ι is induced by the map which sends $x \in C$ to the class of the divisor $x - x_0$. There is a moral to this: the "formal" way of adding points on C is closely related to that afforded by the group law of the Jacobian variety. In any case, the main point is that the Picard group, as well as the principal divisors, can be defined[14] for any algebraic variety over a field. To define the Jacobian variety without recourse to analysis, it is enough to find a way of putting the structure of an abelian variety on the abstract group $\mathrm{Pic}^0(C)$! This is by no means easy but it is essentially the way that Weil proceeded. As a starting point for the investigation of $J(C)$, when enough machinery of algebraic geometry is available, it can be surprisingly efficient to deduce properties of $J(C)$ by simply assuming that some abelian variety does exist which has this property (see [Har, pages 323–325]).

The following statement summarizes some notable geometric facts, which will not be used but deserve mention.

Proposition 1. *Let k be an algebraically closed field $k \subset \mathbf{C}$.*

(1) Let C be a smooth projective algebraic curve of genus $g \geqslant 1$ over k. Then the Abel-Jacobi map $C \to J(C)$ is a closed embedding.

---

[13] It is "Yes" when the genus is 0.

[14] With some care in the definitions.

(2) Let A/k be any abelian variety. There exists a curve C/k and a surjective map $J(C) \to A$ (this can be used in many cases to reduce problems about abelian varieties to the corresponding problems restricted to Jacobians; it is not true however that all abelian varieties are Jacobians, and it is a very interesting problem to characterize those that are among all abelian varieties).

(3) (Torelli's theorem) Let C/k be a smooth projective algebraic curve of genus $g \geqslant 1$, and let $(J(C), H)$ be the pair consisting of the Jacobian variety of C with its canonical polarization (see above). Then C is determined, up to k-isomorphism, by $(J(C), H)$: if $(J(C'), H')$ is another such pair and there exists an isomorphism

$$J(C') \simeq J(C)$$

compatible with the polarizations (in an obvious sense), then C is isomorphic to C' over k. However, if one discards the polarizations, there exist non-isomorphic curves C and C' over $\mathbf{C}$ with $J(C) \simeq J(C')$ as abelian varieties.

If C is an algebraic curve defined over a number field k, knowing that its Jacobian variety is also defined over k allows us to ask questions about its arithmetic. In particular, the rank of the Mordell-Weil group of $J(C)$ is defined, and the Birch and Swinnerton-Dyer conjecture should apply to it. It is possible to express the L-function of a Jacobian variety in terms of the original curve in a simple way. Let $\mathfrak{p} \subset \mathcal{O}_k$ be a prime ideal at which the curve C has good reduction, so the curve reduced modulo $\mathfrak{p}$ is a smooth projective curve $C_\mathfrak{p}$ over the residue field $\mathbf{F}_q = \mathcal{O}_k/\mathfrak{p}$ with $q = N\mathfrak{p}$ elements. The congruence zeta function $Z(C_\mathfrak{p})$ of this reduced curve is defined by the formal power series

$$Z(C_\mathfrak{p}) = \exp\Big(\sum_{n \geqslant 1} |C_\mathfrak{p}(\mathbf{F}_{q^n})| \frac{T^n}{n}\Big)$$

(which is a way – eccentric, it might seem at first glance – of encoding the number of points of $C_\mathfrak{p}$ in all finite degree extensions of the residue field).

Using the Riemann-Roch theorem, Schmidt proved in 1931 (Artin had done it in special cases) that $Z(C_\mathfrak{p})$ is the Taylor expansion of a rational function. Following Weil's proof of the Riemann Hypothesis for curves over finite fields, this rational function can be written in the form

$$Z(C_\mathfrak{p}) = \frac{P_1}{(1-X)(1-qX)}$$

where P_1 is a monic polynomial with integral coefficients, of degree $2g$, such that the reciprocals α_i of its complex roots ($1 \leqslant i \leqslant 2g$) satisfy the Riemann Hypothesis (proved by Weil in 1946, using heavily his algebraic theory of the Jacobian, in fact):

$$|\alpha_i| = N\mathfrak{p}^{1/2} = q^{1/2}.$$

Comparing this expression with the definition of $Z(C_\mathfrak{p})$, one derives easily a formula, and a sharp asymptotic, for the number of points of $C_\mathfrak{p}$ in finite fields:

$$|C_\mathfrak{p}(\mathbf{F}_{q^n})| = q^n + 1 + \sum_{1 \leqslant i \leqslant 2g} \alpha_i^n = q^n + O(q^{n/2}).$$

Theorem 4. *(Weil) For all but finitely many $\mathfrak{p}$, the $\mathfrak{p}$-factor of the L-function of the Jacobian $J(C)$ is equal to the polynomial P_1 (in other words, the Hasse-Weil zeta function of $J(C)$ is equal to that of the curve C).*

Remark 1. It is useful to notice here that, in the statement of the Birch and Swinnerton-Dyer conjecture, it is not necessary to use the complete L-function: if S is any finite set of places of k, and $L_S(A,s)$ is the Euler product extended only to primes $\mathfrak{p}$ not in S, then – since the Euler factors themselves do not vanish in $\mathbf{C}$ – the order of vanishing of $L(A,s)$ is the same as that of $L_S(A,s)$. This is significant because in many cases it is easier to determine the L-function up to finitely many "bad" primes. Such a change is also usually innocuous from the analytic point of view.

Remark 2. As a rather humorous observation, it is possible to give a very quick and simple definition of the Jacobian, by an abstract universal property; fix a base point $x \in C$, then $J(C)$ is characterized by the following: it is an abelian variety such that there is a morphism $\iota : C \to J(C)$ sending x to $0 \in J(C)$, which is universal for morphisms of C into abelian varieties sending x to 0: for every abelian variety A/k and every morphism $f : C \to A$ with $f(x) = 0$, there is a unique morphism of abelian varieties $\varphi : J(C) \to A$ with $\varphi \circ \iota = f$, or in other words the diagram

$$\begin{array}{ccc} C & \xrightarrow{\iota} & J(C) \\ & {\scriptstyle f}\searrow & \downarrow{\scriptstyle \varphi} \\ & & A \end{array}$$

commutes.

1.3 The modular curves and their Jacobians

The subject of this thesis is the study of the rank of some abelian varieties by analytic means. The only assumption that we are willing to admit at any point is the Birch and Swinnerton-Dyer conjecture itself, because we want to explore the possibilities it offers for such study. Furthermore, because our techniques will be analytic, it is pretty much necessary to consider an infinite family and not only an individual abelian variety. As mentioned in Section 1.1, the restriction to varieties for which the Hasse-Weil zeta function is known today to be entire limits our choice. The most intriguing, certainly, would be the family (or a large subfamily, say the semi-stable ones) of modular elliptic curves: the behavior of the rank of elliptic curves over $\mathbf{Q}$, on average, is a subject of many guesses and wonder, and any unconditional result would be of great interest. Unfortunately, the current state of knowledge of the analytic properties of the L-functions of elliptic curves is not sufficient to approach this case unconditionally. So we turn to one of the other families: the Jacobians of modular curves.

It is now time to define those curves and discuss the properties of their Jacobians which make them more amenable to today's techniques of analytic number theory.

They are the curves defined over $\mathbf{Q}$ which are usually denoted by $X_0(q)$. The parameter q, an integer $q \geqslant 1$, is called the level (or conductor). We start as usual by discussing them over $\mathbf{C}$, i.e. by considering them as Riemann surfaces. The theory is then very explicit.

Let $\mathbf{H}$ denote the Poincaré upper half-plane,

$$\mathbf{H} = \{z \in \mathbf{C} \mid z = x + iy,\, y > 0\}$$

which is a model of the hyperbolic plane of constant negative curvature when equipped with the Riemannian metric

$$ds^2 = \frac{dx^2 + dy^2}{y^2}.$$

The group $G = GL(2,\mathbf{R})^0$ of 2 by 2 matrices with real coefficients and positive determinant acts isometrically on $\mathbf{H}$ by linear fractional transformations, namely

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d}.$$

The discrete subgroup $SL(2,\mathbf{Z}) < G$ of integer matrices with determinant 1 acts properly discontinuously on $\mathbf{H}$, and its quotient $PSL(2,\mathbf{Z})$ by the center $\{\pm \mathrm{Id}\}$ acts faithfully. The Hecke congruence subgroup of level q, $q \geqslant 1$ an integer, is the subgroup of $SL(2,\mathbf{Z})$ defined by

$$\Gamma_0(q) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbf{Z}) \mid c \equiv 0 \bmod q \right\}$$

(in what follows, when referring to an element $\gamma \in G$, we always mean by a, b, c, and d the four coordinates as in this matrix). Hence $\Gamma_0(1) = SL(2,\mathbf{Z})$ and $\Gamma_0(q)$ is a finite index subgroup of $SL(2,\mathbf{Z})$, with

$$[SL(2,\mathbf{Z}) : \Gamma_0(q)] = q \prod_{p|q} (1 + p^{-1}). \tag{1.8}$$

The Riemann surface $Y_0(q)$ is by definition the quotient space $Y_0(q) = \Gamma_0(q)\backslash\mathbf{H}$, with the complex structure induced by the one on $\mathbf{H}$ (except for some subtlety at the fixed points of some $\gamma \in \Gamma_0(q)$, which are however finite in number (modulo $\Gamma_0(q)$), for which the uniformizer has to be different from the identity function z in order to avoid singularities, see [Sh1]). This quotient is a connected Riemann surface, but it is not compact, because of the presence of cusps. We only define what cusps are geometrically: the best known picture in the theory (see [Se1, ch. 7]) is that of the "standard" fundamental domain F for $SL(2,\mathbf{Z})$ acting on $\mathbf{H}$: let

$$F = \{z \in \mathbf{H} \mid -1 < \mathrm{Re}\,(z) \leqslant 1, \quad |z| \geqslant 1 \ (> 1 \text{ if } \mathrm{Re}\,(z) < 0)\}$$

then F "represents" the quotient space $SL(2,\mathbf{Z})\backslash\mathbf{H}$. This means that for every z in $\mathbf{H}$ there exists a unique $\gamma \in SL(2,\mathbf{Z})$ with $\gamma z \in F$, and if z, w are in F and $\gamma \in SL(2,\mathbf{Z})$ satisfies $\gamma z = w$, then $\gamma = 1$.

It is obvious that F is not compact, because the y coordinate in F can go to infinity: this infinity is interpreted as a point on the boundary $\mathbf{P}^1_{\mathbf{R}} = \mathbf{R} \cup \{\infty\}$ of $\mathbf{H}$, and one says that ∞ is a cusp (the only cusp) of $Y_0(1)$. On the other hand, F has finite volume: since it is a hyperbolic triangle (with a vertex at infinity), the Gauss-Bonnet formula implies

$$\mathrm{Vol}\,(F) = \frac{\pi}{3}$$

where the volume refers to the (hyperbolic) measure

$$d\mu(z) = \frac{dxdy}{y^2}$$

which is invariant under the action of G. Now for any $q \geqslant 1$, since $\Gamma_0(q)$ is of finite index in $SL(2,\mathbf{Z})$, one sees easily that for any finite set (γ_j) of representatives of

$SL(2,\mathbf{Z})/\Gamma_0(q)$, the set $\bigcup_j \gamma_j F$ has the same properties of a fundamental domain for $\Gamma_0(q)$. Hence it has finitely many cusps, the images of ∞ by the γ_j, and finite volume

$$\mathrm{Vol}\,(\Gamma_0(q)\backslash\mathbf{H}) = \frac{\pi}{3}[SL(2,\mathbf{Z}) : \Gamma_0(q)].$$

As a Riemann surface, $X_0(q)$ is obtained by compactifying $Y_0(q)$ by adjoining the cusps. It is useful to notice that the set of cusps – in this case – can be identified with $\mathbf{P}^1_{\mathbf{Q}} = \mathbf{Q}\cup\{\infty\}$ modulo the natural action of $\Gamma_0(q)$. The uniformizer at a cusp is deduced by conjugation from the one for the cusp at infinity, which is $q = e(z) = \exp(2\pi i z)$.

It is usually possible, as far as analysis goes, to work almost exclusively with $Y_0(q)$, looking at the behavior at the cusps when needed, or even with $\mathbf{H}$ and the group action.

The interest of number theorists for curves such as $Y_0(q)$ or $X_0(q)$ came originally from the remarkable properties of automorphic forms with respect to $\Gamma_0(q)$, which can be thought of as objects living on the quotient space. We only require the special case of forms of weight 2, but will give the definition for any (even) integer $k \geqslant 2$. An automorphic form (also called a modular form) f of weight k and level q is a function $f : \mathbf{H} \to \mathbf{C}$ which satisfies the automorphy relation

$$f(\gamma z) = (cz + d)^k f(z)$$

for all $\gamma \in \Gamma_0(q)$ and is holomorphic on $\mathbf{H}$, and also at all cusps. We recall the meaning of the latter, for the cusp at infinity: from the automorphy relation for the special element (a generator of the stabilizer of ∞ in $\Gamma_0(q)$)

$$\gamma = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

it follows that $f(z+1) = f(z)$. Hence there is a holomorphic function

$$g : \{z \in \mathbf{C} \mid 0 < |z| < 1\} \longrightarrow \mathbf{C}$$

such that $f(z) = g(e(z))$ for all $z \in \mathbf{H}$. The form f is said to be holomorphic at infinity if and only if g extends to a function holomorphic in the whole unit disc $\{z \in \mathbf{C} \mid 0 \leqslant |z| < 1\}$. Writing its Taylor expansion around 0, we conclude that f has a so-called Fourier expansion at infinity of the form

$$f(z) = \sum_{n \geqslant 0} a_f(n) e(nz)$$

for some $a_f(n) \in \mathbf{C}$. Then f is further said to vanish at infinity if $a_f(0) = 0$. A similar definition, reducing by conjugacy to this case, holds for all cusps.

Definition 3. The vector space $S_k(q)$ of cusp forms of weight k and level q is the space of all automorphic forms f which vanish at all cusps.

The link with the compactified curve $X_0(q)$, and with its Jacobian, is first hinted at in the next proposition.

Proposition 2. *Let $q \geqslant 1$ be an integer. There is an isomorphism between the space of weight 2 cusp forms and the space $\Omega^1(X_0(q))$ of holomorphic differentials on $X_0(q)$, given by*

$$\begin{cases} S_2(q) & \to & \Omega^1(q) \\ f & \mapsto & f(z)dz \end{cases}$$

Proof. For any holomorphic function f on $\mathbf{H}$, the differential form $f(z)dz$ is well-defined on $\mathbf{H}$. But because of the chain-rule formula

$$d(\gamma z) = (cz+d)^{-2}dz$$

for all $\gamma \in SL(2,\mathbf{R})$, it follows that f is automorphic of weight 2 if and only if $f(z)dz$ is invariant by $\Gamma_0(q)$, hence descends to a differential form on the quotient $Y_0(q)$:

$$f(\gamma z)d(\gamma z) = f(z)dz.$$

Moreover, let

$$f(z) = \sum_{n\geqslant 0} a_f(n)e(nz) = \sum_{n\geqslant 0} a_f(n)q^n$$

be the Fourier expansion of f at infinity. Then since $\frac{dq}{q} = 2\pi i dz$ one has, in terms of the uniformizer q at infinity

$$f(z)dz = \frac{1}{2\pi i}\Big(\sum_{n\geqslant 0} a_f(n)q^{n-1}\Big)dq$$

so that $f(z)dz$ is holomorphic at infinity if and only if $a_f(0) = 0$. A similar reasoning holds at all other cusps. Conversely, starting from a holomorphic differential ω on $X_0(q)$, its pull-back to $\mathbf{H}$ must be of the form $f(z)dz$, and the same reasoning shows that f must be in $S_2(q)$, thereby finishing the proof. □

In particular, $S_2(q)$ is finite dimensional, of dimension equal to the genus of $X_0(q)$, which is also the dimension of its Jacobian variety. This can be computed explicitly, and we quote the result from [Sh1], Propositions 1.40 and 1.43:

$$\dim \Omega^1(X_0(q)) = \frac{1}{12}[SL(2,\mathbf{Z}) : \Gamma_0(q)] + 1 - \frac{1}{2}\sum_{d|q}\varphi\Big((d,\frac{q}{d})\Big) - \frac{\nu_2(q)}{4} - \frac{\nu_3(q)}{3} \qquad (1.9)$$

where ν_2 and ν_3 are defined by

$$\nu_2(q) = \begin{cases} 0 & \text{if 4 divides } q \\ \displaystyle\prod_{p|q}\Big(1+\Big(\frac{-1}{p}\Big)\Big) & \text{otherwise} \end{cases}$$

$$\nu_3(q) = \begin{cases} 0 & \text{if 9 divides } q \\ \displaystyle\prod_{p|q}\Big(1+\Big(\frac{-3}{p}\Big)\Big) & \text{otherwise} \end{cases}$$

and this simplifies, for q prime, to the simpler formula

$$\dim \Omega^1(X_0(q)) = \frac{q+1}{12} - \frac{1}{4}\Big(1+\Big(\frac{-1}{q}\Big)\Big) - \frac{1}{3}\Big(1+\Big(\frac{-3}{q}\Big)\Big). \qquad (1.10)$$

Notice in particular that the genus tends to infinity as q tends to infinity, more precisely it grows like Vol $X_0(q)$:

$$g = \frac{q}{12}\prod_{p|q}(1+p^{-1}) + O_\varepsilon(q^{\frac{1}{2}+\varepsilon}) = \frac{1}{4\pi}\text{Vol } X_0(q) + O_\varepsilon(q^{\frac{1}{2}+\varepsilon}) \qquad (1.11)$$

and for q prime the approximation is even better:

$$g = \frac{q}{12} + O(1). \tag{1.12}$$

So for any fixed integer $g \geqslant 1$, there are only finitely many values of q with $X_0(q)$ of genus g. Here is a table listing the genus for $q = 1$ and for the first few primes:

$$\left[\begin{array}{cccccccccccccc} q = & 1 & 2 & 3 & 5 & 7 & 11 & 13 & 17 & 19 & 23 & 29 & 31 & 37 & 41 \\ g = & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 2 & 2 & 2 & 2 & 3 \end{array}\right] \tag{1.13}$$

The smallest value of q for which the genus is 1 (so $X_0(q)$ is an elliptic curve) is $q = 11$, and the largest is $q = 49$. There are, in principle, algorithms to find explicit polynomial equations for $X_0(q)$ (but they rapidly become intractable). All this is quite classical, and was extensively studied in the early twentieth century by Fricke, Klein, Weber and many others. For $q = 11$, one finds the elliptic curve

$$X_0(11) \,:\, y^2 = 4x^3 - \frac{4.31}{3}x - \frac{41.61}{27};$$

the coefficients are rational, hence $X_0(11)$ is defined over $\mathbf{Q}$. As we shall see a little later, this is a general fact.

The space $S_2(q)$ also acquires a Hilbert space structure with the Petersson inner product (f, g) defined by

$$(f, g) = \int_{\Gamma_0(q)\backslash \mathbf{H}} f(z)\overline{g(z)}d\mu(z)$$

(where the integration can also be performed on any fundamental domain in $\mathbf{H}$ for $\Gamma_0(q)$). This is a very important analytical fact.

It is possible to write down an explicit generating set of cusp forms. They are the Poincaré series P_m ($m \geqslant 1$ an integer) defined by averaging over the group

$$P_m(z) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma_0(q)} e(m\gamma z)(cz + d)^{-2} \tag{1.14}$$

where

$$\Gamma_\infty = \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \in SL(2, \mathbf{Z}) \mid b \in \mathbf{Z} \right\}$$

is the stabilizer of ∞ in $\Gamma_0(q)$.

That the P_m, $m \geqslant 1$, span $S_2(q)$ can be proved rather more easily than one might expect, using decisively the Hilbert space structure (see [Iw2, ch. 3]). Unfortunately, this generating set is infinite and there is no simple way of finding the relations between its elements. Nevertheless, it can be very useful.

In the next chapter, we will see how automorphic forms can be linked more deeply to the arithmetic of $X_0(q)$. But before that we must show that this makes sense. It is again not obvious that the Riemann surfaces $X_0(q)$, or the automorphic forms of level q, are arithmetic in nature. That this is so is seen by means of their modular interpretation[15]:

---

[15]The word "modular", which comes from this source, refers to the terminology used by Cayley for parameters that can be used to classify algebraic varieties, in particular algebraic curves.

it is possible to interpret the points of $Y_0(q)$ as parameterizing isomorphism classes of elliptic curves over $\mathbf{C}$ (i.e. simply quotients $\mathbf{C}/\Lambda$ of $\mathbf{C}$ by a lattice) with some additional "level q structure". Since it is possible to speak about elliptic curves defined over any subfield $k \subset \mathbf{C}$, it is plausible that there is some arithmetic in this definition.

Let $E = \mathbf{C}/\Lambda$ an elliptic curve over $\mathbf{C}$, $q \geqslant 1$ an integer. A level q structure on E is a cyclic subgroup H of E of order q.

Proposition 3. *Let $q \geqslant 1$ be an integer. There is a bijection between $Y_0(q)$ and the set of isomorphism classes of elliptic curves with a level q structure (E, H), where we say that (E, H) is isomorphic to (E', H') if there exists an isomorphism $f : E \to E'$ such that $H' = f(H)$. This bijection is induced by the map*

$$\tau \mapsto \big(\mathbf{C}/(\mathbf{Z} \oplus \tau\mathbf{Z}), <q^{-1}>\big).$$

Proof. Consider first the case $q = 1$: then 0 is the only level 1-structure on any elliptic curve, the set considered is simply the set, say *Ell*, of isomorphism classes of elliptic curves over $\mathbf{C}$. Each is isomorphic to $\mathbf{C}/\Lambda$ for some lattice $\Lambda \subset \mathbf{C}$. Moreover, $\mathbf{C}/\Lambda_1 \simeq \mathbf{C}/\Lambda_2$ if and only if there exists $\alpha \in \mathbf{C}^\times$ such that $\Lambda_1 = \alpha\Lambda_2$, where the multiplication of a lattice by a complex number is defined in the obvious way. Hence $\omega_1\mathbf{Z} \oplus \omega_2\mathbf{Z} = \omega_1\big(\mathbf{Z} \oplus \tau\mathbf{Z}\big)$ for $\tau = \frac{\omega_2}{\omega_1}$. Since ω_1 and ω_2 are $\mathbf{R}$-linearly independent, $\mathrm{Im}\,(\tau) \neq 0$. Switching ω_1 and ω_2 if necessary we thus see that every elliptic curve is isomorphic to one of the form $E_\tau = \mathbf{C}/\Lambda_\tau$ with $\Lambda_\tau = \mathbf{Z} \oplus \tau\mathbf{Z}$ and $\tau \in \mathbf{H}$. This gives a surjective map $\mathbf{H} \to \mathit{Ell}$.

Now suppose $E_\tau \simeq E_{\tau'}$: this means there exists $\alpha \in \mathbf{C}^\times$ satisfying

$$\alpha\big(\mathbf{Z} \oplus \tau\mathbf{Z}\big) = \mathbf{Z} \oplus \tau'\mathbf{Z},$$

in particular there exist integers a, b, c and d such that

$$\begin{cases} \alpha = d + c\tau' \\ \alpha\tau = b + a\tau' \end{cases}$$

so, dividing out

$$\tau = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tau'.$$

Exchanging the role of τ and τ', we find that this integral matrix has an integral inverse. It follows that $E_\tau \simeq E_{\tau'}$ if and only if there exists $\gamma \in SL(2, \mathbf{Z})$ with $\gamma\tau = \tau'$ (the determinant can not be -1 because τ and τ' are both in $\mathbf{H}$). Hence the above map yields the bijection of the proposition when $q = 1$.

For $q \geqslant 2$, the reasoning is the same, but it is necessary to keep track of the level q structure. First one shows that every (E, H) is isomorphic to one of the form $(E_\tau, <q^{-1}>)$: let Λ be a lattice corresponding to E, take a generator ω of the subgroup H, and extend $\omega_2 = q\omega$, which is in Λ and non-zero by definition of a level q structure, to a basis (ω_1, ω_2) of Λ; then multiply by the inverse of ω_1.

Therefore there is again a surjective map from $\mathbf{H}$ to our set of elliptic curves with level q structures. But the condition for the isomorphism $(E_\tau, <q^{-1}>) \simeq (E_{\tau'}, <q^{-1}>)$ is more stringent than before, and becomes that there exists γ in the subgroup $\Gamma_0(q) < SL(2, \mathbf{Z})$ such that $\gamma\tau = \tau'$ (γ has to leave the subgroup $<q^{-1}>$ invariant, which means that it has to be upper triangular modulo q). This completes the proof. □

This proposition makes the next one more believable.

Proposition 4. *Let $q \geqslant 1$ an integer. The curve $Y_0(q)$ has a structure of a smooth affine algebraic curve defined over $\mathbf{Q}$. The compactification $X_0(q)$ has a structure of a smooth projective algebraic curve over $\mathbf{Q}$. Moreover, $Y_0(q)$ has the following modular interpretation: for every algebraically closed field $k \subset \mathbf{C}$, the set $Y_0(q)(k)$ of k-valued points of $\mathbf{C}$ is naturally isomorphic[16] to the set of pairs (E, H), where E/k is an elliptic curve defined over k and H is a cyclic subgroup of order q of E which is defined over k. If $q \geqslant 3$, the same holds for any field $k \subset \mathbf{C}$. The points of $X_0(q)$ have a similar interpretation, the points at the boundary being identified with generalized elliptic curves (certain non-smooth algebraic curves) with level q structures.*

Remark Actually, rather more is true: it is possible to define "integral models" of $X_0(q)$ defined over the ring $\mathbf{Z}[q^{-1}]$; this is necessary to show that $X_0(q)$ has good reduction at all primes p not dividing q (instead of merely at all but finitely many primes), and is therefore important for the fine arithmetic properties of $X_0(q)$. But we have seen that for the Birch and Swinnerton-Dyer conjecture (in the form stated here, at least), this is not crucial.

As an example, take $q = 1$. The genus of $X_0(1)$ is 0 and $X_0(1)_{\mathbf{Q}} \simeq \mathbf{P}^1_{\mathbf{Q}}$; the isomorphism is given by the classical j-invariant map, which analytically is the holomorphic isomorphism induced by the modular function j (see [Se1] for instance) given by the formula (involving Ramanujan's Δ-function and the Eisenstein series E_4 of weight 4 for $SL(2, \mathbf{Z})$)

$$j : \begin{cases} \mathbf{H} & \to & \mathbf{P}^1_{\mathbf{C}} \\ z & \mapsto & \dfrac{1728(60E_4(z))^3}{\Delta(z)} = \dfrac{1}{q} + 744 + 196884q + 21493760q^2 + \dots \end{cases}$$

while for an elliptic curve E given by a Weierstrass equation (1.1), it is

$$j(E) = \frac{1728(4a)^3}{\Delta(E)}.$$

Finally comes the definition we had been aiming at all along.

Definition 4. Let $q \geqslant 1$ an integer. The abelian variety $J_0(q)$ is the Jacobian of the curve $X_0(q)$; it is defined over $\mathbf{Q}$ of dimension given by (1.9).

The Abel-Jacobi map $X_0(q) \to J_0(q)$ is always understood to take the cusp ∞, which is shown to be a $\mathbf{Q}$-rational point, to 0. Hence it is also defined over $\mathbf{Q}$.

Remark Here and hereafter, the interest lies in large values of q; whenever $J_0(q)$ occurs, it is implicit that q is such that the genus of $X_0(q)$ is not zero (if q is prime, this means $q = 11$ or $q \geqslant 17$). Otherwise, take simply $J_0(q) = 0$ for the other values, and the statements should not be too much in error.

The main results of this thesis are the following two theorems, proved in collaboration with P. Michel. They will appear in three papers [KM1], [KM2], [KM3].

---

[16] Naturally means that if $k \subset k'$, then the inclusion $Y_0(q)(k) \subset Y_0(q)(k')$ corresponds to the obvious inclusion of classes of pairs (E, H).

Theorem 5. *Assume the Birch and Swinnerton-Dyer conjecture.*[17] *There exists an absolute and effective constant $C > 0$ such that for any prime number q we have*

$$\operatorname{rank} J_0(q) \leqslant C \dim J_0(q).$$

Theorem 6. Without any hypothesis*, for any $\varepsilon > 0$, and for any prime number q large enough in terms of ε, it holds*

$$\operatorname{rank} J_0(q) \geqslant \Big(\frac{19}{54} - \varepsilon\Big) \dim J_0(q).$$

(The second result has been proved independently by J. Vanderkam [Vdk], with a smaller constant in place of $\frac{19}{54}$).

We will discuss the results previously known, and make a number of comments in the next chapter, after explaining the link between these statements and results of vanishing or non-vanishing of automorphic L-functions and their derivatives at the central critical point.

As a first remark related to the arithmetic of $J_0(q)$, recall that there exists a morphism (defined over $\mathbf{Q}$) $X_0(q) \to J_0(q)$ taking the cusp ∞ to 0. This induces in particular a map from $X_0(q)(\mathbf{Q})$ to the Mordell-Weil group of $J_0(q)$. This might seem to be a good way of producing points on $J_0(q)$, hopefully of infinite order. But this hope is futile: although some rational points do indeed exist, Mazur showed that the only ones are among the cusps, except for very few exceptions that he described. Moreover, Manin and Drinfeld have proved that the image of the cusps are all torsion points of $J_0(q)$.

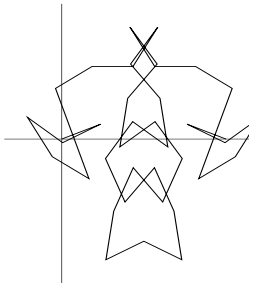
Bibliographical notes

The following references can be read as sources for the material in this chapter:

- [Si1] contains a simple introduction to the basic theory of elliptic curves, over any field, as well as deeper analysis of the arithmetic of elliptic curves.
- [C-S] contains one survey article by M. Rosen about the analytic theory of abelian varieties, and two by J. S. Milne about the algebraic theory and the theory of Jacobian varieties; the historical and bibliographical sketch at the end of Chapter 7 is very interesting. One discovers that the notion of abelian variety, as we know (or learn) it is quite recent.
- [Mu1] starts with the discussion of the analytic theory and then goes to the algebraic theory.
- [Sh1], [Miy] are two good references for the analytic theory of modular curves; Shimura's book has then much more about the more algebraic side, including the very important Eichler-Shimura theory described in Chapter 2. Edixhoven's notes [Edx] are a very good introduction to many aspects of the modular curves $X_0(q)$, over $\mathbf{Q}$ and even $\mathbf{Z}[q^{-1}]$.

[17] More precisely: that the rank be smaller than the order of vanishing of the L-function.

- [Iw2] covers automorphic forms, emphasizing the point of view of analytic number theory.
- [Kob] is a nice introduction to modular forms and the Birch and Swinnerton-Dyer conjecture, using the congruent number problem as *fil rouge*.
- [I-R], especially Chapters 19 and 20 of the second edition, is a very good survey of the Birch and Swinnerton-Dyer conjecture and of the progress made towards it, with special attention to Goldfeld's work on the class number problem.
- Finally, the theory and language of schemes is now the necessary prerequisite to the deeper studies of algebraic geometry and its arithmetic applications; the standard book for this is [Har].



Chapter 2
The analytic side of the theorems

Deux ou trois choses que je sais d'L...
Un film de Jean-Luc Godard.

2.1 Reducing to modular forms

Why is it that the Jacobians of the modular curves $X_0(q)$ lend themselves to so precise an analysis of their rank, via the Birch and Swinnerton-Dyer conjecture, while other abelian varieties do not? The answer lies in their intimate relationship with cusp forms of weight 2. It has already been noticed that there is an isomorphism between the space $\Omega^1(X_0(q))$ of holomorphic 1-forms, which is the analytic "building block" of the Jacobian, and the space $S_2(q)$ of weight 2 cusp forms of level q. This is an analytic link, but there is a much deeper arithmetic link which is the crucial starting point for the proof. This is provided by Eichler-Shimura theory, which computes exactly the Hasse-Weil zeta function of $J_0(q)$ in terms of some specific modular forms.

2.1.1 Hecke theory, primitive forms

We (very) briefly recall the basic theory of Hecke operators and Hecke forms. For any $n \geqslant 1$, there is an Hecke operator $T(n)$, which is an endomorphism of the Hilbert space $S_2(q)$. The subalgebra of $\mathrm{End}\, S_2(q)$ generated by all $T(n)$ is commutative, and moreover the operators $T(n)$ with $(n, q) = 1$ are normal with respect to the Petersson inner product. Hence they are simultaneously diagonalizable. A Hecke form $f \in S_2(q)$ is defined to be a form which is a simultaneous eigenvalue of all those operators $T(n)$, n coprime with q, so there is a basis of $S_2(q)$ consisting of Hecke forms. But one would prefer having simultaneous eigenvalues of all the $T(n)$, without exceptions: only the L-functions of those can have an Euler product and a good functional equation, for instance. Unfortunately, as was discovered by Hecke, there does not always exist a basis of $S_2(q)$ whose elements have this property. Atkin-Lehner theory, however, provides a canonical subspace (usually large) for which this is true. There is an orthogonal decomposition (for the Petersson inner product)

$$S_2(q) = S_2(q)^{old} \oplus S_2(q)^{new}$$

where the space $S_2(q)^{old}$ of oldforms is generated by all the forms of type $f(z) = g(dz)$, where g is a Hecke form of level q/δ for some $\delta \mid q$, $\delta > 1$, and d is a divisor of q/δ, and the space $S_2(q)^{new}$ is simply its orthogonal complement. It is shown that there is a canonical basis $S_2(q)^*$ of this space, characterized by the fact that every $f \in S_2(q)^*$ is an eigenvalue of all the Hecke operators $T(n)$, and its first Fourier coefficient is equal to 1. Such forms are called primitive forms (to emphasize the close analogy with primitive Dirichlet characters).

We now list notations and facts which will be used extensively in the sequel. Let $f \in S_2(q)^*$ be given. We write $\lambda_f(n)$ for its Hecke eigenvalues, which also give the Fourier expansion of f at infinity:

$$f(z) = \sum_{n \geqslant 1} n^{1/2} \lambda_f(n) e(nz). \tag{2.1}$$

The $\lambda_f(n)$ are real algebraic numbers, and the Fourier coefficients $n^{1/2}\lambda_f(n)$ are even algebraic integers. Extracting this factor $n^{1/2}$ from the Fourier expansion, although unsound from the algebraic point of view (for instance, the $\lambda_f(n)$ generate an infinite extension of $\mathbf{Q}$, whereas the Fourier coefficients generate a number field) is convenient for analysis, in particular when considering averages over forms of different weights.

Deligne's bound (the former Ramanujan-Petersson conjecture) for the coefficients of holomorphic cusp forms takes the form

$$|\lambda_f(n)| \leqslant \tau(n) \tag{2.2}$$

for all $n \geqslant 1$ (for weight 2 it is actually a corollary of the Eichler-Shimura formula (2.17) below and Weil's proof of the Riemann Hypothesis for curves).

The Hecke L-function of f is

$$L(f, s) = \sum_{n \geqslant 1} \lambda_f(n) n^{-s}; \tag{2.3}$$

by Deligne's bound it is absolutely convergent for $\mathrm{Re}\,(s) > 1$. Hecke proved the analytic continuation and functional equation: let

$$\Lambda(f, s) = \Big(\frac{\sqrt{q}}{2\pi}\Big)^s \Gamma(s + \tfrac{1}{2}) L(f, s) \tag{2.4}$$

be the completed L-function. Then $\Lambda(f, s)$ (hence also $L(f, s)$) has analytic continuation to an entire function and satisfies the functional equation

$$\Lambda(f, s) = \varepsilon_f \Lambda(f, 1 - s) \tag{2.5}$$

for some $\varepsilon_f \in \{\pm 1\}$ which is called the sign of the functional equation. Thus the critical line for $L(f, s)$ is the line $\mathrm{Re}\,(s) = \frac{1}{2}$, and the center is at $s = \frac{1}{2}$. The opposite of the sign $-\varepsilon_f$ is also the eigenvalue of f for the Atkin-Lehner involution w_q:

$$(f \mid w_q)(z) := f\Big(-\frac{1}{qz}\Big) = -\varepsilon_f f(z).$$

If q is squarefree, ε_f can be computed from the Fourier coefficients by the following formula

$$\varepsilon_f = -\mu(q) q^{1/2} \lambda_f(q). \tag{2.6}$$

We say that a primitive form $f \in S_2(q)^*$ is odd (resp. even) if $\varepsilon_f = -1$ (resp. $\varepsilon_f = 1$). We will write

$$\varepsilon_f^+ = \frac{1 + \varepsilon_f}{2}, \; \varepsilon_f^- = \frac{1 - \varepsilon_f}{2} \tag{2.7}$$

so $f \mapsto \varepsilon_f^+ f$ is, in the basis $S_2(q)^*$ of $S_2(q)^{new}$, the projection operator of the space of primitive forms onto the space of even forms, and correspondingly with ε_f^- for the odd ones. In particular, we have

$$(\varepsilon_f^\pm)^2 = \varepsilon_f^\pm. \tag{2.8}$$

The fact that f is primitive implies that its L-function $L(f,s)$ has an expression as an Euler product, which is of degree 2:

$$L(f,s) = \prod_p (1 - \lambda_f(p)p^{-s} + \varepsilon_q(p)p^{-2s})^{-1} \tag{2.9}$$

and the product converges absolutely for $\mathrm{Re}\,(s) > 1$. This Euler product representation is equivalent with the multiplicativity property of the coefficients $\lambda_f(n)$: for any integers $n \geqslant 1$, $m \geqslant 1$,

$$\lambda_f(n)\lambda_f(m) = \sum_{d|(n,m)} \varepsilon_q(d)\lambda_f\Big(\frac{nm}{d^2}\Big) \tag{2.10}$$

and in particular λ_f is a multiplicative arithmetic function: $\lambda_f(nm) = \lambda_f(n)\lambda_f(m)$ for $(n,m) = 1$. If δ divides q, furthermore, one has for every integer $m \geqslant 1$

$$\lambda_f(\delta m) = \lambda_f(\delta)\lambda_f(m). \tag{2.11}$$

The formula (2.10), by Möbius inversion, yields another useful formula

$$\lambda_f(nm) = \sum_{d|(n,m)} \varepsilon_q(d)\mu(d)\lambda_f\Big(\frac{n}{d}\Big)\lambda_f\Big(\frac{m}{d}\Big). \tag{2.12}$$

If p is a prime not dividing the level q, we factor the polynomial in the p-factor of $L(f,s)$ as follows

$$1 - \lambda_f(p)X + X^2 = (1 - \alpha_p X)(1 - \beta_p X) \tag{2.13}$$

(and sometimes use $\alpha_p(f)$, $\beta_p(f)$ when the dependence on f is important). The bound (2.2) is equivalent (for n coprime with the level) with the assertion that $|\alpha_p| = |\beta_p| = 1$ for all p not dividing q. For p dividing the level, so the p-factor is of degree at most 1, we let $\alpha_p = \lambda_f(p)$, which is shown to be of modulus at most 1 (actually, smaller), and $\beta_p = 0$ (it is also possible that $\alpha_p = 0$), so the polynomial representing the p-factor is still expressed as $(1 - \alpha_p X)(1 - \beta_p X)$.

In addition we require the Dirichlet series expansion for the logarithmic derivative of $L(f,s)$. From the Euler product, using the factorization of the local factors, it follows

$$-\frac{L'}{L}(f,s) = \sum_{n \geqslant 1} b_f(n)\Lambda(n)n^{-s} \tag{2.14}$$

with coefficients given by

$$b_f(n) = \begin{cases} 0, \text{ if } n \text{ is not a power of a prime,} \\ \alpha_p^m + \beta_p^m, \text{ if } n = p^m. \end{cases} \tag{2.15}$$

Finally, assume that q is a prime number. Then, since $X_0(1)$ is of genus 0 (1.13), we have $S_2(1) \simeq \Omega^1(X_0(1)) = 0$, and hence $S_2(q)^{old} = 0$, $S_2(q) = S_2(q)^{new}$. The set of primitive forms $S_2(q)^*$ is therefore a basis of the whole space $S_2(q)$, and

$$|S_2(q)^*| = \dim S_2(q) = \dim J_0(q).$$

This is the main simplification arising from the assumption that q is prime in Theorems 5 and 6; see also the Remark below.

We introduce the notation

$$\omega_f = \frac{1}{4\pi(f,f)} \tag{2.16}$$

for any non-zero cusp form $f \in S_2(q)$ (we call this the "harmonic weight") and define the summation symbol $\sum^h$ by

$$\sum_{f \in S_2(q)^*}^h \alpha_f = \sum_{f \in S_2(q)^*} \omega_f \alpha_f$$

for any family (α_f) of complex numbers. Because (f,f) is of size about $\operatorname{Vol} X_0(q)$, so about $g = \dim J_0(q)$, this behaves like a probability measure, i.e. we have

$$\sum_{f \in S_2(q)^*}^h 1 \sim 1$$

as q tends to infinity (see the Petersson formula (2.29) for $m = n = 1$).

2.1.2 Eichler-Shimura theory and corollaries

Eichler-Shimura theory computes the L-function of certain quotients of the Jacobians of modular curves by means of the L-functions of modular forms. A particular case applies to $J_0(q)$ when q is prime.

Theorem 7. *Let q be a prime number. The L-function of the Jacobian variety $J_0(q)$ is given by* [1]

$$L(J_0(q), s) = \prod_{f \in S_2(q)^*} L(f, s - \tfrac{1}{2}). \tag{2.17}$$

Remark Actually, Eichler and Shimura only proved that for all but finitely many p the p-factors of the Euler products on both sides coincide. Igusa showed that the exceptional p had to be among those dividing the level q, the fact that even those p-factors are the same is due to Carayol. Notice however (as mentioned in the previous chapter) that it would not matter to the applications to the Birch and Swinnerton-Dyer conjecture, in the form we take it, if the equality was known only for all but finitely many primes.

This factorization is analogous to the factorization of the Dedekind Zeta function of the cyclotomic field $\mathbf{Q}(\zeta_n)$ (generated by the n-th roots of unity in $\mathbf{C}$) in terms of Dirichlet L-functions,

$$\zeta_K(s) = \prod_{\chi \bmod n} L(\chi, s).$$

A word about the proof of (2.17): it proceeds one prime at a time, as is natural. The main step is to find a relation between the Frobenius automorphism at p (and its powers) acting on the curve $X_0(q)$ modulo p and the Hecke operator T_p. This requires the geometric interpretation of the Hecke operators, as well as the modular interpretation of $X_0(q)$.

We deduce from the theorem:

---

[1] The shift of $\frac{1}{2}$ occurs because of the normalization chosen for the Hecke L-functions, which have the critical line at $\operatorname{Re}(s) = \frac{1}{2}$.

Corollary 1. *Let q be a prime number. The order of vanishing of the L-function of $J_0(q)$ at $s = 1$ is the sum of the order of vanishing of the Hecke L-functions at $s = \frac{1}{2}$:*

$$\operatorname{ord}_{s=1} L(J_0(q), s) = \sum_{f \in S_2(q)^*} \operatorname{ord}_{s=\frac{1}{2}} L(f, s).$$

If the Birch and Swinnerton-Dyer conjecture holds, then

$$\operatorname{rank} J_0(q) = \sum_{f \in S_2(q)^*} \operatorname{ord}_{s=\frac{1}{2}} L(f, s).$$

This is of great consequence: we see that what the Birch and Swinnerton-Dyer conjecture purports to be the rank of $J_0(q)$ appears as an average of the order of vanishing of automorphic L-functions at the central critical point. Now, the study of the average order of vanishing of various families of L-functions is an old and esteemed part of analytic number theory: this particular case could well be studied independently of any motivation from diophantine geometry, although it does acquire stronger status, and a definite cachet, from this association.

Brumer was the first to study the rank of $J_0(q)$ along those lines [Br1]. He proposed the following conjecture:

Conjecture 2. *As q tends to infinity, the rank of $J_0(q)$ satisfies the asymptotic*

$$\operatorname{rank} J_0(q) \sim \frac{1}{2} \dim J_0(q).$$

This was justified by a simple heuristic, based on the signs of the functional equations for the automorphic L-functions: from (2.5) and the non-vanishing of the Gamma factor at $s = \frac{1}{2}$, it follows that the parity of the order of vanishing of $L(f, s)$ is the same as that of f. So whenever $\varepsilon_f = -1$, the L-function has a zero of order at least 1, and consequently

$$\sum_{f \in S_2(q)^*} \operatorname{ord}_{s=\frac{1}{2}} L(f, s) \geqslant |\{f \in S_2(q)^* \mid \varepsilon_f = -1\}|. \tag{2.18}$$

But it is quite easy to prove that there are about as many odd forms as even ones: for instance when q is squarefree, from the formula (2.6) the difference between the number of even forms and that of odd forms is essentially the trace of the Hecke operator $T(q)$, and one can appeal to the Selberg trace formula (or make a direct computation using the modular interpretation of $X_0(q)$ and the Lefschetz trace formula instead; see also the Petersson formula (2.29) for $m = 1$, $n = q$). Hence, on the Birch and Swinnerton-Dyer conjecture, the lower-bound

$$\operatorname{rank} J_0(q) \geqslant \Big(\frac{1}{2} + o(1)\Big) \dim J_0(q)$$

is immediate. Now the heuristic alluded to is that an L-function will only vanish with "good reasons", such as the one imposed by the sign of the functional equation. In their outstanding majority, the forms f should behave in such a way that the order of vanishing of $L(f, s)$ at $s = \frac{1}{2}$ is the smallest compatible with this symmetry condition: if f is even, $L(f, \frac{1}{2})$ should be non-zero, and if f is odd, the derivative $L'(f, \frac{1}{2})$ should be non-zero. If this is true, then the conjecture follows.

Assuming in addition the Generalized Riemann Hypothesis for the $L(f,s)$, Brumer proved, as a partial confirmation of this heuristic, that the order of magnitude of the rank was not larger than predicted: in the simplest case with which we are concerned here, for q prime, he obtained the bound

$$\sum_{f\in S_2(q)^*} \mathrm{ord}_{s=\frac{1}{2}} L(f,s) \leqslant \Big(\frac{3}{2}+o(1)\Big)\dim J_0(q),$$

and even managed to improve this on average over q. There has been much progress in this direction recently: the constant $\frac{3}{2}$ was improved to $\frac{23}{22}$ in [KM1], then (unpublished) to 1 (assuming moreover the Riemann Hypothesis for Dirichlet L-functions), but Luo-Iwaniec-Sarnak [LIS] have gone much further to obtain a constant $c<1$.

Going beyond 1 turns out to be quite significant: in a larger picture, it means getting beyond the first discontinuity of the Fourier transform of the density of the pair-correlation measure, and this is very strong evidence in favor of the general conjectures of Katz-Sarnak about the distribution of zeros of families of L-functions [K-S], [Sar].

In view of Theorem 7, and since it allows the use of the Birch and Swinnerton-Dyer conjecture, in order to prove the upper bound for $\mathrm{rank}\, J_0(q)$ of Theorem 5 it suffices to establish the following theorem where all mention of algebraic geometry has disappeared.

Theorem 8. *There exists an absolute and effective constant $C>0$ such that for any prime number q we have*

$$\sum_{f\in S_2(q)^*} \mathrm{ord}_{s=\frac{1}{2}} L(f,s) \leqslant C\dim J_0(q).$$

In contrast to this, the lower bound claimed in Theorem 6 scorns the assistance of the Birch and Swinnerton-Dyer conjecture (as it had better do, since the simple consideration of the signs of the functional equations of Hecke L-functions would imply it instantly otherwise) and the beautiful work of Gross and Zagier in the rank 1 case is called for to bring such a translation to analytic number theory.

2.1.3 The Gross-Zagier formula and consequences

The following discussion of the remarkable formula of Gross-Zagier [G-Z] is clearly biased towards its application to the the rank of $J_0(q)$, which is our primary concern, as it seems to clarify the flow of the argument. Also we sketch only the result required for this special case, which is far from being the most general considered in [G-Z].

We are looking for a lower bound for $\mathrm{rank}\, J_0(q)$: this means we want to prove the existence of many independent rational points on $J_0(q)$. Using the modular interpretation of $X_0(q)$, and the group law on the Jacobian, a rather large supply of points will be found, one for each $f\in S_2(q)^*$, say y_f, but only defined over some auxiliary quadratic extension $K/\mathbf{Q}$, there being infinitely many choices of K leading to different points y_f. The y_f are independent in $J_0(q)(K)$ if they are not torsion points, and a criterion is found to ensure this. It is stated, agreeably, in terms of non-vanishing of the derivative of the L-function of f (more precisely, of a lift of f to K). Finally for f odd with $L'(f,\frac{1}{2})\neq 0$, it is shown that $y_f+\overline{y}_f$ is a rational point which is still of

infinite order. This gives the reduction to a non-vanishing problem for special values of L-functions.

First we construct the points y_f, which are called Heegner points (see [Gro] for more details). We first appeal to the modular interpretation of $X_0(q)$ to construct some points in $X_0(q)$. Let K be an imaginary quadratic field of discriminant $D < 0$ [2] such that the prime q splits in K. This condition, in terms of the Kronecker symbol χ_D, is simply $\chi_D(q) = 1$. This is a congruence condition on D, so there exist infinitely many quadratic fields K to which this is applicable. Then q factorizes in K, $q\mathcal{O}_K = \mathfrak{p}\overline{\mathfrak{p}}$, for some prime ideal $\mathfrak{p}$ of $\mathcal{O}_K$. It follows that $\mathcal{O}_K/\mathfrak{p} \simeq \mathbf{Z}/q\mathbf{Z}$.

As an abelian group, $\mathcal{O}_K$ is a free $\mathbf{Z}$-module of rank 2, so it can be thought of as a lattice in $\mathbf{C}$, and the quotient group $\mathbf{C}/\mathcal{O}_K$ is therefore an elliptic curve. This is an example of what are called elliptic curves with complex multiplication, or CM elliptic curves: those are the curves E with the property that the ring of endomorphisms of E is larger than the subring $\mathbf{Z}$ which is obtained by considering the endomorphisms $x \mapsto nx$ of multiplication by integers $n \in \mathbf{Z}$. Here, obviously, $\mathrm{End}(E)$ is isomorphic to $\mathcal{O}_K$. All CM curves can be obtained by a process slightly more general than the one considered here.

Now, a priori $\mathbf{C}/\mathcal{O}_K$ is defined over $\mathbf{C}$ only, but from the theory of complex multiplication ([Si2, II], or [Sh1]) we see that $\mathbf{C}/\mathcal{O}_K$ is in fact defined over the Hilbert class field H_K of K, which we recall is the largest abelian extension of K which is everywhere unramified over K. This field, by Class Field Theory, has degree $h = h(D)$ over K, and more precisely there is a canonical isomorphism (the Artin map)

$$\sigma \,:\, H(K) \longrightarrow \mathrm{Gal}(H_K/K)$$

between the Galois group of H_K and the ideal class-group $H(K)$ of K, which is induced from the map on fractional ideals defined by multiplicativity from $\sigma(\mathfrak{p}) = \mathrm{Fr}_\mathfrak{p}$ for every prime ideal $\mathfrak{p}$ of $\mathcal{O}_K$, $\mathrm{Fr}_\mathfrak{p}$ being the Frobenius endomorphism at $\mathfrak{p}$, which is well-defined for every $\mathfrak{p}$ because the extension is abelian and everywhere unramified.

Let $\mathfrak{p}$ be again one of the two prime ideals dividing the principal ideal $q\mathcal{O}_K$, and let $\mathfrak{p}^{-1}$ be the fractional ideal in K inverse of $\mathfrak{p}$ ($\mathfrak{p}^{-1} = \{x \in K \mid xy \in \mathcal{O}_K \text{ for all } y \in \mathfrak{p}\}$). Then $\mathfrak{p}^{-1}$ is another lattice in $\mathbf{C}$, which obviously contains $\mathcal{O}_K$, and from $\mathcal{O}_K/\mathfrak{p} \simeq \mathbf{Z}/q\mathbf{Z}$ it follows $\mathfrak{p}^{-1}/\mathcal{O}_K \simeq \mathbf{Z}/q\mathbf{Z}$. Hence, by Proposition 3, the pair $(\mathbf{C}/\mathcal{O}_K, \mathfrak{p}^{-1}/\mathcal{O}_K)$ defines a point $x_K \in X_0(q)$ (actually, $Y_0(q)$). Again, the theory of complex multiplication shows that this point is in $Y_0(q)(H_K)$.

This gives us algebraic points on the modular curve, but the field H_K has large degree over $\mathbf{C}$ (by Siegel's theorem, $[H_K : \mathbf{Q}] = 2h(D) \gg_\varepsilon |D|^{\frac{1}{2}-\varepsilon}$ for any $\varepsilon > 0$, the implied constant being ineffective). From an algebraic number $x \in \overline{\mathbf{Q}}$, a way of getting a rational number is to take the trace: by Galois theory, the sum $\sum_\sigma x^\sigma$ of all distinct conjugates of x in $\overline{\mathbf{Q}}$ is a rational number $\mathbf{Q}$. We can not do the same on $Y_0(q)$, the sum doesn't make sense, but of course we can do it on the Jacobian variety! Let $\iota : X_0(q) \to J_0(q)$ be the Abel-Jacobi map (with $\iota(\infty) = 0$). Then the Heegner point associated to K is

$$y_K = \sum_{\sigma \in \mathrm{Gal}(H_K/K)} \iota(x_K)^\sigma \in J_0(q)(\mathbf{C}) \tag{2.19}$$

---

[2] So $K = \mathbf{Q}(\sqrt{d})$ for some squarefree integer $d < 0$ and $D = d$ if $d \equiv 1 \bmod 4$, $D = 4d$ otherwise.

which, by linearity of the action of Galois on points of $J_0(q)$ and Galois theory, is actually a point $y_K \in J_0(q)(K)$. Thus, we have descended to a quadratic extension of $\mathbf{Q}$, but by doing this trace, we might have unfortunately lost everything, if y_K turns out to be of finite order, and this is certainly possible.

For each K (satisfying the conditions described before), we thus obtain a single point y_K. But it is possible to "split" it wondrously, producing one point for each $f \in S_2(q)^*$, by decomposing y_K into its components under the action of the Hecke algebra on the Jacobian $J_0(q)$. More precisely, let $\mathbf{T}$ be the subalgebra of $\mathrm{End}(S_2(q))$ generated by all Hecke operators $T(n)$, $(n,q)=1$. Recall the isomorphisms

$$S_2(q) \simeq \Omega^1(X_0(q))$$

and (see (1.7))

$$\Omega^1(X_0(q)) \simeq T_0^* J_0(q),$$

which show that $\mathbf{T}$ acts on the cotangent space at 0 of the Jacobian variety. In fact, this action is induced (by differentiation) from an action of $\mathbf{T}$ on the Jacobian variety itself, in a way compatible with these isomorphisms: in the description of the points of $J_0(q)$ given by the Abel-Jacobi theorem, as divisor classes, and with the identification of points on $X_0(q)$ as pairs (E,H) of an elliptic curve with a level q structure, the definition is given by

$$T(n)(\iota(E, <x>)) = \sum_G \iota(E/G, <\overline{x}>)$$

where the sum runs over all subgroups of order n in E which have trivial intersection with $H = <x>$, the cyclic subgroup defining the level q structure (so if $(n,q)=1$, G runs over all subgroups of order n), and $\overline{x}$ is the image in E/G of the generator x of H, which is still of order q exactly (for more information on this, see [Edx] for instance).

Moreover (and this is quite clear from this last description) this action of $\mathbf{T}$ is rational over $\mathbf{Q}$, in particular, it induces an action on the finite dimensional $\mathbf{C}$-vector space $V = J_0(q)(K) \otimes \mathbf{C}$ obtained from the Mordell-Weil group of $J_0(q)$ over K.

Hence, for any algebra homomorphism $\lambda : \mathbf{T} \longrightarrow \mathbf{C}$, one can define the λ-isotypical component V^λ of V:

$$V^\lambda = \{y \in V \mid T(y) = \lambda(T)y, \text{ for all } T \in \mathbf{T}\}.$$

For instance, a Hecke form f gives rise to a morphism λ_f characterized by

$$\lambda_f(T(n)) = \lambda_f(n)$$

(remember that the $T(n)$ generate $\mathbf{T}$).

It follows from the existence of the height pairing (described below), for which the action on $\mathbf{T}$ is self-adjoint, that there is a direct sum decomposition

$$V = \bigoplus_\lambda V^\lambda \tag{2.20}$$

of V into its isotypical components. In particular, given the point $y_K \in J_0(q)(K)$, and the corresponding element $y_K \otimes 1 \in V$, one can define the point $y_f \in V$, for any primitive form $f \in S_2(q)^*$, as being the λ_f-isotypical component of y_K in this decomposition.

For us, the Gross-Zagier formula will give a criterion, in terms of L-functions, by which to determine whether a given y_f is non-zero, hence contributes to the rank of $J_0(q)$. This is done by computing the canonical height of y_f. In general, for any abelian variety A/k over a number field, Néron and Tate have constructed a bilinear pairing

$$<\cdot,\cdot> \,:\, A(\overline{k}) \times A(\overline{k}) \longrightarrow \mathbf{C}$$

which is positive, and has the property that the height $h(x) = <x, x>$ of an algebraic point $x \in A(\overline{k})$ vanishes if and only if x is a torsion point (see [C-S], or [Si1] for the case of elliptic curves). Hence, for any extension field k' of k, this Néron-Tate form induces a positive definite hermitian form on the vector space $A(k') \otimes \mathbf{C}$. Now we can state the Gross-Zagier formula.

Theorem 9. *(Gross-Zagier). With notations as above, let $f \in S_2(q)^*$ be given. Put*

$$L_K(f, s) = L(f, s)L(f \otimes \chi_D, s).$$

Then

$$L'_K(f, \tfrac{1}{2}) = \frac{8\pi^2(f, f)}{hw^2|D|^{1/2}} h(y_f). \tag{2.21}$$

Here $2w$ is the number of units in K and $h = h(D)$ is the class number.

Suppose now that $f \in S_2(q)^*$ is odd. Then, by the functional equation, the special value $L(f, \frac{1}{2})$ vanishes, and for any discriminant D with $\chi_D(q) = 1$, the Gross-Zagier formula applies and gives

$$L'_K(f, \tfrac{1}{2}) = L'(f, \tfrac{1}{2})L(f \otimes \chi_D, \tfrac{1}{2}) = \frac{8\pi^2(f, f)}{hw^2|D|^{1/2}} h(y_f). \tag{2.22}$$

Now, it is known that among the discriminants D with $\chi_D(q) = 1$ there exist infinitely many with

$$L(f \otimes \chi_D, \tfrac{1}{2}) \neq 0.$$

This was proved first by Waldspurger, as a consequence of his celebrated theorem relating this special value of the quadratic twist $L(f \otimes \chi_D, s)$ to (the square of) a Fourier coefficient of a half-integral weight modular form related to f by the Shimura lift, but it is much simpler to derive it from scratch by averaging techniques, similar in principle to the ones applied in Chapters 5 and 6 (if only the existence of those discriminants is sought, it is possible to give a very clean and short proof, see [Iw1]).

Fix a discriminant D with this property. Then the formula (2.22), and the definiteness of the height pairing h, imply that $y_f \in V$ is non-zero if and only if $L'(f, \frac{1}{2}) \neq 0$.

Moreover, a simple argument with Heegner points (see [Gro]) shows that the Atkin-Lehner involution w_q, which acts also on $J_0(q)$, acts on the Heegner points like complex conjugation, and it respects the decomposition into isotypical components.

In particular, since an odd form is an eigenfunction of w_q with eigenvalue $-\varepsilon_f = 1$, the f-isotypical component of the $\mathbf{Q}$-rational point $y_K + \overline{y}_K$ in the vector space $W = J_0(q)(\mathbf{Q}) \otimes \mathbf{C} \subset V$ is $2y_f$, which means that y_f is actually in the subspace W. From the decomposition of W into its isotypical components we see that the Gross-Zagier formula implies the unconditional inequality

$$\begin{aligned} \text{rank } J_0(q) = \dim_{\mathbf{C}} W &\geqslant |\{f \in S_2(q)^* \mid W^f \neq 0\}| \quad \text{(obviously)} \\ &\geqslant |\{f \in S_2(q)^* \mid \varepsilon_f = -1 \text{ and } L'(f, \tfrac{1}{2}) \neq 0\}|. \\ &= |\{f \in S_2(q)^* \mid L(f, \tfrac{1}{2}) = 0,\ L'(f, \tfrac{1}{2}) \neq 0\}| \end{aligned}$$

since $L(f, \frac{1}{2}) = 0$ for any odd form f, and the order at $\frac{1}{2}$ of an even form is even.

So Theorem 6 is a consequence of the following non-vanishing theorem for the derivatives of automorphic L-functions, and again all algebraic geometry has silently vanished away.

Theorem 10. *For any $\varepsilon > 0$, and for any prime q large enough in terms of ε, we have*

$$|\{f \in S_2(q)^* \mid L(f, \tfrac{1}{2}) = 0,\ L'(f, \tfrac{1}{2}) \neq 0\}| \geqslant \Big(\frac{19}{54} - \varepsilon\Big) \dim J_0(q).$$

Remark 1. This theorem has been proved independently by J. Vanderkam [Vdk], using a different method, but with a smaller constant. Before, the same non-vanishing problem was considered by Duke in [Du1], who obtained the lower bound

$$|\{f \in S_2(q)^* \mid L(f, \tfrac{1}{2}) = 0,\ L'(f, \tfrac{1}{2}) \neq 0\}| \gg \frac{q}{(\log q)^{10}}$$

with some absolute implied constant. Of course, one can also consider the non-vanishing of the values $L(f, \frac{1}{2})$ themselves: see Remark 3 about this problem.

Remark 2. The proof of this theorem can be extended immediately to apply to higher even weights k such that there are no weight k forms of level 1, namely $k \leqslant 10$ and $k = 14$. It can also be made to apply for weight 12 or $\geqslant 16$, by taking into account, by inclusion-exclusion, the non-primitive forms arising from level 1. It is then also possible, for all weights including 2, to go beyond prime levels q (at least to squarefree levels). This becomes however technically very involved. For the non-vanishing of the special values $L(f, \frac{1}{2})$, this is done in the forthcoming work of Iwaniec and Sarnak [IS1], [IS2]. An extension of the Theorem with respect to the level is actually, in a sense, of more interest than the extension to higher weights, because the latter doesn't have an interaction with algebraic geometry similar to that offered by the Birch and Swinnerton-Dyer conjecture.

Remark 3. Numerically, one has $\frac{19}{54} = 0.35\ldots$ Notice that the only forms f to contribute in the theorem are the odd ones, although we state the density among all primitive forms. Therefore we say, with an obvious abuse of language, that "at least 70 percent" of the odd forms have a zero of order exactly 1 at the critical point. This compares quite favorably with the heuristic behind Brumer's conjecture. There doesn't seem to be any significance attached to the value of the constant, and this is in sharp contradistinction with the corresponding problem of non-vanishing of the values of the L-functions themselves: in the recent work of Iwaniec and Sarnak [IS1] (see also [Sar]), it is shown (among other things) that for any $\varepsilon > 0$ and q large enough in terms of ε, one has

$$|\{f \in S_2(q)^* \mid L(f, \tfrac{1}{2}) > (\log q)^{-2}\}| \geqslant \Big(\frac{1}{4} - \varepsilon\Big) \dim J_0(q). \tag{2.23}$$

which says in particular, for the same reason as before, that about half of the L-functions of even forms do not vanish at $\frac{1}{2}$ (and are even as large as $(\log q)^{-2}$; implicit here is the use of the remarkable fact that $L(f, \frac{1}{2}) \geqslant 0$: this, which of course follows from the Riemann Hypothesis, is true unconditionally by the same theorem of Waldspurger previously mentioned). Furthermore, they show[3] that if one could prove (2.23) with

[3]This was their original motivation for studying this non-vanishing problem.

any better constant $c > \frac{1}{4}$, then it would follow an *effective* lower bound

$$L(1, \chi_D) \gg (\log D)^{-2}$$

for the value at 1 of the L-function of real primitive Dirichlet characters χ_D, or in other words there is no Landau-Siegel zero for real Dirichlet characters; this (yet unproved) statement is one of the major unsolved problems of analytic number theory. As a consequence, by Dirichlet's class number formula, an effective lower bound

$$h(-D) \gg \sqrt{D}(\log D)^{-2}$$

would be obtained for the class number of imaginary quadratic fields $\mathbf{Q}(\sqrt{-D})$ (compare with Goldfeld's bound (1.4), today the best known effective result).

The absence of a barrier for the derivatives similar to that one-fourth for the L-functions themselves (or, maybe more accurately, the fact that this barrier is located much further than at $\frac{1}{4}$) can be explained, at the heuristic/conjectural level at least, by the prediction of Katz-Sarnak [K-S] that the density of the measure which should regulate the imaginary part of the "lowest lying" zero of $L(f, s)$, f odd, vanishes at the origin (to second order), while the corresponding one for even forms doesn't.

Remark 4. And what about non-vanishing problems for higher order derivatives? This is a very different problem, as the heuristic showing that this should be a rare occurrence already makes clear. Analytically, this can be seen from the fact that by computing the number of *odd* forms with $L'(f, \frac{1}{2}) \neq 0$, we are already computing in fact the number of forms with $L(f, s)$ having a zero of order *exactly one* at $s = \frac{1}{2}$: that is to say,

$$\{f \in S_2(q)^* \mid \varepsilon_f = -1 \text{ and } L'(f, \tfrac{1}{2}) \neq 0\} = \{f \in S_2(q)^* \mid L(f, \tfrac{1}{2}) = 0,\ L'(f, \tfrac{1}{2}) \neq 0\},$$

a fact which doesn't generalize to higher orders of vanishing. As the sketch of the proof below will show, this means that to estimate from below

$$|\{f \in S_2(q)^* \mid \operatorname{ord}_{s=\frac{1}{2}} L(f, s) = k\}|$$

(for any $k \geqslant 2$) by the same methods, we would have to consider averages over forms f such that $L^{(k-1)}(f, \frac{1}{2}) = 0$, but this set is very mysterious and there is no good "spectral completeness" to perform the averages, compared to that which exists for all forms.

2.2 Sketch: the upper bound

We will give a fairly detailed sketch of the proof to orient the reader. This section is not completely rigorous but aims for an understanding of the underlying ideas. It provides pointers to the complete proof in Chapter 5.

The principle is the same as the one used by Mestre and Brumer; simply, the analysis has to be done without appealing to the Riemann Hypothesis.

The goal is to bound the sum of the multiplicities of the zeros of the L-functions $L(f, s)$ at $s = \frac{1}{2}$. The standard strategy to do this, exploiting the holomorphy of the L-functions, is to overcount, counting zeros (with multiplicities) in a neighborhood of $\frac{1}{2}$, which is nicely done analytically by means of a contour integration of the logarithmic

derivative of $L(f,s)$. Then one tries to evaluate or estimate the latter. The delicate point is to be able to do this with a neighborhood which is not too large, so that even with the overcounting the result will not be ruined, while the uncertainty principle of harmonic analysis makes it impossible in practice to take too small a neighborhood.

This is actually done by using the so-called "explicit formulae" of analytic number theory. This name refers to formulae of the form

$$\sum_{L(f,\rho)=0} \hat{\psi}(\rho) = \sum_p \lambda_f(p)\psi(p) + \text{other terms} \tag{2.24}$$

relating a sum over primes of the coefficients of the L-function, weighted by some test function ψ, with a sum over the non-trivial zeros of $L(f,s)$, weighted by some integral transform of ψ, the Mellin transform for instance (there are also other terms which are not important, in this case). The first instance of this appears already in Riemann's memoir on the Zeta function, and special cases were extensively used by analytic number theorists. The general form was presented by Weil for Dirichlet L-functions, and then extended and applied by Mestre for abelian varieties.

One wants to take ψ such that $\hat{\psi}$ "approximates" the characteristic function of a suitable small neighborhood of $\frac{1}{2}$, say a circle of radius r_0, while the sum over primes remains manageable. On the Riemann Hypothesis, of course, this is a problem of real variables since the ρ are all on the line $\mathrm{Re}\,(s) = \frac{1}{2}$. It is then not hard to predict that $r_0 = (\log q)^{-1}$ is the limit of the "localization" that one can achieve without getting embroiled in much harder (and usually unworkable) analysis of the sum over primes, which would become "too long" to be dealt with. This barrier is analytic in nature: it is the expression of the uncertainty principle, and boils down to the simple fact that the Fourier transform of the delta function at 0 is the constant function 1.

On the other hand, the classical methods of analytic number theory show that the number $N(f;T)$ of zeros ρ of $L(f,s)$ such that

$$\begin{cases} |\mathrm{Im}\,(\rho)| \leqslant T \\ 0 \leqslant \mathrm{Re}\,(\rho) \leqslant 1 \end{cases}$$

satisfies the asymptotic

$$N(f;T) = \frac{T}{\pi}\Big(\log\frac{qT}{2\pi e}\Big) + O(\log qT)$$

Hence, intuitively at least, one can expect that the number of non-trivial zeros of $L(f,s)$ with imaginary part less than $(\log q)^{-1}$ is – on average – absolutely bounded. If we can reach the limit of the resolution afforded by harmonic analysis, and justify this intuitive argument, it will be possible to deduce that there exists an absolute constant $C > 0$ such that, indeed,

$$\sum_{f\in S_2(q)^*} \mathrm{ord}_{s=\frac{1}{2}}\, L(f,s) \leqslant \sum_{f\in S_2(q)^*} N(f;(\log q)^{-1}) \leqslant C|S_2(q)^*|,$$

which is the statement we are looking for.[4]

[4]While if we took a circle of fixed radius, say 1, the number of zeros would be too large by a factor $\log q$.

This is just a preliminary analysis, which only says that the objective is not utterly hopeless; it could apply, up to now, to any family of automorphic L-functions. But it is not a given either: the corresponding analysis applied to the Selberg trace formula (instead of (2.24)), with the thought of estimating the multiplicity of the eigenvalues of the hyperbolic Laplacian on $\Gamma_0(q)\backslash\mathbf{H}$ (another major problem of analytic number theory), shows that current techniques are doomed to fail, because there are way too many eigenvalues around a fixed $\lambda = \frac{1}{4} + r^2 > 0$ (about r of them in the segment $[r-1, r+1]$) to apply this kind of harmonic analysis.

The challenge is now to prove that the expected behavior of the zeros in this small neighborhood, on average over $f \in S_2(q)^*$, holds. On the Riemann Hypothesis, it is easy to find a test function which localizes as desired on the left-hand side of (2.24); so it is from the right-hand side that we must deduce the validity of this surmise. This is possible because, in a certain precise sense, the Hecke eigenvalues $\lambda_f(n)$ of the forms f are independent enough of each other. This property, a kind of spectral completeness, echoes the fact that the primitive forms are an orthogonal basis of the Hilbert space $S_2(q)$. When summing over f and exchanging the order of summation on the right, one recovers in the inner summation the so-called "Delta symbol" of $S_2(q)^*$:

$$\Delta(m,n) = \sum_{f \in S_2(q)^*} \lambda_f(m)\lambda_f(n)$$

which, *in the range of m and n involved here*, is sufficiently close to the Kronecker delta-symbol $\delta(m,n)$ for this purpose.

But without the Riemann Hypothesis, there is much additional trouble with the zeros on the left-hand side: when their real part can vary, constructing a suitable test function becomes a problem.[5] This is solved by using a test function with a positivity property allowing us to drop the contribution to the sum of the zeros with $|\mathrm{Re}\,(\rho) - \frac{1}{2}| < (\log q)^{-1}$ (their number should be such that this doesn't affect the result), while the possible other zeros with larger real part are handled by means of a density theorem which says, roughly, that on average the $L(f,s)$ have very few zeros ρ with $\mathrm{Re}\,(\rho) \geqslant \frac{1}{2} + \delta$, where δ is about $(\log q)^{-1}$ (see Theorem 14 for the exact statement).

Density theorems of various kinds have a long history in analytic number theory, the first one, for the zeros of $\zeta(s)$ when the imaginary part increases, going back to Bohr and Landau. They are extremely useful to avoid the use of the Riemann Hypothesis. The result we prove is among the more delicate ones, because it must be very efficient near the critical line (other contexts, such as Linnik's theorem on the least prime in an arithmetic progression, require on the contrary density theorems near $\mathrm{Re}\,(s) = 1$ for Dirichlet L-functions; the paper [KM2] also contains a density theorem of this type for automorphic L-functions). Selberg [Sel] proved a result of the kind we need for Dirichlet characters, and we borrow some of his techniques, especially a lemma which reduces the theorem to a bound of the kind

$$\sum_{f \in S_2(q)^*} |M(f, \tfrac{1}{2} + \delta + it)L(f, \tfrac{1}{2} + \delta + it)|^2 \ll |S_2(q)^*|(1+|t|)^B \tag{2.25}$$

for some $B > 0$ (where $\delta = (\log q)^{-1}$, $t \in \mathbf{R}$, and $M(f,s)$ is a partial sum of length q^Δ for some $\Delta > 0$ of the inverse Dirichlet series $L(f,s)^{-1}$). Coincidentally, sums

[5] The transform $\hat{\psi}$ is holomorphic, how can it behave like a characteristic function, even smoothed, of an open subset?

of the same general shape appear prominently in the proof of Theorem 6. In both cases, the "independence" (or lack thereof) of the primitive cusp forms plays again an essential part. We refer to the discussion in the next section, or to the detailed proof in Chapter 5, for further discussion.

2.3 Sketch: the lower bound

Again, this section only provides a non-rigorous overview of the proof of Theorem 10. The statements and formulas written down here should not be taken at face value.

Theorem 10, which estimates odd forms with $L'(f, \frac{1}{2}) \neq 0$, will be proved by comparing the asymptotics of two different moments of the values $L'(f, \frac{1}{2})$, f running over odd forms.

Such a procedure is quite natural, since the theorem can be interpreted as a statement about the distribution of the values of $L'(f, \frac{1}{2})$ among odd forms f: this strategy is a special case of the principle that a measure is completely determined by its moments, and that furthermore partial knowledge of even a few of those moments already contains significant information about it (a principle most famously expounded by Serre in studying the Sato-Tate conjecture from the known cases of analytic continuation of the symmetric powers L-functions). It was used in similar contexts in [Iw1], [Du1], for instance.

In this case, we consider mollified first and second moments

$$M_1 = \sum_{f \in S_2(q)^*} \varepsilon_f^- M(f) L'(f, \tfrac{1}{2})$$
$$M_2 = \sum_{f \in S_2(q)^*} \varepsilon_f^- |M(f) L'(f, \tfrac{1}{2})|^2.$$

where, for the time being, $M(f)$ is simply any complex number.

Remark The last sum is very similar to the second moment (with the L-function itself) considered for the proof of the upper bound (2.25); this coincidence is rather hard to explain. The analysis below applies equally well to this case, with little modifications and actually a number of simplifications (in the former case the $M(f)$ are actually fixed at the outset).

By Cauchy's inequality, we have immediately

$$M_1^2 \leqslant M_2 \times |\{f \in S_2(q)^* \mid \ f \text{ is odd and } L'(f, \tfrac{1}{2}) \neq 0\}|$$

hence, if $M_2 \neq 0$, the lower bound

$$|\{f \in S_2(q)^* \mid \ f \text{ is odd and } L'(f, \tfrac{1}{2}) \neq 0\}| \geqslant \frac{M_1^2}{M_2} \tag{2.26}$$

for the desired cardinality. The choice of the weights $M(f)$ is entirely at our disposal, subject only to M_2 being non-zero, and one can of course try to optimize this choice so as to get the best possible bound. One may also prefer to dispense with it altogether, putting $M(f) = 1$. This seems indeed the most reasonable thing to do, until proved otherwise, but this will turn out to be inefficient: it leads to asymptotics of the type

$$\begin{aligned} M_1 &\sim c_1 (\log q) |S_2(q)^*| \\ M_2 &\sim c_2 (\log q)^3 |S_2(q)^*| \end{aligned} \tag{2.27}$$

for some (explicit) positive constants c_1 and c_2, so (2.26) falls short of the goal by a factor $\log q$. Of course, this is not simply due to some clumsy dealing in the computations, since we have true asymptotics. Rather, this indicates (in itself, a nice fact to know) that while $L'(f, \frac{1}{2})$ is of order of magnitude about $\log q$ on average for f odd, it is quite frequently as large as $(\log q)^{3/2}$. Also worthy of some notice is the fact that since $c_1 > 0$, it follows that $L'(f, \frac{1}{2})$ is positive (still on average). This is of course individually a consequence of the Riemann Hypothesis (a negative derivative at $\frac{1}{2}$ where $L(f, s)$ vanishes would mean that $L(f, s)$ would be negative for s real a little larger than $\frac{1}{2}$, and since it tends to 1 as s tends to $+\infty$, there would have to be a zero not on the critical line), and in this case it is unconditionally known from the Gross-Zagier formula (2.21) but it is also nice to see it come directly from the computation which leads to (2.27).

This brings us back to choosing $M(f)$ non-trivially (and non-artificially: the sum must be manageable) to improve, hopefully, on those asymptotics. The heuristic suggests to try to dampen the oscillations of $L'(f, \frac{1}{2})$ by taking $M(f)$ approximating the inverse $L'(f, \frac{1}{2})^{-1}$. This practice also carries a long history in analytic number theory, first in Bohr-Landau [B-L], and later most notably in the works of Selberg. We select $M(f)$ as a sum

$$M(f) = \sum_{m \leqslant M} \frac{x_m}{\sqrt{m}} \lambda_f(m) \tag{2.28}$$

where $M \geqslant 1$ is a parameter, and the x_m, $m \leqslant M$, are real numbers (not too large, with $x_1 = 1$). Thus M_1 is a linear form in the x_m and M_2 a quadratic form, and the vector (x_m) will be chosen, inasmuch as it is possible, in order to optimize the lower bound for this particular kind of mollifier (the role of M is different, but also crucial: it is expected that the longer the mollifier is, the better the resulting bound – $M(f)$ can better approximate the inverse, in a sense –, while if M is too large, the sums become analytically unmanageable; a correct range has to be found, the existence of which is not entirely clear beforehand).

To handle the sums M_1 and M_2, we use the standard techniques of analytic number theory and the functional equation to express $L'(f, \frac{1}{2})$, which is defined by analytic continuation, as a sum of two short partial sums of the Dirichlet series, and similarly for the square $L'(f, \frac{1}{2})^2$. The partial sums must be of length $\sqrt{q}$ so we obtain, roughly speaking

$$\varepsilon_f^- L'(f, \tfrac{1}{2}) \approx 2 \sum_{n < \sqrt{q}} \frac{\lambda_f(n)}{\sqrt{n}} \Big(\log \frac{\sqrt{q}}{n} \Big).$$

(the formula we get is valid only for odd forms f, otherwise it spells $0 = 0$).

So, summing over m and f and exchanging the order of summation, we arrive at

$$2 \sum_{m \leqslant M} \sum_{n < \sqrt{q}} \frac{x_m}{\sqrt{mn}} \Big(\log \frac{\sqrt{q}}{n} \Big) \Delta^-(m, n)$$

where Δ_- is the Delta-symbol associated to odd cusp forms:

$$\Delta_-(m, n) = \sum_{f \in S_2(q)^*} \varepsilon_f^- \lambda_f(m) \lambda_f(n).$$

At this point the nature of the Jacobian variety, by means of the spectral completeness of the cusp forms, comes into play, as observed before in Section 2.2. The heuristic

is that $\Delta_-(m,n)$ should behave like $\frac{|S_2(q)^*|}{2}\delta(m,n)$, at least in some ranges of m and n. Consider for instance the case of Dirichlet characters modulo q: by orthogonality of characters, the corresponding expression is (for m, n coprime with q, say)

$$\sum_{\chi \bmod q} \chi(m)\chi(n) = \varphi(q)\delta_q(m,n)$$

where $\delta_q(a,b) = 1$ if $a \equiv b \bmod q$ and 0 otherwise. Hence if m and n are both less than q, this is exactly the Kronecker symbol (times the number of characters).

The case of cusp forms can be studied in two ways (at least): in one, the Selberg trace formula is used to express this Delta symbol in closed form, and the "error term" is a sum of class numbers of definite binary quadratic forms of various discriminants (Hurwitz class numbers). This is the method in [Vdk]; it has the disadvantage that the class numbers are not very easy to manipulate analytically, which restricts the efficiency of the final result.

Another method requires to change a little bit the setting: there is a marvelous formula, called the Petersson formula which expresses the Delta symbol *for an orthonormal basis* of $S_2(q)$ in a form which is very congenial to further treatment by methods of analytic number theory. Its proof is also much simpler than that of the Selberg trace formula (see [Iw2]); it relies heavily on the Hilbert space structure of $S_2(q)$, and on the Poincaré series P_m which span it.

Since $S_2(q)^*$ is only an orthogonal basis of $S_2(q)$, it must be normalized before applying the formula, which gives therefore an expression for a weighted Delta-symbol (see the end of Section 2.1.1 for the notation $\sum^h$), namely

$$\sum_{f\in S_2(q)^*}^{h} \lambda_f(m)\lambda_f(n) = \delta(m,n) - 2\pi \sum_{c\equiv 0 \bmod q} \frac{1}{c} S(m,n;c) J_1\Big(\frac{4\pi\sqrt{mn}}{c}\Big) \tag{2.29}$$

where J_1 is a Bessel function, and $S(m,n;c)$ a Kloosterman sum – a most delightful sight to analytic number theorists. Moreover, the factor ε_f^- can be inserted using (2.7), and (2.6), keeping it within reach of hand. Hence the proof of Theorem 10 proceeds in two steps: [6]

(1) We first prove the theorem in a weighted version:

$$\sum_{\substack{\varepsilon_f=-1\\ L'(f,\frac12)\neq 0}}^{h} 1 \geqslant \Big(\frac{19}{54} - \varepsilon\Big)$$

for any $\varepsilon > 0$ and q large enough, by using the method outlined above with weighted moments

$$M_1 = \sum_{f\in S_2(q)^*}^{h} \varepsilon_f^- M(f) L'(f,\tfrac12)$$

and

$$M_2 = \sum_{f\in S_2(q)^*}^{h} \varepsilon_f^- |M(f) L'(f,\tfrac12)|^2$$

[6] Similarly for the case of the upper bound.

(recall that $\sum^h$ behaves like a probability measure).

(2) We show how to "remove" the harmonic weight, getting the same bound for the natural average. This is done by interpreting the Petersson norm (f,f) as a special value at $s=1$ (the edge of the critical strip) of the symmetric square L-function $L(\mathrm{Sym}^2 f, s)$ of f, and by writing the natural sum as an harmonic one with this special value inserted

$$\sum_f \alpha_f = \sum_f^h 4\pi(f,f)\alpha_f.$$

The removal of (f,f) goes by replacing it by a partial sum of the special value $L(\mathrm{Sym}^2 f, 1)$, and dividing this into two parts, a head and a tail. The head is short and is treated by refining the arguments leading to the weighted theorem of Step 1; the tail is shown to be small, in effect by applying the Lindelöf Hypothesis, but precisely – since such a tool is not at our disposal – by using a mean-value estimate for the symmetric square, which has the same power as the Lindelöf Hypothesis on average over f (the first step is again used here). For more details on this last part of the argument, see also Section 3.5.

What do we gain by using this roundabout route (apart from the scenery)? Mainly, flexibility on the side of the series of Kloosterman sums, which in the simple-minded heuristic is supposed to be an error term. Actually the easy bound

$$J_1(x) \ll x$$

and Weil's bound for Kloosterman sums

$$S(m,n;c) \leqslant \tau(c)(m,n,c)^{1/2}c^{1/2}$$

suffice to estimate this series (say $\mathcal{J}(m,n)$) quite sharply,

$$\mathcal{J}(m,n) \ll (m,n,q)(\log(m,n))^2\frac{(mn)^{1/2}}{q^{3/2}},$$

and if m and n are not too large, the heuristic of independence is easily justified.

The flexibility will be especially useful when, for the second moment, it will turn out that in the range we are working with (and which is imposed on us by the nature of the problem) this heuristic *doesn't* hold for odd forms: besides the diagonal term $\delta(m,n)$, a second term will emerge from the series of Kloosterman sums, of the same order of magnitude as the diagonal. Such phenomena have now been observed and exploited in many different contexts, for instance by Duke, Friedlander and Iwaniec ([DFI], among other papers) in their study of automorphic L-functions by the amplification technique, and also by Iwaniec-Sarnak and Luo-Iwaniec-Sarnak in the articles mentioned earlier. The case involved here is actually somewhat simpler (it does not go beyond the range of the large-sieve), but the treatment will nevertheless be quite involved.

After the Petersson formula has been applied, what appears is that if M is small enough ($M < q^{1/4}$ in this case) then, up to smaller contributions, M_1 is a linear form in the x_m which looks like

$$M_1 = \sum_{m\leqslant M}\frac{x_m}{m}\Big(\log\frac{\sqrt{q}}{m}\Big)$$

and M_2 a quadratic form which looks like

$$M_2 = \sum_b \frac{1}{b} \sum_{m_1,m_2} \frac{\tau(m_1 m_2)}{m_1 m_2} x_{bm_1} x_{bm_2} \Big(\log \frac{q}{m_1 m_2}\Big)^3 + \text{other terms.}$$

The problem of performing the choice of x_m to optimize the bound (2.26) is then, theoretically speaking, a "simple" problem of linear algebra. However, it is by no means easy to solve it as explicitly as we want – a closer look at the quadratic form provides some indication. It is necessary to diagonalize M_2, as much as possible, and the computations, if left to grow unchecked, can quickly become unwieldy. However, after some fancy footwork, it is done.

To see quickly how this will give a positive proportion (the precise value, in this case, seems very hard to predict beforehand, although it is a fact that it is the same as the proportion of simple zeros of the Riemann zeta function obtained by Conrey, Ghosh and Gonek !), we wave hands, doing as if the divisor function τ was completely multiplicative, and putting $\log q$ simply in place of $\log \frac{\sqrt{q}}{m}$ (this retains the true order of magnitude, as may be expected). Then

$$M_1 \approx (\log q) \sum_m \frac{x_m}{m}$$

and

$$M_2 \approx (\log q)^3 \sum_b \frac{1}{b} \Big|\sum_m \frac{\tau(m)}{m} x_m\Big|^2 = (\log q)^3 \sum_b \frac{1}{b} |y_b|^2$$

on doing the linear change of variable

$$y_b = \sum_m \frac{\tau(m)}{m} x_{bm}, \text{ for } b \leqslant M.$$

In terms of these variables, by Möbius inversion, M_1 becomes

$$M_1 \approx (\log q) \sum_{k \leqslant M} \frac{\mu(k)}{k} y_k$$

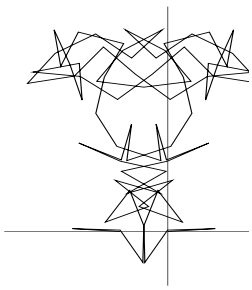
($\mu(k)$ arises as the Dirichlet convolution corresponding to $\zeta(s).\zeta(s)^{-2} = \zeta(s)^{-1}$). Take now $y_k = \mu(k)$ for $k \leqslant M$ (this is optimal by Cauchy's inequality; this confirms the intuition that the mollifier is closely related to the inverse, although the change of variable is needed to make it appear so transparently). Then *if M is a positive power of q*, M_1 has order of magnitude $(\log q)^2$ and M_2 has order of magnitude $(\log q)^4$. Hence the lower bound (2.26) gives a positive proportion. Notice indeed that M of smaller order (a constant, or a power of $\log q$), would not yield this. The value $\frac{19}{54}$ arises as M approaches the limit $q^{1/4}$ allowed in our analysis. In an Appendix to Chapter 6, we sketch a way of lengthening the mollifier up to $q^{1/2}$, following ideas due to Iwaniec and Sarnak [IS1]. The proportion of non-vanishing which comes out of this is $\frac{7}{16}$ instead of $\frac{19}{54}$.

Remark A concrete value of x_m will be chosen to derive Theorem 10. It would be possible to start the computations immediately by plugging those values and ignoring the intermediate step of arriving at them, or even to try to guess what they are (since

the heuristic of the inverse is basically correct). However, this would be quite artificial and misleading. The computations involved in the optimization are quite clean, thanks to changes of variables which are not obvious at the start and might be missed in another derivation.

Also[7], in cold logic, the "two-step" process of treating the harmonic counterparts of M_1 and M_2 first and then removing the weight, is redundant: the straight road to the goal would be to study the sums with a short partial sum of the symmetric square (of length N) inserted, deduce from this that the head of the weight can be handled, then apply the case $N = 1$ to deduce the harmonic result, and use this with the method of the next Chapter to show that the tail of the weight is likewise removable. But again, entangling the first and second moments with this extra partial sum from the start would complicate matters quite a bit, while the present derivation gives the opportunity to get acquainted with the kind of objects and terms which appear, before getting to the worst part of the argument.

Moreover, the proofs are arranged so as to maximize the amount of work done for the harmonic moments which can be reused when the second step is started.



[7]This remark will be clearer after reading the next Chapter.

Chapter 3
A mean-value estimate for the symmetric square of modular forms

"The Romans," Roger and the Reverend Dr. Paul de la Nuit were drunk together one night, or the vicar was, "the ancient Roman priests laid a sieve in the road, and then waited to see which stalks of grass would come up through the holes."
Thomas Pynchon, *"Gravity's Rainbow"*

3.1 The symmetric square of modular forms

The symmetric square L-function of a modular form f was introduced by Shimura in [Sh2]. It has been a very useful tool in many contexts of analytic number theory, and also, in its algebraic form, for algebraic number theory (for instance, it is crucial to Wiles's proof of Fermat's Great Theorem).

Roughly, the symmetric square L-function of f is the Dirichlet series associated to the Fourier coefficients $\lambda_f(n^2)$ at the squares. However, to obtain an entire function, some correction has to be made.

Definition 5. Let $q \geqslant 1$ and $f \in S_2(q)^*$ a primitive form of level q. The symmetric square L-function of f is the Dirichlet series $L(\mathrm{Sym}^2 f, s)$ defined by

$$L(\mathrm{Sym}^2 f, s) = \zeta_q(2s) \sum_{n \geqslant 1} \lambda_f(n^2) n^{-s}. \tag{3.1}$$

We write $\rho_f(n)$ for the coefficients of this Dirichlet series.

The analytic properties of the symmetric square were proved essentially by Shimura (completed by work of Gelbart-Jacquet [G-J], or Coates-Schmidt [CoS] for the functional equation).

Theorem 11. *(Shimura) Let $f \in S_2(q)^*$ be a primitive modular form. The symmetric square $L(\mathrm{Sym}^2 f, s)$ admits an analytic continuation to an entire function. If q is squarefree,*[1] *the completed L-function*

$$\Lambda(\mathrm{Sym}^2 f, s) = \pi^{-3s/2} q^s \Gamma\Big(\frac{s+1}{2}\Big)^2 \Gamma\Big(\frac{s}{2}+1\Big) L(\mathrm{Sym}^2 f, s)$$

satisfies the functional equation

$$\Lambda(\mathrm{Sym}^2 f, s) = \Lambda(\mathrm{Sym}^2 f, 1-s).$$

[1] For q not squarefree, the "correct" symmetric square is not our $L(\mathrm{Sym}^2 f, s)$, as the ramified primes have to be treated more precisely. The functional equation is also more complicated to state.

The following lemma summarizes some properties of the coefficients $\rho_f(n)$.

Lemma 1. *For any $n \geqslant 1$ we have*

$$\rho_f(n) = \sum_{\ell m^2 = n} \varepsilon_q(m)\lambda_f(\ell^2) \tag{3.2}$$

$$\lambda_f(n^2) = \sum_{\ell m^2 = n} \mu(m)\varepsilon_q(m)\rho_f(\ell) \tag{3.3}$$

and in particular $\rho_f(n) = \lambda_f(n^2)$ for n squarefree. Moreover, $L(\mathrm{Sym}^2 f, s)$ has an Euler product expansion of degree 3

$$L(\mathrm{Sym}^2 f, s) = \prod_{(p,q)=1} (1-\alpha_p^2 p^{-s})^{-1}(1-p^{-s})^{-1}(1-\alpha_p^{-2}p^{-s})^{-1} \prod_{p|q}(1-\alpha_p^2 p^{-s})^{-1}$$

where α_p is as in (2.13). Finally, for all $n \geqslant 1$ we have

$$|\rho_f(n)| \leqslant \tau(n)^2. \tag{3.4}$$

The last estimate is proved using Deligne's bound $|\lambda_f(n)| \leqslant \tau(n)$ and the Euler product.

We will need also estimates for $L(\mathrm{Sym}^2 f, 1)$.

Lemma 2. *For all $q \geqslant 1$ and all $f \in S_2(q)^*$, we have*

$$L(\mathrm{Sym}^2 f, 1) \ll (\log q)^3 \tag{3.5}$$

and if q is squarefree

$$L(\mathrm{Sym}^2 f, 1) \gg (\log q)^{-1} \tag{3.6}$$

where the implied constants are absolute in both cases.

Sketch of the proof. The (deeper) lower-bound is the main result of [GHL]; the fact that q is squarefree ensures that f is not a monomial form. The upper bound is much easier and well-known. For q squarefree (the case which will be used), it can be recovered as follows: from the functional equation of $L(\mathrm{Sym}^2 f, s)$, we derive an approximation of the symmetric square at $s = 1$ by a partial sum

$$L(\mathrm{Sym}^2 f, 1) = \sum_{n \leqslant y} \rho_f(n) n^{-1} + O(q^2 y^{-1})$$

and taking $y = q^3$ gives

$$L(\mathrm{Sym}^2 f, 1) \ll \sum_{n \leqslant q^3} \sum_{\ell m^2 = n} \frac{\tau(\ell^2)}{\ell m^2} \ll (\log q)^3$$

because the Dirichlet series

$$\sum_{n \geqslant 1} \tau(n^2) n^{-s} = \frac{\zeta(s)^3}{\zeta(2s)}$$

has a pole of order 3 at $s = 1$ (see Lemma 20).

One can also improve this to $L(\mathrm{Sym}^2 f, 1) \ll \log q$, by using the methods of [GHL]. □

A natural question is to ask when two primitive forms f and g in $S_2(q)^*$ can have the same symmetric square L-function. From the definition of $\rho_f(n)$, $L(\mathrm{Sym}^2 f, s) = L(\mathrm{Sym}^2 g, s)$ is equivalent to $\lambda_f(d^2) = \lambda_g(d^2)$ for all $d \geqslant 1$ which, in turn, is equivalent by multiplicativity to $\lambda_f(d)^2 = \lambda_g(d)^2$ for all $d \geqslant 1$. Since the Fourier coefficients of f and g are real, this means that $\lambda_f(n) = \pm\lambda_g(n)$ for all $n \geqslant 1$. Intuitively, it is natural to expect that those signs $\pm$ correspond to a quadratic Dirichlet character χ, so that f should be a quadratic twist of g (or more precisely, the primitive form associated to a quadratic twist, since the twisted form itself may not be primitive). However, this is quite hard to confirm by a proof, although it seems to have been well-known for some time. The precise result we need is proved by D. Ramakrishnan in the Appendix to [D-K], using properties of the Galois representations associated to holomorphic modular forms (in the case of representations associated to elliptic curves, it is actually proved by Serre, at least for non-integral j-invariant, in [Se2, page 324]). We need a special case only.

Proposition 5. *(Ramakrishnan). Let $f \in S_2(q)^*$ and $g \in S_2(q')^*$ be primitive forms of level q and q' respectively, with q and q' squarefree. Then $L(\mathrm{Sym}^2 f, s) = L(\mathrm{Sym}^2 g, s)$ if and only if $q = q'$ and $f = g$.*

This is contained in the corollary to Theorem A of the appendix to [D-K]. The fact that the levels are squarefree ensures that f can not be a quadratic twist of g, since the conductor of a quadratic twist always has square factors.

3.2 The mean-value estimate

For our purpose, we only require a property of "almost-orthogonality" of the coefficients of the symmetric square L-functions of the forms $f \in S_2(q)^*$. It is implicitly contained in the second part of [D-K], where it was developed for other applications.

Proposition 6. *Let $q \geqslant 1$ be any squarefree integer, and let $N \geqslant q^9$ a real number. The inequality*

$$\sum_{f \in S_2(q)^*} \Big| \sum_{n \leqslant N} a_n \rho_f(n) \Big|^2 \ll N(\log N)^{15} \sum_{n \leqslant N} |a_n|^2 \tag{3.7}$$

holds for any finite family $(a_n)_{1 \leqslant n \leqslant N}$ of complex numbers (with an absolute implied constant).

We will spend a few minutes explaining the meaning of this statement. It is a weaker form of what would be a large-sieve inequality for the symmetric square L-functions. The classical large-sieve inequality, in its sharp multiplicative form due to Gallagher, Bombieri and others, is

$$\sum_{q \leqslant Q} \sideset{}{^*}\sum_{\chi \bmod q} \Big| \sum_{n \leqslant N} a_n \chi(n) \Big|^2 \leqslant (Q^2 + N) \sum_{n \leqslant N} |a_n|^2 \tag{3.8}$$

(the optimal form has $Q^2 + N - 1$ in place of $Q^2 + N$, but this has no consequence for applications).

This has proved to be a very powerful and versatile tool in analytic number theory, essential for instance to the proof of the Bombieri-Vinogradov theorem about the

equidistribution of primes in arithmetic progressions, which in turn can replace the Riemann Hypothesis for Dirichlet characters in many applications.

To see that this inequality can be as strong as the Riemann Hypothesis on average over characters, we apply it to the special coefficients $a_n = \mu(n)$ for $n \leqslant N$, so that the inner sum is

$$M(\chi, N) = \sum_{n \leqslant N} \mu(n)\chi(n).$$

There are about Q^2 characters on the left-hand side. The large-sieve inequality gives

$$\frac{1}{Q^2} \sum_{q \leqslant Q} \sideset{}{^*}\sum_{\chi \bmod q} \Big| M(\chi, N) \Big|^2 \ll \Big(1 + \frac{N}{Q^2}\Big) N$$

and we see that if $N \leqslant Q^2$ then, in mean-square average over primitive characters of conductor at most Q, $M(\chi, N)$ is bounded by $\sqrt{N}$. This is even stronger than what the Riemann Hypothesis for χ would imply, since it is equivalent with the estimate

$$M(\chi, N) \ll_\varepsilon N^{\frac{1}{2}+\varepsilon}$$

for any $\varepsilon > 0$ and $N \geqslant 1$. Another way of saying this is that the large-sieve inequality shows that for bounded coefficients a_n, the sums

$$\sum_{n \leqslant N} a_n \chi(n)$$

are of order of magnitude $\sqrt{N}$ on average, if $N \leqslant Q^2$, so there is as much cancellation as can be expected.

The analogue of (3.8) for the symmetric square would (essentially) be the inequality

$$\sum_{f \in S_2(q)^*} \Big| \sum_{n \leqslant N} a_n \rho_f(n) \Big|^2 \ll (q + N) \sum_n |a_n|^2$$

for any real number $N > 0$. However, the statement of Proposition 6 is quite far from this because it requires a much longer inner sum over n than there are modular forms[2] $f \in S_2(q)^*$. Sharp forms of the large-sieve have been proved for the Fourier coefficients of modular forms by Iwaniec, for level one forms, with respect to the weight, and by Deshouillers and Iwaniec, for forms of fixed level q, with respect to the weight or eigenvalue for Maass forms, and also for fixed weight and level going to infinity. The latter result states that

$$\sum_{f \in \mathcal{F}} \Big| \sum_{n \leqslant N} a_n \lambda_f(n) \Big|^2 \ll \Big(1 + \frac{N}{q}\Big) \sum_n |a_n|^2$$

where $\mathcal{F}$ is any orthonormal basis of the space $S_k(q)$ of holomorphic cusp forms of weight k and level q (which is of dimension about q). Unfortunately, it is not possible today to prove a corresponding statement when averaging over the level q, which requires to

[2]Using a variant of a trick of Viola and Forti, it would be sufficient to prove the proposition with $N > q$ replacing $N > q^9$ to obtain the last inequality.

restrict the outer summation to primitive modular forms (as in the case of Dirichlet characters).

Another interpretation of the large-sieve inequality that is sometimes very fruitful is to consider the inner sum

$$L(\chi) = \sum_{n \leqslant N} a_n \chi(n)$$

as a linear form constructed from the values of a character, and then to employ (3.8) as a way to estimate how many times this can be "large", say $|L(\chi)| \geqslant V$. Indeed, let $N(V)$ be the number of primitive characters χ modulo q with $q \leqslant Q$ such that $|L(\chi)| \geqslant V$. By positivity we derive now

$$V^2 N(V) \leqslant (Q^2 + N) \sum_{n \leqslant N} |a_n|^2.$$

In the case of modular forms, this technique is used for instance in [Du2] to give the first non-trivial upper-bound on the number of non-dihedral weight 1 forms of prime level: the fact that they have associated Galois representations (by the Deligne-Serre theorem) with finite image is used to construct linear forms which are indeed "large" for such forms.

To conclude this section we state an easy corollary for the coefficients $\lambda_f(n^2)$ instead of the $\rho_f(n)$.

Corollary 2. *Let $N \geqslant q^9$ be a real number and $(a(n))_{n \sim N}$ any complex numbers which satisfy*

$$a(n) \ll \frac{(\tau(n) \log n)^A}{n}$$

for some constant $A > 0$. There exists a constant $D = D(A) \geqslant 0$ such that

$$\sum_{f \in S_2(q)^*} \Big| \sum_{n \sim N} a(n) \lambda_f(n^2) \Big|^2 \ll (\log N)^D$$

(with an absolute implied constant).

Proof of Lemma 31.

The point is, of course, that the assumption on the a_n means that we are essentially "on the line $\mathrm{Re}\,(s) = 1$" (or beyond), and in this region the symmetric square behaves as the series

$$\sum_{n \geqslant 1} \lambda_f(n^2) n^{-s}.$$

In exacting details, we have from (3.2)

$$\begin{aligned} \sum_{f \in S_2(q)^*} \Big| \sum_{n \sim N} a(n) \lambda_f(n^2) \Big|^2 &= \sum_{f \in S_2(q)^*} \Big| \sum_{n \sim N} \sum_{\ell m^2 = n} \mu(m) \varepsilon_q(m) \rho_f(\ell) a(n) \Big|^2 \\ &= \sum_{f \in S_2(q)^*} \Big| \sum_{\ell \leqslant 2N} \rho_f(\ell) \tilde{a}(\ell) \Big|^2 \end{aligned}$$

where

$$\tilde{a}(\ell) = \sum_{\sqrt{\frac{N}{\ell}} < m \leqslant \sqrt{\frac{2N}{\ell}}} \mu(m) \varepsilon_q(m) a(\ell m^2).$$

Now we derive from the assumption a bound

$$\tilde{a}(\ell) \ll (N\ell)^{-1/2}(\log \ell)^A,$$

(for some $D \geqslant 0$, with an absolute implied constant), hence the result on applying the mean-value estimate of Proposition 6 to the coefficients $\tilde{a}(\ell)$.

□

3.3 Proof of the mean-value estimate

The proof of Proposition 6 requires more detailed information about the symmetric square. This is best expressed in the language of automorphic representations. We therefore recall the translation between classical modular forms and automorphic representations on $GL(2)$; a readable reference for this is [Gel], and a more recent account is contained in [Bum].

There is an injective map $f \mapsto \pi_f$ from $S_2(q)^*$ to a certain subset of the set of cuspidal automorphic representations of $GL(2)$ over $\mathbf{Q}$ (see [Del], or [Gel]), and this map is compatible with L-functions, in the sense that $L(f,s) = L^f(\pi_f, s)$, where $L(\pi_f, s)$ is the Jacquet-Langlands L-function (complete with the Gamma factor at infinity), which is defined in terms of representation theory (here and elsewhere in this section, L^f, for automorphic-representation L-functions, denotes the finite part of such an L-function, and L is the complete one).

Moreover, Gelbart and Jacquet [G-J] have constructed a symmetric square map $\pi \mapsto \mathrm{Sym}^2 \pi$ associating an automorphic representation of $GL(3)$ to a cuspidal automorphic representation of $GL(2)$ (this is an instance of the Langlands functoriality principle). This representation-theoretic symmetric square is, for $\pi = \pi_f$ associated to a modular form f, related to Shimura's symmetric square L-function in the way that one expects, namely (with the above definition) we have the identity

$$L^f(\mathrm{Sym}^2 \pi_f, s) = L(\mathrm{Sym}^2 f, s)$$

(recall that q is squarefree; Shimura's original definition had slightly different Euler factors at the ramified primes, and others are also needed for q non-squarefree).

We now begin the proof of the mean-value estimate. Observe first that q being squarefree implies that all the symmetric squares $\mathrm{Sym}^2 \pi_f$, $f \in S_2(q)^*$, are actually cuspidal (this generalizes Shimura's result that $L(\mathrm{Sym}^2 f, s)$ is entire, and is needed further on). Indeed, Gelbart and Jacquet have shown that $\mathrm{Sym}^2 \pi$ is cuspidal if and only if π is not monomial, i.e. if π is not obtained from a Hecke character of a quadratic field by automorphic induction. This is impossible here because the level of a monomial form is never squarefree.

The inequality we claim is equivalent to the estimate

$$||T||^2 \ll N(\log N)^{15}$$

(with an absolute implied constant) for the norm of the linear operator

$$T : (a_n)_{n \leqslant N} \mapsto \Big(\sum_{n \leqslant N} a_n \rho_f(n) \Big)_{f \in S_2(q)^*}$$

where both the domain and range are finite dimensional Hilbert spaces (with the canonical hermitian form). We now make use of the duality principle (the norm of an operator is the same as that of its adjoint), but in a somewhat unusual form for variety sake. By general Hilbert-space theory, we have

$$||T||^2 = ||TT^*||$$

and TT^* is the linear operator (endomorphism) defined by

$$(\alpha_f)_{f\in S_2(q)^*} \mapsto \Big(\sum_g \alpha_g K(f,g)\Big)_{f\in S_2(q)^*}$$

where the "kernel" K is (remember that $\lambda_f(n)$, hence $\rho_f(n)$, is always real)

$$K(f,g) = \sum_{n\leqslant N} \rho_f(n)\rho_g(n).$$

Now we conclude immediately from this that

$$||T||^2 \leqslant \operatorname*{Max}_{f\in S_2(q)^*} \sum_{g\in S_2(q)^*} |K(f,g)|,$$

and it is enough to estimate the sums $K(f,g)$.

Actually, one can be analytically more efficient and show the alternate bound

$$||T||^2 \leqslant \operatorname*{Max}_{f\in S_2(q)^*} \sum_{g\in S_2(q)^*} |k(f,g)| \tag{3.9}$$

where the modified kernel k is obtained by choosing (once and for all) a smooth test function $\psi : [0,+\infty[\longrightarrow [0,1]$, compactly supported in $[0,2]$ and identically equal to 1 between 0 and 1, and taking a smoothed version of K:

$$k(f,g) = \sum_{n\geqslant 1} \rho_f(n)\rho_g(n)\psi(n/N)$$

(we leave the proof of this to the reader, or refer to [D-K] where a more classical derivation is given).

We will study the sums $k(f,g)$ by studying the analytic properties of the Dirichlet series

$$L_b(f\otimes g, s) = \sum_{n\geqslant 1} \rho_f(n)\rho_g(n)n^{-s}$$

(which might be called the "bilinear" convolution of the symmetric squares) and expressing the sums as Mellin transforms.

The necessary properties of L_b are consequences of a result which compare it to the Rankin-Selberg convolution of $\mathrm{Sym}^2\,\pi_f$ and $\mathrm{Sym}^2\,\pi_g$. In complete generality, Jacquet, Piatetskii-Shapiro and Shalika have developed a theory of Rankin-Selberg convolutions of automorphic representations on $GL(m_1)$ and $GL(m_2)$ ([JPS] and other papers) over arbitrary global fields. In particular, they have defined a corresponding L-function and studied its properties (analytic continuation and functional equation). Some points which they didn't treat have been established by various other authors (among whom Shahidi, and Moeglin-Waldspurger).

This allows us to consider the L-function $L(\mathrm{Sym}^2\,\pi_f \otimes \mathrm{Sym}^2\,\pi_g, s)$ of the representation-theoretic convolution of the symmetric squares $\mathrm{Sym}^2\,\pi_f$ and $\mathrm{Sym}^2\,\pi_g$. [3] We will prove below a lemma comparing it with L_b, which we state in greater generality (as in [D-K]), since the proof is not harder.

For any automorphic representation π over $\mathbf{Q}$, of conductor $q \geqslant 1$, we denote by $\lambda_\pi(n)$ the coefficients of the finite part of its L-function. We say that π satisfies the Ramanujan-Petersson bound if the bound

$$\lambda_\pi(n) \ll_\varepsilon n^\varepsilon$$

holds for any $\varepsilon > 0$. Because of the Euler product expansion, this will then actually hold uniformly in π (it means that the roots of the local factors of L^f, at the unramified primes, are all of absolute value 1). For $f \in S_2(q)^*$, this is Deligne's bound, and by the Euler product for the symmetric square, it is also true for $\mathrm{Sym}^2\,\pi_f$. The bilinear convolution for automorphic representations π_1 and π_2 is defined as before,

$$L_b(\pi_1 \otimes \pi_2, s) = \sum_{n \geqslant 1} \lambda_{\pi_1}(n)\lambda_{\pi_2}(n) n^{-s}.$$

Lemma 3. *Let π_1 and π_2 be automorphic representations of $GL(n_1)$ and $GL(n_2)$ with conductors q_1 and q_2 respectively, which satisfy the Ramanujan-Petersson bound. There exists an Euler product*

$$H(\pi_1, \pi_2; s) = \prod_p H_p(\pi_1, \pi_2; p^{-s})$$

where $H_p(\pi_1, \pi_2)$ is a rational function for all p and a polynomial (of degree bounded by a constant depending only on n_1 and n_2) for almost all p, such that $H(\pi_1, \pi_2)$ converges absolutely for $\mathrm{Re}(s) > \frac{1}{2}$ (in particular, has no poles in this region), and

$$L_b(\pi_1 \otimes \pi_2, s) = H(\pi_1, \pi_2; s) L^f(\pi_1 \otimes \pi_2, s).$$

Moreover, we have for any $\varepsilon > 0$ and uniformly for $\mathrm{Re}(s) = \sigma > \frac{1}{2}$ a bound

$$H(\pi_1, \pi_2; s) \ll [q_1, q_2]^\varepsilon\, H(\sigma)$$

where H is a fixed Dirichlet series absolutely convergent for $\mathrm{Re}(s) > \frac{1}{2}$ satisfying in this region

$$H(\sigma) \ll (\sigma - \tfrac{1}{2})^{-A}$$

for some $A > 0$ depending only on n_1 and n_2.

This lemma simply reflects the fact that the coefficients of $L^f(\pi_1 \otimes \pi_2, s)$ and $L_b(\pi_1 \otimes \pi_2, s)$ are the same for squarefree integers n, so their analytic behavior is the same up to the critical line.

Coming back to the proposition, we derive from the lemma the analytic continuation of L_b up to the critical line, because $L^f(\pi_1 \otimes \pi_2, s)$ has a meromorphic continuation

[3] This convolution has already been used in other contexts of analytic number theory by Hoffstein and Lockhart [H-L] and by Luo, Rudnick, Sarnak [LRS] to obtain deep results about $GL(2)$ automorphic forms, especially non-holomorphic Maass forms.

to $\mathbf{C}$ by the work of Jacquet, Piatetskii-Shapiro and Shalika. Now we apply Mellin inversion and derive

$$k(f,g) = \frac{1}{2i\pi}\int\limits_{(3)} N^s\hat{\psi}(s)L_b(\mathrm{Sym}^2\,\pi_f\otimes\mathrm{Sym}^2\,\pi_g,s)ds$$
$$= \frac{1}{2i\pi}\int\limits_{(3)} N^s\hat{\psi}(s)H(\mathrm{Sym}^2\,\pi_f,\mathrm{Sym}^2\,\pi_g;s)L^f(\mathrm{Sym}^2\,\pi_f\otimes\mathrm{Sym}^2\,\pi_g,s)ds,$$

where $\hat{\psi}(s)=\int_0^{+\infty}\psi(x)x^s\frac{dx}{x}$ is the Mellin transform of ψ.

Move the line of integration to $\mathrm{Re}(s)=\frac{1}{2}+c$, where $c<\frac{1}{2}$ will be chosen later; the Mellin transform $\hat{\psi}$ is holomorphic for $\mathrm{Re}(s)>0$ and quickly decreasing in any vertical strip $\delta<\mathrm{Re}(s)<b$ with $\delta>0$; the other terms in the integral being at most of polynomial growth (see [R-S] for the Rankin-Selberg convolution), shifting the contour is possible and the only singularities we can pick up by doing so are those of $L^f(\mathrm{Sym}^2\,\pi_f\otimes\mathrm{Sym}^2\,\pi_g,s)$. They are known from the Rankin-Selberg theory. Precisely, it is proved in [M-W]:

Theorem 12. *Let π_1 and π_2 be cuspidal automorphic representations of $GL(n_1)$ and $GL(n_2)$ respectively (over a number field). If there are no $t\in\mathbf{C}$ such that $\pi_1=\pi_2\otimes|\cdot|^t$, then $L(\pi_1\otimes\tilde{\pi}_2,s)$ is entire.*

If $\pi_1=\pi_2$, then $L(\pi_1\otimes\tilde{\pi}_2,s)$ has two simple poles at 0 and 1 and is holomorphic outside those points.

In our case, $\mathrm{Sym}^2\,\pi_f$ and $\mathrm{Sym}^2\,\pi_g$ both have trivial central character so we have $\mathrm{Sym}^2\,\pi_f=\mathrm{Sym}^2\,\pi_g\otimes|\cdot|^t$ for $t=0$ only, and this theorem says that poles arise only when $\mathrm{Sym}^2\,\pi_f=\mathrm{Sym}^2\,\pi_g$.

Keeping this in mind, we estimate the integral on the other line, namely

$$\frac{1}{2i\pi}\int\limits_{(\frac{1}{2}+c)} N^s\hat{\psi}(s)H(\mathrm{Sym}^2\,\pi_f,\mathrm{Sym}^2\,\pi_g;s)L^f(\mathrm{Sym}^2\,\pi_f\otimes\mathrm{Sym}^2\,\pi_g,s)ds.$$

We are only interested in the q-aspect of the matters. By the bounds for H in Lemma 3, for any $\varepsilon>0$ we have

$$H(\pi_1,\tilde{\pi}_2;\tfrac{1}{2}+c+it)\ll_\varepsilon q^\varepsilon c^{-A}.$$

As for the Rankin-Selberg convolution, after inserting the correct Gamma factors it has a functional equation relating its value at s with that at $1-s$ (because it is self-contragredient; for this, see the references to several articles of Shahidi in [M-W]):

$$L(\mathrm{Sym}^2\,\pi_f\otimes\mathrm{Sym}^2\,\pi_g,s)=\tau(f,g)q(f,g)^{\frac{1}{2}-s}L(\mathrm{Sym}^2\,\pi_f\otimes\mathrm{Sym}^2\,\pi_g,1-s)$$

where $\tau(f,g)$ is a complex number of absolute value 1 and $q(f,g)=q(\mathrm{Sym}^2\,\pi_f\otimes\mathrm{Sym}^2\,\pi_g)$ is the conductor of $\mathrm{Sym}^2\,\pi_f\otimes\mathrm{Sym}^2\,\pi_g$. By a theorem of Bushnell and Henniart [B-H], it is bounded by the product of the cubes of the conductors of $\mathrm{Sym}^2\,\pi_f$ and $\mathrm{Sym}^2\,\pi_g$, each of which is at most q^2, so $q(f,g)\leqslant(q^4)^3=q^{12}$.

From the functional equation, Stirling's formula and the convexity principle of Phragmen-Lindelöf, this implies in turn

$$L^f(\mathrm{Sym}^2\,\pi_f\otimes\mathrm{Sym}^2\,\pi_g,\tfrac{1}{2}+c+it)\ll q^{12(1/4-c/2)}|t|^E=q^{3-6c}|t|^E$$

for some (absolute) $E > 0$. Taken together with the previous bound for H, and using the fact that $\hat{\psi}$ decreases faster than any polynomial on the line, we see that the integral on $\mathrm{Re}\,(s) = \frac{1}{2} + c$ is $\ll_\varepsilon c^{-A} N^{\frac{1}{2}+c} q^{3-6c+\varepsilon}$. Since by assumption $N > q^9$, we deduce from this by taking $c = (\log q)^{-1}$ (so that $1 \ll q^c \ll 1$, $N^c \ll 1$ and $c^{-A} = (\log q)^A$) that for any $\varepsilon > 0$

$$k(f,g) = \delta(\mathrm{Sym}^2\,\pi_f, \mathrm{Sym}^2\,\pi_g)\hat{\psi}(1)NR_f + O_\varepsilon(N^{5/6+\varepsilon}) \tag{3.10}$$

where $\delta(\cdot,\cdot)$ is the Kronecker delta, and R_f is the residue at $s = 1$ of the bilinear convolution $L_b(\mathrm{Sym}^2\,\pi_f \otimes \mathrm{Sym}^2\,\pi_f, s)$. To determine that only the diagonal $f = g$ contributes to the main term, we appeal to Proposition 5: since the level q is squarefree, $\mathrm{Sym}^2\,\pi_f = \mathrm{Sym}^2\,\pi_g$, which implies $L(\mathrm{Sym}^2\,f, s) = L(\mathrm{Sym}^2\,g, s)$, is equivalent to $f = g$. Then knowing that $f = g$ we can go around R_f easily by coming back to the definition

$$\begin{aligned} k(f,f) &= \sum_{n \geqslant 1} |\rho_f(n)|^2 \psi(n/N) \leqslant \sum_{n \leqslant 2N} |\rho_f(n)|^2 \\ &\leqslant \sum_{n \leqslant 2N} \tau(n)^4 \ll N(\log N)^{15} \end{aligned}$$

(where the implied constant is absolute). From this and (3.9) we therefore get

$$||T||^2 \ll_\varepsilon N(\log N)^{15} + qN^{5/6+\varepsilon} \ll N(\log N)^{15}$$

by taking ε small enough, with an absolute constant. There is actually some margin left in the bound, but it doesn't seem to matter in the current applications; it is still very far away from a sharp large-sieve type inequality, which would be much more widely applicable.

We now prove Lemma 3. We write

$$L^f(\pi_i, s) = \sum_{n \geqslant 1} \lambda_i(n) n^{-s}$$

for the finite part of the standard L-functions, so

$$L_b(\pi_1 \otimes \pi_2, s) = \sum_{n \geqslant 1} \lambda_1(n)\lambda_2(n) n^{-s};$$

we have to compare this Dirichlet series and the Rankin-Selberg convolution $L^f(\pi_1 \otimes \pi_2, s)$.

The Rankin-Selberg convolution has an Euler product by the general theory, and the bilinear convolution also has one because it's a Dirichlet series whose coefficients are multiplicative:

$$L_b(\pi_1 \otimes \pi_2, s) = \prod_p \sum_{k \geqslant 0} \lambda_1(p^k)\lambda_2(p^k) p^{-ks}.$$

Therefore, since we claim the existence of an Euler product

$$H(\pi_1, \pi_2) = \prod_p H_p(\pi_1, \pi_2)$$

relating the two, we can proceed locally for each prime p.

For any automorphic L-function, we denote by L_p its p-factor, considered as a polynomial (in p^{-s}) with complex coefficients.

Assume first that p is an unramified prime of the Rankin-Selberg convolution. This is true for almost all p, and we will prove now the existence of a polynomial $H_p(\pi_1, \pi_2)$ such that

$$\sum_{k\geqslant 0} \lambda_1(p^k)\lambda_2(p^k)X^k = H_p(\pi_1, \pi_2)L_p(\pi_1 \otimes \pi_2). \tag{3.11}$$

We know that p is unramified for both π_1 and π_2, so that the p-factor of the standard L-function is

$$L_p(\pi_i)^{-1} = \prod_{1\leqslant j\leqslant n_i} (1 - \alpha_{i,j}X) \tag{3.12}$$

where $\alpha_{i,j}$ are the Satake parameters of the local representation for π_i at p. Again, the general theory gives the p-factor of the Rankin-Selberg convolution

$$L_p(\pi_1 \otimes \pi_2)^{-1} = \prod_{\substack{1\leqslant j\leqslant n_1\\ 1\leqslant k\leqslant n_2}} (1 - \alpha_{1,j}\alpha_{2,k}X).$$

Assume, to begin with, that the $\alpha_{i,j}$ are all distinct and the $\alpha_{1,j}\alpha_{2,k}$ also. Coming then to the p-factor of the bilinear convolution, we deduce from the Dirichlet series for $L^f(\pi_i)$

$$\begin{aligned}\sum_{k\geqslant 0} \lambda_i(p^k)X^k &= \prod_{1\leqslant j\leqslant n_i} (1 - \alpha_{i,j}X)^{-1}\\ &= \sum_{1\leqslant j\leqslant n_i} \frac{r_{i,j}}{1 - \alpha_{i,j}X}\end{aligned}$$

for some complex numbers $r_{i,j}$ (partial fraction expansion, since the α's are distinct), whence

$$\lambda_i(p^k) = \sum_{1\leqslant j\leqslant n_i} r_{i,j}\alpha_{i,j}^k.$$

This implies

$$\begin{aligned}\sum_{k\geqslant 0} \lambda_1(p^k)\lambda_2(p^k)X^k &= \sum_{k\geqslant 0} \Big(\sum_{\substack{1\leqslant i\leqslant n_1\\ 1\leqslant j\leqslant n_2}} r_{1,i}r_{2,j}\alpha_{1,i}^k\alpha_{2,j}^k \Big) X^k\\ &= \sum_{i,j} \frac{r_{1,i}r_{2,j}}{1 - \alpha_{1,i}\alpha_{2,j}X}.\end{aligned}$$

Reducing to a common denominator, which is exactly $L_p(\pi_1 \otimes \pi_2)$, we get the required formula (3.11).

Moreover, it is obvious that the coefficients of $H_p(\pi_1, \pi_2)$ are polynomials in the α's and since the Ramanujan bound implies $|\alpha_{i,j}| \leqslant 1$ it follows that those coefficients are bounded by some constants depending only on n_1 and n_2. Hence the absolute convergence (and the absence of poles) in $\mathrm{Re}(s) > \frac{1}{2}$ of the product over the unramified primes will follow if we can show that the coefficient of X of $H_p(\pi_1, \pi_2)$ vanishes, since there is no term in p^{-s} then.

But for any rational function

$$r = \frac{f}{g}$$

with polynomials f and g, satisfying $r(0) = 1$, the coefficient of X of the numerator f of r is $f'(0)$, and so equals $g(0)r'(0) + g'(0)$. If $r = \sum_k b_k X^k$ is the power series expansion of r, we have therefore

$$f'(0) = g(0)b_1 + g'(0).$$

Assume moreover that $g = \prod_j (1 - \beta_j X)$. Then

$$f'(0) = b_1 - \sum_j \beta_j.$$

Applying this to the local factor of L_b, which is of this form, we see that the corresponding coefficient is indeed zero since

$$\lambda_1(p)\lambda_2(p) = \sum_{i,j} \alpha_{1,i}\alpha_{2,j}, \tag{3.13}$$

(which means simply that the bilinear and the true convolution have the same coefficients for squarefree integers).

We can now use a continuity argument to deduce that the existence of the polynomial H_p satisfying formula (3.11) and the vanishing of the coefficient of X remain valid when some of the roots of the local L-functions are the same.

It remains to treat the case of the ramified primes. The local factor at p of the L-functions of π_1 and π_2 is still of the form

$$L_p(\pi_i) = \prod_{1 \leqslant j \leqslant n_i'} (1 - \alpha_{i,j} X)^{-1}$$

for some $n_i' \leqslant n_i$. The same proof as the unramified case shows again that the local factor of the bilinear convolution is a rational function which has poles only among the reciprocals of the products $\alpha_{1,j}\alpha_{2,k}$. So we can define $H_p(\pi_1, \pi_2)$ by

$$H_p(\pi_1, \pi_2) = \Big(\sum_{k \geqslant 0} \lambda_1(p^k)\lambda_2(p^k)X^k\Big) L_p(\pi_1 \otimes \pi_2)^{-1} \tag{3.14}$$

and it's also a rational function.

It remains to establish that the finite product over the ramified primes has no pole for $\mathrm{Re}(s) > \frac{1}{2}$. But a pole s_0 of $H_p(\pi_1, \pi_2, p^{-s})$ must satisfy

$$\alpha_{1,j}\alpha_{2,k}p^{-s_0} = 1$$

(for some j and k), so by the Ramanujan bound again we get $\mathrm{Re}(s_0) \leqslant 0$.

As for bounding $H(\pi_1, \pi_2; s)$, clearly by the Ramanujan bound the product over the unramified primes is absolutely convergent for $\mathrm{Re}(s) > \frac{1}{2}$. It is dominated (termwise) by the Euler product H whose factors are obtained by taking the corresponding factor of H_p and replacing each coefficient of the polynomial by its absolute value, which in turn, since the coefficient of X^2 is absolutely bounded (say by A), is dominated by an

Euler product which may be written (by factoring by force $\zeta(2s)$) as $\zeta(2s)^A J(s)$ where $J(s)$ is absolutely convergent for $\mathrm{Re}(s) > \frac{1}{3}$. The estimate

$$H(\sigma) \ll (\sigma - \tfrac{1}{2})^{-A}$$

then follows directly.

We now estimate the product over the ramified primes

$$\prod_{p|[q_1,q_2]} H_p(\pi_1, \pi_2; p^{-s})$$

using (3.14). For $L_p(\pi_1 \otimes \pi_2)^{-1}$, which is a polynomial of degree at most $n_1 n_2$, we write by the Ramanujan bound:

$$\prod_{p|[q_1,q_2]} L_p(\pi_1 \otimes \pi_2; p^{-s})^{-1} \leqslant \prod_{p|[q_1,q_2]} (1 + p^{-\sigma})^{n_1 n_2} \leqslant (\prod_{p|[q_1,q_2]} 2)^{n_1 n_2} \ll_\varepsilon [q_1, q_2]^\varepsilon$$

for any $\varepsilon > 0$, since the number of prime divisors of an integer n is $O(\log n / \log\log n)$.

On the other hand, still by Ramanujan, for any $\varepsilon > 0$

$$\sum_{k \geqslant 0} \lambda_1(p^k)\lambda_2(p^k) p^{-ks} \ll_\varepsilon \sum_{k \geqslant 0} p^{k(\varepsilon - s)} = \frac{1}{1 - p^{-s+\varepsilon}}$$

so that taking the product over $p \mid [q_1, q_2]$ we obtain by the same reasoning the same bound as above for the product of those terms, and finally

$$\prod_{p|[q_1,q_2]} H_p(\pi_1, \pi_2; p^{-s}) \ll_\varepsilon [q_1, q_2]^\varepsilon .$$

This concludes the proof of the mean-value estimate. As mentioned, it is very far from what one might expect by analogy with other large-sieve inequalities. We end this section by showing that the proof given, relying on the duality principle and the analytic properties of the Rankin-Selberg convolutions, can not yield any improvement in the range of N for which the mean-value estimate is valid.

Indeed, for the special case of the symmetric squares of two forms in $S_2(q)^*$, we can make the comparison in Lemma 3 much more explicit. Take a prime p, unramified for f and g, and consider the p-factors of the symmetric squares. They can be written

$$L_p(\mathrm{Sym}^2 f) = (1 - \alpha^2 X)(1 - X)(1 - \alpha^{-2} X)$$

and

$$L_p(\mathrm{Sym}^2 g) = (1 - \beta^2 X)(1 - X)(1 - \beta^{-2} X),$$

say, for some α and β of absolute value 1. Using patience and cunning, or any symbolic computation software, we find from this (the formula $\lambda_f(p^2) = \alpha^2 + 1 + \alpha^{-2}$ is used repeatedly) the p-factor H_p of the lemma, namely

$$H_p = 1 - \lambda_f(p^2)\lambda_g(p^2) X^2 + \big(\lambda_f(p^2)^2 + \lambda_g(p^2)^2 - 2\big) X^3 - \lambda_f(p^2)\lambda_g(p^2) X^4 + X^6.$$

Disregarding the ramified primes, which will not interfere with this discussion, we see that, up to another Euler product which is absolutely convergent (and uniformly

bounded) for $\mathrm{Re}\,(s) > 1/3$, the comparison function $H(\mathrm{Sym}^2 f, \mathrm{Sym}^2 g)$ is equal to the inverse of the Rankin-Selberg convolution $L^f(\mathrm{Sym}^2 f \otimes \mathrm{Sym}^2 g, 2s)^{-1}$ (or of L_b, which is the same in that sense). Therefore, we can not go beyond the critical line without encountering poles from the zeros of this convolution: the Riemann Hypothesis enters the game, and there is no hope of going further in this easy way.

A curious fact perhaps deserves mention here: the polynomial H_p is self-reciprocal and even, so if x is a root of H_p, then $-x$, x^{-1} and $-x^{-1}$ are also roots. But there is no reason to expect that H_p satisfies the local Riemann Hypothesis, in other words that all its roots are on the unit circle. However, because of the symmetry, if the roots are distinct, then one being on the unit circle implies that four of them at least are. Actually, if one considers H_p as a polynomial depending on two parameters α and β (on the unit circle), the set R of points (α, β) such that all the roots of H_p are unitary is a rather large subset of $\mathbf{S}^1 \times \mathbf{S}^1$.

And in practice, working numerically with the L-functions of some concrete elliptic curves (without CM), one finds that a large percentage of primes p are such that H_p does satisfy the local Riemann Hypothesis. This is accounted for by the Sato-Tate conjecture on the distribution of the arguments of the α_p, and in fact the proportion of p agrees (experimentally) with the measure of the set R (with respect to the product of the Sato-Tate measure on each component).

If a practical analytic interpretation of the roots of the local factors H_p could be found, it might be possible to provide a rigorous proof that this agreement holds as p tends to infinity, and this would provide some theoretical evidence for the Sate-Tate conjecture.

3.4 Notational matters

We will be dealing quite extensively with sums over $f \in S_2(q)^*$. The following notations are designed to emphasize the underlying structure. We usually suppose given a family $\alpha = (\alpha_f)$ of complex numbers, defined for all forms $f \in S_2(q)^*$, q being any level, or maybe restricted to squarefree or prime levels. We then introduce the "natural" averaging operator

$$A[\alpha] = \sum_{f \in S_2(q)^*} \alpha_f$$

where we only sum over forms of a fixed level, and consider the behavior of $A[\alpha]$ as a function of the level q, asymptotically as q gets large. So the interpretation of an inequality which is written

$$A[\alpha] \leqslant f(q)$$

(respectively, $\geqslant$, $\ll \dots$), for any function f, is the following: for q large enough (possibly, q prime large enough), it holds

$$\sum_{f \in S_2(q)^*} \alpha_f \leqslant f(q)$$

(respectively, $\geqslant$, or there exists an absolute constant $C > 0$ with

$$\sum_{f \in S_2(q)^*} \alpha_f \leqslant C f(q)$$

for all q large enough).

Similarly we define the "harmonic" averaging operator

$$A^h[\alpha] = \sum_{f \in S_2(q)^*}^{h} \alpha_f$$

recalling (see (2.16)) that $\sum^h$ means

$$\sum_f^h \alpha_f = \sum_f \omega_f \alpha_f$$

with

$$\omega_f = \frac{1}{4\pi(f,f)}.$$

As a matter of notational convenience, we often write somewhat loosely $A^h[\alpha_f]$ instead of $A^h[\alpha]$, which avoids losing a letter to denote each family (α_f) we wish to average.

3.5 Removing the harmonic weight: the tail

The relevance of Proposition 6 to this thesis, as we mentioned at the end of Section 2.3, is that the harmonic weight $\omega_f = \frac{1}{4\pi(f,f)}$, which is required to express the Δ-symbol of the modular forms by Petersson's formula, is related to the special value of the symmetric square L-function at $s = 1$, which is the edge of the critical strip (in our "analytic" normalization). This is essentially due to Shimura [Sh3].

Proposition 7. *Let q be squarefree and $f \in S_2(q)^*$ a primitive form. Then*

$$4\pi(f,f) = \frac{q}{12\zeta(2)} L(\mathrm{Sym}^2 f, 1) + O_\varepsilon(q^{1/2+\varepsilon})$$

for any $\varepsilon > 0$, and if q is prime

$$4\pi(f,f) = \frac{\dim J_0(q)}{\zeta(2)} L(\mathrm{Sym}^2 f, 1) + O((\log q)^3) \tag{3.15}$$

uniformly in f as q tends to infinity. In particular we have

$$\omega_f \ll \frac{\log q}{q} \tag{3.16}$$

uniformly for $f \in S_2(q)^*$.

Sketch of the proof. We introduce the Rankin-Selberg convolution of f with itself

$$L(f \otimes f, s) = \zeta_q(2s) \sum_{n \geqslant 1} \lambda_f(n)^2 n^{-s}.$$

By a simple formal computation using the multiplicativity of the Fourier coefficients, we have the identity

$$L(f \otimes f, s) = \zeta_q(s) L(\mathrm{Sym}^2 f, s).$$

On the other hand, $L(f \otimes f, s)$ is given by the Rankin-Selberg integral (see [Iw2, page 245])

$$(4\pi)^{-1-s}\Gamma(s+1)\zeta_q(2s)^{-1}L(f\otimes f,s) = \int_{\Gamma_0(q)\backslash \mathbf{H}} |f(z)|^2 E(z,s)dxdy$$

where $E(z,s)$ is the non-holomorphic Eisenstein series for $\Gamma_0(q)$.

The right-hand side of this equation is meromorphic with a simple pole at $s=1$, from the known analytic properties of $E(z,s)$: this already shows that $L(f\otimes f,s)$ is meromorphic, hence $L(\mathrm{Sym}^2 f, s)$ also (the point in Shimura's theorem is that the symmetric square is actually entire).

The residue of $E(z,s)$ at $s=1$ is the constant function Vol $X_0(q)^{-1}$, so the computation of the residue at $s=1$ on both sides yields

$$\frac{1}{(4\pi)^2\zeta(2)}\prod_{p|q}(1+p^{-1})^{-1}L(\mathrm{Sym}^2 f,1) = \frac{(f,f)}{\mathrm{Vol}\ X_0(q)}$$

hence

$$\begin{aligned} 4\pi(f,f) &= \frac{\mathrm{Vol}\ X_0(q)}{4\pi\zeta(2)}\prod_{p|q}(1+p^{-1})^{-1}L(\mathrm{Sym}^2 f,1) \\ &= \frac{q}{12\zeta(2)}L(\mathrm{Sym}^2 f,1) + O_\varepsilon(q^{\frac{1}{2}+\varepsilon}) \quad \text{(by (1.11) and (3.5))}. \end{aligned}$$

Moreover, the more precise formula (1.12) gives indeed, for q prime

$$4\pi(f,f) = \frac{\dim J_0(q)}{\zeta(2)}L(\mathrm{Sym}^2 f,1) + O((\log q)^3)$$

and the last statement is a consequence of the previous ones and the lower bound (3.6) by the very definition $\omega_f^{-1} = 4\pi(f,f)$. □

Suppose we have a family $\alpha = (\alpha_f)$ of complex numbers, for all $f \in S_2(q)^*$ with prime level q, and that we know the behavior of the weighted sum

$$A^h[\alpha] = \sum_{f\in S_2(q)^*}^{h} \alpha_f$$

(for instance, we have an asymptotic formula for q going to infinity), but wish to obtain the same information for the natural sum

$$A[\alpha] = \sum_{f\in S_2(q)^*} \alpha_f.$$

Since, by Petersson's formula for $m=n=1$, $A^h[1] = 1 + O(q^{-3/2})$, we expect that when α is well-distributed and not biased against the Petersson inner-product (or, what amounts to the same thing, against the value of the symmetric square at $s=1$), we should have

$$A[\alpha] \sim \dim J_0(q) A^h[\alpha]$$

meaning that $L(\mathrm{Sym}^2 f, 1)$ and α_f act here as independent random variables would, with the average of $L(\mathrm{Sym}^2 f, 1)$ equal to the obvious constant factor $\zeta_q(2)$ (which is equivalent to $\zeta(2)$ as q tends to infinity).

In this section we build a method to approach this problem, and – using the mean-value estimate established before – prove a result which solves part of the problem for quite general vectors α. This reduces to another estimate which has to be supplied independently in each case, as we will do in the next chapters when the time comes to conclude the proofs of theorems 8 and 10.

3.5.1 Sketch of the idea

We sketch the idea first, since the technical details will tend to obscure it. We make the assumption that $\alpha = (\alpha_f)$ satisfies the conditions

$$A^h[|\alpha_f|] \ll (\log q)^A \qquad \text{(for some absolute } A > 0) \tag{3.17}$$

$$\operatorname*{Max}_{f \in S_2(q)^*} |\omega_f \alpha_f| \ll q^{-\delta} \qquad \text{(for some } \delta > 0) \tag{3.18}$$

as the level q (prime) tends to infinity. Neither of these conditions is very restrictive in practice: the first one is interpreted as saying that $|\alpha_f|$ is "almost" bounded, and can often be achieved by some normalization. If this is true, the second condition is fairly reasonable since we have shown in (3.16) that $\omega_f \ll (\log q) q^{-1}$. In other words, by normalizing if necessary, both conditions can be expected to hold whenever the size of α_f doesn't increase or oscillate wildly.

We write the unweighted average as a weighted one and replace the Petersson inner product by the special value of the symmetric square (3.15):

$$\begin{aligned} A[\alpha] &= \sum_{f \in S_2(q)^*}^{h} 4\pi(f,f)\alpha_f \qquad (3.19)\\ &= \frac{\dim J_0(q)}{\zeta(2)} \sum_{f \in S_2(q)^*}^{h} L(\mathrm{Sym}^2 f, 1)\alpha_f + O((\log q)^3 A^h[|\alpha_f|]). \end{aligned}$$

We wish to replace the value of the symmetric square by a partial sum of the Dirichlet series. This can be done for a long enough sum, say of length y. Thus we define

$$\omega_f(y) = \sum_{n \leqslant y} \rho_f(n) n^{-1} \tag{3.20}$$

and the sum above is, up to a very small quantity, equal to

$$\sum_{f \in S_2(q)^*}^{h} \omega_f(y)\alpha_f.$$

This is now a finite sum of averages over the α_f, twisted by symmetric square coefficients $\rho_f(n)$:

$$\sum_{f \in S_2(q)^*}^{h} \omega_f(y)\alpha_f = \sum_{n \leqslant y} \frac{1}{n} A^h[\rho_f(n)\alpha_f].$$

If by any chance the methods which give us control over the average $A^h[\alpha_f]$ (corresponding to $n = 1$) also apply to the twisted ones, *in the range* $n < y$, then we are

done. Unfortunately, in applications this will only be the case for very small values of y. For instance, in the cases we will consider, we can control the twists only for $y < q^\varepsilon$, where $\varepsilon > 0$ is very small (indeed, arbitrarily so). So, can we recover the L-function from such a short sum (3.20)?

On the Riemann hypothesis (more precisely, the Lindelöf hypothesis suffices for this purpose) this is indeed true: then we have, for any real number σ and any s with $\mathrm{Re}\,(s) > \sigma > \frac{1}{2}$

$$L(\mathrm{Sym}^2 f, s) = \sum_{n<q^\varepsilon} \rho_f(n) n^{-s} + O(q^{-\delta}) \tag{3.21}$$

for any $\varepsilon > 0$, and some $\delta(\sigma, \varepsilon) > 0$, with an absolute implied constant. This is an encouraging sign, but of course this statement is out of reach, and individually we can only take y much larger ($y = q^2$ or maybe $y = q$), and indeed too large for our application.

But we can again exploit the average over f that we have to (even want to) perform. As discussed in Section 3.2, a sharp form of the mean-value estimate of Proposition 6 would be as powerful as the Lindelöf hypothesis on average over f, and it might be used to implement (3.21) on average. But of course, only the weak form of the proposition is at our disposal. Thankfully, we are still saved because we are looking at the symmetric square at a point on the edge of the critical strip, where the Dirichlet series almost converges absolutely. Then the "extra length" needed to enter the effective range of n for the mean-value estimate will not matter, much as the partial sums

$$\sum_{n<q^\delta} n^{-1}$$

of the harmonic series are of the same size as q tends to infinity for any fixed $\delta > 0$.

3.5.2 The tail of the series

We come to the implementation of this idea. Let therefore $\alpha = (\alpha_f)_{f\in S_2(q)^*}$ be given for all q prime, satisfying the conditions (3.17). First, since the conductor of $\mathrm{Sym}^2 f$ for $f \in S_2(q)^*$ is q^2, the functional equation and the usual estimates give the approximation

$$L(\mathrm{Sym}^2 f, 1) = \omega_f(y) + O(q^2 y^{-1}) \tag{3.22}$$

(with an absolute implied constant). We assume $\log y = O(\log q)$, say $y < q^{10}$.

Now let $x < y$ be given. The partial sum is further decomposed as

$$\omega_f(y) = \omega_f(x) + \omega_f(x, y)$$

where

$$\omega_f(x, y) = \sum_{x<n\leqslant y} \rho_f(n) n^{-1}.$$

We consider here the weighted average built with the tail, namely

$$A^h[\omega_f(x, y)\alpha_f] = \sum_{f\in S_2(q)^*}^h \omega_f(x, y)\alpha_f.$$

We will use Hölder's inequality to separate $\omega_f(x, y)$ and α_f. The former is handled by the following lemma.

Lemma 4. *Let $r \geqslant 1$ be an integer, such that $x^r \geqslant q^{11}$. There exists a positive constant $C = C(r) > 0$ such that*

$$A[\omega_f(x,y)^{2r}] \ll (\log q)^C$$

where the implied constant is absolute.

The proof starts with some other lemmas. We say that an integer n is squarefull if for any prime p dividing n, p^2 divides n; in other words, for all p dividing n, the valuation of p in n is at least 2. Notice that

$$\sum_{n \text{ squarefull}} n^{-s} = \prod_p (1 + p^{-2s} + p^{-3s} + \ldots)$$

which converges absolutely for $\mathrm{Re}\,(s) > \frac{1}{2}$, hence we have

$$\sum_{\substack{n \text{ squarefull} \\ n > z}} n^{-1} \ll z^{-1/2} \tag{3.23}$$

with an absolute implied constant.

Lemma 5. *For any integer $r \geqslant 1$ and any $f \in S_2(q)^*$, we can write*

$$\omega_f(x,y)^r = \sum_{x^r < mn \leqslant y^r} \lambda_f(m^2) \frac{c(m,n)}{mn} \tag{3.24}$$

with $c(m,n) = 0$ unless n can be written

$$n = dn_1, \text{ with } d \mid m,\ n_1 \text{ squarefull.} \tag{3.25}$$

and there exists $\gamma = \gamma(r) > 0$ such that

$$|c(m,n)| \leqslant \tau(mn)^\gamma.$$

Moreover, the coefficients c depend on r, x and y but not on the form f.

Proof. We proceed by induction on r. For $r = 1$, we write by (3.2)

$$\begin{aligned} \omega_f(x,y) &= \sum_{x < n \leqslant y} \frac{1}{n} \sum_{\ell m^2 = n} \varepsilon_q(m) \lambda_f(\ell^2) \\ &= \sum_{x < \ell m^2 \leqslant y} \lambda_f(\ell^2) \frac{\varepsilon_q(m)}{\ell m^2} \end{aligned}$$

so we can take $c(\ell, m) = 0$ unless m is square and $c(\ell, m^2) = \varepsilon_q(m)$.

Assume that (3.24) holds for some r and s as claimed, with coefficients c (for r) and c' (for s). Then

$$\begin{aligned} \omega_f(x,y)^{r+s} &= \sum_{\substack{x^r < m_1 n_1 \leqslant y^r \\ x^s < m_2 n_2 \leqslant y^s}} \lambda_f(m_1^2)\lambda_f(m_2^2) \frac{c(m_1,n_1)c'(m_2,n_2)}{m_1 n_1 m_2 n_2} \\ &= \sum_{\substack{x^r < m_1 n_1 \leqslant y^r \\ x^s < m_2 n_2 \leqslant y^s}} \sum_{\substack{d \mid m_1^2 \\ d \mid m_2^2}} \lambda_f\Big(\frac{m_1^2 m_2^2}{d^2}\Big) \frac{\varepsilon_q(d) c(m_1,n_1) c'(m_2,n_2)}{m_1 n_1 m_2 n_2} \end{aligned}$$

by multiplicativity for λ_f.

Now d can be written uniquely as $d = d_1 d_2^2$ with d_1 squarefree and then we have $d \mid m^2$ if and only if $d_1 d_2 \mid m$. Therefore we can write

$$\begin{cases} m_1 = d_1 d_2 m_1' \\ m_2 = d_1 d_2 m_2' \end{cases}$$

and then

$$\omega_f(x,y)^{r+s} = \sum_{\substack{x^r < d_1 d_2 m_1' n_1 \leqslant y^r \\ x^s < d_1 d_2 m_2' n_2 \leqslant y^s}} \lambda_f\big((d_1 m_1' m_2')^2\big) \frac{\varepsilon_q(d_1 d_2) c(d_1 d_2 m_1', n_1) c'(d_1 d_2 m_2', n_2)}{(d_1 d_2)^2 m_1' m_2' n_1 n_2}.$$

Now write $m_0 = d_1 m_1' m_2'$, $n_0 = d_1 d_2^2 n_1 n_2$. By the induction hypothesis we see that if $c(m_1, n_1) \neq 0$ and $c'(m_2, n_2) \neq 0$, then n_0 can be written as $\delta n_0'$ with $\delta \mid m_0$ and n_0' squarefull (this is not absolutely obvious because $m_1 m_2$ does not divide m_0, but the extra prime divisors can be pushed to the squarefull part).

Estimating rather trivially the multiplicity of representation of m_0, we find the desired representation. This immediately concludes the induction. □

Lemma 6. *Let $z \geqslant 1$ be given and the coefficients $c(m,n)$ be as in lemma 5 for r. Then there exists $A = A(r) > 0$ such that*

$$\sum_{\substack{x^r < mn \leqslant y^r \\ n > z}} \lambda_f(m^2) \frac{c(m,n)}{mn} = O\big(z^{-1/2} (\log qz)^A\big).$$

Proof. By Deligne's bound we have

$$\sum_{\substack{x^r < mn \leqslant y^r \\ n > z}} \lambda_f(m^2) \frac{c(m,n)}{mn} \leqslant \sum_{x^r < m \leqslant y^r} \frac{\tau(m)}{m} \sum_{\substack{x^r m^{-1} < n \leqslant y^r m^{-1} \\ n > z}} \frac{|c(m,n)|}{n}$$

but using the condition on the support of $c(m,n)$, the inner sum is

$$\begin{aligned} \sum_{\substack{x^r m^{-1} < n \leqslant y^r m^{-1} \\ n > z}} \frac{|c(m,n)|}{n} &\leqslant \tau(m)^\gamma \sum_{d \mid m} \frac{1}{d} \sum_{\substack{n \text{ squarefull} \\ dn > z}} \frac{\tau(n)^\gamma}{n} \\ &\ll \tau(m)^{\gamma+1} z^{-1/2} (\log z)^A \end{aligned}$$

(by (3.23) and the result follows. □

Lemma 7. *There exists a real number M such that $x^r z^{-1} < M \leqslant y^r z$, and real numbers $c(m)$ such that we have*

$$\sum_{f \in S_2(q)^*} \omega_f(x,y)^{2r} \ll (\log qz)^B \sum_{f \in S_2(q)^*} \Big| \sum_{m \sim M} \lambda_f(m^2) \frac{c(m)}{m} \Big|^2 + O(qz^{-1/2} (\log qz)^B)$$

and

$$|c(m)| \leqslant \tau(m)^C (\log qm)^C$$

for some $C > 0$.

Proof. By the previous lemma

$$\omega_f(x,y)^{2r} = \sum_{n\leqslant z}\Big|\sum_{x^r<mn\leqslant y^r}\lambda_f(m^2)\frac{c(m,n)}{mn}\Big| + O(qz^{-1/2}(\log qz)^A).$$

Write $\xi_n = \operatorname{sign}\Big(\sum_{x^r<mn\leqslant y^r}\lambda_f(m^2)\frac{c(m,n)}{mn}\Big)$, split the summation over dyadic intervals in m, then use Cauchy's inequality and sum over f: the result follows for some M with

$$c(m) = \sum_{x^r m^{-1}<n\leqslant z}\xi_n\frac{c(m,n)}{n} \ll \tau(m)^C(\log qm)^C$$

for some $C>0$, as desired. □

This now easily implies Lemma 4: we take $z=q^2$, then the assumption $x^r \geqslant q^{11}$ implies that $M\geqslant q^9$ and we may appeal to the mean-value estimate of Corollary 2 to bound the first term, with $\log M \ll \log q$.

Proposition 8. *Let (α_f) be complex numbers satisfying conditions (3.17), and $x=q^\kappa$ for some $\kappa>0$. There exists an absolute constant $\gamma=\gamma(\kappa,\delta)>0$ (δ the exponent in (3.17)) such that*

$$A^h[\omega_f(x,y)\alpha_f] \ll q^{-\gamma}$$

and

$$A[\alpha_f] = \frac{\dim J_0(q)}{\zeta(2)}A^h[\omega_f(x)\alpha_f] + O(q^{1-\gamma}).$$

Proof. Let $r\geqslant 1$ be any integer. By Hölder's inequality we have (with s the complementary exponent to $2r$, $(2r)^{-1}+s^{-1}=1$)

$$\begin{aligned}
A^h[\omega_f(x,y)\alpha_f] &= \sum_{f\in S_2(q)^*}^{h}\omega_f(x,y)\alpha_f\\
&= \sum_{f\in S_2(q)^*}\omega_f\omega_f(x,y)\alpha_f\\
&\leqslant A[\omega_f(x,y)^{2r}]^{\frac{1}{2r}}\Big(\sum_{f\in S_2(q)^*}(\omega_f|\alpha_f|)^s\Big)^{\frac{1}{s}}\\
&\leqslant A^{\frac{1}{2r}}A[\omega_f(x,y)^{2r}]^{\frac{1}{2r}}A^h[|\alpha_f|]^{\frac{1}{s}}
\end{aligned}$$

where we have denoted

$$A = \operatorname*{Max}_{f\in S_2(q)^*}\omega_f|\alpha_f|.$$

Take now r large enough so that $x^r\geqslant q^{11}$ ($r=[11\kappa^{-1}]+1$ suffices). Then Lemma 4 gives

$$A[\omega_f(x,y)^{2r}]^{\frac{1}{2r}} \ll (\log q)^D$$

for some $D=D(\kappa)>0$, while we have, from (3.18) and (3.17) respectively,

$$\begin{aligned}
A^{\frac{1}{2r}} &\ll q^{-\gamma_0} \quad \text{for some } \gamma_0=\gamma_0(\kappa,\delta)>0,\\
A^h[|\alpha_f|] &\ll (\log q)^C \quad \text{for some absolute constant } C>0.
\end{aligned}$$

Hence the proposition, the last equality being an immediate corollary of the formula

$$A[\alpha_f] = \frac{\dim J_0(q)}{\zeta(2)} A^h[L(\mathrm{Sym}^2 f, 1)\alpha_f] + O((\log q)^3 A^h[|\alpha_f|])$$

and the decomposition

$$L(\mathrm{Sym}^2 f, 1) = \omega_f(x) + \omega_f(x, y) + O(q^2 y^{-1})$$

applied with $y = q^3$. □

When the average $A^h[\alpha_f]$ is of smaller order of magnitude (being part of an error term), it is often possible to use the known individual upper bounds on ω_f to show that the corresponding $A[\alpha_f]$ is also small enough. Such a result is contained in the following lemma. It will not be used in the sequel, but is mentioned here for completeness.

Lemma 8. *Let (α_f) be complex numbers (defined for $f \in S_2(q)^*$ for all q prime) satisfying*

$$A^h[|\alpha_f|] \ll (\log q)^{-(3+\delta)}$$

for some $\delta > 0$. Then

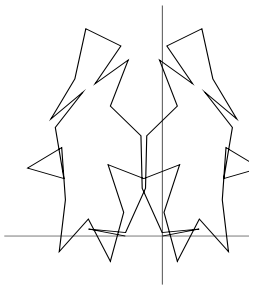
$$A[|\alpha_f|] \ll q(\log q)^{-\delta}$$

(for all q prime) with an implied constant only depending on the one in the assumption.

Proof. Using (3.19), it is enough to quote again

$$L(\mathrm{Sym}^2 f, 1) \ll (\log q)^3$$

from Lemma 2. □



Chapter 4
The Delta symbol for automorphic forms of fixed level

This short chapter studies the fundamental properties of the Delta symbol of primitive forms of prime level in various circumstances, each of which will be used in the course of the last two chapters. For a related treatment of the Delta symbol in weight aspect, see Section 5.5 of [Iw2].

4.1 The Delta symbol for primitive forms

For q prime we define the Delta symbol $\Delta(m,n)$ associated to primitive weight 2 forms of level q by

$$\Delta(m,n) = \sum_{f\in S_2(q)^*}^{h} \lambda_f(m)\lambda_f(n). \tag{4.1}$$

We first state, and sketch of proof of, the Petersson formula.

Theorem 13. *(Petersson). The Delta symbol $\Delta(m,n)$ is given by*

$$\Delta(m,n) = \delta(m,n) - \mathcal{J}(m,n)$$

where

$$\mathcal{J}(m,n) = 2\pi \sum_{c\equiv 0 \bmod q} \frac{1}{c} S(m,n;c) J_1\Big(\frac{4\pi\sqrt{mn}}{c}\Big) \tag{4.2}$$

and $S(m,n;c)$ is a Kloosterman sum, J_1 is a Bessel function of the first kind.

Sketch of the proof.

In abstract principle, this is easily understood. Let $a_f(n)$ denote the Fourier coefficients of a cusp form $f \in S_2(q)$. The map

$$f \mapsto n^{-1/2} a_f(n)$$

is a linear form on the Hilbert space $S_2(q)$ (in the basis $S_2(q)^*$, it is simply $f \mapsto \lambda_f(n)$, see (2.1)), and the latter is finite dimensional, hence by the Riesz representation theorem there exists a unique cusp form $p_n \in S_2(q)$ such that

$$(f, p_n) = n^{-1/2} a_f(n) \tag{4.3}$$

for every $f \in S_2(q)$. We can decompose p_n spectrally in terms of the basis $S_2(q)^*$, so

$$p_n = \sum_{f\in S_2(q)^*}^{h} (p_n, f) f = \sum_{f\in S_2(q)^*}^{h} \lambda_f(n) f$$

since f has real coefficients. Taking the m-th Fourier coefficient of both sides (precisely, computing the scalar product with p_m) yields

$$(p_n, p_m) = \sum_{f \in S_2(q)^*}^{h} \lambda_f(n)\lambda_f(m) = \Delta(m,n).$$

This is the abstract form of the Petersson formula. To give it a concrete shape, it is necessary to identify the forms p_n and compute their mutual inner products (p_m, p_n): this is done by a simple computation of the inner product of a form f with the Poincaré series P_n (1.14) which shows that a suitable multiple of P_n has the required property (4.3). For the details, see [Iw2], Sections 3.2 and 3.3.

□

The first lemma gives the basic estimate on the complementary term $\mathcal{J}(m,n)$, testing in effect the range of m and n for which the forms f are really "independent of each other".

Lemma 9. *For any $m \geqslant 1$, $n \geqslant 1$, we have*

$$\mathcal{J}(m,n) \ll (m,n,q)(mn)^{1/2}(\log(m,n))^2 q^{-3/2} \tag{4.4}$$

with an absolute implied constant, hence for m and n less than q,

$$\Delta(m,n) = \delta(m,n) + O((\log q)^2 q^{-1/2}).$$

Proof. We use Weil's bound for Kloosterman sums ([Iw2, 4.3] for example)

$$S(m,n;c) \leqslant \tau(c)(m,n,c)^{1/2}\sqrt{c} \tag{4.5}$$

and the easy bound (from the series expansion; we do not need any more precise information such as $J_1(x) \ll x^{-1/2}$ for x large; small values of r are not very important here)

$$J_1(x) \ll x$$

valid for all $x \geqslant 0$ with an absolute implied constant. Directly from this we have

$$\begin{aligned}\mathcal{J}(m,n) &\ll \frac{\sqrt{mn}}{q^{3/2}} \sum_{r \geqslant 1} (m,n,qr)^{1/2} \tau(qr) \frac{1}{r^{3/2}} \\ &\ll (m,n,q)^{1/2}(mn)^{1/2}(\log(m,n))^2 q^{-3/2}\end{aligned}$$

using $(m,n,qr) \mid (m,n,q)(m,n,r)$ and a crude estimate

$$\sum_{r \geqslant 1} \frac{\tau(r)(m,n,r)^{1/2}}{r^{3/2}} \ll \sum_{d \mid (m,n)} \frac{\tau(d)\varphi(d)^{1/2}}{d^{3/2}} \ll (\log(m,n))^2$$

which is sufficient to get the equally crude factor $(\log(m,n))^2$. □

This factor $(\log(m,n))^2$ would not appear if the weight were greater than 2, as the Bessel function J_{k-1} satisfies

$$J_{k-1}(x) \ll x^{k-1}$$

in general, which would make the series over r converge absolutely. However, (4.4) will only be applied in situations where we seek (and obtain) a power saving in q with m and n powers of q, so it is only an unimportant annoyance.

4.2 The Delta symbol for odd primitive forms

Keeping q prime, we now consider the Delta symbol associated to primitive forms $f \in S_2(q)^*$ with a given parity. We define Δ_+ and Δ_- by

$$\Delta_\pm(m,n) = 2 \sum_{\substack{f\in S_2(q)^* \\ \varepsilon_f=\pm 1}}^{h} \lambda_f(m)\lambda_f(n) \tag{4.6}$$

so $\Delta(m,n) = \frac{1}{2}(\Delta_+(m,n) + \Delta_-(m,n))$. From (2.7) we see that

$$\begin{aligned}
\Delta_\pm(m,n) &= 2 \sum_{f\in S_2(q)^*}^{h} \varepsilon_f^{\pm} \lambda_f(m)\lambda_f(n) \\
&= \sum_{f\in S_2(q)^*}^{h} (1 \pm \varepsilon_f)\lambda_f(m)\lambda_f(n) \\
&= \Delta(m,n) \pm q^{1/2} \sum_{f\in S_2(q)^*}^{h} \lambda_f(q)\lambda_f(m)\lambda_f(n) \qquad \text{(by (2.6))} \\
&= \Delta(m,n) \pm q^{1/2}\Delta(m,nq) \qquad (4.7)
\end{aligned}$$

(by (2.11)), which shows how the restriction of the summation affects the analytic properties of $\Delta_\pm$ through the insertion of a factor q in the sums. It is not possible to get a good bound for $\Delta_\pm$ by applying Lemma 9 to this decomposition.

Lemma 10. *Let $m < q$, $n \geqslant 1$ and define*

$$\mathcal{J}'(m,n) = -\frac{2\pi}{\sqrt{q}} \sum_{(r,q)=1} \frac{1}{r} S(m\overline{q}, n; r) J_1\Big(\frac{4\pi}{r}\sqrt{\frac{mn}{q}}\Big). \tag{4.8}$$

Then

$$q^{1/2}\mathcal{J}(m,qn) = \mathcal{J}'(m,n) + O\Big(\frac{\sqrt{mn}}{q^2}(\log q)^2\Big)$$

with an absolute implied constant.

Proof. We write

$$\mathcal{J}(m,qn) = \frac{2\pi}{q} \sum_{r\geqslant 1} \frac{1}{r} S(m, qn; qr) J_1\Big(\frac{4\pi}{r}\frac{\sqrt{mn}}{q}\Big)$$

and we estimate first the contribution of those r with $(r,q) > 1$ (so $q \mid r$ since q is prime), by the same method as in the previous lemma:

$$\frac{2\pi}{q} \sum_{(r,q)>1} \frac{1}{r} S(m, qn; qr) J_1\Big(\frac{4\pi}{r}\frac{\sqrt{mn}}{q}\Big) \ll \frac{\sqrt{mn}}{q^{5/2}}(\log q)^2$$

(using here that $m < q$ hence $(m,q) = 1$). The part of the sum which remains, with $(r,q) = 1$, is equal to

$$\frac{2\pi}{q} \sum_{(r,q)=1} \frac{1}{r} S(m, qn; qr) J_1\Big(\frac{4\pi}{r}\frac{\sqrt{mn}}{q}\Big)$$

but by the twisted multiplicativity of Kloosterman sums ([Iw2, 4.3]) we have, for r coprime with q

$$S(m, qn; qr) = S(m\overline{q}, n; r)S(m\overline{r}, 0; q) = -S(m\overline{q}, r; r)$$

since, again using $(m, q) = 1$, $S(m\overline{r}, 0; q) = S(1, 0; q)$ is a Ramanujan sum with q prime, hence $S(1, 0; q) = \mu(q) = -1$. This gives the desired expression on multiplying by $q^{1/2}$. □

We derive from this another basic decomposition and estimate.

Lemma 11. *Let $m < q$ and $n \geqslant 1$. Then*

$$\Delta_{\pm}(m, n) = \delta(m, n) \mp \mathcal{J}'(m, n) + O\Big(\frac{\sqrt{mn}}{q^{3/2}}(\log q)^2\Big) \tag{4.9}$$

$$= \delta(m, n) + O\Big(\frac{\sqrt{mn}}{q}(\log q)^2\Big). \tag{4.10}$$

Proof. From the Petersson formula and (4.7) we obtain

$$\Delta_{\pm}(m, n) = \delta(m, n) \pm q^{1/2}\delta(m, qn) \ - \mathcal{J}(m, n) \mp q^{1/2}\mathcal{J}(m, qn).$$

Since $m < q$, this gives from Lemma 9 and 10

$$\Delta_{\pm}(m, n) = \delta(m, n) \mp \mathcal{J}'(m, n) + O\Big(\frac{\sqrt{mn}}{q^{3/2}}(\log q)^2\Big).$$

Yet again we apply the same method as in Lemma 9 to estimate

$$\mathcal{J}'(m, n) \ll \frac{\sqrt{mn}}{q}(\log q)^2$$

and therefore the second bound follows. □

In Chapter 5, only $\Delta(m, n)$ appears, in ranges such that Lemma 9 is adequate, but in treating the second moment of mollified derivatives in Chapter 6, the corresponding (4.10) is not, and a precise analysis of the contribution arising from $\mathcal{J}'(m, n)$ will be required.

4.3 The Delta-symbol without weight

When applying the technique of Chapter 3 to go from an average $A^h[\alpha_f]$ weighted by ω_f to the natural average $A[\alpha_f]$, we will use variants of the Delta-symbol, twisted by symmetric-square coefficients.

Let $x < \sqrt{q}$ be given. Then we define

$$\Delta^n(m, n) = \sum_{\ell \leqslant x} A^h[\rho_f(\ell)\lambda_f(m)\lambda_f(n)]$$

$$= \sum_{d\ell^2 \leqslant x} \frac{1}{d\ell^2} \sum_{f \in S_2(q)^*}^{h} \lambda_f(d^2)\lambda_f(m)\lambda_f(n) \tag{4.11}$$

(and similarly for $\Delta^n_{\pm}$).

Lemma 12. *For all m, n, we have*

$$\Delta^n(m,n) = \sum_{d\ell^2 \leqslant x} \frac{1}{d\ell^2} \sum_{r|(d^2,m)} \Delta\Big(\frac{md^2}{r^2}, n\Big)$$

(and the same holds with $\Delta^n_\pm$ in place of Δ^n), and if $m < q$, $n < q$,

$$\Delta^n(m,n) = \sum_{d\ell^2 \leqslant x} \frac{1}{d\ell^2} \sum_{r|(d^2,m)} \delta\Big(\frac{md^2}{r^2}, n\Big) + O\Big((\log q)^3 \frac{x\sqrt{mn}}{q^{3/2}}\Big) \tag{4.12}$$

$$\Delta^n_\pm(m,n) = \sum_{d\ell^2 \leqslant x} \frac{1}{d\ell^2} \sum_{r|(d^2,m)} \delta\Big(\frac{md^2}{r^2}, n\Big) + O\Big((\log q)^3 \frac{x\sqrt{mn}}{q}\Big). \tag{4.13}$$

Proof. The first statement is an immediate consequence of the definition of Δ^n and the multiplicativity formula

$$\lambda_f(d^2)\lambda_f(m) = \sum_{r|(d^2,n)} \lambda_f\Big(\frac{md^2}{r}\Big)$$

and the second is a consequence of this, and Lemma 9 for Δ^n, or Lemma 11 for $\Delta^n_\pm$. □

We see that, in first approximation, the difference is that the natural Delta symbol $\Delta^n(m,n)$ detects the condition that n and m differ by a square, instead of being equal.

Appendix: Digression on multiplicativity

The natural Delta-symbol Δ^n, when applied, will need further treatment, in particular in Chapter 6. For lack of a better place, we offer here a digression with the intent of clarifying the apparently complicated details which will be involved in the computations. The situation will be the following: we wish to diagonalize a quadratic form Q which is of the form

$$Q(y) = \sum_{m_1,m_2} \frac{g(m_1m_2)}{m_1m_2} y_{m_1} y_{m_2}$$

with some arithmetic function g, and the variables m_1, m_2 are restricted by m_1, $m_2 \leqslant M$, for some $M > 0$, and m_1 and m_2 squarefree.

If g were totally multiplicative, the form Q would be diagonalized by making the change of variable

$$z_k = \sum_m \frac{g(m)}{m} y_m$$

so that

$$Q(y) = \sum_k |y_k|^2.$$

In practice, g will not satisfy this strong condition, but will retain some multiplicative property. In greater generality, this can be defined as follows, to present a clear picture.

Definition 6. Let g be an arithmetic function. We say that g is *mutative* (for k factors) if and only[1] if the arithmetic function

$$(m_1, \dots, m_k) \mapsto g(m_1 \cdots m_k)$$

defined for all integers $m_i \geqslant 1$ coprime in pairs is a finite sum of product functions

$$(m_1, \dots, m_k) \mapsto h_1(m_1) \cdots h_k(m_k).$$

If this is so for every $k \geqslant 1$, then g is simply called mutative. If $D \geqslant 1$ is the least integer D such that one can write

$$g(m_1 \cdots m_k) = \sum_{1 \leqslant i \leqslant D} h_{1,i}(m_1) \cdots h_{k,i}(m_k)$$

(for m_i coprime in pairs) then g is called D-mutative (for k factors).

In practice, the functions in the decomposition are explicitly known for 2-factors, and are themselves 2-mutative, so g is automatically mutative for any number of factors.

Here are a few (obvious) examples and properties properties of this fancy notion:

1. Every multiplicative function is clearly 1-mutative. Actually, every 1-mutative function g is the product of a constant (the value of g at 1) and a multiplicative function.

2. Sums and products of mutative functions are again mutative, and the same applies to Dirichlet convolutions.

3. Additive functions are 2-mutative for 2 factors:

   $$g(m_1 m_2) = g(m_1) + g(m_2)$$

   and take $h_{1,1} = h_{2,2} = g$, $h_{1,2} = h_{2,1} = 1$. Hence they are mutative for any number of factors. In particular, $n \mapsto (\log Q/n)^k$ is mutative for any $k \geqslant 1$ and any $Q > 0$.

4. As a consequence of the two previous remarks, mutative functions arise naturally as coefficients of derivatives of Dirichlet series with multiplicative coefficients.

5. The function $n \mapsto (n-1)^{-1}$ is not mutative for 2 factors: indeed, if it were, then the vector space generated by the rational functions $(aX - 1)^{-1}$ for all positive integers a would be of finite dimension.

Now consider a quadratic form Q as above, with g mutative (for 3 factors would suffice). Then we "diagonalize"[2] Q by introducing $a = (m_1, m_2)$, so $m_1 = an_1$, $m_2 = an_2$ for some integers n_1 and n_2. Since m_1 and m_2 are squarefree, we have $(a, n_1) =$

[1] Only by the greatest force of will does the author manage to refrain giving a French-style definition in terms of tensor products.

[2] This method will only produce a true diagonalization in special cases, such as g multiplicative, but for convenience the name is retained.

$(a, n_2) = 1$, hence by the mutativity of g, Q is a finite sum of quadratic forms of the type

$$\sum_a \frac{h_1(a^2)}{a^2} \sum_{(n_1,n_2)=1} \frac{h_2(n_1)h_3(n_2)}{n_1 n_2} y_{an_1} y_{an_2}$$

and h_2 and h_3 are still 2-mutative.

Then we remove the condition $(n_1, n_2) = 1$ by Möbius inversion, exploiting again the fact that n_1 and n_2 are squarefree and the mutativity to write Q as a sum of quadratic forms of the type

$$\sum_a \frac{h_1(a^2)}{a^2} \sum_d \frac{\mu(d)h_2(d)^2}{d^2} \sum_{n_1,n_2} \frac{h_3(n_1)h_4(n_2)}{n_1 n_2} y_{adn_1} y_{adn_2}$$

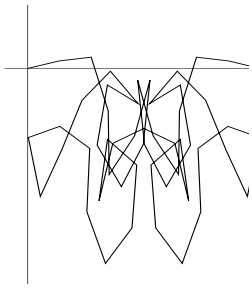
(while h_1 is the same as before, h_2, h_3 might be different), which are "diagonalized" as

$$\sum_k \nu(k) w_k z_k \tag{4.14}$$

with

$$\begin{aligned} \nu(k) &= \frac{1}{k^2} \sum_{ad=k} \mu(d)h_2(d)^2 h_1(a^2), \\ w_k &= \sum_n \frac{h_3(n)}{n} y_{kn}, \\ z_k &= \sum_n \frac{h_4(n)}{n} y_{kn}. \end{aligned}$$

In practice, the function g is D-mutative for small D (often $D = 1$, or $D = 2$), and there are not many terms involved, but it is best to get a feeling for the general process. The ultimate goal will be to estimate the value of Q for a specific choice of (y_m), and most of the terms can be seen very quickly to be less than what the main term is expected to be, so it is not even necessary to write the full decomposition explicitly.



Chapter 5
Proof of the upper bound

A monk asked Joshu: "Does a dog
have the nature of Buddha, or not?"
Joshu answered: "μ."
Mumon, *"The Gateless Gate"*

We recall the statement to be proved: there exists an absolute and effective constant $C > 0$ such that for any prime number q

$$\sum_{f \in S_2(q)^*} \mathrm{ord}_{s=\frac{1}{2}} L(f,s) \leqslant C \dim J_0(q). \tag{5.1}$$

In this chapter, q is always a fixed prime number.

5.1 The explicit formula: reduction to a density theorem

The explicit formulae, discovered in essence by Riemann, and later extended and formalized by Weil, have been used first by Mestre in studying abelian varieties. We choose the following variant.

Proposition 9. *Let $\psi :]0, +\infty[\longrightarrow \mathbf{R}$ be a C^∞ function with compact support, satisfying $\psi(x) = \psi(x^{-1})$ for all x, and $\hat{\psi}$ its Mellin transform, which is an entire function. Then for any primitive form $f \in S_2(q)^*$*

$$\sum_\rho \hat{\psi}(\rho - \tfrac{1}{2}) = \psi(1)\log q - 2\sum_{n \geqslant 1} \frac{b_f(n)}{\sqrt{n}} \Lambda(n)\psi(n) + \frac{1}{2i\pi} \int_{(1/2)} 2\Big(\frac{\Gamma'}{\Gamma}(s+\tfrac{1}{2}) - \log 2\pi\Big)\hat{\psi}(s-\tfrac{1}{2})ds \tag{5.2}$$

the summation on the left-hand side being extended over all zeros ρ of $L(f,s)$ in the critical strip – those with $0 \leqslant \mathrm{Re}\,(s) \leqslant 1$ – counted with multiplicity. The coefficients $b_f(n)$ are defined in (2.15).

Proof. We only give a sketch, as this is well-known (see [KM1], [Br1], [P-P]...) in one form or another.

We define $\varphi(x) = x^{-1/2}\psi(x)$ so that $\hat{\varphi}(s) = \hat{\psi}(s - \frac{1}{2})$ for all $s \in \mathbf{C}$. The condition $\psi(x) = \psi(x^{-1})$ means $\varphi(x) = x^{-1}\varphi(x^{-1})$, or $\hat{\varphi}(s) = \hat{\varphi}(1-s)$.

The logarithmic derivative of the functional equation

$$\Big(\frac{\sqrt{q}}{2\pi}\Big)^s \Gamma(s+\tfrac{1}{2})L(f,s) = \varepsilon_f \Big(\frac{\sqrt{q}}{2\pi}\Big)^{1-s} \Gamma(\tfrac{3}{2}-s)L(f,1-s)$$

gives the identity

$$-\frac{L'}{L}(f,s)-\frac{L'}{L}(f,1-s)=\log q+\frac{\Gamma'}{\Gamma}(s+\tfrac{1}{2})+\frac{\Gamma'}{\Gamma}(\tfrac{3}{2}-s)-2\log 2\pi.$$

Multiplying by $\hat{\varphi}(s)$, which is rapidly decreasing in vertical strips, and integrating, we obtain (using the symmetry $\hat{\varphi}(s)=\hat{\varphi}(1-s)$)

$$\begin{aligned}&\frac{1}{2i\pi}\int\limits_{(\sigma)}-\frac{L'}{L}(f,s)\hat{\varphi}(s)ds+\frac{1}{2i\pi}\int\limits_{(\sigma)}-\frac{L'}{L}(f,1-s)\hat{\varphi}(s)ds\\&=\varphi(1)\log q+\frac{1}{2i\pi}\int\limits_{(1/2)}2\Big(\frac{\Gamma'}{\Gamma}(s+\tfrac{1}{2})-\log 2\pi\Big)\hat{\varphi}(s)ds\end{aligned}$$

for any $\sigma\in\mathbf{R}$ (provided $L(f,s)$ doesn't vanish on $\mathrm{Re}\,(s)=\sigma$). We take $\sigma=-\frac{1}{4}$, then $L(f,s)$ has no trivial zeros with $\mathrm{Re}\,(s)\geqslant\sigma$ by the functional equation and the Euler product. In the first integral, shifting the contour to $\mathrm{Re}\,(s)=\frac{5}{4}$ gives

$$\frac{1}{2i\pi}\int\limits_{(\sigma)}-\frac{L'}{L}(f,s)\hat{\varphi}(s)ds=\sum_{\rho}\hat{\varphi}(\rho)+\frac{1}{2i\pi}\int\limits_{(1-\sigma)}-\frac{L'}{L}(f,s)\hat{\varphi}(s)ds$$

and the latter integral is the same as the second one, and from the Dirichlet series expansion (2.14) of the logarithmic derivative of $L(f,s)$, each is equal to

$$\frac{1}{2i\pi}\int\limits_{(1-\sigma)}-\frac{L'}{L}(f,s)\hat{\varphi}(s)ds=\sum_{n\geqslant 1}b_f(n)\Lambda(n)\varphi(n).$$

Putting all together, the proposition is proved. □

In this chapter, ρ will always designate such a "non-trivial" zero of $L(f,s)$, and we always write

$$\rho=\beta+i\gamma$$

so $\gamma=\mathrm{Im}\,(\rho)$, $\beta=\mathrm{Re}\,(\rho)$. For any α with $0\leqslant\alpha\leqslant 1$, and any real numbers $t_1\leqslant t_2$, we define $N(f;\alpha,t_1,t_2)$ to be the number of zeros $\rho=\beta+i\gamma$ of $L(f,s)$, counted with multiplicity, such that

$$\beta\geqslant\alpha,\ t_1\leqslant\gamma\leqslant t_2,$$

and for any $T>0$ we let

$$N(f;\alpha,T)=N(f;\alpha,-T,T),\quad N(f,T)=N(f;0,T).$$

It is a basic fact of the standard theory of L-functions that $N(f,T)$ has an asymptotic

$$N(f,T)=T(\log\frac{qT}{2\pi e})+O(\log qT).\tag{5.3}$$

We need a test function ψ with certain good properties, which we now describe as we assert its existence.

Proposition 10. *There exists a C^∞ function $\psi :]0,+\infty[\longrightarrow \mathbf{R}^+$ which satisfies:*

(1) It has compact support in $[e^{-1}, e]$ and $\psi(1) = 1$, $\hat{\psi}(0) > 0$.

(2) For all $x \in]0,+\infty[$,

$$\psi(x) = \psi(x^{-1}).$$

(3) For all $s \in \mathbf{C}$ with $|\mathrm{Re}\,(s)| \leqslant 1$,

$$\mathrm{Re}\,(\hat{\psi}(s)) \geqslant 0. \tag{5.4}$$

Proof. The crucial part is of course (5.4); functions of this type were constructed by Poitou and others for the purpose of obtaining lower-bounds for the discriminant of number fields [Poi] (they use the Laplace transform instead of the Mellin transform, so ψ is not exactly as stated, but $\hat{\psi}$ is the same as their Laplace transform). We sketch the principle of such constructions: define $f(y) = \psi(e^y)$, which is an even, compactly supported, C^∞ function on $\mathbf{R}$, and

$$\hat{\psi}(s) = \int_{\mathbf{R}} f(y) e^{sy} dy$$

for all $s \in \mathbf{C}$, hence

$$\begin{aligned}\mathrm{Re}\,(\hat{\psi}(s)) &= \int_{\mathbf{R}} f(y) e^{\sigma y} \cos(ty) dy \\ &= \int_{\mathbf{R}} f(y) \cosh(\sigma y) \cos(ty) dy \quad \text{(by symmetry)}\end{aligned}$$

where $\sigma = \mathrm{Re}\,(s)$. From the maximum modulus principle for harmonic functions, the inequality $\mathrm{Re}\,(\hat{\psi}(s)) \geqslant 0$ will hold for $|\sigma| \leqslant 1$ if and only if the Fourier transform of the even function $g(y) = f(y)\cosh(y)$ is non-negative. Conversely, if a function g (even, smooth and compactly supported) with this property is given, a suitable test function ψ is easily obtained by reversing this procedure. Now if g_0 is any smooth compactly supported positive function on $\mathbf{R}$, then the convolution square $g = g_0 \star g_0$ will work, since $\hat{g} = \hat{g}_0^2$.

The support of ψ and its value at 1 can be easily adjusted by homogeneity. □

Remark In [KM1], a specific test function F is used, which had been constructed previously by Perelli and Pomykala [P-P], which is more subtle. However in our situation, it is not actually necessary to use it (this was observed by Pomykala).

Henceforth we fix a test function ψ as given by the proposition, and for any real number $\lambda > 0$ we define the function ψ_λ by

$$\psi_\lambda(x) = \psi(x^{\lambda^{-1}})$$

so that its Mellin transform is

$$\hat{\psi}_\lambda(s) = \lambda \hat{\psi}(\lambda s).$$

The parameter λ will be used to effect a localization in detecting the zeros around $\frac{1}{2}$ in the explicit formula.

Lemma 13. *For any $s \in \mathbf{C}$, and any integer $k \geqslant 1$, it holds*

$$\hat{\psi}(s) \ll_k \frac{1}{|\mathrm{Im}\,(s)|^k} e^{\mathrm{Re}\,(s)} \tag{5.5}$$

where the implied constant depends only on k (and on the specific choice of ψ).

Proof. This is quite clear by successive integrations by parts, since ψ has compact support in $[e^{-1}, e]$:

$$\hat{\psi}(s) = \frac{(-1)^k}{s(s+1)\dots(s+k)} \int_0^{+\infty} \psi(x) x^{s+k} \frac{dx}{x}$$

hence the result. □

Let q be a prime number and $f \in S_2(q)^*$ a primitive form of level q. We take $\lambda = \theta \log q$, where θ is a constant,[1] a sufficiently small real number which will be chosen later. This choice agrees with the heuristic that we should not lose the result we seek by counting zeros with imaginary part at most λ. Applying the explicit formula (5.2) to f with the test function ψ_λ, we obtain

$$\sum_\rho \hat{\psi}_\lambda(\rho - \tfrac{1}{2}) = \log q - 2 \sum_{n \geqslant 1} \frac{b_f(n)}{\sqrt{n}} \Lambda(n) \psi_\lambda(n) + O(1)$$

after having estimated the integral in (5.2) by

$$\frac{1}{2i\pi} \int_{(1/2)} 2\Big(\frac{\Gamma'}{\Gamma}(s + \tfrac{1}{2}) - \log 2\pi\Big) \hat{\psi}_\lambda(s - \tfrac{1}{2}) ds \ll \lambda \int_{-\infty}^{+\infty} (1 + |u|) \hat{\psi}(\lambda i u) du \ll 1$$

uniformly in λ (we have used

$$\frac{\Gamma'}{\Gamma}(s) \ll |s|$$

for $\mathrm{Re}\,(s) = 1$, and Lemma 13). Let $T > 1$ be a parameter which will be fixed later on. Using (5.3), we estimate first the sum over zeros with $\gamma = \mathrm{Im}\,(\rho) > T$:

$$\begin{aligned}
\sum_{\gamma > T} \hat{\psi}_\lambda(\rho - \tfrac{1}{2}) &\ll \sum_{t=[T]}^{+\infty} N(f, 0; t, t+1) \sup_{s \in [0,1] \times [t, t+1]} \hat{\psi}_\lambda(\rho - \tfrac{1}{2}) \\
&= \sum_{t=[T]}^{+\infty} \lambda N(f, 0; t, t+1) \sup_{s \in [0,1] \times [t, t+1]} \hat{\psi}(\lambda(\rho - \tfrac{1}{2})) \\
&\ll_k \sum_{t=[T]}^{+\infty} \lambda (t+1) (\log q(t+1)) (\lambda t)^{-k} e^{\lambda/2} \\
&\ll_k q^{\theta/2} (\log qT) (\lambda T)^{-(k-2)}
\end{aligned}$$

for any integer $k \geqslant 1$, by Lemma 13. The same holds, of course, for zeros with $\gamma < -T$.

Then we isolate the multiplicity of the zero at $\frac{1}{2}$, and further distinguish among the remaining zeros ρ between those which are close to $\frac{1}{2}$, precisely those with $|\beta - \frac{1}{2}| \leqslant \lambda^{-1}$, and the others. On the other side we use the fact that Λ is supported on powers of primes, and put the primes apart from the squares and higher powers. This way we rewrite the outcome of the explicit formula:

$$\begin{aligned}
\lambda \hat{\psi}(0) \mathrm{ord}_{s=\frac{1}{2}} L(f, s) + \Xi_1(f, \lambda) + \Xi_2(f, \lambda) = \log q - 2S_1(f, \lambda) - 2S_2(f, \lambda) \\
+ O_k(q^{\theta/2} (\log qT) (\lambda T)^{-k}) + O(1) \quad (5.6)
\end{aligned}$$

[1] Also, q is large enough that $\lambda > 1$.

with:

$$\Xi_1(f,\lambda) = \lambda \sum_{\substack{|\gamma|\leqslant T \\ |\beta-\frac{1}{2}|\leqslant\lambda^{-1}}} \hat{\psi}(\lambda(\rho-\tfrac{1}{2})) \tag{5.7}$$

$$\Xi_2(f,\lambda) = \lambda \sum_{\substack{|\gamma|\leqslant T \\ |\beta-\frac{1}{2}|>\lambda^{-1}}} \hat{\psi}(\lambda(\rho-\tfrac{1}{2})) \tag{5.8}$$

$$S_1(f,\lambda) = \sum_p \frac{\lambda_f(p)}{\sqrt{p}}(\log p)\psi_\lambda(p) \tag{5.9}$$

$$S_1(f,\lambda) = \sum_{n\geqslant 2}\sum_p \frac{\lambda_f(p^n)}{p^{n/2}}(\log p)\psi_\lambda(p^n). \tag{5.10}$$

The various terms will be treated differently.

Lemma 14. *For all $f \in S_2(q)^*$, we have*

$$S_2(f,\lambda) \ll \lambda.$$

Proof. Since ψ has compact support in $[e^{-1}, e]$, the sum over n-th powers of primes is void as soon as $p^{n/\lambda} > e$, namely as soon as $n\log p > \lambda$. Since ψ is bounded and

$$|b_f(n)| \leqslant 2$$

for all n, this yields

$$\begin{aligned} S_2(f,\lambda) &\ll \sum_{p\leqslant\exp(\lambda/2)} \frac{\log p}{p} + \sum_{3\leqslant n\leqslant\lambda}\sum_{\log p\leqslant\lambda/n} \frac{\log p}{p^{n/2}} \\ &\ll \lambda. \end{aligned}$$

□

Now, in (5.6), we take the real part. For a zero ρ appearing in $\Xi_1(f,\lambda)$, we have $|\operatorname{Re}\lambda(\rho-\frac{1}{2})| \leqslant 1$, hence

$$\Xi_1(f,\lambda) \geqslant 0$$

by the positivity property (5.4) of the test function ψ. Therefore we can drop this term by positivity and get

$$\begin{aligned} \lambda\hat{\psi}(0)\operatorname{ord}_{s=\frac{1}{2}}L(f,s) \leqslant \log q - 2S_1(f,\lambda) + \operatorname{Re}\left(\Xi_2(f,\lambda)\right) \\ + O(\lambda) + O_k(q^{\theta/2}(\log qT)(\lambda T)^{-k}). \end{aligned}$$

Again, intuitively, this application of positivity should not affect the chances of proving the result being sought, since the number of zeros dropped in the sum $\Xi_1(f,\lambda)$, on average over f, should be bounded.

Performing the average over f, we have consequently

$$\begin{aligned} \lambda\hat{\psi}(0)\sum_{f\in S_2(q)^*}\operatorname{ord}_{s=\frac{1}{2}}L(f,s) \leqslant (\log q)\dim J_0(q) - 2\sum_{f\in S_2(q)^*} S_1(f,\lambda) \\ + \sum_{f\in S_2(q)^*}\operatorname{Re}\left(\Xi_2(f,\lambda)\right) \\ + O(\lambda q) + O_k(q^{1+\theta/2}(\log qT)(\lambda T)^{-k}). \end{aligned} \tag{5.11}$$

Lemma 15. *Assume $\theta < \frac{3}{4}$. There exists a constant $\delta = \delta(\theta) > 0$ such that*

$$\sum_{f \in S_2(q)^*} S_1(f,\lambda) \ll q^{1-\delta}. \tag{5.12}$$

Proof. We write

$$\sum_{f \in S_2(q)^*} S_1(f,\lambda) = A[S_1(f,\lambda)]$$

and proceed to estimate this by the method of Chapter 3. We need to check the conditions (3.17) to apply Proposition 8. The individual bound (3.18) is easy: since

$$S_1(f,\lambda) = \sum_p \frac{\lambda_f(p)}{\sqrt{p}} \psi_\lambda(p)$$

and the support of ψ limits the summation to primes with $\log p \leqslant \lambda$, i.e. $p \leqslant q^\theta$, it holds

$$S_1(f,\lambda) \ll \sum_{p \leqslant q^\theta} p^{-1/2} \ll q^{\theta/2}$$

while $\omega_f \ll (\log q) q^{-1}$ by (3.16).

We next estimate the harmonic average $A^h[|S_1(f,\lambda)|]$. Since the sum $S_1(f,\lambda)$ is real, Cauchy's inequality implies

$$A^h[|S_1(f,\lambda)|] \leqslant A^h[S_1(f,\lambda)^2]^{1/2} A^h[1]^{1/2} \ll A^h[S_1(f,\lambda)^2]^{1/2}.$$

Now we compute

$$\begin{aligned} S_1(f,\lambda)^2 &= \sum_{p,p'} \frac{\lambda_f(p)\lambda_f(p')}{\sqrt{pp'}} (\log p)(\log p') \psi_\lambda(p)\psi_\lambda(p') \\ &= \sum_p \frac{\lambda_f(p)^2}{p} (\log p)^2 \psi_\lambda(p)^2 + \sum_{p \neq p'} \frac{\lambda_f(pp')}{\sqrt{pp'}} (\log p)(\log p') \psi_\lambda(p)\psi_\lambda(p') \end{aligned}$$

so (using $\lambda_f(p)^2 = \lambda_f(p^2) + 1$, which is true for all primes p occurring in the summation since $\psi_\lambda(p) = 0$ for $\log q > \lambda$, i.e. for $p > q^\theta$) we get

$$\begin{aligned} A^h[S_1(f,\lambda)^2] = &\sum_p \frac{(\log p)^2}{p} \psi_\lambda(p)^2 (\Delta(1,1) + \Delta(1,p^2)) \\ &+ \sum_{p \neq p'} \frac{(\log p)(\log p')}{\sqrt{pp'}} \psi_\lambda(p)\psi_\lambda(p') \Delta(1,pp'). \end{aligned}$$

Since evidently $pp' \neq 1$, we obtain from Lemma 9

$$A^h[S_1(f,\lambda)^2] \ll \sum_{p \leqslant q^\theta} \frac{(\log p)^2}{p} + \frac{(\log q)^2}{q^{3/2}} \Big| \sum_{p \leqslant q^\theta} (\log p) \Big|^2 \ll (\log q)^2$$

since $2\theta < \frac{3}{2}$.

From Proposition 8 of Chapter 3, we conclude that there exists a constant $\delta = \delta(\theta)$ such that

$$A[S_1(f,\lambda)] = \frac{\dim J_0(q)}{\zeta(2)} A^h[\omega_f(x) S_1(f,\lambda)] + O(q^{1-\delta})$$

with $x = q^\kappa$, $\kappa < \frac{1}{4}$ being a (small) parameter to be chosen below.

From the definition of $\omega_f(x)$, we derive

$$\begin{aligned} A^h[\omega_f(x)S_1(f,\lambda)] &= \sum_{d\ell^2\leqslant x} \frac{\lambda_f(d^2)}{d\ell^2} \sum_p \frac{\lambda_f(p)}{\sqrt{p}}(\log p)\psi_\lambda(p) \\ &= \sum_{d\ell^2\leqslant x} \frac{1}{d\ell^2} \sum_p \frac{(\log p)}{\sqrt{p}}\psi_\lambda(p)\Delta(p,d^2) \\ &\ll \frac{(\log q)^2}{q^{3/2}} \sum_{d\ell^2\leqslant x} \frac{1}{\ell^2} \sum_{p\leqslant q^\theta} (\log p) \\ &\ll (\log q)^2 q^{\theta+\kappa-\frac{3}{2}} \end{aligned}$$

and the lemma follows by taking κ small enough that the exponent is negative. □

Remark In [KM1], the analogue of this Lemma is quoted from [Br1], where it was proved by means of the Selberg trace formula. The present approach is probably somewhat simpler, and at least more self-contained.

Thus it only remains to estimate the contribution of Ξ_2, the sum over zeros not too close to $\frac{1}{2}$. Of course, on the Generalized Riemann Hypothesis, those do not exist, and we see, taking $T = q$ and then k large enough, that the upper bound above (5.11) immediately implies a weak form of Brumer's result, namely

$$\sum_{f\in S_2(q)^*} \operatorname{ord}_{s=\frac{1}{2}} L(f,s) \ll \dim J_0(q)$$

for q prime. Indeed, up to this point, the treatment is basically the same as Brumer's. But handling Ξ_2 without appealing to the Riemann Hypothesis is precisely the crux of the matter. It will be possible to show that if there are zeros in the region $|\beta-\frac{1}{2}| > \lambda^{-1}$, then they are very few in number, in a very precise sense, which we now describe.

Theorem 14. *Let q be a prime number. There exists an absolute constant $A > 0$ such that for any $T \geqslant 0$ and any real numbers t_1, t_2 with*

$$-T \leqslant t_1 < t_2 \leqslant T$$

$$t_2 - t_1 \geqslant \frac{1}{\log q},$$

for any $\alpha \geqslant \frac{1}{2} + (\log q)^{-1}$ and any c, $0 < c < \frac{1}{4}$, it holds

$$\sum_{f\in S_2(q)^*} N(f;\alpha,t_1,t_2) \ll (1+T)^A q^{1-c(\alpha-\frac{1}{2})}(\log q)(t_2-t_1), \tag{5.13}$$

the implied constant depending only on c.

The bulk of this chapter will be devoted to proving this result.

Remark In this density theorem, only the q-aspect is taken into consideration, and this statement is indeed trivial with respect to T. However, it is important (as the deduction of the upper bound from the density theorem shows) that the bounds

obtained be at most polynomial in the imaginary part T. Thus, in the rest of this chapter, inequalities of the form

$$f(q,T) \ll (1+|T|)^B g(q)$$

will often be encountered; the constant $B \geqslant 0$ may appear, or its value may change, from line to line without further comment.

Assuming Theorem 14, we can now estimate Ξ_2. We argue for each quadrant separately. Subdividing the region $[\lambda^{-1}, \frac{1}{2}] \times [0,T]$ into small squares of side λ^{-1}

$$R(m,n) = \Big[\frac{m}{\lambda}, \frac{m+1}{\lambda}\Big] \times \Big[\frac{n}{\lambda}, \frac{n+1}{\lambda}\Big]$$

with $1 \leqslant m \leqslant \lambda$, $0 \leqslant n \leqslant \lambda T$, we estimate the contribution Ξ_2^1 of those zeros:

$$\begin{aligned}
\sum_{f \in S_2(q)^*} \mathrm{Re}\,(\Xi_2^1(f,\lambda)) &\leqslant \lambda \sum_{m=1}^{\lambda} \sum_{n=0}^{\lambda T} N(f; \tfrac{1}{2} + \tfrac{n}{\lambda}, \tfrac{m}{\lambda}, \tfrac{m+1}{\lambda}) \sup_{s \in R(m,n)} |\hat{\psi}(\lambda s)| \\
&\ll \lambda \sum_{m=1}^{\lambda} \sum_{n=0}^{\lambda T} (1 + \tfrac{n+1}{\lambda})^A q^{1-c\frac{n}{\lambda}} (\log q) \lambda^{-1} \times \sup_{s \in R(m,n)} |\hat{\psi}(\lambda s)| \\
&\ll_k q(\log q) \sum_{m=1}^{\lambda} \sum_{n=1}^{\lambda T} (1 + \tfrac{n+1}{\lambda})^A q^{-cm/\lambda} e^{m+1} n^{-k} \\
&\quad + q(\log q) \sum_{m=1}^{\lambda} q^{-cm/\lambda} e^{m+1} \\
&\ll q(\log q)
\end{aligned}$$

if we choose $\theta < c$, and k large enough, so that the sum over m is a convergent geometric series of the form

$$\sum_m q^{-cm/\lambda} e^m = \sum_m \exp(m(1 - c\theta^{-1})) \leqslant \frac{1}{1 - \exp(1 - c\theta^{-1})}.$$

We can deal by similar dissections with the three other quadrants in Ξ_2, hence from (5.11), taking $T = q$, and dividing out by λ, we deduce the desired inequality (5.1), hence Theorem 8.

5.2 The density theorem

Theorem 14 is the analogue of a result proved by Selberg ([Sel], Theorem 4) for Dirichlet characters in 1946, itself the q-analogue of one of his previous results on the zeros of $\zeta(s)$ near the critical line. We will borrow the general principle from this paper (with some simplifications also found in [Luo]), starting with a crucial lemma which will reduce the theorem to some estimates of a mollified second moment of values of $L(f,s)$, $f \in S_2(q)^*$.

Lemma 16. *(Selberg, [Sel, Lemma 14]). Let h be a function holomorphic in the region*

$$\{s \in \mathbf{C} \mid \mathrm{Re}\,(s) \geqslant \alpha,\ t_1 \leqslant \mathrm{Im}\,(s) \leqslant t_2\}$$

satisfying

$$h(s) = 1 + o\Big(\exp(-\frac{\pi}{t_2 - t_1}\mathrm{Re}\,(s)\Big) \tag{5.14}$$

in this region, uniformly as $\mathrm{Re}\,(s) \to +\infty$. *Denoting the zeros of* f *(in the interior of this region) by* $\rho = \beta + i\gamma$, *we have*

$$2(t_2 - t_1)\sum_{\rho} \sin\Big(\pi\frac{\gamma - t_1}{t_2 - t_1}\Big)\sinh\Big(\pi\frac{\beta - \alpha}{t_2 - t_1}\Big) = \int_{t_1}^{t_2} \sin\Big(\pi\frac{t - t_1}{t_2 - t_1}\Big)\log|h(\alpha + it)|dt$$
$$+ \int_{\alpha}^{+\infty} \sinh\Big(\pi\frac{\sigma - \alpha}{t_2 - t_1}\Big)(\log|h(\sigma + it_1)| + \log|h(\sigma + it_2)|)d\sigma$$

(where the zeros are also summed with multiplicity).

We refer to [Sel] for the proof, a rather clever exercise in complex integration.

This lemma will be applied to the functions $1 - (M(f,s)L(f,s) - 1)^2$, where $M(f,s)$ is a suitable mollifier for which (5.14) holds, for α equal to $\frac{1}{2} + (\log q)^{-1}$. This means that $M(f,s)$ must approximate quite closely the inverse of $L(f,s)$.

Lemma 17. *The inverse* $L(f,s)^{-1}$ *is given by the Dirichlet series*

$$L(f,s)^{-1} = \sum_{m,n \geqslant 1} \varepsilon_q(n)\mu(m)\mu(mn)^2\lambda_f(m)(mn^2)^{-s}$$

which is absolutely convergent for $\mathrm{Re}\,(s) > 1$.

Proof. This is an immediate consequence of the Euler product expansion

$$L(f,s)^{-1} = \prod_p (1 - \lambda_f(p)p^{-s} + \varepsilon_q(p)p^{-2s})$$

by multiplicativity (every integer $\ell \geqslant 1$ has a unique expression as $\ell = mn^2r$ with m, n, r coprime in pairs, m and n squarefree and r cubefull). □

We also define, for every $M \geqslant 1$, a function g_M by

$$g_M(x) = \begin{cases} 1, & \text{if } x \leqslant \sqrt{M} \\ \dfrac{\log M/x}{\log\sqrt{M}}, & \text{if } \sqrt{M} \leqslant x \leqslant M \\ 0, & \text{if } x > M. \end{cases} \tag{5.15}$$

Then for M fixed and any integer $1 \leqslant m \leqslant M$, we let

$$x_m(s) = \frac{\mu(m)}{m^{s-\frac{1}{2}}}\sum_{n \geqslant 1} \frac{\varepsilon_q(n)\mu(mn)^2}{n^{2s}} g_M(mn) \tag{5.16}$$

and we define the mollifier

$$M(f,s) = \sum_{m \leqslant M} \frac{x_m(s)}{\sqrt{m}}\lambda_f(m) \tag{5.17}$$

(compare (6.4)). We observe that $M(f,s)$ is a Dirichlet polynomial of length at most M, with coefficients

$$c_f(\ell) = \sum_{mn^2=\ell} \varepsilon_q(n)\mu(m)\mu(mn)^2\lambda_f(m)g_M(mn) \tag{5.18}$$

and by Deligne's bound, they are bounded by

$$|c_f(\ell)| \leqslant \sum_{m|\ell} \tau(m) \leqslant \tau(\ell)^2. \tag{5.19}$$

As in the next chapter, the length M will be a power of q (here any positive power would suffice; this would also be the case in Chapter 6, if we only wanted to prove the lower bound on the rank with some constant in place of the more precise $\frac{19}{54}$).

Lemma 18. *Let $M = q^\Delta$ with $\Delta > 0$. We have*

$$M(f,s)L(f,s) = 1 + O((\log q)^{15} q^{\Delta(1-\sigma)/2})$$

uniformly for $\mathrm{Re}\,(s) = \sigma \to +\infty$.

Proof. By the definition of g_M, and the Dirichlet series for $L(f,s)^{-1}$, the Dirichlet polynomial $M(f,s)$ has the same coefficients of n^{-s} as $L(f,s)^{-1}$ for all $n \leqslant \sqrt{M}$, hence the product $M(f,s)L(f,s)$ is, for $\mathrm{Re}\,(s) = \sigma > 1$, of the form

$$M(f,s)L(f,s) = 1 + \sum_{c>\sqrt{M}} d_f(n)n^{-s}$$

with coefficients $d_f(n)$ bounded by

$$|d_f(n)| = \Big|\sum_{ab=n} c_f(a)\lambda_f(b)\Big| \leqslant \tau(n)^4$$

from Deligne's bound and (5.18). The result follows immediately. □

The density theorem requires a good estimate for the average of the second moment of $M(f,s)L(f,s)$, $\mathrm{Re}\,(s) \geqslant \frac{1}{2} + (\log q)^{-1}$.

Proposition 11. *Let $M = q^\Delta$ with $\Delta < \frac{1}{4}$, and let c be any positive real number with $c < \Delta$. Then there exists a constant $B > 0$ such that for all q prime large enough*

$$\sum_{f\in S_2(q)^*} |M(f,\beta+it)L(f,\beta+it) - 1|^2 \ll (1+|t|)^B q^{1-c(\beta-\frac{1}{2})}. \tag{5.20}$$

uniformly for $\beta \geqslant \frac{1}{2} + (\log q)^{-1}$ and $t \in \mathbf{R}$, the implied constant depending only on Δ and c.

This is the decisive ingredient, which will be the subject matter of the next two section. Assuming this, the proof of Theorem 14 can be completed, again following Selberg's argument.

Thus let α, t_1, t_2 be as in the statement of the theorem. It is obviously enough to consider the case $t_2 - t_1 = (\log q)^{-1}$. As in the proposition, we let $M = q^{\Delta}$ with $\Delta < \frac{1}{4}$, and suppose given $c < \Delta$. Write

$$t_1' = t_1 - \frac{\gamma}{\log q}, \quad t_2' = t_2 + \frac{\gamma}{\log q}, \quad \alpha' = \alpha - \frac{2}{\log q},$$

where γ is a positive real number such that

$$c > \frac{\pi}{2\gamma + 1}. \tag{5.21}$$

Let $f \in S_2(q)^*$, and $\rho = \beta + i\gamma$ one of the zeros of $L(f, s)$ we wish to count, so

$$\beta \geqslant \alpha, \quad t_1 \leqslant \gamma \leqslant t_2$$

and therefore

$$\beta - \alpha' \geqslant \frac{1}{2\log q};$$

and (from the inequality $\sinh(\pi x) \geqslant \pi x$)

$$\frac{2}{\pi}(\log q)(t_2' - t_1') \sinh\Big(\pi \frac{\beta - \alpha'}{t_2' - t_1'}\Big) \geqslant 1.$$

Moreover, since $t_1 \leqslant \gamma \leqslant t_2$, we check (from the definition of t_1', t_2') that γ is not too close from t_1' and t_2', so that

$$\sin\Big(\pi \frac{\gamma - t_1'}{t_2' - t_1'}\Big) \geqslant \frac{1}{\pi}$$

and we obtain the "zero-detecting" inequality: for any zero ρ in the region concerned

$$1 \leqslant 2(\log q)(t_2' - t_1') \sin\Big(\pi \frac{\gamma - t_1'}{t_2' - t_1'}\Big) \sinh\Big(\pi \frac{\beta - \alpha'}{t_2' - t_1'}\Big)$$

so by summing over the zeros ρ we derive[2]

$$N(f; \alpha, t_1, t_2) \leqslant 2(\log q)(t_2' - t_1') \sum_{\rho} \sin\Big(\pi \frac{\gamma - t_1'}{t_2' - t_1'}\Big) \sinh\Big(\pi \frac{\beta - \alpha'}{t_2' - t_1'}\Big).$$

We can now further extend by positivity the sum to include all the zeros in the larger range

$$\sigma \geqslant \alpha', \quad t_1' \leqslant \gamma \leqslant t_2'$$

so $N(f; \alpha, t_1, t_2)$ is bounded by $\log q$ times the exact expression occurring on the left-hand side of Lemma 16. Since zeros of $L(f, s)$ are zeros (with the same multiplicity or higher) of

$$h(f, s) = 1 - (M(f, s)L(f, s) - 1)^2,$$

[2]We extend the summation to include multiplicity in case there are multiple zeros.

to which Selberg's Lemma is applicable, we obtain

$$N(f,\alpha;t_1,t_2) \leqslant (\log q)\int_{t_1'}^{t_2'} \sin\Big(\pi\frac{t-t_1'}{t_2'-t_1'}\Big)\log|h(f,\alpha'+it)|dt$$
$$+ (\log q)\int_{\alpha'}^{+\infty} \sinh\Big(\pi\frac{\sigma-\alpha'}{t_2'-t_1'}\Big)(\log|h(f,\sigma+it_1')| + \log|h(f,\sigma+it_2')|)d\sigma$$

and now from the inequality

$$\log(1-|x|) \leqslant |x|$$

valid for all x, this is bounded in turn by

$$N(f,\alpha;t_1,t_2) \leqslant \log q\int_{t_1'}^{t_2'} \sin\Big(\pi\frac{t-t_1'}{t_2'-t_1'}\Big)|M(f,\alpha'+it)L(f,\alpha'+it)-1|^2dt$$
$$+ \log q\int_{\alpha'}^{+\infty} \sinh\Big(\pi\frac{\sigma-\alpha'}{t_2'-t_1'}\Big)\Big(|M(f,\sigma+it_1')L(f,\sigma+it_1')-1|^2$$
$$+ |M(f,\sigma+it_2')L(f,\sigma+it_2')-1|^2\Big)d\sigma.$$

We now average over f and exchange the order of summation, obtaining inner sums over f to which the estimate for the second moment applies. By (5.20), the first term is estimated by

$$(\log q)q^{1-c(\alpha'-\frac{1}{2})}(t_2'-t_1')$$

and the second by

$$(\log q)q^{1-c(\alpha'-\frac{1}{2})}\int_0^{+\infty} q^{\sigma(\pi/(2\gamma+1)-c)}d\sigma \ll (\log q)q^{1-c(\alpha'-\frac{1}{2})}$$

by the assumption (5.21) on γ.

This concludes the proof of the density theorem, and thereby of the upper bound for the rank of $J_0(q)$, Theorem 5, on the Birch and Swinnerton-Dyer conjecture.

5.3 The harmonic second moment

Proposition 11 will be proved by the method of Chapter 3, going through a corresponding weighted result first.

Proposition 12. *Let $M = q^{\Delta}$ with $\Delta < \frac{1}{4}$, and $\beta = \frac{1}{2} + b(\log q)^{-1}$, where $b > 0$ is any constant. For all q prime large enough we have*

$$\sum_{f\in S_2(q)^*}^{h} |M(f,\beta+it)L(f,\beta+it)|^2 \ll (1+|t|)^B \tag{5.22}$$

for some absolute constant $B > 0$. The implied constant depends only on b and Δ.

The proof is along lines similar to that followed in Chapter 6 when dealing with the second moment in Section 6.1.3.

We write $\beta = \frac{1}{2}+\delta$ and assume only $\delta \geqslant b(\log q)^{-1}$; the actual case in the proposition is $\delta = b(\log q)^{-1}$, but we need not assume this so soon. Then we define for simplicity

$$M_2(\delta) = \sum_{f \in S_2(q)^*}^{h} |M(f, \beta + it)L(f, \beta + it)|^2 \tag{5.23}$$

which we consider as a quadratic form in the coefficients $x_m = x_m(\beta + it)$ of the mollifier. To emphasize this viewpoint, it will be convenient to simply write x_m and $M(f)$ while performing transformations to facilitate the ultimate estimations. This is again in accordance with the principles followed in Chapter 6 for the special value at $\frac{1}{2}$ of the derivatives of the L-functions.

5.3.1 The square of the L-function

This section is very similar to the computation of $L'(f, \frac{1}{2})^2$ in Chapter 6. Let $f \in S_2(q)^*$ and $\beta = \frac{1}{2} + \delta$ with $0 < \delta < \frac{1}{2}$ be given.

Choose an integer $N \geqslant 1$ (which will have to be large enough, $N = 2$ works already) and a real polynomial G satisfying

$$G(-s) = G(s), \text{ and } G(0) = 1 \tag{5.24}$$

$$G(-N) = \ldots = G(-1) = 0 \tag{5.25}$$

and having no zeros for $-\frac{1}{2} \leqslant \mathrm{Re}\,(s) \leqslant \frac{1}{2}$.

Let $t \in \mathbf{R}$ be a fixed real number. Define the entire function $Z(f, s)$ by

$$Z(f, s) = \Lambda(f, s + \tfrac{1}{2} + it)\Lambda(f, s + \tfrac{1}{2} - it)$$

which satisfies the functional equation

$$Z(f, s) = Z(f, -s).$$

Since the Fourier coefficients $\lambda_f(n)$ of f are real, we have

$$|\Lambda(f, \beta + it)|^2 = Z(f, \beta). \tag{5.26}$$

We now consider the complex integral

$$\begin{aligned} I_\delta &= \frac{1}{2i\pi} \int_{(2)} Z(f, s)G(s + it)G(s - it)\frac{ds}{s - \delta} \\ &= \frac{1}{2i\pi} \int_{(2)} L(f, s + \tfrac{1}{2} + it)L(f, s + \tfrac{1}{2} - it)H(s)\Big(\frac{\sqrt{q}}{2\pi}\Big)^{2s+1} \frac{ds}{s - \delta} \end{aligned}$$

(defined, as a function of δ, for all $\delta \in \mathbf{R}$) with

$$H(s) = G(s + it)G(s - it)\Gamma(s + 1 + it)\Gamma(s + 1 - it).$$

From (5.25), zeros of the polynomial G cancel the first poles of the Γ function, so H is holomorphic for $\mathrm{Re}\,(s) > -N - 1$. Moreover, the Gamma function has exponential

decay in vertical strips, while G has polynomial growth, and more precisely, by Stirling's formula, $H(s)$ satisfies

$$H(s) \ll (1+|t|+|\mathrm{Im}\,(s)|)^B e^{-\pi(|t|+|\mathrm{Im}\,(s)|)}$$

for some constant $B > 0$. Moreover, uniformly for $0 \leqslant \delta \leqslant \frac{1}{2}$, we have also

$$H(\delta) \gg e^{-\pi|t|} \tag{5.27}$$

which will be useful when dividing by this quantity later.

Since $L(f,s)$ itself has at most polynomial growth, the integral I_δ is absolutely convergent. To compute it, we can shift the contour of integration to the line $\mathrm{Re}\,(s) = -2$; only a simple pole at $s = \delta$ appears while shifting, with residue

$$\mathrm{Res}_{s=\delta} Z(f,s)G(s+it)G(s-it)\frac{1}{s-\delta} = \Big(\frac{q}{4\pi^2}\Big)^\beta H(\delta)|L(f,\beta+it)|^2$$

by (5.26).

On the line $\mathrm{Re}\,(s) = -2$, the integral is seen to be

$$\frac{1}{2i\pi}\int\limits_{(-2)} Z(f,s)G(s+it)G(s-it)\frac{ds}{s-\delta} = -I_{-\delta}$$

by the change of variable $s \mapsto -s$, using the functional equation of $Z(f,s)$ and the symmetry $G(s) = G(-s)$. Hence we have the formula

$$\Big(\frac{q}{4\pi^2}\Big)^\beta H(\delta)|L(f,\beta+it)|^2 = I_\delta + I_{-\delta}. \tag{5.28}$$

On the other hand, in the region of absolute convergence we can expand the product $L(f,s+it)L(f,s-it)$ into a Dirichlet series and integrate term by term, which gives

$$I_\delta = \Big(\frac{q}{4\pi^2}\Big)^\beta \sum_{l_1,l_2\geqslant 1} \lambda_f(l_1)\lambda_f(l_2)(l_1l_2)^{-\beta}\Big(\frac{l_1}{l_2}\Big)^{it} W_\delta\Big(\frac{4\pi^2 l_1 l_2}{q}\Big)$$

where

$$W_\delta(y) = \frac{1}{2i\pi}\int\limits_{(2)} H(s+\delta)y^{-s}\frac{ds}{s}. \tag{5.29}$$

If we define

$$V(y) = y^{-\delta}W_\delta(y) + y^\delta W_{-\delta}(y) \tag{5.30}$$

then (5.28) gives

$$\Big(\frac{q}{4\pi^2}\Big)^\delta H(\delta)|L(f,\beta+it)|^2 = \sum_{l_1,l_2\geqslant 1} \lambda_f(l_1)\lambda_f(l_2)(l_1l_2)^{-\beta}\Big(\frac{l_1}{l_2}\Big)^{it} V\Big(\frac{4\pi^2 l_1 l_2}{q}\Big)$$

which is further transformed, using the Hecke relations (2.10) and collecting the variable $n = l_1l_2$, to give finally

$$\Big(\frac{q}{4\pi^2}\Big)^\delta H(\delta)|L(f,\beta+it)|^2 = \sum_{n\geqslant 1}\frac{\lambda_f(n)}{\sqrt{n}}\eta_t(n)U\Big(\frac{4\pi^2 n}{q}\Big) \tag{5.31}$$

where we have defined

$$U(y) = \sum_{d \geqslant 1} \frac{\varepsilon_q(d)}{d} V(yd^2) \tag{5.32}$$

and

$$\eta_t(n) = \sum_{ab=n} \left(\frac{a}{b}\right)^{it} \tag{5.33}$$

(this notation is slightly different from that used in Chapter 6: what is here η_t is there η_{it}; but there should be no confusion).

We conclude this section by listing the basic properties of the test function U and the arithmetic function η_t. These should be skipped and consulted when referred to later.

Lemma 19. *For $\delta \neq 0$, we have*

$$U(y) = H(\delta)\zeta_q(1+2\delta)y^{-\delta} + H(-\delta)\zeta_q(1-2\delta)y^{\delta} + O(y^N(1+|t|)^B e^{-\pi|t|}) \tag{5.34}$$

for $0 \leqslant y \leqslant 1$ and

$$U(y) \ll_j y^{-j}(1+|t|)^B e^{-\pi|t|}, \text{ for all } j \geqslant 1 \tag{5.35}$$

for $y \geqslant 1$.

Proof. For the first part, we write

$$\begin{aligned} U(y) = &\frac{1}{2i\pi} \int_{(2)} H(s+\delta)\zeta_q(1+2s+2\delta)y^{-s-\delta}\frac{ds}{s} \\ &+ \frac{1}{2i\pi} \int_{(2)} H(s-\delta)\zeta_q(1+2s-2\delta)y^{-s+\delta}\frac{ds}{s} \end{aligned}$$

and we move the contour of integration to the line $\mathrm{Re}\,(s) = -N-\delta$. In the region thus covered, both $H(s+\delta)$ and $H(s-\delta)$ are holomorphic. We only encounter two simple poles at $s=0$ and $s=-\delta$ (from ζ) in the first integral, and two simple poles at $s=\delta$ and $s=0$ in the second. The sum of the residues at $s=0$ is

$$H(\delta)\zeta_q(1+2\delta)y^{-\delta} + H(-\delta)\zeta_q(1-2\delta)y^{\delta}$$

while the residue at $s=-\delta$ for the first is

$$-\delta^{-1}(1-q^{-1})H(0)$$

which exactly cancels out with the residue at $s=\delta$ of the second integral. Now the estimate of the integral on $\mathrm{Re}\,(s) = -N-\delta$ gives (5.34).

The second part is easily obtained by shifting the contour as far to the right as necessary. □

Lemma 20. *For all $t \in \mathbf{R}$, the arithmetic function η_t is real valued. It satisfies the identities*

$$\eta_t(n)\eta_t(m) = \sum_{d|(n,m)} \eta_t\Big(\frac{nm}{d^2}\Big) \tag{5.36}$$

$$\eta_t(nm) = \sum_{d|(n,m)} \mu(d)\eta_t\Big(\frac{n}{d}\Big)\eta_t\Big(\frac{m}{d}\Big) \tag{5.37}$$

$$\sum_{n\geqslant 1} \eta_t(n)n^{-s} = \zeta(s-it)\zeta(s+it) \tag{5.38}$$

$$\sum_{n\geqslant 1} \eta_t(n)^2 n^{-s} = \frac{\zeta(s-2it)\zeta(s)^2\zeta(s+2it)}{\zeta(2s)} \tag{5.39}$$

$$\sum_{n\geqslant 1} \eta_t(n^2)n^{-s} = \frac{\zeta(s-2it)\zeta(s)\zeta(s+2it)}{\zeta(2s)} \tag{5.40}$$

and the estimate

$$|\eta_t(n)| \leqslant \tau(n). \tag{5.41}$$

Proof. Everything can be checked elementarily by direct computations, but it may as well be deduced from the fact that $\eta_t(n)$ is a Hecke eigenvalue for the operator $T(n)$ acting on the derivative at $s = \frac{1}{2}$ of the non-holomorphic Eisenstein series $E(z, s)$ of level 1, see [Iw3, page 68] for example. □

5.3.2 Computation of the harmonic second moment

By multiplicativity of the coefficients $\lambda_f(n)$, once more, we have

$$|M(f)|^2 = \sum_b \frac{1}{b} \sum_{m_1,m_2} \frac{\lambda_f(m_1m_2)}{\sqrt{m_1m_2}} x_{bm_1}\overline{x_{bm_2}}$$

so that by (5.31), the second moment $M_2(\delta)$ of (5.23) satisfies

$$\Big(\frac{q}{4\pi^2}\Big)^{\delta} H(\delta)M_2(\delta) = \sum_b \frac{1}{b} \sum_{n\geqslant 1} \sum_{m_1,m_2} \frac{\eta_t(n)}{\sqrt{m_1m_2n}} x_{bm_1}\overline{x_{bm_2}} U\Big(\frac{4\pi^2 n}{q}\Big)\Delta(m_1m_2, n)$$

where Δ is the Delta-symbol of Chapter 4. Recall from Lemma 9 that

$$\Delta(m, n) = \delta(m, n) + O((mn)^{1/2}(\log q)^2 q^{-3/2})$$

(for $m,\ n \leqslant q$) where the implied constant is absolute: this is the simplest estimate based on a "trivial" treatment of the remainder term in the Petersson formula, and it will suffice here, in sharp contrast with Chapter 6, where this remainder term will require a non-trivial treatment. The reason is that we are performing the average over all forms $f \in S_2(q)^*$, whereas in the other case only odd forms are involved, the spectral completeness of which, in a sense, is not quite as good.

Using (5.16) to estimate that

$$x_m \ll \zeta(1 + 2\delta)m^{-\delta}$$

the contribution to $M_2(\delta)$ of the remainder term of $\Delta(m,n)$ is at most

$$\begin{aligned}
&\frac{(\log q)^2}{q^{3/2}}\sum_{b\leqslant M}\frac{1}{b}\Big|\sum_m \tau(m)x_{bm}\Big|^2\Big|\sum_{n\geqslant 1}U\Big(\frac{4\pi^2 n}{q}\Big)\Big|\\
&\ll (1+|t|)^B e^{-\pi|t|}\zeta(1+2\delta)^2\frac{(\log q)^2}{\sqrt{q}}\sum_{b\leqslant M}b^{-1-2\delta}\Big|\sum_{bm\leqslant M}\tau(m)m^{-\delta}\Big|^2\\
&\ll \zeta(1+2\delta)^2(\log q)^4 q^{-1/2}M^{2(1-\delta)}(1+|t|)^B e^{-\pi|t|} \qquad (5.42)
\end{aligned}$$

where we have used Lemma 19 to get

$$\begin{aligned}
\sum_{n\geqslant 1}U\Big(\frac{4\pi^2 n}{q}\Big)&=\sum_{n\leqslant q}U\Big(\frac{4\pi^2 n}{q}\Big)+\sum_{n>q}U\Big(\frac{4\pi^2 n}{q}\Big)\\
&\ll H(\delta)q^{\delta}\sum_{n<q}n^{-\delta}+H(-\delta)q^{-\delta}\sum_{n<q}n^{\delta}+(1+|t|)^B e^{-\pi|t|}q^2\sum_{n>q}n^{-2}\\
&\ll q(1+|t|)^B e^{-\pi|t|}.
\end{aligned}$$

We now study the "diagonal contribution" where $n=m_1m_2$, namely the sum $M'(\delta)$ defined by the equality

$$\Big(\frac{q}{4\pi^2}\Big)^{\delta}H(\delta)M'(\delta)=\sum_b\frac{1}{b}\sum_{m_1,m_2}\frac{\eta_t(m_1m_2)}{m_1m_2}x_{bm_1}\overline{x_{bm_2}}U\Big(\frac{4\pi^2 m_1m_2}{q}\Big).$$

Inserting (5.34), we have

$$\begin{aligned}
\Big(\frac{q}{4\pi^2}\Big)^{\delta}H(\delta)M'(\delta)&=\Big(\frac{q}{4\pi^2}\Big)^{\delta}H(\delta)M''(\delta)\\
+O((1+|t|)^B e^{-\pi|t|}\zeta(1+2\delta)^2(\log q)^2 q^{-1/2}M^{2(1-\beta)}) \qquad (5.43)
\end{aligned}$$

where the sum $M''(\delta)$ is given by

$$\begin{aligned}
\Big(\frac{q}{4\pi^2}\Big)^{\delta}H(\delta)M''(\delta)&=\Big(\frac{q}{4\pi^2}\Big)^{\delta}H(\delta)\zeta_q(1+2\delta)\sum_b\frac{1}{b}\sum_{m_1,m_2}\frac{\eta_t(m_1m_2)}{(m_1m_2)^{1+\delta}}x_{bm_1}\overline{x_{bm_2}}\\
&+\Big(\frac{q}{4\pi^2}\Big)^{-\delta}H(-\delta)\zeta_q(1-2\delta)\sum_b\frac{1}{b}\sum_{m_1,m_2}\frac{\eta_t(m_1m_2)}{(m_1m_2)^{1-\delta}}x_{bm_1}\overline{x_{bm_2}} \qquad (5.44)
\end{aligned}$$

and the error term has been estimated by

$$\begin{aligned}
&(1+|t|)^B e^{-\pi|t|}\frac{1}{\sqrt{q}}\sum_{b\leqslant M}\frac{1}{b}\Big|\sum_{bm\leqslant M}\frac{\tau(m)}{\sqrt{m}}x_{bm}\Big|^2\\
&\ll (1+|t|)^B e^{-\pi|t|}\zeta(1+2\delta)^2(\log q)^2 q^{-1/2}M^{2(1-\beta)}.
\end{aligned}$$

5.3.3 Diagonalization

The sum $M''(\delta)$ is now ready for diagonalization, again a process similar to the one which will be used in Chapter 6, but here much simpler.

First, m_1 and m_2 can be separated in (5.44) by means of the Möbius inversion formula (5.37), so

$$\Big(\frac{q}{4\pi^2}\Big)^{\delta} H(\delta)M''(\delta) =$$
$$\Big(\frac{q}{4\pi^2}\Big)^{\delta} H(\delta)\zeta_q(1+2\delta)\sum_b \frac{1}{b}\sum_a \frac{\mu(a)}{a^{2(1+\delta)}}\sum_{m_1,m_2}\frac{\eta_t(m_1)\eta_t(m_2)}{(m_1m_2)^{1+\delta}}x_{abm_1}\overline{x_{abm_2}}$$
$$+\Big(\frac{q}{4\pi^2}\Big)^{-\delta} H(-\delta)\zeta_q(1-2\delta)\sum_b \frac{1}{b}\sum_a \frac{\mu(a)}{a^{2(1-\delta)}}\sum_{m_1,m_2}\frac{\eta_t(m_1)\eta_t(m_2)}{(m_1m_2)^{1-\delta}}x_{abm_1}\overline{x_{abm_2}}$$

and we can collect the single variable $k = ab$, introducing the arithmetic function

$$\nu_\delta(k) = \sum_{ab=k}\frac{\mu(a)}{a^{1+2\delta}}$$

to derive

$$\begin{aligned}\Big(\frac{q}{4\pi^2}\Big)^{\delta} H(\delta)M''(\delta) &= \Big(\frac{q}{4\pi^2}\Big)^{\delta} H(\delta)\zeta_q(1+2\delta)\sum_k \frac{\nu_\delta(k)}{k}\Big|\sum_m \frac{\eta_t(m)}{m^{1+\delta}}x_{km}\Big|^2 \\ &+ \Big(\frac{q}{4\pi^2}\Big)^{-\delta} H(-\delta)\zeta_q(1-2\delta)\sum_k \frac{\nu_{-\delta}(k)}{k}\Big|\sum_m \frac{\eta_t(m)}{m^{1-\delta}}x_{km}\Big|^2.\end{aligned} \tag{5.45}$$

5.3.4 Estimation of the harmonic second moment

Following Selberg, we notice that for $0 < \delta < \frac{1}{2}$ the inequalities

$$\zeta_q(1-2\delta) \leqslant 0$$
$$H(-\delta) = |\Gamma(1-\delta+it)G(-\delta+it)|^2 > 0$$
$$\nu_{-\delta}(k) = \prod_{p|k}(1-p^{-1+2\delta}) \geqslant 0$$

hold. Hence, by positivity

$$M''(\delta) \leqslant \zeta_q(1+2\delta)\sum_k \frac{\nu_\delta(k)}{k}\Big|\sum_m \frac{\eta_t(m)}{m^{1+\delta}}x_{km}\Big|^2 \tag{5.46}$$

after dividing out by $H(\delta)$.

Let

$$y_k = \sum_m \frac{\eta_t(m)}{m^{1+\delta}}x_{km} \tag{5.47}$$

(which is supported on squarefree integers $k \leqslant M$).

Proposition 13. *Assume $\delta = b(\log q)^{-1}$ for some (absolute) constant $b > 0$ and $M = q^{\Delta}$ with $\Delta < \frac{1}{4}$. Then for k squarefree, $k \leqslant M$, we have*

$$k^{\delta+it}\xi(k)y_k \ll \frac{1}{\log q}$$

where

$$\xi(k) = \prod_{p|k}(1-p^{-1/2}).$$

Remark This saving of a factor $\log q$ is the critical moment. It will come essentially from cancellation due to the oscillations of the Möbius function, or in other words, from the Prime Number Theorem.

Proof. We proceed as in [Luo]. From the definition (5.16), for $s = \beta + it = \frac{1}{2} + \delta + it$, we have

$$x_{km} = \frac{\mu(k)}{k^{\delta+it}} \times \frac{\mu(m)}{m^{\delta+it}} \sum_n \frac{\mu(kmn)^2}{n^{1+2\delta+2it}} g_M(kmn)$$

(there is no $\varepsilon_q(n)$ since $n \leqslant kmn \leqslant M < q$). Therefore

$$y_k = \frac{\mu(k)}{k^{\delta+it}} \sum_{m,n} \frac{\mu(kmn)^2 \mu(m) \eta_t(m) n^{-it}}{(mn)^{1+2\delta+it}} g_M(kmn).$$

Assume first that $1 \leqslant k \leqslant \sqrt{M}$ (and of course k is squarefree).

Claim. For all integers $\ell \geqslant 1$, we have

$$g_M(k\ell) = \frac{1}{2i\pi} \int_{(2)} \frac{(\sqrt{M}/k)^s (M^{s/2} - 1)}{\log \sqrt{M}} \ell^{-s} \frac{ds}{s^2}. \tag{5.48}$$

This follows by a direct computation from the well-known formula

$$\frac{1}{2i\pi} \int_{(2)} y^s \frac{ds}{s^2} = \begin{cases} \log y, & \text{if } y \geqslant 1, \\ 0, & \text{if } 0 < y \leqslant 1. \end{cases}$$

From this, by complex integration, it holds

$$k^{\delta+it} y_k = \frac{1}{2i\pi} \int_{(2)} L_k(s + 1 + 2\delta + it) \frac{(\sqrt{M}/k)^s (M^{s/2} - 1)}{\log \sqrt{M}} \frac{ds}{s^2} \tag{5.49}$$

with the *ad-hoc* Dirichlet series

$$L_k(s) = \sum_{\ell \geqslant 1} \mu(k\ell)^2 \Big(\sum_{mn=\ell} \mu(m) \eta_t(m) n^{-it} \Big) \ell^{-s}$$

which is easily computed. Indeed, the inner sum is the coefficient of ℓ^{-s} in the product

$$\begin{aligned} \zeta(s+it) \sum_{m \geqslant 1} \mu(m) \eta_t(m) m^{-s} &= \prod_p (1 - p^{-s-it})^{-1} (1 - p^{-s}(p^{it} + p^{-it})) \\ &\qquad \text{(by multiplicativity and the definition of } \eta_t) \\ &= \prod_p \Big(1 - p^{-s+it} (1 - p^{-s-it})^{-1} \Big) \\ &= \prod_p (1 - p^{-s+it} \sum_{j \geqslant 0} p^{-j(s+it)}) \end{aligned}$$

and $L_k(s)$ is obtained from this Dirichlet series by taking the subseries restricted to integers prime to k and squarefree (this is the effect of inserting $\mu(k\ell)^2$ in a Dirichlet series). This gives the very simple answer

$$L_k(s) = \zeta_k(s - it)^{-1}.$$

From the theorems of Hadamard and de la Vallée-Poussin, $\zeta(s)$ has no zeros on the line $\mathrm{Re}\,(s) = 1$ and more precisely the estimate

$$\zeta(s)^{-1} \ll \log(2 + |\mathrm{Im}\,(s)|) \tag{5.50}$$

holds with an absolute implied constant (see [Tit, ch. 3]) uniformly for

$$\mathrm{Re}\,(s) \geqslant 1 - \frac{D}{\log(2 + |\mathrm{Im}\,(s)|)}$$

($D > 0$ being another absolute constant).

Let r be small enough so that the circle $|s| \leqslant r$ is included in this zero-free region, and $0 < r < \frac{1}{2}$ (of course, any $r < \frac{1}{2}$ will do, the Riemann Hypothesis being numerically valid in such a range!). In (5.49), we shift the integration to the contour C consisting of the vertical line $\mathrm{Re}\,(s) = 0$ from $-i\infty$ to $-ir$, followed by the half-circle $s = re(x)$ for $-\frac{\pi}{2} \leqslant x \leqslant \frac{\pi}{2}$, and then again the line $\mathrm{Re}\,(s) = 0$ from ir to $i\infty$. By the choice of r, this is permissible; the contour shift passes through a unique simple pole at $s = 0$ (simple because of the zero of $s \mapsto M^{s/2} - 1$), and from the residue and the formula for $L_k(s)$ we get

$$k^{\delta+it} y_k = \zeta_k (1 + 2\delta)^{-1} + \frac{1}{2i\pi} \int_C \frac{\zeta(s + 1 + 2\delta + it)^{-1}}{\prod_{p|k} (1 - p^{-(s+1+2\delta)})} \frac{(\sqrt{M}/k)^s (M^{s/2} - 1)}{\log \sqrt{M}} \frac{ds}{s^2}.$$

The integral over C is now estimated. Using (5.50), the part from ir to $i\infty$ is dominated by

$$\frac{1}{\log M} \Big| \int_r^{+\infty} \frac{\zeta(1 + 2\delta + iu)^{-1}}{\prod_{p|k} (1 - p^{-(1+2\delta+iu)})} (\sqrt{M}/k)^{iu} (M^{iu/2} - 1) \frac{du}{u^2} \Big| \ll \frac{1}{\log q} \frac{1}{\xi(k)}$$

since clearly

$$\prod_{p|k} (1 - p^{-1-2\delta})^{-1} \leqslant \xi(k)^{-1}.$$

The same holds without change for the other vertical half-line. For the semi-circle, we use the fact that $k \leqslant \sqrt{M}$ so that

$$(\sqrt{M}/k)^s (M^{s/2} - 1) \ll 1$$

on this semi-circle where $\mathrm{Re}\,(s) < 0$, and similarly the product over primes dividing k is dominated by its value at $s = -r$ which is

$$\prod_{p|k} (1 - p^{r-1-\delta})^{-1} \leqslant \xi(k)^{-1}$$

since $r < \frac{1}{2}$. Hence the same bound holds again.

In the case $\sqrt{M} \leqslant k \leqslant M$, we use a similar reasoning, replacing (5.48) by the other formula

$$g_M(k\ell) = \frac{1}{2i\pi} \int_{(2)} \frac{(M/k)^s}{\log \sqrt{M}} \ell^{-s} \frac{ds}{s^2}$$

and using the same contour shift. The corresponding integral over C is estimated exactly as before, but a double pole is now present at $s = 0$, with residue

$$\frac{1}{\log\sqrt{M}}\mathrm{Res}_{s=0}\,\frac{\zeta_k(s+1+2\delta)^{-1}(M/k)^s}{s^2}$$

which is equal to

$$\frac{1}{\log\sqrt{M}}\Big(\zeta_k(1+2\delta)^{-1}(\log\frac{M}{k}) + \zeta_k'(1+2\delta)\Big).$$

An easy computation gives

$$\zeta_k'(1+2\delta) = \frac{1}{\nu_\delta(k)}\frac{\zeta'(1+2\delta)}{\zeta(1+2\delta)^2} + \frac{1}{\zeta(1+2\delta)\nu_\delta(k)}\sum_{p|k}\frac{\log p}{p^{1+2\delta}-1}.$$

Now we collect the results and we use the assumption that $\delta = b(\log q)^{-1}$, which implies

$$\zeta(1+2\delta)^{-1} \ll (\log q)^{-1}.$$

Hence for $k \leqslant \sqrt{M}$ we obtain immediately the desired bound

$$\xi(k)k^{1+\delta}y_k \ll \frac{1}{\log q}.$$

This is also true for $\sqrt{M} \leqslant k \leqslant M$: in the residue, the saving of $\log q$ comes from $\zeta(1+2\delta)^{-1}$ for the first term, and from $(\log\sqrt{M})^{-1}$ in the second, while the last is actually smaller, since

$$\sum_{p|k}\frac{\log p}{p^{1+2\delta}-1} \ll \log\log k$$

and both $\zeta(1+2\delta)^{-1}$ and $(\log\sqrt{M})^{-1}$ are present (we use again $\nu_\delta(k)^{-1} \leqslant \xi(k)^{-1}$).

This finishes the proof. □

Corollary 3. *Proposition 12 is true: if $M = q^\Delta$ with $\Delta < \frac{1}{4}$ and $\delta = b(\log q)^{-1}$, then*

$$M_2(\delta) \ll (1+|t|)^B$$

for some absolute constant $B > 0$, the implied constant depending only on b and Δ.

Proof. From the previous proposition and (5.46) we have

$$\begin{aligned} M''(\delta) &\leqslant \zeta_q(1+2\delta)\sum_{k\leqslant M}\frac{\nu_\delta(k)}{k}|y_k|^2 \\ &\ll \frac{\zeta_q(1+2\delta)}{(\log q)^2}\sum_{k\leqslant M}\frac{\mu(k)^2\nu_\delta(k)}{\xi(k)^2}k^{-(1+2\delta)} \\ &\ll \frac{1}{\log q}\sum_{k\leqslant M}\frac{\mu(k)^2\nu_\delta(k)}{\xi(k)^2}k^{-(1+2\delta)} \end{aligned}$$

since

$$\zeta_q(1+2\delta) \ll \log q$$

for $\delta = b(\log q)^{-1}$. Now we have

$$\nu_\delta(k) = \prod_{p|k}(1 - p^{-(1+2\delta)}) \leqslant 1$$

for all k, so

$$M''(\delta) \ll \frac{1}{\log q}\sum_{k\leqslant M}\frac{\mu(k)^2}{\xi(k)^2}k^{-(1+2\delta)}.$$

The Dirichlet series

$$\sum_{k\geqslant 1}\frac{\mu(k)^2}{\xi(k)^2}k^{-s} = \prod_p\Big(1 + \frac{p^{-s}}{(1-p^{-1/2})^2}\Big)$$

is absolutely convergent for $\mathrm{Re}\,(s) > 1$. Clearly it has analytic continuation to the region $\mathrm{Re}\,(s) > \frac{1}{2}$ with a simple pole at $s = 1$, and therefore

$$\sum_{k\leqslant M}\frac{\mu(k)^2}{\xi(k)^2}k^{-(1+2\delta)} \ll \log q.$$

To conclude, we look back to the error terms (5.42) and (5.43) introduced in going from the original second moment $M_2(\delta)$ to $M'(\delta)$ and then to $M''(\delta)$, and we see that they bring a total contribution which is

$$\ll q^{-\gamma}(1+|t|)^B$$

for some $\gamma = \gamma(\Delta)$ if $M = q^\Delta$ with $\Delta < \frac{1}{4}$. The polynomial bound in t is correct: when dividing by $H(\delta)$ (as must be done), we have the bound $H(\delta)^{-1} \ll e^{\pi|t|}$ previously observed, and the factor $e^{-\pi|t|}$ in the error terms cancels it. □

5.4 Removing the harmonic weight: the head, I

Having estimated $A^h[|M(f,\beta+it)L(f,\beta+it)|^2]$, we now apply Proposition 8 to study $A[|M(f,\beta+it)L(f,\beta+it)|^2]$. The notations and assumptions are the same as at the beginning of the previous Section: recall that $\beta = \frac{1}{2}+\delta$, and $M = q^\Delta$ with $\Delta < \frac{1}{4}$.

First we check the conditions; (3.17) is contained in Proposition 12, while for (3.18), we have

Lemma 21. *For all $f \in S_2(q)^*$, it holds*

$$\omega_f|M(f,\beta+it)L(f,\beta+it)|^2 \ll q^{-\frac{1}{4}}(1+|t|)^2$$

for all β with $\beta \geqslant \frac{1}{2}$, the implied constant being absolute.

Proof. Using (5.19), the trivial bound for $M(f,\beta+it)$ is

$$M(f,\beta+it) \ll \sqrt{M}(\log q)^3$$

while the convexity bound for $L(f,s)$ on the critical line gives

$$L(f,\beta+it) \ll_\varepsilon q^{\frac{1}{4}+\varepsilon}(1+|t|)^{\frac{1}{2}+\varepsilon}$$

for $\beta \geqslant \frac{1}{2}$. Since on the other hand we have $\omega_f \ll (\log q)q^{-1}$ from (3.16), the result follows. □

Hence Proposition 8 with $x = q^\kappa$, for any $\kappa > 0$, gives the equality

$$A[|M(f,\beta+it)L(f,\beta+it)|^2] = \frac{\dim J_0(q)}{\zeta(2)} A^h[\omega_f(x)|M(f,\beta+it)L(f,\beta+it)|^2] + O((1+|t|)^B q^{1-\gamma}) \tag{5.51}$$

for some $\gamma = \gamma(\Delta,\kappa) > 0$ (the dependence in t of the error term has to be checked by looking back at the proof of the proposition).

We let

$$\begin{aligned}\mathcal{M}_2(\delta) &= A^h[\omega_f(x)|M(f,\beta+it)L(f,\beta+it)|^2] \\ &= \sum_{d\ell^2\leqslant x} \frac{1}{d\ell^2} \sum_{f\in S_2(q)^*}^h \lambda_f(d^2)|M(f,\beta+it)L(f,\beta+it)|^2.\end{aligned}$$

Computing as in Section 5.3.2, we get

$$\Big(\frac{q}{4\pi^2}\Big)^\delta H(\delta)\mathcal{M}_2(\delta) = \sum_b \frac{1}{b} \sum_{n\geqslant 1} \sum_{m_1,m_2} \frac{\eta_t(n)}{\sqrt{m_1m_2n}} x_{bm_1}\overline{x_{bm_2}} U\Big(\frac{4\pi^2 n}{q}\Big)\Delta^n(m_1m_2,n).$$

From Lemma 12 we have

$$\Delta^n(m,n) = \sum_{d\ell^2\leqslant x} \frac{1}{d\ell^2} \sum_{r|(d^2,m)} \delta\Big(\frac{md^2}{r^2},n\Big) + O\Big((\log q)^3 \frac{x\sqrt{mn}}{q^{3/2}}\Big)$$

and the error term yields a contribution which, by the same computation as in (5.42), is at most

$$\zeta(1+2\delta)^2(\log q)^7 \frac{xM^{2(1-\delta)}}{\sqrt{q}}(1+|t|)^B e^{-\pi|t|} \ll (1+|t|)^B e^{-\pi|t|} q^{-\gamma} \tag{5.52}$$

for some $\gamma > 0$, if κ is taken small enough.

The diagonal contribution $n = md^2r^{-2}$ is

$$\sum_b \frac{1}{b} \sum_{m_1,m_2} \frac{x_{bm_1}\overline{x_{bm_2}}}{m_1m_2} \sum_{d\ell^2\leqslant x} (d\ell)^{-2} \sum_{r|(m_1m_2,d^2)} r\eta_t\Big(\frac{m_1m_2d^2}{r^2}\Big) U\Big(\frac{4\pi^2 m_1m_2d^2}{qr^2}\Big)$$

and we use (5.34) to get

$$\begin{aligned}\Big(\frac{q}{4\pi^2}\Big)^\delta H(\delta)\mathcal{M}_2(\delta) &= \Big(\frac{q}{4\pi^2}\Big)^\delta H(\delta)\zeta_q(1+2\delta)\mathcal{M}(\delta) \\ &+ \Big(\frac{q}{4\pi^2}\Big)^{-\delta} H(-\delta)\zeta_q(1-2\delta)\mathcal{M}(-\delta) + O((1+|t|)^B e^{\pi|t|} q^{-\gamma})\end{aligned} \tag{5.53}$$

for some $\gamma > 0$, where the sum $\mathcal{M}(\delta)$ is

$$\mathcal{M}(\delta) = \sum_b \frac{1}{b} \sum_{m_1,m_2} \frac{x_{bm_1}\overline{x_{bm_2}}}{(m_1m_2)^{1+\delta}} \sum_{d\ell^2\leqslant x} \frac{1}{d^{2+2\delta}\ell^2} \sum_{r|(m_1m_2,d^2)} r^{1+2\delta}\eta_t\Big(\frac{m_1m_2d^2}{r^2}\Big). \tag{5.54}$$

We will first compute the inner sum, showing in particular that we can now again extend the summation over d, ℓ to all integers, and then we compute this complete series.

We define a function $u(s,r)$ for $s \in \mathbf{C}$ and $r \geqslant 1$ an integer by

$$u(s,r) = \sum_{ab=r} \mu(a) b^s = \prod_{p|r} (p^s - 1)$$

and a function $v_x(s,r)$, supported on cubefree integers r, by

$$v_x(s,r) = \sum_{\substack{d\ell^2 \leqslant x \\ r | d^2}} \ell^{-2} d^{-2s} \eta_t\Big(\frac{d^2}{r}\Big) \tag{5.55}$$

we also denote by $v(s,r)$ the function obtained by removing the constraint $d\ell^2 \leqslant x$ in the definition of $v(s,r)$.

From the formula (5.37), we have for every integers m and n

$$\sum_{r|(m,n)} r^s \eta_t\Big(\frac{mn}{r^2}\Big) = \sum_{r|(m,n)} u(s,r) \eta_t\Big(\frac{m}{r}\Big) \eta_t\Big(\frac{n}{r}\Big)$$

hence

$$\sum_{d\ell^2 \leqslant x} \frac{1}{d^{2+2\delta}\ell^2} \sum_{r|(m_1 m_2, d^2)} r^{1+2\delta} \eta_t\Big(\frac{m_1 m_2 d^2}{r^2}\Big) = \sum_{r|m_1 m_2} \eta_t\Big(\frac{m_1 m_2}{r}\Big) u(1+2\delta, r) v_x(1+\delta, r). \tag{5.56}$$

We recall that the multiplicative functions N and M are defined by

$$N(r) = \prod_{p|r} p, \quad M(r) = \prod_{p||r} p.$$

Lemma 22. *For all cubefree integers $r \geqslant 1$, and s with $\mathrm{Re}\,(s) = \sigma > \frac{1}{2}$, we have*

$$v_x(s,r) = v(s,r) + O\Big(\frac{(\log x)^3 \tau(r)}{N(r)^{2\sigma - \frac{1}{2}} \sqrt{x}}\Big). \tag{5.57}$$

Moreover

$$v(s,1) = \frac{\zeta(2)\zeta(2s)\zeta(2s+2it)\zeta(2s-2it)}{\zeta(4s)}$$

and for all $r \geqslant 1$

$$v(s,r) = v(s,1) N(r)^{-2s} \prod_{p||r} \frac{\eta_t(p)}{1 + p^{-2s}}$$

(in the language of Section 4.3, we see that for s fixed the arithmetic function $v(s,r)$ is 1-mutative).

Proof. The point is that for a cubefree integer r and any $d \geqslant 1$, we have $r \mid d^2$ if and only if $N(r) \mid d$. Since

$$r = M(r) \frac{N(r)^2}{M(r)^2} = \frac{N(r)^2}{M(r)}$$

we can write

$$v_x(s,r) = \sum_{\substack{d\ell^2 \leqslant x \\ N(r)|d}} \ell^{-2} d^{-2s} \eta_t\Big(\frac{d^2}{r}\Big)$$
$$= N(r)^{-2s} \sum_{d\ell^2 \leqslant x/N(r)} \ell^{-2} d^{-2s} \eta_t(M(r)d^2).$$

and similarly without constraint for $v(s,r)$. Now, putting $y = x/N(r)$

$$\sum_{d\ell^2 > x/N(r)} \ell^{-2} d^{-2s} \eta_t(M(r)d^2) \ll \tau(M(r))\Big(\sum_{\ell^2<y} \ell^{-2} \sum_{d>y/\ell^2} \tau(d^2) d^{-2\sigma} + \sum_{\ell^2>y} \ell^{-2}\Big)$$
$$\ll \tau(r)(\log x)^3 y^{-1/2}$$

and this gives the first formula.

To compute $v(s,r)$ (which is a kind of "non-primitive" symmetric square for η_t), we define

$$v'(s,r) = \sum_{d\geqslant 1} \eta_t(M(r)d^2) d^{-2s}$$

so that $v(s,r) = \zeta(2)N(r)^{-2s} v'(s,r)$. We denote by $Z(s)$ the full symmetric square given by (5.40), and by Z_p its p-factor.

Every integer d has a unique expression $d = d_1 d_2$ with $d_1 \mid M(r)^\infty$ and $(d_2, M(r)) = 1$ so by multiplicativity we get

$$v'(s,r) = \Big(\sum_{(d,M(r))=1} \eta_t(d^2) d^{-2s}\Big)\Big(\sum_{d|M(r)^\infty} \eta_t(M(r)d^2) d^{-2s}\Big)$$
$$= Z(2s) \prod_{p||r} Z_p(2s)^{-1} \times \prod_{p||r} \sum_{k\geqslant 0} \eta_t(p^{2k+1}) p^{-2ks}.$$

Again by multiplicativity,

$$\eta_t(p^{2k+1}) = \eta_t(p)\eta_t(p^{2k}) - \eta_t(p^{2k-1})$$

for $k \geqslant 1$, so that

$$(1+p^{-2s}) \sum_{k\geqslant 0} \eta_t(p^{2k+1}) p^{-2ks} = \eta_t(p) Z_p(2s)$$

which yields

$$v'(s,r) = Z(2s) \prod_{p||r} \frac{\eta_t(p)}{1+p^{-2s}}.$$

This gives the lemma, since

$$v(s,1) = Z(2s) = \frac{\zeta(2)\zeta(2s)\zeta(2s+2it)\zeta(2s-2it)}{\zeta(4s)}$$

from (5.40). □

Let now $w_x(s,m)$ be the function defined for $s \in \mathbf{C}$ and $m \geqslant 1$ by

$$w_x(s,r) = \sum_{r|m} \eta_t\Big(\frac{m}{r}\Big) u(2s-1,r) v_x(s,r), \tag{5.58}$$

and let $w(s,m)$ be the same with $v(s,r)$ replacing $v_x(s,r)$. Then from (5.56) and (5.54) comes the formula

$$\mathcal{M}(\delta) = \sum_b \frac{1}{b} \sum_{m_1,m_2} \frac{w_x(1+\delta,m)}{(m_1m_2)^{1+\delta}} x_{bm_1} \overline{x_{bm_2}}. \tag{5.59}$$

Lemma 23. *Assume that $\delta = b(\log q)^{-1}$ for any constant $b > 0$. Then*

$$\mathcal{M}(\delta) = \sum_b \frac{1}{b} \sum_{m_1,m_2} \frac{w(1+\delta,m_1m_2)}{(m_1m_2)^{1+\delta}} x_{bm_1} \overline{x_{bm_2}} + O(q^{-\gamma})$$

for some $\gamma = \gamma(\kappa,\Delta) > 0$.

Proof. Since m_1 and m_2 are squarefree, the product m_1m_2, and any divisor thereof, is always cubefree. So we use (5.57) to replace $v_x(1+\delta,r)$ by $v(1+\delta,r)$. This gives first

$$w(1+\delta,m) = w_x(1+\delta,m) + O\Big(\frac{\tau(m)^3(\log x)^3}{\sqrt{x}}\Big)$$

because the error term is bounded by

$$\sum_{r|m} \tau\Big(\frac{m}{r}\Big) |u(1+2\delta,r)| \frac{(\log x)^3 \tau(r)}{N(r)^{\frac{3}{2}+2\delta}\sqrt{x}} \ll \frac{\tau(m)^3(\log x)^3}{\sqrt{x}}$$

by the estimate

$$|u(1+2\delta,r)| = \Big|\prod_{p|r} (p^{1+2\delta}-1)\Big| \leqslant N(r)^{1+2\delta}.$$

Then inserting this inside $\mathcal{M}(\delta)$ gives the result. □

After all those transformations, we are with $\mathcal{M}(\delta)$ (more precisely, its main term) in the situation highlighted in the Appendix to Section 4.3 (except that we have a sum over b of quadratic forms as described there). Indeed, we have seen in Lemma 22 that $v(1+\delta,r)$ is 1-mutative (ie the product of a constant and a multiplicative function), and by Dirichlet convolution it follows that also has $w(1+\delta,m)$ this property:

$$w(1+\delta,m) = v(1+\delta,1)\overline{w}(m)$$

with $\overline{w}$ multiplicative.

Of course, this is a very trivial case, and the diagonalization of $\mathcal{M}(\delta)$ can be performed simply by doing the transformations, but the corresponding moment in Section 6.2 has more sophisticated coefficients. We recall quickly the yoga of extracting

the common divisor of m_1 and m_2 and removing the coprimality condition by Möbius inversion:

$$\begin{aligned}\sum_b \frac{1}{b}\sum_{m_1,m_2}\frac{w(1+\delta,m_1m_2)}{(m_1m_2)^{1+\delta}}x_{bm_1}\overline{x_{bm_2}} &= v(1+\delta,1)\sum_b\frac{1}{b}\sum_{m_1,m_2}\frac{\overline{w}(m_1m_2)}{(m_1m_2)^{1+\delta}}x_{bm_1}\overline{x_{bm_2}}\\ &= v(1+\delta,1)\sum_b\frac{1}{b}\sum_a\frac{\overline{w}(a^2)}{a^{2(1+\delta)}}\sum_{(m_1,m_2)=1}\frac{\overline{w}(m_1)\overline{w}(m_2)}{(m_1m_2)^{1+\delta}}x_{abm_1}\overline{x_{abm_2}}\\ &= v(1+\delta,1)\sum_b\frac{1}{b}\sum_a\frac{\overline{w}(a^2)}{a^{2(1+\delta)}}\sum_d\frac{\mu(d)\overline{w}(d)^2}{d^{2(1+\delta)}}\sum_{m_1,m_2}\frac{\overline{w}(m_1)\overline{w}(m_2)}{(m_1m_2)^{1+\delta}}x_{adbm_1}\overline{x_{adbm_2}}\\ &= v(1+\delta,1)\sum_k\frac{\tilde{\nu}_\delta(k)}{k}\Big|\sum_m\frac{\overline{w}(m)}{m^{1+\delta}}x_{km}\Big|^2 \end{aligned} \tag{5.60}$$

with

$$\tilde{\nu}_\delta(k)=\sum_{abd=k}\frac{\mu(d)\overline{w}(d)^2\overline{w}(a^2)}{(ad)^{1+2\delta}}.$$

Remember that $\tilde{\nu}(k)$ also depends on t (through η_t involved in $\overline{w}$).

Lemma 24. *There exists an absolute constant $b>0$ such that if $|\delta|\leqslant b(\log q)^{-1}$ then*

$$\tilde{\nu}_\delta(k)\geqslant 0$$

for all $t\in\mathbf{R}$ and all $k<q$, and

$$v(1+\delta,1)\gg 1$$

for all $t\in\mathbf{R}$.

Proof. By multiplicativity it is enough to consider $k=p$ prime, $p<q$. Then

$$e^{-2b}\leqslant p^{2\delta}\leqslant e^{2b}.$$

But

$$\nu_\delta(p)=1+\frac{1}{p^{1+2\delta}}(\overline{w}(p^2)-\overline{w}(p)^2)$$

and by direct computation, from Lemma 22 and the definition of $w(1+\delta,m)$, we have

$$\begin{aligned}\overline{w}(p)&=\eta_t(p)+(p^{1+2\delta}-1)p^{-2(1+\delta)}\frac{\eta_t(p)}{1+p^{-2(1+\delta)}}\\ &=\eta_t(p)\frac{p^{2+2\delta}+p^{1+2\delta}}{p^{2+2\delta}+1}\end{aligned}$$

and similarly

$$\begin{aligned}\overline{w}(p^2)&=\eta_t(p^2)+\eta_t(p)^2\frac{p^{1+2\delta}-1}{p^{2+2\delta}+1}+\frac{p^{1+2\delta}-1}{p^{2+2\delta}}\\ &=\eta_t(p)^2\frac{p^{2+2\delta}+p^{1+2\delta}}{p^{2+2\delta}+1}-1+\frac{p^{1+2\delta}-1}{p^{2+2\delta}}\end{aligned}$$

hence

$$\overline{w}(p^2) - \overline{w}(p)^2 = -\eta_t(p)^2 \frac{p^{2+2\delta} + p^{1+2\delta}}{p^{2+2\delta}+1} \frac{p^{1+2\delta}-1}{p^{2+2\delta}+1} - 1 + \frac{p^{1+2\delta}-1}{p^{2+2\delta}}.$$

For given p we can estimate from below using $0 \leqslant \eta_t(p)^2 \leqslant 4$

$$\nu_\delta(k) \geqslant 1 - 4\frac{p+1}{p^2e^{-2b}+1}\frac{pe^{-2b}-1}{p^2e^{-2b}+1} - \frac{e^{2b}}{p} + \frac{pe^{-2b}-1}{p^3e^{-4b}}$$

(if b is small enough). As a function of p, for b fixed, this is continuous. For $b = 0$, it is increasing as a function of p and is positive for $p = 2$ (where its value is $29/200$). Hence the existence of b follows.

As for $v(1+\delta, 1)$, we have by Lemma 22

$$v(1+\delta,1) = \frac{\zeta(2)\zeta(2+2\delta)|\zeta(2+2\delta+it)|^2}{\zeta(4+4\delta)} \gg 1$$

uniformly in t for all $\delta > -\frac{1}{4}$ (for instance). □

Remark Of course, we do not depend on the positivity of $\tilde{\nu}_\delta(k)$ for small primes for the validity of the argument. It is clear from the beginning that for δ of size $(\log q)^{-1}$, we have $\tilde{\nu}_\delta(p) > 0$ for p large enough, independently of q. The remaining small primes can then be sifted out from the start. But it is best to avoid such technical complications.

We can now conclude this part of the argument.

Proposition 14. *Assume that $\delta = b(\log q)^{-1}$ with $b > 0$ a fixed constant such that the previous lemma applies. Then*

$$\sum_{f \in S_2(q)^*} |M(f, \beta+it)L(f,\beta+it)|^2 \ll q(1+|t|)^B$$

for some absolute constant $B \geqslant 0$. The implied constant depends now only on Δ.

Proof. From Lemma 24, and the computation of $\mathcal{M}(\delta)$ and the subsequent diagonalization of the main term, we see that for q large enough we have

$$\mathcal{M}(-\delta) \geqslant 0$$

hence, using the same trick as before that $\zeta_q(1-2\delta) \leqslant 0$, we get by positivity the inequality

$$\mathcal{M}_2(\delta) \leqslant \zeta(1+2\delta)\mathcal{M}(\delta) \ll v(1+\delta,1)\zeta(1+2\delta)\sum_k \frac{\tilde{\nu}_\delta(k)}{k}\Big|\sum_m \frac{\overline{w}(m)}{m^{1+\delta}}x_{km}\Big|^2.$$

Now, in terms of the linear forms y_k introduced in (5.47), we can write

$$\sum_m \frac{\overline{w}(m)}{m^{1+\delta}}x_{km} = \sum_{a,b}\frac{\eta_t(a)\eta_t(b)u(1+\delta,b)}{(ab)^{1+\delta}(b^{2(1+\delta)}+1)}x_{abk} = \sum_b \frac{u(1+\delta,b)\eta_t(b)}{b^{1+\delta}(b^{2(1+\delta)}+1)}y_{bk}$$

since for squarefree n we have $N(n) = n$. But $u(1+\delta, b) \leqslant b^{1+2\delta}$ for b squarefree, and Proposition 13 gives immediately

$$\sum_m \frac{\overline{w}(m)}{m^{1+\delta}}x_{km} \ll \frac{1}{k^{(1+\delta)}\xi(k)}\frac{1}{\log q}$$

and then the proof is completed as for Corollary 3, going back to (5.51) to conclude. □

Proposition 5.20 is an easy consequence of this estimate near the critical line. Indeed, it first immediately provides the bound

$$\sum_{f\in S_2(q)^*} |M(f,\beta+it)L(f,\beta+it)-1|^2 \ll q(1+|t|)^B \tag{5.61}$$

for $\beta = \frac{1}{2} + b(\log q)^{-1}$. On the other hand, for $\mathrm{Re}\,(s) = \sigma > 1$, we have

$$\sum_{f\in S_2(q)^*} |M(f,s)L(f,s)-1|^2 \ll q^{1-\Delta(1-\sigma)}(\log q)^{30} \tag{5.62}$$

as a consequence of the trivial individual bound of Lemma 18.

Consider any $\sigma_0 \geqslant 2$. We will interpolate by convexity between the bound (5.61) near $\sigma = \frac{1}{2}$ and this bound for $\sigma = \sigma_0$, by means of a simple (and well-known) extension of the classical convexity principle of Phragmen-Lindelöf.

Recall first that a function f is called subharmonic if its Laplacian Δf is non-negative. For example, if f is holomorphic, then $|f|^2$ is subharmonic since

$$\Delta|f|^2 = 2(u_x^2+v_x^2+u_y^2+v_y^2) + 2u\Delta v + 2v\Delta u = 2(u_x^2+v_x^2+u_y^2+v_y^2) \geqslant 0$$

(where $u = \mathrm{Re}\,(f)$, $v = \mathrm{Im}\,(f)$, both harmonic). The fundamental property of subharmonic functions is that they satisfy the maximum modulus principle.

Lemma 25. *Let $f_1,\dots,f_j$ be finitely many functions holomorphic inside the strip*

$$0 < a - \vartheta < \mathrm{Re}\,(s) = \sigma < b + \vartheta$$

(for some $\vartheta > 0$), such that the function

$$h = |f_1|^2 + \cdots + |f_j|^2$$

has at most polynomial growth in the strip and satisfies

$$|h(s)| \leqslant Cq^c|s|^B, \text{ for } \mathrm{Re}\,(s) = a$$
$$|h(s)| \leqslant Cq^d|s|^B, \text{ for } \mathrm{Re}\,(s) = b,$$

where c, d, B, and C are real numbers with $B \geqslant 0$, $C \geqslant 0$. Then for any s in the strip $a \leqslant \mathrm{Re}\,(s) \leqslant b$, we have

$$|h(s)| \leqslant Cq^{\alpha(\sigma)}|s|^B$$

where α is the linear function with $\alpha(a) = c$, $\alpha(b) = d$.

Proof. For any $\varepsilon > 0$, let $g_\varepsilon(s)$ be the function

$$g_\varepsilon(s) = q^{-\alpha(\sigma)}|s|^{-B}|\Gamma(1+\varepsilon s)|^2 h(s)$$

defined in the strip. The point is that

$$g_\varepsilon(s) = \sum_i |q^{-\alpha(s)/2}s^{-B/2}\Gamma(1+\varepsilon s)f_i(s)|^2$$

so g_ε, as a sum of subharmonic functions, is a subharmonic function. We have moreover

$$|g_\varepsilon(s)| \leqslant C\Gamma(1+\varepsilon b)^2$$

for $\mathrm{Re}\,(s) = a$ or $\mathrm{Re}\,(s) = b$ since $|\Gamma(s)| \leqslant \Gamma(\mathrm{Re}\,(s))$ for $\mathrm{Re}\,(s) > 0$. The Gamma function decreases exponentially, and h has polynomial growth, so $g_\varepsilon(s)$ tends to zero for $|\mathrm{Im}\,(s)|$ tending to infinity, uniformly for $a \leqslant \mathrm{Re}\,(s) \leqslant b$. Hence by the maximum modulus principle applied in a sufficiently large rectangle $[a, b] \times [-T, T]$, it must hold

$$|g_\varepsilon(s)| \leqslant C\Gamma(1 + \varepsilon b)^2$$

for $a \leqslant \mathrm{Re}\,(s) \leqslant b$, or

$$|h(s)| \leqslant C|s|^B q^{\alpha(\sigma)} \frac{\Gamma(1 + \varepsilon b)^2}{\Gamma(1 + \varepsilon s)^2},$$

and letting ε tend to zero gives the required inequality. □

We deduce from this lemma and (5.61), (5.62) that for

$$\frac{1}{2} + \frac{1}{\log q} \leqslant \sigma \leqslant \sigma_0$$

we have

$$\sum_{f \in S_2(q)^*} |M(f, s)L(f, s) - 1|^2 \ll_\varepsilon q^{1 - c(\varepsilon)(\sigma - \frac{1}{2})}(1 + |t|)^B$$

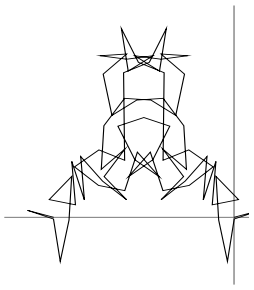
with $c(\varepsilon)$ given by

$$c(\varepsilon) = \frac{\Delta(\sigma_0 - 1) - \varepsilon}{\sigma_0 - \frac{1}{2}}.$$

This inequality remains valid for $\mathrm{Re}\,(s) > \sigma_0$, because it is weaker than the trivial bound in this region. Now taking σ_0 large enough and ε small enough, we will obtain

$$\sum_{f \in S_2(q)^*} |M(f, s)L(f, s) - 1|^2 \ll q^{1 - c(\sigma - \frac{1}{2})}(1 + |t|)^B$$

uniformly for $\mathrm{Re}\,(s) \geqslant \frac{1}{2} + b(\log q)^{-1}$, for any $c < \Delta$. This proves Proposition 11, since we can obviously assume that $b < 1$.



Chapter 6
Proof of the lower bound

"Oh, in that case I guess it's okay,"
Colonel Korn said, mollified.
Joseph Heller, *"Catch-22"*

We recall the statement of Theorem 10 to be proved: for any $\varepsilon > 0$, and for any prime number q large enough in terms of ε, we have

$$|\{f \in S_2(q)^* \mid L(f, \tfrac{1}{2}) = 0,\ L'(f, \tfrac{1}{2}) \neq 0\}| \geqslant \Big(\frac{19}{54} - \varepsilon\Big) \dim J_0(q). \tag{6.1}$$

and this quantity is the same as

$$|\{f \in S_2(q)^* \mid \ f \text{ is odd and } L'(f, \tfrac{1}{2}) \neq 0\}|.$$

In this chapter again, q is a fixed (large) prime number.

6.1 Non-vanishing in harmonic average

As described in Section 2.3, we will first prove the "harmonic" version of (6.1).

Theorem 15. *For any $\varepsilon > 0$, and for any prime number q large enough in terms of ε*

$$\sum_{\substack{\varepsilon_f = -1 \\ L'(f,\frac{1}{2}) \neq 0}}^{h} 1 \geqslant \Big(\frac{19}{54} - \varepsilon\Big). \tag{6.2}$$

6.1.1 Preliminary: a refined statement

We consider a mollified first moment

$$M_1 = \sum_{f \in S_2(q)^*}^{h} \varepsilon_f^- M(f) L'(f, \tfrac{1}{2})$$

and a mollified second moment

$$M_2 = \sum_{f \in S_2(q)^*}^{h} \varepsilon_f^- |M(f) L'(f, \tfrac{1}{2})|^2$$

where the complex numbers $M(f)$ (the "mollifier") are at our disposal. Comparison of an upper bound for M_2 and a lower bound for M_1 yields a lower bound for the quantity in (6.2): by Cauchy's inequality

$$M_1 \leqslant \Big(\sum_{\substack{\varepsilon_f = -1 \\ L'(f,\frac{1}{2}) \neq 0}}^{h} 1 \Big)^{1/2} M_2^{1/2}$$

so that if the mollifier is such that $M_2 \neq 0$ we have

$$\sum_{\substack{\varepsilon_f=-1\\ L'(f,\frac{1}{2})\neq 0}}^{h} 1 \geqslant \frac{M_1^2}{M_2}. \tag{6.3}$$

In order to achieve the best possible estimate, we will seek asymptotics for M_1 and M_2. If the mollifier is ignored (take $M(f) = 1$), a factor $\log q$ is lost in the final estimate.

To make the sums manageable, and in accordance with the principle that the mollifier should dampen the effect of large values of $L'(f, \frac{1}{2})$, we choose $M(f)$ of the shape (compare (5.17))

$$M(f) = \sum_{m\leqslant M} \frac{x_m}{\sqrt{m}}\lambda_f(m) \tag{6.4}$$

for real numbers (x_m) (and a parameter $M > 0$) which we will try to choose to optimize the resulting bound (6.3). If $m > M$, it is understood that $x_m = 0$. Now we only impose that the x_m be supported on squarefree integers[1] and satisfy

$$x_m \ll \big(\tau(m)(\log qm)\big)^A \tag{6.5}$$

for some absolute constant $A > 0$. Henceforth we write $M = q^\Delta$, and will assume $0 \leqslant \Delta < \frac{1}{2}$, so any m appearing in the mollifier, or any product m_1m_2, is less than q, hence coprime with q since q is prime. These assumption imply the growth estimate

$$\sum_{m\leqslant M} |x_m| \ll (\log q)^{A+2^A-1}M. \tag{6.6}$$

The precise form of Theorem 15 that we will prove is

Theorem 16. *Let $\mathfrak{m}_1$, $\mathfrak{m}_{21}$, $\mathfrak{m}_{22}$ and $\mathfrak{m}_2$ be the real polynomials in the indeterminate Δ given by*

$$\mathfrak{m}_1 = \Delta\Big(\frac{\Delta^2}{2} + \frac{\Delta}{2} + \frac{1}{4}\Big) \tag{6.7}$$

$$\begin{aligned}\mathfrak{m}_{21} = \mathfrak{m}_1 &+ 6\Delta^2\Big((\tfrac{1}{2}+\Delta)^2 - \Delta(\tfrac{1}{2}+\Delta) + \frac{\Delta^2}{4}\Big)\\ &+ 6\Delta^3\Big(\frac{4}{3}(\tfrac{1}{2}+\Delta)^2 - \Delta(\tfrac{1}{2}+\Delta) + \frac{\Delta^2}{5}\Big)\\ &+ 2\Delta^3\Big((\tfrac{1}{2}+\Delta)^2 - \Delta(\tfrac{1}{2}+\Delta) + \frac{\Delta^2}{5}\Big)\\ &+ 2\Delta^4\Big(\frac{3}{2}(\tfrac{1}{2}+\Delta)^2 - \Delta(\tfrac{1}{2}+\Delta) + \frac{\Delta^2}{6}\Big)\end{aligned} \tag{6.8}$$

$$\begin{aligned}\mathfrak{m}_{22} = &-\frac{2}{3}\Delta^3\Big((\tfrac{1}{2}+\Delta)^2 - \Delta(\tfrac{1}{2}+\Delta) + \frac{\Delta^2}{5}\Big)\\ &- \frac{2}{3}\Delta^4\Big(\frac{3}{2}(\tfrac{1}{2}+\Delta)^2 - \Delta(\tfrac{1}{2}+\Delta) + \frac{\Delta^2}{6}\Big)\end{aligned} \tag{6.9}$$

$$\mathfrak{m}_2 = \frac{1}{12}\mathfrak{m}_{21} - \frac{1}{4}\mathfrak{m}_{22}. \tag{6.10}$$

[1]This is convenient, although the assumption could be dispensed with, but only to be recovered as a consequence of the choice of x_m.

Then the function

$$\Delta \mapsto \frac{\mathfrak{m}_1(\Delta)^2}{\mathfrak{m}_2(\Delta)}$$

is defined and increasing on the real segment $[0, \frac{1}{2}[$ *and for all* $\Delta \in [0, \frac{1}{4}[$ *it holds*

$$\sum_{\substack{\varepsilon_f=-1 \\ L'(f,\frac{1}{2})\neq 0}}^{h} 1 \geqslant \frac{\mathfrak{m}_1(\Delta)^2}{\mathfrak{m}_2(\Delta)} + O\Big(\frac{(\log\log q)^4}{\log q}\Big).$$

Moreover

$$\frac{\mathfrak{m}_1(\frac{1}{4})^2}{\mathfrak{m}_2(\frac{1}{4})} = \frac{19}{54}.$$

Clearly, this implies Theorem 15. We now start proving this result.

6.1.2 Computation of the first moment

Recall that

$$M_1 = \sum_{f}^{h} \varepsilon_f^- M(f) L'(f, \tfrac{1}{2}).$$

As in the previous chapter we express $L'(f, \frac{1}{2})$ as a rapidly convergent series using contour integration and the functional equation. Choose an integer $N \geqslant 1$ (which will have to be large enough, $N = 2$ works already) and a real polynomial G satisfying

$$G(-s) = G(s), \text{ and } G(0) = 1 \tag{6.11}$$

$$G(-N) = \ldots = G(-1) = 0. \tag{6.12}$$

Notice that from the first of these, we obtain also

$$G'(0) = 0,\ G^{(3)}(0) = 0. \tag{6.13}$$

Then consider the integral

$$I = \frac{1}{2i\pi} \int_{(2)} \Lambda(f, s + \tfrac{1}{2}) G(s) \frac{ds}{s^2}.$$

In I, the Gamma factor $\Gamma(s+1)$ involved in $\Lambda(f, s+\frac{1}{2})$ decreases exponentially in vertical strips, whereas both the L-function and the polynomial G have at most polynomial growth, making it possible to shift the contour of integration to the left, to the line $\mathrm{Re}\,(s) = -2$. A double pole at $s = 0$ is picked up, hence

$$I = \mathrm{Res}_{s=0} \frac{\Lambda(f, s+\frac{1}{2}) G(s)}{s^2} + \frac{1}{2i\pi} \int_{(-2)} \Lambda(f, s+\tfrac{1}{2}) G(s) \frac{ds}{s^2},$$

but, by applying the functional equation and (6.11), the integral on $\mathrm{Re}\,(s) = -2$ is equal to $\varepsilon_f I$, so this yields

$$2\varepsilon_f^- I = \mathrm{Res}_{s=0} \frac{\Lambda(f, s+\frac{1}{2}) G(s)}{s^2}.$$

The residue is computed by writing the Taylor expansions

$$\Lambda(f, s+\tfrac{1}{2}) = \Lambda(f, \tfrac{1}{2}) + s\Lambda'(f, \tfrac{1}{2}) + O(s^2)$$

and (from (6.13), (6.11))

$$G(s) = 1 + O(s^2)$$

so that we derive

$$2\varepsilon_f^- I = \Lambda'(f, \tfrac{1}{2})$$

whence, multiplying through by ε_f^-

$$2\varepsilon_f^- I = \varepsilon_f^- \Big(\frac{\sqrt{q}}{2\pi}\Big)^{1/2} L'(f, \tfrac{1}{2}).$$

On the other hand we can compute I by expanding the L-function in an absolutely convergent Dirichlet series on the line $\mathrm{Re}\,(s) = 2$,

$$\begin{aligned} I &= \frac{1}{2i\pi} \int_{(2)} L(f, s+\tfrac{1}{2}) \Big(\frac{\sqrt{q}}{2\pi}\Big)^{s+1/2} \Gamma(s+1) G(s) \frac{ds}{s^2} \\ &= \Big(\frac{\sqrt{q}}{2\pi}\Big)^{1/2} \sum_{l \geqslant 1} \lambda_f(l) l^{-1/2} \frac{1}{2i\pi} \int_{(2)} n^{-s} \Big(\frac{\sqrt{q}}{2\pi}\Big)^{s} \Gamma(s+1) G(s) \frac{ds}{s^2} \\ &= \Big(\frac{\sqrt{q}}{2\pi}\Big)^{1/2} \sum_{l \geqslant 1} \lambda_f(l) l^{-1/2} V\Big(\frac{2\pi l}{\sqrt{q}}\Big) \end{aligned}$$

where the function V is defined by

$$V(y) = \frac{1}{2i\pi} \int_{(3/2)} \Gamma(s+1) G(s) y^{-s} \frac{ds}{s^2}. \tag{6.14}$$

Putting both computations together we get the desired expression

$$\varepsilon_f^- L'(f, \tfrac{1}{2}) = 2\varepsilon_f^- \sum_{l \geqslant 1} \lambda_f(l) l^{-1/2} V\Big(\frac{2\pi l}{\sqrt{q}}\Big). \tag{6.15}$$

We estimate V easily by shifting the contour to the left, or right.

Lemma 26. *The function V satisfies*

$$V(y) = -\log y - \gamma + O(y^N) \tag{6.16}$$

($\gamma = -\Gamma'(1)$ is Euler's constant) and

$$V(y) \ll_j y^{-j} \quad \text{for all } j \geqslant 1 \tag{6.17}$$

(which are all valid for $y > 0$).

Using the definition of M_1 and the expression (6.4) for $M(f)$, we obtain at once from (6.15)

$$M_1 = \sum_{l,m} x_m (lm)^{-1/2} V\Big(\frac{2\pi l}{\sqrt{q}}\Big) \times \Delta_-(l,m) \tag{6.18}$$

where Δ_- is the Delta symbol for odd forms considered in Section 4.2, namely

$$\Delta_-(l,m) = 2\sum_f^h \varepsilon_f^- \lambda_f(l)\lambda_f(m).$$

By (4.10) of Lemma 11, Δ_- approximates the Kronecker delta

$$\Delta_-(l,m) = \delta(l,m) + O\Big(\frac{\sqrt{lm}}{q}(\log q)^2\Big)$$

and thus we have

$$\begin{aligned} M_1 &= \sum_m \frac{x_m}{m} V\Big(\frac{2\pi m}{\sqrt{q}}\Big) + O\Big(\frac{(\log q)^2}{q}\sum_{l,m}|x_m|\Big|V\Big(\frac{2\pi l}{\sqrt{q}}\Big)\Big|\Big) \\ &= \sum_m \frac{x_m}{m} V\Big(\frac{2\pi m}{\sqrt{q}}\Big) + O\Big(\frac{M}{\sqrt{q}}(\log q)^B\Big) \end{aligned}$$

for some $B > 0$, depending only on the constant A of (6.5), since

$$\begin{aligned} \frac{(\log q)^2}{q}\sum_{l,m}|x_m|\Big|V\Big(\frac{2\pi l}{\sqrt{q}}\Big)\Big| &\ll \frac{M}{q}(\log q)^B\Big\{\sum_{l<\sqrt{q}}\Big|V\Big(\frac{2\pi l}{\sqrt{q}}\Big)\Big| + \sum_{l\geqslant\sqrt{q}}\Big|V\Big(\frac{2\pi l}{\sqrt{q}}\Big)\Big|\Big\} \\ &\ll \frac{M}{q}(\log q)^B \times (\sqrt{q}(\log q) + \sqrt{q}) \end{aligned}$$

by (6.6) and Lemma 26, in particular, (6.12) with $j = 2$ for $l > \sqrt{q}$. Using now (6.16), with $N = 1$, and again (6.6), we have

$$\begin{aligned} \sum_m \frac{x_m}{m} V\Big(\frac{2\pi m}{\sqrt{q}}\Big) &= \sum_{m\leqslant M}\frac{x_m}{m}\Big(\log\frac{\sqrt{q}}{2\pi m} - \gamma\Big) + O\Big(\frac{1}{\sqrt{q}}\sum_{m\leqslant M}|x_m|\Big) \\ &= \sum_{m\leqslant M}\frac{x_m}{m}\Big(\log\frac{\sqrt{q}}{2\pi m} - \gamma\Big) + O\Big(\frac{M}{\sqrt{q}}(\log q)^B\Big) \end{aligned}$$

which establishes the next proposition.

Proposition 15. *Let $M = q^\Delta$ with $\Delta < \frac{1}{2}$. Define the real number $\hat{q}$ by*

$$\log\hat{q} = -\log\frac{2\pi}{\sqrt{q}} - \gamma,$$

then, for some positive constant $\delta = \delta(\Delta) > 0$, we have

$$M_1 = \sum_{m\leqslant M}\frac{x_m}{m}\Big(\log\frac{\hat{q}}{m}\Big) + O(q^{-\delta}). \tag{6.19}$$

Remark . Here one can take, by the above,

$$\delta = \tfrac{1}{2} - \Delta - \varepsilon$$

for any $\varepsilon > 0$ small enough so that this is positive. In the following, when we write an error term of the form $O(q^{-\delta})$, it is implied that $\delta > 0$, δ only depends on Δ, and the value of δ may change from line to line.

6.1.3 Computation of the second moment

We now wish to get an expression for M_2 as a quadratic form in the x_m. A new phenomenon appears, however, at the point where we would like to appeal to Lemma 11, as the remainder term in the Petersson formula (the series of Kloosterman sums $\mathcal{J}(m,n)$) cannot be ignored, and has to be analyzed to yield a contribution to the main term.

The square of the special value

The procedure is similar to that followed to express $L'(f,\frac{1}{2})$. We consider this time the integral

$$J = \frac{1}{2i\pi}\int\limits_{(2)} \Lambda(f, s+\tfrac{1}{2})^2 G(s)\frac{ds}{s^3}$$

and proceed to evaluate it as before. By shifting the contour to $\mathrm{Re}\,(s) = -2$ and applying the functional equation for the square of the L-function

$$\Lambda(f,s)^2 = \Lambda(f,1-s)^2$$

(notice the sign is always $+1$ in this case), we have

$$2J = \mathrm{Res}_{s=0}\frac{\Lambda(f, s+\frac{1}{2})^2 G(s)}{s^3}.$$

Further, from the multiplicativity of the Hecke eigenvalues (2.10), we derive the Dirichlet series expansion

$$L(f,s)^2 = \zeta_q(2s)\sum_{n\geqslant 1}\tau(n)\lambda_f(n)n^{-s}$$

so the term by term integration yields

$$J = \frac{\sqrt{q}}{2\pi}\sum_{n\geqslant 1}\frac{\lambda_f(n)}{\sqrt{n}}\tau(n)W\Big(\frac{4\pi^2 n}{q}\Big)$$

where

$$W(y) = \frac{1}{2i\pi}\int\limits_{(1/2)} \zeta_q(1+2s)\Gamma(s)^2 G(s) y^{-s}\frac{ds}{s} \tag{6.20}$$

(the integration on $\mathrm{Re}\,(s) = 2$ can be shifted to $\mathrm{Re}\,(s) = \frac{1}{2}$ in defining the function W since no poles appear between those two lines and the integrand is exponentially decreasing in vertical strips). Comparing we deduce the equality

$$2\times\frac{\sqrt{q}}{2\pi}\sum_{n\geqslant 1}\frac{\lambda_f(n)}{\sqrt{n}}\tau(n)W\Big(\frac{4\pi^2 n}{q}\Big) = \mathrm{Res}_{s=0}\frac{\Lambda(f, s+\frac{1}{2})^2 G(s)}{s^3}. \tag{6.21}$$

Now if f is odd, we have $L(f,\frac{1}{2}) = \Lambda(f,\frac{1}{2}) = 0$ and then we compute the residue

$$\mathrm{Res}_{s=0}\frac{\Lambda(f, s+\frac{1}{2})^2 G(s)}{s^3} = \Lambda'(f,\tfrac{1}{2})^2 = \frac{\sqrt{q}}{2\pi}L'(f,\tfrac{1}{2})^2$$

from the Taylor expansions

$$\Lambda(f, s+\tfrac{1}{2})^2 = s^2\Lambda'(f,\tfrac{1}{2})^2 + O(s^3)$$
$$G(s) = 1 + \frac{s^2}{2}G''(0) + O(s^4),$$

so (6.21) furnishes a formula for $L'(f,\frac{1}{2})$, valid for odd forms:

$$L'(f,\tfrac{1}{2})^2 = 2\sum_{n\geqslant 1}\frac{\lambda_f(n)}{\sqrt{n}}\tau(n)W\Big(\frac{4\pi^2 n}{q}\Big). \tag{6.22}$$

For our purpose, W is basically a "cut-off" function, which restricts the summation to $n \leqslant q$. Indeed, we have the following

Lemma 27. *The function W satisfies*

$$y^i W^{(j)}(y) \ll_{i,j} \log(y+1/y)^3, \quad \text{for all } i \geqslant j \geqslant 0 \tag{6.23}$$
$$y^i W^{(i)}(y) \ll_j y^{-j}, \quad \text{for all } i \geqslant 0,\ j \geqslant 1. \tag{6.24}$$

Moreover, there exists a polynomial P, independent of q, of degree at most 2, such that

$$W(y) = -\frac{1}{12}(\log y)^3 + P(\log y) + O(q^{-1}(\log y)^2 + y^N). \tag{6.25}$$

Proof. The first two inequalities are obtained by the usual contour shifts and differentiating under the integral sign. As for the last, we have

$$W(y) = \mathrm{Res}_{s=0}\frac{G(s)\Gamma(s)^2\zeta_q(1+2s)y^{-s}}{s} + O(y^N)$$

again by shifting, and the residue is computed by using Taylor expansions, writing

$$\zeta_q(1+2s) = (1-q^{-1-2s})\zeta(2s)$$

where the first factor contributes to the error term in (6.25), hence we get the polynomial P independent of q. □

As for V before, we will use this lemma in the following form:

$$\sum_{n\geqslant 1} a(n)W\Big(\frac{4\pi^2 n}{q}\Big) \ll (\log q)^3\sum_{n<q}|a(n)| + q^2\sum_{n\geqslant q}\frac{|a(n)|}{n^2} \tag{6.26}$$

so, roughly speaking, if the complex numbers $a(n)$ are "almost bounded", the sum will be of size about q, as if the summation had extended only to $n \leqslant q$.

Applying Petersson's formula

The goal now is to compute M_2. From the definition (6.4) of $M(f)$, and by multiplicativity, the second moment M_2 can be written

$$M_2 = \sum_{\varepsilon_f=-1}^{h} L'(f,\tfrac{1}{2})^2 \sum_{m_1,m_2}\frac{x_{m_1}x_{m_2}}{\sqrt{m_1m_2}}\lambda_f(m_1)\lambda_f(m_2)$$
$$= \sum_b \frac{1}{b}\sum_{m_1,m_2}\frac{x_{bm_1}x_{bm_2}}{\sqrt{m_1m_2}}A^h[\varepsilon_f^-\lambda_f(m_1m_2)L'(f,\tfrac{1}{2})^2] \tag{6.27}$$

(there is no $\varepsilon_q(b)$ because $m < q$ by assumption).

We start by investigating the inner sum. Therefore fix some $0 \leqslant \Delta < 1$ and an integer m, $1 \leqslant m \leqslant q^\Delta < q$. We consider the average

$$A^h[\varepsilon_f^- \lambda_f(m) L'(f, \tfrac{1}{2})^2] = \sum_{f \in S_2(q)^*}^h \varepsilon_f^- \lambda_f(m) L'(f, \tfrac{1}{2})^2. \tag{6.28}$$

From (6.22) we obtain

$$\begin{aligned} A^h[\varepsilon_f^- \lambda_f(m) L'(f, \tfrac{1}{2})^2] &= \sum_{n \geqslant 1} \frac{\tau(n)}{\sqrt{n}} W\Big(\frac{4\pi^2 n}{q}\Big) A^h[2\varepsilon_f^- \lambda_f(m)\lambda_f(n)] \\ &= \sum_{n \geqslant 1} \frac{\tau(n)}{\sqrt{n}} W\Big(\frac{4\pi^2 n}{q}\Big) \Delta_-(m, n). \end{aligned}$$

From Lemma 11, we have

$$\Delta_-(m, n) = \delta(m, n) + \mathcal{J}'(m, n) + O\Big(\frac{\sqrt{mn}}{q^{3/2}} (\log q)^2\Big) \tag{6.29}$$

hence

$$\begin{aligned} A^h[\varepsilon_f^- \lambda_f(m) L'(f, \tfrac{1}{2})^2] = \frac{\tau(m)}{\sqrt{m}} W\Big(\frac{4\pi^2 m}{q}\Big) + \sum_{n \geqslant 1} \frac{\tau(n)}{\sqrt{n}} \mathcal{J}'(m, n) W\Big(\frac{4\pi^2 n}{q}\Big) \\ + O\Big(\sqrt{\frac{m}{q}} (\log q)^6\Big) \end{aligned} \tag{6.30}$$

where we have estimated that the error term is

$$\ll \frac{\sqrt{m}}{q^{3/2}} (\log q)^2 \sum_{n \geqslant 1} \tau(n) \Big| W\Big(\frac{4\pi^2 n}{q}\Big)\Big|$$

and the inner sum is estimated by the method of (6.26), as follows

$$\begin{aligned} \sum_{n \geqslant 1} \tau(n) \Big| W\Big(\frac{4\pi^2 n}{q}\Big)\Big| &\ll (\log q)^3 \sum_{n < q} \tau(n) + q^2 \sum_{n \geqslant q} \frac{\tau(n)}{n^2} \\ &\ll q (\log q)^4 \end{aligned}$$

whence (6.30). Notice already that if $M = q^\Delta$ with $\Delta < \frac{1}{4}$, this error term is good enough to put back into M_2 where – as in M_1 – we expect a main term which is roughly a power of $\log q$: by (6.27) it yields a contribution to M_2 which is

$$\ll \frac{(\log q)^6}{\sqrt{q}} \sum_{b \leqslant M} \frac{1}{b} \Big| \sum_m x_m \Big|^2 \ll \frac{M^2}{\sqrt{q}} (\log q)^C$$

for some $C > 0$ by (6.6), and this is $\ll q^\delta$ for some $\delta > 0$ if $\Delta < \frac{1}{4}$.

The first term of (6.27) is further decomposed by means of (6.25) (with $N = 1$): define $\hat{Q}$ by

$$\log \hat{Q} = \log \frac{q}{4\pi^2},$$

then

$$\frac{\tau(m)}{\sqrt{m}}W\Big(\frac{4\pi^2 m}{q}\Big) = \frac{1}{12}\frac{\tau(m)}{\sqrt{m}}\Big(\log\frac{\hat{Q}}{m}\Big)^3 + \frac{\tau(m)}{\sqrt{m}}P\Big(\log\frac{\hat{Q}}{m}\Big) + O\Big(\frac{m}{q}\Big). \tag{6.31}$$

The second term in (6.30) is now the focus of our attention. We call it $X(m)$. From the definition (4.8) of $\mathcal{J}'(m,n)$, this is a sum over integers $r \geqslant 1$ coprime with q, which we write

$$X(m) = \frac{2\pi}{\sqrt{q}}\sum_{(r,q)=1}\frac{1}{r}X_r \tag{6.32}$$

where the term X_r is a weighted sum of Kloosterman sums twisted by the divisor function

$$X_r = -\sum_{n\geqslant 1}\frac{\tau(n)}{\sqrt{n}}S(m\overline{q},n;r)J_1\Big(\frac{4\pi}{r}\sqrt{\frac{mn}{q}}\Big)W\Big(\frac{4\pi^2 n}{q}\Big)\xi(n). \tag{6.33}$$

For technical reasons we have chosen a fixed test function $\xi : \mathbf{R}^+ \longrightarrow [0,1]$, which is C^∞ and satisfies

$$\xi(x) = 0,\ 0 \leqslant x \leqslant \tfrac{1}{2}, \quad \xi(x) = 1,\ x \geqslant 1,$$

and we have inserted in the summation the weight $\xi(n)$: this obviously doesn't affect the sum (all positive integers are at least 1!), but will be useful to gain convergence in some series appearing later (this is required only because the weight is 2).

We separate the sum in r in (6.32) in two parts, $r \leqslant R$ and $r > R$, $R > 0$ being a parameter to be fixed later, but assumed to satisfy $\log R \ll \log q$ with some absolute implied constant (see the choice below). The part with r large is handled directly.

Lemma 28. *In this situation, it holds*

$$\frac{2\pi}{\sqrt{q}}\sum_{\substack{r>R\\(r,q)=1}}\frac{1}{r}X_r \ll \sqrt{\frac{m}{R}}(\log q)^{11}. \tag{6.34}$$

Proof. The computations are similar to those previously done. For Kloosterman sums we use Weil's bound (4.5) and for the Bessel function $J_1(x) \ll x$, and the sum over n is subdivided into $n < q$ and $n \geqslant q$, and accordingly (6.23) or (6.24) is used. Quickly, for instance the part with $n \leqslant q$ is

$$\ll \frac{\sqrt{m}}{q}(\log q)^3\sum_{n<q}\tau(n)\sum_{r>R}\frac{\tau(r)(m,n,r)^{1/2}}{r^{3/2}}$$

and the common divisor is handled by writing

$$(m,n,r)^{1/2} \leqslant (m,n,r) = \sum_{\substack{d|(m,n)\\ d|r}}\varphi(d)$$

so

$$\begin{aligned}\sum_{r>R}\frac{\tau(r)(m,n,r)^{1/2}}{r^{3/2}} &\leqslant \sum_{d|(m,n)}\frac{\tau(d)\varphi(d)}{d^{3/2}}\sum_{rd>R}\frac{\tau(r)}{r^{3/2}}\\ &\ll \frac{1}{\sqrt{R}}(\log q)\sum_{d|(m,n)}\frac{\tau(d)\varphi(d)}{d}\\ &\ll \frac{1}{\sqrt{R}}(\log q)\tau(n)^2.\end{aligned}$$

The part with $n \geqslant q$ is handled similarly, and both together produce the error term announced in (6.34). $\square$

We denote now $X'(m)$ the remaining part of the sum in $X(m)$:

$$X'(m) = \frac{2\pi}{\sqrt{q}} \sum_{\substack{r \leqslant R \\ (r,q)=1}} \frac{1}{r} X_r.$$

Extraction of the main contribution

Let now $r < R$. In X_r (see (6.33)), we now open the Kloosterman sum

$$S(m\overline{q}, n; r) = \sum_{d \bmod r}^{*} e\Big(\frac{m\overline{qd} + nd}{r}\Big)$$

and take the summation over d outside, so

$$X_r = -\sum_{d \bmod r}^{*} e\Big(\frac{m\overline{qd}}{r}\Big) \sum_{n \geqslant 1} \tau(n) e\Big(\frac{nd}{r}\Big) t(n)$$

where the weight function $t : \mathbf{R}^+ \longrightarrow \mathbf{R}$ is

$$t(x) = J_1\Big(\frac{4\pi}{r}\sqrt{\frac{mx}{q}}\Big) W\Big(\frac{4\pi^2 x}{q}\Big) \frac{\xi(x)}{\sqrt{x}}. \tag{6.35}$$

For each d, the summation formula for the divisor function twisted by additive characters (an extension of Voronoï's formula, see [Jut, th. 1.7] for instance) can be applied.

Proposition 16. *Let $t : \mathbf{R}^+ \to \mathbf{C}$ be a C^∞ function which vanishes in the neighborhood of 0 and is rapidly decreasing at infinity. Then for $c \geqslant 1$ and d coprime with c, we have*

$$\begin{aligned}\sum_{m \geqslant 1} \tau(m) e\Big(\frac{dm}{c}\Big) t(m) = \frac{2}{c} \int_0^{+\infty} (\log \frac{\sqrt{x}}{c} + \gamma) t(x) dx \\ - \frac{2\pi}{c} \sum_{h \geqslant 1} \tau(h) e\Big(-\frac{\overline{d}h}{c}\Big) \int_0^{+\infty} Y_0\Big(\frac{4\pi\sqrt{hx}}{c}\Big) t(x) dx \\ + \frac{4}{c} \sum_{h \geqslant 1} \tau(h) e\Big(\frac{\overline{d}h}{c}\Big) \int_0^{+\infty} K_0\Big(\frac{4\pi\sqrt{hx}}{c}\Big) t(x) dx.\end{aligned}$$

Hence, after exchanging again the order of summation, this yields

$$\begin{aligned}X_r = -\frac{2}{r} S(m, 0; r) \int_0^{\infty} (\log \frac{\sqrt{x}}{r} + \gamma) t(x) dx \\ + \frac{2\pi}{r} \sum_{h \geqslant 1} \tau(h) S(hq - m, 0; r) \int_0^{+\infty} Y_0\Big(\frac{4\pi\sqrt{hx}}{r}\Big) t(x) dx \\ - \frac{4}{r} \sum_{h \geqslant 1} \tau(h) S(hq + m, 0; r) \int_0^{+\infty} K_0\Big(\frac{4\pi\sqrt{hx}}{r}\Big) t(x) dx.\end{aligned} \tag{6.36}$$

Let $L(r)$ be the first integral without the test function $\xi(x)$:

$$L(r) = \int_0^\infty (\log\frac{\sqrt{x}}{r} + \gamma)J_1\Big(\frac{4\pi}{r}\sqrt{\frac{mx}{q}}\Big)W\Big(\frac{4\pi^2 x}{q}\Big)\frac{dx}{\sqrt{x}}$$

and for any integer $h \geqslant 1$ let

$$y(h) = \int_0^{+\infty} Y_0\Big(\frac{4\pi\sqrt{hx}}{r}\Big)t(x)dx \tag{6.37}$$

$$k(h) = \int_0^{+\infty} K_0\Big(\frac{4\pi\sqrt{hx}}{r}\Big)t(x)dx \tag{6.38}$$

With these definitions, and by removing the weight ξ in the first term of the summation formula, we get

$$\begin{aligned} X'(m) = {} & \frac{4\pi}{\sqrt{q}} \sum_{\substack{r\leqslant R\\(r,q)=1}} \frac{1}{r^2}S(m,0;r)L(r) \\ & + \frac{4\pi^2}{\sqrt{q}} \sum_{\substack{r\leqslant R\\(r,q)=1}} \frac{1}{r^2}\sum_{h\geqslant 1}\tau(h)S(hq-m,0;r)y(h) \\ & - \frac{8\pi}{\sqrt{q}} \sum_{\substack{r\leqslant R\\(r,q)=1}} \frac{1}{r^2}\sum_{h\geqslant 1}\tau(h)S(hq+m,0;r)k(h) + O\Big(\frac{(\log q)^5}{\sqrt{q}}\Big) \end{aligned} \tag{6.39}$$

because the difference arising from this removal is at most

$$\ll \frac{1}{\sqrt{q}}\sum_{r<R}\frac{1}{r^2}|S(m,0;r)|\int_0^1 |\log\frac{\sqrt{x}}{r} + \gamma||t(x)|dx \ll \frac{(\log q)^5}{\sqrt{q}}$$

since $t(x) \ll (\log q)^3 x^{-1/2}$ for $0 \leqslant x \leqslant 1$ by (6.23), and crudely $|S(m,0;r)| \leqslant r$.

We reserve for later consideration the last two sums (see Section 6.1.3), and evaluate exactly the first one, which we call $X''(m)$. We have

$$\begin{aligned} X''(m) &= -\frac{4\pi}{\sqrt{q}} \sum_{\substack{r\leqslant R\\(r,q)=1}} \frac{1}{r^2}S(m,0;r)\int_0^\infty (\log\frac{\sqrt{x}}{r} + \gamma)J_1\Big(\frac{4\pi}{r}\sqrt{\frac{mx}{q}}\Big)W\Big(\frac{4\pi^2 x}{q}\Big)\frac{dx}{\sqrt{x}} \\ &= -2\sum_{\substack{r\leqslant R\\(r,q)=1}} \frac{1}{r}S(m,0;r)\int_0^\infty (\log\frac{\sqrt{qx}}{2\pi} + \gamma)J_1(2\sqrt{mx})W(r^2x)\frac{dx}{\sqrt{x}} \end{aligned}$$

by the change of variable $x \mapsto \dfrac{r^2}{4\pi^2}qy$. By the definition of W, see (6.20), this is equal to the complex integral

$$X''(m) = \frac{1}{2i\pi}\int_{(1/2)} (-2)Z_m^R(1+2s)\zeta_q(1+2s)s^{-1}\Gamma(s)^2 G(s)L(s)ds, \tag{6.40}$$

with

$$Z_m^R(s) = \sum_{\substack{r\leqslant R\\(r,q)=1}} S(m,0;r)r^{-s},$$

$$L(s) = \int_0^{+\infty} (\log\frac{\sqrt{qx}}{2\pi}+\gamma)J_1(2\sqrt{mx})x^{-s-1/2}dx.$$

Both Z_m^R and L can be computed, the former by extending again the sum over r to infinity.

Lemma 29. *We have for* $\mathrm{Re}\,(s)=\sigma>1$

$$Z_m^R(s) = \zeta_q(s)^{-1}\sum_{d|m} d^{1-s} + O_\sigma(\tau(m)R^{1-\sigma}).$$

Proof. By the formula giving the Ramanujan sum (the $\sum^*$ refers of course to integers coprime with q)

$$\begin{aligned} Z_m^R(s) &= \sum_{r\leqslant R}^* r^{-s}\sum_{d|(m,r)} d\mu\Big(\frac{r}{d}\Big)\\ &= \sum_{d|m} d\sum_{fd\leqslant R}^* \mu(f)(fd)^{-s}\\ &= \sum_{d|m} d^{1-s}\Big\{\zeta_q(s)^{-1}+O\Big(\Big(\frac{R}{d}\Big)^{1-\sigma}\Big)\Big\}\\ &= \zeta_q(s)^{-1}\sum_{d|m} d^{1-s}+O(\tau(m)R^{1-\sigma}) \end{aligned}$$

□

Lemma 30. *Recall that* $\log\hat{Q}=\log\frac{q}{4\pi^2}$. *For all* s *with* $\frac14<\mathrm{Re}\,(s)<1$, *we have*

$$L(s) = -\frac12 m^{s-1/2}\Gamma(-s)\Gamma(s)^{-1}\Big(\log\frac{\hat{Q}}{m}+2\gamma+\psi(1+s)+\psi(1-s)\Big)$$

where $\psi=\Gamma'/\Gamma$.

Proof. The following formula is valid for $-2<\mathrm{Re}\,(s)<-\frac12$ (see [G-R, 6.561.14]):

$$\ell(s) := \int_0^{+\infty} J_1(x)x^s dx = 2^s\Gamma\Big(1+\frac{s}{2}\Big)\Gamma\Big(1-\frac{s}{2}\Big)^{-1} \tag{6.41}$$

and putting $y=2\sqrt{mx}$ in $L(s)$ gives

$$L(s) = 4^s m^{s-1/2}\Big((\frac12\log\frac{\hat{Q}}{m}+\gamma)\ell(-2s)+\ell'(-2s)\Big).$$

From (6.41) we deduce

$$\ell'(s) = 2^s\Gamma\Big(1+\frac{s}{2}\Big)\Gamma\Big(1-\frac{s}{2}\Big)^{-1}\Big(\log 2+\frac12\psi\Big(1+\frac{s}{2}\Big)+\frac12\psi\Big(1-\frac{s}{2}\Big)\Big)$$

and the result follows. □

From Lemma 29 we obtain

$$X''(m) = \frac{1}{2i\pi} \int\limits_{(1/2)} (-2)\sigma_{-2s}(m)s^{-1}\Gamma(s)^2 G(s)L(s)ds + O\Big(\frac{\tau(m)}{R}(\log q)\Big)$$
$$= \frac{1}{2i\pi} \int\limits_{(1/2)} F(s)ds + O\Big(\frac{\tau(m)}{R}(\log q)\Big), \quad \text{say} \tag{6.42}$$

since $1+2s$ is on the line $\mathrm{Re}\,(s) = 2$, and

$$\zeta_q(1+2s)\Gamma(s)^2 G(s)L(s) \ll |\Gamma(s)\Gamma(-s)s^{-1}|(\log q + |\psi(1+s)| + |\psi(1-s)|)$$

on $\mathrm{Re}\,(s) = \frac{1}{2}$, and this decreases exponentially on the line.

The lemmas show that the integrand $F(s)$ in (6.42) is

$$F(s) = m^{-1/2}s^{-1}G(s)\eta_s(m)\Gamma(s)\Gamma(-s)\Big(\log\frac{\hat{Q}}{m} + 2\gamma + \psi(1+s) + \psi(1-s)\Big)$$

where η_s is the arithmetic function defined by

$$\eta_s(m) = \sum_{ab=m}\Big(\frac{a}{b}\Big)^s.$$

Thus, $F(s)$ is seen to be an *odd* function of s, which is moreover holomorphic in the strip $|\mathrm{Re}\,(s)| < 1$, except for a triple pole at $s = 0$, and decreases exponentially in vertical strips. Shifting the contour to $\mathrm{Re}\,(s) = -\frac{1}{2}$ and changing then s into $-s$ allows us to conclude that

$$X''(m) = \frac{1}{2}\mathrm{Res}_{s=0}F(s) + O\Big(\frac{\tau(m)}{R}(\log q)\Big). \tag{6.43}$$

Around $s = 0$, the following expansions hold:

$$s^{-1}\Gamma(s)\Gamma(-s) = -\frac{1}{s^3} + \frac{\gamma^2 - \Gamma''(1)}{s} + O(s)$$
$$2\gamma + \psi(1+s) + \psi(1-s) = \psi''(0)s^2 + O(s^4)$$
$$G(s) = 1 + \frac{1}{2}G''(0)s^2 + O(s^3)$$
$$\eta_s(m) = \tau(m) + \frac{1}{2}T(m)s^2 + O(s^3)$$

where T is the arithmetic function defined by

$$T(m) = \sum_{ab=m}\Big(\log\frac{a}{b}\Big)^2.$$

Combining those, we obtain

$$\frac{1}{2}\mathrm{Res}_{s=0}F(s) = -\frac{1}{4}\frac{T(m)}{\sqrt{m}}\Big(\log\frac{\hat{Q}}{m}\Big) + \alpha\frac{\tau(m)}{\sqrt{m}}\Big(\log\frac{\hat{Q}}{m}\Big), \tag{6.44}$$

where we have set

$$\alpha = \frac{1}{2}\Big(\gamma^2 - \Gamma''(1) - \frac{G''(0)}{2} - \psi''(0)\Big)$$

(a constant).

Coming back to the two other integrals contributing to X_r, hence to $X'(m)$, the following lemmas will be proved in Section 6.1.3, showing that they are of smaller order of magnitude than the term just evaluated. We denote those contributions by $Y(m)$ and $K(m)$:

$$Y(m) = \frac{4\pi^2}{\sqrt{q}} \sideset{}{^*}\sum_{r\leqslant R} \frac{1}{r^2} \sum_{h\geqslant 1} \tau(h) S(hq-m, 0; r) y(h) \tag{6.45}$$

and

$$K(m) = -\frac{8\pi}{\sqrt{q}} \sideset{}{^*}\sum_{r\leqslant R} \frac{1}{r^2} \sum_{h\geqslant 1} \tau(h) S(hq+m, 0; r) k(h). \tag{6.46}$$

Lemma 31. *For all $m < q$ we have*

$$K(m) \ll_\varepsilon \frac{\sqrt{m}}{q} q^\varepsilon$$

for all $\varepsilon > 0$.

Lemma 32. *Assume $R \leqslant q^2$. For all $m < q$ we have*

$$Y(m) \ll_\varepsilon \Big(\frac{\sqrt{m}}{q} + \frac{1}{\sqrt{q}}\Big) q^\varepsilon$$

for all $\varepsilon > 0$.

We choose $R = q^2$, and put together all the information gathered, thereby proving an approximate formula for $X(m)$: starting from (6.33), we have in turn

$$\begin{aligned}
X(m) &= X'(m) + O\Big(\sqrt{\frac{m}{R}} (\log q)^{11}\Big) \quad \text{by (6.34)} \\
&= X''(m) + K(m) + Y(m) + O\Big(\frac{(\log q)^5}{\sqrt{q}}\Big) + O\Big(\sqrt{\frac{m}{R}} (\log q)^{11}\Big) \quad \text{by (6.39)} \\
&= -\frac{1}{4} \frac{T(m)}{\sqrt{m}} \Big(\log \frac{\hat{Q}}{m}\Big) + \alpha \frac{\tau(m)}{\sqrt{m}} \Big(\log \frac{\hat{Q}}{m}\Big) + O\Big(\frac{\tau(m)}{R} (\log q)\Big) \quad \text{by (6.43), (6.44)} \\
&\quad + O_\varepsilon\Big(\frac{\sqrt{m}}{q} q^\varepsilon\Big) + O_\varepsilon\Big(\Big(\frac{\sqrt{m}}{q} + \frac{1}{\sqrt{q}}\Big) q^\varepsilon\Big) \quad \text{by Lemmas 32 and 31} \\
&\quad + O\Big(\frac{(\log q)^5}{\sqrt{q}}\Big) + O\Big(\sqrt{\frac{m}{R}} (\log q)^{11}\Big).
\end{aligned}$$

Cleaning up, we state this formally.

Proposition 17. *Let $0 \leqslant \Delta < 1$ and $1 \leqslant m \leqslant q^\Delta$. For any $\varepsilon > 0$*

$$X(m) = -\frac{1}{4} \frac{T(m)}{\sqrt{m}} \Big(\log \frac{\hat{Q}}{m}\Big) + \alpha \frac{\tau(m)}{\sqrt{m}} \Big(\log \frac{\hat{Q}}{m}\Big) + O_{\Delta,\varepsilon}\Big(\Big(\frac{\sqrt{m}}{q} + \frac{1}{\sqrt{q}}\Big) q^\varepsilon\Big).$$

Together with (6.31), this gives an approximate formula for $A^h[\varepsilon_f^- \lambda_f(m) L'(f, \frac{1}{2})]$.

Proposition 18. *Set $P_1 = P + \alpha X$. Then for $0 \leqslant \Delta < \frac{1}{2}$, and $1 \leqslant m \leqslant q^{\Delta}$, we have for any $\varepsilon > 0$,*

$$A^h[\varepsilon_f^- \lambda_f(m) L'(f, \tfrac{1}{2})] = \frac{1}{12}\frac{\tau(m)}{\sqrt{m}}\Big(\log\frac{\hat{Q}}{m}\Big)^3 - \frac{1}{4}\frac{T(m)}{\sqrt{m}}\Big(\log\frac{\hat{Q}}{m}\Big) + \frac{\tau(m)}{\sqrt{m}}P_1\Big(\log\frac{\hat{Q}}{m}\Big) + O_{\Delta,\varepsilon}\Big(\sqrt{\frac{m}{q}}q^{\varepsilon}\Big).$$

For later use, we record a few properties of the function T.

Lemma 33. *Let $\tau^{(i)}$ be defined for $i \geqslant 0$ by*

$$\tau^{(i)}(m) = \sum_{d|m} (\log d)^i.$$

Then we have

$$T(m) = 4\tau^{(2)}(m) - 2(\log m)\tau^{(1)}(m). \tag{6.47}$$

Moreover, T satisfies

$$T(m_1 m_2) = \tau(m_1)T(m_2) + \tau(m_2)T(m_1) \tag{6.48}$$

for $(m_1, m_2) = 1$ and more generally

$$T(m_1 m_2) = \sum_{d|(m_1,m_2)} \mu(d)\Big(\tau\Big(\frac{m_1}{d}\Big)T\Big(\frac{m_2}{d}\Big) + \tau\Big(\frac{m_2}{d}\Big)T\Big(\frac{m_1}{d}\Big)\Big) \tag{6.49}$$

for all integers m_1, $m_2 \geqslant 1$.

Proof. The first formula is immediate, and the second follows from the third, which is obtained by differentiating the corresponding identity (see (5.37)) for η_s, remarking that

$$\sum_{ab=m}\Big(\log\frac{a}{b}\Big) = 0$$

for any integer m. □

Estimation of the integrals

We now proceed to prove Lemmas 31 and 32, vindicating our contention that $K(m)$ and $Y(m)$ are of smaller order of magnitude (in our situation) than the main term isolated in the previous section.

We need some facts about the Bessel functions Y_0 (and Y_n, $n \geqslant 0$ an integer), J_n, and K_0, which we now quote:

$$K_0(y) \ll y^{-1/2}e^{-y}, \quad \text{for all } y > 0 \tag{6.50}$$

$$Y_0(y) \ll \log y \quad \text{for all } y > 0 \tag{6.51}$$

$$Y_\nu(y) \ll_\nu 1 \quad \text{for all } \nu \geqslant 1,\ y > 0. \tag{6.52}$$

Proof of Lemma 31. Because K_0 has exponential decay at infinity and ξ cuts off the small values of x, this is easy.

We start with $k(h)$, which is given by (6.38), so using (6.50) and $J_1(x) \ll 1$, we estimate

$$\begin{aligned} k(h) &= \int_0^{+\infty} K_0\Big(\frac{4\pi\sqrt{hx}}{r}\Big) J_1\Big(\frac{4\pi}{r}\sqrt{\frac{mx}{q}}\Big) W\Big(\frac{4\pi^2 x}{q}\Big)\xi(x)\frac{dx}{\sqrt{x}} \\ &= \frac{r}{\sqrt{h}}\int_0^{+\infty} K_0(y) J_1\Big(\sqrt{\frac{m}{hq}}y\Big) W\Big(\frac{r^2y^2}{4qh}\Big)\xi\Big(\frac{r^2y^2}{16\pi^2 h}\Big)dy \\ &\ll \frac{r}{h}\sqrt{\frac{m}{q}}(\log q)^3 \int_{\sqrt{h}r^{-1}}^{+\infty} y^{1/2}e^{-y}dy \\ &\ll \frac{r}{h}\sqrt{\frac{m}{q}}(\log q)^3 \exp\Big(-\frac{\sqrt{h}}{2r}\Big). \end{aligned}$$

Furthermore this implies

$$\begin{aligned} K(m) &= -\frac{8\pi}{\sqrt{q}}\sideset{}{^*}\sum_{r\leqslant R}\frac{1}{r^2}\sum_{h\geqslant 1}\tau(h)S(hq+m,0;r)k(h) \\ &\ll \frac{\sqrt{m}}{q}(\log q)^3\sum_{h\geqslant 1}\frac{\tau(h)}{h}\exp\Big(-\frac{\sqrt{h}}{2R}\Big)\sideset{}{^*}\sum_{r\leqslant R}\frac{(r,hq+m)}{r} \\ &\ll \frac{\sqrt{m}}{q}(\log q)^4\sum_{h\geqslant 1}\frac{\tau(h)\tau(hq+m)}{h}\exp\Big(-\frac{\sqrt{h}}{2R}\Big) \ll_\varepsilon \frac{q^\varepsilon\sqrt{m}}{q}. \end{aligned}$$

□

The estimate involving Y_0 is slightly more complicated because Y_0 is not decreasing very fast at infinity, but instead is oscillating: indeed, it satisfies the asymptotic ([G-R, 8.451.2])

$$Y_0(x) \sim \sqrt{\frac{2}{\pi x}}\sin(x-\frac{\pi}{4})$$

as $x \to +\infty$. Hence, if $Y(m)$ is small, this is because of cancellation in the oscillatory integral $y(h)$. This is similar to the Riemann-Lebesgue Theorem, according to which the Fourier coefficients of a C^∞ periodic function, which are integrals against the oscillatory exponential functions $n \mapsto e(nx)$, tend to zero faster than any polynomial, and the mainspring of the proof is successive integration by parts, exploiting the recurrence relations satisfied by Bessel functions (instead of using directly the asymptotic expansions).

We first prove a general lemma, which is quite standard.

Lemma 34. *Let $\nu \geqslant 0$ be a real number, $J \geqslant 0$ an integer, f a C^∞ test function compactly supported on the interval $[Y, 2Y]$ and let $\beta > 0$, $\vartheta > 0$ be real numbers such that the bounds*

$$y^j f^{(j)}(y) \ll_j \vartheta(1+\beta Y)^j \tag{6.53}$$

hold for $0 \leqslant j \leqslant J$, the implied constants depending on j alone. Then for any $\alpha > 0$ such that $\alpha Y \geqslant 1$, we have

$$\int_0^{+\infty} Y_\nu(\alpha y)f(y)dy \ll_J \vartheta\Big(\frac{1+\beta Y}{1+\alpha Y}\Big)^J Y \tag{6.54}$$

where the implied constant depends only on J and on the implied constants in (6.53).

Proof. One could write the asymptotic development of Y_0 to show the oscillating behavior and integrate by parts, but it is cleaner (and amounts to the same thing) to make use of the recurrence formula

$$(y^\nu Y_\nu(y))' = y^\nu Y_{\nu-1}(y)$$

to get (also by integration by part)

$$\int_0^{+\infty} Y_\nu(\alpha y) f(y) dy = \frac{1}{\alpha} \int_0^\infty Y_{\nu+1}(\alpha y) \Big(f'(y) + \frac{f(y)}{y} \Big) dy.$$

Let $g(y) = f'(y) + f(y)/y$; by Leibniz's rule and the assumption, g satisfies

$$\begin{aligned} y^{j+1} g^{(j)}(y) &= y^{j+1} f^{(j)}(y) + \sum_{k=0}^{j} \binom{j}{k} (-1)^{j-k} (j-k)! y^k f^{(k)}(y) \\ &\ll_j \vartheta (1+\beta Y)^{j+1} \end{aligned}$$

for $0 \leqslant j \leqslant J-1$. Hence, iterating this procedure, we obtain

$$\int_0^{+\infty} Y_\nu(\alpha y) f(y) dy = \frac{1}{\alpha^J} \int_0^{+\infty} Y_{\nu+J}(\alpha y) h(y) dy,$$

where the function h is such that

$$y^J h(y) \ll_J \vartheta (1+\beta Y)^J$$

and therefore the result follows by using $Y_{\nu+J}(y) \ll_{J+\nu} 1$. □

Proof of Lemma 32. We have

$$Y(m) = \frac{4\pi^2}{\sqrt{q}} \sideset{}{^*}\sum_{r \leqslant R} \frac{1}{r^2} \sum_h \tau(h) S(hq-m, 0; r) y(h).$$

We make a smooth dyadic partition of unity, so

$$\xi = \sum_{k \geqslant 1} \xi_k$$

where each ξ_k is a C^∞ function with compact support in a dyadic interval $[X_k, 2X_k]$ that satisfies

$$x^j \xi_k^{(j)}(x) \ll_j 1, \text{ for all } j \geqslant 0, \tag{6.55}$$

the implied constants depending on j alone, in particular, they are uniform in k. We study each ξ_k individually, but we keep writing ξ instead of ξ_k, and accordingly we use X rather than X_k.

By the change of variable $2r^{-1}\sqrt{x} = y$, the integral is

$$y(h) = r \int_0^{+\infty} Y_0(2\pi\sqrt{h}x) J_1\Big(2\pi\sqrt{\frac{m}{q}}x\Big) W\Big(\frac{\pi^2 r^2 x^2}{q}\Big) \xi\Big(\frac{r^2x^2}{4}\Big) dx, \tag{6.56}$$

so we define the test function f by

$$f(x) = J_1\Big(2\pi\sqrt{\frac{m}{q}}x\Big)W\Big(\frac{\pi^2 r^2 x^2}{q}\Big)\xi\Big(\frac{r^2x^2}{4}\Big).$$

This is a C^∞ function compactly supported in the dyadic interval $[\rho, 2\rho]$, with

$$\rho = 2\frac{\sqrt{X}}{r}. \tag{6.57}$$

We first treat the case $\frac{1}{2} \leqslant X \leqslant q^2$ (which involves $\ll \log q$ terms).
Claim: f satisfies the hypothesis of Lemma 34 with

$$Y = \rho, \quad \alpha = 2\pi\sqrt{h}, \quad \beta = 2\pi\sqrt{\frac{m}{q}}, \quad \vartheta = (\log q)^3$$

and any (fixed) positive integer $J \geqslant 1$.

This follows from the bound

$$x^j W^{(j)}(x) \ll_j (\log q)^3, \text{ for all } j \geqslant 0,$$

of (6.25), valid for $1/q \ll x \ll q^2$, the analogue bound for ξ in (6.55), the recurrence relation

$$(x^\nu J_\nu(x))' = x^\nu J_{\nu-1}(x)$$

and some elementary, although somewhat lengthy induction arguments and manipulations with inequalities. The intuitive reason is that the function W and ξ are "flat", while $J_1(\alpha x)$ oscillates somewhat like $e(\alpha x)$. For a completely detailed proof, proceed step by step with

$$f_1(x) = \xi\Big(\frac{r^2x^2}{4}\Big), \qquad f_2(x) = W\Big(\frac{\pi^2r^2x^2}{q}\Big),$$
$$f_3(x) = f_1(x)f_2(x), \qquad f_4(x) = J_1(\alpha x),$$
$$f_5(x) = f_3(x)f_4(x) = f(x).$$

Thus, we are in a position to apply the preceding lemma to f, provided that

$$2\pi\rho\sqrt{h} \geqslant 1. \tag{6.58}$$

This restriction will complicate things, unfortunately, and it will be necessary to split into different cases. If (6.58) holds, we obtain

$$y(h) \ll_J r\rho\frac{\Big(1+\rho\sqrt{\frac{m}{q}}\Big)^J}{(1+\rho\sqrt{h})^J}(\log q)^3. \tag{6.59}$$

Consider first the case $\rho > 1$, or $r < \sqrt{X}$, in which case, since $h \geqslant 1$, the condition is satisfied: applying (6.59) with $J \geqslant 3$ (to win convergence in h) yields a contribution

in (6.37) which is therefore

$$\begin{aligned}
&\ll_J \frac{(\log q)^3}{\sqrt{q}} \sideset{}{^*}\sum_{r<\sqrt{X}} \frac{1}{r^2} r\rho^{-(J-1)}\Big(1+\rho\sqrt{\frac{m}{q}}\Big)^J \tau(r)\\
&\ll_J \frac{(\log q)^3}{\sqrt{q}}\Big(\sideset{}{^*}\sum_{r<\sqrt{\frac{mX}{q}}} \rho\frac{\tau(r)}{r}\Big(\sqrt{\frac{m}{q}}\Big)^J + \sideset{}{^*}\sum_{\sqrt{\frac{mX}{q}}\leqslant r<\sqrt{X}} \frac{\tau(r)}{r}\Big)\\
&\ll_J \frac{(\log q)^{5+J}}{\sqrt{q}} q^{1+J(\Delta-1)/2}, \text{ since } m/q \leqslant q^{\Delta-1}
\end{aligned}$$

at which point, since $\Delta < 1$, we can choose J large enough so that $1 + J(\Delta-1)/2 \leqslant 0$ to conclude that this part is

$$\ll_{\Delta,\varepsilon} \frac{q^\varepsilon}{\sqrt{q}}. \tag{6.60}$$

On the other hand, for $\rho \leqslant 1$, we split the summation in h in the following way

$$\sum_{h\geqslant 1} = \sum_{h\leqslant \rho^{-2(1+\kappa)}} + \sum_{h>\rho^{-2(1+\kappa)}}$$

where $\kappa > 0$ will be chosen (sufficiently small) a little later.

For the first sum, where the condition (6.58) is not valid, we come back to (6.37), using again $J_1(x) \ll x$, $Y_0(x) \ll (\log x)$ and $S(hq-m,0;r) \leqslant (hq-m,r)$ to derive

$$\begin{aligned}
&\frac{4\pi^2}{\sqrt{q}} \sideset{}{^*}\sum_{\sqrt{X}\leqslant r\leqslant R} \frac{1}{r^2} \sum_{h\leqslant\rho^{-2(1+\kappa)}} \tau(h) S(hq-m,0;r) y(h)\\
&\ll \frac{(\log q)^3}{\sqrt{q}} \sideset{}{^*}\sum_{\sqrt{X}\leqslant r\leqslant R} \frac{1}{r^2} \sum_{h\leqslant\rho^{-2(1+\kappa)}} \tau(h)(hq-m,r)\frac{X}{r}\sqrt{\frac{m}{q}}\\
&\ll \frac{\sqrt{m}}{q}(\log q)^3 X \sum_{h\leqslant (R^2/X)^{1+\kappa}} \tau(h) \sideset{}{^*}\sum_{\sqrt{X}h^\theta\leqslant r\leqslant R} \frac{(hq-m,r)}{r^3}, \text{where } \theta = (2+2\kappa)^{-1}\\
&\ll \frac{\sqrt{m}}{q}(\log q)^3 X \sum_{h\leqslant (R^2/X)^{1+\kappa}} \tau(h) \sum_{d\mid hq-m} \frac{\varphi(d)}{d^3} \sideset{}{^*}\sum_{\sqrt{X}h^\theta\leqslant dr\leqslant R} \frac{1}{r^3}\\
&\ll \frac{\sqrt{m}}{q}(\log q)^3 X \sum_{h\leqslant (R^2/X)^{1+\kappa}} \tau(h)\tau(hq-m) X^{-1} h^{-1+\kappa/(1+\kappa)}\\
&\ll_\varepsilon \frac{\sqrt{m}}{q} q^\epsilon R^{2\kappa/(1+\kappa)} \ll_\varepsilon \frac{\sqrt{m}}{q} q^\epsilon R^{2\kappa}, \text{ for all } \varepsilon > 0.
\end{aligned} \tag{6.61}$$

For the second sum, we have $\sqrt{h} > \rho^{-2}$ hence $\rho\sqrt{h} > \rho^{-1} > 1$, so applying (6.59) again for $J \geqslant 3$ entails (recall $\rho \leqslant 1$)

$$y(h) \ll_J \sqrt{X}\rho^{-J} h^{-J/2} (\log q)^3$$

and

$$\begin{aligned}
&\frac{4\pi^2}{\sqrt{q}}\sideset{}{^*}\sum_{\sqrt{X}\leqslant r\leqslant R}\frac{1}{r^2}\sum_{h>\rho^{-2(1+\kappa)}}\tau(h)S(hq-m,0;r)y(h)\\
&\ll_J \frac{(\log q)^3}{\sqrt{q}}\sideset{}{^*}\sum_{\sqrt{X}\leqslant r\leqslant R}\frac{1}{r^2}\sqrt{X}\rho^{-J}\rho^{-2(1+\kappa)(1-J/2)}\tau(r)\\
&\ll \frac{(\log q)^3}{\sqrt{q}}\sideset{}{^*}\sum_{\sqrt{X}\leqslant r\leqslant R}\frac{\sqrt{X}}{r^2}\rho^{-2+\kappa(J-2)}\tau(r). \qquad (6.62)
\end{aligned}$$

We choose $\kappa=\varepsilon/4$, then J large enough so that $-2+\kappa(J-2)>0$ (in addition to the previous conditions that $j\geqslant 3$, $1+J(\Delta-1)/2\leqslant 0$); then (6.61) and (6.62) together are

$$\ll_{\Delta,\varepsilon} q^{\varepsilon}\Big(\frac{m^{1/2}}{q}+\frac{1}{\sqrt{q}}\Big)$$

(using $\rho^{-2+\kappa(J-2)}\leqslant 1$; recall also that $R\leqslant q^2$ was assumed in the statement of Lemma 32).

Finally, we return to the case $X>q^2$ which remains. We appeal to (6.24) (for $j=2$), and again use elementary estimations to prove that for $X>q$ the function f satisfies the better bound

$$x^j f^{(j)}(x)\ll \Big(1+x\sqrt{\frac{m}{q}}\Big)^j q^2(rx)^{-4}$$

namely, Lemma 34 can be applied now with $\vartheta=q^2(r\rho)^{-4}=16q^2X^{-2}$. Hence, when applicable, we get

$$y(h)\ll r\rho\frac{\Big(1+\rho\sqrt{\frac{m}{q}}\Big)^J}{(1+\rho\sqrt{h})^J}\frac{q^2}{X^2}(\log q)^3$$

in addition to the bound (6.59).

Since $X>q^2$, the quantity saved is

$$\frac{q^2}{X^2}\ll X^{-1}$$

which is more than sufficient to allow for the sum over the dyadic values of X involved to converge, and proves that all the previous bounds where (6.59) was used remain valid. The only place where this is not the case is the inequality (6.61), but this part of the sum is void for $\sqrt{X}>R$ and the former estimate works in the larger interval $X\leqslant R^2$. □

A formula for the second moment

We can at last reap the fruits of those efforts.

Proposition 19. *Assume $M=q^{\Delta}$ with $\Delta<\frac{1}{4}$. Then there exists $\delta>0$ such that*

$$M_2=\frac{1}{12}M_{21}-\frac{1}{4}M_{22}+M_3+O(q^{-\delta}) \qquad (6.63)$$

where M_{21}, M_{22} and M_3 are quadratic forms in the variables x_m given by

$$M_{21} = \sum_b \frac{1}{b} \sum_{m_1,m_2} \frac{\tau(m_1m_2)}{m_1m_2} x_{bm_1}x_{bm_2}\Big(\log\frac{\hat{Q}}{m_1m_2}\Big)^3 \tag{6.64}$$

$$M_{22} = \sum_b \frac{1}{b} \sum_{m_1,m_2} \frac{T(m_1m_2)}{m_1m_2} x_{bm_1}x_{bm_2}\Big(\log\frac{\hat{Q}}{m_1m_2}\Big) \tag{6.65}$$

$$M_3 = \sum_b \frac{1}{b} \sum_{m_1,m_2} \frac{\tau(m_1m_2)}{m_1m_2} P_1\Big(\log\frac{\hat{Q}}{m_1m_2}\Big). \tag{6.66}$$

Proof. We appeal to (6.27) and then apply Proposition 18. The first three terms give exactly the three quadratic forms M_{21}, M_{22} and M_3. Moreover, using (6.6), the error term is, for any $\varepsilon > 0$

$$\ll q^{-1/2+\varepsilon}\Big|\sum_{m\leqslant M} x_m\Big|^2 \ll M^2q^{-1/2+2\varepsilon}.$$

If $\Delta < \frac{1}{4}$, we can take ε small enough so that this is $O(q^{-\delta})$ for some $\delta > 0$. □

Now, with the expressions of M_1 (Proposition 15) and M_2 (Proposition 19) as rather simple linear and quadratic forms in the coefficients x_m of the mollifier, it remains to optimize the bound (6.3), namely to maximize the quadratic form M_2 under the linear constraint given by M_1.

Such a problem is solved in principle, for a diagonalized quadratic form, by the following well-known lemma.

Lemma 35. *Let*

$$L = \sum_{1\leqslant k\leqslant n} j(k)X_k, \quad Q = \sum_{1\leqslant k\leqslant n} \nu(k)X_k^2$$

be a linear form and a positive definite quadratic form on $\mathbf{R}^n$ (so $\nu(k) > 0$). Then

$$\sup_{\substack{x\in\mathbf{R}^n\\ x\neq 0}} \frac{L(x)^2}{Q(x)} = J$$

with

$$J = \sum_{1\leqslant k\leqslant n} \frac{j(k)^2}{\nu(k)}.$$

Proof. By Cauchy's inequality

$$\begin{aligned} L(x)^2 &= \Big(\sum_{1\leqslant k\leqslant n} j(k)x_k\Big)^2 \\ &= \Big(\sum_{1\leqslant k\leqslant n} \frac{j(k)}{\sqrt{\nu(k)}}\sqrt{\nu(k)}x_k\Big)^2 \\ &\leqslant J\sum_{1\leqslant k\leqslant n} \nu(k)x_k^2 = JQ(x) \end{aligned}$$

for all $x = (x_k) \in \mathbf{R}^n$, so the supremum is at most J, but the case of equality in Cauchy's inequality shows that the bound is achieved for

$$x_k = \frac{j(k)}{\nu(k)},$$

hence the result. □

Unfortunately, the quadratic form M_2 as given by the Proposition is not diagonalized. The strategy of the proof of Theorem 16 is now to write M_{21} as a linear combination of easily diagonalized quadratic forms; the simplest in shape, say Π, is chosen and we are able to select (x_m) to optimize the value of Π with respect to M_1. Then the remaining terms in M_{21} are evaluated, and so is M_{22}. Both are of the same order of magnitude, so our choice may not be perfectly optimal. On the other hand, with our specific choice of x_m, we finally prove that M_3 gives a smaller contribution, namely that

$$M_3 = O\Big(M_{21}\frac{(\log\log q)^4}{\log q}\Big). \tag{6.67}$$

6.1.4 The preferred quadratic form, I

Separating m_1 and m_2 in (6.69) by means of the formula (compare (5.37)

$$\tau(m_1m_2) = \sum_{a|(m_1,m_2)} \mu(a)\tau\Big(\frac{m_1}{a}\Big)\tau\Big(\frac{m_2}{a}\Big) \tag{6.68}$$

we get

$$M_{21} = \sum_b \frac{1}{b}\sum_a \frac{\mu(a)}{a^2}\sum_{m_1,m_2}\frac{\tau(m_1)\tau(m_2)}{m_1m_2}x_{abm_1}x_{abm_2}\Big(\log\frac{\hat{Q}}{a^2m_1m_2}\Big)^3 \tag{6.69}$$

We define the following arithmetic functions

$$\nu_t(k) = \frac{1}{k}\sum_{ab=k}\frac{\mu(a)(\log a)^t}{a}, \quad \text{for } t = 1, 2, 3. \tag{6.70}$$

Then expanding the logarithm in (6.69) and rearranging, we see that M_{21} is a linear combination of the quadratic forms $\Pi(t, u, v, w)$ in the x_m's defined by

$$\Pi(t, u, v, w) = (\log\hat{Q})^u\sum_k \nu_t(k)y_k^{(v)}y_k^{(w)} \tag{6.71}$$

where the new variables $y_k^{(i)}$, for $i \geqslant 1$, are defined by

$$y_k^{(i)} = \sum_m \frac{\tau(m)}{m}(\log m)^i x_{km} \tag{6.72}$$

and t, u, v and w are non-negative integers such that $t + u + v + w = 3$.[2]

[2]Actually, M_3 is also such a linear combination with the difference that $t + u + v + w \leqslant 2$. This will explain (6.67).

We further restrict our attention to $\Pi(u,v,w) := \Pi(0,u,v,w)$; again it will be seen that for the chosen (x_m)

$$\Pi(t,u,v,w) = O\Big(\Pi(0,u,v,w)\frac{(\log\log q)^{t+2}}{\log q}\Big) \tag{6.73}$$

which justifies this restriction. Accordingly we write ν for ν_0, for which we have the formula

$$\nu(k) = \frac{\varphi(k)}{k^2}, \text{ for } k \leqslant M. \tag{6.74}$$

The part of the expansion of M_{21} involving those $\Pi(u,v,w)$ is then (using the obvious symmetry $\Pi(u,v,w) = \Pi(u,w,v)$) denoted by m_{21}:

$$m_{21} = \Pi(3,0,0) - 6\Pi(2,1,0) + 6\Pi(1,1,1) + 6\Pi(1,2,0) - 6\Pi(0,1,2) - 2\Pi(0,0,3). \tag{6.75}$$

Finally, we select the one quadratic form $\Pi := \Pi(3,0,0)$ as reference: we will choose (x_m) to optimize Π and evaluate afterwards the other $\Pi(u,v,w)$, for this choice, before doing the same with M_{22}.

Optimization of the preferred form

By definition, Π is in the desired diagonalized form

$$\Pi = (\log \hat{Q})^3 \sum_k \nu(k) y_k^2. \tag{6.76}$$

Conversely, let $g = \mu \star \mu$ be the Dirichlet convolution inverse of τ, then

$$x_m = \sum_k \frac{g(k)}{k} y_{km}. \tag{6.77}$$

From this we express the linear form[3] in (6.19) in terms of y_k

$$M_1 = \sum_m \frac{x_m}{m} \log\frac{\hat{q}}{m} = \sum_k j(k) y_k \tag{6.78}$$

where

$$j(k) = \frac{1}{k} \sum_{ab=k} g(a) \Big(\log\frac{\hat{q}}{b}\Big).$$

Lemma 36. *For any integer $k \geqslant 1$ we have*

$$j(k) = \frac{\mu(k)}{k}(\log \hat{q}k).$$

Proof. We have

$$\sum_{k\geqslant 1} g(k) k^{-s} = \zeta(s)^{-2}$$

[3]Strictly speaking, the main term of the linear form, but we will keep the same notation.

and therefore

$$\sum_{k\geqslant 1} j(k)k^{-s} = \zeta(s+1)^{-2} \times \Big((\log \hat{q})\zeta(s+1) + \zeta'(s+1)\Big)$$
$$= (\log \hat{q})\zeta(s+1)^{-1} - (\zeta^{-1})'(s+1)$$

whence the result. □

By Lemma 35, the best choice to optimize Π with respect to M_1 is

$$y_k = \begin{cases} \dfrac{j(k)}{\nu(k)} = \dfrac{k\mu(k)}{\varphi(k)}(\log \hat{q}k), \text{ if } k \leqslant M \\ 0, \text{ if } k > M \end{cases} \tag{6.79}$$

and x_m is given by (6.77), from which (and the lemma) the conditions required in section 6.1.2 are immediately verified: obviously, y_k is supported on squarefree integers $k \leqslant M$, hence so is x_m. Moreover

$$g(k) = \sum_{ab=k} \mu(a)\mu(b) \ll \tau(k)$$

and

$$y_k \ll (\log q)(\log\log k)$$

so (very crudely)

$$|x_m| = \Big|\sum_k \frac{g(k)}{k} y_{km}\Big| \ll (\log q)^4.$$

We now compute the various terms in (6.75) to apply the estimate (6.3). [4]

Lemma 37. *With the previous notations and hypothesis, with $M = q^{\Delta}$, we have*

$$M_1 = (\log q)^3 \times \Delta\Big(\frac{\Delta^2}{3} + \frac{\Delta}{2} + \frac{1}{4}\Big) + O\big((\log q)^2\big)$$
$$= \mathfrak{m}_1(\Delta)(\log q)^3 + O((\log q)^2),$$

where $\mathfrak{m}_1$ is the polynomial (6.7), and

$$\Pi = (\log q)^6 \times \Delta\Big(\frac{\Delta^2}{3} + \frac{\Delta}{2} + \frac{1}{4}\Big) + O\big((\log q)^5\big).$$

Proof. By the choice of (y_k) and Lemma 36, we have

$$(\log \hat{Q})^{-3}\Pi = M_1 = \sum_k \frac{j(k)^2}{\nu(k)} = \sum_k \frac{\mu(k)^2}{\varphi(k)}(\log \hat{q}k)^2$$

and therefore the result follows by partial summation from

$$\sum_{k\leqslant K} \frac{\mu(k)^2}{\varphi(k)} = \log K + O(1)$$

---

[4] Since $j(k)$ is about $(\log k)/k$ and ν is about k^{-1}, it is already quite clear that we will get a positive (harmonic) proportion if $M = q^{\Delta}$ for any $\Delta > 0$.

which is well-known, and immediately derived from the residue at $s = 0$ of the Dirichlet series

$$\sum_{n\geqslant 1} \frac{\mu(n)^2}{\varphi(n)} n^{-s} = \prod_p (1 + p^{-s-1}(1-p^{-1})^{-1}) \\ = \frac{\zeta(s+1)}{\zeta(2(s+1))} \prod_p \Big(1 + \frac{1}{(1-p^{-1})(1+p^{-s-1})}\Big).$$

□

Estimation of the other quadratic forms

For the other quadratic forms, we have

$$\Pi(u,v,w) = (\log \hat{Q})^u \sum_k \nu(k) y_k^{(v)} y_k^{(w)}$$

where $y_k^{(i)}$ is given by (6.72),

$$y_k^{(i)} = \sum_m \frac{\tau(m)}{m} (\log m)^i x_{km}.$$

We can express $y_k^{(i)}$ in terms of (y_k) using the higher Von Mangoldt function Λ_i, which is defined by the Dirichlet convolution

$$\Lambda_i = \mu \star (\log)^i,$$

so that

$$(\log m)^i = \sum_{ab=m} \Lambda_i(a).$$

From this, and the fact that the x_m's are supported on squarefree integers, we derive

$$y_k^{(i)} = \sum_{\ell \leqslant M/k} \frac{\tau(\ell)}{\ell} \Lambda_i(\ell) y_{k\ell}. \tag{6.80}$$

We state the properties of Λ_i which we will use.

1. $\Lambda_1 = \Lambda$, the usual Van-Mangoldt function.
2. Λ_i is supported on integers having at most i distinct prime factors.
3. If $m = p_1 \dots p_i$, for distinct primes $p_1, \dots, p_i$, then
$$\Lambda_i(m) = i!(\log p_1) \dots (\log p_i).$$
4. If p_1 and p_2 are distinct primes, then
$$\Lambda_i(p_1) = (\log p_1)^i \\ \Lambda_3(p_1 p_2) = 3(\log p_1)(\log p_2)(\log p_1 p_2).$$

All of these are well known and (or) easy to prove from the recurrence relation

$$\Lambda_{i+1} = (\log)\Lambda_i + \Lambda \star \Lambda_i.$$

In (6.80) we are thus actually dealing with a sum over squarefree ℓ having at most i prime factors, and $i \leqslant 3$. We separate the sum into the parts with a fixed number of prime factors, which produces multiple (at most triple) sums over primes (of Mertens type since $\tau(\ell)\ell^{-1} = 2^j\ell^{-1}$ for such ℓ with $\omega(\ell) = j$ prime factors).

The subsum with i distinct prime factors is, by the above

$$\begin{aligned}
&2^i i! \sum_{\substack{\ell \leqslant M/k \\ \omega(\ell)=i}} \frac{\Lambda_i(\ell)}{\ell}\mu(k\ell)(\log \hat{q}k\ell)\frac{k\ell}{\varphi(k\ell)} \\
&= (-2)^i i! \frac{k\mu(k)}{\varphi(k)} \sum_{\substack{p_1<\ldots<p_i \\ p_1\ldots p_i \leqslant M/k \\ (p_1\ldots p_i,k)=1}} \frac{(\log p_1)\ldots(\log p_i)}{p_1\ldots p_i}(\log \hat{q}kp_1\ldots p_i) + O\big((\log q)^i \frac{k}{\varphi(k)}\big) \\
&= (-2)^i i! \frac{k\mu(k)}{\varphi(k)} \sum_{\substack{p_1<\ldots<p_i \\ p_1\ldots p_i \leqslant M/k}} \frac{(\log p_1)\ldots(\log p_i)}{p_1\ldots p_i}(\log \hat{q}kp_1\ldots p_i) + O\big((\log q)^i (\log_2 q)\frac{k}{\varphi(k)}\big) \\
&= (-2)^i \frac{k\mu(k)}{\varphi(k)} \sum_{\substack{p_1,\ldots,p_i \\ p_1\ldots p_i \leqslant M/k}} \frac{(\log p_1)\ldots(\log p_i)}{p_1\ldots p_i}(\log \hat{q}kp_1\ldots p_i) + O\big((\log q)^i (\log_2 q)\frac{k}{\varphi(k)}\big)
\end{aligned}$$

the error term arising from neglecting the smaller contribution from the primes dividing k and replacing $\varphi(p)^{-1}$ by p^{-1} using the fact that

$$\sum_p \frac{(\log p)^A}{p(p-1)} < +\infty.$$

From Mertens's formula, the last sum is equal, up to $O\big((\log q)^i\big)$, to the integral

$$\begin{aligned}
&\int_{\substack{y_1 \geqslant 0,\ldots,y_i \geqslant 0 \\ y_1+\ldots+y_i \leqslant (\log M/k)}} (\log \hat{q}k + y_1 + \ldots + y_i)dy \\
&= (\log \hat{q}k)\Big(\log \frac{M}{k}\Big)^i \int_{S_i} dx + i\Big(\log \frac{M}{k}\Big)^{i+1} \int_{S_i} x_1 dx
\end{aligned}$$

Here $S_i = \{(x_1, \ldots, x_i) \mid x_j \geqslant 0,\ x_1 + \ldots + x_i \leqslant 1\}$ is the standard i-simplex. By induction, one gets immediately

$$\int_{S_i} dx = \frac{1}{i!},\ \int_{S_i} x_1 dx = \frac{1}{(i+1)!}$$

so this contribution to the sum (6.80) can be written as

$$\frac{(-2)^i\mu(k)}{(i+1)!}\Big(\log \frac{M}{k}\Big)^i (\log \hat{q}^{i+1}M^i k) + O\big((\log q)^i(\log_2 q)\frac{k}{\varphi(k)}\big). \tag{6.81}$$

This is enough to give $y_k^{(1)}$; for $y_k^{(2)}$ there is an additional sum over primes which, by similar computations, is

$$\begin{aligned}&-2\frac{k\mu(k)}{\varphi(k)}\sum_{p\leqslant M/K}\frac{(\log p)^2}{p}(\log \hat{q}kp)+O\big((\log q)(\log_2 q)^2\frac{k}{\varphi(k)}\big)\\&=-\frac{1}{3}\frac{k\mu(k)}{\varphi(k)}\Big(\log\frac{M}{k}\Big)^2(\log \hat{q}^3M^2k)+O\big((\log q)^2\frac{k}{\varphi(k)}\big);\end{aligned}$$

and for $y_k^{(3)}$ there are two other sums, first

$$\begin{aligned}&-2\frac{k\mu(k)}{\varphi(k)}\sum_{p\leqslant M/K}\frac{(\log p)^3}{p}(\log \hat{q}kp)+O\big((\log q)(\log_2 q)^3\frac{k}{\varphi(k)}\big)\\&=-\frac{1}{6}\frac{k\mu(k)}{\varphi(k)}\Big(\log\frac{M}{k}\Big)^3(\log \hat{q}^4M^3k)+O\big((\log q)^3\frac{k}{\varphi(k)}\big);\end{aligned}$$

and finally

$$\begin{aligned}&12\frac{k\mu(k)}{\varphi(k)}\sum_{\substack{p_1<p_2\\p_1p_2\leqslant M/k}}\frac{(\log p_1p_2)(\log p_1)(\log p_2)}{p_1p_2}(\log \hat{q}kp_1p_2)+O\big((\log q)^2(\log_2 q)^2\frac{k}{\varphi(k)}\big)\\&=12\frac{k\mu(k)}{\varphi(k)}\sum_{p_1p_2\leqslant M/k}\frac{(\log p_1)^2(\log p_2)}{p_1p_2}(\log \hat{q}kp_1p_2)+O\big((\log q)^2(\log_2 q)^2\frac{k}{\varphi(k)}\big)\\&=\frac{1}{2}\frac{k\mu(k)}{\varphi(k)}\Big(\log\frac{M}{k}\Big)^3(\log \hat{q}^4M^3k)+O\big((\log q)^3\frac{k}{\varphi(k)}\big).\end{aligned}$$

From all this we conclude:

Lemma 38. *For $i=1,\ 2,\ 3$, we have*

$$y_k^{(i)}=c_i\frac{k\mu(k)}{\varphi(k)}\Big(\log\frac{M}{k}\Big)^i(\log \hat{q}^{i+1}M^ik)+O\big((\log q)^i(\log_2 q)\frac{k}{\varphi(k)}\big)\tag{6.82}$$

with

$$c_1=-1,\ c_2=\frac{1}{3},\ c_3=0.\tag{6.83}$$

It is now easy to finish the computation of the quadratic form m_{21} for our choice of y_k.

Lemma 39. *With notations as in Lemma 37*

$$\begin{aligned}\Pi(2,1,0)&=-(\log q)^6\times\Delta^2\Big(\big(\frac{1}{2}+\Delta\big)^2-\Delta\big(\frac{1}{2}+\Delta\big)+\frac{\Delta^2}{4}\Big)+O\big((\log q)^5\log_2 q\big)\\\Pi(1,1,1)&=(\log q)^6\times\Delta^3\Big(\frac{4}{3}\big(\frac{1}{2}+\Delta\big)^2-\Delta\big(\frac{1}{2}+\Delta\big)+\frac{\Delta^2}{5}\Big)+O\big((\log q)^5\log_2 q\big)\\\Pi(1,2,0)&=\frac{1}{3}(\log q)^6\times\Delta^3\Big(\big(\frac{1}{2}+\Delta\big)^2-\Delta\big(\frac{1}{2}+\Delta\big)+\frac{\Delta^2}{5}\Big)+O\big((\log q)^5\log_2 q\big)\\\Pi(0,1,2)&=-\frac{1}{3}(\log q)^6\times\Delta^4\Big(\frac{3}{2}\big(\frac{1}{2}+\Delta\big)^2-\Delta\big(\frac{1}{2}+\Delta\big)+\frac{\Delta^2}{6}\Big)+O\big((\log q)^5\log_2^3 q\big)\\\Pi(0,0,3)&=O\big((\log q)^5\log_2^3 q\big).\end{aligned}$$

Proof. All are similar, so take for instance $\Pi(0,1,2)$; from the previous lemma

$$\Pi(0,1,2) = -\frac{1}{3}\sum_{k\leqslant M}\frac{\mu(k)^2}{\varphi(k)}\Big(\log\frac{M}{k}\Big)^3(\log\hat{q}^3M^2k)(\log\hat{q}^2Mk) + O\big((\log q)^5(\log_2 q)^3\big)$$

and the sum, by summation by parts again, is – up to $O\big((\log q)^5\big)$ – the same as the integral

$$\int_1^M\Big(\log\frac{M}{x}\Big)^3(\log\hat{q}^3M^2x)(\log\hat{q}^2Mx)\frac{dx}{x} = \int_0^{\log M} y^3(3\log\hat{q}M - y)(2\log\hat{q}M - y)dy$$

from which the result follows, since moreover $\log\hat{q} = \log\sqrt{q} + O(1)$. □

Therefore, from the definition (6.75), we get:

Corollary 4. *We have*

$$m_{21} = \mathfrak{m}_{21}(\Delta)(\log q)^6 + O((\log q)^5(\log_2 q)^3)$$

where $\mathfrak{m}_{21}$ *is the polynomial (6.8).*

The other contribution to the second moment

Recall that

$$M_{22} = \sum_b\frac{1}{b}\sum_{m_1,m_2}\frac{T(m_1m_2)}{m_1m_2}x_{bm_1}x_{bm_2}\Big(\log\frac{\hat{Q}}{m_1m_2}\Big).$$

The formula (6.49) again separates m_1 and m_2, so

$$\begin{aligned}M_{22} &= 2\sum_b\frac{1}{b}\sum_a\frac{\mu(a)}{a^2}\sum_{m_1,m_2}\frac{\tau(m_1)T(m_2)}{m_1m_2}x_{abm_1}x_{abm_2}\Big(\log\frac{\hat{Q}}{a^2m_1m_2}\Big)\\ &= 2\sum_k\nu(k)\sum_{m_1,m_2}\frac{\tau(m_1)T(m_2)}{m_1m_2}x_{km_1}x_{km_2}\Big(\log\frac{\hat{Q}}{m_1m_2}\Big)\\ &- 4\sum_k\nu_1(k)\sum_{m_1,m_2}\frac{\tau(m_1)T(m_2)}{m_1m_2}x_{km_1}x_{km_2}.\end{aligned}$$

Let m_{22} denote the first term; this will be the main contribution. The treatment is now similar to that of m_{21}: define

$$\begin{aligned}z_k = z_k^{(0)} &= \sum_m\frac{T(m)}{m}x_{km}\\ z_k^{(1)} &= \sum_m\frac{T(m)}{m}(\log m)x_{km}\end{aligned}$$

and

$$\tilde{\Pi}(a,b,c) = (\log\hat{Q})^a\sum_k\nu(k)y_k^{(b)}z_k^{(c)};$$

then

$$m_{22} = 2\Big(\tilde{\Pi}(1,0,0) - \tilde{\Pi}(0,1,0) - \tilde{\Pi}(0,0,1)\Big) \tag{6.84}$$

$$M_{22} = m_{22} - 4\sum_k \nu_1(k) y_k z_k. \tag{6.85}$$

Lemma 40. *We have*

$$z_k = 2\sum_{\ell \leqslant M/k} \frac{(\log \ell)\Lambda(\ell)}{\ell} y_{k\ell}$$

$$z_k^{(1)} = \sum_{\ell \leqslant M/k} \frac{\tau(\ell)\Lambda(\ell)}{\ell} z_{k\ell} + \sum_{\ell \leqslant M/k} \frac{T(\ell)\Lambda(\ell)}{\ell} y_{k\ell}.$$

Proof. For the first one, (6.77) implies

$$z_k = \sum_\ell \Big(\sum_{mn=\ell} \frac{T(m)}{m} g(n) \Big) y_{k\ell}$$

and the Dirichlet generating series for the coefficient of ℓ is $L(s+1)$ where

$$L(s) = \zeta(s)^{-2} \sum_n T(n) n^{-s}.$$

From the first part of lemma 33, we get

$$\sum_n T(n) n^{-s} = 4\zeta\zeta'' - 2(\zeta\zeta')' = 2(\zeta\zeta'' - (\zeta')^2)$$

so $L(s) = 2(\zeta'\zeta^{-1})'$. As to $z_k^{(1)}$, write

$$\log m = \sum_{\ell b = m} \Lambda(\ell),$$

and then

$$\begin{aligned} z_k^{(1)} &= \sum_m \frac{T(m)}{m} \sum_{\ell b = m} \Lambda(\ell) x_{km} = \sum_\ell \frac{\Lambda(\ell)}{\ell} \sum_{m \leqslant M/\ell} \frac{T(\ell m)}{m} x_{k\ell m} \\ &= \sum_{\ell \leqslant M/k} \frac{\tau(\ell)\Lambda(\ell)}{\ell} z_{k\ell} + \sum_{\ell \leqslant M/k} \frac{T(\ell)\Lambda(\ell)}{\ell} y_{k\ell} \end{aligned}$$

by the multiplicative property of T. □

We can now evaluate the quadratic form m_{22}. The mollifier was defined by (6.79).

Lemma 41. *We have*

$$\begin{aligned} z_k &= -\frac{1}{3} \frac{k\mu(k)}{\varphi(k)} \Big(\log \frac{M}{k} \Big)^2 (\log \hat{q}^3 M^2 k) + O((\log q)^2 \frac{k}{\varphi(k)}) \\ &= -y_k^{(2)} + O\big(\frac{k}{\varphi(k)} (\log q)^2 \log_2 q\big) \end{aligned}$$

and

$$z_k^{(1)} = O\big(\frac{k}{\varphi(k)} (\log q)^3\big).$$

Proof. We will be brief : on the one hand

$$\begin{aligned} z_k &= -2\frac{k\mu(k)}{\varphi(k)}\sum_{p\leqslant M/k}\frac{(\log p)^2}{p}\log \hat{q}kp + O(\frac{k}{\varphi(k)}(\log_2 q)^3) \\ &= -2\frac{k\mu(k)}{\varphi(k)}\int_0^{\log M/k} y(y+\log \hat{q}k)dy + O\big(\frac{k}{\varphi(k)}(\log q)^2\big) \\ &= -\frac{1}{3}\frac{k\mu(k)}{\varphi(k)}\Big(\log\frac{M}{k}\Big)^2(\log \hat{q}^3M^2k) + O\big(\frac{k}{\varphi(k)}(\log q)^2\big) \end{aligned}$$

and on the other hand the two contributions to $z_k^{(1)}$ are respectively (using the previous computation)

$$\frac{1}{3}\frac{k\mu(k)}{\varphi(k)}\sum_{p\leqslant M/k}\frac{2\log p}{p}\Big(\log\frac{M}{p}\Big)^2(\log \hat{q}^3M^2p) = \frac{1}{6}\frac{k\mu(k)}{\varphi(k)}\Big(\log\frac{M}{k}\Big)^3(\log \hat{q}^4M^3k) + O\big(\frac{k}{\varphi(k)}(\log q)^3\big)$$

and (this is the same as one of the sums considered in $y_k^{(3)}$)

$$-\frac{k\mu(k)}{\varphi(k)}\sum_{p\leqslant M/k}\frac{2(\log p)^3}{p}(\log \hat{q}kp) = -\frac{1}{6}\frac{k\mu(k)}{\varphi(K)}\Big(\log\frac{M}{k}\Big)^3(\log \hat{q}^4M^3k) + O\big(\frac{k}{\varphi(k)}(\log q)^3\big).$$

□

From this (referring to lemma 39), we obtain

$$\begin{aligned} \tilde{\Pi}(1,0,0) &= -(\log\hat{Q})\sum_k \nu(k)y_k y_k^{(2)} + O\big((\log q)^5\big) \\ &= -\Pi(1,2,0) + O\big((\log q)^5\big) \qquad (6.86) \\ \tilde{\Pi}(0,1,0) &= -\sum_k \nu(k)y_k^{(1)}y_k^{(2)} + O\big((\log q)^5\big) \\ &= -\Pi(0,1,2) + O\big((\log q)^5\big) \qquad (6.87) \\ \tilde{\Pi}(0,0,1) &= O\big((\log q)^5\big). \end{aligned}$$

Hence from (6.84) we derive the final estimate for m_{22}.

Corollary 5. *We have*

$$m_{22} = \mathfrak{m}_{22}(\Delta)(\log q)^6 + O((\log q)^5)$$

where $\mathfrak{m}_{22}$ *is the polynomial (6.9).*

6.1.5 Harmonic non-vanishing

Let us now dispose of the residual quadratic forms. Those are the forms $\Pi(t,u,v,w)$ with $t+u+v+w \leqslant 2$, which enter into the quadratic forms M_{21} and M_3, and the form without name in (6.85) which enters in M_{22}.

Lemma 42. *Let t, u, v, w be non-negative integers with*

$$t + u + v + w \leqslant 2.$$

Then

$$\Pi(t, u, v, w) \ll (\log q)^{u+v+w+2}(\log\log q)^{t+2} \ll (\log q)^5(\log\log q)^4$$

(where $\Pi(t, u, v, w)$ refers to the value of the quadratic form for the vector (x_m) previously chosen). Moreover

$$\sum_k \nu_1(k) y_k z_k \ll (\log q)^5(\log\log q)^3.$$

Proof. By Lemma 38 we have

$$y_k^{(i)} \ll (\log\log k)(\log k)^{i+1}$$

and therefore, directly from the definition

$$\begin{aligned}\Pi(t, u, v, w) &= (\log \hat{Q})^u \sum_k \nu_t(k) y_k^{(v)} y_k^{(w)} \\ &\ll (\log q)^{u+v+w+2} \sum_{k \leqslant M} \frac{(\log\log k)^{t+2}}{k} \\ &\ll (\log q)^{u+v+w+3}(\log\log q)^{t+2}\end{aligned}$$

and similarly from Lemma 41

$$\sum_k \nu_1(k) y_k z_k \ll (\log q)^5(\log\log q)^3.$$

□

We can now summarize our computations, referring to the definition of the polynomials $\mathfrak{m}_1$ and $\mathfrak{m}_2$ in Theorem 16, by saying that it follows from the decomposition (6.84 and the results of Lemma 37, Corollaries 4 and 5 (together with Lemma 42 which confirms (6.67) and (6.73)), that for $\Delta < \frac{1}{4}$ the asymptotic formula

$$M_2 = \mathfrak{m}_2(\Delta)(\log q)^6 + O\big((\log q)^5(\log\log q)^4\big)$$

holds, where $\mathfrak{m}_2$ is defined in (6.10).

Thus, applying (6.3), we obtain the estimate claimed in Theorem 16, valid for $\Delta < \frac{1}{4}$:

$$\sum_{\substack{\varepsilon_f = -1 \\ L'(f, \frac{1}{2}) \neq 0}}^{h} 1 \geqslant \frac{\mathfrak{m}_1(\Delta)^2}{\mathfrak{m}_2(\Delta)} + O\Big(\frac{(\log\log q)^4}{\log q}\Big).$$

An explicit calculation shows that

$$\frac{\mathfrak{m}_1(\frac{1}{4})^2}{\mathfrak{m}_2(\frac{1}{4})} = \frac{19}{54}$$

and this establishes Theorem 15, the analogue of Theorem 10 for the harmonic average.
Remark A point worth noticing, in comparison with Chapter 5, is that we have not used any deeper knowledge of the primes than Tchebychef's estimate

$$\sum_{p\leqslant x}\frac{\log p}{p}=\log x+O(1) \tag{6.88}$$

and in particular we did not need the Prime Number Theorem. This is an advantage of working with an "abstract" mollifier, and optimizing it after changing variables. Unfortunately, this could not be imitated in the context of Chapter 5, because the mollifier obtained would not be adequate for the application of Selberg's lemma 16; compare with [Vdk], where a zero-free region of the Riemann zeta function is also required.

6.2 Removing the harmonic weight: the head, II

We now proceed to prove Theorem 10 and the technique of Chapter 3 to go from an harmonic average to the natural one.

Let again $M=q^{\Delta}$ with $\Delta<\frac{1}{4}$. We consider the first and second moments for the natural average

$$M_1^n=\sum_{f\in S_2(q)^*}\varepsilon_f^-M(f)L'(f,\tfrac{1}{2})=A[\varepsilon_f^-M(f)L'(f,\tfrac{1}{2})],$$
$$M_2^n=\sum_{f\in S_2(q)^*}\varepsilon_f^-|M(f)L'(f,\tfrac{1}{2})|^2=A[\varepsilon_f^-|M(f)L'(f,\tfrac{1}{2})|^2],$$

the mollifier $M(f)$ being again of the form (6.4), for some real numbers (x_m) supported on squarefree integers m (this time this will be a necessary assumption from the start) with the same growth condition

$$x_m\ll(\tau(m)(\log qm))^A$$

(for some absolute constant $A>0$).

Let $\kappa>0$ be such that $\Delta+2\kappa<\frac{1}{4}$. We wish to apply Proposition 8 to M_1^n and M_2^n.

Lemma 43. *For any odd primitive form $f\in S_2(q)^*$, we have the bounds*

$$\omega_fM(f)L'(f,\tfrac{1}{2})\ll q^{-5/8},$$
$$\omega_f|M(f)L'(f,\tfrac{1}{2})|^2\ll q^{-1/4}.$$

Moreover

$$A^h[\varepsilon_f^-|M(f)L'(f,\tfrac{1}{2})|]\ll(\log q)^{C_1}, \tag{6.89}$$
$$A^h[\varepsilon_f^-|M(f)L'(f,\tfrac{1}{2})|^2]\ll(\log q)^{C_2}. \tag{6.90}$$

The constants C_1 and C_2 depend only on the constant $A>0$.

Proof. Let f be an odd primitive form. From the definition of $M(f)$ and the growth condition on x_m we derive

$$M(f) \ll (\log q)^B M^{1/2}$$

for some constant $B = B(A)$. Moreover, for the special value of the derivative of the L-function, we have the classical convexity bound (in q-aspect)

$$L'(f, \tfrac{1}{2}) \ll q^{1/4}(\log q)^2$$

which we can reprove from (6.15), estimating by means of Lemma 26 and Deligne's bound:

$$L'(f, \tfrac{1}{2}) \ll \log q \sum_{l \leqslant \sqrt{q}} \frac{\tau(l)}{\sqrt{l}} + \sqrt{q} \sum_{l > \sqrt{q}} \frac{\tau(l)}{l^{3/2}} \ll q^{1/4}(\log q)^2.$$

On the other hand, we use again $\omega_f \ll (\log q)q^{-1}$ from (3.16), and so

$$\omega_f |M(f)L'(f, \tfrac{1}{2})|^2 \ll (\log q)^{2B+5} q^{\Delta - 1/2} \ll q^{-1/4}$$

and similarly for $\omega_f M(f)L'(f, \frac{1}{2})$.

As for the averages, the second one is the same as the moment M_2 previously considered,[5] and we immediately obtain (6.90) from Proposition 19 using the growth condition. Then (6.89) is deduced from (6.90) by Cauchy's inequality. □

Remark The bound for the derivative of the L-function is of course far from the conjectured truth

$$L'(f, \tfrac{1}{2}) \ll_\varepsilon q^\varepsilon$$

(the Lindelöf Hypothesis), but it will suffice here. However, it is important to mention that this convexity bound has been improved by Duke, Friedlander, Iwaniec [DFI], who showed that

$$L'(f, \tfrac{1}{2}) \ll_\varepsilon q^{47/192 + \varepsilon}.$$

The proof is based on the so-called "amplification technique", and involves also very deep treatments of the complementary term in the Petersson formula.

The Lemma verifies the conditions (3.17) and (3.18) and with $x = q^\kappa$, we conclude from Proposition 8 that there exists $\delta = \delta(\kappa, \Delta) > 0$ such that

$$M_1^n = \frac{\dim J_0(q)}{\zeta(2)} \sum_{f \in S_2(q)^*}^h \omega_f(x) \varepsilon_f^- M(f) L'(f, \tfrac{1}{2}) + O(q^{1-\delta}), \tag{6.91}$$

$$M_2^n = \frac{\dim J_0(q)}{\zeta(2)} \sum_{f \in S_2(q)^*}^h \omega_f(x) \varepsilon_f^- |M(f) L'(f, \tfrac{1}{2})|^2 + O(q^{1-\delta}), \tag{6.92}$$

where $\omega_f(x)$ is the partial sum of length x of the symmetric square, see (3.20)

$$\omega_f(x) = \sum_{n \leqslant x} \frac{\rho_f(n)}{n} = \sum_{d\ell^2 \leqslant x} \frac{\lambda_f(d^2)}{d\ell^2}.$$

[5] We avoid the notation, because it must be emphasized that the x_m now considered are different from those used at the end of the previous section.

We denote the moments with $\omega_f(x)$ by $\mathcal{M}_i$:

$$\mathcal{M}_1 = \sum_{f\in S_2(q)^*}^{h} \omega_f(x)\varepsilon_f^- M(f)L'(f,\tfrac{1}{2})$$
$$\mathcal{M}_2 = \sum_{f\in S_2(q)^*}^{h} \omega_f(x)\varepsilon_f^- |M(f)L'(f,\tfrac{1}{2})|^2.$$

This last section will be extremely technical: the combinatorics of the various forms into which the second moment is decomposed, pretty much kept under control in the former Section, are now much harder to follow because of the loss of multiplicativity in the coefficients. The reader should probably confine her interest to the main term, since it is much the same (of course) as in the case of M_2. Justification for this is that, if something was to go wrong, we would discover a correlation between the values of $L(\mathrm{Sym}^2 f, 1)$ and $L'(f,\frac{1}{2})^2$, and this would be even more interesting than the theorem claimed! It could be that we can not prove the theorem, but without establishing this connection: such would be the case, for instance, if some residual estimate was seen to depend on the existence of an exceptional (Landau-Siegel) zero for Dirichlet characters. Indeed, a correlation of this kind is one of the many strange effects of this unlikely event. But the reader should be able to convince himself easily that the computations required to take care of all details involve no deeper facts than before, namely Tchebychef's estimate (6.88) suffices to close the books.

6.2.1 Computation of the first moment

From (6.15), incorporating directly the mollifier $M(f)$, we get

$$\mathcal{M}_1 = \sum_{n,m} \frac{x_m}{\sqrt{nm}} V\Big(\frac{2\pi n}{\sqrt{q}}\Big) \times \Delta_-^n(m,n)$$

where Δ_-^n is the Delta-symbol without weight for odd forms of Section 4.3. As in the computation of M_1 in Section 6.1.2, the basic estimate (4.13) is sufficient to get a good approximation, namely

$$\mathcal{M}_1 = \sum_m \frac{x_m}{m} \sum_{d\ell^2\leqslant x} \frac{1}{(d\ell)^2} \sum_{r|(d^2,m)} rV\Big(\frac{2\pi m d^2}{r^2\sqrt{q}}\Big) + O\Big(\frac{xM}{\sqrt{q}}(\log q)^B\Big)$$

and similarly (6.16) gives now

$$\begin{aligned}\mathcal{M}_1 &= \sum_m \frac{x_m}{m} \sum_{d\ell^2\leqslant x} \frac{1}{(d\ell)^2} \sum_{r|(d^2,m)} r\Big(\log\frac{\hat{q}r^2}{md^2}\Big) + O(q^{-\delta}) \\ &= \sum_m \frac{x_m}{m} \sum_{r|m} r \sum_{\substack{d\ell^2\leqslant x\\ r|d^2}} \frac{1}{(d\ell)^2}\Big(\log\frac{\hat{q}r^2}{md^2}\Big) + O(q^{-\delta})\end{aligned}$$

for some $\delta > 0$ (since $xM = q^{\kappa+\Delta}$ and $\kappa + \Delta < \frac{1}{4}$).

The summation over m is supported on squarefree integers, hence that over the divisors r also. But if r is squarefree, then r divides d^2 if and only if r divides d.

We use this information in the first place, as in the previous Chapter, to remove the constraint $d\ell^2 \leqslant x$ in the inner summation: the difference is at most

$$(\log q)\sum_m \frac{|x_m|}{m}\sum_{r|m} r \sum_{\substack{d\ell^2>x\\ r|d}} (d\ell)^{-2} \ll \frac{\log q}{\sqrt{x}}\sum_m \frac{|x_m|}{m}\sum_{r|m}\frac{1}{\sqrt{r}} \ll \frac{(\log q)^B}{\sqrt{x}}$$

by a calculation similar to that of Lemma 22 and the growth assumption (6.6).

Hence, with the summation over d and ℓ now free, we further have

$$\begin{aligned}\mathcal{M}_1 &= \zeta(2)\sum_m \frac{x_m}{m}\sum_{r|m} r\sum_{r|d}\frac{1}{d^2}\Big(\log\frac{\hat{q}r^2}{md^2}\Big) + O(q^\delta)\\ &= \zeta(2)^2\sum_m \frac{d_{-1}(m)x_m}{m}\Big(\log\frac{\hat{q}}{m}\Big) + 2\zeta(2)\zeta'(2)\sum_m\frac{d_{-1}(m)x_m}{m} + O(q^\delta)\end{aligned} \tag{6.93}$$

for some $\delta = \delta(\Delta) > 0$.

6.2.2 Computation of the second moment

By the definition of $\omega_f(x)$ we have

$$\begin{aligned}\mathcal{M}_2 &= \sum_b \frac{1}{b}\sum_{m_1,m_2}\frac{x_{bm_1}x_{bm_2}}{\sqrt{m_1m_2}}\sum_{d\ell^2\leqslant x}\frac{1}{d\ell^2}A^h[\varepsilon_f^-\lambda_f(d^2)\lambda_f(m_1m_2)L'(f,\tfrac12)^2]\\ &= \sum_b \frac{1}{b}\sum_{m_1,m_2}\frac{x_{bm_1}x_{bm_2}}{\sqrt{m_1m_2}}\sum_{d\ell^2\leqslant x}\frac{1}{d\ell^2}\sum_{r|(m_1m_2,d^2)} A^h[\varepsilon_f^-\lambda_f\Big(\frac{m_1m_2d^2}{r^2}\Big)L'(f,\tfrac12)^2].\end{aligned}$$

Since

$$\frac{m_1m_2d^2}{r^2} \leqslant M^2x^2 \leqslant q^{2\kappa+2\Delta} < q^{1/2},$$

Proposition 18 is applicable, and we get a decomposition of $\mathcal{M}_2$ which is exactly similar to the one of M_2 given by Proposition 19: for some $\delta > 0$,

$$\mathcal{M}_2 = \frac{1}{12}\mathcal{M}_{21} - \frac14\mathcal{M}_{22} + \mathcal{M}_3 + O(q^{-\delta}) \tag{6.94}$$

where

$$\mathcal{M}_{21} = \sum_b\frac1b\sum_{m_1,m_2}\frac{x_{bm_1}x_{bm_2}}{m_1m_2}\sum_{d\ell^2\leqslant x}\frac{1}{(d\ell)^2}\sum_{r|(d^2,m_1m_2)} r\tau\Big(\frac{m_1m_2d^2}{r^2}\Big)\Big(\log\frac{\hat{Q}r^2}{m_1m_2d^2}\Big)^3 \tag{6.95}$$

$$\mathcal{M}_{22} = \sum_b\frac1b\sum_{m_1,m_2}\frac{x_{bm_1}x_{bm_2}}{m_1m_2}\sum_{d\ell^2\leqslant x}\frac{1}{(d\ell)^2}\sum_{r|(d^2,m_1m_2)} rT\Big(\frac{m_1m_2d^2}{r^2}\Big)\Big(\log\frac{\hat{Q}r^2}{m_1m_2d^2}\Big) \tag{6.96}$$

$$\mathcal{M}_3 = \sum_b\frac1b\sum_{m_1,m_2}\frac{x_{bm_1}x_{bm_2}}{m_1m_2}\sum_{d\ell^2\leqslant x}\frac{1}{(d\ell)^2}\sum_{r|(d^2,m_1m_2)} r\tau\Big(\frac{m_1m_2d^2}{r^2}\Big)P_1\Big(\log\frac{\hat{Q}r^2}{m_1m_2d^2}\Big). \tag{6.97}$$

6.2.3 Mutations of the second moment

Those quadratic forms $\mathcal{M}_{21}$, $\mathcal{M}_{22}$, $\mathcal{M}_3$ should be compared to $\mathcal{M}(\delta)$ of (5.54). The strategy to deal with them will be the same: after preliminary transformations, we remove the condition $d\ell^2 \leqslant x$ which brings a decomposition in terms of quadratic forms with mutative coefficients, as discussed in the Appendix to Section 4.3. After that, the path is the same as for M_2: a preferred term is selected, and optimization is performed for this term and $\mathcal{M}_1$, before final estimations produce the no less final result.

The alphabet has only 26 letters: lest notations become unwieldy, we recycle some of those of the previous sections, and enlist the fraktur letters $\mathfrak{i}$, $\mathfrak{j}$, $\mathfrak{k}$, $\mathfrak{l}$ as indices (the letters u, v, w will appear as coefficients, as in Chapter 5).

We define for convenience $t_0 = \tau$, $t_2 = T$. First, observe that each of $\mathcal{M}_{21}$, $\mathcal{M}_{22}$ and $\mathcal{M}_3$, after expanding the powers of logarithms, can be expressed as a sum over b of quadratic forms

$$\sum_{m_1,m_2} \frac{x_{bm_1}x_{bm_2}}{m_1m_2}\Big(\log\frac{\hat{Q}}{m_1m_2}\Big)^{\mathfrak{i}} \sum_{d\ell^2\leqslant x} \frac{(\log d^2)^{\mathfrak{k}}}{(d\ell)^2} \sum_{r\mid(m_1m_2,d^2)} r(\log r)^{\mathfrak{l}} t_{\mathfrak{j}}\Big(\frac{m_1m_2d^2}{r^2}\Big) \tag{6.98}$$

which we denote by $\mathcal{M}_x(\mathfrak{i},\mathfrak{j},\mathfrak{k},\mathfrak{l})$. The parameters satisfy $0 \leqslant \mathfrak{i}+\mathfrak{j}+\mathfrak{k}+\mathfrak{l} \leqslant 3$, $\mathfrak{j}$ is 0 or 2.

Although this looks forbidding enough, the main contribution will come from the case $\mathfrak{i}+\mathfrak{j} = 3$, $\mathfrak{k} = \mathfrak{l} = 0$, in which case the inner sums involve only multiplicative functions (except for T when $\mathfrak{j} = 2$), and for the others, mutativity will create sundry quadratic forms during the process of "diagonalization", as in (4.14), with very complicated coefficients ν, but yet small, and the loss of logarithms in the variables w_k, y_k will make all these terms smaller, as happened for M_2 before.

First we remove the constraint $d\ell^2 \leqslant x$. To do this, we separate m_1m_2/r and d^2/r using the identities

$$\sum_{r\mid(m,n)} r(\log r)^{\mathfrak{l}}\tau\Big(\frac{mn}{r}\Big) = \sum_{r\mid(m,n)} u_{\mathfrak{l}}(r)\tau\Big(\frac{m}{r}\Big)\tau\Big(\frac{n}{r}\Big)$$
$$\sum_{r\mid(m,n)} r(\log r)^{\mathfrak{l}}T\Big(\frac{mn}{r}\Big) = \sum_{r\mid(m,n)} u_{\mathfrak{l}}(r)\Big\{\tau\Big(\frac{m}{r}\Big)T\Big(\frac{n}{r}\Big) + T\Big(\frac{m}{r}\Big)\tau\Big(\frac{n}{r}\Big)\Big\}$$

where (compare with the function $u(s,r)$ of Chapter 5) we have put

$$u_{\mathfrak{l}}(r) = \sum_{ab=r} \mu(a)b(\log b)^{\mathfrak{l}};$$

the first of these is a consequence of (6.68), and the second is one of (6.49).

Therefore $\mathcal{M}_x(\mathfrak{i},0,\mathfrak{k},\mathfrak{l})$ is equal to

$$\sum_{m_1,m_2} \frac{x_{bm_1}x_{bm_2}}{m_1m_2}\Big(\log\frac{\hat{Q}}{m_1m_2}\Big)^{\mathfrak{i}} \sum_{r\mid m_1m_2} u_{\mathfrak{l}}(r)\tau\Big(\frac{m_1m_2}{r}\Big) \sum_{\substack{d\ell^2\leqslant x\\ r\mid d^2}} \frac{(\log d^2)^{\mathfrak{k}}}{(d\ell)^2}\tau\Big(\frac{d^2}{r}\Big)$$

and $\mathcal{M}_x(\mathfrak{i},2,\mathfrak{k},\mathfrak{l})$ is a sum of two terms, one of which is

$$\sum_{m_1,m_2} \frac{x_{bm_1}x_{bm_2}}{m_1m_2}\Big(\log\frac{\hat{Q}}{m_1m_2}\Big)^{\mathfrak{i}} \sum_{r\mid m_1m_2} u_{\mathfrak{l}}(r)T\Big(\frac{m_1m_2}{r}\Big) \sum_{\substack{d\ell^2\leqslant x\\ r\mid d^2}} \frac{(\log d^2)^{\mathfrak{k}}}{(d\ell)^2}\tau\Big(\frac{d^2}{r}\Big) \tag{6.99}$$

and the other the same with τ and T interchanged.

It is now clear that for each of the quadratic forms we can extend the summation in d and ℓ to infinity: indeed, we recognize

$$\sum_{\substack{d\ell^2\leqslant x\\ r|d^2}} \frac{(\log d^2)^{\mathfrak{k}}}{(d\ell)^2}\tau\Big(\frac{d^2}{r}\Big) = (-1)^{\mathfrak{k}} v_x^{(\mathfrak{k})}(1,r)\Big|_{t=0}$$

where $v_x(s,r)$ is defined in (5.55) and the estimation (5.57) of Lemma 22 extends to the derivatives (up to some logarithms) to put this removal into effect. Similarly,

$$\sum_{\substack{d\ell^2\leqslant x\\ r|d^2}} \frac{(\log d^2)^{\mathfrak{k}}}{(d\ell)^2}T\Big(\frac{d^2}{r}\Big) = (-1)^{\mathfrak{k}} \frac{d^2}{dt^2} v_x^{(\mathfrak{k})}(1,r)\Big|_{t=0}$$

(the dependence on t of $v_x(s,r)$ is not displayed in the notation), and the same remark applies.

Hence we remove x and get quadratic forms $\mathcal{M}(\mathfrak{i},\mathfrak{j},\mathfrak{k},\mathfrak{l})$. We write simply

$$v^{(\mathfrak{k})}(r) = v^{(\mathfrak{k})}(1,r), \quad v^{[\mathfrak{k}]}(r) = \frac{d^2}{dt^2} v^{(\mathfrak{k})}(1,r)\Big|_{t=0}.$$

(the latter can only occur with $\mathfrak{k}=1$ when $\mathfrak{i}=0$, $\mathfrak{j}=2$, $\mathfrak{l}=0$, and contributes to the error term at the end; it can be safely ignored by the reader). With these notations, it holds

$$\mathcal{M}(\mathfrak{i},0,\mathfrak{k},\mathfrak{l}) = \sum_{m_1,m_2} \frac{x_{bm_2}x_{bm_2}}{m_1m_2}\Big(\log\frac{\hat{Q}}{m_1m_2}\Big)^{\mathfrak{i}} \sum_{r|m_1m_2} \tau\Big(\frac{m_1m_2}{r}\Big) u_{\mathfrak{l}}(r) v^{(\mathfrak{k})}(r)$$

(and a cognate expression for $\mathcal{M}(\mathfrak{i},2,\mathfrak{k},\mathfrak{l})$).

Since $v(s,r)$ was computed in Lemma 22, and found to be the product of $v(s,1)$ and a multiplicative function, it follows from Leibniz's rule that the derivatives $v^{(\mathfrak{k})}(1,r)$ are mutative in the sense of the Appendix to Section 4.3, and the yoga described there can be used to diagonalize the $\mathcal{M}(\mathfrak{i},\mathfrak{j},\mathfrak{k},\mathfrak{l})$.

These mutativity properties of $u_{\mathfrak{l}}$ and $v^{(\mathfrak{k})}$ take the following simple form. For $u_{\mathfrak{l}}$, from the definition we have

$$u_{\mathfrak{l}}(mn) = \sum_{i+j=\mathfrak{l}} \binom{\mathfrak{l}}{i} u_i(n) u_j(n).$$

for $(m,n)=1$. On the other hand, writing

$$v(s,r) = v(s,1)\overline{v}(s,r)$$

with $\overline{v}$ multiplicative, we see that $v^{(\mathfrak{k})}$ is first a linear combination of the functions

$$\overline{v}^{(j)}(r) = \overline{v}^{(j)}(1,r)$$

for $j\leqslant\mathfrak{k}$, and each $\overline{v}^{(\mathfrak{k})}(r)$ satisfies in turn

$$\overline{v}^{(\mathfrak{k})}(mn) = \sum_{i+j=\mathfrak{k}} \binom{\mathfrak{k}}{i} \overline{v}^{(i)}(m)\overline{v}^{(j)}(n)$$

for $(m,n)=1$, by differentiating at $s=1$ the formula

$$\overline{v}(s,mn)=\overline{v}(s,m)\overline{v}(s,n).$$

Similar mutations are observed for $v^{[\mathfrak{k}]}$. The following lemma will be useful.

Lemma 44. *For all $\mathfrak{k}\geqslant 0$, $\mathfrak{l}\geqslant 0$, we have*

$$u_{\mathfrak{l}}(r)\overline{v}^{(\mathfrak{k})}(r)\ll\frac{r\tau(r)^2(\log r)^{\mathfrak{k}+\mathfrak{l}}}{N(r)^2}$$
$$u_{\mathfrak{l}}(r)\overline{v}^{[\mathfrak{k}]}(r)\ll\frac{r\tau(r)^2(\log r)^{\mathfrak{k}+\mathfrak{l}}}{N(r)^2}$$

the implied constants depending only on $\mathfrak{k}$ and $\mathfrak{l}$.

Proof. This follows very quickly from the computation of $\overline{v}$ in Lemma 22 and the definition of $u_{\mathfrak{l}}$. □

This mutativity enables us to further decompose $\mathcal{M}(\mathfrak{i},0,\mathfrak{k},\mathfrak{l})$ as a linear combination of forms of the type

$$\sum_{m_1,m_2}\frac{x_{bm_1}x_{bm_2}}{m_1m_2}\Big(\log\frac{\hat{Q}}{m_1m_2}\Big)^{\mathfrak{i}}\overline{w}(\mathfrak{k},\mathfrak{l};m_1m_2)\tag{6.100}$$

where we have defined

$$\overline{w}(\mathfrak{k},\mathfrak{l};m)=\sum_{ab=m}\tau(a)u_{\mathfrak{l}}(b)\overline{v}^{(\mathfrak{k})}(b)\tag{6.101}$$

which has the multiplicativity property that $\overline{w}(\mathfrak{k},\mathfrak{l};mn)$, for $(m,n)=1$, is a linear combination of the functions $\overline{w}(\mathfrak{k}_1,\mathfrak{l}_1;m)\overline{w}(\mathfrak{k}_2,\mathfrak{l}_2;n)$ with $\mathfrak{k}_1+\mathfrak{k}_2=\mathfrak{k}$, $\mathfrak{l}_1+\mathfrak{l}_2=\mathfrak{l}$.

Thus, in (6.100), extracting the common divisor of m_1 and m_2, then removing the resulting coprimality condition by Möbius inversion, and collecting the variables, all as in the Appendix to Section 4.3 (except that we keep the logarithm attached to the variables x_{bm_1}, x_{bm_2} until the very end, and then expand again the power) will bring for $\mathcal{M}(\mathfrak{i},0,\mathfrak{k},\mathfrak{l})$ a linear combination of expressions of the type

$$(\log\hat{Q})^{\mathfrak{i}}\sum_k\nu(k)w_ky_k$$

with

$$\nu(k)=\frac{1}{k}\sum_{ad=k}\frac{\mu(d)h_1(d)^2h_2(a^2)}{ad}(\log a)^{\mathfrak{i}_1}(\log d)^{\mathfrak{i}_2}$$
$$w_k=\sum_m h_3(m)(\log m)^{\mathfrak{i}_3}x_{bkm}$$
$$y_k=\sum_m h_4(m)(\log m)^{\mathfrak{i}_4}x_{bkm}$$

for some $\mathfrak{i}_1$, $\mathfrak{i}_2$, $\mathfrak{i}_3$, $\mathfrak{i}_4$ with

$$0\leqslant\mathfrak{i}_1+\mathfrak{i}_2\leqslant\mathfrak{i},$$

and $h_i = \overline{w}(\mathfrak{k}_i, \mathfrak{l}_i)$ for some $\mathfrak{k}_i \leqslant \mathfrak{k}$, $\mathfrak{l}_i \leqslant \mathfrak{l}$. The total "weight" in logarithms must be at most 3,

$$\mathfrak{i} + \mathfrak{i}_1 + \mathfrak{i}_2 + \mathfrak{i}_3 + \mathfrak{i}_4 + \mathfrak{k}_1 + \mathfrak{l}_1 + \mathfrak{k}_2 + \mathfrak{l}_2 + \mathfrak{k}_3 + \mathfrak{l}_3 + \mathfrak{k}_4 + \mathfrak{l}_4 \leqslant 3.$$

In practice, every logarithm from the store of 3 available which is diverted into one of the coefficient functions h_i (so $\mathfrak{k}_i + \mathfrak{l}_i > 0$) or ν (so $\mathfrak{i}_1 + \mathfrak{i}_2 > 0$), has the effect of making the given form of lower order of magnitude than the main term Π corresponding to $\mathfrak{i} + \mathfrak{i}_3 + \mathfrak{i}_4 = 3$, when evaluated for the chosen mollifier (x_m): the estimate

$$(\log \hat{Q})^{\mathfrak{i}} \sum_k \nu(k) w_k y_k = O\Big(\Pi \frac{(\log\log q)^B}{\log q}\Big) \tag{6.102}$$

will hold, for some absolute constant $B > 0$, except in this case.

6.2.4 The preferred quadratic form, II

In this section, we will incorporate back the sum over b into the quadratic forms; since this is everywhere present and propagates into all computations, the adaptation of the formulae and principles of the previous section is immediate.

We now start again by selecting a preferred quadratic form: in $\mathcal{M}_{21}$, we select $\mathcal{M}(3,0,0,0)$. This only involves the multiplicative function $u = u_0$ and the function $v = v^{(0)}$, which by Lemma 22 is given by

$$v(r) = \frac{\zeta(2)^4}{\zeta(4)} \overline{v}(r), \quad \overline{v}(r) = N(r)^{-2} \prod_{p \| r} \frac{\tau(p)}{1 + p^{-2}}.$$

Thus, only the constant factor comes out when performing the steps described previously, and we have the expression

$$\Pi = \mathcal{M}(3,0,0,0) = \frac{\zeta(2)^4}{\zeta(4)} \sum_b \frac{1}{b} \sum_{m_1, m_2} \frac{x_{bm_1} x_{bm_2}}{m_1 m_2} \Big(\log \frac{\hat{Q}}{m_1 m_2}\Big)^3 \overline{w}(m_1 m_2)$$

where

$$\overline{w}(m) = \overline{w}(0,0;m)$$

(compare (5.58)). Recall that

$$\overline{w}(m) = \sum_{ab=m} \tau(a) u(b) \overline{v}(b).$$

We expand the logarithm

$$\Big(\log \frac{\hat{Q}}{m_1 m_2}\Big)^3$$

in Π and thus – as in the case of M_2 –, we obtain the decomposition

$$\Pi = \tilde{\Pi}(3,0,0) - 6\tilde{\Pi}(2,1,0) + 6\tilde{\Pi}(1,1,1) + 6\tilde{\Pi}(1,2,0) - 6\tilde{\Pi}(0,1,2) - 2\tilde{\Pi}(0,0,3). \tag{6.103}$$

with

$$\tilde{\Pi}(\mathfrak{i},\mathfrak{j},\mathfrak{k}) = \frac{\zeta(2)^4}{\zeta(4)} (\log \hat{Q})^{\mathfrak{i}} \sum_b \frac{1}{b} \sum_{m_1, m_2} \frac{\overline{w}(m_1 m_2)}{m_1 m_2} (\log m_1)^{\mathfrak{j}} (\log m_2)^{\mathfrak{k}} x_{bm_1} x_{bm_2}.$$

We state the output of the diagonalization procedure.

Lemma 45. *For all* $\mathfrak{i}$, $\mathfrak{j}$, $\mathfrak{k}$, *we have*

$$\tilde{\Pi}(\mathfrak{i},\mathfrak{j},\mathfrak{k}) = \frac{\zeta(2)^4}{\zeta(4)}(\log \hat{Q})^{\mathfrak{i}} \sum_k \tilde{\nu}(k) y_k^{(\mathfrak{j})} y_k^{(\mathfrak{k})}$$

with

$$\tilde{\nu}(k) = \frac{1}{k}\sum_{abd=k} \frac{\mu(d)\overline{w}(d)^2\overline{w}(a^2)}{ad}$$
$$y_k^{(\mathfrak{j})} = \sum_m \frac{\overline{w}(m)}{m}(\log m)^{\mathfrak{j}} x_{km},$$

(and we let as usual $y_k = y_k^{(0)}$*).*

Proof. This is what has been described before. □

Notice that (x_m) is supported on squarefree integers if and only if (y_k) is also.

We will choose the mollifier by optimizing $\tilde{\Pi}(3,0,0)$, which we simply denote by $\tilde{\Pi}$, with respect to the linear form $\mathcal{M}_1$.

6.2.5 Optimization of the preferred form

We will first prove some preliminary lemmas, and some more notation is needed. Let g be the Dirichlet convolution inverse of $\overline{w}$. The change of variable from x_m to y_k is inverted by

$$x_m = \sum_k \frac{g(k)}{k} y_{km}. \tag{6.104}$$

Furthermore, let j and j_1 be the arithmetic functions

$$j(k) = \frac{1}{k}\sum_{ab=k} g(a)d_{-1}(b), \quad j_1(k) = \frac{1}{k}\sum_{ab=k} g(a)d_{-1}(b)\Big(\log\frac{\hat{q}}{b}\Big) \tag{6.105}$$

so that the (main term of the) linear form $\mathcal{M}_1$ is expressed by

$$\mathcal{M}_1 = \zeta(2)^2 \sum_k j_1(k)y_k + 2\zeta(2)\zeta'(2)\sum_k j(k)y_k. \tag{6.106}$$

Lemma 46. *The multiplicative function* $\overline{w}$ *satisfies*

$$\overline{w}(k) = \tau(k)\prod_{p|k}\frac{1+p^{-1}}{1+p^{-2}} \tag{6.107}$$

for all squarefree integers k *and*

$$\overline{w}(p^2) = 3 + \frac{1}{p} - \frac{1}{p^2} + 4\frac{p-1}{p^2-1} \tag{6.108}$$

for all primes p.

The multiplicative function $\tilde{\nu}$ satisfies

$$\tilde{\nu}(k) = \frac{1}{k}\prod_{p|k} A(p^{-1}) \tag{6.109}$$

for all squarefree integers k, where A is the rational function

$$A = \frac{(1-X^2)^3}{(1+X^2)^2}.$$

Proof. Multiplicativity is obvious in all cases, and the rest is a matter of computing without mistakes. Note that the first two statements are also derived in the proof of Lemma 24.

The last computation gives a mysteriously clean result; recall that

$$\tilde{\nu}(p) = \frac{1}{p}\Big(1 + \frac{\overline{w}(p^2)}{p} - \frac{\overline{w}(p)^2}{p}\Big)$$

and it doesn't appear at first sight that (6.107) and (6.108) will give such a simplification. □

Lemma 47. *The arithmetic function j is multiplicative with*

$$j(k) = \frac{\mu(k)}{k}\prod_{p|k} B(p^{-1}) \tag{6.110}$$

for all squarefree integers k, where B is the rational function

$$B = \frac{(1-X^2)(1+X)}{1+X^2}.$$

The function j_1 satisfies

$$j_1(k) = j(k)(\log \hat{q}k + O(1)). \tag{6.111}$$

Proof. The first statement is another direct computation, using the fact that

$$g(p) = -\overline{w}(p) = -2\frac{1+p^{-1}}{1+p^{-2}}.$$

We have $j_1(k) = (\log \hat{q})j(k) - j_2(k)$ where

$$j_2(k) = \frac{1}{k}\sum_{ab=k} g(a)d_{-1}(b)(\log b)$$

and from the multiplicativity of j, we see that j_2 satisfies the mutativity property

$$j_2(mn) = j(m)j_2(n) + j_2(m)j(n)$$

hence by induction, for k squarefree,

$$\begin{aligned} j_2(k) &= \sum_{p|k} j_2(p)j\Big(\frac{k}{p}\Big) = j(k)\sum_{p|k}(1+p^{-1})(\log p)j(p)^{-1} \\ &= -j(k)\sum_{p|k}(\log p)\frac{1+p^{-2}}{1-p^{-2}} = -j(k)(\log k + O(1)) \end{aligned}$$

using (6.110), which was just proved, along the way. □

The last lemma will be used for summation by part.

Lemma 48. *The Dirichlet series*

$$\sum_{k\geqslant 1}\frac{\mu(k)^2 j(k)^2}{\tilde{\nu}(k)}k^{-s}$$

is absolutely convergent for $\mathrm{Re}\,(s) > 0$ *and extends to a meromorphic function for* $\mathrm{Re}\,(s) > -1$ *with a simple pole of order* 1 *at* $s = 0$ *with residue*

$$R = \frac{\zeta(2)}{\zeta(4)}.$$

Proof. The previous efforts give the Euler product

$$\begin{aligned}\sum_{k\geqslant 1}\frac{\mu(k)^2 j(k)^2}{\tilde{\nu}(k)}k^{-s} &= \prod_p\Big(1+p^{-(s+1)}\frac{(1+p^{-2})^2}{(1-p^{-2})^3}\cdot\frac{(1-p^{-2})^2(1+p^{-1})^2}{(1+p^{-2})^2}\Big)\\ &= \prod_p\Big(1+p^{-(s+1)}\frac{1+p^{-1}}{1-p^{-1}}\Big)\\ &= \frac{\zeta(s+1)}{\zeta(2(s+1))}\prod_p\Big(1+\frac{2p^{-(s+1)}}{(1+p^{-(s+1)})(p-1)}\Big)\end{aligned}$$

and the second Euler product is indeed absolutely convergent for $\mathrm{Re}\,(s) > -1$. Hence the lemma holds, with

$$\begin{aligned}R &= \frac{1}{\zeta(2)}\prod_p\Big(1+\frac{2}{p^2-1}\Big) = \frac{1}{\zeta(2)}\prod_p\frac{p^2+1}{p^2-1}\\ &= \frac{1}{\zeta(2)}\cdot\frac{\zeta(2)^2}{\zeta(4)}\end{aligned}$$

as announced. □

We can see from these computations that

$$\overline{w}(p) = \tau(p) + O(p^{-1}),\quad j_1(p) = \mu(p)(\log\hat{q}p) + O(1),\quad \tilde{\nu}(p) = \frac{1}{p} + O(p^{-2})$$

so those coefficients are very close to the coefficients used in proving the harmonic case of the theorem. This is of course not much of a surprise.

We now choose (y_k) to optimize $\tilde{\Pi}$ with respect to $\mathcal{M}_1$: let

$$y_k = \begin{cases}\mu(k)^2\dfrac{j_1(k)}{\tilde{\nu}(k)},\ \text{for } k\leqslant M\\ 0,\ \text{for } k > M.\end{cases}\tag{6.112}$$

This is supported on squarefree numbers and we immediately see that the corresponding (x_m) satisfies condition (6.5).

Proposition 20. *For this choice of mollifier, it holds*

$$\mathcal{M}_1 = \frac{\zeta(2)^3}{\zeta(4)}\mathfrak{m}_1(\Delta)(\log q)^3 + O((\log q)^2) \tag{6.113}$$

where $\mathfrak{m}_1$ *is the polynomial defined in (6.7), and*

$$\Pi = \frac{\zeta(2)^5}{\zeta(4)^2}\mathfrak{m}_{21}(\Delta)(\log q)^6 + O((\log q)^5(\log\log q)) \tag{6.114}$$

where $\mathfrak{m}_{21}$ *is the polynomial defined in (6.10).*

In other words, apart from constant factors, the first moment and the chosen main term of the second moment are the same as the corresponding harmonic ones: compare this with Lemma 37 and Corollary 4.

Proof. **First step.** From (6.106), we compute

$$\begin{aligned}
\mathcal{M}_{21} &= \zeta(2)^2 \sum_{k\leqslant M} j_1(k)y_k + 2\zeta(2)\zeta'(2)\sum_{k\leqslant M} j(k)y_k \\
&= \zeta(2)^2 \sum_{k\leqslant M} \frac{\mu(k)^2 j_1(k)^2}{\tilde\nu(k)} + 2\zeta(2)\zeta'(2)\sum_{k\leqslant M}\frac{\mu(k)^2 j(k) j_1(k)}{\tilde\nu(k)} \\
&= \zeta(2)^2 \sum_{k\leqslant M} \frac{\mu(k)^2 j(k)^2}{\tilde\nu(k)}\big((\log \hat q k)^2 + O(\log \hat q k)\big) \\
&\qquad + 2\zeta(2)\zeta'(2)\sum_{k\leqslant M}\frac{\mu(k)^2 j(k)^2}{\tilde\nu(k)}(\log \hat q k + O(1))
\end{aligned}$$

by Lemma 47

$$= \frac{\zeta(2)^3}{\zeta(4)}\mathfrak{m}_1(\Delta)(\log q)^3 + O((\log q)^2)$$

(by partial summation from Lemma 48, see Lemma 37). This proves the first statement.

Second step. We claim that for $\mathfrak{i} = 1, 2, 3$, it holds

$$y_k^{(\mathfrak{i})} = c_i \frac{\mu(k)j(k)}{\tilde\nu(k)}\Big(\log\frac{M}{k}\Big)^{\mathfrak{i}}(\log \hat q^{\mathfrak{i}+1} M^{\mathfrak{i}} k) + O\Big(\frac{j(k)}{\tilde\nu(k)}(\log q)^{\mathfrak{i}}(\log\log q)^B\Big) \tag{6.115}$$

(for some absolute constant B), where $c_1 = -1$, $c_2 = \frac{1}{3}$, $c_3 = 0$, namely the variables $y_k^{(\mathfrak{i})}$ behave just like their harmonic analogues, see Lemma 38.

We write again

$$y_k^{(\mathfrak{i})} = \sum_{\ell\leqslant M/k}\frac{\overline{w}(\ell)}{\ell}\Lambda_i(\ell)y_{k\ell}$$

and perform the same calculations leading to Lemma 38, involving multiple sums over primes. But since

$$\overline{w}(\ell) = \tau(\ell) + O(\ell^{-1}), \quad j_1(\ell) = \mu(\ell)(\log \hat q\ell), \quad \tilde\nu(\ell) = \ell^{-1} + O(\ell^{-2})$$

we see that applying those approximation in a sum over numbers with exactly $\mathfrak{j} \leqslant \mathfrak{i}$ prime factors provides the same first term as there, while in the others the sum over at least one of the prime variables involved becomes convergent, so it can be summed trivially, or one logarithm disappears, and in any case, those other terms are at least one logarithm smaller. For instance, with one prime ($\mathfrak{i} = 1$)

$$y_k^{(1)} = \sum_{\ell \leqslant M/k} \frac{\tau(\ell)}{\ell} (\log \ell) \frac{\mu(k\ell) j_1(k\ell)}{\tilde{\nu}(k\ell)} + O\Big(\sum_{\ell \leqslant M/k} \frac{\tau(\ell)(\log \ell)}{\ell^2} \frac{\mu(k\ell) j_1(k\ell)}{\tilde{\nu}(k\ell)} \Big);$$

the estimated term is clearly

$$O\Big(\frac{j_1(k)}{\tilde{\nu}(k)}\Big) = O\Big(\frac{j(k)}{\tilde{\nu}(k)} (\log q)\Big),$$

now continue with j_1, etc...

Third step. The formula (6.114) holds.

Using the result of the second step (and the first step for $\tilde{\Pi}$ itself), this is the same as the proof of Lemma 39, using Lemma 48 to perform the summation by parts, and the decomposition (6.103) to assemble the various parts. □

6.2.6 The second part of the main term

We now turn our attention to the quadratic form $\mathcal{M}_{22}$ given by (6.96); it gives rise, by the transformations described in Section 6.2.3, to various terms, only one of which will contribute to the main term, namely that derived from (6.99) with $\mathfrak{i} = 1$ and $\mathfrak{k} = \mathfrak{l} = 0$, which we denote by $\mathcal{M}'_{22}$:

$$\mathcal{M}'_{22} = \frac{\zeta(2)^4}{\zeta(4)} \sum_b \frac{1}{b} \sum_{m_1, m_2} \frac{x_{bm_1} x_{bm_2}}{m_1 m_2} \Big(\log \frac{\hat{Q}}{m_1 m_2} \Big) \varpi(m_1 m_2)$$

where the arithmetic function ϖ is

$$\varpi(m) = \sum_{ab=m} T(a) u(b) v(b).$$

Using the mutativity of ϖ

$$\varpi(m_1 m_2) = \varpi(m_1) \overline{w}(m_2) + \overline{w}(m_1) \varpi(m_2)$$

which follows from (6.48), the diagonalization yields

$$\mathcal{M}'_{22} = 2 \frac{\zeta(2)^4}{\zeta(4)} \sum_k \tilde{\nu}(k) \sum_{m_1, m_2} \frac{\overline{w}(m_1) \varpi(m_2)}{m_1 m_2} x_{km_1} x_{km_2} \Big(\log \frac{\hat{Q}}{m_1 m_2} \Big) + (\text{other terms})$$

where the other contributions will be of smaller order (the function T, of "weight" 2, appears only in the coefficient function $\nu(k)$ attached to k). We keep the notation $\mathcal{M}'_{22}$ for the first part.

This decomposes, exactly as in the harmonic context, as

$$\mathcal{M}'_{22} = 2(\tilde{\Pi}(1,0,0) - \tilde{\Pi}(0,1,0) - \tilde{\Pi}(0,0,1))$$

where

$$\tilde{\Pi}(\mathfrak{i},\mathfrak{j},\mathfrak{k}) = \frac{\zeta(2)^4}{\zeta(4)}(\log \hat{Q})^{\mathfrak{i}} \sum_k \tilde{\nu}(k) y_k^{(\mathfrak{j})} z_k^{(\mathfrak{k})}$$

and

$$z_k^{(\mathfrak{k})} = \sum_m \frac{\varpi(m)}{m}(\log m)^{\mathfrak{k}} x_{km}.$$

Proposition 21. *With the mollifier given by (6.112), we have*

$$\mathcal{M}_{22}' = \frac{\zeta(2)^5}{\zeta(4)^2}\mathfrak{m}_{22}(\Delta)(\log q)^6 + O((\log q)^5(\log\log q)^B)$$

for some (absolute) $B > 0$.

So this term behaves again as its harmonic counterpart (counterpoint?).

Proof. **First step.** We express $z_k = z_k^{(0)}$ and $z_k^{(1)}$ in terms of y_k:

$$z_k = 2\sum_{\ell \leqslant M/k} \frac{(\log \ell)\Lambda(\ell)}{\ell} y_{kl}$$

$$z_k^{(1)} = \sum_{\ell \leqslant M/k} \frac{\overline{w}(\ell)\Lambda(\ell)}{\ell} z_{k\ell} + \sum_{\ell \leqslant M/k} \frac{\varpi(\ell)\Lambda(\ell)}{\ell} y_{k\ell}$$

(compare Lemma 40).

The first result, exactly the same as the corresponding one in Lemma 40, derives from the Dirichlet series identities

$$\sum_m \varpi(m)m^{-s} = \Big(\sum_m T(m)m^{-s}\Big)\Big(\sum_m u(m)v(m)m^{-s}\Big),$$
$$\sum_m \overline{w}(m)m^{-s} = \Big(\sum_m \tau(m)m^{-s}\Big)\Big(\sum_m u(m)v(m)m^{-s}\Big),$$

since to express z_k in terms of y_k we need only compute the Dirichlet convolution $\varpi \star g$ corresponding to dividing those two series. But then the Dirichlet series for uv simplifies, which means the coefficients for z_k are indeed as in Lemma 40.

The computation of $z_k^{(1)}$ is the also the same as there.

Second step. We have

$$z_k = -y_k^{(2)} + O\Big(\frac{j(k)}{\tilde{\nu}(k)}(\log q)^2 \log\log q\Big)$$
$$z_k^{(1)} = O\Big(\frac{j(k)}{\tilde{\nu}(k)}(\log q)^3 \log\log q\Big)$$

(compare Lemma 41).

Observe that for p prime

$$\varpi(\ell) = T(\ell), \quad \overline{w}(\ell) = \tau(\ell) + O(\ell^{-1})$$

and proceed as in the proof of that lemma: the claim follows for the same reason explained in the proof of Proposition 20.

Third step. The lemma holds: this is a consequence of the previous step, and the method of proof in the harmonic case, with Lemma 48 for the summations by part. □

6.2.7 The residual quadratic forms

It remains to prove that the (many) quadratic forms that were disregarded during the transformation process are, when evaluated for the chosen mollifier (x_m), of smaller order of magnitude.

They are the quadratic forms described in (6.100), with the sum over b added, plus some other similar ones coming from $\mathcal{M}_{22}$, which we have not written down explicitly.

To estimate their value, one has to express in terms of y_k the linear forms

$$\sum_m h(m)(\log m)^{\mathfrak{i}} x_{km}$$

described below (6.100), with the sum over b inserted. This amounts to computing the Dirichlet convolutions $g \star h(\log)^{\mathfrak{i}}$, and the claim that those residual terms are smaller is proved by showing that the Dirichlet generating series has a pole of order at most $\mathfrak{i}$ at $s = 0$. Since g, being the inverse of $\overline{w}$, has a zero of order 2 at $s = 0$ (see 46), and the product $h(\log)^{\mathfrak{i}}$ corresponds to the $\mathfrak{i}$-th derivative of $h = \overline{w}(\mathfrak{k}, \mathfrak{l})$, only the latter needs to be computed, and it will be found to have a pole of order 2 at $s = 0$, just as $\overline{w}$ does, because in the definition (6.101) of $\overline{w}(\mathfrak{k}, \mathfrak{l})$, the factor $u_{\mathfrak{l}}(b)\overline{v}^{(\mathfrak{k})}(b)$ is quite small (by Lemma 44), and only perturbs the behavior of $\overline{w}(\mathfrak{k}, \mathfrak{l})$, compared to that of $\overline{w}$, beyond the line $\mathrm{Re}\,(s) = 0$.

Then the coefficients $\tilde{\nu}(k)$ are bounded by

$$\tilde{\nu}(k) \ll \frac{(\log\log k)^B}{k}$$

for some absolute $B \geqslant 0$, using again Lemma 44, and from this, the definition of the mollifier and Lemma 48, the required estimate will follow by summation by part. The quadratic forms coming from $\mathcal{M}_{22}$ are treated exactly in the same way.

The task of giving any supplementary details is left to the reader, if more is felt to be needed at this point: as already mentioned, getting a supplementary contribution would be quite remarkable, and this sketch shows that everything can be checked.

6.2.8 Conclusion

From (6.93) and Proposition 20, for the first moment, and (6.94), Propositions 20 and 21 for the second moment, we now conclude that for $\Delta + 2\kappa < \frac{1}{4}$ we have

$$\mathcal{M}_1 = \frac{\zeta(2)^3}{\zeta(4)}\mathfrak{m}_1(\Delta)(\log q)^3 + O((\log q)^2(\log\log q)^B)$$
$$\mathcal{M}_2 = \frac{\zeta(2)^5}{\zeta(4)^2}\mathfrak{m}_2(\Delta)(\log q)^6 + O((\log q)^5(\log\log q)^B)$$

for some $B > 0$. Therefore, by (6.91) and (6.92)

$$M_1^n = \frac{\zeta(2)^2}{\zeta(4)}\dim J_0(q)\mathfrak{m}_1(\Delta)(\log q)^3 + O(q(\log q)^2(\log\log q)^B)$$
$$M_2^n = \frac{\zeta(2)^4}{\zeta(4)^2}\dim J_0(q)\mathfrak{m}_2(\Delta)(\log q)^6 + O(q(\log q)^5(\log\log q)^B)$$

and finally, for all q prime (large enough), we obtain

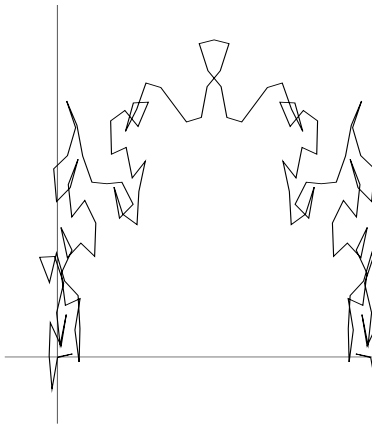
$$|\{f \in S_2(q)^* \mid \varepsilon_f = -1,\ L'(f, \tfrac{1}{2}) \neq 0\}| \geqslant \frac{(M_1^n)^2}{M_2^n}$$
$$= \frac{\mathfrak{m}_1(\Delta)^2}{\mathfrak{m}_2(\Delta)} \dim J_0(q) + O\Big(q \frac{(\log\log q)^B}{\log q}\Big).$$

We know already that

$$\frac{\mathfrak{m}_1(\frac{1}{4})^2}{\mathfrak{m}_2(\frac{1}{4})} = \frac{19}{54}$$

and so by letting Δ go to $\frac{1}{4}$, this inequality is the goal (6.1) of this chapter (in a stronger version), and thus the proof of Theorem 6 is completed, through the non-vanishing Theorem 10.

This probably calls for a celebratory drink.



Appendix: Extending the mollifier

In this brief appendix, we sketch how some of the arguments of the previous chapter can be extended, using techniques of Iwaniec-Sarnak [IS2]. The effect of this is that a mollifier of length q^Δ with $\Delta < \frac{1}{2}$ can be used, instead of merely $\Delta < \frac{1}{4}$; this yields a further improvement of the constant in the non-vanishing theorem. Precisely, we have

Proposition 22. *The approximate formula of Proposition 19 holds for all $M = q^\Delta$ with $\Delta < \frac{1}{2}$, possibly with a different polynomial P_1 in M_3 (still of degree at most 2).*

From this we deduce immediately:

Corollary 6. *For any $\varepsilon > 0$, and any prime number q large enough in terms of ε, we have*

$$|\{f \in S_2(q)^* \mid L(f, \tfrac{1}{2}) = 0,\ L'(f, \tfrac{1}{2}) \neq 0\}| \geqslant \Big(\frac{7}{16} - \varepsilon\Big) \dim J_0(q),$$

and

$$\operatorname{rank} J_0(q) \geqslant \Big(\frac{7}{16} - \varepsilon\Big) \dim J_0(q).$$

Indeed, the approximate formula for M_1 was already valid in the larger range $\Delta < \frac{1}{2}$, and the deduction of (6.1) from Proposition 19 did not use the hypothesis $\Delta < \frac{1}{4}$ in any other way (except in Lemma 43, which checks the applicability of the weight lifting technique of Chapter 3, and which can be immediately extended to this situation). Now we simply compute that

$$\frac{\mathfrak{m}_1(\frac{1}{2})^2}{\mathfrak{m}_2(\frac{1}{2})} = \frac{7}{16} = 0.4375.$$

Sketch of the proof.

Only the use of Lemma 11 in (6.29) (when considering $A^h[\varepsilon_f^- \lambda_f(m) L'(f, \frac{1}{2})^2]$), which leads to (6.30), has to be revised, since the approximate formula (17) for $X(m)$ is meaningful for $\Delta < \frac{1}{2}$ (recall that is is applied for $m = m_1 m_2$, $m_i \leqslant M$).

This means we must reconsider the decomposition of the Delta symbol for odd forms: for $m < q$, from the Petersson formula and Lemma 10 we have

$$\Delta_-(m,n) = \delta(m,n) - \mathcal{J}(m,n) + \mathcal{J}'(m,n) + O\Big(\frac{\sqrt{mn}}{q^2}(\log q)^2\Big)$$

and the error term is now good enough for $\Delta < \frac{1}{2}$. Of course, the part involving $\mathcal{J}'(m,n)$ is the one leading to $X(m)$, and this doesn't require any modification. So we investigate more carefully the remainder term with $\mathcal{J}(m,n)$.

Let us denote this by $\mathcal{X}(m)$, so

$$\begin{aligned}\mathcal{X}(m) &= \sum_{n \geqslant 1} \frac{\tau(n)}{\sqrt{n}} W\Big(\frac{4\pi^2 n}{q}\Big) \mathcal{J}(n,m) \\ &= \frac{2\pi}{q} \sum_{r \geqslant 1} \frac{1}{r} \sum_{n \geqslant 1} \tau(n) S(m,n;qr) t(n)\end{aligned}$$

with

$$t(x) = \frac{1}{\sqrt{x}} J_1\Big(\frac{4\pi\sqrt{mx}}{qr}\Big) W\Big(\frac{4\pi^2 x}{q}\Big).$$

Opening the Kloosterman sum $S(m,n;qr)$, we get

$$\mathcal{X}(m) = \frac{2\pi}{q}\sum_{r\geqslant 1}\frac{1}{r}\sideset{}{^*}\sum_{x \bmod qr} e\Big(\frac{m\overline{x}}{qr}\Big)\sum_{n\geqslant 1}\tau(n)e\Big(\frac{nx}{qr}\Big)t(n)$$

and to the inner sum over n we apply Proposition 16 and get

$$\begin{aligned}\mathcal{X}(m) = \frac{2\pi}{q}\sum_{r\geqslant 1}\frac{1}{r}\times\Big\{&\frac{2}{qr}S(m,0;qr)\int_0^{+\infty}(\log\frac{\sqrt{x}}{qr}+\gamma)t(x)dx\\ &-\frac{2\pi}{qr}\sum_{h\geqslant 1}\tau(h)S(m-h,0;qr)\int_0^{+\infty}Y_0\Big(\frac{4\pi\sqrt{hx}}{qr}\Big)t(x)dx\\ &+\frac{4}{qr}\sum_{h\geqslant 1}\tau(h)S(m+h,0;qr)\int_0^{+\infty}K_0\Big(\frac{4\pi\sqrt{hx}}{qr}\Big)t(x)dx\Big\}.\end{aligned}$$

This is reminiscent of the treatment of $X'(m)$ in Section 6.1.3, but now the main contribution will not come from the first integral, but from the frequency $h=m$ in the second sum, where the Ramanujan sum degenerates completely

$$S(0,0;qr)=\varphi(qr)$$

and thus no cancellation occurs. We will compute exactly this contribution, but will not justify completely here that the remaining terms are smaller. This is done (in more precise form) in the forthcoming paper [IS2], for the case of the second moment of the special values $L(f,\frac{1}{2})$, but there is no important difference. Actually, there is again a technical difficulty of convergence due to the fact that the weight is 2 and J_1 doesn't go to zero quickly enough near 0; it can be treated by inserting in $t(x)$ a test function $\xi(x)$ vanishing near 0 before performing the Voronoï summation, exactly as before.

However we can show quickly that the first term is small: using simply

$$|S(m,0;qr)|\leqslant qr,\quad J_1(x)\ll x$$

and Lemma 27 we find

$$\begin{aligned}\frac{4\pi}{q^2}\sum_{r\geqslant 1}\frac{1}{r^2}S(m,0;qr)\int_0^{+\infty}(\log\frac{\sqrt{x}}{qr}+\gamma)t(x)dx &\ll \frac{(\log q)^4}{q}\sum_{r\geqslant 1}\frac{1}{r}\int_0^q\frac{\sqrt{mx}}{qr}\frac{dx}{\sqrt{x}}\\ &\ll\frac{\sqrt{m}}{q}(\log q)^4\end{aligned}$$

which (although very crude) is good enough to extend the mollifier to $\Delta=\frac{1}{2}$.

The frequency $h=m$ that we mentioned is given by

$$\begin{aligned}\mathcal{Y}(m)&=-\frac{4\pi^2}{q}\tau(m)\sum_{r\geqslant 1}\frac{\varphi(qr)}{r^2}\int_0^{+\infty}Y_0\Big(\frac{4\pi\sqrt{mx}}{qr}\Big)J_1\Big(\frac{4\pi\sqrt{mx}}{qr}\Big)W\Big(\frac{4\pi^2x}{q}\Big)\frac{dx}{\sqrt{x}}\\ &=-\frac{2\pi}{q}\frac{\tau(m)}{\sqrt{m}}\sum_{r\geqslant 1}\frac{\varphi(qr)}{r}\int_0^{+\infty}Y_0(x)J_1(x)W\Big(\frac{qr^2x^2}{4m}\Big)dx\end{aligned}$$

which is rewritten as a complex integral using the definition (6.20) of the function W

$$\mathcal{Y}(m) = -\frac{2\pi}{q}\frac{\tau(m)}{\sqrt{m}} \times \frac{1}{2i\pi}\int\limits_{(1/2)} \zeta_q(1+2s)Z(s)\Big(\frac{4m}{q}\Big)^s \Gamma(s+1)^2 G(s)H(s)\frac{ds}{s^3}$$

where

$$Z(s) = \sum_{r\geqslant 1}\varphi(qr)r^{-(1+2s)}$$
$$H(s) = \int_0^{+\infty} Y_0(x)J_1(x)x^{-2s}dx.$$

Lemma 49. *We have for all q prime*

$$Z(s) = \varphi(q)\frac{\zeta(2s)}{\zeta_q(1+2s)}$$
$$H(s) = \frac{1}{2\sqrt{\pi}}\frac{\Gamma(s)}{\Gamma(\frac{1}{2}-s)}\frac{\Gamma(1-s)^2}{\Gamma(1+s)^2}.$$

Proof. For the first:

$$\begin{aligned} Z(s) &= \sum_{n\geqslant 0}\sum_{\substack{r\geqslant 1\\(r,q)=1}} \varphi(q^{n+1}r)(q^n r)^{-(1+2s)} \\ &= \sum_{n\geqslant 0}\varphi(q^{n+1})q^{-n(1+2s)}\sum_{(r,q)=1}\varphi(r)r^{-(1+2s)} \\ &= \frac{\zeta_q(2s)}{\zeta_q(1+2s)}(q-1)(1-q^{-2s})^{-1} \\ &= \varphi(q)\frac{\zeta(2s)}{\zeta_q(1+2s)} \end{aligned}$$

while for the second we have by [G-R, 6.576.5, 6.576.6]

$$H(s) = \frac{1}{\pi}2^{-2s}\cos(\pi s)\Gamma(\tfrac{1}{2}-s)^2 F(\tfrac{1}{2}-s,\tfrac{1}{2}-s;2;1)$$

where F is the Gauss hypergeometric function, which is here an elementary function ([G-R, 9.122.1])

$$F(\tfrac{1}{2}-s,\tfrac{1}{2}-s;2;1) = \frac{\Gamma(2s)}{\Gamma(\frac{1}{2}+s)^2}$$

and the formula announced follows from this and

$$\Gamma(s)\Gamma(s+\tfrac{1}{2}) = \pi^{1/2}2^{1-2s}\Gamma(2s), \quad \Gamma(\tfrac{1}{2}-s)\Gamma(\tfrac{1}{2}+s)\cos(\pi s) = \pi.$$

□

Hence we arrive at

$$\mathcal{Y}(m) = -\frac{2\pi\varphi(q)}{q}\frac{\tau(m)}{\sqrt{m}} \times \frac{1}{2i\pi}\int\limits_{(1/2)} \frac{1}{2\sqrt{\pi}}\zeta(2s)\frac{\Gamma(2s)}{\Gamma(\frac{1}{2}-s)}\Big(\frac{4m}{q}\Big)^s G(s)\Gamma(1-s)^2\frac{ds}{s^3}.$$

The functional equation of the Riemann zeta function yields

$$\zeta(2s)\Gamma(2s) = \pi^{2s-\frac{1}{2}}\zeta(1-2s)\Gamma(\tfrac{1}{2}-s)$$

so that

$$\mathcal{Y}(m) = -\frac{\varphi(q)}{q}\frac{\tau(m)}{\sqrt{m}} \times \frac{1}{2i\pi}\int\limits_{(1/2)} \zeta(1-2s)\Big(\frac{4\pi^2 m}{q}\Big)^s G(s)\Gamma(1-s)^2\frac{ds}{s^3}.$$

Moving the integration to the line $\mathrm{Re}\,(s) = 1$ (this is cosmetic; remember that $G(1) = 0$), the integral can be estimated directly

$$\frac{1}{2i\pi}\int\limits_{(1)} \zeta(1-2s)\Big(\frac{4\pi^2 m}{q}\Big)^s G(s)\Gamma(1-s)^2\frac{ds}{s^3} \ll \frac{\sqrt{m}}{q}\tau(m)$$

which gives the bound

$$\mathcal{Y}(m) \ll \frac{\sqrt{m}}{q}\tau(m).$$

This shows that this degenerate frequency is also small enough for a mollifier of length up to $q^{\frac{1}{2}}$. However, instead of estimating, a much better treatment comes from remarking that by shifting the contour of integration to $\mathrm{Re}\,(s) = -\frac{1}{2}$, and changing s into $-s$, the resulting integral is (up to the change of ζ into ζ_q) the same exactly as the one giving $W(4\pi^2 m/q)$, with the result that this will essentially cancel the diagonal term arising from the Kronecker symbol $\delta(m,n)$ in the Petersson formula: the "true" main term is then the residue at $s = 0$ of the function being integrated. This approach is further developed and polished in [IS2].

To finish the proof of Proposition 22, we would have to estimate the contributions of the other frequencies, as well as that involving the K_0 function. But for this we refer the reader to [IS2] again, or urge him to exercise his own talents...

Conclusion

Lear: *First let me talk with this philosopher.*
What is the cause of thunder?
William Shakespeare, *"King Lear"*

What have we learned? Concerning the lower-bound, the result is quite satisfying. Even if the Birch and Swinnerton-Dyer conjecture was proved for $J_0(q)$, making a good lower bound obvious by the simple argument of the sign of the functional equation, the non-vanishing theorem for the special values of the derivative of the L-functions would retain its value, and its interpretation as progress towards Brumer's conjecture.

On the other hand, the upper bound for the rank of $J_0(q)$ still assumes that the Birch and Swinnerton-Dyer conjecture holds. Of course, this seems inevitable if analytic methods are to be applied. But how far are algebraic methods from proving the unconditional inequality

$$\text{rank } J_0(q) \ll \dim J_0(q) \tag{6.116}$$

with an absolute implied constant? Answering, or exploring, this question is important for two reasons: to gauge precisely what is the price paid by assuming the truth of the conjecture; and because if it turned out that (6.116) was within reach, then together with the analytic upper bound on the order of vanishing of the L-function of $J_0(q)$ at $s = \frac{1}{2}$, it would also bring fresh evidence to the Birch and Swinnerton-Dyer conjecture by showing how, in the case of $J_0(q)$, the rank and the order of vanishing are of the same order of magnitude.

The exploration of this matter is naturally more algebraic and, because of my limited knowledge, more naïve; I wish to thank here B. Edixhoven and J.L. Colliot-Thélène for discussing this with me.

The current algebraic methods to estimate the rank of abelian varieties start from the proof of the Mordell-Weil theorem. Let A/k be an abelian variety of dimension d defined over a number field k, and let $\bar{k}$ be an algebraic closure of k. The first step of this proof is to show that for any integer $n \geqslant 2$, the quotient $A(k)/nA(k)$ is finite (as is has to be if $A(k)$ is to be of finite rank), using either Galois or étale cohomology to reduce to the fundamental finiteness statements in algebraic number theory.

Precisely, let ℓ be any prime number. It is then shown quite easily (essentially by Kummer theory, see [Si1, X-4]) that there is a short exact sequence

$$0 \longrightarrow A(k)/\ell A(k) \longrightarrow \text{Sel}_\ell(A) \longrightarrow Ш_\ell(A) \longrightarrow 0.$$

$\text{Sel}_\ell(A)$ is the ℓ-Selmer group of A and $Ш_\ell(A)$ the ℓ-part of the Tate-Shafarevitch group of A. Those are defined by

$$\begin{aligned} \text{Sel}_\ell(A) &= \{x \in H^1(G_k, A[\ell]) \mid \text{Res}_v\, x = 0 \in H^1(G_v, A(\bar{k})), \text{ for all place } v \text{ of } k\} \\ Ш_\ell(A) &= \{x \in H^1(G_k, A(\bar{k}))[\ell] \mid \text{Res}_v\, x = 0 \in H^1(G_v, A(\bar{k})), \text{ for all place } v \text{ of } k\} \end{aligned} \tag{6.117}$$

where we have denoted by G_k the absolute Galois group of k and by G_v the decomposition group at v for a place v of k, which is isomorphic to the Galois group of the local field k_v, and the H^1 refer to Galois cohomology groups.

Writing the Mordell-Weil group of A as a direct sum of the finite torsion subgroup and the free part

$$A(k) \simeq A(k)_{tors} \oplus \mathbf{Z}^r$$

it follows that for almost all prime numbers ℓ we have

$$A(k)/\ell A(k) \simeq (\mathbf{Z}/\ell\mathbf{Z})^r$$

hence

$$\text{rank } A = \dim_{\mathbf{F}_\ell} A(k)/\ell A(k) \leqslant \dim \text{Sel}_\ell(A). \tag{6.118}$$

(and this is even an equality, for almost all ℓ, if the conjecture according to which the full Tate-Shafarevitch group $Ш(A)$ is finite is true).

The usual way of showing that the Selmer group is finite is by observing that one can make any extension field (this can only increase the rank, and the Selmer group). In particular one can replace k by a field K containing the coordinates of all the ℓ-torsion points of A, in which case $A[\ell]$ is a trivial G_K-module. Then the Galois cohomology group is simply

$$H^1(G_K, A[\ell]) = \text{Hom}(G_K, A[\ell]) = \text{Hom}(G_K, (\mathbf{Z}/\ell\mathbf{Z})^{2d})$$

since, by the analytic uniformization of A as a complex torus, the group structure of $A[\ell]$ is known.

Now for each homomorphism $\rho : G_K \longrightarrow (\mathbf{Z}/\ell\mathbf{Z})^{2d}$, the fixed field of the kernel of ρ is an abelian Galois extension of K with Galois group isomorphic to the image of ρ, a subgroup of $(\mathbf{Z}/\ell\mathbf{Z})^{2d}$, and conversely to each such extension field there can correspond only finitely many morphisms ρ. Moreover, and here arithmetic enters the fray, one shows that there exists a fixed set Σ of places of K such that all such extensions are unramified outside Σ (this is because A itself is unramified at all but finitely many places). But there are only finitely many extensions of K which are unramified outside Σ and of degree bounded by ℓ^{2d}. Therefore, we find that the Selmer group is finite, and recover the weak Mordell-Weil theorem.

Consider again the case $A = J_0(q)$ over $\mathbf{Q}$, and let's see what this gives. The problem with the above scheme of proof is the moment when the extension to the field K is performed. It means that in (6.118), it is really the rank of A over K which is estimated. But for a high-dimensional variety, as ours are, the field K will usually be of very large degree over $\mathbf{Q}$, because the lowest-degree possible K must be the fixed field of the kernel of the representation ρ_ℓ of the Galois group of $\mathbf{Q}$ on ℓ-torsion points

$$\rho_\ell : G_{\mathbf{Q}} \longrightarrow GL(2d, \mathbf{F}_\ell),$$

which is expected to have a large image. Indeed, it is proved in [Se2] that, in the case of elliptic curves without complex multiplication, the corresponding representation is surjective except for finitely many ℓ. Then the rank of $J_0(q)$ over K will most likely be of larger order of magnitude than $\dim J_0(q)$, so the bound (6.118) cannot get near (6.116).

Heuristically, this can be confirmed as follows. Say the degree of the field K is $d = d(q)$. One can guess that the exceptional set Σ should consist of primes above

ℓ and primes above q (since $J_0(q)$ has good reduction outside q), but in any case the abelian extensions of K unramified outside Σ include at least those which are actually unramified everywhere. Those are classified by the ideal class group of K (by class field theory); we need the ℓ-rank of the Selmer group, and thus the first guess, which cannot be obviously improved, is at least the ℓ-rank of the class group. Again in a worst case scenario that is hard to reject out of hand, this could be as large the logarithm of the discriminant of K. But then the (most optimistic this time!) lower bound on the discriminant of a number field of degree d over $\mathbf{Q}$ is already exponential in d as d tends to infinity, and we obtain something (which is neither lower nor upper bound, but the optimistic worse case, as it were) of the order of the degree d of K. If ρ_ℓ happens to be surjective, this is $|GL(2d, \mathbf{F}_\ell)| \approx \ell^{(2d)^2} \dots$

Whatever the value of this kind of argument, to get an unconditional upper bound, it seems necessary to find a way of estimating the dimension over $\mathbf{F}_\ell$ of the Selmer group $\mathrm{Sel}_\ell(J_0(q))$ without making any extension of the ground field $\mathbf{Q}$. Moreover, the particular geometric properties of $J_0(q)$ should be exploited, short of proving the Birch and Swinnerton-Dyer conjecture in general. And this has been the case as far as the analytic upper bound proved in Chapter 5 is concerned: without the very special link between $J_0(q)$ and weight 2 forms, the results would be much poorer, or nonexistent, as they are today for modular elliptic curves for instance. This offers some hope, because the geometry of the modular curves and their Jacobians has been extensively studied, with considerable success, by Mazur, Wiles, Ribet and many others. Moreover, in the work of Wiles leading to Fermat's theorem, the Selmer groups of the symmetric square of the Galois representations associated to modular forms is of paramount importance (see [Wil], [DDT]), and its cardinality is computed, at least in many cases. From the analytic point of view, this is related ([DDT, page 96]) to the special value at the edge of the critical strip of the symmetric square L-function, which we have also encountered in Chapter 3. Thus it is again clear that the Selmer group occurring in the study of the rank is a more difficult invariant.

There is a possible starting point, exploiting the fact that $J_0(q)$ is a Jacobian, and not simply any abelian variety, in the cohomological nature of the Jacobian, which gives an isomorphism ([Mil, page 126])

$$J_0(q)[\ell] \simeq H^1_{\text{ét}}(X_0(q)_{\bar{\mathbf{Q}}}, \mu_\ell)$$

of $G_{\mathbf{Q}}$-modules. Then, using the Hochschild-Serre spectral sequence

$$H^p(G_{\mathbf{Q}}, H^q_{\text{ét}}(X_0(q)_{\bar{\mathbf{Q}}}, \mu_\ell)) \Rightarrow H^{p+q}_{\text{ét}}(X_0(q), \mu_\ell)$$

of the covering $X_0(q)_{\bar{\mathbf{Q}}} \to X_0(q)$ (the latter designates here the natural model over $\mathbf{Q}$), and more precisely the exact sequence of small order terms ([Mil, page 308]), it transpires that the group $H^1(G_{\mathbf{Q}}, J_0(q)[\ell])$ which contains the Selmer group is essentially (up to other groups which do not depend on q) isomorphic to $H^2_{\text{ét}}(X_0(q), \mu_\ell)$. Then the local conditions defining the Selmer group (6.117) must be reinterpreted in this setting (or, according to Colliot-Thélène, it should be possible to start working directly with an integral model of $X_0(q)$, and the same kind of arguments would yield directly a finite group, possibly larger than the Selmer group however, as the behavior at the ramified primes would be under less control). But of course the difficulty starts after that: what to do next with this? If the problem is reduced in this way to one about the curve $X_0(q)$, it remains unclear whether it is simpler or more amenable to

further treatment. Indeed, over $\mathbf{Q}$ we must see $X_0(q)$ more as an arithmetic surface, of dimension 2 (the other dimension being the primes, the spectrum of $\mathbf{Z}$), and the second cohomology group of a projective surface is very much the most interesting and most subtle in general (see [Mil, chapter 5] for instance).

We can now see quite clearly the depth of the upper bound (6.116) that has been proved on the Birch and Swinnerton-Dyer conjecture. Yet, to the optimistic eye, there seems to be a context to work in here, in which progress might not be impossible to contemplate.

But this is another story, although one which it would be nice to write, or read, some day.

References

[B-H] Bushnell, C. J. and Henniart, G.: *An upper bound on conductors for pairs*, J. Number Theory 65 no. 2 (1997), 183–196.

[B-L] Bohr, H. and Landau, E.: *Sur les zéros de la fonction $\zeta(s)$ de Riemann*, Compte Rendus de l'Acad. des Sciences (Paris) 158 (1914), 106–110.

[Br1] Brumer, A.: *The rank of $J_0(N)$*, Astérisque 228, SMF (1995), 41–68.

[Bum] Bump, D.: *Automorphic forms and representations*, Cambridge Studies in Advanced Mathematics, 55, Cambridge University Press, 1997.

[CoS] Coates, J. and Schmidt, C-G.: *Iwasawa theory for the symmetric square of an elliptic curve*, J. Reine Angew. Math. 375/376 (1987), 104–156.

[C-S] Cornell, G. and Silverman, J. (editors): *Arithmetic Geometry*, Springer Verlag (1986).

[DDT] Darmon, H., Diamond, F. and Taylor, M.: *Fermat's Last Theorem*, Current Developments in Math., International Press (1995), 1–154.

[Del] Deligne, P.: *Formes modulaires et représentations de $GL(2)$*, Modular Forms in One Variable IV, Springer Lecture Notes 749 (1972), 55–105.

[DFI] Duke, W., Friedlander, J. and Iwaniec, H.: *Bounds for automorphic L-functions, II*, Invent. Math. 115 (1994), 219–239.

[Du1] Duke, W.: *The critical order of vanishing of automorphic L-functions with high level*, Invent. Math. 119 (1995), 165–174.

[Du2] Duke, W.: *The dimension of the space of cusp forms of weight one*, Internat. Math. Res. Notices 2 (1995), 99–109.

[D-K] Duke, W. and Kowalski, E.: *A problem of Linnik for elliptic curves and mean-value estimates for automorphic representations*, to appear in Invent. Math.

[Edx] Edixhoven, B.: *The modular curves $X_0(N)$*, notes for the ICTP Summer School, August 1997.

[Gel] Gelbart, S.: *Automorphic forms on adèle groups*, Annals of Math. Studies 83, Princeton Univ. Press (1975).

[GHL] Goldfeld, D., Hoffstein, J. and Lieman, D.: *An effective zero-free region*, Ann. of Math. 140 (1994), 177–181.

[G-J] Gelbart, S. and Jacquet, H.: *A relation between automorphic representations of $GL(2)$ and $GL(3)$*, Ann. Sci. E.N.S 4ème série 11 (1978), 471–552.

[Gol] Goldfeld, D.: *The class number of quadratic fields and the conjectures of Birch and Swinnerton-Dyer*, Ann. Scuola Norm. Sup. Pisa 3, 4 (1976), 623–663.

[Gro] Gross, B.: *Heegner points on $X_0(N)$*, in Modular Forms, (R.A. Rankin, editor), Ellis Horwood (1984), 87–106.

[G-R] Gradshteyn, I. S.; Ryzhik, I. M. *Table of integrals, series, and products*, Fifth edition, Academic Press, 1994.

[G-Z] Gross, B. and Zagier, D.: *Heegner points and derivatives of L-series*, Invent. Math., 84 (1986), 225–320.

[Har] Hartshorne, R.: *Algebraic geometry*, Grad. Texts in Math. 52, Springer Verlag (1977).

[H-L] Hoffstein, J. and Lockhart, P.: *Coefficients of Maass forms and the Siegel zero* (with an appendix by D. Goldfeld, J. Hoffstein and D. Lieman), Ann. of Math. (2) 140 (1994), 161–181.

[I-R] Ireland, K. and Rosen, M.: *A classical introduction to modern number theory*, Second edition, Grad. Texts in Math. 84, Springer Verlag (1990).

[IS1] Iwaniec, H. and Sarnak, P.: *The non-vanishing of central values of automorphic L-functions and Siegel's zeros*, preprint (1997).

[IS2] Iwaniec, H. and Sarnak, P.: *The non-vanishing of central values of automorphic L-functions and Siegel's zeros*, in preparation.

[Iw1] Iwaniec, H.: *On the order of vanishing of modular L-functions at the critical point*, Sém. Théor. Nombres Bordeaux 2 (1990), 365–376.

[Iw2] Iwaniec, H.: *Topics in classical modular forms*, Grad. Studies in Math. 17, A.M.S (1997).

[Iw3] Iwaniec, H.: *Introduction to the Spectral Theory of Automorphic Forms*, Biblioteca de la Revista Matemática Iberoamericana (1995).

[JPS] Jacquet, H., Piatetskii-Shapiro, I. I. and Shalika, J. A.: *Rankin-Selberg convolutions*, Amer. Jour. of Math. 105 (1983), 367–464.

[Jut] Jutila, M.: *Lectures on a method in the theory of exponential sums*, Tata Lectures on Mathematics and Physics 80, Springer-Verlag, 1987.

[K-L] Kolyvagin, V. and Logachev, D.: *Finiteness of the Shafarevich-Tate group and the group of rational points for some modular abelian varieties*, Leningrad Math. J., 1, No. 5 (1990), 1229–1253.

[KM1] Kowalski, E. and Michel, P.: *Sur le rang de $J_0(q)$*, Préprint de l'Université d'Orsay, 53 (1997).

[KM2] Kowalski, E. and Michel, P.: *Sur les zéros des fonctions L automorphes de grand niveau*, Préprint de l'Université d'Orsay, 54 (1997).

[KM3] Kowalski, E. and Michel, P.: *A lower bound for the rank of $J_0(q)$*, Préprint de l'Université d'Orsay, 69 (1997).

[Kob] Koblitz, N.: *Introduction to elliptic curves and modular forms*, Second edition, Grad. Texts in Math. 97, Springer Verlag (1993).

[K-S] Katz, N. and Sarnak, P.: *Random matrices, Frobenius eigenvalues, and monodromy*, to appear.

[LIS] Luo, W., Iwaniec, H. and Sarnak, P.: *Low lying zeros for families of L-functions*, preprint (1998).

[LRS] Luo, W., Rudnick, Z. and Sarnak, P.: *On Selberg's eigenvalue conjecture*, Geom. Funct. Anal. 5 (1995), 387–401.

[Luo] Luo, W.:*Zeros of Hecke L-functions associated with cusp forms*, Acta Arith. 71, No.2 (1995), 139–158.

[Maz] Mazur, B.: *On the passage from local to global in number theory*, Bull. Amer. Math. Soc. (N.S.) 29 (1993), no. 1, 14–50.

[Me] Merel, L.: *Bornes pour la torsion des courbes elliptiques sur les corps de nombres*, Invent. Math 124 (1996), 437–450.

[Mes] Mestre, J.-F.: *Formules explicites et minorations de conducteurs de variétés algébriques*, Comp. Math. 58 (1986), 209–232.

[Mil] Milne, J.S.: *Étale cohomology*, Princeton Mathematical Series 33, Princeton Univ. Press, 1980.

[Miy] Miyake, T.: *Modular Forms*, Springer Verlag, 1989.

[M-W] Moeglin, C. and Waldspurger, J.L.: *Pôles des fonctions L de paires pour $GL(N)$*, appendix to *Le spectre résiduel de $GL(n)$*, Ann. Sci. ENS (4ème série) 22 (1989), 605–674.

[Mu1] Mumford, D.: *Abelian varieties*, Oxford University Press, 1970.

[Mu2] Mumford, D.: *Tata lectures on Theta, II*, Progress in Math. 43, Birkhäuser, 1984.

[Poi] Poitou, G.: *Sur les petits discriminants*, Séminaire Delange-Pisot-Poitou, 18e année (1976/77), Théorie des nombres, Fasc. 1, Exp. No. 6, 18 pp., Secrétariat Math., Paris, 1977.

[P-P] Perelli, A. and Pomykala, J.: *Averages over twisted elliptic L-functions*, Acta Arith. 80, No 2 (1997), 149–163.

[R-S] Rudnick, Z. and Sarnak, P.: *Zeros of principal L-functions and random matrix theory*, A celebration of John F. Nash, Jr., Duke Math. J. 81, no. 2 (1996), 269–322.

[Sar] Sarnak, P.: *Quantum Chaos, Symmetry and Zeta Functions*, Current Developments in Math., International Press 1997.

[Sel] Selberg, A.: *Contributions to the theory of Dirichlet's L-functions*, Skr. Norske Vid. Akad. Oslo. I. (1946), 1–62, or Collected Papers, vol. 1, Springer Verlag, Berlin, (1989), 281–340.

[Se1] Serre, J.-P: *Cours d'Arithmétique*, 3ème édition, P.U.F (1988).

[Se2] Serre, J.-P.: *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Invent. Math. 15 (1972), 259–331.

[Sh1] Shimura, G.: *Introduction to the arithmetic theory of automorphic functions*, Iwanami Shoten and Princeton Univ. Press (1971).

[Sh2] Shimura, G.: *On the holomorphy of certain Dirichlet series*, Proc. of the London Math. Soc (3) 31 (1975), 79–95.

[Sh3] Shimura, G.: *The special values of zeta functions associated with cusp forms*, Comm. Pure and Appl. Math. 29 (1976), 783–804.

[Si1] Silverman, J.: *The arithmetic of elliptic curves*, Grad. Texts in Math. 106, Springer Verlag (1986).

[Si2] Silverman, J.: *Advanced topics in the arithmetic of elliptic curves*, Grad. Texts in Math. 151, Springer Verlag (1994).

[Tit] Titchmarsh, E.C.: *The theory of the Riemann Zeta-function*, Second edition (revised by D. R. Heath-Brown), Oxford University Press, 1986.

[Vdk] Vanderkam, J.: *The rank of quotients of $J_0(N)$*, preprint 1997.

[Wil] Wiles, A.: *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) 141 (1995), 443–551.