BEYOND ENDOSCOPY[1]

R. P. Langlands[2]

Ya tutarsa – Nasreddin Hoca

Introduction[3]

0.1 Functoriality and related matters. The notion of L-group and the principle of functoriality appeared in [L] and were explained at more length in [Cor] and elsewhere. The principle of functoriality, which is now widely believed but is very far from being established in general, can be roughly stated as follows.

(I) *If H and G are two reductive groups over the global field F and the group G is quasi-split then to each homomorphism*

$$\phi: \quad {}^L H \longrightarrow {}^L G$$

there is associated a transfer *of automorphic representations of H to automorphic representations of G.*

A second problem that arose some time after functoriality is that of associating to an automorphic representation π, now on the group G, an algebraic subgroup ${}^\lambda H_\pi$ of ${}^L G$ that would at best be defined up to conjugacy, although even that might often fail, and would have the following property.[4]

(II) *If ρ is a representation of ${}^L G$ then the multiplicity $m_H(\rho)$ of the trivial representation of ${}^\lambda H_\pi$ in the restriction of ρ to ${}^\lambda H_\pi$ is the order $m_\pi(\rho)$ of the pole of $L(s,\pi,\rho)$ at $s=1$.* Once again, this is not intended as an absolutely precise statement.

---

[1] Dedicated to Joseph Shalika on the occasion of his sixtieth birthday.

[2] I would like to thank James Arthur, who once again guided me through the subtleties of weighted orbital integrals, Erez Lapid and Peter Sarnak for useful conversations related to the material of this paper and Werner Hoffmann for his comments on [H] and on Appendix C and D.

[3] There is available at `http://SunSITE.UBC.CA/DigitalMathArchive/Langlands` the text of a lecture *Endoscopy and beyond* that can also serve as an introduction to this paper. It has the advantage of being informal, but there are misprints and some suggestions towards the end are red herrings. The present paper may well turn out to have the same defects!

[4] I use the notation ${}^\lambda H$ to stress that we are dealing with a subgroup of the L-group ${}^L G$ that may not itself be an L-group, but is close to one. Although there is not yet a group H attached to ${}^\lambda H$, I use, for simplicity, in the next statement and subsequently, the notation $m_H(\rho)$ or $m_{H_\pi}(\rho)$ rather than $m_{{}^\lambda H}(\rho)$ or $m_{{}^\lambda H_\pi}(\rho)$

0.2 Some touchstones. There are three. The first two form a part of functoriality. The third does not. It is a question raised by a theorem of Deligne-Serre ([DL]). I take for expository purposes the ground field F to be an arbitrary number field (of finite degree).

(T1) *Take H to be $GL(2)$, G to be $GL(m+1)$ and ϕ to be the m^{th} symmetric power representation.*

(T2) *Take H to be the group consisting of a single element and G to be $GL(2)$. Then ${}^LH$ is a Galois group and problem* (I) *is that of associating an automorphic form to a two-dimensional Galois representation.*

(T3) *Take G to be $GL(2)$ and π to be an automorphic representation such that at every infinite place v of the π_v is associated to a two-dimensional representation not merely of the Weil group but of the Galois group over F_v. Show that H_π is finite.*

A positive solution of the first problem has as consequence the Ramanujan-Petersson conjecture and the Selberg conjecture in their strongest forms; the Artin conjecture follows from the second. As is well-known, all these problems have been partially solved; some striking results for the first problem are very recent. For various reasons, the partial solutions all leave from a methodological point of view a great deal to be desired. Although none of these problems refer to the existence of ${}^\lambda H_\pi$, I am now inclined to the view that the key to the solution of the first two and of functoriality in general lies in the problem (II), whose ultimate formulation will include functoriality. Moreover, as I shall observe at the very end of the paper, the problem (T3) can be approached in the same spirit.

I by no means want to suggest that I believe the solution to (II) is imminent. What I want to suggest rather, and to establish on the basis of the concrete calculations in this paper, is that reflecting on the problem of attacking (II) with the help of the trace formula, in my opinion the only analytic tool of any substantial promise available for either (I) or (II), one is led to concrete problems in analytic number theory. They are difficult; but an often successful strategy, even though slow and usually inglorious, for breaching an otherwise unassailable mathematical problem is to reduce some aspect of it to a concrete, accessible form on which at least small inroads can be made and some experience acquired. The calculations, tentative as they are, described in the third part of this paper are intended as a first step in this direction for problems (I) and (II). I concentrate on (T2), for which G is $GL(2)$ and on π for which ${}^\lambda H_\pi$ is finite. The same approach applied to (T1) would entail dealing with $GL(m+1)$ and π for which ${}^\lambda H$ was the image of $GL(2)$ under the mth symmetric power. This would require the use of the trace formula for $GL(m+1)$, much more sophisticated than that for $GL(2)$ although perhaps not completely inaccessible to numerical investigation for very small m.

Part I: Formal structure

1.1 The group ${}^\lambda H_\pi$. We might take (II) as a definition of ${}^\lambda H_\pi$, but there are several difficulties. It is, first of all, perhaps best not to try to define ${}^\lambda H_\pi$ for all π. Arthur in his study of the trace formula has been led to a classification of automorphic representations that in spite of its apparent reliance on objects whose existence is not established can, in fact, in the context of the trace formula usually be formulated in decidable terms. The classification is above all a separation into representations that are of *Ramanujan type* and those that are not. It is of conceptual significance that one expects to prove ultimately that the representations of Ramanujan type are exactly those that satisfy the general form of the Ramanujan conjecture, but that is not essential to the classification. The point is that a given trace formula will give a sum over both types of automorphic representation but the contribution to the formula of the representations that are not of Ramanujan type will be expressible in terms of traces from groups of lower dimension, so that the remainder can be regarded as the sum over the representations of Ramanujan type. We shall see a simple application of this principle to $GL(2)$. If π is not of Ramanujan type, it will be natural to define ${}^\lambda H_\pi$ as the product ${}^\lambda H_{\pi'} \times S$ of a group ${}^\lambda H_{\pi'}$ defined by an ancillary π' of Ramanujan type with an image S of $SL(2,\mathbb{C})$, but this is a matter for which any great concern would be premature.

The other difficulties are more severe. The first is that even though we may expect that when π is of Ramanujan type the functions $L(s,\pi,\rho)$ are analytic on $\mathrm{Re}(s) \geq 1$ except perhaps for a finite number of poles on $\mathrm{Re}(s)=1$ we are in no position to prove it. So an alternative definition of $m_\pi(\rho)$ is called for, even though, as must be stressed, the definition need at first only be used operationally – as a guide to the construction of various algebraic and analytic expressions whose meaning will be clear and unambiguous.

There are two more difficulties: given π (implicitly of Ramanujan type) why should there exist an ${}^\lambda H$ (implicitly a reductive, but often not a connected, group) such that

$$m_H(\rho) = m_\pi(\rho)$$

for all ρ; even if there is such an ${}^\lambda H$, why should it be unique, or rather why should its conjugacy class under $\hat{G}$ be unique? Recall that the L-group is the semidirect product of its connected component $\hat{G}$

with the Galois group $\mathrm{Gal}(K/F)$ of a finite Galois extension of F that has to be allowed to be arbitrarily large, so that the L-group is really an inverse sequence of groups with a common connected component. It normally suffices, however, to fix a K large enough for the purposes at hand.

The second of these difficulties is easily resolved. The conjugacy class may not be unique and there may be several groups to be denoted ${}^\lambda H_\pi$. This is related to the multiplicity problem for automorphic representations. It will, however, be important to establish that if the function $\rho \to m_H(\rho)$ is given then there are only finitely many possibilities for the conjugacy class of ${}^\lambda H$. Jean-Pierre Wintenberger has pointed out to me that as a consequence of a theorem of Larsen-Pink ([LP]) the group ${}^\lambda H$ is uniquely determined by the numbers $m_H(\rho)$ if ${}^L G$ is $GL(n, \mathbb{C})$, thus if G is $GL(n)$ over F and the Galois extension of F used to define the L-group is F itself.[5]

In so far as the condition that the function m_π be an m_H is a linear condition – thus in so far as (in some sense!) $m_\pi(\rho) = \operatorname{tr}\pi(f^\rho)$, where f^ρ is some kind of generalized function on $G(\mathbb{A}_F)$ – the existence of ${}^\lambda H_\pi$ is something to be verified by the trace formula. In the simplest of cases, there would be a linear form

$$\sum \alpha_\rho m_\pi(\rho), \qquad \alpha_\rho = \alpha_\rho^H, \tag{1}$$

which is 0 if ${}^\lambda H_\pi$ is not conjugate to a given ${}^\lambda H$ but is 1 if it is. The trace formula will, with any luck, yield an expression for the sum over all π with appropriate multiplicities of (1) and will thus select exactly those π attached to ${}^\lambda H$, but a similar sum that selected exactly, perhaps with multiplicity, those π such that ${}^\lambda H_\pi$ lies in a given ${}^\lambda H$ would be better. Thus $\sum \alpha_\rho m_\pi(\rho)$ is to be 0 if none of the possible ${}^\lambda H_\pi$ is conjugate to a subgroup of ${}^\lambda H$ but is otherwise to be $\beta_\pi^H \neq 0$, where β_π^H depends only on the collection of possible ${}^\lambda H_\pi$ and is to be 1 if ${}^\lambda H_\pi = {}^\lambda H$.

If we admit the possibility that there is a second group ${}^\lambda H'$ such that $m_{H'}(\rho) = m_H(\rho)$ for all ρ, then we see that we are demanding too much from the form (1). We might rather introduce a partial ordering on the collection of ${}^\lambda H$, writing

$$\lambda_{H'} \prec_{LP} \lambda_H$$

if $m_{H'}(\rho) \geq m_H(\rho)$ for all ρ. Then we could at best hope that (1) would be different from 0 only if ${}^\lambda H_\pi \prec_{LP} {}^\lambda H$ and that it would be 1 if ${}^\lambda H_\pi \sim_{LP} {}^\lambda H$, thus if $m_{H_\pi}(\rho) = m_H(\rho)$ for all ρ. We would then, for each ${}^\lambda H_\pi$, try to obtain from the trace formula an expression for

$$\sum_{{}^\lambda H_\pi \prec_{LP} {}^\lambda H} \sum_\rho \alpha_\rho^H m_\pi(\rho). \tag{2}$$

[5] There are certain supplementary conditions to be taken into account even in this case.

It is best, however, to admit frankly that the first of the two difficulties, which amounts to understanding the conditions on the linear form $\rho \to m(\rho)$ that guarantee it is given by a subgroup ${}^\lambda H$ and to showing that m_π satisfies these conditions, is a very serious problem that is not broached here. I content myself with a basic example or two that suggest it is prudent to keep an open mind about the properties to be possessed by (1) and about the final structure of the arguments. So (1) and (2) are at best provisional approximations to what is to be investigated.

1.2 A simple observation. Not only is the L-group an inverse sequence but so is, implicitly, each ${}^\lambda H$. If the occasion arises to distinguish the group in the sequence that lies in ${}^L G^K = \hat{G} \rtimes \mathrm{Gal}(K/F)$, we denote it ${}^\lambda H^K$. If $K \subset K'$, there is a surjective map

$$ {}^\lambda H^{K'} \to {}^\lambda H^K. $$

Among the representations ρ are those that factor through the projection of ${}^L G$ on the Galois group, $\mathrm{Gal}(K/F)$. Since $L(s,\pi,\rho)$ is, for such a representation, equal to the Artin L-function $L(s,\rho)$, the number $m_\pi(\rho) = m_{H_\pi}(\rho)$ is just the multiplicity with which the trivial representation occurs in ρ. If $\mathfrak{H}$ is the image of ${}^\lambda H_\pi$ in $\mathfrak{G} = \mathrm{Gal}(K/F)$, it is also $m_{\mathfrak{H}}(\rho)$, calculated with respect to $\mathfrak{G}$. This is clearly possible for all ρ only if $\mathfrak{H} = \mathfrak{G}$. Thus if ${}^\lambda H_\pi$ exists it will have to be such that its projection on $\mathrm{Gal}(K/F)$ is the full group. We shall implicitly assume throughout the paper that any group ${}^\lambda H$ appearing has this property.

1.3 Calculation of $m_H(\rho)$ in some simple cases. In the second part of the paper, I shall consider only the group $G = GL(2)$ and it only over the base field $\mathbb{Q}$. I have not reflected on any other cases. I shall also often consider only π whose central character is trivial, so that π is an automorphic representation of $PGL(2)$. Then $m_\pi(\rho)$ will not change when ρ is multiplied by any one-dimensional representation of $GL(2,\mathbb{C})$ and ${}^\lambda H_\pi$ will lie in $SL(2,\mathbb{C})$ or, to be more precise, in the family $\{SL(2,\mathbb{C}) \times \mathrm{Gal}(K/\mathbb{Q})\}$. It is instructive to compute $m_H(\rho)$ for a few ${}^\lambda H^K$ in $SL(2,\mathbb{C}) \times \mathrm{Gal}(K/\mathbb{Q})$ and a few ρ. We may as well confine ourselves to the standard symmetric powers σ_m, $m = 1, 2, \ldots$ of dimension $m+1$ and to their tensor products with irreducible Galois representations τ.

If ${}^\lambda H \subset SL(2,\mathbb{C}) \times \mathrm{Gal}(K/\mathbb{Q})$, the multiplicity $m_H(\sigma_1)$ is 2 if the projection of ${}^\lambda H$ on the first factor is $\{1\}$ and is 0 otherwise. Thus if we confine ourselves to groups ${}^\lambda H$ that project onto $\mathrm{Gal}(K/\mathbb{Q})$, then

$$ (A) \qquad a_1 m_H(\rho_1), \quad a_1 = \frac{1}{2}, \quad \rho_1 = \sigma_1, $$

is 1 if ${}^\lambda H = \{1\} \times \mathrm{Gal}(K/\mathbb{Q})$ and 0 otherwise. On the other hand,

$$(B) \qquad a_1 m_{H'}(\rho_1), \quad a_1 = 1, \quad \rho_1 = \det,$$

is 1 for all subgroups ${}^\lambda H'$ of ${}^\lambda H = SL(2,\mathbb{C}) \times \mathrm{Gal}(K/\mathbb{Q})$ but 0 for groups that are not contained in ${}^\lambda H$. When and if the occasion arises for a precise reference, we denote the groups in these two cases by ${}^\lambda H_A$ and ${}^\lambda H_B$.

In general, as in (1) and (2), given ${}^\lambda H$, we would like to find a collection $\rho_1, \dots, \rho_n$ of representations and a collection $a_1, \dots, a_n$ of real numbers such that

$$\sum_k a_k m_{H'}(\rho_k) = 1$$

if ${}^\lambda H' \subset {}^\lambda H$ and 0 if it is not. We will normally want to consider only ${}^\lambda H$ and ${}^\lambda H'$ defined with respect to a given K. To make clear to which group given collections are associated I sometimes write as before $\rho_k = \rho_k^H$, $a_k = a_k^H$.

If the kernel of the projection of ${}^\lambda H$ to $\mathrm{Gal}(K/\mathbb{Q})$ is infinite, it is either $SL(2,\mathbb{C})$, a trivial case already treated, or contains the group

$$\hat{H} = \{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \mid a \in \mathbb{C}^\times \}$$

as a normal subgroup of index 1 or 2. The group of outer automorphisms of $\hat{H}$, through which the action of ${}^\lambda H$ on $\hat{H}$ factors, is of order two and the image of ${}^\lambda H$ in it may or may not be trivial. If it is trivial, then ${}^\lambda H = \hat{H} \times \mathrm{Gal}(K/\mathbb{Q})$ and $m_H(\sigma_m \otimes \tau)$ is 1 if m is even and τ is trivial and otherwise 0. We take

$$(C) \qquad a_1 = 1 \quad \rho_1 = \sigma_2,$$

and denote the pertinent group by ${}^\lambda H_C$.

If the image of ${}^\lambda H$ in the group of outer automorphisms, identified with $\mathbb{Z}_2$, is not trivial the map ${}^\lambda H \to \mathbb{Z}_2$ may or may not factor through the Galois group. If it does not then $\hat{H} \backslash {}^\lambda H$ is isomorphic to $\mathbb{Z}_2 \times \mathrm{Gal}(K/\mathbb{Q})$ and ${}^\lambda H$ contains the normalizer of $\hat{H}$ in $SL(2,\mathbb{C})$. Moreover $m_H(\sigma_m \otimes \tau) = 0$ unless $m \equiv 0 \pmod 4$ and τ is trivial, when it is 1. If the map ${}^\lambda H \to \mathbb{Z}_2$ factors through the Galois group then $\hat{H} \backslash H$ is isomorphic to $\mathrm{Gal}(K/\mathbb{Q})$ and $m_H(\sigma_m \otimes \tau)$ is 1 if and only if $m \equiv 0 \pmod 4$ and τ is trivial or $m \equiv 2 \pmod 4$ is even and τ is the one-dimensional representation τ_0 of $\mathrm{Gal}(K/\mathbb{Q})$

obtained by projecting onto the group $\mathbb{Z}_2$ and then taking the nontrivial character of this group, which is of order two. Otherwise $m_H(\sigma_m \otimes \tau)$ is 0. We take in these two cases:

$$(D) \qquad a_1 = 1, \quad \rho_1 = \sigma_4;$$

$$(E) \qquad a_1 = 1, \quad \rho_1 = \sigma_2 \otimes \tau_0.$$

The two groups will of course be denoted by ${}^\lambda H_D$ and ${}^\lambda H_E$.

If ${}^\lambda H'$ and ${}^\lambda H$ are each one of the five groups just described, then

$$\sum_k a_k^H m_{H'}(\rho_k^H)$$

is different from 0 only if ${}^\lambda H'$ is conjugate to a subgroup of ${}^\lambda H$ and is 1 if ${}^\lambda H' = {}^\lambda H$. Observe as well that in each of these cases, $m_H(\sigma_m \otimes \tau)$, depends only on τ and on m modulo 4.

The only remaining possibility is that ${}^\lambda H$ projects to a finite nontrivial subgroup in $SL(2,\mathbb{C})$. The projection is either abelian, dihedral, tetrahedral, octahedral or icosahedral. For the last three cases, the numbers $m_H(\sigma_m)$ are calculated for $m = 1, \dots, 30$ to be the following.

$$\begin{array}{ll} \text{Tetrahedral}: & 0,0,0,0,0,1,0,1,0,0,0,2,0,1,0,1,0,2,0,2,0,1,0,3,0,2,0,2,0,3; \\ \text{Octahedral}: & 0,0,0,0,0,0,0,1,0,0,0,1,0,0,0,1,0,1,0,1,0,0,0,2,0,1,0,1,0,1; \\ \text{Icosahedral}: & 0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,1,0,0,0,1,0,0,0,0,0,1. \end{array}$$

As a consequence, if we take K to be $\mathbb{Q}$ and let ${}^\lambda H_T$, ${}^\lambda H_O$ and ${}^\lambda H_I$ be the three subgroups of $SL(2,\mathbb{C})$ corresponding to the regular solids and if we set

$$\begin{array}{llll} a_1^T = 1, & a_2^T = -1, & \rho_1^T = \sigma_6, & \rho_2^T = \sigma_2, \\ a_1^O = 1, & a_2^O = -1, & \rho_1^O = \sigma_8, & \rho_2^O = \sigma_4, \\ a_1^I = 1, & a_2^I = -1, & \rho_1^I = \sigma_{12}, & \rho_2^I = \sigma_8, \end{array}$$

then, for ${}^\lambda H'$ infinite or equal to one of the same three groups,

$$\sum_k a_k^H m_{H'}(\rho_k^H)$$

is 0 if ${}^\lambda H'$ is not conjugate to a subgroup of ${}^\lambda H$ and is 1 if ${}^\lambda H' = {}^\lambda H$.

On the other hand, if the projection on $SL(2,\mathbb{C})$ is abelian of order l, then $m_H(\sigma_m)$ is the number N of integers in $\{m, m-2, \dots, -m\}$ divisible by l, and if it is dihedral with center of order $l \geq 3$ then $m_H(\sigma)$ is $N/2$ if m is odd and $(N+1)/2$ if $m \equiv 0 \pmod 4$ and $(N-1)/2$ if $m \equiv 2 \pmod 4$.

Suppose, for example, that it is dihedral with center of order 6. Then $N = 3$ for $m = 6$ and $N = 1$ for $m = 2$. Thus

$$a_1^T m_H(\rho_1^T) - a_2^T m_H(\rho_2^T) = 1 \neq 0,$$

but the group H is not contained in the tetrahedral group. If we try exclude the group H by adding other representations to the sequence ρ_1^T, ρ_2^T, for example σ_{10}, then we will introduce other groups, like the abelian group of order 10, that should be, but will not be, subgroups of the tetrahedral group. So we are still hoping for too much from the form (1). It looks as though we will have to accept in (2) groups that are not subgroups of the tetrahedral group, but that are finite dihedral groups or abelian. Since ${}^\lambda H_\pi$ is abelian only if π is associated to Eisenstein series, we can envisage treating them by first treating the infinite dihedral groups along the lines of (1) and (2), and then treating dihedral ${}^\lambda H$ as subgroups of the L-group of the group defined by the elements of norm 1 in a quadratic extension. This is clumsier than one might hope.[6]

Suppose the group ${}^\lambda H^{\mathbb{Q}} = {}^\lambda H = {}^\lambda H_\pi$ is defined and finite for $K = \mathbb{Q}$. Then for an extension K the projection of the group ${}^\lambda H^K$ on $SL(2,\mathbb{C})$ will be ${}^\lambda H^{\mathbb{Q}}$ and

$${}^\lambda H^K \subset {}^\lambda H^{\mathbb{Q}} \times \mathrm{Gal}(K/\mathbb{Q}).$$

There are two possibilities: there exists a K such that the projection of ${}^\lambda H^K$ onto $\mathrm{Gal}(K/\mathbb{Q})$ is an isomorphism or there does not, so that the kernel is never trivial. If our definitions are correct, it should be possible to decide which from the behavior of the $m_H(\rho)$ as K and ρ vary.

Take as an example the case that ${}^\lambda H^{\mathbb{Q}}$ is a cyclic group of odd prime order l, a possibility that will certainly arise. Then ${}^\lambda H^K$ will be a subgroup of $\mathbb{Z}/l\mathbb{Z} \times \mathrm{Gal}(K/\mathbb{Q})$. If it is a proper subgroup, then its projection to $\mathrm{Gal}(K/\mathbb{Q})$ is an isomorphism. If it is not, the case to be considered, then it is the full product. In both cases, $m_H(\sigma_l) = 2$, $m_H(\sigma_{l-2}) = 0$ and $m_H(\rho_a) = 2$ if

$$\rho_a = \sigma_l - \sigma_{l-2}$$

is defined as a virtual representation.

The numbers

$$l-2, l-4, \ldots, 1, -1, \ldots, 2-l$$

[6] On the other hand, we would be using these arguments in combination with the trace formula, in which there is always an implicit upper bound on the ramification of the π that occur. Since π with large finite ${}^\lambda H_\pi$ would, in all likelihood, necessarily have large ramification, we can imagine that these two contrary influences might allow us to remove the unwanted groups from (2).

run over all the nonzero residues of l, so that every nontrivial character of $\mathbb{Z}/l\mathbb{Z} = {}^\lambda H^{\mathbb{Q}}$ appears exactly once in the restriction of the representation σ_{l-2} to ${}^\lambda H^{\mathbb{Q}}$. Suppose that τ is a character of the Galois group of order l and consider the representation,

$$\rho_b = \sigma_{l-2} \otimes \tau.$$

If ${}^\lambda H$ is the full group ${}^\lambda H^{\mathbb{Q}} \times \mathrm{Gal}(K/\mathbb{Q})$, then $m_H(\rho_b) = 0$ because ρ_b does not contain the trivial representation of ${}^\lambda H$. If, on the other hand, it is not the full group and τ factors through $\mathrm{Gal}(K/\mathbb{Q}) \simeq {}^\lambda H^K \to {}^\lambda H^{\mathbb{Q}}$, then it contains the trivial representation exactly once and $m_H(\rho_b) = 1$. Thus

$$\frac{1}{2}\, m_H(\rho_a) - m_H(\rho_b) \neq 0 \tag{3}$$

if ${}^\lambda H^K$ is the full group, but can be 0 if it is not.

The question with which we began is very difficult but an obvious hypothesis lies at hand.

1.4 A splitting hypothesis. Suppose that for some automorphic representation π of Ramanujan type the group ${}^\lambda H_\pi = {}^\lambda H_\pi^K \subset \hat{G} \rtimes \mathrm{Gal}(K/F)$, whose existence is only hypothetical, were finite. Then I expect – and there is no reason to believe that I am alone – that for a perhaps larger extension L and the group ${}^\lambda H_\pi^L$ in $\hat{G} \rtimes \mathrm{Gal}(L/F)$ the projection of ${}^\lambda H_\pi^L$ to $\mathrm{Gal}(L/F)$ will be an isomorphism and that this will then continue to be true for all Galois extensions of F that contain L. Moreover if π is unramified outside a finite set S it is natural to suppose that L can also be taken unramified outside of S and of a degree that is bounded by an integer determined by the order of the intersection of ${}^\lambda H_\pi$ with $\hat{G}$. Thus L could be chosen among one of a finite number of fields.

In general, even when ${}^\lambda H_\pi \cap \hat{G}$ is not finite, we can expect that for some sufficiently large L, the group ${}^\lambda H_\pi^L \cap \hat{G}$ will be connected and that L can be taken unramified where π is unramified and of a degree over K bounded by an integer determined by the number of connected components of ${}^\lambda H_\pi^K \cap \hat{G}$. The observations at the end of the previous section indicate what, at least from the point of view of this paper, the proof of the hypothesis will entail in a special case: it must be shown that the expression (3), which we still do not know how we might calculate, is 0 for at least one of the finitely many cyclic extensions of $\mathbb{Q}$ of order p unramified outside a finite set that depends on the original π. One might expect that the general hypothesis, or rather each case of it, reduces to similar statements.

1.5 Alternative definition of m_π. The integers $m_\pi(\rho)$ have been defined by residues of the logarithmic derivatives of automorphic L-functions at a point $s = 1$ outside the region at which they are known to be absolutely convergent. So it is not clear how this definition might be implemented. Since these integers have been introduced in the hope of broaching the problem of functoriality and thus that of analytic continuation, an alternative definition has to be found that better lends itself to harmonic analysis and to numerical investigation. For this purpose, I recall some familiar basic principles of analytic number theory. Since the extension of the principles and the definitions to other number fields will be patent, I confine myself for simplicity to the rationals.

If $c > 0$ is sufficiently large and $X > 0$, then

$$-\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{L'}{L}(s,\pi,\rho)X^s\frac{ds}{s} \tag{4}$$

is equal to

$$\frac{1}{2\pi i}\sum_p\sum \ln(p)\int_{c-i\infty}^{c+i\infty}\frac{\operatorname{tr}(\rho(A(\pi_p)^k))}{p^{ks}}X^s\frac{ds}{s}. \tag{5}$$

This expansion shows that the integral (4) converges at least conditionally. Those terms of (5) for which $X < p^k$ are 0 as is shown by moving the contour to the right. The finite number of terms for which $X > p^k$ are calculated by moving the integral to the left as a residue at $s = 0$. So (5) is equal to

$$\sum_{p^k<X}\ln(p)\operatorname{tr}(\rho(A(\pi_p)^k))$$

On the other hand, if the L-function can be analytically continued to a region containing the closed half-plane $\operatorname{Re}(s) \geq 1$ where it has no poles except for a finite number at points $1+i\rho_l$, $l = 1,\ldots,n$ and if its behavior in $\operatorname{Im}(s)$ permits a deformation of the contour of integration in (4) to a contour C that except for small semi-circles skirting these points on the left runs directly from $1-i\infty$ to $1+i\infty$ on the line $\operatorname{Re}(s) = 1$, then (4) is (morally) equal to

$$\sum_l\frac{m_{1+i\rho_l}}{1+i\rho_l}X^{1+i\rho_l}+o(X).$$

As a consequence

$$m_\pi(\rho) = m_1 = \lim_{M\to\infty,\,X\to\infty}\frac{1}{M}\int_X^{X+M}\frac{\sum_{p^k<Y}\ln(p)\operatorname{tr}(\rho(A(\pi_p)^k))}{Y}dY. \tag{6}$$

If, for whatever reason, we know that the only possible pole is at 1, then this may be simplified to

$$m_\pi(\rho) = \lim_{X\to\infty}\frac{\sum_{p^k<X}\ln(p)\operatorname{tr}(\rho(A(\pi_p)^k))}{X}. \tag{7}$$

The possible appearance of other poles and thus the introduction of M are simply nuisances that we could well do without.

For summation over primes, the sums ([Lan])

$$\vartheta(X) = \sum_{p<X} \ln(p)$$

are the analogues of the sums over all positive integers

$$\sum_{1 \leq n<X} 1.$$

In particular, $\vartheta(X) = X + o(X)$. Moreover,

$$\psi(X) = \sum_{p^k<X} \ln(p) = \vartheta(X) + o(X).$$

Since it is expected that for π of Ramanujan type the eigenvalues of $\rho(A(\pi_p))$ all have absolute value equal to 1, it is therefore not unreasonable in a tentative treatment to replace (6) and (7), both nothing but possible definitions, by

$$m_\pi(\rho) = \lim_{M\to\infty,\, X\to\infty} \frac{1}{M} \int_X^{X+M} \frac{\sum_{p<Y} \ln(p)\,\mathrm{tr}(\rho(A(\pi_p)))}{Y} dY \tag{8}$$

and by

$$m_\pi(\rho) = \lim_{X\to\infty} \frac{\sum_{p<X} \ln(p)\,\mathrm{tr}(\rho(A(\pi_p)))}{X}. \tag{9}$$

We want to see to what extent these definitions can be given real content and how.

We could modify (5) by replacing the denominator s by $s(s+1)$. The residues at $s = 1 + i\rho_l$ become $1/(1+i\rho_l)(2+i\rho_l)$ and the residue at $s=0$ is replaced by residues at $s=0$ and $s=1$. The result is that

$$m_\pi(\rho) = \lim_{M\to\infty,\, X\to\infty} \frac{2}{M} \int_X^{X+M} \frac{\sum_{p^k<X} \ln(p)(1-p/X)\,\mathrm{tr}(\rho(A(\pi_p)^k))}{X} dX, \tag{6$'$}$$

or in the favorable case that there is only a pole at $s=1$,

$$m_\pi(\rho) = \lim_{X\to\infty} \frac{2\sum_{p^k<X} \ln(p)(1-p/X)\,\mathrm{tr}(\rho(A(\pi_p)^k))}{X}. \tag{7$'$}$$

The two formulas (8) and (9) can be similarly modified. Some of the experiments have been made using (7′), on the somewhat doubtful and certainly untested assumption that this improves convergence.

1.6 The role of the trace formula. As we have already stressed, in the general theory of automorphic forms it is usually unwise to attempt to calculate directly any invariant associated to individual automorphic representations. Rather one calculates the sum – often weighted as, for example, in endoscopy – of the invariants over all automorphic representations of one group and compares them with an analogous sum for a second group, establishing by a term-by-term comparison their equality. For present purposes, what we might hope to calculate from the trace formula is[7]

$$\sum_{\pi} \mu_\pi m_\pi(\rho) \prod_{v\in S} \mathrm{tr}(\pi_v(f_v)). \tag{10}$$

(We have to expect that it will at first be unknown whether the $m_p i(\rho)$ are integers. To show that they are integers comparisons like those envisaged in (15) will very likely be necessary.) The finite-dimensional complex-analytic representation ρ of ${}^LG$ is arbitrary. The set S is a finite set of places of the base field F, including all archimedean places and all places where the group G is not quasi-split and split over an unramified extension, and f_v is a suitable function on $G(F_v)$. Implicitly we also fix a hyperspecial maximal compact subgroup at each place outside of S. The coefficient μ_π is usually a multiplicity; the sum is over automorphic representations of Ramanujan type unramified outside of S, ultimately perhaps only over the cuspidal ones, although it is best not to try without more experience to anticipate exactly what will be most useful – or the exact nature of μ_π.

For the base field $F = \mathbb{Q}$, at this stage an adequate representative of the general case, to arrive at (10) we choose, for each a given prime $p \notin S$, f_q, $q \notin S$ and $q \neq p$ to be the unit element of the Hecke algebra at q and we choose f_p in the Hecke algebra to be such that

$$\mathrm{tr}(\pi_p(f_p)) = \mathrm{tr}(\rho(A(\pi_p))) \tag{11}$$

if π_p is unramified. Then we take $f^p(g) = \prod_v f_v(g_v)$, where f_v, $v \in S$, is given in (10). If R is the representation of G on the space of cuspidal automorphic forms of Ramanujan type and if we can get away with (9), then (10) is equal to

$$\lim_{X\to\infty} \sum_{\pi} \mu_\pi \frac{\sum_{p<X} \ln(p)\,\mathrm{tr}(R(f^p))}{X}. \tag{12}$$

If we use (7′) then (12) is replaced by

$$2 \lim_{X\to\infty} \sum_{\pi} \mu_\pi \frac{\sum_{p<X} \ln(p)(1-p/X)\,\mathrm{tr}(R(f^p))}{X}. \tag{12′}$$

[7] We have to expect that it will at first be unknown whether the $m_\pi(\rho)$ are integers. To show that they are integers, comparisons like those envisaged in (15) will very likely be necessary.

Not only is it not clear at this stage whether it is the representation on the space of cuspidal automorphic forms that is most appropriate or whether it might not be better to include some noncuspidal representations but it is also not clear whether it is best to take the ordinary trace or the stable trace. Such questions are premature. The important questions are whether we can hope to prove that the limit of (12) exists and whether we can find a useful, concrete expression for it.

We shall address some very particular cases of this question in the second part of this paper. Grant for the moment that we have such a representation for representations ρ_k, $1 \le k \le n$. Then for any coefficients a_k we also have an expression for

$$\sum_{\pi} \mu_\pi \sum_{k=1}^{n} a_k m_\pi(\rho_k) \prod_{v \in S} \mathrm{tr}(\pi_v(f_v))$$

If we could find a_k such that

$$\sum_k a_k m_\pi(\rho_k) \tag{13}$$

is equal to 1 if and only if ${}^\lambda H_\pi$ is IP-dominated by a given group ${}^\lambda H$ and is otherwise 0. Then we would have an expression for

$$\sum_{{}^\lambda H_\pi \prec {}^\lambda H} \mu_\pi \prod_{v \in S} \mathrm{tr}(\pi_v(f_v)), \tag{14}$$

the sum being over automorphic representations of G unramified outside of S, principally over cuspidal but perhaps with some noncuspidal terms present as well. The multiplicities μ_π could be ordinary multiplicities, but they will more likely be stable multiplicities and may even depend on ${}^\lambda H_\pi$. As we observed, we will have to content ourselves with satisfying the conditions imposed on (13) approximately; some of the representations for which it is not zero may have to be dealt with separately by an iterative procedure or the argument modified.

The existence of coefficients a_k for which (13) has the desired properties, exactly or approximately, is an algebraic question that I have not broached except for $GL(2)$. The group $GL(2)$ has a center, so that the representation of $GL(2,\mathbb{A})$ on the space of cusp forms is not the direct sum of irreducible representations. To achieve this it is necessary, as usual, to consider the cusp forms transforming under a given character of $Z_+ = \mathbb{R}^+$ and there is no good reason at this stage not to suppose that this character is the trivial character. So we treat the representation on the space of functions on $GL(2,\mathbb{Q})Z_+\backslash GL(2,\mathbb{A})$. Then f_∞ will be a smooth function of compact support on $Z_+\backslash GL(2,\mathbb{R})$ and the only π to be considered are those for which π_∞ is trivial on Z_+. This implies that the central character of π is trivial on $\mathbb{R}^+$. If in addition we suppose that S consists of the infinite place alone – this is an assumption to be made purely for convenience as it removes inessential complications from the preliminary algebra and from the experiments – then we conclude that the only π to be considered are those whose central character is trivial on $\mathbb{R}^+$ and unramified and thus trivial. Since the central character controls the group $\det(H_\pi)$, this means that we are taking only π with ${}^\lambda H_\pi \subset SL(2,\mathbb{C})$, or, more precisely, ${}^\lambda H_\pi \subset SL(2,\mathbb{C}) \times \mathrm{Gal}(K\backslash\mathbb{Q})$. These are the very simple groups that we considered in a previous section and for which we are in a position to find – insoar as they are available – the coefficients of (13).

1.7 Comparison. If we managed by a combination of the trace formula with various limiting processes to obtain a formula for (14), then we would want to compare it with the trace formula on ${}^\lambda H$ itself, except that ${}^\lambda H$ may not be an L-group, for it may not be defined by a semidirect product. When, however, the kernel of ${}^\lambda H^K \to \mathrm{Gal}(K/\mathbb{Q})$ is connected, it is possible as a consequence of, for example, Prop. 4 of [L1] to imbed the center $\hat{Z}$ of $\hat{H}$, the connected component of the identity in ${}^\lambda H$, in the connected dual $\hat{T}$ of a product of tori, $T = \prod_i K_i^\times$, where each K_i is a field over F, and to imbed it in such a way that ${}^L\tilde{H}$, the quotient of the semidirect product $\hat{T} \rtimes {}^\lambda H$ by the diagonally imbedded $\hat{Z}$ becomes an L-group.[8] Notice that the Galois group $\mathrm{Gal}(K/\mathbb{Q})$ acts on $\hat{T}$, so that ${}^\lambda H$ does as well. Maps ϕ into ${}^\lambda H$ may be identified with maps into ${}^L\tilde{H}$ that correspond to automorphic representations of $\tilde{H}$ whose central character is prescribed by the structure of ${}^\lambda H$. They can be presumably be identified in the context of the trace formula.

Then to make use of (14), we would have to introduce a transfer $f \to f^H$ from functions on $G(\mathbb{A}_F)$ to functions on $H(\mathbb{A}_F)$ (if ${}^\lambda H = {}^L H$ is an L-group but to functions on $\tilde{H}(\mathbb{A}_F)$ in the general case) and compare (14) with

$$\sum \prod_{v \in S} \mathrm{tr}(\pi'_v(f_v^H)), \tag{15}$$

the sum being over automorphic representations of H of, say, Ramanujan type unramified outside of S, so that there will also be a formula for (15) which is to be compared with that for (14). The difference between IP-domination and inclusion will undoubtedly complicate this comparison.

There is no reason not to admit the possibility that (14) is replaced by a sum over groups ${}^\lambda H$,

$$\sum_{{}^\lambda H} \sum_{{}^\lambda H_\pi \prec {}^\lambda H} \mu_\pi \prod_{v \in S} \mathrm{tr}(\pi_v(f_v)). \tag{14$'$}$$

[8] The L-group may have to be defined by the Weil group and not by the Galois group, but that is of no import.

Then (15) would be replaced by a similar sum (15′).

It is perhaps well to underline explicitly the differences between the comparison envisaged here and endoscopic comparison. For endoscopy the transfer $f \to f^H$ is defined in terms of a correspondence between conjugacy classes. In general, the transfer $f \to f^H$, which is defined locally, will be much less simple. There will already be much more knowledge of local harmonic analysis, especially of irreducible characters, implicit in its definition. Secondly, there will be difficult *analytic* problems to overcome in taking the limit of the trace formula on G. Thirdly, the groups ${}^\lambda H$ that occur are essentially arbitrary subgroups of ${}^L G$, not just those defined by endoscopic conditions.

1.8 Further concrete cases. I consider $GL(2)$ and icosahedral representations but in two different ways. The ground field F may as well be taken to be $\mathbb{Q}$. Suppose $K/\mathbb{Q}$ is a Galois extension and $\mathrm{Gal}(K/\mathbb{Q})$ admits an imbedding τ in $GL(2,\mathbb{C})$ as an icosahedral representation. Thus $\mathrm{Gal}(K/\mathbb{Q})$ is an extension of the icosahedral group by $\mathbb{Z}_2$. Take ${}^L G = {}^L G^K$ and consider $\rho = \sigma_1 \otimes \tilde{\tau}$, where $\tilde{\tau}$ is the contragedient of τ. If $m_H(\rho) \neq 0$, then σ_1 and τ define the same representation of ${}^\lambda H$. Therefore the kernels of ${}^\lambda H \to \mathrm{Gal}(K/\mathbb{Q})$ and ${}^\lambda H \to GL(2,\mathbb{C})$ are the same and thus $\{1\}$. So the projection of ${}^\lambda H$ to $\mathrm{Gal}(K/\mathbb{Q})$ is an isomorphism; ${}^\lambda H$ is an L-group, that attached to the group $H = \{1\}$; and σ_1 restricted to ${}^\lambda H$ is τ, or rather the composition of τ with the isomorphism ${}^\lambda H \to \mathrm{Gal}(K/\mathbb{Q})$.

Thus we can expect that $m_\pi(\rho) \neq 0$ if and only if $\pi = \pi(\tau)$ is the automorphic representation attached to τ by functoriality. To compare (14), provided we can find such a formula, and (15) we will need to define the local transfer $f_v \to f_v^H$ by means of the characters of $\pi_v(\tau)$.

On the other hand, define the L-group ${}^L G$ to be ${}^L G^{\mathbb{Q}}$ and take $\rho = \sigma_{12} - \sigma_8$. We have seen that $m_H(\rho)$ is nonzero only if ${}^\lambda H$ is a subgroup of the icosahedral group or perhaps a finite abelian or dihedral group that can be treated independently. Then $m_\pi(\rho)$ will be nonzero only if ${}^\lambda H_\pi$ is such a subgroup. There will be many such π and although (15) will not have to take them all into account, it will have to contain a sum over all icosahedral extensions unramified outside a given set of places.

So the first approach has at least one advantage: it singles out a unique π. It may have another. Numerical experiments involving σ_1 are manageable. Those for σ_m quickly become impossible as m grows. Even $m = 3$ is very slow. On the other hand, the first approach alone cannot, so far as I can see, assure us that if ${}^\lambda H_\pi \subset {}^L G^{\mathbb{Q}}$ is an icosahedral group, then π is associated to an icosahedral representation of the Galois group. No matter what τ we choose, it necessarily overlooks π for which this is false.

1.9 A cautionary example. Take the group G to be $GL(1)$ over $\mathbb{Q}$ and take ${}^\lambda H$ to be the finite group of order m in ${}^L G = \mathbb{C}^\times$. If ρ is the representation $z \to z^m$, then $m_{H'}(\rho) = 1$ if ${}^\lambda H'$lies in ${}^\lambda H$ and is otherwise 0. Let S be, as usual, a finite set of places containing the infinite places. In order to have a discrete spectrum under the action of $G(\mathbb{A})$, we consider functions, thus automorphic forms, on $\mathbb{R}^+\mathbb{Q}^\times\backslash I$, I being the group of idèles. This is the space $R^+G(\mathbb{Q})\backslash G(\mathbb{A})$. The function $f = \prod_v f_v$ will be such that f_∞ is in fact a function on $\mathbb{R}^+\backslash\mathbb{R}^\times = \{\pm 1\}$. The function f_p is the characteristic function of the set of integral γ with $|\gamma| = p^{-m}$.

If we take the measure on $\mathbb{R}^+\backslash G(\mathbb{A})$ to be a product measure, with the measure of $G(\mathbb{Z}_p)$ and of $\mathbb{R}^+\backslash\mathbb{R}^+$ equal to 1, then $\mu(\mathbb{R}^+G(\mathbb{Q})\backslash G(\mathbb{A}))$ is equal to 1 and

$$\operatorname{tr} R(f) = \sum_\pi \operatorname{tr} \pi(f) = \sum_{\gamma\in\mathbb{Q}^\times} f(\gamma). \tag{16}$$

The element γ must be equal to ap^m, where

$$a = \pm \prod_{q\in S} q^{\alpha_q}. \tag{17}$$

Thus the expression (16) is equal to $g(p^m)$, where g is the function on $\prod_{q\in S}\mathbb{Z}_q$ given by $g(x) = \sum f(ax)$, the sum being over all a of the form (17).

Thus (12) is

$$\lim_{X\to\infty} \frac{\sum_{p<X} \ln(p) g(p^m)}{X},$$

which is equal to

$$\frac{\sum_{x \bmod M} g(x^m)}{\varphi(M)},$$

where M is a positive integer that is divisible only by primes in S and that depends on the collection of functions f_q, $q \in S$, each of them being smooth. The number $\varphi(M)$ is the order of the multiplicative group of $\mathbb{Z}/M\mathbb{Z}$. In terms of f, this is

$$\frac{\int_{\mathbb{R}^+\mathbb{Q}_S I_S^m} f(x)dx}{\int_{\mathbb{R}^+\mathbb{Q}_S I_S^m} dx}. \tag{18}$$

where $\mathbb{Q}_S$ is the set of nonzero rational numbers that are units outside of S and I_S is the product $\prod_{v\in S}\mathbb{Q}_v$, the first regarded as a subgroup of the second.

The expression (18) is certainly in an appropriate form and is equal to

$$\sum_\chi \chi(f),$$

where χ runs over all characters of $\mathbb{R}^+\mathbb{Q}_S I_S$ of order dividing m. This, however, is pretty much the point from which we began. We are still left, as in class-field theory, with the problem of showing that these characters can be deduced from characters of the Galois group. Thus we cannot expect that the trace formula will spare us the arithmetical investigations. It will, at best, make it clear what these must be.

Part II: Preliminary analysis

2.1 Measures and orbital integrals. In this part of the paper, we shall review the trace formula for $GL(2)$, the only group with which we are seriously concerned at present, and examine the possibility of obtaining an expression for (14) or (15′). It would be worthwhile to undertake a similar study of the trace formula for other groups. If the general trace formula admits a similar analysis and transformation, it will be an encouraging sign.

To obtain expressions that can then be used for numerical purposes, we have to be clear about the conventions. As we already observed, we shall consider automorphic forms on $G(\mathbb{Q})Z_+\backslash G(\mathbb{A})$, $\mathbb{A} = \mathbb{A}_\mathbb{Q}$ and $G = GL(2)$. The functions whose trace is to be calculated are functions on $Z_+\backslash G(\mathbb{A})$ and are taken to be products $f(g) = \prod_v f_v(g_v)$. The measure on $Z_+\backslash G(\mathbb{A})$ is to be a product measure as is the measure on $Z_+\backslash G_\gamma(\mathbb{A})$ if γ is regular and semisimple. The group G_γ is then defined by the multiplicative group of a ring E_γ, the centralizer of γ in the ring of 2×2 matrices. At a nonarchimedean place p, the subgroup $G_\gamma(\mathbb{Z}_p)$ has a natural definition and we normalize the local measures by the conditions:

$$\mu(G_\gamma(\mathbb{Z}_p)) = 1, \quad \mu(G(\mathbb{Z}_p)) = 1.$$

At infinity, the choice of measure on $Z_+\backslash G(\mathbb{R})$ is not important, nor is that on $Z_+\backslash G_\gamma(\mathbb{R})$. It is not necessary to be explicit about the first, but it is best to be explicit about the second.

(a) *Elliptic torus.* Here I mean that the torus is elliptic at infinity and thus that $E = E_\gamma$ is an imaginary quadratic extension. I assume, for simplicity, that it is neither $\mathbb{Q}(\sqrt{-2})$ nor $\mathbb{Q}(\sqrt{-3})$. An element in $G_\gamma(\mathbb{R})$ is given by its eigenvalues, $\sigma e^{i\theta}$ and $\sigma e^{-i\theta}$. The value of $\sigma > 0$ is irrelevant and I take the measure to be $d\theta$. The volume of

$$Z_+G_\gamma(\mathbb{Q})\backslash G_\gamma(\mathbb{A}) = Z_+E^\times\backslash I_E \tag{19}$$

is the class number C_E times the measure of

$$\pm Z_+\backslash\mathbb{C}^\times \times \prod_p G_\gamma(\mathbb{Z}_p),$$

which, according to the conventions chosen, is the measure of $\pm Z_+\backslash\mathbb{C}^\times$ or

$$\int_0^\pi d\theta = \pi.$$

(b) *Split torus.* Once again, the torus is only to be split at infinity, so that E_γ is a real quadratic field. If the eigenvalues of an element δ are α and β, set

$$\begin{aligned}
r &= \alpha+\beta, \\
N &= 4\alpha\beta, \\
\frac{r}{\sqrt{|N|}} &= \frac{1}{2}(\operatorname{sgn}\alpha\sqrt{\frac{|\alpha|}{|\beta|}} + \operatorname{sgn}\beta\sqrt{\frac{|\beta|}{|\alpha|}}) = \pm\frac{1}{2}(\lambda\pm\lambda^{-1}), \\
\lambda &= \sqrt{\frac{|\alpha|}{|\beta|}}, \quad \sigma = \sqrt{|\alpha\beta|}, \\
\alpha &= \pm\sigma\lambda, \quad \beta = \pm\frac{\sigma}{\lambda}.
\end{aligned} \tag{20}$$

The value of σ is irrelevant and I take the measure to be $\frac{d\lambda}{\lambda}$. Notice that

$$d(\frac{r}{\sqrt{|N|}}) = \frac{1}{2}(1\mp\lambda^{-2})d\lambda = \frac{1}{2}(1-\frac{\beta}{\alpha})d\lambda. \tag{21}$$

The upper sign is that of N. The parameters $r = \operatorname{tr}\delta$ and $N = 4\det\delta$ can also be defined when the torus is elliptic at infinity or globally or at any other place. When the torus is elliptic at infinity,

$$d(\frac{r}{\sqrt{|N|}}) = d\cos\theta = \frac{i}{2}(1-\lambda^{-2})d\lambda, \quad \lambda = e^{i\theta}. \tag{22}$$

The fundamental unit ϵ can be taken to be the unit with the smallest absolute value $|\epsilon| > 1$. Thus $\ln|\epsilon|$ is the regulator as it appears in [C]. The measure of the quotient (19) is now the class number times the measure of

$$\pm\mathbb{R}^+\backslash\mathbb{R}^\times \times \mathbb{R}^\times/\{\epsilon^k | k\in\mathbb{Z}\}. \tag{23}$$

Since $\pm\mathbb{R}^+\backslash\mathbb{R}^\times \times \mathbb{R}^\times$ can be identified with $\mathbb{R}^\times$ by projecting on the first factor, the measure of (23) is $2\ln|\epsilon|$, in the notation of [C], $2h(D)R(D)$ if D is the discriminant of the field E_γ.

There is a very small point to which attention has to be paid when computing with the trace formula. Locally there are two measures to be normalized, that on $G_\gamma(\mathbb{Q}_v)\backslash G(\mathbb{Q}_v)$ and that on $Z_+\backslash G_\gamma(\mathbb{R})$ or $G_\gamma(\mathbb{Q}_p)$. They appear in two ways in the measure on $G(\mathbb{Q}_v)$: once when fixing it, as

$d\delta d\bar{g}$, by the measure on the subgroup $G_\gamma(\mathbb{Q}_p)$ (or $Z_+\backslash G_\gamma(\mathbb{R})$) and the measure on the quotient space $G_\gamma(\mathbb{Q}_v)\backslash G(\mathbb{Q}_v)$; and once, as in the Weyl integration formula, when fixing the measure on

$$\{g^{-1}\delta g \mid \delta \in G_\gamma(\mathbb{Q}_v),\ g \in G(\mathbb{Q}_v)\}$$

by means of the map

$$(\delta, g) \to g^{-1}\delta g, \quad G_\gamma(\mathbb{Q}_v) \times (G_\gamma(\mathbb{Q}_v)\backslash G(\mathbb{Q}_v)) \to G(\mathbb{Q}_v). \tag{24}$$

Since (24) is a double covering, the measure to be used in the Weyl integration formula is

$$\frac{1}{2}\prod_\alpha |1-\alpha(\delta)| d\delta d\bar{g},$$

the product over α being a product over the two roots of the torus.

If m is a nonnegative integer, let T_p^m be the characteristic function of

$$\{X \in \mathrm{Mat}(\mathbb{Z}_p)\big| \quad |\det X| = p^{-m}\},$$

where $\mathrm{Mat}(\mathbb{Z}_p)$ is the algebra of 2×2-matrices over $\mathbb{Z}_p$. If $\rho = \sigma_m$, then $T_p^m/p^{m/2}$ is the function f_p of (11). In other words,

$$\mathrm{tr}\,\pi_p(T_p^m) = p^{m/2}\sum_{k=0}^m \alpha^{m-k}(\pi_p)\beta^k(\pi_p)$$

if π_p is unramified and $\alpha(\pi_p)$ and $\beta(\pi_p)$ are the eigenvalues of $A(\pi_p)$. I recall the standard calculation.

Take π_p to be the usual induced representation, so that the vector fixed by $GL(2,\mathbb{Z}_p)$ is the function

$$\phi(ntk) = |a|^{-s_1+1/2}|b|^{-s_2-1/2}, \qquad t = \begin{pmatrix} a & 0 \\ 0 & b\end{pmatrix},$$

and $\{\alpha(\pi_p), \beta(\pi_p)\} = \{p^{s_1}, p^{s_2}\}$. Then

$$\int \phi(g)T_p^m(g)dg = \sum_{k=0}^m p^{(m-k)s_1+ks_2}p^{(2k-m)/2}\int_{|x|\,\leq\, p^{m-k}} dx = p^{m/2}\sum_{k=0}^m p^{(m-k)s_1+ks_2}.$$

We shall need the orbital integrals of the functions $T_p^m/p^{m/2}$ for all m, but $m=0$ is particularly important as it is the unit element of the Hecke algebra. The pertinent calculations can be found in [JL] but there is no harm in repeating them here. If γ is a regular semisimple element in $G(\mathbb{Q}_p)$, set

$$U^m(\gamma) = \int_{G_\gamma(\mathbb{Q}_p)\backslash G(\mathbb{Q}_p)} T_p^m(g^{-1}\gamma g). \tag{25}$$

Denote the two eigenvalues of γ by γ_1 and γ_2 and extend the usual norm on $\mathbb{Q}_p$ to $\mathbb{Q}_p(\gamma_1,\gamma_2)$ or to $E_p = E_\gamma \otimes \mathbb{Q}_p$, which we identify, taking $\gamma_1 = \gamma$. The ring of integral elements in E_p is of the form $\mathbb{Z}_p \oplus \mathbb{Z}_p\Delta$. If $\bar\Delta$ is the conjugate of Δ, so that $\Delta + \bar\Delta = \mathrm{tr}(\Delta)$, set $\delta_\gamma = p|\Delta - \bar\Delta|$. Let $\gamma_1 - \gamma_2 = b(\Delta - \bar\Delta)$ with $|b| = p^{-k}$, $k = k_\gamma$.

Lemma 1. *$U^m(\gamma)$ is 0 unless γ_1 and γ_2 are integral and $|\gamma_1\gamma_2| = p^{-m}$, when it is given by the following formulas.*

(a) *If γ is split then (25) is*

$$p^k = \frac{1}{|\gamma_1 - \gamma_2|}.$$

(b) *If γ is not split and E_γ is unramified then (2) is*

$$p^k \frac{p+1}{p-1} - \frac{2}{p-1}.$$

(c) *If γ is not split and E_γ is ramified then (25) is*

$$\frac{p^{k+1}}{p-1} - \frac{1}{p-1}.$$

The proof is familiar and easy. As the lemma is basic to our calculations, I repeat it. The value of the characteristic function $T_p^m(g^{-1}\gamma g)$ is 1 if and only if $g^{-1}\gamma g$ takes the lattice $L_0 = \mathbb{Z}_p \oplus \mathbb{Z}_p$ into itself and has determinant with absolute value p^{-m}, thus only if it stabilizes the lattice and $|\det(\gamma)| = p^{-m}$. Thus, assuming this last condition, if and only if γ stabilizes $L = gL_0$. Knowing L is equivalent to knowing g modulo $G(\mathbb{Z}_p)$ on the right. Multiplying g on the left by an element of $G_\gamma(\mathbb{Q}_p) = E_p^\times$ is equivalent to multiplying L by the same element.

If E_p is split, then we can normalize L up to such a multiplication by demanding that

$$L \cap \{(0,z) \,|\, z \in \mathbb{Q}_p\} = \{(0,z) \,|\, z \in \mathbb{Z}_p\}$$

and that its projection onto the first factor is Z_p. Then the x such that $(1,x)$ lies in L are determined modulo Z_p by L. Multiplying by

$$\begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}, \quad |\alpha| = |\beta| = 1,$$

we replace $(1,x)$ by $(1,\beta x/\alpha)$, so that only the absolute value $|x|$ counts. The measure in $G_\gamma(\mathbb{Q}_p)\backslash G(\mathbb{Q}_p)$ of the set of g giving the lattice L is the index in $G_\gamma(\mathbb{Z}_p)$ of the stabilizer of L. This is just the number of y modulo $\mathbb{Z}_p$ with the same absolute value as x (or with $|y| \le 1$ if $|x| \le 1$), thus the number of lattices that can be obtained from the given one by multiplying by an element of $G_\gamma(\mathbb{Z}_p)$. The condition that L be fixed by $\gamma = (\gamma_1, \gamma_2)$ is that γ_1 and γ_2 be integral and that

$$(\gamma_1, \gamma_2 x) = \gamma_1(1,x) + (0, (\gamma_2 - \gamma_1)x)$$

lie in L, thus that $(\gamma_1 - \gamma_2)x$ be integral. We conclude finally that (25) is equal to $1/|\gamma_1 - \gamma_2|$.

The argument is the same in the remaining cases. Identifying $G(\mathbb{Q}_p)$ with the automorphisms of the vector space E_p, we identify the quotient $G(\mathbb{Q}_p)/G(\mathbb{Z}_p)$ with the lattices in E_p. Modulo the action of $G_\gamma(\mathbb{Q}_p)$, these can be put in the form $\mathbb{Z}_p + \mathbb{Z}_p p^j\Delta$, $j \geq 0$. Such a lattice is fixed by γ if and only if $k \geq j$. In the unramified case, the stabilizer of the lattice in $G_\gamma(\mathbb{Z}_p)$ has index 1 if $j = 1$ and index $p^j(1 + 1/p)$ otherwise. So (25) is equal to

$$1 + \sum_{j=1}^{k} p^j(1 + \frac{1}{p}) = p^k\frac{p+1}{p-1} - \frac{2}{p-1},$$

as asserted by (b). If E_p is ramified, the stabilizer has index 1 if $j = 0$ and index p^j otherwise. So (25) is now equal to

$$1 + \sum_{j=1}^{k} p^j = \frac{p^{k+1}}{p-1} - \frac{1}{p-1}.$$

This is (c).

This lemma provided us with the orbital integrals that we need outside of S. The discussion inside of S is quite different. Since we are going to take, for the present purposes, $S = \{\infty\}$, I confine myself to this case. The same principles apply in all cases. Over the field $\mathbb{R} = \mathbb{Q}_\infty$, the necessary information is in the discussion of HCS-families in Chap. 6 of [L2] although it is not elegantly expressed. Let $\mathrm{ch}(\gamma) = (4\,\mathrm{Nm}(\gamma), \mathrm{tr}(\gamma))$. For any γ in $GL(2, \mathbb{R})$,

$$\int f_\infty(g^{-1}\gamma g)dg = \psi(\mathrm{ch}(\gamma)) = \psi'_\infty(\mathrm{ch}(\gamma)) + \psi''_\infty(\mathrm{ch}(\gamma))\frac{|\,\mathrm{Nm}\,\gamma|^{1/2}}{|\gamma_1 - \gamma_2|} \tag{26}$$

where ψ'_∞ and ψ''_∞ depend on f_∞. The second is a smooth function on the plane with the y-axis removed. The first is 0 outside the parabola $y^2 - x \leq 0$, but inside and up to the boundary of this parabola, it is a smooth function of x and $y^2 - x$. The functions ψ' and ψ'' are not uniquely determined. Since we have taken f positively homogeneous, the function ψ is positively homogeneous, $\psi(\lambda^2 N, \lambda r) = \psi(N, r)$ for $\lambda > 0$. Thus it is determined by the two functions $\psi(\pm 1, r)$ on the line. The function $\psi_- = \psi(-1, r)$ is smooth; the function $\psi_+ = \psi(1, r)$ may not be. They are both compactly supported.

If $\theta(\gamma) = \theta(\mathrm{ch}(\gamma))$ is any positively homogeneous class function on $G(\mathbb{R})$, the Weyl integration formula and formulas (21) and (22) give

$$\begin{aligned}\int_{Z_+\backslash G(\mathbb{R})} \theta(g)f(g)dg &= \frac{1}{2}\sum \int_{Z_+\backslash T(R)} \theta(\mathrm{ch}(\gamma))\psi(\mathrm{ch}(\gamma))|1 - \frac{\alpha}{\beta}||1 - \frac{\beta}{\alpha}|d\lambda|\lambda| \\ &= 4\sum \int_{-\infty}^{\infty} \psi_\pm(r)\theta(\pm 1, r)\sqrt{|r^2 \mp 1|}dr\end{aligned} \tag{27}$$

because

$$|1-\frac{\alpha}{\beta}||1-\frac{\beta}{\alpha}|\frac{d\lambda}{|\lambda|}=2|1-\frac{\alpha}{\beta}||\lambda|\frac{dr}{\sqrt{|N|}},$$
$$\lambda(1-\frac{\alpha}{\beta})=\lambda\mp\lambda^{-1},$$
$$\frac{r^2}{N}-1=\pm(\lambda\pm\lambda^{-1})^2-1=\pm\frac{(\lambda\mp\lambda^{-1})^2}{4},$$

and

$$2|\lambda\mp\lambda^{-1}|\frac{dr}{\sqrt{|N|}}=4\sqrt{|\frac{r^2}{N}-1|}\frac{dr}{\sqrt{|N|}}.$$

The sums in (27) are over the two tori and then, in the last line, over the two possible signs. The elliptic torus corresponds to the region $-1<r<1$, $N=1$; the split torus to the rest. The factor $1/2$ is removed in the passage from the first to the second line of (27) because the map $\gamma\to\mathrm{ch}(\gamma)$ from each of the tori to the plane is also a double covering.

The formula (27) is applicable if θ is a one-dimensional representation of $G(\mathbb{R})$, in particular if it is identically equal to 1, and then (27) yields

$$\mathrm{tr}(\theta(f))=4\sum\int_{-\infty}^{\infty}\psi_\pm(r)\sqrt{|r^2\mp1|}dr \tag{28}$$

Another possibility is to take θ to be the character of the representation π_χ unitarily induced from a character

$$\chi:\begin{pmatrix}\alpha&0\\0&\beta\end{pmatrix}\to\mathrm{sgn}\,\alpha^k\,\mathrm{sgn}\,\beta^l$$

of the diagonal matrices. Only the parities of k and l matter. The character is 0 on the elliptic elements, where $N>0$ and $r^2<N$. Otherwise it is constant on the four sets determined by fixing the signs of N and r, where it is given by

$$(\mathrm{sgn}\,N+1)\frac{\mathrm{sgn}(r)}{\sqrt{|1-\alpha/\beta||1-\beta/\alpha|}} \tag{29.a}$$

if $k\neq l$ and by

$$2\frac{\mathrm{sgn}(N)^l}{\sqrt{|1-\alpha/\beta||1-\beta/\alpha|}} \tag{29.b}$$

otherwise. The eigenvalues α and β of γ with $\mathrm{ch}(\gamma)=(N,r)$ are of course one-half the roots of $x^2-2rx+N=0$. Since

$$\frac{|\gamma_1-\gamma_2|^2}{|\,\mathrm{Nm}\,\gamma|}=|1-\alpha/\beta||1-\beta/\alpha|=|\lambda|^2|1\mp\lambda^{-2}|^2=4|\frac{r^2}{N}-1|,$$

we conclude that $\psi_\pm$, although not necessarily bounded, are integrable functions of r and that $\mathrm{tr}(\pi_\chi(f))$ is given by

$$4\{\int_{-\infty}^{-1}\epsilon_{+-}\psi_+(r)+\int_1^\infty\epsilon_{++}\psi_+(r)+\int_{-\infty}^0\epsilon_{--}\psi_-(r)+\int_0^\infty\epsilon_{-+}\psi_-(r)\},\tag{30}$$

where the constants $\epsilon_{\pm\pm}$, which are ±1 or 0, are to be chosen as prescribed by (29).

2.2 Calculating with the trace formula. Rather than refer to Arthur's general trace formula as I should if I were intent on preparing for the general case, I prefer to appeal to the formula on pp. 516-517 of [JL] with which I am more at ease and to which the reader is requested to refer. There are eight terms in that formula, but for a base field of characteristic zero the term (iii) is absent. We shall also only consider, for reasons already given, automorphic representations whose central character is trivial on $\mathbb{R}^+$. The formula of [JL] gives the sum of the traces of $\pi(f)$ over all automorphic representations occurring discretely in $L^2(Z_+G(\mathbb{Q})\backslash G(\mathbb{A}))$. So we need to subtract those representations that are not of Ramanujan type. For $G = GL(2)$, these are the one-dimensional representations. Their traces will be subtracted from the term (ii) of [JL] and the difference will be more important than (ii) itself. We refer to the difference as the elliptic term. It is the most difficult and will be discussed – not treated – in §2.5.

The principal question that concerns us is whether there are any possible developments in analytic number theory that might enable us to find an explicit expression for (12). The numbers μ_π are here equal to 1 (or, for those π that are absent from the sum, 0). Lacking all experience, I fell back on the obvious and made explicit calculations. For $\rho = \sigma_m \otimes \tau$ they are feasible and not all too slow for $m = 1$. For $m = 2, 3$ something can still be done, but for higher m, at least with my inefficient programs, they are too slow to provide any useful information. On the other hand, as the first problem of §1.8 demonstrates, calculations for $m = 1$ are of considerable interest provided that we take the tensor product of σ_m with a general τ or even just a τ of icosahedral type. Although taking such a tensor product demands a simultaneous study of icosahedral representations or other Galois representations, there is no reason not to expect that the important features of the problem are not already present for τ trivial and that they persist. Of course, there may be accidental features, but these the wise student should recognize and resolutely ignore. The addition of the Galois representation will add to the labor but should not put additional demands on raw computer power, only on the skill of the programmer. So I confine myself to trivial τ and, by and large, to $m = 1$. Although it is important for theoretical purposes to envisage taking S arbitrarily large, computations and theory for larger S should not differ essentially from the case that S consists of the infinite place alone, although there will be many more terms in the trace formula to be taken into account.

The sum over $r \in \mathbb{Z}$ that occurs in the elliptic term will be replaced by sums over r satisfying a congruence condition. This will entail that whatever behavior we find for $S = \{\infty\}$ should remain valid when congruence conditions are imposed. Such an assertion, which implies a greater theoretical regularity that may make the proofs easier to come by, has to be tested further, but these are the principles that justify confining myself at first to $m = 1$ and to representations unramified at all finite places. We know of course a good deal about such representations. In particular, there are none of Galois type, but this is an accidental circumstance that we will use to verify that the programs are functioning well but that will be otherwise irrelevant to our conclusions.

The representations are to be unramified at all finite places; so the central character η of [JL] is trivial. Since we will also examine, at least briefly, some $m > 1$, I do not fix m to be 1. The representation τ will be, however, trivial. The trace formula replaces the expression (12) by a sum of seven terms, corresponding to its seven terms. The function Φ of [JL] is now being denoted f, f^p or even $f^{p,m}$ and

$$f^{p,m}(g) = f_\infty(g_\infty) f_p^m(g_p) \prod_{q \neq p} f_q(g_q),$$

in which f_∞ is a variable function, f_p^m depends on m, but all the other f_q do not depend on m. Thus the function Φ does not satisfy the conditions of [JL]; it does not transform according to a character of the center $Z(\mathbb{A})$ of $G(\mathbb{A})$ and the resulting trace formula is different, but not very different. In (i) there is a sum over the scalar matrices. In (ii) and (iv) there are sums over the full tori, not just over the tori divided by the center. In (v) there is also a sum over the scalar matrices, the n_0 defining $\theta(s, f_v)$ being replaced by zn_0. In principle, (vi), (vii) and (viii) are different, but because f_q is a spherical function for all q, the sum over (μ, ν) implicit in these expressions reduces to the single term $\mu = \nu = 1$.

Since we are in a situation where (7) is appropriate and (6) unnecessary, the contribution of the first term of the trace formula to (12) is given by

$$(TF.1) \qquad \sum_{Z(\mathbb{Q})} \frac{\mu(Z_+ G(\mathbb{Q}) \backslash G(\mathbb{A}))}{X} \sum \ln(p) f^m(z).$$

Since

$$f^m(z) = \frac{f_\infty(z)}{p^{m/2}},$$

if $z = \pm p^{m/2}$ and 0 otherwise, the limit that appears in (12) or (12′) will be 0.

The second term is the elliptic term to be treated in the next section. None of the terms (iv), (v) and (viii) of [JL], is invariant on its own, so that some recombination of these terms is necessary. The terms (vi) and (vii) can, however, be treated directly.

I begin with (vi), which yields a contribution that is not in general 0. Since f_q is a spherical function for all q, the only pair (μ,ν) that contributes to (vi) or to (vii) is the pair of trivial characters and $\rho(\cdot,s)$, denoted ξ_s in this paper to avoid a conflict of notation, is the global (or local) representation unitarily induced from the representation

$$\begin{pmatrix} \alpha & x \\ 0 & \beta \end{pmatrix} \to |\alpha|^{\frac{s}{2}}|\beta|^{-\frac{s}{2}}$$

of the adelic superdiagonal matrices. It is, moreover, easily verified that $M(0)$ is the operator $-I$. Thus the contribution of (vi) to (12) is

$$(TF.2) \qquad \frac{1}{4X}\sum \ln(p)\,\mathrm{tr}(\xi_0(f_\infty))\,\mathrm{tr}(\xi_0(f_p^m)).$$

Since $\mathrm{tr}(\xi_0(f_p^m)) = m+1$, the limit as $X\to\infty$ is

$$(31) \qquad \frac{m+1}{4}\,\mathrm{tr}(\xi_0(f_\infty)).$$

From (30) we conclude that for $m=1$, this is

$$(32) \qquad 2\left\{\int_{-\infty}^{-1}\psi_+(r)dr + \int_1^\infty \psi_+(r)dr + \int_{-\infty}^\infty \psi_-(r)dr\right\}.$$

Apart from the elliptic term, this will be the only nonzero contribution to the limit. Since the standard automorphic L-function $L(s,\pi,\sigma_1)$ does not have poles on $\mathrm{Re}(s)=1$, we expect that (32) will be cancelled by the elliptic term. This is accidental and will not be for us, even numerically, the principal feature of the elliptic term.

The function $m(s)$ that appears in (vii) is

$$\pi\frac{\Gamma((1-s)/2)}{\Gamma((1+s)/2)}\frac{\zeta(1-s)}{\zeta(1+s)}.$$

Thus

$$(33) \qquad \frac{m'(s)}{m(s)} = -\frac{1}{2}\frac{\Gamma'((1-s)/2)}{\Gamma((1-s)/2)} - \frac{1}{2}\frac{\Gamma'((1+s)/2)}{\Gamma((1+s)/2)} - \frac{\zeta'(1-s)}{\zeta(1-s)} - \frac{\zeta'(1+s)}{\zeta(1+s)}.$$

It is to be multiplied by the product of

$$(34) \qquad \mathrm{tr}(\xi_s(f_\infty))$$

and $\mathrm{tr}(\xi_s(f_p^m)$. The first of these two functions, as a function on $(-i\infty, i\infty)$, is the Fourier transform of a smooth function of compact support. The second is equal to

$$p^{im\frac{s}{2}} + p^{i(m-2)\frac{s}{2}} + \ldots + p^{i(2-m)\frac{s}{2}} + p^{-im\frac{s}{2}}. \tag{35}$$

The estimates of §48 and §77 of [Lan] assure us that the product of (33) and (34) is an L^1-function on $(-i\infty, i\infty)$. From the Riemann-Lebesgue lemma we then conclude that, for odd m, the integral of the product of (33), (34) and (35) over that line approaches 0 as p approaches infinity. So, for m odd, (vii) does not contribute to the limit in (12) or (12′).

2.3 The noninvariant terms. Both $\omega(\gamma, f_v)$ and $\omega_1(\gamma, f_v)$ are 0 unless there is a matrix

$$n = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \quad x \in F_v,$$

and a matrix k in the maximal compact subgroup of $GL(2, \mathbb{Q}_v)$ such that the element $k^{-1}n^{-1}\gamma nk$ lies in the support of f_v. Since f_∞ is fixed at present, this means that the two eigenvalues α and β of γ are units away from p and that there is a fixed $\delta > 0$ such that $\delta < |\alpha/\beta|_\infty < 1/\delta$. From the product formula we conclude that $\delta < |\alpha/\beta|_p < 1/\delta$. Since $\alpha\beta = \alpha^2(\beta/\alpha) = \pm p^m$ if $\omega(\gamma, f_v)$ or $\omega_1(\gamma, f_v)$ is not 0, we conclude that (iv) is 0 for all but a finite number of p if m is odd and thus does not contribute to (12) or (12′). If m is even, there are only a finite number of γ that yield a nonzero contribution to (iv). Indeed, such γ have to be of the form

$$\gamma = \begin{pmatrix} \pm p^k & 0 \\ 0 & \pm p^l \end{pmatrix}, \quad k + l = m.$$

Since $\alpha/\beta = \pm p^{k-m}$ is bounded in absolute value, for all but a finite number of p only

$$\gamma \begin{pmatrix} \pm p^{m/2} & 0 \\ 0 & \pm p^{m/2} \end{pmatrix}$$

contribute. Since γ is not central, the signs must be different.

In the new form of (v), $\theta(s, f_v)$ depends upon a nonzero scalar z,

$$\theta_z(s, f_v) = \int\int f_v(k_v^{-1}a_v^{-1}zn_0a_vk_v)|\frac{\alpha_v}{\beta_v}|^{-1-s}da_vdk_v.$$

So the only contribution to (v) will be from $z = \pm p^{m/2}$ and it will only occur for even m.

At finite places q, the operator $R'(\mu_q, \nu_q, s)$ that occurs in (viii) annihilates the vector fixed by $G(\mathbb{Z}_q)$. So, with our assumptions, (viii) reduces to

$$\frac{1}{4\pi}\int_{-i\infty}^{i\infty} \mathrm{tr}(R^{-1}(s)R'(s)\xi_s(f_\infty))\Big(\sum_{k=0}^{m} p^{i(m-2k)s}\Big)d|s|$$

in which R is the local intertwining operator at infinity normalized as in [JL] and in which it is implicit that $\mu_\infty = \nu_\infty = 1$. According to the estimates of [A],

$$|\operatorname{tr}\left(R^{-1}(s)R'(s)\xi_s(f_\infty)\right)| = O(\frac{1}{s^2}), s \to \infty.$$

Thus we can once again apply the Riemann-Lebesgue lemma to conclude that, for m odd, there will be no contribution to the limits (12) or (12′) from (viii).

2.4 The case of even m.[9] For even m, there are several contributions in addition to those from (ii) that survive when we take the limit in X. Since the term p^0 occurs in (35), the expression (vii) contributes

$$\frac{1}{4\pi}\int_{-i\infty}^{i\infty} \frac{m'(s)}{m(s)} \operatorname{tr}(\xi_s(f_\infty))d|s| \tag{36}$$

to (12) or (12′). From (viii) we have

$$\frac{1}{4\pi}\int_{-i\infty}^{i\infty} \operatorname{tr}(R^{-1}(s)R'(s)\xi_s(f_\infty))d|s| \tag{37}$$

To treat (36), or at least part of it, we deform the contour from $\mathrm{Re}(s) = 0$ to $\mathrm{Re}(s) > 0$ or to $\mathrm{Re}(s) < 0$, as the usual estimates permit ([Lan]), expand

$$-\frac{\zeta'(1-s)}{\zeta(1-s)} - \frac{\zeta'(1+s)}{\zeta(1+s)}$$

as

$$\sum_q \sum_{n>0} \frac{\ln q}{q^{n(1-s)}} + \sum_q \sum_{n>0} \frac{\ln q}{q^{n(1+s)}} \tag{38}$$

and integrate term by term, deforming the contours of the individual integrals back to $\mathrm{Re}(s) = 0$. In fact, because of the pole of

$$\frac{\zeta'(1\pm s)}{\zeta(1\pm s)}$$

at $s = 0$, we have first to move the contour to the right and then, for the contribution from $\zeta'(1-s)/\zeta(1-s)$, move it back to the left. The result is that we pick up a supplementary contribution $-\operatorname{tr}(\xi_0(f_\infty))/2$.

[9] There are four sections devoted to even m, by and large to $m = 2$, or to weighted orbital integrals: §2.4, §4.3, and Appendices B and C. They are not used in this paper and are best omitted on a first reading. The formulas for even m are given for almost no other purpose than to make clear that for odd m many significant simplifications occur. Since the formulas are not elegant and are applied neither theoretically nor numerically, I very much fear that errors may have slipped in and advise the reader to be cautious.

Since the character of ξ_s is the function

$$\frac{|\alpha/\beta|^{s/2} + |\beta/\alpha|^{s/2}}{\sqrt{|1-\alpha/\beta||1-\beta/\alpha|}},$$

the calculation that led to (30) shows that

$$\operatorname{tr}(\xi_s(f_\infty)) = 2\int (\lambda^{s/2} + \lambda^{-s/2})\psi_\pm(r)dr \tag{39}$$

where the integral is to be taken over the set of $(\pm 1, r)$ with the interval

$$\{(1,r)\} \mid -1 \le r \le 1\}$$

removed. This may be rewritten as

$$2\int_{-\infty}^{\infty} |t|^s\{|t-t^{-1}|\psi_+(t+1/t) + |t+t^{-1}|\psi_-(t-1/t)\}\frac{dt}{|t|},$$

so that, for s purely imaginary, $\operatorname{tr}(\xi_s(f_\infty))$ is the Fourier transform of

$$2\{|e^x - e^{-x}|\psi_+(e^x + e^{-x}) + |e^x + e^{-x}|\psi_-(e^x - e^{-x})\}. \tag{40}$$

As a result, the contribution of (38) to (36) is

$$\sum_q \sum_{n>0} \frac{\ln q}{q^n}\{|q^n - q^{-n}|\psi_+(q^n + q^{-n}) + |q^n + q^{-n}|\psi_-(q^n - q^{-n})\} - \frac{\operatorname{tr}(\xi_0(f_\infty))}{2}, \tag{41}$$

in which the terms for large q or large n are 0. Since this expression occurs for every p, it remains in the average, as in part a sum of atomic measures that may well be finally cancelled by a contribution from the elliptic term, but it is hard to see at present how this will occur!

Although the local normalization of the intertwining operators to R used in [JL] is necessary if the products and sums appearing in the trace formula are to converge, or at least if the global contribution (vii), which entails no study of local harmonic analysis, is to be clearly separated from the contributions (viii) for which the primary difficulty lies in the local harmonic analysis. Nonetheless it is best that, having separated (38) from (33) to obtain a term that could be analyzed more easily, we combine what remains of (36) with (37) so that we can more readily appeal to known results on weighted orbital integrals.

Since

$$-\frac{1}{2}\frac{\Gamma'((1-s)/2)}{\Gamma((1-s)/2)} - \frac{1}{2}\frac{\Gamma'((1+s)/2)}{\Gamma((1+s)/2)} \tag{42}$$

is the logarithmic derivative of

$$\frac{\pi^{(s-1)/2}\Gamma((1-s)/2)}{\pi^{(s+1)/2}\Gamma((1+s)/2)},$$

the combination of the two terms amounts to multiplying the unnormalized operator

$$J_s : \phi \to J_s\phi, \quad J_s\phi(g) = \int_{\mathbb{R}} \phi(\bar{n}(x)g)dx \tag{43}$$

on the space of the induced representation ξ_s by

$$\frac{\pi^{(s-1)/2}\Gamma((1-s)/2)}{\pi^{s/2}\Gamma(s/2)}. \tag{44}$$

I set

$$n(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \quad \bar{n}(x) = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix}.$$

The choice of measure is irrelevant, because a logarithmic derivative is to be taken. Moreover there is a slight difference between (43) and the intertwining operator of [JL], but this difference too disappears when the logarithmic derivative is taken. So we use (43), which is the definition used in [H].

The logarithmic derivatives of both (44) and J_s now have a pole at $s = 0$, the poles cancelling, so that, when we replace the sum of (37) and the contribution of (42) to (36) by the integrals of the logarithmic derivative of (43) and of (44), the contour of integration has to be deformed whenever we want to discuss them separately. Hoffmann prefers to avoid 0 by skirting it to the right. I follow his convention. So if C is the new contour, we are left with two terms,

$$\frac{1}{4\pi i}\int_C \left\{-\frac{1}{2}\frac{\Gamma'((1-s)/2)}{\Gamma((1-s)/2)} - \frac{1}{2}\frac{\Gamma'(s/2)}{\Gamma(s/2)}\right\} \operatorname{tr}\xi_s(f_\infty)ds \tag{45}$$

and

$$\frac{1}{4\pi i}\int_C \operatorname{tr}(J_s^{-1}J_s'\xi_s(f_\infty))ds. \tag{46}$$

The contribution (46) is not invariant and must be paired with terms from (iv) and (v) to obtain an invariant distribution, the only kind that is useful in our context, for it is the only kind expressible in terms of ψ alone.

The two expressions will be, however, ultimately combined. Indeed, there is a danger in discussing them separately. We need an explicit expression for the sum as

$$\frac{1}{4\pi i}\int_{-i\infty}^{i\infty} \Omega(s) \operatorname{tr}\xi_s(f_\infty)ds.$$

Since $\operatorname{tr}\xi_s(f_\infty)$ is even but otherwise essentially arbitrary, the function (or distribution) Ω will be unique if it is assumed even. The integrands of (36) and (37) are even, so that if we stay with them it is easier to use parity to monitor the manipulations. On the other hand, the factor multiplying the trace in (45) is not even; nor is the integrand of (46). Since Hoffmann's results for (46) are in a form that is not only transparent but also symmetric and since we can easily put (45) in symmetric form, we can readily restore the symmetry, the only cost being the replacement of (45) by a somewhat lengthier expression, in which there is one surprise, the final term in the following formula. If we avoid 0 by a small semi-circle of radius ϵ then (45) becomes, up to a term of order $O(\epsilon)$

$$\frac{1}{4\pi i}\int_{-i\infty}^{-i\epsilon}+\int_{i\epsilon}^{i\infty}\left\{-\frac{1}{2}\frac{\Gamma'((1-s)/2)}{\Gamma((1-s)/2)}-\frac{1}{2}\frac{\Gamma'(s/2)}{\Gamma(s/2)}\right\}\operatorname{tr}\xi_s(f_\infty)ds+\frac{\operatorname{tr}\xi_0(f_\infty)}{4}.$$

The first factor in the integrand may be symmetrized, so that the singularity $1/s$ at $s=0$ disappears, and then ϵ allowed to go to 0. The result is the sum of

$$\frac{1}{16\pi i}\int_{-i\infty}^{i\infty}\left\{-\frac{\Gamma'((1-s)/2)}{\Gamma((1-s)/2)}-\frac{\Gamma'(s/2)}{\Gamma(s/2)}-\frac{\Gamma'((1+s)/2)}{\Gamma((1+s)/2)}-\frac{\Gamma'(-s/2)}{\Gamma(-s/2)}\right\}\operatorname{tr}\xi_s(f_\infty)ds,$$

or better, since

$$\frac{\Gamma'(s/2)}{\Gamma(s/2)}=\frac{\Gamma'(1+s/2)}{\Gamma(1+s/2)}-\frac{2}{s},$$

of

$$\frac{-1}{16\pi i}\int\left\{\frac{\Gamma'((1-s)/2)}{\Gamma((1-s)/2)}+\frac{\Gamma'(1+s/2)}{\Gamma(1+s/2)}+\frac{\Gamma'((1+s)/2)}{\Gamma((1+s)/2)}+\frac{\Gamma'(1-s/2)}{\Gamma(1-s/2)}\right\}\operatorname{tr}\xi_s(f_\infty) \tag{47}$$

and

$$\frac{\operatorname{tr}\xi_0(f_\infty)}{4} \tag{48}$$

When considering (iv) and (v), we suppose that $p\neq 2$ since, as we observed, the values for a particular p have no influence on the limit. We may also suppose that $\gamma=\pm p^{m/2}\delta$, where δ is a matrix with eigenvalues ± 1. The signs are equal for (v) and different for (iv).

We begin with[10] (iv), inverting the order of summation and discarding all terms that do not contribute to the average, so that it becomes a sum over just two γ followed by a sum over the places of $\mathbb{Q}$. If v is a finite place q different from p, then

$$n(x)^{-1}\gamma n(x)=\pm p^{m/2}\begin{pmatrix}1 & 2x\\ 0 & -1\end{pmatrix}$$

[10] There appears to be a factor of $1/2$ missing in (iv). It was lost on passing from p. 530 to p. 531 of [JL], but is included below.

is integral in $\mathbb{Q}_v$ if and only if $2x$ is integral. Consequently, $\omega_1(\gamma, f_q)$, $q \neq p$, is 0 except for $q = 2$, but for $q = 2$,

$$\omega_1(\gamma, f_2) = -\frac{\ln(2^2)}{2} = -\ln 2$$

Moreover,

$$\omega(\gamma, f_q) = \begin{cases} 1, & q \neq 2, p; \\ 2, & q = 2 \end{cases}.$$

On the other hand,

$$\omega(\gamma, f_p) = 1,$$

if $p \neq 2$. Finally

$$\omega_1(\gamma, f_p) = -\frac{\int_{1<|x| \le p^{m/2}} \ln |x|^2 dx}{p^{m/2}} = -(1 - \frac{1}{p}) \frac{\ln p}{p^{m/2}} \sum_{j=1}^{m/2} 2jp^j$$

The integral is taken in $\mathbb{Q}_p$.

Thus the sum over v in (iv) reduces to three terms, those for $v = \infty$, $v = 2$ and $v = p$. Since our emphasis is on f_∞, the only variable part of f, the first plays a different role than the last two. It is invariant only in combination with (46). The first two are already invariant as functions of f_∞.

Before continuing, we give the values of the three constants c, λ_0 and λ_{-1} appearing in the trace formula as given in [JL]. First of all, $\lambda_{-1} = 1$ and λ_0 is Euler's constant. The constant c is the ratio between two measures, the numerator being the measure introduced in §2.1 and used to define the operators

$$R(f) = \int_{Z_+ \backslash G(\mathbb{A})} f(g) dg$$

appearing in the trace formula[11] and the denominator being that given locally and globally as $d(ank) = dadndk$, $g = ank$ being the Iwasawa decomposition. Thus both measures are product measures, so that $c = \prod c_v$. If we choose, as we implicitly do, the measures da and dn so that $A(\mathbb{Z}_q)$ and $N(\mathbb{Z}_q)$ have measure 1 for all q, then $c_q = 1$ at all finite places. On the other hand, we have not been explicit about the measure on $Z_+ \backslash G(\mathbb{R})$. There was no need for it. We may as well suppose that it is taken to be $dadndk$, where a now belongs to $Z_+ \backslash A(\mathbb{R})$. Then $c = c_\infty = 1$.

[11] So the symbol R has two different roles. It is not the only symbol of the paper whose meaning depends on the context.

The measure on $Z_+\backslash A(\mathbb{R})$ has already been fixed, but the choice of measures on N_∞ and K_∞ do not enter the formulas explicitly. We have

$$\omega(\gamma, f_\infty) = \psi(\mathrm{ch}(\gamma)).$$

Thus the contribution from (iv) is the sum of two terms. The first

$$-\sum_\gamma \omega_1(\gamma, f_\infty) \prod_{q\neq\infty} \omega(\gamma, f_q), \tag{49}$$

in which only two γ appear,

$$\gamma = \pm p^{m/2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

is to be combined with (46). Since $\omega(\gamma, f_q)$ is just the orbital integral of f_q, it is calculated by Lemma 1 and, as $|\gamma_1 - \gamma_2| = |2p^{m/2}|$ for the γ in question,

$$\prod_{q\neq\infty} \omega(\gamma, f_q) = \frac{1}{p^{m/2}} 2p^{m/2} = 2. \tag{50}$$

The second stands alone and is

$$\psi_-(0)\left\{\ln 2 + \left(1 - \frac{1}{p}\right) \frac{\ln p}{p^{m/2}} \sum_{j=1}^{m/2} 2jp^j\right\}, \tag{51}$$

an expression that is about $\ln p$ in size. Its occurrence is certainly unexpected, as it is not bounded in p, so that the elliptic term will have to contain something that compensates for it. The source of this atomic contribution to the elliptic term – if it is present – should not be hard to find, but I have not yet searched for it.

We verify immediately that

$$\theta_z(0, f_v) = \int_{\mathbb{Q}_v} \int_{K_v} f_v(\pm p^{m/2} k_v^{-1} n(x) k_v) dk_v dx, \quad z = \pm p^{m/2},$$

if $v = q$ is finite. If $q \neq p$, this is equal to 1. If $q = p$, it is equal to 1 because $p^{m/2}x$ is integral for $|x| \leq p^{m/2}$. Since (the notation is that of [JL], p. 194)

$$L(1) = \pi^{-1/2}\Gamma(\frac{1}{2}) = 1$$

and f_∞ is positively homogeneous, the first term of (v) contributes

$$\sum_\pm \lambda_0 \theta_{\pm 1}(0, f_\infty) = \lambda_0 \sum \int_{\mathbb{R}} \int_{K_\infty} f_\infty(\pm k^{-1} n(x) k) dk dx. \tag{52}$$

to the average.

The expression (52) can be calculated easily in terms of ψ_+. We take

$$\gamma = \pm \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix},$$

and let t approach 0. Then $\mathrm{ch}(\gamma) = (N, r) = (4, \pm 2\cosh t)$, $r/\sqrt{|N|}$ approaches ± 1, and, as a simple change of variables shows,

$$2\left(\sqrt{\frac{r^2}{N} - 1}\right)\psi_+(r) = 2|\sinh t| \int_{\mathbb{R}} \int_{K_\infty} f_\infty(k^{-1}\gamma n((1-e^{-2t})x)k)dxdk \tag{53}$$

approaches the integral of (52). According to (26), the limit of (53) is $\psi''_\infty(4, \pm 2)$. It is nonzero only when ψ_+ is singular at 1 or -1.

The derivative

$$\theta'_z(0, f_v) = -\frac{\ln q}{q-1} \int f_v(k^{-1}zn(x)k)dxdk + \int f_v(k^{-1}zn(x)k)\ln|x|dxdk \tag{54}$$

if $v = q$ is nonarchimedean. If it is archimedean, then

$$\theta'_z(0, f_v) = \kappa \int f_\infty(k^{-1}zn(x)k)dxdk + \int f_\infty(k^{-1}zn(x)k)\ln|x|dxdk, \tag{55}$$

where

$$\kappa = -\frac{\pi^{-1/2}\Gamma(1/2)}{2}\ln \pi + \pi^{-1/2}\Gamma'(1/2) = -\frac{\lambda_0}{2} - \frac{\ln \pi}{2} - \ln 2,$$

a result of

$$\Gamma'(\frac{1}{2}) = (-\lambda_0 - 2\ln 2)\sqrt{\pi}, \quad \text{(cf. [N], p. 15).}$$

The expression (54) is deceptive. If $v \neq p$ and $z = \pm p^{m/2}$, then $\theta_z(s, f_q)$ is identically 1 and its derivative 0. If $q = p$, then

$$\theta_z(s, f_q) = \frac{1}{p^{m/2}L_q(1+s, 1)} \int_{|\beta| \le p^{m/2}} |\beta|^{1+s}\frac{d\beta}{|\beta|} = p^{ms/2},$$

so that

$$\theta'_z(0, f_p) = \frac{m \ln p}{2}.$$

The sum in (v) is a double sum, over $\gamma = \pm p^{m/2}$ and over v. Only $v = \infty$ and $v = p$ yield a contribution different from 0. The first will be combined with (52) to give the sum of

$$\kappa_1 \sum \int_{\mathbb{R}} \int_{K_\infty} f_\infty(\pm k^{-1}n(x)k)dx, \quad \kappa_1 = \frac{\lambda_0}{2} - \frac{\ln \pi}{2} - \ln 2 \tag{56}$$

and the noninvariant expression

$$\sum \int f_\infty(k^{-1}zn(x)k)\ln|x|dxdk, \tag{57}$$

which will have to be combined with (46). The second is

$$\sum \frac{m\ln p}{2}\int_{\mathbb{R}}\int_{K_\infty} f_\infty(\pm k^{-1}n(x)k)dx, \tag{58}$$

in which two disagreeable features appear: the logarithm of p which cannot possibly have an average and the integral that is expressible only in terms of the singularities of ψ_+ at ± 1. So there is no question of the logarithmic terms in (51) and (58) cancelling.

2.5 The elliptic term. The sum in the expression (ii) from [JL] is over the global regular elliptic elements γ, each γ being determined by its trace, which we have denoted r and by 4 times its determinant, $N = 4\det(\gamma)$. Only γ for which r is integral and $N = \pm 4p^m$ appear. The eigenvalues of γ are

$$\frac{r}{2} \pm \frac{\sqrt{r^2-N}}{2}.$$

Their difference is $\pm\sqrt{r^2-N}$. Thus γ will be elliptic if and only if r^2-N is not a square. We write[12] $r^2-N = s^2D$, where D is a fundamental discriminant, thus $D \equiv 0,1 \pmod 4$. Both D and s are understood to be functions of r and N. If $r^2 = N$, then D is taken to be 0; if it is a square then $D = 1$.

I claim that

$$\sum_{f|s} f \prod_{q|f}\left(1-\frac{\left(\frac{D}{q}\right)}{q}\right), \tag{59}$$

in which $\left(\frac{D}{p}\right)$ is the Kronecker symbol ([C]), is equal to the product of $U^m(\gamma)$ taken at p with the product over $q \neq p$ of $U^1(\gamma)$. By multiplicativity, it is enough to consider

$$1+\sum_{j=1}^{k} q^j\left(1-\frac{\left(\frac{D}{q}\right)}{q}\right)$$

for each prime q. If $\left(\frac{D}{q}\right) = 1$, this is $1+q^k-1$, but if $\left(\frac{D}{q}\right) = -1$, it is

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

[12] So the symbol s appears in the paper in two quite different ways: here and elsewhere, as an integer by whose square we divide to obtain the fundamental discriminant; previously and also below, as a variable parametrizing the characters of $\mathbb{R}^+$.

Finally, if $\left(\frac{D}{q}\right) = 0$, it is

$$\sum_{j=0}^{k} q^j = \frac{q^{k+1}}{q-1} - \frac{1}{q-1}.$$

So we have only to appeal to Lemma 1.

Let μ_D be the volume $\mu(Z_+G_\gamma(\mathbb{Q})\backslash G_\gamma(\mathbb{A}))$ if $D \neq 1$. Then the uncorrected elliptic term is the sum

$$\sum_{N=\pm 4p^m} \sum_r \mu_D \frac{\psi(N,r)}{p^{m/2}} \sum_{f|s} f \prod_{q|f} \left(1 - \frac{\left(\frac{D}{q}\right)}{q}\right), \tag{60}$$

in which the function ψ continues to be defined as in (26). The factor $1/2$ in (ii) has been removed because each r accounts for two γ. Because of the presence of the term $\psi(N,r)$, the sum is finite, the number of terms being of order $\sqrt{|N|}$. The terms with $D = 0, 1$ are excluded because they do not correspond to regular elliptic γ. Moreover p is fixed for the moment.

We now make use of formulas from [C] (§5.3.3 and §5.6.2 – the general form of the second formula is stated incorrectly but we do not need the general form) for μ_D. If n is a positive integer and x a real number, define the function $\varphi(x,n)$ by the following formulas.

$x < 0$:

$$\varphi(x,n) = \pi \operatorname{erfc}\left(\frac{n\sqrt{\pi}}{\sqrt{|x|}}\right) + \frac{\sqrt{|x|}}{n} \exp(-\pi n^2/|x|).$$

$x > 0$:

$$\varphi(x,n) = \frac{\sqrt{x}}{n} \operatorname{erfc}\left(\frac{n\sqrt{\pi}}{\sqrt{|x|}}\right) + \mathrm{E}_1\left(\frac{\pi n^2}{x}\right),$$

where E_1 is defined to be the function

$$-\gamma - \ln(x) + \sum_{k\,\geq\,1} (-1)^{k-1} \frac{x^k}{k!k},$$

γ being Euler's constant. Then, on making use of the formulas in §2.1 for μ_D in terms of the class number, we obtain

$$\mu_D = \sum_{n=1}^{\infty} \left(\frac{D}{n}\right) \varphi(D,n). \tag{61}$$

The series (61) is absolutely convergent and we substitute it in (60) to obtain

$$2\sum_n \sum_f \sum_{\{(r,N)\,|\,f|s\}} f \left(\frac{D}{n}\right) \varphi(D,n) \frac{\psi(N,r)}{\sqrt{|N|}} \prod_{q|f} \left(1 - \frac{\left(\frac{D}{q}\right)}{q}\right). \tag{62}$$

Recall that N assumes only two values $\pm 4p^m$, but that r runs over all integers except the very few for which $D = 0, 1$. By homogeneity, we may replace $\psi(N, r)$ by $\psi_\pm(x_r)$, where for brevity of notation I set $x_r = r/\sqrt{|N|}$. I rewrite (62) as

$$(63) \qquad 2\sum_f \sum_{\{n\,|\,(n,f)=1\}} \sum_{\{(r,N)\,|\,f|s\}} \sum_{f'} f \left(\frac{D}{nf'}\right) \varphi(D, nf') \frac{\psi_\pm(x_r)}{\sqrt{|N|}} \prod_{q|f} \left(1 - \frac{\left(\frac{D}{q}\right)}{q}\right).$$

The sum over f' is over all positive numbers all of whose prime divisors are prime divisors of f. The sum over (r, N) is over those pairs for which $f|s$, s continuing to be defined by $r^2 - N = s^2 D$. In principle, we want to examine the individual terms

$$(64) \qquad 2\sum_{f|s} \sum_{f'} f \left(\frac{D}{nf'}\right) \varphi(D, nf') \frac{\psi_\pm(x_r)}{\sqrt{|N|}} \prod_{q|f} \left(1 - \frac{\left(\frac{D}{q}\right)}{q}\right),$$

the outer sum being a sum over r and the two possible N, but we must first subtract the contribution (28) from the trivial representation. So we have to express it too as a sum over n and f.

The contribution from the trivial representation θ is the product of (28) with

$$\operatorname{tr}\theta(f_p^m) = \sum_{k=0}^{m} p^{(m-2k)/2} = p^{m/2}\frac{1 - p^{-m}}{1 - p^{-1}} = \frac{\sqrt{|N|}}{2}\frac{1 - p^{-m}}{1 - p^{-1}}$$

So it is

$$(65) \qquad 2\sqrt{|N|}\frac{1 - p^{-m}}{1 - p^{-1}} \sum \int \psi_\pm(x)\sqrt{|x^2 \mp 1|}dx,$$

the sum being over the set $\{+, -\}$. To see how this is to be expressed as a sum over n and f, we observe that $\varphi(D, n)$ behaves for large $|D|$ like $\sqrt{|D|}/n = \sqrt{|r^2 - N|}/sn$, so that, for a rough analysis, (63) may be replaced by

$$(66) \qquad 2\sqrt{|N|}\sum_f \sum_{\{n\,|\,(n,f)=1\}} \sum_{f|s} \sum_{f'} \frac{f}{snf'} \left(\frac{D}{nf'}\right) \psi_\pm(x_r) \frac{\sqrt{|x_r^2 \mp 1|}}{\sqrt{|N|}} \prod_{q|f} \left(1 - \frac{\left(\frac{D}{q}\right)}{q}\right).$$

Suppose we replace each of the factors

$$(67) \qquad \sum_{f'} \frac{f}{snf'} \left(\frac{D}{nf'}\right) \prod_{q|f} \left(1 - \frac{\left(\frac{D}{q}\right)}{q}\right) = \frac{f}{sn}\left(\frac{D}{n}\right)$$

by a number $\epsilon_{n,f}(N)$, an approximation to its average value on intervals long with respect to n but short with respect to $\sqrt{|N|}$. Then (63) is replaced by

$$2\sqrt{|N|}\sum_f \sum_{\{n\,|\,(n,f)=1\}} \sum_{\pm} \epsilon_{n,f}(N)\int \psi_\pm(x)\sqrt{|x^2\mp 1|}dx. \tag{68}$$

The inner sum is over the two possible values of N.

The exact sense in which $\epsilon_{n,f}(N)$ is an approximation to the average is not important, provided the choice works, but we do need to show that

$$\begin{aligned}\sum_{n,f}\epsilon_{n,f}(N) &= \frac{1}{1-p^{-1}}+O(|N|^{-1}),\\ &= \frac{1-p^{-m-1}}{1-p^{-1}}+O(|N|^{-1}),\end{aligned} \tag{69}$$

so that (65) is equal to (68) and the difference between (63) and (65) has some chance of being $o(|N|^{1/2})$. For the purposes of further examination, we write this difference as the sum over n and f, $\gcd(n,f)=1$, of[13]

$$2\{\sum f\left(\frac{D}{nf'}\right)\varphi(D,nf')\frac{\psi_\pm(x_r)}{\sqrt{|N|}}\Phi-\sqrt{|N|}\sum_\pm \epsilon_{n,f}(N)\int\psi_\pm(x)\sqrt{|x^2\mp 1|}dx\}, \tag{70}$$

with

$$\Phi=\Phi_f=\prod_{q|f}\left(1-\frac{\left(\frac{D}{q}\right)}{q}\right).$$

The first sum in (70) is over r, f' and $\pm$.

I now explain how we choose $\epsilon_{n,f}(N)$. Let t be an integer large with respect to $\sqrt{|N|}$ and divisible by a multiplicatively very large square. The average of (68) is to be first calculated on $[0,t)$. The divisibility of r^2-N by s^2 is then decided for all s up to a certain point by the residue of r modulo t. Whether $(r^2-N)/s^2$ is then divisible by further squares is not, but it is except for squares that are only divisible by very large primes. There will be very few r for which this occurs. Otherwise f divides s if and only if f^2 divides r^2-N with a remainder congruent to 0 or 1 modulo 4. Then, for $(n,f)=1$,

$$\left(\frac{(r^2-N)f^2/s^2}{n}\right)=\left(\frac{D}{n}\right).$$

[13] Notice that the sum (70) has a simpler mathematical structure than (60), especially for $f=1$ for then the sum over f' is absent. The only element that varies irregularly with r is the square s^2 dividing r^2-N.

Thus $\epsilon_{n,f}(N)$ will be an approximation to the average value of

$$\frac{f}{sn}\left(\frac{(r^2-N)f^2/s^2}{n}\right). \tag{71}$$

If $\gcd(n,f)$ were not 1, these expressions would be 0 and it is useful to set $\epsilon_{n,f}(N)=0$ if $\gcd(n,f)\neq 1$. The calculation of these factors is long and tedious, but their values are needed for the numerical experiments, and (69) is a confirmation of the correctness of the calculation. So I present the calculation in an appendix.

Part III: Numerical experiments

3.1 A first test. We observed that (32) was, apart from the elliptic term, the only nonzero contribution to the limit. Since $L(s,\pi,\sigma_1)$ is regular and nonzero on $\mathrm{Re}(s)=1$ for all cuspidal automorphic π, we expect that the limit (12′) is 0. For $m=1$, we have calculated explicitly all contributions to the limit (12′) except for the difference between the elliptic term (60) and the contribution (65) from the one-dimensional representations. So we have to show that this difference, or rather its average in the sense of (12′) over $p<X$, cancels the simple, but in our context fundamental, distribution (32). A first test is numeric.

Both (60) and (65) are distributions, even measures, on the pair (ψ_+,ψ_-). The first is a sum of atomic measures. The second is absolutely continuous with respect to Lebesgue measure. So their difference and the average over p is also a measure, symmetric with respect to $r\to -r$. I divide the interval from -3 to 3 on each of the lines $N=\pm 1$ into 60 equal parts of length .1 and calculate numerically for each p the measure of each interval. In the unlikely event that a point common to two intervals has nonzero mass, I assign half of this mass to each of the two intervals. Then I average over the first n primes in the sense of (12′). The result should be approaching $-.2$ on each of the intervals except those on $N=+1$ between -1 and 1, where it should approach 0. From Table 3.1 at the end of the paper in which the first two columns refer to the average over the first 200 primes, the second to that over the first 3600 primes, and the third to that over the first 9400, we see that the average is almost immediately approximately correct at least for the intervals closer to 0, that it does seem to converge to the correct values, but that the convergence is slow, sometimes even doubtfully slow.

Thanks to the symmetry, only the results for the intervals from -3 to 0 need be given. In each set of two columns, the numbers in the first column are for intervals of r with $N=-1$, and those in the second for $N = 1$. Once the results get within about .007 of the expected values they cease to improve. I assume they would with better programming.

3.2 A rough estimate. For m odd and in particular for $m = 1$, the elliptic term, or more precisely the difference between the elliptic term and the contribution from the one-dimensional representations, is a formidable expression, with which it is difficult, probably very difficult, to deal. The limit of the average is nevertheless expected to exist and is, moreover, expected to be, even if S contains finite places, a linear combination of the distributions (30), of which there are three, because $\epsilon_{--} = \epsilon_{-+}$. The coefficients will depend on the functions f_q, $q \neq \infty$, $q \in S$, thus on congruence conditions. So there is a great deal of uniformity present in the limit, and it is fair to assume that it will influence the structure of the proofs.

On the other hand, the average of the difference, with the elliptic term expressed as in (62) and (65), decomposed with the help of (69), will be a sum over three parameters, r, p and n. More precisely, the last sum is over n and f, but the additional sum over f may be little more than an unfortunate complication, whose implications are limited, of the sum over n. At the moment, I am not concerned with it. There are also sums over $\pm$ and f' that occur simultaneously with the sum over r and are understood to be part of it.

The sum over r has a simple structure, except for the dependence on s. The use of the logarithmic derivatives that leads us to an average over p with the factor $\ln p$ is alarming as any incautious move puts us dangerously close to the mathematics of the Riemann hypothesis, but there is nothing to be done about it. The structure suggested by functoriality and the L-group imposes the use of the logarithmic derivative on us, and any attempt to avoid it for specious (in the sense of MacAulay[14])

[14] Although in the middle of the nineteenth century the word did not yet have its present thoroughly pejorative sense, it did evoke doubt; so a few words of explanation are in order. The only tool presently in sight for passing from π to ${}^{\lambda}H$ are the functions m_π and m_H, which are linear in ρ. We are already familiar, thanks to basic results for the Artin L-functions, with the importance of the linearity in ρ of the order of the pole of the L-functions at $s = 1$. The linearity in ρ is naturally accompanied by a linearity in π. The functions m_π not only incorporate the structural advantages suggested by functoriality that will be of great importance when we pass to groups of large dimension but also are fully adapted to the trace formula, provided we take them as defined by the logarithm of the L-function $L(s, \pi, \rho)$ or its derivative. On the other hand, we have none of the necessary analytic experience. We are faced with sums and limits in which we do not know what is large, what is small, what converges, what does not, and we desperately need insight. If some experience and some feeling for the analysis can be acquired by a modification of the problem in special cases in which the sums over primes that appear in the logarithmic derivative are replaced by easier sums over integers, then common sense suggests that we start there.

technical advantages is likely to lead us away from our goal, not toward it. We do want to discover something about the behaviour of automorphic L-functions near $s = 1$ but if we are careful we should not otherwise find ourselves inside the critical strip.

By passing from (60) to (62) we remove the class number, an almost intractable factor, but at the cost of the additional sum over n and f. Although the contribution from the one-dimensional representations is not at first expressed as a sum over n and f, we observed in §2.5 that there was a natural way so to express it, so that the difference becomes the sum over n and f of (70).

If it turned out that for each n and f, the sum over r and p behaved well, then we would, it seems to me, have a much better chance of dealing with the elliptic term. More precisely, it would be a real windfall if the average of (70) approached a limit for each n and f and if the sum over n and f followed by the average could be replaced by the average followed by the sum. The most important observation of this paper is that preliminary numerical investigations suggest that the average of (70) does indeed have a regular behavior, but there are no windfalls. Since my experience as a programmer is limited and mistakes are easy to make, either outright blunders or a careless analysis of possible systematic errors in what are necessarily approximate calculations, I very much hope that others will find the results sufficiently curious to be worthy of their attention. Not only should my conclusions be examined again and more extensively, but, apart from any theoretical efforts, higher m, especially $m = 2, 3$, need to be considered as does the effect of congruence conditions or of characters of the Galois group.

Interchanging the order of summation and the passage to the limit is another matter. In the summation there are three ranges: n substantially smaller than $\sqrt{|N|}$; n about equal to $\sqrt{|N|}$; n substantially larger than $\sqrt{|N|}$. We can expect that the interchange picks out the first range. The function $\varphi(x, n)$ is such that we can expect the last range to contribute nothing. This leaves the intermediate range, which may very well contribute but about which nothing is said in this paper, whose tentative explorations, instructive though they are, stop short of all difficult analytic problems.[15]

[15] Since $\varphi(x, n)$ is a function of $n/\sqrt{x}$, the factor $\varphi(D, n)$ in (62) can almost be treated as a constant when $n \sim \sqrt{N}$. Thus, in so far as D is just $r^2 - N$, the pertinent expression in the intermediate range is pretty much

$$(F.1) \qquad \sum_{-cn \le r \le cn} \left(\frac{r^2 - N}{n}\right) \psi_{\pm}\left(\frac{r}{n}\right).$$

More extensive investigations, which I have not yet undertaken, would examine, at least numerically but also theoretically if this is possible, the sum over the intermediate range in this light as well as the validity of the separation into three ranges. Is it possible to hope that the average over $p < X$ of (F.1) will have features like those described in §3.3 for the average over the first range? Can the separation be

Although we persuaded ourselves that (70) might very well be $o(|N|^{1/2})$, so that it is smaller than the two expressions of which it is a difference, we made no effort to see what size it might be. Its average over p is intended to have a limit, thus, in particular, to be $O(1)$, but that does not prevent violent oscillations in the individual terms. Besides the existence of a limit may be too much to expect. The numerical results described later in this section suggest that (70) is $O(\ln^2|N|)$, but I have not yet even been able to show that it is $O(\ln^c|N|)$ for some exponent c. Before coming to the experiments, I describe briefly the difficulties that I met in trying to estimate (70) directly. I have not yet made a serious attempt to overcome them.

Recall first that $\psi_\pm$ in (70) are zero outside some interval $[c_1, c_2]$, so that r need be summed only over $c_1\sqrt{|N|} \leq r \leq c_2\sqrt{|N|}$. To simplify the – in any case rough – analysis, I suppose that both $\psi_\pm$ are bounded; thus I ignore the possible singularity of ψ_+ at $r = \pm 1$. Observe that, for large N,

$$\varphi(x,n) = \frac{\sqrt{|x|}}{n} + A + B\ln|x| + O(|x|^{-1/2}),$$

where A and B are well-determined constants that depend only on the sign of x. The constant implicit in the error term depends on n. Since

$$\sum_{c_1\sqrt{|N|} \leq r \leq c_2\sqrt{|N|}} \frac{1}{\sqrt{|N|}} = O(1),$$

we can replace $\varphi(x,n)$ by $\sqrt{|x|}/n$ at a cost that is $O(\ln|N|)$, a price that we are willing to pay.

Thus, at that level of precision, we can make the same modifications as led from (63) to (66) and replace (70) by twice the sum over $\pm$ of the difference

$$\sum_r \frac{f}{sn}\left(\frac{D}{n}\right)\psi_\pm(x_r)\sqrt{|x_r^2 \mp 1|\Phi} - \sqrt{|N|}\epsilon_{n,f}(N)\int \psi_\pm(x)\sqrt{|x^2\mp 1|}dx. \tag{72}$$

In (72) there is no longer a sum over f' and no need to sum over $\pm$, as we can simply fix the sign.

To simplify further, I take $n = 1$. In so far as there is any real argument in the following discussion, it can easily be extended to an arbitrary n. This is just a matter of imposing further congruence conditions on r modulo primes dividing $2n$. For similar reasons, I also take $f = 1$. Then (72) becomes

$$\sum_r \psi_\pm(x_r)\frac{\sqrt{|D|}}{\sqrt{|N|}} - \sqrt{|N|}\epsilon_{1,1}(N)\int \psi_\pm(x)\sqrt{x^2\mp 1}dx.$$

made cleanly so that any contributions from intermediate domains on the marches of the three ranges are small?

It might be better to take $\psi_\pm$ to be the characteristic function of an interval and to attack this fairly simple expression directly. I tried a different approach.

I indicate explicitly the dependence of s on r by setting $s = s_r$ and then write the first term of the difference, with n now equal to 1, as

$$\sum_s \frac{1}{s} \sum_{s_r = s} \psi_\pm(x_r)\sqrt{|x_r^2 \mp 1|}. \tag{73}$$

We then compare

$$\sum_{s_r = s} \psi_\pm(x_r)\sqrt{|x_r^2 \mp 1|} \tag{74}$$

with

$$\frac{\sqrt{|N|}}{s^2}\int_{-\infty}^{\infty} \psi_\pm(x)\sqrt{|x^2 \mp 1|}g_s(x)dx, \tag{75}$$

where $g_s(x)$ is constant on each interval $[ks^2/\sqrt{|N|}, (k+1)s^2/\sqrt{|N|})$ and equal to the number $C_s(k)$ of integral points r in $[ks^2, (k+1)s^2)$ such that $r^2 - N$ divided by s^2 is a fundamental discriminant. The sum (73) is compared with

$$\sum_s \frac{\sqrt{|N|}}{s^3}\int_{-\infty}^{\infty} \psi_\pm(x)\sqrt{|x^2 \mp 1|}g_s(x)dx \tag{76}$$

We have to compare (76) not only with (73) but also with the second term of (72), which is, for $n = f = 1$,

$$\sqrt{|N|}\epsilon_{1,1}(N)\int \psi_\pm(x)\sqrt{|x^2 \mp 1|}dx. \tag{77}$$

I truncate both (73) and (76) at $s \le M = |N|^{1/4}$. An integer r contributes to the number $C_s(k)$ only if s^2 divides $r^2 - N$. This already fixes r up to a number of possibilities modulo s^2 bounded by $2^{\#(s)}$, where $\#(s)$ is the number of prime divisors of s. Thus the truncation of (76) leads to an error whose order is no larger than

$$\sum_{s > M} \frac{\sqrt{|N|}2^{\#(s)}}{s^3} = O(\ln^{2c-1}|N|),$$

as in Lemma B.1 of Appendix B. As observed there, the constant c may very well be 1.

For $s > M$, the number of r in $[c_1\sqrt{|N|}, c_2\sqrt{|N|})$ such that s^2 divides $r^2 - N$ is $O(2^{\#(s)})$, because $\sqrt{|N|}/s^2$ is bounded by 1. Thus, according to Lemma B.3, the error entailed by the truncation of (73) is of order no worse than

$$\sum_{C\sqrt{|N|} > s > M} \frac{2^{\#(s)}}{s} = O(\ln^{2c}|N|).$$

To estimate the difference between (74) and (75), we regard

$$\psi_\pm(x_r)\sqrt{|x_r^2 \mp 1|} = \frac{\sqrt{|N|}}{s^2}\psi_\pm(x_r)\sqrt{|x_r^2 \mp 1|}\frac{s^2}{\sqrt{|N|}},$$

$ks^2 \le r < (k+1)s^2$, as an approximation to

$$\frac{\sqrt{|N|}}{s^2}\int_{\frac{ks^2}{\sqrt{|N|}}}^{\frac{(k+1)s^2}{\sqrt{|N|}}} \psi_\pm(x)\sqrt{|x^2 \mp 1|}dx.$$

Difficulties around $x = \pm 1$ – where ψ_+ may not be bounded, much less smooth – aside, the approximation will be good to within

$$\frac{\sqrt{|N|}}{s^2}O\left(\left(\frac{s^2}{\sqrt{|N|}}\right)^2\right) = O\left(\frac{s^2}{\sqrt{|N|}}\right).$$

Multiplying by $1/s$ and summing up to M, we obtain as an estimate for the truncated difference between (73) and (76)

$$\frac{1}{\sqrt{|N|}}O\Big(\sum_{s \le M} 2^{\#(s)}s\Big),$$

which is estimated according to Corollary B.4 as $O(\ln^{2c}|N|)$.

For a given N and a given natural number s, the condition that $(r^2 - N)/s^2$ be integral but divisible by the square of no odd prime dividing s and that $(r^2 - N)/4$ have some specified residue modulo 4, or any given higher power of 2 is a condition on r modulo $4s^4$ or some multiple of this by a power of 2, so that it makes sense to speak of the average number $\alpha'(N, s)$ of such r. The number $\alpha'(N, s)$ is $O(2^{\#(s)}/s^2)$. If q is odd and prime to s then the average number of r for which, in addition, $r^2 - N$ is not divisible by q^2 is

$$1 - \frac{1 + \left(\frac{N}{q}\right)}{q^2}.$$

Thus the average number of r for which $r^2 - N$ divided by s^2 is a fundamental discriminant can be defined as

$$\alpha(N, s) = \alpha'(N, s)\prod_{\gcd(q,2s)=1}\left(1 - \frac{1 + \left(\frac{N}{q}\right)}{q^2}\right).$$

As in the first appendix,

$$\epsilon_{1,1}(N) = \sum_{s=1}^{\infty}\frac{\alpha(N, s)}{s}.$$

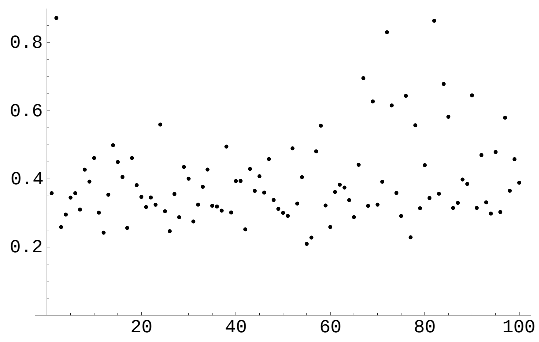

Diagram 3.2.A

It remains to compare (75) with

$$\sqrt{|N|}\alpha(N,s)\int_{-\infty}^{\infty}\psi_\pm(x)\sqrt{|x^2\mp 1|}dx \tag{78}$$

remembering that their difference is to be divided by s and then summed over s, although by Lemma B.1, the sum can be truncated at $s \leq M$. I had difficulties with the estimates that I have not yet been able to overcome. I describe them.

Let $\bar{g}_s$ be the average of g_s on some interval $[-C,C]$ large enough to contain in its interior the support of $\psi_\pm$. The difference between (75) and (78) divided by s is the sum of two terms. First of all,

$$\frac{\sqrt{|N|}}{s^3}\int_{-C}^{C}\psi_\pm(x)\sqrt{|x^2\mp 1|}(g_s(x)-\bar{g}_s(x))dx; \tag{79}$$

and secondly,

$$\frac{\sqrt{|N|}}{s^3}\int_{-C}^{C}\psi_\pm(x)\sqrt{|x^2\mp 1|}(\bar{g}_s(x)-s^2\alpha(N,s))dx. \tag{80}$$

The first should be smallest when $\psi_\pm$ is very flat; the second when its mean is 0. So it appears they are to be estimated separately.

First of all, to calculate $C_s(k)$ and thus g_s, we have to examine the $O(2^{\#(s)})$ integers r in the pertinent interval such that s^2 divides r^2-N. For simplicity, rather than work with g_s and $\bar{g}_s$, I work with the contributions to $C_s(k)$ from a single residue class $\bar{r}$ modulo s^2, but without changing the notation. As a result, the estimates obtained will have to be multiplied by the familiar factor $2^{\#(s)}$. Moreover, the definition of $\alpha(N,s)$ will have to be modified according to the same principle.

If r lies in $[-C\sqrt{|N|}, C\sqrt{|N|}]$ and has residue $\bar{r}$, then we attach to r the set $p_1,p_2,\ldots,p_l$ such that $s^2p_i^2$ divides r^2-N and is congruent to 0 or 1 modulo 4. Then $s(r)$ is divisible by $sp_1\ldots p_l$

and $s^2p_1^2 \dots p_l^2 \le (C^2+1)|N|$. So there are only a finite number of sets $\{p_1, \dots, p_l\}$ that arise. Let $A(p_1, \dots, p_l)$ be the set of k such that $r \in [ks^2, (k+1)s^2) \subset [-C\sqrt{|N|}, C\sqrt{|N|}]$ with the given residue $\bar{r}$ has $s(r)$ divisible by $sp_1 \dots p_l$ and by no prime but those in $\{p_1, \dots, p_l\}$. Let $|A|$ be the total number of elements in all the $A(p_1, \dots, p_l)$, $l \ge 0$. Then $|A| - 2C\sqrt{|N|}/s^2$ is $O(1)$ and $1/|A| = s^2/2C\sqrt{|N|} + O(1/|A|^2)$. Let $A(+)$ be the union of $A(p_1, \dots, p_l)$, $l > 0$.

Set $\Psi_\pm(k)$ equal to the integral over the interval $[ks^2/\sqrt{|N|}, (k+1)s^2/\sqrt{|N|}]$ of $\psi_\pm(x)\sqrt{|x^2 \pm 1|}$. Then with our new conventions, the integral in (79) becomes

$$\sum_{k \in A()} \Psi_\pm(k)(1 - \sum_{i \in A()} 1/|A|) - \sum_{k \in A(+)} \sum_{i \in A()} \Psi_\pm(k)/|A| + O(s^2/\sqrt{|N|}).$$

Thanks to (B.10) we may ignore the error term. The main term is

$$\frac{1}{|A|} \sum_{k \in A()} \sum_{i \in A(+)} (\Psi_\pm(k) - \Psi_\pm(i)) \tag{81}$$

Each term $\Psi_\pm(k)$ that appears in (81) is assigned not only to a $k \in A()$ but also to an $i \in A(+)$, say $i \in A(p_1, \dots, p_l)$. We can change the assignation and thus rearrange the sum by decomposing the integers into intervals $I_m = [ms^2p_1^2 \dots p_l^2, (m+1)s^2p_1^2 \dots p_l^2)$, choosing for each of these intervals an i' in it such that $i' \equiv i \pmod{s^2p_1^2 \dots p_l^2}$ and assigning $\Psi_\pm(k)$ to k and to that i' lying in the same interval I_m as k. For this to be effective, we introduce sets $B(p_1, \dots, p_l)$, defined as the set of k such that the $r \in [ks^2, (k+1)s^2)$ with the given residue $\bar{r}$ modulo s has $s(r)$ divisible by $sp_1 \dots p_l$. Then i' necessarily lies in $B(p_1, \dots, p_l)$, although it may not lie in $A(p_1, \dots, p_l)$. Then the union $B(+)$ of all the $B(p_1, \dots, p_l)$, $l > 0$, is again $A(+)$ but these sets are no longer disjoint. The number of times $Q(k, i')$ that k is assigned to a given i' is clearly $O(\sqrt{|N|}/s^2p_1^2 \dots p_l^2)$.

If we change notation, replacing i' by i, the sum (81) becomes

$$\frac{1}{|A|} \sum_{i \in A(+)} \sum_{k \in A()} Q(k,i)(\Psi_\pm(k) - \Psi_\pm(i)). \tag{82}$$

If $\psi_\pm$ is continuously differentiable and if $i \in A(p_1, \dots, p_l)$ and $Q(k,i) \neq 0$, then

$$\Psi_\pm(k) - \Psi_\pm(i) = O(\frac{s^2}{\sqrt{|N|}})O(\frac{s^2\prod_{j=1}^l p_j^2}{\sqrt{|N|}}),$$

the first factor coming from the length of the interval, the second from the difference of the functions $\psi_\pm$ on the two intervals. Since the number of elements in $B(p_1, \dots, p_l)$ is $O(2^{l'}\sqrt{|N|}/s^2p_1^2 \dots p_l^2)$, l' being the number of p_j^2, $1 \le j \le l$, that do not divide s, (81) is estimated as

$$\frac{1}{|A|} \sum_{l>0} \sum_{p_1, \dots, p_l} 2^{l'} O(\frac{s^2}{\sqrt{|N|}})O(\frac{s^2\prod_{j=1}^l p_j^2}{\sqrt{|N|}})O(\frac{\sqrt{|N|}}{s^2p_1^2 \dots p_l^2})^2 O(p_1^2 \dots p_l^2),$$

the final factor being the number of intervals of length s^2 in an interval I_m. This expression is

$$O(\frac{s^2}{\sqrt{|N|}})O(\sum_{l>0}\sum_{p_1,\dots,p_l} 2^{l'}), \tag{83}$$

which multiplied by $\sqrt{|N|}/s^3$ yields

$$O(\frac{1}{s})O(\sum_{l>0}\sum_{p_1,\dots,p_l} 2^{l'}). \tag{84}$$

Were it not for the second factor, we could appeal to (B.10). Even though this factor is a finite sum because $s^2p_1^2\dots p_l^2 \le (C^2+1)\sqrt{|N|}$, it is far too large to be useful. It is likely to have been very wasteful to estimate the terms in (82) individually. We can after all expect that if $0 \le \bar{i} < s^2p_1^2\dots p_l^2$ is the residue of i in $A(p_1,\dots,p_l)$ then $\bar{i}/s^2p_1^2\dots p_l^2$ is distributed fairly uniformly over $[0,1)$ as $p_1,\dots,p_l$ vary but at the moment I do not know how to establish or to use this. So the poor estimate (83) is one obstacle to establishing a reasonable estimate for (72).

As in the analysis of (79), we may calculate, with an error that is easily estimated as $O(s^2/\sqrt{|N|})$, $\bar{g}_s$ as $|A()|/|A|$ or, better, $s^2|A()|/2C\sqrt{|N|}$. It is clear that

$$|A()| = |B()| - \sum_{p_1}|B(p_1)| + \sum_{p_1,p_2}|B(p_1,p_2)| - + \dots, \tag{85}$$

in which the sum is over $s^2p_1^2\dots p_l^2 \le C^2+1$. Each set $B(p_1,\dots,p_l)$ corresponds to an interval $[ks^2p_1^2\dots p_l^2, (k+1)s^2p_1^2\dots p_l^2)$ and it is implicit in the definition that this interval must meet $[-C\sqrt{|N|}, C\sqrt{|N|}]$. It is not, however, necessary that it be contained in the larger interval. Then

$$|B(p_1,\dots,p_l)| = \alpha(p_1,\dots,p_l)(\frac{2C\sqrt{N}}{s^2p_1^2\dots p_l^2} + \epsilon(p_1,\dots,p_l)). \tag{86}$$

Here

$$\alpha(p_1,\dots,p_l) = \prod_{j=1}^{l}\alpha(p_j).$$

If p is odd, $\alpha(p)$, which is 0, 1 or 2, is the number of solutions of $r^2 - N \equiv 0 \pmod{s^2p^2}$, $(r^2 - N)/s^2p^2 \equiv 0,1 \pmod 4$ with the condition that the residue of r modulo s^2 is $\bar{r}$. If $p=2$, it is $1/4$ the number of solutions of the same conditions but with r taken modulo $4s^2p^2$. Because those intervals $[ks^2p_1^2\dots p_l^2, (k+1)s^2p_1^2\dots p_l^2)$ that lie partly inside $[-C\sqrt{|N|}, C\sqrt{|N|}]$ and partly outside may or not belong to $B(p_1,\dots,p_l)$ the number $\epsilon(p_1,\dots,p_l)$ lies between -1 and 1 if no p_j is 2. Otherwise it lies between -4 and 4.

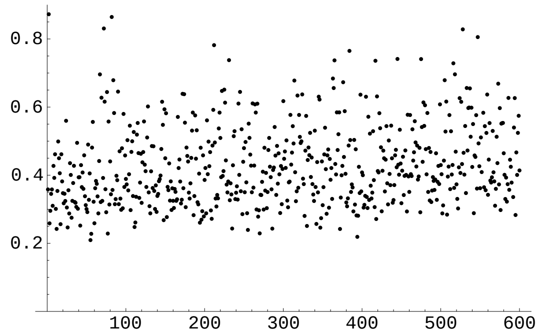

Diagram 3.2.B

We can also calculate the modified $\alpha(N, s)$ as

$$\sum_{p_1,\ldots,p_l} (-1)^l \frac{\alpha(p_1,\ldots,p_l)}{s^2 p_1^2 \ldots p_l^2}.$$

We conclude from (85) and (86) that, apart from an error that we can allow ourselves, the difference $\bar{g}_s - s^2\alpha$ is

$$\sum_{l\,\geq\,0} \sum_{p_1,\ldots,p_l} (-1)^l \alpha(p_1,\ldots,p_l)\epsilon(p_1,\ldots,p_l) \frac{s^2}{2C\sqrt{|N|}}.$$

So once again, we have to deal with

$$\sum_s \frac{1}{s} \Big\{ \sum_{l\,\geq\,0} \sum_{p_1,\ldots,p_l} (-1)^l \alpha(p_1,\ldots,p_l)\epsilon(p_1,\ldots,p_l) \Big\}. \tag{87}$$

0.356779,	−0.136681,	0.358137,	−0.135542,	0.358815,	−0.136301,
−0.305870,	0.089564,	−0.304567,	0.090622,	−0.303917,	0.089916,
0.102864,	0.114616,	0.104113,	0.115592,	0.104736,	0.114941,
−0.108330,	0.054990,	−0.107135,	0.055883,	−0.106538,	0.055287,
−0.027594,	0.071878,	−0.026452,	0.072683,	−0.025881,	0.072146,
−0.212283,	0.083968,	−0.211192,	0.084682,	−0.210647,	0.084206,
0.117788,	−0.003163,	0.118829,	−0.002547,	0.119348,	−0.002958,
0.091523,	−0.015066,	0.092514,	−0.014557,	0.093010,	−0.014897,
0.020256,	−0.084660,	0.021200,	−0.084275,	0.021671,	−0.084532,
−0.252761,	−0.016231,	−0.251863,	−0.016025,	−0.251414,	−0.016162,
0.133049,	0.067864,	0.133903,	0.068064,	0.134330,	0.067930,
0.088015,	0.014307,	0.088828,	0.014663,	0.089234,	0.014425,
−0.081067,	−0.030958,	−0.080293,	−0.030509,	−0.079906,	−0.030808,
0.017027,	0.076392,	0.017766,	0.076908,	0.018135,	0.076564,
0.121633,	0.025750,	0.122340,	0.026318,	0.122693,	0.025939,
−0.081617,	0.053260,	−0.080938,	0.053867,	−0.080599,	0.053463,
−0.002066,	−0.126718,	−0.001409,	−0.126081,	−0.001082,	−0.126505,
0.068478,	0.082597,	0.069116,	0.083256,	0.069435,	0.082816,
−0.004951,	0.124929,	−0.004325,	0.125601,	−0.004012,	0.125153,
−0.239656,	0.099643,	−0.239035,	0.100322,	−0.238725,	0.099869.

Table 3.2.A: Part 1: $p = 59369$

The expression in parentheses in (87) depends strongly on s and is, once again, apparently far too large, a coarse estimate suggesting that the inner sum is of magnitude

$$\sum_{p_1^2 \cdots p_l^2 \le (C^2+1)\sqrt{|N|}/s^2} 2^{l'}, \tag{88}$$

where l' is once again the number of j, $1 \le j \le l$ such that p_j does not divide s. Perhaps we have to take into account that the signs of the factors $\epsilon(p_1, \ldots, p_l)$ vary and cancel each other. I have not tried to do this.

-0.219263,	0.068058,	-0.215435,	0.071714,	-0.208291,	0.070964,
0.143721,	0.001004,	0.146339,	0.004404,	0.153195,	0.003706,
-0.020184,	0.035014,	-0.016663,	-0.065354,	-0.010093,	-0.065998,
-0.091281,	-0.098202,	-0.087911,	-0.095335,	-0.081622,	-0.095923,
0.003985,	-0.017991,	0.007207,	-0.015405,	0.013220,	-0.015936,
0.087754,	0.076422,	-0.042187,	0.078715,	-0.036444,	0.078245,
0.180775,	-0.060344,	0.183710,	-0.058364,	0.189187,	-0.058771,
0.107415,	0.038662,	0.110212,	0.013086,	0.115430,	0.012750,
-0.058412,	-0.094389,	-0.167936,	-0.093154,	-0.162967,	-0.093407,
-0.073063,	0.007405,	-0.070531,	0.008066,	-0.065803,	0.007930,
0.070096,	0.045556,	0.072506,	0.046198,	0.077003,	0.046066,
-0.023146,	0.060006,	-0.020854,	0.061151,	-0.016574,	0.060916,
0.129606,	-0.071048,	0.084723,	-0.117606,	0.088799,	-0.117902,
-0.044026,	0.110108,	-0.041942,	0.111766,	-0.038054,	0.111426,
-0.015731,	-0.062830,	-0.013737,	-0.061007,	-0.010015,	-0.061381,
-0.180290,	0.079119,	-0.258895,	0.081069,	-0.255319,	0.080669,
0.077034,	-0.078929,	0.078885,	-0.076884,	0.082340,	-0.077304,
-0.011454,	0.060534,	-0.009653,	-0.008355,	-0.006292,	-0.008789,
-0.003079,	0.082049,	-0.001312,	0.084207,	0.001986,	0.083764,
0.031046,	0.121403,	-0.041293,	0.123583,	-0.038028,	0.123136.

Table 3.2.A: Part 2: $p = 746777$

Our estimate of (70) is unsatisfactory, so that at this stage it is useful to examine it numerically. The numerical results that I now describe suggest strongly that all estimates that look, for one reason or another, weak are indeed so and that (70) is $O(\ln^2 |N|)$. The experimental results, too, leave a good deal to be desired, partly because it is impossible to detect slowly growing coefficients but also because it is inconvenient (for me with my limited programming skills) to work with integers greater than $2^{31} = 2147483648$. For example, when testing the divisibility properties of $r^2 - N$ by s^2, it is inconvenient to take s greater than 2^{15}. Since we can work with remainders when taking squares, we can let r be as large as 2^{31}. Nonethless, if we do not want to take more time with the programming and do not want the machine to be too long with the calculations, there are limits on the accuracy with which we can calculate the $s = s_r$ appearing in (72). We can calculate a large divisor of s, for example the largest prime divisor that is the product of powers $q^a = q^{a_q}$ of the first Q primes, where Q is at our disposition and where q^{a_q} is at most 2^{15}. The same limitations apply to the calculation of $\epsilon_{n,f}(N)$ and in particular of $\epsilon_{1,1}(N)$. So we can only approximate (72), the approximation depending also on Q.

$0.065614,$	$-0.026226,$	$0.077929,$	$-0.016149,$	$0.086276,$	$-0.021717,$
$-0.099652,$	$-0.151652,$	$-0.087834,$	$-0.142283,$	$-0.079825,$	$-0.147459,$
$-0.148913,$	$0.172582,$	$-0.137585,$	$0.181226,$	$-0.129909,$	$0.176450,$
$0.403084,$	$-0.036168,$	$0.413927,$	$-0.028269,$	$0.421275,$	$-0.032633,$
$-0.156494,$	$0.060108,$	$-0.146127,$	$0.067236,$	$-0.139102,$	$0.063297,$
$0.016583,$	$0.038520,$	$0.026483,$	$0.044839,$	$0.033191,$	$0.041348,$
$0.273885,$	$0.195733,$	$0.283327,$	$0.201189,$	$0.289726,$	$0.198174,$
$-0.074761,$	$-0.041762,$	$-0.065764,$	$-0.037253,$	$-0.059667,$	$-0.039744,$
$0.092875,$	$0.010855,$	$0.101440,$	$0.014259,$	$0.107244,$	$0.012378,$
$-0.126962,$	$0.015473,$	$-0.118812,$	$0.017294,$	$-0.113290,$	$0.016288,$
$0.065242,$	$0.009336,$	$0.072994,$	$0.011104,$	$0.078248,$	$0.010127,$
$0.258894,$	$0.023944,$	$0.266270,$	$0.027098,$	$0.271269,$	$0.025356,$
$-0.120552,$	$-0.028011,$	$-0.113526,$	$-0.024038,$	$-0.108765,$	$-0.026233,$
$-0.059059,$	$0.034596,$	$-0.052355,$	$0.039166,$	$-0.047812,$	$0.036641,$
$-0.106739,$	$-0.015397,$	$-0.100324,$	$-0.010374,$	$-0.095977,$	$-0.013149,$
$0.093762,$	$-0.044505,$	$0.099926,$	$-0.039132,$	$0.104104,$	$-0.042101,$
$-0.101929,$	$0.219005,$	$-0.095973,$	$0.224641,$	$-0.091937,$	$0.221527,$
$0.101531,$	$0.011202,$	$0.107326,$	$0.017028,$	$0.111253,$	$0.013809,$
$0.036660,$	$-0.033043,$	$0.042345,$	$-0.027093,$	$0.046198,$	$-0.030380,$
$-0.010312,$	$-0.008676,$	$-0.004683,$	$-0.002666,$	$-0.000869,$	$-0.005986.$

Table 3.2.A: Part 3: $p = 8960467$

In Table 3.2.A, which has three parts, we give three approximations not to the difference itself but to the difference divided by $\ln p$. Each is for $n = f = 1$ and for three different primes of quite different sizes, the 6000th, $p = 59369$, the 60000th, $p = 746777$, and the 600000th, $p = 8960467$. The three approximations are for $Q = 80, 160, 320$. They give not (72) itself, but the measure implicit in it, thus the mass with respect to the measure of the twenty intervals of length .1 between -2 and 0, a point mass falling between exactly at the point separating two intervals being assigned half to one and half to the other interval. All these masses are divided by $\ln p$. For the smallest of the three primes, all approximations give similar results. For the largest of the primes, even the best two are only close to another. For numbers with any claim to precision, either a larger value of Q or a larger bound on the powers of the primes would be necessary. Nevertheless, the change in the numbers with increasing Q is far, far less than suggested by (84) and (88).

In each part of the table one of the three primes is considered. Each part has three double columns, each of them corresponding to one value of Q. For a given Q, the first element of the double column is the measure for ψ_- and the second for ψ_+. The interval in the first row is $[-2,-1.9]$ and in the last is $[-.1,0]$. Notice that the mass divided by $\ln p$ does not seem to grow much or to decrease much but does behave irregularly. Thus the mass itself at first glance seems to be about $O(\ln p)$, but, as already suggested, this is not the correct conclusion.

$0.065424,$	$0.139319,$	$0.009127,$	$0.020117,$
$-0.339535,$	$0.095697,$	$0.244114,$	$0.028861,$
$0.064939,$	$-0.125936,$	$-0.242039,$	$0.013350,$
$0.227577,$	$0.215971,$	$0.068047,$	$0.008164,$
$-0.077371,$	$-0.069531,$	$0.011311,$	$0.034411,$
$-0.210577,$	$-0.056532,$	$0.159543,$	$-0.064707,$
$0.373104,$	$-0.083462,$	$-0.002606,$	$0.046361,$
$-0.279405,$	$0.042905,$	$0.147926,$	$0.197334,$
$0.176183,$	$0.114329,$	$-0.041558,$	$0.238247,$
$-0.049009,$	$-0.016793,$	$-0.180851,$	$-0.082724.$

Table 3.2.B

To exhibit the fluctuating character of these numbers, a similar table for the 6001st prime $p = 746791$ is included as Table 3.2.B, but I only give the results for $Q = 320$. Once again, they come in pairs, for $N = -1$ and $N = 1$, but there are two columns, the first for the interval from -2 to -1 and the second for the interval from -1 to 0. Table 3.2.B can be compared with Table 3.2.A, Part 1 to see the change on moving from one prime to the next.

As a further test, I took the largest of the absolute values of the masses of the 2 times 20 intervals for the 1000th, the 2000th, and so on up to the 100000th prime and divided it by $\ln p$. The one hundred numbers so obtained appear as Table 3.2.C, which is to be read like a normal text, from left to right and then from top to bottom. In the calculations, the integer Q was taken to be 160, but doubling this has only a slight effect. Although at first glance, there is no obvious sign in the table of any increase, a plot of the numbers, as in[16] Diagram 3.2.A, suggests that they do increase and rather dramatically.

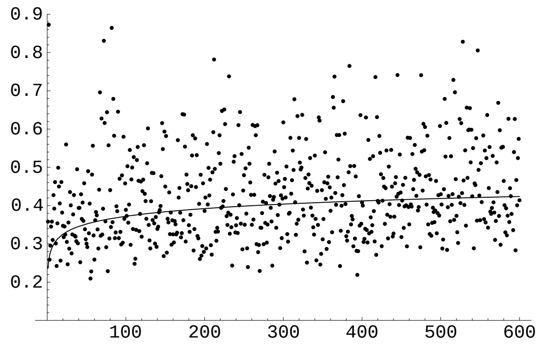

Diagram 3.2.C

On the other hand, if we continue up to the 600000th prime we obtain the results of Diagram 3.2.B, where once again $Q = 160$ and where once again doubling Q leads to essentially the same scattering with only a slight displacement of the points. So Diagram 3.2.A is misleading and there is no dramatic rise! A second, more careful glance at the diagram suggests, however, that a slow movement of the points upward, perhaps compatible with the $O(\ln^2 |N|) = O(\ln^2 p)$ hypothesis, is not out of the question. We will return to this point when we have more and different data at our disposition. As a convenient comparison, Diagram 3.2.C superposes the points of Diagram 3.2.B on the graph of the curve $.4\ln(1000x\ln(1000x))/15, 1 \leq x \leq 600$. The diagram confirms, to the extent it can, the hypothesis.

[16] Unfortunately, it was not always convenient to insert the tables and the diagrams at the points where they are discussed in the text.

0.358120,	0.872623,	0.258640,	0.295414,	0.345120,
0.358137,	0.310121,	0.427307,	0.392101,	0.461449,
0.301074,	0.242229,	0.353707,	0.498970,	0.449767,
0.405747,	0.256198,	0.461453,	0.381769,	0.347241,
0.317558,	0.345492,	0.324273,	0.559732,	0.305104,
0.246601,	0.355806,	0.287550,	0.435331,	0.400707,
0.275095,	0.324584,	0.376984,	0.427550,	0.321304,
0.319035,	0.306974,	0.494958,	0.301518,	0.393844,
0.394138,	0.252000,	0.429559,	0.365034,	0.407917,
0.359968,	0.458391,	0.338244,	0.312106,	0.300587,
0.291630,	0.489896,	0.327670,	0.405218,	0.209386,
0.227849,	0.481018,	0.556393,	0.322056,	0.258895,
0.361781,	0.383069,	0.374638,	0.337790,	0.287852,
0.441601,	0.695974,	0.321117,	0.627571,	0.324480,
0.391816,	0.830769,	0.615896,	0.358815,	0.291243,
0.644122,	0.228597,	0.557525,	0.313941,	0.440433,
0.343996,	0.864512,	0.356637,	0.678889,	0.582523,
0.314871,	0.329813,	0.398283,	0.385383,	0.645377,
0.314966,	0.470168,	0.331259,	0.298338,	0.479059,
0.302799,	0.579901,	0.365380,	0.457965,	0.388941.

Table 3.2.C

3.3 Some suggestive phenomena. The previous section does not establish beyond doubt that (70) is $O(\ln^2 p)$ or even the slightly weaker hypothesis, that, for some integer l the expression (70) is $O(\ln^l p)$. We now, consider fixing n and f and taking the average of (70), in the sense of (12′) over the primes up to X. If $X = x\ln x$, then, under the hypothesis that (70) is $O(\ln p)$, the order of the average will be majorized by a constant times

$$\frac{\sum_{n<x} \ln^2(n\ln n)}{x\ln x}. \tag{89}$$

This is approximately

$$\frac{\int_2^x \ln^2(t\ln t)}{x\ln x} \sim \ln x \sim \ln(x\ln x) = \ln X.$$

If the order were $\ln^l p$, then (89) would be majorized by a constant times $\ln^l X$. The average is a measure $\nu_{n,f,X}$, which we may also consider as a distribution on the set of possible $\psi_\pm$. Suppose

$$\nu_{n,f,X} = \alpha_{n,f} + \beta_{n,f} \ln X + o_{n,f}(1),$$

where $\alpha_{n,f}$ and $\beta_{n,f}$ are two measures or distributions. Then interchanging the order of the sum up to X and the sum over n, we find that we are to take the limit of

$$\sum_{n,f} \alpha_{n,f} + \sum_{n,f} \beta_{n,f} \ln X + \sum_{n,f} o_{n,f}(1). \tag{90}$$

If there were no contributions from the other two ranges and if the third sum was itself $o(1)$, then the sum

$$\sum_{n,f} \beta_{n,f} \tag{91}$$

would have to be 0, and the sum

$$\sum_{n,f} \alpha_{n,f} \tag{92}$$

the limit for which we are looking, thus the contribution from the first range of summation where n is smaller than $\sqrt{|N|}$. There is, however, no good reason to expect that (91) is 0. It may be cancelled by a contribution from the intermediate range.

We can certainly envisage polynomials of higher degree in (90). For such asymptotic behavior to make sense, it is best that the change in $\ln^l X$ be $o(1)$, as X changes from $n \ln n$ to $(n+1)\ln(n+1)$, thus in essence as we pass from one prime to the next. Since

$$(n+1)\ln(n+1) = (n+1)\ln n + O(1) = n \ln n(1 + O(\frac{1}{n})),$$

we have

$$\ln((n+1)\ln(n+1)) = \ln(n \ln n) + O(\frac{1}{n}).$$

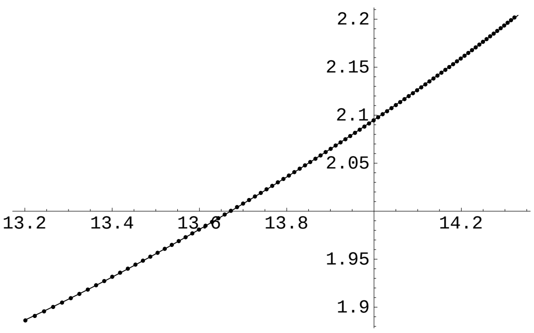

Diagram 3.3.E

I examined the behavior of the average of the sum (70) for $n = 1, 3, 5, 15$ and $f = 1$, treating it again as a measure on the two lines $N = \pm 1$ and plotting the average, in the sense of (12′), over the first $1000k$ primes for $1 \le k \le 60$ against $1000k \ln(1000k)$. The results are given at the end of the paper in Diagrams 3.3.A to 3.3.D. The results are not so simple as (90), although they do make it clear that the average behaves regularly and is naturally expressed as a quadratic function of $\ln(X)$, so that $\nu_{f,n,X}$ would be a quadratic function of $\ln X$ with a small remainder and there would be another sum in (91) that would have to vanish.[17] I divided the interval $[-3, 0]$ into six intervals of length .5, each column of each diagram contains the six graphs for the six intervals, the first column for $N = -1$ and the second for $N = 1$. They are close to linear as (90) suggests, but not exactly linear.

So I redid the experiments for $\det N > 0$ and $n = 1$ on the intervals in $[-1, 0]$ for primes up to 140000, using a slightly better approximation to the integral

$$\int \sqrt{1 - x^2} dx$$

over the two intervals but continuing to use only 320 primes to compute the various factors. Since $\ln(k \ln k)$ is 13.4 for $k = 60000$, 14.32 for $k = 140000$ and 15.7 for $k = 600000$, not much is gained by taking even more primes. The two resulting curves, but only for $50000 \le k \le 140000$, together with quadratic approximations to them are shown in Diagrams 3.3.E, for the first interval, and 3.3.F, for the second. The quadratic approximations are

$$-1.37552 + 0.24677x + 0.06329(x - 13.5)^2$$

for the first interval and

$$-1.97565 + 0.33297x + 0.10656(x - 13.5)^2$$

for the second. The quadratic term looks to be definitely present. There may even be terms of higher order, but there appears to be a little question that we are dealing with a function that as a function of $\ln X$ is essentially polynomial. Thus the natural parameter is $\ln X$ and not some power of X.

[17] It is perfectly clear to me that these suggestions are far-fetched. I feel, nevertheless, that they are worth pursuing.

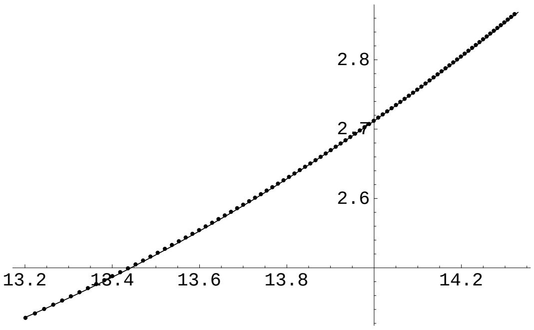

Diagram 3.3.F

Part IV: Supplementary remarks

4.1 Quaternion algebras. There is some advantage in treating quaternion algebras as similar results are to be expected, but only the terms (i) and (ii) appear in the trace formula. The disadvantage, especially for numerical purposes, is that some ramification has to be admitted immediately. Apart from that, the only formal difference in the elliptic term is that the discriminant D is subject to the condition that $\left(\frac{D}{q}\right) = 1$ for those q that ramify in the quaternion algebra. Moreover, if the algebra is ramified at infinity then only $D < 0$ are allowed.

4.2 Transfer from elliptic tori. The representation σ_2 is of course the representation

(93) $$X \to AXA^t$$

on the space of symmetric matrices. Thus if a reductive subgroup ${}^\lambda H^{\mathbb{Q}}$ of ${}^L G^{\mathbb{Q}} = GL(2,\mathbb{C})$ is not abelian but has a fixed vector in the representation, it is contained in an orthogonal group. Observe that the condition of §1.3 may no longer be fulfilled: the group ${}^\lambda H$ may not lie in $SL(2,\mathbb{C}) \times \mathrm{Gal}(K/F)$. If ${}^\lambda H^{\mathbb{Q}}$ is the first term of an inverse system ${}^\lambda H$ in the system ${}^L G$, then ${}^\lambda H^{\mathbb{Q}}$ is contained in the usual image in ${}^L G^{\mathbb{Q}}$ of the L-group of an elliptic torus. Thus, if we take the ρ implicit in (12′) to be σ_2, then we can expect to single out in the limit those cuspidal representations π that are transfers from elliptic tori. They will, however, have an additional property. If the torus is associated to the quadratic extension E with associated character χ_E, then χ_E will be the central character of π. Since we can, in the context of the trace formula, fix the central character of the representations π to be considered in any way we like, we can in fact single out those representations that are transfers from a given elliptic torus. Then the sum in (14′) will be a sum over a single torus.

If we want an arbitrary central character, then we have to replace (93) by the tensor product of σ_4 with $\det^{-2}$. Thus the sum in (14′) will be an infinite sum, over all elliptic tori. Moreover there will in all likelihood be no choice but to let the transfer $f \to f^H$ reflect the reality of the situation. It will have to be defined by the condition that

$$\operatorname{tr}\theta(f^H) = \operatorname{tr}\Theta(f)$$

if Θ is the transfer of the character θ. These transfers are certainly known to exist, but the relation between the characters of θ and Θ remains obscure. So the definition of f^H, which is to be made locally is by no means clear.

If the base field is $\mathbb{Q}$, we cannot take f_v to be unramified at all finite places, because f_v^H would then necessarily be 0 at those places where the quadratic field defining the torus H was ramified. So for experimental purposes, some ramification in f has to be admitted.

If we consider only representations trivial on Z_+, then (14′) will be

$$\sum_H \sum_\theta \operatorname{tr}\theta(f^H), \tag{94}$$

with those θ that lead to noncuspidal representations excluded. Since they can be taken care of separately, it is best to include them. Then (94) can be written as

$$\sum_H \mu(Z_+H(\mathbb{Q})\backslash H(\mathbb{A})) \sum_{\gamma\in H(\mathbb{Q})} f^H(\gamma). \tag{95}$$

Although this sum appears infinite, it will not be, because f^H will necessarily be 0 for those H that ramify where f_v is unramifed. The sum (95) is very much like the elliptic term of the trace formula, except that the γ in the center appear more than once.

The transfer $\theta \to \Theta$ is well understood at infinity. There, at least, the inverse tranfer $f \to f^H$ differs in an important way from endoscopic transfer. Endoscopic transfer is local in the sense that the support of (the orbital integrals of) f^H is, in the stable sense, the same as the support of (those of) f^H. In contrast, even if the orbital integrals of f_∞ are supported on hyperbolic elements, f_∞^H may be nonzero for tori elliptic at infinity. This does not prevent a comparison between (14′) and (17′), but does suggest that it may have a number of novel elements not present for endoscopy.

The first, simplest test offers itself for the representations unramified everywhere. Since every quadratic extension of $\mathbb{Q}$ is ramified somewhere, there are no unramified representations arising from elliptic tori. Thus the limit (12′) should be 0 for $\rho = \sigma_m$, $m = 2$. This is even less obvious than for $m = 1$ and everything will depend on the elliptic contribution to the trace formula. It must cancel all the others. I have made no attempt to understand numerically how this might function, but it would be very useful to do so. A distillation that separates the different kinds of contribution in the elliptic term term may be necessary. It would then be useful to understand clearly the orders of magnitude of these contributions.

As a convenient reference for myself, and for anyone else who might be inclined to pursue the matter, I apply the formulas of Appendix B to the conclusions of §2.4 to obtain a list of all the contributions to be cancelled. As it stands, the list has no structure and the terms no meaning. Until they do, §4.3 has to be treated with scepticism.

4.3 Contributions for even m. I consider all contributions but the elliptic. The first is made up of (31) from the term (ii), corrected by the last term in (41) and by (48) to yield

$$(a) \qquad \frac{m}{4}\,\mathrm{tr}(\xi_0(f_\infty)).$$

The second is the sum of atomic measures in (41):

$$(b) \qquad \sum_q \sum_{n>0} \{|q^n - q^{-n}|\psi_+(q^n + q^{-n}) + |q^n + q^{-n}|\psi_-(q^n - q^{-n})\},$$

The third arises from (51) which is equal to

$$\psi_-(0)\{\ln 2 + m \ln p(1 + O(\frac{1}{p}))\}$$

and whose average is

$$(c) \qquad \psi_-(0)(\ln 2 + m \ln X).$$

As was already suggested, this means that for $m = 2$ the analogue of (91) will not be 0, but will have to cancel, among other things, (c), at least when there is no ramification.

The contributions from (56) and (48) yield together, in the notation of Appendix C,

$$\sum_{\pm}(\kappa_1 + \frac{m \ln p}{2})\hat{f}_\infty(a(1, \pm 1)),$$

or when averaged

$$(d) \qquad \sum_{\pm}(\kappa_1 + \frac{m \ln X}{2})\hat{f}_\infty(a(1, \pm 1)).$$

I offer no guarantee for the constants in (c) and (d).

All that remains are the terms resulting from the combination of (49) and (57) with (46) and an application of Hoffmann's formula. There is, first of all, the contribution from (C.13) (which must be multiplied by $1/2$)

$$(e)\qquad -\frac{1}{2}\sum\int_{-\infty}^{\infty}\frac{e^{-|x|}}{1+e^{-|x|}}\hat{f}_\infty(a)dx,$$

where the sum is over the arbitrary sign before the matrix

$$a=\pm\begin{pmatrix} e^x & 0 \\ 0 & -e^{-x}\end{pmatrix},$$

and, from (C.17) and (C.18),

$$(f)\qquad \frac{1}{2}\sum\int_{-\infty}^{\infty}(\frac{e^{-|x|}}{1-e^{-|x|}}-\frac{1}{|x|})\hat{f}_\infty(a)dx,$$

in which

$$a=a(x)=\pm\begin{pmatrix} e^x & 0 \\ 0 & e^{-x}\end{pmatrix},$$

the sum being again over the sign, and

$$(g)\qquad -\frac{1}{2}\int_{-\infty}^{\infty}\ln|x|\frac{d\hat{f}_\infty}{dx}(a)dx,$$

which according to the formula of Appendix D is equal to

$$\int_{-i\infty}^{i\infty}(\ln|s|+\lambda_0)\operatorname{tr}\xi_s(f_\infty).$$

From (47) we have

$$(h)\qquad \frac{1}{16\pi i}\int_{-i\infty}^{i\infty}\{-\frac{\Gamma'((1-s)/2)}{\Gamma((1-s)/2)}-\frac{\Gamma'(s/2)}{\Gamma(s/2)}-\frac{\Gamma'((1+s)/2)}{\Gamma((1+s)/2)}-\frac{\Gamma'(-s/2)}{\Gamma(-s/2)}\}\operatorname{tr}\xi_s(f_\infty)ds.$$

Finally, from (D.19) there is the completely different contribution

$$(i)\qquad -\frac{1}{2}\sum_{k=0}^{\infty}(\pm 1)^{k-1}\Theta_{\pi_k}(f).$$

The usual formulas ([N,§72] for the logarithmic derivative of the Γ-function suggest that there should be cancellation among (f), (g) and (h). The Fourier transform of $\xi_s(f_\infty)$ is, however, a function on all four components of the group of diagonal matrices, each component determined by the signs in

$$a=a(x)=\begin{pmatrix} \pm e^x & 0 \\ 0 & \pm e^{-x}\end{pmatrix}.$$

So any cancellation between (h) and (f) would also have to involve (e). I am not familiar with any formula that relates (e) to the Γ-function and have not searched for one.

4.4 The third touchstone. The problem (T3) is, on the face of it, different than the first two, but may be amenable to the same kind of arguments. If the base field is $\mathbb{Q}$, the pertinent representations of $GL(2,\mathbb{R})$ are those obtained by induction from the representations

$$\begin{pmatrix} a & x \\ 0 & b \end{pmatrix} \to (\operatorname{sgn} a)^k (\operatorname{sgn} b)^l |\frac{a}{b}|^{s/2}, \quad k,l = \pm 1.$$

We can try to isolate them by a function f_∞ such that $\operatorname{tr}\pi(f_\infty)$ is 0 if π lies in the discrete series and $\operatorname{tr}\xi_s^{k,l}(f_\infty)$ is independent of k,l but, as a function of s, is an approximation to the δ-function at $s=0$. This means that $\hat{f}_\infty$ is concentrated on a with positive eigenvalues and that it is approaching the function identically equal to 1. Thus ψ_- will be 0 and ψ_+ will be 0 for $x<-1$. For $x>1$, it will be approaching

$$\psi_+(x) = \frac{1}{e^t - e^{-t}} = \frac{1}{\sqrt{|x^2-1|}}, \quad r = e^t + e^{-t}, \quad x = \frac{r}{2}.$$

What will happen on the range $-1<x<1$ remains to be worked out.

Since the approximation at infinity would be occurring while f_q remained fixed at the other places, the sum over r in the elliptic term of the trace formula would be a sum over a fixed lattice – the lattice of integral r if f_q were the unit element of the Hecke algebra everywhere. So the problems that arise look to be different than those for (T2): the limits to be taken are of a different nature. They are perhaps easier, perhaps more difficult; but I have not examined the matter. I have also not examined the role of the other terms in the trace formula.

4.5 General groups. Is there an obvious obstacle to extending the considerations of this paper to general groups? Recall that the structure of the trace formula is the equality of a spectral side and a geometric side. The principal term of the spectral side is the sum over the representations occurring discretely in $L^2(G(\mathbb{Q})\backslash G(\mathbb{A}))$ of $\operatorname{tr}\pi(f))$. As for $GL(2)$, we will expect that an inductive procedure will be necessary to remove the contributions from representations that are not of Ramanujan type. This will leave

$$\sum_{\pi}^{R} \operatorname{tr}\pi(f)$$

in which to substitute appropriate f before passing to the limit.

On the geometric side, there will also be a main term, the sum over the elliptic elements. For $GL(k)$ an elliptic element γ corresponds to a monic polynomial

$$x^k + a_1 x^{k-1} + \ldots a_{k-1} x + a_k.$$

For $GL(2)$, $a_1 = -r$, $a_2 = N/4$. Of course, for γ to be regular certain degenerate sequences $a_1, a_2, \ldots, a_k$ will have to be excluded. For $GL(2)$, not only is $N \neq 0$ but $r^2 - N \neq 0$. In addition split γ are excluded. We should like to say that for a general group, an elliptic element is defined, after the exclusion of singular or partially split elements, by the values of a similar sequence $a_1, a_2, \ldots, a_k$. If the group is semisimple and simply connected, these could be the characters of the representations with highest weight λ_i, $(\lambda_i, \alpha_j) = \delta_{i,j}$, but only if we deal not with conjugacy classes in the usual sense but with stable conjugacy classes, as is perfectly reasonable if we first stabilize the trace formula. For groups that are not semisimple or not simply connected, something can surely be arranged. So we can expect in general a sum over a lattice, analogous either to the lattice of integral (r, N), or, if we recognize that the values of the rational characters of G on those γ that yield a contribution different from 0 will be determined up to a finite number of possibilities by f, over an analogue of the lattice of r. As for $GL(2)$, it will be appropriate to allow a fixed denominator or to impose congruence conditions.

The limits of the remaining terms, either on the spectral side or on the geometric side, we can hope to treat by induction. So the question arises during these preliminary reflections whether the terms in the sum over the lattice have the same structural features as for $GL(2)$. If so and if there is a procedure for passing rigorously to the limit in the sum over $p < X$, either one in the spirit of the remarks in Part III or some quite different method, then we can continue to hope that the constructs of this paper have some general validity.

There are several factors in the sum: the volume μ_γ of $G_\gamma(\mathbb{Q})\backslash G_\gamma(\mathbb{A})$; the orbital integral at infinity, a function of $a_1, \ldots, a_k$ and the analogue of ψ; the orbital integrals at the finite number of finite primes in S that give congruence conditions and conditions on the denominators; the orbital integrals at the primes outside of S. These latter account for the contribution

(96) $$\sum_{f|s} f \prod_{q|f} \left(1 - \frac{\left(\frac{D}{q}\right)}{q}\right)$$

of (59).

The usual calculations of the volume of $T(\mathbb{Q})\backslash T(\mathbb{A})$ (see Ono's appendix to [W]) show that it is expressible as the value of an L-function at $s = 1$ so that it will be given by an expression similar to (61).

There will be changes. In particular, the L-function will be a product of nonabelian Artin L-functions. For $GL(k)$ the Kronecker symbols $\left(\frac{D}{n}\right)$ will be replaced by an expression determined by the behavior of $x^k + a_1x^{k-1} + \ldots \equiv 0$ in the local fields defined by this equation and associated to the primes dividing n. This behavior is periodic in $a_1, \ldots, a_k$ with period given by some bounded power of the primes dividing n, so that the nature of the contribution of μ_γ to the numerical analysis appears to be unchanged. For other groups the relation between the coefficients $a_1, a_2, \ldots$ and the stable conjugacy class will be less simple, but the principle is the same.

The contribution of the orbital integrals for places outside S will not be so simple as that given by Lemma 1. It has still to be examined, but it will have similar features. Lemma 1 expresses, among other things, a simple form of the Shalika germ expansion, and it may very well be that this structural feature of orbital integrals will be pertinent to the general analysis. It is reassuring for those who have struggled with the fundamental lemma and other aspects of orbital integrals to see that the arithmetic structure of the orbital integrals of functions in the Hecke algebra, especially of the unit element, may have an even deeper signifance than yet appreciated.

It remains, however, to be seen whether anything serious along these lines can be accomplished!

−0.686858,	−0.010181,	−0.232848,	−0.181406,	−0.207348,	−0.213738,
0.186493,	−0.509315,	−0.143169,	−0.267028,	−0.160988,	−0.214750,
−0.291132,	−0.231973,	−0.268148,	−0.267438,	−0.226005,	−0.252182,
−0.199118,	−0.079383,	−0.183245,	−0.132645,	−0.202296,	−0.171024,
−1.025438,	−0.527494,	−0.233874,	−0.247813,	−0.248718,	−0.231158,
0.017161,	−0.245755,	−0.271121,	−0.200476,	−0.211613,	−0.186097,
0.057466,	−0.073425,	−0.243888,	−0.170394,	−0.227328,	−0.194990,
−0.604006,	−0.449603,	−0.106058,	−0.211854,	−0.123694,	−0.198008,
−0.147666,	−0.198848,	−0.244547,	−0.154175,	−0.267187,	−0.169350,
−0.232995,	−0.460777,	−0.199048,	−0.265850,	−0.163095,	−0.238796,
−0.352068,	0.088846,	−0.154301,	−0.183977,	−0.186853,	−0.186285,
−0.183918,	−0.399292,	−0.319741,	−0.250550,	−0.273583,	−0.235420,
−0.218394,	−0.137239,	−0.112422,	−0.158514,	−0.140592,	−0.170782,
−0.331184,	−0.328770,	−0.223462,	−0.234503,	−0.237538,	−0.201874,
−0.330528,	−0.277603,	−0.236737,	−0.227831,	−0.191458,	−0.218459,
−0.107266,	−0.126031,	−0.195398,	−0.185583,	−0.213405,	−0.202951,
−0.138815,	−0.188041,	−0.211230,	−0.181449,	−0.197189,	−0.192290,
−0.388114,	−0.267641,	−0.236432,	−0.223636,	−0.239041,	−0.204733,
−0.285824,	−0.179338,	−0.159782,	−0.179327,	−0.176213,	−0.194733,
−0.147042,	−0.182515,	−0.213875,	−0.195378,	−0.207439,	−0.192548,
−0.137437,	−0.008915,	−0.225177,	−0.013033,	−0.211749,	−0.000857,
−0.413068,	−0.056893,	−0.220921,	0.007333,	−0.201721,	0.003851,
−0.080076,	−0.004867,	−0.169043,	−0.005863,	−0.182254,	−0.007391,
−0.270411,	−0.037313,	−0.235472,	−0.020661,	−0.224686,	−0.006618,
−0.282038,	−0.001461,	−0.188183,	0.019859,	−0.193986,	0.005356,
−0.232331,	−0.095297,	−0.217932,	−0.011204,	−0.224461,	0.004303,
−0.125913,	0.028871,	−0.208961,	−0.020619,	−0.194989,	−0.011606,
−0.238424,	−0.026239,	−0.177729,	0.004167,	−0.197464,	−0.006211,
−0.175674,	−0.020565,	−0.215916,	0.008770,	−0.192947,	−0.007468,
−0.249100,	0.015408,	−0.184689,	−0.009332,	−0.198297,	0.006565.

Table 3.1

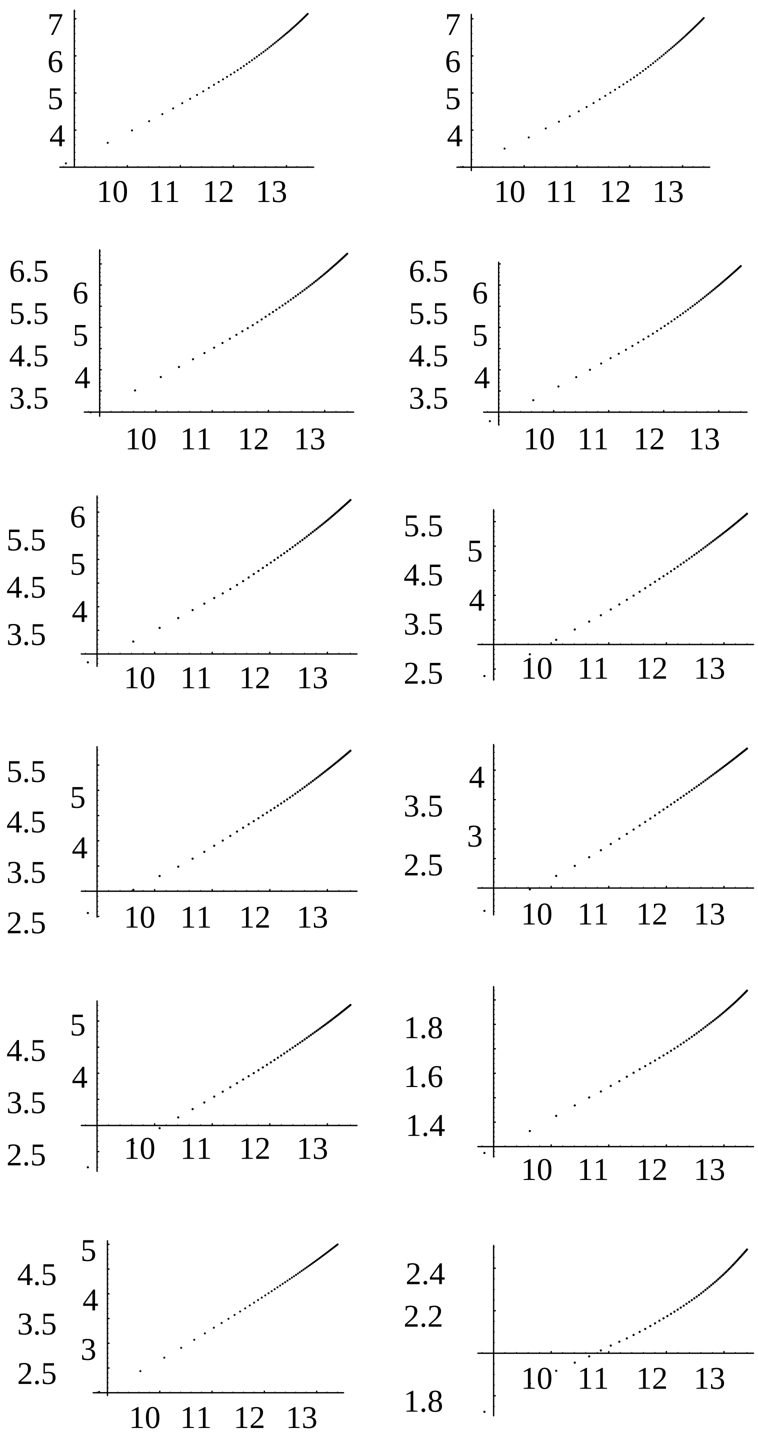
7 6 5 4
10 11 12 13
7 6 5 4
10 11 12 13
6.5 5.5 4.5 3.5 6 5 4
10 11 12 13
6.5 5.5 4.5 3.5 6 5 4
10 11 12 13
5.5 4.5 3.5 6 5 4
10 11 12 13
5.5 4.5 3.5 2.5 5 4
10 11 12 13
5.5 4.5 3.5 2.5 5 4
10 11 12 13
3.5 2.5 4 3
10 11 12 13
4.5 3.5 2.5 5 4
10 11 12 13
1.8 1.6 1.4
10 11 12 13
4.5 3.5 2.5 5 4 3
10 11 12 13
2.4 2.2 1.8
10 11 12 13

Diagram 3.3.A

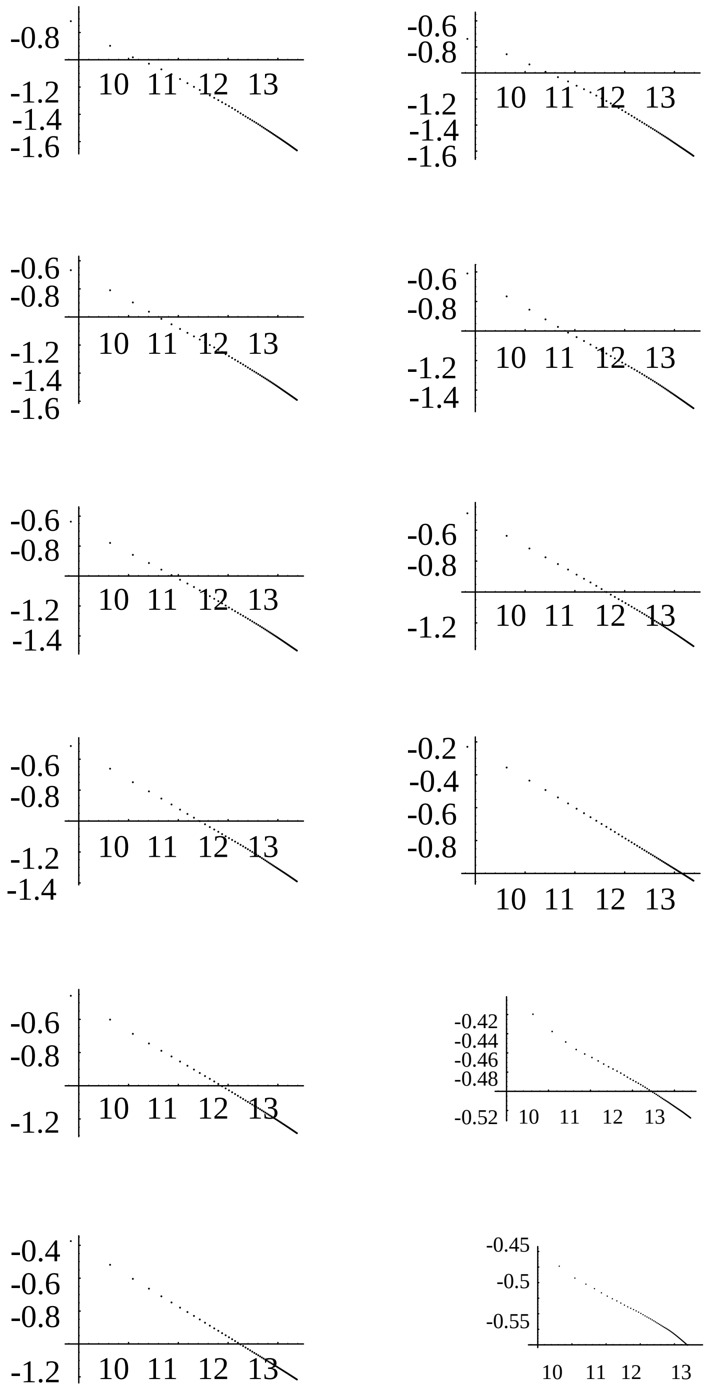
-0.8 -1.2 -1.4 -1.6
10 11 12 13
-0.6 -0.8 -1.2 -1.4 -1.6
10 11 12 13
-0.6 -0.8 -1.2 -1.4 -1.6
10 11 12 13
-0.6 -0.8 -1.2 -1.4
10 11 12 13
-0.6 -0.8 -1.2 -1.4
10 11 12 13
-0.6 -0.8 -1.2
10 11 12 13
-0.6 -0.8 -1.2 -1.4
10 11 12 13
-0.2 -0.4 -0.6 -0.8
10 11 12 13
-0.6 -0.8 -1.2
10 11 12 13
-0.42 -0.44 -0.46 -0.48 -0.52
10 11 12 13
-0.4 -0.6 -0.8 -1.2
10 11 12 13
-0.45 -0.5 -0.55
10 11 12 13

Diagram 3.3.B

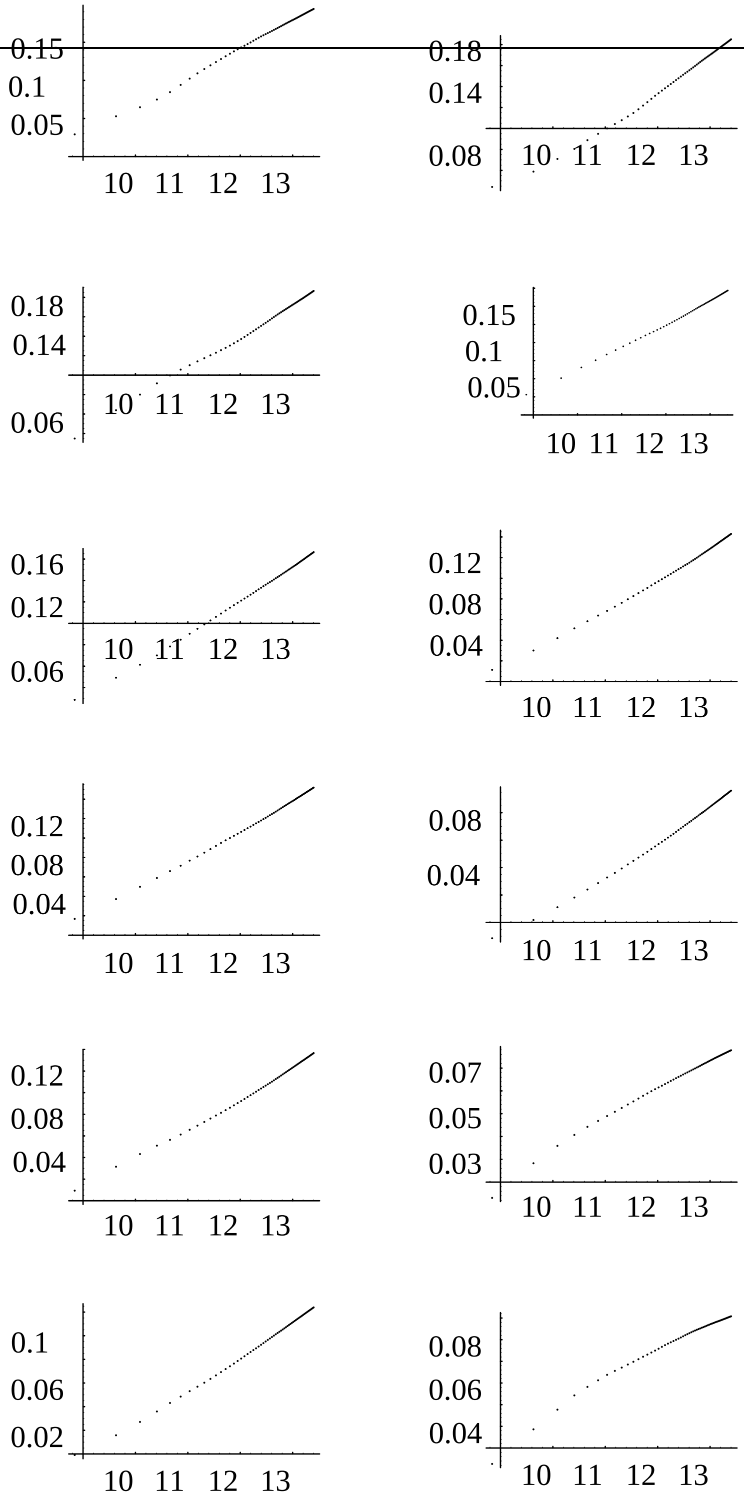
0.15
0.1
0.05
10 11 12 13
0.18
0.14
0.08
10 11 12 13
0.18
0.14
0.06
10 11 12 13
0.15
0.1
0.05
10 11 12 13
0.16
0.12
0.06
10 11 12 13
0.12
0.08
0.04
10 11 12 13
0.12
0.08
0.04
10 11 12 13
0.08
0.04
10 11 12 13
0.12
0.08
0.04
10 11 12 13
0.07
0.05
0.03
10 11 12 13
0.1
0.06
0.02
10 11 12 13
0.08
0.06
0.04
10 11 12 13

Diagram 3.3.C

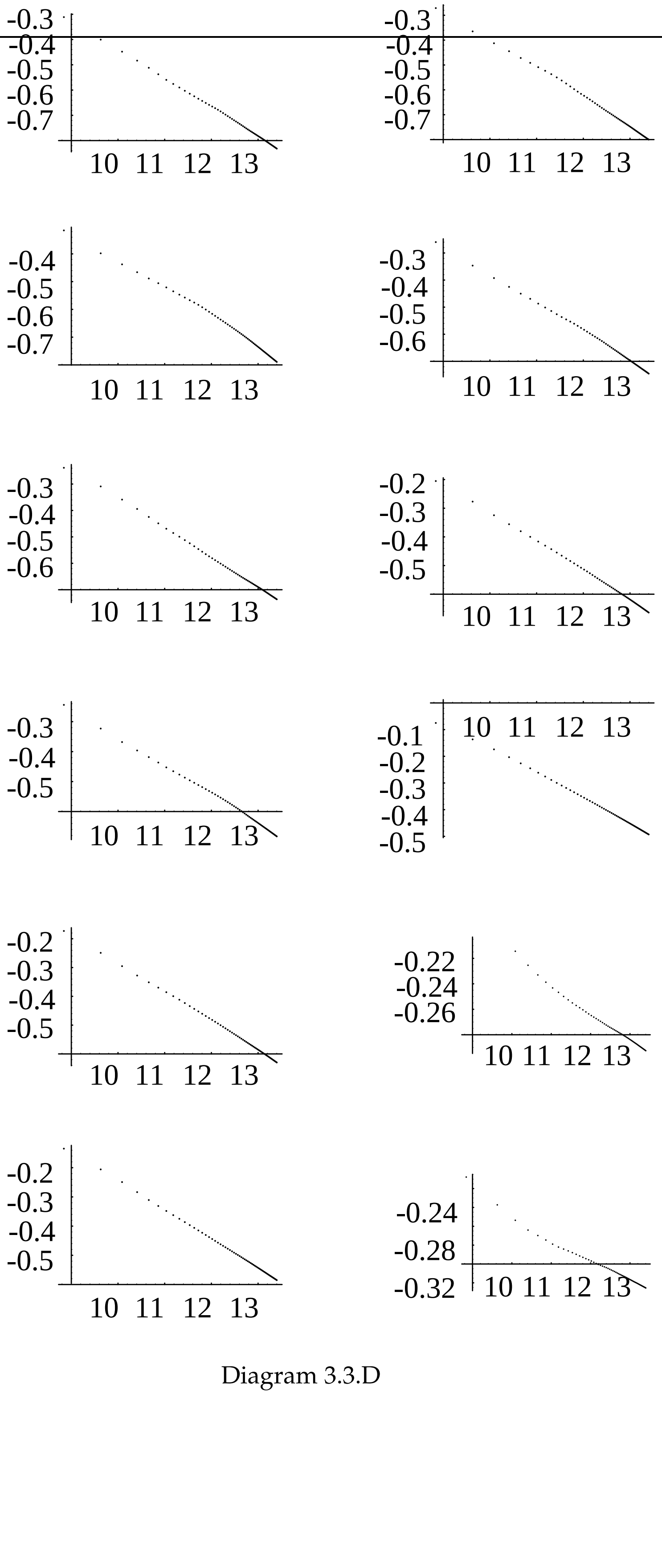
-0.3 -0.4 -0.5 -0.6 -0.7
10 11 12 13
-0.3 -0.4 -0.5 -0.6 -0.7
10 11 12 13
-0.4 -0.5 -0.6 -0.7
10 11 12 13
-0.3 -0.4 -0.5 -0.6
10 11 12 13
-0.3 -0.4 -0.5 -0.6
10 11 12 13
-0.2 -0.3 -0.4 -0.5
10 11 12 13
-0.3 -0.4 -0.5
10 11 12 13
10 11 12 13
-0.1 -0.2 -0.3 -0.4 -0.5
-0.2 -0.3 -0.4 -0.5
10 11 12 13
-0.22 -0.24 -0.26
10 11 12 13
-0.2 -0.3 -0.4 -0.5
10 11 12 13
-0.24 -0.28 -0.32
10 11 12 13

Diagram 3.3.D

Appendix A: Calculation of $\epsilon_{n,f}(N)$.

Both n and f are products of prime powers, $n = \prod q^a$ and $f = \prod q^b$. Thanks to the Chinese remainder theorem,

$$\epsilon_{n,f}(N) = \prod_q \epsilon_{q^a,q^b}(N).$$

It will suffice to show that

$$\text{(A.1)}\qquad \begin{aligned} \sum_{a,b=0}^{\infty} \epsilon_{q^a,q^b}(N) &= 1, && q \neq p, \\ &= \tfrac{1}{1-p^{-1}} + O(|N|^{-1/2}), && q = p. \end{aligned}$$

When q is fixed, we set for brevity $\epsilon_{q^a,q^b}(N) = \Lambda_{a,b}$. It will be more convenient to define $\Lambda_{a,b,c}$, $c \geq b$, as the product of the average value of

$$\left(\frac{(r^2-N)q^{2b}/q^{2c}}{q^{2a}}\right)$$

on the set of r for which q^{2c} is the highest even power of q dividing $r^2 - N$ with a remainder congruent to 0 or 1 modulo 4 with the density of the set, and to calculate $\Lambda_{a,b}$ as

$$(A.2)\qquad \sum_{c\,\geq\,b} \frac{q^b}{q^{a+c}} \Lambda_{a,b,c}.$$

That the $\Lambda_{a,b,c}$ are at least as natural to calculate as the $\Lambda_{a,b}$ suggests that rather than expressing the elliptic term as a sum over f and n as in the experiments to be described, one might want to express it as a sum over f, n and s. This would mean that a,$d = c - b$ and c were as good a choice of parameters as a, b and c, or that (71) could be replaced by

$$(A.3)\qquad \frac{1}{gn}\left(\frac{(r^2-N)/g^2}{n}\right), \quad g = \frac{s}{f}.$$

A direct analytic attack on the problems leads to (A.3) and not to (71).

Suppose first that q is odd and not equal to p. Then N is prime to q. If t is a high power of q, then the density of r modulo t such that $r^2 - N$ is divisible by q^c is

$$(1 + \left(\frac{N}{q}\right))q^{-c}.$$

if $c > 0$. Thus the density μ_c of r such that it is divisible by q^{2c} and not by q^{2c+2} is

$$
\text{(A.4)}\qquad
\begin{aligned}
&1 - \frac{1}{q^2}\Big(1 + \Big(\frac{N}{q}\Big)\Big), && c = 0\\
&\frac{1}{q^{2c}}\Big(1 + \Big(\frac{N}{q}\Big)\Big)\Big(1 - \frac{1}{q^2}\Big), && c > 0
\end{aligned}
$$

For positive even a it is the density ν_c of r such that $r^2 - N$ is divisible by q^{2c} and not by q^{2c+1} that is pertinent. This is

$$
\text{(A.5)}\qquad
\begin{aligned}
&1 - \frac{1}{q}\Big(1 + \Big(\frac{N}{q}\Big)\Big), && c = 0\\
&\frac{1}{q^{2c}}\Big(1 + \Big(\frac{N}{q}\Big)\Big)\Big(1 - \frac{1}{q}\Big), && c > 0
\end{aligned}
$$

When $c > 0$, if $r^2 = N + uq^c$, then

$$(r + vq^c)^2 \equiv N + (u + 2v)q^c \pmod{q^{c+1}}.$$

Thus, the average value of

$$\left(\frac{(r^2 - N)/q^{2c}}{q^a}\right)$$

on those r for which $r^2 - N$ is divisible by q^{2c} and not by q^{2c+2} is 0 if a is odd. For $c = 0$ and a odd, we have a simple lemma that shows that the average is $-1/q$.

Lemma 2 *The sum A of*

$$\left(\frac{r^2 - N}{q}\right)$$

over r modulo q is -1.

Since the number of solutions of

$$(A.6)\qquad y^2 = x^2 - N$$

for a given value of x is

$$\left(\frac{x^2 - N}{q}\right) + 1,$$

the number $A + q$ is just the number of points on the rational curve (A.6) modulo q whose coordinates are finite. The lemma follows.

The value of all $\Lambda_{a,b,c}$ and $\Lambda_{a,b}$ can now be calculated. For $a = b = 0$, $\Lambda_{0,0,c} = \mu_c$ and

$$\begin{aligned}
\Lambda_{0,0} &= \sum \frac{1}{q^c}\mu_c \\
&= 1 - \frac{1}{q^2}(1 + \left(\frac{N}{q}\right)) + \sum_{c=1}^{\infty} \frac{1}{q^{3c}}(1 + \left(\frac{N}{q}\right))(1 - \frac{1}{q^2}) \\
&= 1 - (1 + \left(\frac{N}{q}\right))\{\frac{1}{q^2} - \frac{1}{q^2}\frac{q^2-1}{q^3-1}\} \\
&= 1 - (1 + \left(\frac{N}{q}\right))\frac{q-1}{q^3-1}
\end{aligned}$$

If $a > 0$ and $b > 0$ then $\Lambda_{a,b,c} = 0$ and $\Lambda_{a,b} = 0$. For $b > 0$,

$$\begin{aligned}
\Lambda_{0,b} &= \sum_{c=b}^{\infty} \frac{q^b}{q^c}\Lambda_{0,b,c} \\
&= \sum_{c=b}^{\infty} \frac{q^b}{q^c}\mu_c \\
&= (1 - \frac{1}{q^2})(1 + \left(\frac{N}{q}\right))\sum_{c=b}^{\infty} \frac{q^b}{q^{3c}} \\
&= (1 - \frac{1}{q^2})(1 + \left(\frac{N}{q}\right))\frac{1}{q^{2b}}\frac{1}{1-q^{-3}}.
\end{aligned}$$

Thus

$$(A.7) \qquad \sum_{b=1}^{\infty} \Lambda_{0,b} = \frac{1}{q^2}(1 + \left(\frac{N}{q}\right))\frac{1}{1-q^{-3}} = \frac{q}{q^3-1}(1 + \left(\frac{N}{q}\right))$$

If $a > 0$ is even,

$$\Lambda_{a,0} = \frac{1}{q^a}\sum_{c=0}^{\infty} \frac{\nu_c}{q^c} = \frac{1}{q^a}\{1 - \frac{1}{q}(1 + \left(\frac{N}{q}\right)) + \frac{1}{q^3-1}(1 + \left(\frac{N}{q}\right))(1 - \frac{1}{q})\}.$$

The sum of this over all positive even integers is

$$\begin{aligned}
\sum_{a=1}^{\infty} \Lambda_{2a,0} &= \frac{1}{q^2-1}\{1 - \frac{1}{q}(1 + \left(\frac{N}{q}\right)) + \frac{1}{q^3-1}(1 + \left(\frac{N}{q}\right))(1 - \frac{1}{q})\} \\
&= \frac{1}{q^2-1} - \frac{1}{q^3-1}(1 + \left(\frac{N}{q}\right))
\end{aligned}.$$

If a is odd

$$\Lambda_{a,0} = -\frac{1}{q^{a+1}}.$$

Thus
$$\sum_{a=0}^{\infty}\Lambda_{2a+1}=-\frac{1}{q^2}\frac{1}{1-\frac{1}{q^2}}=-\frac{1}{q^2-1}.$$

Examining the previous calculations, we conclude that

$$\sum_{a,b=0}^{\infty}\Lambda_{a,b}=1$$

We now consider $q=2\neq p$, calculating first of all for each r the highest even power 2^{2c} of 2 that divides r^2-N with a remainder congruent to 0 or 1 modulo 4. We begin by observing that 4 divides r^2-N if and only if $r=2t$ is even and then

$$\frac{r^2-N}{4}=t^2-M,\quad M=\pm p^m,$$

which is congruent to 0 modulo 4 if and only if t is odd and $\left(\frac{-1}{M}\right)=1$ and to 1 if and only if t is even and $\left(\frac{-1}{M}\right)=-1$. In the first of these two cases, $c>0$; in the second $c=1$. Otherwise $c=0$.

There are thus two ways in which c can be 0. Either r is odd or r is even. Since r^2-N is odd if and only r is odd and is then congruent to $1-N$ modulo 8,

$$\begin{aligned}
\Lambda_{0,0,0}&=\frac{1}{2}+\frac{1}{4}=\frac{3}{4},\\
\Lambda_{a,0,0}&=\frac{1}{2}\left(\frac{1-N}{2}\right)=\frac{1}{2}\left(\frac{5}{2}\right)=-\frac{1}{2},\quad && a>0,\quad a\text{ odd},\\
\Lambda_{a,0,0}&=\frac{1}{2}, && a>0,\quad a\text{ even}.
\end{aligned}$$

If $\left(\frac{-1}{M}\right)=-1$ and $c>0$, then c is necessarily 1. Thus for such M,

$$\Lambda_{a,b,c}=0,\quad c>1.$$

Moreover, recalling that the Kronecker symbol $\left(\frac{n}{2}\right)$ is 0 for n even, 1 for $n\equiv 1,7\pmod 8$, and -1 for $n\equiv 3,5\pmod 8$ and that t^2-M is odd only for t even and then takes on the values $-M, 4-M$ modulo 8 with equal frequency, we see that for the same M,

$$\begin{aligned}
\Lambda_{0,b,1}&=\frac{1}{4}=\frac{1}{4}-\frac{1}{16}(1+\left(\frac{-1}{M}\right)),\quad && 1\geq b\geq 0,\\
\Lambda_{a,0,1}&=0, && a>0,\quad a\text{ odd},\\
\Lambda_{a,0,1}&=\frac{1}{4}, && a>0,\quad a\text{ even}.
\end{aligned}$$

Now suppose that $\left(\frac{-1}{M}\right) = 1$ and $c > 0$. Then, as observed, 4 divides $t^2 - M$ if and only if $t = 2u + 1$. The integer $u^2 + u$ is necessarily even and for any even v and any $d \geq 2$, $u^2 + u \equiv v \pmod{2^d}$ has exactly two solutions modulo 2^d. In particular, $u \equiv 0, 1, 2, 3 \pmod 4$ yield respectively $u^2 + u \equiv 0, 2, 2, 0 \pmod 4$. Since

$$\frac{t^2 - M}{4} = u^2 + u + \frac{1 - M}{4}, \tag{A.8}$$

and we conclude that $c = 1$ for $1/2$ of the possible values of u and $c = 2$ for the other half when $M \equiv 5 \pmod 8$. Thus, in this case,

$$\begin{array}{ll}
\Lambda_{0,b,1} = \frac{1}{8} = \frac{1}{4} - \frac{1}{16}(1 + \left(\frac{-1}{M}\right)), & 1 \geq b \geq 0, \\
\Lambda_{0,b,2} = \frac{1}{8} = \frac{1}{32}(1 + \left(\frac{-1}{M}\right))(1 - \left(\frac{M}{2}\right)), & 2 \geq b \geq 0, \\
\Lambda_{a,0,1} = 0, \quad \Lambda_{a,0,2} = 0, & a > 0, \quad a \text{ odd}, \\
\Lambda_{a,0,1} = 0, \quad \Lambda_{a,0,2} = \frac{1}{8}. & a > 0, \quad a \text{ even}
\end{array}$$

These numbers are to be incorporated with the factor

$$\frac{1}{4}(1 + \left(\frac{-1}{M}\right))(1 - \left(\frac{M}{2}\right))$$

in so far as it is not already present.

For $M \equiv 1 \pmod 8$, (A.8) can be any even number and the density of u for which it can be divided by 2^{2d}, $d \geq 0$, to give a number congruent to $0, 1$ modulo 4 is $1/2$ if $d = 0$ and $1/2^{2d}$ if $d > 0$. Since d will be $c - 2$, this is $1/2^{2c-4}$. On the other hand, the density is multiplied by $1/4$ when we pass from u to r, so that

$$\begin{array}{ll}
\Lambda_{0,b,1} = \dfrac{1}{8} = \dfrac{1}{4} - \dfrac{1}{16}(1 + \left(\dfrac{-1}{M}\right)), & 1 \geq b \geq 0, \\
\Lambda_{a,0,1} = \Lambda_{a,0,2} = 0, & a > 0, \quad a \text{ odd}, \\
\Lambda_{a,0,1} = \Lambda_{a,0,2} = 0, & a > 0, \quad a \text{ even}, \\
\Lambda_{0,b,2} = \dfrac{1}{16}, & 2 \geq b \geq 0, \\
\Lambda_{0,b,c} = \dfrac{1}{2^{2c-2}}(1 - \dfrac{1}{4}), & c \geq b \geq 0, \quad c > 2, \\
\Lambda_{a,0,c} = 0, & a > 0, \quad a \text{ odd}, \quad c > 2, \\
\Lambda_{a,0,c} = \dfrac{1}{2^{2c-1}}, & a > 0, \quad a \text{ even}, \quad c > 2.
\end{array}$$

These numbers are to be incorporated with the factor

$$\frac{1}{4}(1 + \left(\frac{-1}{M}\right))(1 + \left(\frac{M}{2}\right)).$$

Then $\Lambda_{0,0}$ is the sum of

$$\frac{7}{8}-\frac{1}{32}(1+\left(\frac{-1}{M}\right))+\frac{1}{2^7}(1+\left(\frac{-1}{M}\right))(1-\left(\frac{M}{2}\right)) \tag{A.9$'$}$$

and

$$\{\frac{1}{2^8}+\frac{3}{2^4}\sum_{c=3}^{\infty}\frac{1}{2^{3c-2}}\}(1+\left(\frac{-1}{M}\right))(1+\left(\frac{M}{2}\right))$$

or

$$\Lambda'_{0,0}=\frac{1}{2^8}(1+\frac{3}{7})(1+\left(\frac{-1}{M}\right))(1+\left(\frac{M}{2}\right)) \tag{A.9$''$}$$

For $b>0$,

$$\Lambda_{0,b}=\sum_{c=b}^{\infty}\frac{2^b}{2^c}\Lambda_{0,b,c}.$$

Thus $\Lambda_{0,1}$ is

$$\frac{1}{4}-\frac{1}{16}(1+\left(\frac{-1}{M}\right))+\frac{1}{64}(1+\left(\frac{-1}{M}\right))(1-\left(\frac{M}{2}\right))+\frac{1}{2^7}(1+\frac{3}{7})(1+\left(\frac{-1}{M}\right))(1+\left(\frac{M}{2}\right)),$$

while

$$\Lambda_{0,2}=\frac{1}{32}(1+\left(\frac{-1}{M}\right))(1-\left(\frac{M}{2}\right))+\frac{1}{2^6}(1+\frac{3}{7})(1+\left(\frac{-1}{M}\right))(1+\left(\frac{M}{2}\right))$$

and

$$\Lambda_{0,b}=\frac{3}{7}\frac{1}{2^{2b-1}}(1+\left(\frac{-1}{M}\right))(1+\left(\frac{M}{2}\right)),\qquad b>2,$$

because

$$\frac{3}{16}\sum_{c=b}^{\infty}\frac{2^b}{2^{3c-2}}=\frac{3}{7}\frac{1}{2^{2b-1}}.$$

Thus

$$\sum_{b>2}\Lambda_{0,b}=\frac{1}{2^3}\frac{1}{7}(1+\left(\frac{-1}{M}\right))(1+\left(\frac{M}{2}\right))$$

and

$$\Lambda'_{0,0}+\sum_{b>0}\Lambda_{0,b}$$

is equal to the sum of

$$\frac{1}{4}-\frac{1}{16}(1+\left(\frac{-1}{M}\right)) \tag{A.10$'$}$$

and

$$\frac{3}{64}(1+\left(\frac{-1}{M}\right))(1-\left(\frac{M}{2}\right)) \tag{A.10$''$}$$

and

$$(\mathrm{A}.10''') \qquad \{\frac{5}{2^7}+\frac{1}{7}\frac{1}{2^3}\}(1+\left(\frac{-1}{M}\right))(1+\left(\frac{M}{2}\right)),$$

because

$$\frac{5}{7}(\frac{1}{2^7}+\frac{1}{2^6}+\frac{1}{2^5})+\frac{1}{7}\frac{1}{2^3}=\frac{5}{2^7}+\frac{1}{7}\frac{1}{2^3}.$$

Finally

$$\Lambda_{a,0}=\sum_{c=0}^{\infty}\frac{1}{2^{a+c}}\Lambda_{a,0,c}, \quad a>0.$$

I express it as a sum of three terms, the first of which is

$$\Lambda'_{a,0}=-\frac{1}{2^{a+1}}, \quad a \text{ odd},$$

or

$$\Lambda'_{a,0}=\frac{1}{2^{a+1}}+\frac{1}{2^{a+4}}(1-\left(\frac{-1}{M}\right)), \quad a \text{ even}.$$

The other two, $\Lambda''_{a,0}$ and $\Lambda'''_{a,0}$, will be multiples of $(1+\left(\frac{-1}{M}\right))(1+\left(\frac{M}{2}\right))$ and $(1+\left(\frac{-1}{M}\right))(1-\left(\frac{M}{2}\right))$ respectively. Observe that

$$(A.11) \qquad \begin{aligned}\sum_{a=1}^{\infty}\Lambda'_{a,0}&=-\frac{1}{3}+\frac{1}{6}+\frac{1}{48}(1-\left(\frac{-1}{M}\right))\\&=-\frac{1}{16}(1-\left(\frac{-1}{M}\right))-\frac{1}{12}(1+\left(\frac{-1}{M}\right)).\end{aligned}$$

Since

$$(A.12) \qquad \frac{1}{8}=\frac{1}{16}(1+\left(\frac{-1}{M}\right))+\frac{1}{16}(1-\left(\frac{-1}{M}\right)),$$

we can conclude at least that $\sum\Lambda_{a,b}-1$ is a multiple of $(1+\left(\frac{-1}{M}\right))$.

The terms that involve $(1+\left(\frac{-1}{M}\right))$ alone without a second factor $(1\pm\left(\frac{M}{2}\right))$ come from (A.9′), (A.10′), (A.11) and (A.12).

$$(A.13) \qquad \{-\frac{1}{32}-\frac{1}{16}+\frac{1}{16}-\frac{1}{12}\}(1+\left(\frac{-1}{M}\right)).$$

Since all other terms involve the second factor, I multiply (A.13) by

$$\frac{1}{2}(1+\left(\frac{M}{2}\right))+\frac{1}{2}(1-\left(\frac{M}{2}\right)).$$

To establish the first equality of (A.1) for $q = 2$, we have to show that the coefficients of the two expressions $(1 + \left(\frac{-1}{M}\right))(1 \pm \left(\frac{M}{2}\right))$ add up to 0.

The remaining terms that involve the product $(1 + \left(\frac{-1}{M}\right))(1 - \left(\frac{M}{2}\right))$ come from (A.9′), (A.10″) and

$$\sum_{a>0} \frac{1}{2^{a+2}} \Lambda''_{a,0} = \sum_{a>0} \frac{1}{2^{2a+7}} (1 + \left(\frac{-1}{M}\right))(1 - \left(\frac{M}{2}\right))$$
$$= \frac{1}{2^7}\frac{1}{3}(1 + \left(\frac{-1}{M}\right))(1 - \left(\frac{M}{2}\right)).$$

They multiply it by the factor

$$\frac{1}{2^7} + \frac{3}{64} + \frac{1}{2^7}\frac{1}{3}.$$

The sum of this factor and $1/2$ of that of (A.13) is

$$\frac{1}{32} - \frac{1}{24} + \frac{1}{2^5}\frac{1}{3} = 0.$$

Since

$$\sum_{a>0} \Lambda'''_{a,0} = \frac{1}{4}(1 + \left(\frac{-1}{M}\right))(1 + \left(\frac{M}{2}\right)) \sum_{a=2d>0} \sum_{c=3}^{\infty} \frac{1}{2^{a+3c-1}}$$
$$= \frac{1}{2^7}\frac{1}{3}\frac{1}{7}(1 + \left(\frac{-1}{M}\right))(1 + \left(\frac{M}{2}\right)),$$

the terms involving the factor $(1 + \left(\frac{-1}{M}\right))(1 + \left(\frac{M}{2}\right))$ yield

$$(A.14) \qquad \{\frac{5}{2^7} + \frac{1}{7}\frac{1}{2^3} + \frac{1}{2^7}\frac{1}{3}\frac{1}{7}\}(1 + \left(\frac{-1}{M}\right)(1 + \left(\frac{M}{2}\right) = \frac{11}{2^6 3}(1 + \left(\frac{-1}{M}\right)(1 + \left(\frac{M}{2}\right)$$

Since

$$-\frac{1}{32} - \frac{1}{12} = -\frac{11}{2^5 3},$$

the term (A.14) cancels the contribution from (A.13).

I treat the second equality of (A.1) only for q odd as this suffices for our purposes. We calculate $\Lambda_{a,b}$ using (A.2). Since

$$\sum_{a\,\geq\,0} \sum_{c\,\geq\,b} \sum_{c\,\geq\,m/2} \frac{q^b}{q^{a+c}} O(\frac{1}{q^c}) = \{\sum_{a\,\geq\,0} \sum_{d\,\geq\,0} \frac{1}{q^{a+d}}\} \sum_{c\,\geq\,m/2} O(\frac{1}{q^c}),$$

we need not use the exact value of $\Lambda_{a,b,c}$ for $2c \geq m$. We need only approximate it uniformly within $O(\frac{1}{q^c})$.

For $2c < m$, the density of r for which $r^2 - N$ is exactly divisible by q^{2c} is $(1-1/q)/q^c$. For $2c \geq m$, it is $O(1/q^c)$. Thus, as an approximation,

$$\Lambda_{0,0} \sim (1-\frac{1}{q})\sum_{c=0}^{\infty}\frac{1}{q^{2c}} = \frac{q}{q+1}.$$

Moreover, again as an approximation,

$$\Lambda_{0,b} \sim (1-\frac{1}{q})\sum_{c=b}^{\infty}\frac{q^b}{q^{2c}} = \frac{q}{q+1}\frac{1}{q^b}, \quad b>0,$$

so that

$$\sum_{b>0}\Lambda_{0,b} \sim \frac{q}{q+1}\frac{1}{q-1}.$$

If $2c < m$ and $r = q^c t$, $(q,t) = 1$, then

$$\frac{r^2-N}{q^2 c} \equiv t^2 \pmod q$$

and

$$\left(\frac{(r^2-N)/q^{2c}}{q}\right) = 1.$$

Thus, the approximation is

$$\Lambda_{a,0} \sim (1-\frac{1}{q})\sum_{c=0}^{\infty}\frac{1}{q^{a+2c}} = \frac{1}{q^a}\frac{q}{q+1},$$

and

$$\sum_{a>0}\Lambda_{a,0} \sim \frac{q}{q+1}\frac{1}{q-1}.$$

Finally

$$\Lambda_{0,0} + \sum_{b>0}\Lambda_{0,b} + \sum_{a>0}\Lambda_{a,0} \sim \frac{1}{1-q^{-1}}.$$

Appendix B: Some estimates.

I collect here a few simple estimates needed in Section 3.2. They are provisional and made without any effort to search the literature. To simplify the notation, take N to be positive and $M = N^{1/4}$. If s is a positive integer, let $\#(s)$ be the number of distinct prime divisors of s.

Lemma B.1 *There is a constant* $c \geq 1$ *such that*

$$\sqrt{N}\sum_{s>M}\frac{2^{\#(s)}}{s^3} = O(\ln^{2c-1} N).$$

There is a chance that the constant c is 1. It is even very likely, but I make no effort to prove it here. The analysis would certainly be more difficult. For the lemma as stated it is sufficient to use the well-known Tchebychef estimate ([HW], p. 10) for the n-th prime number[18] $p(n) \asymp n\ln 2n$. I have used $n\ln 2n$ rather than $n\ln n$ only to avoid dividing by $\ln 1 = 0$. To verify the lemma with $c = 1$ would undoubtedly entail the use the prime number theorem, thus the asymptotic relation $p(n) \sim n/\ln 2n \sim n/\ln n$, and a different, more incisive treatment of the sums that appear.

Let $q(n)$ be the n^{th} element of the sequence of prime powers $\{2,3,4,5,7,8,9,\ldots\}$ and $\sigma(x)$ the number of prime powers less than x. I observe first that the Tchebychef estimate $\pi(x) \asymp x\ln x$ implies that $\sigma(x) \asymp x\ln x$ as well and thus that $q(n) \asymp n\ln n \asymp n\ln 2n$.

Indeed,

$$\sigma(x) = \pi(x) + \pi(x^{1/2}) + \ldots \pi(x^{1/D}) + O(1), \quad D = [\ln x],$$

and

$$\sum_{j=2}^{D}\pi(x^{1/j}) \leq C\int_{t=1}^{D}\frac{x^{1/t}}{\ln x^{1/t}}dt \leq C\int_{t=1}^{\ln x}\frac{x^{1/t}}{\ln x^{1/t}}dt,$$

because $y/\ln y$ is an increasing function for $y \geq e$. The integral is

$$\begin{aligned}\frac{1}{\ln x}\int_1^{\ln x} te^{\ln x/t}dt &= \frac{1}{\ln x}\int_{1/\ln x}^{1} e^{t\ln x}\frac{dt}{t^3}\\ &= \ln x\int_1^{\ln x} e^t\frac{dt}{t^3}\\ &\leq \ln x\int_1^{\ln x/2} e^t\frac{dt}{t^3} + \frac{8}{\ln^2 x}\int_{\ln x/2}^{\ln x} e^t dt = O\left(\frac{x}{\ln^2 x}\right)\end{aligned}$$

Thus $\sigma(x) \asymp \pi(x)$.

To prove the lemma we write s as $s = p_1^{a_1}\ldots p_l^{a_l}$, where all the primes $p_1,\ldots,p_l$, are different. At first, take $p_1 < p_2 < \ldots < p_l$. The expression of the lemma may be written as

$$\sqrt{N}\{\sum_{l>0}\sum_{p_1,p_2,\ldots,p_l}\frac{2^l}{p_1^{3a_1}\ldots p_l^{3a_l}}\}.$$

[18] Following [HW], I use the notation $p(n) \asymp n\ln 2n$ to mean that $C_1 n\ln 2n \leq p(n) \leq C_2 n\ln 2n$, with positive constants C_1 and C_2.

There is certainly a sequence $1' < \ldots < k'$, $k' \leq l$ such that $p_{1'}^{a_{1'}} \ldots p_{k'}^{a_{k'}} > M$ while $p_{1'}^{a_{1'}} \ldots \hat{p}_{i'}^{a_{i'}} \ldots p_{k'}^{a_{k'}} \leq M$ for any i', $1 \leq i' \leq k'$. The notation signifies that $p_{i'}^{a_{i'}}$ is removed from the product. Thus the expression of the lemma is bounded by

$$\sqrt{N}\{\sum_{k>0}\sum_{p_1,p_2,\ldots,p_k}\frac{2^k}{p_1^{3a_1}\ldots p_k^{3a_k}}(\sum_t \frac{2^{\#(t)}}{t^3})\},$$

where t is allowed to run over all integers prime to $p_1,\ldots,p_k$, but where $p_1 < \ldots < p_k$, $p_1^{a_1}\ldots p_k^{a_k} > M$, and $p_1^{a_1}\ldots\hat{p}_i^{a_i}\ldots p_k^{a_k} \leq M$.

I next allow $p_1,\ldots,p_k$ to appear in any order, so that I have to divide by $k!$. It is still the case, however, that $p_1^{a_1}\ldots p_k^{a_k} > M$ and that $p_1^{a_1}\ldots p_{k-1}^{a_{k-1}} \leq M$. Since

$$\sum_t \frac{2^{\#(t)}}{t^3} \leq \prod_p (1 + \frac{2}{p^3} + \frac{2}{p^6} + \ldots)$$

is finite, we may drop it from the expression and consider

$$(B.1) \qquad \sqrt{N}\sum_{k=1}^{\infty}\frac{2^k}{k!}\sum_{\substack{p_1^{a_1}\ldots p_k^{a_k}>M \\ p_1^{a_1}\ldots p_{k-1}^{a_{k-1}} \leq M}}\frac{1}{p_1^{3a_1}\ldots p_k^{3a_k}} = \sqrt{N}\sum_{k=1}^{\infty}\frac{2^k}{k!}\sum_{\substack{q_1\ldots q_k>M \\ q_1\ldots q_{k-1}\leq M}}\frac{1}{q_1^3\ldots q_k^3},$$

where $q_1,\ldots q_k$ are prime powers. It is this sum that is to be estimated. In it,

$$q_k > A = \frac{M}{q_1\ldots q_{k-1}} \geq 1.$$

So in general, as a first step, we need to estimate, for any $A \geq 1$,

$$(B.2) \qquad \sum_{q>A}\frac{1}{q^3}.$$

We apply the Tchebychef estimate. Thus, if C is taken to be an appropriate positive constant independent of A, (B.2) is majorized by a constant times

$$\begin{aligned}\sum_{n>C\frac{A}{\ln 2A}}\frac{1}{n^3\ln^3 2n} \leq \frac{1}{\ln^3(CA/\ln 2A)}\sum_{n>C\frac{A}{\ln 2A}}\frac{1}{n^3} \\ = O(\frac{1}{(A/\ln 2A)^2}\frac{1}{\ln^3(2A/\ln 2A)}) \\ = O(\frac{1}{A^2\ln 2A}).\end{aligned}$$

Although the argument itself is doubtful for small A, especially if C is also small, the conclusion is not.

As a result, (B.1) is bounded by

$$C\sqrt{N}\sum_{k=1}^{\infty}\frac{2^k}{k!}\sum_{q_1\dots q_{k-1}\,\le\,M}\frac{1}{q_1^3\dots q_{k-1}^3}\frac{q_1^2\dots q_{k-1}^2}{M^2}\frac{1}{\ln(2M/q_1\dots q_{k-1})},$$

with perhaps a new constant C. Since $M^2=\sqrt{N}$, this is

$$C\sum_k\frac{2^k}{k!}\sum_{q_1\dots q_{k-1}\,\le\,M}\frac{1}{q_1\dots q_{k-1}}\frac{1}{\ln(2M/q_1\dots q_{k-1})}. \tag{$B.3$}$$

To complete the proof of Lemma B.1, we shall use another lemma.

Lemma B.2. *If* $A\ge 1$, *then*

$$\sum_{q\,\le\,A}\frac{1}{q\ln(2A/q)}\le c\frac{\ln\ln A}{\ln 2A}, \tag{$B.4$}$$

the sum running over prime powers.

The constant of this lemma is the constant that appears in Lemma B.1. So it is Lemma B.2 that will have to be improved.

Before proving the lemma, we complete the proof of Lemma B.1. Set $A=M/p_1\dots p_{k-2}$. Then

$$\sum_{q_1\dots q_{k-1}\,\le\,M}\frac{1}{q_1\dots q_{k-1}}\frac{1}{\ln(2M/q_1\dots q_{k-1})} \tag{$B.5$}$$

may be rewritten as

$$\sum_{q_1\dots q_{k-2}\,\le\,M}\frac{1}{q_1\dots q_{k-2}}\sum_{q_{k-1}\,\le\,A}\frac{1}{q_{k-1}}\frac{1}{\ln(2A/q_{k-1})},$$

which, by Lemma B.2, is at most

$$c\sum_{q_1\dots q_{k-2}\,\le\,M}\frac{1}{q_1\dots q_{k-2}}\frac{\ln\ln A}{\ln 2A}\le c\ln\ln M\sum_{q_1\dots q_{k-2}\,\le\,M}\frac{1}{q_1\dots q_{k-2}}\frac{1}{\ln 2A}$$

It is clear that (B.5) is $O((c\ln\ln M)^{k-1}/\ln 2M)$ for $k=1$, and this estimate now follows readily by induction for all k uniformly in k. As a result (B.3) is

$$O(\sum_{k=1}^{\infty}\frac{2^k(c\ln\ln M)^{k-1}}{k!\ln M})=O(\frac{e^{2c\ln\ln M}}{\ln\ln M\ln M})=O(\ln^{2c-1}M),$$

where we have discarded a $\ln\ln M$ in the denominator that is of no help.

If we are willing to accept a very large constant c in (B.4), then we can replace $\ln 2A/q$ in the denominator by $\ln CA/\ln p$, where C is any given constant greater than 1 or by $CA/n\ln 2n$, if $q = q(n)$ is the nth prime power and C is chosen sufficiently large in comparison to the constant in the Tchebychef inequality. We can also replace the $p(n)$ in the denominator by $n\ln 2n$. Thus, at the cost of adding some terms, we may replace the sum (B.4) by

$$(B.6) \qquad \sum_{n\ln 2n\,\leq\, C'A} \frac{1}{n\ln 2n}\frac{1}{\ln(CA/n\ln 2n)}.$$

There is no harm in supposing that $C' = 1$, Clearly, we can demand in addition that the sum run over $n\ln 2n \geq C_1$, where C_1 is a fixed arbitrary constant, because the sum

$$\sum_{n\ln 2n\,\leq\, C_1} \frac{1}{n\ln 2n}\frac{1}{\ln(CA/n\ln 2n)}$$

is certainly $O(1/\ln A)$. Set

$$\frac{CA}{n\ln 2n} = A^{1-\alpha}, \qquad \alpha = e^{-a}.$$

If $C_1 \geq C$, $\alpha = \alpha(n) \geq 0$. Moreover, as we have agreed to exclude the initial terms of the original sum, $\alpha < 1$ and $a > 0$. If β is some fixed number less than 1, then

$$\sum_{\alpha(n)\,\leq\,\beta} \frac{1}{n\ln 2n}\frac{1}{\ln(CA/n\ln 2n)} \leq C_2\frac{1}{\ln 2A}\sum_{\alpha(n)\,\leq\,\beta}\frac{1}{n\ln 2n} \leq C_3\frac{\ln\ln A}{\ln 2A}.$$

So we may sum over $\alpha(n) > \beta$ or $a = a(n) < b$, $b = -\ln\beta$. We now confine ourselves to this range.

In addition $(1-\alpha)\ln A \geq \ln C$, so that $(1-\alpha) \geq C_4/\ln A$ and

$$a \geq C_5/\ln A$$

Let $\epsilon > 0$ and set $b(k) = b(1+\epsilon/\ln A)^{-k}$. I shall decompose the sum into sums over the intervals $b(k+1) \leq a(n) < b(k)$, for all those k such that $b(k+2) \geq C_5/\ln A$ and into one last interval $C_5/\ln A \leq a(n) < b(k)$, where k is the first integer such that $b(k+2) < C_5/\ln A$. I shall denote these intervals by I and use the Hardy-Wright notation to indicate uniformity with respect to I.

Notice first that

$$A^{\alpha(n+1)-\alpha(n)} = \frac{(n+1)\ln 2(n+1)}{n\ln 2n} = 1 + O(\frac{1}{n}) = 1 + O(\frac{1}{\ln 2n}).$$

Thus $\alpha(n+1) - \alpha(n) \leq C_6/\ln^2 A$ when $\alpha(n) > \beta$. As a result, on the same range $a(n) - a(n+1) \leq C_7/\ln^2 A$. Moreover

$$b(k) - b(k+1) \geq b(k)\frac{\epsilon}{\ln A} > \frac{C_5\epsilon}{\ln^2 A}.$$

Thus each of these intervals contains at least two terms of our sum provided that $C_5\epsilon > 2C_7$, as we assume. Moreover, if a' and a lie in the same interval, then $a'/a \asymp 1$ and $(1-\alpha')/(1-\alpha) \asymp 1$, so that

$$\frac{\ln(CA/n'\ln 2n')}{\ln(CA/2n\ln 2n)} \asymp 1,$$

where $n' = n(\alpha')$ and $n = n(\alpha)$ are not necessarily integers.

We conclude first of all that, for any point a_I in I,

$$\sum_{a(n)\in I} \frac{1}{n\ln 2n}\frac{1}{\ln(CA/n\ln 2n)} \asymp \frac{1}{a_I \ln 2A}\sum_{a(n)\in I}\frac{1}{n\ln 2n}$$

and that

$$\int_I \frac{1}{a}da \asymp \frac{1}{a_I}\int_I da.$$

So, if we can show that

$$(B.7) \qquad \sum_{a(n)\in I}\frac{1}{n\ln 2n} \asymp \int_I da$$

the lemma will follow, because

$$\sum_I \int_I \frac{1}{a}da = \int_{C_5/\ln A}^{b} \frac{1}{a}da = O(\ln\ln A).$$

Since

$$\frac{(n+1)\ln 2(n+1)}{n\ln 2n} = O((1+\frac{1}{n})^2),$$

the sum in (B.7) may be replaced by the integral with respect to dn from n_1 to n_2 if $a_2 = a(n_2-1)$ and $a_1 = a(n_1)$ are the first and last points in the interval associated to integers. The integral is equal to

$$\int_{n_1}^{n_2} \frac{1}{n\ln 2n}dn = \ln\ln 2n_2 - \ln\ln 2n_1.$$

We show that the right-hand side is equivalent in the sense of Hardy-Wright to $a_2 - a_1$ or, what is the same on the range in question, to $\alpha_1 - \alpha_2$. Thus all three are of comparable magnitudes uniformly in I. Since $a_2 - a_1$ is equivalent, again in the sense of Hardy-Wright, to the length of I, the relation (B.7) will follow.

Since $n\ln 2n = CA^\alpha$, $\ln n + \ln\ln 2n = \ln C + \alpha\ln A$,

$$\ln n + \ln\ln 2n = \ln n + (1 + \frac{\ln\ln 2n}{\ln n}),$$

and $\alpha = \alpha(n)$ is bounded below by $-\ln b$, we infer that $\ln n \asymp \ln A$. Moreover

$$(B.8)\qquad \begin{aligned}\ln\ln n + \ln\big(1 + \frac{\ln\ln n}{\ln 2n}\big) &= \ln(\alpha \ln A\big(1 + \frac{\ln C}{\alpha \ln A}\big))\\ &= \ln\alpha + \ln\ln A + \ln\big(1 + \frac{\ln C}{\alpha\ln A}\big).\end{aligned}$$

Since a difference between the values of a continuously differentiable function at two values of the argument is equal to the difference of the arguments times the derivative at some intermediate point,

$$(B.9)\qquad \ln(1 + \frac{\ln C}{\alpha_2 \ln A}) - \ln\big(1 + \frac{\ln C}{\alpha_1 \ln A}\big) = O(\frac{\ln C}{\ln A}(\alpha_1 - \alpha_2)) = O(\frac{1}{\ln A}(a_2 - a_1)).$$

The expression

$$\ln\big(1 + \frac{\ln\ln 2n}{\ln n}\big) = \ln\big(1 + \frac{\ln X}{X - \ln 2}\big), \qquad X = \ln 2n.$$

So the difference

$$\ln\big(1 + \frac{\ln_2 \ln 2n_2}{\ln n_2}\big) - \ln\big(1 + \frac{\ln_1 \ln 2n_1}{\ln n_1}\big) = O(\frac{\ln\ln A}{\ln^2 A}(\ln n_2 - \ln n_1))$$

Since

$$\ln\ln n_2 - \ln\ln n_1 \asymp \frac{1}{\ln A}(\ln n_2 - \ln n_1),$$

we conclude from (B.8) and (B.9)that

$$\ln\ln 2n_2 - \ln\ln 2n_1 \asymp \ln\ln n_2 - \ln\ln n_1 \asymp \alpha_2 - \alpha_1.$$

The next lemma is similar to Lemma B.1.

Lemma B.3. *There is a positive constant $c \geq 1$ such that for any positive constant C,*

$$\sum_{C\sqrt{N} > s > M} \frac{2^{\#(s)}}{s} = O(\ln^{2c} N).$$

It is again very likely that c may be taken equal to 1, but once again our proof will squander a good deal of the force even of the Tchebychef inequality.

I have stated the lemma in the way it will be used, but the constant C is clearly neither here nor there. Moreover, we prove the stronger statement

$$(B.10)\qquad \sum_{s \leq \sqrt{N}} \frac{2^{\#(s)}}{s} = O(\ln^{2c} N).$$

Thus the lower bound on s in the sum is unnecessary. We take $A = \sqrt{N}$ and write $s = p_1 \dots p_l t$, where t is prime to $p_1, \dots, p_l$ and where $p|t$ implies that $p^2|t$. So the left side of (B.10) is majorized by

$$(\sum_{l \geq 0} \sum_{p_1 \dots p_l < A} \frac{2^l}{p_1 \dots p_l}) \prod_p (1 + \frac{2}{p^2} + \frac{2}{p^3} + \dots).$$

The product is a constant factor and can be dropped for purposes of the estimation. So we are left with

$$(B.11) \qquad \sum_{l \geq 0} \sum_{p_1 \dots p_l < A} \frac{2^l}{p_1 \dots p_l} = \sum_{l \geq 0} \sum_{p_1 \dots p_l < A} \frac{2^l}{l!} \frac{1}{p_1 \dots p_l},$$

the difference between the left and the right sides being that the first is over $p_1 < \dots < p_l$, whereas in the second the primes are different but the order arbitrary.

It is clear that

$$\sum_{p<A} \frac{1}{p} = O(\sum_{n \ln 2n < CA} \frac{1}{n \ln 2n}) = O(\ln \ln (\frac{A}{\ln A})) = O(\ln \ln A).$$

Thus,

$$\sum_{p_1 \dots p_l < A} \frac{1}{p_1 \dots p_l} \leq (\sum_{p<A} \frac{1}{p})^l \leq (c \ln \ln A)^l,$$

uniformly in l. The estimate (B.10) follows from (B.11).

Applying Lemma B.3 with N replaced by $\sqrt{N}$ we obtain

Corollary B.4. *There is a constant $c \geq 1$ such that*

$$\frac{1}{\sqrt{N}} \sum_{s \leq M} 2^{\#(s)} s = O(\ln^{2c} N).$$

Appendix C: Weighted orbital integrals.

This is largely a matter of recollecting results from [H] and earlier papers, amply acknowledged in [H]. More must be said than would be necessary had the author, W. Hoffmann, not assumed that his groups were connected, for, like many groups that arise in the arithmetic theory of automorphic forms, $Z_+ \backslash GL(2, \mathbb{R})$ is unfortunately disconnected, but there is no real difficulty and I shall be as brief as possible. The goal of §2.4 and §4.3, for which we need these results, is just to make clear what terms in addition to the elliptic term contribute to the limit (12′) when m is even and how. We first establish the relation between the notation of this paper and that of [H], as well as the connection between $\omega_1(\gamma, f_\infty)$ and $\theta'_z(0, f_\infty)$, or rather, on referring to (55), between $\omega_1(\gamma, f_\infty)$ and

$$\int f_\infty(k^{-1}zn(x)k)\ln|x|dxdk. \tag{$C.1$}$$

Let

$$\gamma = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}.$$

According to its definition in [JL],

$$\begin{aligned}\omega_1(\gamma, f_\infty) &= -\int\int f_\infty(k^{-1}n^{-1}(x)\gamma n(x)k)\ln(1+x^2)dxdk \\ &= -\int\int f_\infty(k^{-1}\gamma n((1-\beta/\alpha)x)k)\ln(1+x^2)dxdk\end{aligned}$$

which is equal to

$$-\frac{1}{|1-\beta/\alpha|}\int\int f_\infty(k^{-1}\gamma n(x)k)\{\ln((1-\beta/\alpha)^2+x^2)-\ln(1-\beta/\alpha)^2\}dxdk.$$

Thus

$$|1-\beta/\alpha|\omega_1(\gamma, f_\infty) - \ln(1-\beta/\alpha)^2\omega(\gamma, f_\infty) \tag{$C.2$}$$

approaches -2 times (C.1) as α and β approach z. So we shall be able to deduce a convenient expression for (C.1) from Hoffmann's formulas, which are valid for $\alpha\beta > 0$. Since the singularity of $|1-\beta/\alpha|\omega_1(\gamma, f_\infty)$ at $\alpha = \beta$ is only logarithmic, we may multiply it in (C.2) by any smooth function that assumes the value 1 for $\alpha = \beta$.

Because

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

does not lie in the connected component of $Z_+\backslash GL(2,\mathbb{R})$, Hoffmann's arguments do not apply directly to $\omega_1(\gamma, f_\infty)$ for this γ.

When comparing the notation of this paper with that of Hoffmann, it is best to replace, without comment, all of Hoffmann's group elements by their inverses. Otherwise the conventions are not those of number-theorists and not those of this paper. For him maximal compact subgroups operate on the left, and parabolic and discrete groups on the right.

The group P of Hoffmann is for us the group of upper-triangular matrices, $\bar{P}$ the group of lower-triangular matrices, and M is the quotient of the group of diagonal matrices by Z_+ and has as Lie algebra $\mathfrak{a}_{\mathbb{R}}$. His map λ_P, which is determined by the weight in the noninvariant orbital integral defining ω_1, we take to be

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \to a - b,$$

and the λ defining his σ to be $s/2$ times λ_P. In addition, his $d\lambda$ is $ds/2$. Then, as a result of the transfer of the parabolic subgroup to the right in [H], Hoffmann's $v(n(x))$ is $\ln(1+x^2)$ and is, as he observes, positive.[19] Since

$$D_G(m) = (1 - \frac{\beta}{\alpha})(1 - \frac{\alpha}{\beta}), \quad m = \gamma,$$

we conclude that

$$J_M(m, f_\infty) = -\frac{|\alpha - \beta|}{|\alpha\beta|^{\frac{1}{2}}}\omega_1(\gamma, f_\infty).$$

So we may replace $|1 - \beta/\alpha|\omega_1(\gamma, f_\infty)$ in (C.2) by $-J_M(m, f_\infty)$. Here and elsewhere in this appendix I freely use the symbol m as it is used by Hoffmann. Elsewhere in the paper, the symbol m is reserved for the degree of the symmetric power.

Before entering into further comparisons between our notation and that of Hoffmann, I review my understanding of his conventions about the measure on M and on its dual. He takes the two measures to be dual with respect to the Fourier transform. So when they both appear, the normalization is immaterial. On the other hand, only one may appear; moreover, there is a second choice, that of λ_P, which is fixed by the weighting factor v. Hoffmann's I_P is a linear combination of $J_M(m, f)$ and an integral over the dual $\hat{M}$. $J_M(m, f)$ depends directly on λ_P but not on the two Haar measures. There is a further dependence on the measure on $M\backslash G$, but this dependence is the same in every pertinent expression in his paper and can be ignored. The integral over the dual depends directly on the measure on $\hat{M}$ and directly on the measure on M because of the presence of $\pi_{P,\sigma}(f)$ which depends directly on the measure on G, thus on the measures on M and $M\backslash G$; because of the derivative δ_P it depends directly on λ_P as well. Since the measures on M and its dual are inversely proportional, the dependence on the two measures is cancelled and both terms of the sum depend on λ_P alone.

This must therefore be the case for the right side of the formula in his Theorem 1 as well. In the second term, the integral over $\hat{M}$, this is clear, because Θ_{π_σ} depends directly on the measure on M and $\Omega_{P,\Sigma}$ depends depends directly on λ_P. In the first term, however, the only dependence is through $\Omega_\pi(f)$ and is a direct dependence on the measure on M. If the theorem is to be valid, this measure

[19] What with signs and factors of 2, there is considerable room for error when attempting to reconcile conventions from various sources.

must be defined directly in terms of the form λ_P. This Hoffmann does in a straightforward manner. I refer to his paper for more precision. For the group $SL(2,\mathbb{R}) = Z_+\backslash G^+$, with $G^+ = \{g \in GL(2,\mathbb{R}) \mid \det(g) > 0\}$ and with our parameters, s, for the characters of M and t for $a = a(x)$ as in §4.3, the measures are $d\sigma = d|s|/2$ and $da = dx$, which is also the measure $d\lambda/\lambda$ of §2.1.

The collection $\hat{M}$ of unitary representations of M has four connected components, corresponding to the four choices of $k, l = 0, 1$,

$$\sigma : \gamma \to \operatorname{sgn}(\alpha)^k \operatorname{sgn}(\beta)^l |\frac{\alpha}{\beta}|^{s/2},$$

with s purely imaginary. Although J_s and $\operatorname{tr}(J_s^{-1}J_s'\xi_s(f_\infty))$ were defined in §2.3 only for $k = l = 0$, they are defined for all choices of k and l and Hoffmann's $-J_P(\sigma, f_\infty)$ is nothing but $2\operatorname{tr}(J_s^{-1}J_s'\xi_s^{k,l}(f_\infty))$, an expression in which all implicit dependence on k and l is not indicated. Earlier in the paper, $\xi_s^{0,0}$ appeared simply as ξ_s. The factor 2 is a result of the relation $\lambda = s\lambda_P/2$.

Recalling that $D_M(m) = 1$, we consider

$$(C.3) \qquad J_M(m, f_\infty) + \frac{1}{8\pi i}\sum\int_C \operatorname{sgn}(\alpha)^k \operatorname{sgn}(\beta)^l |\frac{\alpha}{\beta}|^{-s/2} \operatorname{tr}(J_s^{-1}J_s'\xi_s^{k,l}(f_\infty))ds.$$

The sum before the integration is over the four possible choices for the pair (k, l). If f is supported on G^+ and if $\det(m) > 0$, then the integrand does not change when k, l are replaced modulo 2 by $k+1, l+1$. So the sum over l can be dropped, l can be taken to be 0 and the 8 becomes 4. So (C.3) would reproduce Hoffmann's definition if we were concerned with G^+ alone.

We will, in general, be summing (C.3) over $\pm m$, so that the total contribution from the integrals for $k \neq l$ will be 0 and for $k = l$ the $8\pi i$ in the denominator will be replaced by $4\pi i$. Moreover replacing $k = l = 0$ by $k = l = 1$ has the effect of replacing $\xi_s(g)$ by $\operatorname{sgn}(\det(g))\xi_s(g)$ and has no effect on J_s. For the contribution from (iv), we shall be concerned with $\alpha = -\beta$ and, for such an m, $\operatorname{sgn}\alpha^k \operatorname{sgn}\beta^k$ is 1 for $k = 0$ and -1 for $k = 1$. The sum of (C.3) over $\pm m$ therefore reduces to

$$(C.4) \qquad J_M(m, f_\infty) + J_M(-m, f_\infty) + \frac{1}{2\pi i}\int_C \operatorname{tr}(J_s^{-1}J_s'\xi_s(f_\infty^-))ds,$$

where f_∞^- is the product of f_∞ with the characteristic function of the component of $Z_+\backslash GL(2,\mathbb{R})$ defined by $\det(g) = -1$. The analogous f_∞^+ will appear below. For the m in question, the factor $|D_G(m)|^{1/2}$ is equal to 2. This is the factor coming from $\omega(\gamma, f_2)$. Thus (C.4) is twice the negative of the sum of the contribution to the limit (12′) of (iv), in which there is yet another minus sign, and of that part of (viii) associated to f_∞^-.

The expression (C.3) has no meaning for the γ that are pertinent in the contribution of (v) to the limit (12′) for even symmetric powers, namely for $\alpha = \beta$. We may however consider it for α unequal but close to β. Once again we consider the sum over $\pm m$. Then only the terms with $k = l$ remain. Since $\operatorname{sgn}\alpha$ will be equal to $\operatorname{sgn}\beta$, we obtain

$$(C.5) \qquad J_M(m, f_\infty) + J_M(-m, f_\infty) + \frac{1}{2\pi i}\int_C |\frac{\alpha}{\beta}|^{-s/2}\operatorname{tr}(J_s^{-1}J_s'\xi_s(f_\infty^+))ds.$$

We add to this

$$\ln(1-\beta/\alpha)^2\{\omega(\gamma, f_\infty) + \omega(-\gamma, f_\infty)\}, \quad \gamma = m.$$

Since the second term in (C.3) is well behaved as $\alpha \to \beta$, the result will have a limit as α and β approach a common value z because the integrals themselves will have a limit. The limit is

$$(C.6) \qquad 2\sum_{j=0}^{1}\int f_\infty(k^{-1}(-1)^j zn(x)k)\ln|x|dxdk + \frac{1}{2\pi i}\int_C \operatorname{tr}(J_s^{-1}J_s'\xi_s(f_\infty^+))ds.$$

This is twice the contribution of (57) and of that part of (viii) associated to f_∞^+ to the limit (12′),

Although the results of Hoffmann cannot be applied directly to the general form of (C.3) or (C.4), they can be applied to (C.5). In fact, the material necessary for extending his arguments is available, although not all in print. The principal ingredients are the differential equation for the weighted orbital integrals and an analysis of their asymptotic behavior. The first is available in general ([A1]) and the second will appear in the course of time in a paper by the same author. Since irreducible representations of $Z_+\backslash GL(2,\mathbb{R})$ are obtained by decomposing – into at most two irreducible constituents – representations induced from its connected component $SL(2,\mathbb{R})$, the Plancherel measure of the larger group is, at least for the discrete series, the same as that of the smaller one. So I feel free to apply Hoffmann's results to (C.3) and (C.4) as well, taking care that the measures used are compatible on restriction to functions supported on G^+ with his.

For any diagonal matrix m with diagonal entries of different absolute value, Hoffmann ([H], Th. 1) finds – at least for f supported on G^+ – that $I_P(m, f_\infty)$ is equal to

$$(C.7) \qquad -\frac{|\alpha-\beta|}{|\alpha\beta|^{\frac{1}{2}}}\sum_\pi \Theta_{\tilde\pi}(m)\Theta_\pi(f) + \frac{1}{8\pi i}\sum_{k,l}\int_{-i\infty}^{i\infty}\Omega(m,s)\operatorname{tr}\xi_s^{k,l}(f_\infty)ds,$$

where

$$\Omega(m,s) = \eta_{k,l}(m,s) + \eta_{l,k}(m,-s)$$

and

$$(C.8) \qquad \eta_{k,l}(m,s) = \operatorname{sgn}\alpha^k \operatorname{sgn}\beta^l e^{ts} \begin{cases} \sum_{n=1}^{\infty} \frac{(\alpha/\beta)^{-n}}{n-s}, & t>0 \\ \sum_{n=0}^{\infty} \frac{(\alpha/\beta)^{n}}{n+s} + \frac{\pi(-1)^{k+l}}{\sin(\pi s)}, & t<0, \end{cases}$$

if

$$\begin{aligned} m = m(t) &= \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \\ &= \begin{pmatrix} \pm e^t & 0 \\ 0 & \pm e^{-t} \end{pmatrix}, \end{aligned}$$

the two signs being chosen independently. The factor $\lambda_P(H_\alpha)/2$ that appears in [H] is 1.

There are two observations to be made. First of all, Ω depends not only on m and s, but also on k and l, which determine the character of M_I. Secondly, $\Omega(m,s)$ is, for a given m, symmetric in s and, despite appearances, does not have a singularity at $s=0$, so that the contour of integration can pass through that point.

For our purposes, it is best to represent (C.7) in terms of the Fourier transform of $\xi_s^{k,l}(f_\infty)$. We begin with the case that $\det m$ is negative, for the passage to the limit $|\alpha|=|\beta|$ is then more direct. We refer to the first term in (C.7) as the elliptic contribution and to the second as the hyperbolic contribution. If $\det m$ is negative, then the character of a discrete-series representation vanishes at m, $\Theta_{\tilde\pi}(m)=0$. So the elliptic contribution is 0.

The character of the representation $\xi_s^{k,l}$ is 0 on the elliptic elements of $GL(2,\mathbb{R})$, but on a hyperbolic element

$$(C.9) \qquad a = a(x) = \epsilon \begin{pmatrix} e^x & 0 \\ 0 & \delta e^{-x} \end{pmatrix}, \qquad \delta, \epsilon = \pm 1,$$

it is equal to

$$(C.10) \qquad \epsilon^{k+l} \frac{\delta^l e^{sx} + \delta^k e^{-sx}}{\sqrt{|1-\alpha/\beta||1-\beta/\alpha|}},$$

where the signs are that appearing in the matrix. Since

$$r = \epsilon(e^x \pm e^{-x}),$$

the numbers e^x and e^{-x} can of course be recovered from r and the sign. The measure $d\lambda/\lambda$ is in this new notation dx. If

$$\hat f_\infty(a) = \sqrt{|1-\alpha/\beta||1-\beta/\alpha|} \int_{M\backslash GL(2,\mathbb{R})} f_\infty(g^{-1}ag)dg,$$

then, by the Weyl integration formula,

$$\operatorname{tr}\xi_s^{k,l}(f_\infty) = \sum \int_{-\infty}^{\infty} \epsilon^{k+l}\delta^l e^{sx}\hat{f}_\infty(a)dx,$$

where a is given by (C.9) and there is a sum over the two free signs in a. Thus $\mathrm{tr}(\xi_s^{k,l}(f_\infty))$ is expressed as the Fourier transform of the functions $\hat{f}_\infty(a)$, although the formula (C.10) and the calculations that led to (30) allow us to express this immediately as an integral of the two functions $\psi_\pm$. It is, however, too soon to pass to the variable r.

What we want to do is to express the hyperbolic contribution to (C.7), for $|\alpha| \neq |\beta|$, in terms not of $\operatorname{tr}\xi_s^{k,l}(f_\infty)$ but in terms of its Fourier transform, then to pass to $\alpha = -\beta$, and at this point and for this particular choice to express the result in terms of $\psi_\pm$. I stop short of this final transformation.

Since we shall be taking the limit $t \to 0$, it suffices to take $t > 0$. Since the signs of α and β are supposed different, the function $\eta(m, s)$ is the Fourier transform of the function that is

$$\operatorname{sgn}\alpha^k \operatorname{sgn}\beta^l \sum_{n=1}^{\infty}(-1)^n e^{-n(t+x)} = -\operatorname{sgn}\alpha^k \operatorname{sgn}\beta^l \frac{e^{-(t+x)}}{1+e^{-(t+x)}}$$

for $x > t$ and 0 for $x < t$. Thus, the hyperbolic contribution is

$$-\frac{1}{4}\sum_{k,l}\sum \int_t^{\infty} \operatorname{sgn}\alpha^k \operatorname{sgn}\beta^l \epsilon^{k+l}\delta^l \frac{e^{-(t+x)}}{1+e^{-(t+x)}}\hat{f}_\infty(a)dx,$$

in which the inner sum is over the free signs in a. The effect of the sum over k and l together with the factor $1/4$ is to remove all terms of the inner sum except the one for which $\epsilon = \alpha$ and $\delta\epsilon = \beta$, as we could have predicted. Thus the signs of a are those of m and the hyperbolic contribution is

$$(C.11) \qquad -\sum \int_t^{\infty} \frac{e^{-(t+x)}}{1+e^{-(t+x)}}\hat{f}_\infty(a)dx.$$

The limit as $t \to 0$ can be taken without further ado and gives

$$(C.12) \qquad -\sum \int_0^{\infty} \frac{e^{-x}}{1+e^{-x}}\hat{f}_\infty(a)dx,$$

where a has eigenvalues of opposite sign. Which is positive and which is negative does not matter because of the summation over the two possible opposing signs. When we take $\eta(m, -s)$ into account as well, we obtain in addition

$$(\text{C.12}') \qquad -\sum \int_{-\infty}^{0} \frac{e^{-x}}{1+e^{-x}}\hat{f}_\infty(a)dx,$$

The two are to be added together. Since we take the sum of $I_P(m, f_\infty)$ and $I_P(-m, f_\infty)$, it is probably best to represent it as the sum of (C.12) (together with (C.12′)),

$$(C.13) \qquad -\sum \int_{-\infty}^{\infty} \frac{e^{-|x|}}{1+e^{-|x|}} \hat{f}_\infty(a)dx, \quad a = a(x)$$

If $\det m$ is positive, then, the sign no longer appearing, (C.11) is replaced by

$$(C.14) \qquad \sum \int_t^{\infty} \frac{e^{-(t+x)}}{1-e^{-(t+x)}} \hat{f}_\infty(a)dx,$$

where, of course, the signs of a are those of m. When we need to be explicit, we denote by $a(x, \epsilon)$ the diagonal matrix with eigenvalues ϵe^x and ϵe^{-x}, ϵ being ± 1. For the passage to the limit,[20] we replace (C.14) by the sum of

$$(C.15) \qquad \sum \int_t^{\infty} \left(\frac{e^{-(t+x)}}{1-e^{-(t+x)}} - \frac{1}{t+x}\right) \hat{f}_\infty(a)dx,$$

whose limit is obtained by setting $t = 0$, and

$$(C.16) \qquad \frac{1}{2}\sum \int_t^{\infty} \frac{1}{t+x} \hat{f}_\infty(a)dx = -\sum \hat{f}_\infty(\pm m)\ln(2t) - \int_t^{\infty} \ln(t+x)\frac{d\hat{f}_\infty}{dx}(a)dx,$$

where we have integrated by parts. Once again, there will be similar terms arising from $\eta(m, -s)$. The first term is an even function of x and will thus contribute

$$-2\sum \hat{f}_\infty(\pm m)\ln(2t).$$

Since $1 - \beta/\alpha \sim 2t$, we are to add to this

$$\ln(4t^2)\omega(\gamma, f_\infty) = 2\ln(2t)\hat{f}_\infty(m), \qquad \gamma = m,$$

because in spite of our notation, taken as it is from a variety of sources, $\hat{f}(m) = \omega(\gamma, f_\infty)$.

So the limit as $t \to 0$ of the sum over m and $-m$ is the sum of

$$(C.17) \qquad \sum \int_{-\infty}^{\infty} \left(\frac{e^{-|x|}}{1-e^{-|x|}} - \frac{1}{|x|}\right) \hat{f}_\infty(a)dx$$

[20] The formulas here are variants of those to be found in [H], especially Lemma 6. They are not necessarily more useful.

and

$$(C.18) \qquad -\sum \int_{-\infty}^{\infty} \ln|x| \operatorname{sgn} x \frac{d\hat{f}_\infty}{dx}(a)dx.$$

In both (C.17) and (C.18) there is a sum over a and $-a$, as in (C.13).

For the elliptic contribution, we recall that from the formula for the discrete-series character with parameter $k \geq 0$, as found, for example, in [K]

$$-\frac{\alpha-\beta|}{|\alpha\beta|^{\frac{1}{2}}}\Theta_{\tilde{\pi}}(m) = -(\pm 1)^{k-1}e^{-kt}, \quad m = a(t, \pm 1),\ t > 0.$$

This has a limit as $t \to 0$. It is $-(\pm 1)^{k-1}$. Since

$$(C.19) \qquad -\sum_{k=0}^{\infty}(\pm 1)^{k-1}\Theta_{\pi_k}(f)$$

is absolutely convergent, we can provisionally take (C.19) as the contribution of the elliptic term of Hoffmann's formula. The contribution (C.19) does not appear to be expressible as an integral of the pair of functions $\psi_\pm$ against a measure. So for the moment I prefer to leave it as it stands.

Appendix D: A Fourier transform.

The Fourier transform of the distribution

$$(D.1) \qquad h \to \int_0^\infty \ln x \frac{dh}{dx}(x)dx$$

is calculated by treating the distribution as minus the derivative with respect to the purely imaginary Fourier transform variable s of

$$\lim_{\epsilon \to 0} \frac{d}{dt} \int_0^\infty x^t e^{-\epsilon x} h(x)dx$$

for $s = 0$. The Fourier transform of the distribution without either the derivative or the limit is calculated directly as

$$\int_0^\infty x^t e^{-\epsilon x} e^{sx} dx = (\epsilon - s)^{-1-t}\Gamma(t+1),$$

where s is purely imaginary. Differentiating, setting $t = 0$, and multiplying by $-s$, we obtain

$$\frac{s}{\epsilon - s}\Gamma(1)\ln(\epsilon + s) - \Gamma'(1)\frac{s}{\epsilon - s}.$$

Careful attention to the real content of this formal argument reveals that $\ln(\epsilon + s)$ is to be chosen between $-\pi/2$ and $\pi/2$. Letting ϵ approach 0, this becomes

$$-\ln s + \Gamma'(1), \tag{D.2}$$

where $\ln s$ is $\ln|s| + \frac{\pi}{2}\operatorname{sgn} s$. The symmetric form of (D.1) is

$$\int_{-\infty}^{\infty} \ln|x| \operatorname{sgn} x \frac{dh}{dx}(x)dx \tag{D.1′)}$$

and the symmetric form of (D.2) is $-2\ln|s| + 2\Gamma'(1)$. Recall from [N, p. 15] that $\Gamma'(1) = -\lambda_0$ is the negative of Euler's constant.

References

[A] J. Arthur, *On the Fourier transforms of weighted orbital integrals*, J. reine angew. Math. **451** (1994), 163–217.

[A1] ________ , *Some tempered distributions on groups of real rank one*, Ann. of Math. **100** (1974), 553–584; *The local behaviour of weighted orbital integrals*, Duke Math. J. **56** (1988), 223–293.

[C] H. Cohen, *A course in computational number theory*, Springer Verlag, 1996.

[Cor] A. Borel and W. Casselman, *Automorphic forms, representations and L-functions*, Proc. Symposia in Pure Math. Amer. Math. Soc., 1979.

[DS] P. Deligne and J.-P. Serre, *Formes modulaires de poids* 1, Ann. Sc. Éc. Norm. Sup.(4) **7** (1974), 507–530.

[HW] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford at the Clarendon Press,

[H] W. Hoffmann, *The Fourier transforms of weighted orbital integrals on semisimple groups of real rank one*, J. reine angew. Math. **489** (1997), 53–97.

[JL] H. Jacquet and R. P. Langlands, *Automorphic forms on $GL(2)$*, SLM 114, Springer Verlag, 1970.

[K] A. W. Knapp, *Representation theory of semisimple groups*, Princeton University Press, 1986.

[Lan] E. Landau, *Primzahlen*, Chelsea Publishing Co.

[L] R. P. Langlands, *Problems in the theory of automorphic forms*, Lectures in Modern Analysis and Applications, vol III, SLM 170, 1970, pp. 18-61.

[L1] ________ , *Stable conjugacy: definitions and lemmas*, Can. J. Math. **31** (1979), 700–725.

[L2] ________ , *Base change for $GL(2)$*, Ann. Math Studies vol. 96, Princeton University Press, 1980.

[LP] M. Larsen and R. Pink, *Determining representations from invariant dimensions*, Invent. Math. **102** (1990), 377–398.

[N] N. Nielsen, *Die Gammafunktion*, Chelsea Publishing Co.

[W] A. Weil, *Adèles and algebraic groups*, Birkhäuser, 1982.

Added in Proof. As the reader was cautioned, much of the material of the paper was provisional: the analysis rough and numerical experiments preliminary. Although further reflection confirms so far the general conclusions, details will have to be modified. There is little point in precise explanations until arguments or experiments have reached a more mature stage. There are only three observations to make. First of all, the $O(\ln^l(p))$ hypothesis for the size of (70) appears more than doubtful. On the other hand, the averages continue to behave, with better code, as described, but better, for the quadratic term in $\ln p$ seems to have coefficient O. Finally formula (81) is not correct as it stands.